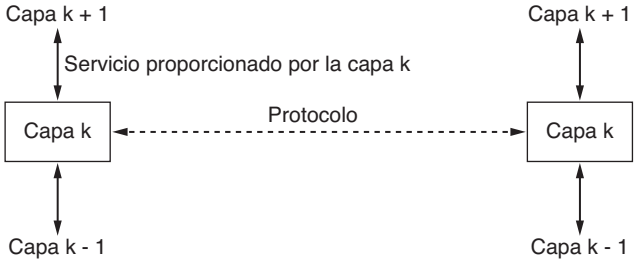

Figura 1-19. La relación entre un servicio y un protocolo.

Vale la pena mencionar una analogía con los lenguajes de programación. Un servicio es como un tipo de datos abstracto o un objeto en un lenguaje orientado a objetos. Define las operaciones que se pueden realizar en un objeto, pero no especifica cómo se implementan estas operaciones. En contraste, un protocolo se relaciona con la *implementación* del servicio y como tal, no es visible al usuario del mismo.

Muchos protocolos antiguos no diferenciaban el servicio del protocolo. En efecto, una capa típica podría tener una primitiva de servicio SEND PACKET en donde el usuario proporcionaba un apuntador hacia un paquete completamente ensamblado. Este arreglo significaba que los usuarios podían ver de inmediato todos los cambios en el protocolo. Ahora, la mayoría de los diseñadores de redes consideran dicho diseño como un error garrafal.

1.4 MODELOS DE REFERENCIA

Ahora que hemos analizado en lo abstracto las redes basadas en capas, es tiempo de ver algunos ejemplos. Analizaremos dos arquitecturas de redes importantes: el modelo de referencia OSI y el modelo de referencia TCP/IP. Aunque ya casi no se utilizan los *protocolos* asociados con el modelo OSI, el *modelo* en sí es bastante general y sigue siendo válido; asimismo, las características en cada nivel siguen siendo muy importantes. El modelo TCP/IP tiene las propiedades opuestas: el modelo en sí no se utiliza mucho, pero los protocolos son usados ampliamente. Por esta razón veremos ambos elementos con detalle. Además, algunas veces podemos aprender más de los fracasos que de los éxitos.

1.4.1 El modelo de referencia OSI

El modelo OSI se muestra en la figura 1-20 (sin el medio físico). Este modelo se basa en una propuesta desarrollada por la Organización Internacional de Normas (ISO) como el primer paso hacia la estandarización internacional de los protocolos utilizados en las diversas capas (Day y Zimmerman, 1983). Este modelo se revisó en 1995 (Day, 1995) y se le llama **Modelo de referencia OSI** (**Interconexión de Sistemas Abiertos**, del inglés *Open Systems Interconnection*) de la ISO puesto que se ocupa de la conexión de sistemas abiertos; esto es, sistemas que están abiertos a la comunicación con otros sistemas. Para abreviar, lo llamaremos **modelo OSI**.

El modelo OSI tiene siete capas. Los principios que se aplicaron para llegar a las siete capas se pueden resumir de la siguiente manera:

1. Se debe crear una capa en donde se requiera un nivel diferente de abstracción.
2. Cada capa debe realizar una función bien definida.
3. La función de cada capa se debe elegir teniendo en cuenta la definición de protocolos estandarizados internacionalmente.

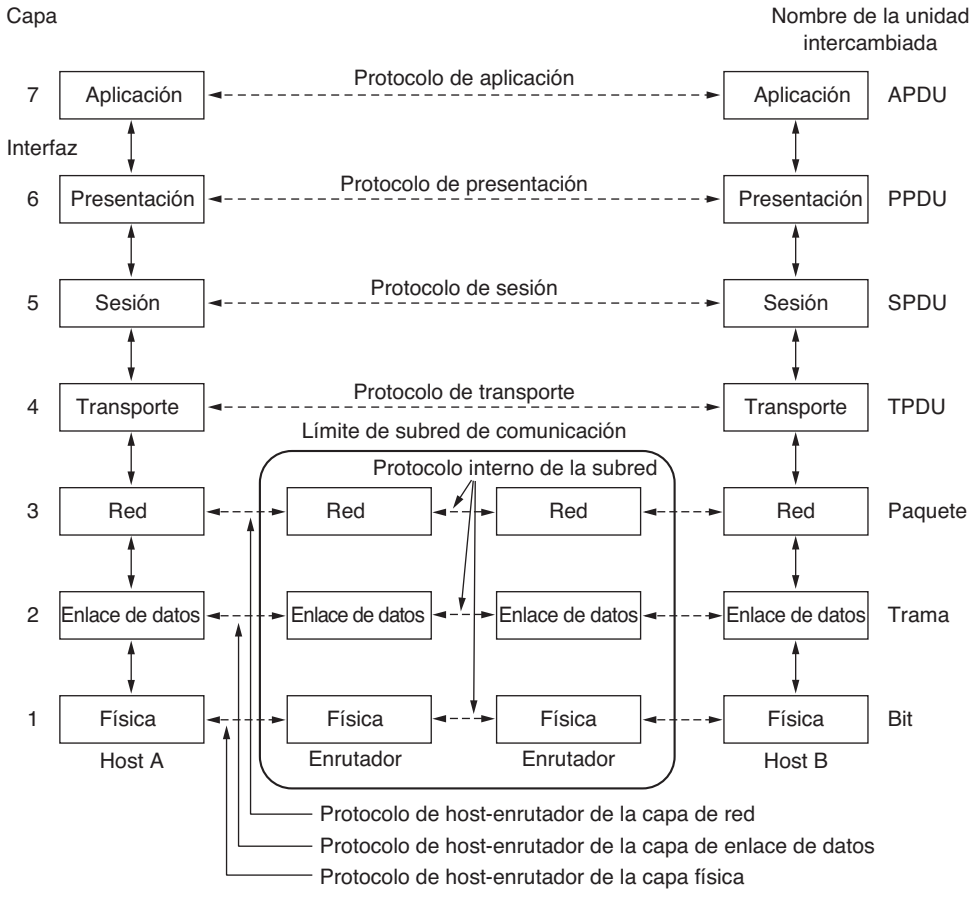

Figura 1-20. El modelo de referencia OSI.

4. Es necesario elegir los límites de las capas de modo que se minimice el flujo de información a través de las interfaces.
5. La cantidad de capas debe ser suficiente como para no tener que agrupar funciones distintas en la misma capa; además, debe ser lo bastante pequeña como para que la arquitectura no se vuelva inmanejable.

A continuación estudiaremos cada capa del modelo en orden, empezando por la capa inferior. Tenga en cuenta que el modelo OSI en sí no es una arquitectura de red, ya que no especifica los servicios y protocolos exactos que se van a utilizar en cada capa. Sólo indica lo que una debe hacer. Sin embargo, la ISO también ha elaborado estándares para todas las capas, aunque no son parte del modelo de referencia en sí. Cada uno se publicó como un estándar internacional separado. Aunque el *modelo* (en parte) es muy usado, los protocolos asociados han estado en el olvido desde hace tiempo.

La capa física

La **capa física** se relaciona con la transmisión de bits puros a través de un canal de transmisión. Los aspectos de diseño tienen que ver con la acción de asegurarse que cuando uno de los lados envíe un bit 1 el otro lado lo reciba como un bit 1, no como un bit 0. En este caso las preguntas típicas son: ¿que señales

eléctricas se deben usar para representar un 1 y un 0?, ¿cuántos nanosegundos dura un bit?, ¿la transmisión puede proceder de manera simultánea en ambas direcciones?, ¿cómo se establece la conexión inicial y cómo se interrumpe cuando ambos lados han terminado?, ¿cuántos pines tiene el conector de red y para qué sirve cada uno? Los aspectos de diseño tienen que ver con las interfaces mecánica, eléctrica y de temporización, así como con el medio de transmisión físico que se encuentra bajo la capa física.

La capa de enlace de datos

La principal tarea de la **capa de enlace de datos** es transformar un medio de transmisión puro en una línea que esté libre de errores de transmisión. Enmascara los errores reales, de manera que la capa de red no los vea. Para lograr esta tarea, el emisor divide los datos de entrada en **tramas de datos** (por lo general, de algunos cientos o miles de bytes) y transmite las tramas en forma secuencial. Si el servicio es confiable, para confirmar la recepción correcta de cada trama, el receptor devuelve una **trama de confirmación de recepción**.

Otra cuestión que surge en la capa de enlace de datos (y en la mayoría de las capas superiores) es cómo evitar que un transmisor rápido inunde de datos a un receptor lento. Tal vez sea necesario algún mecanismo de regulación de tráfico para notificar al transmisor cuando el receptor puede aceptar más datos.

Las redes de difusión tienen una consideración adicional en la capa de enlace de datos: cómo controlar el acceso al canal compartido. Una subcapa especial de la capa de enlace de datos, conocida como subcapa de **control de acceso al medio**, es la que se encarga de este problema.

La capa de red

La **capa de red** controla la operación de la subred. Una cuestión clave de diseño es determinar cómo se encaminan los paquetes desde el origen hasta el destino. Las rutas se pueden basar en tablas estáticas que se "codifican" en la red y rara vez cambian, aunque es más común que se actualicen de manera automática para evitar las fallas en los componentes. También se pueden determinar el inicio de cada conversación; por ejemplo, en una sesión de terminal al iniciar sesión en una máquina remota. Por último, pueden ser muy dinámicas y determinarse de nuevo para cada paquete, de manera que se pueda reflejar la carga actual en la red.

Si hay demasiados paquetes en la subred al mismo tiempo, se interpondrán en el camino unos con otros y formarán cuellos de botella. El manejo de la congestión también es responsabilidad de la capa de red, en conjunto con las capas superiores que adaptan la carga que colocan en la red. Otra cuestión más general de la capa de red es la calidad del servicio proporcionado (retardo, tiempo de tránsito, variaciones, etcétera).

Cuando un paquete tiene que viajar de una red a otra para llegar a su destino, pueden surgir muchos problemas. El direccionamiento utilizado por la segunda red puede ser distinto del que utiliza la primera. La segunda red tal vez no acepte el paquete debido a que es demasiado grande. Los protocolos pueden ser diferentes, etc. Es responsabilidad de la capa de red solucionar todos estos problemas para permitir la interconexión de redes heterogéneas.

En las redes de difusión, el problema de encaminamiento es simple, por lo que con frecuencia la capa de red es delgada o incluso inexistente.

La capa de transporte

La función básica de la **capa de transporte** es aceptar datos de la capa superior, dividirlos en unidades más pequeñas si es necesario, pasar estos datos a la capa de red y asegurar que todas las piezas lleguen

correctamente al otro extremo. Además, todo esto se debe realizar con eficiencia y de una manera que aísle las capas superiores de los inevitables cambios en la tecnología de hardware que se dan con el transcurso del tiempo.

La capa de transporte también determina el tipo de servicio que debe proveer a la capa de sesión y, en última instancia, a los usuarios de la red. El tipo más popular de conexión de transporte es un canal punto a punto libre de errores que entrega los mensajes o bytes en el orden en el que se enviaron. Sin embargo existen otros posibles tipos de servicio de transporte, como el de mensajes aislados sin garantía sobre el orden de la entrega y la difusión de mensajes a múltiples destinos. El tipo de servicio se determina al establecer la conexión (cabe mencionar que es imposible lograr un canal libre de errores; lo que se quiere decir en realidad con este término es que la tasa de errores es lo bastante baja como para ignorarla en la práctica).

La capa de transporte es una verdadera capa de extremo a extremo; lleva los datos por toda la ruta desde el origen hasta el destino. En otras palabras, un programa en la máquina de origen lleva a cabo una conversación con un programa similar en la máquina de destino mediante el uso de los encabezados en los mensajes y los mensajes de control. En las capas inferiores cada uno de los protocolos está entre una máquina y sus vecinos inmediatos, no entre las verdaderas máquinas de origen y de destino, que pueden estar separadas por muchos enrutadores. En la figura 1-20 se muestra la diferencia entre las capas de la 1 a la 3, que están encadenadas, y entre las capas de la 4 a la 7, que son de extremo a extremo.

La capa de sesión

La capa de sesión permite a los usuarios en distintas máquinas establecer **sesiones** entre ellos. Las sesiones ofrecen varios servicios, incluyendo el **control del diálogo** (llevar el control de quién va a transmitir), el **manejo de tokens** (evitar que dos partes intenten la misma operación crítica al mismo tiempo) y la **sincronización** (usar puntos de referencia en las transmisiones extensas para reanudar desde el último punto de referencia en caso de una interrupción).

La capa de presentación

A diferencia de las capas inferiores, que se enfocan principalmente en mover los bits de un lado a otro, la **capa de presentación** se enfoca en la sintaxis y la semántica de la información transmitida. Para hacer posible la comunicación entre computadoras con distintas representaciones internas de datos, podemos definir de una manera abstracta las estructuras de datos que se van a intercambiar, junto con una codificación estándar que se use "en el cable". La capa de presentación maneja estas estructuras de datos abstractas y permite definir e intercambiar estructuras de datos de mayor nivel (por ejemplo, registros bancarios).

La capa de aplicación

La **capa de aplicación** contiene una variedad de protocolos que los usuarios necesitan con frecuencia. Un protocolo de aplicación muy utilizado es **HTTP** (**Protocolo de Transferencia de Hipertexto**, del inglés *HyperText Transfer Protocol*), el cual forma la base para la World Wide Web. Cuando un navegador desea una página web, envía el nombre de la página que quiere al servidor que la hospeda mediante el uso de HTTP. Después el servidor envía la página de vuelta. Hay otros protocolos de aplicación que se utilizan para transferir archivos, enviar y recibir correo electrónico y noticias.

1.4.2 El modelo de referencia TCP/IP

Pasemos ahora del modelo de referencia OSI al modelo de referencia que se utiliza en la más vieja de todas las redes de computadoras de área amplia: ARPANET y su sucesora, Internet. Aunque más adelante veremos una breve historia de ARPANET, es conveniente mencionar ahora unos cuantos aspectos de esta red. ARPANET era una red de investigación patrocinada por el **DoD** (**Departamento de Defensa de Estados Unidos**, del inglés *U.S. Department of the Defense*). En un momento dado llegó a conectar cientos de universidades e instalaciones gubernamentales mediante el uso de líneas telefónicas rentadas. Cuando después se le unieron las redes de satélites y de radio, los protocolos existentes tuvieron problemas para interactuar con ellas, de modo que se necesitaba una nueva arquitectura de referencia. Así, casi desde el principio la habilidad de conectar varias redes sin problemas fue uno de los principales objetivos de diseño. Posteriormente esta arquitectura se dio a conocer como el **Modelo de referencia TCP/IP**, debido a sus dos protocolos primarios. Este modelo se definió por primera vez en Cerf y Kahn (1974); después se refinó y definió como estándar en la comunidad de Internet (Braden, 1989). Clark (1988) describe la filosofía de diseño detrás de este modelo.

Debido a la preocupación del DoD de que alguno de sus valiosos hosts, enrutadores y puertas de enlace de interredes pudieran ser volados en pedazos en cualquier momento por un ataque de la antigua Unión Soviética, otro de los objetivos principales fue que la red pudiera sobrevivir a la pérdida de hardware de la subred sin que se interrumpieran las conversaciones existentes. En otras palabras, el DoD quería que las conexiones permanecieran intactas mientras las máquinas de origen y de destino estuvieran funcionando, incluso aunque algunas de las máquinas o líneas de transmisión en el trayecto dejaran de funcionar en forma repentina. Además, como se tenían en mente aplicaciones con requerimientos divergentes que abarcaban desde la transferencia de archivos hasta la transmisión de voz en tiempo real, se necesitaba una arquitectura flexible.

La capa de enlace

Todos estos requerimientos condujeron a la elección de una red de conmutación de paquetes basada en una capa sin conexión que opera a través de distintas redes. La capa más baja en este modelo es la **capa de enlace**; ésta describe qué enlaces (como las líneas seriales y Ethernet clásica) se deben llevar a cabo para cumplir con las necesidades de esta capa de interred sin conexión. En realidad no es una capa en el sentido común del término, sino una interfaz entre los hosts y los enlaces de transmisión. El primer material sobre el modelo TCP/IP tiene poco que decir sobre ello.

La capa de interred

Esta capa es el eje que mantiene unida a toda la arquitectura. Aparece en la figura 1-21 con una correspondencia aproximada a la capa de red de OSI. Su trabajo es permitir que los hosts inyecten paquetes en cualquier red y que viajen de manera independiente hacia el destino (que puede estar en una red distinta). Incluso pueden llegar en un orden totalmente diferente al orden en que se enviaron, en cuyo caso es responsabilidad de las capas más altas volver a ordenarlos, si se desea una entrega en orden. Tenga en cuenta que aquí utilizamos "interred" en un sentido genérico, aunque esta capa esté presente en la Internet.

La analogía aquí es con el sistema de correos convencional (lento). Una persona puede dejar una secuencia de cartas internacionales en un buzón en un país y, con un poco de suerte, la mayoría de ellas se entregarán a la dirección correcta en el país de destino. Es probable que las cartas pasen a través de una o más puertas de enlace de correo internacionales en su trayecto, pero esto es transparente a los usuarios. Además, los usuarios no necesitan saber que cada país (es decir, cada red) tiene sus propias estampillas, tamaños de sobre preferidos y reglas de entrega.

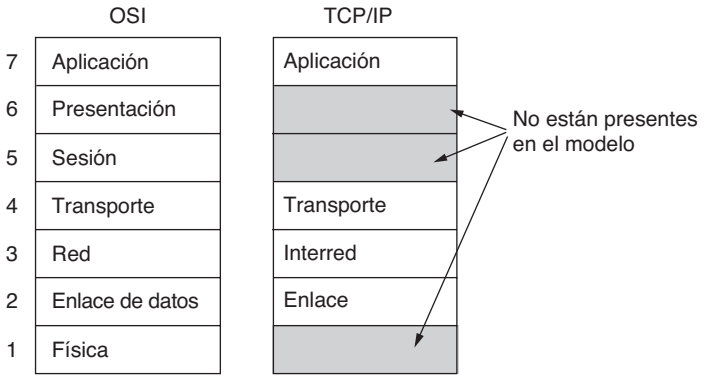

Figura 1-21. El modelo de referencia TCP/IP.

La capa de interred define un formato de paquete y un protocolo oficial llamado **IP** (**Protocolo de Internet**, del inglés *Internet Protocol*), además de un protocolo complementario llamado **ICMP** (**Protocolo de Mensajes de Control de Internet**, del inglés *Internet Control Message Protocol*) que le ayuda a funcionar. La tarea de la capa de interred es entregar los paquetes IP a donde se supone que deben ir. Aquí el ruteo de los paquetes es sin duda el principal aspecto, al igual que la congestión (aunque el IP no ha demostrado ser efectivo para evitar la congestión).

La capa de transporte

Por lo general, a la capa que está arriba de la capa de interred en el modelo TCP/IP se le conoce como **capa de transporte**; y está diseñada para permitir que las entidades pares, en los nodos de origen y de destino, lleven a cabo una conversación, al igual que en la capa de transporte de OSI. Aquí se definieron dos protocolos de transporte de extremo a extremo. El primero, **TCP** (**Protocolo de Control de la Transmisión**, del inglés *Transmission Control Protocol*), es un protocolo confiable orientado a la conexión que permite que un flujo de bytes originado en una máquina se entregue sin errores a cualquier otra máquina en la interred. Este protocolo segmenta el flujo de bytes entrante en mensajes discretos y pasa cada uno a la capa de interred. En el destino, el proceso TCP receptor vuelve a ensamblar los mensajes recibidos para formar el flujo de salida. El TCP también maneja el control de flujo para asegurar que un emisor rápido no pueda inundar a un receptor lento con más mensajes de los que pueda manejar.

El segundo protocolo en esta capa, **UDP** (**Protocolo de Datagrama de Usuario**, del inglés *User Datagram Protocol*), es un protocolo sin conexión, no confiable para aplicaciones que no desean la asignación de secuencia o el control de flujo de TCP y prefieren proveerlos por su cuenta. También se utiliza mucho en las consultas de petición-respuesta de una sola ocasión del tipo cliente-servidor, y en las aplicaciones en las que es más importante una entrega oportuna que una entrega precisa, como en la transmisión de voz o video. En la figura 1-22 se muestra la relación entre IP, TCP y UDP. Desde que se desarrolló el modelo, el IP se ha implementado en muchas otras redes.

La capa de aplicación

El modelo TCP/IP no tiene capas de sesión o de presentación, ya que no se consideraron necesarias. Las aplicaciones simplemente incluyen cualquier función de sesión y de presentación que requieran. La experiencia con el modelo OSI ha demostrado que esta visión fue correcta: estas capas se utilizan muy poco en la mayoría de las aplicaciones.

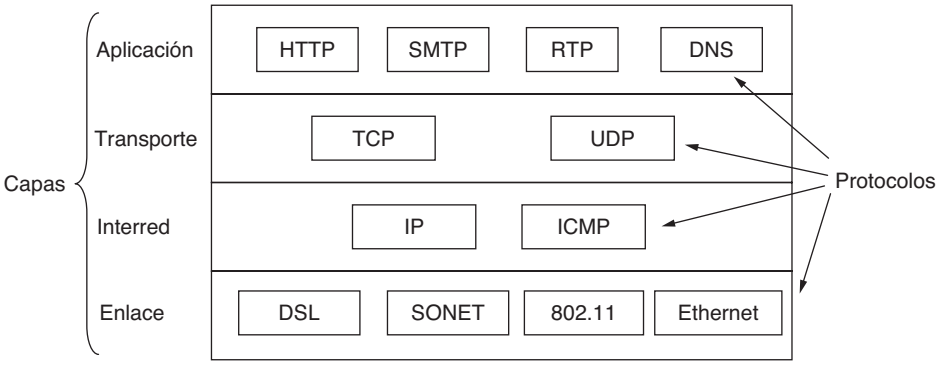

Figura 1-22. El modelo TCP/IP con algunos de los protocolos.

Encima de la capa de transporte se encuentra la **capa de aplicación**. Ésta contiene todos los protocolos de alto nivel. Entre los primeros protocolos están el de terminal virtual (TELNET), transferencia de archivos (FTP) y correo electrónico (SMTP). A través de los años se han agregado muchos otros protocolos. En la figura 1-22 se muestran algunos de los más importantes que veremos más adelante: el Sistema de nombres de dominio (DNS) para resolución de nombres de hosts a sus direcciones de red; HTTP, el protocolo para recuperar páginas de la World Wide Web; y RTP, el protocolo para transmitir medios en tiempo real, como voz o películas.

1.4.3 El modelo utilizado en este libro

Como dijimos antes, la fortaleza del modelo de referencia OSI es el *modelo* en sí (excepto las capas de presentación y de sesión), el cual ha demostrado ser excepcionalmente útil para hablar sobre redes de computadoras. En contraste, la fortaleza del modelo de referencia TCP/IP son los *protocolos*, que se han utilizado mucho durante varios años. Como a los científicos de computadoras les gusta hacer sus propias herramientas, utilizaremos el modelo híbrido de la figura 1-23 como marco de trabajo para este libro.

Figura 1-23. El modelo de referencia que usaremos en este libro.

Este modelo tiene cinco capas, empezando por la capa física, pasando por las capas de enlace, red y transporte hasta llegar a la capa de aplicación. La capa física especifica cómo transmitir bits a través de distintos tipos de medios como señales eléctricas (u otras señales analógicas). La capa de enlace trata sobre cómo enviar mensajes de longitud finita entre computadoras conectadas de manera directa con niveles específicos de confiabilidad. Ethernet y 802.11 son ejemplos de protocolos de capa de enlace.

La capa de red se encarga de combinar varios enlaces múltiples en redes, y redes de redes en interredes, de manera que podamos enviar paquetes entre computadoras distantes. Aquí se incluye la tarea de buscar la ruta por la cual enviarán los paquetes. IP es el principal protocolo de ejemplo que estudiaremos para esta capa. La capa de transporte fortalece las garantías de entrega de la capa de Red, por lo general con una mayor confiabilidad, además provee abstracciones en la entrega, como un flujo de bytes confiable, que coincida con las necesidades de las distintas aplicaciones. TCP es un importante ejemplo de un protocolo de capa de transporte.

Por último, la capa de aplicación contiene programas que hacen uso de la red. Muchas aplicaciones en red tienen interfaces de usuario, como un navegador web. Sin embargo, nuestro interés está en la parte del programa que utiliza la red. En el caso del navegador web se trata del protocolo HTTP. También hay programas de soporte importantes en la capa de aplicación, como el DNS, que muchas aplicaciones utilizan.

La secuencia de nuestros capítulos se basa en este modelo. De esta forma, retenemos el valor del modelo OSI para comprender las arquitecturas de red al tiempo que nos concentramos principalmente en los protocolos que son importantes en la práctica, desde TCP/IP y los protocolos relacionados hasta los más recientes como 802.11, SONET y Bluetooth.

1.4.4 Comparación de los modelos de referencia OSI y TCP/IP

Los modelos de referencia OSI y TCP/IP tienen mucho en común. Ambos se basan en el concepto de una pila de protocolos independientes. Además, la funcionalidad de las capas es muy similar. Por ejemplo, en ambos modelos las capas por encima de la de transporte, incluyendo ésta, se encuentran ahí para proporcionar un servicio de transporte independiente de la red, de extremo a extremo, para los procesos que desean comunicarse. Estas capas forman el proveedor de transporte. También en ambos modelos, las capas que están arriba de la de transporte son usuarias orientadas a la aplicación del servicio de transporte.

A pesar de estas similitudes fundamentales, los dos modelos también tienen muchas diferencias. En esta sección nos enfocaremos en las diferencias clave entre los dos modelos de referencia. Es importante tener en cuenta que aquí compararemos los *modelos de referencia* y no las *pilas de protocolos* correspondientes. Más adelante estudiaremos los protocolos en sí. Un libro completo dedicado a comparar y contrastar TCP/IP y OSI es el de Piscitello y Chapin (1993).

Hay tres conceptos básicos para el modelo OSI:

1. Servicios.
2. Interfaces.
3. Protocolos.

Quizá, la mayor contribución del modelo OSI es que hace explícita la distinción entre estos tres conceptos. Cada capa desempeña ciertos *servicios* para la capa que está sobre ella. La definición del servicio indica lo que hace la capa, no cómo acceden a ella las entidades superiores ni cómo funciona. Define la semántica de la capa.

La *interfaz* de una capa indica a los procesos superiores cómo pueden acceder a ella. Especifica cuáles son los parámetros y qué resultados se pueden esperar. Pero no dice nada sobre su funcionamiento interno.

Por último, la capa es la que debe decidir qué *protocolos* de iguales utilizar. Puede usar los protocolos que quiera, siempre y cuando realice el trabajo (es decir, que provea los servicios ofrecidos). También los puede cambiar a voluntad sin afectar el software de las capas superiores.

Estas ideas encajan muy bien con las ideas modernas sobre la programación orientada a objetos. Al igual que una capa, un objeto tiene un conjunto de métodos (operaciones) que los procesos fuera

del objeto pueden invocar. La semántica de estos métodos define el conjunto de servicios que ofrece el objeto. Los parámetros y resultados de los métodos forman la interfaz del objeto. El código interno del objeto es su protocolo y no se puede ver ni es de la incumbencia de las entidades externas al objeto.

Al principio, el modelo TCP/IP no tenía una distinción clara entre los servicios, las interfaces y los protocolos, aunque las personas han tratado de reajustarlo a fin de hacerlo más parecido al OSI. Por ejemplo, los únicos servicios que realmente ofrece la capa de interred son SEND IP PACKET y RECEIVE IP PACKET. Como consecuencia, los protocolos en el modelo OSI están ocultos de una mejor forma que en el modelo TCP/IP, además se pueden reemplazar con relativa facilidad a medida que la tecnología cambia. La capacidad de realizar dichos cambios con transparencia es uno de los principales propósitos de tener protocolos en capas en primer lugar.

El modelo de referencia OSI se ideó *antes* de que se inventaran los protocolos correspondientes. Este orden significa que el modelo no estaba orientado hacia un conjunto específico de protocolos, un hecho que lo hizo bastante general. La desventaja de este orden fue que los diseñadores no tenían mucha experiencia con el tema y no supieron bien qué funcionalidad debían colocar en cada una de las capas.

Por ejemplo, en un principio la capa de enlace de datos trabajaba sólo con redes de punto a punto. Cuando surgieron las redes de difusión, fue necesario insertar una nueva subcapa al modelo. Además, cuando las personas empezaron a construir redes reales mediante el modelo OSI y los protocolos existentes, se descubrió que estas redes no coincidían con las especificaciones de los servicios requeridos, de modo que tuvieron que integrar en el modelo subcapas convergentes que permitieran cubrir las diferencias. Finalmente, el comité en un principio esperaba que cada país tuviera una red operada por el gobierno en la que se utilizaran los protocolos OSI, por lo que no se tomó en cuenta la interconexión de redes. Para no hacer el cuento largo, las cosas no salieron como se esperaba.

Con TCP/IP sucedió lo contrario: primero llegaron los protocolos y el modelo era en realidad sólo una descripción de los protocolos existentes. No hubo problema para que los protocolos se ajustaran al modelo. Encajaron a la perfección. El único problema fue que el *modelo* no encajaba en ninguna otra pila de protocolos. En consecuencia, no era útil para describir otras redes que no fueran TCP/IP.

Pasando de las cuestiones filosóficas a las más específicas, una diferencia obvia entre los dos modelos está en el número de capas: el modelo OSI tiene siete capas, mientras que el modelo TCP/IP tiene cuatro. Ambos tienen capas de (inter)red, transporte y aplicación, pero las demás capas son distintas.

Hay otra diferencia en el área de la comunicación sin conexión frente a la comunicación orientada a conexión. El modelo OSI soporta ambos tipos de comunicación en la capa de red, pero sólo la comunicación orientada a conexión en la capa de transporte, en donde es más importante (ya que el servicio de transporte es visible a los usuarios). El modelo TCP/IP sólo soporta un modo en la capa de red (sin conexión) pero soporta ambos en la capa de transporte, de manera que los usuarios tienen una alternativa, que es muy importante para los protocolos simples de petición-respuesta.

1.4.5 Una crítica al modelo y los protocolos OSI

Ni el modelo OSI y sus protocolos, ni el modelo TCP/IP y sus protocolos son perfectos. Ambos pueden recibir bastantes críticas, y así se ha hecho. En ésta y en la siguiente sección analizaremos algunas de ellas. Empezaremos con el modelo OSI y después examinaremos el modelo TCP/IP.

Para cuando se publicó la segunda edición de este libro (1989), a muchos expertos en el campo les pareció que el modelo OSI y sus protocolos iban a adueñarse del mundo y sacar todo lo demás a su paso.

Pero esto no fue así. ¿Por qué? Tal vez sea útil analizar en retrospectiva algunas de las razones. Podemos resumirlas de la siguiente manera:

1. Mala sincronización.
2. Mala tecnología.
3. Malas implementaciones.
4. Mala política.

Mala sincronización

Veamos la razón número uno: mala sincronización. El tiempo en el cual se establece un estándar es absolutamente imprescindible para su éxito. David Clark, del Massachusetts Institute of Technology (MIT), tiene una teoría de estándares a la que llama el *apocalipsis de los dos elefantes*, la cual se ilustra en la figura 1-24.

Esta figura muestra la cantidad de actividad alrededor de un nuevo tema. Cuando se descubre el tema por primera vez, hay una ráfaga de actividades de investigación en forma de discusiones, artículos y reuniones. Después de cierto tiempo esta actividad disminuye, las corporaciones descubren el tema y llega la ola de inversión de miles de millones de dólares.

Es imprescindible que los estándares se escriban en el intermedio entre los dos "elefantes". Si se escriben demasiado pronto (antes de que los resultados de la investigación estén bien establecidos), tal vez el tema no se entienda bien todavía; el resultado es un estándar malo. Si se escriben demasiado tarde, es probable que muchas empresas hayan hecho ya importantes inversiones en distintas maneras de hacer las cosas, de modo que los estándares se ignorarán en la práctica. Si el intervalo entre los dos elefantes es muy corto (ya que todos tienen prisa por empezar), la gente que desarrolla los estándares podría quedar aplastada.

En la actualidad, parece que los protocolos estándar de OSI quedaron aplastados. Para cuando aparecieron los protocolos de OSI, los protocolos TCP/IP competidores ya se utilizaban mucho en universidades que hacían investigaciones. Aunque todavía no llegaba la ola de inversión de miles de millones de dólares, el mercado académico era lo bastante grande como para que muchos distribuidores empezaran a ofrecer con cautela los productos TCP/IP. Para cuando llegó el modelo OSI, los distribuidores no quisieron apoyar una segunda pila de protocolos hasta que se vieron obligados a hacerlo, de modo que no hubo ofertas iniciales. Como cada empresa estaba esperando a que otra tomara la iniciativa, ninguna lo hizo y OSI nunca se llevó a cabo.

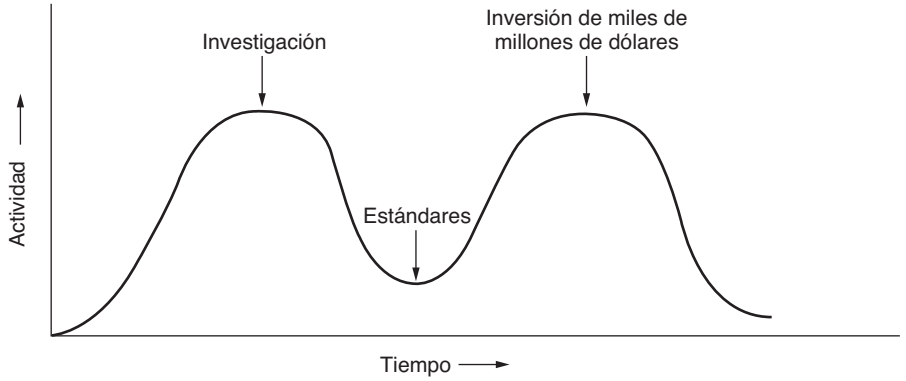

Figura 1-24. El apocalipsis de los dos elefantes.

Mala tecnología

La segunda razón por la que OSI nunca tuvo éxito fue que tanto el modelo como los protocolos tienen fallas. La opción de siete capas era más política que técnica, además de que dos de las capas (sesión y presentación) están casi vacías, mientras que otras dos (enlace de datos y red) están demasiado llenas.

El modelo OSI, junto con sus correspondientes definiciones y protocolos de servicios, es muy complejo. Si se apilan, los estándares impresos ocupan una fracción considerable de un metro de papel. Además son difíciles de implementar e ineficientes en su operación. En este contexto nos viene a la mente un acertijo propuesto por Paul Mockapetris y citado por Rose (1993):

P: ¿Qué obtenemos al cruzar un pandillero con un estándar internacional?
R: Alguien que le hará una oferta que no podrá comprender.

Además de ser incomprensible, otro problema con el modelo OSI es que algunas funciones como el direccionamiento, el control de flujo y el control de errores, vuelven a aparecer una y otra vez en cada capa. Por ejemplo, Saltzer y sus colaboradores (1984) han señalado que para ser efectivo, hay que llevar a cabo el control de errores en la capa más alta, por lo que repetirlo una y otra vez en cada una de las capas más bajas es con frecuencia innecesario e ineficiente.

Malas implementaciones

Dada la enorme complejidad del modelo y los protocolos, no es sorprendente que las implementaciones iniciales fueran enormes, pesadas y lentas. Todos los que las probaron se arrepintieron. No tuvo que pasar mucho tiempo para que las personas asociaran "OSI" con la "mala calidad". Aunque los productos mejoraron con el tiempo, la imagen perduró.

En contraste, una de las primeras implementaciones de TCP/IP fue parte del UNIX, de Berkeley, y era bastante buena (y además, gratuita). Las personas empezaron a utilizarla rápidamente, lo cual provocó que se formara una extensa comunidad de usuarios, lo que condujo a mejoras, lo que llevó a una comunidad todavía mayor. En este caso la espiral fue hacia arriba, en vez de ir hacia abajo.

Malas políticas

Gracias a la implementación inicial, mucha gente (en especial los académicos) pensaba que TCP/IP era parte de UNIX, y UNIX en la década de 1980 para los académicos era algo así como la paternidad (que en ese entonces se consideraba erróneamente como maternidad) y el pay de manzana para los estadounidenses comunes.

Por otro lado, OSI se consideraba en muchas partes como la invención de los ministerios europeos de telecomunicaciones, de la Comunidad Europea y después, del gobierno de Estados Unidos. Esta creencia no era del todo justificada, pero la simple idea de un grupo de burócratas gubernamentales que trataban de obligar a los pobres investigadores y programadores que estaban en las trincheras desarrollando verdaderas redes de computadoras a que adoptaran un estándar técnicamente inferior no fue de mucha utilidad para la causa de OSI. Algunas personas vieron este suceso como algo similar a cuando IBM anunció en la década de 1960 que PL/I era el lenguaje del futuro, o cuando luego el DoD corrigió esto para anunciar que en realidad el lenguaje era Ada.

1.4.6 Una crítica al modelo de referencia TCP/IP

El modelo y los protocolos de TCP/IP también tienen sus problemas. Primero, el modelo no diferencia con claridad los conceptos de servicios, interfaces y protocolos. La buena práctica de la ingeniería

de software requiere una distinción entre la especificación y la implementación, algo que OSI hace con mucho cuidado y que TCP/IP no. En consecuencia, el modelo TCP/IP no sirve mucho de guía para diseñar modernas redes que utilicen nuevas tecnologías.

Segundo, el modelo TCP/IP no es nada general y no es muy apropiado para describir cualquier pila de protocolos aparte de TCP/IP. Por ejemplo, es imposible tratar de usar el modelo TCP/IP para describir Bluetooth.

Tercero, la capa de enlace en realidad no es una capa en el sentido normal del término como se utiliza en el contexto de los protocolos en capas. Es una interfaz (entre las capas de red y de enlace de datos). La diferencia entre una interfaz y una capa es crucial, y hay que tener mucho cuidado al respecto.

Cuarto, el modelo TCP/IP no distingue entre la capa física y la de enlace de datos. Éstas son completamente distintas. La capa física trata sobre las características de transmisión del cable de cobre, la fibra óptica y la comunicación inalámbrica. La tarea de la capa de enlace de datos es delimitar el inicio y el fin de las tramas, además de transmitirlas de un extremo al otro con el grado deseado de confiabilidad. Un modelo apropiado debe incluir ambas capas por separado. El modelo TCP/IP no hace esto.

Por último, aunque los protocolos IP y TCP se diseñaron e implementaron con sumo cuidado, muchos de los otros protocolos se fueron creando según las necesidades del momento, producidos generalmente por un par de estudiantes de licenciatura que los mejoraban hasta fastidiarse. Después las implementaciones de los protocolos se distribuían en forma gratuita, lo cual trajo como consecuencia que se utilizaran amplia y profundamente en muchas partes y, por ende, eran difíciles de reemplazar. Algunos de ellos son un poco vergonzosos en la actualidad. Por ejemplo, el protocolo de terminal virtual TELNET se diseñó para una terminal de Teletipo mecánica de 10 caracteres por segundo. No sabe nada sobre las interfaces gráficas de usuario y los ratones. Sin embargo, aún se sigue usando a 30 años de su creación.

1.5 REDES DE EJEMPLO

El tema de las redes de computadoras cubre muchos tipos distintos de redes, grandes y pequeñas, populares y no tanto. Tienen distintos objetivos, escalas y tecnologías. En las siguientes secciones analizaremos algunos ejemplos para tener una idea de la variedad que podemos encontrar en el área de las redes de computadoras.

Empezaremos con Internet, que tal vez sea la red más popular; analizaremos su historia, evolución y tecnología. Después consideraremos la red de teléfonos móviles. Técnicamente es muy distinta de Internet y contrasta muy bien con ella. Más adelante introduciremos el IEEE 802.11, el estándar dominante para las redes LAN inalámbricas. Por último, analizaremos las redes RFID y de sensores, tecnologías que extienden el alcance de la red para incluir al mundo físico y los objetos cotidianos.

1.5.1 Internet

En realidad Internet no es una red, sino una enorme colección de distintas redes que utilizan ciertos protocolos comunes y proveen ciertos servicios comunes. Es un sistema inusual en cuanto a que nadie la planeó y nadie la controla. Para comprender mejor esto, empecemos desde el inicio para ver cómo se ha desarrollado y por qué. Si desea leer una maravillosa historia de Internet, le recomendamos ampliamente el libro de Jim Naughton (2000). Es uno de esos libros inusuales que no sólo son divertidos, sino que también cuenta con 20 páginas de *ibídems* y *obras citadas* (*ob. cit.*) para el verdadero historiador. Una parte del material de esta sección se basa en ese libro.

Claro que también se han escrito innumerables libros sobre Internet y sus protocolos. Para obtener más información puede consultar a Maufer (1999).

ARPANET

La historia empieza a finales de la década de 1950. En la cúspide de la Guerra Fría, el DoD de Estados Unidos quería una red de comando y control que pudiera sobrevivir a una guerra nuclear. En ese tiempo todas las comunicaciones militares utilizaban la red telefónica pública, que se consideraba vulnerable. Podemos ver la razón de esta creencia en la figura 1-25(a). Los puntos negros representan las oficinas de conmutación telefónica, cada una de las cuales se conectaba a miles de teléfonos. Estas oficinas de conmutación se conectaban a su vez con oficinas de conmutación de mayor nivel (oficinas interurbanas), para formar una jerarquía nacional con sólo una pequeña cantidad de redundancia. La vulnerabilidad del sistema era que, si se destruían unas cuantas oficinas interurbanas clave, se podía fragmentar el sistema en muchas islas aisladas.

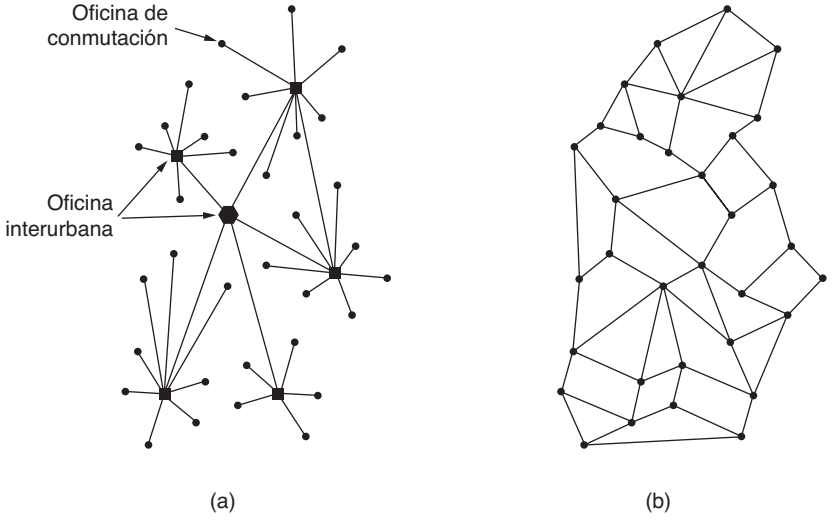

Figura 1-25. (a) Estructura de un sistema telefónico. (b) El sistema de conmutación distribuida propuesto por Baran.

Alrededor de la década de 1960, el DoD otorgó un contrato a la empresa RAND Corporation para buscar una solución. Uno de sus empleados, Paul Baran, ideó el diseño tolerante a fallas altamente distribuido de la figura 1-25(b). Como las rutas entre dos oficinas de conmutación cualesquiera eran ahora mucho más largas de lo que las señales análogas podían viajar sin distorsión, Baran propuso el uso de la tecnología de conmutación de paquetes digital, y escribió varios informes para el DoD en donde describió sus ideas con detalle (Baran, 1964). A los oficiales del Pentágono les gustó el concepto y pidieron a AT&T, que en ese entonces era el monopolio telefónico nacional en Estados Unidos, que construyera un prototipo. Pero AT&T hizo caso omiso de las ideas de Baran. La corporación más grande y opulenta del mundo no iba a permitir que un joven impertinente les dijera cómo construir un sistema telefónico. Dijeron que la idea de Baran no se podía construir y se desechó.

Pasaron otros siete años y el DoD seguía sin poder obtener un mejor sistema de comando y control. Para comprender lo que ocurrió después tenemos que remontarnos hasta octubre de 1957, cuando la antigua Unión Soviética venció a Estados Unidos en la carrera espacial con el lanzamiento del primer satélite artificial, Sputnik. Cuando el presidente Eisenhower trató de averiguar quién se había quedado dormido en los controles, quedó consternado al descubrir que el Ejército, la Marina y la Fuerza Aérea estaban riñendo por el presupuesto de investigación del Pentágono. Su respuesta inmediata fue crear una sola

organización de investigación de defensa, **ARPA** (**Agencia de Proyectos de Investigación Avanzados**, del inglés *Advanced Research Projects Agency*). La ARPA no tenía científicos ni laboratorios; de hecho, sólo tenía una oficina y un pequeño presupuesto (según los estándares del Pentágono). Para realizar su trabajo otorgaba concesiones y contratos a las universidades y las compañías cuyas ideas fueran prometedoras.

Durante los primeros años, la ARPA trató de averiguar cuál debería ser su misión. En 1967 Larry Roberts, director de la ARPA, quien trataba de averiguar cómo proveer acceso remoto a las computadoras, giró su atención a las redes. Contactó a varios expertos para decidir qué hacer. Uno de ellos de nombre Wesley Clark, sugirió construir una subred de conmutación de paquetes y conectar cada host a su propio enrutador.

Después de cierto escepticismo inicial, Roberts aceptó la idea y presentó un documento algo impreciso sobre ella en el Simposio SIGOPS de la ACM sobre Principios de Sistemas Operativos que se llevó a cabo en Gatlinburg, Tennessee, a finales de 1967 (Roberts, 1967). Para gran sorpresa de Roberts había otro documento en la conferencia que describía un sistema similar que no sólo se había diseñado, sino que también se había implementado por completo bajo la dirección de Donald Davies en el Laboratorio Nacional de Física (NPL), en Inglaterra. El sistema del NPL no era un sistema nacional (sólo conectaba varias computadoras en su campus), pero demostraba que la conmutación de paquetes podía funcionar. Además citaba el trabajo anterior de Baran que había sido descartado. Roberts regresó de Gatlinburg determinado a construir lo que después se convirtió en **ARPANET**.

La subred consistiría de minicomputadoras llamadas **IMP** (**Procesadores de Mensajes de Interfaz**, del inglés *Interface Message Processors*), conectadas por líneas de transmisión de 56 kbps. Para una confiabilidad alta, cada IMP se conectaría por lo menos a otras dos. La subred sería de datagramas, de manera que si se destruían algunas líneas e IMP, los mensajes se podrían encaminar nuevamente de manera automática a través de rutas alternativas.

Cada nodo de la red debía estar constituido por una IMP y un host, en el mismo cuarto, conectados por un cable corto. Un host podía enviar mensajes de hasta 8 063 bits a su IMP, que a su vez los descompondría en paquetes de 1 008 bits a lo más y los enviaría de manera independiente a su destino. Cada paquete se recibía en su totalidad antes de enviarlo, por lo que la subred fue la primera red electrónica de conmutación de paquetes de almacenamiento y envío.

Entonces ARPA lanzó una convocatoria para construir la subred y fueron 12 compañías las que licitaron. Después de evaluar todas las propuestas, la ARPA seleccionó a BBN, una empresa de consultoría con base en Cambridge, Massachusetts, y en diciembre de 1968 le otorgó un contrato para construir la subred y escribir el software. BBN optó por usar como IMP las minicomputadoras Honeywell DDP-316 modificadas de manera especial con palabras de 16 bits y 12 KB de memoria básica. Los IMP no tenían discos, ya que las partes móviles se consideraban no confiables. Los IMP se interconectaron mediante líneas de 56 kbps que se rentaban a las compañías telefónicas. Aunque ahora 56 kbps son la opción para los adolescentes que no pueden pagar DSL o cable, en ese entonces era lo mejor que el dinero podía comprar.

El software se dividió en dos partes: subred y host. El software de subred consistía del extremo IMP de la conexión host a IMP, del protocolo IMP a IMP y de un protocolo de IMP de origen a IMP de destino diseñado para mejorar la confiabilidad. En la figura 1-26 se muestra el diseño original de la ARPANET.

Fuera de la subred también se necesitaba software, es decir, el extremo host de la conexión host a IMP, el protocolo host a host y el software de aplicación. Pronto quedó claro que BBN consideraba que al aceptar un mensaje en un cable host a IMP y colocarlo en el cable host a IMP de destino, su trabajo estaba terminado.

Pero Roberts tenía un problema: los hosts también necesitaban software. Para lidiar con ello, convocó una junta de investigadores de redes, que en su mayor parte eran estudiantes de licenciatura, en Snowbird, Utah, en el verano de 1969. Los estudiantes esperaban que un experto en redes les explicara el gran diseño de la red y su software, y que después les asignara la tarea de escribir parte de ella. Quedaron pasmados

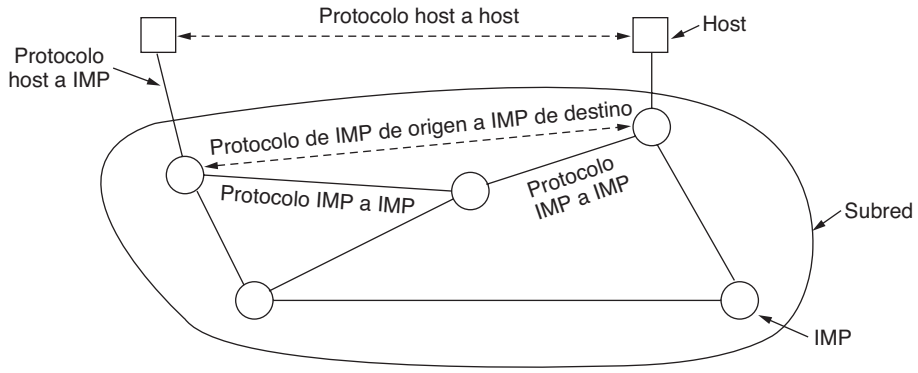

Figura 1-26. Diseño original de ARPANET.

al descubrir que no había ningún experto en redes ni un gran diseño. Tuvieron que averiguar qué hacer por su cuenta.

Sin embargo, de alguna forma una red experimental se puso en línea en diciembre de 1969 con cuatro nodos: en UCLA, UCSB, SRI y la Universidad de Utah. Se eligieron estos cuatro nodos debido a que todos tenían una gran cantidad de contratos de ARPA y todos tenían computadoras host distintas y totalmente incompatibles (sólo para hacerlo más divertido). Dos meses antes se había enviado el primer mensaje de host a host desde el nodo de UCLA por un equipo dirigido por Len Kleinrock (pionero de la teoría de conmutación de paquetes), hasta el nodo de SRI. La red creció con rapidez a medida que se entregaban e instalaban más equipos IMP; pronto abarcó Estados Unidos. En la figura 1-27 se muestra qué tan rápido creció ARPANET durante los primeros tres años.

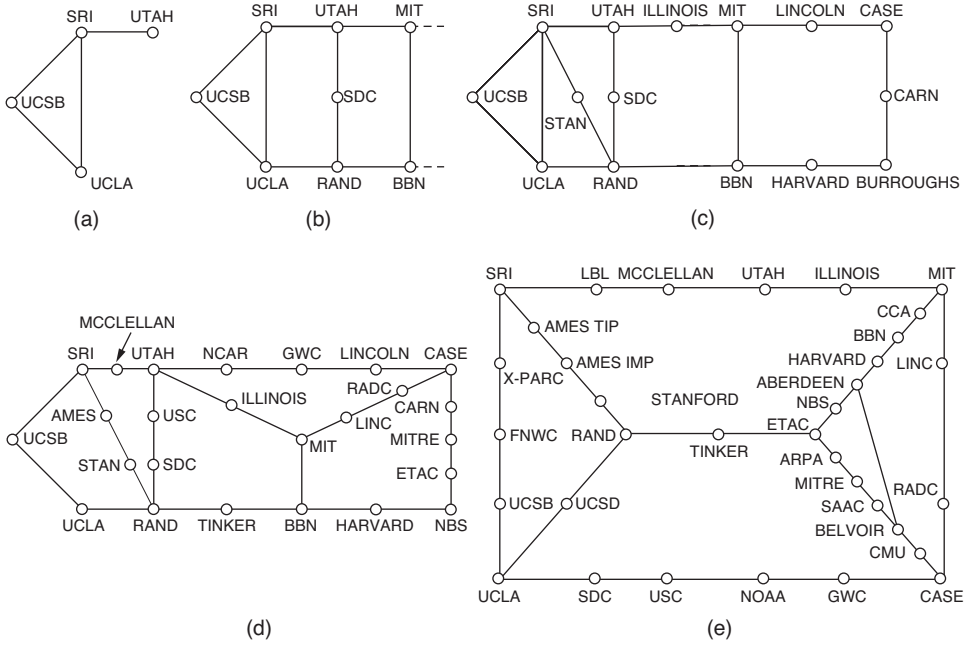

Figura 1-27. Crecimiento de ARPANET. (a) Diciembre de 1969. (b) Julio de 1970. (c) Marzo de 1971. (d) Abril de 1972. (e) Septiembre de 1972.

Además de ayudar al crecimiento de la recién creada ARPANET, la ARPA también patrocinó la investigación sobre el uso de las redes satelitales y las redes de radio de paquetes móviles. En una famosa demostración, un camión que recorría California usó la red de radio de paquetes para enviar mensajes a SRI, que a su vez los envió a través de ARPANET a la Costa Este, en donde se enviaron al Colegio Universitario, en Londres, a través de la red satelital. Gracias a esto, un investigador en el camión pudo utilizar una computadora en Londres mientras conducía por California.

Este experimento también demostró que los protocolos existentes de ARPANET no eran adecuados para trabajar en distintas redes. Esta observación condujo a más investigaciones sobre protocolos, lo que culminó con la invención del modelo y los protocolos TCP/IP (Cerf y Kahn, 1974). El modelo TCP/IP se diseñó de manera específica para manejar la comunicación a través de interredes, algo que se volvía día con día más importante a medida que más redes se conectaban a ARPANET.

Para fomentar la adopción de estos nuevos protocolos, la ARPA otorgó varios contratos para implementar TCP/IP en distintas plataformas de computadora, incluyendo sistemas de IBM, DEC y HP, así como para el UNIX, de Berkeley. Los investigadores de la Universidad de California, en Berkeley, rediseñaron el modelo TCP/IP con una nueva interfaz de programación llamada **sockets** para la futura versión 4.2BSD del UNIX, de Berkeley. También escribieron muchos programas de aplicación, utilería y administración para mostrar lo conveniente que era usar la red con sockets.

La sincronización era perfecta. Muchas universidades acababan de adquirir una segunda o tercera computadoras VAX y una LAN para conectarlas, pero no tenían software de red. Cuando llegó el 4.2BSD junto con TCP/IP, los sockets y muchas utilerías de red, el paquete completo se adoptó de inmediato. Además, con TCP/IP era fácil conectar las redes LAN a ARPANET, y muchas lo hicieron.

Durante la década de 1980 se conectaron redes adicionales (en especial redes LAN) a ARPANET. A medida que aumentó la escala, el proceso de buscar hosts se hizo cada vez más costoso, por lo que se creó el **DNS** (**Sistema de Nombres de Dominio**, del inglés *Domain Name System*) para organizar a las máquinas en dominios y resolver nombres de host en direcciones IP. Desde entonces, el DNS se convirtió en un sistema de base de datos distribuido y generalizado para almacenar una variedad de información relacionada con la asignación de nombres. En el capítulo 7 estudiaremos este sistema con detalle.

NSFNET

A finales de la década de 1970, la **NSF** (**Fundación Nacional de la Ciencia**, del inglés U.S. National Science Foundation) vio el enorme impacto que había tenido ARPANET en la investigación universitaria al permitir que científicos de todo el país compartieran datos y colaboraran en proyectos de investigación. Pero para entrar a ARPANET una universidad tenía que tener un contrato de investigación con el DoD. Como muchas no tenían un contrato, la respuesta inicial de la NSF fue patrocinar la Red de Ciencias Computacionales (**CSNET**, del inglés *Computer Science Network*) en 1981. Esta red conectó los departamentos de ciencias computacionales y los laboratorios de investigación industrial a ARPANET por medio de líneas de marcación y rentadas. A finales de la década de 1980, la NSF fue más allá y decidió diseñar un sucesor para ARPANET que estuviera abierto a todos los grupos universitarios de investigación.

Para tener algo concreto con qué empezar, la NSF decidió construir una red troncal (*backbone*) para conectar sus seis centros de supercomputadoras en San Diego, Boulder, Champaign, Pittsburgh, Ithaca y Princeton. Cada supercomputadora recibió un hermano pequeño que consistía en una microcomputadora LSI-11 llamada **fuzzball**. Las fuzzballs se conectaron a líneas rentadas de 56 kbps para formar la subred, la misma tecnología de hardware que utilizaba ARPANET. Sin embargo, la tecnología de software era diferente: las fuzzballs funcionaban con TCP/IP desde un principio, así que se convirtió en la primera WAN de TCP/IP.

La NSF también patrocinó algunas redes regionales (finalmente fueron cerca de 20) que se conectaban a la red troncal para permitir que los usuarios de miles de universidades, laboratorios de investigación,

bibliotecas y museos tuvieran acceso a cualquiera de las supercomputadoras y se comunicaran entre sí. La red completa, incluyendo la red troncal y las redes regionales, se llamó **NSFNET**. Se conectaba a ARPANET por medio de un enlace entre un IMP y una fuzzball en el cuarto de máquinas de Carnegie-Mellon. En la figura 1-28 se ilustra la primera red troncal de NSFNET, superpuesta en un mapa de Estados Unidos.

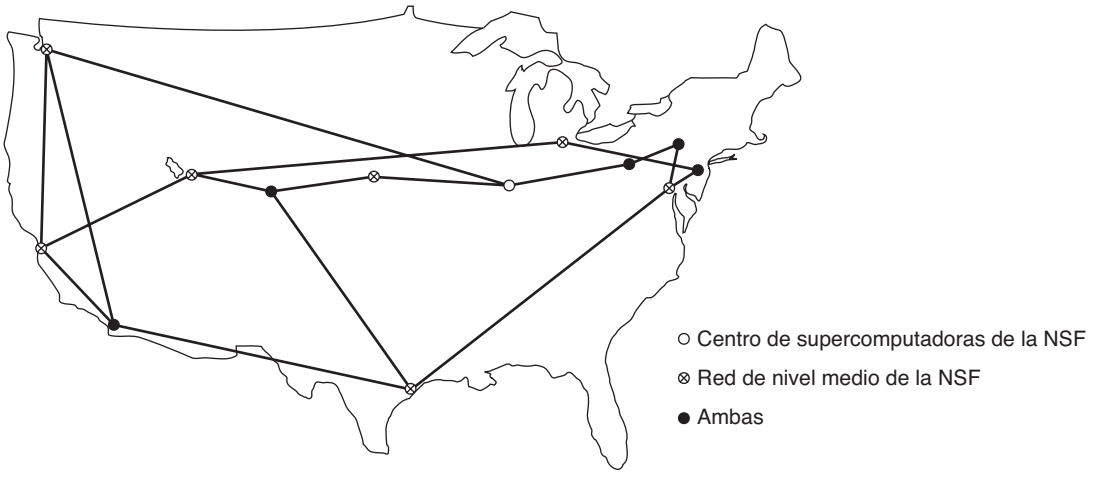

Figura 1-28. La red troncal de NSFNET en 1988.

La NSFNET fue un éxito instantáneo y se sobrecargó desde el principio. La NSF empezó de inmediato a planear su sucesora y otorgó un contrato al consorcio MERIT con base en Michigan para llevar a cabo la tarea. Se rentaron a MCI (que desde entonces se fusionó con WorldCom) unos canales de fibra óptica a 448 kbps para proveer la versión 2 de la red troncal. Se utilizaron equipos PC-RT de IBM como enrutadores. Esta red también se sobrecargó casi de inmediato y, para 1990, la segunda red troncal se actualizó a 1.5 Mbps.

Mientras la red seguía creciendo, la NSF se dio cuenta de que el gobierno no podría seguir financiando el uso de las redes por siempre. Además, las organizaciones comerciales querían unirse pero los estatutos de la NSF les prohibían usar las redes pagadas por la Fundación. En consecuencia, la NSF animó a MERIT, MCI e IBM para que formaran una corporación sin fines de lucro llamada **ANS** (**Redes y Servicios Avanzados**, del inglés *Advanced Networks and Services*), como primer paso en el camino hacia la comercialización. En 1990, ANS se hizo cargo de la NSFNET y actualizó los enlaces de 1.5 Mbps a 45 Mbps para formar la **ANSNET**. Esta red operó durante cinco años y después se vendió a America Online. Pero para entonces, varias empresas estaban ofreciendo el servicio IP comercial y era evidente que el gobierno debía ahora salirse del negocio de las redes.

Para facilitar la transición y asegurarse de que cada red regional se pudiera comunicar con las demás redes regionales, la NSF otorgó contratos a cuatro distintos operadores de red para establecer un **NAP** (**Punto de Acceso a la Red**, del inglés *Network Access Point*). Estos operadores fueron PacBell (San Francisco), Ameritech (Chicago), MFS (Washington, D.C.) y Sprint (Nueva York, en donde para fines de NAP, Pennsauken, Nueva Jersey cuenta como la ciudad de Nueva York). Todos los operadores de redes que quisieran ofrecer el servicio de red troncal a las redes regionales de la NSF se tenían que conectar a todos los NAP.

Este arreglo significaba que un paquete que se originara en cualquier red regional podía elegir entre varias portadoras de red troncal para ir desde su NAP hasta el NAP de destino. En consecuencia, las

portadoras de red troncal se vieron forzadas a competir por el negocio de las redes regionales con base en el servicio y al precio, que desde luego era lo que se pretendía. Como resultado, el concepto de una sola red troncal predeterminada se reemplazó por una infraestructura competitiva impulsada por el comercio. A muchas personas les gusta criticar al gobierno federal por no ser innovador, pero en el área de las redes fueron el DoD y la NSF quienes crearon la infraestructura que formó la base para Internet y después la entregaron a la industria para que la pusiera en funcionamiento.

Durante la década de 1990, muchos otros países y regiones también construyeron redes de investigación nacional, que con frecuencia seguían el patrón de ARPANET y de la NSFNET. Entre éstas tenemos a EuropaNET y EBONE en Europa, que empezaron con líneas de 2 Mbps y después actualizaron a líneas de 34 Mbps. En un momento dado, la infraestructura de red en Europa también se puso en manos de la industria.

Internet ha cambiado mucho desde sus primeros días. Su tamaño se expandió de manera considerable con el surgimiento de la World Wide Web (WWW) a principios de la década de 1990. Datos recientes de Internet Systems Consortium indican que el número de hosts visibles en Internet está cerca de los 600 millones. Ésta es una estimación baja, pero excede por mucho los varios millones de hosts que había cuando se sostuvo la primera conferencia sobre la WWW en el CERN en 1994.

También ha cambiado mucho la forma en que usamos Internet. Al principio dominaban las aplicaciones como el correo electrónico para los académicos, los grupos de noticias, inicios remotos de sesión y transferencias de archivos. Después cambió a correo para todos, luego la web y la distribución de contenido de igual a igual, como el servicio Napster que está cerrado en la actualidad. Ahora están empezando a tomar popularidad la distribución de medios en tiempo real, las redes sociales (como Facebook) y los microblogs (como Twitter). Estos cambios trajeron a Internet tipos de medios más complejos, y por ende, mucho más tráfico. De hecho, el tráfico dominante en Internet parece cambiar con cierta regularidad puesto que, por ejemplo, las nuevas y mejores formas de trabajar con la música o las películas se pueden volver muy populares con gran rapidez.

Arquitectura de Internet

La arquitectura de Internet también cambió mucho debido a que creció en forma explosiva. En esta sección trataremos de analizar de manera breve las generalidades sobre cómo se ve Internet en la actualidad. La imagen se complica debido a las continuas fusiones en los negocios de las compañías telefónicas (telcos), las compañías de cable y los ISP, y por lo que es difícil distinguir quién hace cada cosa. Uno de los impulsores de esta confusión es la convergencia de las telecomunicaciones, en donde una red se utiliza para distintos servicios que antes realizaban distintas compañías. Por ejemplo, en un "triple play", una compañía le puede vender telefonía, TV y servicio de Internet a través de la misma conexión de red, con el supuesto de que usted ahorrará dinero. En consecuencia, la descripción aquí proporcionada será algo más simple que la realidad. Y lo que es verdad hoy tal vez no lo sea mañana.

En la figura 1-29 se muestra el panorama completo. Examinaremos esta figura pieza por pieza, empezando con una computadora en el hogar (en los extremos de la figura). Para unirse a Internet, la computadora se conecta a un **Proveedor de servicios de Internet**, o simplemente **ISP**, a quien el usuario compra **acceso** o **conectividad a Internet**. Esto permite a la computadora intercambiar paquetes con todos los demás hosts accesibles en Internet. El usuario podría enviar paquetes para navegar por la web o para cualquiera de los otros miles de usos, en realidad no importa. Hay muchos tipos de acceso a Internet y por lo general se distinguen con base en el ancho de banda que se ofrece además de su costo, pero el atributo más importante es la conectividad.

Una manera común de conectar un ISP es mediante la línea telefónica, en cuyo caso su compañía telefónica será su ISP. La tecnología **DSL** (**Línea de Suscriptor Digital**, del inglés *Digital Subscriber Line*) reutiliza la línea telefónica que se conecta a su casa para obtener una transmisión de datos digital. La

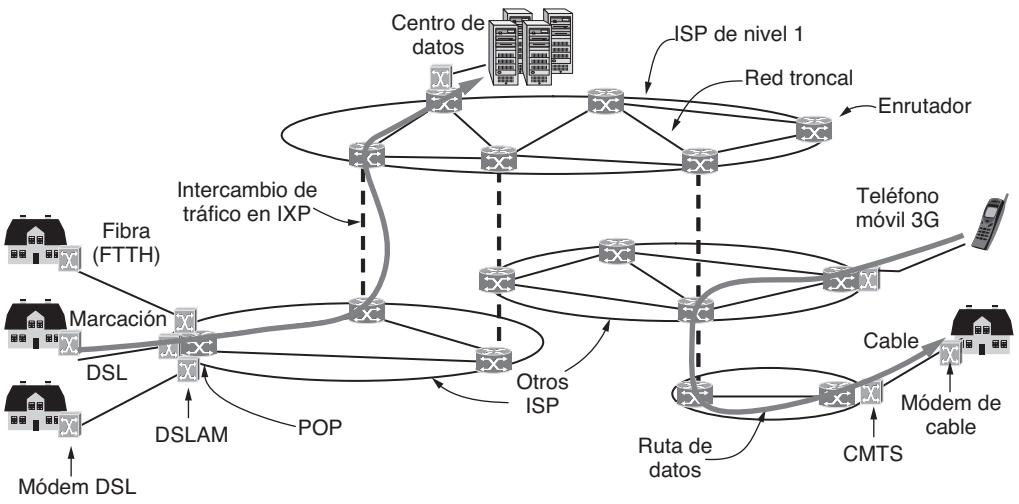

Figura 1-29. Generalidades sobre la arquitectura de Internet.

computadora se conecta a un dispositivo conocido como **módem DSL**, el cual realiza la conversión entre los paquetes digitales y las señales analógicas que pueden pasar libremente a través de la línea telefónica. En el otro extremo hay un dispositivo llamado **DSLAM** (**Multiplexor de Acceso a la Línea de Suscriptor Digital**, del inglés *Digital Subscriber Line Access Multiplexer*) que realiza la conversión entre señales y paquetes.

Hay otras formas populares de conectarse a un ISP, las cuales se muestran en la figura 1-29. DSL es una opción de utilizar la línea telefónica local con más ancho de banda que la acción de enviar bits a través de una llamada telefónica tradicional en vez de una conversación de voz. A esto último se le conoce como **marcación** y se lleva a cabo con un tipo distinto de módem en ambos extremos. La palabra **módem** es la abreviación de "*mo*dulador *dem*odulador" y se refiere a cualquier dispositivo que realiza conversiones entre bits digitales y señales analógicas.

Otro método es enviar señales a través del sistema de TV por cable. Al igual que DSL, ésta es una forma de reutilizar la infraestructura existente, que en este caso es a través de los canales de TV por cable que no se utilizan. El dispositivo en el extremo conectado a la casa se llama **módem de cable** y el dispositivo en la **cabecera del cable** se llama **CMTS** (**Sistema de Terminación del Módem de Cable**, del inglés *Cable Modem Termination System*).

Las tecnologías DSL y de TV por cable proveen acceso a Internet con velocidades que varían desde una pequeña fracción de un megabit/segundo hasta varios megabits/segundo, dependiendo del sistema. Estas velocidades son mucho mayores que en las líneas de marcación, las cuales se limitan a 56 kbps debido al estrecho ancho de banda que se utiliza para las llamadas de voz. Al acceso a Internet con una velocidad mucho mayor que la de marcación se le llama **banda ancha**. El nombre hace referencia al ancho de banda más amplio que se utiliza para redes más veloces, en vez de hacer referencia a una velocidad específica.

Los métodos de acceso mencionados hasta ahora se limitan con base en el ancho de banda de la "última milla" o último tramo de transmisión. Al usar cable de fibra óptica en las residencias, se puede proveer un acceso más rápido a Internet con velocidades en el orden de 10 a 100 Mbps. A este diseño se le conoce como **FTTH** (**Fibra para el Hogar**, del inglés *Fiber To The Home*). Para los negocios en áreas comerciales tal vez tenga sentido rentar una línea de transmisión de alta velocidad de las oficinas hasta el ISP más cercano. Por ejemplo, en Estados Unidos una línea T3 opera aproximadamente a 45 Mbps.

La tecnología inalámbrica también se utiliza para acceder a Internet. Un ejemplo que veremos en breve es el de las redes de teléfonos móviles 3G. Estas redes pueden proveer una transmisión de datos a velocidades de 1 Mbps o mayores para los teléfonos móviles y los suscriptores fijos que se encuentren en el área de cobertura.

Ahora podemos mover los paquetes entre el hogar y el ISP. A la ubicación en la que los paquetes entran a la red del ISP para que se les dé servicio le llamamos el **POP** (**Punto De Presencia**, del inglés *Point Of Presence*) del ISP. A continuación explicaremos cómo se mueven los paquetes entre los POP de distintos ISP. De aquí en adelante, el sistema es totalmente digital y utiliza la conmutación de paquetes.

Las redes de ISP pueden ser de alcance regional, nacional o internacional. Ya hemos visto que su arquitectura está compuesta de líneas de transmisión de larga distancia que interconectan enrutadores en los POP de las distintas ciudades a las que los ISP dan servicio. A este equipo se le denomina la **red troncal** (*backbone*) del ISP. Si un paquete está destinado a un host al que el ISP da servicio directo, ese paquete se encamina a través de la red troncal y se entrega al host. En caso contrario, se debe entregar a otro ISP.

Los ISP conectan sus redes para intercambiar tráfico en lo que llamamos un **IXP** (**Punto de Intercambio en Internet**, del inglés *Internet eXchange Points*). Se dice que los ISP conectados **intercambian tráfico** entre sí. Hay muchos IXP en ciudades de todo el mundo. Se dibujan en sentido vertical en la figura 1-29 debido a que las redes de ISP se traslapan geográficamente. En esencia, un IXP es un cuarto lleno de enrutadores, por lo menos uno por ISP. Una LAN en el cuarto conecta a todos los enrutadores, de modo que los paquetes se pueden reenviar desde cualquier red troncal de ISP a cualquier otra red troncal de ISP. Los IXP pueden ser instalaciones extensas pertenecientes a entidades independientes. Uno de los más grandes es Amsterdam Internet Exchange, en donde se conectan cientos de ISP y a través del cual intercambian cientos de gigabits/segundo de tráfico.

El intercambio de tráfico (*peering*) que ocurre en los IXP depende de las relaciones comerciales entre los ISP. Hay muchas relaciones posibles. Por ejemplo, un ISP pequeño podría pagar a un ISP más grande para obtener conectividad a Internet para alcanzar hosts distantes, así como cuando un cliente compra servicio a un proveedor de Internet. En este caso, se dice que el ISP pequeño paga por el **tránsito**. O tal vez dos ISP grandes decidan intercambiar tráfico de manera que cada ISP pueda entregar cierto tráfico al otro ISP sin tener que pagar por el tránsito. Una de las diversas paradojas de Internet es que los ISP que compiten públicamente por los clientes, cooperan con frecuencia en forma privada para intercambiar tráfico (Metz, 2001).

La ruta que toma un paquete por Internet depende de las opciones de intercambio de tráfico de los ISP. Si el ISP que va a entregar un paquete intercambia tráfico con el ISP de destino, podría entregar el paquete directamente a su igual. En caso contrario, podría encaminar el paquete hasta el lugar más cercano en donde se conecte con un proveedor de tránsito pagado, de manera que éste pueda entregar el paquete. En la figura 1-29 se dibujan dos rutas de ejemplo a través de los ISP. Es muy común que la ruta que toma un paquete no sea la ruta más corta a través de Internet.

En la parte superior de la "cadena alimenticia" se encuentra un pequeño grupo de empresas, como AT&T y Sprint, que operan extensas redes troncales internacionales con miles de enrutadores conectados mediante enlaces de fibra óptica con un extenso ancho de banda. Estos ISP no pagan por el tránsito. Por lo general se les denomina ISP de **nivel 1** y se dice que forman la red troncal de Internet, ya que todos los demás se tienen que conectar a ellos para poder llegar a toda la Internet.

Las empresas que proveen mucho contenido, como Google y Yahoo!, tienen sus computadoras en **centros de datos** que están bien conectados al resto de Internet. Estos centros de datos están diseñados para computadoras, no para humanos, y pueden contener estante (*rack*) tras estante de máquinas, a lo que llamamos **granja de servidores**. Los centros de datos de **colocación** u **hospedaje** permiten a los clientes tener equipo como servidores en los POP de un ISP, de manera que se puedan realizar conexiones cortas y rápidas entre los servidores y las redes troncales del ISP. La industria de hospedaje en Internet se

está virtualizando cada vez más, de modo que ahora es común rentar una máquina virtual que se ejecuta en una granja de servidores en vez de instalar una computadora física. Estos centros de datos son tan grandes (decenas o cientos de miles de máquinas) que la electricidad es uno de los principales costos, por lo que algunas veces estos centros de datos se construyen en áreas en donde el costo de la electricidad sea más económico.

Con esto terminamos nuestra breve introducción a Internet. En los siguientes capítulos tendremos mucho qué decir sobre los componentes individuales y su diseño, los algoritmos y los protocolos. Algo más que vale la pena mencionar aquí es que el significado de estar en Internet está cambiando. Antes se decía que una máquina estaba en Internet si: (1) ejecutaba la pila de protocolos TCP/IP; (2) tenía una dirección IP; y (3) podía enviar paquetes IP a todas las demás máquinas en Internet. Sin embargo, a menudo los ISP reutilizan las direcciones dependiendo de las computadoras que se estén utilizando en un momento dado, y es común que las redes domésticas compartan una dirección IP entre varias computadoras. Esta práctica quebranta la segunda condición. Las medidas de seguridad, como los firewalls, también pueden bloquear en parte las computadoras para que no reciban paquetes, con lo cual se quebranta la tercera condición. A pesar de estas dificultades, tiene sentido decir que esas máquinas estarán en Internet mientras permanezcan conectadas a sus ISP.

También vale la pena mencionar que algunas compañías han interconectado todas sus redes internas existentes, y con frecuencia usan la misma tecnología que Internet. Por lo general, se puede acceder a estas **intranets** sólo desde las premisas de la compañía o desde computadoras notebook de la empresa, pero en los demás aspectos funcionan de la misma manera que Internet.

1.5.2 Redes de teléfonos móviles de tercera generación

A las personas les encanta hablar por teléfono mucho más de lo que les gusta navegar en Internet, y esto ha logrado que la red de teléfonos móviles sea la más exitosa del mundo. Tiene más de cuatro mil millones de suscriptores a nivel mundial. Para poner esta cantidad en perspectiva, digamos que constituye aproximadamente 60% de la población mundial y es mucho más que la cantidad de hosts de Internet y líneas telefónicas fijas combinadas (ITU, 2009).

La arquitectura de la red de teléfonos móviles ha cambiado y ha crecido de manera considerable durante los últimos 40 años. Los sistemas de telefonía móvil de primera generación transmitían las llamadas de voz como señales de variación continua (analógicas) en vez de secuencias de bits (digitales). El sistema **AMPS** (**Sistema Telefónico Móvil Avanzado**, del inglés *Advanced Mobile Phone System*), que se desarrolló en Estados Unidos en 1982, fue un sistema de primera generación muy popular. Los sistemas de teléfonos móviles de segunda generación cambiaron a la transmisión de las llamadas de voz en formato digital para aumentar su capacidad, mejorar la seguridad y ofrecer mensajería de texto. El sistema **GSM** (**Sistema Global para Comunicaciones Móviles**, del inglés *Global System for Mobile communications*), que se implementó a partir de 1991 y se convirtió en el sistema de telefonía móvil más utilizado en el mundo, es un sistema 2G.

Los sistemas de tercera generación (o 3G) comenzaron a implementarse en el año 2001 y ofrecen servicios de datos tanto de voz digital como de datos digitales de banda ancha. También vienen con mucho lenguaje tecnológico y distintos estándares a elegir. La ITU (una organización internacional de estándares de la que hablaremos en la siguiente sección) define al estándar 3G en sentido general como un servicio que ofrece velocidades de por lo menos 2 Mbps para usuarios estacionarios o móviles, y de 384 kbps en un vehículo en movimiento. El sistema **UMTS** (**Sistema Universal de Telecomunicaciones Móviles**, del inglés *Universal Mobile Telecommunications System*), también conocido como **WCDMA** (**Acceso Múltiple por División de Código de Banda Ancha**, del inglés *Wideband Code Division Multiple Access*), es el principal sistema 3G que se está implementando con rapidez en todo el mundo. Puede proveer hasta

14 Mbps en el enlace de bajada y casi 6 Mbps en el enlace de subida. Las futuras versiones utilizarán varias antenas y radios para proveer velocidades aún mayores para los usuarios.

El recurso escaso en los sistemas 3G, al igual que en los sistemas 2G y 1G anteriores, es el espectro de radio. Los gobiernos conceden el derecho de usar partes del espectro a los operadores de la red de telefonía móvil, a menudo mediante una subasta de espectro en donde los operadores de red realizan ofertas. Es más fácil diseñar y operar sistemas cuando se tiene una parte del espectro con licencia, ya que a nadie más se le permite transmitir en ese espectro, pero la mayoría de las veces es algo muy costoso. Por ejemplo, en el Reino Unido en el año 2000, se subastaron cinco licencias para 3G por un total aproximado de $40 mil millones de dólares.

Esta escasez del espectro es la que condujo al diseño de la **red celular** que se muestra en la figura 1-30 y que ahora se utiliza en las redes de telefonía móvil. Para manejar la interferencia de radio entre los usuarios, el área de cobertura se divide en celdas. Dentro de una celda, a los usuarios se les asignan canales que no interfieren entre sí y que no provocan mucha interferencia para las celdas adyacentes. Esto permite una reutilización eficiente del espectro, o **reutilización de frecuencia**, en las celdas adyacentes, lo cual incrementa la capacidad de la red. En los sistemas 1G, que transmitían cada llamada de voz en una banda de frecuencia específica, las frecuencias se elegían con cuidado de modo que no tuvieran conflictos con las celdas adyacentes. De esta forma, una frecuencia dada sólo se podría reutilizar una vez en varias celdas. Los sistemas 3G modernos permiten que cada celda utilice todas las frecuencias, pero de una manera que resulte en un nivel tolerable de interferencia para las celdas adyacentes. Existen variaciones en el diseño celular, incluyendo el uso de antenas direccionales o sectorizadas en torres de celdas para reducir aún más la interferencia, pero la idea básica es la misma.

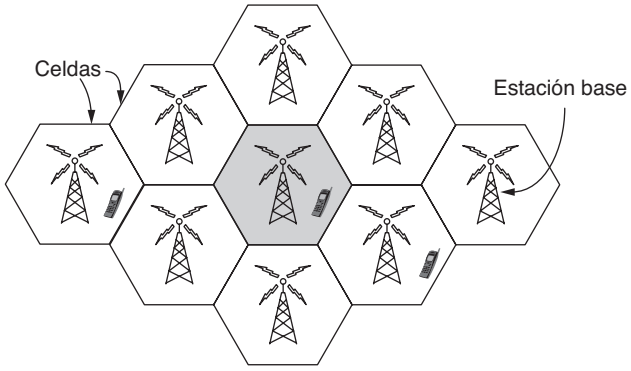

Figura 1-30. Diseño celular de las redes de telefonía móvil.

La arquitectura de la red de telefonía móvil es muy distinta a la de Internet. Tiene varias partes, como se muestra en la versión simplificada de la arquitectura UMTS en la figura 1-31. Primero tenemos a la **interfaz aérea**. Éste es un término elegante para el protocolo de radiocomunicación que se utiliza a través del aire entre el dispositivo móvil (como el teléfono celular) y la **estación base celular**. Los avances en la interfaz aérea durante las últimas décadas han aumentado en forma considerable las velocidades de datos inalámbricas. La interfaz aérea de UTMS se basa en el **Acceso Múltiple por División de Código** (**CDMA**, del inglés *Code Division Multiple Access*), una técnica que estudiaremos en el capítulo 2.

La estación base celular forma junto con su controlador la **red de acceso por radio**. Esta parte constituye el lado inalámbrico de la red de telefonía móvil. El nodo controlador o **RNC** (**Controlador de la**

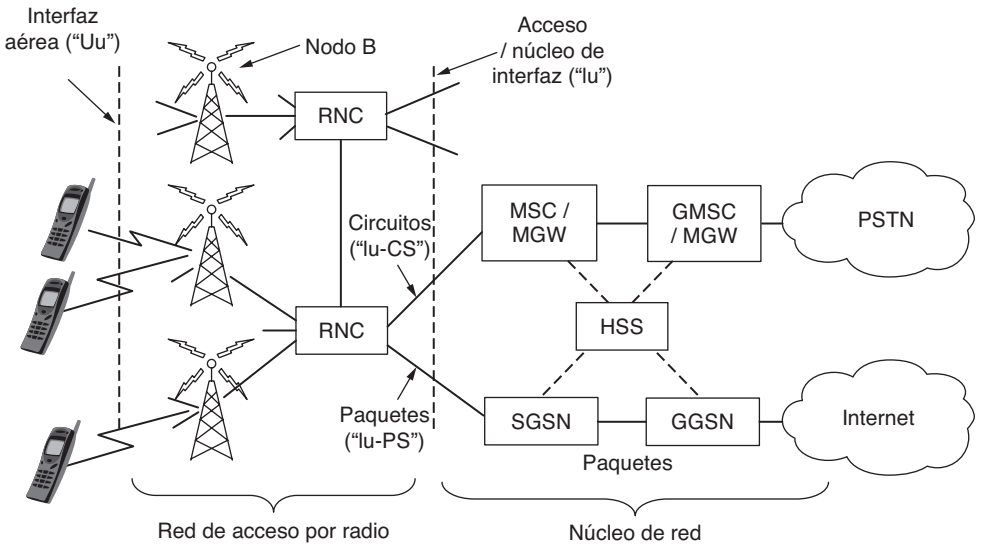

Figura 1-31. Arquitectura de la red de telefonía móvil 3G UTMS.

Red de Radio, del inglés *Radio Network Controller*) controla la forma en que se utiliza el espectro. La estación base implementa a la interfaz aérea. A ésta se le conoce como **Nodo B**, una etiqueta temporal que se quedó para siempre.

El resto de la red de telefonía móvil transporta el tráfico para la red de acceso por radio. A esto se le conoce como **núcleo de red**. La red básica UMTS evolucionó a partir de la red básica que se utilizaba para el sistema GSM 2G anterior. Sin embargo, algo sorprendente está ocurriendo en la red básica UMTS.

Desde los inicios de las redes se ha venido desatando una guerra entre las personas que apoyan las redes de paquetes (es decir, subredes sin conexión) y las personas que apoyan las redes de circuitos (es decir, redes orientadas a conexión). Los principales defensores de los paquetes provienen de la comunidad de Internet. En un diseño sin conexión, cada paquete se encamina de manera independiente a los demás paquetes. Como consecuencia, si algunos enrutadores fallan durante una sesión, no habrá daño alguno siempre y cuando el sistema pueda reconfigurarse a sí mismo en forma dinámica, de modo que los siguientes paquetes puedan encontrar una ruta a su destino, aun cuando sea distinta a la que hayan utilizado los paquetes anteriores.

El campo de circuitos proviene del mundo de las compañías telefónicas. En el sistema telefónico, un usuario debe marcar el número de la parte a la que va a llamar y esperar una conexión antes de poder hablar o enviar datos. Esta forma de realizar la conexión establece una ruta a través del sistema telefónico que se mantiene hasta terminar la llamada. Todas las palabras o paquetes siguen la misma ruta. Si falla una línea o un interruptor en la ruta se aborta la llamada, es decir, es un método menos tolerante a las fallas en comparación con el diseño sin conexión.

La ventaja de los circuitos es que soportan la calidad del servicio con más facilidad. Al establecer una conexión por adelantado, la subred puede reservar recursos como el ancho de banda del enlace, el espacio de búfer de los switches, y tiempo de la CPU. Si alguien intenta hacer una llamada y no hay suficientes recursos disponibles, la llamada se rechaza y el usuario recibe una señal de ocupado. De esta forma, una vez establecida la conexión, recibirá un buen servicio.

Con una red sin conexión, si llegan demasiados paquetes al mismo enrutador en el mismo momento, es probable que pierda algunos. El emisor se dará cuenta de esto en un momento dado y volverá a enviarlos, pero la calidad del servicio será intermitente e inadecuada para transmitir audio o video, a menos que

la red tenga una carga ligera. Sin necesidad de decirlo, proveer una calidad adecuada de audio y video es algo por lo que las compañías telefónicas se preocupan mucho, de aquí que prefieran un servicio orientado a la conexión.

La sorpresa en la figura 1-31 es que hay equipo tanto de paquetes como de conmutación de circuitos en el núcleo de red. Esto muestra a la red de telefonía móvil en transición, en donde las compañías de telefonía móvil pueden implementar una o, en ocasiones, ambas alternativas. Las redes de telefonía móvil antiguas usaban un núcleo de conmutación de paquetes al estilo de la red telefónica tradicional para transmitir las llamadas de voz. Esta herencia se puede ver en la red UMTS con los elementos **MSC** (**Centro de Conmutación Móvil**, del inglés *Mobile Switching Center*), **GMSC** (**Centro de Conmutación Móvil de Puerta de Enlace**, del inglés *Gateway Mobile Switching Center*) y **MGW** (**Puerta de Enlace de Medios**, del inglés *Media Gateway*) que establecen conexiones a través de un núcleo de red con conmutación de paquetes como **PSTN** (**Red Telefónica Pública Conmutada**, del inglés *Public Switched Telephone Network*).

Los servicios de datos se han convertido en una parte de la red de telefonía móvil mucho más importante de lo que solían ser, empezando con la mensajería de texto y los primeros servicios de datos de paquetes, como **GPRS** (**Servicio General de Paquetes de Radio**, del inglés *General Packet Radio Service*) en el sistema GSM. Estos servicios de datos antiguos operaban a decenas de kbps, pero los usuarios querían más. En comparación, una llamada de voz se transmite a una velocidad de 64 kbps, comúnmente de 3 a 4 veces menos con compresión.

Para transmitir todos estos datos, los nodos del núcleo de red UMTS se conectan directamente a una red de conmutación de paquetes. El **SGSN** (**Nodo de Soporte del Servicio GPRS**, del inglés *Serving GPRS Support Node*) y el **GGSN** (**Nodo de Soporte de la Puerta de Enlace de GPRS***, del inglés Gateway GPRS Support Node*) transmiten paquetes de datos hacia y desde dispositivos móviles y hacen interfaz con redes de paquetes externas, como Internet.

Esta transición está destinada a continuar en las redes de telefonía móvil que se planean e implementan en la actualidad. Incluso se utilizan protocolos de Internet en dispositivos móviles para establecer conexiones para llamadas de voz a través de una red de paquetes de datos, en forma de voz sobre IP. El protocolo IP y los paquetes se utilizan en todo el camino, desde el acceso por radio hasta el núcleo de red. Desde luego que también se están haciendo cambios en el diseño de las redes IP para soportar una mejor calidad de servicio. Si no fuera así, los problemas con la señal entrecortada de audio y video no impresionarían a los clientes y dejarían de pagar. En el capítulo 5 retomaremos este tema.

Otra diferencia entre las redes de telefonía móvil y la Internet tradicional es la movilidad. Cuando un usuario se sale del rango de una estación base celular y entra al rango de otra, el flujo de datos se debe encaminar nuevamente desde la estación antigua hasta la nueva estación base celular. A esta técnica se le conoce como **traspaso** (*handover*) o **entrega** (*handoff*) y se ilustra en la figura 1-32.

El dispositivo móvil o la estación base pueden solicitar un traspaso si disminuye la calidad de la señal. En algunas redes celulares (por lo general las que están basadas en tecnología CDMA) es posible conec-

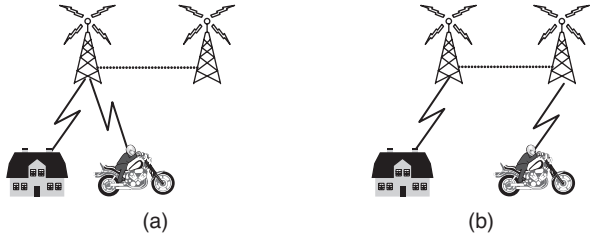

Figura 1-32. Traspaso de telefonía móvil (a) antes, (b) después.

tarse a la nueva estación base antes de desconectarse de la estación anterior. Esto mejora la calidad de la conexión para el dispositivo móvil, ya que no se interrumpe el servicio; el dispositivo móvil se conecta a dos estaciones base por un breve instante. A esta manera de realizar un traspaso se le llama **traspaso suave** para diferenciarla de un **traspaso duro**, en donde el dispositivo móvil se desconecta de la estación base anterior antes de conectarse a la nueva estación.

Una cuestión relacionada es cómo buscar un móvil en primer lugar cuando hay una llamada entrante. Cada red de telefonía móvil tiene un **HSS** (**Servidor de Suscriptores Locales**, del inglés *Home Subscriber Server*) en el núcleo de red, el cual conoce la ubicación de cada suscriptor así como demás información de perfil que se utiliza para la autenticación y la autorización. De esta forma, para encontrar un dispositivo móvil hay que ponerse en contacto con el HSS.

El último tema en cuestión es la seguridad. A través de la historia, las compañías telefónicas han tomado la seguridad mucho más en serio que las compañías de Internet por mucho tiempo, debido a la necesidad de cobrar por el servicio y evitar el fraude (en los pagos). Por desgracia, esto no dice mucho. Sin embargo, en la evolución de la tecnología 1G a la 3G, las compañías de telefonía móvil han sido capaces de desarrollar varios mecanismos básicos de seguridad para dispositivos móviles.

A partir del sistema GSM 2G, el teléfono móvil se dividió en una terminal y un chip removible que contenía la identidad del suscriptor y la información de su cuenta. Al chip se le conoce de manera informal como **tarjeta SIM** (**Módulo de Identidad del Suscriptor**, del inglés *Subscriber Identity Module*). Las tarjetas SIM se pueden usar en distintas terminales para activarlas, además de que proveen una seguridad básica. Cuando los clientes de GSM viajan a otros países por motivos de negocios o de placer, a menudo traen consigo sus terminales pero compran una nueva tarjeta SIM por unos cuantos dólares al llegar, para poder hacer llamadas locales sin cargos de roaming.

Para reducir los fraudes, la red de telefonía móvil también usa la información en las tarjetas SIM para autenticar a los suscriptores y verificar que puedan usar la red. Con el sistema UTMS, el dispositivo móvil también usa la información en la tarjeta SIM para verificar que está hablando con una red legítima.

La privacidad es otro aspecto de la seguridad. Las señales inalámbricas se difunden a todos los receptores cercanos, por lo que para evitar que alguien pueda espiar las conversaciones se utilizan claves criptográficas en la tarjeta SIM para cifrar las transmisiones. Esta metodología ofrece una mayor privacidad que en los sistemas 1G, que se podían intervenir fácilmente, pero no es una panacea debido a las debilidades en los esquemas de cifrado.

Las redes de telefonía móvil están destinadas a desempeñar un papel central en las futuras redes. Ahora tratan más sobre aplicaciones móviles de banda ancha que sobre llamadas de voz, y esto tiene implicaciones importantes para las interfaces aéreas, la arquitectura del núcleo de red y la seguridad de las futuras redes. Las tecnologías 4G que son más veloces y mejores ya están en fase de diseño bajo el nombre de **LTE** (**Evolución a Largo Plazo**, del inglés *Long Term Evolution*), incluso a medida que continúa el diseño y el desarrollo de la tecnología 3G. Hay otras tecnologías inalámbricas que también ofrecen acceso a Internet de banda ancha para clientes fijos y móviles, en particular las redes 802.16 bajo el nombre común de **WiMAX**. Es totalmente posible que LTE y WiMAX vayan a chocar en un futuro y es difícil predecir qué les ocurrirá.

1.5.3 Redes LAN inalámbricas: 802.11

Casi al mismo tiempo en que aparecieron las computadoras laptop, muchas personas soñaban con entrar a una oficina y que su laptop se conectara mágicamente a Internet. En consecuencia, varios grupos empezaron a trabajar en formas para lograr este objetivo. La metodología más práctica consiste en equipar tanto a la oficina como las computadoras laptop con transmisores de radio de corto alcance y receptores para que se puedan comunicar.

El trabajo en este campo condujo rápidamente a que varias empresas empezaran con la comercialización de las redes LAN inalámbricas. El problema era que ni siquiera había dos de ellas que fueran compatibles. La proliferación de estándares implicaba que una computadora equipada con un radio marca *X* no trabajaría en un cuarto equipado con una estación base marca *Y*. A mediados de la década de 1990, la industria decidió que sería muy conveniente tener un estándar para las redes LAN inalámbricas, de modo que el comité IEEE que había estandarizado las redes LAN alámbricas recibió la tarea de idear un estándar para redes LAN inalámbricas.

La primera decisión fue la más sencilla: cómo llamar a este estándar. Todos los demás estándares de LAN tenían números como 802.1, 802.2 y 802.3 hasta 802.10, así que al estándar de LAN inalámbrica se le dio el número 802.11. En la jerga computacional a este estándar se le conoce con el nombre de **WiFi**, pero es un estándar importante y merece respeto, de modo que lo llamaremos por su nombre: 802.11.

El resto fue más difícil. El primer problema era hallar una banda de frecuencia adecuada que estuviera disponible, de preferencia a nivel mundial. La metodología utilizada fue contraria a la que se utilizó en las redes de telefonía móvil. En vez de un espectro costoso bajo licencia, los sistemas 802.11 operan en bandas sin licencia como las bandas **ISM** (**Industriales, Científicas y Médicas**, del inglés *Industrial, Scientific, and Medical*) definidas por el ITU-R (por ejemplo, 902-929 MHz, 2.4-2.5 GHz, 5.725-5.825 GHz). Todos los dispositivos pueden usar este espectro siempre y cuando limiten su potencia de transmisión para dejar que coexistan distintos dispositivos. Desde luego que esto significa que los radios 802.11 podrían entrar en competencia con los teléfonos inalámbricos, los abridores de puertas de garaje y los hornos de microondas.

Las redes 802.11 están compuestas de clientes (como laptops y teléfonos móviles) y de una infraestructura llamada **AP** (**Puntos de Acceso**) que se instala en los edificios. Algunas veces a los puntos de acceso se les llama **estaciones base**. Los puntos de acceso se conectan a la red alámbrica y toda la comunicación entre los clientes se lleva a cabo a través de un punto de acceso. También es posible que los clientes que están dentro del rango del radio se comuniquen en forma directa, como en el caso de dos computadoras en una oficina sin un punto de acceso. A este arreglo se le conoce como **red *ad hoc***. Se utiliza con menor frecuencia que el modo de punto de acceso. En la figura 1-33 se muestran ambos modos.

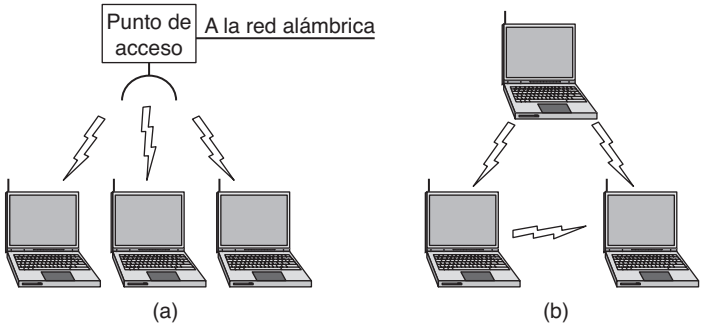

Figura 1-33. (a) Red inalámbrica con un punto de acceso. (b) Red *ad hoc*.

La transmisión 802.11 se complica debido a las condiciones inalámbricas que varían incluso con pequeños cambios en el entorno. En las frecuencias usadas para 802.11 las señales de radio pueden

rebotar de objetos sólidos, de modo que varios ecos de una transmisión podrían llegar a un receptor a través de distintas rutas. Los ecos se pueden cancelar o reforzar unos a otros y provocar que la señal recibida fluctúe de manera considerable. Este fenómeno se llama **desvanecimiento multitrayectoria** y se muestra en la figura 1-34.

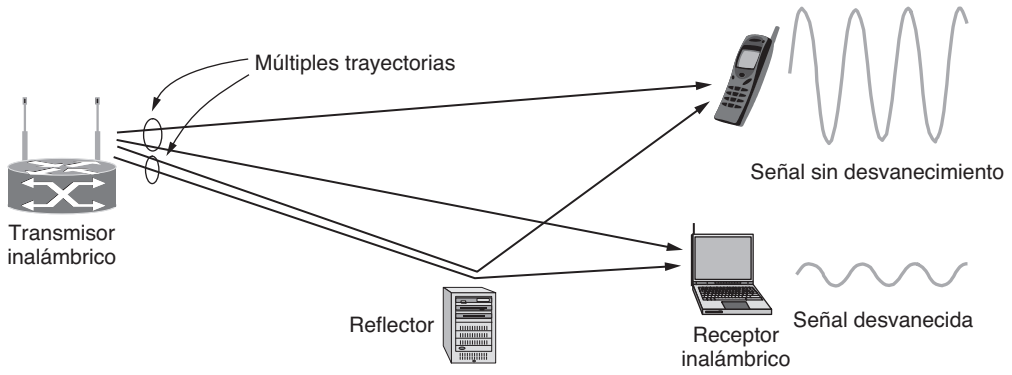

Figura 1-34. Desvanecimiento multitrayectorias.

La idea clave para solventar las condiciones inalámbricas variables es la **diversidad de rutas**, o el envío de información a través de múltiples rutas independientes. De esta forma, es probable que la información se reciba incluso si una de las rutas resulta ser pobre debido a un desvanecimiento. Por lo general estas rutas independientes están integradas al esquema de modulación digital en la capa física. Las opciones incluyen el uso de distintas frecuencias a lo largo de la banda permitida, en donde se siguen distintas rutas espaciales entre los distintos pares de antenas o se repiten bits durante distintos periodos.

Las distintas versiones de 802.11 han usado todas estas técnicas. El estándar inicial (1997) definió una LAN inalámbrica que podía operar a 1 Mbps o 2 Mbps mediante saltos entre frecuencias o también se podía extender la señal a lo largo del espectro permitido. Casi de inmediato surgieron las quejas de las personas diciendo que era muy lenta, por lo que se empezó a trabajar en estándares más veloces. El diseño de espectro extendido se amplió y convirtió en el estándar 802.11b (1999) que operaba a velocidades de hasta 11 Mbps. Los estándares 802.11a (1999) y 802.11g (2003) cambiaron a un esquema de modulación distinto llamado **OFDM** (**Multiplexado por División de Frecuencias Ortogonales**, del inglés *Orthogonal Frequency Division Multiplexing*). Este esquema divide una banda amplia de espectro en muchas fracciones estrechas, a través de las cuales se envían distintos bits en paralelo. Este esquema mejorado, que estudiaremos en el capítulo 2, logró aumentar las velocidades en bits de los estándares 802.11a/g hasta 54 Mbps. Es un aumento considerable, pero las personas querían una velocidad aún mayor para soportar usos más demandantes. La versión más reciente es 802.11n (2009), la cual utiliza bandas de frecuencia más amplias y hasta cuatro antenas por computadora para alcanzar velocidades de hasta 450 Mbps.

Como la tecnología inalámbrica es un medio de difusión por naturaleza, los radios 802.11 también tienen que lidiar con el problema de que las múltiples transmisiones que se envían al mismo tiempo tendrán colisiones, lo cual puede interferir con la recepción. Para encargarse de este problema, 802.11 utiliza un esquema **CSMA** (**Acceso Múltiple por Detección de Portadora**, del inglés *Carrier Sense Multiple Access*) basado en ideas provenientes de la Ethernet alámbrica que, irónicamente, se basó en una de las primeras redes inalámbricas desarrolladas en Hawai, llamada **ALOHA**. Las computadoras esperan durante un intervalo corto y aleatorio antes de transmitir, y difieren sus transmisiones si escuchan que hay alguien más transmitiendo. Este esquema reduce la probabilidad de que dos computadoras envíen datos al mismo tiempo, pero no funciona tan bien como en el caso de las computadoras conectadas por cables.

Para ver por qué, examine la figura 1-35. Suponga que la computadora *A* está transmitiendo datos a la computadora *B*, pero el rango de radio del transmisor de *A* es demasiado corto como para llegar a la computadora *C*. Si *C* desea transmitir a *B* puede escuchar antes de empezar, pero el hecho de que no escuche nada no significa que su transmisión vaya a tener éxito. La incapacidad de *C* de escuchar a *A* antes de empezar provoca algunas colisiones. Después de una colisión, el emisor espera durante un retardo aleatorio más largo y vuelve a transmitir el paquete. A pesar de ésta y de otras cuestiones, el esquema funciona bastante bien en la práctica.

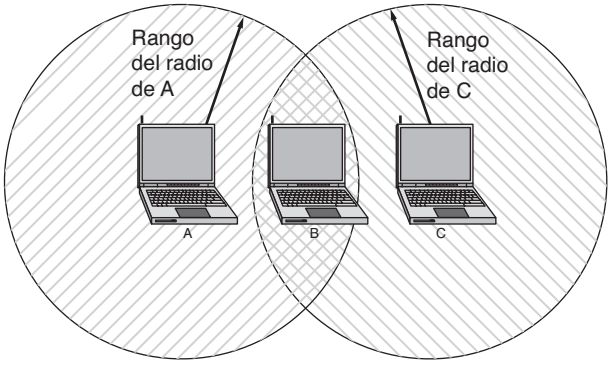

Figura 1-35. El rango de un solo radio tal vez no cubra todo el sistema.

Otro problema es la movilidad. Si un cliente móvil se aleja del punto de acceso que utiliza y entra en el rango de un punto de acceso distinto, se requiere alguna forma de entrega. La solución es que una red 802.11 puede consistir de múltiples celdas, cada una con su propio punto de acceso, y de un sistema de distribución que las conecte. Con frecuencia el sistema de distribución es Ethernet conmutada, pero puede usar cualquier tecnología. A medida que los clientes se desplazan, tal vez encuentren otro punto de acceso con una mejor señal que la que tienen en ese momento y pueden cambiar su asociación. Desde el exterior, el sistema completo se ve como una sola LAN alámbrica.

Aclarado el punto, la movilidad en el estándar 802.11 ha sido de un valor limitado si se le compara con la movilidad disponible en la red de telefonía móvil. Por lo general, el 802.11 lo utilizan los clientes nómadas que van de una ubicación fija a otra, en vez de usarlo en el camino. Estos clientes en realidad no necesitan movilidad. Incluso cuando se utiliza la movilidad que ofrece el estándar 802.11, se extiende sobre una sola red 802.11, que podría cubrir cuando mucho un edificio extenso. Los esquemas en lo futuro tendrán que proveer movilidad a través de distintas redes y diferentes tecnologías (por ejemplo, 802.21).

Por último tenemos el problema de la seguridad. Como las transmisiones inalámbricas son difundidas, es fácil que las computadoras cercanas reciban paquetes de información que no estaban destinados para ellas. Para evitar esto, el estándar 802.11 incluyó un esquema de cifrado conocido como **WEP** (**Privacidad Equivalente a Cableado**, del inglés *Wired Equivalent Privacy*). La idea era lograr que la seguridad inalámbrica fuera igual a la seguridad alámbrica. Es una buena idea, pero por desgracia el esquema era imperfecto y no pasó mucho tiempo para que fallara (Borisov y colaboradores, 2001). Desde entonces se reemplazó con esquemas más recientes que tienen distintos detalles criptográficos en el estándar 802.11i, conocido también como **Acceso protegido WiFi**, que en un principio se llamó **WPA** pero ahora se reemplazó por el **WPA2**.

El estándar 802.11 provocó una revolución en las redes inalámbricas que está destinada a continuar. Aparte de los edificios, se ha empezado a instalar en trenes, aviones, botes y automóviles de modo que las

personas puedan navegar por Internet en cualquier parte a donde vayan. Los teléfonos móviles y todo tipo de electrodomésticos, desde las consolas de juego hasta las cámaras digitales, se pueden comunicar con este estándar. En el capítulo 4 hablaremos detalladamente sobre este estándar.

1.5.3 Redes RFID y de sensores

Las redes que hemos estudiado hasta ahora están compuestas de dispositivos de cómputo fáciles de reconocer, desde computadoras hasta teléfonos móviles. Gracias a la **Identificación por Radio Frecuencia** (**RFID**), los objetos cotidianos también pueden formar parte de una red de computadoras.

Una etiqueta RFID tiene la apariencia de una calcomanía del tamaño de una estampilla postal que se puede pegar (o incrustar) en un objeto, de modo que se pueda rastrear. El objeto podría ser una vaca, un pasaporte o un libro. La etiqueta consiste en un pequeño microchip con un identificador único y una antena que recibe transmisiones por radio. Los lectores RFID instalados en puntos de rastreo encuentran las etiquetas cuando están dentro del rango y las interrogan para obtener su información como se muestra en la figura 1-36. Las aplicaciones incluyen: verificar identidades, administrar la cadena de suministro, carreras de sincronización y reemplazar códigos de barras.

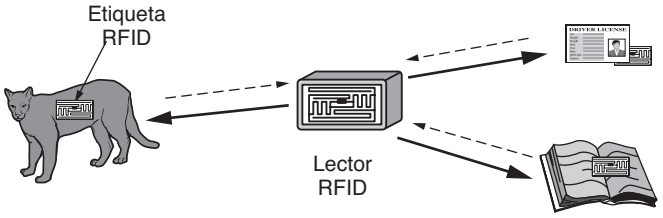

Figura 1-36. La tecnología RFID se utiliza para conectar objetos cotidianos en red.

Hay muchos tipos de RFID, cada uno con distintas propiedades, pero tal vez el aspecto más fascinante de la tecnología RFID sea que la mayoría de las etiquetas RFID no tienen enchufe eléctrico ni batería, sino que toda la energía necesaria para operarlos se suministra en forma de ondas de radio a través de los lectores RFID. A esta tecnología se le denomina **RFID pasiva** para diferenciarla de la **RFID activa** (menos común), en la cual hay una fuente de energía en la etiqueta.

La **RFID de UHF** (**RFID de Ultra Alta Frecuencia**, del inglés *Ultra-High Frequency RFID*) es una forma común de RFID que se utiliza en algunas licencias de conducir. Los lectores envían señales en la banda de 902-928 MHz en Estados Unidos. Las etiquetas se pueden comunicar a distancias de varios metros al cambiar la forma en que reflejan las señales de los lectores; el lector es capaz de recuperar estas reflexiones. A esta forma de operar se le conoce como **retrodispersión** (*backscatter*).

La **RFID de HF** (**RFID de Alta Frecuencia**, del inglés *High Frequency RFID*) es otro tipo popular de RFID que opera a 13.56 MHz y se utiliza por lo general en pasaportes, tarjetas de crédito, libros y sistemas de pago sin contacto. La RFID de HF tiene un rango corto, por lo común de un metro o menos, debido a que el mecanismo físico se basa en la inducción en vez de la retrodispersión. Existen también otras formas de RFID que utilizan otras frecuencias, como la **RFID de LF** (**RFID de Baja Frecuencia**, del inglés *Low Frequency RFID*) que se desarrolló antes de la RFID de HF y se utilizaba para rastrear animales. Es el tipo de RFID que podría llegar a estar en su gato.

Los lectores RFID deben resolver de alguna manera el problema de lidiar con varias etiquetas dentro del rango de lectura. Esto significa que una etiqueta no puede simplemente responder cuando escucha a un lector, o que puede haber colisiones entre las señales de varias etiquetas. La solución es similar a la

metodología aplicada en el estándar 802.11: las etiquetas esperan durante un intervalo corto y aleatorio antes de responder con su identificación, lo cual permite al lector reducir el número de etiquetas individuales e interrogarlas más.

La seguridad es otro problema. La habilidad de los lectores RFID de rastrear con facilidad un objeto, y por ende a la persona que lo utiliza, puede representar una invasión a la privacidad. Por desgracia es difícil asegurar las etiquetas RFID debido a que carecen del poder de cómputo y de comunicación requerido para ejecutar algoritmos criptográficos sólidos. En vez de ello se utilizan medidas débiles como las contraseñas (que se pueden quebrantar con facilidad). Si un oficial en una aduana puede leer de manera remota una tarjeta de identificación, ¿qué puede evitar que otras personas rastreen esa misma tarjeta sin que usted lo sepa? No mucho.

Las etiquetas RFID empezaron como chips de identificación, pero se están convirtiendo con rapidez en computadoras completas. Por ejemplo, muchas etiquetas tienen memoria que se puede actualizar y que podemos consultar después, de modo que se puede almacenar información sobre lo que ocurra con el objeto etiquetado. Reiback y colaboradores (2006) demostraron que esto significa que se aplican todos los problemas comunes del software malicioso de computadora, sólo que ahora sería posible usar su gato o su pasaporte para esparcir un virus de RFID.

La **red de sensores** va un paso más allá en cuanto a capacidad, en comparación con la RFID. Las redes de sensores se implementan para vigilar los aspectos del mundo físico. Hasta ahora se han utilizado en su mayor parte para la experimentación científica, como el monitoreo de los hábitats de las aves, la actividad volcánica y la migración de las cebras, pero es probable que pronto surjan aplicaciones para el cuidado de la salud, equipo de monitoreo de vibraciones y rastreo de artículos congelados, refrigerados u otro tipo de perecederos.

Los nodos sensores son pequeñas computadoras, por lo general del tamaño de un control de llave, que tienen sensores de temperatura, vibración y demás. Muchos nodos se colocan en el entorno que se va a vigilar. Por lo general tienen baterías, aunque también pueden obtener energía de las vibraciones del Sol. Al igual que la RFID, tener suficiente energía es un reto clave por lo que los nodos deben comunicarse con cuidado para transmitir la información de sus sensores a un punto externo de recolección. Una estrategia común es que los nodos se autoorganicen para transmitir mensajes unos de otros, como se muestra en la figura 1-37. Este diseño se conoce como **red multisaltos**.

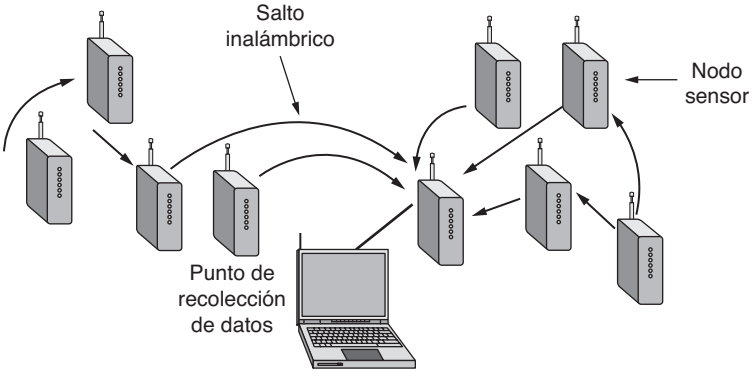

Figura 1-37. Topología multisaltos de una red de sensores.

Es probable que las redes RFID y de sensores sean mucho más capaces y dominantes en el futuro. Los investigadores ya han combinado lo mejor de ambas tecnologías al crear prototipos de etiquetas RFID con sensores de luz, movimiento y otros sensores (Sample y colaboradores, 2008).

1.6 ESTANDARIZACIÓN DE REDES

Existen muchos distribuidores y proveedores de servicios de red, cada uno con sus propias ideas de cómo hacer las cosas. Sin coordinación existiría un caos completo y los usuarios nunca lograrían hacer nada. La única salida es acordar ciertos estándares de redes. Los buenos estándares no sólo permiten que distintas computadoras se comuniquen, sino que también incrementan el mercado para los productos que se adhieren a estos estándares. Un mercado más grande conduce a la producción en masa, economías de escala en la fabricación, mejores implementaciones y otros beneficios que reducen el precio y aumentan más la aceptación.

En esta sección veremos las generalidades sobre el importante pero poco conocido mundo de la estandarización internacional. Pero primero hablaremos sobre lo que debe incluir un estándar. Una persona razonable podría suponer que un estándar nos dice cómo debe funcionar un protocolo, de modo que podamos hacer un buen trabajo al implementarlo. Esa persona estaría equivocada.

Los estándares definen lo que se requiere para la interoperabilidad y nada más. Esto permite que emerja un mercado más grande y también deja que las empresas compitan con base en qué tan buenos son sus productos. Por ejemplo, el estándar 802.11 define muchas velocidades de transmisión pero no dice cuándo un emisor debe utilizar cierta velocidad, lo cual es un factor clave para un buen desempeño. Esto queda a criterio del fabricante del producto. A menudo es difícil obtener una interoperabilidad de esta forma, ya que hay muchas opciones de implementación y los estándares por lo general definen muchas opciones. Para el 802.11 había tantos problemas que, en una estrategia que se convirtió en práctica común, un grupo llamado **Alianza WiFi** empezó a trabajar en la interoperabilidad con el estándar 802.11.

De manera similar, un estándar de protocolos define el protocolo que se va a usar a través del cable pero no la interfaz de servicio dentro de la caja, excepto para ayudar a explicar el protocolo. A menudo las interfaces de servicio reales son de marca registrada. Por ejemplo, la manera en que TCP hace interfaz con IP dentro de una computadora no importa para comunicarse con un host remoto. Sólo importa que el host remoto utilice TCP/IP. De hecho, TCP e IP se implementan juntos con frecuencia sin ninguna interfaz distinta. Habiendo dicho esto, las buenas interfaces de servicio (al igual que las buenas API) son valiosas para lograr que se utilicen los protocolos, además de que las mejores (como los sockets de Berkeley) se pueden volver muy populares.

Los estándares se dividen en dos categorías: de facto y de jure. Los estándares *de facto* (del latín "del hecho") son aquellos que simplemente aparecieron, sin ningún plan formal. El protocolo HTTP con el que opera la web empezó como un estándar de facto. Era parte de los primeros navegadores WWW desarrollados por Tim Berners-Lee en CERN y su uso se popularizó debido al crecimiento de la web. Bluetooth es otro ejemplo. En un principio fue desarrollado por Ericsson, pero ahora todo el mundo lo utiliza.

En contraste, los estándares *de jure* (del latín "por ley") se adoptan por medio de las reglas de alguna organización formal de estandarización. Por lo general las autoridades de estandarización internacionales se dividen en dos clases: las que se establecieron mediante un tratado entre gobiernos nacionales y las conformadas por organizaciones voluntarias que no surgieron de un tratado. En el área de los estándares de redes de computadoras hay varias organizaciones de cada tipo, en especial: ITU, ISO, IETF e IEEE, de las cuales hablaremos a continuación.

En la práctica, las relaciones entre los estándares, las empresas y los organismos de estándares son complicadas. A menudo los estándares de facto evolucionan para convertirse en estándares *de jure*, en especial si tienen éxito. Esto ocurrió en el caso de HTTP, que fue elegido rápidamente por el IETF. Es común que los organismos de estándares ratifiquen los estándares de otros organismos, dando la impresión de aprobarse unos a otros, en un esfuerzo por incrementar el mercado para una tecnología. En estos días, muchas alianzas de negocios *ad hoc* que se forman con base en tecnologías específicas también desempeñan un papel considerable en el desarrollo y refinamiento de los estándares de redes. Por ejemplo, **3GPP**

(**Proyecto de Sociedad de Tercera Generación**, del inglés *Third Generation Parthnership Project*) es una colaboración entre asociaciones de telecomunicaciones que controla los estándares de la telefonía móvil 3GS UMTS.

1.6.1 Quién es quién en el mundo de las telecomunicaciones

El estado legal de las compañías telefónicas del mundo varía de manera considerable de un país a otro. En un extremo se encuentra Estados Unidos, que tiene cerca de 200 compañías privadas telefónicas separadas (la mayoría muy pequeñas). Con la disolución de AT&T en 1984 (que en ese entonces era la corporación más grande del mundo que proveía servicio a cerca del 80% de los teléfonos en América) surgieron unas cuantas compañías más, junto con la Ley de Telecomunicaciones en 1996 que replanteó las reglamentaciones para fomentar la competencia.

Al otro extremo están los países en donde el gobierno nacional tiene un total monopolio sobre toda la comunicación, incluyendo el correo, telégrafo, teléfono y a menudo la radio y televisión. Una gran parte del mundo entra en esta categoría. En algunos casos la autoridad de telecomunicaciones es una compañía nacionalizada, y en otros es simplemente una rama del gobierno, por lo general conocida como **PTT** (**Oficina de Correos, Telegrafía y Teléfonos**, del inglés *Port, Telegraph & Telephone*). A nivel mundial la tendencia es ir hacia la liberalización y la competencia para alejarse del monopolio gubernamental. La mayoría de los países europeos han privatizado ya (en forma parcial) sus oficinas PTT, pero en otras partes el proceso apenas si va ganando fuerza lentamente.

Con todos estos diferentes proveedores de servicios, existe sin duda la necesidad de proveer compatibilidad a escala mundial para asegurar que las personas (y computadoras) en un país puedan llamar a sus contrapartes en otro país. En realidad, esta necesidad ha existido desde hace un buen tiempo. En 1865, los representantes de muchos gobiernos europeos se reunieron para formar el predecesor de lo que hoy es **ITU** (**Unión Internacional de Telecomunicaciones**, del inglés *International Telecommunication Union*). Su tarea era estandarizar las telecomunicaciones internacionales, que en esos días consistían en la telegrafía. Aun en ese entonces era evidente que si la mitad de los países utilizaban código Morse y la otra mitad utilizaban algún otro código, iba a haber problemas. Cuando el teléfono entró a dar servicio internacional, la ITU también se hizo cargo de la tarea de estandarizar la telefonía. En 1947 la ITU se convirtió en una agencia de las Naciones Unidas.

La ITU tiene cerca de 200 miembros gubernamentales, incluyendo casi todos los miembros de las Naciones Unidas. Como Estados Unidos no cuenta con una PTT, alguien más tuvo que representar a este país en la ITU. Esta tarea repercutió en el Departamento de Estado, probablemente con la justificación de que la ITU tenía que lidiar con países extranjeros, lo cual era la especialidad de este departamento. La ITU cuenta también con más de 700 miembros de sectores y asociados. Entre ellos se incluyen las compañías telefónicas (como AT&T, Vodafone, Sprint), los fabricantes de equipo de telecomunicaciones (como Cisco, Nokia, Nortel), los distribuidores de computadoras (como Microsoft, Agilent, Toshiba), los fabricantes de chips (como Intel, Motorola, TI) y demás compañías interesadas (como Boeing, CBS, VeriSign).

La ITU tiene tres sectores principales. Nos enfocaremos principalmente en **ITU-T**, Sector de estandarización de telecomunicaciones, que se encarga de los sistemas de telefonía y comunicaciones de datos. Antes de 1993 a este sector se le llamaba **CCITT**, siglas de su nombre en francés, Comité Consultatif International Telégraphique et Teléphonique. **ITU-R**, sector de radiocomunicaciones, se encarga de coordinar el uso de las radiofrecuencias a nivel mundial por parte de los grupos de interés competidores. El otro sector es ITU-D, sector de desarrollo que promueve el desarrollo de las tecnologías de información y comunicación para estrechar la "división digital" entre los países con acceso efectivo a las tecnologías de información y los países con acceso limitado.

La tarea del sector ITU-T es hacer recomendaciones técnicas sobre las interfaces de telefonía, telegrafía y comunicación de datos. A menudo estas recomendaciones se convierten en estándares con reconocimiento internacional, aunque técnicamente las recomendaciones son sólo sugerencias que los gobiernos pueden adoptar o ignorar según lo deseen (porque los gobiernos son como niños de 13 años; no les gusta recibir órdenes). En la práctica, un país que desee adoptar un estándar de telefonía distinto al utilizado por el resto del mundo tiene la libertad de hacerlo, pero es a costa de quedar aislado de todos los demás. Esto podría funcionar para Corea del Norte, pero en cualquier otra parte sería un verdadero problema.

El verdadero trabajo del sector ITU-T se lleva a cabo en sus **Grupos de estudio** (*Study Groups*, o **SG**). En la actualidad hay 10 grupos de estudio de hasta 400 personas cada uno, en donde se tratan temas que varían desde la facturación telefónica y los servicios multimedia hasta la seguridad. Por ejemplo, el SG 15 estandariza las tecnologías DSL que son muy populares para conectarse a Internet. Para que sea posible realizar su trabajo, los grupos de estudio se dividen en **Equipos de trabajo** (*Working Parties*), que a su vez se dividen en **Equipos de expertos** (*Expert Teams*), los que a su vez se dividen en grupos ad hoc. La burocracia siempre será burocracia.

A pesar de todo esto, el sector ITU-T realmente hace su trabajo. Desde su creación ha producido más de 3 000 recomendaciones, muchas de las cuales son de uso popular en la práctica. Por ejemplo, la recomendación H.264 (que también es un estándar de ISO conocido como MPEG-4 AVC) es muy utilizada para la compresión de video, y los certificados de claves públicas X.509 se utilizan para la navegación web segura y el correo con firma digital.

A medida que el campo de las telecomunicaciones completa la transición iniciada en la década de 1980 para dejar de ser totalmente nacional y pasar a ser totalmente global, los estándares serán cada vez más importantes y cada vez más organizaciones querrán involucrarse en el proceso de establecer estos estándares. Para obtener más información sobre la ITU, consulte a Irmer (1994).

1.6.2 Quién es quién en el mundo de los estándares internacionales

Los estándares internacionales son producidos por la **ISO** (**Organización Internacional de Estándares**, del inglés *International Standards Organization*†), una organización voluntaria no surgida de un tratado y fundada en 1946. Sus miembros son las organizaciones nacionales de estándares de los 157 países miembros. Entre estos miembros están ANSI (Estados Unidos), BSI (Inglaterra), AFNOR (Francia), DIN (Alemania) y otras 153 organizaciones más.

La ISO emite estándares sobre una gran variedad de temas, que varían desde tuercas y pernos (literalmente) hasta los recubrimientos de los postes telefónicos [sin mencionar los granos de cacao (ISO 2451), las redes de pescar (ISO 1530), la ropa interior femenina (ISO 4416) y muchos otros temas más que no parecieran estar sujetos a la estandarización]. En cuestiones de estándares de telecomunicaciones, la ISO y el ITU-T cooperan con frecuencia (ISO es miembro del ITU-T) para evitar la ironía de dos estándares internacionales oficiales y mutuamente incompatibles.

Se han emitido más de 17 000 estándares, incluyendo los estándares OSI. La ISO tiene más de 200 Comités Técnicos (TC) enumerados en el orden de su creación, cada uno trata un tema específico. El TC1 trata con las tuercas y tornillos (la estandarización de los pasos de rosca de los tornillos). El JTC1 trata con la tecnología de información, incluyendo las redes, las computadoras y el software. Es el primer (y hasta ahora el único) Comité Técnico unido, el cual se creó en 1987 al fusionar el TC97 con las actividades en el IEC, otro organismo de estandarización. Cada TC tiene subcomités (SC), los que a su vez se dividen en grupos de trabajo (WG).

† Para los puristas, el verdadero nombre de ISO es Organización Internacional para la Estandarización.

El verdadero trabajo se hace en gran parte en los WG a través de los más de 100 000 voluntarios en todo el mundo. Muchos de estos "voluntarios" se asignan para trabajar en cuestiones de la ISO por sus patrones, cuyos productos se están estandarizando. Otros voluntarios son funcionarios de gobierno interesados en que la forma en que se hacen las cosas en su país llegue a ser el estándar internacional. También participan expertos académicos en muchos de los WG.

El procedimiento que utiliza la ISO para adoptar estándares se ha diseñado para lograr un consenso tan amplio como sea posible. El proceso empieza cuando una de las organizaciones nacionales de estándares siente la necesidad de un estándar internacional en cierta área. Después se forma un grupo de trabajo para proponer un **CD** (**Borrador de Comité**, del inglés *Committee Draft*). Después se circula el CD a todos los miembros, quienes tienen seis meses para criticarlo. Si una mayoría considerable lo aprueba, se produce un documento revisado llamado **DIS** (**Borrador de Estándar Internacional**, del inglés *Draft International Standard*), y se circula para que los miembros comenten y voten. Con base en los resultados de esta ronda, se prepara, aprueba y publica el texto final del **IS** (**Estándar Internacional**, del inglés *International Standard*). En áreas de mucha controversia, tal vez un CD o DIS tenga que pasar por varias versiones antes de adquirir suficientes votos, y el proceso completo puede tardar años.

El **NIST** (**Instituto Nacional de Estándares y Tecnología**, del inglés *National Institute of Standards and Technology*) forma parte del Departamento de Comercio. Solía llamarse Oficina Nacional de Estándares. Este organismo emite estándares obligatorios para las compras hechas por el gobierno de Estados Unidos, excepto las que realiza el Departamento de Defensa, el cual define sus propios estándares.

Otro protagonista importante en el mundo de los estándares es el **IEEE** (**Instituto de Ingenieros Eléctricos y Electrónicos**, del inglés *Institute of Electrical and Electronics Engineers*), la organización profesional más grande del mundo. Además de publicar muchas revistas y organizar numerosas conferencias cada año, el IEEE tiene un grupo de estandarización que desarrolla parámetros en el área de la ingeniería eléctrica y la computación. El comité 802 del IEEE ha estandarizado muchos tipos de redes LAN. Más adelante en el libro estudiaremos algunos de sus logros. El verdadero trabajo se realiza a través de una colección de grupos de trabajo, los cuales se muestran en la figura 1-38. El índice de éxito de los diversos grupos de trabajo del comité 802 ha sido bajo; tener un número 802.x no es garantía de éxito. Aun así, el impacto de las historias exitosas (en especial 802.3 y 802.11) en la industria y el mundo ha sido enorme.

1.6.3 Quién es quién en el mundo de estándares de Internet

El amplio mundo de Internet tiene sus propios mecanismos de estandarización, muy distintos a los de ITU-T e ISO. Para resumir en forma burda la diferencia, podemos decir que las personas que van a las reuniones de estandarización de la ITU o la ISO usan trajes, mientras que las personas que van a las reuniones de estandarización de Internet usan jeans (excepto cuando se reúnen en San Diego, en donde usan pantalones cortos y camisetas).

Las reuniones de la ITU-T y la ISO están pobladas de oficiales corporativos y burócratas para quienes la estandarización es su trabajo. Consideran la estandarización como algo positivo y dedican sus vidas a ella. Por otra parte, las personas de Internet prefieren la anarquía como cuestión de principios. Sin embargo, con cientos de millones de personas, cada una se ocupa de sus propios asuntos, no puede haber mucha comunicación. Por ende, algunas veces se necesitan los estándares por más lamentables que sean. En este contexto, una vez David Clark, del MIT, hizo un, ahora famoso, comentario acerca de que la estandarización de Internet consistía en "consenso aproximado y código en ejecución".

Cuando se inició ARPANET, el DoD creó un comité informal para supervisarla. En 1983 el comité cambió su nombre a **IAB** (**Consejo de Actividades de Internet**, del inglés *Internet Activities Board*) y recibió una misión un poco más amplia: mantener a los investigadores involucrados con ARPANET e

Número	Tema
802.1	Generalidades y arquitectura de redes LAN.
802.2 ↓	Control de enlaces lógicos.
802.3 *	Ethernet.
802.4 ↓	Token bus (se utilizó brevemente en las plantas de producción).
802.5	Token ring (la aportación de IBM al mundo de las redes LAN).
802.6 ↓	Bus doble de cola distribuida (la primera red de área metropolitana).
802.7 ↓	Grupo asesor técnico sobre tecnologías de banda ancha.
802.8 †	Grupo asesor técnico sobre tecnologías de fibra óptica.
802.9 ↓	Redes LAN isocrónicas (para aplicaciones en tiempo real).
802.10 ↓	Redes LAN virtuales y seguridad.
802.11 *	Redes LAN inalámbricas (WiFi).
802.12 ↓	Prioridad de demanda (AnyLAN, de Hewlett-Packard).
802.13	Número de mala suerte; nadie lo quiso.
802.14 ↓	Módems de cable (extinto: un consorcio industrial llegó primero).
802.15 *	Redes de área personal (Bluetooth, Zigbee).
802.16 *	Banda ancha inalámbrica (WIMAX).
802.17	Anillo de paquete elástico.
802.18	Grupo asesor técnico sobre cuestiones regulatorias de radio.
802.19	Grupo asesor técnico sobre la coexistencia de todos estos estándares.
802.20	Banda ancha móvil inalámbrica (similar a 802.16e).
802.21	Entrega independiente de los medios (para recorrer las tecnologías).
802.22	Red de área regional inalámbrica.

Figura 1-38. Los grupos de trabajo 802. Los importantes están marcados con *. Los que están marcados con ↓ están en hibernación. El que está marcado con † se dio por vencido y se deshizo.

Internet apuntando más o menos en la misma dirección, una actividad parecida a controlar una manada de gatos. El significado de las siglas "IAB" se cambió más adelante a **Consejo de Arquitectura de Internet**.

Cada uno de los aproximadamente 10 miembros del IAB encabezó una fuerza de trabajo sobre algún aspecto de importancia. El IAB se reunió varias veces al año para comentar sobre los resultados y brindar retroalimentación al DoD y la NSF, quienes proporcionaban la mayor parte de los fondos en esa época. Cuando se necesitaba un estándar (por ejemplo, un nuevo algoritmo de enrutamiento), los miembros del IAB lo discutían y después anunciaban el cambio de manera que los estudiantes de licenciatura, quienes eran el corazón del esfuerzo de software, pudieran implementarlo. La comunicación se llevaba a cabo mediante una serie de informes técnicos llamados **RFC** (**Petición de Comentarios**, del inglés *Request For Comments*). Los RFC se guardan en línea y cualquiera que se interese en ellos puede obtenerlos en *www.ietf.org/rfc*. Se enumeran en orden cronológico de creación. En la actualidad existen más de 5 000. Nos referiremos a muchos RFC en este libro.

Para 1989 Internet había crecido tanto que este estilo altamente informal ya no era funcional. Para entonces muchos distribuidores ofrecían productos TCP/IP y no querían cambiarlos sólo porque los investigadores habían tenido una mejor idea. En el verano de 1989, el IAB se volvió a organizar. Los investigadores pasaron a la **IRTF** (**Fuerza de Trabajo de Investigación de Internet**, del inglés *Internet Research Task Force*), la cual se hizo subsidiaria del IAB, junto con la **IETF** (**Fuerza de Trabajo de Ingeniería de Internet**, del inglés *Internet Engineering Task Force*). El IAB se repobló con gente que representaba un rango más amplio de organizaciones, no sólo la comunidad de investigación. En un principio fue un grupo que se perpetuaba a sí mismo, pues sus miembros servían por un término de dos años y los nuevos miembros eran designados por los antiguos. Más tarde se creó la **Sociedad de Internet** (*Internet Society*), formada por gente interesada en Internet. Así, podemos comparar en cierto sentido a la Sociedad de Internet con la ACM o el IEEE, ya que está gobernada por administradores elegidos, quienes designan a los miembros de la IAB.

El objetivo de esta división era hacer que la IRTF se concentrara en investigaciones a largo plazo, mientras que la IETF se encargaba de los problemas de ingeniería a corto plazo. La IETF se dividió en grupos de trabajo, cada uno con un problema específico por resolver. En un principio los presidentes de estos grupos de trabajo se reunieron como un comité de conducción para dirigir los trabajos de ingeniería. Los temas del grupo de trabajo incluyen nuevas aplicaciones, información de usuarios, integración de OSI, enrutamiento y direccionamiento, seguridad, administración de redes y estándares. En un momento dado se llegaron a formar tantos grupos de trabajo (más de 70) que se agruparon en áreas, en donde presidentes de cada una se reunía como el comité de conducción.

Además se adoptó un proceso de estandarización más formal con base en los patrones de la ISO. Para convertirse en una **Propuesta de estándar**, la idea básica se debe explicar en un RFC y debe generar suficiente interés en la comunidad para justificar su consideración. Para avanzar a la etapa de **Borrador de estándar**, una implementación funcional se debe probar rigorosamente por al menos dos sitios independientes durante cuatro meses como mínimo. Si el IAB se convence de que la idea es buena y el software funciona, puede declarar que el RFC es un **Estándar de Internet**. Algunos estándares de Internet se han convertido en estándares del DoD (MIL-STD), los cuales son obligatorios para los proveedores del DoD.

En cuanto a los estándares de la web, el **Consorcio World Wide Web** (*W3C*) desarrolla protocolos y lineamientos para facilitar el crecimiento a largo plazo de la web. Es un consorcio industrial encabezado por Tim Berners-Lee que se estableció en 1994, cuando la web realmente había empezado a despegar. Ahora el W3C tiene más de 300 miembros de todo el mundo y ha producido más de 100 Recomendaciones W3C, como se les dice a sus estándares, que tratan sobre temas tales como HTML y la privacidad en la web.

1.7 UNIDADES MÉTRICAS

Para evitar cualquier confusión, vale la pena indicar de manera explícita que en este libro, al igual que en la ciencia computacional en general, se utilizan medidas métricas en vez de unidades inglesas tradicionales (el sistema *furlong-stone-fortnight*). En la figura 1-39 se muestran los principales prefijos métricos. Por lo general se abrevian con base en sus primeras letras, y las unidades mayores a 1 se escriben en mayúsculas (KB, MB, etc.). Una excepción (por razones históricas) es kbps para kilobits/segundo. Así, una línea de comunicación de 1 Mbps transmite 10^6 bits/segundo y un reloj de 100 pseg (o 100 ps) genera un tic cada 10^{-10} segundos. Como mili y micro empiezan con la letra "m", hubo que tomar una decisión. Por lo general, "m" se utiliza para mili y "μ" (la letra griega mu) para micro.