# MBONE UNICAUCA Y SU CONEXIÓN AL MBONE INTERNET


**Maritza Piedad Joaquí Chimachaná**

**Fabio Andrés Rodríguez Tejada**

**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Departamento de Telecomunicaciones**

**Popayán**

**2004**

# MBONE UNICAUCA Y SU CONEXIÓN AL MBONE INTERNET

**MARITZA PIEDAD JOAQUÍ CHIMACHANÁ**
**FABIO ANDRÉS RODRÍGUEZ TEJADA**

**Monografía de Grado**

**Director: Ing. Francisco Javier Terán**

**Universidad del Cauca**
**Facultad de Ingeniería Electrónica y Telecomunicaciones**
**Departamento de Telecomunicaciones**
**Popayán**
**2004**

# TABLA DE CONTENIDO

# LISTA DE FIGURAS

# LISTA DE TABLAS

# INTRODUCCIÓN

La tecnología Multicast es un área relativamente nueva en el campo de las redes de telecomunicaciones. Sus orígenes se remontan a principios de los 80's y en los años 90's ya existía un backbone dedicado exclusivamente a su uso montado sobre el backbone de Internet. En este momento a nivel mundial se encuentran diferentes organizaciones trabajando alrededor de ésta área tanto en infraestructura de redes como en aplicaciones, entre ellas la Universidad del Cauca. Ha sido tanta la fuerza de esta línea de trabajo que es una de las principales características del nuevo Protocolo de Internet IPv6 y cualquier interfaz de red que pretenda ser compatible con IPv6 es implícitamente compatible con Multicast. Es tal su importancia que reemplaza al mecanismo clásico de difusión: el broadcast; además se utiliza como soporte de algunos protocolos en la pila TCP/IP.

La tecnología Multicast es un mecanismo que ofrece algunas ventajas que pueden ser aprovechadas para enriquecer los procesos de comunicación de la comunidad universitaria, por consiguiente, la Red de Datos de la Universidad del Cauca debe apropiarse de esta tecnología como parte de su continua evolución para mejorar su desempeño y la calidad de los servicios que presta a los usuarios de la Universidad.

Las Telecomunicaciones se encargan de crear plataformas que sirven de soporte para desarrollo de aplicaciones telemáticas, el backbone Multicast es una de estas plataformas que ofrece características especiales en el manejo de información concurrente; el impacto inmediato del proyecto está condicionado por el software existente, pero se abre una gran posibilidad para crear nuevas herramientas que se adecuen a las necesidades del entorno de la Universidad del Cauca. Las aplicaciones de voz y video en tiempo real, de transferencia de datos y de administración constituirán una nueva forma de comunicación en la

Universidad, generando mecanismos que aportan al desarrollo de proyectos de tele-medicina, tele-educación, entre otros. Toda la infraestructura de la Red de Datos de la Universidad está cubierta por el diseño de este backbone, pero por razones administrativas la implementación sólo se lleva a cabo en el edificio de ingenierías que se conecta directamente con el Mbone Internet.

La necesidad de evolución de la red de datos de la Universidad y las ventajas de la tecnología Multicast hacen que este proyecto sea práctico y oportuno para la comunidad universitaria.

# 1 INTRODUCCIÓN A MULTICAST SOBRE IP

En la comunicación IP unicast, un host fuente envía paquetes a un host destino específico. En este caso, la dirección de destino en el paquete IP es la dirección de un único y exclusivo host en la red IP. Estos p
aquetes IP se reenvían a través de la red desde el host fuente hasta el host destino por medio de los enrutadores y demás dispositivos de red. Los enrutadores en cada punto, entre la fuente y el destino, utilizan sus bases de información de enrutamiento unicast (RIB, Routing Information Base) para tomar decisiones de envío unicast basadas en la dirección IP destino del paquete.

En IP Broadcast un host fuente envía paquetes a todos los hosts sobre un segmento de red. La dirección de destino de un paquete broadcast IP tiene los bits correspondientes al host fijados a 1 y los bits correspondientes a la red con el valor de la dirección de red. Los hosts IP (incluyendo a los enrutadores) entienden que los paquetes que contienen direcciones broadcast IP como dirección de destino están dirigidos a todos los hosts en la subred. A menos que específicamente se configure, los enrutadores no reenviarán paquetes broadcast IP y por lo tanto la comunicación broadcast IP normalmente se limita a la subred local. Si el objetivo es permitir que un host envíe paquetes IP a otros hosts que no estén en la subred local, entonces la difusión broadcast no es suficiente para cumplir este objetivo.

La difusión Multicast se encuentra entre las comunicaciones IP unicast y broadcast; permite a un host enviar paquetes IP a un grupo de hosts donde quiera que ellos estén dentro de la red IP; la dirección de destino en un paquete multicast IP es una forma especial de dirección llamada "IP Multicast Group Address". Los enrutadores multicast IP (también conocidos como "mrouters") tienen que reenviar los paquetes multicast entrantes a todas las interfaces que

les permitan alcanzar a miembros del grupo IP multicast. La dirección del grupo multicast se especifica en el campo de dirección de destino en el paquete IP.A principios de los 80's en la Universidad de Stanford un estudiante de doctorado, Steve Deering, trabajó en el proyecto de un sistema operativo distribuido llamado Vsystem compuesto de muchos computadores unidos en un sistema de multiprocesamiento acoplado a través de un segmento ethernet común. Los computadores en este segmento ethernet trabajaban conjuntamente y se comunicaban a nivel de sistema operativo a través de mensajes especiales enviados sobre el segmento ethernet común; una de las primitivas del sistema operativo permitía a un computador enviar mensajes a un grupo de computadores en el segmento ethernet local utilizando multicast a nivel MAC.

Con el tiempo creció la necesidad de añadir mas computadores al sistema de multiprocesamiento. El problema era que los únicos computadores disponibles estaban al otro lado del campus con enrutadores entre las dos redes; en consecuencia, la transmisión de mensajes multicast a nivel MAC tendría que ser extendida para que trabajara a nivel 3 en las redes que hacen uso de enrutamiento, de tal manera que los computadores al otro lado del campus pudieran funcionar como parte del sistema.

Después de estudiar los protocolos de enrutamiento OSPF (Open Shortest Path First) y RIP (Routing Information Protocol), se concluyó que el mecanismo de estado de enlace del protocolo OSPF podía ser extendido para soportar multicast y que el mecanismo básico de RIP serviría como base de un nuevo protocolo de enrutamiento multicast basado en vector distancia; la idea condujo a más investigaciones en el área de IP Multicast convirtiéndose en la temática de la tesis doctoral de Steve Deering "Multicast Routing in a Datagram Network" publicada en diciembre de 1991.

En la tesis de Steve Deering también se describió el Protocolo de Afiliación de Host, (Host Membership Protocol), que sirvió de base para el Protocolo de Gestión de Grupos Internet (IGMP, Internet Group Management Protocolo), que es utilizado por los hosts multicast para informarle al enrutador de la red su interés de unirse a un grupo multicast. Adicionalmente se describía el Protocolo de Enrutamiento IP Multicast basado en Vector Distancia que fue la base del Protocolo DVMRP (Distance Vector Multicast Routing Protocol) desarrollado pocos

años después. Estos dos protocolos fueron la base en la extensión de multicast a nivel 3 del modelo OSI.

Desde entonces los avances de la tecnología multicast IP han continuado y se han desarrollado protocolos como PIM (Protocol Independent Multicasting) y MBGP (Multiprotocol Border Gateway Protocol).

## 1.1 VENTAJAS DE MULTICAST SOBRE IP

En el caso de Internet y muchas compañías que han crecido en términos del número de usuarios conectados, un amplio número de ellos frecuentemente desean acceder casi al mismo tiempo a la misma información. Utilizando las técnicas de multicast sobre IP para distribuir esta información puede reducirse en gran medida la demanda de ancho de banda total de la red. Un buen ejemplo es el rápido crecimiento de contenido de audio y video en la Web. La tecnología multicast provee las siguientes ventajas.

### 1.1.1 Ancho de banda

Si en una compañía se desea transmitir flujo de audio en tiempo real a 8 Kbps, la línea sólida de la figura 1.1 muestra que a mayor número de suscriptores unicast, la cantidad de ancho de banda de la red se incrementa linealmente; por otra parte, si se utiliza multicast para el mismo propósito (representado por la línea discontinua), un flujo de datos único multicast puede entregar el audio a la misma cantidad de suscriptores.

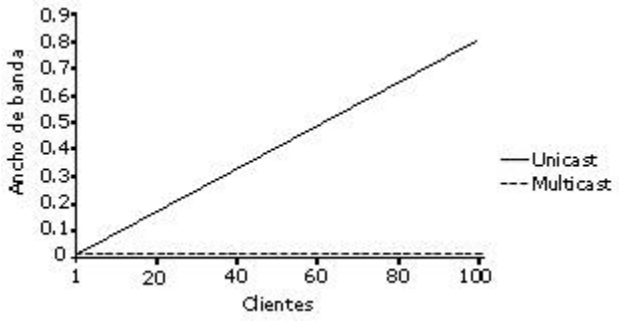


**Figura 1.1 Ancho de Banda para audio, Unicast vs. Multicast**

### 1.1.2 Carga del Servidor

Al mismo tiempo que el número de clientes conectados aumenta, la carga en el servidor se va a incrementar hasta el punto en el que no va a ser capaz de entregar el flujo de datos de 8 Kbps necesarios para distribuir audio sin interrupciones, lo que no es conveniente para una empresa que ofrezca este servicio; si este fuera el caso, la empresa tendría que aumentar el poder de CPU del servidor y acomodar el ancho de banda de su interfaz para servir a más y más clientes; si esto no fuese suficiente, se tendría que proveer múltiples servidores de audio de tiempo real para satisfacer la demanda.

Por otra parte, si se utilizara multicast sobre IP para entregar el flujo de audio, sólo se necesitaría un servidor para alcanzar a todos los clientes. Debido a esto no se necesitaría adquirir más servidores o aumentar el poder de CPU a medida que el número de clientes crece. Es claro que multicast ofrece una ventaja significativa en la reducción de la potencia de máquina de los servidores.

### 1.1.3 Carga de la Red

Debido a que el uso de multicast puede reducir significativamente los requerimientos de ancho de banda cuando se entrega el mismo contenido a múltiples clientes, también se reduce la carga de los enrutadores dentro de la red.

## 1.2 DESVENTAJAS DE MULTICAST SOBRE IP

A pesar de que hay buenas razones para utilizar multicast, es necesario tener en cuenta que también hay limitaciones y desventajas de esta tecnología.

### 1.2.1 Entrega de paquetes no confiable

Tanto la tecnología multicast como unicast son inherentemente no confiables. Sólo a través de TCP (a nivel 4) es posible que un flujo de información sea confiable. Sin embargo, como multicast asume un comportamiento de uno a muchos, no fue diseñado para utilizar el mecanismo punto a punto implícito de TCP. Los paquetes multicast generalmente utilizan UDP, cuya naturaleza es la del mejor esfuerzo, por lo tanto una aplicación que use multicast tiene que soportar

la pérdida de paquetes ocasionalmente y de alguna manera manejar esto a través de la capa de aplicación.

### 1.2.2 Duplicación de paquetes

La duplicación de paquetes es un hecho, al igual que en unicast UDP; sin embargo, una diferencia entre el enrutamiento unicast y multicast es que este último envía copias de los paquetes multicast por varias interfaces intencionalmente. Este comportamiento aumenta la probabilidad de que múltiples copias de un paquete multicast puedan alcanzar a un receptor. Por ejemplo, en algunas topologías de redes redundantes en donde existen múltiples rutas hacia un receptor, se pueden presentar paquetes duplicados antes de que el protocolo de enrutamiento multicast converja y elimine la ruta redundante.

### 1.2.3 Congestión en la red

En TCP el mecanismo de ventanas ajusta automáticamente la velocidad de la transferencia de datos y por lo tanto provee un mecanismo para la disminución de la congestión. Dado que multicast no utiliza TCP, no existe ningún método para prevenir la congestión de un enlace saturado o de otros recursos críticos de un enrutador. Además, no existe un método para prevenir que un host se una a un grupo multicast que esté enviando información a una velocidad que exceda el ancho de banda total disponible en la red a la que pertenece el host.

## 1.3 APLICACIONES MULTICAST

Es común para la gente pensar que multicast sobre IP y video-conferencia es casi lo mismo; no obstante la primera aplicación a utilizar en una red habilitada para multicast es generalmente la video-conferencia, el video es sólo una de las muchas aplicaciones multicast que le pueden dar un valor agregado al modelo de negocios de una compañía.

A continuación se expondrá las diferentes aplicaciones de la tecnología Multicast.

### 1.3.1 Conferencia Multimedia

Sobre el ambiente Unix se crearon algunas herramientas para conferencia multimedia que se pueden utilizar sobre el Mbone. Estas herramientas (algunas de las cuales han sido portadas recientemente a sistemas Windows) permiten

una conferencia de sólo audio ó de audio y video de muchos a muchos. Además de las herramientas de audio y video, se desarrolló una herramienta para Unix que permite a los usuarios compartir una pizarra electrónica común.
Mucha gente realiza conferencias de audio y video porque el video es una nueva forma de comunicarse dentro de una red pero cuando se hace evidente el consumo de ancho de banda y de los recursos de máquina, en particular cuando todos los participantes de la video-conferencia son emisores, es común que los usuarios decidan establecer conferencias de sólo audio.

Si una conferencia de sólo audio se complementa con una aplicación de información compartida (como la pizarra electrónica) que le permite a los miembros de la conferencia compartir información gráfica, el resultado es una potente forma de conferencia multimedia que no consume mucho ancho de banda.

### 1.3.2 Multidifusión de información en tiempo real

La entrega de información en tiempo real a un grupo numeroso de hosts es otra área importante de aplicación multicast. Un buen ejemplo es la entrega de información crítica en el sector financiero; muchas firmas están investigando el uso de multicast sobre IP para distribuir información, de importancia significativa, en tiempo real y utilizarlo como otro servicio del negocio.

Asignando diferentes categorías financieras (bonos, transporte, farmacéutica, etc.) a diferentes grupos multicast, los comerciantes pueden utilizar sus estaciones de trabajo para recibir sólo la información financiera de tiempo real en la cual están interesados.

### 1.3.3 Juegos y simulaciones

Multicast sobre IP es muy utilizada en aplicaciones de juegos y simulaciones en red, aunque muchas de estas aplicaciones hacen uso de la difusión unicast para transportar información. Generalmente una aplicación de juego o simulación tiene que aprender de los otros participantes a través de una configuración manual o a través de un mecanismo especial de notificación de participantes. Cuando se da esta notificación cada PC hace una conexión unicast IP a cada uno de los otros PCs en el juego o simulación, lo que origina un problema del orden de $N^2$ , es decir, que se requieren $N^2$ interconexiones unicast entre las N

estaciones por lo que no es escalable a un gran número de participantes. El límite superior de esta clase de juegos depende en gran medida de la potencia de las estaciones de trabajo utilizadas, por lo general está entre 5 y 10 participantes.

Otro método que se utiliza frecuentemente es un servidor de juegos central al cual todos los participantes se conectan a través de una conexión unicast. Esto pone toda la carga de la distribución de la información del juego en el servidor; dependiendo de la potencia del servidor esta solución puede escalar generalmente a unos 100 participantes.
Multicast es una opción para extender este tipo de juegos a un gran número de participantes; las estaciones de trabajo simplemente se unen a un grupo multicast y empiezan a enviar y recibir información del juego. Se puede ampliar este concepto dividiendo la información de juego en varios flujos de información, cada uno asociado a un grupo multicast, de manera que cada participante se una al flujo de información que necesita en determinado momento, disminuyendo la cantidad de información que debe procesar la máquina.

## 1.4 DIRECCIONAMIENTO MULTICAST A NIVEL IP

A diferencia de las direcciones IP unicast que identifican a un único host, las direcciones IP multicast identifican a un grupo arbitrario de hosts que se han unido a un grupo y desean recibir tráfico enviado a este grupo.

### 1.4.1 Direcciones IP clase D

Las direcciones multicast han sido asignadas por la IANA (*Internet Assigned Number Authority*) al antiguo espacio de direcciones clase D. Las direcciones en este espacio se denotan con un prefijo binario de ***1110*** en los 4 primeros bits del primer byte; estas direcciones multicast cubren el rango 224.0.0.0 - 239.255.255.255.

### 1.4.2 Direcciones multicast asignadas

La IANA controla la asignación de direcciones IP multicast debido a que es un recurso limitado y no asignará bloques de direcciones o direcciones individuales a aplicaciones multicast sin una justificación técnica para hacerlo, en vez de esto,

la IANA asigna direcciones multicast a protocolos de red específicos, lo que implica que el resto de direcciones se deben asignar dinámicamente de alguna forma para su uso en la Internet.

Actualmente el método más utilizado para asignar direcciones multicast dinámicamente es el programa SDR (*Session Directory*). La técnica utilizada por el SDR para evitar colisiones de direcciones multicast no fue diseñada para escalar a miles de grupos multicast, en este momento la IETF planea modificar el SDR y lograr mayor escalabilidad o definir una nueva forma de asignación dinámica de direcciones multicast.

## 1.5 DIRECCIONES MULTICAST DE ENLACE LOCAL

La IANA ha reservado el rango de direcciones desde la 224.0.0.0 hasta la 224.0.0.255 para los protocolos de red en el segmento de red local. Los paquetes que contengan este rango de direcciones son locales en su ámbito y no son reenviados por los enrutadores (sin importar su valor de TTL).

La tabla 1.1 es una lista parcial de las direcciones multicast reservadas por la IANA; en la tabla se muestra la dirección de ámbito local reservada, la función del protocolo de red al cual fue asignado y la persona que solicitó la dirección o el RFC asociado con el protocolo.

| Dirección | Uso | Referencia |
|---|---|---|
| 224.0.0.1 | Todos los hosts | [RFC 1112, JBP] |
| 224.0.0.2 | Todos los enrutadores multicast | [JBP] |
| 224.0.0.3 | No asignado | [JBP] |
| 224.0.0.4 | Enrutadores DVMRP | [RFC 1075, JBP] |
| 224.0.0.5 | Enrutadores OSPF | [RFC 1583, JXM1] |
| 224.0.0.6 | Enrutadores designados OSPF | [RFC 1582, JXM1] |
| 224.0.0.7 | Enrutadores ST | [RFC 1190, KS14] |
| 224.0.0.8 | Hosts ST | [RFC 1190, KS14] |
| 224.0.0.9 | Enrutadores RIP2 | [RFC 1723, SM11] |
| 224.0.0.10 | Enrutadores IGRP | [Farinacci] |
| 224.0.0.11 | Agentes móviles | [Bill Simpson] |
| 224.0.0.12 | Agentes Server/Relay DHCP | [RFC 1884] |
| 224.0.0.13 | Enrutadores PIM | [Farinacci] |

| | | |
|---|---|---|
| 224.0.0.14 | Encapsulamiento RSVP | [Braden] |
| 224.0.0.15 | Enrutadores CBT | [Ballardie] |
| 224.0.0.16 | SBM designados | [Baker] |
| 224.0.0.17 | Todos los SBMS | [Baker] |
| 224.0.0.18 | VRRP | [Hinden] |
| 224.0.0.22 | IGMPv3 (Membership Report) | [RFC 3376] |

**Tabla 1.1 Direcciones multicast de ámbito local**

## 1.6 OTRAS DIRECCIONES RESERVADAS

La IANA periódicamente asigna direcciones multicast únicas a peticiones para protocolos de red, aplicaciones de red en el rango de direcciones 224.0.1.xxxx. Los mrouters reenviarán estas direcciones multicast, al contrario de las direcciones en el rango 224.0.0.xxxx.

| Dirección | Uso | Referencia |
|---|---|---|
| 224.0.1.0 | Grupo de administradores VMTP | [RFC 1045, DRC3] |
| 224.0.1.1 | Protocolo NTP-Network Time | [RFC 1119, DLM1] |
| 224.0.1.2 | SGI-Dogfight | [AXC] |
| 224.0.1.3 | Rwhod | [SXD] |
| 224.0.1.6 | Servidor de servicio de nombrado-NSS | [BXS2] |
| 224.0.1.8 | SUN NIS+Servicio de información | [CXM3] |
| 224.0.1.20 | Cualquier experimento privado | [JBP] |
| 224.0.1.21 | DVMRP sobre MOSPF | [John Moy] |
| 224.0.1.32 | Mtrace | [Casner] |
| 224.0.1.33 | RSVP-encap-1 | [Braden] |
| 224.0.1.34 | RSVP-encap-2 | [Braden] |
| 224.0.1.39 | Cisco-RP-Announce | [Farinacci] |
| 224.0.1.40 | Cisco-RP-Discovery | [Farinacci] |
| 224.0.1.52 | MBone-VCR-Directory | [Holfelder] |
| 224.0.1.78 | Tibco Multicast1 | [Shum] |
| 224.0.1.79 | Tibco Multicast2 | [Shum] |
| 224.0.1.80 - 224.0.1.255 | No asignado | [JBP] |

**Tabla 1.2 Otras direcciones multicast reservadas**

### 1.6.1 Administración de ámbitos con direcciones Multicast

Además del rango de direcciones multicast descritas previamente, la IANA ha reservado el rango desde 239.0.0.0 hasta 239.255.255.255 como direcciones de ámbito administrativo para utilizarlas en dominios multicast privados. Estas direcciones son similares en su naturaleza a los rangos unicast IP reservados, como el 10.0.0.0/8, definido en el RFC 1918 que no son asignados por la IANA a ningún grupo o protocolo; el uso de las direcciones de ámbito administrativo también ayuda a conservar el límite del espacio de direcciones multicast ya que pueden ser reutilizadas en diferentes regiones de la red. En realidad, los administradores de la red tienen que configurar sus mrouters para asegurar que el tráfico multicast dentro de este rango de direcciones no cruce desde ni hacia fuera de su dominio multicast.

## 1.7 DIRECCIONES MULTICAST A NIVEL MAC

La especificación original Ethernet esta prevista para la transmisión de paquetes broadcast y/o multicast. Como se muestra en la figura 1.2 el bit 0 del primer byte en la dirección MAC indica si la dirección de destino es una dirección


broadcast/multicast o una dirección unicast.

**Figura 1.2 Formato de Dirección MAC IEEE802.3**

Si este bit es puesto a uno, entonces la trama MAC se envía a un grupo arbitrario de hosts o a todos los hosts en la red (cuando la dirección de destino MAC es una dirección Broadcast, 0xFFFF.FFFF.FFFF). Multicast a nivel 2 hace uso de esta capacidad de transmitir paquetes multicast IP a un grupo de hosts en un segmento de LAN. A continuación se explica como se mapean las direcciones multicast de nivel 3 en direcciones multicast a nivel 2.

En el caso de Ethernet, las tramas IP multicast utilizan las direcciones de nivel MAC que comienzan con el prefijo de 24 bits de 0x0100.5Exx.xxxx; pero solamente la mitad de estas direcciones MAC están disponibles para el uso de IP multicast. Esto deja 23 bits del espacio de dirección MAC para mapear las direcciones IP multicast de nivel 3 dentro de las direcciones MAC de nivel 2. Como todas las direcciones multicast de nivel 3 tienen los 4 primeros bits de los 32 bits en total fijados a 0x1110, esto deja 28 bits de información de dirección IP multicast. Estos 28 bits tienen que mapearse dentro de sólo 23 bits de la dirección MAC disponible. La figura 1.3 ilustra esta situación.

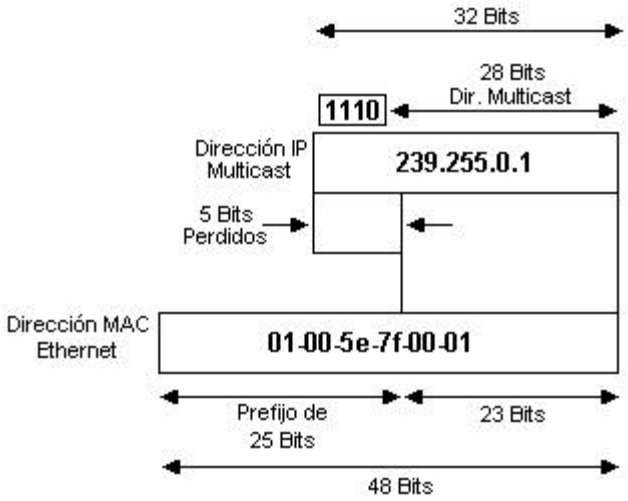


**Figura 1.3 Mapeado de direcciones IP multicast en direcciones MAC multicast**

Debido a que no todos los 28 bits de la dirección IP multicast de nivel 3 pueden ser mapeados dentro de los 23 bits del espacio de dirección MAC, 5 bits de información de dirección se pierden en el proceso de mapeo. Esto resulta en una ambigüedad de direcciones de $2^5$ o 32:1 cuando una dirección multicast nivel 3 se mapea en una dirección IEEE MAC de nivel 2. Esto significa que cada dirección MAC multicast puede representar 32 direcciones IP multicast.

Esta ambigüedad de direcciones de 32:1 puede causar algunos problemas, como por ejemplo, un host que quiere recibir el grupo multicast 224.1.1.1 programará el hardware de la NIC para que interrumpa la CPU cuando se recibe una trama con una dirección MAC multicast de destino de 0x0100.5E00.0101. Esta dirección MAC multicast es la misma utilizada por otros 31 grupos multicast. Si alguno de

estos otros 31 grupos está activo en la LAN local, la CPU del host recibirá interrupciones cada vez que se recibe una trama de alguno de estos otros grupos. La CPU tendrá que examinar la porción IP de cada trama recibida para determinar si hace parte del grupo deseado; esto tiene un impacto en la potencia de la CPU disponible del host si la carga del tráfico de los demás grupos es suficientemente alta.

Además de tener un posible impacto negativo en la CPU del host, esta ambigüedad puede causar problemas cuando se trata de limitar la inundación de tráfico multicast en las LAN basadas en switches nivel 2.

## 1.8 ÁRBOLES DE DISTRIBUCIÓN MULTICAST

Para entender el modelo de IP multicast es necesario un buen conocimiento de cómo trabajan los árboles de distribución multicast. En el modelo unicast, el tráfico se encamina a través de la red a lo largo de una ruta simple desde la fuente hacia el host destino; en el modelo multicast, sin embargo, la fuente envía el tráfico a un grupo arbitrario de hosts que están representados por una dirección de grupo multicast.

Para entregar el tráfico multicast a todos los receptores, los árboles de distribución multicast se utilizan para describir la ruta que toma el tráfico a través de la red. Existen 2 tipos básicos de árboles de distribución multicast que son: ***árboles de distribución de fuente*** y ***árboles compartidos*** (que de ahora en adelante se nombrarán como *source trees* y *shared trees*, respectivamente [*]).

### 1.8.1 Source Trees

Esta forma simplificada de árboles de distribución multicast está basada en una fuente, cuya raíz es la fuente del tráfico multicast en el cual sus ramas forman un árbol expansivo a través de la red hacia los receptores. Debido a que este árbol utiliza la ruta mas corta a través de la red, también se conoce como *árbol de ruta de acceso mas corta* (*SPT*, *Shortest Path Tree*). La figura 1.4 muestra un ejemplo de un SPT para el grupo 224.1.1.1 cuya raíz, el host A, es la fuente y se conectan 2 receptores, host B y host C.

[*] Se utilizarán los términos en inglés para una mayor comprensión de su significado.

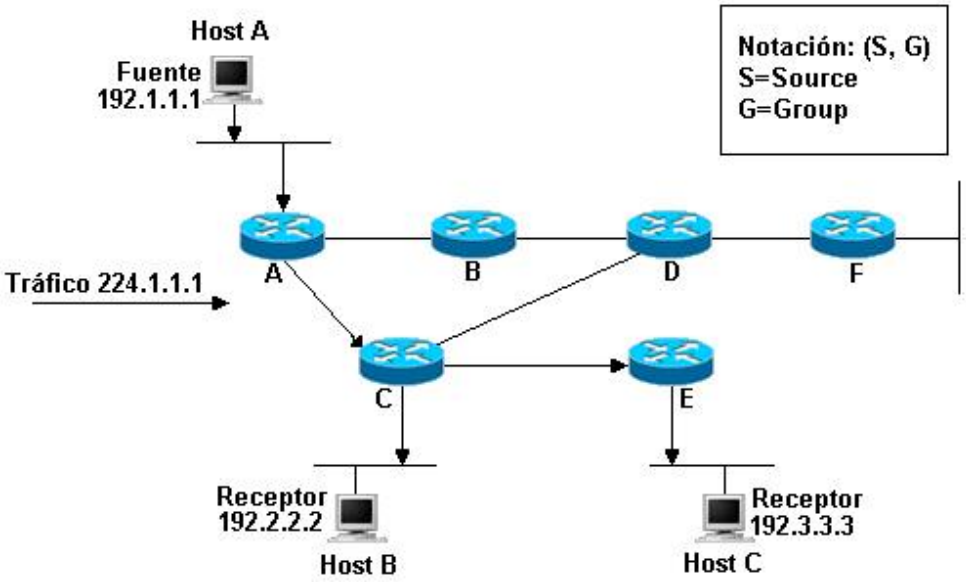


**Figura 1.4 SPT del host A**

En esta clase de árboles se utiliza una notación especial: (S, G) donde S es la dirección IP de la fuente y G es la dirección IP del grupo multicast. Utilizando esta notación, el SPT del ejemplo de la figura 1.4 se escribirá como (192.1.1.1, 224.1.1.1).

Esta notación implica que existe un SPT separado para cada fuente que envíe tráfico a un grupo. Por lo tanto, si el host B también envía tráfico al grupo 224.1.1.1 y los hosts A y C son receptores, entonces existirá un SPT (S, G) separado.

### 1.8.2 Shared Trees

Contrario a los SPTs que tienen su raíz en la fuente, los árboles compartidos utilizan una raíz única y común ubicada en algún punto escogido sobre la red. Dependiendo del protocolo de enrutamiento multicast, esta raíz se conoce comúnmente como ***Rendezvous Point*** (RP) o ***Core***, por esto los árboles compartidos también se conocen como: RPT (*RP-trees*) o CBT (*core-based trees*).

La figura 1.5 muestra un árbol compartido para el grupo 224.2.2.2 con la raíz ubicada en el enrutador D. Cuando se utiliza un shared tree, la fuente tiene que mandar su tráfico a la raíz para que desde ahí pueda alcanzar a todos los receptores.

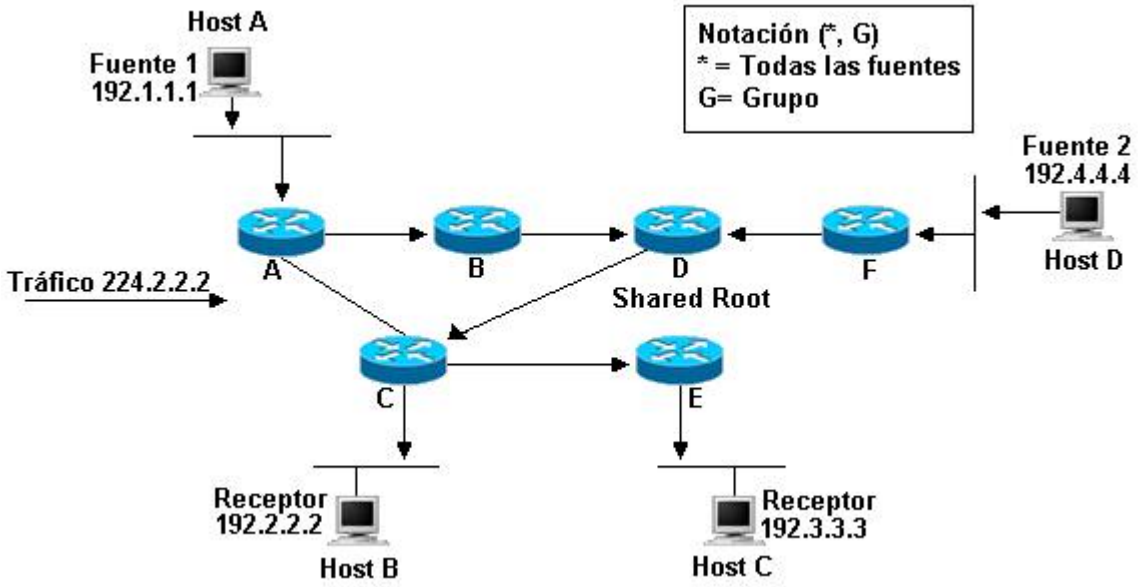


**Figura 1.5 Árbol compartido para el grupo 224.2.2.2 con D como raíz del árbol**

En este ejemplo, el tráfico del grupo multicast enviado por los hosts fuentes A y D viaja hacia la raíz (enrutador D) y a continuación baja a través del árbol compartido hacia los hosts receptores B y C. Como todas las fuentes en el grupo multicast utilizan un árbol compartido, la notación (*, G) representa el árbol. En este caso, el * significa todas las fuentes y la G representa el grupo multicast. Por lo tanto, el shared tree que se muestra en la figura 1.5 se denota como (*, 224.2.2.2). Hay dos clases de árboles compartidos: unidireccionales y bidireccionales.

### *Árboles compartidos bidireccionales*

En este caso el tráfico multicast puede fluir hacia arriba o hacia abajo del árbol para alcanzar a todos los receptores. La figura 1.6 muestra un ejemplo de un árbol compartido bidireccional.

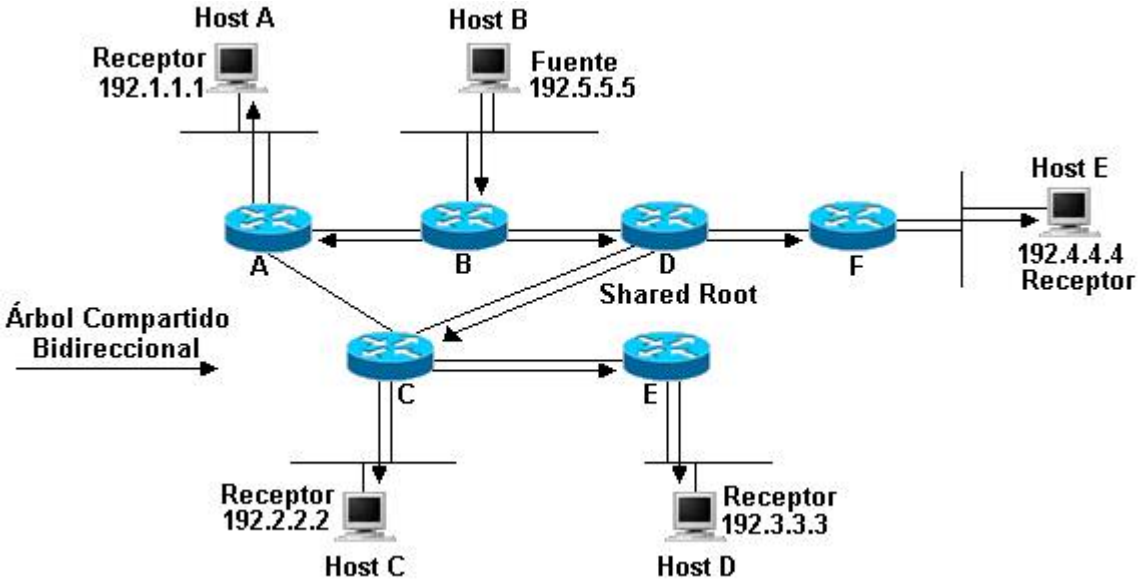


**Figura 1.6 Shared tree bidireccional**

El tráfico multicast desde el host B se envía hacia arriba en dirección a la raíz del árbol a través de su primer enrutador y hacia abajo del árbol en dirección a los otros receptores.

### *Árboles compartidos unidireccionales*

Estos árboles sólo permiten que el tráfico multicast fluya hacia abajo del árbol compartido, desde la raíz hasta los receptores. Por lo tanto, las fuentes del tráfico multicast tienen que utilizar otra forma para lograr que el tráfico llegue primero a la raíz y después se envíe hacia abajo del árbol.

Un método que se puede utilizar es tener a la raíz del árbol compartido unida a un SPT que tiene como raíz a la fuente de tráfico multicast, de tal manera que por el SPT se envíe dicho tráfico hacia la raíz del árbol compartido y de ahí sea distribuido a los receptores. El Protocolo Multicast Independiente (PIM) utiliza este método para hacer llegar el tráfico multicast de la fuente a la raíz o RP.

Otro método para enviar tráfico multicast de la fuente a la raíz es hacer que el primer enrutador envíe el tráfico a través de unicast directamente hacia la raíz. El protocolo de enrutamiento multicast CBT utiliza este método cuando un host fuente quiere enviar tráfico a un grupo como se muestra en la figura 1.7.

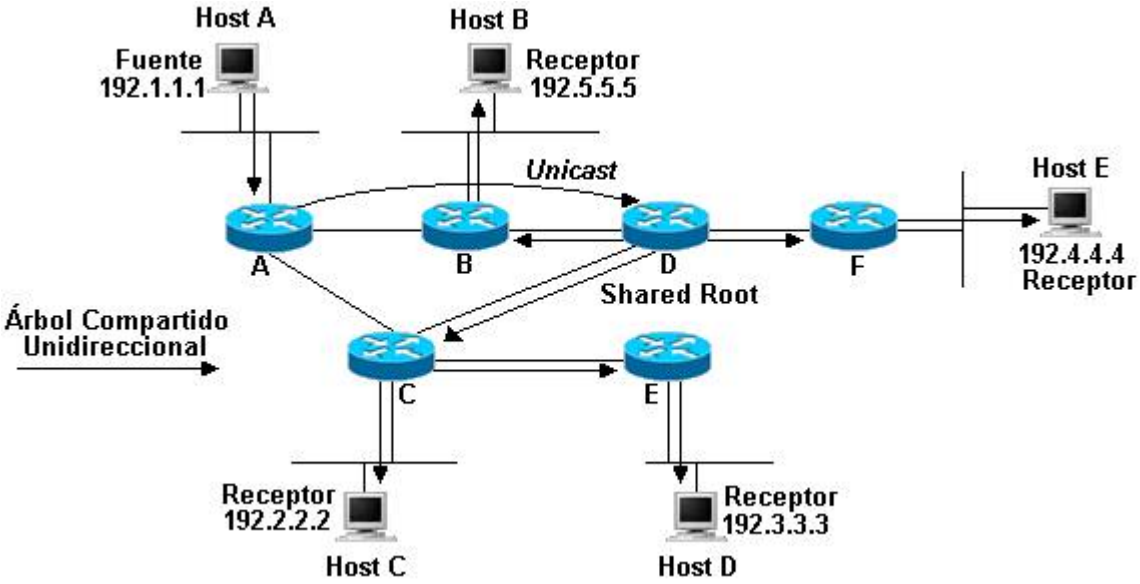


**Figura 1.7 Shared tree unidireccional**

En este ejemplo host A es la fuente y el host B es un receptor. El enrutador A encapsula el tráfico multicast recibido de la fuente y lo envía por medio de unicast directamente hacia la raíz a través de un túnel IP–IP. La raíz desencapsula el paquete y lo envía hacia abajo del árbol compartido.

## 1.9 ENVÍO MULTICAST

En el modelo unicast los enrutadores envían el tráfico a través de la red a lo largo de una ruta simple desde la fuente hacia el host destino en donde la dirección IP aparece en el campo de dirección de destino del paquete IP. Cada enrutador en el trayecto toma una decisión de encaminamiento unicast, utilizando la dirección IP de destino del paquete, buscando la dirección de destino en la tabla de enrutamiento unicast y luego reenviando el paquete hacia el siguiente salto, a través de la interfaz indicada, en dirección al destino.

En el modelo multicast, la fuente envía tráfico hacia un grupo arbitrario de hosts representados por una dirección de grupo multicast en el campo de dirección de destino del paquete IP. En contraste con el modelo unicast, el mrouter no puede basar su decisión de encaminamiento en la dirección de destino del paquete; generalmente estos enrutadores tienen que reenviar el paquete multicast por

múltiples interfaces para alcanzar a todos los receptores. Este requerimiento hace que el proceso de reenvío multicast sea más complejo que el que se utiliza en el encaminamiento unicast.

***Reenvío por la ruta inversa (RPF, Reverse Path Forwarding)***

Implícitamente todos los protocolos de enrutamiento multicast hacen uso de alguna forma de RPF o chequeo de la interfaz entrante como el mecanismo primario para determinar si se debe enviar o eliminar un paquete multicast entrante. Cuando un paquete multicast llega al enrutador, este lleva a cabo un chequeo RPF sobre el paquete. Si el chequeo RPF es exitoso el paquete se reenvía, de otro modo se elimina.

Para el tráfico que fluye hacia abajo del SPT, el mecanismo de chequeo RPF funciona de la siguiente manera:

1. El enrutador examina la dirección de la fuente del paquete multicast entrante para determinar si el paquete llegó a través de la interfaz correcta, es decir, la interfaz que está en la ruta inversa hacia la fuente.
2. Si el paquete llega por la interfaz correcta, el chequeo RPF es exitoso y el paquete es reenviado.
3. Si el chequeo RPF falla, se descarta el paquete.

La manera como un mrouter determina la interfaz de la ruta inversa hacia la fuente depende del protocolo de enrutamiento en uso. En algunos casos, el protocolo de enrutamiento multicast mantiene una tabla de enrutamiento multicast separada y la utiliza para su chequeo RPF, como por ejemplo el protocolo de enrutamiento multicast DVMRP. En otros casos, el protocolo multicast utiliza la tabla de enrutamiento unicast existente para determinar la interfaz de la ruta inversa hacia la fuente, los protocolos de enrutamiento multicast PIM y CBT son un ejemplo de esto.

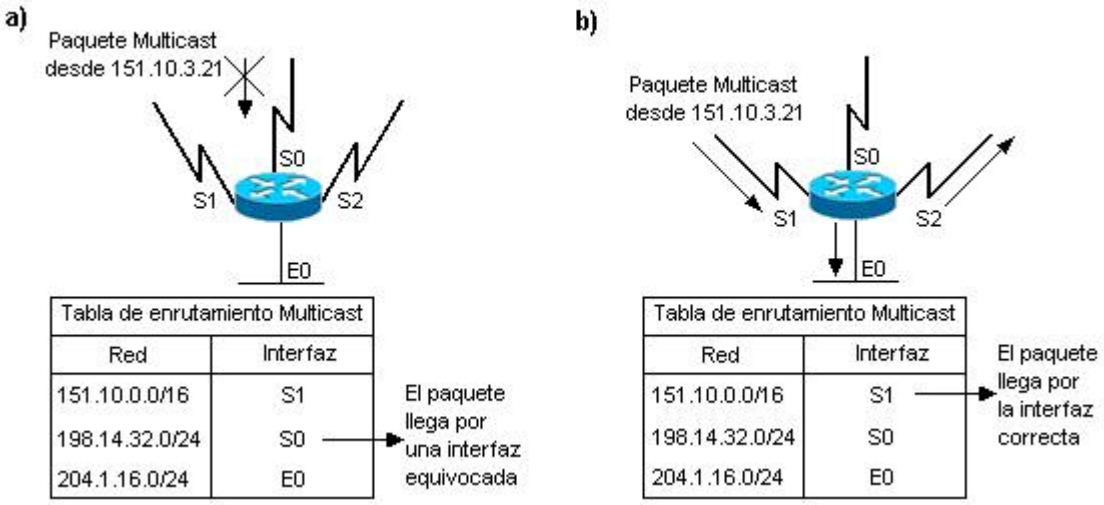


**Figura 1.8 Chequeo RPF**

En la parte (a) de la figura 1.8 un paquete multicast desde la fuente 151.10.3.21 se recibe sobre la interfaz S0. Un chequeo de la tabla de enrutamiento multicast muestra que la interfaz de la ruta inversa hacia la fuente es S1 y no S0. Por lo tanto, el chequeo RPF falla y el paquete es descartado.

En la parte (b) el paquete llega por la interfaz S1 y el cheque RPF es exitoso y por lo tanto el paquete es reenviado hacia todas las interfaces en la lista de interfaces salientes. No todas las interfaces del enrutador tienen que estar en la lista de interfaces de salida.

Teniendo claros los conceptos básicos de multicast se describe en el siguiente capítulo la forma de interacción entre los enrutadores multicast y los hosts, a través del Protocolo de Gestión de Grupos de Internet.

# 2 PROTOCOLO DE GESTIÓN DE GRUPOS DE INTERNET IGMP

El Protocolo de Gestión de Grupos de Internet nació del *Host Membership Protocol* desarrollado en la tesis de doctorado del Dr. Steve Deering; La primera versión, IGMPv1, fue definida en el RFC 1112; IGMPv2 fue ratificado en noviembre de 1997 como estándar por la IETF y está documentado en el RFC 2236; la versión más reciente es IGMPv3 definida en el RFC 3376 de octubre de 2002.

Los mensajes IGMP son utilizados básicamente por los hosts multicast para indicar a los enrutadores multicast cuándo desean unirse a un grupo multicast específico y empezar a recibir tráfico de ese grupo. Los hosts también deben indicar a los mrouters que desean abandonar un grupo multicast y, por lo tanto, no están interesados en seguir recibiendo el trafico multicast dirigido a ese grupo.

Utilizando la información obtenida a través de IGMP, los enrutadores mantienen una lista del número de grupos multicast que tienen receptores en cada interfaz. Un grupo multicast está activo en una interfaz si por lo menos un host sobre esa interfaz ha señalado su interés de recibir el tráfico de ese grupo multicast por medio de IGMP.

A continuación se hace referencia al funcionamiento del protocolo IGMP, tanto la versión 1 como la 2, al igual que se hará una pequeña descripción de IGMPv3 y los cambios mas importantes respecto a la versión anterior.

## 2.1 IGMP VERSIÓN 1

Aunque en este momento el estándar de IGMP es la versión 2 muchos sistemas operativos todavía utilizan la versión 1, como Windows 95 (a menos que se baje la versión actualizada de la DLL Winsock de Microsoft) y muchas versiones antiguas de Unix. IGMPv2 fue desarrollada haciendo extensiones y mejoras de su predecesora, por lo tanto se necesita una buena comprensión de su funcionamiento para la construcción del marco teórico de este proyecto.

### 2.1.1 Formato de los mensajes IGMPv1

Los mensajes IGMP se transmiten dentro de datagramas IP denotados por el número ***2*** en el campo *IP protocol*. Los mensajes IGMP tienen el campo IP time-to-live (TTL) fijado a 1 indicando un ámbito local de manera que no son reenviados por los enrutadores. La figura 2.1 muestra el formato del mensaje IGMPv1.

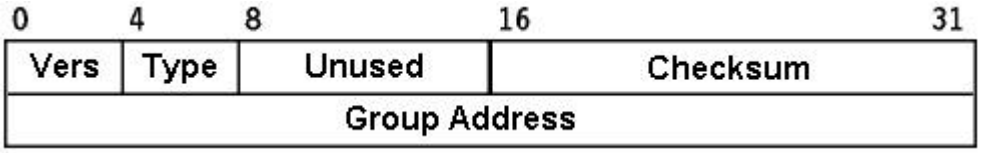


**Figura 2.1 Formato del Mensaje IGMPv1**

***Campo Versión.*** Este campo contiene la identificación de la versión de IGMP, en este caso esta campo será 1. En IGMPv2 este campo ha sido eliminado. El antecesor de IGMPv1 llevaba este campo en 0 y fue especificado en el RFC 988.

***Campo Type.*** En IGMPv1 se utilizaban sólo dos tipos de mensajes entre los hosts y los enrutadores.

- Membership Query
- Membership Report

***Campo Checksum.*** Este es un campo de 16 bits de longitud, es el complemento a uno de la suma del complemento a uno del mensaje IGMP. El campo Checksum se fija a cero para llevar a cabo este cálculo.

***Campo Group Address.*** Contiene la dirección del grupo multicast cuando se envía un Membership Report. Este campo es cero en un Membership Query y debe ser ignorado por los hosts.

### 2.1.2 Proceso de Petición en IGMPv1

IGMP utiliza un modelo de *Pregunta-Respuesta* (Query-Response) que permite al enrutador multicast averiguar cuales grupos multicast están activos (esto es, que tengan uno o más hosts interesados en un grupo multicast) en la subred local.

A manera de ejemplo se tiene la red de la figura 2.2 en donde los hosts H1 y H2 están interesados en recibir tráfico multicast del grupo 224.1.1.1. Además el host H3 quiere recibir tráfico multicast del grupo 224.2.2.2. El enrutador A es el IGMP Querier y es el responsable de atender las peticiones. Existe un enrutador B non-querier que solamente escucha y graba las respuestas de los hosts.

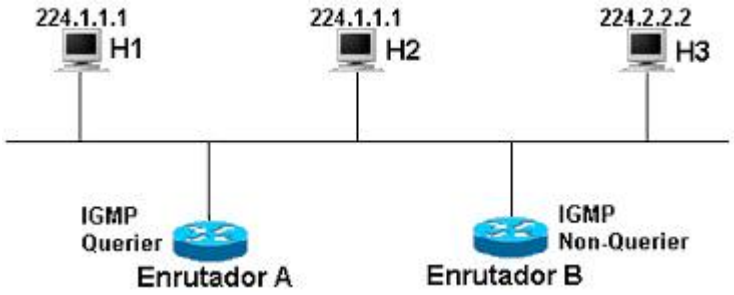


**Figura 2.2 IGMPv1**

El mecanismo de *Pregunta-Respuesta* funciona de la siguiente manera:

- El enrutador A, IGMP Querier, multidifunde periódicamente (cada 60 segundos por defecto) un Membership Query IGMPv1 al grupo multicast de todos los hosts (224.0.0.1) en la subred local. Todos los hosts escuchan este Query siempre y cuando estén habilitados para multicast.
- Todos los hosts reciben el IGMPv1 Membership Query y un solo host (H2 por ejemplo) responde primero difundiendo un IGMPv1 Membership Report al grupo multicast 224.1.1.1 del cual es miembro. Este mensaje informa a los enrutadores en la subred que un host está interesado en recibir tráfico multicast del grupo 224.1.1.1.
- Como el host H1 está escuchando al grupo multicast 224.1.1.1, escucha el IGMPv1 Membership Report multidifundido por el host H2. Por lo tanto el host

H1 suprime su respuesta al grupo 224.1.1.1 porque el host H2 ya ha informado a los enrutadores dentro de la subred que existe por lo menos un host interesado en recibir tráfico multicast para ese grupo. Este mecanismo de supresión de reportes ayuda a repartir la carga de tráfico de la red local.

- El host H3 también ha recibido el IGMPv1 Membership Query y responde con un IGMPv1 Membership Report al grupo multicast 224.2.2.2 del cual es miembro. Este reporte informa a los enrutadores de la subred que un host esta interesado en recibir tráfico multicast del grupo 224.2.2.2.

Como resultado del intercambio Pregunta-Respuesta, el enrutador A sabe que hay receptores activos para los grupos multicast 224.1.1.1 y 224.2.2.2 en la subred local. Adicionalmente el enrutador B ha estado escuchando pasivamente durante todo el proceso y por lo tanto conoce la misma información.

### 2.1.3 Mecanismo de supresión de Reporte en IGMPv1

El mecanismo de supresión de reporte IGMP ayuda a reducir la carga del tráfico IGMP en una subred al mínimo necesario para mantener el estado de los grupos multicast. A continuación se describe este mecanismo con mayor detalle.

- Cuando un host recibe un IGMP Membership Query, el host empieza un temporizador de reporte en cuenta regresiva para cada grupo multicast al cual se ha unido. Cada temporizador de reporte se inicializa en un valor aleatorio entre cero y el intervalo máximo de respuesta (por defecto es de 10 segundos).
- Si el temporizador de reporte expira, el host multidifunde un IGMP Membership Report al grupo multicast activo asociado al temporizador de reporte.
- Si el host escucha que otro host envía un IGMP Membership Report, este cancela su reporte asociado con el Membership Report recibido, suprimiendo el envió del Membership Report para ese grupo.

### 2.1.4 Enrutador Querier en IGMPv1

Si hay múltiples enrutadores multicast en una subred, se presentará un desperdicio de ancho de banda si se tiene a más de uno de ellos haciendo consultas. En dicho caso, se hace esencial un enrutador Querier IGMPv1, responsable de enviar todas las preguntas (Queries) IGMPv1 en una subred. El

RFC 1112 no especifica como se elige el IGMPv1 Querier. Por el contrario, el protocolo IGMPv1 delega esta función al protocolo de enrutamiento multicast IP (PIM, DVMRP, etc.) para resolver este conflicto eligiendo un *Enrutador Designado-DR* para la subred.

### 2.1.5 Proceso de registro en IGMPv1

Para reducir la latencia en el proceso de registro (particularmente cuando un host es el primero en registrarse o afiliarse a un grupo multicast en una subred), no es necesario esperar al siguiente Membership Query antes de poder enviar un Membership Report para unirse a un grupo multicast. Cuando un host desea unirse a un grupo multicast, el host inmediatamente enviará uno o mas Membership Report *no solicitados* para el grupo multicast al cual desea unirse.

Es importante notar que un host sólo utiliza el protocolo IGMP para informar al enrutador multicast local su deseo de recibir o detener el tráfico multicast, de un grupo en particular, que esta dirigido a él.

También se debe tener en cuenta que si un host desea empezar a enviar tráfico multicast no es estrictamente necesario que el host se una a un grupo multicast.

### 2.1.6 Proceso de abandono en IGMPv1

IGMPv1 tiene un método muy simple para que los hosts abandonen un grupo multicast, ellos simplemente lo dejan. En IGMPv1 no hay un mensaje de abandono de grupo para notificar a los enrutadores en la subred que un host no desea recibir mas tráfico de un grupo específico. El host simplemente deja de procesar tráfico del grupo multicast y no envía más respuestas cuando recibe los Queries IGMP enviados por el enrutador, es decir, que la única manera en que los enrutadores IGMPv1 saben que ya no hay más receptores activos para un grupo particular es cuando dejan de recibir Membership Reports. Para facilitar este proceso, los enrutadores IGMPv1 asocian un temporizador en cuenta regresiva a un grupo IGMP dentro de una subred. Cuando se recibe un Membership Report para el grupo dentro de la subred, se restablece el temporizador. Para los enrutadores IGMPv1, este temporizador es generalmente 3 veces el intervalo de pregunta ó 3 minutos. Este tiempo de expiración significa que el enrutador debe continuar reenviando tráfico multicast dentro de la subred por 3 minutos más, una vez todos los hosts hayan dejado el grupo multicast.

## 2.2 IGMP VERSIÓN 2

En noviembre de 1997, IGMPv2 fue ratificado como estándar por la IETF en el RFC 2236, el cual sirve como actualización del RFC 1112. IGMPv2 fue desarrollado principalmente para manejar algunos inconvenientes de IGMPv1.

Los mensajes Query y Membership Report en IGMPv2 son idénticos a los mensajes de IGMPv1 con 2 excepciones.

La primera diferencia es que el mensaje Query de IGMPv2 se divide en dos categorías: ***General Queries***, que cumplen la misma función que los mensajes Query antiguos de IGMPv1; y los ***Group-Specific Queries***, que son preguntas directas a un grupo específico.

La segunda diferencia es que los Membership Report de IGMPv1 e IGMPv2 tienen diferentes códigos en el campo Type.

El proceso de Pregunta-Respuesta en IGMPv2 es el mismo que en la versión anterior.

En la versión 2 de IGMP se agregaron una algunas mejoras:

***Proceso de elección del Querier.*** Provee, a los enrutadores IGMPv2, la capacidad de elegir el enrutador Querier sin tener que delegar esta función a algún protocolo de enrutamiento multicast.

***Campo de Tiempo máximo de Respuesta.*** Un nuevo campo en los mensajes Query le permite al enrutador Querier especificar el tiempo de Pregunta-Respuesta máximo. Este campo permite la depuración del proceso de Pregunta-Respuesta para controlar el tiempo de respuesta y para ajustar la latencia de abandono a un grupo.

***Mensaje Group-Specific Query.*** Permite al enrutador Querier llevar a cabo la operación de consulta sobre un grupo específico en vez de todos los grupos.

***Mensajes Leave Group.*** Proveen, a los hosts, un método para notificar a los enrutadores en la red el deseo de abandonar un grupo.
Los mensajes Group-Specific Query y Leave Group permiten la reducción de la latencia de minutos a segundos.

### 2.2.1 Formato del Mensaje IGMPv2

El formato de los mensajes IGMPv2 se muestra en la figura 2.3. El campo *Type* y el campo *Versión* (del formato IGMPv1) se fusionan en el formato del mensaje IGMPv2 y ahora ocupan un octeto completo. Los valores asignados a los diferentes tipos de mensajes se escogieron de tal manera que fueran compatibles con IGMPv1.

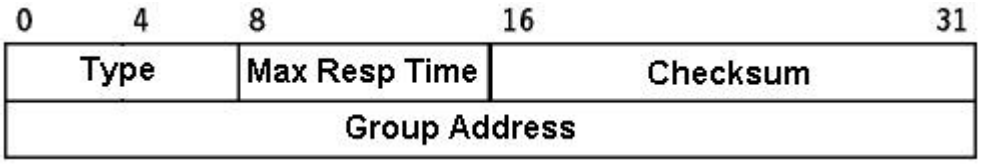


**Figura 2.3 Formato del mensaje IGMPv2**

***Campo Type.*** En la versión 2 de IGMP, los siguientes 4 tipos de mensajes se utilizan entre los hosts y los enrutadores.

- Membership Query (código=x11), existen dos subtipos de mensajes Membership Query.
  - General Query, utilizado para determinar cuales grupos multicast están activos de la misma manera que lo hacia IGMPv1. Un General Query se denota por ceros en el campo Group Address.
  - Group-Specific Query, se emplea para determinar si un grupo multicast específico todavía tiene miembros activos. Este mensaje contiene la dirección del grupo solicitado.
- Versión 1 del Membership Report (código=0x12), este tipo de mensaje facilita la compatibilidad con IGMPv1.
- Version 2 del Membership Report (código=0x16).
- Leave Group (código=0x17).

***Campo Maximum Response Time.*** Este campo no se utilizaba en los mensajes IGMPv1. Se emplea únicamente en los mensajes Membership Query y especifica

el tiempo máximo, en unidades de décimas de segundo, que un host debe esperar para responder a un mensaje Query, el valor por defecto es de 100 (10 segundos). Los hosts utilizan este Tiempo de Respuesta Máxima como el límite superior del valor aleatorio de su temporizador de reporte de grupo, utilizado por el mecanismo de supresión de reporte.

***Campo Checksum.*** Este es un campo de 16 bits de longitud, es el complemento a uno de la suma del complemento a uno del mensaje IGMP. El campo Checksum se fija a cero para llevar a cabo este cálculo.

***Campo Group Address.*** Cuando se envía un General Query, el campo Group Address se fija a cero para diferenciarlo del Group-Specific Query, el cual contiene el grupo multicast del grupo solicitado. Cuando se envía un mensaje Membership Report o Leave Group, este campo se fija con la dirección del grupo multicast objetivo.

### 2.2.2 Depuración del proceso Pregunta-Respuesta

El campo Maximum Response Time se adicionó al mensaje Query de IGMPv2 para que un host IGMPv2 no use un valor configurado estáticamente para el intervalo del proceso Pregunta-Respuesta. Este campo permite que el tiempo de respuesta sea configurado por el IGMP Querier, que indica a todos los hosts el límite superior para el retardo de sus respuestas al mensaje Query, colocando dicho valor en el campo Maximum Response Time.

### 2.2.3 Mensajes Leave Group en IGMPv2

IGMPv2 define un nuevo tipo de mensaje, Leave Group, utilizado por los hosts cuando desean abandonar un grupo. Cuando un host abandona un grupo multicast y fue el último host en responder a un mensaje Query con un Membership Report, el host debe enviar un mensaje de Leave Group al grupo multicast de todos los enrutadores (224.0.0.2). El RFC dice: "Un host siempre ***puede*** enviar un mensaje Leave Group cuando abandona un grupo"; la palabra utilizada en el RFC es ***puede*** (may) y no ***tiene*** (must), es decir, que es opcional enviar o no el mensaje Leave Group, lo cual indica que la falta de mensajes Leave Group en IGMPv2 puede derivar en los mismos problemas de latencia experimentados en IGMPv1. La mayoría de sistemas, tanto Windows como Unix,

que utilizan IGMPv2 creen más conveniente utilizar siempre el proceso de abandono de un grupo enviando los mensajes Leave Group.

### 2.2.4 Mensajes Group-Specific Query en IGMPv2

Este es otro de los nuevos mensajes de IGMPv2, que es enviado por el enrutador IGMP Querier y cuyo propósito es consultar por un grupo específico en vez de hacerlo para todos los grupos. En un Group-Specific Query, el campo Group Address contiene el grupo objetivo; los hosts que reciben este mensaje responden de la misma manera como lo hacen a un General Query.

Otra diferencia entre el Group-Specific Query y el General Query es que el primero utiliza un valor más pequeño en el campo Maximum Response Time, ayudando a disminuir la latencia en el proceso de abandono de grupo; por defecto el valor de este campo es 1 segundo.

### 2.2.5 El proceso de abandono en IGMPv2

La adición de los mensajes Leave Group y Group-Specific Query en IGMPv2, complementada con el campo Maximum Response Time, permite a IGMPv2 reducir la latencia debida al abandono de un grupo a sólo unos pocos segundos.

Retomando como ejemplo la red de la figura 2.2 donde se tienen los host H1 y H2 miembros del grupo 224.1.1.1; también existe un enrutador A que esta designado como IGMP Querier; el host H2 desea dejar el grupo y para ello se deben seguir los siguientes pasos:

- El host H2 multidifunde un mensaje IGMPv2 Leave Group al grupo multicast de todos los enrutadores (224.0.0.2) para informar que está abandonando el grupo.
- El enrutador A escucha el mensaje Leave Group del host H2. Sin embargo, como los enrutadores mantienen una lista sólo de los grupos miembros que están activos en una subred (no de los host individuales que son miembros) el enrutador A envía un Group-Specific Query para determinar si todavía hay miembros presentes para el grupo 224.1.1.1 en esa interfaz. Cabe anotar que el Group-Specific Query se multidifunde hacia el grupo objetivo, por lo tanto sólo los que son miembros de este grupo responderán.

- El host H1 todavía es miembro del grupo 224.1.1.1 y, por lo tanto, escucha el Group-Specific Query y responde a esta pregunta con un IGMPv2 Membership Report para informar al enrutador dentro de la subred que un miembro de este grupo todavía está presente. Si más de un host en esta subred hubiera estado presente, el mecanismo de supresión de reporte también sería utilizado para evitar un desperdicio de ancho de banda.
- Ahora el host H1 es el último miembro del grupo 224.1.1.1 en esta subred, si H1 quiere abandonar el grupo, multidifunde un mensaje Leave Group al grupo multicast de todos los enrutadores para informales su deseo de abandonar el grupo.
- De nuevo, el enrutador A escucha el mensaje Leave Group (esta vez del host H1) y envía un Group-Specific Query para determinar si existen miembros restantes para el grupo 224.1.1.1.
- Ahora no existen miembros restantes para el grupo 224.1.1.1 en la subred, por lo tanto, ningún host responde al Group-Specific Query. Al no obtener respuesta, el enrutador A espera un tiempo llamado *Last Member Query Interval* (por defecto es de 1 segundo) y envía otro Group-Specific Query, sin obtener respuesta (por defecto el enrutador intenta dos veces). En este momento el enrutador A se da cuenta de que no hay más miembros del grupo y suspende el reenvío de tráfico dentro de la subred.

### 2.2.6 Proceso de elección del Querier

IGMPv2 no depende del protocolo de enrutamiento multicast para la elección del enrutador Querier, en cambio utiliza la dirección IP del mensaje General Query para elegir el enrutador IGMP Querier a través del siguiente procedimiento:

- Cuando los enrutadores IGMPv2 arrancan, cada uno multidifunde un mensaje IGMPv2 General Query hacia el grupo de todos los sistemas multicast (224.0.0.1) con su dirección de interfaz en el campo *Source IP Address* del mensaje.
- Cuando un enrutador IGMPv2 recibe el mensaje General Query, el enrutador compara la dirección IP de origen del mensaje con su propia dirección de interfaz. El enrutador con la dirección IP más pequeña en la subred es elegido como el IGMP Querier.
- Todos los enrutadores non-Querier empiezan un temporizador que se restablece cada vez que se recibe un mensaje General Query desde el IGMP

Querier. La duración por defecto de este temporizador es dos veces el *Query Interval* (250 segundos). Si el temporizador expira, se asume que el enrutador Querier se ha caído y se corre nuevamente el proceso de elección para designar un nuevo IGMP Querier.

## 2.3 IGMP VERSIÓN 3

Un problema muy común, particularmente en conferencias multimedia, es que cuando un host se une a un grupo multicast, está pidiendo que el tráfico de todas las fuentes de ese grupo sea entregado hacia la subred del host. Si bien varias fuentes pueden estar enviando tráfico hacia el grupo, lo más común es que un host desee recibir tráfico sólo de un fuente. La versión 3 de IGMP extiende el mecanismo de *registro/abandono* permitiendo que los registros y los abandonos sean emitidos para un par fuente/grupo específico a través de los mensajes (S, G) *Join/Leave* de IGMPv3.

La principal característica de IGMPv3 es la capacidad de filtraje de fuente, pero también hay otros cambios, algunos de los más importantes se mencionan a continuación:

- Los estados de grupo se definen como Grupo + Lista de fuentes, no simplemente como Grupo en IGMPv2.
- IGMPv3 define operaciones que permiten la interoperabilidad con IGMPv1 e IGMPv2.
- Los enrutadores Querier incluyen el Intervalo de Petición (Query Interval) en los mensajes Query para permitir la sincronización de esta variable en los enrutadores Non-Querier.
- El tiempo máximo de respuesta en los mensajes Query tiene un rango exponencial, pasa de 25.5 segundos a cerca de 53 minutos, que se emplea en enlaces con un gran número de sistemas.
- Se definen secciones de información adicionales para futuras extensiones.
- Los mensajes de Reporte se envían a la dirección 224.0.0.22, para ayudar a los Switches nivel 2 en el "snooping".

- Los hosts ya no utilizan el mecanismo de supresión y de esta manera se simplifican las implementaciones y se permite el rastreo de afiliación.

Una vez establecida la comunicación entre los hosts y los enrutadores multicast, hay protocolos encargados del transporte de la información acordada entre ellos, como se describe en el siguiente capítulo.

# 3 PROTOCOLOS A NIVEL DE APLICACIÓN EN IP MULTICAST

Este capítulo expondrá los protocolos empleados por las aplicaciones de conferencia multimedia. El Protocolo de Tiempo Real (RTP, Real Time Protocol) y el Protocolo de Control de Tiempo Real (RTCP, Real Time Control Protocol) se utilizan para encapsular los flujos de información de video y audio de las conferencias multimedia y para monitorear la entrega de la información a las estaciones finales. Más adelante se describirá el Protocolo de Anuncio de Sesión (SAP, Session Announcement Protocol) y el Protocolo de Descripción de Sesión (SDP, Session Description Protocol). Las aplicaciones de Directorio de Sesión utilizan estos protocolos para anunciar y aprender acerca de la existencia de sesiones de conferencias multimedia en la red.

RTP consiste de dos componentes**:**

- El componente RTP, que lleva la información de tiempo real.
- EL protocolo de control RTP (RTCP), que provee información de los participantes de la sesión y monitorea la entrega de información empleando algunas medidas de calidad del servicio simples como pérdida de paquetes y jitter.

## 3.1 PROTOCOLO DE TIEMPO REAL – RTP

El RTP está definido en el RFC 1889 y permite a las aplicaciones transmitir varios tipos de flujos de tiempo real como audio, video u otro tipo de información con características de tiempo real. Generalmente UDP transporta al protocolo RTP y se puede utilizar tanto en flujo de datos unicast como multicast. El protocolo también provee identificación del tipo de carga, numeración de secuencia, y

marca de tiempo (timestamping), así como mecanismos para monitorear la entrega de información.

RTP por sí mismo no provee ningún mecanismo de entrega confiable y normalmente relega esta función a las capas inferiores. Aunque RTP utiliza UDP e IP, este depende de la aplicación para manejar los problemas de datagramas perdidos y orden de entrega. Estas condiciones se pueden detectar con el uso del campo del número de secuencia en el encabezado RTP.

Las aplicaciones multicast multimedia generalmente asignan una dirección de grupo multicast y dos puertos; un puerto para el flujo de información RTP y el otro para el control del flujo a través de RTCP. En la mayoría de los casos el puerto utilizado para el control es numéricamente mayor en 1 que el puerto de datos.

La señal de información entrante se muestrea en intervalos de tiempo pequeños y fijos por la aplicación para después ser codificada y encapsulada dentro de los paquetes RTP. Los encabezados de los paquetes RTP contienen el número de secuencia y la marca de tiempo al igual que el esquema de codificación utilizado.

Cuando la aplicación recibe paquetes RTP, el número de secuencia y la marca de tiempo sirven para recobrar información de la fuente y determinar cuantos paquetes se han perdido. Las muestras de información de la señal codificada en el paquete RTP se colocan luego en orden, en un buffer de salida, basado en el número de secuencia y la marca de tiempo de manera que se puedan decodificar y reproducir en el destino.

La congestión en la red puede derivar en la llegada de paquetes en tiempos variables, resultando en la reproducción distorsionada de la información. Utilizando un buffer grande de salida y luego retardando la reproducción de la información antes de que el buffer esté casi lleno, se puede disminuir el jitter. La desventaja de utilizar un buffer grande de salida es que se introduce un retardo en el flujo de información. El retardo no es un inconveniente en las comunicaciones de un solo sentido pero se puede volver un problema en aplicaciones interactivas.

## 3.2 PROTOCOLO DE CONTROL DE TIEMPO REAL – RTCP

Ya que es útil saber quien está participando en la conferencia y la calidad en la recepción, las aplicaciones multidifunden periódicamente un Reporte de Receptor (RR, Receiver Report) en un paquete RTCP en el puerto de control. Estos reportes de receptor contienen el nombre del usuario e información del número de paquetes perdidos y el jitter para cada fuente en la conferencia. Las fuentes pueden utilizar esta información para determinar si sus transmisiones están siendo recibidas por cada receptor y en algunos casos cambiar a otro método de codificación para tratar de mejorar la recepción.

Las fuentes también multidifunden periódicamente Reportes de Fuente (SR, Source Report) en los paquetes RTCP hacia el mismo puerto de control. Estos reportes de fuente contienen la misma información que los reportes del receptor pero además incluyen una sección de 20 bytes de información de fuente que contienen las marcas de tiempo, bytes enviados y paquetes enviados al puerto de datos. Los miembros del grupo pueden utilizar esta información para calcular el Tiempo de Viaje Completo (RTT, Round Trip Time) y otras estadísticas sobre el flujo de tráfico.

Las aplicaciones basadas en RTP utilizan el RTCP para transmitir periódicamente información de control de sesión a todos los participantes de la conferencia para cumplir las siguientes funciones:

1. Proveer retroalimentación en la calidad de recepción de la información y en muchos casos, modificar los esquemas de codificación para mejorar la calidad de recepción total. Aplicaciones externas pueden utilizar esta información para diagnosticar problemas de entrega y determinar áreas de la red que sufren de una baja calidad en la recepción.
2. Identificar cada fuente de la capa de transporte en la conferencia mediante un nombre canónico (CNAME). El CNAME puede ser utilizado para asociar varios flujos de información de un participante como parte de una sola sesión multimedia. Esto puede ser importante si se trata de sincronizar flujos de información de audio y video.

3. Transmitir paquetes RTCP de tal manera que se pueda determinar el número de participantes; es necesario para que todos los participantes puedan cumplir las funciones 1 y 2 y para que la velocidad de transmisión de los datos de control RTCP se pueda ajustar al porcentaje adecuado de ancho de banda total de la sesión.
4. Distribuir información (nombre de usuario, ubicación, etc.) que identifican a los participantes en una sesión de una manera más amigable para el usuario. Esta información normalmente se despliega en la interfaz de usuario de la aplicación.

Si se utiliza el modelo RTP sobre IP Multicast, las funciones 1, 2 y 3 son obligatorias para permitir que la aplicación sea escalable a un gran número de participantes. El hecho de que muchas de las aplicaciones multimedia multicast mas populares usen este modelo tiene la siguiente implicación en el diseño de redes multicast:

*Cada una de las estaciones finales en una sesión multicast multimedia basada en RTP es una fuente de tráfico multicast.*

La información adicional generada por estos receptores se debe considerar cuando se hace el diseño de una red multicast porque algunos protocolos multicast no son escalables cuando se presenta un gran número de fuentes.

## 3.3 PROTOCOLO DE ANUNCIO DE SESIÓN – SAP

SAP es un protocolo de anunciación para las sesiones de conferencia multicast y fue desarrollado por el grupo de trabajo Multiparty Multimedia Session Control (MMUSIC) de la IETF y está definido en el RFC 2974.

Los clientes SAP anuncian periódicamente sesiones de conferencia multidifundiendo paquetes SAP que contienen información de sesión a una dirección y puerto multicast bien conocidos. La información de sesión dentro del paquete utiliza el SDP (Session Description Protocol). Cuando se desea privacidad la información SDP se puede encriptar para evitar que sea leída sin autorización.

La dirección y el puerto multicast bien conocidos para los anuncios SAP dependen del mecanismo de ámbito multicast utilizado en el cliente SAP. El ámbito de una sesión multicast se basa en el valor del TTL de la sesión o en el rango de direcciones administrativas que caen dentro del rango de direcciones 239.0.0.0 hasta 239.255.255.255.

Si se utilizan los anuncios basados en el ámbito TTL, la dirección bien conocida es la 224.2.127.254 y el puerto UDP es el 9875. Los anuncios de sesión basados en el ámbito TTL siempre se multidifunden con el mismo TTL con el cual se difunde la sesión.

Debido a la complejidad del descubrimiento de zonas de ámbito administrativo, los anuncios basados en ámbitos administrativos son poco utilizados.

## 3.4 PROTOCOLO DE DESCRIPCIÓN DE SESIÓN – SDP

Este protocolo se emplea para codificar la información de la sesión actual. Fue desarrollado por el grupo de trabajo MMUSIC de la IETF y está definido en el RFC 2327. SDP no es realmente un protocolo de transporte de la misma manera como lo es el SAP, en realidad SDP es la especificación de un formato basado en texto ASCII que utiliza un número de líneas de texto para describir una sesión, cada línea de texto tiene el siguiente formato: `type = value`

A continuación se presenta una tabla con los descriptores a nivel de sesión de este protocolo.

| Tipo | Descripción |
|---|---|
| v = | Versión de protocolo |
| o = | Propietario e identificador de sesión |
| s = | Nombre de la sesión |
| i = | Información de la sesión |
| u = | URL de descripción |
| e = | Dirección electrónica |
| p = | Número telefónico |
| c = | Información de conexión |
| b = | Información de ancho de banda |

| | |
|---|---|
| t = | Tiempo activo de la sesión |
| r = | Número de repeticiones |
| z = | Ajustes a la zona horaria |
| k = | Clave de encriptación |
| a = | Cero o más líneas de atributos |

**Tabla 3.1 Tipos de descriptores de sesión SDP**

En un sesión multicast pueden haber varios flujos de información correspondientes a varias aplicaciones, para cada uno de los cuales se define líneas de atributo a nivel de medio de comunicación. El descriptor "a" a nivel de sesión identifica cero o mas descriptores a nivel de medio de comunicación que corresponden a los flujos de información pertenecientes a la sesión. La tabla 3.2 muestra estos descriptores.

| Tipo | Descripción |
|---|---|
| m = | Nombre del medio de comunicación y dirección de transporte |
| c = | Información de conexión |
| b = | Información de ancho de banda |
| k = | Clave de encriptación |
| a = | Cero o más líneas de atributos |

**Tabla 3.2 Tipos de descriptores de medios de SDP**

Después de estudiar la tecnología multicast en los niveles de enlace de datos, de red y de transporte, se abre paso a un área fundamental y extensa como son los protocolos de enrutamiento multicast.

# 4 PROTOCOLOS DE ENRUTAMIENTO MULTICAST

En la actualidad los protocolos de enrutamiento multicast se pueden dividir en 3 categorías:

- Protocolos de modo denso como DVMRP y PIM-DM;
- Protocolos de modo disperso como PIM-SM y CBT;
- Protocolos de estado de enlace como MOSPF.

Algunos protocolos como PIM son capaces de operar en modo denso y en modo disperso dependiendo de cómo esté configurado el enrutador .

Los protocolos de modo denso utilizan solamente SPTs para entregar tráfico multicast (S, G) utilizando el principio de inundación que asume que cada subred tiene por lo menos un receptor de tráfico multicast (S, G) y por lo tanto el tráfico tiene que ser entregado a todos los puntos de la red.

Los protocolos de modo disperso utilizan los árboles compartidos y ocasionalmente SPTs para distribuir el tráfico multicast en la red. En vez de utilizar el modelo de inundación los enrutadores esperan solicitudes de los receptores para adicionar las ramas downstream al árbol compartido.

El funcionamiento de los protocolos de estado de enlace es parecido a los protocolos de modo denso por el hecho de que ambos utilizan SPTs para distribuir el tráfico multicast aunque no utilizan el mecanismo de inundación. En vez de esto inundan información de estado de enlace que indica en donde hay miembros de grupo (receptores multicast) en la red. Todos los enrutadores en la red utilizan esta información de afiliación para construir los árboles de ruta de acceso más corta desde cada fuente hacia los receptores del grupo.

## 4.1 PROTOCOLO DE ENRUTAMIENTO MULTICAST DE VECTOR DISTANCIA – DVMRP

DVMRP (Distance Vector Multicast Routing Protocol) fue el primer protocolo de enrutamiento multicast que se utilizó ampliamente. Está basado en el trabajo de Steve Deering y presenta muchas similitudes con RIP adicionándole algunas pequeñas variaciones para soporte multicast.

Algunas de las características importantes de DVMRP son:

- Basado en vector distancia (Similar a RIP).
- Actualización periódica de rutas (cada 60 segundos).
- Red inalcanzable después de 32 saltos (RIP utiliza 16 saltos).
- El Poison Reverse tiene un significado especial.
- Es classless, lo que significa que la actualización de rutas lleva máscaras.

Actualmente no todos los proveedores de enrutadores implementan los mismos protocolos de enrutamiento multicast; sin embargo la mayoría de los proveedores soportan DVMRP en cierto grado de manera que puede ser usado virtualmente entre todos los enrutadores. A continuación se presentará en más detalle su funcionamiento.

### 4.1.1 Descubrimiento de Vecino en DVMRP

Este mecanismo es importante porque los enrutadores DVMRP tienen que mantener una base de datos de las adyacencias con otros enrutadores DVMRP, especialmente cuando el protocolo está operando sobre redes multiacceso (como Ethernet, FDDI, etc.) porque la red puede tener varios enrutadores DVMRP. Para el buen funcionamiento de este protocolo cada enrutador debe conocer exactamente sus enrutadores DVMRP vecinos en cada interfaz.

Periódicamente se envían mensajes de prueba (Probe Message) a la dirección de grupo de todos los enrutadores DVMRP, 224.0.0.4. En la figura 4.1 se muestra el mecanismo de descubrimiento de vecino.

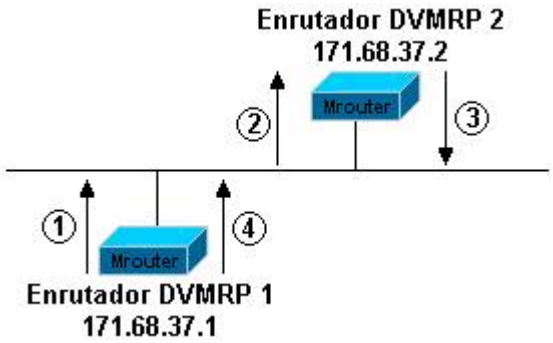


**Figura 4.1 Mecanismo de descubrimiento de vecino**

1. El enrutador 1 envía un mensaje Probe. Como el enrutador 1 no ha escuchado ningún mensaje Probe de otro enrutador, la lista de vecinos del mensaje Probe está vacía.
2. El enrutador 2 escucha el mensaje Probe enviado por el enrutador 1 y añade la dirección IP del enrutador 1 a su lista interna de vecinos DVMRP para esa interfaz.
3. El enrutador 2 envía un mensaje Probe con la dirección del enrutador 1 en la lista de vecinos DVMRP.
4. El enrutador 1 escucha el mensaje Probe del enrutador 2 y añade la dirección IP del enrutador 2 a su lista interna de vecinos DVMRP para esa interfaz. El siguiente mensaje Probe del enrutador 1 será enviado con la dirección del enrutador 2 en la lista de vecinos DVMRP.

Cuando un enrutador DVMRP recibe un mensaje Probe con su propia dirección IP en la lista de vecino DVMRP, el enrutador sabe que se ha establecido una adyacencia mutua entre él y el vecino que envió el mensaje Probe.

La figura 4.2 muestra la porción de una red multicast conformada por dos enrutadores DVMRP, en este caso dos estaciones UNIX corriendo el proceso de enrutamiento Mrouted, conectados a través de Ethernet.

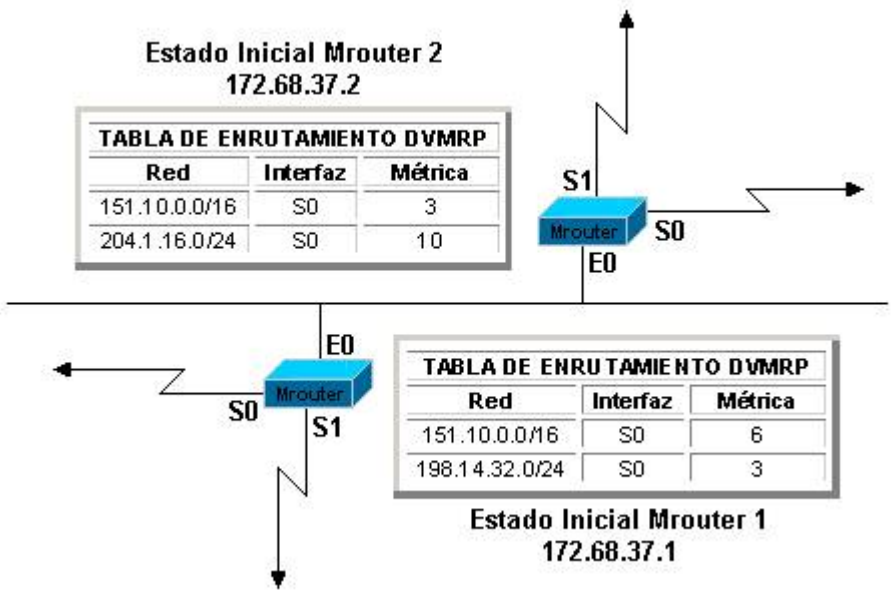


**Figura 4.2 Tablas de enrutamiento DVMRP**

El contenido de las tablas de enrutamiento refleja las rutas DVMRP que han sido aprendidas a través de los enlaces seriales y se muestran antes de que cualquier reporte de enrutamiento haya sido intercambiado. Cabe anotar que ambos enrutadores tienen una entrada para la red 151.10.0.0/16.

El enrutador 2 envía primero su reporte de enrutamiento con dos rutas, el enrutador 1 recibe este reporte y responde adicionando una nueva entrada para la red 204.1.16.0/24 a su tabla de enrutamiento DVMRP. Como el enrutador 2 tiene una mejor métrica hacia la red 151.10.0.0/16, el enrutador 1 actualiza la entrada para esta red con una nueva métrica (4) y el próximo salto a la interfaz Eth0. La tabla de enrutamiento del enrutador 1 queda de la siguiente manera, tabla 4.1.

| TABLA DE ENRUTAMIENTO DVMRP | | | |
|---|---|---|---|
| Red | Interfaz | Métrica | |
| 151.10.0.0/16 | Eth0 | 4 | → Actualizada |
| 198.14.32.0/24 | S0 | 3 | |
| 204.1.16.0/24 | Eth0 | 11 | → Adicionada |

**Tabla 4.1 Tabla de enrutamiento DVMRP actualizada del enrutador 1**

Por último el enrutador 1 responde enviando su propio reporte de enrutamiento al enrutador 2 aplicando el mecanismo de Poison Reverse a las redes que se

alcance a través de esta interfaz, esto es, se le suma a la métrica actual de la ruta el número 32 (ruta inalcanzable), este cambio le informa al enrutador 2 que el enrutador 1 es downstream o que esta debajo del enrutador 2 y que espera recibir tráfico multicast de esa red fuente a través del enrutador 2.

El enrutador 2 recibe este reporte de enrutamiento y actualiza su propia tabla de enrutamiento DVMRP adicionando una nueva entrada para la red 198.14.32.0/24 como se muestra en la tabla 4.2.

| TABLA DE ENRUTAMIENTO DVMRP | | |
|---|---|---|
| Red | Interfaz | Métrica |
| 151.10.0.0/16 | S0 | 3 |
| 204.1.16.0/24 | S0 | 10 |
| 198.14.32.0/24 | Eth0 | 4 |


**Tabla 4.2 Tabla de enrutamiento DVMRP actualizada del enrutador 2**

Cuando el enrutador 2 envíe nuevamente su reporte de enrutamiento, tendrá que utilizar el mecanismo de Poison Reverse para la red 198.14.32.0/24 que aprendió del reporte de enrutamiento del enrutador 1.

### 4.1.2 Árboles Broadcast Truncados DVMRP

DVMRP es un protocolo de modo denso que utiliza SPT (Shortest Path Tree) para enviar el tráfico multicast. Los árboles SPT construidos por los enrutadores DVMRP se basan en los árboles broadcast truncados que utilizan las métricas de las tablas de enrutamiento de los enrutadores DVMRP. Para construir estos árboles broadcast truncados, los enrutadores indican al enrutador upstream (el vecino que tiene la mejor métrica hacia una red fuente) que están debajo de él a través del mecanismo de Poison Reverse.

Estos anuncios Poison Reverse le indican al enrutador upstream que por la interfaz donde fueron recibidos debe enviar el tráfico multicast que proviene de la red envenenada; de tal manera que los enrutadores downstream puedan recibir este tráfico multicast.

Si un enrutador tiene acceso a una red por medio de dos enrutadores vecinos, con la misma métrica, se escogerá como enrutador upstream el que tenga la dirección IP más baja.

Cada subred tendrá su propio árbol broadcast truncado que está definido por las métricas de enrutamiento DVMRP y el mecanismo de Poison Reverse.

Los árboles broadcast truncados se generan en cada enrutador una vez que se ha recibido el primer paquete con la dirección de un grupo multicast y su correspondiente red fuente, esto con la idea de ahorrar recursos en los enrutadores. Se puede considerar este un mecanismo bajo demanda. Una vez el árbol broadcast truncado es creado la red es inundada por el tráfico multicast.

### 4.1.3 Reenvío multicast en DVMRP

A diferencia del enrutamiento unicast (que se interesa hacia donde van los paquetes) el enrutamiento multicast se interesa por la fuente de los paquetes, es decir, de donde vienen. La información en la tabla de enrutamiento DVMRP se utiliza para determinar si un paquete multicast entrante fue recibido por la interfaz correcta; si no, se descarta el paquete para prevenir loops multicast. El reenvío basado en la interfaz entrante se conoce como reenvío por la ruta inversa (RPF, Reverse Path Forwarding) y la prueba para verificar que un paquete llegó por la interfaz correcta se conoce como Chequeo RPF.

Utilizando el ejemplo de la figura 4.2 y la tabla del enrutador 1, un paquete con dirección de origen 151.10.40.3 llega por la interfaz S1, aplicando el concepto de árbol broadcast truncado se conoce que la interfaz adecuada para alcanzar esta red es la interfaz S0 por lo tanto falla el chequeo RPF y se descarta el paquete. Si por el contrario el paquete hubiera llegado por la interfaz S0, por donde debe ser esperado, el chequeo RPF hubiera sido exitoso y el paquete se reenviaría hacia los demás vecinos DVMRP.

### 4.1.4 Podado DVMRP

Como todos los protocolos de modo denso, DVMRP utiliza el mecanismo de inundación y podado (flood-and-prune) para poder entregar inicialmente el tráfico multicast a todos los enrutadores en la red. En el caso de DVMRP el tráfico

es inundado a través del árbol broadcast truncado hacia todos los posibles receptores.

Sin embargo, para preservar los recursos de la red, se necesita podar el tráfico en las ramas donde no haya receptores. Por lo tanto, enrutadores finales en el árbol broadcast truncado (enrutadores que no son upstream para ese árbol broadcast truncado) que no tienen receptores directos envían un mensaje de podado DVMRP hacia arriba del árbol broadcast truncado para detener el flujo de tráfico multicast no deseado y eliminar las ramas no deseadas del árbol. Lo que queda después del podado DVMRP es un SPT para esa fuente específica.

Como DVMRP es un protocolo de inundación y podado, el SPT que resulta del podado DVMRP se revierte a un árbol broadcast truncado tan pronto como los mensajes de podado expiren. Los mensajes de podado tienen vigencia por 2 minutos aproximadamente y luego toda la red se inunda nuevamente con tráfico multicast.

Las tablas de enrutamiento DVMRP no son modificadas por los mensajes de podado DVMRP ya que estas describen el árbol broadcast truncado y no el SPT, en vez de esto, la información recibida a través de los mensajes de podado DVMRP se mantiene por separado en el enrutador en una estructura diferente; esta información es utilizada para modificar el flujo del trafico (S, G) hacia abajo del árbol broadcast truncado.

### 4.1.5 Enganche DVMRP

DVMRP soporta un mecanismo confiable de enganche (Grafting) que une nuevamente ramas de un árbol que han sido podadas. Sin este mecanismo, la latencia producto de unirse a un grupo se incrementaría severamente si se tuviera que esperar a que el estado de podado de los enrutadores upstream expire. Dependiendo del número de enrutadores a lo largo de la rama podada y los valores de expiración utilizados, pueden pasar varios minutos antes de que un host empiece a recibir tráfico multicast. Utilizando el mecanismo de enganche, DVMRP reduce esta latencia a unos pocos milisegundos.

Se dice que el mecanismo de enganche es confiable por el uso de mensajes de confirmación (Graft ACK) para los mensajes de enganche (Graft). Estos mensajes

son enviados por el enrutador upstream en respuesta a un mensaje Graft recibido, evitando su pérdida por congestión lo que podría causar fallas en el proceso de enganche.

### 4.1.6 Escalabilidad DVMRP

A pesar de que DVMRP fue utilizado en el Mbone y en otras redes multicast de intradominio, algunos inconvenientes de escalabilidad limitan su aplicabilidad en ambientes multicast de gran escala.

Como DVMRP utiliza el número de saltos como métrica, con un valor de 32 para redes inalcanzables, es lógico que no se pueda utilizar en redes con un diámetro mayor a 31 saltos sin utilizar tunelaje. Por esta razón DVMRP no se puede utilizar como el protocolo multicast que interconecte a Internet. Además de esto, DVMRP tiene todos los inconvenientes de un protocolo de vector distancia, incluyendo convergencia lenta y un mecanismo de actualización de enrutamiento periódico que no puede manejar el número cada vez más elevado de prefijos activos en la Internet.

## 4.2 PROTOCOLO MULTICAST INDEPENDIENTE DE MODO DENSO – PIM-DM

PIM (Protocol Independent Multicast) como su nombre lo indica es un protocolo de enrutamiento multicast IP independiente. Esto quiere decir que es independiente de cual(es) protocolo(s) de enrutamiento unicast se esté(n) utilizando para llenar la tabla de enrutamiento unicast, incluyendo las rutas estáticas; PIM utiliza esta información para realizar el reenvío multicast. Aunque PIM sea un protocolo de enrutamiento multicast, emplea la tabla de enrutamiento unicast para realizar el chequeo RPF (Reverse Path Forwarding) en vez de mantener una tabla de enrutamiento multicast aparte. Debido a esto PIM no necesita enviar y/o recibir actualizaciones de enrutamiento multicast como en otros protocolos (MOSPF o DVMRP, por ejemplo), disminuyendo significativamente la información de los mensajes PIM en comparación con otros protocolos.

### 4.2.1 Descubrimiento de vecino PIM

Al igual que DVMRP, PIM emplea un mecanismo de descubrimiento de vecino para establecer las adyacencias PIM, es decir, que un mrouter multidifunde un mensaje Hello PIM (cada 30 segundos por defecto, Periodo-Hello) a todos los enrutadores PIM (224.0.0.13) por cada una de sus interfaces habilitadas para multicast.

En PIMv1, estos mensajes Hello algunas veces se nombran como Mensajes Query PIM. Como todos los mensajes PIMv1, estos paquetes se multidifunden a la dirección de grupo multicast de todos los enrutadores (224.0.0.2) y van dentro de paquetes IGMP que tienen códigos especiales. Por otra parte, PIMv2 tiene su propio número de protocolo asignado (103), y por lo tanto no necesita ir dentro de paquetes IGMP.

***Mensajes Hello PIM***

Los mensajes Hello PIM contienen un valor Holdtime, que le dice al receptor cuando expira la adyacencia de vecino asociada con el emisor si no se reciben más mensajes Hello. El valor que se envía en el campo Holdtime generalmente es 3 veces el Periodo-Hello del emisor, o 90 segundos si se toma el intervalo por defecto de 30 segundos.

Mantener el estado de los enrutadores PIM adyacentes es muy importante para construir y mantener los árboles de distribución de fuente (SPT), este aspecto se tratará más adelante.

***Enrutador Designado PIM-DM***

Los mensajes Hello también se utilizan para elegir al Enrutador Designado (DR) en una red multiacceso. Los enrutadores PIM registran, por medio de los mensajes Hello, el enrutador con la dirección IP más baja en la red, que será elegido como el DR.

El mecanismo de DR se aplica sobre todo en redes de modo disperso y tiene poco significado en redes de modo denso. La única excepción a esta regla es cuando se está tranajando con IGMPv1 en una interfaz, en este caso, el DR-PIM también funciona como enrutador Querier debido a que IGMPv1 no tiene un mecanismo de elección del enrutador Querier.

PIMv2 cuenta con una opción de prioridad-DR, muy conveniente en redes multiacceso, esta prioridad la asigna el administrador de la red a cada uno de los enrutadores de manera que el enrutador con mayor prioridad será el DR; si hay varios enrutadores con la misma prioridad entonces el DR será el que tenga la dirección IP más alta.

### 4.2.2 Árboles de distribución de fuente o árboles de ruta de acceso más corta – SPT

PIM-DM sólo utiliza SPTs para distribuir tráfico multicast a los receptores en la red. Estos SPTs se construyen bajo demanda, utilizando el mecanismo de inundación y podado (flood-and-prune) tan pronto como una fuente multicast empieza a transmitir.

A diferencia de DVMRP, que usa su propia tabla de enrutamiento multicast y un mecanismo de Poison Reverse para construir inicialmente un árbol de distribución mínimo, PIM-DM usa su propia información de vecino para construir un SPT similar. En PIM-DM, inicialmente se considera que los vecinos están en el SPT, donde la interfaz de entrada es la que está en la dirección de la fuente (basado en la tabla de enrutamiento unicast) y los demás son vecinos PIM-DM downstream para esta fuente. Esta forma inicial de SPT se conoce como Árbol Broadcast porque un enrutador envía el tráfico multicast a todos los vecinos como Broadcast; en contraste, los enrutadores DVMRP usan Árboles Broadcast Truncados para enviar inicialmente tráfico multicast únicamente a los enrutadores downstream.

Esta es una pequeña modificación de una técnica llamada ***Reverse Path Flooding*** donde el tráfico entrante que pasa el chequeo RPF se inunda a todas las interfaces. La diferencia aquí es que la inundación ocurre solo en las interfaces de salida donde se ha detectado al menos un vecino PIM-DM o que tienen receptores conectados directamente.

La figura 4.3 muestra un ejemplo de inundación inicial de tráfico multicast en una red PIM-DM hacia abajo del árbol Broadcast.

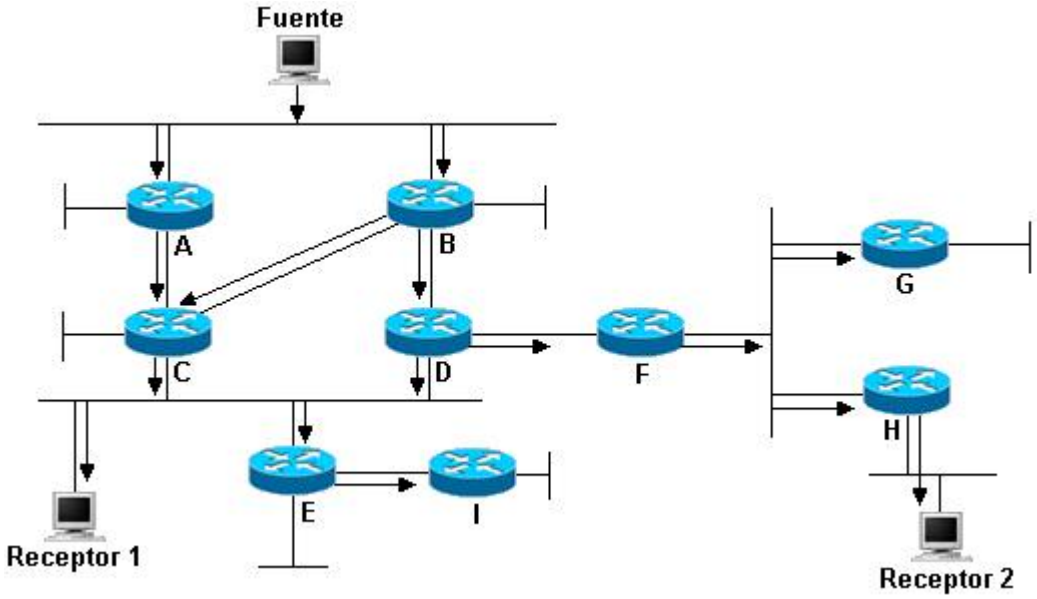


**Figura 4.3 Árbol de distribución PIM-DM (Inundación inicial)**

### 4.2.3 Reenvío multicast PIM-DM

Cuando un enrutador PIM-DM recibe un paquete multicast, se realiza el chequeo RPF al paquete para asegurar que llegó por la interfaz correcta. El enrutador PIM explora la tabla de enrutamiento unicast buscando la mayor concordancia de la dirección IP de la fuente del paquete y utiliza esta información para determinar la interfaz de entrada para el tráfico multicast de esa fuente. Si hay varias entradas para la red fuente en la tabla de enrutamiento (lo que puede ocurrir si hay rutas con igual costo para esa red), el enrutador escoge solo una interfaz, lo que implica una regla muy importante no sólo para PIM sino también para multicast en general: *Un enrutador puede tener sólo una interfaz de entrada para cualquier entrada en su tabla de enrutamiento multicast.*

En PIM (tanto modo denso como modo disperso), cuando existen varias entradas en la tabla de enrutamiento unicast, la entrada con la dirección IP de próximo salto más alta, se usa para el chequeo RPF y por consiguiente se designaría como la interfaz de entrada.

### 4.2.4 Podado PIM-DM

PIM-DM envía mensajes de podado *Prune* bajo las siguientes condiciones:

- Si se presenta tráfico entrante en interfaces no RPF punto a punto.
- En un enrutador Leaf* sin receptores conectados directamente.
- En un enrutador non-Leaf** en un enlace punto a punto que ha recibido un mensaje Prune desde su vecino.
- Cundo un enrutador non-Leaf, en un segmento LAN (sin receptores conectados directamente) recibe un mensaje Prune desde un vecino y el mensaje no ha sido cancelado por otro de sus vecinos.

Tomando la red ejemplo de la figura 4.3, se pueden analizar los casos en los cuales se hace uso de los mensajes Prune.

- Asumiendo que la métrica a la fuente multicast es mejor por medio del enlace entre A y C que entre B y C, el enrutador C envía un mensaje Prune al enrutador B al recibir tráfico desde la fuente por una interfaz no-RPF. De forma que el enrutador B responde al Prune podando su enlace a C.
- Por otra parte, I es un enrutador Leaf (sin receptores conectados directamente), así que debe enviar un mensaje Prune al enrutador E. El enrutador E responde podando su enlace a I; igualmente, el enrutador E debe enviar un mensaje Prune a los enrutadores C y D debido a que no tiene receptores conectados directamente y sus enlaces downstrem han sido podados.
- Sin embargo, debido a que el receptor 1 esta conectado directamente a la misma interfaz que E, los enrutadores C y D ignoran el Prune enviado por E. Mientras la fuente continúe enviando tráfico multicast, el enrutador E continuará enviando mensajes Prune; no obstante, este limita los mensajes Prune, enviándolos aproximadamente cada tres minutos, lo que evita una inundación de mensajes Prune que consumirían ancho de banda en este segmento de red multiacceso.

* Un enrutador Leaf es el último salto en un árbol de distribución
** Un enrutador non-leaf se encuentra en un punto intermedio entre el enrutador de primer salto de la fuente y el último salto

***Anulación de los mensajes Prune – Prune Override***

El mecanismo Prune en redes multiacceso trabaja de diferente forma en PIM y en DVMRP. DVMRP mantiene el rastro de sus vecinos en una interfaz y registra el envío de mensajes Prune desde un vecino; PIM espera recibir Joins desde los vecinos downstream que desean continuar recibiendo tráfico como respuesta a los mensajes Prune enviados por otros vecinos en la misma interfaz.

El caso de los enrutadores G y H de la figura 4.3 es buen ejemplo de la anulación de los mensajes Prune.

- G es un enrutador Leaf sin receptores directamente conectados, así que envía un mensaje Prune al enrutador F.
- Debido a que los mensajes Prune se multidifunden a todos los enrutadores PIM (224.0.0.13), el enrutador H alcanza a oír el mensaje Prune enviado al enrutador E.
- Como H tiene receptores conectados directamente, envía un mensaje Join para anular el Prune enviado por el enrutador G.

***Acumulación de retardos***

Para que el mecanismo Prune funcione adecuadamente, cuando un enrutador PIM recibe un mensaje Prune en una red multiacceso se inicia un Temporizador de Retardo de 3 segundos. Si no se recibe un mensaje Join que anule el Prune y cancele el temporizador, se aceptará el mensaje Prune recibido.

En las redes multiacceso son de gran importancia los retardos Prune, en algunos casos ese retardo se puede acumular cuando fluye tráfico multicast indeseado a través de varias redes multiacceso.

### 4.2.5 Elección del enrutador designado en PIM-DM – Mecanismo Assert

Volviendo a la red de la figura 4.3, y después de los procesos de Prune y de Override mencionados anteriormente, el tráfico de la fuente sólo fluiría a los puntos de la red donde haya receptores, pero se presenta un problema más: el tráfico duplicado que están entregando C y D en el segmento Ethernet donde está conectado el receptor 1.

Para solucionar este problema e interrumpir uno de esos flujos, PIM usa un mecanismo ***Assert*** para elegir el enrutador designado de una fuente multicast particular; este mecanismo se rige por la siguiente regla:

*Si un enrutador recibe un paquete multicast por medio de una interfaz perteneciente a la lista de interfaces de salida asociado con una fuente multicast, envía un Mensaje Assert a la interfaz para resolver cual enrutador continuará enviando ese tráfico.*

Cuando se da inicio al mecanismo Assert en una interfaz, un enrutador PIM envía un mensaje Assert que contiene su métrica a la fuente. Todos los enrutadores en la red examinan la métrica en el mensaje y determinan cual enrutador tiene la mejor métrica hacia la fuente multicast. El enrutador con la mejor métrica continúa enviando el tráfico de la fuente a la red y los demás enrutadores PIM suspenden el envío de tráfico multicast por sus interfaces. Si hay métricas iguales, entonces se escoge el enrutador con la dirección IP más alta.

En la red de la figura 4.3, los enrutadores C y D envían mensajes Assert ya que están recibiendo paquetes multicast de la fuente por medio de una de sus interfaces de salida (para dicha fuente). Si C y D tienen la misma métrica y C tiene la dirección IP más alta, entonces C será quien continúe enviando tráfico multicast de la fuente en el segmento Ethernet y D poda su interfaz.

### 4.2.6 Enganche PIM-DM

PIM-DM tiene la capacidad de enganchar (Graft) nuevamente ramas previamente podadas del árbol de distribución, de manera que el flujo de tráfico multicast se reinicia con un retardo mínimo. La figura 4.4 muestra la red mostrada en la figura 4.3 con un nuevo receptor conectado al enrutador I.

Cuando el enrutador I recibe el mensaje IGMP Membership Report desde el receptor 3, I sabe que debe enviar un mensaje Graft al enrutador E para reiniciar el flujo del tráfico multicast (lo sabe porque aún mantiene el estado multicast, a pesar del Prune, para esa fuente). Cuando el enrutador E recibe el mensaje Graft del enrutador I, responde enviando un mensaje Graft-Ack.

Debido a que el enrutador E ha enviado un Prune previamente para esta fuente (aunque fue ignorado porque el receptor 1 esta conectado directamente a la red), E también envía un mensaje Graft al enrutador C, el cual inmediatamente responde enviando un mensaje Graft-Ack, así el tráfico multicast empieza a fluir a través de E y de I al receptor 3.

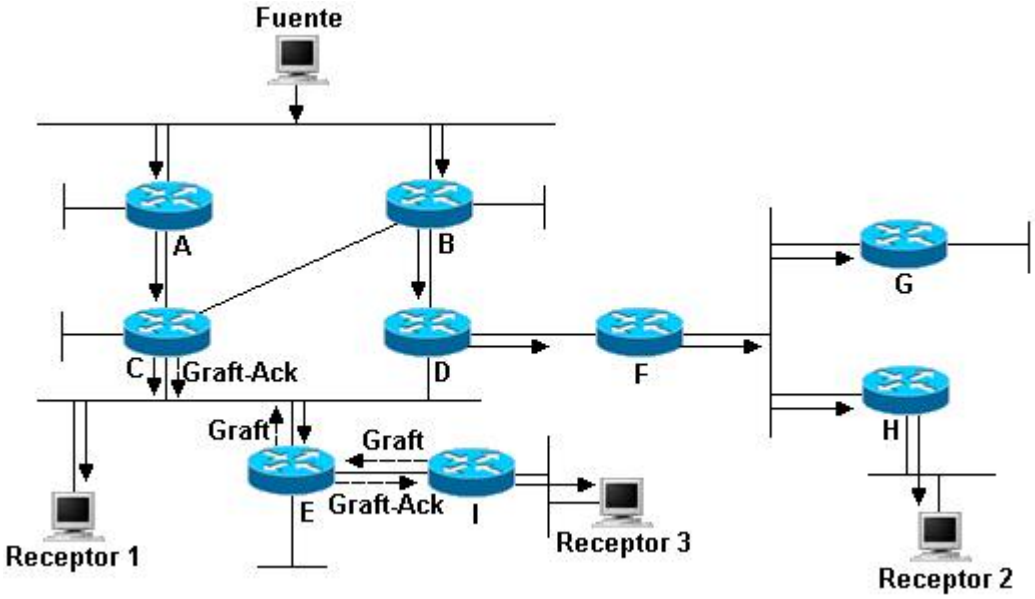


**Figura 4.4 Enganche PIM-DM**

### 4.2.7 Mejora PIM-DM: State-Refresh

La IETF ha estado trabajando en el desarrollo del protocolo PIM-DM y en este momento, la especificación de este protocolo se encuentra en el draft-ietf-pim-dm-new-v2-03.txt de la IETF; este trabajo esta encaminado a tratar de eliminar el comportamiento de inundación-podado de PIM-DM. Esta extensión es la *Actualización de estado* (State-Refresh) cuya función es refrescar los estados de los enrutadores downstream de manera que las ramas podadas del árbol broadcast no expiren. Esta función la cumplen los enrutadores de primer salto enviando periódicamente un mensaje de actualización (S, G) hacia abajo del árbol broadcast mientras la fuente S continúe enviando tráfico al grupo G.

### 4.2.8 Escalabilidad de PIM-DM

Si una red unicast esta bien diseñada y utiliza de forma óptima una asignación jerárquica de direccionamiento IP y agregación de rutas, PIM-DM tiene una mejor escalabilidad que DVMRP. La razón es que PIM usa la tabla de enrutamiento unicast en los enrutadores para hacer el chequeo RPF, a diferencia de DVMRP, no

envía tablas de enrutamiento multicast separadas. Sin embargo, PIM-DM tiene el mismo comportamiento de inundación-podado que DVMRP, y por consiguiente, puede sufrir el flujo periódico de tráfico indeseado a través del dominio de modo denso. La extensión State-Refresh de la especificación PIM previene el flujo de tráfico indeseado.

## 4.3 PROTOCOLO MULTICAST INDEPENDIENTE DE MODO DISPERSO – PIM-SM

El protocolo Multicast Independiente de modo disperso PIM-SM, al igual que PIM-DM usa la tabla de enrutamiento unicast para realizar el chequeo RPF. Es independientemente de los protocolos de enrutamiento unicast con los cuales se esté trabajando en el enrutador; PIM-SM utiliza esta información para realizar el reenvío multicast.

### 4.3.1 Modelo de unión explícita – Explicit join model

El tráfico multicast sólo se envía a partes de la red que específicamente lo requieran, por medio de mensajes de unión PIM (*PIM Joins*), que se envían salto a salto hasta la raíz del árbol (RP en un SPT); cuando no se desea recibir más tráfico multicast se envía un mensaje de Prune a través del árbol hasta el RP. Se puede ver una diferencia básica entre PIM-DM y PIM-SM, en el primero el reenvío del tráfico multicast por parte del enrutador depende del tráfico que llegue como en todos los protocolos flood-and-prune, por otra parte en el modelo de unión explícito de PIM-SM los enrutadores envían tráfico multicast como resultado de los mensajes Join y Prune que reciban.

### 4.3.2 Árboles compartidos PIM-SM

El funcionamiento de PIM-SM se centra en un único árbol unidireccional en el cual el nodo raíz se llama ***Rendezvous Point – RP***.

Los shared trees o RP trees se conocen como RPTs para evitar confusión con los source trees (o SPT, shortest path trees).

Los enrutadores que tienen receptores conectados directamente para un grupo o fuente multicast, se unen a este árbol compartido; el enrutador por sí mismo se poda del árbol compartido cuando ya no necesite recibir tráfico multicast.

PIM-SM utiliza un árbol compartido unidireccional, donde el tráfico sólo fluye hacia abajo del árbol, por esta razón las fuentes de tráfico multicast deben registrarse con el RP para enviar su tráfico multicast por el árbol, pasando primero por el RP. El proceso de registro da inicio a una unión SPT del RP hacia la fuente, siempre y cuando haya receptores activos para el grupo en la red.

***Mensajes de unión en el árbol compartido - Shared tree Joins***

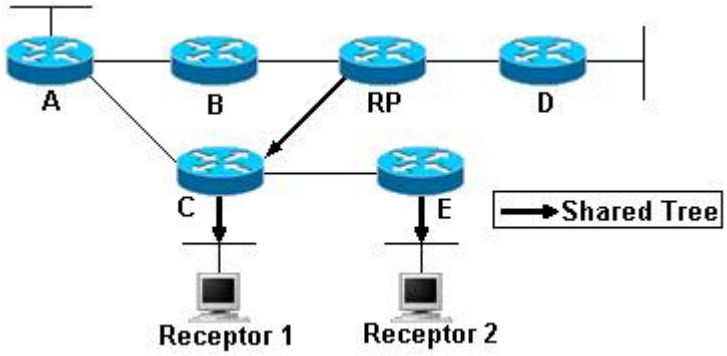


**Figura 4.5 Mensajes Join en el árbol compartido PIM**

En la figura 4.5, el receptor 1 se une a un grupo multicast G enviando un mensaje IGMP membership report. El Enrutador C al cual está conectado el receptor 1 debe crear una entrada (*, G) en su tabla de enrutamiento multicast para dicho grupo multicast. El enrutador C coloca su interfaz Ethernet en la lista de interfaces de salida para el grupo (*, G); el enrutador C también debe enviar un mensaje Join (*, G) hacia el RP para unirse al árbol compartido. (El enrutador C usa su tabla de enrutamiento unicast para determinar la interfaz hacia el RP.)

El RP recibe el Join (*, G) y como inicialmente no tiene ninguna condición previa para el grupo multicast G, crea su entrada (*, G) en la tabla de enrutamiento multicast y adiciona el enlace al enrutador C (que le envío el Join) a la lista de interfaces de salida. Es decir que se ha construido un árbol compartido para el grupo G desde el RP al enrutador C y al receptor 1, de tal manera que cualquier

tráfico, para el grupo G, que llegue al RP puede distribuirse a través del shared tree hacia el receptor 1.

Otro host, el receptor 2 se une al grupo G enviando un mensaje IGMP que escucha el enrutador E (que tiene un enlace al enrutador C); como el enrutador E no tiene ningún estado previo para el grupo multicast G, crea una entrada (*, G) en la tabla de enrutamiento multicast y adiciona la interfaz ethernet a su lista de interfaces de salida. El enrutador E a su vez envía un Join (*, G) hacia el RP para unirse al árbol compartido para el grupo multicast G.

Cuando el enrutador C recibe el Join (*, G) desde el enrutador E, encuentra que ya tiene un estado previo para el grupo G (es decir, que ya está unido al árbol compartido para ese grupo), por lo que simplemente adiciona el enlace al enrutador E a la lista de interfaces de salida para su entrada (*, G). Ahora el árbol irá del RP al enrutador C y luego al receptor 1 y a E, y del enrutador E al receptor 2.

***Mensajes de podado en el árbol compartido - Shared Tree Prunes***

Debido a que PIM-SM usa el modelo de unión explícito para construir los árboles de distribución, también usa los Prunes para desmontar los árboles cuando ya no se necesitan. Para esto se podría sólo detener el envío periódico de Joins que actualizan el árbol de manera que las ramas del árbol expiren, pero no se usarían eficientemente los recursos de la red.

En la figura 4.5, si el receptor 2 deja el grupo multicast G enviando un mensaje IGMP Leave, dado que es el único host unido al grupo en la interfaz ethernet del enrutador E, esta interfaz se remueve de la lista de interfaces de salida en su entrada (*, G) así que la lista queda vacía (NULL), lo que indica que el enrutador E no necesita más tráfico del grupo G; el enrutador E entonces envía un Prune (*, G) hacia el RP para podarse del árbol compartido.

Cuando el enrutador C recibe este Prune, remueve el enlace al enrutador E de su lista de interfaces de salida para la entrada (*, G), pero como el enrutador C tiene al receptor 1 conectado directamente, su lista de interfaces de salida para la entrada (*, G) no es nula y por lo tanto el enrutador C seguirá en el árbol (no envía un Prune al RP).

Cuando se envía un Prune (*, G) en una red multiacceso con varios enrutadores PIM-SM unidos al mismo árbol se usa el mismo mecanismo de anulación de mensajes Prune o Prune Override que en PIM-DM evitando el podado prematuro del árbol compartido.

### 4.3.3 Árboles de ruta de acceso más corta PIM-SM

La ventaja de PIM-SM sobre otros protocolos de modo disperso (como CBT) es que no se limita a recibir tráfico solamente por medio del árbol compartido. Se puede utilizar el mismo mecanismo de unión explícito para unirse al SPT donde la raíz es cualquier fuente particular. Uniéndose al SPT, el tráfico multicast se envía directamente a los receptores sin tener que pasar por el RP, por lo que se reduce la latencia de la red y la posible congestión en el RP, pero también se presenta una desventaja al unirse al SPT debido a que los enrutadores deben crear y mantener entradas (S, G) en sus tablas de enrutamiento multicast a lo largo del SPT (S, G), lo que consume más recursos del enrutador.

Aún así, la cantidad total de información (S, G) que deben almacenar los enrutadores PIM-SM generalmente es mucho menor que la necesaria para protocolos de modo denso; la razón de esto es que con el mecanismo de Flood-and-Prune que usan los protocolos de modo denso todos los enrutadores en la red mantienen entradas (S, G) en sus tablas de enrutamiento multicast para todas las fuentes activas, inclusive cuando no hay receptores para los grupos a los cuales las fuentes están transmitiendo.

Uniéndose a un SPT, se trabaja con un árbol de distribución óptimo sin sufrir la sobrecarga de información e ineficiencias asociadas con otros protocolos de modo denso como DVMRP, PIM-DM y MOSPF.

Aunque es mejor trabajar con SPTs en PIM-SM en lugar de árboles compartidos, es obligatorio unirse al árbol compartido para distribuir los primeros paquetes multicast, si no fuera así los enrutadores no tendrían manera de saber que hay una fuente activa.

Se han propuesto varios métodos (incluyendo entradas dinámicas DNS) para avisar a los enrutadores cuales fuentes están activas para dichos grupos. Usando esta información, un enrutador podría unirse al SPT de todas las fuentes activas

inmediata y directamente, esto eliminaría la necesidad de un árbol compartido y su núcleo (core) o RP asociado, aunque ninguno de los métodos propuestos han contado con mucha aceptación en la comunidad Internet.

***Mensajes de unión al SPT***

Enviando un Join (S, G) hasta la fuente S, un enrutador puede unirse al SPT para la fuente S y recibir directamente el tráfico multicast que envíe la fuente.

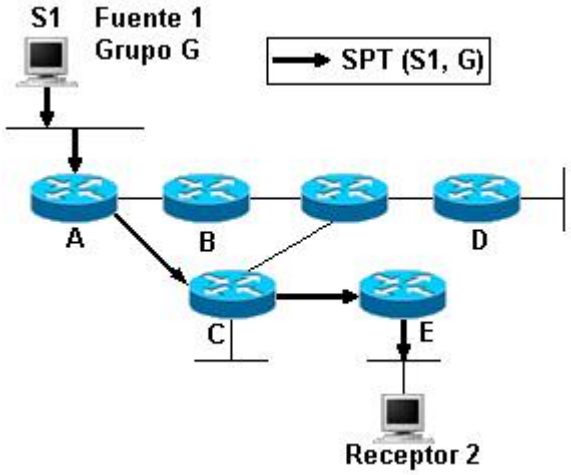


**Figura 4.6 Unión al SPT**

Suponiendo en la figura 4.6 que de alguna manera el enrutador E sabe que la fuente S1 (conectada directamente al enrutador A) está activa para el grupo G; el enrutador E envía un Join (S1, G) hacia la fuente S1 para unirse al SPT. En realidad el enrutador E debería haber aprendido que S1 está activa al recibir un paquete de la fuente por medio del árbol compartido.

El enrutador E envía un Join (S1, G) hacia la fuente por la interfaz correcta calculando la interfaz RPF hacia la fuente S1 (para calcular el RPF se usa la tabla de enrutamiento unicast, que le indicará al enrutador E que C es el próximo salto a la fuente S1).

Cuando el enrutador C recibe el Join (S1, G) crea una entrada (S1, G) en su tabla de envío multicast y adiciona la interfaz por la cual recibió el Join a la lista de interfaces de salida. Como el enrutador C tuvo que crear esta entrada, también envía un Join (S1, G) hacia la fuente, el enrutador A.

Finalmente, cuando el enrutador A recibe el Join (S1, G) adiciona el enlace al enrutador C a la lista de interfaces de salida de la entrada (S1, G). El enrutador A crea la entrada (S1, G) tan pronto como reciba el primer paquete desde la fuente, ya que es el router del primer salto para S1.

***Mensajes de podado en el SPT***

Los SPTs se pueden podar usando mensajes de podado Prune (S1, G) de la misma manera que en los árboles compartidos. El mecanismo Override también aplica en los mensajes Prune SPT.
Cuando el enrutador E ya no necesite mas tráfico (S1, G), envía un Prune (S1, G) hacia la fuente S1; cuando el enrutador C recibe este Prune, remueve la interfaz, por la cual recibió el Prune, de la lista de interfaces de salida para la entrada (S1, G), en este caso la lista queda vacía y el enrutador C tiene que enviar un Prune (S1, G) hacia la fuente S1.

Cuando el enrutador A recibe el Prune del enrutador C, remueve la interfaz por la que recibió el Prune de la lista de interfaces de salida para dicha entrada; pero como el enrutador A tiene a la fuente S1 directamente conectada no toma ninguna otra acción, continua omitiendo cualquier paquete desde S1 debido a que la lista de interfaces de salida para (S1, G) está vacía.

### 4.3.4 Mensajes PIM Join/Prune

Los mensajes Join y Prune son un solo tipo de mensaje. Cada mensaje Join/Prune contiene tanto una lista Join como una lista Prune, que pueden estar vacías, dependiendo de la información que se esté transmitiendo hacia arriba del árbol de distribución. Incluyendo múltiples entradas en las listas Join y/o Prune, un enrutador puede Unir/Podar varias fuentes y/o grupos con un solo mensaje Join/Prune mejorando la eficiencia del mecanismo periódico de actualización, ya que sólo se necesita un mensaje para actualizar el estado del enrutador upstream.

Las entradas en las listas Join y Prune de los mensajes Join/Prune comparten un formato común que contienen la siguiente información:

- Dirección de la fuente multicast, dirección IP de la fuente multicast para Unir/Podar (si la bandera Wildcard está en 1, ésta es la dirección del RP).

- Dirección del grupo multicast, dirección clase D del grupo multicast al que se desea unir o podar.
- Bit WC (Bandera Wildcard), indica si el mensaje Join/Prune va dirigido a la fuente de un SPT o hacia el RP.
- Bit RP (Bandera RP Tree), indica que esta información Join/Prune se debe enviar hacia arriba del árbol compartido. Manipulando esta información en cada lista de entradas Join/Prune, se pueden señalar varios requerimientos a un enrutador upstream.

### 4.3.5 Actualización de estados PIM-SM

Para evitar la expiración de los estados de reenvío en los enrutadores PIM, esta información tiene un tiempo de vida de aproximadamente 3 minutos, después de los cuales se borra.

Se asocia un temporizador con cada entrada (*, G) y (S, G) en la tabla de enrutamiento multicast. Cuando estos temporizadores expiran la entrada se borra, de tal manera que, los enrutadores downstream deben actualizar su estado de reenvío periódicamente para evitar ser borrados cuando expiren en el enrutador upstream, enviando un mensaje Join/Prune al vecino upstream apropiado cada 60 segundos. Cuando el vecino upstream recibe el mensaje Join/Prune actualiza su estado de reenvío multicast existente y restablece el temporizador a 3 minutos nuevamente.

Los enrutadores actualizan los árboles compartidos cada 60 segundos enviando mensajes Join (*, G) al vecino upstream en dirección al RP. Adicionalmente, los enrutadores envían mensajes Join (S, G), para actualizar los SPTs cada 60 segundos, al vecino upstream en dirección a la fuente.

### 4.3.6 Registro de la fuente

El proceso de registro de fuentes multicast ante el RP se hace por medio de mensajes de *Registro* y mensajes de *Detención de Registro*. Un concepto erróneo común es que una fuente debe registrarse antes de que cualquier receptor pueda unirse al árbol compartido. Sin embargo, los receptores pueden unirse al árbol compartido aunque no haya fuentes activas; cuando una fuente empiece a transmitir, el RP une el SPT a la fuente y empieza a enviar este tráfico hacia abajo del árbol compartido. Igualmente, las fuentes pueden registrarse aún si no

hay receptores en la red. Después, cuando un receptor se une al grupo, el RP se une a los SPTs hacia todas las fuentes en el grupo y empieza a enviar el tráfico al grupo por el árbol compartido.

***Mensajes de Registro***

Los mensajes de registro PIM son enviados por el DR de primer salto (un DR directamente conectado a la fuente multicast) hacia el RP. Los propósitos del mensaje de Registro PIM son:

1. Notificar al RP que la fuente S1 está activa y puede enviar tráfico al grupo G.
2. Entregar los paquetes multicast iniciales enviados por la fuente S1 (c/u encapsulado en un solo mensaje de Registro PIM) al RP para ser entregados al árbol compartido.

Cuando una fuente multicast empieza a transmitir, el DR de primer salto recibe los paquetes multicast enviados por la fuente y crea una entrada (S, G) en su tabla de enrutamiento multicast. Este DR encapsula cada paquete multicast en un mensaje de Registro PIM separado y lo envía al RP utilizando unicast. El DR aprende la dirección del RP por medio del procedimiento de descubrimiento de RP (*RP Discovery*).

A diferencia de los otros mensajes PIM que se difunden con multicast en un segmento local y viajan salto a salto a través de la red, los mensajes de Registro PIM se envían utilizando unicast entre el DR de primer salto y el RP.

Cuando un RP recibe un mensaje de registro PIM, desencapsula el mensaje para poder examinar el paquete multicast dentro de él. Si el paquete es para un grupo multicast activo, el RP envía el paquete por el árbol compartido. El RP entonces se une al SPT para la fuente S1 de manera que pueda recibir tráfico (S1, G) de forma nativa y no encapsulado dentro de los mensajes de registro. Si no hay un árbol compartido para el grupo, el RP simplemente descarta los paquetes multicast y no envía el mensaje Join hacia la fuente.

***Mensajes de Detención de registro***

El RP, utilizando unicast, envía mensajes de detención de registro al DR de primer salto, indicándole que detenga el envío de mensajes de registro (S1, G) porque se está presentando cualquiera de las siguientes situaciones:

- Si el RP empieza a recibir tráfico multicast desde la fuente S1 por medio del SPT (S1, G) entre la fuente y el RP.
- Si el RP no necesita este tráfico porque no tiene un árbol compartido activo para el grupo.

Cuando el DR de primer salto recibe un mensaje de detención de registro sabe que el RP recibió exitosamente el mensaje de registro y se ha presentado una de las 2 situaciones. En ambos casos, el DR termina el proceso de registro y no encapsula más paquetes (S1, G) en mensajes registro.

### 4.3.7 Conmutación al SPT

PIM-SM habilita al enrutador de último salto (enrutador que tiene hosts, pertenecientes a un grupo multicast, conectados directamente) para conmutar de un árbol compartido al SPT para una fuente específica. Para llevar a cabo la conmutación se especifica un umbral SPT en términos del ancho de banda. Si este umbral se excede, el enrutador de último salto se une al SPT. Para ilustrar de manera más clara esta conmutación se analiza la red de la figura 4.7.

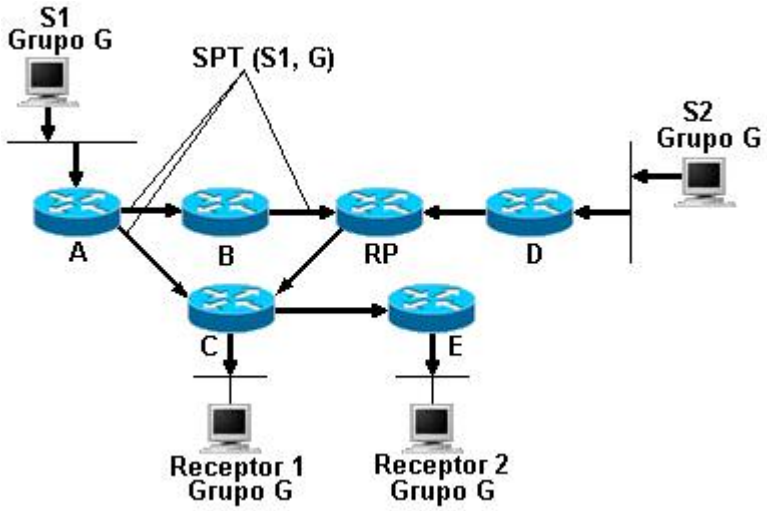


**Figura 4.7 Conmutación al árbol de ruta de acceso más corta**

El enrutador C tiene la opción de conmutar a los SPTs para la fuente S1; C tendría que enviar un mensaje Join (S1, G) hacia la fuente S1, es decir, al enrutador A; cuando el enrutador A recibe este Join adiciona la interfaz por la cual lo recibió a la lista de interfaces de salida para la entrada (S1, G) en su tabla de envío multicast. Esto adiciona el enlace entre A y C al SPT (S1, G), como se muestra en la figura 4.7, así que el tráfico multicast (S1, G) puede fluir directamente hacia el enrutador C por medio del SPT (S1, G).

Normalmente el grupo de umbrales SPT se configuran de forma consistente en todos los enrutadores de la red, por lo que el enrutador E también iniciaría una conmutación al SPT enviando un Join (S1, G) al enrutador upstream (C) hacia la fuente. Los enrutadores son los que inician la conmutación al SPT no los hosts.

Una vez se haya terminado este procedimiento, hay 2 rutas por los cuales puede fluir el tráfico multicast (S1, G) para llegar al enrutador C: el árbol compartido y el SPT, es decir, que se están entregando paquetes duplicados entregados al enrutador C, presentándose un desperdicio de ancho de banda de la red, así que es necesario informarle al RP que pode el tráfico multicast (S1, G) que está enviando por el árbol compartido.

***Podado de fuentes del árbol compartido***

Cuando se presenta el caso expuesto anteriormente se usa un tipo especial de mensaje Prune para decirle al RP que pode este tráfico enviado por el árbol compartido: el mensaje ***Prune Bit-RP (S, G)*** tiene la bandera RP en 1 en la lista de entrada (esta bandera RP indica que este mensaje se debe enviar hacia el RP por medio del árbol compartido). Colocando este bit en 1 en un mensaje Prune (S1, G) y enviándolo hacia arriba por el árbol compartido, se les dice a los enrutadores a lo largo del árbol que poden el tráfico multicast de S1.

En la figura 4.7, el enrutador C envía un mensaje Prune RP-Bit (S1, G) hacia el RP por el árbol compartido para podar el tráfico de S1. Después de recibir este Prune especial, el RP actualiza su estado de envío multicast, así que el tráfico (S1, G) no se envía al enrutador C. Sin embargo, como este enlace al enrutador C era la única interfaz en el árbol compartido que necesitaba tráfico multicast (S1, G), el RP no necesita más de este tráfico, de tal manera que envía un mensaje Prune (S1, G) hacia S1 para detener ese tráfico.

Como resultado se poda el árbol compartido (S1, G), dejando sólo el enlace entre los enrutadores A y C. El enrutador E sigue recibiendo tráfico multicast (S1, G) del enrutador C, pero no se da por enterado que el enrutador C ha conmutado al SPT para S1.

### 4.3.8 Enrutador Designado PIM-SM

PIM-SM elige un DR en cada red multiacceso usando mensajes Hello. En PIM-DM el DR tenía significado sólo si se usaba IGMPv1 en la red ya que esta versión no tiene un mecanismo de elección de Querier, pero en PIM-SM el DR tiene mucha más importancia.

***La función del DR***

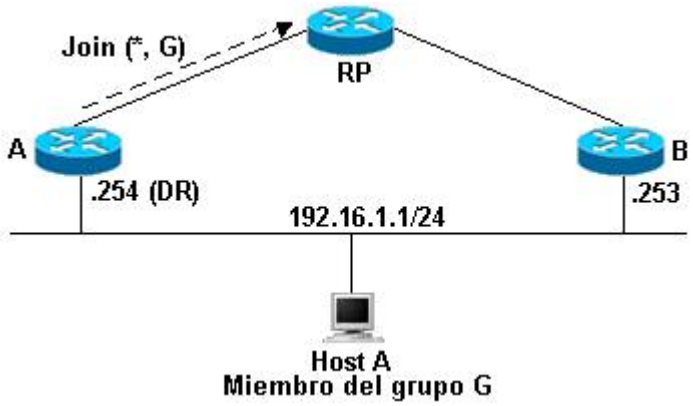


**Figura 4.8 Enrutador designado PIM-SM**

En la figura 4.8 hay dos enrutadores conectados a una red multiacceso común con un receptor activo para el grupo G. Como se utiliza el Modelo de Unión Explicito, sólo el DR (Enrutador A) debería enviar mensajes Join al RP para construir el árbol compartido para el grupo G. Si se les permite a ambos enrutadores enviar mensajes Join (*, G) al RP, se pueden crear rutas paralelas y el host A recibiría paquetes multicast duplicados. Si el host A empieza a transmitir tráfico multicast al grupo, el DR es el responsable de enviar mensajes al RP. Si los enrutadores A y B envían mensajes de registro al RP, este último recibiría paquetes multicast duplicados.

***Enrutador de redundancia designado***

PIM-SM no sólo permite el mecanismo de elección del DR sino también la manera de detectar si el DR falla; el enrutador B detectaría esta situación cuando su

adyacencia de vecino con el enrutador A expire, lo que llevaría a una nueva elección de DR (enrutador B). Si esto sucede, el enrutador B está enterado que hay un receptor activo (host A) en la red porque ha estado escuchando mensajes de reporte de membresía IGMP desde el host. Así que el enrutador B tiene una entrada creada para el grupo G en esta interfaz; el enrutador B enviaría un mensaje Join al RP tan pronto sea elegido como el nuevo DR.

### 4.3.9 Descubrimiento del punto de reunión RP

Para que PIM-SM trabaje apropiadamente, todos los enrutadores dentro del dominio deben conocer la dirección del RP. En redes pequeñas que usan un sólo RP para todos los grupos multicast, es posible configurar manualmente la dirección IP del RP en cada enrutador.

Sin embargo, si el tamaño de la red crece o si el RP cambia frecuentemente, la configuración manual de cada enrutador no sería lo más conveniente. Este problema es peor si los diferentes grupos usan diferentes puntos de reunión en otras locaciones en el dominio.

PIMv2 define un mecanismo Bootstrap que permite a todos los enrutadores PIM-SM dentro de un dominio aprender dinámicamente todos los mapeos *Group-to-RP* y evitar el problema de la configuración manual del RP. Además, la implementación PIM de Cisco tiene otro mecanismo, *Auto-RP*, que cumple la misma tarea. (Auto-RP de Cisco se desarrolló antes de que la especificación PIMv2 se escribiera).

## 4.4 ÁRBOLES BASADOS EN NÚCLEO - CBT

El protocolo de enrutamiento multicast Core-Based Trees (CBT) es un protocolo en desarrollo y lo ha sido por muchos años. La versión original CBTv1 fue remplazada por CBTv2 que está definida en el RFC 2189 y no es compatible con la versión anterior; sin embargo CBTv1 nunca fue implementada en ninguna red productiva por lo que la compatibilidad no es un inconveniente. Hasta la fecha CBTv2 no ha tenido ninguna implementación importante, lo que posiblemente sea bueno ya que el draft de CBTv3 esta a punto de convertirse en una especificación y no es compatible con la versión 2.

### 4.4.1 Funcionamiento de CBT

Uno de los aspectos importantes en el diseño de CBT fue la escalabilidad a medida que crece el número de grupos activos en la red. Por lo tanto uno de los objetivos de CBT es la reducción de la información de estado de los grupos multicast que el enrutador mantiene e intercambia. Para lograr esto CBT fue diseñado como un protocolo en modo disperso que sólo utiliza árboles de distribución compartidos para entregar tráfico de los grupos multicast a las porciones de la red que se han unido explícitamente al grupo. Los árboles compartidos tienen su raíz en un enrutador llamado ***Core*** y permiten que el tráfico multicast fluya en ambas direcciones, hacia arriba o hacia abajo. Debido a la naturaleza bidireccional de estos árboles compartidos, los enrutadores sobre el árbol no tienen que ejecutar ninguna tarea especial para enviar el tráfico multicast de origen local al Core. El primer enrutador CBT puede enviar el tráfico hacia arriba del árbol. Cada enrutador en el árbol simplemente reenvía el tráfico por todas las interfaces menos por donde fue recibido el paquete. La figura 4.9 es un ejemplo de un árbol bidireccional, los hosts M1 a M7 están unidos al árbol y el host M3 también es una fuente de tráfico multicast para el grupo (cuando un host es un miembro y una fuente para el grupo se le llama *miembro fuente*).

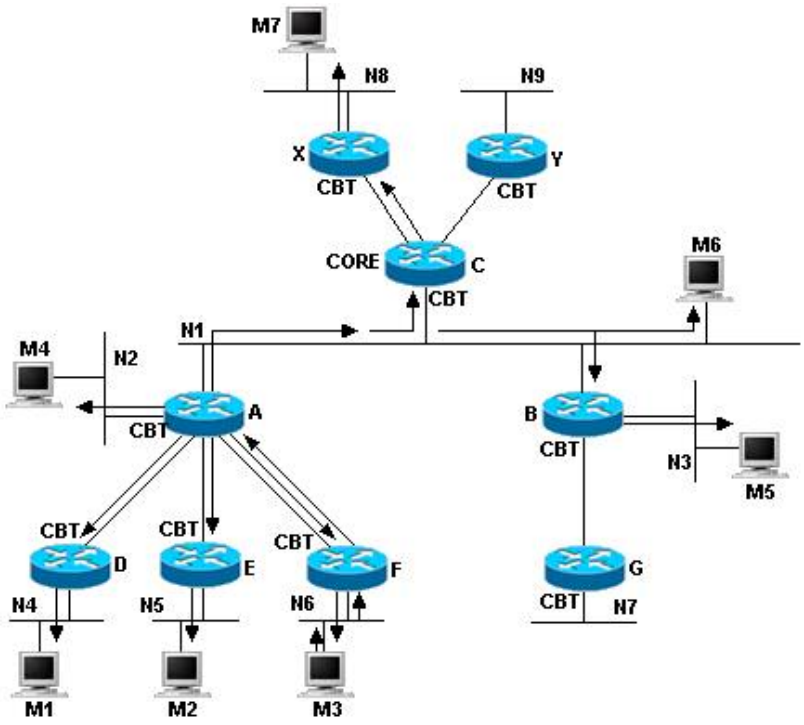


**Figura 4.9 Árbol bidireccional CBT**

El flujo generado por el host M3 puede fluir hacia arriba y hacia abajo del árbol compartido; los enrutadores no necesitan de estados de reenvío adicional si tienen miembros fuente.

Contrario a PIM-SM los enrutadores que no poseen miembros fuente (lo que significa que el enrutador no esta en el árbol compartido) deben mandar la información de la fuente conectada a ellos por medio de un túnel IP-IP para que puedan alcanzar al Core y llevar el tráfico hacia abajo del árbol. La única forma de eliminar el encapsulamiento IP-IP es que el enrutador se una al árbol compartido, lo que implica que un host conectado a él se una al grupo multicast.

La idea de utilizar árboles compartidos con el fin de minimizar la información de estado en los enrutadores no permite que CBT haga uso de los SPT; en algunas topologías el árbol compartido desde el *Core* hasta los receptores y el SPT desde la fuete hacia los mismos receptores puede ser diferente. El árbol compartido no siempre es la ruta mas optima haciendo que una latencia innecesaria aparezca en el flujo de trafico de un grupo multicast.

### 4.4.2 Versión 3 de CBT

CBTv3 utiliza mucha más información de estado que su predecesor porque se han agregado extensiones para manejar sus deficiencias y agregarle otras características.

Una de las principales características es la definición de dos nuevos estados: (*, C*ore*) y (S, G), con la idea de manejar el tráfico multicast interdominio usando enrutadores de frontera CBT (BRs). Estas adiciones hacen que CBTv3 no sea compatible con la versión 2.

## 4.5 MULTICAST OPEN SHORTEST PATH FIRST - MOSPF

OSPFv2 (definido en el RFC 2328) es un protocolo de enrutamiento unicast basado en el estado de enlace y utiliza una jerarquía de red de dos niveles. La primera es la jerarquía 0, también conocida como el área backbone, a la cual todas las áreas de segundo nivel están conectadas a través de enlaces físicos o virtuales. OSPF es jerárquico porque todo el tráfico inter-área tiene que pasar por

el área 0. Los enrutadores OSPF inundan cada área con información de estado de enlace que describe la topología de la red dentro del área y cada enrutador mantiene una copia de esta información en su base de datos de área. Cada vez que hay un cambio en la topología, la nueva información de estado de enlace es inundada por la red para que cada enrutador tenga una topología actualizada de la red. Utilizando la información de topología de la base de datos del área cada enrutador construye un árbol de distribución de menor costo de la red dentro del área (siendo él mismo la raíz) utilizando el algoritmo de Dijkstra. Este árbol de distribución es utilizado para construir una tabla de reenvío unicast. Enrutadores especiales en los bordes de cada área (ABRs) conectan las áreas de segundo nivel con el área 0 y mantienen una base de datos de área separada para cada área a la cual están conectados (incluyendo una para el área 0).
Multicast Open Shortest Path First (MOSPF) esta definido en el RFC 1584 "Multicast Extensions to OSPF" y proporciona definiciones de formato para información OSPF adicional y especificaciones de operación basadas en gran medida en el trabajo de Steve Deering. Las extensiones al protocolo OSPF permiten que el tráfico multicast pueda ser reenviado dentro de una red unicast OSPF usando SPTs a través de un conjunto parcial o de todos los enrutadores OSPF.

La ventaja más significativa de MOSPF es que comparte la capacidad de OSPF para responder rápidamente a los cambios en la topología de la red ya que utiliza métodos de enrutamiento basados en estados de enlace para calcular los árboles de distribución multicast. Sin embargo esta capacidad aumenta el gasto de los recursos de CPU a medida que aumenta el número de pares (S, G) en la red ya que también aumenta el número de cálculos del algoritmo Dijkstra para crear cada SPT.

MOSPF es probablemente el protocolo más adecuado para redes multicast de propósitos específicos en donde el administrador de la red tiene control absoluto sobre los factores que afectan el desempeño y la escalabilidad de MOSFP como:

- Ubicación de las fuentes
- Número de fuentes
- Número de grupos
- Pertenencia a grupos

## 4.6 PROTOCOLO DE PASARELA DE FRONTERA MULTIPROTOCOLO – MBGP

MBGP, Multiprotocol Border Gateway Protocol también conocido como Multicast Border Gateway Protocol, esta definido en el RFC 2858 (Multiprotocol extensions for BGP-4) que adiciona capacidades para que BGP-4 pueda transportar información de enrutamiento de diferentes protocolos de nivel de red como: IPv6, IPX, etc. También se define un mecanismo para la propagación de información de enrutamiento multicast a través de diferentes Sistemas Autónomos.

MBGP es un BGP mejorado que transporta dos grupos de rutas, uno correspondiente al enrutamiento unicast y otro para el enrutamiento multicast. Las rutas asociadas al enrutamiento multicast son empleadas por el protocolo PIM para crear los árboles de distribución de datos.

Así como lo hace BGP en el caso del enrutamiento unicast, MBGP provee información acerca de la accesibilidad de varias redes multicast e implementa una política de control para el enrutamiento multicast. Los enrutadores que intercambian información de enrutamiento multicast, propagan sus tablas de enrutamiento local, conocidas como tablas MBGP, utilizando actualizaciones de enrutamiento que ellos obtienen de sus vecinos. Cada entrada en la tabla MBGP corresponde a la ruta hacia una red. Uno de los campos es la dirección IP del enrutador del siguiente salto y el otro, llamado camino AS, es una secuencia de Sistemas Autónomos que el paquete tiene que atravesar para alcanzar la red destino.

Las rutas MBGP, son los trayectos más cortos desde el enrutador hacia la red destino. La diferencia básica entre BGP y MBGP radica en la manera como se utilizan estas rutas. Para el enrutamiento unicast se emplean las rutas BGP para decidir la interfaz por la que los paquetes serán enviados, contrario al enrutamiento multicast el chequeo RPF se realiza con las rutas MBGP.

## 4.7 PROTOCOLO DE DESCUBRIMIENTO DE FUENTE MULTICAST

El Protocolo de Descubrimiento de Fuente Multicast (MSDP, Multicast Source Discovery Protocol), definido en el RFC 3618, tiene representantes en cada dominio que anuncian a otros dominios la existencia de fuentes activas.

El modo de funcionamiento de MSDP se describe en los siguientes pasos:

- Cuando hay una nueva fuente activa para un grupo el enrutador MSDP en el dominio detectará la existencia de la nueva fuente y enviará un mensaje de Fuente activa (SA, Source Active) a todos los enrutadores MSDP conectados directamente.
- La inundación de mensajes MSDP:
  - Los enrutadores MSDP que reciben el mensaje SA revisarán si el mensaje llegó por la ruta correcta. Estos chequeos RPF son necesarios para prevenir bucles de mensajes SA.
  - Si un enrutador MSDP recibe una mensaje SA por la interfaz correcta, el mensaje se reenvía a todos los enrutadores MSDP a excepción del que envió el mensaje.
- Dentro de un dominio, un enrutador MSDP (al igual que el RP) revisará si tiene creada una entrada para algún miembro del grupo en el dominio. Si no existe ninguna entrada, el RP enviará un mensaje PIM Join a la dirección de la fuente anunciada en el mensaje SA.
- Si el mensaje contiene tráfico multicast, el RP lo reenvía al árbol multicast. Una vez los miembros del grupo reciban la información, pueden escoger conmutar al SPT usando las convenciones de PIM-SM.
- Los últimos tres pasos se repiten hasta que todos los enrutadores MSDP han recibido el mensaje SA y todos los miembros del grupo empiezan a recibir información desde la fuente.

Cuando las fuentes multicast empiezan a transmitir, la red necesita crear algún tipo de estado de enrutamiento para controlar el flujo de paquetes. En este capítulo se han tratado diferentes tipos de protocolos de enrutamiento multicast que cumplen con esta función; sin embargo, en el caso de MSDP, la información de las fuentes existentes tiene que ser transmitida con anterioridad para crear los

estados de enrutamiento. Esta tarea adicional incrementa la carga del manejo de grupos. Cuando los grupos son dinámicos, debido a fuentes intermitentes o a los eventos de Unión/Abandono de miembros del grupo, la carga de la gestión de grupos puede ser significativa. Dos problemas específicos relacionados con las fuentes y grupos dinámicos son:

- Latencia de Unión. Debido a que los mensajes SA sólo se envían periódicamente, puede haber un retardo significativo entre el momento en que los receptores se unen y cuando escuchan el próximo mensaje SA; para resolver este problema los enrutadores MSDP se deben configurar para guardar en cache mensajes SA; un enrutador MSDP que no tiene esta capacidad puede solicitar un mensaje SA-Request a un par MSDP que pueda almacenar los mensajes SA. Esto proporciona a los enrutadores MSDP un mecanismo para conocer las fuentes activas reduciendo la latencia. La desventaja es el almacenamiento extra y la complejidad de mantener una memoria cache.
- Fuentes intermitentes. El problema ocurre cuando se trata de establecer un árbol multicast para esta clase de fuentes, cuando uno o algunos paquetes se envían al RP. El RP escuchará el paquete e inundará la red con un mensaje SA, los RPs en otros dominios enviarán hacia la fuente mensajes Join. Sin embargo, como no existe ningún estado de reenvío multicast cuando el paquete fue enviado originalmente y como toma tiempo reenviar mensajes SA y conocer los estados de reenvío de los otros RPs, la ráfaga original no alcanzará nuevos receptores. Una vez se establece el estado de reenvío, todos los paquetes subsecuentes deben alcanzar a los receptores. Pero si el periodo de silencio entre ráfagas de paquetes excede el valor de expiración de una entrada de reenvío (generalmente 3 minutos) , esta entrada se descarta. Cuando se envía otro anuncio de sesión, se repite el mismo proceso de establecer una entrada de estado, pero perdiendo la primera ráfaga de información. De esta manera, los paquetes enviados por una fuente intermitente nunca alcanzarán a los miembros de un grupo. La solución es que los mensajes SA transporten los primeros n paquetes de información.

### *Escalabilidad MSDP*

Es muy importante considerar la escalabilidad para MSDP; por la forma de operación de MSDP, si el tráfico multicast sufre un incremento grande, la carga de MSDP podría ser muy alta, esto ocurriría si se tiene miles de fuentes multicast. El número de mensajes SA (más la información) que se inundan por toda la red podría ser muy grande. La conclusión a la cual se llega es que MSDP no es una solución suficientemente escalable por lo que no se califica como una solución a largo plazo; pero debido a que las soluciones a largo plazo no están listas aún para su puesta en marcha, MSDP es una solución inmediata a una necesidad inmediata.

Los enrutadores multicast se comunican entre ellos por medio de los protocolos expuestos en este capítulo y cuando varios de ellos se unen aparece un nuevo concepto de backbone multicast. El siguiente capítulo se enfoca en la historia y evolción del Backbone Multicast de Internet.

# 5 MBONE - BACKBONE MULTICAST DE INTERNET

El Backbone Multicast de Internet (MBone) era un conjunto pequeño de enrutadores y hosts interconectados, capaces de enviar tráfico Multicast. Hay que aclarar que el trafico multicast todavía no alcanza a todos los puntos de la Internet.

Se piensa equivocadamente que si se tiene conexión a Internet entonces se puede recibir tráfico multicast; hace falta activar el enrutamiento IP Multicast en los enrutadores de frontera o instalar algún software especial en el PC para recibir tráfico multicast a través de un proveedor ISP.

## 5.1 LAS SESIONES DEL MBONE

La idea de la difusión multicast es poder transmitir información de uno a muchos y para hacer esto se debe acordar un espacio para que la fuente y los receptores envíen y reciban información de manera sincronizada.

A este tipo de reunión se le llama *sesión* y se define por el tiempo de duración de la reunión, la periodicidad de la misma, el tipo de aplicación que se utiliza, el tipo de protocolo a diferentes niveles y tipo de compresión si es el caso, además de la información de la fuente y el grupo multicast objetivo.

Existen varias aplicaciones que permiten manejar las sesiones multicast como el gwTTS de la universidad de Virginia, el ISC, NSDR, entre otros, aunque el más común de todos es el SDR; a través de estas herramientas se pueden conocer y tratar diferentes tipos de eventos que se difunden a través del Mbone.

## 5.2 LA HISTORIA DEL MBONE

A principios de 1990, muchos miembros de la comunidad investigativa se quejaron ante el DARPA (Defense Advanced Research Projects Agency) que la Internet se había vuelto una red de producción y no tenía una disposición investigativa y experimental de las nuevas tecnologías de red. Debido a esto, el gobierno de los Estados Unidos proporcionó la red DARPA de prueba (DARTNet - DARPA Tested Network) para darle a los investigadores una red donde ellos pudieran probar y evaluar nuevas tecnologías y herramientas sin afectar la producción de Internet.

La DARTNet se compuso inicialmente por varios sitios conectados a través de líneas T1 que incluían a Xerox PARC, Lawrence Berkley Labs, SRI, ISI, BBN, MIT y la Universidad de Dalaware. Estos sitios usaban estaciones SUN SPARC corriendo el demonio de protocolo de enrutamiento unicast *routed* y el demonio *mrouted* como protocolo de enrutamiento multicast. De esta manera, DARTNet soportaba de forma nativa multicast sobre IP entre todos los sitios. Semanalmente se llevaban a cabo audio-conferencias entre investigadores ubicados en diferentes sitios de la DARTNet alrededor de los Estados Unidos.

A principios de 1992, la IETF hizo planes para llevar a cabo su siguiente reunión en el mes de marzo en San Diego, California; Una de las investigadoras de la DARTNet no podía estar presente. Muchos de los investigadores de la DARTNet, incluyendo Steve Deering y Steve Casner, decidieron difundir el flujo de audio de la conferencia de la IETF a través de multicast.

Steve Deering y Steve Casner organizaron la conferencia de audio a través de un servidor SUN Sparc; para difundir el audio multicast en la DARTNet, se configuró un túnel DVMRP entre la estación Sparc de la IETF y el backbone de la DARTNet. También se repartieron invitaciones para participar en la conferencia multicast de audio de la IETF a varias organizaciones de investigación de los Estados Unidos, Australia, Suecia e Inglaterra, junto con la información de cómo configurar una estación Sun Sparc con un túnel DVMRP a través de la Internet hacia el backbone de la DARTNet. Muchos de estas organizaciones respondieron a la invitación y

configuraron túneles DVMRP hacia el backbone de la DARTNet; el resultado fue la primera audio-conferencia multicast de la IETF para varios sitios en la Internet alrededor del mundo.

A finales de Marzo de 1992, los túneles DVMRP se desconfiguraron y la DARTNet retornó a la normalidad. Sin embargo la audio-conferencia de la IETF fue tan exitosa que se planeó multidifundir tanto el audio como el video de la próxima convención de la IETF.

Los administradores de la DARTNet y otros sitios participantes de las conferencias de la IETF decidieron dejar los túneles DVMRP para llevar a cabo conferencias multimedia sobre Internet. Estos primeros túneles complementados con la DARTNet sirvieron como la red núcleo multicast inicial, que posteriormente pasó a ser el Mbone.

Además de los túneles DVMRP entre estaciones, ahora el Mbone posee la capacidad de soportar multicast nativo, es decir, los enrutadores son capaces de manejar paquetes multicast. Además, la continuidad en las investigaciones ha llevado a desarrollar y emplear dos protocolos de modo denso adicionales: MOSPF y PIM-DM. Igualmente se desarrollaron dos protocolos de modo disperso como PIM-SM y CBT para solucionar los problemas de escalabilidad de los protocolos de modo denso.

No sólo se han hecho progresos en el área de desarrollo de protocolos, también el crecimiento del Mbone ha repercutido en un mayor conocimiento de los usuarios acerca de la tecnología multicast, lo que generó una demanda de nuevas aplicaciones y mejor soporte para la información de tiempo real. Se han hecho mejoras en los protocolos de nivel de transporte, como por ejemplo, el Protocolo de Tiempo Real (RTP).

Para interconectar dominios administrados localmente (Sistemas Autónomos) que soportan multicast nativo se ha desarrollado lo que se conoce como Multicast Interdominio dada la necesidad de proveer multicast escalable y jerárquico en la Internet.

La solución inmediata para el enrutamiento multicast interdominio consta de tres partes. La primera es una extensión del protocolo unicast BGP; la segunda y la tercera son protocolos adicionales necesarios para construir e interconectar árboles a través de los límites de un dominio.

Con la utilización de túneles DVMRP, el Mbone era una red de jerarquía plana, primer inconveniente a superar por los problemas de escalabilidad. La agregación y abstracción de rutas así como el enrutamiento salto a salto, son provistos por BGP en el enrutamiento unicast. BGP ofrece un control y abstracción substancial de rutas entre dominios; dentro del dominio un administrador de red puede utilizar cualquier protocolo de enrutamiento deseado, para alcanzar un host en un dominio externo simplemente es cuestión de escoger el enlace apropiado. BGP soporta enrutamiento interdominio, intercambiando de manera confiable información de asequibilidad. Esta información se utiliza para calcular una ruta de estilo vector distancia punto a punto entre los sistemas autónomos. Cada sistema autónomo anuncia el conjunto de rutas que puede alcanzar y el costo asociado. Cada enrutador de frontera puede calcular el conjunto de sistemas autónomos que debe atravesar para alcanzar cualquier red. El uso de un algoritmo de vector distancia junto con la información completa de la ruta le permite a BGP superar muchas limitaciones de los algoritmos tradicionales de vector distancia. Los paquetes también se envían salto a salto, pero se necesita menos información y se toman mejores decisiones de enrutamiento.

La funcionalidad provista por BGP y su bien conocido paradigma de conexión de sistemas autónomos, son importantes catalizadores para el soporte de multicast interdominio. Una versión de BGP capaz de transportar rutas multicast no sólo proporcionaría enrutamiento jerárquico y políticas de decisión, sino que también permitiría a un proveedor de servicios utilizar diferentes topologías para tráfico unicast y multicast.

La extensión de BGP que permite transportar rutas multicast se ha llamado MBGP (Multiprotocol Extensions to BGP4). MBGP es capaz de transportar rutas multiprotocolo adicionando el Identificador de Familia de Dirección Subsecuente (SAFI, Subsequent Address Family Identifier) a dos mensajes BGP4: MP_REACH_NLRI y MP_UNREACH_NLRI.

Con MBGP en vez de que cada enrutador deba saber la totalidad de la topología plana multicast, sólo necesitan saber la topología de su propio dominio y las rutas para alcanzar cada uno de los otros dominios.

Para describir de manera más clara el servicio que provee MBGP, se debe tener presente que si un dominio anuncia asequibilidad para multicast, debe notificar que tiene rutas para las fuentes que están en las redes que puede alcanzar. Los mensajes MBGP no llevan información de grupos multicast (las direcciones clase D nunca se envían en un mensaje MBGP). Los árboles multicast se construyen usando una ruta inversa hacia la fuente, sin embargo, se utiliza la información MBGP cuando se envía un mensaje Join desde un RP o receptor hacia la fuente. MBGP provee la información del próximo salto entre dominios.

MBGP es el primer paso para lograr multicast interdominio, pero no es toda la solución. MBGP puede determinar el siguiente salto a un host, pero no provee la función de construir un árbol multicast. Para esto se utilizó PIM-SM, para establecer un árbol multicast entre dominios que contengan miembros de un grupo.

Hasta ahora se tiene un protocolo que intercambia rutas para soporte multicast y un protocolo para conectar receptores y fuentes a través de los límites de un dominio, PIM-SM. Pero todavía falta una función para conectar dominios de modo disperso juntos. El problema es básicamente cómo informar a un RP en un dominio que hay fuentes en otros dominios. El inconveniente ahora es que un grupo puede tener varios RPs; sin embargo, en realidad sólo hay un RP por dominio, pero ahora se ven involucrados varios dominios.

El problema surge cuando los miembros de un grupo están dispersos alrededor de varios dominios. No hay ningún mecanismo para conectar varios árboles multicast intradominio. El tráfico de todas las fuentes para un grupo dentro de un dominio alcanzará a todos los receptores pero cualquier fuente fuera del dominio quedará fuera del árbol. Esto debido a que los receptores envían mensajes de unión hacia el RP y las fuentes envían mensajes de registro hacia el mismo RP. Sin embargo, no hay forma de que un RP dentro de un dominio encuentre fuentes en otros dominios que estén usando otros RPs. No hay un mecanismo

para que los RPs se comuniquen cuando reciben un mensaje de registro de una fuente.

La solución adoptada para solucionar este problema es un nuevo protocolo llamado MSDP (Multlicast Source Discovery Protocol), que posee representantes en cada dominio que anuncian a los demás dominios la existencia de fuentes activas. MSDP se ejecuta en el mismo enrutador que funciona como RP. MSDP hace uso de TCP para el intercambio confiable de mensajes de sesión.

Ahora que es posible el enrutamiento multicast interdominio, el problema es como migrar de la topología plana del Mbone a esta nueva infraestructura. La solución fue asignar al Mbone su propio sistema autónomo, llamado AS10888. Todos los túneles Mbone y los sitios conectados por túneles fueron relegados al AS10888. La conectividad entre AS10888 y otros sistemas autónomos multicast fue provista por el NASA Ames[*] MIX[**]. La NASA Ames MIX brindó conectividad entre el Mbone (AS10888) y otros sistemas autónomos que tuvieran implementado MBGP/PIM-SM/MSDP. La puesta en marcha de multicast interdominio continuó creciendo y se eliminó la topología de enrutamiento plana.

El desarrollo de multicast en Internet 2 sigue las pautas trazadas por el Grupo de Trabajo Multicast de Internet 2, de modo que se trabaje con multicast nativo y protocolos de modo disperso. No se permiten túneles y todos los enrutadores deben soportar enrutamiento multicast interdominio usando MBGP/MSDP. El desarrollo de multicast en Intenert2 ha tenido una buena acogida, que condujo al impulso de su backbone, a la conexión de redes de alta velocidad y el trabajo con aplicaciones multicast de gran ancho de banda (~30Mbps).

## 5.3 ESTUDIO DE PROTOCOLOS UTILIZADOS EN EL MBONE

Mientras MBGP/PIM-SM/MSDP es una buena solución a corto plazo, todavía se necesitan soluciones a largo plazo que pueden ser divididas en dos grupos:

[*] AMES es un centro de investigación de la NASA localizado en Moffett Field, California en Silicon Valley. Fue fundado el 20 de Diciembre de 1939 como un laboratorio de investigación de naves espaciales por el Comité Asesor Nacional Aeronáutico (NACA) y en 1958 se hizo parte de la NASA.

[**] MIX (Multicast-friendly Internet eXchange) es una arquitectura diseñada con el objetivo de estandarizar el enrutamiento multicast nativo, las políticas de intercambio de rutas para un dominio escalable y permitir una variedad de protocolos IGP y topologías para uso intradominio. Está definido en el draft-ietf-mboned-mix-02.txt.

algunas basadas en la filosofía tradicional de IP Multicast y las otras que buscan un cambio de la tecnología multicast para simplificar los problemas asociados.

### 5.3.1 Protocolo Multicast de Pasarela de Frontera

El Protocolo Multicast de Pasarela de Frontera (BGMP, Border Gateway Multicast Protocol) fue la primera solución que se dio a largo plazo para multicast interdominio en Internet. La idea principal de BGMP es construir árboles compartidos bidireccionales entre dominios utilizando una única raíz. Una de las funciones de BGMP es decidir en que dominio estará la raíz del árbol compartido. BGMP se basa en la premisa de que las dependencias interdominio se pueden evitar usando un esquema de direccionamiento estricto; es decir, cada dominio tendrá asignado direcciones específicas o un rango de direcciones, la idea es que si un dominio particular posee la dirección de un grupo específico, este se verá involucrado en la prestación del servicio.

El esquema de asignación de direcciones de BGMP se debe definir claramente para evitar colisiones. Por lo tanto, el mecanismo utilizado por el SDR, que escoge aleatoriamente las direcciones, no es apropiado. BGMP es relativamente flexible y puede utilizar cualquier esquema de direccionamiento siempre y cuando sea estricto.

#### *Protocolo MASC*

El protocolo MASC (Multicast Address-Set Claim) definido en el RFC 2909, soporta la asignación de direcciones entre dominios. Este protocolo incluye mecanismos para prevenir la duplicación de direcciones utilizando la Arquitectura de Asignación de Direcciones Multicast (MAAA, Multicast Address Allocation Architecture). MAAA es una especificación de la IETF (RFC 2908) que presenta 3 niveles de asignación de direcciones: a nivel de dominio, dentro de un dominio y entre los hosts y la red. MASC actúa como un protocolo de asignación de direcciones de alto nivel y opera entre dominios; el Protocolo de Asignación de direcciones (AAP, Allocation Address Protocol, especificado en el draft-ietf-malloc-aap-04.txt) es utilizado dentro de un dominio y el Protocolo MADCAP (Multicast Address Dynamic Client Allocation) definido en el RFC 2730, es utilizado por los hosts para solicitar direcciones a un Servidor de Asignación de Direcciones Multicast (MAAS, Multicast Address Allocation Server).

***Protocolo GLOP***

El protocolo de direccionamiento GLOP, definido en el RFC 2770, es mucho más simple ya que propone la asignación estática de direcciones multicast a cada sistema autónomo. Se asigna un "glop" de direcciones a cada AS. El número de AS se codifica como parte de la dirección. La primera versión de GLOP se evaluó en sólo una parte del rango de direcciones 224/4, utilizando el rango 233/8: el primer octeto es estático, los dos siguientes octetos codifican el número de AS y el octeto final provee un rango de direcciones para ser asignadas. Este propuesta tiene dos limitaciones; la primera, debido a que sólo 8 bits, o 256 direcciones, están disponibles para cada AS, es probable que este número de direcciones sea insuficiente para un AS, lo que podría resolverse utilizando más espacio del rango de direcciones clase D ó migrando a IPv6; el segundo inconveniente es que GLOP no especifica un mecanismo con el cual asignar las direcciones dentro del dominio, esto podría solucionarse utilizando un simple procedimiento administrativo, por ejemplo un protocolo dinámico como AAP/MADCAP o una versión modificada del SDR para intradominio.

### 5.3.2 Arquitectura Multicast direccionada en la raíz

Dada la complejidad de MBGP/PIM-SM/MSDP y BGMP además de la necesidad de manejar problemas adicionales relacionados con la seguridad, facturación y gestión, algunos miembros de la comunidad multicast están esperando realizar cambios fundamentales en el modelo multicast. Una propuesta es la arquitectura RAMA (Root Addressed Multicast Architecture). La premisa de los protocolos asociados a la arquitectura RAMA es que la mayoría de las aplicaciones multicast utilizan una sola fuente o tienen una fuente primaria fácilmente identificable. Haciendo esta fuente la raíz del árbol, se puede eliminar la complejidad de la ubicación del núcleo (Core) de otros protocolos de enrutamiento. Existen dos protocolos relacionados a la arquitectura RAMA.

***Multicast Express***

Express se diseñó específicamente como un protocolo de fuente única. La raíz del árbol se sitúa en la fuente y los miembros del grupo envían mensajes a lo largo de la ruta de regreso a la fuente. Express también provee mecanismos para recolectar de manera efectiva información de los suscriptores.

***Multicast Simple***

Multicast Simple y Multicast Express son similares, pero la diferencia es que el primero brinda mayor flexibilidad soportando múltiples fuentes por grupo. Una única fuente debe ser escogida como primaria y la raíz del árbol será situada en este nodo. Los receptores envían mensajes Join hacia la fuente para construir el árbol bidireccional. Las fuentes adicionales envían paquetes hacia la fuente primaria. Debido a que el árbol es bidireccional, tan pronto los paquetes alcanzan un enrutador en el árbol, estos son reenviados en ambos sentidos: downstream hacia los receptores y upstream hacia el Core. Las personas a favor de esta propuesta creen que de esta forma se elimina el problema de asignación de direcciones y la necesidad de situar y localizar RPs. La asignación de direcciones se hace empleando las direcciones del Core y del grupo multicast para identificar exclusivamente a un grupo.

Se ha hecho un estudio de los protocolos y de la interconexión de redes de área extensa, quedando por examinar los mecanismos que afectan a las redes internas y mas específicamente a las redes conmutadas, como en el caso de la Red de Datos de la Universidad del Cauca; por lo tanto, el siguiente capítulo describe multicast en las redes nivel 2.

# 6 MULTICAST EN REDES CONMUTADAS

Las redes conmutadas o llamadas también redes de nivel 2 del modelo OSI, se han convertido en los últimos años en la alternativa más utilizada en la construcción de redes de área local. Switches que se interconectan entre sí a través de líneas de alta velocidad y que también interconectan otros dispositivos, como hubs, PCs o switches; es una solución económica y atractiva para empresas e instituciones que desean conformar una red de computadores.

Este capítulo esta destinado a estudiar el comportamiento del tráfico multicast en los dispositivos de red que trabajan a nivel 2, sus problemas y como solucionarlos.

## 6.1 FUNCIONAMIENTO DE LOS SWITCHES LAN

Cuando una trama MAC llega al puerto de un switch, éste observa la dirección MAC de destino y la relaciona con la tabla de reenvío MAC que posee para determinar el puerto por el cual se debe reenviar la trama. Para que un switch pueda construir su tabla de reenvío tiene que aprender las direcciones MAC de las estaciones y los puertos a los cuales están conectadas. Esto lo hace examinando las direcciones de origen de las tramas MAC enviadas entre estaciones de la LAN y anotando el puerto del switch por el que fueron recibidas.

Cuando los switches empezaron a salir al mercado, la CPU principal del switch se encargaba de todas las tareas. Cuando se recibía una trama en un puerto, la CPU examinaba la dirección de destino y la comparaba con la tabla de reenvió para transmitirla por el puerto correcto. Para aumentar la velocidad de este proceso,

la mayoría de estos switches almacenaban la tabla de reenvío en una memoria llamada CAM (Content-Addressable Memory) que permitía usar la dirección MAC como puntero hacia la salida correspondiente.

A medida que creció la demanda de un mayor rendimiento en los switches, se decidió disminuir la carga del trabajo de la CPU principal y construir algún tipo de dispositivo de conmutación especial que pudiera acceder a la CAM y conmutar las tramas desde su puerto de entrada al de salida a la velocidad a la que se trasmite la información.

Para lograr que la velocidad de conmutación fuera igual a la velocidad a la que los paquetes se transmiten, la mayoría de los switches emplean un dispositivo de conmutación basado en circuitos integrados de aplicación específica, llamados ASIC (Application-Specific Integrated Circuits). Estos ASICs pueden examinar la dirección MAC de destino de una trama, buscarla en la tabla CAM y conmutar la trama al puerto de salida indicado. Ahora los switches cuentan con un puerto interno que lo comunican con el dispositivo de conmutación.

## 6.2 INUNDACIÓN DEL TRÁFICO BROADCAST Y MULTICAST

Si un switch recibe una trama sin ninguna concordancia de su dirección de destino en la tabla CAM, no tiene otra alternativa más que reenviar la trama a todos los puertos con la esperanza de que la trama alcance el destino correcto. Este comportamiento generalmente ocurre en las siguientes situaciones:

- La dirección MAC de destino no ha sido aprendida aún;
- La dirección MAC de destino es una dirección broadcast o multicast.

La primera situación no es un problema ya que el switch adicionará esta dirección MAC en la tabla CAM en el momento en que la estación transmita una trama; desde este momento todas las transmisiones hacia esta estación no inundarán toda la red.

En la segunda situación, las tramas broadcast deben ser retransmitidas a todos los puertos, a excepción del entrante. De la misma manera el switch no tiene ninguna manera de saber en que puertos se encuentran miembros de un grupo multicast por lo que las tramas multicast deben inundar todos los puertos del switch. Este comportamiento de los switches es el principal problema de las redes conmutadas respecto al tráfico multicast ya que inundará todos los rincones de la red, aún si no es deseado, consumiendo todos sus recursos.

## 6.3 REDUCIENDO LA INUNDACIÓN DEL TRÁFICO MULTICAST

A medida que las redes conmutadas crecen en popularidad, se han hecho esfuerzos para controlar la inundación de las tramas multicast/broadcast. El primer intento fue introducir un límite en la velocidad de transmisión, es decir, colocar un límite configurable al ancho de banda consumido antes de que las tramas empezaran a ser descartadas. Esta no fue la mejor solución a este problema, ya que en el momento de empezar a descartar indiscriminadamente tramas multicast/broadcast puede resultar en la inestabilidad de la red, que en algunos casos puede llegar a tumbarla.

El segundo intento, fue extender el formato de la tabla CAM para permitir que un grupo de puertos estuviera asociado con una dirección MAC. La idea es configurar manualmente un switch para relacionar un rango de direcciones multicast a un grupo de puertos. Ahora, cuando el switch recibe una trama multicast que concuerda con una dirección en la tabla CAM, la trama sólo se reenvía a los puertos indicados manualmente, reduciendo la inundación solo a algunos sectores de la red. Esta solución funcionó bien inicialmente para algunos grupos multicast estáticos, como el grupo de todos los enrutadores OSPF, en el cual no se esperan cambios frecuentes en su ubicación. Desafortunadamente este mecanismo no es muy útil en tratamiento de grupos multicast en donde las direcciones se asignan dinámicamente.

De esta manera se deben utilizar otros mecanismos para evitar la inundación multicast:

- IGMP snooping;
- CGMP (Cisco Group Management Protocol);
- GARP (Generic Attribute Resolution Protocol) de la IEEE especificado en el estándar 802.1p

De estos tres métodos los dos primeros han sido los más utilizados; CGMP es un protocolo propietario de Cisco, por lo que sólo se tratará IGMP Snooping.

## 6.4 IGMP SNOOPING

La solución más lógica para reducir la inundación multicast es que el switch vigile el diálogo IGMP entre el host y el enrutador. Cuando el switch escucha un IGMP Report de un host para un grupo multicast, este agrega el puerto del host a la entrada correspondiente de la tabla CAM. Cuando el switch escucha un mensaje IGMP Leave de un host se remueve el puerto donde se recibió el mensaje de la entrada correspondiente de la tabla CAM.

Aparentemente esta solución parece ser bastante simple para poner en práctica, sin embargo, dependiendo de la arquitectura del switch, IGMP snooping puede ser difícil de implementar sin degradar el desempeño del switch.

### 6.4.1 Unirse a un grupo utilizando IGMP Snooping

La figura 6.1 muestra una topología muy común en las redes de área local. A continuación se examinará lo que sucede en un switch cuando un host se une a un grupo multicast por medio de la conversación IGMP con un enrutador.

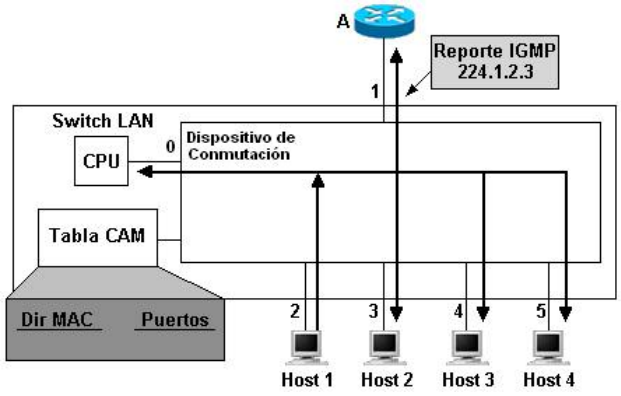


**Figura 6.1 Funcionamiento IGMP Snooping en un switch**

1. El host 1 desea unirse al grupo multicast 224.1.2.3 y por lo tanto multidifunde un mensaje IGMP Membership Report no solicitado con la dirección MAC de destino 0x0100.5E01.0203. Como inicialmente no hay entradas en la tabla CAM para esta dirección MAC, se envía el mensaje a todos los puertos del switch.
2. Cuando la CPU del switch recibe el reporte IGMP enviado por el host 1, la CPU utiliza la información de este mensaje para crear una entrada en la tabla CAM que incluye el puerto del host 1 (puerto 2), el puerto del enrutador (puerto 1) y el puerto interno de la CPU del switch (puerto 0).

Como consecuencia, de ahí en adelante todas las tramas dirigidas a la dirección MAC 0x0100.5E01.0203 serán reenviadas sólo a los puertos 0, 1 y 2 del switch.

Si el host 4 quisiera unirse al grupo multicast y envía un mensaje IGMP Report no solicitado, el switch reenvía la trama a los puertos 1 y 2 basado en la tabla CAM. Como la CPU del switch también recibe el mensaje IGMP Report, éste añade el puerto, por el que fue escuchado el Report (puerto 5), a la entrada correspondiente de la tabla CAM. De esta manera, cualquier paquete con dirección destino 0x0100.5E01.0203 será enviado únicamente a los host 1, 2 y 4, al enrutador y a la CPU del switch.

### 6.4.2 Abandonar un grupo a través de IGMP Snooping

Si el host 1 de la figura 6.1 quisiera dejar el grupo, se llevaría a cabo el siguiente procedimiento:

1. El host 1 envía un mensaje IGMP Leave Group a la dirección del grupo de todos los enrutadores multicast 224.0.0.2 (dirección MAC 0x0100.5E00.0002). La CPU del switch no reenvía este mensaje a ningún otro puerto del switch. Los mensajes IGMP Leave Group son los únicos cuya dirección de destino no es la del grupo objetivo, en este caso un switch nivel 2 puede utilizar una entrada en la tabla CAM para que intercepte este tipo de mensajes IGMP porque sólo estos mensajes están destinados a la dirección MAC 0x0100.5E00.0002.
2. La CPU del switch responde a este mensaje enviando un IGMP General Query al puerto 2 para averiguar si hay algún otro host miembro de ese grupo en

este puerto (en el caso de que varios hosts estén conectados a un mismo puerto a través de un hub).

3. Si se recibe algún IGMP Report de otro host conectado al puerto 2 entonces la CPU descarta el mensaje IGMP Leave Group del host 1. Si por el contrario no se recibe ningún mensaje IGMP Report en este puerto, la CPU borra el puerto de la entrada correspondiente de la tabla CAM. Como todavía existen puertos activos (que pertenecen a miembros no enrutadores) en la entrada de la tabla CAM para este grupo, no se envía ningún mensaje al enrutador.
4. Ahora el host 4 desea abandonar el grupo mandando un mensaje IGMP Leave Group.
5. De nuevo el mensaje es interceptado por la CPU del switch y responde mandando un mensaje IGMP General Query por el puerto 5.
6. Si el host 4 es el único miembro de este grupo en el puerto 5 ningún IGMP Report será recibido para este grupo, por lo que el switch borrará el puerto 5 de la entrada correspondiente de la tabla CAM. Como este era el último puerto de un miembro no enrutador para la dirección 0x0100.5E00.0002 la CPU del switch borra la entrada de la tabla CAM para este grupo y envía un mensaje IGMP Leave Group al puerto del enrutador para que tome las medidas necesarias.

### 6.4.3 Funcionamiento general de IGMP Snooping

En algunas ocasiones el proceso de abandono de grupo no funciona exactamente como se describió en la sección anterior debido a varios factores como:

- No todos los hosts envían un mensaje IGMP Leave Group ya que el RFC de IGMPv2 no especifica que sea obligatorio, como se describió en el capitulo 2.
- El mensaje IGMP Leave Group puede perderse debido a congestiones en la red; IGMP es un protocolo no confiable por lo que no devuelve mensajes ACK que confirmen la recepción de paquetes.
- Los hosts que utilizan la versión 1 de IGMP no hacen uso del mensaje IGMP Leave Group.

Esta clase de comportamiento se maneja de la misma manera como se hace convencionalmente, esperando que el enrutador envíe un mensaje IGMP General Query sin respuesta por parte de los hosts y se desencadene el mecanismo explicado en el capitulo 2.

Por otro lado, volviendo al ejemplo de la figura 6.1 y asumiendo que los host 1 y 4 se han unido al grupo multicast 224.1.2.3, el funcionamiento general del switch sería como se describe a continuación:

1. El enrutador periódicamente multidifunde un mensaje General Query al grupo de todos los host 224.0.0.1 (dirección MAC 0x0100.5E00.0001). La CPU del switch intercepta el mensaje General Query y lo reenvía a todos los puertos.
2. Todos los hosts que son miembros de un grupo (en este ejemplo los host 1 y 4) envían un IGMP Report. La CPU del switch intercepta todos los mensajes IGMP y por tal motivo los hosts no pueden escuchar los mensajes IGMP Report de los demás lo que elimina el mecanismo de supresión de reporte. Esto se hace necesario para que la CPU del switch sepa en que puertos hay miembros del grupo y pueda mantener correctamente la tabla CAM.
3. Para mantener el estado de pertenencia a un grupo activo en el enrutador, el switch tiene que enviar uno o mas (preferiblemente uno) mensajes IGMP Report hacia el enrutador.

### 6.4.4 Impacto en el rendimiento de un switch que utiliza IGMP Snooping

Como se mencionó anteriormente, el switch posee un puerto interno que permite la comunicación del dispositivo de conmutación con la CPU principal. Este puerto está presente en la tabla CAM para todas las entradas que posean una dirección multicast para que la CPU pueda escuchar todos los mensajes IGMP entre los hosts y el enrutador; todos los paquetes que posean una dirección multicast serán enviados al puerto interno del switch y la CPU tendrá que procesar todas las tramas para saber cuales son mensajes IGMP y cuales no, lo que causa un impacto en el desempeño del switch.

La mayoría de los switches de nivel 2 sufren una reducción drástica en su rendimiento al implementar IGMP snooping que conduce a la pérdida ocasional de tráfico multicast y unicast. En otros casos el dispositivo de conmutación continúa enviando el tráfico multicast y unicast sin ningún tipo de pérdida, sin embargo, la CPU principal empezará a descartar paquetes que pueden resultar en mensajes IGMP perdidos afectando la latencia en el abandono y en la unión de los hosts a un grupo.

Para solucionar este problema es necesario rediseñar el dispositivo de conmutación para que pueda usar nuevos ASICs y tablas CAM que puedan analizar en más detalle las tramas y examinar información a nivel de la capa 3, entregándole a la CPU sólo las tramas que posean mensajes IGMP. Switches que utilizan este tipo de dispositivos se conocen como Switches nivel 3.

### 6.4.5 Detección de enrutadores con IGMP Snooping

En los ejemplos anteriores el puerto al que el mrouter esta conectado se adiciona automáticamente en la tabla CAM para las direcciones multicast, esto debido a que el enrutador tiene que recibir indiscriminadamente el tráfico multicast de todos los grupos. El switch que utiliza IGMP snooping debe tener un mecanismo para saber en que puerto está conectado el mrouter.

Un switch que utiliza IGMP snooping debe escuchar los menajes IGMP General Query de los enrutadores y recordar el puerto por el cual fueron escuchados; pero no es suficiente, en el caso de que exista más de un mrouter sólo uno de ellos actuará como IGMP Querier, mientras los otros estarán en modo pasivo, lo que indica que no enviarán ningún tipo de mensaje por lo que el switch creerá que no existe otro enrutador presente. Los enrutadores non-Querier necesitan escuchar toda la información igual que el enrutador designado si se desea tener un mecanismo de redundancia, pero si el switch no detecta a través de los mensajes IGMP que existe un enrutador presente, sencillamente no enviará ningún tipo de información al puerto del enrutador non-Querier.

Es por esto que el switch además de escuchar los mensajes IGMP debe escuchar paquetes especiales de protocolos de enrutamiento que puedan indicar la presencia de mrouters. En algunos switches se interceptan mensajes Hello de los protocolos PIMv1 y PIMv2, pruebas de vecinos DVMRP, IGMP Queries y mensajes CGMP, entre otros, que pueden fácilmente indicar la presencia de enrutadores multicast.

Hasta el momento se ha hecho un estudio del estado del arte de multicast en IPv4, el siguiente capítulo describe el futuro de multicast en IPv6 y su backbone.

# 7 MULTICAST EN IPv6

Debido a la preocupación reciente por el agotamiento inminente del conjunto actual de direcciones de Internet y el deseo de proporcionar funcionalidad adicional para dispositivos modernos, se encuentra en proceso de normalización una actualización de la versión del Protocolo Internet IPv4. La nueva versión, denominada IP versión 6 (IPv6), resuelve problemas de diseño no previstos en IPv4. En este capítulo se va a hacer especial énfasis en el área multicast en Ipv6.

## 7.1 ESPACIO DE DIRECCIONES DE IPV6

La característica distintiva más evidente de IPv6 es el uso de direcciones mucho mayores. El tamaño de una dirección en IPv6 es de 128 bits, cuatro veces mayor que el de una dirección de IPv4. El espacio de direcciones de 32 bits permite hasta 4.294.967.296 direcciones. Un espacio de direcciones de 128 bits permite hasta 340.282.266.920.938.463.463.374.607.431.768.211.465 (o $3{,}4 \times 10^{38}$) direcciones. El tamaño relativamente grande de una dirección IPv6 se ha diseñado así para que se pueda subdividir en dominios de enrutamiento jerárquico que reflejen la topología de Internet actual. El uso de 128 bits permite varios niveles de jerarquía y ofrece flexibilidad para diseñar un enrutamiento y un direccionamiento jerárquico, algo que actualmente no ofrece la tecnología IPv4. La arquitectura de direccionamiento de IPv6 se describe en RFC 2373.

De modo similar al que se utiliza para dividir el espacio de direcciones de IPv4, el espacio de direcciones de IPv6 se divide según el valor de los bits de orden superior. Los bits de orden superior y su valor fijo se conocen como prefijo de formato (FP, Format Prefix). En la tabla 7.1 se muestra la asignación del espacio de direcciones de IPv6 por FP.

| Asignación | Prefijo de formato (FP) | Fracción del espacio de direcciones |
|---|---|---|
| Reservado | 0000 0000 | 1/256 |
| Sin asignar | 0000 0001 | 1/256 |
| Reservado para la asignación de NSAP | 0000 001 | 1/128 |
| Reservado para la asignación IPX | 0000 010 | 1/128 |
| Sin asignar | 0000 011 | 1/128 |
| Sin asignar | 0000 1 | 1/32 |
| Sin asignar | 0001 | 1/16 |
| Direcciones de unidifusión global agregables | 001 | 1/8 |
| Sin asignar | 010 | 1/8 |
| Sin asignar | 011 | 1/8 |
| Sin asignar | 100 | 1/8 |
| Sin asignar | 101 | 1/8 |
| Sin asignar | 110 | 1/8 |
| Sin asignar | 1110 | 1/16 |
| Sin asignar | 1111 0 | 1/32 |
| Sin asignar | 1111 10 | 1/64 |
| Sin asignar | 1111 110 | 1/128 |
| Sin asignar | 1111 1110 0 | 1/512 |
| Direcciones de unidifusión local de vínculo | 1111 1110 10 | 1/1024 |
| Direcciones de unidifusión local de sitio | 1111 1110 11 | 1/1024 |
| Direcciones de multidifusión | 1111 1111 | 1/256 |

**Tabla 7.1 Asignación de direcciones IPv6**

## 7.2 TIPOS DE DIRECCIONES DE IPV6

En el rango de direcciones de IPv6 se distinguen tres tipos de direcciones, las direcciones broadcast ya no se utilizan, fueron reemplazadas por las direcciones multicast y específicamente por la dirección de todos los hosts dándole mayor importancia a este rango de direcciones; además, se introduce un nuevo concepto de difusión: Anycast.

- **Unicast**. Una dirección unicast identifica a una sola interfaz. Con la topología de enrutamiento de unicast apropiada, los paquetes dirigidos a una dirección de unicast se entregan a una sola interfaz. Para ajustarse a los sistemas de equilibrio de carga, el RFC 2373 permite que varias interfaces utilicen la misma dirección, siempre y cuando las distintas interfaces aparezcan como una sola interfaz para la implementación de IPv6 en el host.

- **Multicast**. Una dirección multicast identifica varias interfaces. Con la topología de enrutamiento multicast apropiada, los paquetes dirigidos a una dirección multicast se entregan a todas las interfaces identificadas por la dirección.
- **Anycast**. Una dirección anycast identifica varias interfaces. Los paquetes dirigidos a una dirección anycast se entregan a una sola interfaz, la más próxima que identifica la dirección. La interfaz "más próxima" se define como la más cercana en términos de distancia de enrutamiento. Una dirección multicast se utiliza para la comunicación "de uno a muchos", con entrega a varias interfaces. Una dirección anycast se utiliza para la comunicación "de uno a uno de muchos", con entrega a una sola interfaz.

En todos los casos, las direcciones IPv6 identifican interfaces, no nodos. Un nodo se identifica mediante cualquier dirección unicast asignada a una de sus interfaces.

***Direcciones IPv6 multicast***

En IPv6, el envío de tráfico multicast funciona del mismo modo que en IPv4. Los nodos IPv6 ubicados arbitrariamente pueden atender al tráfico multicast en una dirección multicast IPv6 arbitraria. Los nodos IPv6 pueden escuchar a varias direcciones multicast simultáneamente. Los nodos pueden unirse a un grupo multicast o abandonarlo en cualquier momento.

Las direcciones multicast utilizan el FP 11111111 (hxFF). Las direcciones multicast no se pueden utilizar como direcciones de origen o como destinos intermedios en el encabezado de un protocolo de enrutamiento.

Además del FP, las direcciones multicast incluyen una estructura adicional para identificar sus indicadores, ámbito y grupo multicast. En la figura 7.1 se muestra la dirección multicast IPv6.

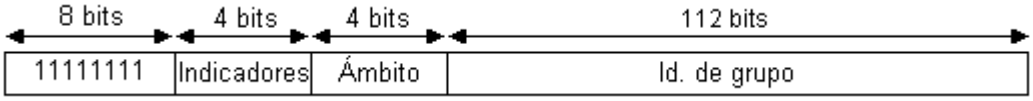


**Figura 7.1 Dirección multicast IPv6**

Los campos del encabezado son los siguientes:

**Flags (Indicadores)**: muestra los indicadores establecidos en la dirección. El tamaño de este campo es de 4 bits. Según el RFC 2373, el único indicador definido es el indicador de provisionalidad, Transient (T). El indicador T utiliza el bit de orden inferior del campo Flags. Cuando se establece en el valor 0, el indicador T indica que la dirección multicast es una dirección asignada de forma definitiva (bien conocida) por la IANA. Cuando se establece en el valor 1, el indicador T especifica que la dirección multicast es transitoria (no está asignada definitivamente).

**Scope (Ámbito)**: indica el ámbito de la red interna de IPv6 para la que está previsto el tráfico multicast. El tamaño de este campo es de 4 bits. Además de la información proporcionada por los protocolos de enrutamiento multicast, los enrutadores utilizan el ámbito multicast para determinar si se puede reenviar el tráfico multicast. En la tabla 7.2 se muestran los valores definidos para el campo Scope.

| Valor | Ámbito |
|---|---|
| 0 | Reservado |
| 1 | Ámbito local de nodo |
| 2 | Ámbito local de enlace |
| 5 | Ámbito local de sitio |
| 8 | Ámbito local de organización |
| E | Ámbito global |
| F | Reservado |

**Tabla 7.2 Valores definidos para el campo Scope**

Por ejemplo, el tráfico con la dirección multicast FF02:: tiene un ámbito local de enlace. Un enrutador IPv6 nunca reenvía este tráfico más allá del enlace local.

**Id. de grupo**: identifica el grupo multicast y es único en el ámbito. El tamaño de este campo es de 112 bits. Los identificadores de grupo asignados definitivamente son independientes del ámbito. Los identificadores de grupo transitorios sólo son relevantes para un ámbito determinado. Las direcciones multicast comprendidas entre FF01:: y FF0F:: son direcciones bien conocidas y reservadas. Para identificar todos los nodos de los ámbitos locales de nodo y de interfaz, se definen las siguientes direcciones:

- FF01::1, dirección multicast para todos los nodos del ámbito local de nodo.
- FF02::1, dirección multicast para todos los nodos del ámbito local de interfaz.

Para identificar todos los enrutadores de los ámbitos locales de nodo, de interfaz y de sitio, se definen las siguientes direcciones:

- FF01::2, dirección multicast para todos los enrutadores del ámbito local de nodo.
- FF02::2, dirección multicast para todos los enrutadores del ámbito local de interfaz.
- FF05::2, dirección multicast para todos los enrutadores del ámbito local de sitio.

Con 112 bits en el Id. de grupo, es posible tener $2^{112}$ identificadores. Sin embargo, debido a la forma en la que las direcciones multicast IPv6 se asignan a las direcciones MAC multicast Ethernet, el RFC 2373 recomienda asignar el identificador de grupo a partir de los 32 bits de orden inferior de la dirección multicast IPv6 y establecer en cero los demás bits del identificador de grupo original. Al utilizarse únicamente los 32 bits de orden inferior, cada identificador de grupo se asigna a una dirección MAC multicast Ethernet única. En la figura 7.2 se muestra la dirección multicast IPv6 modificada.

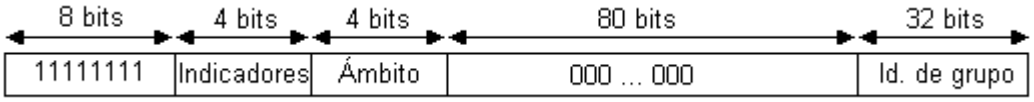


**Figura 7.2 Dirección multicast IPv6 modificada con un Id. de grupo de 32 bits.**

## 7.3 DESCUBRIMIETNO DE ESCUCHA MULTICAST

Multicast Listener Discovery (MLD) es el equivalente en IPv6 de la versión 2 del Protocolo de gestión de grupos de Internet (IGMPv2) para IPv4. MLD es un conjunto de mensajes que se intercambian enrutadores y nodos, que permite a los enrutadores descubrir el conjunto de direcciones multicast para las que hay nodos interesados en cada interfaz conectada. Al igual que IGMPv2, MLD sólo descubre la lista de direcciones multicast para las que hay al menos un host interesado, no la lista de hosts multicast para cada dirección multicast. MLD está documentado en el RFC 2710.

A diferencia de IGMPv2, MLD utiliza mensajes ICMPv6 en vez de definir su propia estructura de mensajes. Todos los mensajes MLD son mensajes ICMPv6 de los tipos 130, 131 y 132. Los tres tipos de mensajes MLD son:

- **Consulta de escucha multicast (Multicast Listener Query).** Los enrutadores utilizan los mensajes Multicast Listener Query para consultar en una interfaz los hosts multicast. Existen dos tipos de mensajes Multicast Listener Query: ***Consulta general (General Query)*** y ***Consulta específica de dirección multicast (Multicast-Address-Specific Query)***. El mensaje General Query se utiliza para consultar a los hosts multicast de todas las direcciones multicast. El mensaje Multicast-Address-Specific Query se utiliza para consultar hosts multicast de una dirección multicast específica. Estos dos tipos de mensajes se distinguen mediante la dirección de destino multicast en el encabezado IPv6 y una dirección multicast en el mensaje Multicast Listener Query.
- **Reporte de host multicast (Multicast Listener Report).** Un host multicast utiliza estos reportes para informar el interés por recibir tráfico multicast para una dirección multicast determinada o para responder a un mensaje Multicast Listener Query.
- **Escucha multicast terminada (Multicast Listener Done).** Un host multicast utiliza Multicast Listener Done para informar que ya no está interesado en recibir tráfico multicast para una dirección multicast determinada.

El paquete de un mensaje MLD consta de un encabezado IPv6, un encabezado de extensión Opciones de salto a salto (Hop-by-Hop Options) y el mensaje MLD. El encabezado de extensión de Opciones de salto a salto contiene la opción Alerta de enrutador (Router Alert) de IPv6, documentada en RFC 2711, que se utiliza para asegurar que los enrutadores procesan los mensajes MLD enviados a direcciones multicast en las que el enrutador no está interesado. En la figura 7.3 se muestra el formato de un paquete de mensaje MLD.

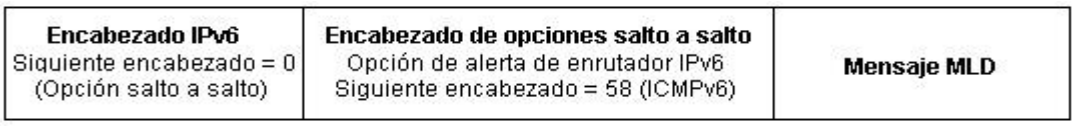


**Figura 7.3 Formato de un paquete de mensaje MLD**

### 7.3.1 Consulta de escucha multicast (Multicast Listener Query)

Un mensaje MLD Multicast Listener Query equivale al mensaje IGMPv2 Host Membership Query. Lo utiliza un enrutador para consultar el interés de los hosts en un grupo multicast.

En el encabezado IPv6, la dirección de origen es la dirección local de la interfaz en la que se envía la consulta. El campo Límite de saltos (Hop Limit) se establece en el valor 1. Para General Query, la dirección de destino es la dirección multicast de todos los nodos de ámbito local de enlace (FF02::1). Para Multicast-Address-Specific Query, la dirección de destino es la dirección multicast específica que se consulta.
En la figura 7.4 se muestra el mensaje MLD, donde el campo Tipo es el que indica el tipo de mensaje MLD.

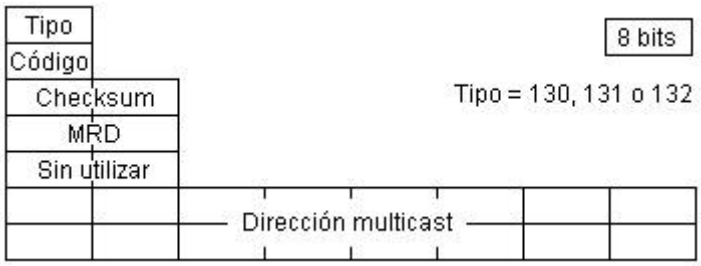


**Figura 7.4 Mensaje MLD**

En el mensaje MLD Multicast Listener Query, el campo Tipo (Type) se establece en el valor 130 y el campo Código (Code) se establece en 0. Después del campo Suma de comprobación (Checksum), se encuentran los campos de 16 bits Retardo máximo de respuesta (Maximum Response Delay) y Sin utilizar (Reserved). Maximum Response Delay especifica la cantidad de tiempo máxima en milisegundos en la que un miembro del grupo multicast debe informar de su pertenencia al grupo mediante un mensaje MLD Multicast Listener Report. En General Query, el campo Dirección multicast (Multicast Address) se establece en la dirección no especificada (::). En Multicast-Address-Specific Query, el campo Multicast Address se establece en la dirección multicast específica que se consulta.

### 7.3.2 Informe de escucha multicast (Multicast Listener Report)

Un mensaje MLD Multicast Listener Report equivale al mensaje IGMPv2 Host Membership Report (Pertenencia a un grupo de hosts). Los hosts lo utilizan para informar de su interés en recibir tráfico multicast en una dirección multicast específica o responder a un mensaje MLD General Query o Multicast-Address-Specific Query. En el encabezado IPv6, la dirección de origen es la dirección local de la interfaz en la que se envía el informe. El campo Límite de saltos (Hop Limit) se establece en el valor 1 y la dirección de destino es la dirección multicast sobre la que trata el informe.

En el mensaje MLD Multicast Listener Report, el campo Tipo se establece en 131 y el campo Código se establece en el valor 0. El campo MRD no se utiliza en un mensaje Multicast Listener Report y se establece en 0. El campo Dirección multicast se configura con la dirección multicast específica sobre la que trata el informe.

### 7.3.3 Escucha multicast terminada (Multicast Listener Done)

Un mensaje MLD Multicast Listener Done equivale al mensaje IGMPv2 Leave Group. Lo utiliza un host para informar a los enrutadores locales que ya no está interesado en una dirección multicast específica.

En el encabezado IPv6, la dirección de origen es la dirección de la interfaz en la que se envía el informe. El campo Límite de saltos se establece en el valor 1 y la dirección de destino es la dirección multicast de todos los enrutadores de ámbito local de enlace (FF02::2).

En el mensaje MLD Multicast Listener Done, el campo Tipo se establece en el valor 132 y el campo Código se establece en el valor 0. El campo MRD no se utiliza en un mensaje Multicast Listener Done y se establece en 0. El campo Dirección multicast se configura con la dirección multicast específica para la cual el host que ya no está interesado.

## 7.4 EL BACKBONE MULTICAST DE IPv6 – M6BONE

El M6Bone es una red de prueba multicast que empezó en julio del 2001 con el soporte de la Asociación Aristóteles, el Grupo G6 y la red Renater (Francia). La meta de este proyecto es la de ofrecer conexión multicast sobre IPv6 a los sitios interesados para probar y desarrollar software y equipos relacionados con la tecnología IPv6 multicast.

La topología del M6Bone consta hoy en día de más de 25 nodos, desafortunadamente hay muy pocos equipos que implementan los protocolos de enrutamiento multicast sobre IPv6 lo que hace que sea una red superpuesta. Se utilizan túneles IPv6/IPv6 e IPv6/Pv4 para conectar a los enrutadores del M6Bone. La figura 7.5 muestra la topología mundial del M6Bone, el RP está ubicado en el enrutador multicast de Renater, que también actúa como enrutador Bootstrap[*] para el dominio; Un segundo RP de menor prioridad en el CSC/Funet actúa como RP de redundancia.

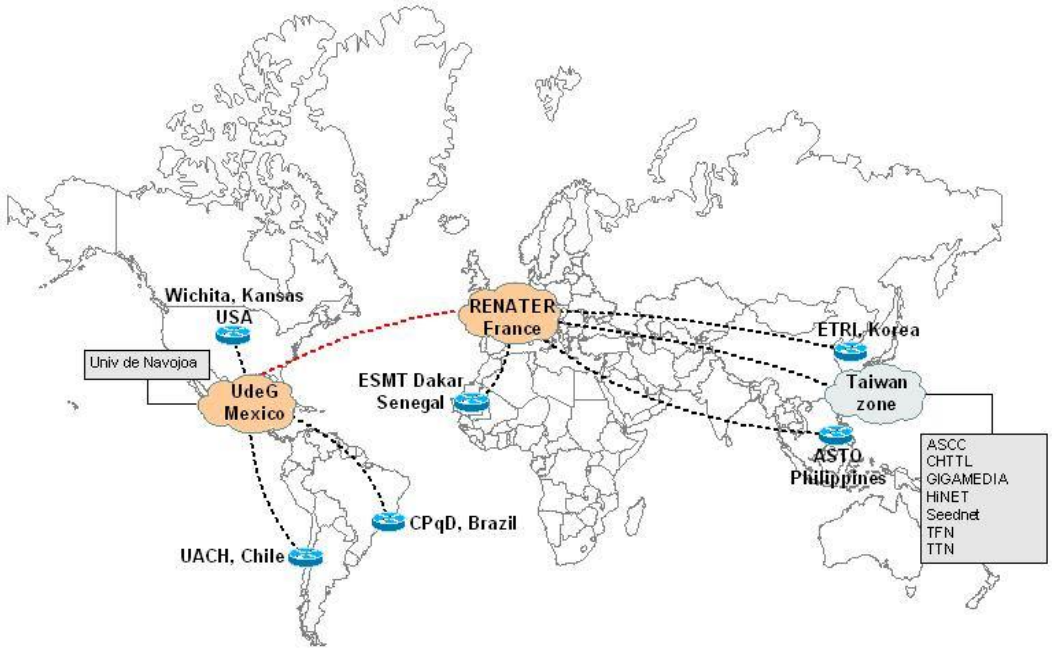


**Figura 7.5 Mapa mundial del M6Bone**

[*] El enrutador Bootstrap es el encargado de manejar el mecanismo de elección del RP por medio de la prioridad mas alta, si dos enrutadores tienen la misma prioridad se escoge el de la dirección IP mas alta; esta información debe ser difundida por el enrutador Bootstrap a todos los enrutadores de la red.

Muchos de los participantes del M6Bone son miembros del proyecto 6NET. Sin embargo, mientras que 6NET desarrolló una red IPv6 nativa incluyendo a 15 Redes de Educación e Investigación Nacionales a mediados de 2002, no soportaba multicast nativo en IPv6 (a la espera de implementaciones de PIM-SM en enrutadores Cisco de la serie GSR utilizados en el backbone de 6NET).

Por lo que fue necesario construir una red IPv6 multicast de prueba que utilizara las implementaciones multicast de IPv6 existentes. La red utiliza PIM-SM, que está implementado para sistemas BSD, Hitachi (en el GR2000) y 6WIND (en el 6WINDGate 6200).

El M6Bone utiliza una topología diferente para tráfico IPv6 multicast y unicast; cuando un enrutador IPv6 multicast recibe un paquete, lleva a cabo un chequeo RPF. Si las topologías IPv6 unicast y multicast no son las mismas, el protocolo de enrutamiento multicast IPv6 realiza el chequeo RPF utilizando la tabla de enrutamiento multicast. Debido a que todavía no se ha implementado este mecanismo es necesario realizar el chequeo RFP utilizando la tabla de enrutamiento unicast IPv6, por lo que las topologías IPv6 para unicast y multicast deben ser las mismas.

Hasta el momento la solución a este inconveniente es tener un equipo de enrutamiento dedicado para multicast IPv6. Estos enrutadores intercambian sus tablas de enrutamiento unicast que se utilizarán para el chequeo RPF empleando el protocolo RIPng, definido en los RFCs 2080 y 2081, cada sitio anuncia su prefijo correspondiente a la subred donde está habilitado multicast IPv6 a través de túneles. Esta política de enrutamiento hace posible anunciar en el M6Bone sólo los prefijos utilizados dentro del M6Bone, por lo que cada sitio tiene que decidir que prefijo(s) (/64) se utilizarán; otros prefijos se filtran en el enrutador core.

Los usuarios del M6Bone han usado clientes Linux, FreeBSD y Windows para las aplicaciones típicas multicast IPv6 como VIC, RAT y NTE, aunque hay otras nuevas aplicaciones IPv6 como radio MP3 e ISABEL.

Es muy importante monitorear y probar la red, para esto existe una página web del M6Bone que muestra el estado de los enlaces conectados directamente al RP,

la tabla de enrutamiento unicast y la tabla de enrutamiento multicast virtual. Además existen reflectores unicast-a-multicast IPv6 y reflectores multicast IPv4-a-IPv6, los primeros se utilizan para inyectar tráfico unicast IPv6 al M6Bone y los segundos para transmitir contenido multicast desde y hacia el mundo multicast IPv4.

Uno de los problemas más importantes es la ausencia de un protocolo multicast interdominio para IPv6, no hay MSDP para IPv6 y probablemente nunca lo habrá y BGMP está todavía en su fase inicial. Debido a esto, PIM-SSM (PIM Source Specific Mode, documentado en el RFC 3569 y los drafts de la IETF: draft-bhattach-pim-ssm-00.txt y draft-ietf-ssm-arch-04.txt) es el mejor camino para multicast IPv6, dado que el escenario de mayor aplicación para multicast es uno a muchos; aunque se ha propuesto un método para insertar la dirección del RP en la dirección multicast.

Se da por hecho que el ámbito local será un requerimiento común, por lo que es necesario contar con la característica de un enrutador Bootstrap. El campo de límite de la dirección IPv6 multicast se pude utilizar para tener diferentes RPs en diferentes ámbitos administrativos. El uso de múltiples RPs en diferentes partes del M6Bone está en estudio. Dentro de una LAN, MLD (sucesor de IGMP) Snooping será necesario en los switches para prevenir la inundación multicast en las redes jerárquicamente planas.

Otros problemas que se han originado son:

- Switches Ethernet antiguos que no soportan multicast IPv6.
- La falta de implementación de MLDv2 para IPv6 PIM-SSM.
- La inexistencia de tablas de enrutamiento multicast IPv6.

Aunque el M6Bone no es una red multicast IPv6 nativa, es muy útil para ayudar a determinar problemas que son independientes de las redes no nativas. La comunidad de usuarios de multicast IPv6 ha ido creciendo, utilizando aplicaciones y herramientas de monitoreo y ganando experiencia en la configuración de enrutadores y hosts.

Teniendo en cuenta todo el soporte teórico de los capítulos anteriores, a continuación se hace una descripción de los resultados obtenidos en el diseño y la implementación del backbone multicast de la Universidad del Cauca.

# 8 BACKBONE MULTICAST UNICAUCA

La red de datos de la Universidad del Cauca está basada en una topología de estrella que tiene su nodo principal en las instalaciones del IPET, de donde se extiende a través de fibra óptica hacia los diferentes edificios que conforman el campus universitario y también de donde sale la conexión a Internet a través de los proveedores: Telecom que provee un enlace de 2 Mbps y Orbitel con un enlace de 2 Mbps.

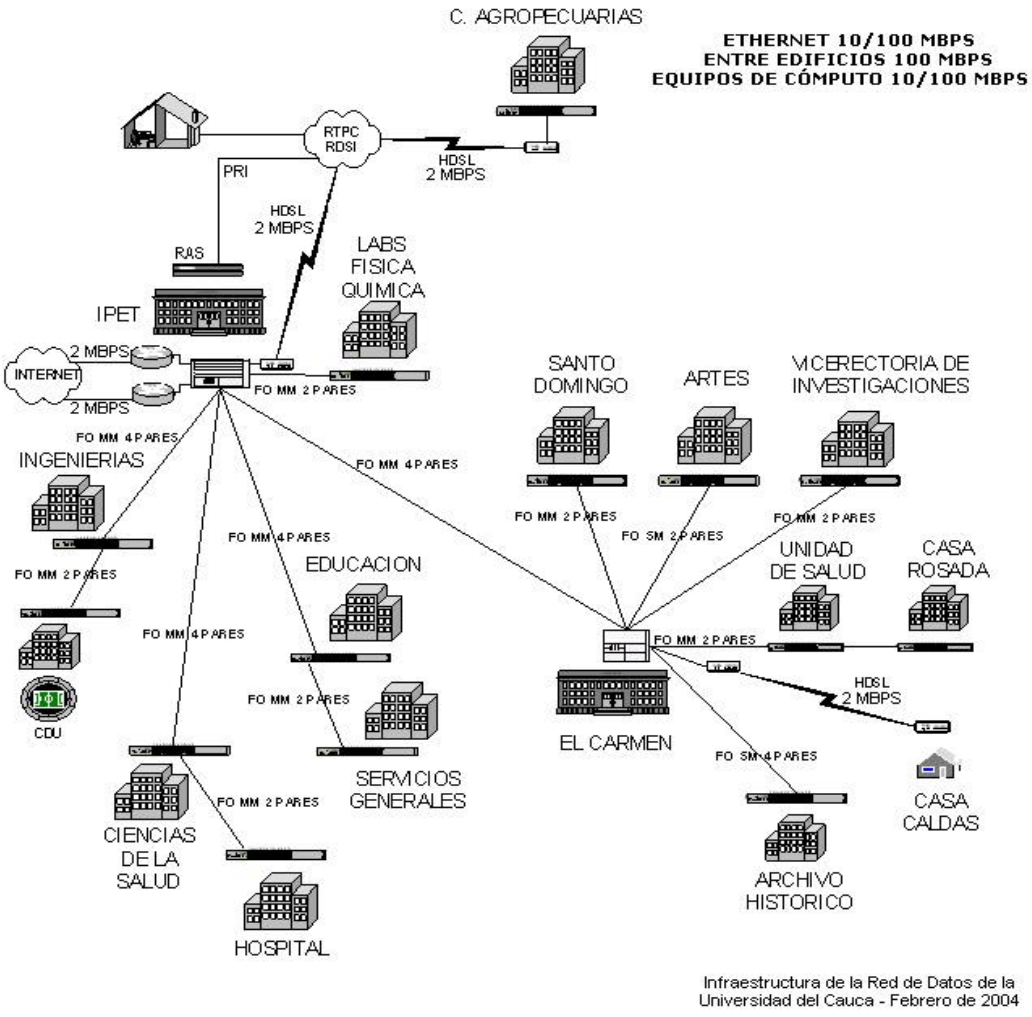


**Figura 8.1 Topología general Red Unicauca**

En la figura 8.1 se puede observar la topoogía general de la red de la Universidad del Cauca. Existen dos dependencias que se conectan a través de Módem HDSL que son La Casa Caldas y la Facultad de Ciencias Agropecuarias. Cada edificio de la Universidad posee un cableado estructurado basado en tecnología Fast Ethernet que provee un ancho de banda de 100 Mbps sobre UTP 5. La Universidad también presta el servicio de conexión a Internet domiciliario a través de un Servidor de Acceso Remoto – RAS.

La red de la Universidad cuenta con dos enrutadores Cisco, un 2509 con un IOS v 11.1 (24a) y otro 3600 con IOS , que dan soporte a la conexión Internet y al enrutamiento de las redes internas de la Universidad; conectado a estos hay un switch nivel 3 BayNetworks Accelar 1200 a donde llegan todas las conexiones de fibra óptica de los edificios. Esto se estudia con mayor detalle en el **Anexo A**.

La red de la Universidad es una LAN conmutada basada en el nivel 2 del modelo OSI por lo que cuenta con varios switches nivel 2, la mayoría de ellos Nortel BayStack 350-24T y 3Com 3300XM, además se cuenta con un Nortel BayNetworks 28200, un 3Com 4400, un BayStack 302F y un Encore. Para darle una mayor cobertura y aprovechar el ancho de banda de Fast Ethernet se utilizan varios concentradores a lo largo de toda la red para conectar a los usuarios finales. Información mas detallada sobre la topología de la red de la Universidad, sus equipos y sus conexiones se presenta en el **Anexo A.**

## 8.1 IGMP SNOOPING EN LA RED UNICAUCA

Los switches que conforman el backbone de la red de la Universidad pueden ser habilitados para trabajar con IGMP Snooping con algunas excepciones como por ejemplo el Switch BayStack 28200 ubicado en el edificio del Carmen, el BayNetworks 302-F, el Switch Encore que se encuentra en la Vicerrectoría de Investigaciones y los switches no gestionables como el que se encuentra en el edificio de Geotecnia.

El switch Accelar 1200 que se encuentra en las instalaciones del IPET provee capacidades adicionales para el tráfico IGMP e incluso capacidades de enrutamiento DVMRP.

En la práctica la capacidad de IGMP Snooping se habilitó únicamente en los switches del edificio de ingenierías, esto es, los que se encuentran en los centros de cableado 1, 3 y 15. La forma detallada y los pasos necesarios para configurar los diferentes switches de la red Unicauca para el funcionamiento de IGMP Snooping se muestran en el **Anexo B**.

## 8.2 CONEXIÓN AL MBONE DE INTERNET

Como se mencionó en el capítulo 5, en la actualidad el Mbone de Internet es una red de varios sistemas autónomos multicast que utilizan el protocolo PIM-SM para enrutamiento intradominio y los protocolos MBGP/MSDP para enrutamiento interdominio.

En el entorno de la Universidad del Cauca no hay ninguna otra organización que esté trabajando en el área de multicast, por lo que no es posible utilizar estos protocolos, siendo necesaria la implementación de protocolos más antiguos como el DVMRP y técnicas de entunelamiento para conectar la red de la Universidad al mundo multicast.

Para poder hacer parte del mundo multicast se necesitó la colaboración de una organización que hiciera parte del Mbone que aceptara la puesta en marcha de un túnel con la Universidad del Cauca.

**RETINA**, la Red Teleinformática de Argentina "es un proyecto de la *Asociación Civil Ciencia Hoy* que tiene como objetivo facilitar la integración de las redes académicas ya existentes y promover el uso de las nuevas tecnologías de la comunicación por parte de investigadores, docentes y personas vinculadas al ámbito académico. Las instituciones se integran a RETINA firmando un convenio en el cual se establece entre otras cosas el carácter cooperativo de la red y su uso con fines no comerciales".

La Universidad del Cauca en comunicación con el grupo trabajo del Mbone (Mbone Working Group) de la IETF y la NSRC (Network Startup Resource Center) de la Universidad de Oregon – EE.UU. pudo establecer un contacto con el personal de RETINA que ofrecieron su colaboración para que la Universidad del Cauca tuviera acceso al tráfico multicast mundial, que llega a todos los usuarios de Internet 2 y a las redes avanzadas de las Américas como REUNA de Chile, RNP de Brasil, CUDI de México, CANARIE de Canadá y del mismo modo a Europa por la red GEANT.

Para la implementación de este túnel la red Unicauca utiliza una máquina con el sistema operativo Linux y mrouted, un demonio de enrutamiento que ejecuta el algoritmo del protocolo DVMRP. Mrouted hace que la máquina sea el enrutador multicast en el extremo Unicauca del túnel y funciona como enrutador IGMP para la red interna de la Universidad. El enrutador IGMP es utilizado por los hosts en el mecanismo join/leave descrito en el capitulo 2 (Protocolo de Gestión de Grupos de Internet - IGMP) y por los switches que soporten el mecanismo de IGMP Snooping descrito en el capitulo 6. En el extremo RETINA del túnel se utiliza un enrutador Cisco de la serie 2600 con IOS 12.1 que hace parte de su red académica y esta conectado a Internet2. El enrutador Cisco utiliza el protocolo de enrutamiento multicast PIM-SM que le permite intercambiar información multicast y ser uno de los nodos en la estructura actual del Mbone; por una de sus interfaces se configura una interfaz virtual tipo túnel que utilice el protocolo DVMRP. Aunque este no es el método más eficiente y la arquitectura y las políticas del Mbone han cambiado bastante como se menciono en el capítulo 5, en el siguiente apartado se expone el diseño que se considera más conveniente para la implementación de la tecnología multicast en la evolución de la red de datos Unicauca.

Los enrutadores Cisco no trabajan con el protocolo DVMRP propiamente dicho, sino que utiliza una simulación del protocolo para cumplir ciertas funciones, como por ejemplo, la interoperabilidad PIM-DVMRP. Un enrutador Cisco encuentra un enrutador DVMRP por medio de los mensajes *Probe* del protocolo y cuando lo hace en una interfaz habilitada para el protocolo PIM automáticamente se habilita la interoperabilidad PIM-DVMRP.

El enrutador multicast (máquina linux) que se incluyó en la topología de la red de la Universidad se puede observar en la figura 8.2.

Para mayor información de la configuración de una máquina Linux como enrutador DVMRP y la configuración de un enrutador Cisco para soporte multicast y túneles DVMRP, consultar el **Anexo C**.

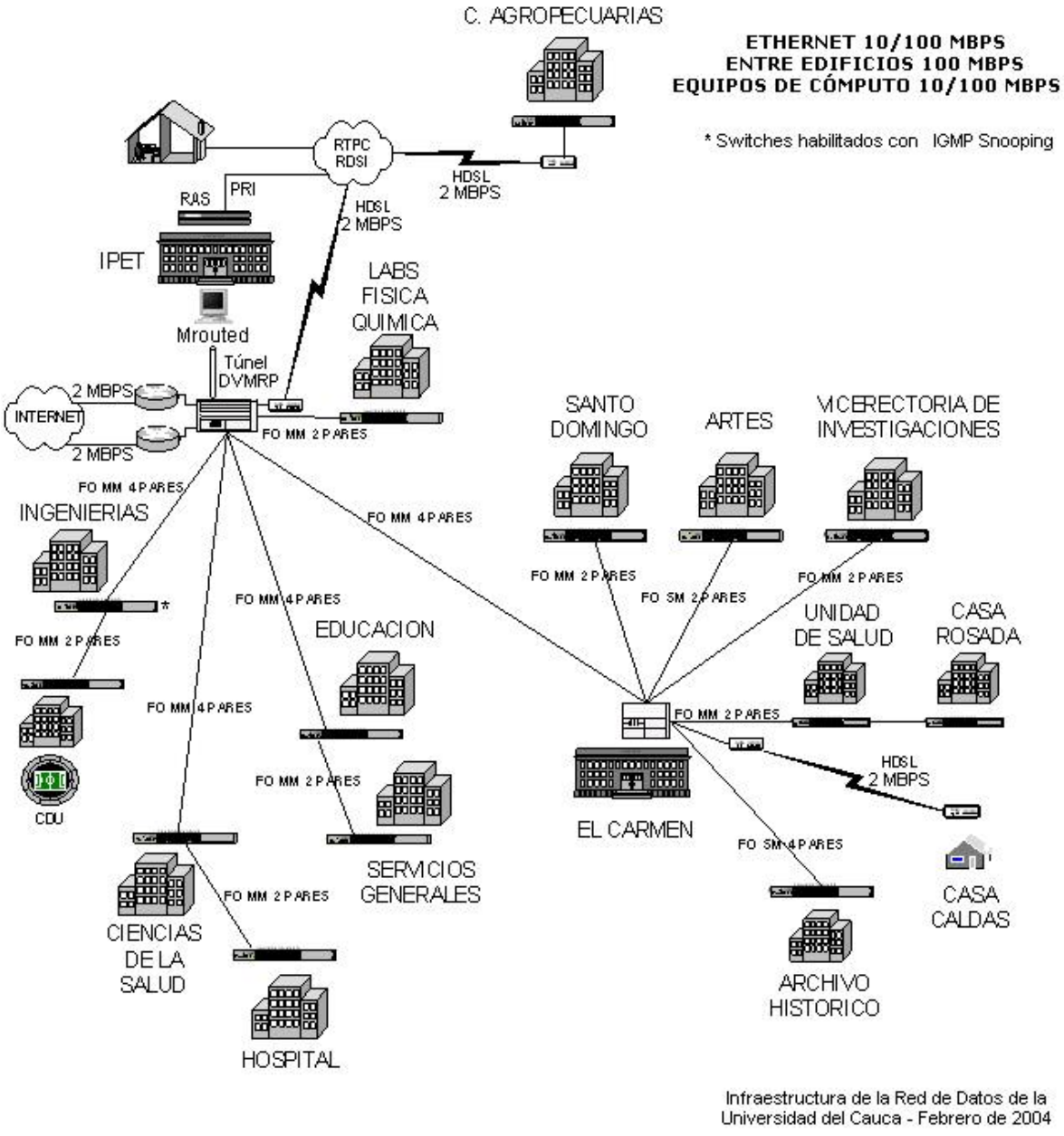


**Figura 8.2 Enrutador multicast en la red de datos Unicauca**

## 8.3 DISEÑO DEL BACKBONE MULTICAST DE UNICAUCA

La información recolectada sobre la topología de la red de datos de la Universidad del Cauca, el estado actual del backbone multicast de Internet y las nuevas tecnologías que se están utilizando y planeando a futuro para el uso de multicast impulsan cambios en la estructura de la red que se deben implementar tanto a nivel 2 como a nivel 3 del modelo de referencia OSI.

### 8.3.1 Evolución Red Unicauca a nivel 3

Los problemas de escalabilidad y de convergencia que presenta DVMRP por ser un protocolo de enrutamiento multicast de modo denso llevan a que el primer cambio hacia un backbone multicast moderno sea el de sustituir el acceso al tráfico multicast por medio de túneles DVMRP por medios nativos como ocurrió en la evolución del Mbone de Internet.

La evolución lógica es configurar los dos enrutadores de la Universidad para que soporte PIM-SM y llegar a un acuerdo con los proveedores para que habiliten sus redes para el tráfico multicast con el objetivo de conformar una nube PIM-SM y crear un sistema autónomo multicast. Si los proveedores acceden a esta propuesta se tendría la posibilidad de que otras redes en la ciudad de Popayán hagan parte de esta nube ya que el manejo y la configuración de equipos no es complicada debido a la topología sencilla con la que cuentan las pequeñas redes de la cuidad. Como se estudió en el capítulo 4, el protocolo de enrutamiento PIM-SM requiere de la ubicación estratégica de un punto de encuentro o RP, que en este caso podría estar situado en alguno de los enrutadores de los proveedores, que igualmente tendría que implementar los protocolos de enrutamiento interdominio MBGP/MSDP.

### 8.3.2 Evolución Red Unicauca a nivel 2

Es conveniente actualizar los equipos de nivel 2 que no soporten el mecanismo de IGMP Snooping para evitar dominios de colisión multicast segmentando el tráfico y haciendo un uso mas eficiente del ancho de banda disponible. El switch BayNetworks 302F y el switch Encore que se encuentran ubicados en Vicerrectoria de Investigaciones y el switch BayNetworks 28200 situado en el

Edificio del Carmen (Ver **Anexo A**) son ejemplos de dispositivos que deben ser cambiados si se quiere un buen manejo del tráfico multicast.

En el estado actual de la red Unicauca sólo los equipos del Edificio de Ingenierías están habilitados con la funcionalidad de IGMP Snooping pero es recomendable hacer el mismo trabajo en el resto de equipos nivel 2 para evitar la inundación de tráfico multicast en toda la red.

También se considera conveniente actualizar el sistema operativo del switch principal para que cumpla funciones de nivel 3.

### 8.3.3 Multicast IPv6 en la red Unicauca

Además de habilitar el tráfico multicast IPv4, también se puede hacer una conexión al M6Bone utilizando una conexión a la Universidad de Guadalajara por medio de un túnel IPv6/IPv4; Los lineamientos que se deben seguir para ser parte del M6Bone se describen a continuación.

- Contar con una organización que ya haga parte del M6Bone y que esté dispuesta a trabajar con la Universidad del Cauca, en este momento la Universidad de Guadalajara es la institución que presta soporte para conexión al M6Bone en el continente Americano y está dispuesta a colaborar con la Universidad del Cauca. La información respecto al trabajo multicast en IPv6 en la Universidad de Guadalajara se puede encontrar en la página http://www.ipv6.udg.mx/.
- Se debe tener una red IPv6 funcionando y se debe escoger que parte de esta será habilitada para el tráfico multicast y que prefijos serán utilizados. La Universidad del Cauca ya cuenta con un rango de direcciones IPv6 proporcionado por la UNAM de México que se podría utilizar para este fin.
- Habilitar un enrutador multicast IPv6, la forma mas fácil de hacer esto es a través de una máquina con el sistema operativo FreeBSD 4.8 que utilice el stack Kame, el cual es un proyecto desarrollado por 6 compañías Japonesas para el soporte de IPv6 e IPsec (para IPv4 e IPv6) para sistemas BSD, el software Kame, la documentación y la información para la configuración de un enrutador multicast IPv6 se puede encontrar en la página http://www.kame.net.

- Configurar los hosts de la red IPv6 para soporte del protocolo MLD e instalar las herramientas multicast para IPv6. Los sistemas operativos que se han probado son: FreeBSD, Linux, Win2000/XP. Las especificaciones para cada sistema operativo se pueden encontrar en la página http://www.m6bone.net/hosts.html
- Se debe llenar un formulario y enviarlo a la dirección de la lista de correo multicast IPv6, como se indica en la página http://www.m6bone.net/form.txt

La página oficial del M6Bone donde se puede encontrar toda la información referente a los trabajos adelantados en este campo y las organizaciones involucradas en este proyecto, así como toda la información relacionada es: http://www.m6bone.net/

Finalmente, si todas las consideraciones vistas anteriormente son tomadas en cuentas e implementadas, la topología final de la red de la Universidad del Cauca se muestra en la figura 8.3.

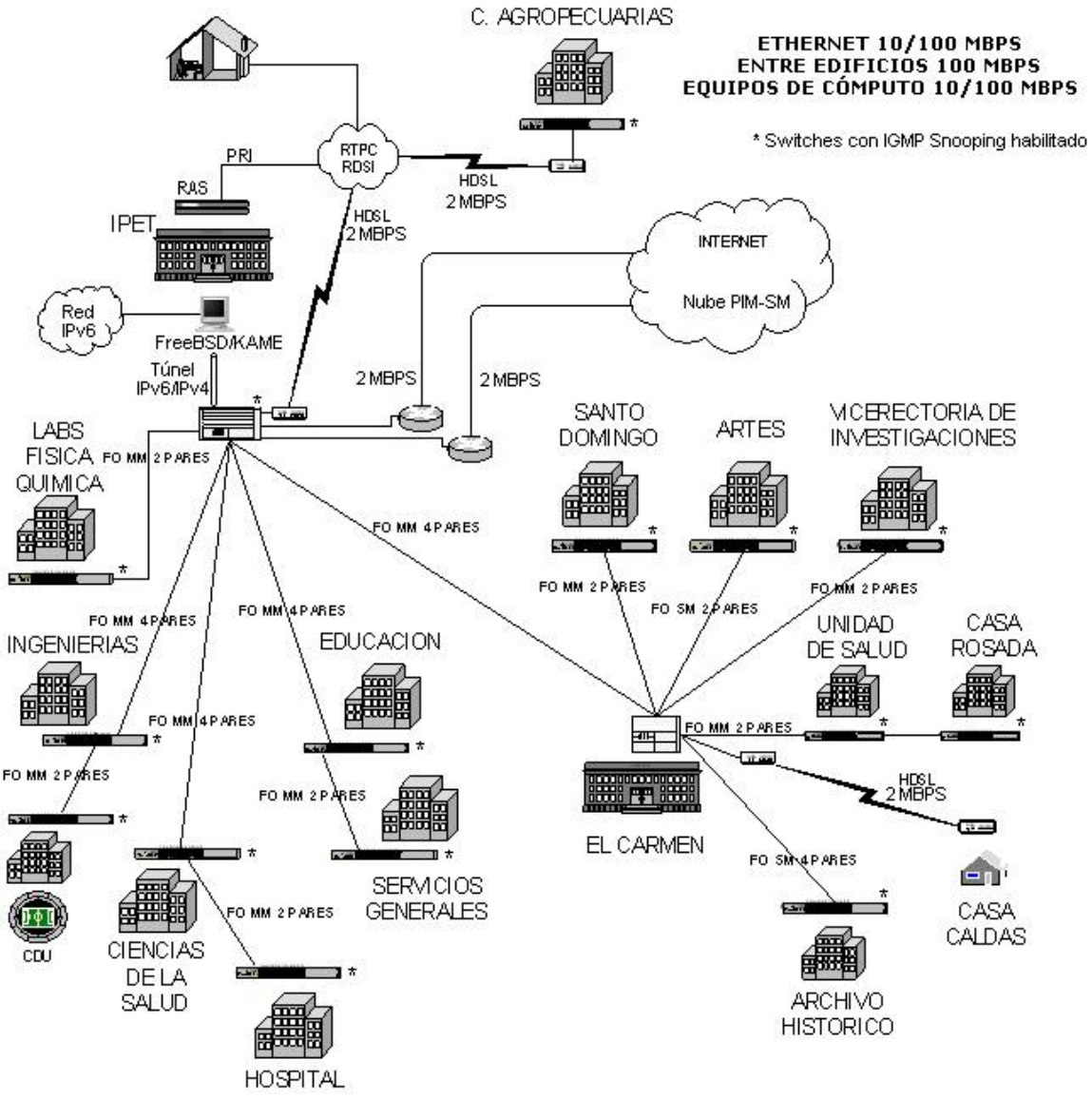


**Figura 8.3 Implementación del diseño multicast de la Red de Unicauca**

# CONCLUSIONES Y RECOMENDACIONES

- Multicast es una tecnología que se puede utilizar en sectores como la tele-educación, la tele-medicina entre otros, desarrollando aplicaciones que aprovechen las ventajas de esta tecnología cuando se requiere difundir información de uno a muchos.

- La tecnología multicast está en continuo desarrollo y todos los esfuerzos estan orientados a que sea un estandar en IPv6 e Internet 2.

- En Colombia no se ha hecho ningún esfuerzo para introducir esta tecnología, perdiendo las capacidades que ofrece, especialmente en las instituciones universitarias donde se podrían llevar a cabo desarrollos y aplicaciones para aprovechar las ventajas de ahorro de ancho de banda en servicios como video y audio conferencias.

- Este proyecto de grado podría utilizarse como base para realizar investigaciones e implementaciones sobre tecnologías más actuales relacionadas con la tecnología multicast.

- Es recomendable desarrollar una red de investigación y desarrollo para la Facultad de Ingeniería Electrónica y Telecomunicaciones para probar tecnologías de última generación que después puedan ser llevadas a la práctica.

# BIBLIOGRAFÍA


- ALMEROTH, Kevin C. "The Evolution of Multicast: From the Mbone to Interdomain Multicast to Internet 2 Deployment". Estados Unidos. IEEE Network. Enero 2000.

- Cisco Systems ."Configuring Logical Interfaces". Disponible en: http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121cgcr/inter_c/icdlogin.pdf

- Cisco Systems. "Configuring a Rendezvous Point". Disponible en: http://www.cisco.com/univercd/cc/td/doc/cisintwk/ intsolns/mcst_sol/rps.pdf

- Cisco Systems. "Configuring DVMRP Interoperability". Disponible en: http://www.cisco.com/univercd/cc/td/doc/product/software/ ios122/122cgcr/fipr_c/ipcpt3/1cfdvmrp.pdf

- Cisco Systems. "Multicast Deployment Made Easy". Disponible en: http://www.cisco.com/warp/public/cc/techno/ tity/ipmu/tech/ipcas_dg.pdf

- Cisco Systems. "Cisco IOS, IP Commando Reference, Volume 3 of 3: Multicast". Disponible en: http://www.cisco.com/warp/public/cc/techno/ tity/ipmu/tech/ipcas_dg.pdf

- Cisco Systems. "Cisco - Multicast Quick - Start Configuration Guide". Disponible en: http://www.cisco.com/warp/public/105/48.pdf

- Cisco Systems. "Interdomain Multicast Solutions Using MSDP". Disponible en: http://www.cisco.com/univercd/cc/td/doc/cisintwk/ intsolns/mcst_p1/mcstmsdp/mcst_p1.pdf

- CUJAR OTERO, Carlos Fernando; NARVÁEZ JOAQUÍ, Oscar Felipe. "Evolución de la Red de datos de la Universidad del Cauca hacia una infraestructura de red de área local de alta velocidad". Popayán, Colombia. 2001.

- GIFFORD, Steve. "Academia de Networking de Cisco Systems: Guía del primer año". Estados Unidos. Pearson Education. 2002.

- Nortel Networks. "Using the BayStack 350 10/100/1000 Series Switch". Estados Unidos. Agosto 2000.

- WILLIAMSON, Beau. "Developing IP Multicast Networks". Estados Unidos. Cisco Press. 2000.

- 3Com. "SuperStack Switch Management Guide". Estados Unidos. Agosto 2000.

- RFC 1112, Host Extensions for IP Multicasting.

- RFC 1584, Multicast Extensions to OSPF
- RFC 1889, RTP: A Transport Protocol for Real-Time Applications
- RFC 1918, Address Allocation for Private Internets
- RFC 2080, RIPng for IPv6
- RFC 2081, RIPng Protocol Applicability Statement
- RFC 2189, Core Based Trees (CBTv2) Multicast Routing—Protocol Specification—
- RFC 2236, Internet Group Management Protocol, Version 2
- RFC 2327, SDP: Session Description Protocol
- RFC 2328, OSPF Versión 2
- RFC 2373, IP Version 6 Addressing Architecture
- RFC 2710, Multicast Listener Discovery (MLD) for IPv6
- RFC 2711, IPv6 Router Alert Option
- RFC 2730, Multicast Address Dynamic Client Allocation Protocol (MADCAP)
- RFC 2770, GLOP Addressing in 233/8
- RFC 2858, Multiprotocol Extensions for BGP-4
- RFC 2908, The Internet Multicast Address Allocation Architecture
- RFC 2974, Session Announcement Protocol
- RFC 3376, Internet Froup Management Protocol, Version 3
- RFC 3569, An Overview of Source-Specific Multicast (SSM)
- RFC 3618, Multicast Source Discovery Protocol MSDP)
- draft-ietf-pim-dm-new-v2-03.txt, Protocol Independent Multicast – Dense Mode (PIM-DM): Protocol Specification
- draft-ietf-malloc-aap-04.txt, Multicast Address Allocation Protocol (AAP)
- draft-bhattach-pim-ssm-00.txt, A Framework for Source-Specific IP Multicast Deployment
- draft-ietf-ssm-arch-04.txt, Source-Specific Multicast for IP
- draft-ietf-mboned-mix-02.txt, Multicast-Friendly Internet Exchange (MIX)

# ANEXOS

Anexo A. "Topología de la red Unicauca"

Anexo B. "Configuración de los switches de la red de datos de Universidad del Cauca"

Anexo C. "Configuración de enrutadores multicast"

Anexo D. "Demostración práctica de la implementación multicast en el edificio de ingenierías y el mbone"

# ACRÓNIMOS

**A**

AAP, Allocation Address Protocol
ABR, Area Border Router
AS, Autonomous System
ASIC, Application-Specific Integrated Circuits

**B**

BGMP, Border Gateway Multicast Protocol
BGP, Border Gateway Protocol
BR, Border Router

**C**

CAM, Content-Addressable Memory
CBT, Core-Based Trees
CGMP, Cisco Group Management Protocol
CNAME, Canonical NAME
CPU, Central Procesor Unit
CUDI, Corporación Universitaria para el Desarrollo de Internet A.C (México)

**D**

DARPA, Defense Advanced Research Projects Agency
DARTNet - DARPA Tested Network
DR, Designated Router
DVMRP, Distance Vector Multicast Routing Protoco

**F**

FDDI, Fiber Distributed Data Interface
FO, Fibra Óptica

**G**

GARP, Generic Attribute Resolution Protocol

**H**

HDSL, High bit-rate Digital Subscriber Line

**I**

IANA, Internet Assigned Number Authority
IEEE, Institute of Electrical and Electronic Engineers
IETF, Internet Engineering Task Force
IGMP. Internet Group Management Protocol
IP, Internet Protocol
IPv6, Internet Protocol Version 6
ISP, Internet Service Provider

**L**

LAN, Local Area Network

**M**

M6Bone, IPv6 Multicast Backbone
MAAA, Multicast Address Allocation Architecture
MAAS, Multicast Address Allocation Server
MAC, Media Access Control
MADCAP, Multicast Address Dynamic Client Allocation
MASC, Multicast Address-Set Claim
MBGP, Multiprotocol Border Gateway Protocol
MBone, Backbone Multicast
MIX, Multicast-friendly Internet eXchange
MLD, Multicast Listener Discovery
MMUSIC, Multiparty MUltimedia Session Control
MOSPF, Mlticast Open Shortest Path First
mrouted, multicast route deamon
mrouter, multicast routeer
MSDP, Multicast Source Discovery Protocol

**N**

NIC, Network Interface Card
NSRC, Network Startup Resource Center
NTE, Network Text Editor

**P**

PIM, Protocol Independent Multicast
PIM-DM, Protocol Independent Multicast - Dense Mode
PIM-SM, Protocol Independent Multicast - Sparse Mode
PIM-SSM, Protocol Independent Multicast - Source Specific Mode

**R**

RAMA, Root Addressed Multicast Architecture
RAS, Remote Access Server
RAT, Robust Audio Tool
RETINA, la Red Teleinformática de Argentina
REUNA, Red Universitaria Nacional (Chile)
RFC, Request For Comment
RIB, Routing Information Base
RIP, Routing Information Protocol
RNP, Red Nacional de Investigación (Brasil)

RP, Rendezvous Point
RPF, Reverse Path Forwarding
RPT, RP-Trees
RR, Receiver Report
RTCP, Real Time Control Protocol
RTP, Real Time Protocol
RTT, Round Trip Time

**S**

SA, Source Active
SAFI, Subsequent Address Family Identifier
SAP, Session Announcement Protocol
SDP, Session Description Protocol
SDR, Session Directory
SPT, Shortest Path Tree
SR, Source Report
SW, Switch

**T**

TCP, Transport Control Protocol
TTL, Time To Live

**O**

OSI, Open System Interconnection
OSPF, Open Shortest Path First

**U**

UDP, User Datagram Protocol

**V**

VIC, Video Conference Tool

**W**

WB, White Board