1

INTRODUCCIÓN

Cada uno de los tres últimos siglos fue dominado por una tecnología. El siglo XVIII fue la era de los grandes sistemas mecánicos que acompañaron la Revolución Industrial. El siglo XIX fue la edad de la máquina de vapor. Durante el siglo XX la tecnología clave fue la obtención, el procesamiento y la distribución de la información. Entre otros acontecimientos, vimos la instalación de redes mundiales de telefonía, la invención de la radio y la televisión, el nacimiento y crecimiento sin precedentes de la industria de la computación, así como el lanzamiento de satélites de comunicaciones.

Como resultado del rápido progreso tecnológico, estas áreas están convergiendo de una manera acelerada y las diferencias entre la recolección, transportación, almacenamiento y procesamiento de la información están desapareciendo rápidamente. Organizaciones con cientos de oficinas dispersas en una amplia área geográfica esperan de manera rutinaria poder examinar el estado actual incluso de su sucursal más distante con sólo oprimir un botón. Al aumentar nuestra capacidad de obtener, procesar y distribuir información, la demanda de procesamiento de información cada vez más complejo crece incluso con más celeridad.

Aunque la industria de la computación aún es joven en comparación con otras industrias (como la automotriz y la aeronáutica), ha progresado espectacularmente en poco tiempo. Durante las dos primeras décadas de su existencia, los sistemas de computación estaban altamente centralizados, por lo general, en una sala grande e independiente. Con frecuencia, estas salas tenían paredes de cristal a través de las cuales los visitantes podían atisbar la maravilla electrónica que encerraban. Las compañías o universidades medianas apenas llegaban a tener una o dos computadoras, en tanto que las

instituciones grandes tenían, cuando mucho, una docena. La idea de que en veinte años se pudieran producir en masa millones de computadoras igualmente poderosas pero más pequeñas que un timbre postal era ciencia-ficción.

La fusión de las computadoras y las comunicaciones ha tenido una influencia profunda en la manera en que están organizados los sistemas computacionales. Actualmente, el concepto de "centro de cómputo" como un espacio amplio con una computadora grande a la que los usuarios llevaban su trabajo a procesar es totalmente obsoleto. El modelo antiguo de una sola computadora que realiza todas las tareas computacionales de una empresa ha sido reemplazado por otro en el que un gran número de computadoras separadas pero interconectadas hacen el trabajo. Estos sistemas se denominan **redes de computadoras.** El diseño y la organización de estas redes es el objetivo de este libro.

A lo largo del libro utilizaremos el término "redes de computadoras" para designar un conjunto de computadoras autónomas interconectadas. Se dice que dos computadoras están interconectadas si pueden intercambiar información. No es necesario que la conexión se realice mediante un cable de cobre; también se pueden utilizar las fibras ópticas, las microondas, los rayos infrarrojos y los satélites de comunicaciones. Las redes tienen varios tamaños, formas y figuras, como veremos más adelante. Aunque a algunas personas les parezca extraño, ni Internet ni Web son una red de computadoras. Este concepto quedará claro al finalizar el libro. La respuesta rápida es: Internet no es una red única, sino una red de redes, y Web es un sistema distribuido que se ejecuta sobre Internet.

Existe una gran confusión entre una red de computadoras y un **sistema distribuido.** La diferencia principal radica en que, en un sistema distribuido, un conjunto de computadoras independientes aparece ante sus usuarios como un sistema consistente y único. Por lo general, tiene un modelo o paradigma único que se presenta a los usuarios. Con frecuencia, una capa de software que se ejecuta sobre el sistema operativo, denominada **middleware,** es la responsable de implementar este modelo. Un ejemplo bien conocido de un sistema distribuido es **World Wide Web,** en la cual todo se ve como un documento (una página Web).

En una red de computadoras no existe esta consistencia, modelo ni software. Los usuarios están expuestos a las máquinas reales, y el sistema no hace ningún intento porque las máquinas se vean y actúen de manera similar. Si las máquinas tienen hardware diferente y distintos sistemas operativos, eso es completamente transparente para los usuarios. Si un usuario desea ejecutar un programa de una máquina remota, debe registrarse en ella y ejecutarlo desde ahí.

De hecho, un sistema distribuido es un sistema de software construido sobre una red. El software le da un alto grado de consistencia y transparencia. De este modo, la diferencia entre una red y un sistema distribuido está en el software (sobre todo en el sistema operativo), más que en el hardware.

No obstante, tienen muchas cosas en común. Por ejemplo, tanto los sistemas distribuidos como las redes de computadoras necesitan mover archivos. La diferencia está en quién invoca el movimiento, el sistema o el usuario. Aunque el objetivo principal de este libro son las redes, muchos de los temas se relacionan con los sistemas distribuidos. Para más información acerca de los sistemas distribuidos, vea (Tanenbaum y Van Steen, 2002).

1.1 USOS DE LAS REDES DE COMPUTADORAS

Antes de empezar a examinar con detalle los elementos técnicos, vale la pena dedicar algo de tiempo a precisar por qué la gente se interesa en las redes de computadoras y para qué se pueden utilizar. Después de todo, si nadie se hubiera interesado en ellas, no se habrían construido tantas. Empezaremos con el uso tradicional que les dan las empresas y los individuos, y luego avanzaremos a los últimos desarrollos con respecto a los usuarios móviles y la conexión de redes domésticas.

1.1.1 Aplicaciones de negocios

Muchas compañías tienen una cantidad considerable de computadoras. Por ejemplo, una compañía podría tener computadoras separadas para supervisar la producción, controlar inventarios y hacer la nómina. Al principio estas computadoras tal vez hayan trabajado por separado pero, en algún momento, la administración decidió conectarlas para extraer y correlacionar información acerca de toda la compañía.

Dicho de una manera más general, el asunto aquí es la **compartición de recursos** y el objetivo es hacer que todos los programas, el equipo y, en particular, los datos estén disponibles para todos los que se conecten a la red, independientemente de la ubicación física del recurso y del usuario. Un ejemplo claro y muy difundido es el de un grupo de oficinistas que comparten una impresora. Ninguno de los individuos necesita una impresora privada, y una impresora de alto volumen en red suele ser más barata, rápida y fácil de mantener que varias impresoras individuales.

Sin embargo, compartir información es tal vez más importante que compartir recursos físicos, como impresoras, escáneres y quemadores de CDs. Para las compañías grandes y medianas, así como para muchas pequeñas, la información computarizada es vital. La mayoría de las compañías tiene en línea registros de clientes, inventarios, cuentas por cobrar, estados financieros, información de impuestos, etcétera. Si todas las computadoras de un banco se cayeran, éste no duraría más de cinco minutos. Una moderna planta manufacturera, con una línea de ensamblado controlada por computadora, ni siquiera duraría ese tiempo. Incluso una pequeña agencia de viajes o un despacho jurídico de tres personas, ahora dependen en gran medida de las redes de computadoras para que sus empleados puedan tener acceso de manera instantánea a la información y a los documentos importantes.

En las compañías más pequeñas, es posible que todas las computadoras estén en una sola oficina o en un solo edificio, pero en las más grandes, las computadoras y los empleados pueden estar dispersos en docenas de oficinas y plantas en varios países. No obstante, un vendedor en Nueva York podría requerir algunas veces tener acceso a una base de datos de inventario de productos que se encuentra en Singapur. En otras palabras, el hecho de que un usuario esté a 15,000 km de sus datos no debe ser impedimento para que utilice esos datos como si fueran locales. Esta meta se podría resumir diciendo que es un intento por acabar con la “tiranía de la geografía”.

En términos aún más sencillos, es posible imaginar el sistema de información de una compañía como si consistiera en una o más bases de datos y algunos empleados que necesitan acceder a

ellas de manera remota. En este modelo, los datos están almacenados en computadoras poderosas que se llaman **servidores.** Con frecuencia, éstos se encuentran alojados en una central y un administrador de sistemas les da mantenimiento. En contraste, los empleados tienen en sus escritorios máquinas más sencillas, llamadas **clientes,** con las que pueden acceder a datos remotos —por ejemplo, para incluirlos en las hojas de cálculo que están elaborando. (Algunas veces nos referiremos a los usuarios de las máquinas como "el cliente", pero debe quedar claro, por el contexto, si el término se refiere a la computadora o a su usuario.) Las máquinas cliente y servidor están conectadas por una red, como se ilustra en la figura 1-1. Observe que hemos representado a la red como un óvalo sencillo, sin detalle alguno. Utilizaremos esta forma cuando nos refiramos a una red en sentido general. Cuando se requieran más detalles, los proporcionaremos.

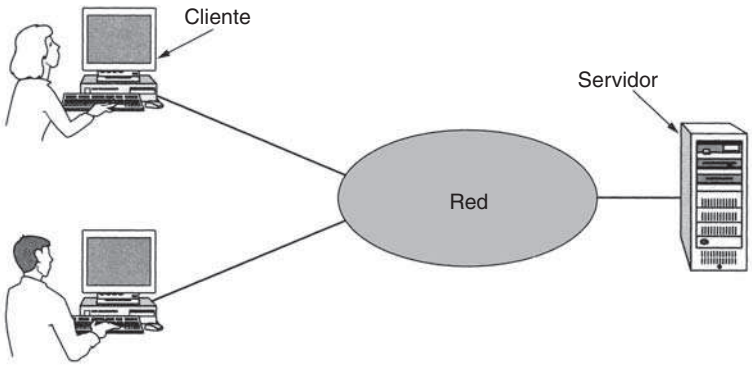

Figura 1-1. Una red con dos clientes y un servidor.

Este conjunto se conoce como **modelo cliente-servidor.** Se utiliza ampliamente y forma la base en gran medida del uso de redes. Es aplicable cuando el cliente y el servidor están en el mismo edificio (por ejemplo, cuando pertenecen a la misma compañía), pero también cuando están bastante retirados. Por ejemplo, cuando una persona en casa accede a una página Web, se emplea el mismo modelo, en el que el servidor remoto de Web es el servidor y la computadora personal del usuario es el cliente. En la mayoría de los casos, un servidor puede manejar una gran cantidad de clientes.

Si vemos el modelo cliente-servidor en detalle, nos daremos cuenta de que hay dos procesos involucrados, uno en la máquina cliente y otro en la máquina servidor. La comunicación toma la siguiente forma: el proceso cliente envía una solicitud a través de la red al proceso servidor y espera una respuesta. Cuando el proceso servidor recibe la solicitud, realiza el trabajo que se le pide o busca los datos solicitados y devuelve una respuesta. Estos mensajes se muestran en la figura 1-2.

Un segundo objetivo de la configuración de una red de computadoras tiene que ver más con la gente que con la información e, incluso, con las computadoras mismas. Una red de computadoras

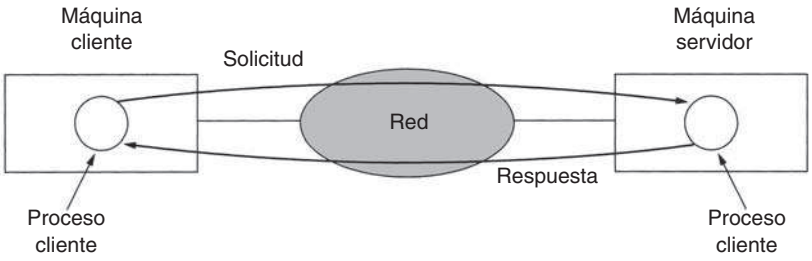

Figura 1-2. El modelo cliente-servidor implica solicitudes y respuestas.

es un poderoso **medio de comunicación** entre los empleados. Casi todas las compañías que tienen dos o más computadoras cuentan con **correo electrónico,** mediante el cual los empleados mantienen generalmente una comunicación diaria. De hecho, una queja común es la gran cantidad de correo electrónico que tenemos que atender, mucho de él sin sentido porque los jefes han descubierto que pueden enviar el mismo mensaje (a menudo sin contenido) a todos sus subordinados con sólo oprimir un botón.

Pero el correo electrónico no es la única forma de comunicación mejorada que las redes de computadoras hacen posible. Con una red es fácil que dos o más personas que trabajan a distancia escriban en conjunto un informe. Si un empleado hace un cambio a un documento en línea, los demás pueden ver el cambio de inmediato, en vez de esperar una carta durante varios días. Esta agilización facilita la cooperación entre grupos de personas que no se encuentran en el mismo lugar, lo cual antes había sido imposible.

Otra forma de comunicación asistida por computadora es la videoconferencia. Con esta tecnología, los empleados en ubicaciones distantes pueden tener una reunión, viéndose y escuchándose unos a otros e incluso escribiendo en una pizarra virtual compartida. La videoconferencia es una herramienta poderosa para eliminar el costo y el tiempo que anteriormente se empleaba en viajar. A veces se dice que la comunicación y el transporte están en competencia, y que el que gane hará obsoleto al otro.

Una tercera meta para cada vez más compañías es hacer negocios de manera electrónica con otras compañías, sobre todo proveedores y clientes. Por ejemplo, los fabricantes de automóviles, de aviones, de computadoras, etcétera, compran subsistemas de diversos proveedores y luego ensamblan las partes. Mediante las redes de computadoras los fabricantes pueden hacer pedidos electrónicamente conforme se requieran. Tener la capacidad de hacer pedidos en tiempo real (es decir, conforme se requieren) reduce la necesidad de tener grandes inventarios y mejora la eficiencia.

Una cuarta meta que se está volviendo más importante es la de hacer negocios con consumidores a través de Internet. Las líneas aéreas, las librerías y los vendedores de música han descubierto que muchos consumidores prefieren realizar sus compras desde casa. Por consiguiente, muchas compañías proporcionan en línea catálogos de sus productos y servicios y levantan pedidos de la misma manera. Se espera que este sector crezca rápidamente en el futuro. Es lo que se conoce como **comercio electrónico.**

1.1.2 Aplicaciones domésticas

En 1977 Ken Olsen era presidente de Digital Equipment Corporation, que en esa época era el segundo proveedor de computadoras en el mundo (después de IBM). Cuando se le preguntó por qué Digital no perseguía el mercado de las computadoras personales en gran volumen, contestó: "No hay razón alguna para que un individuo tenga una computadora en su casa". La historia demostró lo contrario y Digital ya no existe. ¿Por qué la gente compra computadoras para uso doméstico? En principio, para procesamiento de texto y juegos, pero en los últimos años esto ha cambiado radicalmente. Tal vez la razón más importante ahora sea por el acceso a Internet. Algunos de los usos más comunes de Internet por parte de usuarios domésticos son los siguientes:

1. Acceso a información remota.
2. Comunicación de persona a persona.
3. Entretenimiento interactivo.
4. Comercio electrónico.

El acceso a la información remota se puede realizar por diversas razones. Puede ser que navegue por World Wide Web para obtener información o sólo por diversión. La información disponible incluye artes, negocios, cocina, gobiernos, salud, historia, pasatiempos, recreación, ciencia, deportes, viajes y muchas otras cosas más. La diversión viene en demasiadas formas como para mencionarlas, más algunas otras que es mejor no mencionar.

Muchos periódicos ahora están disponibles en línea y pueden personalizarse. Por ejemplo, en algunos casos le puede indicar a un periódico que desea toda la información acerca de políticos corruptos, incendios, escándalos que involucran a las celebridades y epidemias, pero nada sobre fútbol. Incluso puede hacer que los artículos que usted desea se descarguen en su disco duro o se impriman mientras usted duerme, para que cuando se levante a desayunar los tenga disponibles. Mientras continúe esta tendencia, se provocará el desempleo masivo de los niños de 12 años que entregan los diarios, pero los periódicos lo quieren así porque la distribución siempre ha sido el punto débil en la gran cadena de producción.

El tema más importante después de los periódicos (además de las revistas y periódicos científicos) son las bibliotecas digitales en línea. Muchas organizaciones profesionales, como la ACM (*www.acm.org*) y la Sociedad de Computación del IEEE (*www.computer.org*), ya cuentan con muchos periódicos y presentaciones de conferencias en línea. Otros grupos están creciendo de manera rápida. Dependiendo del costo, tamaño y peso de las computadoras portátiles, los libros impresos podrían llegar a ser obsoletos. Los escépticos deben tomar en cuenta el efecto que la imprenta tuvo sobre los manuscritos ilustrados del medioevo.

Todas las aplicaciones anteriores implican las interacciones entre una persona y una base de datos remota llena de información. La segunda gran categoría del uso de redes es la comunicación de persona a persona, básicamente la respuesta del siglo XXI al teléfono del siglo XIX. Millones de personas en todo el mundo utilizan a diario el correo electrónico y su uso está creciendo rápidamente. Ya es muy común que contenga audio y vídeo, así como texto y figuras. Los aromas podrían tardar un poco más.

Muchas personas utilizan los **mensajes instantáneos.** Esta característica, derivada del programa *talk* de UNIX, que se utiliza aproximadamente desde 1970, permite que las personas se escriban mensajes en tiempo real. Una versión, para varias personas, de esta idea es el **salón de conversación** (*chat room*), en el que un grupo de personas puede escribir mensajes para que todos los vean.

Los grupos de noticias mundiales, con debates sobre todo tema imaginable, ya son un lugar común entre un grupo selecto de personas y este fenómeno crecerá para abarcar a la población en general. Estos debates, en los que una persona envía un mensaje y todos los demás suscriptores del grupo de noticias lo pueden leer, van desde los humorísticos hasta los apasionados. A diferencia de los salones de conversación, los grupos de noticias no son en tiempo real y los mensajes se guardan para que cuando alguien vuelva de vacaciones, encuentre todos los mensajes que hayan sido enviados en el ínterin, esperando pacientemente a ser leídos.

Otro tipo de comunicación de persona a persona a menudo se conoce como comunicación de **igual a igual** (*peer to peer*), para distinguirla del modelo cliente-servidor (Parameswaran y cols., 2001). De esta forma, los individuos que forman un grupo esparcido se pueden comunicar con otros del grupo, como se muestra en la figura 1-3. Cada persona puede, en principio, comunicarse con una o más personas; no hay una división fija de clientes y servidores.

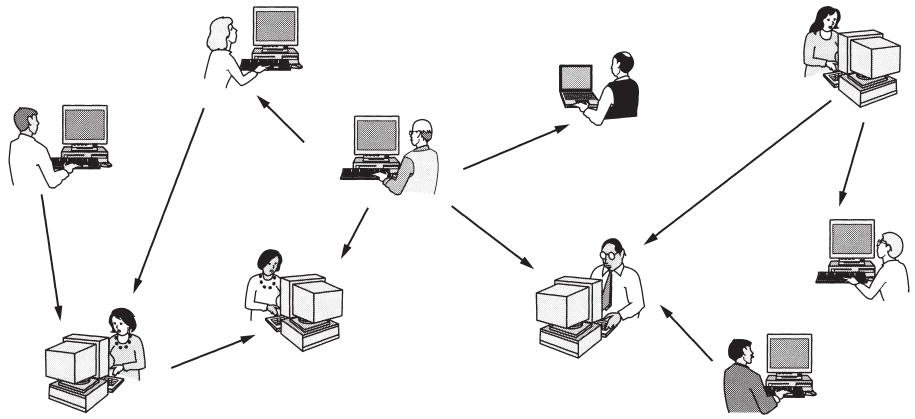

Figura 1-3. En el sistema de igual a igual no hay clientes ni servidores fijos.

La comunicación de igual a igual dominó la mayor parte del 2000 con un servicio llamado Napster, que en su mayor apogeo tenía más de 50 millones de personas canjeando música, lo que fue probablemente la mayor infracción a derechos de autor en toda la historia de la música grabada (Lam y Tan, 2001, y Macedonia, 2000). La idea era muy sencilla. Los miembros registraban en una base de datos central mantenida en el servidor de Napster la música que tenían en sus discos duros. Si un miembro deseaba una canción, verificaba la base de datos para ver quién la tenía e iba directamente ahí para obtenerla. Al no conservar realmente ninguna obra musical en las máquinas, Napster argumentaba que no estaba infringiendo los derechos de autor de nadie. Las cortes no estuvieron de acuerdo y lo clausuraron.

Sin embargo, la siguiente generación de sistemas de igual a igual elimina la base de datos central al hacer que cada usuario mantenga su propia base de datos de manera local, y al proporcionarle una lista de otras personas cercanas que también son miembros del sistema. De esta manera, un nuevo usuario puede ir a cualquiera de ellas para ver qué tiene y obtener una lista de otras más para indagar acerca de más música y más nombres. Este proceso de consulta se puede repetir de manera indefinida hasta construir una enorme base de datos local de lo que hay a disposición. Es una actividad que podría ser tediosa para las personas pero que para las computadoras es muy sencilla.

También existen las aplicaciones legales para la comunicación de igual a igual. Por ejemplo, un club de admiradores que comparte un dominio público de música o cintas de muestra que las nuevas bandas han cedido para efectos de publicidad, familias que comparten fotografías, películas e información genealógica y adolescentes que juegan en línea juegos para varias personas. De hecho, una de las aplicaciones de Internet más populares, el correo electrónico, es esencialmente de igual a igual. Se espera que esta forma de comunicación crezca con rapidez en el futuro.

Los delitos electrónicos no se limitan a la ley de derechos de autor. Otra área activa es la de los juegos electrónicos. Las computadoras han simulado cosas durante décadas. ¿Por qué no simular máquinas tragamonedas, ruedas de la fortuna, repartidores de blackjack y más equipo de juegos electrónicos? El problema es que los juegos electrónicos son legales en muchos lugares (Inglaterra, por ejemplo) y los propietarios de casinos han aprovechado el potencial de los juegos electrónicos por Internet. ¿Qué pasaría si el jugador y el casino estuvieran en países diferentes entre los cuales hay conflicto de leyes? Ésa es una buena pregunta.

Otras aplicaciones orientadas a la comunicación y de rápido crecimiento incluyen el uso de Internet para transportar llamadas telefónicas, el teléfono con vídeo y la radio por Internet. Otra aplicación es el teleaprendizaje, es decir, asistir a clases a las 8:00 A.M. sin el inconveniente de tener que levantarse antes de la cama. A largo plazo, el uso de las redes para mejorar la comunicación de persona a persona puede demostrar que ésta es el área más importante.

Nuestra tercera categoría es el entretenimiento, que es una industria grande y en crecimiento. La aplicación dominante (la que podría impulsar al resto) es el vídeo bajo demanda. De aquí a 10 años, podría seleccionar cualquier película o programa de televisión producido en cualquier país y proyectarlo en su pantalla al instante. Las películas nuevas podrían llegar a ser interactivas, en las que se pediría ocasionalmente al usuario que eligiera el rumbo de la narración, con escenarios alternativos preparados para todos los casos. La televisión en vivo también podría llegar a ser interactiva, permitiendo que la audiencia participe en programas de preguntas, elija entre los competidores, etcétera.

Por otra parte, tal vez el vídeo bajo demanda no sea la aplicación dominante. Podría ser la de los juegos. En la actualidad ya contamos con juegos de simulación de varias personas en tiempo real, como el de las escondidas en un calabozo virtual y simuladores de vuelo en los que los jugadores de un equipo tratan de derribar a los del equipo contrario. Si los juegos se juegan con anteojos y tiempo real tridimensional, con imágenes en movimiento de calidad fotográfica, tenemos un tipo de realidad virtual compartida a nivel mundial.

Nuestra cuarta categoría es el comercio electrónico en el más amplio sentido de la palabra. Comprar desde el hogar ya es una actividad común y permite que los usuarios inspeccionen los

catálogos en línea de miles de compañías. Algunos de estos catálogos proporcionarán pronto la capacidad de obtener un vídeo instantáneo de cualquier producto con sólo hacer clic en el nombre de éste. Si un cliente compra un producto por vía electrónica y no sabe cómo usarlo, podrá consultar el soporte técnico en línea.

Otra área en la que el comercio electrónico ya se está dando es en las instituciones financieras. Mucha gente ya efectúa sus pagos, administra sus cuentas bancarias y maneja sus inversiones de manera electrónica. Seguramente esto crecerá en cuanto las redes sean más seguras.

Un área que prácticamente nadie previó son los mercados de pulgas electrónicos. Las subastas en línea de artículos de segunda mano se han convertido en una industria masiva. A diferencia del comercio electrónico tradicional, que sigue el modelo cliente-servidor, las subastas en línea son más que un sistema de igual a igual, un tipo de sistema de consumidor a consumidor. Algunas de estas formas de comercio electrónico han adoptado una serie de etiquetas con base en que "to" y "2" (en inglés) suenan igual. La figura 1-4 presenta una lista de las abreviaturas más comunes.

Etiqueta	Nombre completo	Ejemplo
B2C	Negocio a consumidor	Pedido de libros en línea
B2B	Negocio a negocio	La fábrica de automóviles hace un pedido de llantas al proveedor
G2C	Gobierno a consumidor	El gobierno distribuye formas fiscales electrónicamente
C2C	Consumidor a consumidor	Subasta en línea de productos de segunda mano
P2P	Igual a igual	Compartición de archivos

Figura 1-4. Algunas formas de comercio electrónico.

Sin duda, el rango de usos de las redes de computadoras crecerá con rapidez y probablemente en formas que nadie puede prever ahora. Después de todo, ¿cuánta gente pudo predecir en 1990 que en diez años las personas podrían escribir mensajes breves en teléfonos celulares durante sus viajes en autobús, lo cual podría ser una forma muy ventajosa para que las compañías telefónicas ganaran dinero? Sin embargo, en la actualidad el servicio de mensajes breves es muy rentable.

Las redes de computadoras podrían llegar a ser sumamente importantes para la gente que no vive en las grandes ciudades, pues les da el mismo acceso a servicios que a las personas que sí viven en ellas. El teleaprendizaje podría afectar radicalmente la educación; las universidades podrían dar servicio a estudiantes nacionales o internacionales. La telemedicina está en inicio (por ejemplo, se utiliza para la supervisión remota de un paciente), pero puede llegar a ser muy importante. Sin embargo, la aplicación clave podría ser algo mundano, como utilizar una *webcam* (cámara conectada a Internet) en su refrigerador, para saber si tiene que comprar leche al regresar del trabajo.

1.1.3 Usuarios móviles

Las computadoras portátiles, como las *notebook* y los asistentes personales digitales (PDAs), son uno de los segmentos de crecimiento más rápido de la industria de la computación. Muchos propietarios de estas computadoras poseen máquinas de escritorio en la oficina y desean estar conectados a su base doméstica cuando están de viaje o fuera de casa. Puesto que no

es posible tener una conexión alámbrica en autos y aviones, hay un gran interés en las redes inalámbricas. En esta sección veremos brevemente algunos usos de ellas.

¿Por qué querría alguien una? Un argumento común es la oficina portátil. Con frecuencia, las personas que están de viaje desean utilizar sus equipos portátiles para enviar y recibir llamadas telefónicas, faxes y correo electrónico, navegar en Web, acceder a archivos remotos e iniciar sesión en máquinas remotas. Y desean hacer esto desde cualquier punto, ya sea por tierra, mar o aire. Por ejemplo, actualmente en las conferencias por computadora, los organizadores suelen configurar una red inalámbrica en el área de la conferencia. Cualquiera que tenga una computadora portátil y un módem inalámbrico puede conectarse a Internet, como si la computadora estuviera conectada a una red alámbrica (cableada). Del mismo modo, algunas universidades han instalado redes inalámbricas en sus campus para que los estudiantes se puedan sentar entre los árboles y consultar los archivos de la biblioteca o leer su correo electrónico.

Las redes inalámbricas son de gran utilidad para las flotas de camiones, taxis, vehículos de entrega y reparadores, para mantenerse en contacto con la casa. Por ejemplo, en muchas ciudades los taxistas trabajan por su cuenta, más que para una empresa de taxis. En algunas de estas ciudades, los taxis tienen una pantalla que el conductor puede ver. Cuando el cliente solicita un servicio, un despachador central escribe los puntos en los que el chofer deberá recoger y dejar al cliente. Esta información se despliega en las pantallas de los conductores y suena un timbre. El conductor que oprima primero un botón en la pantalla recibe la llamada.

Las redes inalámbricas también son importantes para la milicia. Si tiene que estar disponible en breve para pelear una guerra en cualquier parte de la Tierra, probablemente no sea bueno pensar en utilizar la infraestructura de conectividad de redes local. Lo mejor sería tener la propia.

Aunque la conectividad inalámbrica y la computación portátil se relacionan frecuentemente, no son idénticas, como se muestra en la figura 1-5, en la que vemos una diferencia entre **inalámbrica fija** e **inalámbrica móvil.** Incluso en ocasiones las computadoras portátiles son alámbricas. Por ejemplo, si un viajero conecta una portátil a una toma telefónica en su habitación del hotel, tiene movilidad sin una red inalámbrica.

Inalámbrica	Móvil	Aplicaciones
No	No	Computadoras de escritorio en oficinas
No	Sí	Una computadora portátil usada en un cuarto de hotel
Sí	No	Redes en construcciones antiguas sin cableado
Sí	Sí	Oficina portátil; PDA para inventario de almacén

Figura 1-5. Combinaciones de redes inalámbricas y computación móvil.

Por otra parte, algunas computadoras inalámbricas no son móviles. Un ejemplo representativo sería una compañía que posee un edificio antiguo que no tiene cableado de redes y que desea conectar sus computadoras. La instalación de una red inalámbrica podría requerir un poco más que comprar una caja pequeña con algunos aparatos electrónicos, desempacarlos y conectarlos. Sin embargo, esta solución podría ser mucho más barata que contratar trabajadores que coloquen ductos de cable para acondicionar el edificio.

Desde luego, también existen las aplicaciones inalámbricas móviles, que van desde la oficina portátil hasta las personas que pasean por una tienda con un PDA realizando un inventario. En muchos aeropuertos, los empleados de alquiler de coches trabajan en los estacionamientos con computadoras portátiles inalámbricas. Escriben el número de la placa de circulación de los autos alquilados, y su computadora portátil, que tiene una impresora integrada, llama a la computadora principal, obtiene la información del arrendamiento e imprime la factura en el acto.

Conforme se extienda la tecnología inalámbrica, es probable que surjan otras aplicaciones. Echemos un vistazo a algunas de las posibilidades. Los parquímetros inalámbricos tienen ventajas para los usuarios y las autoridades administrativas gubernamentales. Los medidores pueden aceptar tarjetas de crédito o de débito y verificarlas de manera instantánea a través del vínculo inalámbrico. Cuando un medidor expire, se podría verificar la presencia de un auto (emitiendo una señal) y reportar la expiración a la policía. Se ha estimado que con esta medida, los gobiernos de las ciudades de Estados Unidos podrían colectar $10 mil millones adicionales (Harte y cols., 2000). Además, la entrada en vigor del aparcamiento ayudaría al ambiente, debido a que los conductores que al saber que podrían ser detenidos al estacionarse de manera ilegal, utilizarían el transporte público.

Los expendedores automáticos de alimentos, bebidas, etcétera, se encuentran por todas partes. Sin embargo, los alimentos no entran en las máquinas por arte de magia. Periódicamente, alguien va con un camión y las llena. Si los expendedores automáticos emitieran informes periódicos una vez al día en los que indicaran sus inventarios actuales, el conductor del camión sabría qué máquinas necesitan servicio y qué cantidad de qué productos llevar. Esta información podría conducir a una mayor eficiencia en la planeación de las rutas. Desde luego que esta información también se podría enviar a través de un teléfono de línea común, pero proporcionar a cada expendedor automático una conexión fija telefónica para que realice una llamada al día es costoso debido a los cargos fijos mensuales.

Otra área en la que la tecnología inalámbrica podría ahorrar dinero es en la lectura de medidores de servicios públicos. Si los medidores de electricidad, gas, agua y otros servicios domésticos reportaran su uso a través de una red inalámbrica, no habría necesidad de enviar lectores de medidores. Del mismo modo, los detectores inalámbricos de humo podrían comunicarse con el departamento de bomberos en lugar de hacer tanto ruido (lo cual no sirve de nada si no hay nadie en casa). Conforme baje el costo de los dispositivos de radio y el tiempo aire, más y más medidas e informes se harán a través de redes inalámbricas.

Un área de aplicación totalmente diferente para las redes inalámbricas es la fusión esperada de teléfonos celulares y PDAs en computadoras inalámbricas diminutas. Un primer intento fue el de los diminutos PDAs que podían desplegar páginas Web reducidas al mínimo en sus pequeñas pantallas. Este sistema, llamado **WAP 1.0** (**Protocolo de Aplicaciones Inalámbricas**), falló en gran parte debido a sus pantallas microscópicas, bajo ancho de banda y servicio deficiente. Pero con WAP 2.0 serán mejores los dispositivos y servicios nuevos.

La fuerza que impulsa estos dispositivos es la llamada **comercio móvil** (*m-commerce*) (Senn, 2000). La fuerza que impulsa este fenómeno consiste en diversos fabricantes de PDAs inalámbricos y operadores de redes que luchan por descubrir cómo ganar una parte del pastel del comercio móvil. Una de sus esperanzas es utilizar los PDAs inalámbricos para servicios bancarios y de compras. Una idea es utilizar los PDAs inalámbricos como un tipo de cartera electrónica, que

autorice pagos en tiendas como un reemplazo del efectivo y las tarjetas de crédito. De este modo, el cargo aparecerá en la factura del teléfono celular. Desde el punto de vista de la tienda, este esquema le podría ahorrar la mayor parte de la cuota de la empresa de tarjetas de crédito, que puede ser un porcentaje importante. Desde luego, este plan puede resultar contraproducente, puesto que los clientes que están en una tienda podrían utilizar los PDAs para verificar los precios de la competencia antes de comprar. Peor aún, las compañías telefónicas podrían ofrecer PDAs con lectores de códigos de barras que permitan a un cliente rastrear un producto en una tienda y obtener en forma instantánea un informe detallado de dónde más se puede comprar y a qué precio.

Puesto que el operador de redes sabe dónde está el usuario, algunos servicios se hacen intencionalmente dependientes de la ubicación. Por ejemplo, se podría preguntar por una librería cercana o un restaurante chino. Los mapas móviles y los pronósticos meteorológicos muy locales ("¿Cuándo va a dejar de llover en mi traspatio?") son otros candidatos. Sin duda, aparecerán otras muchas aplicaciones en cuanto estos dispositivos se difundan más ampliamente.

Un punto muy importante para el comercio móvil es que los usuarios de teléfonos celulares están acostumbrados a pagar por todo (en contraste con los usuarios de Internet, que esperan recibir prácticamente todo sin costo). Si un sitio Web cobrara una cuota por permitir a sus clientes pagar con tarjeta de crédito, provocaría una reclamación muy ruidosa de los usuarios. Si un operador de telefonía celular permitiera que las personas pagaran artículos en una tienda utilizando el teléfono celular y luego cargara una cuota por este servicio, probablemente sus clientes lo aceptarían como algo normal. Sólo el tiempo lo dirá.

Un poco más lejanas están las redes de área personal y las microcomputadoras personales de bolsillo. IBM ha desarrollado un reloj que ejecuta Linux (el cual incluye el sistema de ventanas X11) y tiene conectividad inalámbrica a Internet para enviar y recibir correo electrónico (Narayanaswami y cols., 2002). En el futuro, las personas podrían intercambiar tarjetas de presentación con sólo exponer sus relojes entre sí. Las computadoras de bolsillo inalámbricas pueden dar acceso a las personas a sitios seguros de la misma manera en que lo hacen las tarjetas de banda magnética (posiblemente en combinación con un código de PIN o medición biométrica). Estos relojes también podrían recuperar información relativa a la ubicación actual del usuario (por ejemplo, restaurantes locales). Las posibilidades son infinitas.

Los relojes inteligentes con radio han sido parte de nuestro espacio mental desde que aparecieron en las tiras cómicas de Dick Tracy, en 1946. Pero, ¿polvo inteligente? Los investigadores en Berkeley han empaquetado una computadora inalámbrica en un cubo de 1 mm por lado (Warneke y cols., 2001). Entre las aplicaciones potenciales se incluyen el seguimiento de inventarios, paquetes e incluso pequeños pájaros, roedores e insectos.

1.1.4 Temas sociales

La amplia introducción de las redes ha presentado problemas sociales, éticos y políticos. Mencionemos brevemente algunos de ellos; un estudio completo requeriría todo un libro, por lo menos. Un rasgo popular de muchas redes son los grupos de noticias o boletines electrónicos mediante los cuales las personas pueden intercambiar mensajes con individuos de los mismos intereses. Siempre y cuando los asuntos se restrinjan a temas técnicos o pasatiempos como la jardinería, no surgirán demasiados problemas.

El problema viene cuando los grupos de noticias se enfocan en temas que las personas en realidad tocan con cuidado, como política, religión o sexo. Los puntos de vista enviados a tales grupos podrían ser ofensivos para algunas personas. Peor aún, podrían no ser políticamente correctos. Además, los mensajes no tienen que limitarse a texto. En la actualidad se pueden enviar fotografías en alta resolución e incluso pequeños videoclips a través de redes de computadoras. Algunas personas practican la filosofía de vive y deja vivir, pero otras sienten que enviar cierto material (por ejemplo, ataques a países o religiones en particular, pornografía, etcétera) es sencillamente inaceptable y debe ser censurado. Los diversos países tienen diferentes y conflictivas leyes al respecto. De esta manera, el debate se aviva.

Las personas han demandado a los operadores de redes, afirmando que son responsables, como sucede en el caso de los periódicos y las revistas, del contenido que transmiten. La respuesta inevitable es que una red es como una compañía de teléfonos o la oficina de correos, por lo que no se puede esperar que vigilen lo que dicen los usuarios. Más aún, si los operadores de redes censuraran los mensajes, borrarían cualquier contenido que contuviera incluso la mínima posibilidad de que se les demandara, pero con esto violarían los derechos de sus usuarios a la libre expresión. Probablemente lo más seguro sería decir que este debate seguirá durante algún tiempo.

Otra área divertida es la de los derechos de los empleados en comparación con los de los empleadores. Muchas personas leen y escriben correo electrónico en el trabajo. Muchos empleadores han exigido el derecho a leer y, posiblemente, censurar los mensajes de los empleados, incluso los enviados desde un equipo doméstico después de las horas de trabajo. No todos los empleados están de acuerdo con esto.

Incluso si los empleadores tienen poder sobre los empleados, ¿esta relación también rige a las universidades y los estudiantes? ¿Qué hay acerca de las escuelas secundarias y los estudiantes? En 1994, la Carnegie-Mellon University decidió suspender el flujo de mensajes entrantes de varios grupos de noticias que trataban sexo porque la universidad sintió que el material era inapropiado para menores (es decir, menores de 18 años). Tomó años recuperarse de este suceso.

Otro tema de importancia es el de los derechos del gobierno y los de los ciudadanos. El FBI ha instalado un sistema en muchos proveedores de servicios de Internet para curiosear entre todos los correos electrónicos en busca de fragmentos que le interesen (Blaze y Bellovin, 2000; Sobel, 2001; Zacks, 2001). El sistema se llamaba originalmente **Carnivore** pero la mala publicidad provocó que se cambiara el nombre por uno menos agresivo que sonara como DCS1000. Pero su objetivo sigue siendo el de espiar a millones de personas con la esperanza de encontrar información acerca de actividades ilegales. Por desgracia, la Cuarta Enmienda de la Constitución de Estados Unidos prohíbe que el gobierno realice investigaciones sin una orden de cateo. Decidir si estas palabras, escritas en el siglo XVIII, aún son válidas en el siglo XXI es un asunto que podría mantener ocupadas a las cortes hasta el siglo XXII.

El gobierno no tiene el monopolio de las amenazas contra la privacidad de una persona. El sector privado también hace su parte. Por ejemplo, los archivos pequeños llamados cookies que los navegadores Web almacenan en las computadoras de los usuarios permiten que las empresas rastreen las actividades de éstos en el ciberespacio, y podrían permitir que los números de tarjeta de crédito, del seguro social y otra información confidencial se divulguen por toda la Internet (Berghel, 2001).

Las redes de computadoras ofrecen la posibilidad de enviar mensajes anónimos. En algunas situaciones esta capacidad podría ser deseable. Por ejemplo, los estudiantes, soldados, empleados y ciudadanos pueden denunciar el comportamiento ilegal de algunos profesores, oficiales, superiores y políticos sin temor a represalias. Por otra parte, en Estados Unidos, y en la mayoría de las democracias, la ley otorga específicamente a una persona acusada el derecho de poder confrontar y desafiar a su acusador en la corte. Las acusaciones anónimas no se pueden usar como evidencia.

En resumen, las redes de computadoras, como la imprenta hace 500 años, permiten que el ciudadano común distribuya sus puntos de vista en diversos modos y a audiencias diferentes, lo cual antes no era posible. Este nuevo fondo de libertad ofrece consigo muchos temas sociales, políticos y morales sin resolver.

Junto con lo bueno viene lo malo. Así parece ser la vida. Internet hace posible encontrar con rapidez información, pero una gran cantidad de ella está mal documentada, es falsa o completamente errónea. El consejo médico que obtuvo en Internet podría haber venido de un ganador del Premio Nobel o de un desertor de la preparatoria. Las redes de computadoras también han introducido nuevos tipos de comportamientos antisociales y criminales. La publicidad no deseada (*spam*) se ha convertido en algo común debido a que algunas personas se dedican a reunir millones de direcciones de correo electrónico y las venden en CD-ROMs a comerciantes. Los mensajes por correo electrónico que contienen elementos activos (básicamente programas o macros que se ejecutan en la máquina del receptor) pueden contener virus potencialmente destructores.

El robo de identidad se ha convertido en un problema grave, ya que los ladrones ahora reúnen información sobre una persona para obtener tarjetas de crédito y otros documentos a nombre de ella. Por último, la capacidad de transmitir música y vídeo de manera digital ha abierto la puerta a violaciones masivas de derechos de autor, que son difíciles de detectar y castigar.

Muchos de estos problemas se podrían resolver si la industria de las computadoras tomara la seguridad de las computadoras con seriedad. Si todos los mensajes se codificaran y autenticaran, sería más difícil que se cometieran delitos. Esta tecnología está bien establecida y la estudiaremos en detalle en el capítulo 8. El problema es que los proveedores de hardware y software saben que poner funciones de seguridad cuesta dinero y que sus clientes no las solicitan. Además, una gran cantidad de los problemas proviene de un software con fallas, debido a que los proveedores saturan de funciones sus programas, lo que implica más código e, inevitablemente, más fallas. Un impuesto a las funciones nuevas podría ayudar, pero eso sería como vender un problema por centavos. Reponer el software defectuoso podría ser bueno, pero eso llevaría a la quiebra a toda la industria del software en el primer año.

1.2 HARDWARE DE REDES

Ya es tiempo de centrar nuevamente la atención en los temas técnicos correspondientes al diseño de redes (la parte de trabajo) y dejar a un lado las aplicaciones y los aspectos sociales de la conectividad (la parte divertida). Por lo general, no hay una sola clasificación aceptada en la que se ajusten todas las redes de computadoras, pero hay dos que destacan de manera importante: la tecnología de transmisión y la escala. Examinaremos cada una a la vez.

En un sentido amplio, hay dos tipos de tecnología de transmisión que se utilizan de manera extensa. Son las siguientes:

1. Enlaces de difusión.
2. Enlaces de punto a punto.

Las **redes de difusión** (*broadcast*) tienen un solo canal de comunicación, por lo que todas las máquinas de la red lo comparten. Si una máquina envía un mensaje corto —en ciertos contextos conocido como **paquete**—, todas las demás lo reciben. Un campo de dirección dentro del paquete especifica el destinatario. Cuando una máquina recibe un paquete, verifica el campo de dirección. Si el paquete va destinado a esa máquina, ésta lo procesa; si va destinado a alguna otra, lo ignora.

En una analogía, imagine a alguien que está parado al final de un corredor con varios cuartos a los lados y que grita: "Jorge, ven. Te necesito". Aunque en realidad el grito (paquete) podría haber sido escuchado (recibido), por muchas personas, sólo Jorge responde (lo procesa). Los demás simplemente lo ignoran. Otra analogía es la de los anuncios en un aeropuerto que piden a todos los pasajeros del vuelo 644 se reporten en la puerta 12 para abordar de inmediato.

Por lo general, los sistemas de difusión también permiten el direccionamiento de un paquete a *todos* los destinos utilizando un código especial en el campo de dirección. Cuando se transmite un paquete con este código, todas las máquinas de la red lo reciben y procesan. Este modo de operación se conoce como **difusión** (*broadcasting*). Algunos sistemas de difusión también soportan la transmisión a un subconjunto de máquinas, algo conocido como **multidifusión** (*multicasting*). Un esquema posible es la reserva de un bit para indicar la multidifusión. Los bits de dirección $n - 1$ restantes pueden contener un número de grupo. Cada máquina puede "suscribirse" a alguno o a todos los grupos. Cuando se envía un paquete a cierto grupo, se distribuye a todas las máquinas que se suscriben a ese grupo.

En contraste, las redes **punto a punto** constan de muchas conexiones entre pares individuales de máquinas. Para ir del origen al destino, un paquete en este tipo de red podría tener que visitar primero una o más máquinas intermedias. A menudo es posible que haya varias rutas o longitudes diferentes, de manera que encontrar las correctas es importante en redes de punto a punto. Por regla general (aunque hay muchas excepciones), las redes más pequeñas localizadas en una misma área geográfica tienden a utilizar la difusión, mientras que las más grandes suelen ser de punto a punto. La transmisión de punto a punto con un emisor y un receptor se conoce como **unidifusión** (*unicasting*).

Un criterio alternativo para la clasificación de las redes es su escala. En la figura 1-6 clasificamos los sistemas de procesadores múltiples por tamaño físico. En la parte superior se muestran las **redes de área personal,** que están destinadas para una sola persona. Por ejemplo, una red inalámbrica que conecta una computadora con su ratón, teclado e impresora, es una red de área personal. Incluso un PDA que controla el audífono o el marcapaso de un usuario encaja en esta categoría. A continuación de las redes de área personal se encuentran redes más grandes. Se pueden dividir en redes de área local, de área metropolitana y de área amplia. Por último, la conexión de dos o más redes se conoce como interred.

Distancia entre procesadores	Procesadores ubicados en el mismo	Ejemplo
1 m	Metro cuadrado	Red de área personal
10 m	Cuarto	Red de área local
100 m	Edificio	
1 km	Campus	
10 km	Ciudad	Red de área metropolitana
100 km	País	Red de área amplia
1,000 km	Continente	
10,000 km	Planeta	Internet

Figura 1-6. Clasificación de procesadores interconectados por escala.

Internet es un ejemplo bien conocido de una interred. La distancia es importante como una clasificación en metros porque se utilizan diferentes técnicas en diferentes escalas. En este libro nos ocuparemos de las redes en todas estas escalas. A continuación se proporciona una breve introducción al hardware de redes.

1.2.1 Redes de área local

Las **redes de área local** (generalmente conocidas como **LANs**) son redes de propiedad privada que se encuentran en un solo edificio o en un campus de pocos kilómetros de longitud. Se utilizan ampliamente para conectar computadoras personales y estaciones de trabajo en oficinas de una empresa y de fábricas para compartir recursos (por ejemplo, impresoras) e intercambiar información. Las LANs son diferentes de otros tipos de redes en tres aspectos: 1) tamaño; 2) tecnología de transmisión, y 3) topología.

Las LANs están restringidas por tamaño, es decir, el tiempo de transmisión en el peor de los casos es limitado y conocido de antemano. El hecho de conocer este límite permite utilizar ciertos tipos de diseño, lo cual no sería posible de otra manera. Esto también simplifica la administración de la red.

Las LANs podrían utilizar una tecnología de transmisión que consiste en un cable al cual están unidas todas las máquinas, como alguna vez lo estuvo parte de las líneas de las compañías telefónicas en áreas rurales. Las LANs tradicionales se ejecutan a una velocidad de 10 a 100 Mbps, tienen un retardo bajo (microsegundos o nanosegundos) y cometen muy pocos errores. Las LANs más nuevas funcionan hasta a 10 Gbps. En este libro continuaremos con lo tradicional y mediremos las velocidades de las líneas en megabits por segundo (1 Mbps es igual a 1,000,000 de bits por segundo) y gigabits por segundo (1 Gbps es igual a 1,000,000,000 de bits por segundo).

Para las LANs de difusión son posibles varias topologías. La figura 1-7 muestra dos de ellas. En una red de bus (es decir, un cable lineal), en cualquier instante al menos una máquina es la maestra y puede transmitir. Todas las demás máquinas se abstienen de enviar. Cuando se presenta el conflicto de que dos o más máquinas desean transmitir al mismo tiempo, se requiere un meca-

nismo de arbitraje. Tal mecanismo podría ser centralizado o distribuido. Por ejemplo, el IEEE 802.3, popularmente conocido como **Ethernet,** es una red de difusión basada en bus con control descentralizado, que por lo general funciona de 10 Mbps a 10 Gbps. Las computadoras que están en una Ethernet pueden transmitir siempre que lo deseen; si dos o más paquetes entran en colisión, cada computadora espera un tiempo aleatorio y lo intenta de nuevo más tarde.

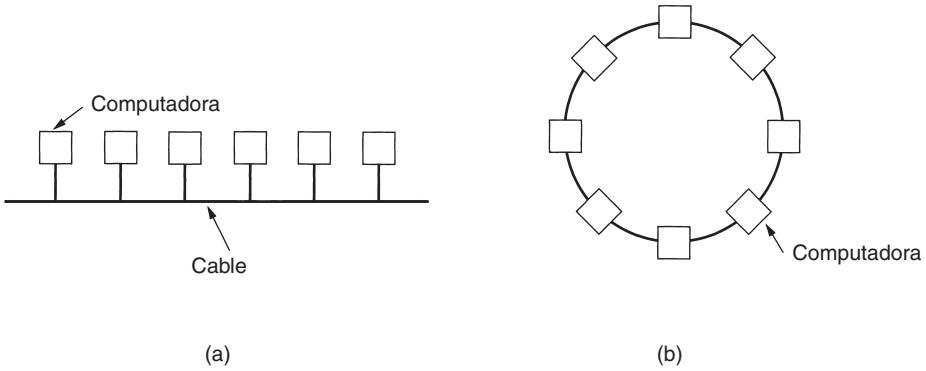

Figura 1-7. Dos redes de difusión. (a) De bus. (b) De anillo.

Un segundo tipo de sistema de difusión es el de anillo. En un anillo, cada bit se propaga por sí mismo, sin esperar al resto del paquete al que pertenece. Por lo común, cada bit navega por todo el anillo en el tiempo que le toma transmitir algunos bits, a veces incluso antes de que se haya transmitido el paquete completo. Al igual que con todos los demás sistemas de difusión, se requieren algunas reglas para controlar los accesos simultáneos al anillo. Se utilizan varios métodos, por ejemplo, el de que las máquinas deben tomar su turno. El IEEE 802.5 (el token ring de IBM) es una LAN basada en anillo que funciona a 4 y 16 Mbps. El FDDI es otro ejemplo de una red de anillo.

Las redes de difusión se pueden dividir aún más en estáticas y dinámicas, dependiendo de cómo se asigne el canal. Una asignación estática típica sería dividir el tiempo en intervalos discretos y utilizar un algoritmo *round-robin*, permitiendo que cada máquina transmita sólo cuando llegue su turno. La asignación estática desperdicia capacidad de canal cuando una máquina no tiene nada que transmitir al llegar su turno, por lo que la mayoría de los sistemas trata de asignar el canal de forma dinámica (es decir, bajo demanda).

Los métodos de asignación dinámica para un canal común pueden ser centralizados o descentralizados. En el método centralizado hay una sola entidad, por ejemplo, una unidad de arbitraje de bus, la cual determina quién sigue. Esto se podría hacer aceptando solicitudes y tomando decisiones de acuerdo con algunos algoritmos internos. En el método descentralizado de asignación de canal no hay una entidad central; cada máquina debe decidir por sí misma cuándo transmitir. Usted podría pensar que esto siempre conduce al caos, pero no es así. Más adelante estudiaremos muchos algoritmos designados para poner orden y evitar el caos potencial.

1.2.2 Redes de área metropolitana

Una **red de área metropolitana** (**MAN**) abarca una ciudad. El ejemplo más conocido de una MAN es la red de televisión por cable disponible en muchas ciudades. Este sistema creció a partir de los primeros sistemas de antena comunitaria en áreas donde la recepción de la televisión al aire era pobre. En dichos sistemas se colocaba una antena grande en la cima de una colina cercana y la señal se canalizaba a las casas de los suscriptores.

Al principio eran sistemas diseñados de manera local con fines específicos. Después las compañías empezaron a pasar a los negocios, y obtuvieron contratos de los gobiernos de las ciudades para cablear toda una ciudad. El siguiente paso fue la programación de televisión e incluso canales designados únicamente para cable. Con frecuencia, éstos emitían programas de un solo tema, como sólo noticias, deportes, cocina, jardinería, etcétera. Sin embargo, desde su inicio y hasta finales de la década de 1990, estaban diseñados únicamente para la recepción de televisión.

A partir de que Internet atrajo una audiencia masiva, los operadores de la red de TV por cable se dieron cuenta de que con algunos cambios al sistema, podrían proporcionar servicio de Internet de dos vías en las partes sin uso del espectro. En ese punto, el sistema de TV por cable empezaba a transformarse de una forma de distribución de televisión a una red de área metropolitana. Para que se dé una idea, una MAN podría verse como el sistema que se muestra en la figura 1-8, donde se aprecia que las señales de TV e Internet se alimentan hacia un **amplificador head end** para enseguida transmitirse a las casas de las personas. En el capítulo 2 trataremos con detalle este tema.

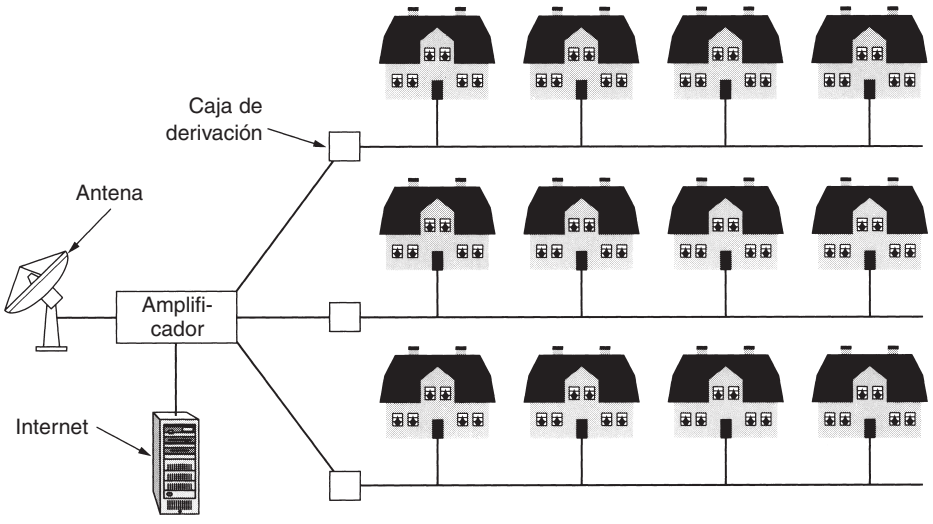

Figura 1-8. Una red de área metropolitana, basada en TV por cable.

La televisión por cable no es solamente una MAN. Desarrollos recientes en el acceso inalámbrico a alta velocidad a Internet dieron como resultado otra MAN, que se estandarizó como IEEE 802.16. En el capítulo 2 veremos esta área.

1.2.3 Redes de área amplia

Una **red de área amplia (WAN),** abarca una gran área geográfica, con frecuencia un país o un continente. Contiene un conjunto de máquinas diseñado para programas (es decir, aplicaciones) de usuario. Seguiremos el uso tradicional y llamaremos **hosts** a estas máquinas. Los *hosts* están conectados por una **subred de comunicación,** o simplemente **subred,** para abreviar. Los clientes son quienes poseen a los *hosts* (es decir, las computadoras personales de los usuarios), mientras que, por lo general, las compañías telefónicas o los proveedores de servicios de Internet poseen y operan la subred de comunicación. La función de una subred es llevar mensajes de un *host* a otro, como lo hace el sistema telefónico con las palabras del que habla al que escucha. La separación de los aspectos de la comunicación pura de la red (la subred) de los aspectos de la aplicación (los *hosts*), simplifica en gran medida todo el diseño de la red.

En la mayoría de las redes de área amplia la subred consta de dos componente distintos: líneas de transmisión y elementos de conmutación. Las **líneas de transmisión** mueven bits entre máquinas. Pueden estar hechas de cable de cobre, fibra óptica o, incluso, radioenlaces. Los **elementos de conmutación** son computadoras especializadas que conectan tres o más líneas de transmisión. Cuando los datos llegan a una línea de entrada, el elemento de conmutación debe elegir una línea de salida en la cual reenviarlos. Estas computadoras de conmutación reciben varios nombres; conmutadores y enrutadores son los más comunes.

En este modelo, que se muestra en la figura 1-9, cada *host* está conectado frecuentemente a una LAN en la que existe un enrutador, aunque en algunos casos un *host* puede estar conectado de manera directa a un enrutador. El conjunto de líneas de comunicación y enrutadores (pero no de *hosts*) forma la subred.

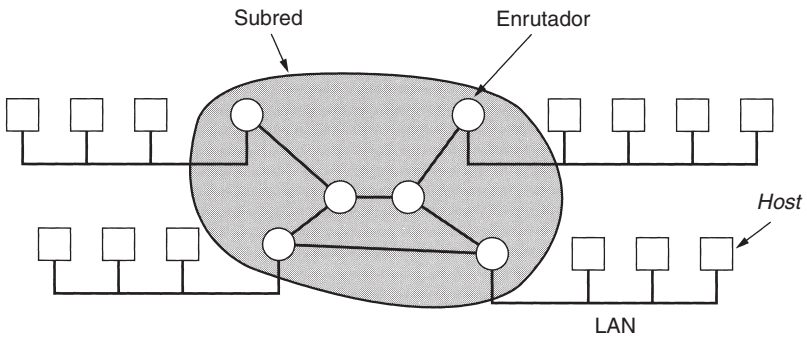

Figura 1-9. Relación entre *hosts* de LANs y la subred.

A continuación se presenta un breve comentario acerca del término "subred". Originalmente, su **único** significado era el conjunto de enrutadores y líneas de comunicación que movía paquetes del *host* de origen al de destino. Sin embargo, algunos años más tarde también adquirió un segundo

significado junto con el direccionamiento de redes (que expondremos en el capítulo 5). Desgraciadamente, no existe una alternativa de amplio uso con respecto a su significado inicial por lo que, con algunas reservas, utilizaremos este término en ambos sentidos. El contexto dejará en claro su significado.

En la mayoría de las WANs, la red contiene numerosas líneas de transmisión, cada una de las cuales conecta un par de enrutadores. Si dos enrutadores que no comparten una línea de transmisión quieren conectarse, deberán hacerlo de manera indirecta, a través de otros enrutadores. Cuando un paquete es enviado desde un enrutador a otro a través de uno o más enrutadores intermedios, el paquete se recibe en cada enrutador intermedio en su totalidad, se almacena ahí hasta que la línea de salida requerida esté libre y, por último, se reenvía. Una subred organizada a partir de este principio se conoce como subred de **almacenamiento y reenvío** (*store and forward*) o de **conmutación de paquetes.** Casi todas las redes de área amplia (excepto las que utilizan satélites) tienen subredes de almacenamiento y reenvío. Cuando los paquetes son pequeños y tienen el mismo tamaño, se les llama **celdas.**

El principio de una WAN de conmutación de paquetes es tan importante que vale la pena dedicarle algunas palabras más. En general, cuando un proceso de cualquier *host* tiene un mensaje que se va a enviar a un proceso de algún otro *host*, el *host* emisor divide primero el mensaje en paquetes, los cuales tienen un número de secuencia. Estos paquetes se envían entonces por la red de uno en uno en una rápida sucesión. Los paquetes se transportan de forma individual a través de la red y se depositan en el *host* receptor, donde se reensamblan en el mensaje original y se entregan al proceso receptor. En la figura 1-10 se ilustra un flujo de paquetes correspondiente a algún mensaje inicial.

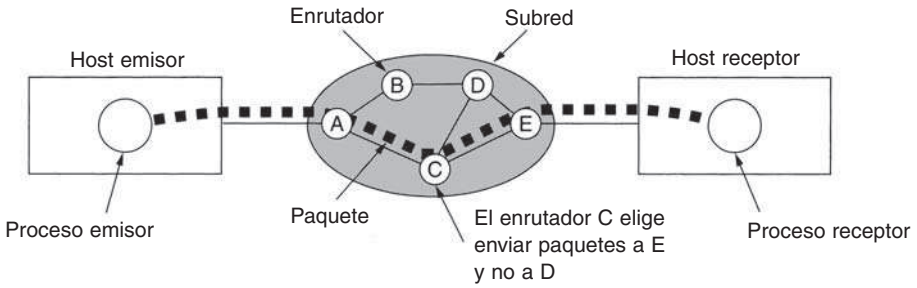

Figura 1-10. Flujo de paquetes desde un emisor a un receptor.

En esta figura todos los paquetes siguen la ruta *ACE* en vez de la *ABDE* o *ACDE*. En algunas redes todos los paquetes de un mensaje determinado *deben* seguir la misma ruta; en otras, cada paquete se enruta por separado. Desde luego, si *ACE* es la mejor ruta, todos los paquetes se podrían enviar a través de ella, incluso si cada paquete se enruta de manera individual.

Las decisiones de enrutamiento se hacen de manera local. Cuando un paquete llega al enrutador *A*, éste debe decidir si el paquete se enviará hacia *B* o hacia *C*. La manera en que el enrutador *A* toma esa decisión se conoce como **algoritmo de enrutamiento.** Existen muchos de ellos. En el capítulo 5 estudiaremos con detalle algunos.

No todas las WANs son de conmutación de paquetes. Una segunda posibilidad para una WAN es un sistema satelital. Cada enrutador tiene una antena a través de la cual puede enviar y recibir. Todos los enrutadores pueden escuchar la salida *desde* el satélite y, en algunos casos, también pueden escuchar las transmisiones de los demás enrutadores *hacia* el satélite. Algunas veces los enrutadores están conectados a una subred de punto a punto elemental, y sólo algunos de ellos tienen una antena de satélite. Por naturaleza, las redes satelital son de difusión y son más útiles cuando la propiedad de difusión es importante.

1.2.4 Redes inalámbricas

La comunicación inalámbrica digital no es una idea nueva. A principios de 1901, el físico italiano Guillermo Marconi demostró un telégrafo inalámbrico desde un barco a tierra utilizando el código Morse (después de todo, los puntos y rayas son binarios). Los sistemas inalámbricos digitales de la actualidad tienen un mejor desempeño, pero la idea básica es la misma.

Como primera aproximación, las redes inalámbricas se pueden dividir en tres categorías principales:

1. Interconexión de sistemas.
2. LANs inalámbricas.
3. WANs inalámbricas.

La interconexión de sistemas se refiere a la interconexión de componentes de una computadora que utiliza radio de corto alcance. La mayoría de las computadoras tiene un monitor, teclado, ratón e impresora, conectados por cables a la unidad central. Son tantos los usuarios nuevos que tienen dificultades para conectar todos los cables en los enchufes correctos (aun cuando suelen estar codificados por colores) que la mayoría de los proveedores de computadoras ofrece la opción de enviar a un técnico a la casa del usuario para que realice esta tarea. En consecuencia, algunas compañías se reunieron para diseñar una red inalámbrica de corto alcance llamada **Bluetooth** para conectar sin cables estos componentes. Bluetooth también permite conectar cámaras digitales, auriculares, escáneres y otros dispositivos a una computadora con el único requisito de que se encuentren dentro del alcance de la red. Sin cables, sin instalación de controladores, simplemente se colocan, se encienden y funcionan. Para muchas personas, esta facilidad de operación es algo grandioso.

En la forma más sencilla, las redes de interconexión de sistemas utilizan el paradigma del maestro y el esclavo de la figura 1-11(a). La unidad del sistema es, por lo general, el maestro que trata al ratón, al teclado, etcétera, como a esclavos. El maestro le dice a los esclavos qué direcciones utilizar, cuándo pueden difundir, durante cuánto tiempo pueden transmitir, qué frecuencias pueden utilizar, etcétera. En el capítulo 4 explicaremos con más detalle el Bluetooth.

El siguiente paso en la conectividad inalámbrica son las LANs inalámbricas. Son sistemas en los que cada computadora tiene un módem de radio y una antena mediante los que se puede comunicar con otros sistemas. En ocasiones, en el techo se coloca una antena con la que las máquinas se comunican, como se ilustra en la figura 1-11(b). Sin embargo, si los sistemas están lo suficientemente cerca, se pueden comunicar de manera directa entre sí en una configuración de

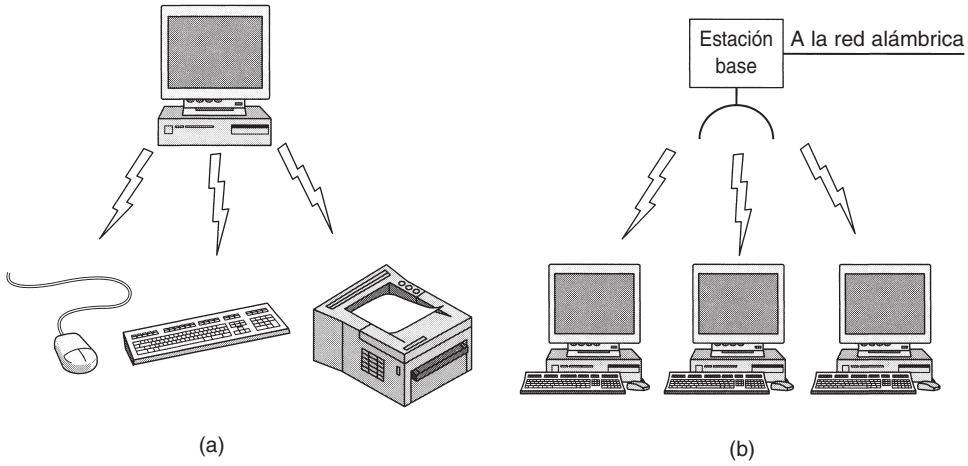

Figura 1-11. (a) Configuración Bluetooth. (b) LAN inalámbrica.

igual a igual. Las LANs inalámbricas se están haciendo cada vez más comunes en casas y oficinas pequeñas, donde instalar Ethernet se considera muy problemático, así como en oficinas ubicadas en edificios antiguos, cafeterías de empresas, salas de conferencias y otros lugares. Existe un estándar para las LANs inalámbricas, llamado **IEEE 802.11,** que la mayoría de los sistemas implementa y que se ha extendido ampliamente. Esto lo explicaremos en el capítulo 4.

El tercer tipo de red inalámbrica se utiliza en sistemas de área amplia. La red de radio utilizada para teléfonos celulares es un ejemplo de un sistema inalámbrico de banda ancha baja. Este sistema ha pasado por tres generaciones. La primera era analógica y sólo para voz. La segunda era digital y sólo para voz. La tercera generación es digital y es tanto para voz como para datos. En cierto sentido, las redes inalámbricas celulares son como las LANs inalámbricas, excepto porque las distancias implicadas son mucho más grandes y las tasas de bits son mucho más bajas. Las LANs inalámbricas pueden funcionar a tasas de hasta 50 Mbps en distancias de decenas de metros. Los sistemas celulares funcionan debajo de 1 Mbps, pero la distancia entre la estación base y la computadora o teléfono se mide en kilómetros más que en metros. En el capítulo 2 hablaremos con mucho detalle sobre estas redes.

Además de estas redes de baja velocidad, también se han desarrollado las redes inalámbricas de área amplia con alto ancho de banda. El enfoque inicial es el acceso inalámbrico a Internet a alta velocidad, desde los hogares y las empresas, dejando a un lado el sistema telefónico. Este servicio se suele llamar servicio de distribución local multipuntos. Lo estudiaremos más adelante. También se ha desarrollado un estándar para éste, llamado IEEE 802.16. Examinaremos dicho estándar en el capítulo 4.

La mayoría de las redes inalámbricas se enlaza a la red alámbrica en algún punto para proporcionar acceso a archivos, bases de datos e Internet. Hay muchas maneras de efectuar estas conexiones, dependiendo de las circunstancias. Por ejemplo, en la figura 1-12(a) mostramos un aeroplano con una serie de personas que utilizan módems y los teléfonos de los respaldos para llamar a la oficina. Cada llamada es independiente de las demás. Sin embargo, una opción mucho

más eficiente es la LAN dentro del avión de la figura 1-12(b), donde cada asiento está equipado con un conector Ethernet al cual los pasajeros pueden acoplar sus computadoras. El avión tiene un solo enrutador, el cual mantiene un enlace de radio con algún enrutador que se encuentre en tierra, y cambia de enrutador conforme avanza el vuelo. Esta configuración es una LAN tradicional, excepto porque su conexión al mundo exterior se da mediante un enlace por radio en lugar de una línea cableada.

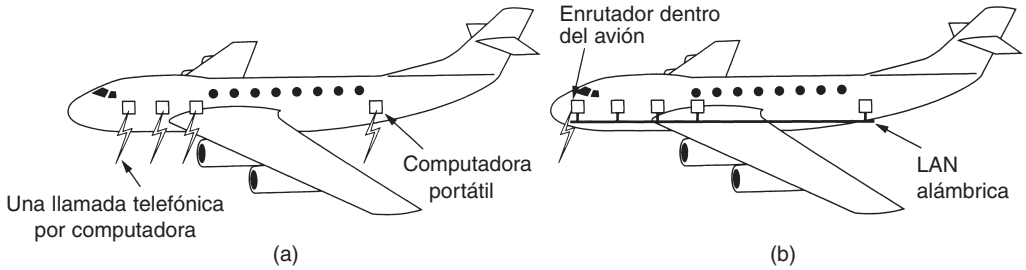

Figura 1-12. (a) Computadoras móviles individuales. (b) LAN dentro del avión.

Muchas personas creen que lo inalámbrico es la onda del futuro (por ejemplo, Bi y cols., 2001; Leeper, 2001; Varshey y Vetter, 2000) pero se ha escuchado una voz disidente. Bob Metcalfe, el inventor de Ethernet, ha escrito: "Las computadoras inalámbricas móviles son como los baños portátiles sin cañería: bacinicas portátiles. Serán muy comunes en los vehículos, en sitios en construcción y conciertos de rock. Mi consejo es que coloque cables en su casa y se quede ahí" (Metcalfe, 1995). La historia podría colocar esta cita en la misma categoría que la explicación de T.J. Watson, presidente de IBM en 1945, de por qué esta empresa no entraba en el negocio de las computadoras: "Cuatro o cinco computadoras deberán ser suficientes para todo el mundo hasta el año 2000".

1.2.5 Redes domésticas

La conectividad doméstica está en el horizonte. La idea fundamental es que en el futuro la mayoría de los hogares estarán preparados para conectividad de redes. Cualquier dispositivo del hogar será capaz de comunicarse con todos los demás dispositivos y todos podrán accederse por Internet. Éste es uno de esos conceptos visionarios que nadie solicitó (como los controles remotos de TV o los teléfonos celulares), pero una vez que han llegado nadie se puede imaginar cómo habían podido vivir sin ellos.

Muchos dispositivos son capaces de estar conectados en red. Algunas de las categorías más evidentes (con ejemplos) son las siguientes:

1. Computadoras (de escritorio, portátiles, PDAs, periféricos compartidos).
2. Entretenimiento (TV, DVD, VCR, videocámara, cámara fotográfica, estereofónicos, MP3).
3. Telecomunicaciones (teléfono, teléfono móvil, intercomunicadores, fax).
4. Aparatos electrodomésticos (horno de microondas, refrigerador, reloj, horno, aire acondicionado, luces).
5. Telemetría (metro utilitario, alarma contra fuego y robo, termostato, cámaras inalámbricas).

La conectividad de computadoras domésticas ya está aquí, aunque limitada. Muchas casas ya cuentan con un dispositivo para conectar varias computadoras para una conexión rápida a Internet. El entretenimiento por red aún no existe, pero cuanto más y más música y películas se puedan descargar de Internet, habrá más demanda para que los equipos de audio y las televisiones se conecten a Internet. Incluso las personas desearán compartir sus propios vídeos con amigos y familiares, por lo que deberá haber una conexión en ambos sentidos. Los dispositivos de telecomunicaciones ya están conectados al mundo exterior, pero pronto serán digitales y tendrán capacidad de funcionar sobre Internet. Un hogar promedio tal vez tiene una docena de relojes (los de los aparatos electrodomésticos), y todos se tienen que reajustar dos veces al año cuando inicia y termina el tiempo de ahorro de luz de día (horario de verano). Si todos los relojes estuvieran conectados a Internet, ese reajuste se haría en forma automática. Por último, el monitoreo remoto de la casa y su contenido es el probable ganador. Es muy factible que muchos padres deseen invertir en monitorear con sus PDAs a sus bebés dormidos cuando van a cenar fuera de casa, aun cuando contraten a una niñera. Si bien podemos imaginar una red separada para cada área de aplicación, la integración de todas en una sola red es probablemente una mejor idea.

La conectividad doméstica tiene algunas propiedades diferentes a las de otro tipo de redes. Primero, la red y los dispositivos deben ser fáciles de instalar. El autor ha instalado numerosas piezas de hardware y software en varias computadoras durante varios años con resultados diferentes. Al realizar una serie de llamadas telefónicas al personal de soporte técnico del proveedor por lo general recibió respuestas como: 1) Lea el manual; 2) Reinicie la computadora; 3) Elimine todo el hardware y software, excepto los nuestros, y pruebe de nuevo; 4) Descargue de nuestro sitio Web el controlador más reciente y, si todo eso falla, 5) Reformatee el disco duro y reinstale Windows desde el CD-ROM. Decirle al comprador de un refrigerador con capacidad de Internet que descargue e instale una nueva versión del sistema operativo del refrigerador, no conduce a tener clientes contentos. Los usuarios de computadoras están acostumbrados a soportar productos que no funcionan; los clientes que compran automóviles, televisiones y refrigeradores son mucho menos tolerantes. Esperan productos que trabajen al 100% desde que se compran.

Segundo, la red y los dispositivos deben estar plenamente probados en operación. Los equipos de aire acondicionado solían tener una perilla con cuatro parámetros: OFF, LOW, MEDIUM y HIGH (apagado, bajo, medio, alto). Ahora tienen manuales de 30 páginas. Una vez que puedan conectarse en red, no se le haga extraño que tan sólo el capítulo de seguridad tenga 30 páginas. Esto estará más allá de la comprensión de prácticamente todos los usuarios.

Tercero, el precio bajo es esencial para el éxito. Muy pocas personas, si no es que ninguna, pagarán un precio adicional de $50 por un termostato con capacidad de Internet, debido a que no considerarán que monitorear la temperatura de sus casas desde sus trabajos sea algo importante. Tal vez por $5 sí lo comprarían.

Cuarto, la principal aplicación podría implicar multimedia, por lo que la red necesita capacidad suficiente. No hay mercado para televisiones conectadas a Internet que proyecten películas inseguras a una resolución de 320×240 píxeles y 10 cuadros por segundo. Fast Ethernet, el caballo de batalla en la mayoría de las oficinas, no es bastante buena para multimedia. En consecuencia, para que las redes domésticas lleguen a ser productos masivos en el mercado, requerirán mejor desempeño que el de las redes de oficina actuales, así como precios más bajos.

Quinto, se podría empezar con uno o dos dispositivos y expandir de manera gradual el alcance de la red. Esto significa que no habrá problemas con el formato. Decir a los consumidores que adquieran periféricos con interfaces IEEE 1394 (FireWire) y años después retractarse y decir que USB 2.0 es la interfaz del mes, es hacer clientes caprichosos. La interfaz de red tendrá que permanecer estable durante muchos años; el cableado (si lo hay) deberá permanecer estable durante décadas.

Sexto, la seguridad y la confianza serán muy importantes. Perder algunos archivos por un virus de correo electrónico es una cosa; que un ladrón desarme su sistema de seguridad desde su PDA y luego saquee su casa es algo muy diferente.

Una pregunta interesante es si las redes domésticas serán alámbricas o inalámbricas. La mayoría de los hogares ya tiene seis redes instaladas: electricidad, teléfono, televisión por cable, agua, gas y alcantarillado. Agregar una séptima durante la construcción de una casa no es difícil, pero acondicionar las casas existentes para agregar dicha red es costoso. Los costos favorecen la conectividad inalámbrica, pero la seguridad favorece la conectividad alámbrica. El problema con la conectividad inalámbrica es que las ondas de radio que utiliza traspasan las paredes con mucha facilidad. No a todos les gusta la idea de que cuando vaya a imprimir, se tope con la conexión de su vecino y pueda leer el correo electrónico de éste. En el capítulo 8 estudiaremos cómo se puede utilizar la encriptación para proporcionar seguridad, pero en el contexto de una red doméstica la seguridad debe estar bien probada, incluso para usuarios inexpertos. Es más fácil decirlo que hacerlo, incluso en el caso de usuarios expertos.

Para abreviar, la conectividad doméstica ofrece muchas oportunidades y retos. La mayoría de ellos se relaciona con la necesidad de que sean fáciles de manejar, confiables y seguros, en particular en manos de usuarios no técnicos, y que al mismo tiempo proporcionen alto desempeño a bajo costo.

1.2.6 Interredes

Existen muchas redes en el mundo, a veces con hardware y software diferentes. Con frecuencia, las personas conectadas a una red desean comunicarse con personas conectadas a otra red diferente. La satisfacción de este deseo requiere que se conecten diferentes redes, con frecuencia incompatibles, a veces mediante máquinas llamadas **puertas de enlace** (*gateways*) para hacer la conexión y proporcionar la traducción necesaria, tanto en términos de hardware como de software. Un conjunto de redes interconectadas se llama **interred.**

Una forma común de interred es el conjunto de LANs conectadas por una WAN. De hecho, si tuviéramos que reemplazar la etiqueta “subred” en la figura 1-9 por “WAN”, no habría nada más que cambiar en la figura. En este caso, la única diferencia técnica real entre una subred y una WAN es si hay *hosts* presentes. Si el sistema que aparece en el área gris contiene solamente enrutadores, es una subred; si contiene enrutadores y *hosts*, es una WAN. Las diferencias reales se relacionan con la propiedad y el uso.

Subredes, redes e interredes con frecuencia se confunden. La subred tiene más sentido en el contexto de una red de área amplia, donde se refiere a un conjunto de enrutadores y líneas de

comunicación poseídas por el operador de redes. Como una analogía, el sistema telefónico consta de oficinas de conmutación telefónica que se conectan entre sí mediante líneas de alta velocidad, y a los hogares y negocios, mediante líneas de baja velocidad. Estas líneas y equipos, poseídas y administradas por la compañía de teléfonos, forman la subred del sistema telefónico. Los teléfonos mismos (los *hosts* en esta analogía) no son parte de la subred. La combinación de una subred y sus *hosts* forma una red. En el caso de una LAN, el cable y los *hosts* forman la red. En realidad, ahí no hay una subred.

Una interred se forma cuando se interconectan redes diferentes. Desde nuestro punto de vista, al conectar una LAN y una WAN o conectar dos LANs se forma una interred, pero existe poco acuerdo en la industria en cuanto a la terminología de esta área. Una regla de oro es que si varias empresas pagaron por la construcción de diversas partes de la red y cada una mantiene su parte, tenemos una interred más que una sola red. Asimismo, si la terminología subyacente es diferente en partes diferentes (por ejemplo, difusión y punto a punto), probablemente tengamos dos redes.

1.3 SOFTWARE DE REDES

Las primeras redes de computadoras se diseñaron teniendo al hardware como punto principal y al software como secundario. Esta estrategia ya no funciona. Actualmente el software de redes está altamente estructurado. En las siguientes secciones examinaremos en detalle la técnica de estructuración de software. El método descrito aquí es la clave de todo el libro y se presentará con mucha frecuencia más adelante.

1.3.1 Jerarquías de protocolos

Para reducir la complejidad de su diseño, la mayoría de las redes está organizada como una pila de **capas** o **niveles,** cada una construida a partir de la que está debajo de ella. El número de capas, así como el nombre, contenido y función de cada una de ellas difieren de red a red. El propósito de cada capa es ofrecer ciertos servicios a las capas superiores, a las cuales no se les muestran los detalles reales de implementación de los servicios ofrecidos.

Este concepto es muy conocido y utilizado en la ciencia computacional, donde se conoce de diversas maneras, como ocultamiento de información, tipos de datos abstractos, encapsulamiento de datos y programación orientada a objetos. La idea básica es que una pieza particular de software (o hardware) proporciona un servicio a sus usuarios pero nunca les muestra los detalles de su estado interno ni sus algoritmos.

La capa *n* de una máquina mantiene una conversación con la capa *n* de otra máquina. Las reglas y convenciones utilizadas en esta conversación se conocen de manera colectiva como protocolo de capa *n*. Básicamente, un **protocolo** es un acuerdo entre las partes en comunicación sobre cómo se debe llevar a cabo la comunicación. Como una analogía, cuando se presenta una mujer con un hombre, ella podría elegir no darle la mano. Él, a su vez, podría decidir saludarla de mano o de beso, dependiendo, por ejemplo, de si es una abogada americana o una princesa europea en

una reunión social formal. Violar el protocolo hará más difícil la comunicación, si no es que imposible.

En la figura 1-13 se ilustra una red de cinco capas. Las entidades que abarcan las capas correspondientes en diferentes máquinas se llaman **iguales** (*peers*). Los iguales podrían ser procesos, dispositivos de hardware o incluso seres humanos. En otras palabras, los iguales son los que se comunican a través del protocolo.

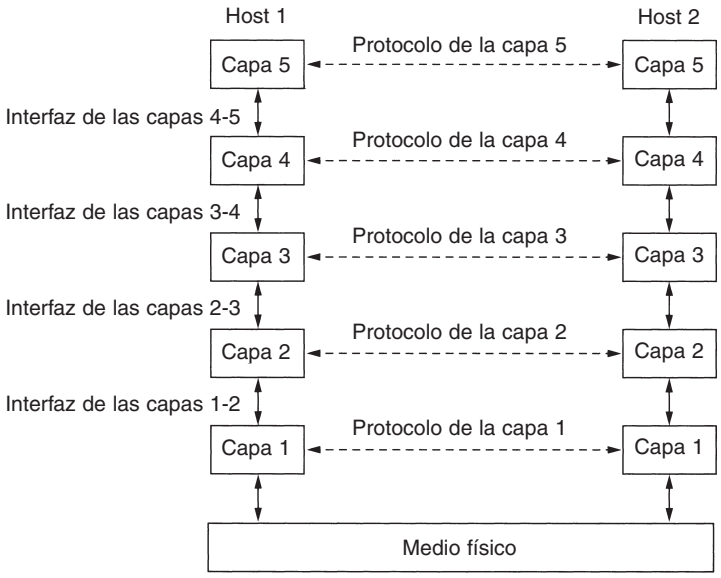

Figura 1-13. Capas, protocolos e interfaces.

En realidad, los datos no se transfieren de manera directa desde la capa *n* de una máquina a la capa *n* de la otra máquina, sino que cada capa pasa los datos y la información de control a la capa inmediatamente inferior, hasta que se alcanza la capa más baja. Debajo de la capa 1 se encuentra el **medio físico** a través del cual ocurre la comunicación real. En la figura 1-13, la comunicación virtual se muestra con líneas punteadas, en tanto que la física, con líneas sólidas.

Entre cada par de capas adyacentes está una **interfaz.** Ésta define qué operaciones y servicios primitivos pone la capa más baja a disposición de la capa superior inmediata. Cuando los diseñadores de redes deciden cuántas capas incluir en una red y qué debe hacer cada una, una de las consideraciones más importantes es definir interfaces limpias entre las capas. Hacerlo así, a su vez, requiere que la capa desempeñe un conjunto específico de funciones bien entendidas. Además de minimizar la cantidad de información que se debe pasar entre las capas, las interfaces bien definidas simplifican el reemplazo de la implementación de una capa con una implementación totalmente diferente (por ejemplo, todas las líneas telefónicas se reemplazan con canales por satélite)

porque todo lo que se pide de la nueva implementación es que ofrezca exactamente el mismo conjunto de servicios a su vecino de arriba, como lo hacía la implementación anterior. De hecho, es muy común que diferentes *hosts* utilicen diferentes implementaciones.

Un conjunto de capas y protocolos se conoce como **arquitectura de red.** La especificación de una arquitectura debe contener información suficiente para permitir que un implementador escriba el programa o construya el hardware para cada capa de modo que se cumpla correctamente con el protocolo apropiado. Ni los detalles de la implementación ni las especificaciones de las interfaces son parte de la arquitectura porque están ocultas en el interior de las máquinas y no son visibles desde el exterior. Incluso, tampoco es necesario que las interfaces de todas las máquinas en una red sean las mismas, siempre y cuando cada máquina pueda utilizar correctamente todos los protocolos. La lista de protocolos utilizados por un sistema, un protocolo por capa, se conoce como **pila de protocolos.** Los aspectos de las arquitecturas de red, las pilas de protocolos y los protocolos mismos son el tema principal de este libro.

Una analogía podría ayudar a explicar la idea de comunicación entre múltiples capas. Imagine a dos filósofos (procesos de iguales en la capa 3), uno de los cuales habla urdu e inglés, y el otro chino y francés. Puesto que no tienen un idioma común, cada uno contrata un traductor (proceso de iguales en la capa 2) y cada uno a su vez contacta a una secretaria (procesos de iguales en la capa 1). El filósofo 1 desea comunicar su afición por el *oryctolagus cuniculus* a su igual. Para eso, le pasa un mensaje (en inglés) a través de la interfaz de las capas 2-3 a su traductor, diciendo: "Me gustan los conejos", como se ilustra en la figura 1-14. Los traductores han acordado un idioma neutral conocido por ambos, el holandés, para que el mensaje se convierta en "Ik vind konijnen leuk". La elección del idioma es el protocolo de la capa 2 y los procesos de iguales de dicha capa son quienes deben realizarla.

Entonces el traductor le da el mensaje a una secretaria para que lo transmita por, digamos, fax (el protocolo de la capa 1). Cuando el mensaje llega, se traduce al francés y se pasa al filósofo 2 a través de la interfaz de las capas 2-3. Observe que cada protocolo es totalmente independiente de los demás en tanto no cambien las interfaces. Los traductores pueden cambiar de holandés a, digamos, finlandés, a voluntad, siempre y cuando los dos estén de acuerdo y no cambien su interfaz con las capas 1 o 3. Del mismo modo, las secretarias pueden cambiar de fax a correo electrónico o teléfono sin molestar (o incluso avisar) a las demás capas. Cada proceso podría agregar alguna información destinada sólo a su igual. Esta información no se pasa a la capa superior.

Ahora veamos un ejemplo más técnico: cómo proporcionar comunicación a la capa superior de la red de cinco capas de la figura 1-15. Un proceso de aplicación que se ejecuta en la capa 5 produce un mensaje, *M,* y lo pasa a la capa 4 para su transmisión.

La capa 4 pone un **encabezado** al frente del mensaje para identificarlo y pasa el resultado a la capa 3. El encabezado incluye información de control, como números de secuencia, para que la capa 4 de la máquina de destino entregue los mensajes en el orden correcto si las capas inferiores no mantienen la secuencia. En algunas capas los encabezados también pueden contener tamaños, medidas y otros campos de control.

En muchas redes no hay límites para el tamaño de mensajes transmitidos en el protocolo de la capa 4, pero casi siempre hay un límite impuesto por el protocolo de la capa 3. En consecuencia,

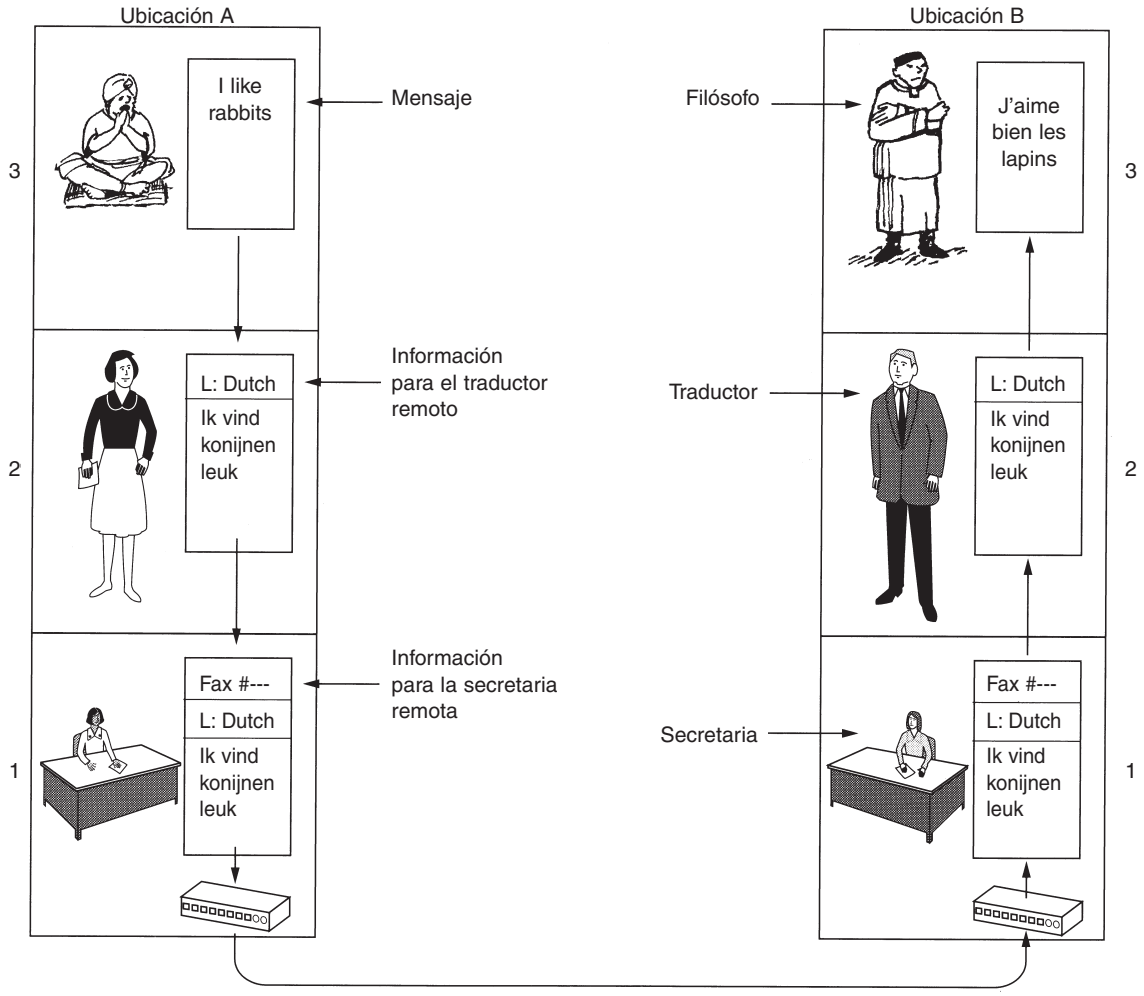

Figura 1-14. Arquitectura filósofo-traductor-secretaria.

la capa 3 debe desintegrar en unidades más pequeñas, paquetes, los mensajes que llegan, y a cada paquete le coloca un encabezado. En este ejemplo, M se divide en dos partes, M_1 y M_2.

La capa 3 decide cuál de las líneas que salen utilizar y pasa los paquetes a la capa 2. Ésta no sólo agrega un encabezado a cada pieza, sino también un terminador, y pasa la unidad resultante a la capa 1 para su transmisión física. En la máquina receptora el mensaje pasa hacia arriba de capa en capa, perdiendo los encabezados conforme avanza. Ninguno de los encabezados de las capas inferiores a n llega a la capa n.

Lo que debe entender en la figura 1-15 es la relación entre las comunicaciones virtual y real, y la diferencia entre protocolos e interfaces. Por ejemplo, los procesos de iguales en la capa 4 piensan

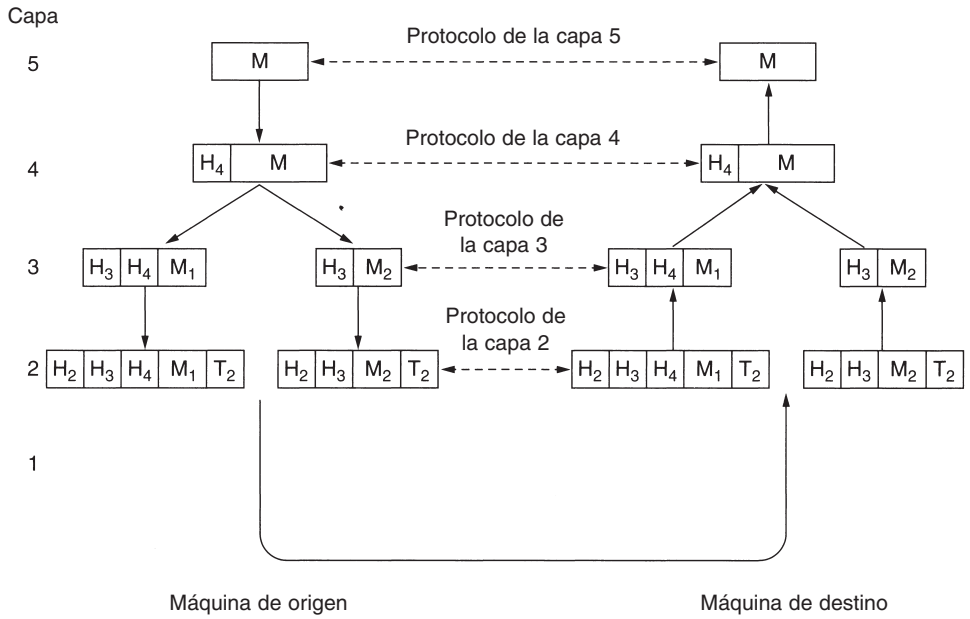

Figura 1-15. Ejemplo de flujo de información que soporta una comunicación virtual en la capa 5.

conceptualmente de su comunicación como si fuera "horizontal", y utilizan el protocolo de la capa 4. Pareciera que cada uno tuviera un procedimiento llamado algo así como *EnviadoalOtroLado* y *RecibidoDesdeElOtroLado*, aun cuando estos procedimientos en realidad se comunican con las capas inferiores a través de la interfaz de las capas 3-4, no con el otro lado.

La abstracción del proceso de iguales es básica para todo diseño de red. Al utilizarla, la inmanejable tarea de diseñar toda la red se puede fragmentar en varios problemas de diseño más pequeños y manejables, es decir, el diseño de las capas individuales.

Aunque la sección 1.3 se llama "Software de redes", vale la pena precisar que las capas inferiores de una jerarquía de protocolos se implementan con frecuencia en el hardware o en el firmware. No obstante, están implicados los algoritmos de protocolo complejos, aun cuando estén integrados (en todo o en parte) en el hardware.

1.3.2 Aspectos de diseño de las capas

Algunos de los aspectos clave de diseño que ocurren en las redes de computadoras están presentes en las diversas capas. Más adelante mencionaremos brevemente algunos de los más importantes.

Cada capa necesita un mecanismo para identificar a los emisores y a los receptores. Puesto que una red por lo general tiene muchas computadoras —algunas de las cuales tienen varios procesos—, se necesita un método para que un proceso en una máquina especifique con cuál de ellas

quiere hablar. Como consecuencia de tener múltiples destinos, se necesita alguna forma de **direccionamiento** a fin de precisar un destino específico.

Otro conjunto de decisiones de diseño concierne a las reglas de la transferencia de datos. En algunos sistemas, los datos viajan sólo en una dirección; en otros, pueden viajar en ambas direcciones. El protocolo también debe determinar a cuántos canales lógicos corresponde la conexión y cuáles son sus prioridades. Muchas redes proporcionan al menos dos canales lógicos por conexión, uno para los datos normales y otro para los urgentes.

El **control de errores** es un aspecto importante porque los circuitos de comunicación física no son perfectos. Muchos códigos de detección y corrección de errores son conocidos, pero los dos extremos de la conexión deben estar de acuerdo en cuál es el que se va a utilizar. Además, el receptor debe tener algún medio de decirle al emisor qué mensajes se han recibido correctamente y cuáles no.

No todos los canales de comunicación conservan el orden en que se les envían los mensajes. Para tratar con una posible pérdida de secuencia, el protocolo debe incluir un mecanismo que permita al receptor volver a unir los pedazos en forma adecuada. Una solución obvia es numerar las piezas, pero esta solución deja abierta la cuestión de qué se debe hacer con las piezas que llegan sin orden.

Un aspecto que ocurre en cada nivel es cómo evitar que un emisor rápido sature de datos a un receptor más lento. Se han propuesto varias soluciones que explicaremos más adelante. Algunas de ellas implican algún tipo de retroalimentación del receptor al emisor, directa o indirectamente, dependiendo de la situación actual del receptor. Otros limitan al emisor a una velocidad de transmisión acordada. Este aspecto se conoce como **control de flujo.**

Otro problema que se debe resolver en algunos niveles es la incapacidad de todos los procesos de aceptar de manera arbitraria mensajes largos. Esta propiedad conduce a mecanismos para desensamblar, transmitir y reensamblar mensajes. Un aspecto relacionado es el problema de qué hacer cuando los procesos insisten en transmitir datos en unidades tan pequeñas que enviarlas por separado es ineficaz. La solución a esto es reunir en un solo mensaje grande varios mensajes pequeños que vayan dirigidos a un destino común y desmembrar dicho mensaje una vez que llegue a su destino.

Cuando es inconveniente o costoso establecer una conexión separada para cada par de procesos de comunicación, la capa subyacente podría decidir utilizar la misma conexión para múltiples conversaciones sin relación entre sí. Siempre y cuando esta **multiplexión** y **desmultiplexión** se realice de manera transparente, cualquier capa la podrá utilizar. La multiplexión se necesita en la capa física, por ejemplo, donde múltiples conversaciones comparten un número limitado de circuitos físicos. Cuando hay múltiples rutas entre el origen y el destino, se debe elegir la mejor o las mejores entre todas ellas. A veces esta decisión se debe dividir en dos o más capas. Por ejemplo, para enviar datos de Londres a Roma, se debe tomar una decisión de alto nivel para pasar por Francia o Alemania, dependiendo de sus respectivas leyes de privacidad. Luego se debe tomar una decisión de bajo nivel para seleccionar uno de los circuitos disponibles dependiendo de la carga de tráfico actual. Este tema se llama **enrutamiento.**

1.3.3 Servicios orientados a la conexión y no orientados a la conexión

Las capas pueden ofrecer dos tipos de servicios a las capas que están sobre ellas: orientados a la conexión y no orientados a la conexión. En esta sección veremos estos dos tipos y examinaremos las diferencias que hay entre ellos.

El **servicio orientado a la conexión** se concibió con base en el sistema telefónico. Para hablar con alguien, usted levanta el teléfono, marca el número, habla y luego cuelga. Del mismo modo, para usar un servicio de red orientado a la conexión, el usuario del servicio primero establece una conexión, la utiliza y luego la abandona. El aspecto esencial de una conexión es que funciona como un tubo: el emisor empuja objetos (bits) en un extremo y el receptor los toma en el otro extremo. En la mayoría de los casos se conserva el orden para que los bits lleguen en el orden en que se enviaron.

En algunos casos, al establecer la conexión, el emisor, el receptor y la subred realizan una **negociación** sobre los parámetros que se van a utilizar, como el tamaño máximo del mensaje, la calidad del servicio solicitado y otros temas. Por lo general, un lado hace una propuesta y el otro la acepta, la rechaza o hace una contrapropuesta.

En contraste, el **servicio no orientado a la conexión** se concibió con base en el sistema postal. Cada mensaje (carta) lleva completa la dirección de destino y cada una se enruta a través del sistema, independientemente de las demás. En general, cuando se envían dos mensajes al mismo destino, el primero que se envíe será el primero en llegar. Sin embargo, es posible que el que se envió primero se dilate tanto que el segundo llegue primero.

Cada servicio se puede clasificar por la **calidad del servicio.** Algunos servicios son confiables en el sentido de que nunca pierden datos. Por lo general, en un servicio confiable el receptor confirma la recepción de cada mensaje para que el emisor esté seguro de que llegó. Este proceso de confirmación de recepción introduce sobrecargas y retardos, que con frecuencia son valiosos pero a veces son indeseables.

Una situación típica en la que un servicio orientado a la conexión es apropiado es en la transferencia de archivos. El propietario del archivo desea estar seguro de que lleguen correctamente todos los bits y en el mismo orden en que se enviaron. Muy pocos clientes que transfieren archivos preferirían un servicio que revuelve o pierde ocasionalmente algunos bits, aunque fuera mucho más rápido.

Un servicio orientado a la conexión confiable tiene dos variantes menores: secuencias de mensaje y flujo de bytes. En la primera variante se conservan los límites del mensaje. Cuando se envían dos mensajes de 1024 bytes, llegan en dos mensajes distintos de 1024 bytes, nunca en un solo mensaje de 2048 bytes. En la segunda, la conexión es simplemente un flujo de bytes, sin límites en el mensaje. Cuando llegan los 2048 bytes al receptor, no hay manera de saber si se enviaron como un mensaje de 2048 bytes o dos mensajes de 1024 bytes o 2048 mensajes de un byte. Si se envían las páginas de un libro en mensajes separados sobre una red a una fotocomponedora, podría ser importante que se conserven los límites de los mensajes. Por otra parte, cuando un usuario inicia sesión en un servidor remoto, todo lo que se necesita es un flujo de bytes desde la computadora del usuario al servidor. Los límites del mensaje no son importantes.

Como lo mencionamos antes, para algunas aplicaciones, los retardos de tránsito ocasionados por las confirmaciones de recepción son inaceptables. Una de estas aplicaciones es el tráfico de voz digitalizada. Es preferible para los usuarios de teléfono escuchar un poco de ruido en la línea de vez en cuando que experimentar un retardo esperando las confirmaciones de recepción. Del mismo modo, tener algunos píxeles erróneos cuando se transmite una videoconferencia no es problema, pero experimentar sacudidas en la imagen cuando se interrumpe el flujo para corregir errores es muy molesto.

No todas las aplicaciones requieren conexiones. Por ejemplo, conforme el correo electrónico se vuelve más común, la basura electrónica también se torna más común. Es probable que el emisor de correo electrónico basura no desee enfrentarse al problema de configurar una conexión y luego desarmarla sólo para enviar un elemento. Tampoco es 100 por ciento confiable enviar lo esencial, sobre todo si eso es más costoso. Todo lo que se necesita es una forma de enviar un mensaje único que tenga una alta, aunque no garantizada, probabilidad de llegar. Al servicio no orientado a la conexión no confiable (es decir, sin confirmación de recepción) se le conoce como **servicio de datagramas,** en analogía con el servicio de telegramas, que tampoco devuelve una confirmación de recepción al emisor.

En otras situaciones se desea la conveniencia de no tener que establecer una conexión para enviar un mensaje corto, pero la confiabilidad es esencial. Para estas aplicaciones se puede proporcionar el **servicio de datagramas confirmados.** Es como enviar una carta certificada y solicitar una confirmación de recepción. Cuando ésta regresa, el emisor está absolutamente seguro de que la carta se ha entregado a la parte destinada y no se ha perdido durante el trayecto.

Otro servicio más es el de **solicitud-respuesta.** En este servicio el emisor transmite un solo datagrama que contiene una solicitud; a continuación el servidor envía la respuesta. Por ejemplo, una solicitud a la biblioteca local preguntando dónde se habla uighur cae dentro de esta categoría. El esquema de solicitud-respuesta se usa comúnmente para implementar la comunicación en el modelo cliente-servidor: el cliente emite una solicitud y el servidor la responde. La figura 1-16 resume los tipos de servicios que se acaban de exponer.

	Servicio	Ejemplo
Orientado a la conexión	Flujo confiable de mensajes	Secuencia de páginas
	Flujo confiable de bytes	Inicio de sesión remoto
	Conexión no confiable	Voz digitalizada
No orientado a la conexión	Datagrama no confiable	Correo electrónico basura
	Datagrama confirmado	Correo certificado
	Solicitud-respuesta	Consulta de base de datos

Figura 1-16. Seis tipos de servicio diferentes.

El concepto del uso de la comunicación no confiable podría ser confuso al principio. Después de todo, en realidad, ¿por qué preferiría alguien la comunicación no confiable a la comunicación

confiable? Antes que nada, la comunicación confiable (en nuestro sentido, es decir, con confirmación de la recepción) podría no estar disponible. Por ejemplo, Ethernet no proporciona comunicación confiable. Ocasionalmente, los paquetes se pueden dañar en el tránsito. Toca al protocolo más alto enfrentar este problema. En segundo lugar, los retardos inherentes al servicio confiable podrían ser inaceptables, en particular para aplicaciones en tiempo real como multimedia. Éstas son las razones de que coexistan la comunicación no confiable y la confiable.

1.3.4 Primitivas de servicio

Un servicio se especifica formalmente como un conjunto de **primitivas** (operaciones) disponibles a un proceso de usuario para que acceda al servicio. Estas primitivas le indican al servicio que desempeñe alguna acción o reporte sobre una acción que ha tomado una entidad igual. Si la pila de protocolos se ubica en el sistema operativo, como suele suceder, por lo general las primitivas son llamadas al sistema. Estas llamadas provocan un salto al modo de *kernel*, que entonces cede el control de la máquina al sistema operativo para enviar los paquetes necesarios.

El conjunto de primitivas disponible depende de la naturaleza del servicio que se va a proporcionar. Las primitivas de servicio orientado a la conexión son diferentes de las del servicio no orientado a la conexión. Como un ejemplo mínimo de las primitivas para servicio que se podrían proporcionar para implementar un flujo de bytes confiable en un ambiente cliente-servidor, considere las primitivas listadas en la figura 1-17.

Primitiva	Significado
LISTEN	Bloquea en espera de una conexión entrante
CONNECT	Establece una conexión con el igual en espera
RECEIVE	Bloquea en espera de un mensaje entrante
SEND	Envía un mensaje al igual
DISCONNECT	Da por terminada una conexión

Figura 1-17. Cinco primitivas de servicio para la implementación de un servicio simple orientado a la conexión.

Estas primitivas se podrían usar como sigue. En primer lugar, el servidor ejecuta LISTEN para indicar que está preparado para aceptar las conexiones entrantes. Una manera común de implementar LISTEN es hacer que bloquee la llamada al sistema. Después de ejecutar la primitiva, el proceso del servidor se bloquea hasta que aparece una solicitud de conexión.

A continuación, el proceso del cliente ejecuta CONNECT para establecer una conexión con el servidor. La llamada CONNECT necesita especificar a quién conecta con quién, así que podría tener un parámetro que diera la dirección del servidor. El sistema operativo, en general, envía un paquete al igual solicitándole que se conecte, como se muestra en (1) en la figura 1-18. El proceso del cliente se suspende hasta que haya una respuesta. Cuando el paquete llega al servidor, es procesado ahí por el sistema operativo. Cuando el sistema ve que el paquete es una solicitud de

conexión, verifica si hay un escuchador. En ese caso hace dos cosas: desbloquea al escuchador y envía de vuelta una confirmación de recepción (2). La llegada de esta confirmación libera entonces al cliente. En este punto tanto el cliente como el servidor están en ejecución y tienen establecida una conexión. Es importante observar que la confirmación de recepción (2) es generada por el código del protocolo mismo, no en respuesta a una primitiva al nivel de usuario. Si llega una solicitud de conexión y no hay un escuchador, el resultado es indefinido. En algunos sistemas el paquete podría ser puesto en cola durante un breve tiempo en espera de un LISTEN.

La analogía obvia entre este protocolo y la vida real es un consumidor (cliente) que llama al gerente de servicios a clientes de una empresa. El gerente de servicios empieza por estar cerca del teléfono en caso de que éste suene. Entonces el cliente hace la llamada. Cuando el gerente levanta el teléfono se establece la conexión.

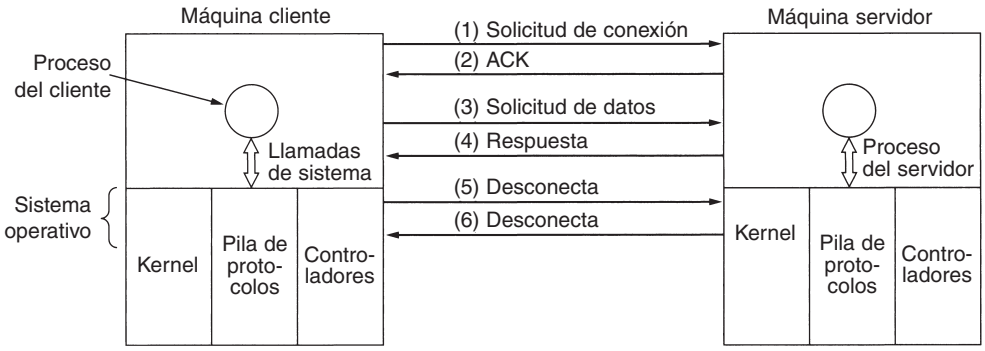

Figura 1-18. Paquetes enviados en una interacción simple cliente-servidor sobre una red orientada a la conexión.

El paso siguiente es que el servidor ejecute RECEIVE para prepararse para aceptar la primera solicitud. Normalmente, el servidor hace esto de inmediato en cuanto está libre de LISTEN, antes de que la confirmación de recepción pueda volver al cliente. La llamada RECEIVE bloquea al servidor.

Entonces el cliente ejecuta SEND para transmitir sus solicitudes (3) seguidas de la ejecución de RECEIVE para obtener la respuesta.

La llegada del paquete de solicitud a la máquina servidor desbloquea el proceso del servidor para que pueda procesar la solicitud. Una vez hecho su trabajo, utiliza SEND para devolver la respuesta al cliente (4). La llegada de este paquete desbloquea al cliente, que ahora puede revisar la respuesta. Si el cliente tiene solicitudes adicionales las puede hacer ahora. Si ha terminado, puede utilizar DISCONNECT para finalizar la conexión. Por lo común, un DISCONNECT inicial es una llamada de bloqueo, que suspende al cliente y envía un paquete al servidor en el cual le indica que ya no es necesaria la conexión (5). Cuando el servidor recibe el paquete también emite un DISCONNECT, enviando la confirmación de recepción al cliente y terminando la conexión. Cuando el paquete del servidor (6) llega a la máquina cliente, el proceso del cliente se libera y finaliza la conexión. En pocas palabras, ésta es la manera en que funciona la comunicación orientada a la conexión.

Desde luego, no todo es tan sencillo. Hay muchas cosas que pueden fallar. La temporización puede estar mal (por ejemplo, CONNECT se hace antes de LISTEN), se pueden perder paquetes, etcétera. Más adelante veremos en detalle estos temas, pero por el momento la figura 1-18 resume cómo podría funcionar la comunicación cliente-servidor en una red orientada a la conexión.

Dado que se requieren seis paquetes para completar este protocolo, cabría preguntarse por qué no se usa en su lugar un protocolo no orientado a la conexión. La respuesta es que en un mundo perfecto podría utilizarse, en cuyo caso bastarían dos paquetes: uno para la solicitud y otro para la respuesta. Sin embargo, en el caso de mensajes grandes en cualquier dirección (por ejemplo, en un archivo de megabytes), errores de transmisión y paquetes perdidos, la situación cambia. Si la respuesta constara de cientos de paquetes, algunos de los cuales se podrían perder durante la transmisión, ¿cómo sabría el cliente si se han perdido algunas piezas? ¿Cómo podría saber que el último paquete que recibió fue realmente el último que se envió? Suponga que el cliente esperaba un segundo archivo. ¿Cómo podría saber que el paquete 1 del segundo archivo de un paquete 1 perdido del primer archivo que de pronto apareció va en camino al cliente? Para abreviar, en el mundo real un simple protocolo de solicitud-respuesta en una red no confiable suele ser inadecuado. En el capítulo 3 estudiaremos en detalle una variedad de protocolos que soluciona éstos y otros problemas. Por el momento, baste decir que a veces es muy conveniente tener un flujo de bytes ordenado y confiable entre procesos.

1.3.5 Relación de servicios a protocolos

Servicios y protocolos son conceptos distintos, aunque con frecuencia se confunden. Sin embargo, esta distinción es tan importante que por esa razón ponemos énfasis de nuevo en ese punto. Un *servicio* es un conjunto de primitivas (operaciones) que una capa proporciona a la capa que está sobre ella. El servicio define qué operaciones puede realizar la capa en beneficio de sus usuarios, pero no dice nada de cómo se implementan tales operaciones. Un servicio está relacionado con la interfaz entre dos capas, donde la capa inferior es la que provee el servicio y la superior, quien lo recibe.

Un *protocolo,* en contraste, es un conjunto de reglas que rigen el formato y el significado de los paquetes, o mensajes, que se intercambiaron las entidades iguales en una capa. Las entidades utilizan protocolos para implementar sus definiciones del servicio. Son libres de cambiar sus protocolos cuando lo deseen, siempre y cuando no cambie el servicio visible a sus usuarios. De esta manera, el servicio y el protocolo no dependen uno del otro.

En otras palabras, los servicios se relacionan con las interacciones entre capas, como se ilustra en la figura 1-19. En contraste, los protocolos se relacionan con los paquetes enviados entre entidades iguales de máquinas diferentes. Es importante no confundir estos dos conceptos.

Vale la pena hacer una analogía con los lenguajes de programación. Un servicio es como un tipo de datos abstractos o un objeto en un lenguaje orientado a objetos. Define operaciones que se deben realizar en un objeto pero no especifica cómo se implementan estas operaciones. Un protocolo se relaciona con la *implementación* del servicio y, como tal, el usuario del servicio no puede verlo.

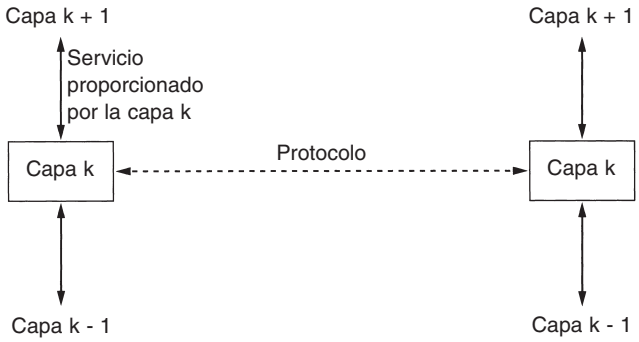

Figura 1-19. La relación entre un servicio y un protocolo.

Muchos protocolos antiguos no distinguían el servicio del protocolo. En efecto, una capa típica podría haber tenido una primitiva de servicio SEND PACKET y el usuario proveía un apuntador a un paquete ensamblado totalmente. Este arreglo significa que el usuario podía ver de inmediato todos los cambios del protocolo. En la actualidad, la mayoría de los diseñadores de redes señalan a este tipo de diseño como un error grave.

1.4 MODELOS DE REFERENCIA

Ahora que hemos visto en teoría las redes con capas, es hora de ver algunos ejemplos. En las dos secciones siguientes veremos dos arquitecturas de redes importantes: los modelos de referencia OSI y TCP/IP. Aunque los *protocolos* asociados con el modelo OSI ya casi no se usan, el *modelo* en sí es muy general y aún es válido, y las características tratadas en cada capa aún son muy importantes. El modelo TCP/IP tiene las propiedades opuestas: el modelo en sí no se utiliza mucho pero los protocolos sí. Por estas razones analizaremos con detalle ambos modelos. Además, a veces podemos aprender más de las fallas que de los aciertos.

1.4.1 El modelo de referencia OSI

El modelo OSI se muestra en la figura 1-20 (sin el medio físico). Este modelo está basado en una propuesta desarrollada por la ISO (Organización Internacional de Estándares) como un primer paso hacia la estandarización internacional de los protocolos utilizados en varias capas (Day y Zimmermann, 1983). Fue revisado en 1995 (Day, 1995). El modelo se llama **OSI** (**Interconexión de Sistemas Abiertos**) de **ISO** porque tiene que ver con la conexión de sistemas abiertos, es decir, sistemas que están abiertos a la comunicación con otros sistemas. Para abreviar, lo llamaremos modelo OSI.

El modelo OSI tiene siete capas. Podemos resumir brevemente los principios que se aplicaron para llegar a dichas capas:

1. Una capa se debe crear donde se necesite una abstracción diferente.
2. Cada capa debe realizar una función bien definida.
3. La función de cada capa se debe elegir con la intención de definir protocolos estandarizados internacionalmente.
4. Los límites de las capas se deben elegir a fin de minimizar el flujo de información a través de las interfaces.
5. La cantidad de capas debe ser suficientemente grande para no tener que agrupar funciones distintas en la misma capa y lo bastante pequeña para que la arquitectura no se vuelva inmanejable.

A continuación analizaremos una por una cada capa del modelo, comenzando con la capa inferior. Observe que el modelo OSI no es en sí una arquitectura de red, debido a que no especifica los servicios y protocolos exactos que se utilizarán en cada capa. Sólo indica lo que debe hacer cada capa. Sin embargo, ISO también ha producido estándares para todas las capas, aunque éstos no son parte del modelo de referencia mismo. Cada uno se ha publicado como un estándar internacional separado.

La capa física

En esta capa se lleva a cabo la transmisión de bits puros a través de un canal de comunicación. Los aspectos del diseño implican asegurarse de que cuando un lado envía un bit 1, éste se reciba en el otro lado como tal, no como bit 0. Las preguntas típicas aquí son: ¿cuántos voltios se deben emplear para representar un 1 y cuántos para representar un 0?, ¿cuántos nanosegundos dura un bit?, ¿la transmisión se debe llevar a cabo en ambas direcciones al mismo tiempo?, ¿cómo se establece la conexión inicial y cómo se finaliza cuando ambos lados terminan?, ¿cuántos pines tiene un conector de red y para qué se utiliza cada uno? Los aspectos de diseño tienen que ver mucho con interfaces mecánicas, eléctricas y de temporización, además del medio físico de transmisión, que está bajo la capa física.

La capa de enlace de datos

La tarea principal de esta capa es transformar un medio de transmisión puro en una línea de comunicación que, al llegar a la capa de red, aparezca libre de errores de transmisión. Logra esta tarea haciendo que el emisor fragmente los datos de entrada en **tramas de datos** (típicamente, de algunos cientos o miles de bytes) y transmitiendo las tramas de manera secuencial. Si el servicio es confiable, el receptor confirma la recepción correcta de cada trama devolviendo 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 hacer que un transmisor rápido no sature de datos a un receptor lento. Por lo general se necesita un mecanismo de regulación de tráfico que indique al transmisor cuánto espacio de búfer tiene el receptor en ese momento. Con frecuencia, esta regulación de flujo y el manejo de errores están integrados.

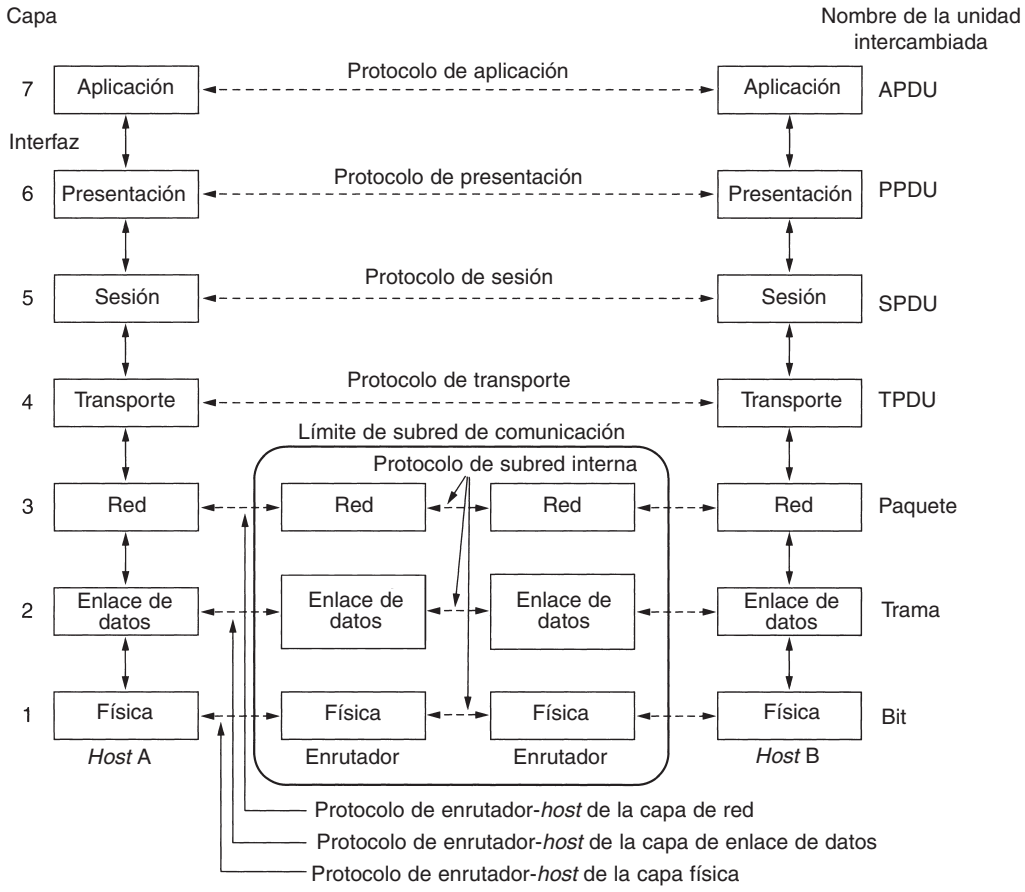

Figura 1-20. El modelo de referencia OSI.

Las redes de difusión tienen un aspecto 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, la subcapa de control de acceso al medio, se encarga de este problema.*

La capa de red

Esta capa controla las operaciones de la subred. Un aspecto clave del diseño es determinar cómo se enrutan los paquetes desde su origen a su destino. Las rutas pueden estar basadas en tablas estáticas (enrutamiento estático) codificadas en la red y que rara vez cambian.**

*En esta capa se define el direccionamiento físico, que permite a los *hosts* identificar las tramas destinadas a ellos. Este direccionamiento es único, identifica el hardware de red que se está usando y el fabricante, y no se puede cambiar. (N. del R.T.)

**En el enrutamiento estático la ruta que seguirán los paquetes hacia un destino particular es determinada por el administrador de la red. Las rutas también pueden determinarse cuando los enrutadores intercambian información de enrutamiento (enrutamiento dinámico). En este tipo de enrutamiento los enrutadores deciden la ruta que seguirán los paquetes hacia un destino sin la intervención del administrador de red. En el enrutamiento dinámico las rutas pueden cambiar para reflejar la topología o el estado de la red. (N. del R.T.)

Si hay demasiados paquetes en la subred al mismo tiempo, se interpondrán en el camino unos y otros, lo que provocará que se formen cuellos de botella. La responsabilidad de controlar esta congestión también pertenece a la capa de red, aunque esta responsabilidad también puede ser compartida por la capa de transmisión. De manera más general, la calidad del servicio proporcionado (retardo, tiempo de tránsito, inestabilidad, etcétera) también corresponde a la capa de red.

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 podría ser diferente del de la primera.* La segunda podría no aceptar todo el paquete porque es demasiado largo. Los protocolos podrían ser diferentes, etcétera. La capa de red tiene que resolver todos estos problemas para que las redes heterogéneas se interconecten.

En las redes de difusión, el problema de enrutamiento es simple, por lo que la capa de red a veces es delgada o, en ocasiones, ni siquiera existe.

La capa de transporte

La función básica de esta capa es aceptar los datos provenientes de las capas superiores, dividirlos en unidades más pequeñas si es necesario, pasar éstas a la capa de red y asegurarse de que todas las piezas lleguen correctamente al otro extremo. Además, todo esto se debe hacer con eficiencia y de manera que aísle a las capas superiores de los cambios inevitables en la tecnología del hardware.

La capa de transporte también determina qué tipo de servicio proporcionar a la capa de sesión y, finalmente, a los usuarios de la red. El tipo de conexión de transporte más popular es un canal punto a punto libre de errores que entrega mensajes o bytes en el orden en que se enviaron. Sin embargo, otros tipos de servicio de transporte posibles son la transportación de mensajes aislados, que no garantiza el orden de entrega, y la difusión de mensajes a múltiples destinos. El tipo de servicio se determina cuando se establece la conexión. (Como observación, es imposible alcanzar un canal libre de errores; lo que se quiere dar a entender con este término es que la tasa de error es tan baja que se puede ignorar en la práctica.)

La capa de transporte es una verdadera conexión de extremo a extremo, en 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, usando los encabezados de mensaje y los mensajes de control. En las capas inferiores, los protocolos operan entre cada máquina y sus vecinos inmediatos, y no entre las máquinas de los extremos, la de origen y la de destino, las cuales podrían estar separadas por muchos enrutadores. En la figura 1-20 se muestra la diferencia entre las capas 1 a 3, que están encadenadas, y las capas 4 a 7, que operan de extremo a extremo.

La capa de sesión

Esta capa permite que los usuarios de máquinas diferentes establezcan **sesiones** entre ellos. Las sesiones ofrecen varios servicios, como el **control de diálogo** (dar seguimiento de a quién le toca

*El direccionamiento usado en esta capa es un direccionamiento lógico, diferente al direccionamiento físico empleado en la capa de enlace de datos. Este direccionamiento lógico permite que una interfaz o puerto pueda tener más de una dirección de capa de red. (N. del R.T.)

transmitir), **administración de token** (que impide que las dos partes traten de realizar la misma operación crítica al mismo tiempo) y **sincronización** (la adición de puntos de referencia a transmisiones largas para permitirles continuar desde donde se encontraban después de una caída).

La capa de presentación

A diferencia de las capas inferiores, a las que les corresponde principalmente mover bits, a la **capa de presentación** le corresponde la sintaxis y la semántica de la información transmitida. A fin de que las computadoras con diferentes representaciones de datos se puedan comunicar, las estructuras de datos que se intercambiarán se pueden definir de una manera abstracta, junto con una codificación estándar para su uso "en el cable". La capa de presentación maneja estas estructuras de datos abstractas y permite definir e intercambiar estructuras de datos de un nivel más alto (por ejemplo, registros bancarios).

La capa de aplicación

Esta capa contiene varios protocolos que los usuarios requieren con frecuencia. Un protocolo de aplicación de amplio uso es **HTTP** (**Protocolo de Transferencia de Hipertexto**), que es la base de World Wide Web. Cuando un navegador desea una página Web, utiliza este protocolo para enviar al servidor el nombre de dicha página. A continuación, el servidor devuelve la página. Otros protocolos de aplicación se utilizan para la transferencia de archivos, correo electrónico y noticias en la red.

1.4.2 El modelo de referencia TCP/IP

Tratemos ahora el modelo de referencia usado en la abuela de todas las redes de computadoras de área amplia, ARPANET, y en su sucesora, la Internet mundial. Aunque daremos más adelante una breve historia de ARPANET, es útil mencionar algunos de sus aspectos ahora. ARPANET fue una red de investigación respaldada por el DoD (Departamento de Defensa de Estados Unidos). Con el tiempo, conectó cientos de universidades e instalaciones gubernamentales mediante líneas telefónicas alquiladas. Posteriormente, cuando se agregaron redes satelitales y de radio, los protocolos existentes tuvieron problemas para interactuar con ellas, por lo que se necesitaba una nueva arquitectura de referencia. De este modo, la capacidad para conectar múltiples redes en una manera sólida fue una de las principales metas de diseño desde sus inicios. Más tarde, esta arquitectura se llegó a conocer como el **modelo de referencia TCP/IP,** de acuerdo con sus dos protocolos primarios. Su primera definición fue en (Cerf y Kahn, 1974). Posteriormente se definió en (Leiner y cols., 1985). La filosofía del diseño que respalda al modelo se explica en (Clark, 1988).

Ante el temor del DoD de que algunos de sus valiosos *hosts*, enrutadores y puertas de enlace de interredes explotaran en un instante, otro objetivo fue que la red pudiera sobrevivir a la pérdida de hardware de la subred, sin que las conversaciones existentes se interrumpieran. En otras palabras, el DoD quería que las conexiones se mantuvieran intactas en tanto las máquinas de origen

y destino estuvieran funcionando, aunque algunas de las máquinas o líneas de transmisión intermedias quedaran fuera de operación repentinamente. Además, se necesitaba una arquitectura flexible debido a que se preveían aplicaciones con requerimientos divergentes, desde transferencia de archivos a transmisión de palabras en tiempo real.

La capa de interred

Todos estos requerimientos condujeron a la elección de una red de conmutación de paquetes basada en una capa de interred no orientada a la conexión. Esta capa, llamada **capa de interred,** es la pieza clave que mantiene unida a la arquitectura. Su trabajo es permitir que los *hosts* inyecten paquetes dentro de cualquier red y que éstos viajen a su destino de manera independiente (podría ser en una red diferente). Tal vez lleguen en un orden diferente al que fueron enviados, en cuyo caso las capas más altas deberán ordenarlos, si se desea una entrega ordenada. Observe que aquí el concepto "interred" se utiliza en un sentido genérico, aun cuando esta capa se presente en Internet.

Aquí la analogía es con el sistema de correo tradicional. Una persona puede depositar una secuencia de cartas internacionales en un buzón y, con un poco de suerte, la mayoría de ellas se entregará en la dirección correcta del país de destino. Es probable que durante el trayecto, las cartas viajen a través de una o más puertas de enlace de correo internacional, pero esto es transparente para los usuarios. Además, para los usuarios también es transparente el hecho de que cada país (es decir, cada red) tiene sus propios timbres postales, tamaños preferidos de sobre y reglas de entrega.

La capa de interred define un paquete de formato y protocolo oficial llamado **IP** (**Protocolo de Internet**)**.** El trabajo de la capa de interred es entregar paquetes IP al destinatario. Aquí, el enrutamiento de paquetes es claramente el aspecto principal, con el propósito de evitar la congestión. Por estas razones es razonable decir que la capa de interred del modelo TCP/IP es similar en funcionalidad a la capa de red del modelo OSI. La figura 1-21 muestra esta correspondencia.

La capa de transporte

La capa que está arriba de la capa de interred en el modelo TCP/IP se llama **capa de transporte.** Está diseñada para permitir que las entidades iguales en los *hosts* de origen y destino puedan llevar a cabo una conversación, tal como lo hace la capa de transporte OSI. Aquí se han definido dos protocolos de transporte de extremo a extremo. El primero, **TCP** (**Protocolo de Control de Transmisión**)**,** es un protocolo confiable, orientado a la conexión, que permite que un flujo de bytes que se origina en una máquina se entregue sin errores en cualquier otra máquina en la interred. Divide el flujo de bytes entrantes en mensajes discretos y pasa cada uno de ellos a la capa de interred. En el destino, el proceso TCP receptor reensambla en el flujo de salida los mensajes recibidos. TCP también maneja el control de flujo para asegurarse de que un emisor rápido no sature a un receptor lento con más mensajes de los que puede manejar.

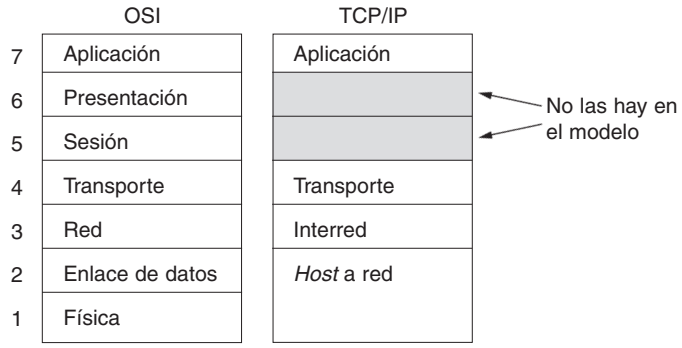

Figura 1-21. El modelo de referencia TCP/IP.

El segundo protocolo de esta capa, **UDP (Protocolo de Datagrama de Usuario),** es un protocolo no confiable y no orientado a la conexión para aplicaciones que no desean la secuenciación o el control de flujo de TCP y que desean proporcionar el suyo. También tiene un amplio uso en consultas únicas de solicitud-respuesta de tipo cliente-servidor en un solo envío, así como aplicaciones en las que la entrega puntual es más importante que la precisa, como en la transmisión de voz o vídeo. La relación de IP, TCP y UDP se muestra en la figura 1-22. Puesto que el modelo se desarrolló, se ha implementado IP en muchas otras redes.

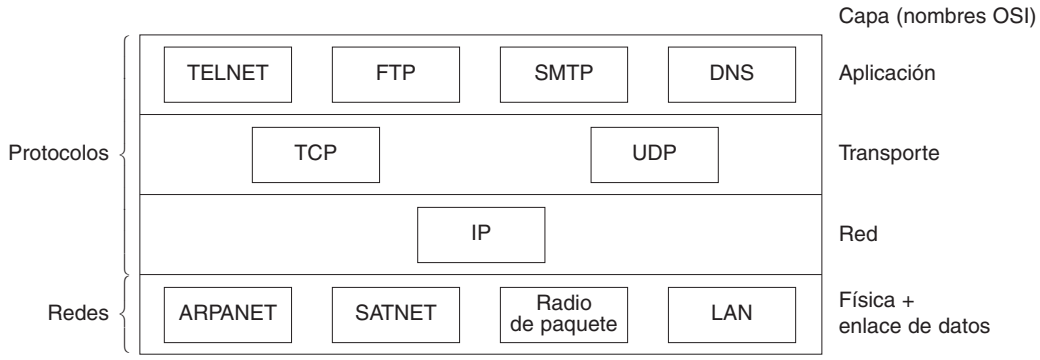

Figura 1-22. Protocolos y redes en el modelo TCP/IP inicialmente.

La capa de aplicación

El modelo TCP/IP no tiene capas de sesión ni de presentación. No se han necesitado, por lo que no se incluyen. La experiencia con el modelo OSI ha probado que este punto de vista es correcto: son de poco uso para la mayoría de las aplicaciones.

Arriba de la capa de transporte está la **capa de aplicación.** Contiene todos los protocolos de nivel más alto. Los primeros incluyeron una terminal virtual (TELNET), transferencia de archivos (FTP) y correo electrónico (SMTP), como se muestra en la figura 1-22. El protocolo de terminal virtual permite que un usuario en una máquina se registre en una máquina remota y trabaje ahí. El protocolo de transferencia de archivos proporciona una manera de mover con eficiencia datos de una máquina a otra. El correo electrónico era originalmente sólo un tipo de transferencia de archivos, pero más tarde se desarrolló un protocolo especializado (SMTP) para él. Con el tiempo, se han agregado muchos otros protocolos: DNS (Sistema de Nombres de Dominio) para la resolución de nombres de *host* en sus direcciones de red; NNTP, para transportar los artículos de noticias de USENET; HTTP, para las páginas de World Wide Web, y muchos otros.

La capa *host* a red

Debajo de la capa de interred hay un gran vacío. El modelo de referencia TCP/IP en realidad no dice mucho acerca de lo que pasa aquí, excepto que puntualiza que el *host* se tiene que conectar a la red mediante el mismo protocolo para que le puedan enviar paquetes IP. Este protocolo no está definido y varía de un *host* a otro y de una red a otra. Este tema rara vez se trata en libros y artículos sobre TCP/IP.

1.4.3 Comparación entre los modelos de referencia OSI y TCP/IP

Los modelos de referencia OSI y TCP/IP tienen mucho en común. Los dos se basan en el concepto de una pila de protocolos independientes. Asimismo, la funcionalidad de las capas es muy parecida. Por ejemplo, en ambos modelos las capas que están arriba de, incluyendo a, la capa de transporte están ahí para proporcionar un servicio de transporte independiente de extremo a extremo a los procesos que desean comunicarse. Estas capas forman el proveedor de transporte. De nuevo, 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 estos dos modelos de referencia. Es importante tener en cuenta que estamos comparando los *modelos de referencia,* no las *pilas de protocolos* correspondientes. Más adelante explicaremos los protocolos. Si desea un libro dedicado a comparar y contrastar TCP/IP y OSI, vea (Piscitello y Chapin, 1993).

Tres conceptos son básicos para el modelo OSI:

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

Probablemente la contribución más grande del modelo OSI es que hace explícita la distinción entre estos tres conceptos. Cada capa desempeña algunos servicios para la capa que está arriba de ella. La definición de *servicio* indica qué hace la capa, no la forma en que la entidad superior tiene acceso a ella, o cómo funciona dicha capa. Define el aspecto semántico de la capa.

La *interfaz* de una capa indica a los procesos que están sobre ella cómo accederla. Especifica cuáles son los parámetros y qué resultados se esperan. Incluso, no dice nada sobre cómo funciona internamente la capa.

Por último, una capa es quien debe decidir qué *protocolos* de iguales utilizar. Puede usar cualesquier protocolos que desee, en tanto consiga que se haga el trabajo (es decir, proporcione los servicios ofrecidos). También puede cambiarlos cuando desee sin afectar el software de las capas superiores.

Estas ideas encajan muy bien con las ideas modernas sobre la programación orientada a objetos. Un objeto, como una capa, cuenta con un conjunto de métodos (operaciones) que pueden ser invocados por procesos que no estén en dicho objeto. 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 es visible o no tiene importancia fuera del objeto.

Originalmente, el modelo TCP/IP no distinguía entre servicio, interfaz y protocolo, aunque las personas han tratado de readaptarlo con el propósito de hacerlo más parecido al OSI. Por ejemplo, los únicos servicios ofrecidos realmente por la capa de interred son SEND IP PACKET y RECEIVE IP PACKET.

Como consecuencia, los protocolos del modelo OSI están mejor ocultos que los del modelo TCPI/IP y se pueden reemplazar fácilmente conforme cambia la tecnología. La facilidad para realizar tales cambios es uno de los objetivos principales de tener protocolos en capas.

El modelo de referencia OSI se vislumbró *antes* de que se inventaran los protocolos correspondientes. Esta clasificación significa que el modelo no estaba diseñado para un conjunto particular de protocolos, un hecho que lo hizo general. Una deficiencia de esta clasificación es que los diseñadores no tenían mucha experiencia con el asunto y no tenían una idea concreta de qué funcionalidad poner en qué capa.

Por ejemplo, originalmente la capa de enlace de datos sólo trataba con redes de punto a punto. Cuando llegaron las redes de difusión, se tuvo que extender una nueva subcapa en el modelo. Cuando las personas empezaron a construir redes reales utilizando el modelo OSI y los protocolos existentes, se descubrió que estas redes no coincidían con las especificaciones de los servicios solicitados (maravilla de maravillas), por lo que se tuvieron que integrar subcapas convergentes en el modelo para proporcionar un espacio para documentar las diferencias. Por último, el comité esperaba en un principio que cada país tuviera una red, controlada por el gobierno y que utilizara los protocolos OSI, pero nunca pensaron en la interconectividad de redes. Para no hacer tan larga la historia, las cosas no sucedieron como se esperaba.

Con TCP/IP sucedió lo contrario: los protocolos llegaron primero y el modelo fue en realidad una descripción de los protocolos existentes. No había problemas para ajustar los protocolos al modelo. Encajaban a la perfección. El único problema era que el *modelo* no aceptaba otras pilas de protocolos. Como consecuencia, no era útil para describir otras redes que no fueran TCP/IP.

Volviendo de los asuntos filosóficos a los más específicos, una diferencia patente entre los dos modelos es el número de capas: el modelo OSI tiene siete y el TCP/IP sólo cuatro. Los dos tienen capas de (inter)red, transporte y aplicación, pero las otras capas son diferentes.

Otra diferencia está en el área de la comunicación orientada a la conexión comparada con la no orientada a la conexión. El modelo OSI soporta ambas comunicaciones en la capa de red, pero sólo la de comunicación orientada a la conexión en la capa de transporte, donde es importante (porque el servicio de transporte es transparente para los usuarios). El modelo TCP/IP sólo tiene un modo en la capa de red (no orientado a la conexión) pero soporta ambos modos en la capa de transporte, lo que da a los usuarios la oportunidad de elegir. Esta elección es importante especialmente para protocolos sencillos de solicitud-respuesta.

1.4.4 Crítica al modelo OSI y los protocolos

Ni el modelo OSI y sus protocolos ni el modelo TCP/IP y sus protocolos son perfectos. Se les pueden hacer, y se les han hecho, críticas. En ésta y en la siguiente sección veremos algunas de estas críticas. Empezaremos con el modelo OSI y más adelante examinaremos el modelo TCP/IP.

En la época en la que 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 dominar el mundo y a desplazar a todos los demás. Eso no sucedió. ¿Por qué? Sería útil echar un vistazo a algunas lecciones. Éstas se pueden resumir así:

1. Aparición inoportuna.
2. Mala tecnología.
3. Malas implementaciones.
4. Malas políticas.

Aparición inoportuna

Primero veamos la razón número uno: aparición inoportuna. El tiempo en que se establece un estándar es absolutamente crítico para el éxito. David Clark, del M.I.T., tiene una teoría de estándares que llama *apocalipsis de los dos elefantes,* la cual se ilustra en la figura 1-23.

Esta figura muestra la cantidad de actividad que rodea a un sujeto nuevo. Cuando se descubre primero el sujeto, hay una explosión de actividad de investigación en forma de exposiciones, documentos y reuniones. Después de un tiempo esta actividad disminuye, las empresas descubren el sujeto y surge la ola de miles de millones de dólares de inversión.

Es esencial que los estándares se escriban en el punto intermedio entre los dos "elefantes". Si los estándares se escriben demasiado pronto, antes de que se termine la investigación, el tema podría no estar entendido por completo; el resultado son malos estándares. Si se escriben demasiado tarde, varias empresas podrían haber hecho ya inversiones importantes en diversas maneras de hacer las cosas que los estándares han ignorado. Si el intervalo entre los dos elefantes es muy corto (porque cada cual tiene prisa por empezar), las personas que están desarrollando los estándares podrían fracasar.

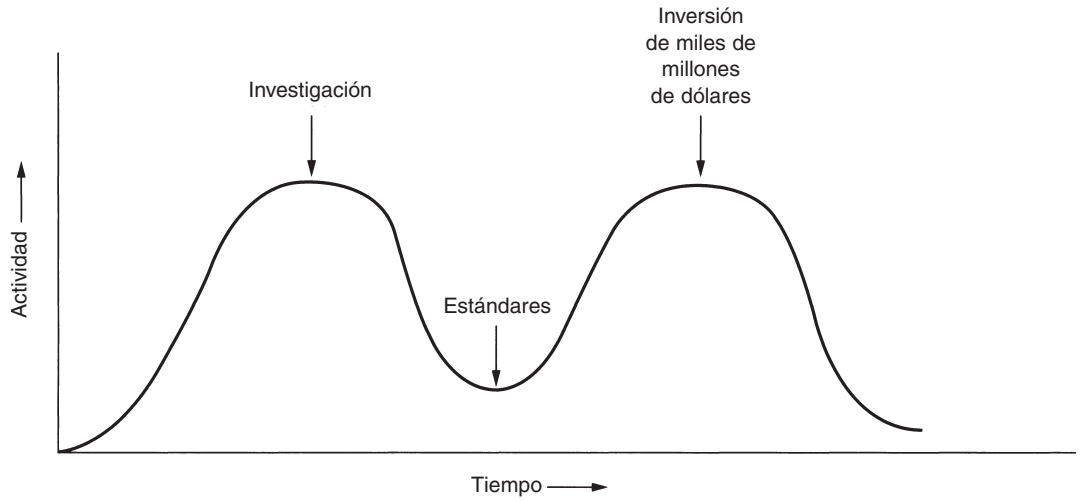

Figura 1-23. El apocalipsis de los dos elefantes.

Al parecer, los protocolos OSI estándar han sido vencidos. Los protocolos TCP/IP competidores ya eran ampliamente utilizados por las universidades investigadoras al momento en que aparecieron los protocolos OSI. Mientras la ola de los miles de millones de inversión aún no golpeaba, el mercado académico era bastante grande para que los proveedores empezaran a hacer ofertas cautas de los productos TCP/IP. Cuando OSI llegó, no quisieron soportar una segunda pila de protocolos hasta que se vieran forzados, por lo que no hubo ofertas iniciales. Puesto que cada empresa esperaba que la otra diera el primer paso, ninguna lo hizo y OSI nunca prosperó.

Mala tecnología

La segunda razón por la que OSI no tuvo éxito es que tanto el modelo como los protocolos tienen defectos. La elección de las siete capas fue más política que técnica, y dos de las capas (la de sesión y la de presentación) están casi vacías, mientras que las otras dos (la de enlace de datos y la de red) están saturadas.

El modelo OSI, junto con el servicio asociado de definiciones y protocolos, es extraordinariamente complejo. Si se apilan, los estándares impresos ocupan una fracción importante de un metro de papel. Incluso son difíciles de implementar y de operación deficiente. En este contexto, nos viene a la mente un enigma propuesto por Paul Mockapetris y citado en (Rose, 1993):

P: ¿Qué obtiene cuando cruza un gángster con un estándar internacional?
R: Alguien que le hace una oferta que usted no entiende.

Además de ser incomprensible, otro problema con OSI es que algunas funciones, como direccionamiento, control de flujo y control de errores, reaparecen una y otra vez en cada capa. Por

ejemplo, Saltzer y cols. (1984) han apuntado que para ser efectivo el control de errores, se debe hacer en la capa superior, puesto que repetirlo una y otra vez en cada una de las capas inferiores suele ser innecesario e ineficaz.

Malas implementaciones

Ante la enorme complejidad del modelo y los protocolos, no es de sorprender que las implementaciones iniciales fueran grandes, pesadas y lentas. Todos los que lo intentaron fracasaron. No le tomó mucho tiempo a las personas asociar OSI con "baja calidad". Aunque los productos mejoraron con el paso del tiempo, la imagen persistió.

En contraste, una de las primeras implementaciones de TCP/IP era parte de UNIX de Berkeley y fue bastante buena (sin mencionar que era gratis). Las personas pronto empezaron a utilizarla, lo que la llevó a un uso mayor por la comunidad, y esto a su vez condujo a mejoras que la llevaron a un mayor uso en la comunidad. Aquí la espiral fue ascendente en vez de descendente.

Malas políticas

A causa de la implementación inicial, muchas personas, sobre todo en el nivel académico, pensaban que TCP/IP era parte de UNIX, y en la década de 1980, UNIX no parecía tener paternidad alguna en la universidad.

Por otra parte, se tenía la idea de que OSI sería la criatura de los ministerios de telecomunicación de Europa, de la comunidad europea y más tarde del gobierno de los Estados Unidos. Esta creencia era cierta en parte, pero no ayudaba mucho la idea de un manojo de burócratas gubernamentales intentando poner en marcha un estándar técnicamente inferior al mando de los investigadores y programadores pobres que estaban en las trincheras desarrollando realmente redes de computadoras. Algunas personas compararon este desarrollo con la ocasión en que IBM anunció, en la década de 1960, que PL/I era el lenguaje del futuro, o cuando más tarde el DoD corregía esto anunciando que en realidad era Ada.

1.4.5 Crítica del modelo de referencia TCP/IP

El modelo de referencia TCP/IP y los protocolos también tienen sus problemas. En primer lugar, el modelo no distingue claramente los conceptos de servicio, interfaz y protocolo. Una buena ingeniería de software requiere la diferenciación entre la especificación y la implementación, algo que OSI hace con mucho cuidado y que TCP/IP no hace. En consecuencia, el modelo TCP/IP no es una guía para diseñar redes nuevas mediante tecnologías nuevas.

En segundo lugar, el modelo TCP/IP no es general del todo y no está bien ajustado para describir ninguna pila de protocolos más que de TCP/IP. Por ejemplo, es completamente imposible tratar de utilizar el modelo TCP/IP para describir Bluetooth.

En tercer lugar, la capa *host* a red no es en realidad una capa del todo en el sentido normal del término, como se utiliza en el contexto de los protocolos de capas. Es una interfaz (entre la capa de red y la de enlace de datos). La distinción entre una interfaz y una capa es crucial y nadie debe ser descuidado al respecto.

En cuarto lugar, el modelo TCP/IP no distingue (ni menciona) las capas física y de enlace de datos. Son completamente diferentes. La capa física tiene que ver con las características de transmisión de comunicación por cable de cobre, por fibra óptica e inalámbrica. El trabajo de la capa de enlace de datos es delimitar el inicio y fin de las tramas y captarlas de uno a otro lado con el grado deseado de confiabilidad. Un modelo adecuado debería incluir ambas como capas separadas. El modelo TCP/IP no hace esto.

Por último, aunque los protocolos IP y TCP se idearon e implementaron con sumo cuidado, muchos de los demás protocolos fueron hechos con fines específicos, producidos por lo general por estudiantes de licenciatura que los mejoraban hasta que se aburrían. Posteriormente, las implementaciones de tales protocolos se distribuyeron de manera gratuita, lo que dio como resultado un uso amplio y profundo y, por lo tanto, que fueran difíciles de reemplazar. Algunos de ellos ahora están en apuros. Por ejemplo, el protocolo de terminal virtual, TELNET, se diseñó para una terminal de teletipo mecánica de 10 caracteres por segundo. No sabe nada de interfaces gráficas de usuario ni de ratones. No obstante, 25 años más tarde aún tiene un amplio uso.

En resumen, a pesar de sus problemas, el *modelo* OSI (excepto las capas de sesión y presentación) ha probado ser excepcionalmente útil en la exposición de redes de computadoras. En contraste, los *protocolos* OSI no han sido muy populares. Sucede lo contrario con TCP/IP: el *modelo* es prácticamente inexistente, pero los *protocolos* tienen un amplio uso. En este libro utilizaremos un modelo OSI modificado pero nos concentraremos principalmente en el modelo TCP/IP y los protocolos relacionados, así como en los novísimos 802, SONET y Bluetooth. En efecto, utilizaremos el modelo híbrido de la figura 1-24 como marco de trabajo para este libro.

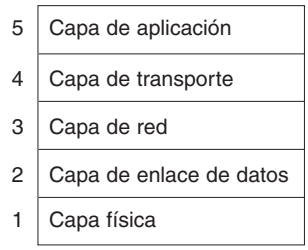

Figura 1-24. Modelo de referencia híbrido que se usará en este libro.

1.5 REDES DE EJEMPLO

El tema de las redes de computadoras cubre muchos y diversos tipos de redes, grandes y pequeñas, bien conocidas y no tan bien conocidas. Tiene diferentes objetivos, escalamientos y tecnologías. En las siguientes secciones veremos algunos ejemplos para tener una idea de la variedad que se puede encontrar en el área de la conectividad de redes.

Empezaremos con Internet, que es probablemente la red más conocida y veremos su historia, evolución y tecnología. Luego consideraremos ATM, cuyo uso es frecuente en el núcleo de redes (telefónicas) grandes. Desde el punto de vista técnico difiere muy poco de Internet, y contrasta gratamente. Después presentaremos Ethernet, la red de área local dominante. Y, por último, veremos el IEEE 802.11, el estándar para las LANs inalámbricas.

1.5.1 Internet

Internet no es del todo una red, sino un inmenso conjunto de redes diferentes que usan ciertos protocolos comunes y proporcionan ciertos servicios comunes. Es un sistema poco común porque nadie lo planeó y nadie lo controla. Para entenderlo mejor, empecemos desde el principio y veamos cómo se desarrolló y por qué. Si desea leer una historia maravillosa sobre Internet, recomendamos ampliamente el libro de John Naughton (2000). Es uno de esos raros libros cuya lectura no sólo es divertida, sino que también contiene 20 páginas de *ibídems* y *op. cits*. para el historiador serio. Parte del material que se muestra a continuación se basa en dicho libro.

Desde luego, se ha escrito una infinidad de libros técnicos sobre Internet y sus protocolos. Para más información, vea (Maufer, 1999).

ARPANET

Nuestro relato empieza a fines de la década de 1950. Durante el auge de la Guerra Fría, el DoD quería una red de control y comando que pudiera sobrevivir a una guerra nuclear. En esa época todas las comunicaciones militares usaban la red telefónica pública, que se consideraba vulnerable. La razón de esta creencia se puede entresacar de la figura 1-25(a). Los puntos negros representan las oficinas de conmutación telefónica, a cada una de las cuales se conectaban miles de teléfonos. Estas oficinas de conmutación estaban, a su vez, conectadas a oficinas de conmutación de más alto nivel (oficinas interurbanas), para conformar una jerarquía nacional con sólo una mínima redundancia. La vulnerabilidad del sistema estaba en que la destrucción de algunas de las oficinas interurbanas clave podía fragmentar el sistema en muchas islas incomunicadas.

Hacia 1960, el DoD firmó un contrato con RAND Corporation para encontrar una solución. Uno de sus empleados, Paul Baran, presentó el diseño de amplia distribución y tolerancia a fallas que se muestra en la figura 1-25(b). Puesto que las trayectorias entre cualquiera de las oficinas de conmutación eran ahora más grandes de lo que las señales análogas podían viajar sin distorsión, Baran propuso que se utilizara la tecnología digital de conmutación de paquetes a través del sistema. Baran escribió varios informes al DoD describiendo en detalle sus ideas. A los oficiales del Pentágono les agradó el concepto y pidieron a AT&T, en ese entonces el monopolio telefónico estadounidense, que construyera un prototipo. AT&T desechó las ideas de Baran. La corporación más grande y rica del mundo no iba a permitir que un jovenzuelo le dijera cómo construir un sistema telefónico. Dijeron que la red de Baran no se podía construir y la idea se desechó.

Pasaron varios años y el DoD aún no tenía un mejor sistema de control y comando. Para entender qué sucedió a continuación, tenemos que volver al 7 de octubre de 1957, cuando la Unión soviética lanzó el Sputnik, su primer satélite artificial, con lo cual se le adelantó a Estados Unidos.

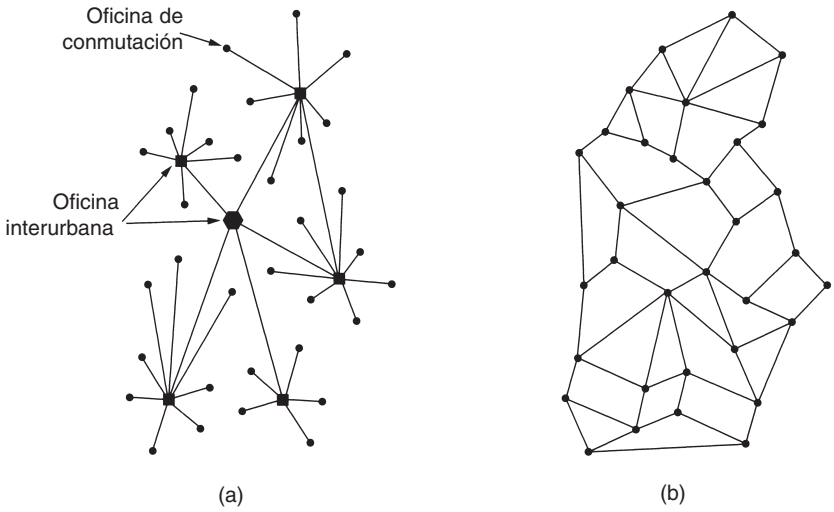

Figura 1-25. (a) Estructura de un sistema telefónico. (b) Sistema de conmutación distribuida propuesto por Baran.

Cuando el presidente Eisenhower trató de encontrar quién estaba dormido en sus laureles, se espantó al encontrarse con que la armada, el ejército y la fuerza aérea se peleaban por el presupuesto de investigación del Pentágono. Su respuesta inmediata fue crear una organización única de investigación para la defensa, **ARPA** (**Agencia de Proyectos de Investigación Avanzada**)**.** Ésta no tenía científicos ni laboratorios; de hecho, no tenía más que una oficina y un presupuesto pequeño (por normas del Pentágono). Hacía su trabajo otorgando subvenciones y contratos a universidades y empresas cuyas ideas le parecían prometedoras.

Durante los primeros años, ARPA trataba de imaginarse cuál sería su misión, pero en 1967 la atención de su entonces director, Larry Roberts, se volvió hacia las redes. Se puso en contacto con varios expertos para decidir qué hacer. Uno de ellos, Wesley Clark, sugirió la construcción de una subred de conmutación de paquetes, dando a cada *host* su propio enrutador, como se ilustra en la figura 1-10.

Después del escepticismo inicial, Roberts aceptó la idea y presentó un documento algo vago al respecto en el Simposio sobre Principios de Sistemas Operativos ACM SIGOPS que se llevó a cabo en Gatlinburg, Tennessee, a fines de 1967 (Roberts, 1967). Para mayor sorpresa de Roberts, otro documento en la conferencia describía un sistema similar que no sólo había sido diseñado, sino que ya estaba implementado bajo la dirección de Donald Davies en el National Physical Laboratory en Inglaterra. El sistema del NPL no era un sistema a nivel nacional (sólo conectaba algunas computadoras en el campus del NPL), pero demostró que era posible hacer que la conmutación de paquetes funcionara. Además, citaba el trabajo anterior de Baran, el cual había sido descartado. Roberts salió de Gatlinburg determinado a construir lo que más tarde se conocería como **ARPANET.**

La subred constaría de minicomputadoras llamadas **IMPs** (**Procesadores de Mensajes de Interfaz**), conectadas por líneas de transmisión de 56 kbps. Para alta confiabilidad, cada IMP estaría conectado al menos a otros dos IMPs. La subred iba a ser de datagramas, de manera que si se destruían algunos IMPs, los mensajes se podrían volver a enrutar de manera automática a otras rutas alternativas.

Cada nodo de la red iba a constar de un IMP y un *host*, en el mismo cuarto, conectados por un cable corto. Un *host* tendría la capacidad de enviar mensajes de más de 8063 bits a su IMP, el cual los fragmentaría en paquetes de, a lo sumo, 1008 bits y los reenviaría de manera independiente hacia el destino. Cada paquete se recibiría íntegro antes de ser reenviado, por lo que la subred sería la primera red electrónica de conmutación de paquetes de almacenamiento y reenvío.

Entonces ARPA lanzó una convocatoria para construir la subred. Doce empresas licitaron. Después de evaluar las propuestas, ARPA seleccionó a BBN, una empresa de consultoría de Cambridge, Massachusetts, y en diciembre de 1968 le otorgó el contrato para construir la subred y escribir el software de ésta. BBN eligió utilizar como IMPs minicomputadoras Honeywell DDP-316 especialmente modificadas con palabras de 16 bits y 12 KB de memoria central. Los IMPs no tenían discos, ya que las partes móviles se consideraban no confiables. Estaban interconectadas por líneas de 56 kbps alquiladas a las compañías telefónicas. Aunque 56 kbps ahora es la elección de las personas que no pueden permitirse ADSL o cable, entonces era la mejor opción.

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

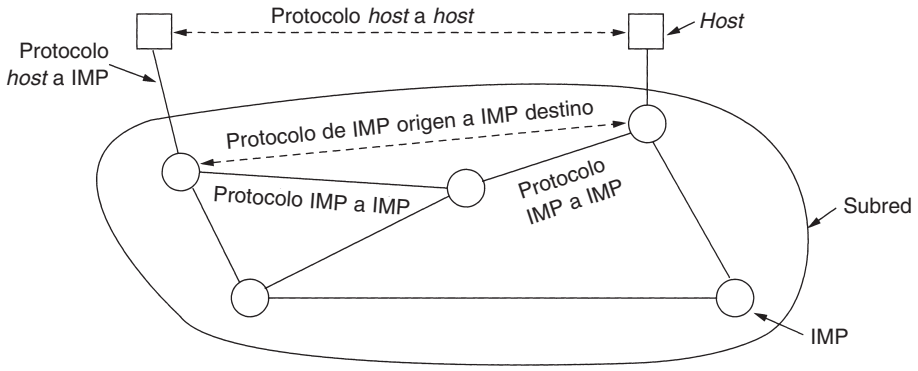

Figura 1-26. Diseño original de ARPANET.

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 sintió que cuando se aceptaba un mensaje en un cable *host* a IMP y se ponía en un cable *host* a IMP en el destino, el trabajo estaba hecho.

Roberts tenía un problema: los *hosts* también necesitaban software. Para resolverlo convocó a una reunión de investigadores de red —en su mayoría estudiantes de licenciatura de Snowbird, Utah— durante el verano de 1969. Los estudiantes esperaban que algún experto en redes les explicara el gran diseño de la red y su software y que luego les asignara el trabajo de escribir parte de él. Se quedaron asombrados al descubrir que no había ningún experto ni un gran diseño. Tenían que averiguar qué era lo que se tenía que hacer.

No obstante, en diciembre de 1969 de alguna manera surgió una red experimental con cuatro nodos: en UCLA, UCSB, SRI y la Universidad de Utah. Se eligieron estas cuatro porque todas tenían un gran número de contratos de ARPA y todas tenían computadoras *host* diferentes incompatibles en su totalidad (precisamente para hacerlo más divertido). La red creció con rapidez a medida que se entregaban e instalaban más IMPs; pronto abarcó Estados Unidos. La figura 1-27 muestra qué tan rápido creció ARPANET en los primeros tres años.

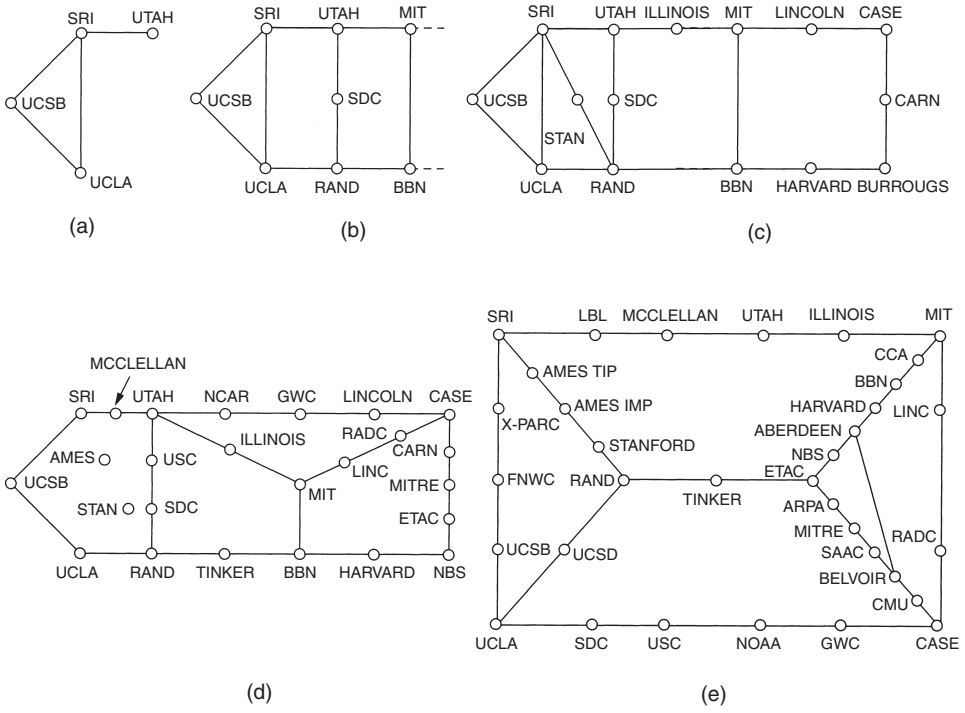

Figura 1-27. Crecimiento de ARPANET: (a) Diciembre de 1969. (b) Julio de 1970. (c) Marzo de 1971. (d) Abril de 1972. (e) Septiembre de 1972.

Además de ayudar al crecimiento de la novel ARPANET, ARPA también proporcionó fondos para la investigación sobre el uso de redes satelitales y redes de radio de paquetes móviles. En una demostración, ahora famosa, un camión que viajaba por California utilizó la red de radio de

paquetes para enviar mensajes a SRI, que luego los reenvió por ARPANET a la Costa Este, donde se expidieron al University College en Londres a través de una red satelital. Esto permitió que el investigador que iba en el camión usara una computadora que se encontraba en Londres mientras manejaba por California.

Este experimento también demostró que los protocolos existentes de ARPANET no eran adecuados para ejecutarse a través de varias redes. Esta observación condujo a más investigación sobre los protocolos, culminando con la invención del modelo y los protocolos de TCP/IP (Cerf y Kahn, 1974). TCP/IP está diseñado de manera específica para manejar comunicación por interredes, aspecto cuya importancia se acrecentó conforme cada vez más y más redes se adhirieron a ARPANET.

Para alentar la adopción de estos nuevos protocolos, ARPA concedió varios contratos a BBN y a la Universidad de California en Berkeley para integrarlos en UNIX de Berkeley. Los investigadores en Berkeley desarrollaron una interfaz de programa adecuada para la red (sockets) y escribieron muchos programas de aplicación, utilería y administración para hacer más fácil la conectividad.

El momento era perfecto. Muchas universidades habían adquirido recientemente una segunda o tercera computadora VAX y una LAN para conectarlas, pero no tenían software de redes. Cuando llegó 4.2BSD junto con TCP/IP, sockets y muchas utilerías de red, el paquete completo se adoptó de inmediato. Además, con TCP/IP, fue fácil para las LANs conectarse a ARPANET y muchas lo hicieron.

Durante la década de 1980, se conectaron redes adicionales, en particular LANs, a ARPANET. Conforme crecía el escalamiento, encontrar *hosts* llegó a ser muy costoso, por lo que se creó el **DNS** (**Sistema de Nombres de Dominio**) para organizar máquinas dentro de dominios y resolver nombres de *host* en direcciones IP. Desde entonces, el DNS ha llegado a ser un sistema de base de datos distribuido generalizado para almacenar una variedad de información relacionada con la elección de un nombre. En el capítulo 7 estudiaremos en detalle este tema.

NSFNET

A finales de la década de 1970, la NFS (Fundación Nacional para las Ciencias, de Estados Unidos) vio el enorme impacto que ARPANET estaba teniendo en la investigación universitaria, permitiendo que científicos de todo el país compartieran datos y colaboraran en proyectos de investigación. Sin embargo, para estar en ARPANET, una universidad debía tener un contrato de investigación con el DoD, lo cual muchas no tenían. La respuesta de la NSF fue diseñar un sucesor de ARPANET que pudiera estar abierto a todos los grupos de investigación de las universidades. Para tener algo concreto con que empezar, la NSF decidió construir una red dorsal (o troncal) para conectar sus seis centros de supercomputadoras en San Diego, Boulder, Champaign, Pittsburgh, Ithaca y Princeton. A cada supercomputadora se le dio un hermano menor, que consistía en una microcomputadora LSI-11 llamada ***fuzzball***. Estas computadoras estaban conectadas a líneas alquiladas de 56 kbps y formaban una subred, utilizando la misma tecnología de hardware que ARPANET. Sin embargo, la tecnología de software era diferente: las *fuzzball* utilizan TCP/IP desde el inicio, creando así la primera WAN TCP/IP.

La NSF también fundó algunas redes regionales (alrededor de 20) que se conectaban a la red dorsal para que los usuarios en miles de universidades, laboratorios de investigación, bibliotecas y museos, tuvieran acceso a cualquiera de las supercomputadoras y se comunicaran entre sí. Toda la red, incluyendo la red dorsal y las redes regionales, se llamó **NSFNET.** Ésta se conectó a ARPANET a través de un enlace entre un IMP y una *fuzzball* en el cuarto de máquinas de Carnegie-Mellon. En la figura 1-28 se muestra la primera red dorsal NSFNET.

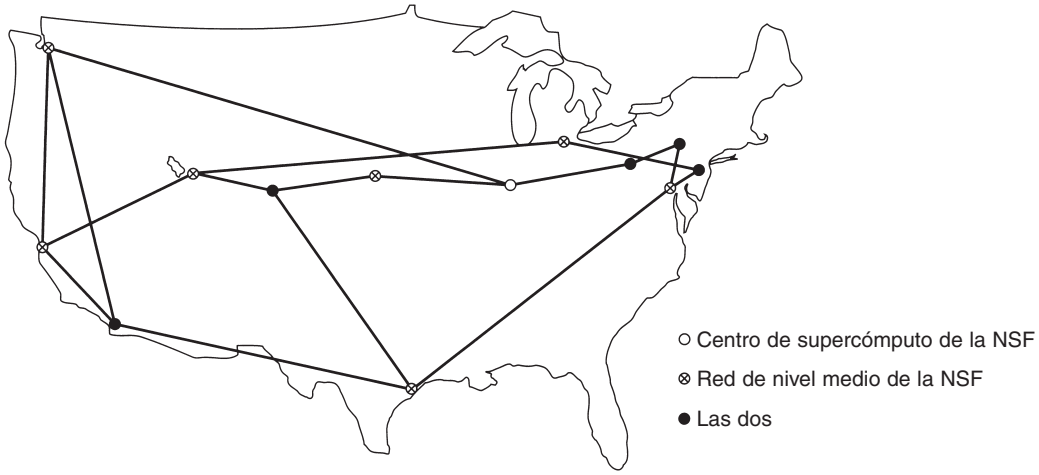

Figura 1-28. La red dorsal NSFNET en 1988.

NSFNET fue un éxito instantáneo y pronto se saturó. Inmediatamente, la NSF empezó a planear su sucesor y otorgó un contrato al consorcio MERIT de Michigan para que lo creara. Se alquilaron a MCI (puesto que se fusionó con WorldCom) canales de fibra óptica a 448 kbps para establecer la versión 2 de la red dorsal. Se utilizaron PC-RTs de IBM como enrutadores. Esta segunda red dorsal también se sobrecargó pronto, y en 1990 se escaló a 1.5 Mbps.

Al continuar el crecimiento, la NSF se percató de que el gobierno no podría financiar por siempre el uso de redes. Además, las empresas comerciales se querían unir, pero los estatutos de la NSF les prohibían utilizar las redes por las que la NSF había pagado. En consecuencia, la NSF alentó a MERIT, MCI e IBM a que formaran una corporación no lucrativa, **ANS (Redes y Servicios Avanzados),** como el primer paso hacia la comercialización. En 1990, ANS adquirió NSFNET y escaló los enlaces de 1.5 Mbps a 45 Mbps para formar **ANSNET.** Esta red operó durante cinco años y luego fue vendida a America Online. Pero para entonces varias empresas estaban ofreciendo servicios IP comerciales y fue evidente que el gobierno se debía retirar del negocio de las redes.

Para facilitar la transición y hacer que todas las redes regionales se pudieran comunicar con las demás redes regionales, la NSF concedió contratos a cuatro diferentes operadores de redes para establecer un **NAP (Punto de Acceso a la Red).** Estos operadores eran PacBell (San Francisco),

Ameritech (Chicago), MFS (Washington, D.C.) y Sprint (Nueva York, donde para efectos de NAP, Pennsauken, Nueva Jersey se toma en cuenta como si fuera la ciudad de Nueva York). Todo operador de red que quisiera proporcionar el servicio de red dorsal a las redes regionales de la NSF se tenía que conectar a todos los NAPs.

Este arreglo significaba que un paquete que se originara en cualquier red regional tenía la opción de contar con operadores de red dorsal desde su NAP al NAP de destino.

En consecuencia, los operadores de red dorsal se vieron forzados a competir por el negocio de las redes regionales con base en el servicio y el precio, que, desde luego, era la idea. Como resultado, el concepto de una única red dorsal predeterminada fue reemplazado por una infraestructura competitiva orientada a la comercialización. A muchas personas les gusta criticar al gobierno federal por no ser innovador, pero en el área de redes, el DoD y la NSF fueron los creadores de la infraestructura que cimentó la base para Internet y luego dejaron que la industria la operara.

Durante la década de 1990, muchos otros países y regiones también construyeron redes nacionales de investigación, con frecuencia siguiendo el patrón de ARPANET y NSFNET. Éstas incluían EuropaNET y EBONE en Europa, que empezaron con líneas de 2 Mbps y luego las escalaron a 34 Mbps. Finalmente, en Europa la infraestructura de redes quedó en manos de la industria.

Uso de Internet

El número de redes, máquinas y usuarios conectados a ARPANET creció rápidamente luego de que TCP/IP se convirtió en el protocolo oficial el 1o. de enero de 1983. Cuando NSFNET y ARPANET estaban interconectadas, el crecimiento se hizo exponencial. Muchas redes regionales se unieron y se hicieron conexiones a redes en Canadá, Europa y el Pacífico.

En algún momento a mediados de la década de 1980, las personas empezaron a ver el conjunto de redes como una interred y más tarde como Internet, aunque no hubo una inauguración oficial con algún político rompiendo una botella de champaña sobre una *fuzzball*.

El aglutinante que mantiene unida la Internet es el modelo de referencia TCP/IP y la pila de protocolos de TCP/IP. TCP/IP hace posible el servicio universal y se puede comparar con la adopción de la medida estándar para el ancho de vía del ferrocarril en el siglo XIX o la adopción de los protocolos de señalización comunes para las compañías telefónicas.

¿Qué significa en realidad estar en Internet? Nuestra definición es que una máquina está en Internet si ejecuta la pila de protocolos de TCP/IP, tiene una dirección IP y puede enviar paquetes IP a todas las demás máquinas en Internet. La sola capacidad para enviar y recibir correo electrónico no basta, puesto que el correo electrónico es la puerta de entrada a muchas redes fuera de Internet. Sin embargo, el aspecto se nubla de alguna manera por el hecho de que millones de computadoras personales pueden llamar a un proveedor de servicios de Internet mediante un módem, recibir direcciones IP temporales y enviar paquetes IP a otros *hosts* de Internet. Tiene sentido decir que tales máquinas están en Internet en tanto estén conectadas al enrutador del proveedor de servicios.

Tradicionalmente (es decir, de 1970 a 1990) Internet y sus predecesores tenían cuatro aplicaciones principales:

1. **Correo electrónico.** La capacidad para redactar, enviar y recibir correo electrónico ha sido posible desde los inicios de ARPANET y su gran popularidad. Muchas personas obtienen docenas de mensajes al día y consideran esto como su primer medio de interactuar con el mundo exterior, más allá del teléfono y el correo caracol que se han quedado atrás. Hoy en día los programas de correo electrónico están disponibles en prácticamente todo tipo de computadora.
2. **Noticias.** Los grupos de noticias son foros especializados en los que los usuarios con un interés común pueden intercambiar mensajes. Existen miles de grupos de noticias, dedicados a temas técnicos y no técnicos, entre ellos computadoras, ciencia, recreación y política. Cada grupo de noticias tiene su propia etiqueta, estilo, hábitos y penas en que se incurre al violarlas.
3. **Inicio remoto de sesión.** Mediante los programas telnet, rlogin o ssh, los usuarios de cualquier parte en Internet pueden iniciar sesión en cualquier otra máquina en la que tengan una cuenta.
4. **Transferencia de archivos.** Con el programa FTP, los usuarios pueden copiar archivos de una máquina en Internet a otra. Por este medio se encuentra disponible una vasta cantidad de artículos, bases de datos y otra información.

Hasta principios de la década de 1990, Internet era muy visitada por investigadores académicos, del gobierno e industriales. Una nueva aplicación, **WWW** (**World Wide Web**) cambió todo eso y trajo millones de usuarios nuevos no académicos a la red. Esta aplicación —inventada por Tim Berners-Lee, físico del CERN— no cambió ninguna de las características subyacentes pero las hizo más fáciles de usar. Junto con el navegador Mosaic, escrito por Marc Andreessen en el Centro Nacional para Aplicaciones de Supercómputo en Urbana, Illinois, WWW hizo posible que un sitio estableciera páginas de información que contienen texto, imágenes, sonido e incluso vídeo, y vínculos integrados a otras páginas. Al hacer clic en un vínculo, el usuario es transportado de inmediato a la página a la que apunta dicho vínculo. Por ejemplo, muchas compañías tienen una página de inicio con entradas que apuntan a otras páginas que contienen información de productos, listas de precios, ventas, soporte técnico, comunicación con empleados, información para los accionistas y más.

En muy poco tiempo han aparecido páginas de otro tipo, incluyendo mapas, tablas del mercado accionario, catálogos de fichas bibliográficas, programas de radio grabados e incluso una página que apunta al texto completo de muchos libros cuyos derechos de autor han expirado (Mark Twain, Charles Dickens, etcétera). Muchas personas también tienen páginas personales (páginas de inicio).

Gran parte de su crecimiento durante la década de 1990 estuvo alimentado por empresas llamadas **ISPs** (**proveedores de servicios de Internet**)**.** Hay compañías que ofrecen a los usuarios individuales domésticos la capacidad de llamar a una de sus máquinas y conectarse a Internet, obteniendo así acceso al correo electrónico, WWW y otros servicios de Internet. Estas compañías suscribieron contratos con decenas de millones de usuarios nuevos por un año durante el final de

la década de 1990, cambiando por completo el carácter de la red de ser un campo de recreo para académicos y militares a uno de utilidad pública, muy semejante al sistema telefónico. Ahora el número de usuarios de Internet se desconoce, pero lo cierto es que son cientos de millones en todo el mundo y tal vez pronto lleguen a rebasar los mil millones.

Arquitectura de Internet

En esta sección trataremos de dar un breve panorama de lo que es Internet hoy. Debido a las muchas fusiones entre compañías telefónicas (telcos) e ISPs, las aguas se han enturbiado y a veces es difícil decir quién hace qué. En consecuencia, la siguiente descripción será, por necesidad, algo más sencilla que la realidad. En la figura 1-29 se muestra el panorama general. Ahora examinemos esta figura parte por parte.

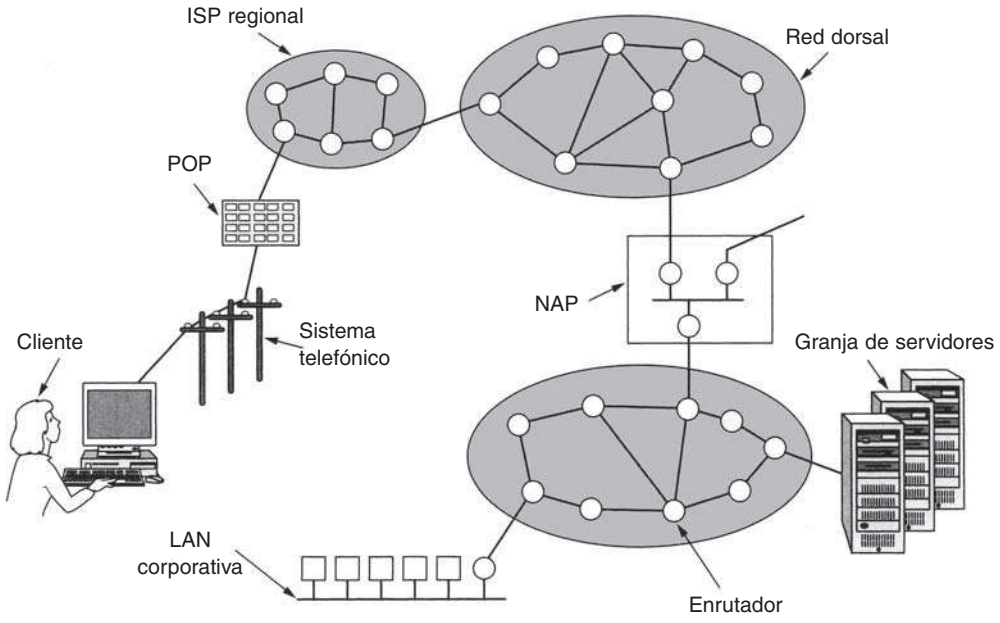

Figura 1-29. Panorama de Internet.

Un buen lugar para empezar es con un cliente en casa. Supongamos que nuestro cliente llama a su ISP desde una conexión de línea telefónica conmutada, como se muestra en la figura 1-29. El módem es una tarjeta dentro de su PC que convierte las señales digitales que la computadora produce en señales análogas que pueden pasar sin obstáculos a través del sistema telefónico. Estas señales se transfieren al **POP** (**Punto de Presencia**) del ISP, donde se retiran del sistema telefónico y se inyectan en la red regional del ISP. A partir de este punto, el sistema es totalmente digital y de conmutación de paquetes. Si el ISP es la telco local, es probable que el POP esté ubicado en la

oficina de conmutación telefónica, donde termina el cableado de teléfono de los clientes. Si el ISP no es la telco local, el POP podría ser alguna de las oficinas de conmutación en el camino.

La red regional de ISPs consta de enrutadores interconectados en las diversas ciudades en las que el ISP opera o da servicio. Si el paquete está destinado a un *host* servido directamente por el ISP, el paquete se entrega al *host*. En caso contrario, se entrega al operador de la red dorsal del ISP.

En la cima de la cadena alimenticia están los operadores de las principales redes dorsales, empresas como AT&T y Sprint. Éstas operan grandes redes de redes dorsales internacionales, con miles de enrutadores conectados por fibra óptica de banda ancha. Grandes corporaciones y servicios de *hosting* que ejecutan granjas de servidores (máquinas que pueden servir miles de páginas Web por segundo) con frecuencia se conectan de manera directa a la red dorsal. Los operadores de redes dorsales alientan esta conexión directa rentando espacio en lo que se llama **hoteles de portadores,** que son básicamente gabinetes de equipos en el mismo cuarto que el enrutador para conexiones cortas y rápidas entre las granjas de servidores y la red dorsal.

Si un paquete dado a la red dorsal se destina a un ISP o a una compañía servida por la red dorsal, se envía al enrutador más cercano y se pierde cualquier responsabilidad por este paquete. Sin embargo, en el mundo hay muchas redes dorsales, de varios tamaños, de manera que un paquete podría tener que ir a una red dorsal competidora. Para que los paquetes viajen entre redes dorsales, todas las redes principales se conectan a los NAPs explicados antes. Básicamente, un NAP es un cuarto lleno de enrutadores, al menos uno por red dorsal. Una LAN en el cuarto conecta todos los enrutadores, de modo que los paquetes se pueden reenviar desde una red dorsal hacia cualquier otra. Además de estar conectadas en los NAPs, las redes dorsales más grandes tienen numerosas conexiones directas entre sus enrutadores, una técnica conocida como **igualdad privada** (*private peering*). Una de las muchas paradojas de Internet es que los ISPs que compiten en público entre sí por clientes, con frecuencia cooperan estableciendo igualdades privadas entre ellos (Metz, 2001).

Aquí termina nuestro rápido viaje por Internet. En los siguientes capítulos tendremos mucho que decir sobre los componentes individuales y su diseño, algoritmos y protocolos. También vale la pena mencionar de paso que algunas empresas tienen interconectadas todas sus redes internas existentes, utilizando con frecuencia la misma tecnología que Internet. Por lo general, estas **intranets** son accesibles sólo dentro de la empresa pero, por otra parte, funcionan del mismo modo que Internet.

1.5.2 Redes orientadas a la conexión: X.25, Frame Relay y ATM

Desde el inicio de la conectividad surgió una guerra entre aquellos que apoyan a las subredes no orientadas a la conexión (es decir, de datagramas) y quienes apoyan a las subredes orientadas a la conexión. Los principales defensores de las subredes no orientadas a la conexión vienen de la comunidad ARPANET/Internet. Recuerde que el deseo original del DoD al fundar y construir ARPANET era tener una red que pudiera funcionar incluso después de que varios impactos de armas nucleares destruyeran numerosos enrutadores y líneas de transmisión. Por lo tanto, la tolerancia a

errores era importante en su lista de prioridades, no tanto lo que pudieran cobrar a los clientes. Este enfoque condujo a un diseño no orientado a la conexión en el que cada paquete se enruta independientemente de cualquier otro paquete. Por lo tanto, si algunos enrutadores se caen durante una sesión, no hay daño puesto que el sistema puede reconfigurarse a sí mismo de manera dinámica para que los paquetes subsiguientes puedan encontrar alguna ruta a su destino, aun cuando sea diferente de la que utilizaron los paquetes anteriores.

El campo orientado a la conexión viene del mundo de las compañías telefónicas. En el sistema telefónico, quien llama debe marcar el número de la parte a la que desea llamar y esperar la conexión antes de poder hablar o enviar los datos. Esta configuración de conexión establece una ruta a través del sistema telefónico que se mantiene hasta que se termina la llamada. Todas las palabras o paquetes siguen la misma ruta. Si una línea o conmutador se cae en el trayecto, la llamada se cancela. Esta propiedad es precisamente lo que al DoD no le gustaba.

Entonces, ¿por qué le gustaba a las compañías telefónicas? Por dos razones:

1. Calidad en el servicio.
2. Facturación.

Al establecer de antemano una conexión, la subred puede reservar recursos como espacio de búfer y capacidad de procesamiento (CPU) en los enrutadores. Si se intenta establecer una llamada y los recursos disponibles son insuficientes, la llamada se rechaza y el invocador recibe una señal de ocupado. De esta manera, una vez que se establece una conexión, ésta da un buen servicio. Con una red no orientada a la conexión, si llegan demasiados paquetes al mismo enrutador al mismo tiempo, el enrutador se saturará y tal vez pierda algunos paquetes. Tal vez el emisor advierta esto y los envíe de nuevo, pero la calidad del servicio será accidentada e inadecuada para audio o vídeo a menos que la red tenga poca carga. No es necesario decir que proveer una calidad de audio adecuada es algo en lo que las compañías telefónicas ponen mucho cuidado, de ahí su preferencia por las conexiones.

La segunda razón por la que las compañías telefónicas prefieren el servicio orientado a la conexión es que están acostumbradas a cobrar por el tiempo de conexión. Cuando hace una llamada de larga distancia (sea nacional o internacional) se le cobra por minuto. Cuando llegaron las redes, se vieron atraídas precisamente hacia un modelo en el que el cobro por minuto fuera fácil de hacer. Si se tiene que establecer una conexión antes de enviar los datos, en ese momento es cuando el reloj de la facturación empieza a correr. Si no hay conexión, no hay cobro.

Irónicamente, mantener registros de facturación es muy costoso. Si una compañía telefónica adoptara una tarifa mensual plana sin límite de llamadas y sin facturación o mantenimiento de un registro, probablemente ahorraría una gran cantidad de dinero, a pesar del incremento en llamadas que generaría esta política. Sin embargo, hay políticas, regulaciones y otros factores que pesan en contra de hacer esto. Curiosamente, el servicio de tarifa plana existe en otros sectores. Por ejemplo, la TV por cable se factura en una tasa mensual plana, independientemente de cuántos programas vea. Podría haberse diseñado con pago por evento como concepto básico, pero no fue así, en parte por lo costoso de la facturación (y dada la calidad de la mayoría de los programas televisivos, la vergüenza no se puede descontar del todo). Asimismo, muchos parques de diversiones

cobran una cuota de admisión por día con acceso ilimitado a los juegos, en contraste con las ferias ambulantes que cobran por juego.

Dicho esto, no nos debería sorprender que todas las redes diseñadas por la industria de la telefonía hayan sido subredes orientadas a la conexión. Lo que sí es de sorprender es que Internet también se está inclinado en esa dirección, a fin de proporcionar un mejor servicio de audio y vídeo, un tema al que volveremos en el capítulo 5. Por ahora examinaremos algunas redes orientadas a la conexión.

X.25 y Frame Relay

Nuestro primer ejemplo de red orientada a la conexión es la **X.25,** que fue la primera red de datos pública. Se desplegó en la década de 1970, cuando el servicio telefónico era un monopolio en todas partes y la compañía telefónica de cada país esperaba que hubiera una red de datos por país —la propia. Para utilizar X.25, una computadora establecía primero una conexión con la computadora remota, es decir, hacía una llamada telefónica. Esta conexión daba un número de conexión para utilizarlo en los paquetes de transferencia de datos (ya que se podían abrir muchas conexiones al mismo tiempo). Los paquetes de datos eran muy sencillos, consistían en un encabezado de 3 bytes y hasta 128 bytes de datos. El encabezado constaba de un número de conexión de 12 bits, un número de secuencia de paquete, un número de confirmación de recepción y algunos bits diversos. Las redes X.25 funcionaron por casi diez años con resultados mixtos.

En la década de 1980, las redes X.25 fueron reemplazadas ampliamente por un nuevo tipo de red llamada **Frame Relay.** Ésta es una red orientada a la conexión sin controles de error ni de flujo. Como era orientada a la conexión, los paquetes se entregaban en orden (en caso de que se entregaran todos). Las propiedades de entrega en orden, sin control de errores ni de flujo hicieron el Frame Relay parecido a la LAN de área amplia. Su aplicación más importante es la interconexión de LANs en múltiples oficinas de una empresa. Frame Relay disfrutó de un éxito modesto y aún se sigue utilizando en algunas partes.

Modo de Transferencia Asíncrona

Otro tipo de red orientada a la conexión, tal vez el más importante, es **ATM** (**Modo de Transferencia Asíncrona**)**.** La razón de tan extraño nombre se debe a que en el sistema telefónico la mayor parte de la transmisión es síncrona (lo más parecido a un reloj), y en ATM no sucede así.

ATM se diseñó a principios de la década de 1990 y se lanzó en medio de una increíble exageración (Ginsburg, 1996; Goralski, 1995; Ibe, 1997; Kimn y cols., 1994, y Stallings, 2000). ATM iba a resolver todos los problemas de conectividad y telecomunicaciones fusionando voz, datos, televisión por cable, télex, telégrafo, palomas mensajeras, botes conectados por cordón, tambores, señales de humo y todo lo demás, en un solo sistema integrado que pudiera proporcionar todos los servicios para todas las necesidades. Eso no sucedió. En gran parte, los problemas fueron semejantes a los ya descritos en el tema de OSI, es decir, una aparición inoportuna, junto con tecnología, implementación y políticas equivocadas. Habiendo noqueado a las compañías telefónicas en el primer asalto, gran parte de la comunidad de Internet vio a ATM como cuando Internet era el contrincante de las telcos: la segunda parte. Pero no fue así en realidad y esta vez incluso los in-

transigentes fanáticos de los datagramas se dieron cuenta de que la calidad de servicio de Internet dejaba mucho que desear. Para no alargar la historia, ATM tuvo mucho más éxito que OSI y actualmente tiene un uso profundo dentro del sistema telefónico, con frecuencia en el transporte de los paquetes IP. Como en la actualidad las empresas portadoras la utilizan principalmente para su transporte interno, los usuarios no se percatan de su existencia pero, definitivamente, vive y goza de salud.

Circuitos virtuales de ATM

Puesto que las redes ATM están orientadas a la conexión, el envío de datos requiere que primero se envíe un paquete para establecer la conexión. Conforme el mensaje de establecimiento sigue su camino a través de la subred, todos los conmutadores que se encuentran en la ruta crean una entrada en sus tablas internas tomando nota de la existencia de la conexión y reservando cualesquier recursos que necesite la conexión. Con frecuencia a las conexiones se les conoce como **circuitos virtuales,** en analogía con los circuitos físicos utilizados en el sistema telefónico. La mayoría de las redes ATM soportan también **circuitos virtuales permanentes,** que son conexiones permanentes entre dos *hosts* (distantes). Son similares a las líneas alquiladas del mundo telefónico. Cada conexión, temporal o permanente, tiene un solo identificador de conexión. En la figura 1-30 se ilustra un circuito virtual.

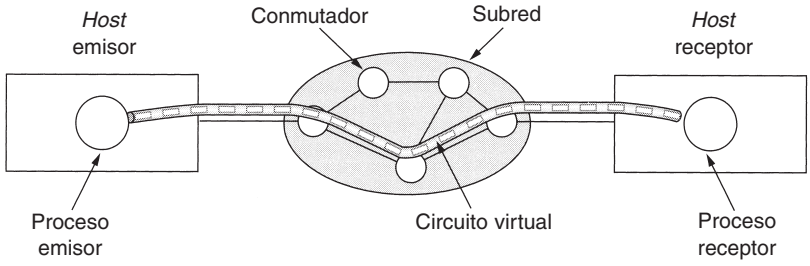

Figura 1-30. Un circuito virtual.

Una vez establecida la conexión, cada lado puede empezar a transmitir datos. La idea básica en que se fundamenta ATM es transmitir toda la información en paquetes pequeños, de tamaño fijo, llamados **celdas.** Las celdas tienen un tamaño de 53 bytes, de los cuales cinco son del encabezado y 48 de carga útil, como se muestra en la figura 1-31. Parte del encabezado es el identificador de la conexión, por lo que los *hosts* emisor y receptor y todos los conmutadores intermedios pueden saber qué celdas pertenecen a qué conexiones. Esta información permite que cada conmutador sepa cómo enviar cada celda entrante. La conmutación de celdas se hace en el hardware, a alta velocidad. De hecho, el principal argumento para tener celdas de tamaño fijo es que así es fácil construir conmutadores de hardware para manejar celdas pequeñas, de longitud fija. Los paquetes de longitud variable de IP se tienen que enrutar mediante software, que es un proceso más lento.

Otro punto a favor de ATM es que el hardware se puede configurar para enviar una celda entrante a múltiples líneas de salida, una propiedad necesaria para el manejo de un programa de televisión que se va a difundir a varios receptores. Por último, las celdas pequeñas no bloquean ninguna línea por mucho tiempo, lo que facilita la garantía en la calidad del servicio.

Todas las celdas siguen la misma ruta al destino. La entrega de celdas no está garantizada, pero el orden sí. Si las celdas 1 y 2 se envían en ese orden, entonces deben arribar en el mismo orden, nunca primero la 2 y luego la 1. No obstante, una de las dos o ambas se pueden perder en el trayecto. A los niveles más altos de protocolos les corresponde la recuperación de celdas perdidas. Observe que aunque esta garantía no es perfecta, es mejor que la de Internet. Ahí los paquetes no sólo se pierden, sino que además se entregan en desorden. ATM, en contraste, garantiza que las celdas nunca se entregarán en desorden.

Figura 1-31. Una celda ATM.

Las redes ATM se organizan como las WANs tradicionales, con líneas y conmutadores (enrutadores). Las velocidades más comunes para las redes ATM son de 155 y 622 Mbps, aunque también se soportan velocidades más altas. Se eligió la velocidad de 155 Mbps porque ésta es la que se requiere para transmitir televisión de alta definición. La elección exacta de 155.52 Mbps se hizo por compatibilidad con el sistema de transmisión SONET de AT&T, punto que estudiaremos en el capítulo 2. La velocidad de 622 Mbps se eligió para que se pudieran enviar cuatro canales de 155 Mbps.

El modelo de referencia ATM

ATM tiene su propio modelo de referencia, el cual es diferente del OSI y también del TCP/IP. En la figura 1.32 se muestra el modelo de referencia ATM. Consta de tres capas: la física, la ATM y la de adaptación ATM, además de lo que el usuario desee poner arriba de ellas.

La capa física tiene que ver con el medio físico: voltajes, temporización de bits y otros aspectos más. ATM no prescribe un conjunto particular de reglas, tan sólo especifica que las celdas ATM se pueden enviar tal cual por cable o fibra, pero también se pueden empacar dentro de la carga útil de otros sistemas de transporte. En otras palabras, ATM se ha diseñado para ser independiente del medio de transmisión.

La **capa ATM** se encarga de las celdas y su transporte. Define la disposición de una celda e indica qué significan los campos del encabezado. También tiene que ver con el establecimiento y la liberación de los circuitos virtuales. El control de congestión también se ubica aquí.

Puesto que la mayoría de las aplicaciones no necesita trabajar de manera directa con las celdas (aunque algunas podrían hacerlo), se ha definido una capa superior a la capa ATM para que

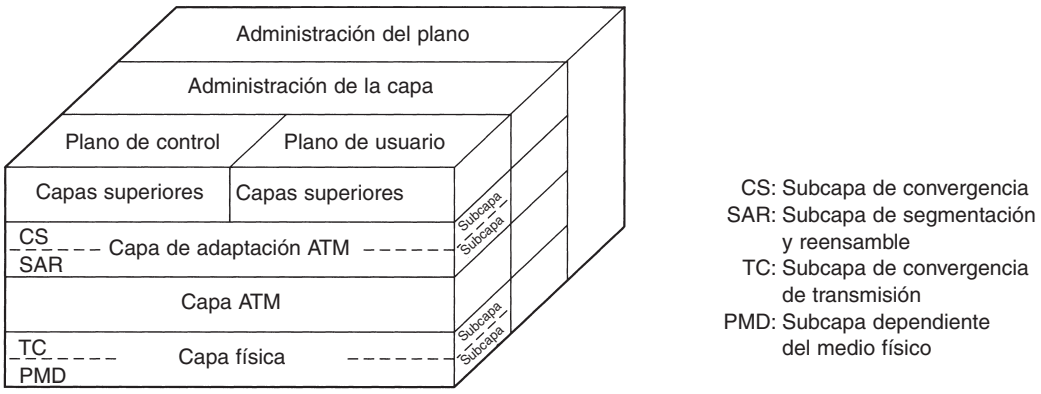

Figura 1-32. El modelo de referencia ATM.

los usuarios envíen paquetes más grandes que una celda. La interfaz de ATM segmenta estos paquetes, transmite de forma individual las celdas y las reensambla en el otro extremo. Esta capa es la **AAL (Capa de Adaptación ATM).**

A diferencia de los primeros modelos de referencia bidimensionales, el modelo ATM se define como si fuera tridimensional, lo que se puede apreciar en la figura 1-32. El **plano de usuario** trata con el transporte de datos, control de flujo, corrección de errores y otras funciones de usuario. En contraste, el **plano de control** se ocupa de la administración de la conexión. Las funciones de administración del plano y de la capa se relacionan con la administración de recursos y coordinación entre capas.

Cada una de las capas física y AAL se dividen en dos subredes, una en la parte inferior que hace el trabajo y en la subcapa de convergencia en la parte superior que proporciona la interfaz propia de la capa superior inmediata. En la figura 1-33 se muestran las funciones de las capas y subcapas.

La subcapa **PMD (Dependiente del Medio Físico**) interactúa con el cable real. Mueve los bits dentro y fuera y maneja la temporización de bits, es decir, el tiempo que existe entre cada bit al transmitirlos. Esta capa será diferente para diferentes transportadoras y cables.

La otra subcapa de la capa física es la subcapa **TC (Convergencia de Transmisión).** Cuando se transmiten las celdas, la capa TC las envía como una cadena de bits a la capa PMD. Esto es sencillo. En el otro extremo, la subcapa TC recibe una serie de bits de entrada de la subcapa PMD. Su trabajo es convertir este flujo de bits en un flujo de celdas para la capa ATM. Maneja todos los aspectos relacionados con las indicaciones de dónde empiezan y terminan las celdas en el flujo de bits. En el modelo ATM, esta funcionalidad se da en la capa física. En el modelo OSI y en gran parte de las demás redes, el trabajo de entramado, es decir, convertir una serie de bits en bruto en una secuencia de tramas o celdas, es la tarea de la capa de enlace de datos.

Como mencionamos antes, la capa ATM maneja celdas, incluyendo su generación y transporte. La mayoría de los aspectos interesantes de ATM se encuentra ubicada aquí. Es una combinación de las capas de enlace de datos y de red del modelo OSI; no hay una división en subcapas.

Capa OSI	Capa ATM	Subcapa ATM	Funcionalidad
3/4	AAL	CS	Provisión de la interfaz estándar (convergencia)
		SAR	Segmentación y reensamblado
2/3	ATM		Control de flujo Generación/extracción de encabezado de celda Circuito virtual/administración de ruta Multiplexión/desmultiplexión de celdas
2	Física	TC	Desacoplamiento de proporción de celdas Generación y comprobación de la suma de verificación de encabezados Generación de celdas Empaque/desempaque de celdas a partir del sobre contenedor Generación de tramas
1		PMD	Temporización de bits Acceso a la red física

Figura 1-33. Las capas y subcapas de ATM y sus funciones.

La capa AAL se divide en una subcapa **SAR** (**Segmentación y Reensamble**) y una **CS** (**Subcapa de Convergencia**)**.** La subcapa inferior fragmenta paquetes en celdas en el lado de transmisión y los une de nuevo en el destino. La subcapa superior permite que los sistemas ATM ofrezcan diversos tipos de servicios a diferentes aplicaciones (por ejemplo, la transferencia de archivos y el vídeo bajo demanda tienen diferentes requerimientos respecto a manejo de errores, temporización, etcétera).

Puesto que quizá ATM esté en declive, no lo explicaremos más en este libro. No obstante, puesto que existe una base instalada considerable, es probable que aún siga en uso durante algunos años. Para más información sobre ATM, vea (Dobrowsky y Grise, 2001, y Gadecki y Heckart, 1997).

1.5.3 Ethernet

Internet y ATM se diseñaron para conectividad de área amplia. Sin embargo, muchas empresas, universidades y otras organizaciones tienen un gran número de computadoras que requieren interconexión. Esta necesidad dio origen a la red de área local. En esta sección diremos algo sobre la LAN más popular: Ethernet.

La historia empieza en la prístina Hawaii a principios de la década de 1970. En este caso, "prístina" se puede interpretar como "que no tiene un sistema telefónico funcional". En tanto los días son más agradables para los vacacionistas cuando no son interrumpidos por el teléfono, no fue así para el investigador Norman Abramson y sus colegas de la Universidad de Hawaii, quienes estuvieron tratando de conectar usuarios de las islas remotas a la computadora principal de Hono-

lulu. Conectar sus propios cables bajo el Océano Pacífico parecía imposible, de modo que buscaron una solución diferente.

La primera que encontraron fueron los radios de onda corta. Cada terminal estaba equipada con un radio pequeño de dos frecuencias: un canal ascendente (a la computadora central) y otro descendente (desde la computadora central). Cuando el usuario deseaba conectarse con la computadora, sólo transmitía por el canal ascendente un paquete que contenía los datos. Si en ese instante nadie más estaba transmitiendo, probablemente el paquete saldría y su recepción sería confirmada en el canal descendente. Si había contención por el canal ascendente, la terminal detectaría la falta de confirmación de recepción y haría otro intento. Puesto que sólo habría un emisor en el canal descendente (la computadora central), nunca habría colisiones ahí. Este sistema, llamado ALOHANET, trabajaba muy bien en condiciones de bajo tráfico pero se caía cuando el flujo de tráfico ascendente era pesado.

En esa misma época, un estudiante llamado Bob Metcalfe hizo su licenciatura en el M.I.T. y luego se mudó para obtener su doctorado en Harvard. Durante sus estudios, conoció el trabajo de Abramson. Se interesó tanto en él que después de graduarse en Harvard decidió pasar el verano en Hawaii trabajando con Abramson antes de empezar a trabajar en el Centro de Investigación de Palo Alto de Xerox (PARC). Cuando llegó al PARC, vio que los investigadores de ahí habían diseñado y construido lo que más tarde se llamarían computadoras personales. Pero las máquinas estaban aisladas. Aplicando su conocimiento del trabajo de Abramson, junto con su colega David Boggs, diseñó e implementó la primera red de área local (Metcalfe y Boggs, 1976).

Llamaron **Ethernet** al sistema, por lo de *luminiferous ether,* a través del cual se pensó alguna vez que se propagaba la radiación electromagnética. (Cuando, en el siglo XIX, el físico inglés James Clerk Maxwell descubrió que la radiación electromagnética se podía describir mediante una ecuación de onda, los científicos supusieron que el espacio debía estar lleno de algún medio etéreo en el cual se propagaba la radiación. Sólo después del famoso experimento de Michelson-Morley en 1887, los físicos descubrieron que la radiación electromagnética se podía propagar por el vacío.)

Aquí el medio de transmisión no era el vacío, sino un cable coaxial grueso (el éter) de más de 2.5 km de largo (con repetidoras cada 500 metros). El sistema podía contener hasta 256 máquinas por medio de transceptores acoplados al cable. Un cable con múltiples máquinas en paralelo se llama **cable de derivación múltiple** (*multidrop*). El sistema se ejecutaba a 2.94 Mbps. En la figura 1-34 se presenta un esbozo de su arquitectura. Ethernet tenía una mejora importante respecto de ALOHANET; antes de transmitir, una computadora tenía que escuchar el cable para ver si había alguien más transmitiendo. En caso de que ya lo hubiera, la computadora se mantenía en espera de que la transmisión actual terminara. Al hacerlo así se evitaba interferir con las transmisiones existentes, dando una mayor eficiencia. ALOHANET no trabajaba de este modo porque para una terminal en una isla era imposible detectar la transmisión de otra terminal en una isla distante. El problema se resolvía con un cable único.

A pesar de que la computadora escucha antes de transmitir, surge un problema: ¿qué sucede si dos o más computadoras esperan hasta que se complete la transmisión actual y luego empiezan

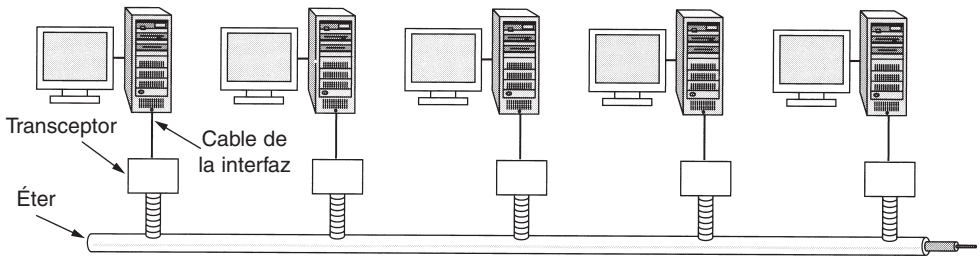

Figura 1-34. Arquitectura de la Ethernet original.

a transmitir al mismo tiempo? La solución es que cada computadora escuche durante su propia transmisión y, si detecta interferencia, mande una señal para poner en alerta a todos los transmisores. Después espera un tiempo aleatorio antes de reintentarlo. Si sucede una colisión, el tiempo aleatorio de espera se duplica y así sucesivamente, para separar las transmisiones que están en competencia y dar a alguna la oportunidad de transmitir primero.

La Ethernet de Xerox fue tan exitosa que DEC, Intel y Xerox diseñaron un estándar en 1978 para una Ethernet de 10 Mbps, llamado **estándar DIX.** Con dos cambios menores, en 1983 el estándar DIX se convirtió en el estándar IEEE 802.3.

Por desgracia para Xerox, ya tenía fama de hacer inventos originales (como el de las computadoras personales) y luego fallar en la comercialización de los mismos, como se menciona en un relato titulado *Fumbling the Future* (Smith y Alexander, 1988). Cuando Xerox mostró poco interés en hacer algo con Ethernet aparte de ayudar a estandarizarlo, Metcalfe formó su propia empresa, 3Com, con el propósito de vender adaptadores Ethernet para PCs. Ha vendido más de 100 millones.

Ethernet continuó su desarrollo y aún está en desarrollo. Han salido nuevas versiones a 100 y 1000 Mbps, e incluso más altas. También se ha mejorado el cableado y se han agregado conmutación y otras características. En el capítulo 4 explicaremos Ethernet en detalle.

De paso, vale la pena mencionar que Ethernet (IEEE 802.3) no es el único estándar de LAN. El comité también estandarizó Token Bus (802.4) y Token Ring (802.5). La necesidad de tres estándares más o menos incompatibles tiene poco que ver con la tecnología y mucho con la política. En el momento de la estandarización, General Motors estaba impulsando una LAN en la que la topología era la misma que la usada en Ethernet (un cable linear), pero las computadoras transmitían por turnos pasando un pequeño paquete de computadora a computadora, llamado **token.** Una computadora podía transmitir sólo si poseía el *token*, lo que evitaba colisiones. General Motors anunció que este esquema era esencial para la manufactura de automóviles y que no estaba preparado para cambiar su postura. No obstante este anuncio, el 802.4 prácticamente desapareció.

Del mismo modo, IBM tenía su favorito: su Token Ring patentado. En este esquema el *token* se pasaba a través del anillo y la computadora que poseyera el *token* podía transmitir antes de poner el *token* de nuevo en el anillo. A diferencia del 802.4, este esquema, estandarizado como 802.5,

aún se usa en algunos sitios de IBM, pero prácticamente en ninguna parte más. Sin embargo, se está desarrollando una versión de 1 gigabit (802.5v), pero parece poco probable que alcance a Ethernet. Resumiendo, había una guerra entre Ethernet, Token Bus y Token Ring, pero Ethernet ganó, en gran medida porque fue la primera y los retadores no pudieron superarlo.

1.5.4 LANs inalámbricas: 802.11

Casi al mismo tiempo que aparecieron las computadoras portátiles, muchas personas tuvieron el sueño de andar por la oficina y poder conectar a Internet su computadora. En consecuencia, varios grupos empezaron a trabajar para cumplir con esta meta. El método más práctico es equipar las computadoras de la oficina y las portátiles con transmisores y receptores de radio de onda corta que les permitan comunicarse. Este trabajo condujo rápidamente a que varias empresas empezaran a comercializar las LANs inalámbricas.

El problema es que no había compatibilidad entre ninguna de ellas. Esta proliferación de estándares implicaba que una computadora equipada con un radio de marca *X* no funcionara en un cuarto equipado con una estación de base marca *Y*. Finalmente, la industria decidió que un estándar de LAN inalámbrica sería una buena idea, por lo que al comité del IEEE que estandarizó las LANs alámbricas se le encargó la tarea de diseñar un estándar para LANs inalámbricas. El estándar resultante se llamó 802.11. En la jerga común se le conoce como **WiFi.** Es un estándar importante y merece respeto, así que lo llamaremos por su nombre propio, 802.11.

El estándar propuesto tenía que trabajar en dos modos:

1. En presencia de una estación base.
2. En ausencia de una estación base.

En el primer caso, toda la comunicación se hacía a través de la estación base, que en la terminología del 802.11 se conoce como **punto de acceso.** En el segundo caso, las computadoras podrían enviarse mensajes entre sí directamente. Este modo se llama a veces **red ad hoc.** Un ejemplo típico es el de dos o más personas que se encuentran juntas en un cuarto no equipado con una LAN inalámbrica y cuyas computadoras se comunican entre sí de manera directa. Los dos modos se ilustran en la figura 1-35.

La primera decisión fue la más sencilla: cómo llamarlo. Todos las otros estándares LAN tenían números como 802.1, 802.2, hasta 802.10, por lo que el estándar LAN se llamó o publicó como 802.11. El resto fue más difícil.

En particular, varios de los diversos retos que había que enfrentar eran: encontrar una banda de frecuencia adecuada, de preferencia mundial; enfrentar el hecho de que las señales de radio tienen un rango finito; asegurarse de que se mantuviera la privacidad de los usuarios; tomar en cuenta la vida limitada de las baterías; preocuparse por la seguridad humana (¿las ondas de radio causan cáncer?); comprender las implicaciones de la movilidad de las computadoras y, por último, construir un sistema con suficiente ancho de banda para que sea económicamente viable.

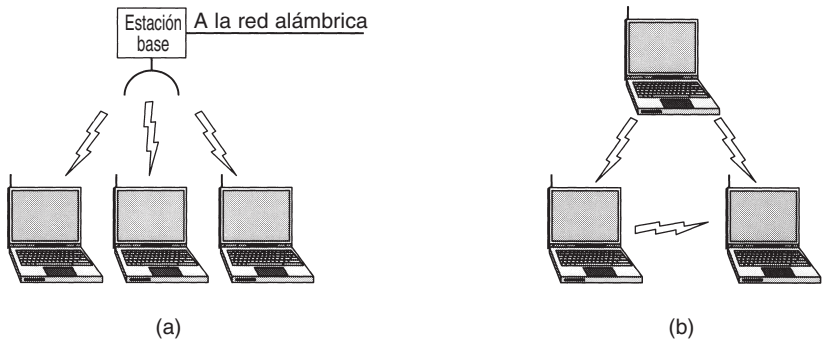

Figura 1-35. (a) Red inalámbrica con una estación base. (b) Red *ad hoc*.

Cuando empezó el proceso de estandarización (a mediados de la década de 1990), Ethernet ya había llegado a dominar las redes de área local, por lo que el comité decidió hacer que el 802.11 fuera compatible con Ethernet sobre la capa de enlace de datos. En particular, se podría enviar un paquete IP sobre la LAN inalámbrica del mismo modo en que una computadora conectada mediante cable enviaba un paquete IP a través de Ethernet. No obstante, existen algunas diferencias inherentes con Ethernet en las capas física y de enlace de datos y tuvieron que manejarse mediante el estándar.

Primero, una computadora en Ethernet siempre escucha el medio antes de transmitir. Sólo si el medio está inactivo la computadora puede empezar a transmitir. Esta idea no funciona igual en las LANs inalámbricas. Para ver por qué, examine la figura 1-36. Suponga que la computadora *A* está transmitiendo a la computadora *B,* pero el alcance del radio del transmisor de *A* es muy corto para encontrar a la computadora *C*. Si *C* desea transmitir a *B* puede escuchar el medio antes de empezar, pero el hecho de que no escuche nada no quiere decir que su transmisión tendrá éxito. El estándar 802.11 tenía que resolver este problema.

El segundo problema que se tenía que resolver es que los objetos sólidos pueden reflejar una señal de radio, por lo que ésta se podría recibir múltiples veces (a través de varias rutas). Esta interferencia da como resultado lo que se llama **desvanecimiento por múltiples trayectorias.**

El tercer problema es que una gran cantidad de software no toma en cuenta la movilidad. Por ejemplo, muchos procesadores de texto tienen una lista de impresoras de entre las cuales los usuarios pueden elegir para imprimir un archivo. Cuando la computadora en la que se ejecuta el procesador de texto se coloca en un nuevo entorno, la lista interna de impresoras ya no es útil.

El cuarto problema es que si una computadora portátil se mueve lejos de la estación base que está usando y dentro del rango de una estación base diferente, se requiere algún tipo de manejo. Aunque este problema ocurre con los teléfonos celulares, eso no sucede con Ethernet y requiere solución. En particular, la red prevista consta de múltiples celdas, cada una con su propia estación base pero con las estaciones base conectadas por Ethernet, como se muestra en la figura 1-37. Desde fuera todo el sistema se vería como una Ethernet sola. La conexión entre el sistema 802.11 y el mundo exterior se conoce como **portal.**

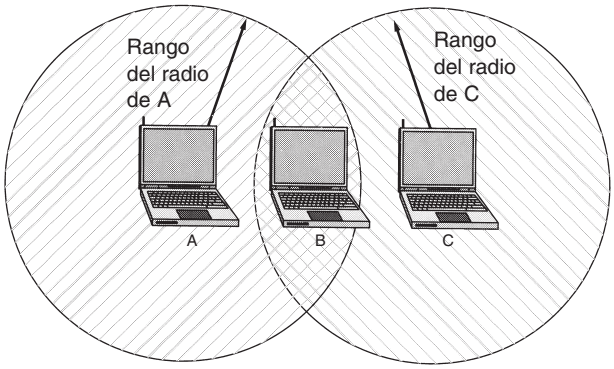

Figura 1-36. El rango de un solo radio no podría cubrir todo el sistema.

Después de algún trabajo, el comité se presentó en 1997 con un estándar que se dirigía a éstos y otros respectos. La LAN inalámbrica descrita se ejecutaba a 1 o 2 Mbps. Casi de inmediato la gente comenzó a quejarse de que era demasiado lenta, de manera que empezaron a trabajar en estándares más rápidos. Una división desarrollada con el comité tuvo como resultado dos nuevos estándares en 1999. El estándar 802.11a utiliza una banda de frecuencia más ancha y se ejecuta a velocidades de hasta 54 Mbps. El estándar 802.11b utiliza la misma banda de frecuencia que el 802.11, pero se vale de una técnica de modulación diferente para alcanzar 11 Mbps. Algunas personas ven esto como un aspecto psicológico importante puesto que 11 Mbps es más rápido que la Ethernet alámbrica original. Es posible que el 802.11 original de 1 Mbps desaparezca con rapidez, pero aún no queda claro cuál de los nuevos estándares será el ganador.

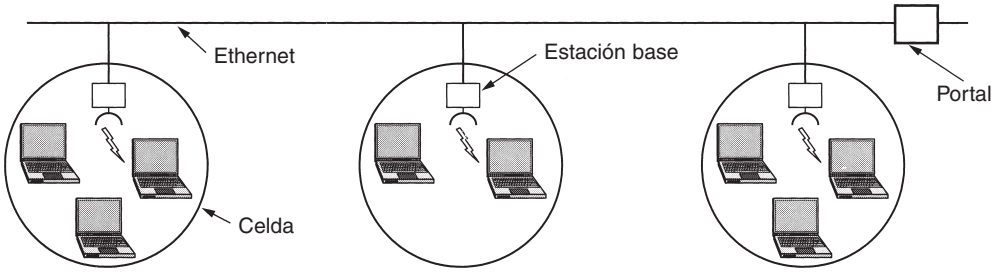

Figura 1-37. Una red 802.11 de múltiples celdas.

Para hacer las cosas todavía más complicadas, el comité 802 ha creado otra variante, el 802.11g, que utiliza la técnica de modulación del 802.11a pero la banda de frecuencia del 802.11b. En el capítulo 4 trataremos en detalle al 802.11.

Sin lugar a dudas, el 802.11 va a causar una revolución en computación y en el acceso a Internet. Aeropuertos, estaciones de trenes, hoteles, centros comerciales y universidades lo están instalando rápidamente. Incluso cafeterías de lujo están instalando el 802.11 para que los *yuppies* que se reúnen puedan navegar en Web mientras toman su café con leche. Es posible que el 802.11 haga por Internet lo que las computadoras portátiles hicieron por la computación: volverla móvil.

1.6 ESTANDARIZACIÓN DE REDES

Existen muchos fabricantes y proveedores de redes, cada uno con sus propias ideas de cómo se deben hacer las cosas. Sin coordinación sería un caos total y los usuarios nunca conseguirían nada. La única manera de resolver esta situación es ponerse de acuerdo en la adopción de algunos estándares para redes.

Los estándares no sólo permiten que computadoras diferentes se comuniquen, sino que también incrementan el mercado de productos que se ajustan al estándar. Un mercado grande conduce a la producción masiva, economías de escala en la producción, implementaciones VLSI y otros beneficios que disminuyen el precio e incrementan aún más la aceptación. En las siguientes secciones daremos un vistazo al importante, pero poco conocido, mundo de la estandarización internacional.

Los estándares se dividen en dos categorías: de facto y de jure. Los estándares ***de facto*** (“de hecho”) son los que simplemente surgieron, sin ningún plan formal. La PC de IBM y sus sucesoras son estándares de facto para oficinas chicas y equipos domésticos porque docenas de fabricantes decidieron copiar exactamente las máquinas de IBM. Del mismo modo, UNIX es el estándar de facto para sistemas operativos en los departamentos de ciencias de la computación de las universidades.

En contraste, los estándares ***de jure*** (“por derecho”), son formales, legales, adoptados por alguna institución de estandarización autorizada. Por lo general, las autoridades de estandarización internacional se dividen en dos clases: las establecidas por acuerdos entre los gobiernos de cada país, y las incluidas de manera voluntaria, sin acuerdos entre organizaciones. En el área de los estándares de redes de computadoras hay varias organizaciones de cada tipo, las cuales explicaremos a continuación.

1.6.1 Quién es quién en el mundo de las telecomunicaciones

La situación legal de las compañías telefónicas del mundo varía considerablemente de un país a otro. En un extremo están los Estados Unidos con sus 1500 empresas telefónicas privadas individuales. Antes de que AT&T se dividiera, en 1984, era la empresa más grande del mundo y dominaba el escenario. Proporcionaba servicio telefónico a casi 80% de los usuarios de Estados

Unidos, con sucursales diseminadas en la mitad del país; las compañías oponentes atendían al resto de los clientes (en su mayor parte rurales). Desde que se dividió, AT&T sigue proporcionando servicio de larga distancia, pero ahora en competencia con otras empresas. Las siete Compañías Operadoras Regionales de Bell que surgieron de la división de AT&T y numerosas empresas independientes proporcionan servicios de telefonía local y celular. Debido a las frecuentes fusiones y otros cambios, la industria está en un estado de movimiento constante.

Las empresas que dan servicios de comunicación al público en Estados Unidos se llaman **portadoras comunes** (*carriers*). Las ofertas y precios se describen en un documento llamado **tarifa,** el cual debe ser aprobado por la Comisión Federal de Comunicaciones para el tráfico interestatal e internacional, pero el tráfico estatal interno lo aprueban las comisiones de servicios públicos.

En el otro extremo están los países en los cuales el gobierno respectivo tiene el monopolio de todas las comunicaciones, como correos, telégrafos, teléfonos y a veces la radio y la televisión. La mayor parte del mundo cae dentro de esta categoría. En algunos casos la autoridad de la telecomunicación es una compañía nacionalizada y en otros es simplemente una rama del gobierno, conocida generalmente como **PTT** (administración de **Correos, Telégrafos y Teléfonos**)**.** La tendencia a nivel mundial es hacia una liberación y competencia, y alejarse del monopolio gubernamental. La mayoría de los países europeos tiene privatizadas (parcialmente) sus PTTs, pero en otras partes el proceso avanza con lentitud.

Con tantos proveedores diferentes de servicios, es claro que se necesita una compatibilidad a escala mundial para asegurarse de que las personas (y las computadoras) de un país puedan llamar a sus contrapartes en otro. En realidad, esta necesidad ha existido desde hace mucho tiempo. En 1865, los representantes de muchos gobiernos de Europa se reunieron para formar el predecesor de la actual **ITU** (**Unión Internacional de Telecomunicaciones**)**.** Su trabajo era estandarizar las telecomunicaciones internacionales, que en esos días se hacían mediante el telégrafo. Incluso entonces era patente que si la mitad de los países utilizaba el código Morse y la otra utilizaba un código diferente, surgiría un problema. Cuando el teléfono entró al servicio internacional, la ITU empezó a trabajar en la estandarización de la telefonía. En 1947 la ITU se convirtió en una agencia de las Naciones Unidas.

La ITU tiene tres sectores principales:

1. Radiocomunicaciones (ITU-R).
2. Estandarización de telecomunicaciones (ITU-T).
3. Desarrollo (ITU-D).

La ITU-R se ocupa de asignar frecuencias de radio en todo el mundo a los grupos de interés en competencia. Nos enfocaremos en primer lugar en la ITU-T, que se ocupa de los sistemas telefónicos y de comunicación de datos. De 1956 a 1993, la ITU-T se conocía como **CCITT** (del francés *Comité Consultatif International Télégraphique et Téléphonique,* Comité Consultivo Internacional para la Telegrafía y Telefonía). El 1o. de marzo de 1993 el CCITT se reorganizó para hacerlo menos burocrático y cambió de nombre para reflejar su nuevo papel. Tanto la ITU-T como el CCITT emitieron recomendaciones en el área de comunicaciones telefónicas y de datos.

Es frecuente encontrar algunas de las recomendaciones del CCITT, como la X.25 del CCITT, aunque desde 1993 las recomendaciones llevan la etiqueta de la ITU-T.

La ITU-T tiene cuatro clases de miembros:

1. Gobiernos nacionales.
2. De sector.
3. Asociados.
4. Agencias reguladoras.

La ITU-T tiene alrededor de 200 miembros gubernamentales, entre ellos casi todos los miembros de las Naciones Unidas. Puesto que Estados Unidos no tiene una PTT, alguien más tenía que representarlos en la ITU-T. Esta tarea recayó en el Departamento de Estado, probablemente porque la ITU-T tenía que ver con los países extranjeros, que era la especialidad del Departamento de Estado.

Hay aproximadamente 500 miembros de sector, incluyendo compañías telefónicas (por ejemplo, AT&T, Vodafone, WorldCom), fabricantes de equipos de telecomunicación (como Cisco, Nokia, Nortel), fabricantes de computadoras (como Compaq, Sun, Toshiba), fabricantes de chips (como Intel, Motorola, TI), compañía de medios (como AOL Time Warner, CBS, Sony) y otras empresas interesadas (como Boeing, Samsung, Xerox). Varias organizaciones científicas no lucrativas y consorcios industriales también son miembros de sector (por ejemplo, IFIP e IATA). Los miembros asociados son organizaciones más pequeñas que se interesan en un grupo de estudio en particular. Las agencias reguladoras son quienes vigilan el negocio de la telecomunicación, como la Comisión Federal de Comunicaciones de Estados Unidos.

La tarea de la ITU-T es hacer recomendaciones técnicas sobre telefonía, telegrafía y las interfaces de comunicación de datos. Estas recomendaciones suelen convertirse en estándares reconocidos internacionalmente, por ejemplo el V.24 (también conocido en Estados Unidos como EIA RS-232), el cual especifica la ubicación y significado de los diversos pines en el conector utilizado para la mayoría de las terminales asíncronas y módems externos.

Es preciso observar que las recomendaciones de la ITU-T técnicamente son sólo sugerencias que los gobiernos pueden adoptar o ignorar (ya que los gobiernos parecen adolescentes de 13 años a quienes no les gusta recibir órdenes). En la práctica, un país que desee adoptar un estándar telefónico diferente del utilizado por el resto del mundo, es libre de hacerlo, pero el precio es el aislamiento. Esto podría funcionar en Corea del Norte, pero fuera de ahí sería un verdadero problema. El sofisma de llamar "recomendaciones" a los estándares de la ITU-T era y es necesario para mantener en calma el nacionalismo de varios países.

El trabajo verdadero de la ITU-T se realiza en sus 14 grupos de estudio, a veces de hasta 400 personas, que abarcan aspectos que van desde la facturación telefónica hasta servicios de multimedia. Para conseguir la realización de los proyectos, los grupos de estudio se dividen en equipos de trabajo, que a su vez se dividen en equipos de expertos, que a su vez se dividen en grupos específicos. Una vez burócrata, jamás se deja de serlo.

A pesar de todo esto, en realidad la ITU-T hace su trabajo. Desde que surgió, ha producido cerca de 3000 recomendaciones que ocupan cerca de 60,000 páginas de papel. Muchas de ellas se han llevado a la práctica en gran medida. Por ejemplo, una de sus recomendaciones es el popular estándar V.90 para módems de 56 kbps.

En tanto las comunicaciones completen la transición, que empezó en la década de 1980, de ser nacionales totalmente a ser globales totalmente, los estándares llegarán a ser más importantes cada vez, y más y más organizaciones querrán estar implicadas en su establecimiento. Para más información sobre la ITU, vea (Irmer, 1994).

1.6.2 Quién es quién en los estándares internacionales

Los estándares internacionales son producidos y publicados por la **ISO** (**Organización de Estándares Internacionales**),† una organización voluntaria no surgida de un acuerdo, fundada en 1946. Sus miembros son las organizaciones de estándares nacionales de los 89 países miembro. Entre ellos se encuentran ANSI (Estados Unidos), BSI (Gran Bretaña), AFNOR (Francia), DIN (Alemania) y otros 85.

La ISO emite estándares sobre una gran cantidad de temas, desde los más básicos (literalmente) como tuercas y pernos, hasta el revestimiento de los postes telefónicos (sin mencionar las semillas de cacao [ISO 2451], las redes de pesca [ISO 1530], ropa interior femenina [ISO 4416] y algunos otros objetos que no se pensaría que fueran sujetos de estandarización). Se han emitido más de 13,000 estándares, entre ellos los estándares de OSI. La ISO tiene casi 200 comités técnicos, numerados por el orden de su creación, refiriéndose cada uno a un objeto específico. El TC1 se ocupa de las tuercas y pernos (estandariza la rosca de los tornillos). El TC97 trata con computadoras y procesamiento de información. Cada TC tiene subcomités (SCs) divididos en grupos de trabajo (WGs).

El trabajo real lo hacen sobre todo los WGs, integrados por más de 100,000 voluntarios en todo el mundo. Muchos de estos "voluntarios" son asignados a trabajar en asuntos de la ISO por sus empleadores, cuyos productos se están estandarizando. Otros son oficiales gubernamentales ansiosos de que lo que se hace en su país llegue a ser estándar internacional. Los expertos académicos también están activos en muchos de los WGs.

En cuanto a estándares de telecomunicación, la ISO y la ITU-T suelen cooperar (la ISO es miembro de la ITU-T), para evitar la ironía de dos estándares internacionales oficiales mutuamente incompatibles.

El representante de Estados Unidos en la ISO es el **ANSI** (**Instituto Estadounidense de Estándares Nacionales**), que a pesar de su nombre es una organización privada no gubernamental y no lucrativa. Sus miembros son fabricantes, empresas portadoras comunes y otras partes interesadas. La ISO suele adoptar los estándares ANSI como estándares internacionales.

El procedimiento seguido por la ISO para adoptar estándares se ha diseñado para obtener el mayor consenso posible. El proceso inicia cuando alguna de las organizaciones de estándares

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

nacionales siente la necesidad de un estándar internacional en un área determinada. Entonces se forma un grupo de trabajo para presentar un **CD** (**Borrador de Comité**)**.** El CD se distribuye a todos los miembros, que tienen seis meses para criticarlo. Si la mayoría lo aprueba, se revisa y distribuye un documento revisado, llamado **DIS** (**Borrador de Estándar Internacional**) para comentarios y votación. Con base en los resultados de esta vuelta, se prepara, aprueba y publica el texto final del **IS** (**Estándar Internacional**)**.** En áreas de gran controversia, un CD o un DIS podría llegar a tener varias versiones antes de lograr suficientes votos y todo el proceso puede llegar a tardar años.

El **NIST** (**Instituto Nacional de Estándares y Tecnología**) es parte del Departamento de Comercio de Estados Unidos. Se llamaba Oficina Nacional de Estándares. Emite estándares que son obligatorios para compras hechas por el gobierno de Estados Unidos, excepto por los del Departamento de Defensa, que tiene sus propios estándares.

Otro representante importante en el mundo de los estándares es el **IEEE** (**Instituto de Ingenieros Eléctricos y Electrónicos**)**,** la mayor organización de profesionales del mundo. Además de publicar multitud de periódicos y organizar cientos de conferencias cada año, el IEEE tiene un grupo de estandarización que desarrolla estándares en el área de ingeniería eléctrica y computación. El comité 802 del IEEE ha estandarizado muchos tipos de LANs. Estudiaremos algunos de sus resultados más adelante. El trabajo real lo hace un conjunto de grupos de trabajo, que se listan en la figura 1.38. La tasa de éxito de los diversos grupos de trabajo del 802 ha sido baja; el hecho de tener un número 802.x no garantiza el éxito. Pero el impacto de las historias de éxito (en especial, del 802.3 y el 802.11) ha sido tremendo.

1.6.3 Quién es quién en el mundo de los estándares de Internet

El amplio mundo de Internet tiene sus propios mecanismos de estandarización, muy diferentes de los de la ITU-T y la ISO. La diferencia se puede resumir diciendo que quienes asisten a las reuniones de estandarización de la ITU o la ISO van de traje, pero las personas que asisten a las juntas de estandarización de Internet van de mezclilla (excepto cuando se reúnen en San Diego, donde van de *short* y camiseta).

En las reuniones de la ITU y la ISO abundan los oficiales corporativos y burócratas, para quienes la estandarización es su trabajo. Se refieren a la estandarización como una Cosa Buena y dedican sus vidas a ella. Por otro lado, la gente de Internet prefiere la anarquía por cuestión de principios. Sin embargo, con cientos de millones de personas haciendo sus propias cosas, la comunicación es escasa. Por lo tanto, los estándares, aunque deplorables, son necesarios.

Cuando se configuró ARPANET, el DoD creó un comité informal para supervisarla. En 1983 se dio otro nombre al comité: **IAB** (**Consejo de Actividades de Internet**) y se le encomendó una misión un poco más amplia, que era la de mantener a los investigadores de ARPANET y de Internet apuntando más o menos en la misma dirección; algo muy parecido a juntar una manada de gatos. El significado del acrónimo "IAB" se cambió a **Consejo para la Arquitectura de Internet.**

Número	Tema
802.1	Supervisión y arquitectura de LANs
802.2 ↓	Control lógico de enlace
802.3 *	Ethernet
802.4 ↓	Token bus (se utilizó por un corto tiempo en plantas manufactureras)
802.5	Token ring (entrada de IBM al mundo de las LANs)
802.6 ↓	Cola dual, bus dual (primera red de área metropolitana)
802.7 ↓	Grupo de consultoría técnico de tecnologías de banda ancha
802.8 †	Grupo de consultoría de tecnologías de fibra óptica
802.9 ↓	LANs síncrona (para aplicaciones de tiempo real)
802.10 ↓	LANs virtuales y seguridad
802.11 *	LANs inalámbricas
802.12 ↓	Demanda de prioridad (AnyLAN de Hewlett-Packard)
802.13	Número de mala suerte. Nadie lo quiso
802.14 ↓	Módems de cable (desaparecido: primero surgió un consorcio en la industria)
802.15 *	Redes de área personal (Bluetooth)
802.16 *	Redes inalámbricas de área ancha
802.17	Anillo de paquete elástico

Figura 1-38. Los grupos de trabajo del 802. Los importantes se marcan con *. Los que se marcan con ↓ están en hibernación. El que tiene la † se desintegró.

Cada uno de los aproximadamente 10 miembros del IAB encabezaba una fuerza de trabajo relacionada con algún asunto importante. El IAB se reunía varias veces al año para discutir los resultados y para dar retroalimentación al DoD y a la NSF, que proporcionaban la mayor parte de los fondos en aquel entonces. Cuando se necesitaba un estándar (por ejemplo, un nuevo algoritmo de enrutamiento), los miembros del IAB le daban solución y después anunciaban los cambios para que los estudiantes que estuvieran a cargo de la implementación del software pudieran realizarlos. La comunicación se llevaba a cabo mediante una serie de informes técnicos denominados **RFCs** (**Solicitudes de Comentarios**). Estos informes se almacenan en línea y cualquiera que esté interesado en ellos puede descargarlos de *www.ietf.org/rfc.* Los RFCs se encuentran organizados por el orden cronológico de su creación. Actualmente existen alrededor de 3000. En este libro mencionaremos muchos RFCs.

Para 1989 Internet había crecido tanto que este estilo sumamente informal dejó de ser funcional. Muchos fabricantes ofrecían productos de TCP/IP en ese entonces y no deseaban cambiarlos tan sólo porque 10 investigadores habían concebido una mejor idea. El IAB fue reorganizado de nueva cuenta en el verano de 1989. Los investigadores fueron transferidos a la **IRTF** (**Fuerza de Trabajo para la Investigación sobre Internet**), que fue puesta bajo el mando del IAB, junto con la **IETF** (**Fuerza de Trabajo para la Ingeniería de Internet**). El IAB se renovó con nuevos

miembros, que representaban a un rango más amplio de organizaciones que tan sólo a la comunidad de investigadores. Inicialmente fue un grupo que se autorrenovaba, cuyos miembros servían durante dos años y ellos mismos designaban a sus sucesores. Más tarde se creó la **Sociedad de Internet,** integrada por gente interesada en Internet. En cierto sentido, esta sociedad se asemeja al ACM o al IEEE. Es dirigida por administradores electos que designan a los miembros del IAB.

El propósito de esta división era que la IRTF se concentrara en proyectos de investigación a largo plazo, en tanto que la IETF se encargaba de proyectos de ingeniería a corto plazo. La IETF se dividió en grupos de trabajo, cada uno a cargo de un problema específico. Inicialmente, los líderes de cada grupo se reunían en un comité para dirigir los proyectos de ingeniería. Entre los temas de los grupos de trabajo están nuevas aplicaciones, información de usuario, integración OSI, enrutamiento y direccionamiento, seguridad, administración de redes y estándares. Con el tiempo se formaron tantos grupos de trabajo (más de 70), que se ordenaron por áreas y el líder de cada área formaba parte del comité directivo.

Además, se adoptó un proceso de estandarización más formal, con base en el ISO. Para convertirse en **Estándar Propuesto,** la idea fundamental debía explicarse completamente en un RFC y despertar suficiente interés en la comunidad. Para avanzar a la etapa de **Estándar Borrador,** una implementación funcional debía haber sido rigurosamente probada por al menos dos sitios independientes durante al menos cuatro meses. Si el IAB se convence de que la idea suena lógica y el software funciona, declara que el RFC es un Estándar de Internet. Algunos de estos estándares se han convertido en estándares del DoD (MIL-STD), con lo cual son obligatorios para los proveedores del DoD. En cierta ocasión, David Clark hizo el siguiente comentario, ahora famoso, acerca del proceso de estandarización de Internet: “consenso apretado y código que funcione”.

1.7 UNIDADES MÉTRICAS

Para evitar confusiones, vale la pena indicar de manera explícita que en este libro, como en las ciencias de la computación en general, se utilizan unidades métricas en lugar de las unidades inglesas tradicionales. En la figura 1.39 se muestran los principales prefijos del sistema métrico. Por lo general, los prefijos se abrevian mediante sus primeras letras, con las unidades mayores que uno en mayúsculas (KB, MB, etcétera). Una excepción (por razones históricas) es kbps, de kilobits por segundo. Por lo tanto, una línea de comunicación de 1 Mbps transmite 10^6 bits por segundo y un reloj de 100 pseg (o 100 ps) marca cada 10^{-10} segundos. Dado que tanto mili como micro empiezan con la letra “m”, se tiene que hacer una distinción. Por lo general, “m” es para mili y “μ” (la letra griega mu) es para micro.

También vale la pena señalar que para medir el tamaño de la memoria, de disco, de archivo y de bases de datos, en la práctica común de la industria de la computación las unidades tienen equivalencias ligeramente distintas. En ésta, kilo equivale a 2^{10} (1024) en vez de 10^3 (1000) porque las memorias siempre son potencias de dos. De esta forma, 1 KB de memoria son 1024 bytes, no 1000 bytes. De manera similar, 1 MB de memoria son 2^{20} (1,048,576) bytes, 1 GB de memoria son 2^{30} (1,073,741,824) bytes, y 1 TB de base de datos son 2^{40} (1,099,511,627,776) bytes. No obstante, una línea de comunicación de 1 kbps transmite 1000 bits por segundo y una LAN de

Exp.	Explícito	Prefijo	Exp.	Explícito	Prefijo
10^{-3}	0.001	mili	10^{3}	1,000	Kilo
10^{-6}	0.000001	micro	10^{6}	1,000,000	Mega
10^{-9}	0.000000001	nano	10^{9}	1,000,000,000	Giga
10^{-12}	0.000000000001	pico	10^{12}	1,000,000,000,000	Tera
10^{-15}	0.000000000000001	femto	10^{15}	1,000,000,000,000,000	Peta
10^{-18}	0.000000000000000001	atto	10^{18}	1,000,000,000,000,000,000	Exa
10^{-21}	0.000000000000000000001	zepto	10^{21}	1,000,000,000,000,000,000,000	Zeta
10^{-24}	0.000000000000000000000001	yocto	10^{24}	1,000,000,000,000,000,000,000,000	Yotta

Figura 1-39. Los principales prefijos métricos.

10 Mbps corre a 10,000,000 de bits por segundo debido a que estas velocidades no son potencias de dos. Desgraciadamente, mucha gente mezcla estos dos sistemas, en particular en lo referente a los tamaños de disco. Para evitar la ambigüedad, en este libro utilizaremos los símbolos KB, MB y GB para 2^{10}, 2^{20} y 2^{30} bytes, respectivamente, y los símbolos kbps, Mbps y Gbps para 10^{3}, 10^{6} y 10^{9} bits por segundo, respectivamente.

1.8 PANORAMA DEL RESTO DEL LIBRO

Este libro estudia tanto los principios como la práctica de las redes de computadoras. La mayoría de los capítulos inician con un análisis de los principios relevantes, seguido por diversos ejemplos que ilustran estos principios. Por lo general, los ejemplos se toman de Internet y de las redes inalámbricas puesto que ambos son importantes y muy distintos. Donde es necesario, se dan otros ejemplos.

El libro se estructura de acuerdo con el modelo híbrido que se presenta en la figura 1-24. El análisis de la jerarquía de protocolos empieza en el capítulo 2, a partir de la capa más baja. El segundo capítulo proporciona algunos antecedentes en el campo de la comunicación de datos. Se presentan sistemas alámbricos, inalámbricos y satelitales. Este material se relaciona con la capa física, aunque veremos únicamente los aspectos de arquitectura y no los de hardware. También se analizan numerosos ejemplos de la capa física, como la red telefónica pública conmutada, la telefonía celular y la red de televisión por cable.

En el capítulo 3 se presenta la capa de enlace de datos y sus protocolos a través de ejemplos que crecen en complejidad. También se cubre el análisis de estos protocolos. Más tarde se examinan algunos protocolos importantes que se usan con mucha frecuencia, entre ellos HDLC (que se emplea en redes de baja y mediana velocidad) y PPP (que se utiliza en Internet).

El capítulo 4 tiene que ver con la subcapa de acceso al medio, que forma parte de la capa de enlace de datos. El aspecto principal al que se enfrenta esta subcapa es cómo determinar quién uti-

lizará la red cuando ésta consiste en un solo canal compartido, como ocurre en la mayoría de las LANs y en algunas redes satelitales. Se dan muchos ejemplos de LANs alámbricas, LANs inalámbricas (en especial Ethernet), MANs inalámbricas, Bluetooth y redes satelitales. También se analizan los puentes y los conmutadores de enlace de datos.

El capítulo 5 aborda la capa de red, en particular el enrutamiento, con muchos algoritmos de enrutamiento, tanto estáticos como dinámicos. Aun con el uso de buenos algoritmos de enrutamiento, si existe más tráfico del que puede manejar la red, se genera congestión, por lo que analizaremos el tema de la congestión y cómo evitarla. Es aún mejor garantizar una calidad específica en el servicio que tan sólo evitar la congestión. También analizaremos este punto. La conexión de redes heterogéneas para conformar interredes acarrea numerosos problemas que también examinaremos. Daremos una amplia cobertura a la capa de red en Internet.

El capítulo 6 se encarga de la capa de transporte. Gran parte del capítulo se dedica a los protocolos orientados a la conexión, puesto que muchas aplicaciones los necesitan. Se analiza en detalle un ejemplo de servicio de transporte y su implementación. Se proporciona el código para este sencillo ejemplo con el propósito de mostrar cómo se puede implementar. Tanto UDP como TCP, protocolos de transporte de Internet, se abordan en detalle, al igual que sus aspectos de desempeño. Asimismo, veremos aspectos relacionados con las redes inalámbricas.

El capítulo 7 presenta la capa de aplicación, sus protocolos y aplicaciones. El primer tema es el DNS, que es el directorio telefónico de Internet. A continuación trataremos el correo electrónico, junto con un análisis de sus protocolos. Más adelante pasaremos a Web, con explicaciones minuciosas sobre contenido estático, contenido dinámico, lo que sucede tanto en el cliente como en el servidor, protocolos, rendimiento, la Web inalámbrica, entre otros temas. Por último, examinaremos la multimedia en red, con temas como audio de flujo continuo, radio en Internet y vídeo bajo demanda.

El capítulo 8 se relaciona con la seguridad de red. Este tema tiene aspectos que se relacionan con todas las capas, por lo cual es más sencillo abordarlo después de haber explicado minuciosamente todas las capas. El capítulo inicia con una introducción a la criptografía. Más adelante muestra cómo se puede utilizar ésta para garantizar la seguridad en las comunicaciones, el correo electrónico y Web. El libro finaliza con un análisis de algunas áreas en las cuales la seguridad afecta la privacidad, la libertad de expresión, la censura y otros aspectos sociales con los cuales choca directamente.

El capítulo 9 contiene listas de lecturas sugeridas, con comentarios, organizadas por capítulo. Su propósito es ayudar a los lectores que deseen llevar más allá el estudio sobre las redes. El capítulo también tiene una bibliografía alfabética de todas las referencias que se dan en el libro.

El sitio Web del autor puede consultarlo desde:

http://www.pearsonedlatino.com/tanenbaum

el cual contiene una página con vínculos a muchos tutoriales, FAQs, compañías, consorcios industriales, organizaciones profesionales, organizaciones de estándares, tecnologías, documentos y muchas cosas más.

1.9 RESUMEN

Las redes de computadoras se pueden utilizar para diversos servicios, tanto para compañías como para individuos. Para las compañías, las redes de computadoras personales que utilizan servidores compartidos con frecuencia dan acceso a información corporativa. Por lo general, estas redes siguen el modelo cliente-servidor, con estaciones de trabajo clientes en los escritorios de los empleados que acceden a servidores instalados en la sala de máquinas. Para los individuos, las redes ofrecen acceso a una diversidad de recursos de información y entretenimiento. Los individuos acceden a Internet mediante una llamada al ISP a través de un módem, aunque una cantidad creciente de usuarios cuenta con una conexión fija en casa. Un área con gran futuro es la de las redes inalámbricas, con nuevas aplicaciones como acceso móvil al correo electrónico y el comercio móvil.

A grandes rasgos, las redes se pueden dividir en LANs, MANs, WANs e interredes, con sus propias características, tecnologías, velocidades y nichos. Las LANs ocupan edificios y operan a altas velocidades. Las MANs abarcan toda una ciudad, por ejemplo, el sistema de televisión por cable, el cual es utilizado por mucha gente para acceder a Internet. Las WANs se extienden por un país o un continente. Las LANs y las MANs pueden ser o no conmutadas (es decir, no tienen enrutadores); las WANs son conmutadas. Las redes inalámbricas se están volviendo sumamente populares, en especial las LANs inalámbricas. Las redes se interconectan para formar interredes.

El software de red consta de protocolos, que son reglas mediante las cuales se comunican los procesos. Los protocolos son de dos tipos: orientados a la conexión y no orientados a la conexión. La mayoría de las redes soporta jerarquías de protocolos, en la cual cada capa proporciona servicios a las capas superiores a ella y las libera de los detalles de los protocolos que se utilizan en las capas inferiores. Las pilas de protocolos se basan generalmente en el modelo OSI o en el modelo TCP/IP. Ambos modelos tienen capas de red, de transporte y de aplicación, pero difieren en las demás capas. Entre los aspectos de diseño están la multiplexión, el control de flujo y el control de errores. Gran parte del libro está dedicada a los protocolos y su diseño.

Las redes ofrecen servicios a sus usuarios. Los servicios pueden ser orientados a la conexión o no orientados a ésta. En algunas redes se proporciona servicio no orientado a la conexión en una capa y servicio orientado a la conexión en la capa superior.

Las redes bien conocidas incluyen Internet, ATM, Ethernet y la LAN IEEE 802.11 inalámbrica. Internet evolucionó de ARPANET, a la cual se agregaron otras redes para conformar una interred. La Internet actual es en realidad un conjunto de miles de redes, más que de una sola red. El aspecto que la distingue es el uso generalizado de la pila de protocolos TCP/IP. ATM tiene un uso muy extendido en los sistemas telefónicos para el tráfico de datos de larga distancia. Ethernet es la LAN más popular y se utiliza en la mayoría de las compañías y universidades. Por último, las LANs inalámbricas a velocidades sorprendentemente altas (hasta 54 Mbps) comienzan a desplegarse en forma masiva.

Para que varias computadoras se comuniquen entre sí es necesaria una gran cantidad de estandarización, tanto en el software como en el hardware. Organizaciones como la ITU-T, el ISO, el IEEE y el IAB manejan diferentes partes del proceso de estandarización.

PROBLEMAS

1. Imagine que ha entrenado a su San Bernardo, Byron, para que transporte una caja con tres cintas de 8 mm en lugar del barrilito de brandy. (Cuando se llene su disco, usted tendrá una emergencia.) Cada una de estas cintas tiene capacidad de 7 gigabytes. El perro puede trasladarse adondequiera que usted vaya, a una velocidad de 18 km/hora. ¿Para cuál rango de distancias tiene Byron una tasa de datos más alta que una línea de transmisión cuya tasa de datos (sin tomar en cuenta la sobrecarga) es de 150 Mbps?
2. Una alternativa a una LAN es simplemente un enorme sistema de compartición de tiempo con terminales para todos los usuarios. Mencione dos ventajas de un sistema cliente-servidor que utilice una LAN.
3. Dos factores de red ejercen influencia en el rendimiento de un sistema cliente-servidor: el ancho de banda de la red (cuántos bits por segundo puede transportar) y la latencia (cuánto tiempo toma al primer bit llegar del cliente al servidor). Mencione un ejemplo de una red que cuente con ancho de banda y latencia altos. A continuación, mencione un ejemplo de una que cuente con ancho de banda y latencia bajos.
4. ¿Además del ancho de banda y la latencia, qué otros parámetros son necesarios para dar un buen ejemplo de la calidad de servicio ofrecida por una red destinada a tráfico de voz digitalizada?
5. Un factor en el retardo de un sistema de conmutación de paquetes de almacenamiento y reenvío es el tiempo que le toma almacenar y reenviar un paquete a través de un conmutador. Si el tiempo de conmutación es de 10 μseg, ¿esto podría ser un factor determinante en la respuesta de un sistema cliente-servidor en el cual el cliente se encuentre en Nueva York y el servidor en California? Suponga que la velocidad de propagación en cobre y fibra es 2/3 de la velocidad de la luz en el vacío.
6. Un sistema cliente-servidor utiliza una red satelital, con el satélite a una altura de 40,000 km. ¿Cuál es el retardo en respuesta a una solicitud, en el mejor de los casos?
7. En el futuro, cuando cada persona tenga una terminal en casa conectada a una red de computadoras, serán posibles las consultas públicas instantáneas sobre asuntos legislativos pendientes. Con el tiempo, las legislaturas existentes podrían eliminarse, para dejar que la voluntad popular se exprese directamente. Los aspectos positivos de una democracia directa como ésta son bastante obvios; analice algunos de los aspectos negativos.
8. Cinco enrutadores se van a conectar en una subred de punto a punto. Los diseñadores podrían poner una línea de alta velocidad, de mediana velocidad, de baja velocidad o ninguna línea, entre cada par de enrutadores. Si toma 100 ms de tiempo de la computadora generar e inspeccionar cada topología, ¿cuánto tiempo tomará inspeccionarlas todas?
9. Un grupo de $2^n - 1$ enrutadores están interconectados en un árbol binario centralizado, con un enrutador en cada nodo del árbol. El enrutador *i* se comunica con el enrutador *j* enviando un mensaje a la raíz del árbol. A continuación, la raíz manda el mensaje al enrutador *j*. Obtenga una expresión aproximada de la cantidad media de saltos por mensaje para un valor grande de *n,* suponiendo que todos los pares de enrutadores son igualmente probables.
10. Una desventaja de una subred de difusión es la capacidad que se desperdicia cuando múltiples *hosts* intentan acceder el canal al mismo tiempo. Suponga, por ejemplo, que el tiempo se divide en ranuras discretas, y que cada uno de los *hosts n* intenta utilizar el canal con probabilidad *p* durante cada parte. ¿Qué fracción de las partes se desperdicia debido a colisiones?

11. Mencione dos razones para utilizar protocolos en capas.

12. Al presidente de Specialty Paint Corp. se le ocurre la idea de trabajar con una compañía cervecera local para producir una lata de cerveza invisible (como medida para reducir los desechos). El presidente indica a su departamento legal que analice la situación, y éste a su vez pide ayuda al departamento de ingeniería. De esta forma, el ingeniero en jefe se reúne con su contraparte de la otra compañía para discutir los aspectos técnicos del proyecto. A continuación, los ingenieros informan los resultados a sus respectivos departamentos legales, los cuales a su vez se comunican vía telefónica para ponerse de acuerdo en los aspectos legales. Por último, los dos presidentes corporativos se ponen de acuerdo en la parte financiera del proyecto. ¿Éste es un ejemplo de protocolo con múltiples capas semejante al modelo OSI?

13. ¿Cuál es la diferencia principal entre comunicación orientada a la conexión y no orientada a ésta?

14. Dos redes proporcionan servicio confiable orientado a la conexión. Una de ellas ofrece un flujo confiable de bytes y la otra un flujo confiable de mensajes. ¿Son idénticas? Si es así, ¿por qué se hace la distinción? Si no son idénticas, mencione un ejemplo de algo en que difieran.

15. ¿Qué significa "negociación" en el contexto de protocolos de red? Dé un ejemplo.

16. En la figura 1-19 se muestra un servicio. ¿Hay algún otro servicio implícito en la figura? Si es así, ¿dónde? Si no lo hay, ¿por qué no?

17. En algunas redes, la capa de enlace de datos maneja los errores de transmisión solicitando que se retransmitan las tramas dañadas. Si la probabilidad de que una trama se dañe es *p,* ¿cuál es la cantidad media de transmisiones requeridas para enviar una trama? Suponga que las confirmaciones de recepción nunca se pierden.

18. ¿Cuál de las capas OSI maneja cada uno de los siguientes aspectos?:
(a) Dividir en tramas el flujo de bits transmitidos.
(b) Determinar la ruta que se utilizará a través de la subred.

19. Si la unidad que se transmite al nivel de enlace de datos se denomina trama y la que se transmite al nivel de red se llama paquete, ¿las tramas encapsulan paquetes o los paquetes encapsulan tramas? Explique su respuesta?

20. Un sistema tiene una jerarquía de protocolos de *n* capas. Las aplicaciones generan mensajes con una longitud de *M* bytes. En cada una de las capas se agrega un encabezado de *h* bytes. ¿Qué fracción del ancho de banda de la red se llena con encabezados?

21. Mencione dos similitudes entre los modelos de referencia OSI y TCP/IP. A continuación mencione dos diferencias entre ellos.

22. ¿Cuál es la principal diferencia entre TCP y UDP?

23. La subred de la figura 1-25(b) se diseñó para resistir una guerra nuclear. ¿Cuántas bombas serían necesarias para partir los nodos en dos conjuntos inconexos? Suponga que cualquier bomba destruye un nodo y todos los enlaces que se conectan a él.

24. Internet está duplicando su tamaño aproximadamente cada 18 meses. Aunque no se sabe a ciencia cierta, una estimación indica que en el 2001 había 100 millones de *hosts* en Internet. Utilice estos datos para calcular la cantidad esperada de *hosts* para el año 2010. ¿Cree que esto es real? Explique por qué.

25. Cuando un archivo se transfiere entre dos computadoras, pueden seguirse dos estrategias de confirmación de recepción. En la primera, el archivo se divide en paquetes, y el receptor confirma la recepción de cada uno de manera individual, aunque no confirma la recepción del archivo como un todo. En contraste, en la segunda estrategia la recepción de los paquetes no se confirma de manera individual, sino la del archivo completo. Comente las dos estrategias.

26. ¿Por qué ATM utiliza celdas pequeñas de longitud fija?

27. ¿Qué tan grande era un bit, en metros, en el estándar 802.3 original? Utilice una velocidad de transmisión de 10 Mbps y suponga que la velocidad de propagación en cable coaxial es 2/3 la velocidad de la luz en el vacío.

28. Una imagen tiene 1024×768 píxeles con 3 bytes/píxel. Suponga que la imagen no se encuentra comprimida. ¿Cuánto tiempo tomará transmitirla sobre un canal de módem de 56 kpbs? ¿Sobre un módem de cable de 1 Mbps? ¿Sobre una red Ethernet a 10 Mbps? ¿Sobre una red Ethernet a 100 Mbps?

29. Ethernet y las redes inalámbricas tienen algunas similitudes y diferencias. Una propiedad de Ethernet es que sólo se puede transmitir una trama a la vez sobre una red de este tipo. ¿El 802.11 comparte esta propiedad con Ethernet? Comente su respuesta.

30. Las redes inalámbricas son fáciles de instalar, y ello las hace muy económicas puesto que los costos de instalación eclipsan por mucho los costos del equipo. No obstante, también tienen algunas desventajas. Mencione dos de ellas.

31. Cite dos ventajas y dos desventajas de contar con estándares internacionales para los protocolos de red.

32. Cuando un sistema tiene una parte fija y una parte removible (como ocurre con una unidad de CD-ROM y el CD-ROM), es importante que exista estandarización en el sistema, con el propósito de que las diferentes compañías puedan fabricar tanto la parte removible como la fija y todo funcione en conjunto. Mencione tres ejemplos ajenos a la industria de la computación en donde existan estándares internacionales. Ahora mencione tres áreas donde no existan.

33. Haga una lista de sus actividades cotidianas en las cuales intervengan las redes de computadoras. ¿De qué manera se alteraría su vida si estas redes fueran súbitamente desconectadas?

34. Averigüe cuáles redes se utilizan en su escuela o lugar de trabajo. Describa los tipos de red, las topologías y los métodos de conmutación que utilizan.

35. El programa *ping* le permite enviar un paquete de prueba a un lugar determinado y medir cuánto tarda en ir y regresar. Utilice *ping* para ver cuánto tiempo toma llegar del lugar donde se encuentra hasta diversos lugares conocidos. Con los resultados, trace el tiempo de tránsito sobre Internet como una función de la distancia. Lo más adecuado es utilizar universidades, puesto que la ubicación de sus servidores se conoce con mucha precisión. Por ejemplo, *berkeley.edu* se encuentra en Berkeley, California; *mit.edu* se localiza en Cambridge, Massachusetts; *vu.nl* está en Amsterdam, Holanda; *www.usyd.edu.au* se encuentra en Sydney, Australia, y *www.uct.ac.za* se localiza en Cape Town, Sudáfrica.

36. Vaya al sitio Web de la IETF, *www.ietf.org,* y entérese de lo que hacen ahí. Elija un proyecto y escriba un informe de media página acerca del problema y la solución que propone.

37. La estandarización es sumamente importante en el mundo de las redes. La ITU y la ISO son las principales organizaciones oficiales encargadas de la estandarización. Vaya a los sitios Web de estas organiza-

ciones, en *www.itu.org* y *www.iso.org,* respectivamente, y analice el trabajo de estandarización que realizan. Escriba un breve informe sobre las cosas que han estandarizado.

38. Internet está conformada por una gran cantidad de redes. Su disposición determina la topología de Internet. En línea se encuentra una cantidad considerable de información acerca de la topología de Internet. Utilice un motor de búsqueda para investigar más sobre la topología de Internet y escriba un breve informe sobre sus resultados.