

Transparencias de Redes de Ordenadores

Tema 12 IP Multicast

Uploaded by

IngTeleco

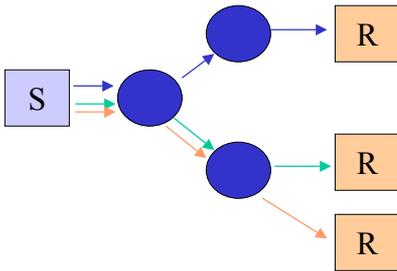
<http://ingteleco.iespana.es>
ingtelecowed@hotmai.com

La dirección URL puede sufrir modificaciones en el futuro. Si no funciona contacta por email

¿Qué es multicast?

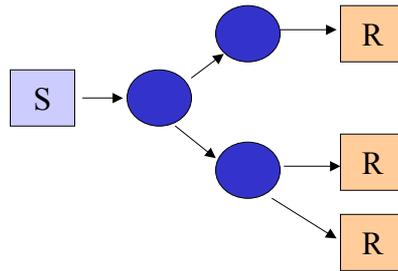
• Servicio Unicast

- Un solo emisor y un solo receptor
- Los datos son enviados desde el emisor al receptor



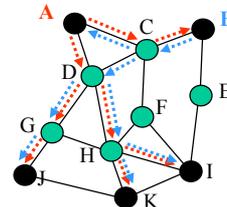
• Servicio multicast

- Un solo emisor y varios receptores
- Los datos son enviados desde el emisor a todos los receptores



¿Por qué multicast?

- Necesario para: *un envío eficiente a múltiples destinos a través de inter/intranets*
- Muchos usuarios desean obtener los mismos datos.
- El objetivo del multicast es enviar los mismos datos a múltiples usuarios de manera eficiente.
- ¿Qué significa eficiente?
 - El número de veces que un mismo paquete circula por un enlace.
 - El número de veces que un emisor debe generar un mismo paquete
 - La complejidad de mantener la lista de a quiénes enviar un paquete
- **A** quiere hablar con **B**, I, J, K y **B** con **A**, I, J, K
- Con unicast, se envían 4 mensajes desde cada emisor
 - Los enlaces AD, BC transportan el mismo paquete tres veces
- Con multicast sólo se enviaría un mensaje desde cada emisor
- El objetivo es minimizar el coste total
 - Un solo paquete desde el emisor alcanza todos los receptores
 - Cada paquete viaja por el número mínimo de enlaces

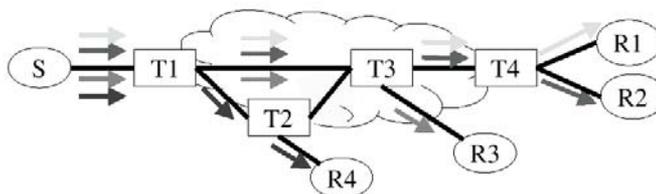


Aplicaciones Multicast

- Actualizaciones de Noticias/deportes/bolsa/tiempo
- Educación a distancia
- Configuración, actualización de routing, localización de servicios
- Aplicaciones "push" Pointcast-type
- Teleconferencia (audio, video, pizarra compartida, editor texto)
- Simulaciones o juegos interactivos distribuidos
- Listas de distribución de Email
- Distribución de Software
- Actualizaciones de Web-cache
- Replicación de Bases de Datos
- Características
 - Número de emisores (simultaneos) hacia el grupo
 - El *tamaño de los grupos*
 - *Numero de miembros* (receptores)
 - Extensión geográfica o *scope*
 - *Diametro* del grupo medido en saltos
 - La *longevidad* del grupo
 - Número de paquetes/segundo
 - El pico/media de uso del emisor
 - Nivel de *interactividad* humana
 - Modo lectura vs interactivo
 - Sólo-Datos (pe replicación de BD) vs multimedia

Multicast mediante unicast del nivel de aplicación

- Un emisor, N receptores
 - El emisor envía N paquetes duplicados unicast, un paquete a cada receptor
- ¿Cuáles son los problemas?
 - Los paquetes duplicados aumentan la "utilización de la red"
 - El enlace de salida del emisor S está inundado con N paquetes duplicados (tantos como destinatarios)

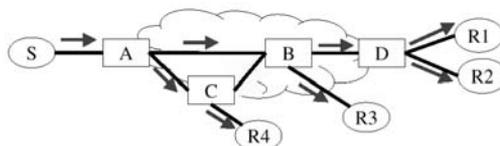


unicast del nivel de aplicación: problemas

- Replicación
 - Algunos enlaces transportan copias duplicadas de los paquetes
 - Inevitable, porque los paquetes no son replicados dentro de la red.
 - Resulta útil para grupos pequeños con bajo ancho de banda
 - Juegos de red
 - Variación: Uso de un Servidor
 - Tener un servidor centralizado que distribuye copias unicast de los paquetes a los miembros del grupo
 - NO escalable a grupos grandes o a muchos grupos
- Gestión de la pertenencia a grupos
 - El emisor debe mantener una copia de la pertenencia al grupo
 - Saber a quien enviar paquetes unicast
 - Grave problema:
 - Los miembros pueden adherirse y abandonar los grupos dinámicamente
 - Los miembros a veces fallan
 - El tamaño de las tablas de miembros puede ser importante
- Denominación de grupo
 - Mediante una lista de las direcciones IP de sus miembros
 - ¿Cómo conseguir dicha lista? ¿Cómo mantener la lista actualizada?

¿Por qué Multicast a nivel de red?

- Soluciona el problema de la duplicación de paquetes
 - Dejar a los routers que repliquen los paquetes cuando sea necesario para hacer llegar una copia a cada miembro del grupo
 - El mismo paquete se envía una vez por cada enlace como máximo (basta con asegurar que todos los miembros reciben UNA copia del paquete)
- Soluciona el problema de la gestión de pertenencia a un grupo
 - Dejar que la red conserve la información sobre la pertenencia a los grupos
- Soluciona el problema de la denominación de los grupos
 - Cada grupo multicast recibe una dirección multicast
 - Desacopla la identidad del grupo de sus miembros y por lo tanto es independiente de estos
 - Escalado



Grupos Multicast

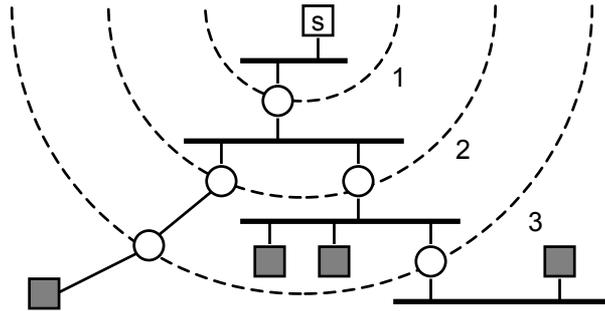
- Entidad lógica que identifica un conjunto de receptores que desean recibir los mismos datos.
 - Independiente del número o identidad de los receptores y los potenciales emisores
 - Los miembros son los receptores
 - Los emisores pueden ser o no miembros del grupo
 - Instanciado (creado en la red) bien cuando un emisor inicia la emisión o cuando un receptor expresa su interés en recibir datos (se suscribe).
 - Los nodos pueden pertenecer a varios grupos
 - Los nodos pueden enviar mensajes a muchos grupos
 - Soporte a la creación dinámica de grupos, pertenencia dinámica, y emisores dinámicos
 - Ninguna entidad centralizada mantiene una lista completa de los receptores suscritos
 - La pertenencia a un grupo se mantiene indirectamente a través del estado de encaminamiento de los routers.
- La dirección de destino es una dirección IP de clase D
 - Es una porción globalmente conocida del espacio de direcciones

Multicast sobre LANs & Scoping

- Los paquetes multicasts se inundan a través de los puentes de nivel MAC a través de un árbol de expansión
 - La inundación puede restar oportunidad de transmitir a estaciones no miembros que desean hacerlo
 - ¡Casi como broadcast!
- Scope: ¿Hasta dónde deben propagarse las transmisiones?
 - Implicit scoping:
 - Direcciones Mcast reservadas \Rightarrow no abandonar la subred.
 - También llamadas direcciones "link-local"
 - TTL-based scoping:
 - Los routers multicast tienen un umbral TTL
 - Los datagramas Mcast se borran si $TTL \leq$ umbral TTL
 - Administrative scoping:
 - Usa una porción del espacio de direcciones de clase D (239.0.0.0 a 239.255.255.255)
 - Auténticamente local al dominio administrado; posible la reutilización de direcciones.
 - En IPv6, scoping es un atributo interno de una dirección IPv6 multicast

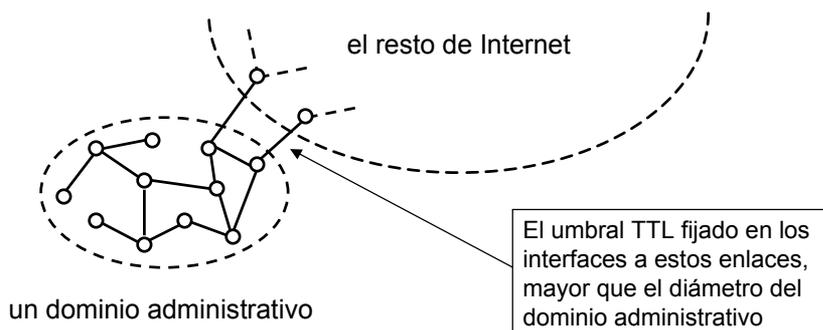
Control del Ambito Multicast: TTLs pequeñas

- Alcanzar o buscar un subconjunto próximo de miembros de un grupo



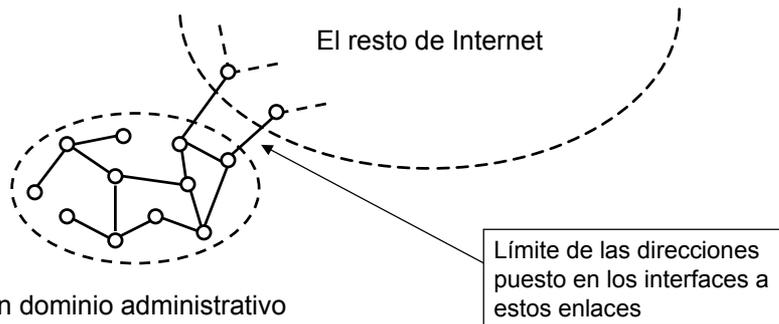
Control de Ambito Multicast: TTLs grandes

Mantener el tráfico multicast dentro de un dominio administrativo, p.e. a efectos de privacidad o consumo de recursos

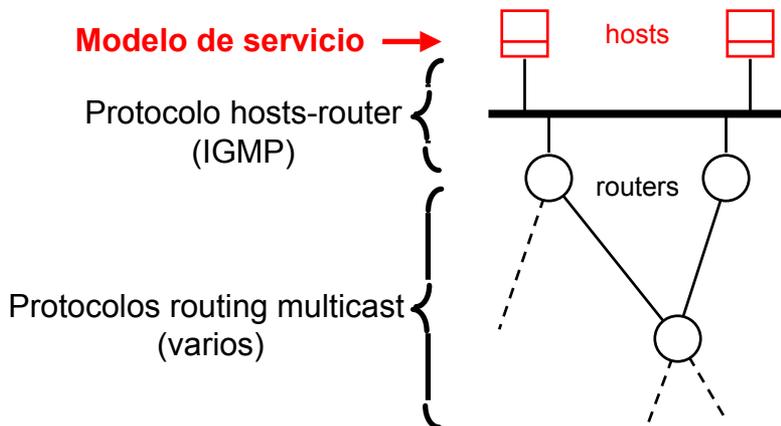


Control del ámbito multicast : Direcciones con ámbito administrado

- RFC 1112
- Utiliza direcciones del rango 239.0.0.0 — 239.255.255.255
- Soporta dominios con solapamiento (no solo anidamiento)



Componentes de la Arquitectura IP Multicast



Modelo de Servicio IP Multicast (RFC-1112)

- Cada grupo se identifica mediante una sola dirección IP de clase D
- Los grupos pueden ser de cualquier tamaño
- Los miembros de los grupos pueden estar localizados en cualquier lugar
- La pertenencia a un grupo no es conocida explícitamente
- Los miembros de un grupo pueden unirse y abandonarlo en cualquier momento
- Los emisores no necesitan ser miembros
- Los routers escuchan las direcciones multicast y utilizan protocolos de routing para gestionar los grupos (IGMP, RFC2236)

Analogía : Cada dirección multicast es como una frecuencia de radio, en la cual cualquiera puede transmitir, y cualquiera puede sintonizar

- Los paquetes no se duplican o envían a destinos fuera del grupo
- Se construye un árbol de distribución para el envío de paquetes
- No aparece más de una copia del paquete en una subred
- Los paquetes se entregan sólo a los receptores "interesados" ⇒ los árboles de distribución multicast cambian dinámicamente
- La red debe descubrir activamente caminos entre emisores y receptores

Servicio IP Multicast: Emisión y Recepción

EMISION

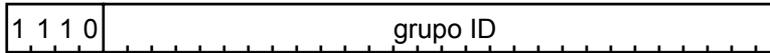
- Utiliza la operación normal de emisión de datagramas IP, con una dirección IP multicast como destino.
 - Se transmite como multicast de enlace de datos por aquellos interfaces que soportan multicast.
- Debe proporcionar a la aplicación emisora un modo de:
 - Especificar el interfaz de red de salida, cuando hay más de uno
 - Especificar el tiempo de vida (TTL) del paquete emitido
 - Habilitar/deshabilitar el "loopback" si el nodo emisor es miembro del grupo de destino en el interfaz de salida

RECEPCION

- Dos operaciones nuevas:
 - Unir_a_grupo_multicast (Dirección de grupo, interfaz)
 - Abandonar_grupo_multicast (Dirección de grupo, interfaz)
- Recibir paquetes multicast para los grupos a los que pertenece mediante la recepción normal de datagramas IP
- Los routers multicast deben recibir todos los IP multicasts por un enlace, sin saber por adelantado a qué grupos serán enviados

Direcciones IP Multicast

Espacio de Direcciones IP de Clase D :



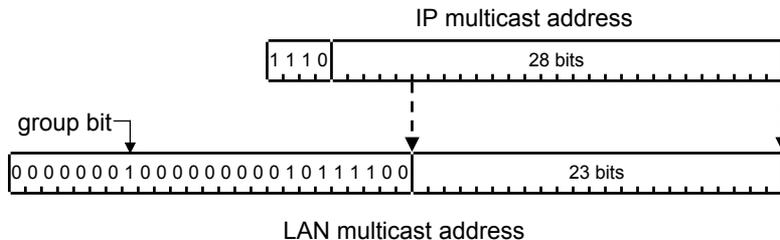
En notación decimal : 224.0.0.0 — 239.255.255.255

Dos categorías administrativas:

- Direcciones multicast "well-known", asignadas por la IANA (RFC1700)
 - 224.0.0.0 – 224.0.0.25
 - 224.0.0.1 (all-systems.mcast.net)
 - 224.0.0.1 (all-routers.mcast.net)
- Direcciones multicast "transitorias", asignadas y abandonadas dinámicamente.

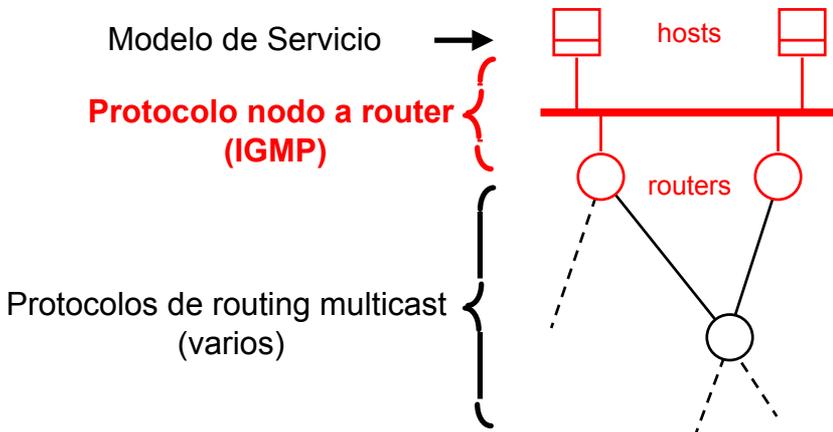
Mapeo de Direcciones IP Multicast a Direcciones MAC

- Para Ethernet y otras LANs que emplean direcciones 802:



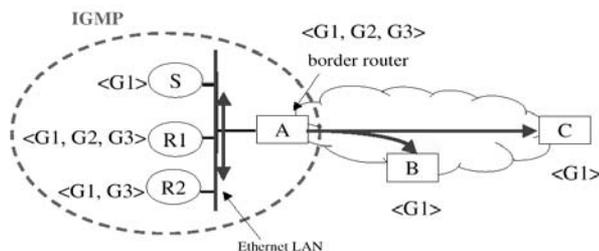
- Varias direcciones de clase D se corresponden con la misma dirección MAC
 - Es poco probable la existencia de dos grupos con los mismos 23 bits finales sobre la misma LAN.
- La traducción automática evita la necesidad de tablas o protocolos
- Para enlace punto a punto no es necesario ningún mapeo

Componentes de la Arquitectura IP Multicast

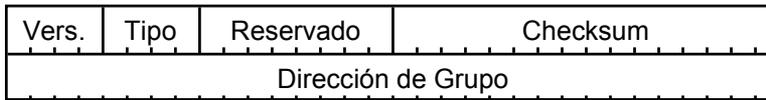


Internet Group Management Protocol (IGMP)

- Protocolo mediante el que los nodos notifican su pertenencia a un grupo a los routers vecinos
 - Permite a los routers aprender qué grupos están presentes en las redes a las que están conectados.
 - Version 1, el actual estándar Internet, especificado en RFC-1112
 - Versión 2, futuro estándar, especificado en RFC-2236
- Opera sobre LANs de difusión y enlaces punto a punto
 - entre sistemas finales y routers
- Ocupa posición y papel similar a ICMP en la pila TCP/IP

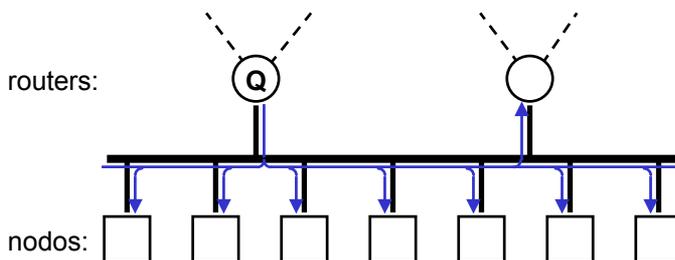


Formato de mensajes IGMP V1



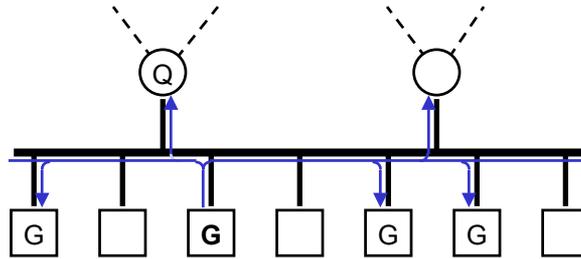
Version 1
Tipo 1 = Pregunta de Pertenencia
2 = Notificación de pertenencia
Checksum comprobación del mensaje IGMP (estilo IP)
Dirección de Grupo Grupo que se reporta (cero en las preguntas)

¿Cómo opera IGMP?



- En cada LAN, se elige un router (router designado) para realizar las preguntas
- Este router, envía periódicamente un mensaje de Pregunta de Pertenencia al grupo de "todos los nodos" (224.0.0.1), con TTL = 1
- El intervalo entre preguntas es normalmente 60—90 s
- Al recibirlo, los nodos arrancan temporizadores aleatorios (entre 0 y 10 segundos) para cada grupo multicast al que pertenecen

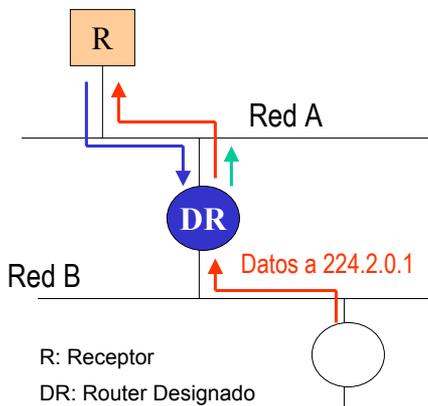
¿Cómo opera IGMP? (ii)



- Cuando en un nodo expira el temporizador para un grupo G, difunde una Notificación de Pertenencia al grupo G (subscription request), con TTL = 1
 - Otros miembros de G escuchan la Notificación y detienen sus temporizadores
 - Se evitan generar duplicados (los routers no necesitan conocer a todos los miembros, sólo si existen o no miembros)
- Los routers escuchan todas las notificaciones, y eliminan los grupos que no responden
- Cuando un nodo se adhiere a un grupo en primer lugar, envía una o dos notificaciones inmediatas, en lugar de esperar una pregunta

IGMP - Adhesión a un grupo

IGMP Notificación de Pertenencia



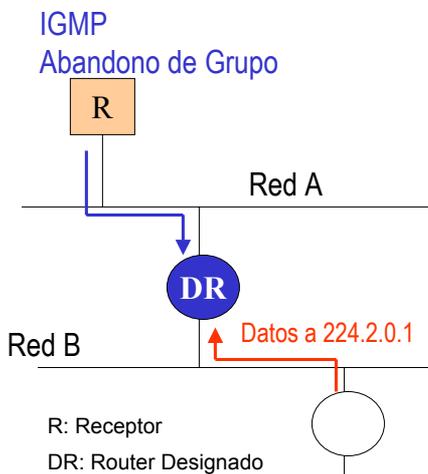
Ejemplo : R se adhiere al Grupo 224.2.0.1

- R envía IGMP Informes de Pertenencia a 224.2.0.1
- DR los recibe. DR enviará paquetes para 224.2.0.1 a la red A
- DR envía periódicamente IGMP Pregunta de Pertenencia a 224.0.0.1 (ALL-SYSTEMS.MCAST.NET)
- R responde IGMP Informes de Pertenencia a 224.2.0.1

IGMP Versión 2

- Cambios respecto de la versión 1:
 - Un mensaje nuevo (abandono de grupo) y procedimientos para reducir la latencia de abandono ("leave latency")
 - Se especifica un método de elección estandar para el router que genera las preguntas
 - Los cambios versión y tipo se unen en un solo campo
- Compatibilidad con la versión 1
- Ya está implementado con profusión

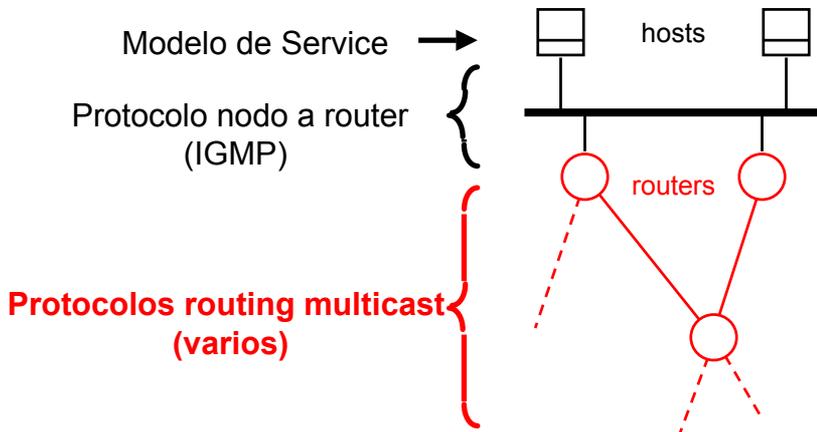
IGMP - Abandono de un grupo



Ejemplo : R abandona el Grupo
224.2.0.1

- R envía IGMP Abandono de Grupo a 224.0.0.2 (ALL-ROUTERS.MCAST.NET)
- DR lo recibe.
- DR deja de enviar paquetes para 224.2.0.1 a la Red A si no hay más miembros del grupo 224.2.0.1 en la Red A.

Componentes de la Arquitectura IP Multicast



Routing Multicast IP

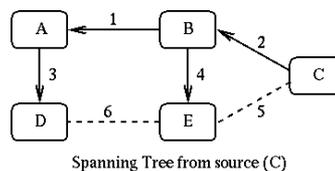
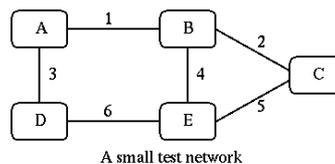
- Hacer llegar los paquetes a los miembros de un grupo.
- Propósito de los protocolos de routing multicast:
 - Compartir entre routers información sobre los grupos dar una solución de encaminamiento mejor para la distribución de datos.
- Algoritmos para construir árboles multicast
 - Inundación
 - Arboles de Expansión
 - Reverse Path Forwarding (RPF)
 - Truncated Reverse Path Forwarding (TRPF)
 - Steiner Trees (ST)
 - Core-Based Trees (CBT).

Inundación

- Operación:
 - Un router recibe un datagrama multicast.
 - Comprueba si lo ha recibido con anterioridad.
 - Si es la primera vez que lo recibe lo reenviará por todos sus interfaces a excepción de aquel por el que lo recibió
 - En caso contrario, el router descartará el datagrama.
- Pros
 - Simple
 - Robusto
- Contras
 - Todos los routers reciben todos los paquetes incluso aunque no estén suscritos al grupo
 - Los routers reciben multiples copias de los paquetes y los enlaces transportan cada paquete al menos dos veces
 - Determinar si un paquete se ha visto con anterioridad es complicado
 - Necesidades de almacenamiento y procesamiento sustanciales
 - El almacenamiento puede ser costoso para sesiones multicast largas
 - El rendimiento del reenvío puede disminuir

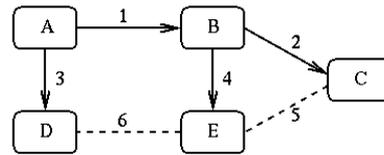
Spanning Tree

- Selecciona un subconjunto de enlaces
 - Definen una estructura de árbol tal que sólo haya un camino activo entre dos routers cualesquiera (grafo libre de bucles).
- Cuando un router recibe un datagrama multicast lo reenvía por todos los enlaces que pertenecen al árbol de expansión, excepto por el que recibió el datagrama.
 - El datagrama alcanza a todos los routers de la red.
- Deficiencias:
 - Centralizan todo el tráfico en un reducido número de enlaces
 - No considera la pertenencia a grupos.

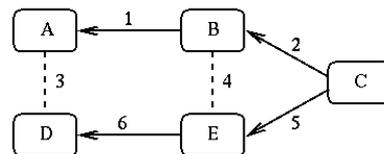


Reverse Path Broadcasting (RPB)

- Construye un árbol de expansión para cada par (emisor, grupo)
 - Arboles de distribución con la raíz en el emisor
- "Enlace padre":
 - Por el que el router espera recibir los datagramas multicast de un emisor.
- Para cada par (emisor, grupo), si llega un paquete a través del enlace padre, el router lo reencamina por todos los interfaces hijos.
- "Enlaces hijos":
 - Los de salida a través de los cuales el router encamina el datagrama multicast.
- RPB+
 - Si un router no está en el camino más corto entre el emisor y un vecino, el datagrama será descartado en el router vecino \Rightarrow no sería necesario reenviar el datagrama
- Deficiencia:
 - No tiene en cuenta la pertenencia a grupos multicast para construir árboles de expansión.



RPB tree from source (A)



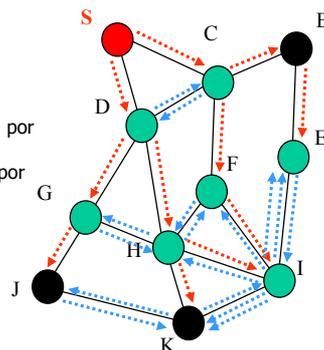
RPB tree from source (C)

RPB (ii)

- Se reenvían paquetes procedente del emisor S por todos los interfaces (salvo el de llegada) sólo si llega por el interfaz que corresponde al camino más corto a S.
 - Paquete recibido a través del interfaz I, desde S (emisor) para G (grupo multicast): paquete (S,G)
 - El router mira en su tabla de routing para encontrar el interfaz usado para enviar paquetes a S: I(padre).
 - Si $I \neq I(\text{padre})$, I es un interfaz erróneo para recibir (S,G).
 - Si $I = I(\text{padre})$, I es un interfaz correcto para recibir (S,G).
- Consecuencias de estar en el camino más corto al emisor:
 - Si se está, es (probablemente) la primera vez que se ve el paquete, de modo que debe reenviarse.
 - Si no se está, (probablemente) ya se ha recibido y reenviado, por tanto debe descartarse.
 - En cualquier caso está garantizado que recibirás el paquetes por el camino(s) más corto(s).
- Beneficios
 - No es necesario recordar los paquetes pasados
 - Ahorro en el número de duplicados, pero no se eliminan (múltiples caminos más cortos)
- Consideraciones
 - Se suponen caminos *simétricos*
 - El reenvío es más complejo que en unicast

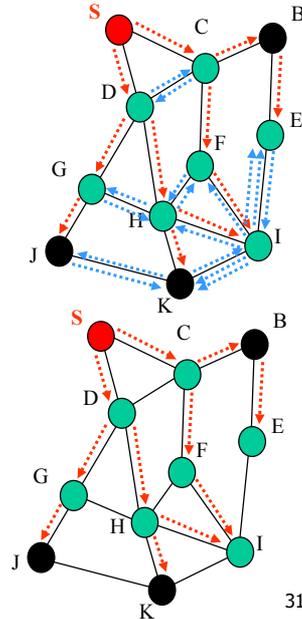
Ejemplo

- C no necesita enviar los paquetes recibidos desde D
- Los paquetes rojos son enviados pero no los azules



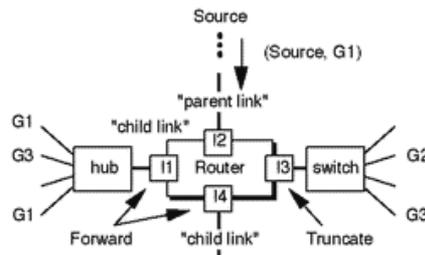
Una mejora a RPB (RPB+)

- Un router no envía un paquete hacia abajo ("downstream") si no se encuentra en el camino más corto al emisor del router destinatario
- Beneficios
 - Elimina la mayor parte de los duplicados
- Desventajas
 - Los duplicados siguen existiendo cuando el router siguiente tiene múltiples caminos más cortos al emisor
 - Complejidad adicional sustancial
 - El router necesita calcular/conocer los caminos más cortos desde TODOS sus vecinos hacia abajo
 - Se precisan comprobaciones adicionales cuando se envían paquetes



Truncated Reverse Path Broadcasting (TRPB)

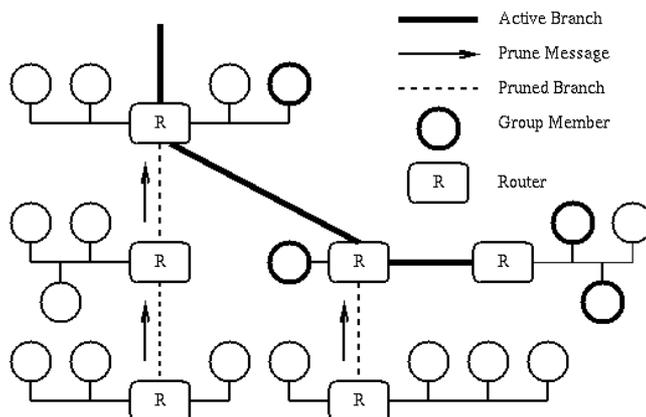
- TRPB ha sido propuesto para solucionar algunas de las limitaciones de RPB.
- Utilizando IGMP un router puede determinar si hay miembros de un grupo multicast presentes en una subred o no.
- Una "subred hoja" (sin ningún otro router conectado a ella) sin miembros de un grupo será truncada por el router del árbol de expansión.



Reverse Path Multicasting (RPM)

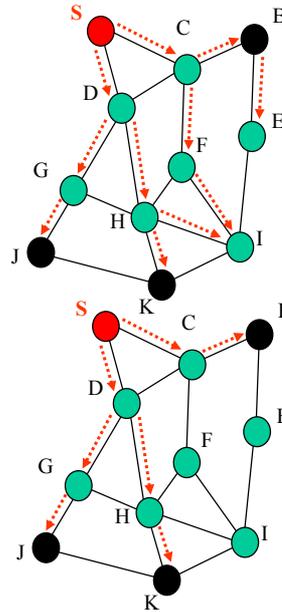
- RPB con podas
- Construye un árbol de envío que expande sólo:
 - 1) subredes con miembros de un grupo
 - 2) routers y subredes en el camino más corto a las subredes con miembros de un grupo.
- El árbol RPM puede ser podado de modo que los datagramas multicast se envíen por enlaces que conducen a los miembros del grupo destino.
- Procedimiento de poda:
 - Para un par dado (emisor, grupo) el primer datagrama multicast se envía basándose en el algoritmo TRPB.
 - Si un "router hoja" recibe un datagrama multicast para un (emisor, grupo) y no tiene ningún miembro del grupo en sus subredes, envía un mensaje de "poda" al router del que recibió el paquete multicast.
 - Los mensajes de poda sólo se envían un salto hacia atrás en dirección al emisor.
 - Si el router por encima no tiene ningún miembro local y recibe mensajes de poda de todos sus "hijos" en el árbol TRPB, enviará a su vez un mensaje de poda a su "padre" en el árbol TRPB.
 - Los mensajes de poda en cascada truncarán el árbol original TRPB de modo que los datagramas multicast serán enviados sólo por aquellos enlaces que conduzcan a un nodo de destino (miembro de un grupo).

Reverse Path Multicasting (RPM)



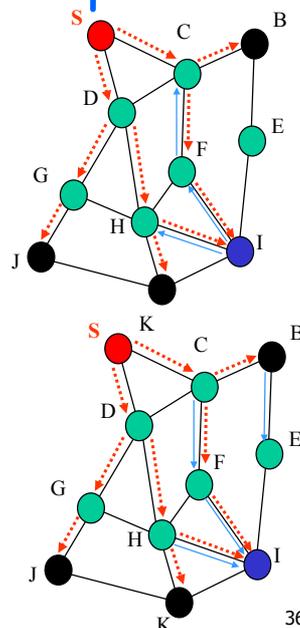
RPM: TRPB con Poda

- Idea básica
 - RPB+ no elimina transmisiones innecesarias
 - E, F, e I reciben paquetes aunque no los necesitan
 - Es necesario algún mecanismo para que los routers indiquen que quieren (o no quieren) recibir paquetes para un grupo multicast
- ⇒ La poda es el proceso de eliminar explícitamente un enlace (router) de la lista de interfaces a los que reenviar un paquete para una dirección multicast
 - Puede ser específico a un grupo multicast o a un emisor y un grupo
- Ejemplo
 - Los routers C, B y H pueden podar los enlaces a los routers F, E, e I respectivamente.



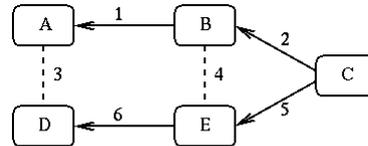
Recuperación de una poda

- Un nodo de la LAN conectada a I se une a un grupo multicast y quiere recibir mensajes de S
- IGMP permite a I conocer el interés del nodo en el grupo multicast de S
 - I puede enviar un mensaje *join(group, A)* a F o H, que lo propagan hasta alcanzar un router que se encuentre ya en el árbol multicast
 - La unión al árbol es rápida, pero produce algún "overhead"
 - Pueden ser necesarias podas subsiguientes
 - Los routers pueden inundar periódicamente mensajes "pruning probe", e I se abstendrá de responder al mensaje de poda.
 - Mecanismo simple
 - Latencia en la "reunión"

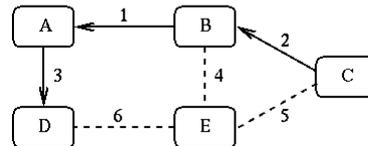


Arboles de Steiner (ST)

- En los algoritmos RPB (RPB, TRPB, y RPM) se utiliza el camino más corto entre el nodo emisor y cada nodo de destino, garantizando que el envío se hace lo más rápidamente posible.
- Ninguno de estos algoritmos trata de minimizar el uso de los recursos de la red.
- El segundo árbol usa menos enlaces, aunque es más lento que el RPB
- Es denominado árbol de Steiner.
- Dificultades para calcularlo han hecho \Rightarrow poco usado.
- Muy inestables.



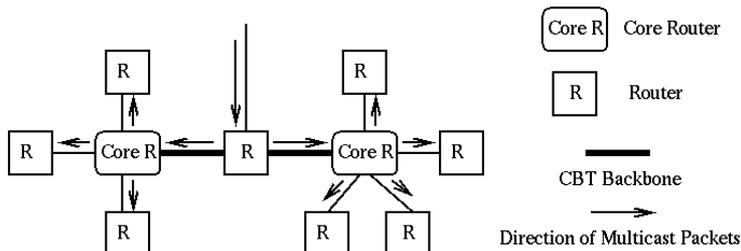
RPB tree from source (C)



ST tree from source (C)

Core-Based Trees (CBT)

- CBT crea un árbol de distribución para cada grupo.
- El árbol utilizado para enviar los datagramas multicast de un grupo particular es un árbol independiente de la localización del nodo emisor.
- Todos los datagramas para un grupo son enviados como mensajes unicast al router "core" hasta que alcanzan un router que pertenece al correspondiente árbol de distribución
 - Entonces el datagrama se envía por todos los interfaces que son parte del árbol de distribución excepto por el que llegó.



Clasificación de los protocolos de Routing Multicast IP

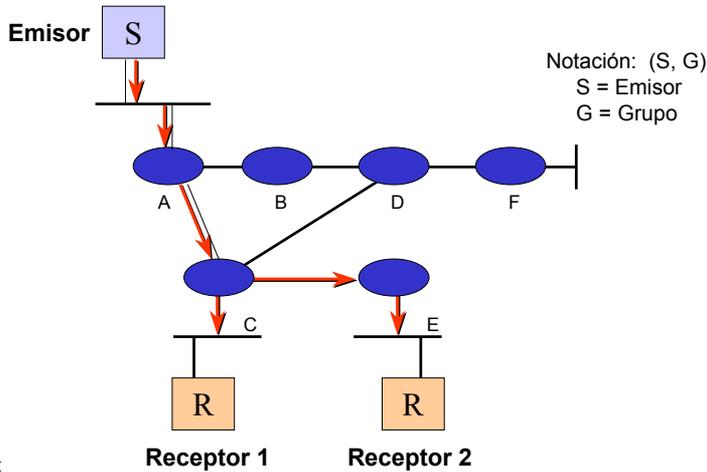
- Estructura del Arbol de Distribución
 - Arboles separados con raíz en cada emisor de datos ("source distribution tree")
 - (DVMRP, MOSPF, PIM-DM, PIM-SM)
 - Arboles con raíz compartida en un punto común ("shared distribution tree")
 - (CBT, PIM-SM)
- Política de Distribución de Datos
 - Tipo NACK
 - Comienza con difusión, después se podan las ramas en las que no hay receptores para crear un árbol de distribución.
 - Inconveniente:
 - Hay mucho tráfico inútil cuando hay pocos receptores en un área amplia
 - Tipo ACK
 - Sólo se envía a los nodos que explícitamente se han adherido al grupo
 - Inconveniente:
 - Latencia por la propagación de las uniones y elevado coste cuando hay muchos receptores

Clasificación de los protocolos de Routing Multicast IP (ii)

- Protocolos de modo denso:
 - Suponen pertenencia densa a grupos
 - "Source distribution tree" y tipo NACK
 - DVMRP
 - Distance-Vector Multicast Routing Protocol draft-ietf-admr-dvmrp-v3-01.txt (not RFC-1075!)
 - MOSPF
 - Extensiones Multicast a "Open Shortest-Path First Protocol" RFC-1584 & RFC-1585
 - PIM-DM
 - Protocol-Independent Multicast, Dense Mode
- Protocolos de modo disperso:
 - Suponen pertenencia dispersa a grupos
 - "Shared distribution tree" y tipo ACK
 - PIM-SM
 - Protocol-Independent Multicast, Sparse Mode
 - CBT
 - Core-Based Trees

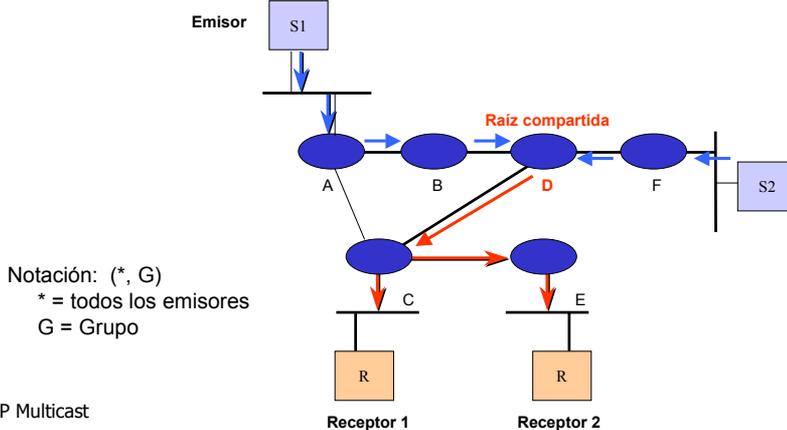
" Source distribution tree "

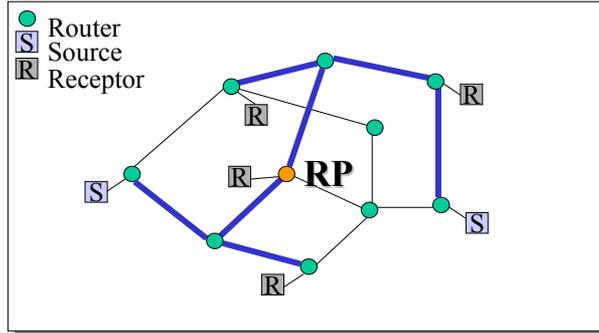
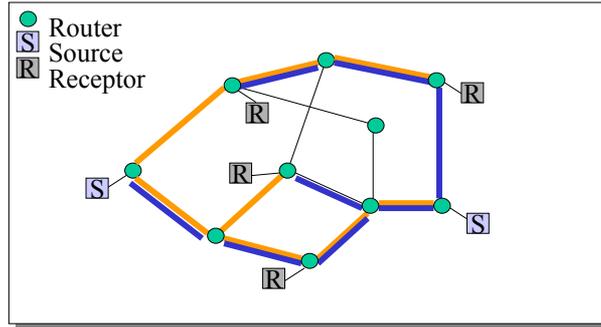
- Source tree
 - Más memoria $O(G \times S)$ en los routers
 - Camino óptimo desde el emisor, minimiza el retraso
 - Bueno para:
 - Pocos emisores y muchos receptores, tales como aplicaciones de difusión de Radio



" Shared distribution tree "

- Shared tree
 - Menos memoria $O(G)$ en los routers
 - Caminos sub-óptimos desde el emisor al receptor y retrasos adicionales (emisor a raíz)
 - Puede haber transferencias duplicadas (posible duplicación de un camino desde el emisor a la raíz y de la raíz a los receptores)
 - Bueno para
 - Muchos emisores con poco ancho de banda
 - Entornos en los que la mayor parte del árbol compartido es idéntico al source tree



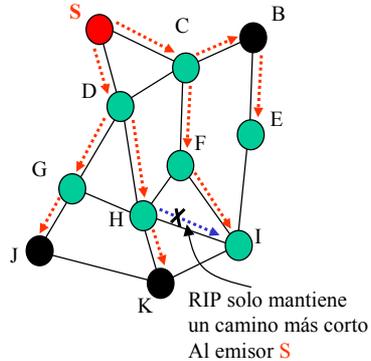


DVMRP

- Protocolo routing Multicast Vector-Distancia
 - Adecuado sólo para routing multicast intra-dominio
 - Sólo es *experimental* (no es un standard)
- RIP como punto de partida
 - Vector Distancia basado en el cálculo de tablas de encaminamiento
 - La distancia es el número de saltos
 - Substancialmente más complejo que el routing unicast
 - Las decisiones de encaminamiento de paquetes se basan en el camino de vuelta (más corto) a los *emisores* de paquetes multicast (a veces llamados *destinos* por consistencia con RIP)
- Extensión de podado
 - Puede notificar explícitamente la no pertenencia a un grupo determinado
 - Propaga hacia arriba desde las LANs no pertenecientes
 - Utiliza mensajes DVMRP *Non Membership Report* (NMR)
 - Cancela el reenvío durante un tiempo fijo
 - Puede cancelar explícitamente el NMR
 - Utiliza el mensaje NMR Cancel message

DVMRP (ii)

- RPB+ se fundamenta en determinar si:
 - Un paquete llega por el interfaz correspondiente al camino más corto al emisor
 - Un router está en el camino más corto desde el siguiente router al emisor
- Enfoque
 - Basarse en cálculos del camino más corto estandar (RIP)
- Pero
 - Los cálculos VD no proporcionan a un router acceso a la tabla de encaminamiento de otros routers
 - Necesario para determinar si un router se encuentra en el camino más corto desde el siguiente router hacia el emisor
- Solución
 - Split horizon con "poison reverse"
 - Yo estoy en tu camino más corto si tú notificas una ruta de costo infinito, y yo tengo una finita en mi tabla de encaminamiento.



Ejemplo

- H recibe un paquete desde D que está en su camino más corto a S
Un paquete desde D sería encaminado
- K notifica a H una ruta de costo ∞ a S, pero F y G no
Un paquete sería encaminado a K pero no a F y G

DVMRP

Intercambio Vector Distancia

- Mantiene una tabla de routing multicast mediante el intercambio de información vector distancia entre routers
 - Visión consistente de un árbol entre routers
 - Conveniente para tener caminos separados para datagramas unicast y multicast
- Información dependiente del router"
 - Información acerca de routers responsables de la transmisión desde un emisor
 - Un router aguas abajo envía Poison Revers a un router elegido aguas arriba para indicar su dependencia
 - Poison Reverse : métrica + 32 (∞)

Source	Subnet	Subnet Mask	From Gateway	Metric	Status	TTL	InPort	OutPorts
128.1.0.0	255.255.0.0	128.7.5.2	3	Up	200	1	2,3	
128.2.0.0	255.255.0.0	128.7.5.2	5	Up	150	2	1	
128.3.0.0	255.255.0.0	128.6.3.1	2	Up	150	2	1,3	
128.4.0.0	255.255.0.0	128.6.3.1	4	Up	200	1	2	

Source	Subnet	Multicast Group	TTL	InPort	OutPorts
128.1.0.0	224.1.1.1	200	1	Pr	2p 3p
	224.2.2.2	100	1		2p 3
	224.3.3.3	250	1		2

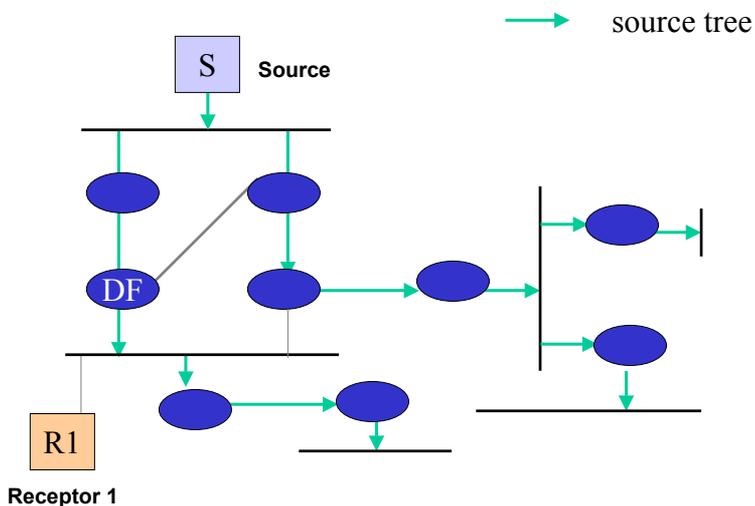
DVMRP

Difusión & poda

- El envío de paquetes multicast se basa en la regla RPB
- Los routers hojas (leaf routers) comprueban y envían mensajes de poda al router aguas arriba cuando no hay miembros de un grupo en la red
- El router aguas arriba poda el interfaz sin ningún router dependiente aguas abajo
- Mensajes Graft para crear una nueva rama para nuevos participantes
- Se reorganiza después del tiempo de poda (estándar 720 minutos)
- Emisor designado
 - Múltiples routers en la misma LAN
 - El de métrica más baja o la dirección IP menor se convierte en emisor designado (se descubre cuando se intercambia información de métrica)
- draft-ietf-idmr-dvmrp-v3-09.txt (Septiembre 1999)

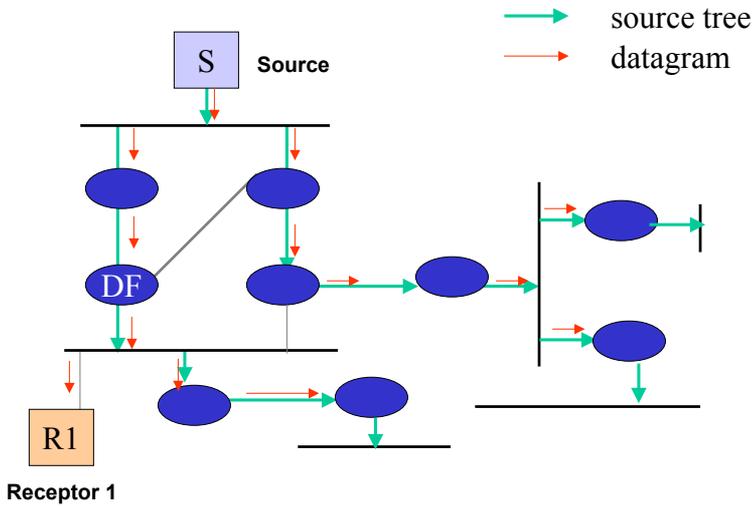
DVMRP (1)

Source tree: intercambio de métrica



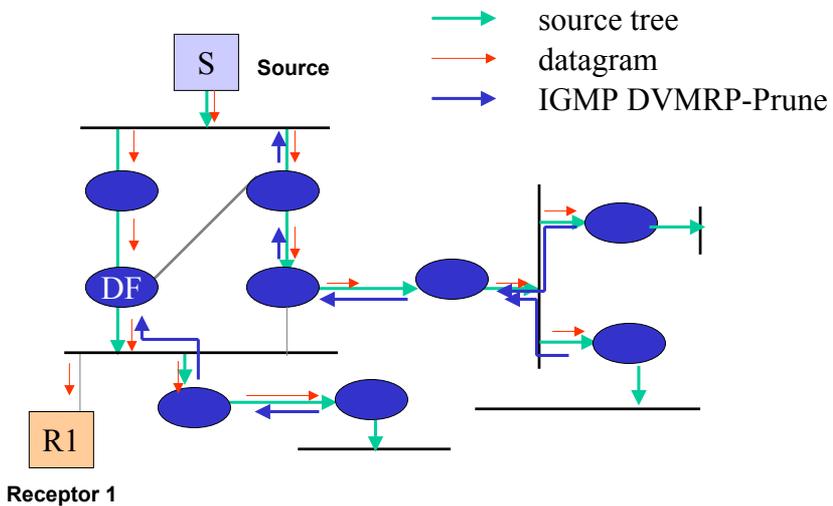
DVMRP (2)

difusión



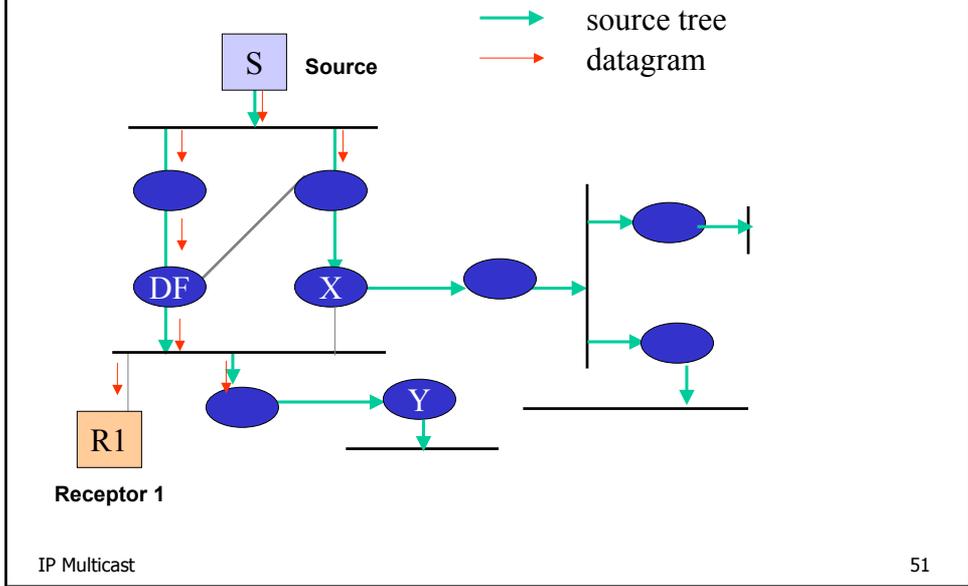
DVMRP (3)

Poda



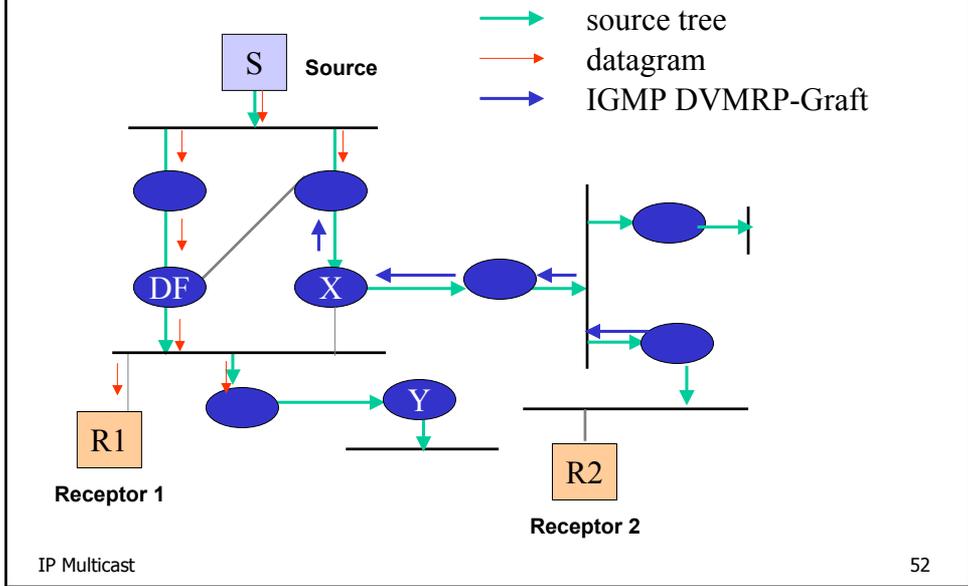
DVMRP (4)

X e Y podados



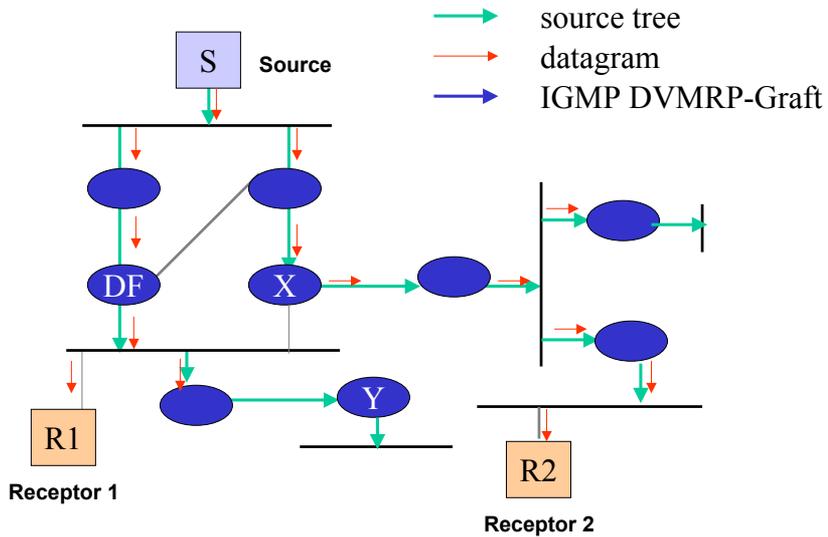
DVMRP (5)

Nuevo miembro (grafting)

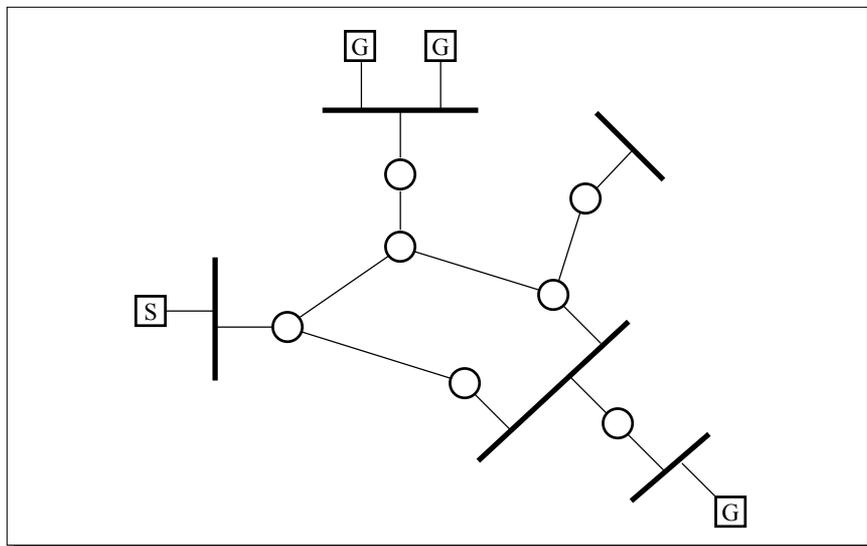


DVMRP (6)

Nueva rama

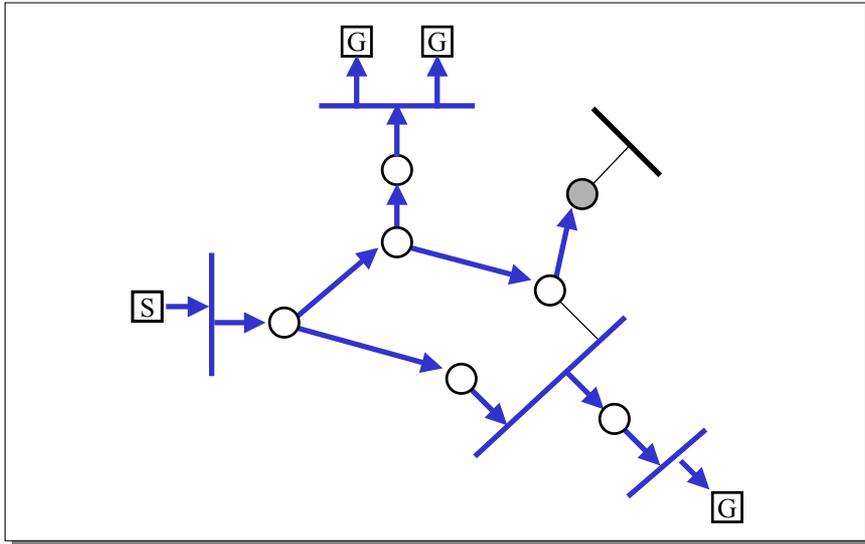


DMRV: Otro ejemplo



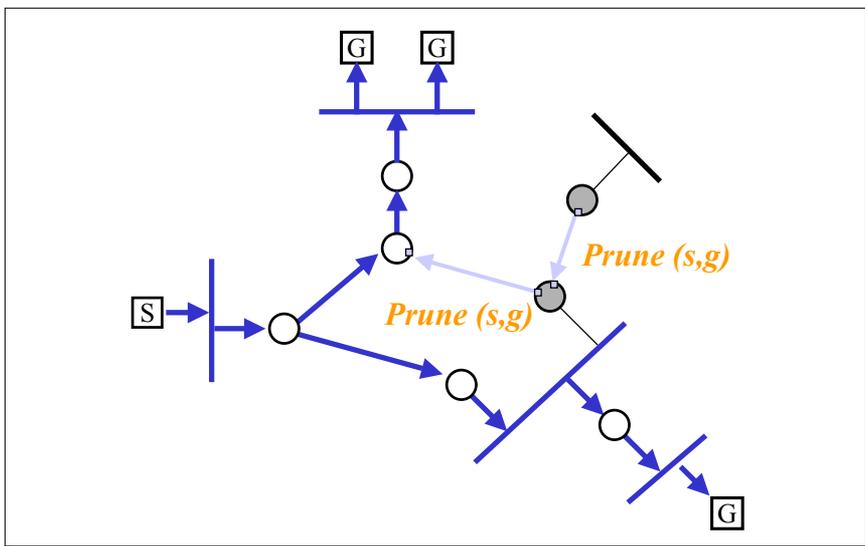
DVMRP (2)

difusión



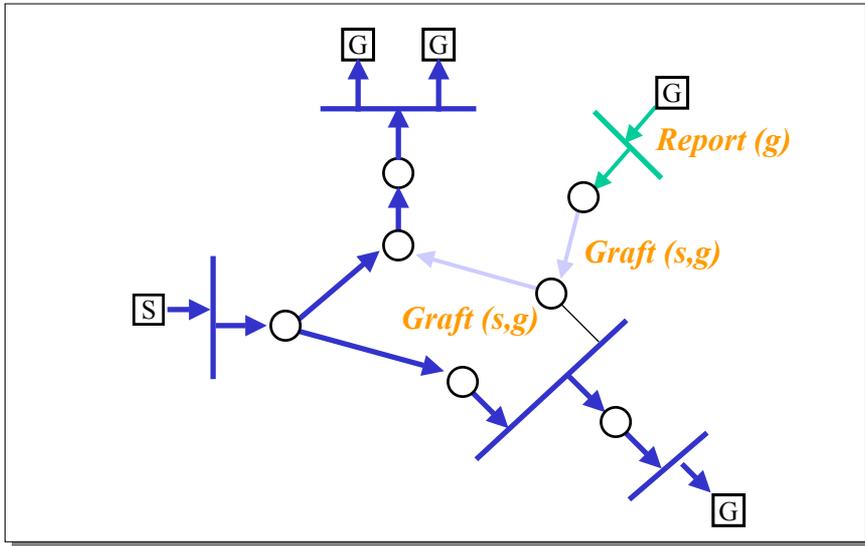
DVMRP (3)

Poda



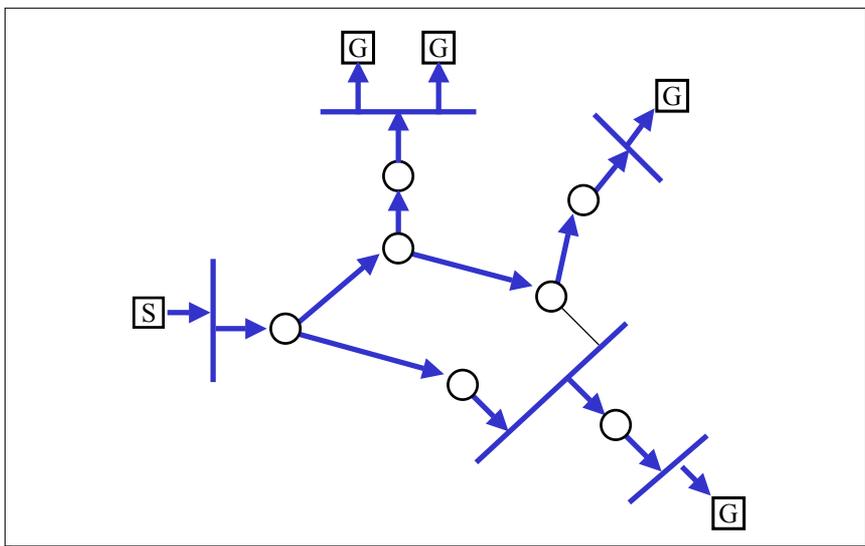
DVMRP (4)

Nuevo miembro (grafting)



DVMRP (5)

Estado estacionario



Mejoras a DVMRP

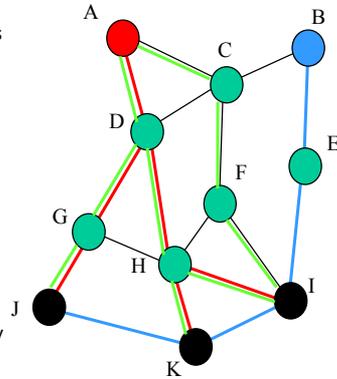
- DVMRP jerárquico
 - Agregación de subredes en prefijos más cortos (cf. CIDR) para reducir el tamaño de la tabla DVMRP
 - Cada dominio de routing ejecuta su propia instancia de un protocolo multicast.
 - Cada router sólo precisa conocer los detalles explícitos de los destinos dentro de su propio dominio.
- Reconocimiento automático de "stub trees"
 - Para reducir el número de mensajes de routing en enlaces lentos
- Soporte de poda específica para nodos, para IGMP v3

Multicast OSPF (MOSPF)

- Añadidos a OSPF (Open Shortest-Path First)
- Routers MOSPF
 - Usan IGMP para monitorizar la pertenencia a grupos multicast en las subredes a las que están directamente conectados.
 - "Base de datos de grupos locales"
 - Determina la responsabilidad del router en el envío de datagramas multicast
 - En un entorno mixto (MOSPF y OSPF) debe elegirse un Router MOSPF como DR de la subred para generar Queries IGMP.
 - El DR es responsable de comunicar la pertenencia de grupos a todos los demás routers en el área MOSPF difundiendo LSAs Group-Membership.
 - Arbol del Camino más corto
 - Para cada par (emisor, grupo) se construye "bajo demanda" cuando un router recibe el primer datagrama multicast para dicho par.
 - Basándose en las Router-LSAs y Network-LSAs, se construye un árbol de camino más corto con raíz en el emisor utilizando el algoritmo de Dijkstra's.
 - Después se utilizan LSAs Group-Membership para podar las ramas que no conducen a subredes con miembros individuales de un grupo.
 - El resultado del cálculo de Dijkstra es una poda del arbol de camino más corto con raíz en el emisor del datagrama.

MOSPF

- Idea básica
 - Aprovechar la ventaja del conocimiento de la topología de la red y las métricas de los enlaces para calcular el árbol de caminos más cortos.
 - Es necesario conocer qué nodos tienen receptores suscritos a un grupo multicast
- Enfoque
 - Los routers difunden información de pertenencia a un grupo mediante LSAs (new group-membership LSA)
 - Los routers calculan independientemente el árbol de camino más corto que sólo incluye routers con capacidad multicast
 - Los árboles son específicos para cada fuente
- Ventaja
 - NO necesita inundación y poda
- Consideraciones
 - Necesita calcular el árbol de camino más corto *por emisor* y grupo
 - Necesita caché en cada router para conseguir eficiencia
 - Pierde precisión para multicast inter-area e inter-dominio

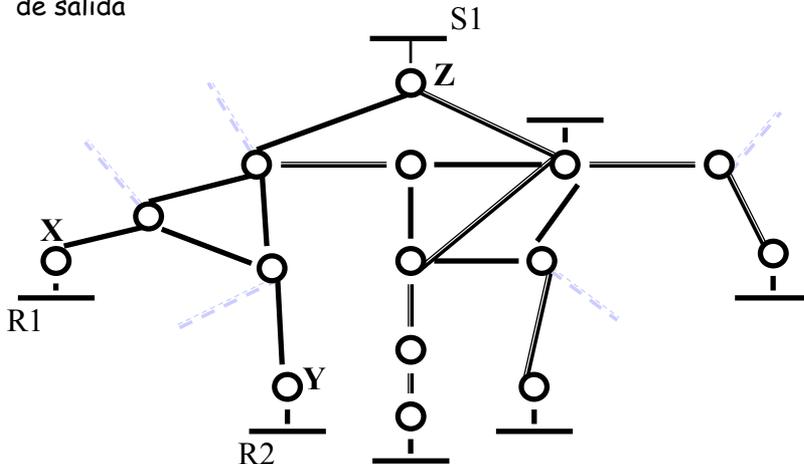


MOSPF (ii)

OSPF: Cada router inunda notificaciones de estado de enlace

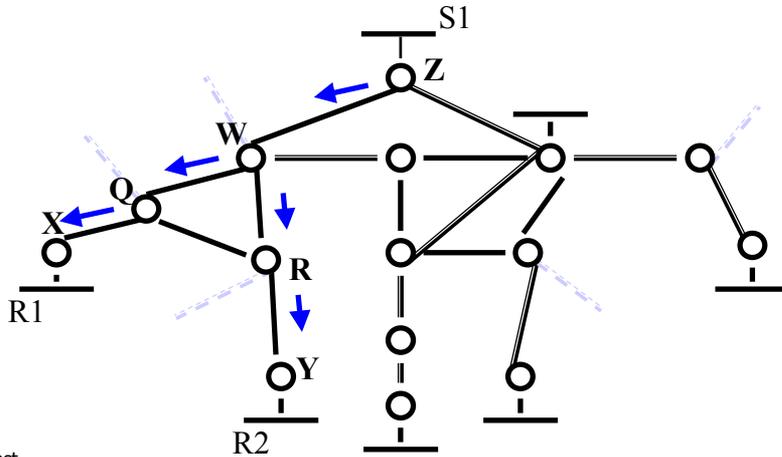
MOSPF: añade información de pertenencia a un grupo al LSA

- Cada router calcula el árbol multicast para cada emisor activo,
- construye entradas de encaminamiento con lista de interfaces de salida



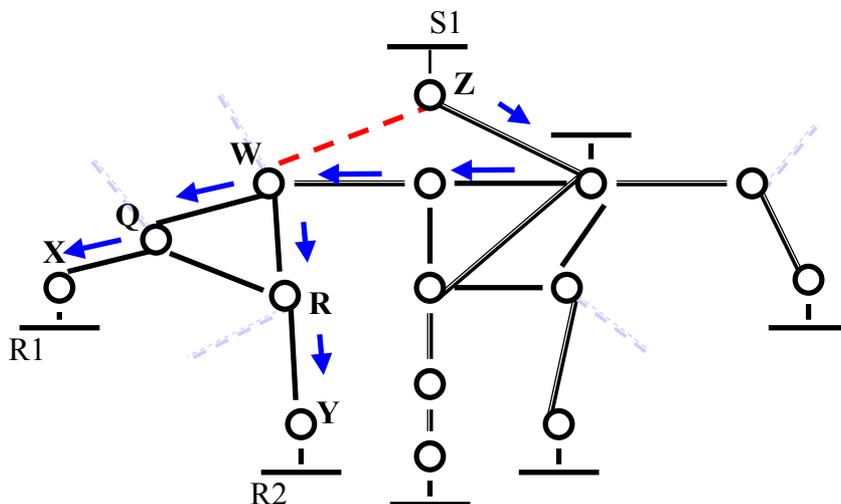
MOSPF (iii)

Z tiene un mapa de red, incluyendo la pertenencia a X e Y
Z calcula el árbol de camino más corto desde S1 a X e Y
Z construye la entrada multicast con un interfaz saliente
W, Q, R, cada uno construyen entradas multicast



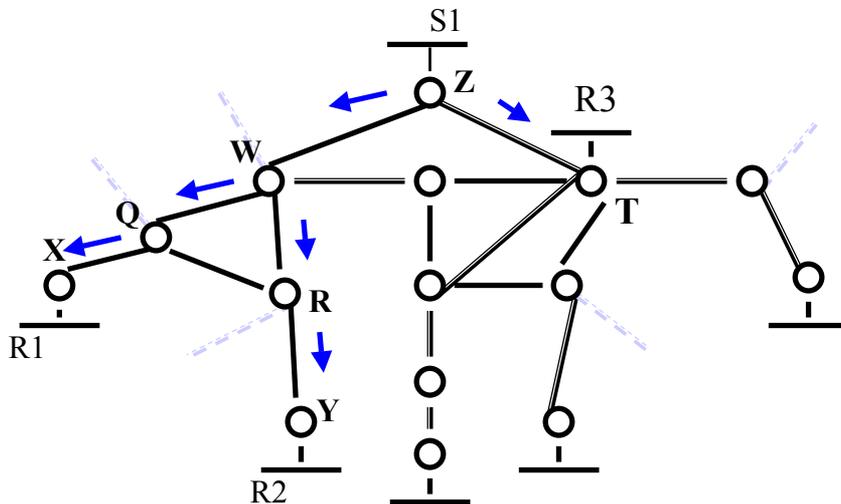
MOSPF (iv)

Una Notificación de Estado de Enlace con nueva topología puede precisar recálculo del árbol y la entrada de encaminamiento



MOSPF (v)

Notificación de enlace (T) con una nueva pertenencia (R3) puede requerir cálculo incremental y adición de un interfaz a la lista de interfaces de salida (Z)

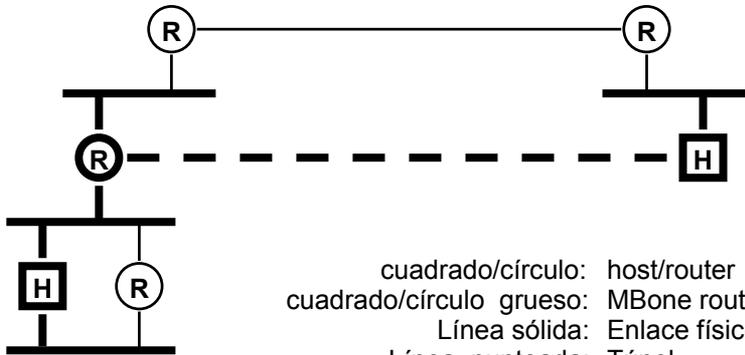


Cálculo de rutas y procesado de LSAs

- No puede precalcular todos los árboles **multicast**
 - Cálculo bajo demanda cuando llega el primer paquete de un emisor S a un grupo G
 - Envía el paquete por todos los interfaces de salida que corresponden a la parte local del árbol
- Procesado de LSAs
 - La llegada de LSAs puede conducir a añadir o eliminar interfaces de salida si contiene diferentes direcciones de grupo
 - La llegada de LSAs puede provocar el cálculo del árbol entero si las rutas entre emisor y destino están afectadas
- Comparte los beneficios de escalado y las limitaciones de OSPF
 - Las áreas proporcionan buena escalabilidad intra-dominios
 - El enfoque de estado de enlace de dos niveles no es escalable a inter-dominios

Multicast backbone: MBone

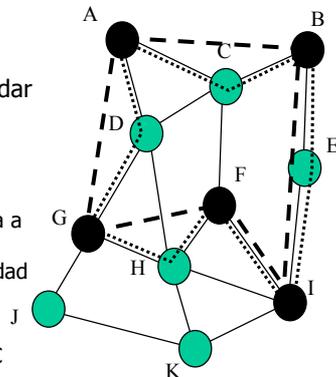
- Una red superpuesta de routers con capacidad multicast usando DVMRP
- herramientas: sdr (session directory), vic, vat, wb



cuadrado/círculo: host/router
cuadrado/círculo grueso: MBone router
Línea sólida: Enlace físico
Línea punteada: Túnel
Línea gruesa: Parte de MBone

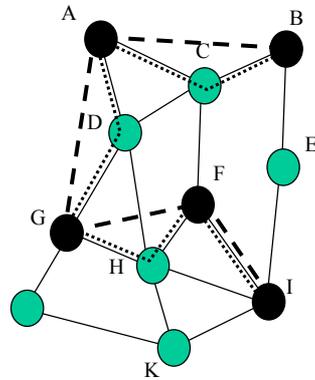
MBone

- MBone = Multicast Backbone
- El encaminamiento multicast es mucho más complejo que el unicast
- ⇒ Además es preciso que *todos* los routers soporten multicast
- Multicast se puede desplegar *sobre* una red estándar unicast
 - Un conjunto de routers con capacidad multicast “interconectados”, que proporcionan el servicio IP multicast en Internet
 - Podemos imaginarlos como una red virtual solapada a Internet (modelo Mbone)
 - Enlaces virtuales (*tuneles*) entre routers con capacidad multicast
- Pero impacta en el cálculo del camino más corto
 - El camino más corto entre A y F no es a través de C sino a través de D
 - El routing multicast necesita utilizar su propio protocolo de routing multicast sobre la red solapada (DVMRP)



Construcción de túneles Multicast

- Propiedades de los túneles
 - Envío de paquetes al extremo del túnel
 - Identificación y reconstrucción del paquete multicast en los extremos
 - Preserva "properties" (actualización de TTL, opciones IP) del paquete original
 - Minimiza el overhead
- Enfoque
 - Se fundamenta en la encapsulación a través de una ruta con encaminamiento flexible (dos elementos)
 - El primero se utiliza transportar la dirección IP del emisor original
 - El segundo elemento transporta la dirección de destino j multicast
 - La dirección del emisor del paquete es la dirección IP del extremo local del túnel
 - La dirección de destino del paquete es la del extremo remoto del túnel.
 - El procesamiento inverso se realiza cuando el paquete se recibe en el extremo remoto



$\langle S, MCAST \rangle$



$\langle A, G \rangle : [S, MCAST]$

Túneles MBone

- Un método para enviar paquetes multicast a través de routers sin capacidad multicast
- Los paquetes IP multicast se encapsulan en paquetes unicast IP (IP-in-IP) dirigidos al extremo del túnel:

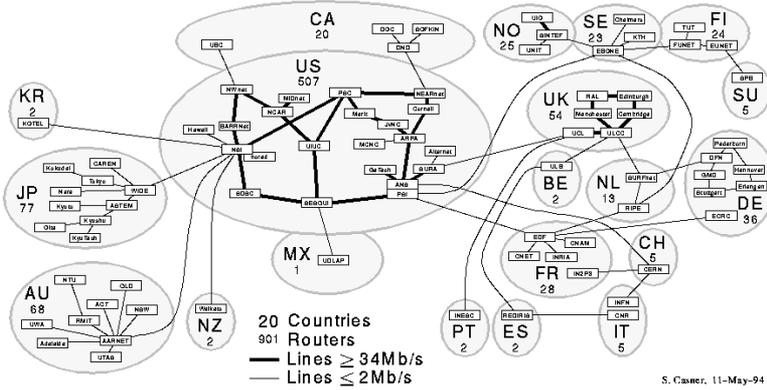
IP header, dest = unicast	IP header, dest = multicast	Transport header and data...
------------------------------	--------------------------------	---------------------------------

- El túnel actúa como un enlace punto a punto virtual
 - Los routers intermedios sólo ven la cabecera exterior
 - El extremo del túnel reconoce IP-in-IP (protocolo = 4) y desencapsula el datagrama para procesarlo
- Cada extremo del túnel se configura manualmente con las direcciones unicast del otro extremo

M-BONE in 1994

- Iniciado en Marzo de 1992 para difusión de audio del IETF meeting (San Diego)
- Ultima topología
 - <ftp://ftp.parcftp.xerox.com/pub/net-research/mbone/maps/mbone-map-big.ps>
 - Alrededor de 6000 (S,G) entradas
- Discusión en mbone@isi.edu

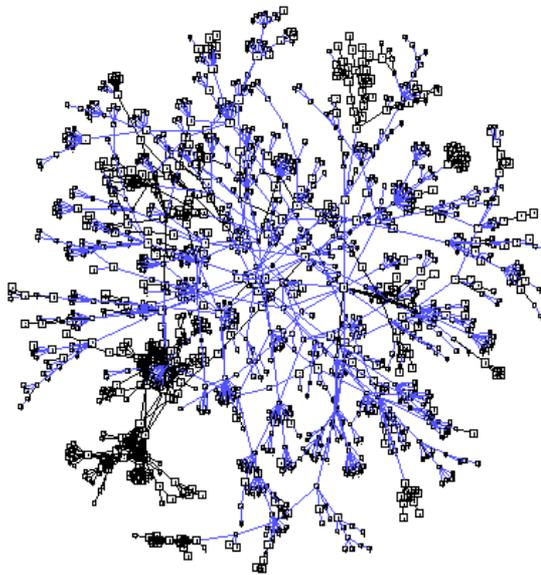
Major MBONE Routers and Links



IP Multicast

71

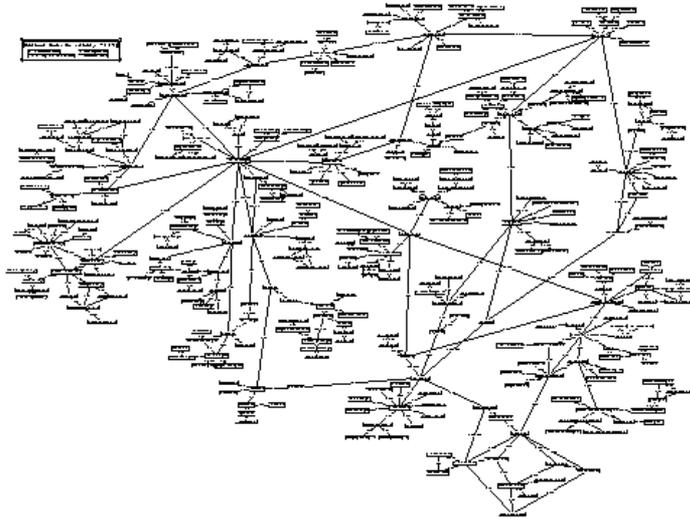
M-BONE in 1996



IP Multicast

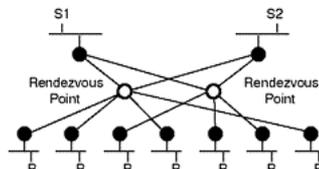
72

M-BONE in 1998



Protocol Independent Multicast (PIM)

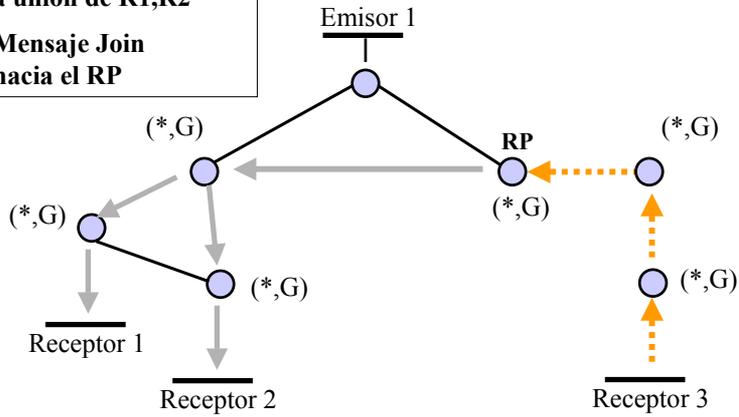
- Soporta árboles compartidos y con raíz en el emisor
 - Modo denso (árboles con raíz en el emisor)
 - Similar a DVMRP
 - Modo disperso (árboles compartidos)
 - Núcleo = rendezvous point (RP)
- Independiente del protocolo de routing unicast
 - Sólo usa tablas de encaminamiento unicast
- Pasos básicos
 - Los routers con miembros locales se unen con el Rendezvous Point (RP) para unirse al árbol
 - Los routers con emisores locales encapsulan los datos en mensajes Register al RP



PIM : Construir el árbol compartido

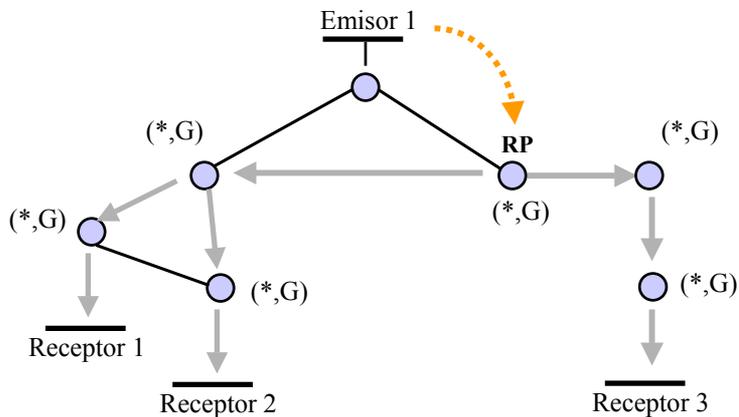
→ Arbol compartido tras la unión de R1,R2

→ Mensaje Join hacia el RP



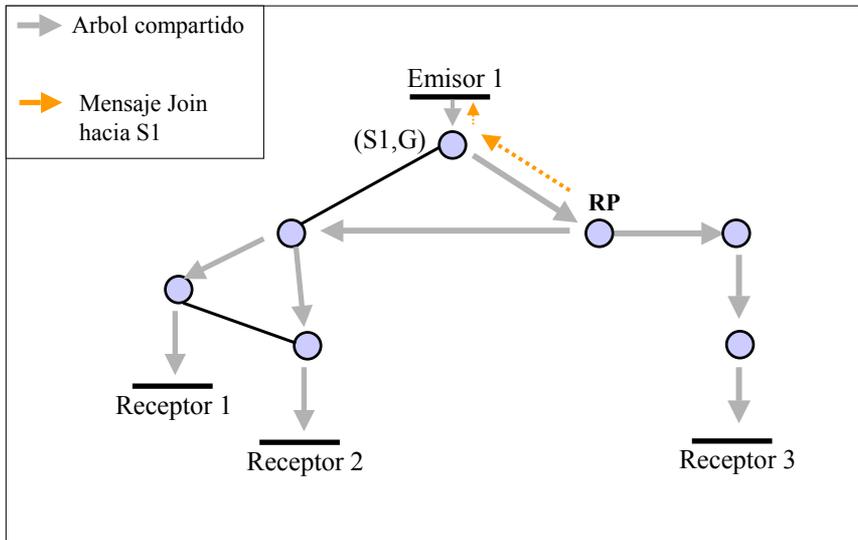
PIM: Datos encapsulados en Register

Datos unicast encapsulados al RP en un Register

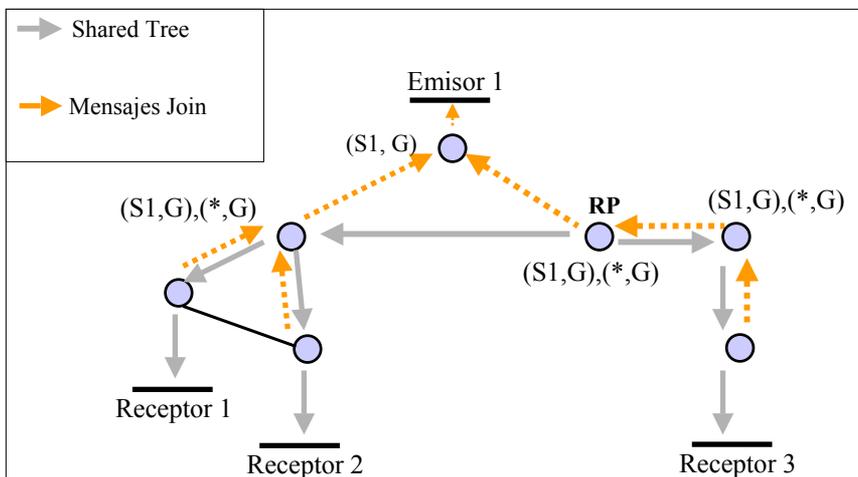


RP desencapsula, hacia abajo en el árbol compartido

PIM: RP envía Join a un emisor de alta velocidad

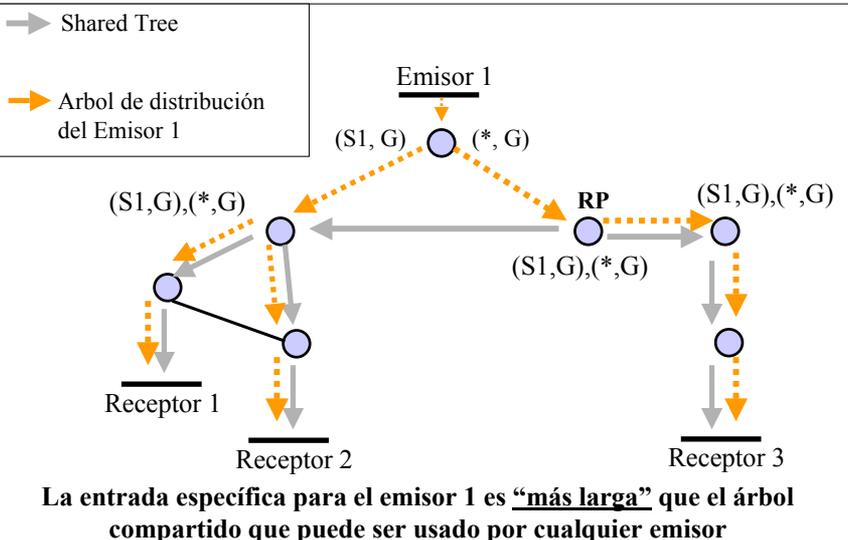


PIM: Construcción de un árbol de distribución específico de un emisor

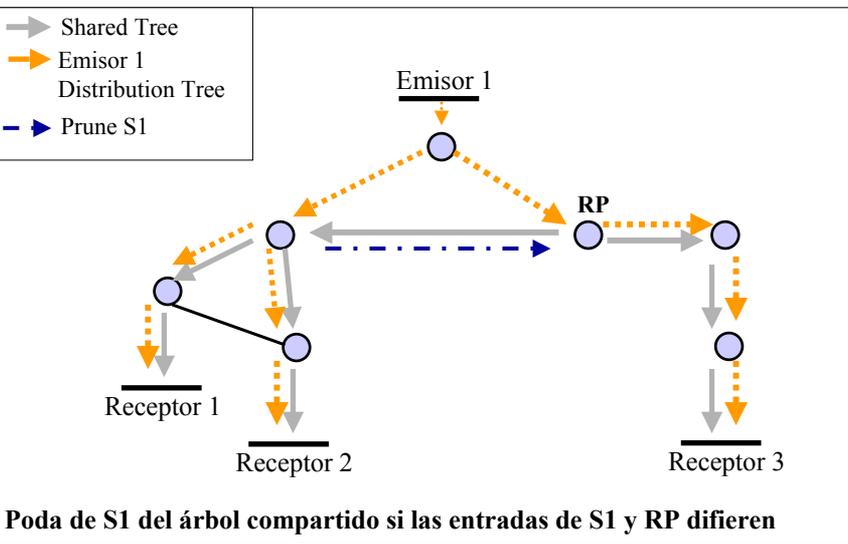


Se construye un árbol específico para un emisor de alta velocidad

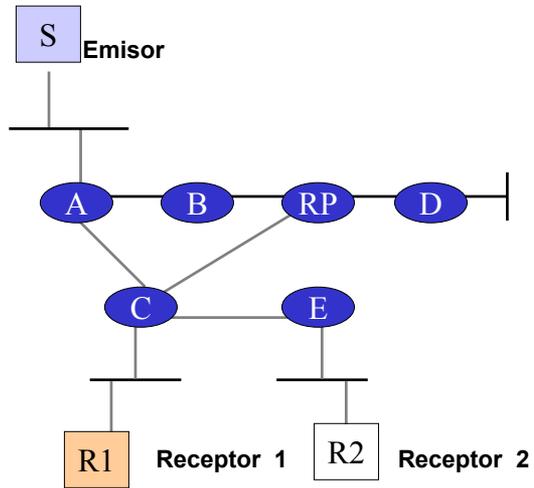
PIM: Encaminamiento por la entrada “coincidente más larga”



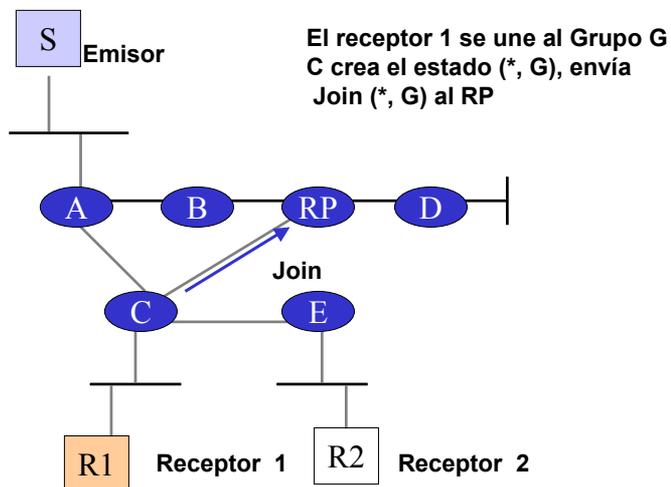
PIM: Poda de S1 del Arbol compartido



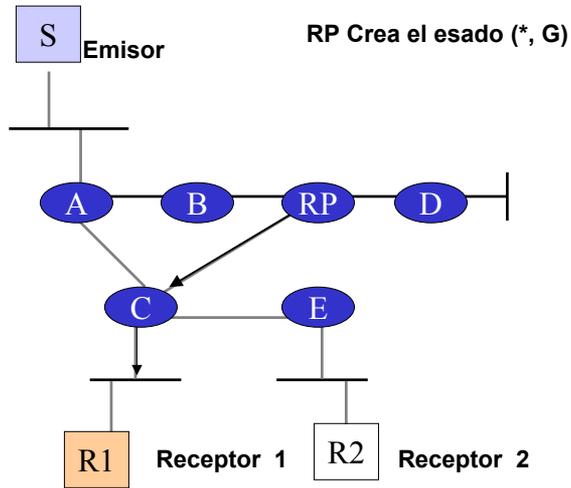
PIM-SM(1)



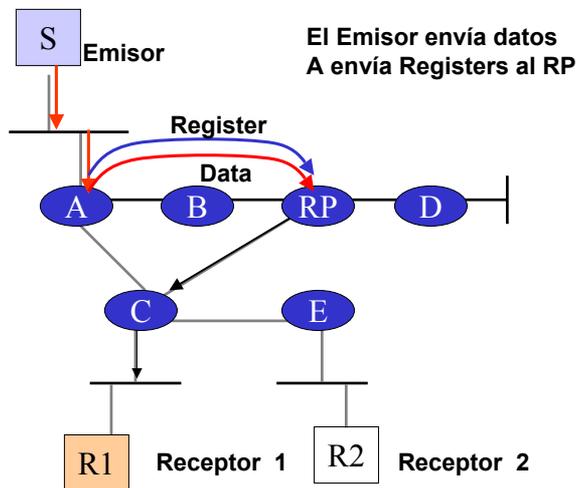
PIM-SM(2)



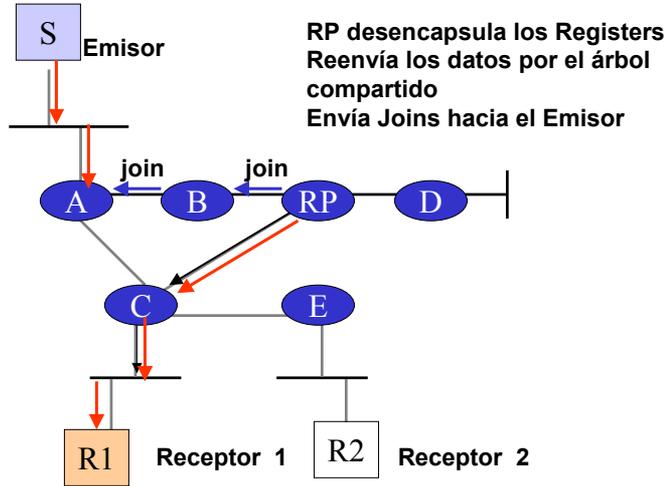
PIM-SM(3)



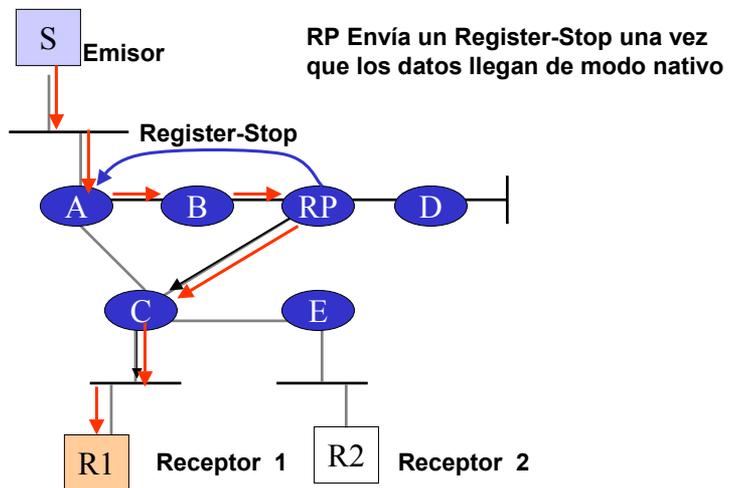
PIM-SM(4)



PIM-SM(5)

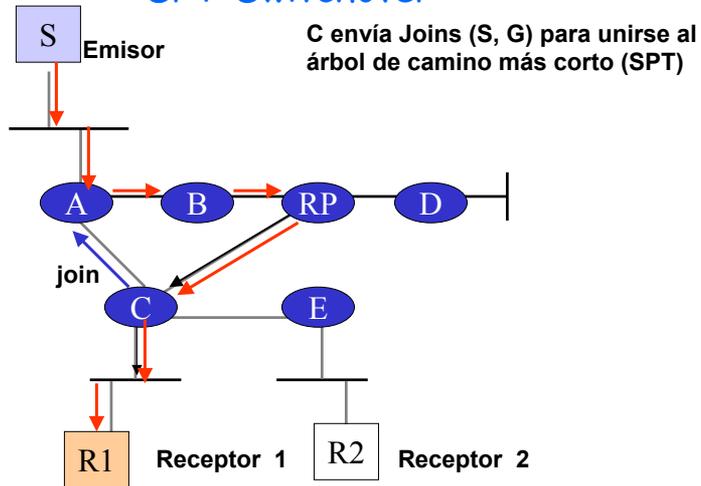


PIM-SM(6)

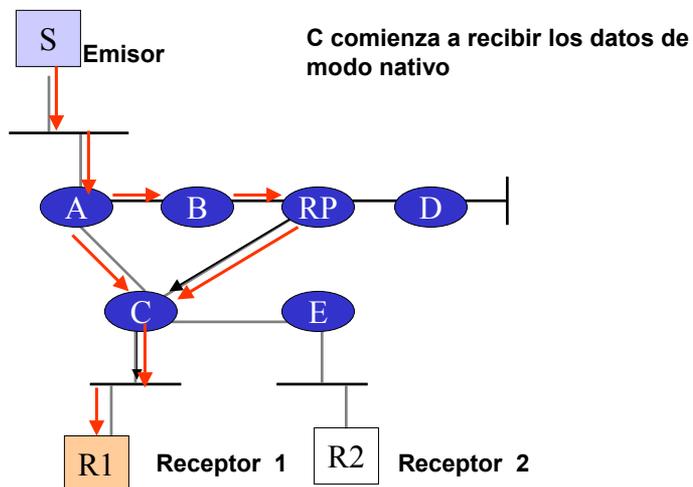


PIM-SM(7)

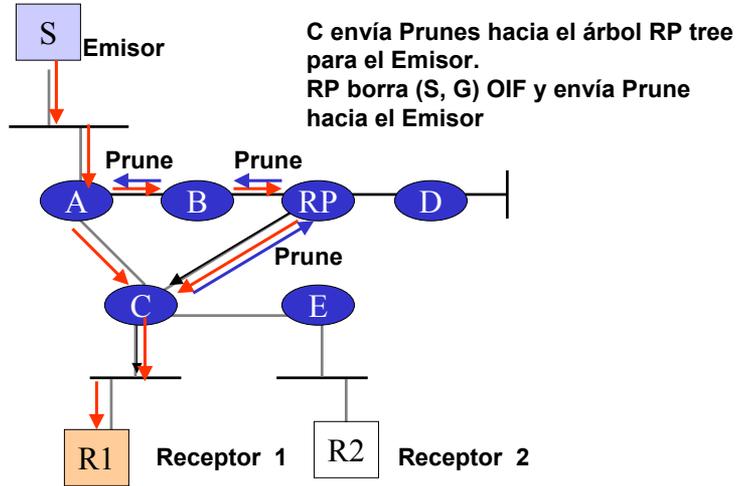
SPT Switchover



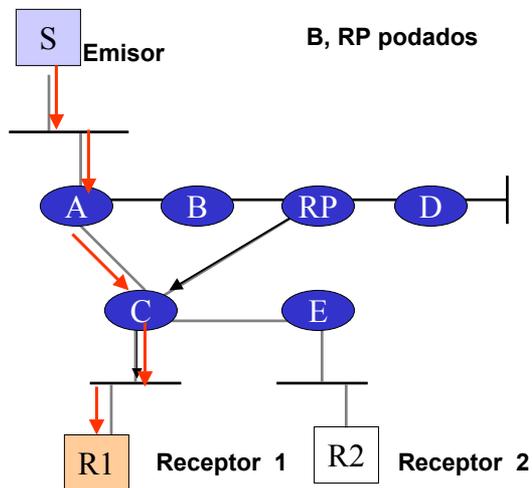
PIM-SM(8)



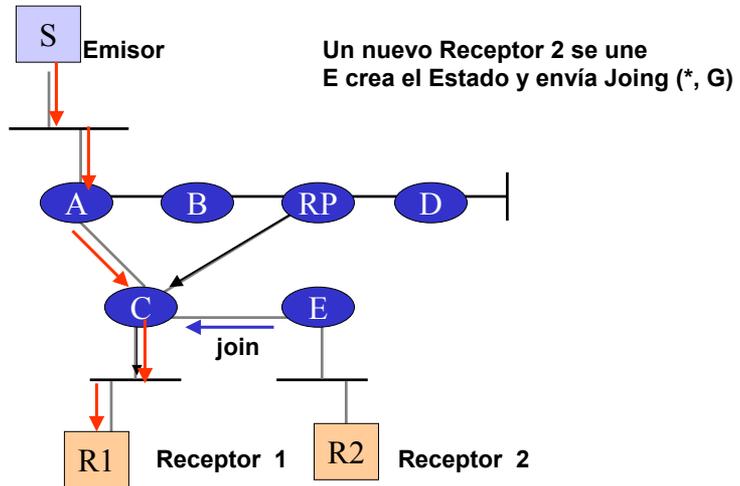
PIM-SM(9)



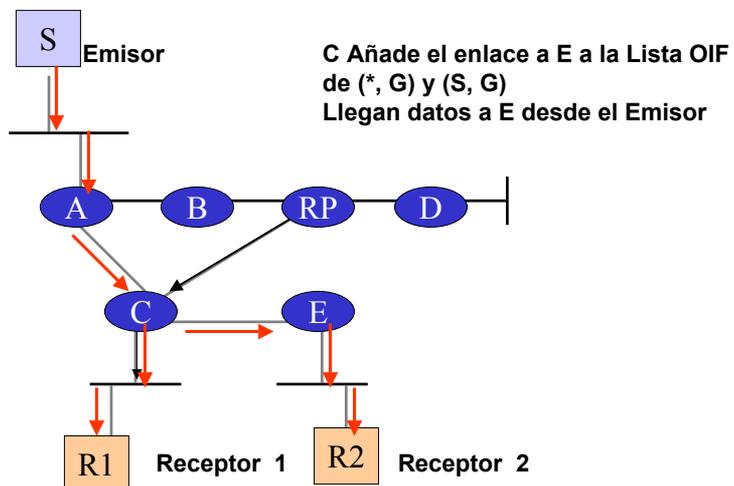
PIM-SM(10)



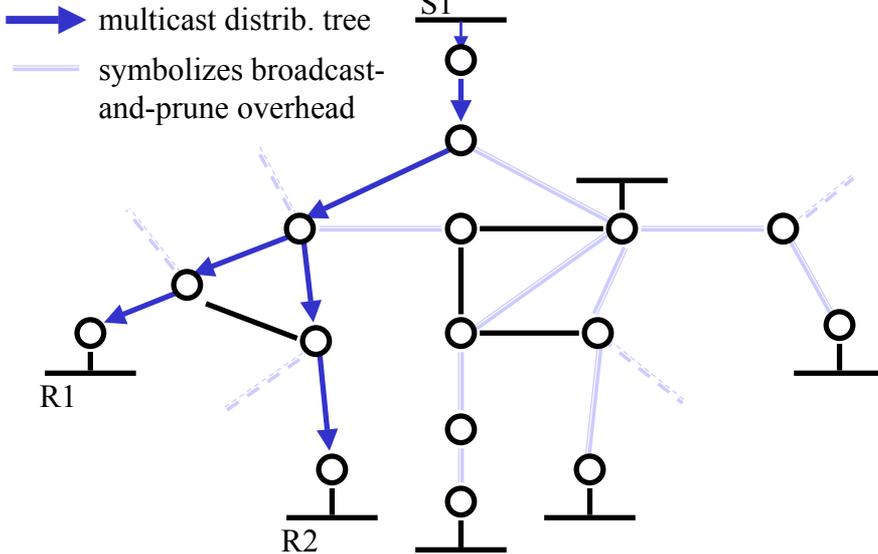
PIM-SM(11)



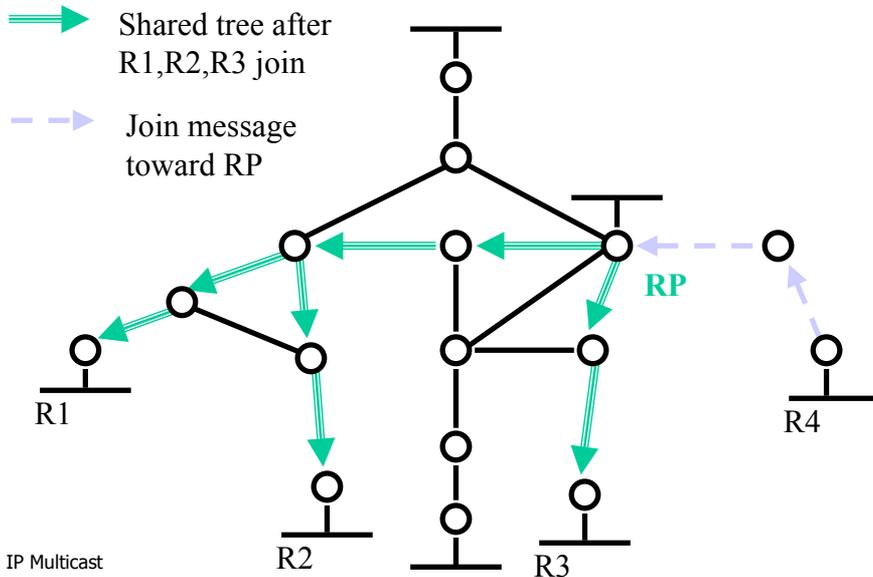
PIM-SM(12)



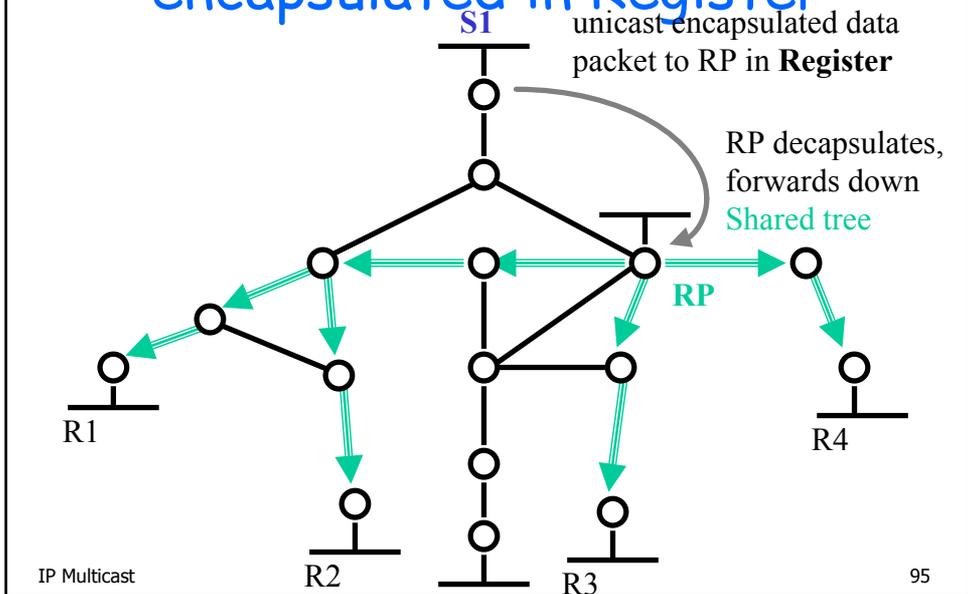
PIM SM Example(2): Broadcast-and-Prune Overhead



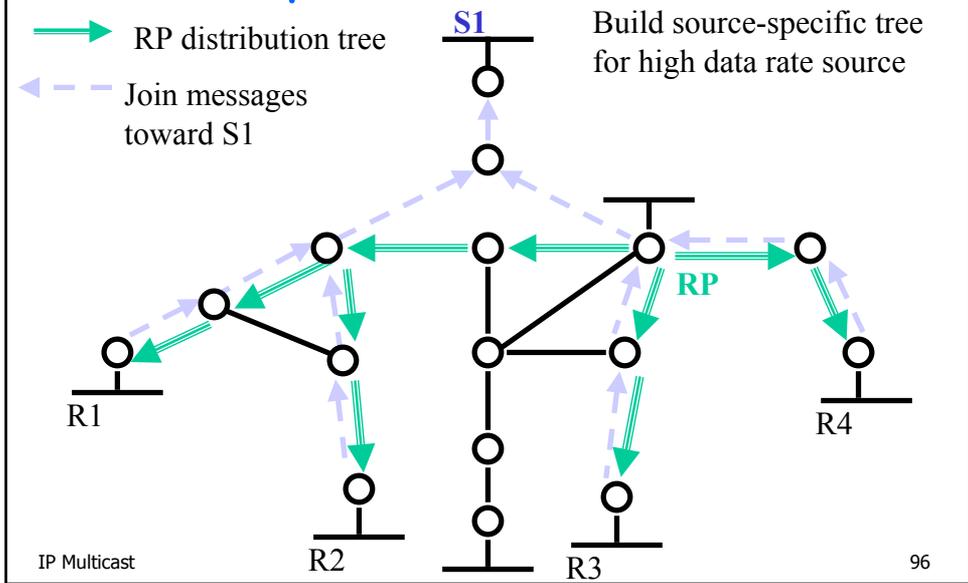
PIM SM Example(2): build Shared tree



PIM SM Example(2):Data encapsulated in Register



PIM SM Example(2):Build Source-specific distribution tree

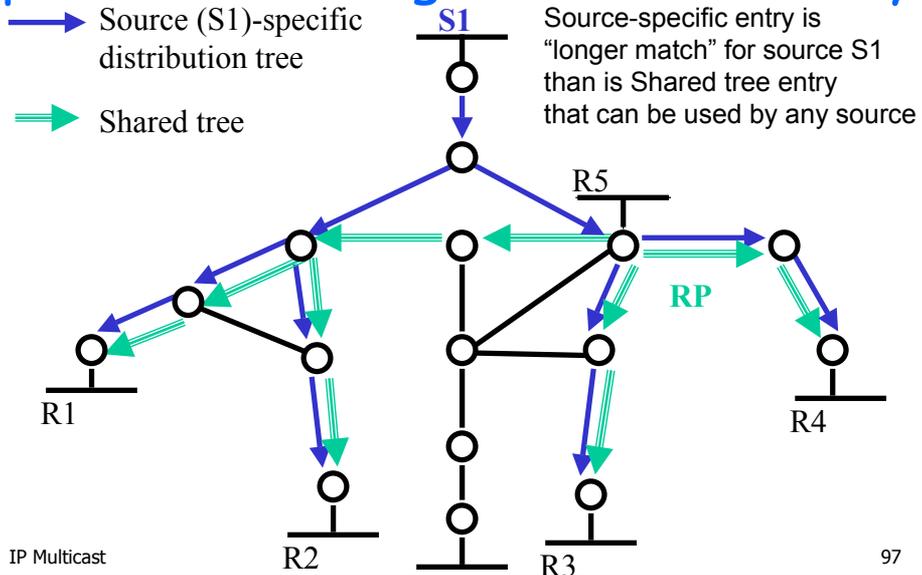


PIM SM Example(2): Forward packets on "longest-match" entry

— Source (S1)-specific distribution tree

— Shared tree

Source-specific entry is "longer match" for source S1 than is Shared tree entry that can be used by any source

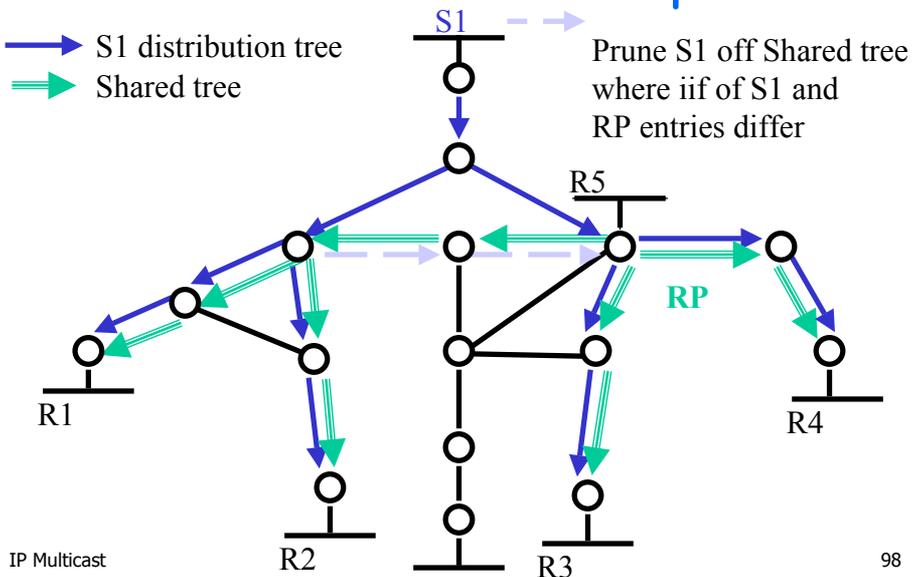


PIM SM Ej(2): Prune S1 off Shared tree to avoid duplicates

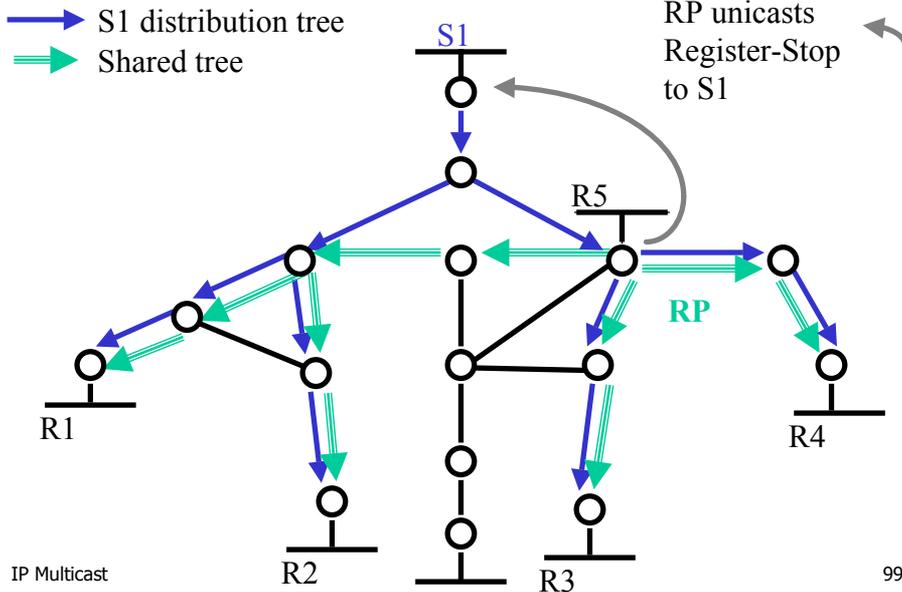
— S1 distribution tree

— Shared tree

Prune S1 off Shared tree where iif of S1 and RP entries differ



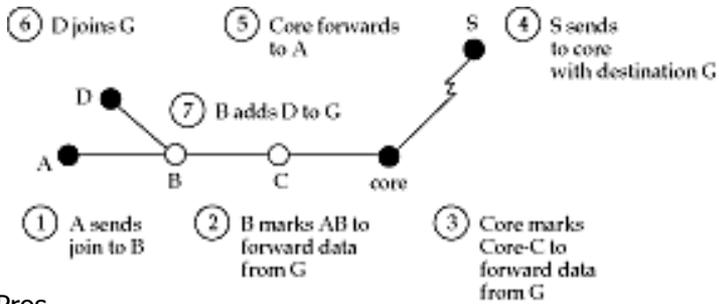
PIM SM EJ(2): RP sends Register-Stop to S1 when packets received natively



Core-Based Trees

- Goal is to avoid
 - Inefficiency of DVMRP
 - Complexity of MOSPF
- Key idea
 - *Core* router is responsible for coordinating multicast forwarding and joining/leaving of multicast group
 - Joining is by having host send a *join* request to core router
 - Join process is used to automatically build the tree
 - Forwarding relies on two components
 - Routers along path followed by the *join* message mark incoming interface for forwarding
 - Two cases to consider for senders
 - Sender attached to a router on the tree can directly send packets on the multicast tree
 - Sender attached to a router not on the tree needs to unicast its packets to the core, which decapsulates them and forwards them on the tree

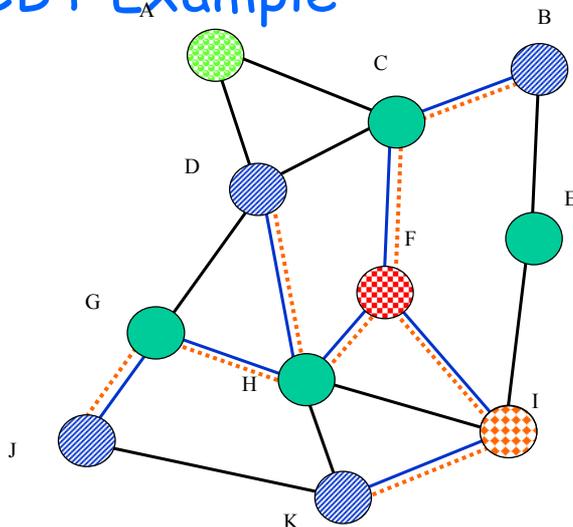
CBT Example



- Pros
 - routers not part of a group are not involved in pruning
 - explicit join/leave makes membership changes faster
 - router needs to store only one record per group
- Cons
 - core handles traffic from all remote senders and may be a bottleneck
 - traffic may travel on non-optimal paths

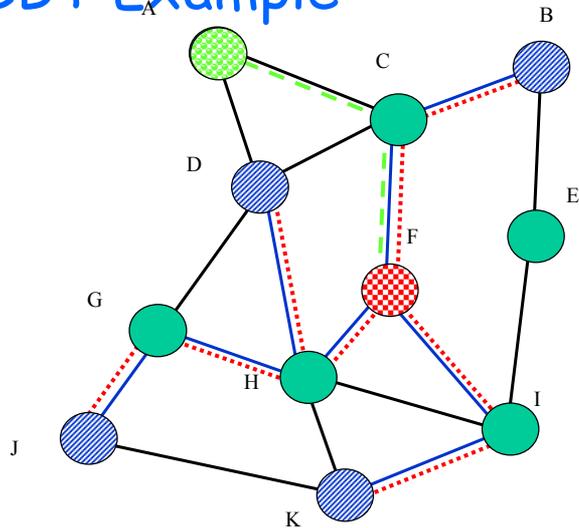
CBT Example

-  Joined receiver
-  On-tree sender
-  Tree link
-  Core router



CBT Example

-  Joined receiver
-  Off-tree sender
-  Tree link
-  Core router



PIM DM

- Difusión y poda resulta ideal para grupos densos
- El árbol con raíz en el emisor se crea bajo demanda usando la regla RPF
- Si el emisor se desactiva, el árbol es "derrivado"
- Fácil plug-and-play
- Las ramas que no quieren datos son podadas
- Se usan "Grafts" para unirse a árboles ya existentes
- Utiliza "Asserts" para determinar el emisor para LAN multiacceso
- Poda en enlaces Punto a punto no-RPF
- Velocidad de poda limitada en enlaces Punto a punto no-RPF

Encaminamiento PIM DM

- Los interfaces PIM DM se ubican en la "oilst" para un grupo multicast si:
 - Se escucha un vecino PIM en el interfaz
 - Un nodo en este interfaz se ha adherido al grupo
 - El Interfaz se ha configurado manualmente para unirse al grupo
- Los paquetes se "difunden" por todos los interfaces de la "oilst"
 - Si hay un vecino PIM presente, DM supone que TODOS quieren recibir el grupo de modo difunde a dicho enlace por definición

Mecanismo "Assert" PIM

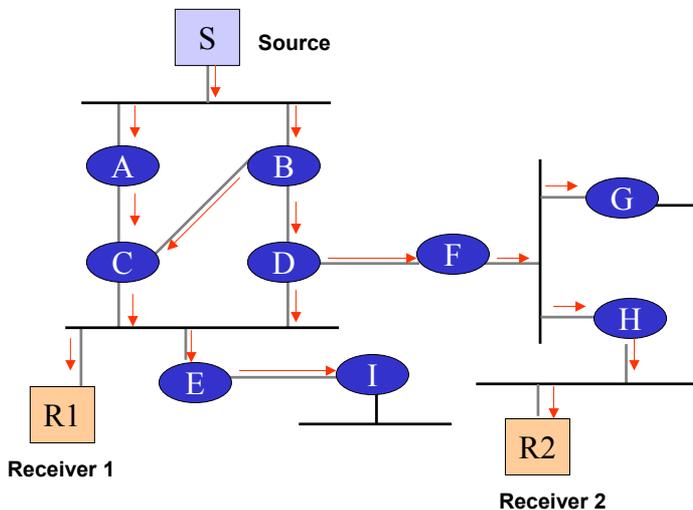
- Los routers reciben paquetes por cualquier interfaz de su "oilst"
 - Sólo un router continuaría enviando para evitar paquetes duplicados
- Los routers envían mensajes "PIM assert"
 - Compara los valores de distancia y métrica
 - El router con mejor ruta al emisor gana
 - Si métrica & distancia son iguales, el de mayor dirección IP gana
 - El router perdedor deja de enviar (poda el interfaz)

Mantenimiento de estados PIM DM

- El estado se mantiene por un comportamiento "difusión y poda" del modo denso.
 - A la recepción de paquetes multicast se reinician los contadores de expiración de la entrada (S,G).
 - Cuando los temporizadores de la entrada (S,G) expiran la entrada se elimina.
- Los temporizadores de estado de poda de interfaz expiran cada 210 seconds originando redifusiones y podas periódicas

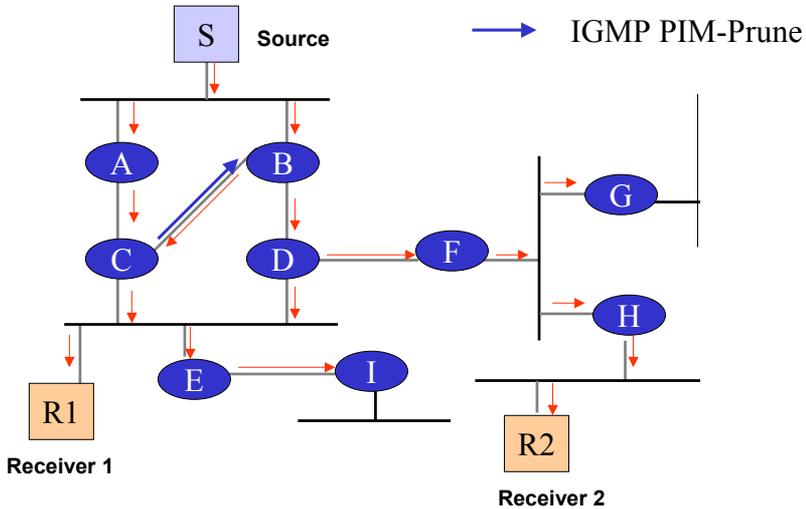
PIM-DM(1)

Initial flood of data



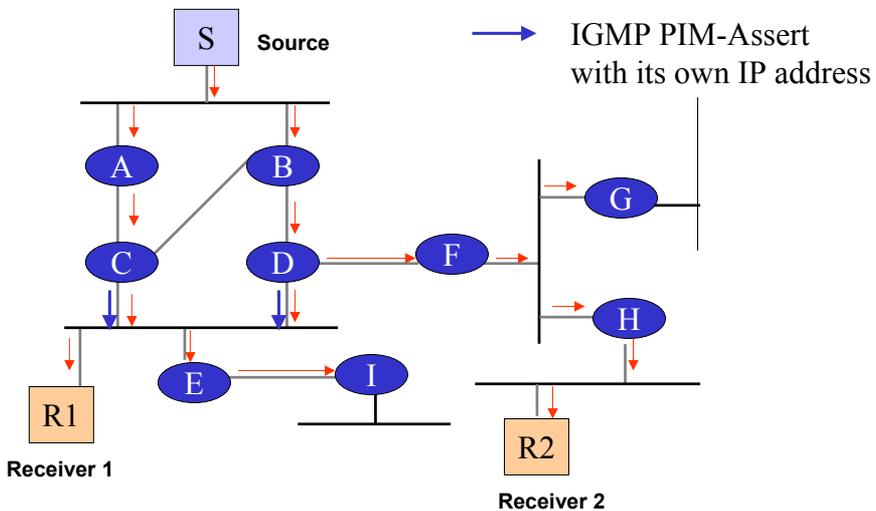
PIM-DM(2)

prune non-RPF p2p link



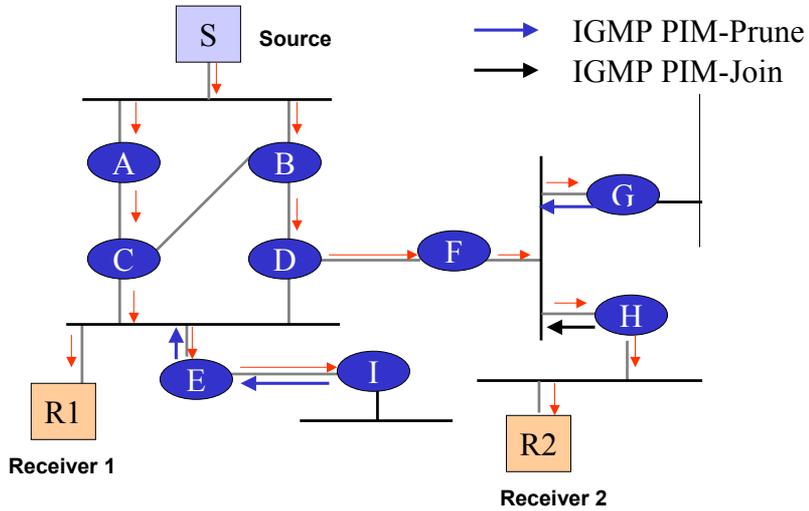
PIM-DM(3)

C and D Assert to Determine Forwarder for the LAN, C Wins



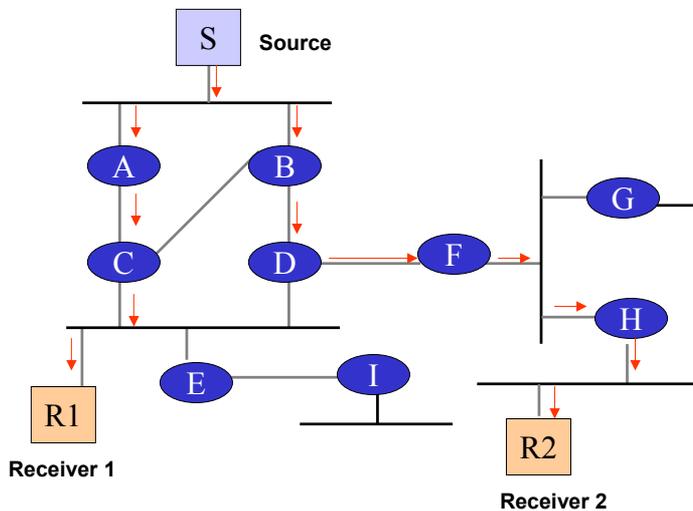
PIM-DM(4)

I, E, G send Prune
H send Join to override G's Prune



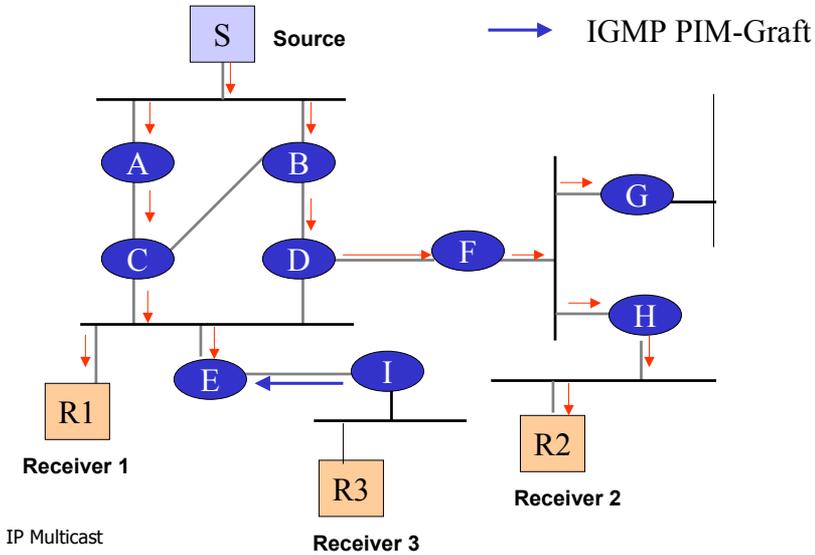
PIM-DM(5)

I Gets Pruned
E's Prune is Ignored
G's Prune is Overridden



PIM-DM(6)

New Receiver, I send Graft



PIM-DM(6)

new branch

