7. Capítulo 7: Protocolos de Ruteo

Uno de los aspectos más complejos y cruciales del diseño de redes de conmutación de paquetes incluidas internet y las redes privadas, es la implementación de una solución eficiente al encaminamiento (ruteo) de datagramas. En términos generales, el ruteo es una función que busca encontrar *"la mejor ruta"* para comunicar dos sistemas finales.

Este capítulo comienza con una breve descripción general de los problemas relacionados con el diseño del ruteo. A continuación, se analizan distintas opciones de ruteo, algoritmos y su implementación a través de protocolos.

7.1. El problema del encaminamiento (ruteo)

La función principal de una red de conmutación de paquetes, internet o una red privada interconectada por enrutadores (routers) es aceptar paquetes procedentes de una estación emisora y enviarlos hacia una estación destino. Para ello se debe determinar una ruta a través de la red, siendo posible generalmente la existencia de más de una. Así pues, se debe realizar una función de encaminamiento, entre cuyos requisitos se encuentran los siguientes: exactitud, imparcialidad, simplicidad, optimización, robustez, eficiencia, y estabilidad.

Las dos primeras características mencionadas se explican por sí mismas. La robustez está relacionada con la habilidad de la red para enviar paquetes de alguna forma ante la aparición de sobrecargas y fallos localizados. Idealmente, la red puede reaccionar ante estas contingencias sin sufrir pérdidas de paquetes o caída de circuitos virtuales. No obstante, la robustez puede implicar cierta inestabilidad. Las técnicas que reaccionan ante condiciones cambiantes presentan una tendencia no deseable a reaccionar demasiado lentamente ante determinados eventos o a experimentar oscilaciones inestables de una situación extrema a otra. Por ejemplo, la red puede reaccionar ante la aparición de congestión en un área desplazando la mayor parte de la carga hacia una segunda zona. Ahora será la segunda región la que estará sobrecargada y la primera infrautilizada, produciéndose un segundo desplazamiento del tráfico. Durante estos desplazamientos puede ocurrir que los paquetes viajen en bucles a través de la red.

También existe un compromiso entre la característica de imparcialidad y el hecho de que el encaminamiento trate de ser óptimo. Algunos criterios de funcionamiento pueden dar prioridad al intercambio de paquetes entre estaciones vecinas frente al intercambio realizado entre estaciones distantes, lo cual puede maximizar la eficiencia promedio, pero será injusto para aquella estación que necesite comunicar principalmente con estaciones lejanas.

Finalmente, una técnica de encaminamiento implica cierto coste de procesamiento en cada nodo y, en ocasiones, también un coste en la transmisión,

impidiéndose en ambos casos el funcionamiento eficiente de la red. Este coste debe ser inferior a los beneficios obtenidos por el uso de una métrica razonable, como la mejora de la robustez o la imparcialidad. A continuación, un resumen de los elementos que hay que tener en cuenta al momento de diseñar una solución de encaminamiento:

- ✓ Criterios de rendimiento
  - Número de saltos
  - Coste
  - Retardo
  - Eficiencia
- ✓ Instante de decisión
  - Paquete (datagrama)
  - Sesión (circuitos virtuales)
8. Lugar de decisión
   - Cada nodo (distribuido)
   - Nodo central (centralizado)
   - Nodo origen (fuente)
- ✓ Fuente de información de la red
  - Ninguna
  - Local
  - Nodo adyacente
  - Nodos a lo largo de la ruta
  - Todos los nodos
- ✓ Tiempo de actualización de la información de la red
  - Continuo
  - Periódico
  - Cambio importante en la carga
  - Cambio en la topología

Criterios de rendimiento

La elección de una ruta se fundamenta generalmente en algún criterio de rendimiento. El más simple consiste en elegir el camino con menor número de saltos (aquel que atraviesa el menor número de nodos) a través de la red. Éste es un criterio que se puede medir fácilmente y que debería minimizar el consumo de recursos de la red. Una generalización del criterio de menor número de saltos lo constituye el encaminamiento de mínimo coste. En este caso se asocia un coste a cada enlace y, para cualesquiera dos estaciones conectadas, se elige aquella ruta a través de la red que implique el coste total mínimo. Por ejemplo, en la Figura 7.1.1 se muestra una red en la que las dos líneas con flecha entre cada par de nodos representan un enlace entre ellos, y los números asociados indican el coste actual del enlace en cada sentido. El camino más corto (menor número de saltos) desde el nodo 1 hasta el 6 es 1-3-6 (costo = 5+5 = 10), pero el de mínimo costo es 1-4-5- 6 (costo = 1+1+2 = 4). La asignación de los costes a los enlaces se hace en función de los objetivos de diseño; por ejemplo, el coste podría estar inversamente relacionado con la velocidad (es decir, a mayor velocidad menor coste) o con el retardo actual de la cola asociada al enlace. En el primer caso, la ruta de mínimo coste maximizaría la eficiencia, mientras que en el segundo minimizaría el retardo.

Tanto en la técnica de menor número de saltos como en la de mínimo coste, el algoritmo para determinar la ruta o camino óptimo entre dos estaciones es relativamente sencillo, siendo el tiempo de procesamiento aproximadamente el mismo en ambos casos. Dada su mayor flexibilidad, el criterio de mínimo coste es más utilizado que el de menor número de saltos.

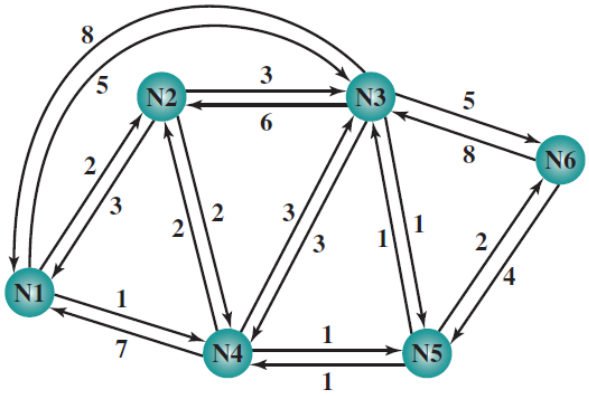

Figura 7.1.1: Red con costos asociados a sus enlaces

Instante y lugar de decisión

Las decisiones de encaminamiento se realizan de acuerdo con algún criterio de rendimiento. Dos cuestiones importantes en la toma de esta decisión son el instante temporal y el lugar en que se toma la decisión.

El instante de decisión viene determinado por el hecho de que la decisión de encaminamiento se hace en base a un paquete o a un circuito virtual. Cuando la operación interna de la red se basa en datagramas, la decisión de encaminamiento se toma de forma individual para cada paquete.

El término lugar de decisión hace referencia al nodo o nodos en la red responsables de la decisión de encaminamiento. El más común es el encaminamiento distribuido, en el que cada nodo de la red tiene la responsabilidad de seleccionar un enlace de salida sobre el que llevar a cabo el envío de los paquetes a medida que éstos se reciben. En el encaminamiento centralizado, la decisión se toma por parte de algún nodo designado al respecto, como puede ser un centro de control de red. El peligro de esta última aproximación es que el fallo del centro de control puede bloquear el funcionamiento de la red; así pues, aunque la aproximación distribuida puede resultar más compleja es también más robusta. Una tercera alternativa empleada en algunas redes es la conocida como encaminamiento desde el origen. En este caso, es la estación origen y no los nodos de la red quien realmente toma la decisión de encaminamiento, comunicándosela a la red. Esto permite al usuario fijar una ruta a través de la red de acuerdo con criterios locales al mismo.

7.2. Estrategias de encaminamiento

Existen numerosas estrategias de encaminamiento para abordar las necesidades de encaminamiento en redes de conmutación de paquetes. Muchas de ellas son aplicables también al encaminamiento en la interconexión de redes.

7.2.1. Encaminamiento estático

En el encaminamiento estático se configura una única ruta permanente para cada par de nodos origen-destino en la red, pudiéndose utilizar para ello cualquiera de los algoritmos de encaminamiento de mínimo coste conocidos (Dijkstra, Bellman-Ford, entre otros), o directamente definir y cargar manualmente las rutas. Las rutas son fijas (al menos mientras lo sea la topología de la red), de modo que los costes de enlace usados para el diseño de las rutas no pueden estar basados en variables dinámicas como el tráfico, aunque sí podrían estarlo en tráfico esperado o en capacidad.

La Figura 7.2.1.1 sugiere cómo se pueden implementar rutas estáticas. Se crea una matriz de encaminamiento central, almacenada, por ejemplo, en un centro de control de red. Esta matriz especifica para cada par de nodos origen-destino, la identidad del siguiente nodo en la ruta. Luego mediante comandos del sistema operativo del router se cargan cada una de las rutas.

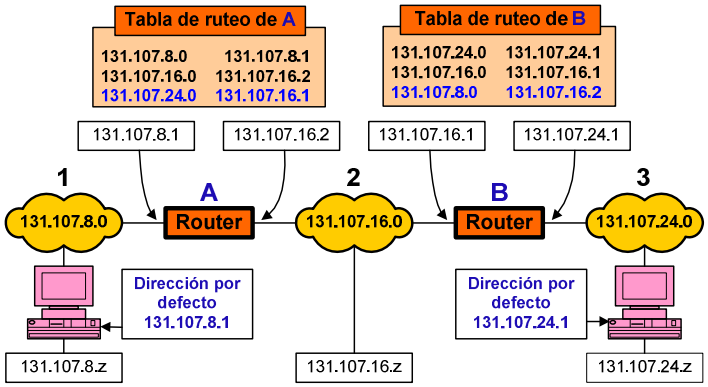

Figura 7.2.1.1: Tablas de ruteo estáticas

7.2.2. Encaminamiento dinámico: Primera Generación (Vector Distancia)

El algoritmo de encaminamiento original, diseñado en 1969, era un algoritmo adaptable distribuido que hacía uso de la estimación de los retardos como criterio de rendimiento y de una versión del algoritmo de Bellman-Ford. Para este algoritmo, cada nodo mantiene dos vectores:

$$D_i = \begin{bmatrix} d_{i1} \\ \vdots \\ d_{iN} \end{bmatrix} \qquad S_i = \begin{bmatrix} s_{i1} \\ \vdots \\ s_{iN} \end{bmatrix}$$

Dónde,

D_i = vector de retardo para el nodo i.

D_{ij} = estimación actual del retardo mínimo desde el nodo i al nodo j (dii%0).

N = número de nodos en la red.

S_i = vector de nodos sucesores para el nodo i.

S_{ij} = nodo siguiente en la ruta actual de mínimo retardo de i a j.

Periódicamente (cada 128 ms), cada nodo intercambia su vector de retardo con todos sus vecinos. A partir de todos los vectores de retardo recibidos, un nodo k actualiza sus dos vectores como sigue:

$$d_{kj} = \min_{i \in A} [d_{ij} + l_{ki}]$$

$s_{kj} = i$, siendo i el que minimiza la expresión anterior

En la Figura 7.2.2.1 se muestra un ejemplo del algoritmo original de ARPANET, usando la red de la Figura 7.2.2.2. En la Figura 7.2.2.1(a) se muestra la tabla de encaminamiento del nodo 1 en un instante de tiempo que refleja los costes asociados a los enlaces de la Figura 7.2.2.2. Para cada destino se especifica un retardo y el nodo siguiente en la ruta que lo produce. En algún momento, los costes de los enlaces cambian a los valores indicados en la Figura 7.1.1. Suponiendo que los vecinos del nodo 1 (nodos 2, 3 y 4) conocieran el cambio antes que él, cada uno de estos nodos actualizará su vector de retardo y enviará una copia a todos sus vecinos, incluyendo el nodo 1 (7.2.2.1(b)). El nodo 1 desecha su tabla de encaminamiento y construye una nueva basándose en los vectores de retardo recibidos y en la propia estimación que él hace del retardo para cada uno de los enlaces de salida a sus vecinos. El resultado obtenido se muestra en la Figura 7.2.2.1(c).

El retardo de enlace estimado no es más que el tamaño o longitud de la cola para el enlace en cuestión. Así, con la construcción de una nueva tabla de encaminamiento el nodo tiende a favorecer aquellos enlaces con menores colas, lo que compensa la carga entre las distintas líneas de salida. Sin embargo, dado que el tamaño de las colas varía rápidamente a lo largo del tiempo, la percepción distribuida de la ruta más corta podría cambiar mientras un paquete se encuentra en tránsito. Esto podría provocar una situación en la que un paquete se encamina hacia un área de baja congestión en lugar de hacia el destino.

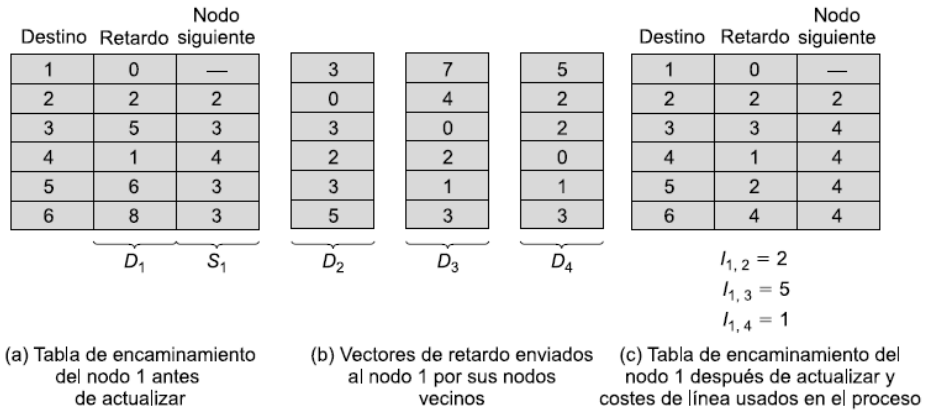

Figura 7.2.2.1: Algoritmo de encaminamiento vector distancia

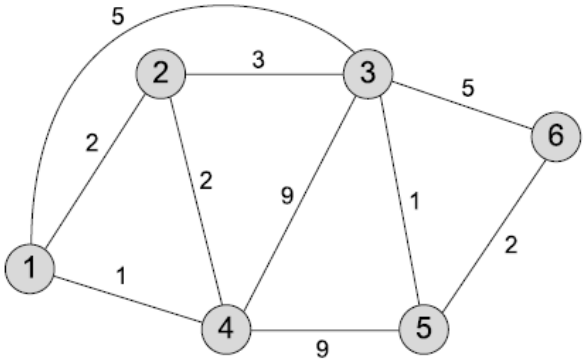

Figura 7.2.2.2: Red para construir la matriz de la figura 7.2.2.1(a)

7.2.3. Encaminamiento dinámico: Segunda Generación (Estado del enlace)

El enrutamiento por vector de distancia se utilizó en ARPANET hasta 1979, cuando se reemplazó por el enrutamiento por estado del enlace. El principal problema que provocó su desaparición era que, con frecuencia, el algoritmo tardaba demasiado en converger una vez que cambiaba la topología de la red (debido al problema del conteo al infinito). En consecuencia, se reemplazó por un algoritmo totalmente nuevo, ahora conocido como enrutamiento por estado del enlace.

La idea detrás del enrutamiento por estado del enlace es bastante simple y se puede enunciar en cinco partes. Cada enrutador debe realizar lo siguiente para hacerlo funcionar:

1. Descubrir a sus vecinos y conocer sus direcciones de red.
2. Establecer la métrica de distancia o de costo para cada uno de sus vecinos.

3. Construir un paquete que indique todo lo que acaba de aprender.
4. Enviar este paquete a todos los demás enrutadores y recibir paquetes de ellos.
5. Calcular la ruta más corta a todos los demás enrutadores.

De hecho, se distribuye la topología completa a todos los enrutadores. Después se puede ejecutar el algoritmo de Dijkstra en cada enrutador para encontrar la ruta más corta a los demás enrutadores. A continuación, veremos con mayor detalle cada uno de estos cinco pasos.

Aprender sobre los vecinos

Cuando un enrutador se pone en funcionamiento, su primera tarea es averiguar quiénes son sus vecinos. Para lograr esto envía un paquete especial HELLO en cada línea punto a punto. Se espera que el enrutador del otro extremo regrese una respuesta en la que indique su nombre. Estos nombres deben ser globalmente únicos puesto que, cuando un enrutador distante escucha después que hay tres enrutadores conectados a F, es indispensable que pueda determinar si los tres se refieren al mismo F.

Establecimiento de los costos de los enlaces

El algoritmo de enrutamiento por estado del enlace requiere que cada enlace tenga una métrica de distancia o costo para encontrar las rutas más cortas. El costo para llegar a los vecinos se puede establecer de modo automático, o el operador de red lo puede configurar. Una elección común es hacer el costo inversamente proporcional al ancho de banda del enlace. Por ejemplo, una red Ethernet de 1 Gbps puede tener un costo de 1 y una red Ethernet de 100 Mbps un costo de 10. Esto hace que las rutas de mayor capacidad sean mejores opciones.

Si la red está geográficamente dispersa, el retardo de los enlaces se puede considerar en el costo, de modo que las rutas a través de enlaces más cortos sean mejores opciones. La manera más directa de determinar este retardo es enviar un paquete especial ECO a través de la línea, que el otro extremo tendrá que regresar de inmediato. Si se mide el tiempo de ida y vuelta, y se divide entre dos, el enrutador emisor puede obtener una estimación razonable del retardo.

Construcción de los paquetes de estado del enlace

Una vez que se ha recabado la información necesaria para el intercambio, el siguiente paso es que cada enrutador construya un paquete que contenga todos los datos. El paquete comienza con la identidad del emisor, seguida de un número de secuencia, una edad (se describirá luego) y una lista de vecinos. También se proporciona el costo para cada vecino. En la figura 7.2.3.1(a) se muestra una red de ejemplo; los costos se muestran como etiquetas en las líneas. En la figura 7.2.3.1(b) se muestran los paquetes de estado del enlace correspondientes para los seis enrutadores.

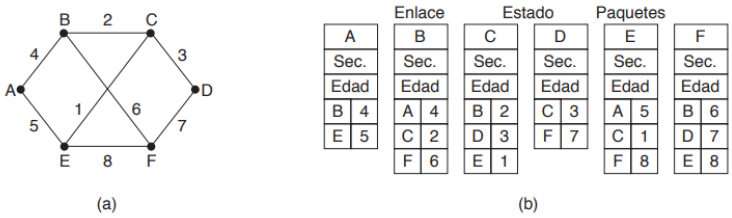

Figura 7.2.3.1: (a) La red, (b) Paquetes de estado de enlace para esta red

Es fácil construir los paquetes de estado del enlace. La parte difícil es determinar cuándo construirlos. Una posibilidad es construirlos de manera periódica; es decir, a intervalos regulares. Otra posibilidad es construirlos cuando ocurra un evento significativo, como la caída o la reactivación de una línea o de un vecino, o cuando sus propiedades cambien en forma considerable.

Distribución de los paquetes de estado del enlace

La parte más complicada del algoritmo es la distribución de los paquetes de estado del enlace. Todos los enrutadores deben recibir todos los paquetes de estado del enlace con rapidez y confiabilidad. Si se utilizan distintas versiones de la topología, las rutas que se calculen podrían tener inconsistencias como ciclos, máquinas inalcanzables y otros problemas.

Primero se describirá el algoritmo básico de distribución. Después se le aplicará algunos refinamientos. La idea fundamental es utilizar inundación para distribuir los paquetes de estado del enlace a todos los enrutadores. Con el fin de mantener controlada la inundación, cada paquete contiene un número de secuencia que se incrementa con cada nuevo paquete enviado. Los enrutadores llevan el registro de todos los pares (enrutador de origen, secuencia) que ven. Cuando llega un nuevo paquete de estado del enlace, se verifica y compara con la lista de paquetes ya vistos. Si es nuevo, se reenvía a través de todas las líneas, excepto aquella por la que llegó. Si es un duplicado, se descarta. Si llega un paquete con número de secuencia menor que el mayor visto hasta el momento, se rechaza como obsoleto debido a que el enrutador tiene datos más recientes.

Este algoritmo tiene algunos problemas, pero son manejables. Primero, si los números de secuencia vuelven a comenzar, reinará la confusión. La solución aquí es utilizar un número de secuencia de 32 bits. Con un paquete de estado del enlace por segundo, el tiempo para volver a empezar será de 137 años, por lo que podemos ignorar esta posibilidad. Segundo, si llega a fallar un enrutador, perderá el registro de su número de secuencia. Si comienza nuevamente en 0, se rechazará como duplicado el siguiente paquete que envíe. Tercero, si llega a corromperse un número de secuencia y se recibe 65540 en vez de 4 (un error de 1 bit), los paquetes 5 a 65 540 se rechazarán como obsoletos, dado que se piensa que el número de secuencia actual es 65 540.

La solución a todos estos problemas es incluir la edad de cada paquete después del número de secuencia y disminuirla una vez cada segundo. Cuando la edad llega a cero, se descarta la información de ese enrutador. Por lo general, un paquete nuevo entra, por ejemplo, cada 10 segundos, por lo que la información de los enrutadores sólo expira cuando un enrutador está caído (o cuando se pierden seis paquetes consecutivos, un evento poco probable). Los enrutadores también decrecen el campo Edad durante el proceso inicial de inundación para asegurar que no pueda perderse ningún paquete y sobrevivir durante un periodo de tiempo indefinido (se descarta el paquete cuya edad sea cero).

Algunos refinamientos a este algoritmo pueden hacerlo más robusto. Una vez que llega un paquete de estado del enlace a un enrutador para ser inundado, no se encola para su transmisión de inmediato, sino que se coloca en un área de almacenamiento para esperar un tiempo corto, en caso de que se activen o desactiven más enlaces. Si llega otro paquete de estado del enlace proveniente del mismo origen antes de que se transmita el primer paquete, se comparan sus números de secuencia. Si son iguales, se descarta el duplicado. Si son diferentes, se desecha el más antiguo. Como protección contra los errores en los enlaces, se confirma la recepción de todos los paquetes de estado del enlace

Cálculo de las nuevas rutas

Una vez que un enrutador ha acumulado un conjunto completo de paquetes de estado del enlace, puede construir el grafo de toda la red debido a que todos los enlaces están simbolizados. De hecho, cada enlace se representa dos veces, una para cada dirección. Las distintas direcciones pueden tener incluso costos diferentes. Así, los cálculos de la ruta más corta pueden encontrar rutas del enrutador A a B que sean distintas a las del enrutador de B a A.

Ahora se puede ejecutar localmente el algoritmo de Dijkstra para construir las rutas más cortas a todos los destinos posibles. Los resultados de este algoritmo indican al enrutador qué enlace debe usar para llegar a cada destino. Esta información se instala en las tablas de enrutamiento y se puede reanudar la operación normal.

En comparación con el enrutamiento por vector de distancia, el enrutamiento por estado del enlace requiere más memoria y poder de cómputo. Para una red con n enrutadores, cada uno de los cuales tiene k vecinos, la memoria requerida para almacenar los datos de entrada es proporcional a kn, que es por lo menos tan grande como una tabla de enrutamiento que lista todos los destinos. Además, el tiempo de cómputo también crece con más rapidez que kn, incluso con las estructuras de datos más eficientes; un problema en las grandes redes. Sin embargo, en muchas situaciones prácticas, el enrutamiento por estado del enlace funciona bien debido a que no sufre de los problemas de convergencia lenta.

7.3. Sistema Autónomo (Autonomous System)

Para continuar con el análisis sobre los protocolos de encaminamiento, se necesita introducir el concepto de sistema autónomo. Un sistema autónomo (AS, Autonomous System) posee las siguientes características:

1. Un AS se compone de un conjunto de encaminadores y redes gestionados por una única organización.
2. Un AS consiste en un grupo de dispositivos de encaminamiento que intercambian información a través de un protocolo de encaminamiento común.
3. Excepto en momentos de avería, un AS está conectado (en un sentido teórico de grafo). Es decir, existe un camino entre cualquier par de nodos.

Un protocolo común de encaminamiento, al que nos referiremos como protocolo de ruteo interior (IRP, Interior Router Protocol), distribuye la información de encaminamiento entre los dispositivos de encaminamiento dentro de un AS. El protocolo que se emplea dentro de un sistema autónomo no necesita ser implementado fuera del sistema. Esta flexibilidad permite que los IRP se hagan a medida para aplicaciones y requisitos específicos.

Puede ocurrir, sin embargo, que una interconexión de redes esté constituida por más de un AS. Por ejemplo, todas las LAN de una organización, como puede ser un complejo de oficinas o un campus, podrían estar enlazadas mediante encaminadores para formar un AS. Este sistema se podría unir a otros AS a través de una red de área amplia. Esta situación se muestra en la Figura 7.3.1.

En este caso, los algoritmos de encaminamiento y la información de las tablas de encaminamiento utilizadas por los encaminadores en los distintos AS pueden ser diferentes. Sin embargo, los encaminadores de un AS necesitan al menos un nivel mínimo de información referente a las redes externas al sistema que puedan alcanzar. El protocolo que se utiliza para pasar información de encaminamiento entre diferentes AS se conoce como protocolo de ruteo exterior (ERP, Exterior Router Protocol)[1].

[1] En la bibliografía relacionada se utilizan a menudo los términos protocolo de pasarela interior (IGP, Interior Gateway Protocol) y protocolo de pasarela exterior (EGP, Exterior Gateway Protocol) para designar lo que aquí denominamos IRP y ERP.

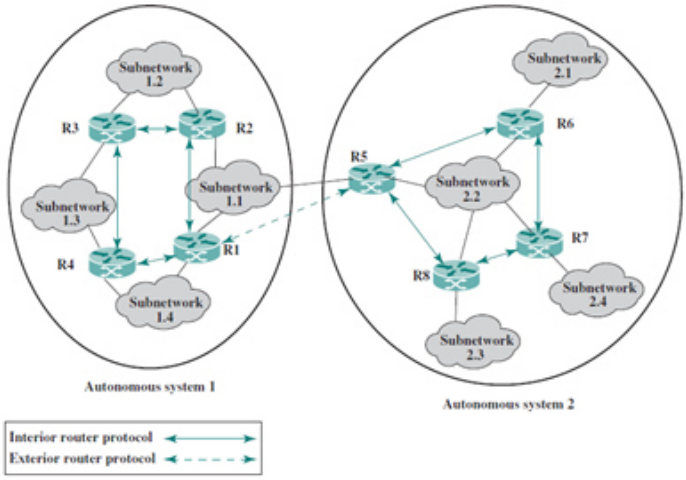

Figura 7.3.1: Protocolos de ruteo interior y exterior

7.4. Protocolo de ruteo interno: OSPF (Open Shortest Path First)

OSPF (Open Shortest Path First) es un protocolo de encaminamiento dinámico de pasarela interior o IGP (Interior Gateway Protocol), que se creó para solucionar las limitaciones que tenía el protocolo RIP (Routing Information Protocol). Este protocolo permite gestionar redes con un diámetro mayor que 16, mejorando además el tiempo de convergencia y la agregación de rutas.

El algoritmo utilizado por OSPF es más complejo que el utilizado por RIP, por lo que se necesitan routers con más potencia de procesador y memoria, y se requiere más tiempo de diseño e implementación. De este modo, se puede afirmar que ambos protocolos se han diseñado para entornos totalmente distintos: OSPF está diseñado para redes grandes y complejas, mientras que RIP está destinado a redes pequeñas con una configuración sencilla.

A diferencia del algoritmo de vector de distancia utilizado por RIP, OSPF se basa en un algoritmo de estado de los enlaces (link state). Por ello, este protocolo no envía a los encaminadores adyacentes el número de saltos que los separa, sino el estado del enlace que los separa. De esta manera, cada router es capaz de construir un mapa completo del estado de la red y, por consecuencia, puede elegir en cada momento la ruta más apropiada para enviar un mensaje a un destino dado.

Como cada router contiene el mismo mapa de la topología de la red, OSPF no requiere que las actualizaciones se envíen a intervalos regulares. De este modo, OSPF

reduce el consumo de red necesario para el intercambio de actualizaciones mediante la multidifusión, enviando una actualización sólo cuando se detecta un cambio (en lugar del envío periódico) y enviando cambios de la tabla de encaminamiento (en vez de la tabla completa) sólo cuando es necesaria una actualización. Además, el hecho de no tener que ir incrementando el número de saltos cada vez que se pasa por un router intermedio se traduce en una cantidad de información a intercambiar mucho menos abundante y, por tanto, en un ancho de banda libre mejor que en el caso de RIP.

La métrica utilizada es más sofisticada que en el caso de RIP, ya que se basa en el ancho de banda del enlace (por omisión, coste = 10^8/ancho de banda (b/s)). Por ejemplo, para el caso de un enlace mediante Ethernet 10Mb/s se tiene un coste de 10.

Existen diferentes versiones de este protocolo: OSPFv1 (RFC1131 y RFC1247); OSPFv2 (RFC2328); y OSPFv3 (RFC2740), el cual está adaptado para IPv6. En las redes encaminadas por OSPF se define la siguiente jerarquía (Figura 7.4.1):

- ✓ Área: Constituye una frontera para el cálculo en la base de datos del estado del enlace. Los routers que están en la misma área contienen la misma base de datos topológica. Un área es en realidad una subdivisión de un AS (Autonomous System). El área principal se denomina backbone y a ésta se conectan el resto de áreas (sean conexiones física o virtualmente). Las diferentes áreas se conectan entre sí mediante unos routers de borde que se encargan de intercambiar las diferentes tablas de encaminamiento.
- ✓ ABR (Area Border Router): Router que contiene enlaces en varias áreas dentro del mismo AS, cuya función consiste en resumir las informaciones de encaminamiento y gestionar los intercambios de rutas entre áreas.
- ✓ IR (Internal Router): Router cuyos enlaces pertenecen todos a un área determinada.
- ✓ DR (Designated Router): Cada segmento de red tiene un DR y un BDR, por lo que un router conectado a múltiples redes puede ser DR de un segmento y un router normal del otro segmento. En realidad, es la interfaz del router la que actúa como DR o BDR. La principal función del DR es minimizar el *"flooding"* (inundación por anuncios) y la sincronización de las DB's (Data Bases) centralizando el intercambio de información. De este modo, los routers de un mismo segmento no intercambian información del estado del enlace entre ellos, sino que lo hacen con el DR.
- ✓ BDR (Backup Designated Router): Es el router elegido como anunciante secundario (generalmente el segundo con la prioridad más alta). El BDR no hace nada mientras haya un DR en la red (solo actúa si el DR falla). Un BDR detecta que un DR falla porque durante un cierto tiempo no escucha LSA's (Link State Advertisements).
- ✓ ASBR (Autonomous System Border Router): Router que contiene enlaces a distintos AS y que sirve de gateway entre OSPF y otros protocolos de encaminamiento (IGRP, EIGRP, ISIS, RIP, BGP, Static).

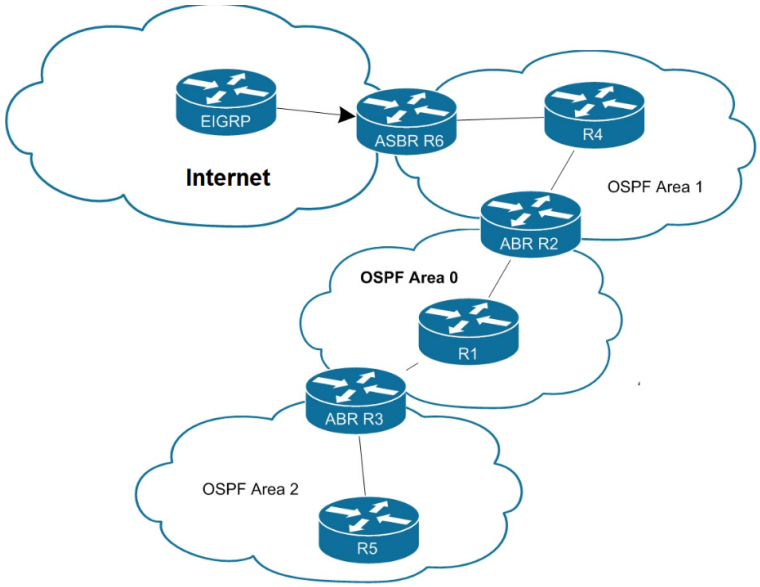

Figura 7.4.1: Topología del protocolo OSPF

En un principio, cada router identifica (o conoce por configuración) a sus vecinos inmediatos. De entre todos los routers de un mismo segmento se eligen un router principal (DR) y un router principal de seguridad (que lo sustituirá en caso de que éste falle). Generalmente, se elige como DR al router con la prioridad más alta de entre todos los routers que pertenecen al mismo segmento de red (la prioridad varía entre 0 y 255). Como la prioridad por defecto suele ser 1, para desempatar se usa el que tenga mayor RID o router ID (donde el router ID suele ser la @IP más alta de una interfaz activa del router). Los routers con prioridad igual a 0 no pueden ser elegidos como DR's.

Una vez que se han elegido el DR y el BDR, se pasa a la fase de descubrimiento de rutas. Para ello, el DR y el DBR forman una adyacencia con cada uno de los routers de su mismo segmento de red. Una adyacencia es una relación que se establece entre un router y su DR y BDR mediante el protocolo *"Hello"*.

En cada adyacencia, uno de los dos routers actúa como master (el de mayor *"routerID"*, que suele ser el DR) y el otro como slave (esclavo). El master envía un resumen de su DB al slave y este la reconoce y viceversa. Luego, el slave compara la información recibida y pide que le envíe aquellas entradas que no tiene. De este modo, cada router difunde al DR los mensajes LSA con los cambios que se producen en la red y el DR se encarga de actualizar la base de datos de cada uno de los routers de su segmento de red haciendo flooding de la información de encaminamiento.

A partir de la base de datos con la topología de la red, en cada router se calculan localmente los caminos más cortos a todos los destinos mediante el algoritmo SPF (Shortest Path First) de Djkstra para poder construir así la tabla de encaminamiento. Cuando se tienen redes con una gran cantidad de routers el número de LSU's enviado produce un gran consumo de ancho de banda, a la vez que se hacen también grandes el tiempo de convergencia y el tamaño de las bases de datos de los routers para guardar la

información sobre la topología de toda la red. Por ello, el encaminamiento OSPF propone como solución la división de la red de una forma jerárquica en áreas, que delimita el dominio de envío de los mensajes LSA a un conjunto de routers y redes en un mismo AS.

Por otra parte, en una red multi área se tienen distintos tipos de mensajes LSA:

- ✓ Router LSA (tipo 1): Cada router genera estos paquetes hacia el resto de routers de su misma área, indicando la lista de sus vecinos inmediatos y el coste (métrica) de sus enlaces.
- ✓ Network LSA (tipo 2): Estos paquetes son generados por los routers DR (de una red BMA) sobre los routers vinculados a esa red BMA y solo se envían dentro del área.
- ✓ Summary LSA (tipo 3): Estos paquetes son generados por los ABR para anunciar las redes internas procedentes de un área específica a otros ABR del mismo AS. Se genera un resumen por cada subred de cada área hacia las demás áreas. Estas informaciones se envían primeramente al backbone (área 0), el cual se encargará después de distribuirlas hacia el resto de áreas del AS.
- ✓ ASBR summary LSA (tipo 4): Generados por los ABR's describen rutas al ASBR's. Los routers ABR deben propagar también las informaciones de encaminamiento hacia los ASBR para que éstos puedan conocer cómo alcanzar los routers externos en otras AS (Autonomous System).
- ✓ AS external LSA (tipo 5): Generados por los ASBR's describen rutas externas al AS (entre ellas la ruta por defecto para salir del AS).

Los mensajes de encaminamiento OSPF se encapsulan como un protocolo de transporte con número 89 (Figura 7.4.2):

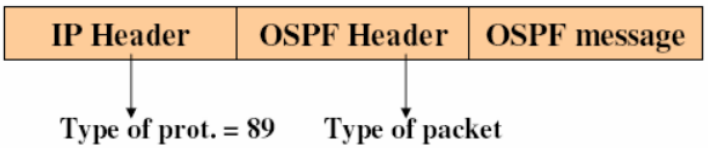

Figura 7.4.2: Encapsulamiento paquete OSPF

Todos los paquetes OSPF incluyen en su cabecera información básica relacionada con el router (Figura 7.4.3):

- ✓ Version: Identifica la versión OSPF.
- ✓ Type: Identifica el tipo de paquete OSPF. Hay 5 tipos de paquetes OSPF:
  - HELLO packets (Type = 1).
  - Database Description (DBD) packets (Type = 2).

- Link-State Request (LSR) packets (Type = 3).
- Link-State Update (LSU) packets (Type = 4).
- Link-State ACK (LSAck) packets (Type = 5).

✓ Packet Length: Longitud del paquete (incluida la cabecera OSPF).

✓ Router ID (RID): Identifica el origen del paquete OSPF (normalmente cada router escoge como RID la @IP mayor entre las @IP activas del mismo).

✓ Area ID: Identifica el área al cual pertenece el paquete OSPF.

✓ Checksum.

✓ Authentication type:

- Type 0: no authentication.
- Type 1: clear-text password or simple authentication.
- Type 2: cryptographic or MD5 authentication.

✓ Authentication information: Contiene la información de autentificación.

✓ Data: Encapsula información de encaminamiento.

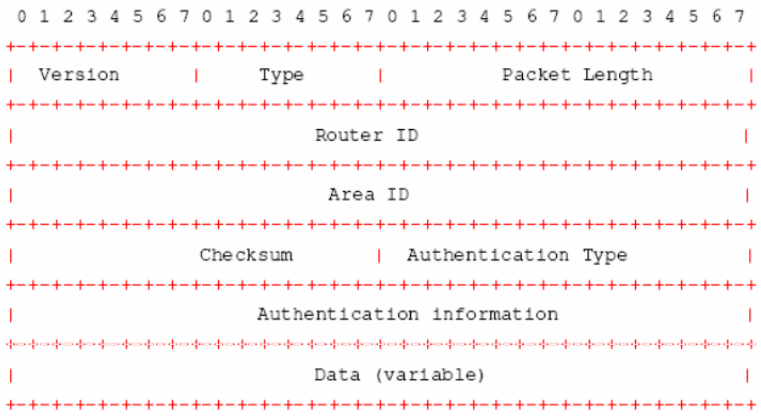

Figura 7.4.3: Cabecera de un paquete OSPF

Paquetes "*Hello*"

Los paquetes Hello (saludo) se utilizan para verificar comunicaciones bidireccionales, anunciar requerimientos de vecindad y elegir routers designados (Figura 7.4.4). Así, los anuncios permiten establecer y mantener una adyacencia, que consiste en una conexión virtual a un vecino para poder transferir anuncios de estado del enlace.

Estos paquetes de saludo se envían a intervalos periódicos de tiempo (HelloInterval = 10 segundos) usando la dirección multicast 224.0.0.5. Los paquetes Hello tienen el formato siguiente:

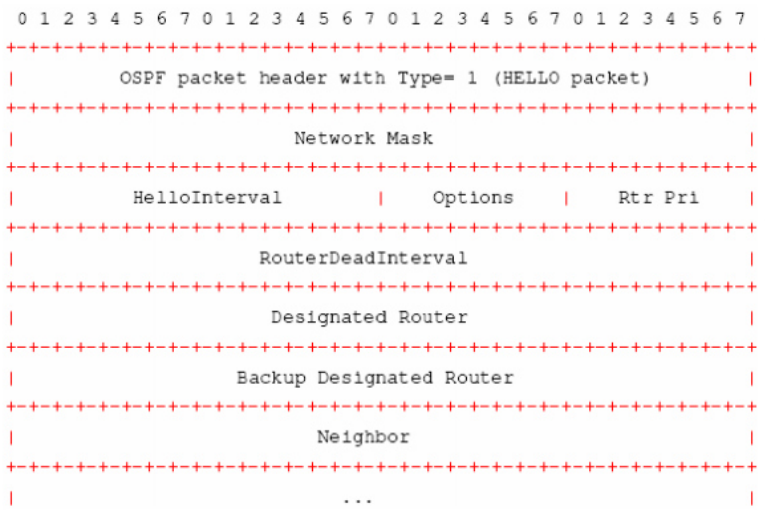

Figura 7.4.4: Cabecera paquete *"Hello"* en OSPF

- ✓ Network Mask: Máscara asociada con esa interfaz.
- ✓ Hello Interval: Intervalo en que se envían paquetes HELLO (10 segundos).
- ✓ Options: Capacidades opcionales que soporta este router.
- ✓ Router Priority: Prioridad (por defecto =1).
- ✓ Router-Dead-Interval: Tiempo que espera un router hasta que deja de considerar que un vecino está activo (4*HelloInterval).
- ✓ DR y BDR: Direcciones IP de ambos (0.0.0.0 si inicialmente desconocidas y hay que descubrirlos).
- ✓ Neighbours: RouteID de cada vecino que ha escuchado durante los últimos RouterDead-Interval segundos.

Por otro lado, se tienen los mensajes DBD, LSR, LSU y LSAck:

Paquetes de descripción de base de datos (DBD): describe el contenido de las DB, incluyendo encabezamientos LSA (no todo el LSA) para que el router receptor confirme que tiene todos los LSA requeridos.

- ✓ Paquetes de petición de estado del enlace (LSR): Solicitan a los vecinos los LSA que están en el listado de petición de estado del enlace. Este listado lleva un registro de los LSA que deben solicitarse porque no se dispone de ellos o porque no se tiene la versión más actualizada. El router sabe qué paquetes le faltan gracias a la recepción de paquetes de petición de otros routers o de paquetes de descripción de base de datos de otros routers.

- ✓ Paquetes de actualización de estado del enlace (LSU): Suministran los LSA (Link State Advertisements) a los routers remotos.
- ✓ Paquetes de acuse de recibo de estado del enlace (LSAck): Sirven de acuse de recibo explícito a uno o más LSU.

Los LSA's (Link-State Advertisements) son unidades de datos que describen el estado local de un router o red. Para un router, esto incluye el estado de las interfaces del router y sus adyacencias. Un LSA va empaquetado en paquetes DBD, LSU, LSR o LSAck.

7.5. Protocolo de ruteo externo: BGP (Border Gateway Protocol)

Dentro de un solo sistema autónomo, OSPF e IS-IS son los protocolos de uso común. Entre los sistemas autónomos se utiliza un protocolo diferente, conocido como BGP (Protocolo de Puerta de Enlace de Frontera, del inglés Border Gateway Protocol). Se necesita un protocolo diferente debido a que los objetivos de un protocolo intra dominio y de un protocolo inter dominio no son los mismos. Todo lo que tiene que hacer un protocolo intra dominio es mover paquetes de la manera más eficiente posible desde el origen hasta el destino. No tiene que preocuparse por las políticas.

En contraste, los protocolos de enrutamiento inter dominio tienen que preocuparse en gran manera por la política. Por ejemplo, tal vez un sistema autónomo corporativo desee la habilidad de enviar paquetes a cualquier sitio de Internet y recibir paquetes de cualquier sitio de Internet. Sin embargo, quizás no esté dispuesto a llevar paquetes de tránsito que se originen en un AS foráneo y estén destinados a un AS foráneo diferente, aun cuando su propio AS se encuentre en la ruta más corta entre los dos sistemas autónomos foráneos ("Ése es su problema, no el nuestro"). Por otro lado, podría estar dispuesto a llevar el tráfico del tránsito para sus vecinos o incluso para otros sistemas autónomos específicos que hayan pagado por este servicio. Por ejemplo, las compañías telefónicas podrían estar contentas de actuar como empresas portadoras para sus clientes, pero no para otros. En general, los protocolos de puerta de enlace exterior (y BGP en particular) se han diseñado para permitir que se implementen muchos tipos de políticas de enrutamiento en el tráfico entre sistemas autónomos.

Las políticas típicas implican consideraciones políticas, de seguridad, o económicas. Algunos ejemplos de posibles restricciones de enrutamiento son:

1. No transportar tráfico comercial en la red educativa.
2. Nunca enviar tráfico del Pentágono por una ruta a través de Irak.
3. Usar TeliaSonera en vez de Verizon porque es más económico.
4. No usar AT&T en Australia porque el desempeño es pobre.
5. El tráfico que empieza o termina en Apple no debe transitar por Google.

Como puede imaginar de esta lista, las políticas de enrutamiento pueden ser muy individuales. A menudo son propietarias pues contienen información comercial delicada. Sin embargo, podemos describir algunos patrones que capturan el razonamiento anterior de la compañía y que se utilizan con frecuencia como un punto de partida.

Para implementar una política de enrutamiento hay que decidir qué tráfico puede fluir a través de cuáles enlaces entre los sistemas autónomos. Una política común es que un ISP cliente pague a otro ISP proveedor por entregar paquetes a cualquier otro destino en Internet y recibir los paquetes enviados desde cualquier otro destino. Se dice que el ISP cliente compra servicio de tránsito al ISP proveedor. Es justo igual que cuando un cliente doméstico compra servicio de acceso a Internet con un ISP. Para que funcione, el proveedor debe anunciar las rutas a todos los destinos en Internet al cliente a través del enlace que los conecta. De esta forma, el cliente tendrá una ruta para enviar paquetes a cualquier parte. Por el contrario, el cliente sólo debe anunciar al proveedor las rutas a los destinos en su red. Esto permitirá al proveedor enviar tráfico al cliente sólo para esas direcciones; al cliente no le conviene manejar el tráfico destinado a otras partes.

En la figura 7.5.1 podemos ver un ejemplo del servicio de tránsito. Hay cuatro sistemas autónomos conectados. Con frecuencia, la conexión se hace mediante un enlace en IXP's (Puntos de Intercambio de Internet, del inglés Internet eXchange Points): instalaciones con las que muchos ISP tienen un enlace para fines de conectarse con otros ISP. AS2, AS3 y AS4 son clientes de AS1, pues le compran servicio de tránsito. Así, cuando la fuente A envía al destino C, los paquetes viajan de AS2 hacia AS1 y finalmente a AS4. Los anuncios de enrutamiento viajan en dirección opuesta a los paquetes. AS4 anuncia a C como un destino a su proveedor de tránsito AS1, para que las fuentes puedan llegar a C por medio de AS1. Después, AS1 anuncia a sus otros clientes, incluyendo AS2, una ruta a C para que éstos sepan que pueden enviar tráfico a C por medio de AS1.

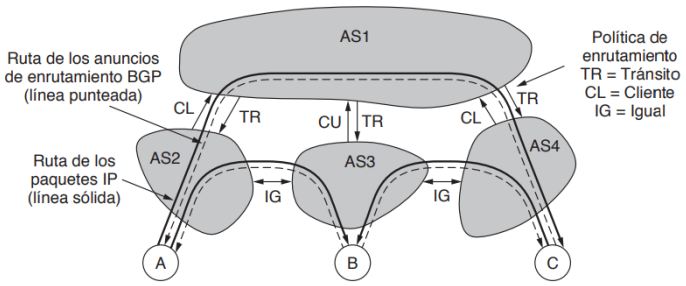

Figura 7.5.1: Política de enrutamiento en cuatro Sistemas Autónomos

En la figura 7.5.1, los demás sistemas autónomos compran servicio de tránsito a AS1. Este servicio les proporciona conectividad para que puedan interactuar con cualquier host en Internet. Sin embargo, tienen que pagar por este privilegio. Suponga que AS2 y AS3 intercambian mucho tráfico. Dado que sus redes ya se encuentran conectadas, si lo desean pueden usar una política diferente: pueden enviar tráfico directamente uno al otro sin costo. Esto reducirá la cantidad de tráfico que debe entregar

AS1 a cuenta de AS2 y AS3, y con suerte reducirá sus facturas. A esta política se le conoce como comunicación entre pares (peering).

Para implementar la comunicación entre pares, dos sistemas autónomos se envían anuncios de enrutamiento entre sí, respecto a las direcciones que residen en sus redes. Al hacer esto, AS2 puede enviar a AS3 paquetes de "A" destinados a "B" y viceversa. Sin embargo, hay que tener en cuenta que la comunicación entre iguales no es transitiva. En la Figura 7.5.1, AS3 y AS4 también se comunican entre sí. Esta comunicación de igual a igual permite que el tráfico de "C", que está destinado a "B", se envíe directamente a AS4. ¿Qué ocurre si "C" envía un paquete a "A"? AS3 sólo está anunciando a AS4 una ruta a "B". No está anunciando una ruta a "A". La consecuencia es que el tráfico no pasará de AS4 a AS3 a AS2, aun cuando existe una ruta física. Esta restricción es justo lo que AS3 quiere. Se comunica con AS4 para intercambiar tráfico, pero no quiere transportar tráfico de AS4 a otras partes de Internet, ya que no se le paga por hacerlo. En cambio, AS4 recibe servicio de tránsito de AS1. Por ende, es AS1quien transportará el paquete de "C" a "A".

Ahora que sabemos sobre el tránsito y la comunicación entre iguales, también podemos ver que, "A", "B" y "C" tienen arreglos de tránsito. Por ejemplo, "A" debe comprar acceso a Internet a AS2. A podría ser una sola computadora doméstica o la red de una compañía con muchas LAN. Sin embargo, no necesita ejecutar BGP debido a que es una red aislada (stub network) que está conectada al resto de Internet sólo mediante un enlace. Por tanto, el único lugar para enviar paquetes destinados a puntos que estén fuera de la red es a través del enlace a AS2. No hay ningún otro lugar a dónde ir. Para arreglar esta ruta, sólo hay que establecer una ruta predeterminada. Por esta razón no hemos mostrado a "A", "B" y "C" como sistemas autónomos que participan en el enrutamiento interdominio.

Por otro lado, las redes de algunas compañías están conectadas a varios ISP. Esta técnica se utiliza para mejorar la confiabilidad, ya que si la ruta a través de un ISP falla, la compañía puede usar la ruta a través del otro ISP. Esta técnica se conoce como multihoming. En este caso, la red de la compañía probablemente ejecute un protocolo de enrutamiento interdominio (como BGP) para indicar a otros sistemas autónomos qué enlaces de ISP pueden llegar a cuáles direcciones.

Hay muchas variaciones posibles de estas políticas de tránsito y comunicación entre iguales, pero todas ilustran cómo las relaciones de negocios y el control sobre el camino que pueden tomar los anuncios de rutas pueden implementar distintos tipos de políticas. Ahora consideraremos con más detalle cómo los enrutadores que ejecutan BGP se anuncian rutas entre sí y seleccionan rutas a través de las cuales pueden reenviar los paquetes.

BGP es una forma de protocolo de vector de distancia, aunque es bastante distinto a los protocolos de vector de distancia intra dominio como RIP. Ya vimos antes que la política, y no la distancia mínima, se utiliza para elegir qué rutas usar. Otra gran diferencia es que, en vez de mantener sólo el costo de la ruta a cada destino, cada enrutador BGP lleva el registro de la ruta utilizada. Esta metodología se conoce como protocolo de vector de ruta. La ruta consiste en el enrutador del siguiente salto (que puede

estar del otro lado del ISP y no necesariamente ser adyacente) y la secuencia de sistemas autónomos (o ruta AS) en el recorrido (se proporciona en orden inverso). Por último, los pares de enrutadores BGP se comunican entre sí mediante el establecimiento de conexiones TCP. Al operar de esta forma se obtiene una comunicación confiable; además se ocultan todos los detalles de la red que se está atravesando.

En la figura 7.5.2 se muestra un ejemplo de cómo se anuncian las rutas BGP. Hay tres sistemas autónomos y el de en medio provee tránsito a los ISP izquierdo y derecho. Un anuncio de ruta para el prefijo C empieza en AS3. Cuando se propaga a través del enlace a R2c en la parte superior de la figura, tiene la ruta AS que consiste tan sólo en AS3 y el enrutador del siguiente salto de R3a. En la parte inferior tiene la misma ruta AS, pero un siguiente salto distinto, debido a que provino de un enlace diferente. Este anuncio continúa su propagación y atraviesa el límite hacia AS1. En el enrutador R1a, en la parte superior de la figura, la ruta AS es AS2, AS3 y el siguiente salto es R2a.

Al transportar la ruta completa con el recorrido, es fácil para el enrutador receptor detectar e interrumpir los ciclos de enrutamiento. La regla es que cada enrutador que envíe un recorrido hacia fuera del AS anteponga su propio número de AS a este recorrido (ésta es la razón por la cual la lista está en orden inverso). Cuando un enrutador recibe un recorrido, verifica que su propio número de AS ya se encuentre en la ruta AS. Si es así, se ha detectado un ciclo y se descarta el anuncio. No obstante (aunque suene un poco irónico), a finales de la década de 1990 se descubrió que, a pesar de esta precaución, BGP sufre de una versión del problema de conteo al infinito (Labovitz, 2000) . No hay ciclos de larga duración, pero algunas veces los recorridos pueden ser lentos para converger y tienen ciclos transitorios.

Proporcionar una lista de sistemas autónomos es una forma muy burda de especificar una ruta. Un AS podría ser una compañía pequeña o una red troncal internacional. No hay forma de saberlo con base en la ruta. BGP ni siquiera lo intenta, ya que los distintos sistemas autónomos pueden usar protocolos intra dominio diferentes, cuyos costos no se puedan comparar. Incluso si se pudieran comparar, tal vez un AS no quiera revelar su métrica interna. Ésta es una de las diferencias entre los protocolos de enrutamiento inter dominio y los protocolos intra dominio.

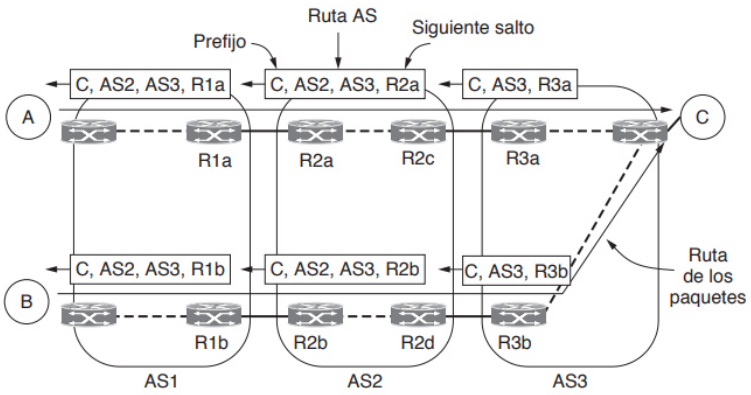

Figura 7.5.2: Anuncia de rutas en BGP

Funciones

Como se mencionó, BGP se diseñó para permitir la cooperación en el intercambio de información de encaminamiento entre dispositivos de encaminamiento de diferentes sistemas autónomos (AS), llamados pasarelas en el estándar. El protocolo opera en términos de mensajes, que se envían utilizando conexiones TCP. El repertorio de mensajes se resume en la Tabla 7.5.3. La versión actual de BGP se conoce como BGP-4 (RFC 1771).

Tabla 7.5.3: Mensajes BGP

Establecer	Utilizado para establecer una relación de vecindad con otro dispositivo de encaminamiento
Actualizar	Utilizado para (1) transmitir información acerca de una única ruta y/o (2) enumerar rutas múltiples que se vayan a eliminar.
Mantener Activa	Utilizado para (1) confirmar un mensaje «establecer» y (2) confirmar periódicamente la relación de vecindad.
Notificación	Se envía cuando se detecta una condición de error.

BGP supone tres procedimientos funcionales, que son:

- ✓ Adquisición de vecino.
- ✓ Detección de vecino alcanzable.
- ✓ Detección de red alcanzable.

Dos dispositivos de encaminamiento se considera que son vecinos si están conectados a la misma subred. Si los dos encaminadores se encuentran en sistemas autónomos diferentes, podrían desear intercambiar información de encaminamiento. Para este cometido, es necesario primero realizar la operación de adquisición de vecino. Básicamente, la adquisición de un vecino se produce cuando dos dispositivos de encaminamiento vecinos de diferentes sistemas autónomos se ponen de acuerdo en intercambiar regularmente información de encaminamiento. Se requiere un procedimiento formal de adquisición debido a que uno de los encaminadores podría no querer participar. Por ejemplo, el dispositivo de encaminamiento puede estar sobresaturado y no querer ser responsable de tráfico que llegue de fuera del sistema. En el proceso de adquisición de un vecino, un dispositivo de encaminamiento envía un mensaje de petición al otro, el cual puede aceptar o rechazar el ofrecimiento. El protocolo no aborda la cuestión de cómo puede un dispositivo de encaminamiento conocer la dirección o incluso la existencia de otro encaminador, ni siquiera de cómo decide que necesita intercambiar información de encaminamiento con un encaminador en particular. Estas cuestiones deben ser tratadas en el momento de establecer la configuración o mediante una intervención activa del administrador de la red.

Para llevar a cabo la adquisición de vecino, un dispositivo de encaminamiento envía a otro un mensaje «establecer» *("Open")*. Si el encaminador

destino acepta la solicitud, devuelve un mensaje «mantener activa» (*"Keepalive"*) como respuesta.

Una vez establecida la relación de vecino, se utiliza el procedimiento detección de vecino alcanzable para mantener la relación. Cada asociado necesita estar seguro de que el otro asociado existe y está todavía comprometido con la relación de vecino. Con este propósito, ambos dispositivos de encaminamiento se envían periódicamente mensajes *"mantener activa"*.

El último procedimiento especificado por BGP es la detección de red alcanzable. Cada dispositivo de encaminamiento mantiene una base de datos con las redes que puede alcanzar y la ruta preferida para ello. Siempre que se produzca un cambio en esta base de datos, el dispositivo de encaminamiento difunde un mensaje *"actualizar"* (*"Update"*) a todos los otros dispositivos de encaminamiento que implementen BGP. Dado que el mensaje *"actualizar"* se envía por difusión, todos los encaminadores BGP pueden generar y mantener su información de encaminamiento.

Mensajes BGP

La Figura 7.5.4 muestra el formato de todos los mensajes BGP. Cada mensaje comienza con una cabecera de 19 bytes que contiene tres campos, como se indica en la parte sombreada en la figura:

- ✓ Marcador: reservado para autenticación. El emisor puede insertar un valor en este campo que se emplearía como parte de un mecanismo de autenticación para permitir al destino verificar la identidad del emisor.
- ✓ Longitud: longitud del mensaje en bytes.
- ✓ Tipo: tipo de mensaje: establecer, actualizar, notificación o mantener activa.

Para adquirir un vecino, un encaminador abre primero una conexión TCP con el encaminador vecino de interés. Entonces envía un mensaje *"establecer"*. Este mensaje identifica al AS al que pertenece el emisor y proporciona la dirección IP del dispositivo de encaminamiento. También incluye un parámetro de tiempo de mantenimiento, que indica el número de segundos que propone el emisor para el temporizador de mantenimiento. Si el destino está preparado para establecer una relación de vecindad, calcula un valor para el temporizador de mantenimiento como el mínimo entre su tiempo de mantenimiento y el valor de tiempo especificado en el mensaje «establecer». El valor calculado representa el máximo número de segundos que pueden transcurrir entre la recepción en el emisor de mensajes «mantener activa» sucesivos y/o mensajes *"actualizar"*.

El mensaje *"mantener activa"* consta sólo de la cabecera. Cada dispositivo de encaminamiento emite estos mensajes a cada uno de sus pares con suficiente regularidad para evitar que expire su temporizador de mantenimiento.

El mensaje *"actualizar"* facilita dos tipos de información:

- ✓ Información sobre una ruta determinada a través de la interconexión de redes. Esta información se puede incorporar a la base de datos de cualquier dispositivo de encaminamiento que la recibe.
- ✓ Una lista de rutas previamente anunciadas por este dispositivo de encaminamiento que van a ser eliminadas.

Un mensaje *"actualizar"* puede contener uno o ambos tipos de información. La información sobre una ruta particular a través de la red incluye tres campos: el campo de información de accesibilidad de la capa de red (NLRI, Network Layer Reachability Information), el campo de longitud total de los atributos de la ruta y el campo de los atributos de la ruta. El campo NLRI contiene una lista de identificadores de redes que se pueden alcanzar por esta ruta. Cada red se identifica por su dirección IP, que es en realidad una parte de una dirección IP completa. Recuerde que una dirección IP es una cantidad de 32 bits de la forma {red, estación}. El prefijo o parte izquierda de esta cantidad identifica a una red concreta.

El campo atributos de la ruta contiene una lista de atributos que se aplican a esta ruta concreta.

Los atributos definidos son los siguientes:

- ✓ Origen: indica si la información fue generada por un protocolo de dispositivo de encaminamiento interior (por ejemplo, OSPF) o por un protocolo de dispositivo de encaminamiento exterior (en particular, BGP).
- ✓ Camino AS: una lista de los AS que son recorridos por la ruta.
- ✓ Siguiente salto: dirección IP del dispositivo de encaminamiento frontera que se debe usar como siguiente salto para alcanzar los destinos indicados en el campo NLRI.
- ✓ Discriminante de salida múltiple: se emplea para comunicar alguna información sobre rutas internas a un AS. Este atributo se describirá más adelante en esta sección.
- ✓ Preferencias locales: empleado por un dispositivo de encaminamiento para informar a otros dispositivos de encaminamiento de su mismo AS de su grado de preferencia por una ruta particular. No tiene significado alguno para los dispositivos de encaminamiento de otros AS.
- ✓ Agregado atómico, Agente unión (Atomic–aggregate, Aggregator): estos dos campos implementan el concepto de unión de rutas. En esencia, un conjunto de redes interconectadas y su espacio de direcciones correspondiente se pueden organizar jerárquicamente (es decir, como un árbol). En este caso, las direcciones de las redes se estructuran en dos o más partes. Todas las redes de un subárbol comparten una dirección de red parcial común. Usando esta dirección parcial común, la cantidad de información que se debe comunicar en el campo NLRI se puede reducir significativamente.

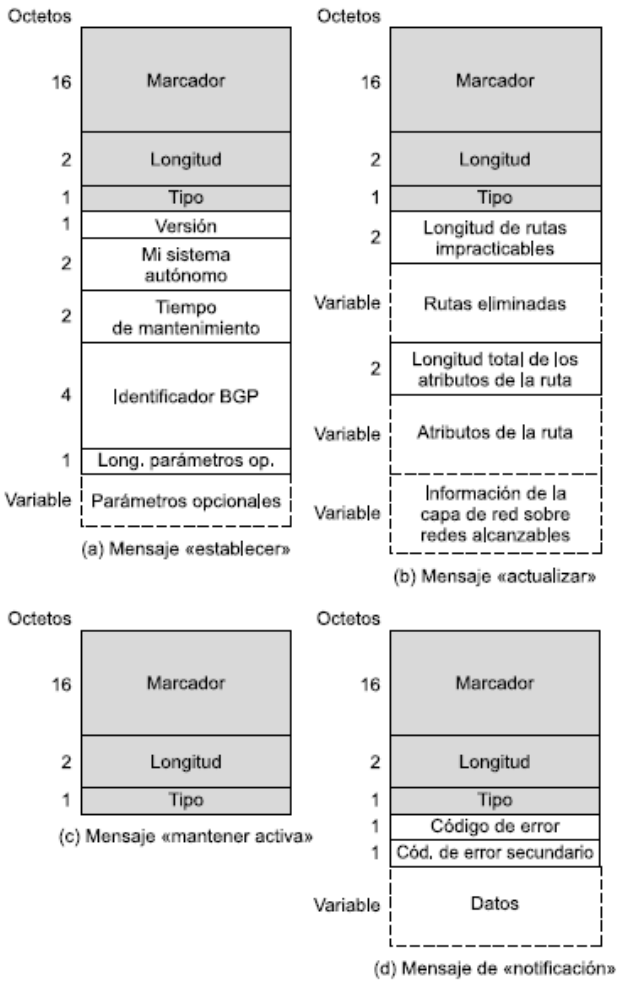

Figura 7.5.4: Formato de mensajes BGP

El atributo *"camino AS"* sirve realmente para dos objetivos. Dado que indica los AS que debe atravesar un datagrama si sigue esta ruta, la información de camino AS permite a un dispositivo de encaminamiento implementar políticas de encaminamiento. Es decir, un dispositivo de encaminamiento puede decidir evitar un camino particular para evitar el paso por un AS concreto. Por ejemplo, la información que es confidencial puede estar limitada a ciertos tipos de AS, o un encaminador puede tener información sobre el rendimiento o calidad de una porción de red que esté incluida en un AS, lo que lleva al encaminador evitar ese AS. Algunos ejemplos de rendimiento o métrica de calidad son: velocidad del enlace, la capacidad, la tendencia a estar congestionado y la calidad global de funcionamiento. Otro criterio que se podría usar es minimizar el número de AS de tránsito.

El lector se puede preguntar por el objetivo del atributo siguiente salto. El dispositivo de encaminamiento que realiza la solicitud querrá conocer necesariamente qué redes se pueden alcanzar a través del encaminador que responde, pero, ¿por qué proporcionar información acerca de otros dispositivos de encaminamiento? Esta cuestión se explica mejor con la ayuda de la Figura 7.3.1. En ese ejemplo, el dispositivo de encaminamiento R1 en el sistema autónomo 1 y el dispositivo de encaminamiento R5 en

el sistema autónomo 2 implementan BGP y establecen una relación de vecindad. R1 envía un mensaje *"actualizar"* a R5 indicando qué redes puede alcanzar y las distancias (saltos de red) implicadas. R1 también proporciona la misma información en representación de R2. Es decir, R1 le dice a R5 qué redes se pueden alcanzar vía R2. En este ejemplo, R2 no implementa BGP. Normalmente, la mayoría de los dispositivos de encaminamiento en un sistema autónomo no implementan BGP. Sólo unos pocos dispositivos de encaminamiento tendrán asignada la responsabilidad de comunicarse con otros encaminadores de otros sistemas autónomos. Un apunte final: R1 tiene la información necesaria sobre R2 debido a que R1 y R2 comparten un protocolo de encaminador interior (IRP).

El segundo tipo de información de actualización consiste en la supresión de una o más rutas. En este caso, la ruta se identifica por la dirección IP de la red destino.

Finalmente, el mensaje de *"notificación"* se envía cuando se detecta una condición de error. Se puede informar de los siguientes errores:

- ✓ Error en la cabecera del mensaje: incluye errores de sintaxis y de autenticación.
- ✓ Error en un mensaje *"establecer"*: incluye errores de sintaxis y opciones no reconocidas en un mensaje *"establecer"*. Este mensaje también se puede utilizar para indicar que el tiempo de mantenimiento en un mensaje *"establecer"* es inaceptable.
- ✓ Error en un mensaje *"actualizar"*: incluye errores de sintaxis y validez en un mensaje *"actualizar"*.
- ✓ Tiempo de mantenimiento expirado: Si el dispositivo de encaminamiento emisor no ha recibido sucesivos mensajes *"mantener activa"* y/o *"actualizar"* y/o mensajes de *"notificación"* durante el tiempo de mantenimiento, entonces se comunica este error y se cierra la conexión.
- ✓ Error en la máquina de estados finitos: Incluye cualquier error de procedimiento.
- ✓ Cese: Utilizado por un dispositivo de encaminamiento para cerrar una conexión con otro encaminador en ausencia de cualquier otro error.

Intercambio de información de encaminamiento de BGP

La esencia de BGP es el intercambio de información de encaminamiento entre dispositivos de encaminamiento participantes en múltiples AS. Este proceso puede ser bastante complejo. A continuación, proporcionaremos una visión simplificada.

Consideremos el dispositivo de encaminamiento R1 en el sistema autónomo 1 (AS1) de la Figura 7.3.1. Para empezar, un encaminador que implemente BGP implementará también un protocolo encaminamiento interno como OSPF. Usando OSPF, R1 puede intercambiar información de encaminamiento con otros dispositivos de encaminamiento dentro de AS1 y construir un esquema de la topología de las redes y dispositivos de encaminamiento en AS1 para construir una tabla de encaminamiento. A

continuación, R1 puede emitir un mensaje «actualizar» a R5 en AS2. El mensaje "actualizar" podría incluir lo siguiente:

- ✓ Camino AS: la identidad de AS1.
- ✓ Siguiente salto: la dirección IP de R1.
- ✓ NLRI: una lista de todas las redes de AS1.

Este mensaje informa a R5 que todas las redes indicadas en NLRI se alcanzan vía R1 y que el único sistema autónomo que hay que atravesar es AS1.

Suponga ahora que R5 también mantiene una relación de vecindad con otro dispositivo de encaminamiento en otro sistema autónomo, digamos R9 en AS3. R5 enviará la información que acaba de recibir de R1 a R9 en un nuevo mensaje *"actualizar"*. Este mensaje incluye lo siguiente:

- ✓ Camino AS: la lista de identificadores {AS2, AS1}.
- ✓ Siguiente salto: la dirección IP de R5.
- ✓ NLRI: una lista de todas las redes en AS1.

Este mensaje informa a R9 de que todas las redes indicadas en NLRI son alcanzables vía R5 y que los sistemas autónomos que hay que atravesar son AS2 y AS1. R9 debe decidir si ésta es ahora su ruta preferida hacia las redes indicadas. R9 podría conocer una ruta alternativa a alguna o a todas esas redes, prefiriéndola por razones de rendimiento o algún otro criterio métrico. Si R9 decide que la ruta proporcionada en el mensaje de actualización de R5 es preferible, entonces incorpora la información de encaminamiento en su base de datos de encaminamiento y propaga la nueva información a otros vecinos. Este mensaje nuevo incluirá un campo camino AS del tipo {AS3,AS2, AS1}.

De esta forma, la información de encaminamiento de actualización se propaga a través de la interconexión de redes mayor, que consta a su vez de varios sistemas autónomos interconectados.

El campo camino AS se emplea para asegurar que el mensaje no circula indefinidamente: si un dispositivo de encaminamiento recibe un mensaje «actualizar» en un AS que esté incluido en el campo camino AS, ese encaminador no enviará la información de actualización a otros encaminadores.

Los dispositivos de encaminamiento de un mismo AS, denominados vecinos internos, pueden intercambiar información BGP. En este caso, el dispositivo de encaminamiento emisor no incorpora el identificador del AS común al campo camino AS. Cuando un encaminador ha seleccionado una ruta a un destino externo como preferida, transmite esta ruta a todos sus vecinos internos. Cada uno de estos dispositivos de encaminamiento decide entonces si la nueva ruta pasa a ser la preferida, en cuyo caso incorpora la nueva ruta a su base de datos y envía un nuevo mensaje *"actualizar"*.

8. Cuando hay disponibles múltiples puntos de entrada a un AS desde un dispositivo de encaminamiento fronterizo de otro AS, el atributo discriminante de salida múltiple puede utilizarse para elegir uno de ellos. Este atributo contiene un número que refleja alguna métrica interna para alcanzar los destinos dentro de un AS. Por ejemplo, suponga que en la Figura 7.3.1 los dispositivos de encaminamiento Rl y R2 implementan BGP, y que ambos tienen una relación de vecindad con R5. Cada uno envía un mensaje *"actualizar"* a R5 para la red 1.3 que incluye una métrica de encaminamiento utilizada internamente en AS1, al igual que la métrica de encaminamiento asociada con el protocolo de encaminador interno OSPF. R5 podría usar entonces estas dos métricas nuevas como criterio para elegir entre las dos Capítulo 7: Protocolos de Ruteo

Uno de los aspectos más complejos y cruciales del diseño de redes de conmutación de paquetes incluidas internet y las redes privadas, es la implementación de una solución eficiente al encaminamiento (ruteo) de datagramas. En términos generales, el ruteo es una función que busca encontrar *"la mejor ruta"* para comunicar dos sistemas finales.

Este capítulo comienza con una breve descripción general de los problemas relacionados con el diseño del ruteo. A continuación, se analizan distintas opciones de ruteo, algoritmos y su implementación a través de protocolos.

8.1. El problema del encaminamiento (ruteo)

La función principal de una red de conmutación de paquetes, internet o una red privada interconectada por enrutadores (routers) es aceptar paquetes procedentes de una estación emisora y enviarlos hacia una estación destino. Para ello se debe determinar una ruta a través de la red, siendo posible generalmente la existencia de más de una. Así pues, se debe realizar una función de encaminamiento, entre cuyos requisitos se encuentran los siguientes: exactitud, imparcialidad, simplicidad, optimización, robustez, eficiencia, y estabilidad.

Las dos primeras características mencionadas se explican por sí mismas. La robustez está relacionada con la habilidad de la red para enviar paquetes de alguna forma ante la aparición de sobrecargas y fallos localizados. Idealmente, la red puede reaccionar ante estas contingencias sin sufrir pérdidas de paquetes o caída de circuitos virtuales. No obstante, la robustez puede implicar cierta inestabilidad. Las técnicas que

reaccionan ante condiciones cambiantes presentan una tendencia no deseable a reaccionar demasiado lentamente ante determinados eventos o a experimentar oscilaciones inestables de una situación extrema a otra. Por ejemplo, la red puede reaccionar ante la aparición de congestión en un área desplazando la mayor parte de la carga hacia una segunda zona. Ahora será la segunda región la que estará sobrecargada y la primera infrautilizada, produciéndose un segundo desplazamiento del tráfico. Durante estos desplazamientos puede ocurrir que los paquetes viajen en bucles a través de la red.

También existe un compromiso entre la característica de imparcialidad y el hecho de que el encaminamiento trate de ser óptimo. Algunos criterios de funcionamiento pueden dar prioridad al intercambio de paquetes entre estaciones vecinas frente al intercambio realizado entre estaciones distantes, lo cual puede maximizar la eficiencia promedio, pero será injusto para aquella estación que necesite comunicar principalmente con estaciones lejanas.

Finalmente, una técnica de encaminamiento implica cierto coste de procesamiento en cada nodo y, en ocasiones, también un coste en la transmisión, impidiéndose en ambos casos el funcionamiento eficiente de la red. Este coste debe ser inferior a los beneficios obtenidos por el uso de una métrica razonable, como la mejora de la robustez o la imparcialidad. A continuación, un resumen de los elementos que hay que tener en cuenta al momento de diseñar una solución de encaminamiento:

- ✓ Criterios de rendimiento
  - Número de saltos
  - Coste
  - Retardo
  - Eficiencia
- ✓ Instante de decisión
  - Paquete (datagrama)
  - Sesión (circuitos virtuales)
9. Lugar de decisión
   - Cada nodo (distribuido)
   - Nodo central (centralizado)
   - Nodo origen (fuente)
- ✓ Fuente de información de la red
  - Ninguna
  - Local
  - Nodo adyacente
  - Nodos a lo largo de la ruta
  - Todos los nodos
- ✓ Tiempo de actualización de la información de la red
  - Continuo
  - Periódico
  - Cambio importante en la carga
  - Cambio en la topología

Criterios de rendimiento

La elección de una ruta se fundamenta generalmente en algún criterio de rendimiento. El más simple consiste en elegir el camino con menor número de saltos (aquel que atraviesa el menor número de nodos) a través de la red. Éste es un criterio que se puede medir fácilmente y que debería minimizar el consumo de recursos de la red. Una generalización del criterio de menor número de saltos lo constituye el encaminamiento de mínimo coste. En este caso se asocia un coste a cada enlace y, para cualesquiera dos estaciones conectadas, se elige aquella ruta a través de la red que implique el coste total mínimo. Por ejemplo, en la Figura 7.1.1 se muestra una red en la que las dos líneas con

flecha entre cada par de nodos representan un enlace entre ellos, y los números asociados indican el coste actual del enlace en cada sentido. El camino más corto (menor número de saltos) desde el nodo 1 hasta el 6 es 1-3-6 (costo = 5+5 = 10), pero el de mínimo costo es 1-4-5- 6 (costo = 1+1+2 = 4). La asignación de los costes a los enlaces se hace en función de los objetivos de diseño; por ejemplo, el coste podría estar inversamente relacionado con la velocidad (es decir, a mayor velocidad menor coste) o con el retardo actual de la cola asociada al enlace. En el primer caso, la ruta de mínimo coste maximizaría la eficiencia, mientras que en el segundo minimizaría el retardo.

Tanto en la técnica de menor número de saltos como en la de mínimo coste, el algoritmo para determinar la ruta o camino óptimo entre dos estaciones es relativamente sencillo, siendo el tiempo de procesamiento aproximadamente el mismo en ambos casos. Dada su mayor flexibilidad, el criterio de mínimo coste es más utilizado que el de menor número de saltos.

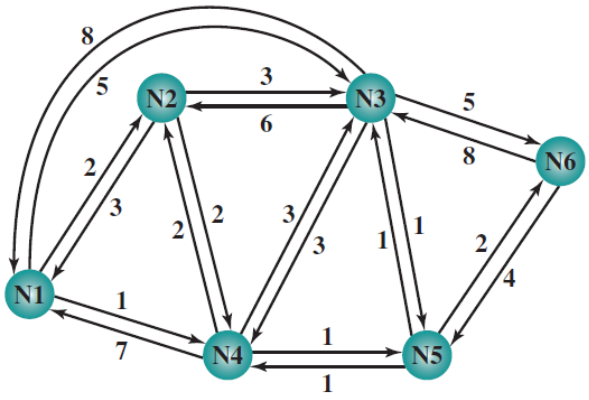

Figura 7.1.1: Red con costos asociados a sus enlaces

Instante y lugar de decisión

Las decisiones de encaminamiento se realizan de acuerdo con algún criterio de rendimiento. Dos cuestiones importantes en la toma de esta decisión son el instante temporal y el lugar en que se toma la decisión.

El instante de decisión viene determinado por el hecho de que la decisión de encaminamiento se hace en base a un paquete o a un circuito virtual. Cuando la operación interna de la red se basa en datagramas, la decisión de encaminamiento se toma de forma individual para cada paquete.

El término lugar de decisión hace referencia al nodo o nodos en la red responsables de la decisión de encaminamiento. El más común es el encaminamiento distribuido, en el que cada nodo de la red tiene la responsabilidad de seleccionar un enlace de salida sobre el que llevar a cabo el envío de los paquetes a medida que éstos se reciben. En el encaminamiento centralizado, la decisión se toma por parte de algún nodo designado al respecto, como puede ser un centro de control de red. El peligro de esta última aproximación es que el fallo del centro de control puede bloquear el funcionamiento de la red; así pues, aunque la aproximación distribuida puede resultar más

compleja es también más robusta. Una tercera alternativa empleada en algunas redes es la conocida como encaminamiento desde el origen. En este caso, es la estación origen y no los nodos de la red quien realmente toma la decisión de encaminamiento, comunicándosela a la red. Esto permite al usuario fijar una ruta a través de la red de acuerdo con criterios locales al mismo.

7.6. Estrategias de encaminamiento

Existen numerosas estrategias de encaminamiento para abordar las necesidades de encaminamiento en redes de conmutación de paquetes. Muchas de ellas son aplicables también al encaminamiento en la interconexión de redes.

7.2.4. Encaminamiento estático

En el encaminamiento estático se configura una única ruta permanente para cada par de nodos origen-destino en la red, pudiéndose utilizar para ello cualquiera de los algoritmos de encaminamiento de mínimo coste conocidos (Dijkstra, Bellman-Ford, entre otros), o directamente definir y cargar manualmente las rutas. Las rutas son fijas (al menos mientras lo sea la topología de la red), de modo que los costes de enlace usados para el diseño de las rutas no pueden estar basados en variables dinámicas como el tráfico, aunque sí podrían estarlo en tráfico esperado o en capacidad.

La Figura 7.2.1.1 sugiere cómo se pueden implementar rutas estáticas. Se crea una matriz de encaminamiento central, almacenada, por ejemplo, en un centro de control de red. Esta matriz especifica para cada par de nodos origen-destino, la identidad del siguiente nodo en la ruta. Luego mediante comandos del sistema operativo del router se cargan cada una de las rutas.

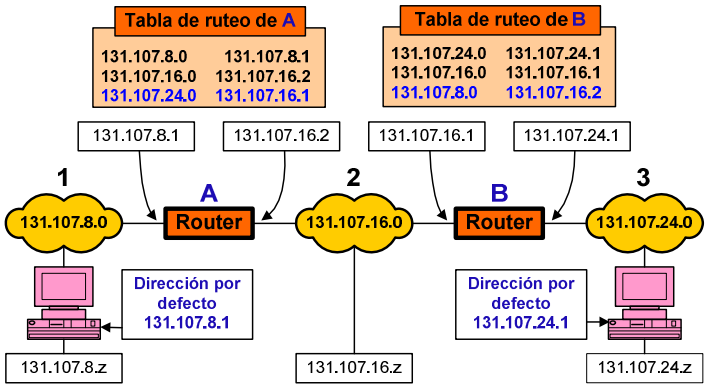

Figura 7.2.1.1: Tablas de ruteo estáticas

7.2.5. Encaminamiento dinámico: Primera Generación (Vector Distancia)

El algoritmo de encaminamiento original, diseñado en 1969, era un algoritmo adaptable distribuido que hacía uso de la estimación de los retardos como criterio de rendimiento y de una versión del algoritmo de Bellman-Ford. Para este algoritmo, cada nodo mantiene dos vectores:

$$D_i = \begin{bmatrix} d_{i1} \\ \vdots \\ d_{iN} \end{bmatrix} \qquad S_i = \begin{bmatrix} s_{i1} \\ \vdots \\ s_{iN} \end{bmatrix}$$

Dónde,

D_i = vector de retardo para el nodo i.

D_{ij} = estimación actual del retardo mínimo desde el nodo i al nodo j (dii%0).

N = número de nodos en la red.

S_i = vector de nodos sucesores para el nodo i.

S_{ij} = nodo siguiente en la ruta actual de mínimo retardo de i a j.

Periódicamente (cada 128 ms), cada nodo intercambia su vector de retardo con todos sus vecinos. A partir de todos los vectores de retardo recibidos, un nodo k actualiza sus dos vectores como sigue:

$$d_{kj} = \min_{i \in A} [d_{ij} + l_{ki}]$$

$s_{kj} = i$, siendo i el que minimiza la expresión anterior

En la Figura 7.2.2.1 se muestra un ejemplo del algoritmo original de ARPANET, usando la red de la Figura 7.2.2.2. En la Figura 7.2.2.1(a) se muestra la tabla de encaminamiento del nodo 1 en un instante de tiempo que refleja los costes asociados a los enlaces de la Figura 7.2.2.2. Para cada destino se especifica un retardo y el nodo siguiente en la ruta que lo produce. En algún momento, los costes de los enlaces cambian a los valores indicados en la Figura 7.1.1. Suponiendo que los vecinos del nodo 1 (nodos 2, 3 y 4) conocieran el cambio antes que él, cada uno de estos nodos actualizará su vector de retardo y enviará una copia a todos sus vecinos, incluyendo el nodo 1 (7.2.2.1(b)). El nodo 1 desecha su tabla de encaminamiento y construye una nueva basándose en los vectores de retardo recibidos y en la propia estimación que él hace del retardo para cada uno de los enlaces de salida a sus vecinos. El resultado obtenido se muestra en la Figura 7.2.2.1(c).

El retardo de enlace estimado no es más que el tamaño o longitud de la cola para el enlace en cuestión. Así, con la construcción de una nueva tabla de

encaminamiento el nodo tiende a favorecer aquellos enlaces con menores colas, lo que compensa la carga entre las distintas líneas de salida. Sin embargo, dado que el tamaño de las colas varía rápidamente a lo largo del tiempo, la percepción distribuida de la ruta más corta podría cambiar mientras un paquete se encuentra en tránsito. Esto podría provocar una situación en la que un paquete se encamina hacia un área de baja congestión en lugar de hacia el destino.

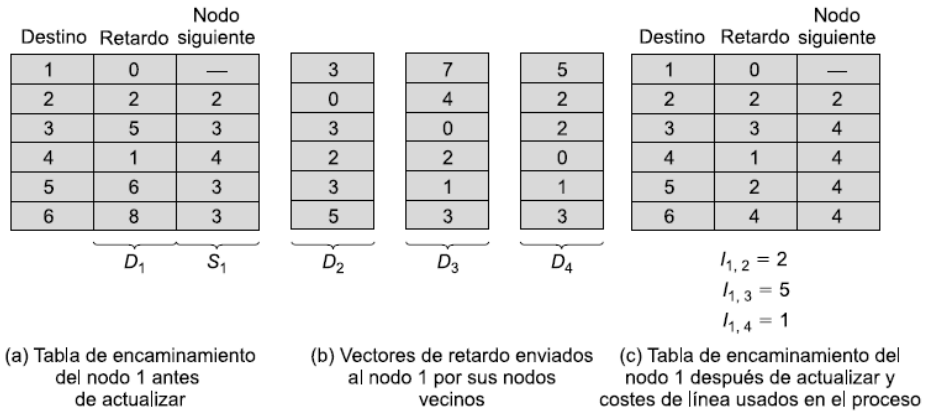

Figura 7.2.2.1: Algoritmo de encaminamiento vector distancia

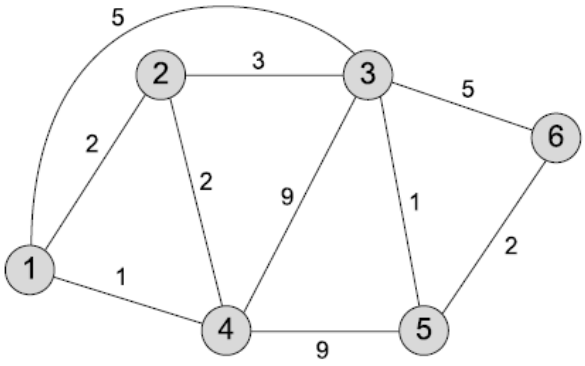

Figura 7.2.2.2: Red para construir la matriz de la figura 7.2.2.1(a)

7.2.6. Encaminamiento dinámico: Segunda Generación (Estado del enlace)

El enrutamiento por vector de distancia se utilizó en ARPANET hasta 1979, cuando se reemplazó por el enrutamiento por estado del enlace. El principal problema que provocó su desaparición era que, con frecuencia, el algoritmo tardaba demasiado en converger una vez que cambiaba la topología de la red (debido al problema del conteo al infinito). En consecuencia, se reemplazó por un algoritmo totalmente nuevo, ahora conocido como enrutamiento por estado del enlace.

La idea detrás del enrutamiento por estado del enlace es bastante simple y se puede enunciar en cinco partes. Cada enrutador debe realizar lo siguiente para hacerlo funcionar:

1. Descubrir a sus vecinos y conocer sus direcciones de red.
2. Establecer la métrica de distancia o de costo para cada uno de sus vecinos.
3. Construir un paquete que indique todo lo que acaba de aprender.
4. Enviar este paquete a todos los demás enrutadores y recibir paquetes de ellos.
5. Calcular la ruta más corta a todos los demás enrutadores.

De hecho, se distribuye la topología completa a todos los enrutadores. Después se puede ejecutar el algoritmo de Dijkstra en cada enrutador para encontrar la ruta más corta a los demás enrutadores. A continuación, veremos con mayor detalle cada uno de estos cinco pasos.

Aprender sobre los vecinos

Cuando un enrutador se pone en funcionamiento, su primera tarea es averiguar quiénes son sus vecinos. Para lograr esto envía un paquete especial HELLO en cada línea punto a punto. Se espera que el enrutador del otro extremo regrese una respuesta en la que indique su nombre. Estos nombres deben ser globalmente únicos puesto que, cuando un enrutador distante escucha después que hay tres enrutadores conectados a F, es indispensable que pueda determinar si los tres se refieren al mismo F.

Establecimiento de los costos de los enlaces

El algoritmo de enrutamiento por estado del enlace requiere que cada enlace tenga una métrica de distancia o costo para encontrar las rutas más cortas. El costo para llegar a los vecinos se puede establecer de modo automático, o el operador de red lo puede configurar. Una elección común es hacer el costo inversamente proporcional al ancho de banda del enlace. Por ejemplo, una red Ethernet de 1 Gbps puede tener un costo de 1 y una red Ethernet de 100 Mbps un costo de 10. Esto hace que las rutas de mayor capacidad sean mejores opciones.

Si la red está geográficamente dispersa, el retardo de los enlaces se puede considerar en el costo, de modo que las rutas a través de enlaces más cortos sean mejores opciones. La manera más directa de determinar este retardo es enviar un paquete especial ECO a través de la línea, que el otro extremo tendrá que regresar de inmediato. Si se mide el tiempo de ida y vuelta, y se divide entre dos, el enrutador emisor puede obtener una estimación razonable del retardo.

Construcción de los paquetes de estado del enlace

Una vez que se ha recabado la información necesaria para el intercambio, el siguiente paso es que cada enrutador construya un paquete que contenga todos los datos. El paquete comienza con la identidad del emisor, seguida de un número de secuencia, una edad (se describirá luego) y una lista de vecinos. También se proporciona el costo para cada vecino. En la figura 7.2.3.1(a) se muestra una red de ejemplo; los costos se muestran como etiquetas en las líneas. En la figura 7.2.3.1(b) se muestran los paquetes de estado del enlace correspondientes para los seis enrutadores.

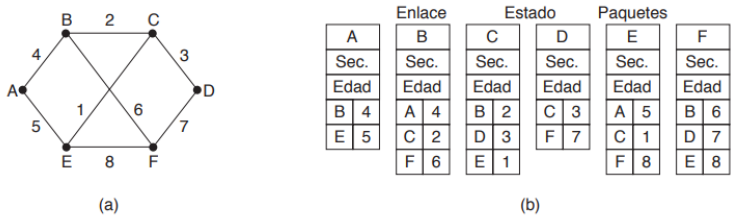

Figura 7.2.3.1: (a) La red, (b) Paquetes de estado de enlace para esta red

Es fácil construir los paquetes de estado del enlace. La parte difícil es determinar cuándo construirlos. Una posibilidad es construirlos de manera periódica; es decir, a intervalos regulares. Otra posibilidad es construirlos cuando ocurra un evento significativo, como la caída o la reactivación de una línea o de un vecino, o cuando sus propiedades cambien en forma considerable.

Distribución de los paquetes de estado del enlace

La parte más complicada del algoritmo es la distribución de los paquetes de estado del enlace. Todos los enrutadores deben recibir todos los paquetes de estado del enlace con rapidez y confiabilidad. Si se utilizan distintas versiones de la topología, las rutas que se calculen podrían tener inconsistencias como ciclos, máquinas inalcanzables y otros problemas.

Primero se describirá el algoritmo básico de distribución. Después se le aplicará algunos refinamientos. La idea fundamental es utilizar inundación para distribuir los paquetes de estado del enlace a todos los enrutadores. Con el fin de mantener controlada la inundación, cada paquete contiene un número de secuencia que se incrementa con cada nuevo paquete enviado. Los enrutadores llevan el registro de todos los pares (enrutador de origen, secuencia) que ven. Cuando llega un nuevo paquete de estado del enlace, se verifica y compara con la lista de paquetes ya vistos. Si es nuevo, se reenvía a través de todas las líneas, excepto aquella por la que llegó. Si es un duplicado, se descarta. Si llega un paquete con número de secuencia menor que el mayor visto hasta el momento, se rechaza como obsoleto debido a que el enrutador tiene datos más recientes.

Este algoritmo tiene algunos problemas, pero son manejables. Primero, si los números de secuencia vuelven a comenzar, reinará la confusión. La solución aquí es utilizar un número de secuencia de 32 bits. Con un paquete de estado del enlace por segundo, el tiempo para volver a empezar será de 137 años, por lo que podemos ignorar esta posibilidad. Segundo, si llega a fallar un enrutador, perderá el registro de su número de secuencia. Si comienza nuevamente en 0, se rechazará como duplicado el siguiente paquete que envíe. Tercero, si llega a corromperse un número de secuencia y se recibe 65540 en vez de 4 (un error de 1 bit), los paquetes 5 a 65 540 se rechazarán como obsoletos, dado que se piensa que el número de secuencia actual es 65 540.

La solución a todos estos problemas es incluir la edad de cada paquete después del número de secuencia y disminuirla una vez cada segundo. Cuando la edad llega a cero, se descarta la información de ese enrutador. Por lo general, un paquete nuevo entra, por ejemplo, cada 10 segundos, por lo que la información de los enrutadores sólo expira cuando un enrutador está caído (o cuando se pierden seis paquetes consecutivos, un evento poco probable). Los enrutadores también decrecen el campo Edad durante el proceso inicial de inundación para asegurar que no pueda perderse ningún paquete y sobrevivir durante un periodo de tiempo indefinido (se descarta el paquete cuya edad sea cero).

Algunos refinamientos a este algoritmo pueden hacerlo más robusto. Una vez que llega un paquete de estado del enlace a un enrutador para ser inundado, no se encola para su transmisión de inmediato, sino que se coloca en un área de almacenamiento para esperar un tiempo corto, en caso de que se activen o desactiven más enlaces. Si llega otro paquete de estado del enlace proveniente del mismo origen antes de que se transmita el primer paquete, se comparan sus números de secuencia. Si son iguales, se descarta el duplicado. Si son diferentes, se desecha el más antiguo. Como protección contra los errores en los enlaces, se confirma la recepción de todos los paquetes de estado del enlace

Cálculo de las nuevas rutas

Una vez que un enrutador ha acumulado un conjunto completo de paquetes de estado del enlace, puede construir el grafo de toda la red debido a que todos los enlaces están simbolizados. De hecho, cada enlace se representa dos veces, una para cada dirección. Las distintas direcciones pueden tener incluso costos diferentes. Así, los cálculos de la ruta más corta pueden encontrar rutas del enrutador A a B que sean distintas a las del enrutador de B a A.

Ahora se puede ejecutar localmente el algoritmo de Dijkstra para construir las rutas más cortas a todos los destinos posibles. Los resultados de este algoritmo indican al enrutador qué enlace debe usar para llegar a cada destino. Esta información se instala en las tablas de enrutamiento y se puede reanudar la operación normal.

En comparación con el enrutamiento por vector de distancia, el enrutamiento por estado del enlace requiere más memoria y poder de cómputo. Para una red con n enrutadores, cada uno de los cuales tiene k vecinos, la memoria requerida para almacenar los datos de entrada es proporcional a kn, que es por lo menos tan grande como

una tabla de enrutamiento que lista todos los destinos. Además, el tiempo de cómputo también crece con más rapidez que kn, incluso con las estructuras de datos más eficientes; un problema en las grandes redes. Sin embargo, en muchas situaciones prácticas, el enrutamiento por estado del enlace funciona bien debido a que no sufre de los problemas de convergencia lenta.

7.7. Sistema Autónomo (Autonomous System)

Para continuar con el análisis sobre los protocolos de encaminamiento, se necesita introducir el concepto de sistema autónomo. Un sistema autónomo (AS, Autonomous System) posee las siguientes características:

1. Un AS se compone de un conjunto de encaminadores y redes gestionados por una única organización.
2. Un AS consiste en un grupo de dispositivos de encaminamiento que intercambian información a través de un protocolo de encaminamiento común.
3. Excepto en momentos de avería, un AS está conectado (en un sentido teórico de grafo). Es decir, existe un camino entre cualquier par de nodos.

Un protocolo común de encaminamiento, al que nos referiremos como protocolo de ruteo interior (IRP, Interior Router Protocol), distribuye la información de encaminamiento entre los dispositivos de encaminamiento dentro de un AS. El protocolo que se emplea dentro de un sistema autónomo no necesita ser implementado fuera del sistema. Esta flexibilidad permite que los IRP se hagan a medida para aplicaciones y requisitos específicos.

Puede ocurrir, sin embargo, que una interconexión de redes esté constituida por más de un AS. Por ejemplo, todas las LAN de una organización, como puede ser un complejo de oficinas o un campus, podrían estar enlazadas mediante encaminadores para formar un AS. Este sistema se podría unir a otros AS a través de una red de área amplia. Esta situación se muestra en la Figura 7.3.1.

En este caso, los algoritmos de encaminamiento y la información de las tablas de encaminamiento utilizadas por los encaminadores en los distintos AS pueden ser diferentes. Sin embargo, los encaminadores de un AS necesitan al menos un nivel mínimo de información referente a las redes externas al sistema que puedan alcanzar. El protocolo que se utiliza para pasar información de encaminamiento entre diferentes AS se conoce como protocolo de ruteo exterior (ERP, Exterior Router Protocol)[2].

[2] En la bibliografía relacionada se utilizan a menudo los términos protocolo de pasarela interior (IGP, Interior Gateway Protocol) y protocolo de pasarela exterior (EGP, Exterior Gateway Protocol) para designar lo que aquí denominamos IRP y ERP.

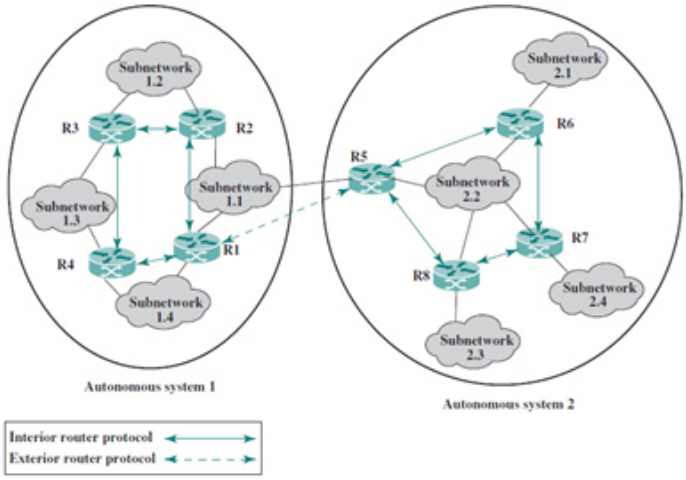

Figura 7.3.1: Protocolos de ruteo interior y exterior

7.8. Protocolo de ruteo interno: OSPF (Open Shortest Path First)

OSPF (Open Shortest Path First) es un protocolo de encaminamiento dinámico de pasarela interior o IGP (Interior Gateway Protocol), que se creó para solucionar las limitaciones que tenía el protocolo RIP (Routing Information Protocol). Este protocolo permite gestionar redes con un diámetro mayor que 16, mejorando además el tiempo de convergencia y la agregación de rutas.

El algoritmo utilizado por OSPF es más complejo que el utilizado por RIP, por lo que se necesitan routers con más potencia de procesador y memoria, y se requiere más tiempo de diseño e implementación. De este modo, se puede afirmar que ambos protocolos se han diseñado para entornos totalmente distintos: OSPF está diseñado para redes grandes y complejas, mientras que RIP está destinado a redes pequeñas con una configuración sencilla.

A diferencia del algoritmo de vector de distancia utilizado por RIP, OSPF se basa en un algoritmo de estado de los enlaces (link state). Por ello, este protocolo no envía a los encaminadores adyacentes el número de saltos que los separa, sino el estado del enlace que los separa. De esta manera, cada router es capaz de construir un mapa completo del estado de la red y, por consecuencia, puede elegir en cada momento la ruta más apropiada para enviar un mensaje a un destino dado.

Como cada router contiene el mismo mapa de la topología de la red, OSPF no requiere que las actualizaciones se envíen a intervalos regulares. De este modo, OSPF

reduce el consumo de red necesario para el intercambio de actualizaciones mediante la multidifusión, enviando una actualización sólo cuando se detecta un cambio (en lugar del envío periódico) y enviando cambios de la tabla de encaminamiento (en vez de la tabla completa) sólo cuando es necesaria una actualización. Además, el hecho de no tener que ir incrementando el número de saltos cada vez que se pasa por un router intermedio se traduce en una cantidad de información a intercambiar mucho menos abundante y, por tanto, en un ancho de banda libre mejor que en el caso de RIP.

La métrica utilizada es más sofisticada que en el caso de RIP, ya que se basa en el ancho de banda del enlace (por omisión, coste = 10^8/ancho de banda (b/s)). Por ejemplo, para el caso de un enlace mediante Ethernet 10Mb/s se tiene un coste de 10.

Existen diferentes versiones de este protocolo: OSPFv1 (RFC1131 y RFC1247); OSPFv2 (RFC2328); y OSPFv3 (RFC2740), el cual está adaptado para IPv6. En las redes encaminadas por OSPF se define la siguiente jerarquía (Figura 7.4.1):

- ✓ Área: Constituye una frontera para el cálculo en la base de datos del estado del enlace. Los routers que están en la misma área contienen la misma base de datos topológica. Un área es en realidad una subdivisión de un AS (Autonomous System). El área principal se denomina backbone y a ésta se conectan el resto de áreas (sean conexiones física o virtualmente). Las diferentes áreas se conectan entre sí mediante unos routers de borde que se encargan de intercambiar las diferentes tablas de encaminamiento.
- ✓ ABR (Area Border Router): Router que contiene enlaces en varias áreas dentro del mismo AS, cuya función consiste en resumir las informaciones de encaminamiento y gestionar los intercambios de rutas entre áreas.
- ✓ IR (Internal Router): Router cuyos enlaces pertenecen todos a un área determinada.
- ✓ DR (Designated Router): Cada segmento de red tiene un DR y un BDR, por lo que un router conectado a múltiples redes puede ser DR de un segmento y un router normal del otro segmento. En realidad, es la interfaz del router la que actúa como DR o BDR. La principal función del DR es minimizar el *"flooding"* (inundación por anuncios) y la sincronización de las DB's (Data Bases) centralizando el intercambio de información. De este modo, los routers de un mismo segmento no intercambian información del estado del enlace entre ellos, sino que lo hacen con el DR.
- ✓ BDR (Backup Designated Router): Es el router elegido como anunciante secundario (generalmente el segundo con la prioridad más alta). El BDR no hace nada mientras haya un DR en la red (solo actúa si el DR falla). Un BDR detecta que un DR falla porque durante un cierto tiempo no escucha LSA's (Link State Advertisements).
- ✓ ASBR (Autonomous System Border Router): Router que contiene enlaces a distintos AS y que sirve de gateway entre OSPF y otros protocolos de encaminamiento (IGRP, EIGRP, ISIS, RIP, BGP, Static).

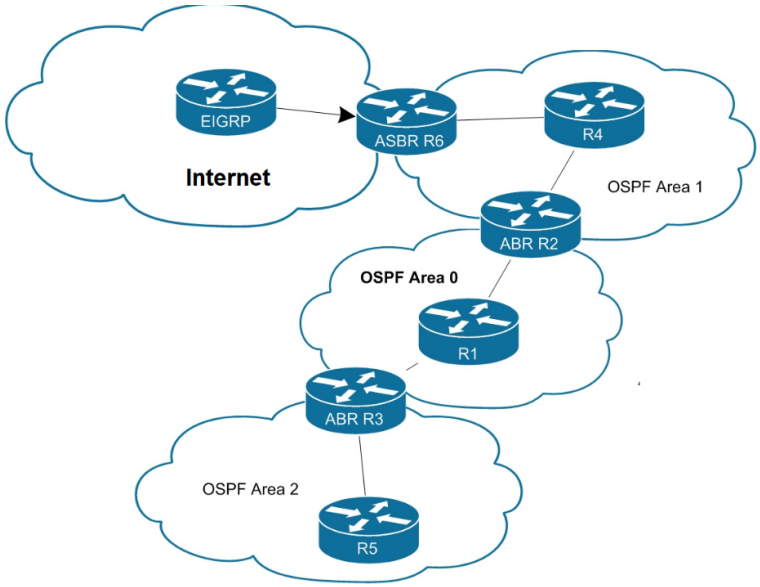

Figura 7.4.1: Topología del protocolo OSPF

En un principio, cada router identifica (o conoce por configuración) a sus vecinos inmediatos. De entre todos los routers de un mismo segmento se eligen un router principal (DR) y un router principal de seguridad (que lo sustituirá en caso de que éste falle). Generalmente, se elige como DR al router con la prioridad más alta de entre todos los routers que pertenecen al mismo segmento de red (la prioridad varía entre 0 y 255). Como la prioridad por defecto suele ser 1, para desempatar se usa el que tenga mayor RID o router ID (donde el router ID suele ser la @IP más alta de una interfaz activa del router). Los routers con prioridad igual a 0 no pueden ser elegidos como DR's.

Una vez que se han elegido el DR y el BDR, se pasa a la fase de descubrimiento de rutas. Para ello, el DR y el DBR forman una adyacencia con cada uno de los routers de su mismo segmento de red. Una adyacencia es una relación que se establece entre un router y su DR y BDR mediante el protocolo *"Hello"*.

En cada adyacencia, uno de los dos routers actúa como master (el de mayor *"routerID"*, que suele ser el DR) y el otro como slave (esclavo). El master envía un resumen de su DB al slave y este la reconoce y viceversa. Luego, el slave compara la información recibida y pide que le envíe aquellas entradas que no tiene. De este modo, cada router difunde al DR los mensajes LSA con los cambios que se producen en la red y el DR se encarga de actualizar la base de datos de cada uno de los routers de su segmento de red haciendo flooding de la información de encaminamiento.

A partir de la base de datos con la topología de la red, en cada router se calculan localmente los caminos más cortos a todos los destinos mediante el algoritmo SPF (Shortest Path First) de Djkstra para poder construir así la tabla de encaminamiento. Cuando se tienen redes con una gran cantidad de routers el número de LSU's enviado produce un gran consumo de ancho de banda, a la vez que se hacen también grandes el tiempo de convergencia y el tamaño de las bases de datos de los routers para guardar la

información sobre la topología de toda la red. Por ello, el encaminamiento OSPF propone como solución la división de la red de una forma jerárquica en áreas, que delimita el dominio de envío de los mensajes LSA a un conjunto de routers y redes en un mismo AS.

Por otra parte, en una red multi área se tienen distintos tipos de mensajes LSA:

- ✓ Router LSA (tipo 1): Cada router genera estos paquetes hacia el resto de routers de su misma área, indicando la lista de sus vecinos inmediatos y el coste (métrica) de sus enlaces.
- ✓ Network LSA (tipo 2): Estos paquetes son generados por los routers DR (de una red BMA) sobre los routers vinculados a esa red BMA y solo se envían dentro del área.
- ✓ Summary LSA (tipo 3): Estos paquetes son generados por los ABR para anunciar las redes internas procedentes de un área específica a otros ABR del mismo AS. Se genera un resumen por cada subred de cada área hacia las demás áreas. Estas informaciones se envían primeramente al backbone (área 0), el cual se encargará después de distribuirlas hacia el resto de áreas del AS.
- ✓ ASBR summary LSA (tipo 4): Generados por los ABR's describen rutas al ASBR's. Los routers ABR deben propagar también las informaciones de encaminamiento hacia los ASBR para que éstos puedan conocer cómo alcanzar los routers externos en otras AS (Autonomous System).
- ✓ AS external LSA (tipo 5): Generados por los ASBR's describen rutas externas al AS (entre ellas la ruta por defecto para salir del AS).

Los mensajes de encaminamiento OSPF se encapsulan como un protocolo de transporte con número 89 (Figura 7.4.2):

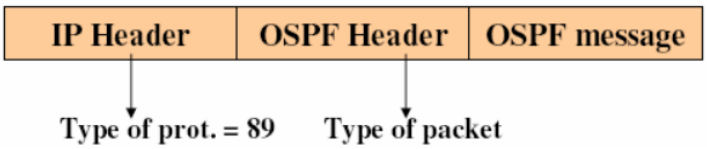

Figura 7.4.2: Encapsulamiento paquete OSPF

Todos los paquetes OSPF incluyen en su cabecera información básica relacionada con el router (Figura 7.4.3):

- ✓ Version: Identifica la versión OSPF.
- ✓ Type: Identifica el tipo de paquete OSPF. Hay 5 tipos de paquetes OSPF:
  - HELLO packets (Type = 1).
  - Database Description (DBD) packets (Type = 2).

- Link-State Request (LSR) packets (Type = 3).
- Link-State Update (LSU) packets (Type = 4).
- Link-State ACK (LSAck) packets (Type = 5).

✓ Packet Length: Longitud del paquete (incluida la cabecera OSPF).

✓ Router ID (RID): Identifica el origen del paquete OSPF (normalmente cada router escoge como RID la @IP mayor entre las @IP activas del mismo).

✓ Area ID: Identifica el área al cual pertenece el paquete OSPF.

✓ Checksum.

✓ Authentication type:

- Type 0: no authentication.
- Type 1: clear-text password or simple authentication.
- Type 2: cryptographic or MD5 authentication.

✓ Authentication information: Contiene la información de autentificación.

✓ Data: Encapsula información de encaminamiento.

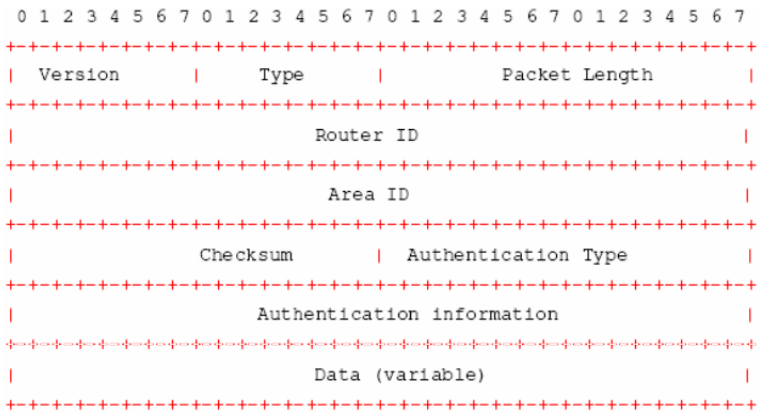

Figura 7.4.3: Cabecera de un paquete OSPF

Paquetes "*Hello*"

Los paquetes Hello (saludo) se utilizan para verificar comunicaciones bidireccionales, anunciar requerimientos de vecindad y elegir routers designados (Figura 7.4.4). Así, los anuncios permiten establecer y mantener una adyacencia, que consiste en una conexión virtual a un vecino para poder transferir anuncios de estado del enlace.

Estos paquetes de saludo se envían a intervalos periódicos de tiempo (HelloInterval = 10 segundos) usando la dirección multicast 224.0.0.5. Los paquetes Hello tienen el formato siguiente:

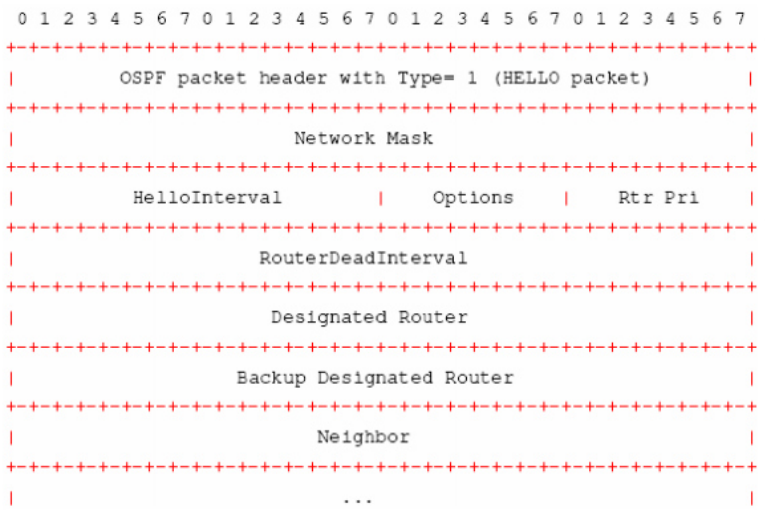

Figura 7.4.4: Cabecera paquete *"Hello"* en OSPF

- ✓ Network Mask: Máscara asociada con esa interfaz.
- ✓ Hello Interval: Intervalo en que se envían paquetes HELLO (10 segundos).
- ✓ Options: Capacidades opcionales que soporta este router.
- ✓ Router Priority: Prioridad (por defecto =1).
- ✓ Router-Dead-Interval: Tiempo que espera un router hasta que deja de considerar que un vecino está activo (4*HelloInterval).
- ✓ DR y BDR: Direcciones IP de ambos (0.0.0.0 si inicialmente desconocidas y hay que descubrirlos).
- ✓ Neighbours: RouteID de cada vecino que ha escuchado durante los últimos RouterDead-Interval segundos.

Por otro lado, se tienen los mensajes DBD, LSR, LSU y LSAck:

Paquetes de descripción de base de datos (DBD): describe el contenido de las DB, incluyendo encabezamientos LSA (no todo el LSA) para que el router receptor confirme que tiene todos los LSA requeridos.

- ✓ Paquetes de petición de estado del enlace (LSR): Solicitan a los vecinos los LSA que están en el listado de petición de estado del enlace. Este listado lleva un registro de los LSA que deben solicitarse porque no se dispone de ellos o porque no se tiene la versión más actualizada. El router sabe qué paquetes le faltan gracias a la recepción de paquetes de petición de otros routers o de paquetes de descripción de base de datos de otros routers.

- ✓ Paquetes de actualización de estado del enlace (LSU): Suministran los LSA (Link State Advertisements) a los routers remotos.
- ✓ Paquetes de acuse de recibo de estado del enlace (LSAck): Sirven de acuse de recibo explícito a uno o más LSU.

Los LSA's (Link-State Advertisements) son unidades de datos que describen el estado local de un router o red. Para un router, esto incluye el estado de las interfaces del router y sus adyacencias. Un LSA va empaquetado en paquetes DBD, LSU, LSR o LSAck.

7.9. Protocolo de ruteo externo: BGP (Border Gateway Protocol)

Dentro de un solo sistema autónomo, OSPF e IS-IS son los protocolos de uso común. Entre los sistemas autónomos se utiliza un protocolo diferente, conocido como BGP (Protocolo de Puerta de Enlace de Frontera, del inglés Border Gateway Protocol). Se necesita un protocolo diferente debido a que los objetivos de un protocolo intra dominio y de un protocolo inter dominio no son los mismos. Todo lo que tiene que hacer un protocolo intra dominio es mover paquetes de la manera más eficiente posible desde el origen hasta el destino. No tiene que preocuparse por las políticas.

En contraste, los protocolos de enrutamiento inter dominio tienen que preocuparse en gran manera por la política. Por ejemplo, tal vez un sistema autónomo corporativo desee la habilidad de enviar paquetes a cualquier sitio de Internet y recibir paquetes de cualquier sitio de Internet. Sin embargo, quizás no esté dispuesto a llevar paquetes de tránsito que se originen en un AS foráneo y estén destinados a un AS foráneo diferente, aun cuando su propio AS se encuentre en la ruta más corta entre los dos sistemas autónomos foráneos ("Ése es su problema, no el nuestro"). Por otro lado, podría estar dispuesto a llevar el tráfico del tránsito para sus vecinos o incluso para otros sistemas autónomos específicos que hayan pagado por este servicio. Por ejemplo, las compañías telefónicas podrían estar contentas de actuar como empresas portadoras para sus clientes, pero no para otros. En general, los protocolos de puerta de enlace exterior (y BGP en particular) se han diseñado para permitir que se implementen muchos tipos de políticas de enrutamiento en el tráfico entre sistemas autónomos.

Las políticas típicas implican consideraciones políticas, de seguridad, o económicas. Algunos ejemplos de posibles restricciones de enrutamiento son:

1. No transportar tráfico comercial en la red educativa.
2. Nunca enviar tráfico del Pentágono por una ruta a través de Irak.
3. Usar TeliaSonera en vez de Verizon porque es más económico.
4. No usar AT&T en Australia porque el desempeño es pobre.
5. El tráfico que empieza o termina en Apple no debe transitar por Google.

Como puede imaginar de esta lista, las políticas de enrutamiento pueden ser muy individuales. A menudo son propietarias pues contienen información comercial delicada. Sin embargo, podemos describir algunos patrones que capturan el razonamiento anterior de la compañía y que se utilizan con frecuencia como un punto de partida.

Para implementar una política de enrutamiento hay que decidir qué tráfico puede fluir a través de cuáles enlaces entre los sistemas autónomos. Una política común es que un ISP cliente pague a otro ISP proveedor por entregar paquetes a cualquier otro destino en Internet y recibir los paquetes enviados desde cualquier otro destino. Se dice que el ISP cliente compra servicio de tránsito al ISP proveedor. Es justo igual que cuando un cliente doméstico compra servicio de acceso a Internet con un ISP. Para que funcione, el proveedor debe anunciar las rutas a todos los destinos en Internet al cliente a través del enlace que los conecta. De esta forma, el cliente tendrá una ruta para enviar paquetes a cualquier parte. Por el contrario, el cliente sólo debe anunciar al proveedor las rutas a los destinos en su red. Esto permitirá al proveedor enviar tráfico al cliente sólo para esas direcciones; al cliente no le conviene manejar el tráfico destinado a otras partes.

En la figura 7.5.1 podemos ver un ejemplo del servicio de tránsito. Hay cuatro sistemas autónomos conectados. Con frecuencia, la conexión se hace mediante un enlace en IXP's (Puntos de Intercambio de Internet, del inglés Internet eXchange Points): instalaciones con las que muchos ISP tienen un enlace para fines de conectarse con otros ISP. AS2, AS3 y AS4 son clientes de AS1, pues le compran servicio de tránsito. Así, cuando la fuente A envía al destino C, los paquetes viajan de AS2 hacia AS1 y finalmente a AS4. Los anuncios de enrutamiento viajan en dirección opuesta a los paquetes. AS4 anuncia a C como un destino a su proveedor de tránsito AS1, para que las fuentes puedan llegar a C por medio de AS1. Después, AS1 anuncia a sus otros clientes, incluyendo AS2, una ruta a C para que éstos sepan que pueden enviar tráfico a C por medio de AS1.

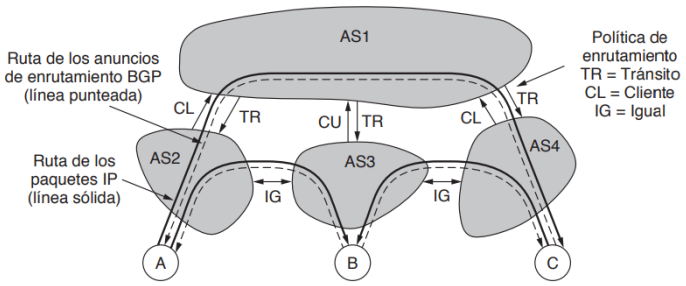

Figura 7.5.1: Política de enrutamiento en cuatro Sistemas Autónomos

En la figura 7.5.1, los demás sistemas autónomos compran servicio de tránsito a AS1. Este servicio les proporciona conectividad para que puedan interactuar con cualquier host en Internet. Sin embargo, tienen que pagar por este privilegio. Suponga que AS2 y AS3 intercambian mucho tráfico. Dado que sus redes ya se encuentran conectadas, si lo desean pueden usar una política diferente: pueden enviar tráfico directamente uno al otro sin costo. Esto reducirá la cantidad de tráfico que debe entregar

AS1 a cuenta de AS2 y AS3, y con suerte reducirá sus facturas. A esta política se le conoce como comunicación entre pares (peering).

Para implementar la comunicación entre pares, dos sistemas autónomos se envían anuncios de enrutamiento entre sí, respecto a las direcciones que residen en sus redes. Al hacer esto, AS2 puede enviar a AS3 paquetes de "A" destinados a "B" y viceversa. Sin embargo, hay que tener en cuenta que la comunicación entre iguales no es transitiva. En la Figura 7.5.1, AS3 y AS4 también se comunican entre sí. Esta comunicación de igual a igual permite que el tráfico de "C", que está destinado a "B", se envíe directamente a AS4. ¿Qué ocurre si "C" envía un paquete a "A"? AS3 sólo está anunciando a AS4 una ruta a "B". No está anunciando una ruta a "A". La consecuencia es que el tráfico no pasará de AS4 a AS3 a AS2, aun cuando existe una ruta física. Esta restricción es justo lo que AS3 quiere. Se comunica con AS4 para intercambiar tráfico, pero no quiere transportar tráfico de AS4 a otras partes de Internet, ya que no se le paga por hacerlo. En cambio, AS4 recibe servicio de tránsito de AS1. Por ende, es AS1quien transportará el paquete de "C" a "A".

Ahora que sabemos sobre el tránsito y la comunicación entre iguales, también podemos ver que, "A", "B" y "C" tienen arreglos de tránsito. Por ejemplo, "A" debe comprar acceso a Internet a AS2. A podría ser una sola computadora doméstica o la red de una compañía con muchas LAN. Sin embargo, no necesita ejecutar BGP debido a que es una red aislada (stub network) que está conectada al resto de Internet sólo mediante un enlace. Por tanto, el único lugar para enviar paquetes destinados a puntos que estén fuera de la red es a través del enlace a AS2. No hay ningún otro lugar a dónde ir. Para arreglar esta ruta, sólo hay que establecer una ruta predeterminada. Por esta razón no hemos mostrado a "A", "B" y "C" como sistemas autónomos que participan en el enrutamiento interdominio.

Por otro lado, las redes de algunas compañías están conectadas a varios ISP. Esta técnica se utiliza para mejorar la confiabilidad, ya que si la ruta a través de un ISP falla, la compañía puede usar la ruta a través del otro ISP. Esta técnica se conoce como multihoming. En este caso, la red de la compañía probablemente ejecute un protocolo de enrutamiento interdominio (como BGP) para indicar a otros sistemas autónomos qué enlaces de ISP pueden llegar a cuáles direcciones.

Hay muchas variaciones posibles de estas políticas de tránsito y comunicación entre iguales, pero todas ilustran cómo las relaciones de negocios y el control sobre el camino que pueden tomar los anuncios de rutas pueden implementar distintos tipos de políticas. Ahora consideraremos con más detalle cómo los enrutadores que ejecutan BGP se anuncian rutas entre sí y seleccionan rutas a través de las cuales pueden reenviar los paquetes.

BGP es una forma de protocolo de vector de distancia, aunque es bastante distinto a los protocolos de vector de distancia intra dominio como RIP. Ya vimos antes que la política, y no la distancia mínima, se utiliza para elegir qué rutas usar. Otra gran diferencia es que, en vez de mantener sólo el costo de la ruta a cada destino, cada enrutador BGP lleva el registro de la ruta utilizada. Esta metodología se conoce como protocolo de vector de ruta. La ruta consiste en el enrutador del siguiente salto (que puede

estar del otro lado del ISP y no necesariamente ser adyacente) y la secuencia de sistemas autónomos (o ruta AS) en el recorrido (se proporciona en orden inverso). Por último, los pares de enrutadores BGP se comunican entre sí mediante el establecimiento de conexiones TCP. Al operar de esta forma se obtiene una comunicación confiable; además se ocultan todos los detalles de la red que se está atravesando.

En la figura 7.5.2 se muestra un ejemplo de cómo se anuncian las rutas BGP. Hay tres sistemas autónomos y el de en medio provee tránsito a los ISP izquierdo y derecho. Un anuncio de ruta para el prefijo C empieza en AS3. Cuando se propaga a través del enlace a R2c en la parte superior de la figura, tiene la ruta AS que consiste tan sólo en AS3 y el enrutador del siguiente salto de R3a. En la parte inferior tiene la misma ruta AS, pero un siguiente salto distinto, debido a que provino de un enlace diferente. Este anuncio continúa su propagación y atraviesa el límite hacia AS1. En el enrutador R1a, en la parte superior de la figura, la ruta AS es AS2, AS3 y el siguiente salto es R2a.

Al transportar la ruta completa con el recorrido, es fácil para el enrutador receptor detectar e interrumpir los ciclos de enrutamiento. La regla es que cada enrutador que envíe un recorrido hacia fuera del AS anteponga su propio número de AS a este recorrido (ésta es la razón por la cual la lista está en orden inverso). Cuando un enrutador recibe un recorrido, verifica que su propio número de AS ya se encuentre en la ruta AS. Si es así, se ha detectado un ciclo y se descarta el anuncio. No obstante (aunque suene un poco irónico), a finales de la década de 1990 se descubrió que, a pesar de esta precaución, BGP sufre de una versión del problema de conteo al infinito (Labovitz, 2000) . No hay ciclos de larga duración, pero algunas veces los recorridos pueden ser lentos para converger y tienen ciclos transitorios.

Proporcionar una lista de sistemas autónomos es una forma muy burda de especificar una ruta. Un AS podría ser una compañía pequeña o una red troncal internacional. No hay forma de saberlo con base en la ruta. BGP ni siquiera lo intenta, ya que los distintos sistemas autónomos pueden usar protocolos intra dominio diferentes, cuyos costos no se puedan comparar. Incluso si se pudieran comparar, tal vez un AS no quiera revelar su métrica interna. Ésta es una de las diferencias entre los protocolos de enrutamiento inter dominio y los protocolos intra dominio.

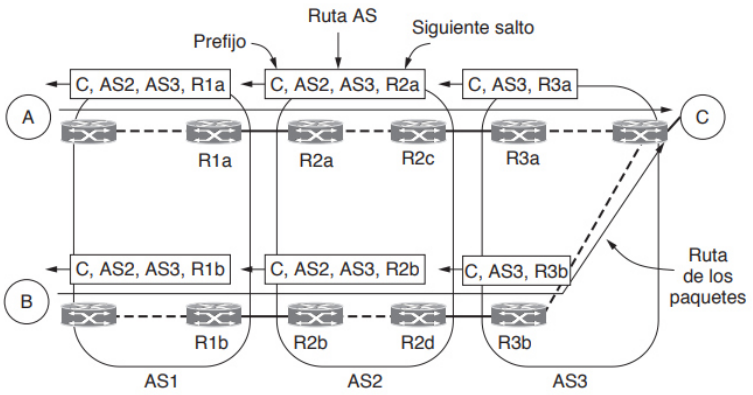

Figura 7.5.2: Anuncia de rutas en BGP

Funciones

Como se mencionó, BGP se diseñó para permitir la cooperación en el intercambio de información de encaminamiento entre dispositivos de encaminamiento de diferentes sistemas autónomos (AS), llamados pasarelas en el estándar. El protocolo opera en términos de mensajes, que se envían utilizando conexiones TCP. El repertorio de mensajes se resume en la Tabla 7.5.3. La versión actual de BGP se conoce como BGP-4 (RFC 1771).

Tabla 7.5.3: Mensajes BGP

Establecer	Utilizado para establecer una relación de vecindad con otro dispositivo de encaminamiento
Actualizar	Utilizado para (1) transmitir información acerca de una única ruta y/o (2) enumerar rutas múltiples que se vayan a eliminar.
Mantener Activa	Utilizado para (1) confirmar un mensaje «establecer» y (2) confirmar periódicamente la relación de vecindad.
Notificación	Se envía cuando se detecta una condición de error.

BGP supone tres procedimientos funcionales, que son:

- ✓ Adquisición de vecino.
- ✓ Detección de vecino alcanzable.
- ✓ Detección de red alcanzable.

Dos dispositivos de encaminamiento se considera que son vecinos si están conectados a la misma subred. Si los dos encaminadores se encuentran en sistemas autónomos diferentes, podrían desear intercambiar información de encaminamiento. Para este cometido, es necesario primero realizar la operación de adquisición de vecino. Básicamente, la adquisición de un vecino se produce cuando dos dispositivos de encaminamiento vecinos de diferentes sistemas autónomos se ponen de acuerdo en intercambiar regularmente información de encaminamiento. Se requiere un procedimiento formal de adquisición debido a que uno de los encaminadores podría no querer participar. Por ejemplo, el dispositivo de encaminamiento puede estar sobresaturado y no querer ser responsable de tráfico que llegue de fuera del sistema. En el proceso de adquisición de un vecino, un dispositivo de encaminamiento envía un mensaje de petición al otro, el cual puede aceptar o rechazar el ofrecimiento. El protocolo no aborda la cuestión de cómo puede un dispositivo de encaminamiento conocer la dirección o incluso la existencia de otro encaminador, ni siquiera de cómo decide que necesita intercambiar información de encaminamiento con un encaminador en particular. Estas cuestiones deben ser tratadas en el momento de establecer la configuración o mediante una intervención activa del administrador de la red.

Para llevar a cabo la adquisición de vecino, un dispositivo de encaminamiento envía a otro un mensaje «establecer» *("Open")*. Si el encaminador

destino acepta la solicitud, devuelve un mensaje «mantener activa» (*"Keepalive"*) como respuesta.

Una vez establecida la relación de vecino, se utiliza el procedimiento detección de vecino alcanzable para mantener la relación. Cada asociado necesita estar seguro de que el otro asociado existe y está todavía comprometido con la relación de vecino. Con este propósito, ambos dispositivos de encaminamiento se envían periódicamente mensajes *"mantener activa"*.

El último procedimiento especificado por BGP es la detección de red alcanzable. Cada dispositivo de encaminamiento mantiene una base de datos con las redes que puede alcanzar y la ruta preferida para ello. Siempre que se produzca un cambio en esta base de datos, el dispositivo de encaminamiento difunde un mensaje *"actualizar"* (*"Update"*) a todos los otros dispositivos de encaminamiento que implementen BGP. Dado que el mensaje *"actualizar"* se envía por difusión, todos los encaminadores BGP pueden generar y mantener su información de encaminamiento.

Mensajes BGP

La Figura 7.5.4 muestra el formato de todos los mensajes BGP. Cada mensaje comienza con una cabecera de 19 bytes que contiene tres campos, como se indica en la parte sombreada en la figura:

- ✓ Marcador: reservado para autenticación. El emisor puede insertar un valor en este campo que se emplearía como parte de un mecanismo de autenticación para permitir al destino verificar la identidad del emisor.
- ✓ Longitud: longitud del mensaje en bytes.
- ✓ Tipo: tipo de mensaje: establecer, actualizar, notificación o mantener activa.

Para adquirir un vecino, un encaminador abre primero una conexión TCP con el encaminador vecino de interés. Entonces envía un mensaje *"establecer"*. Este mensaje identifica al AS al que pertenece el emisor y proporciona la dirección IP del dispositivo de encaminamiento. También incluye un parámetro de tiempo de mantenimiento, que indica el número de segundos que propone el emisor para el temporizador de mantenimiento. Si el destino está preparado para establecer una relación de vecindad, calcula un valor para el temporizador de mantenimiento como el mínimo entre su tiempo de mantenimiento y el valor de tiempo especificado en el mensaje «establecer». El valor calculado representa el máximo número de segundos que pueden transcurrir entre la recepción en el emisor de mensajes «mantener activa» sucesivos y/o mensajes *"actualizar"*.

El mensaje *"mantener activa"* consta sólo de la cabecera. Cada dispositivo de encaminamiento emite estos mensajes a cada uno de sus pares con suficiente regularidad para evitar que expire su temporizador de mantenimiento.

El mensaje *"actualizar"* facilita dos tipos de información:

- ✓ Información sobre una ruta determinada a través de la interconexión de redes. Esta información se puede incorporar a la base de datos de cualquier dispositivo de encaminamiento que la recibe.
- ✓ Una lista de rutas previamente anunciadas por este dispositivo de encaminamiento que van a ser eliminadas.

Un mensaje *"actualizar"* puede contener uno o ambos tipos de información. La información sobre una ruta particular a través de la red incluye tres campos: el campo de información de accesibilidad de la capa de red (NLRI, Network Layer Reachability Information), el campo de longitud total de los atributos de la ruta y el campo de los atributos de la ruta. El campo NLRI contiene una lista de identificadores de redes que se pueden alcanzar por esta ruta. Cada red se identifica por su dirección IP, que es en realidad una parte de una dirección IP completa. Recuerde que una dirección IP es una cantidad de 32 bits de la forma {red, estación}. El prefijo o parte izquierda de esta cantidad identifica a una red concreta.

El campo atributos de la ruta contiene una lista de atributos que se aplican a esta ruta concreta.

Los atributos definidos son los siguientes:

- ✓ Origen: indica si la información fue generada por un protocolo de dispositivo de encaminamiento interior (por ejemplo, OSPF) o por un protocolo de dispositivo de encaminamiento exterior (en particular, BGP).
- ✓ Camino AS: una lista de los AS que son recorridos por la ruta.
- ✓ Siguiente salto: dirección IP del dispositivo de encaminamiento frontera que se debe usar como siguiente salto para alcanzar los destinos indicados en el campo NLRI.
- ✓ Discriminante de salida múltiple: se emplea para comunicar alguna información sobre rutas internas a un AS. Este atributo se describirá más adelante en esta sección.
- ✓ Preferencias locales: empleado por un dispositivo de encaminamiento para informar a otros dispositivos de encaminamiento de su mismo AS de su grado de preferencia por una ruta particular. No tiene significado alguno para los dispositivos de encaminamiento de otros AS.
- ✓ Agregado atómico, Agente unión (Atomic–aggregate, Aggregator): estos dos campos implementan el concepto de unión de rutas. En esencia, un conjunto de redes interconectadas y su espacio de direcciones correspondiente se pueden organizar jerárquicamente (es decir, como un árbol). En este caso, las direcciones de las redes se estructuran en dos o más partes. Todas las redes de un subárbol comparten una dirección de red parcial común. Usando esta dirección parcial común, la cantidad de información que se debe comunicar en el campo NLRI se puede reducir significativamente.

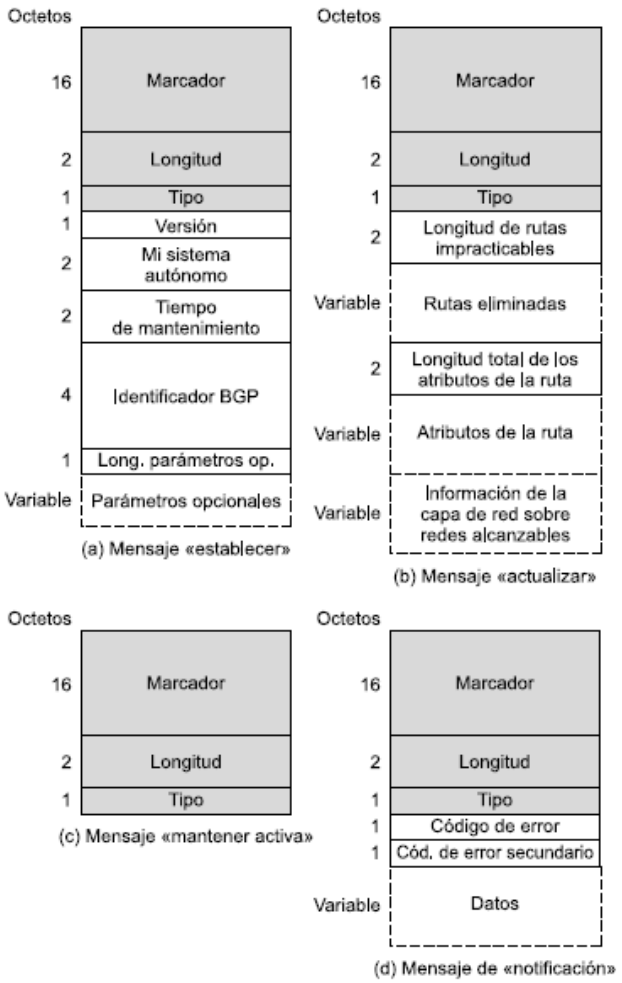

Figura 7.5.4: Formato de mensajes BGP

El atributo *"camino AS"* sirve realmente para dos objetivos. Dado que indica los AS que debe atravesar un datagrama si sigue esta ruta, la información de camino AS permite a un dispositivo de encaminamiento implementar políticas de encaminamiento. Es decir, un dispositivo de encaminamiento puede decidir evitar un camino particular para evitar el paso por un AS concreto. Por ejemplo, la información que es confidencial puede estar limitada a ciertos tipos de AS, o un encaminador puede tener información sobre el rendimiento o calidad de una porción de red que esté incluida en un AS, lo que lleva al encaminador evitar ese AS. Algunos ejemplos de rendimiento o métrica de calidad son: velocidad del enlace, la capacidad, la tendencia a estar congestionado y la calidad global de funcionamiento. Otro criterio que se podría usar es minimizar el número de AS de tránsito.

El lector se puede preguntar por el objetivo del atributo siguiente salto. El dispositivo de encaminamiento que realiza la solicitud querrá conocer necesariamente qué redes se pueden alcanzar a través del encaminador que responde, pero, ¿por qué proporcionar información acerca de otros dispositivos de encaminamiento? Esta cuestión se explica mejor con la ayuda de la Figura 7.3.1. En ese ejemplo, el dispositivo de encaminamiento R1 en el sistema autónomo 1 y el dispositivo de encaminamiento R5 en

el sistema autónomo 2 implementan BGP y establecen una relación de vecindad. R1 envía un mensaje *"actualizar"* a R5 indicando qué redes puede alcanzar y las distancias (saltos de red) implicadas. R1 también proporciona la misma información en representación de R2. Es decir, R1 le dice a R5 qué redes se pueden alcanzar vía R2. En este ejemplo, R2 no implementa BGP. Normalmente, la mayoría de los dispositivos de encaminamiento en un sistema autónomo no implementan BGP. Sólo unos pocos dispositivos de encaminamiento tendrán asignada la responsabilidad de comunicarse con otros encaminadores de otros sistemas autónomos. Un apunte final: R1 tiene la información necesaria sobre R2 debido a que R1 y R2 comparten un protocolo de encaminador interior (IRP).

El segundo tipo de información de actualización consiste en la supresión de una o más rutas. En este caso, la ruta se identifica por la dirección IP de la red destino.

Finalmente, el mensaje de *"notificación"* se envía cuando se detecta una condición de error. Se puede informar de los siguientes errores:

- ✓ Error en la cabecera del mensaje: incluye errores de sintaxis y de autenticación.
- ✓ Error en un mensaje *"establecer"*: incluye errores de sintaxis y opciones no reconocidas en un mensaje *"establecer"*. Este mensaje también se puede utilizar para indicar que el tiempo de mantenimiento en un mensaje *"establecer"* es inaceptable.
- ✓ Error en un mensaje *"actualizar"*: incluye errores de sintaxis y validez en un mensaje *"actualizar"*.
- ✓ Tiempo de mantenimiento expirado: Si el dispositivo de encaminamiento emisor no ha recibido sucesivos mensajes *"mantener activa"* y/o *"actualizar"* y/o mensajes de *"notificación"* durante el tiempo de mantenimiento, entonces se comunica este error y se cierra la conexión.
- ✓ Error en la máquina de estados finitos: Incluye cualquier error de procedimiento.
- ✓ Cese: Utilizado por un dispositivo de encaminamiento para cerrar una conexión con otro encaminador en ausencia de cualquier otro error.

Intercambio de información de encaminamiento de BGP

La esencia de BGP es el intercambio de información de encaminamiento entre dispositivos de encaminamiento participantes en múltiples AS. Este proceso puede ser bastante complejo. A continuación, proporcionaremos una visión simplificada.

Consideremos el dispositivo de encaminamiento R1 en el sistema autónomo 1 (AS1) de la Figura 7.3.1. Para empezar, un encaminador que implemente BGP implementará también un protocolo encaminamiento interno como OSPF. Usando OSPF, R1 puede intercambiar información de encaminamiento con otros dispositivos de encaminamiento dentro de AS1 y construir un esquema de la topología de las redes y dispositivos de encaminamiento en AS1 para construir una tabla de encaminamiento. A

continuación, R1 puede emitir un mensaje «actualizar» a R5 en AS2. El mensaje "actualizar" podría incluir lo siguiente:

- ✓ Camino AS: la identidad de AS1.
- ✓ Siguiente salto: la dirección IP de R1.
- ✓ NLRI: una lista de todas las redes de AS1.

Este mensaje informa a R5 que todas las redes indicadas en NLRI se alcanzan vía R1 y que el único sistema autónomo que hay que atravesar es AS1.

Suponga ahora que R5 también mantiene una relación de vecindad con otro dispositivo de encaminamiento en otro sistema autónomo, digamos R9 en AS3. R5 enviará la información que acaba de recibir de R1 a R9 en un nuevo mensaje *"actualizar"*. Este mensaje incluye lo siguiente:

- ✓ Camino AS: la lista de identificadores {AS2, AS1}.
- ✓ Siguiente salto: la dirección IP de R5.
- ✓ NLRI: una lista de todas las redes en AS1.

Este mensaje informa a R9 de que todas las redes indicadas en NLRI son alcanzables vía R5 y que los sistemas autónomos que hay que atravesar son AS2 y AS1. R9 debe decidir si ésta es ahora su ruta preferida hacia las redes indicadas. R9 podría conocer una ruta alternativa a alguna o a todas esas redes, prefiriéndola por razones de rendimiento o algún otro criterio métrico. Si R9 decide que la ruta proporcionada en el mensaje de actualización de R5 es preferible, entonces incorpora la información de encaminamiento en su base de datos de encaminamiento y propaga la nueva información a otros vecinos. Este mensaje nuevo incluirá un campo camino AS del tipo {AS3,AS2, AS1}.

De esta forma, la información de encaminamiento de actualización se propaga a través de la interconexión de redes mayor, que consta a su vez de varios sistemas autónomos interconectados.

El campo camino AS se emplea para asegurar que el mensaje no circula indefinidamente: si un dispositivo de encaminamiento recibe un mensaje «actualizar» en un AS que esté incluido en el campo camino AS, ese encaminador no enviará la información de actualización a otros encaminadores.

Los dispositivos de encaminamiento de un mismo AS, denominados vecinos internos, pueden intercambiar información BGP. En este caso, el dispositivo de encaminamiento emisor no incorpora el identificador del AS común al campo camino AS. Cuando un encaminador ha seleccionado una ruta a un destino externo como preferida, transmite esta ruta a todos sus vecinos internos. Cada uno de estos dispositivos de encaminamiento decide entonces si la nueva ruta pasa a ser la preferida, en cuyo caso incorpora la nueva ruta a su base de datos y envía un nuevo mensaje *"actualizar"*.

Cuando hay disponibles múltiples puntos de entrada a un AS desde un dispositivo de encaminamiento fronterizo de otro AS, el atributo discriminante de salida múltiple

puede utilizarse para elegir uno de ellos. Este atributo contiene un número que refleja alguna métrica interna para alcanzar los destinos dentro de un AS. Por ejemplo, suponga que en la Figura 7.3.1 los dispositivos de encaminamiento Rl y R2 implementan BGP, y que ambos tienen una relación de vecindad con R5. Cada uno envía un mensaje *"actualizar"* a R5 para la red 1.3 que incluye una métrica de encaminamiento utilizada internamente en AS1, al igual que la métrica de encaminamiento asociada con el protocolo de encaminador interno OSPF. R5 podría usar entonces estas dos métricas nuevas como criterio para elegir entre las dos