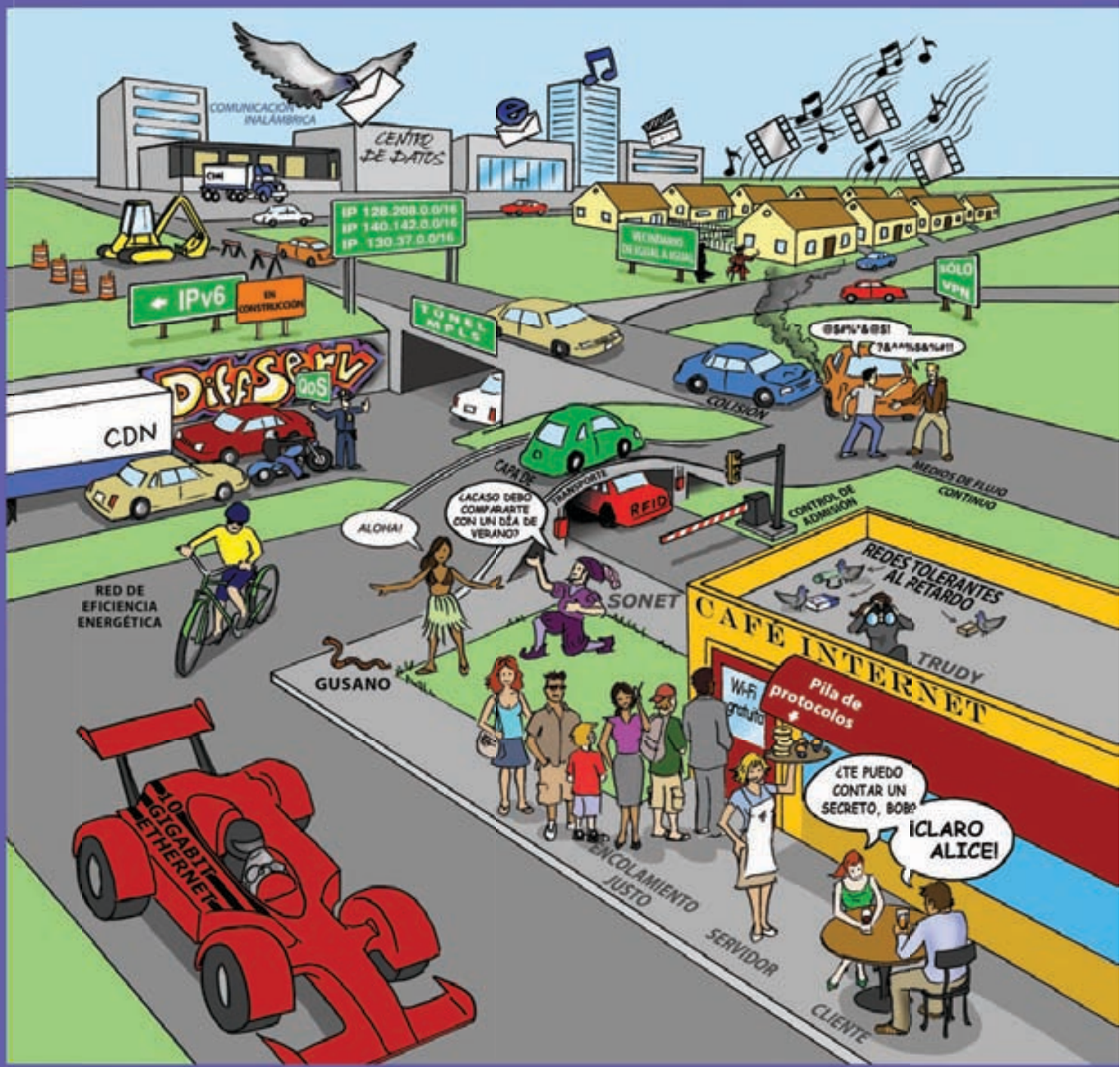
REDES DE COMPUTADORAS

QUINTA EDICIÓN

TANENBAUM | WETHERALL



ALWAYS LEARNING

PEARSON

Redes de computadoras

Quinta edición

Redes de computadoras

Quinta edición

Andrew S. Tanenbaum
Vrije Universiteit
Amsterdam

David J. Wetherall
University of Washington
Seatle, Washington

TRADUCCIÓN

Alfonso Vidal Romero Elizondo
Ingeniero en Sistemas Electrónicos
Instituto Tecnológico y de Estudios
Superiores de Monterrey-Campus Monterrey

REVISIÓN TÉCNICA

M. en C. Cyntia E. Enríquez Ortiz
Escuela Superior de Cómputo
Instituto Politécnico Nacional

PEARSON

Datos de catalogación bibliográfica

ANDREW S. TANENBAUM y DAVID J. WETHERALL

Redes de computadoras
Quinta edición

PEARSON EDUCACIÓN, México, 2012
ISBN: 978-607-32-0817-8
Área: Computación

Formato: 20 × 25.5 cm Páginas: 816

Edición en español
Editor: Luis M. Cruz Castillo
e-mail: luis.cruz@pearson.com
Editor de desarrollo: Bernardino Gutiérrez Hernández
Supervisor de producción: Juan José García Guzmán

QUINTA EDICIÓN, 2012

ISBN VERSIÓN IMPRESA: 978-607-32-0817-8
ISBN VERSIÓN E-BOOK: 978-607-32-0818-5
ISBN E-CHAPTER: 978-607-32-0819-2

Impreso en México. Printed in Mexico.
1 2 3 4 5 6 7 8 9 0 - 14 13 12 11

PEARSON

www.pearsoneducacion.net

Para Suzanne, Barbara, Daniel, Aron, Marvin, Matilde
y a la memoria de Bram y Sweetie
(AST)

Para Katrim, Lucy y Pepper
(DJW)

CONTENIDO

Prefacio

Este libro se encuentra ahora en su quinta edición. Cada edición ha correspondido a una fase distinta en cuanto a la forma en que se utilizaban las redes de computadoras. Cuando apareció la primera edición en 1980, las redes eran una curiosidad académica. Para la segunda edición, en 1988, las redes se utilizaban en las universidades y en las grandes empresas. Cuando salió al mercado la tercera edición, en 1996, las redes de computadoras (en especial Internet) se habían convertido en una realidad diaria para millones de personas. Ya para la cuarta edición, en 2003, las redes inalámbricas y las computadoras móviles se habían vuelto herramientas comunes para acceder a la web e Internet. Ahora, en la quinta edición, las redes tratan sobre la distribución de contenido (en especial los videos que utilizan CDN y redes de igual a igual) y los teléfonos móviles son pequeñas computadoras con Internet.

Novedades de la quinta edición

Entre los diversos cambios que se presentan en este libro, el más importante es la incorporación del profesor David J. Wetherall como coautor. El profesor Wetherall posee una extensa experiencia con las redes, tiene más de 20 años experimentando con las redes de área metropolitana. Desde entonces ha trabajado con las redes inalámbricas e Internet, además de fungir como profesor en la University of Washington, en donde ha enseñado y realizado investigaciones sobre las redes de computadoras y temas relacionados durante la última década.

Desde luego, el libro también incluye cambios sustanciales para estar a la par con el siempre cambiante mundo de las redes computacionales. Algunos de estos cambios incluyen material actualizado y nuevo sobre:

- Redes inalámbricas (802.12 y 802.16).
- Las redes 3G que utilizan los teléfonos inteligentes.

- Redes RFID y de sensores.
- Distribución de contenido mediante el uso de CDN.
- Redes de igual a igual.
- Medios en tiempo real (de fuentes almacenadas, de flujo continuo y en vivo).
- Telefonía por Internet (voz sobre IP).
- Redes tolerantes al retraso.

A continuación encontrará una descripción más detallada por capítulo.

El capítulo 1 tiene la misma función de presentación que en la cuarta edición, pero revisamos y actualizamos el contenido. Aquí hablamos sobre Internet, las redes de teléfonos móviles, 802.11, las redes RFID y de sensores como ejemplos de redes computacionales. Eliminamos el material sobre la Ethernet original (con sus conexiones tipo vampiro), junto con el material sobre ATM.

El capítulo 2, que trata sobre la capa física, cuenta con una cobertura más amplia de la modulación digital (incluyendo la multiplexación OFDM y su popularidad en las redes inalámbricas) y las redes 3G (basadas en CDMA). También se describen las nuevas tecnologías, incluyendo *Fiber to Home* o FTTH (fibra hasta el hogar) y las redes a través del cableado eléctrico.

El capítulo 3, que trata sobre los enlaces punto a punto, se mejoró de dos formas. Se actualizó el material sobre los códigos para detección y corrección de errores, además de incluir una breve descripción de los códigos modernos importantes en la práctica (por ejemplo, los códigos convolucional y LDPC). Los ejemplos de protocolos utilizan ahora un paquete sobre SONET y ADSL. Tuvimos que eliminar el material sobre verificación de protocolos debido a que en la actualidad no se utiliza mucho.

En el capítulo 4, sobre la subcapa MAC, los principios son eternos pero las tecnologías han cambiado. Se rediseñaron las secciones sobre las redes de ejemplo de manera acorde, incluyendo las redes Gigabit Ethernet, 802.11, 802.16, Bluetooth y RFID. También se actualizó la información sobre las redes LAN conmutadas, incluyendo las redes VLAN.

El capítulo 5, que trata sobre la capa de red, cubre los mismos conceptos que en la cuarta edición. Se hicieron revisiones para actualizar el material y agregar más detalles, en especial sobre la calidad del servicio (lo cual es relevante para los medios en tiempo real) y la interconectividad (*internetworking*). Se expandieron las secciones sobre BGP, OSPF y CIDR, así como el material sobre el enrutamiento multidifusión (*multicast*). Ahora se incluye también el enrutamiento *anycast*.

En el capítulo 6, sobre la capa de transporte, se agregó, modificó y eliminó material. El nuevo material describe las redes tolerantes al retardo y el control de congestión en general. El material modificado actualiza y expande la cobertura sobre el control de congestión en TCP. El material eliminado describe las capas de red orientadas a la conexión, algo que se ve rara vez en la actualidad.

En el capítulo 7, que trata sobre aplicaciones, también se actualizó la información y se aumentó el contenido. Aunque el material sobre DNS y correo electrónico es similar al de la cuarta edición, en los últimos años se han suscitado varios acontecimientos en cuanto al uso de web, los medios de flujo continuo y la distribución de contenido. Asimismo, se actualizaron las secciones sobre web y los medios de flujo continuo. Hay una nueva sección sobre la distribución de contenido, incluyendo las redes CDN y de igual a igual.

El capítulo 8, sobre la seguridad, trata aún la criptografía tanto simétrica como de clave pública para la confidencialidad y autenticidad. Se actualizó el material sobre las técnicas utilizadas en la práctica, incluyendo *firewalls* y redes VPN; además se agregó material nuevo sobre la seguridad en redes 802.11 y Kerberos V5.

El capítulo 9 contiene una lista renovada de lecturas sugeridas y una extensa bibliografía con más de 300 citas de literatura actual. Más de la mitad de éstas son de artículos y libros, mientras que el resto son citas de artículos clásicos.

Lista de siglas y acrónimos

Los libros de computación están llenos de acrónimos y éste no es la excepción. Para cuando termine de leerlo, los siguientes acrónimos le serán familiares: ADSL, AES, AJAX, AODV, AP, ARP, ARQ, AS, BGP, BOC, CDMA, CDN, CGI, CIDR, CRL, CSMA, CSS, DCT, DES, DHCP, DHT, DIFS, DMCA, DMT, DMZ, DNS, DOCSIS, DOM, DSLAM, DTN, FCFS, FDD, FDDI, FDM, FEC, FIFO, FSK, FTP, GPRS, GSM, HDTV, HFC, HMAC, HTTP, IAB, ICANN, ICMP, IDEA, IETF, IMAP, IMP, IP, IPTV, IRTF, ISO, ISP, ITU, JPEG, JSP, JVM, LAN, LATA, LEC, LEO, LLC, LSR, LTE, MAN, MFJ, MIME, MPEG, MPLS, MSC, MTSO, MTU, NAP, NAT, NRZ, NSAP, OFDM, OSI, OSPF, PAWS, PCM, PGP, PIM, PKI, POP, POTS, PPP, PSTN, QAM, QPSK, RED, RFC, RFID, RPC, RSA, RTSP, SHA, SIP, SMTP, SNR, SOAP, SONET, SPE, SSL, TCP, TDD, TDM, TSAP, UDP, UMTS, URL, VLAN, VSAT, WAN, WDM y XML. No se preocupe; cada uno aparecerá en **negritas** y lo definiremos con detalle antes de usarlo. Hagamos una prueba divertida: revise cuántas de estas siglas puede identificar *antes* de leer el libro, escriba el número al margen y vuelva a intentarlo *después* de leer el libro.

Cómo usar este libro

Para ayudar a los profesores a utilizar este libro como texto para cursos cuya duración puede variar entre un trimestre o un semestre, estructuramos los subtítulos de los capítulos como: material básico y opcional. Las secciones marcadas con un asterisco (*) en el contenido son *opcionales*. Si hay una sección importante marcada de esta forma (vea por ejemplo la sección 2.7), entonces todas sus subsecciones son opcionales, ya que proveen material sobre tecnologías de redes que es útil pero se puede omitir en un curso corto sin perder la continuidad. Desde luego que hay que animar a los estudiantes a que lean también esas secciones, siempre y cuando tengan tiempo suficiente, ya que todo el material está actualizado y es valioso.

Materiales didácticos para los profesores

Los siguientes materiales didácticos (en inglés) "protegidos para los profesores" están disponibles en *www.pearsoneducacion.net/tanenbaum*, sitio web creado para este libro. Para obtener un nombre de usuario y contraseña, póngase en contacto con su representante local de Pearson.

- Manual de soluciones.
- Diapositivas en PowerPoint.

Materiales didácticos para los estudiantes

Los recursos para los estudiantes (en inglés) están disponibles también en el sitio web de este libro: *www.pearsoneducacion.net/tanenbaum*, a través del vínculo *Companion Website*, e incluyen:

- Recursos web.
- Figuras, tablas y programas del libro.
- Demostración de esteganografía.

AGRADECIMIENTOS

Muchas personas nos ayudaron durante el desarrollo de esta quinta edición. Nos gustaría agradecer en especial a Emmanuel Agu (Worcester Polytechnic Institute), Yoris Au (University of Texas en San Antonio), Nikhil Bhargava (Aircom International, Inc.), Michael Buettner (University of Washington), John Day (Boston University), Kevin Fall (Intel Labs), Ronald Fulle (Rochester Institute of Technology), Ben Greenstein (Intel Labs), Daniel Halperin (University of Washington), Bob Kinicki (Worcester Polytechnic Institute), Tadayoshi Kohno (University of Washington), Sarvish Kulkarni (Villanova University), Hank Levy (University of Washington), Ratul Mahajan (Microsoft Research), Craig Partridge (BBN), Michael Piatek (University of Washington), Joshua Smith (Intel Labs), Neil Spring (University of Maryland), David Teneyuca (University of Texas en San Antonio), Tammy VanDegrift (University of Portland) y Bo Yuan (Rochester Polytechnic Institute), por aportar ideas y retroalimentación. Melody Kadenko y Julie Svendsen brindaron apoyo administrativo al profesor Wetherall.

Shivakant Mishra (University of Colorado en Boulder) y Paul Nagin (Chimborazo Publishing, Inc) idearon muchos de los problemas nuevos y retadores de fin de capítulo. Tracy Dunkelberger, nuestra editora en Pearson, fue tan útil como siempre en muchas tareas tanto grandes como pequeñas. Melinda Haggerty y Jeff Holcomb hicieron un excelente trabajo al cuidar que todo se llevara a cabo sin problemas. Steve Armstrong (LeTourneau University) preparó las diapositivas de PowerPoint. Stephen Turner (University of Michigan en Flint) revisó meticulosamente los recursos web y los simuladores que acompañan el libro. Rachel Head, nuestra correctora de estilo, es una híbrido fuera de lo común: tiene la vista de águila y la memoria de un elefante. Después de leer sus correcciones, ambos autores nos preguntamos cómo fue posible que pasáramos del tercer grado de primaria.

Por último, llegamos a las personas más importantes. Suzanne ha pasado por esto 19 veces hasta ahora y aún sigue con su paciencia y amor interminables. Barbara y Marvin ahora conocen la diferencia entre los buenos libros de texto y los malos libros que siempre son una inspiración para producir buenos. Daniel y Matilde son los miembros más recientes de nuestra familia, a

quienes recibimos con gran afecto. Es poco probable que Aron vaya a leer pronto este libro, pero de todas formas le gustan las bonitas imágenes de la figura 8-54. Katrin y Lucy brindaron su apoyo incondicional y siempre lograron mantener una sonrisa en mi rostro. Gracias.

ANDREW S. TANENBAUM
DAVID J. WETHERALL

ACERCA DE LOS AUTORES

Andrew S. Tanenbaum tiene una licenciatura en Ciencias (S.B.) por el MIT y un doctorado por la University of California, en Berkeley. Actualmente es profesor de Ciencias Computacionales en la Vrije Universiteit, en donde ha enseñado durante más de 30 años sobre sistemas operativos, redes y temas relacionados. Su investigación actual es sobre los sistemas operativos muy confiables, aunque ha trabajado también en compiladores, sistemas distribuidos, segu-ridad y otros temas. Estos proyectos de investigación han generado más de 150 artículos de referencia en publicaciones especializadas e infinidad de conferencias.

El profesor Tanenbaum también ha sido coautor de cinco libros que a la fecha han aparecido en 19 ediciones. Sus libros se han traducido a 21 idiomas, que varían desde el vasco hasta el tailandés y se utilizan en universidades de todo el mundo; en total, hay 159 versiones (combinadas idiomas y ediciones) que se listan en *www.cs.vu.nl/~ast/publications*.

El profesor Tanenbaum también ha producido un volumen considerable de software, incluyendo el paquete de compilador Amsterdam (un compilador portátil reorientable), Amoeba (uno de los primeros sistemas distribuidos utilizados en LAN) y Globe (un sistema distribuido de área amplia).

También es autor de MINIX, un pequeño clon de UNIX que en un principio estaba destinado a usarse en los laboratorios de programación estudiantiles. Fue la inspiración directa de Linux y la plataforma en la que se desarrolló inicialmente este sistema operativo. La versión actual de MINIX, conocida como MINIX3, ahora se enfoca en ser un sistema operativo en extremo confiable y seguro. El profesor Tanenbaum considerará su trabajo terminado cuando las computadoras no estén equipadas con un botón de reinicio y ningún ser humano haya experimentado una falla del sistema. MINIX3 es un proyecto continuo de código fuente abierto, al cual usted está invitado a contribuir. Vaya a *www.minix3.org* para que descargue una copia gratuita y averigüe las novedades.

Tanenbaum es miembro del ACM y del IEEE, además de pertenecer a la Real Academia de Artes y Ciencias de los Países Bajos. También ha ganado numerosos premios científicos, como:

- Premio TAA McGuffey, en 2010, para libros de ciencias computacionales e ingeniería.
- Medalla James H. Mulligan Jr., del IEEE, en 2007, por sus contribuciones a la educación.
- Premio TAA Texty, en 2002, para libros de ciencias computacionales e ingeniería.
- Premio ACM/SIGCSE, en 1997, por sus sorprendentes contribuciones a la educación en las ciencias computacionales.
- Premio Karl V. Karlstrom, del ACM, en 1994, para educadores sobresalientes.

Puede visitar su página de World Wide Web en *http://www.cs.vu.nl/~ast/*.

David J. Wetherall es profesor asociado de Ciencias Computacionales e Ingeniería en la University of Washington, en Seattle, además de trabajar como consultor para Intel Labs, en Seattle. Es originario de Australia, en donde recibió su licenciatura en ingeniería eléctrica por la University of Western Australia, y su doctorado en Ciencias Computacionales del MIT.

El profesor Wetherall ha trabajado en el área de las redes durante las últimas dos décadas. Su investigación se enfoca en los sistemas de red, en especial las redes inalámbricas y la computación móvil, el diseño de protocolos de Internet y la medición en las redes.

Recibió el premio ACM SIGCOMM Test-of-Time por su investigación pionera en las redes activas, una arquitectura para introducir los nuevos servicios de red con rapidez. Recibió el premio William Bennet, del IEEE, por sus descubrimientos en el mapeo de Internet. Su investigación se reconoció con un premio NSF CAREER, en 2002, y fue becario Sloan, en 2004.

Además de impartir clases sobre redes, el profesor Wetherall participa en la comunidad de investigación sobre redes. Ha codirigido los comités de programas de SIGCOMM, NSDI y MobiSys, además de ser cofundador de los talleres de trabajo HotNets, de la ACM. Ha participado en numerosos comités de programas para conferencias sobre redes y es editor de la publicación *Computer Communication Review*, de la ACM.

Puede visitar su página de World Wide Web en *http://djw.cs.washington.edu*.

1

INTRODUCCIÓN

Cada uno de los tres últimos siglos ha estado dominado por una nueva tecnología. El siglo XVIII fue la época de los grandes sistemas mecánicos que dieron paso a la Revolución Industrial. El siglo XIX fue la era de la máquina de vapor. Durante el siglo XX, la tecnología clave fue la recopilación, procesamiento y distribución de información. Entre otros desarrollos vimos la instalación de las redes telefónicas a nivel mundial, la invención de la radio y la televisión, el nacimiento y crecimiento sin precedentes de la industria de la computación, el lanzamiento de satélites de comunicaciones y, desde luego, Internet.

Como resultado del vertiginoso progreso tecnológico, estas áreas están convergiendo con rapidez en el siglo XXI, y las diferencias entre recolectar, transportar, almacenar y procesar información están desapareciendo rápidamente. Las organizaciones con cientos de oficinas esparcidas sobre una amplia área geográfica dan por sentado como algo rutinario la capacidad de examinar el estado actual, aun de su oficina más remota, con sólo presionar un botón. A medida que aumenta nuestra habilidad para recopilar, procesar y distribuir la información, la demanda por un procesamiento aún más complejo de la información aumenta rápidamente.

A pesar de que la industria de la computación es joven si se le compara con otras (como la automotriz y la de transporte aéreo), las computadoras han progresado de manera espectacular en un periodo muy corto. Durante las primeras dos décadas de su existencia, estos sistemas estaban altamente centralizados y por lo general se encontraban dentro de un salón grande e independiente. Era común que este salón tuviera paredes de vidrio, a través de las cuales los visitantes podían mirar boquiabiertos la gran maravilla electrónica que había en su interior. Una empresa o universidad de tamaño mediano apenas lograba tener una o dos computadoras, mientras que las instituciones muy grandes tenían, cuando mucho, unas cuantas docenas. La idea de que en un lapso de 40 años se produjeran en masa miles de millones de computadoras mucho más poderosas y del tamaño de una estampilla postal era en ese entonces mera ciencia ficción.

La fusión de las computadoras y las comunicaciones ha tenido una profunda influencia en cuanto a la manera en que se organizan los sistemas de cómputo. El concepto una vez dominante del "centro de cómputo" como un salón con una gran computadora a la que los usuarios llevaban su trabajo para procesarlo es ahora totalmente obsoleto (aunque los centros de datos que contienen miles de servidores de Internet se están volviendo comunes). El viejo modelo de una sola computadora para atender todas las necesidades computacionales de la organización se ha reemplazado por uno en el que un gran número de computadoras separadas pero interconectadas realizan el trabajo. A estos sistemas se les conoce como **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 "red de computadoras" para referirnos a un conjunto de computadoras autónomas interconectadas mediante una sola tecnología. Se dice que dos computadoras están interconectadas si pueden intercambiar información. La conexión no necesita ser a través de un cable de cobre; también se puede utilizar fibra óptica, microondas, infrarrojos y satélites de comunicaciones. Las redes pueden ser de muchos tamaños, figuras y formas, como veremos más adelante. Por lo general se conectan entre sí para formar redes más grandes, en donde **Internet** es el ejemplo más popular de una red de redes.

Existe una gran confusión en la literatura entre una red de computadoras y un **sistema distribuido**. La diferencia clave está en que en un sistema distribuido, un conjunto de computadoras independientes aparece frente a sus usuarios como un solo sistema coherente. Por lo general, tiene un modelo o paradigma único que se presenta a los usuarios. A menudo se utiliza una capa de software encima del sistema operativo, conocido como **middleware**; esta capa es responsable de implementar este modelo. Un ejemplo reconocido de un sistema distribuido es la **World Wide Web**. Este sistema opera sobre Internet y presenta un modelo en el cual todo se ve como un documento (página web).

En una red de computadoras no existe esta coherencia, modelo ni software. Los usuarios quedan expuestos a las máquinas reales, sin que el sistema haga algún intento por hacer que éstas se vean y actúen de una manera coherente. Si las máquinas tienen distinto hardware y distintos sistemas operativos, es algo que está a la vista de los usuarios. Si un usuario desea ejecutar un programa en un equipo remoto, tiene que iniciar sesión en esa máquina y ejecutarlo ahí.

En efecto, un sistema distribuido es un sistema de software construido sobre una red. El software le ofrece un alto nivel de cohesión y transparencia. Por ende, la distinción entre una red y un sistema distribuido recae en el software (en especial, el sistema operativo) y no en el hardware.

Sin embargo, los dos temas se superponen de manera considerable. Por ejemplo, tanto los sistemas distribuidos como las redes de computadoras necesitan mover archivos. La diferencia recae en quién invoca el movimiento, si el sistema o el usuario. Aunque este libro se enfoca principalmente en las redes, muchos de los temas también son importantes en los sistemas distribuidos. Para obtener más información, vea Tanenbaum y Van Steen (2007).

1.1 USOS DE LAS REDES DE COMPUTADORAS

Antes de examinar las cuestiones técnicas con detalle, vale la pena dedicar cierto tiempo a señalar por qué las personas están interesadas en las redes de computadoras y para qué se pueden utilizar. Después de todo, si nadie estuviera interesado en ellas, se construirían muy pocas. Empezaremos con las cuestiones tradicionales en las empresas, después pasaremos a las redes domésticas y a los acontecimientos recientes en relación con los usuarios móviles, para terminar con las cuestiones sociales.

1.1.1 Aplicaciones de negocios

La mayoría de las empresas tienen una cantidad considerable de computadoras. Por ejemplo, tal vez una empresa tenga una computadora para cada empleado y las utilice para diseñar productos, escribir folletos y llevar la nómina. Al principio, algunas de estas computadoras tal vez hayan trabajado aisladas unas de otras, pero en algún momento, la administración podría decidir que es necesario conectarlas para distribuir la información en toda la empresa.

En términos generales, el asunto es **compartir recursos** y la meta es que todos los programas, equipo y en especial los datos estén disponibles para cualquier persona en la red, sin importar la ubicación física del recurso o del usuario. Un ejemplo obvio y de uso popular es el de un grupo de empleados de oficina que comparten una impresora. Ninguno de los individuos necesita realmente una impresora privada, por otro lado, una impresora en red de alto volumen es más económica, veloz y fácil de mantener que una extensa colección de impresoras individuales.

Pero, probablemente, compartir información sea aún más importante que compartir recursos físicos como impresoras y sistemas de respaldo en cinta magnética. Las empresas tanto pequeñas como grandes dependen vitalmente de la información computarizada. La mayoría tiene registros de clientes, información de productos, inventarios, estados de cuenta, información fiscal y muchos datos más en línea. Si de repente todas sus computadoras se desconectaran de la red, un banco no podría durar más de cinco minutos. Una planta moderna de manufactura con una línea de ensamble controlada por computadora no duraría ni cinco segundos. Incluso una pequeña agencia de viajes o un despacho legal compuesto de tres personas son altamente dependientes de las redes de computadoras para permitir a los empleados acceder a la información y los documentos relevantes de manera instantánea.

En las empresas más pequeñas es probable que todas las computadoras se encuentren en una sola oficina o tal vez en un solo edificio, pero en las empresas más grandes las computadoras y empleados se encuentran esparcidos en docenas de oficinas y plantas en muchos países. Sin embargo, un vendedor en Nueva York podría requerir acceso a una base de datos que se encuentra en Singapur. Las redes conocidas como **VPN** (**Redes Privadas Virtuales**, del inglés *Virtual Private Networks*) se pueden usar para unir las redes individuales, ubicadas en distintos sitios, en una sola red extendida. En otras palabras, el simple hecho de que un usuario esté a 15 000 km de distancia de sus datos no debe ser impedimento para que los utilice como si fueran locales. Podemos sintetizar este objetivo al decir que es un intento por acabar con la "tiranía de la geografía".

En términos más simples, imaginemos el sistema de información de una empresa como si estuviera constituido por una o más bases de datos con información de la empresa y cierto número de empleados que necesitan acceder a esos datos en forma remota. En este modelo, los datos se almacenan en poderosas computadoras denominadas **servidores**. A menudo estos servidores están alojados en una ubicación central y un administrador de sistemas se encarga de su mantenimiento. Por el contrario, los empleados tienen en sus escritorios máquinas más simples conocidas como **clientes**, con las cuales acceden a los datos remotos, por ejemplo, para incluirlos en las hojas de cálculo que desarrollan (algunas veces nos referiremos al usuario humano del equipo cliente como el "cliente", aunque el contexto debe dejar en claro si nos referimos a la computadora o a su usuario). Las máquinas cliente y servidor se conectan mediante una red, como se muestra en la figura 1-1. Observe que mostramos la red como un óvalo simple, sin ningún detalle. Utilizaremos esta forma cuando hablemos de una red en el sentido más abstracto. Proveeremos los detalles según se requieran.

A esta disposición se le conoce como **modelo cliente-servidor**. Es un modelo ampliamente utilizado y forma la base de muchas redes. La realización más popular es la de una **aplicación web**, en la cual el servidor genera páginas web basadas en su base de datos en respuesta a las solicitudes de los clientes que pueden actualizarla. El modelo cliente-servidor es aplicable cuando el cliente y el servidor se encuentran

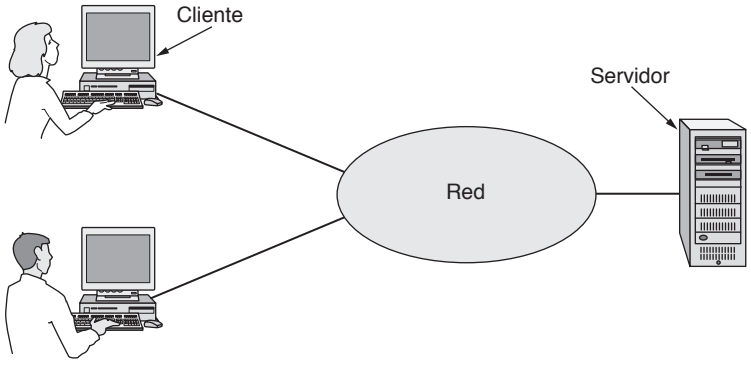


Figura 1-1. Una red con dos clientes y un servidor.

en el mismo edificio (y pertenecen a la misma empresa), pero también cuando están muy alejados. Por ejemplo, cuando una persona accede desde su hogar a una página en la World Wide Web se emplea el mismo modelo, en donde el servidor web remoto representa al servidor y la computadora personal del usuario representa al cliente. En la mayoría de las situaciones un servidor puede manejar un gran número (cientos o miles) de clientes simultáneamente.

Si analizamos detalladamente el modelo cliente-servidor, podremos ver que hay dos procesos (es decir, programas en ejecución) involucrados: uno en la máquina cliente y otro en la máquina servidor. La comunicación ocurre cuando el proceso cliente envía un mensaje a través de la red al proceso servidor. El proceso cliente espera un mensaje de respuesta. Cuando el proceso servidor obtiene la solicitud, lleva a cabo la tarea solicitada o busca los datos solicitados y devuelve una respuesta. Estos mensajes se muestran en la figura 1-2.

Un segundo objetivo al establecer una red de computadoras se relaciona con las personas y no con la información o con las computadoras. Una red de computadoras puede proveer un poderoso **medio de comunicación** entre los empleados. Ahora casi todas las empresas que tienen dos o más computadoras usan el **email** (**correo electrónico**), generalmente para la comunicación diaria. De hecho, una de las quejas comunes que se escucha por parte de los empleados a la hora de sus descansos es la gran cantidad de correos electrónicos con la que tienen que lidiar, pues la mayoría son sin sentido debido a que los jefes han descubierto que pueden enviar el mismo mensaje (a menudo sin contenido) a todos sus subordinados con sólo oprimir un botón.

En algunos casos, las llamadas telefónicas entre los empleados se pueden realizar a través de la red de computadoras en lugar de usar la compañía telefónica. A esta tecnología se le conoce como **telefonía IP** o **Voz sobre IP** (**VoIP**) cuando se utiliza la tecnología de Internet. El micrófono y el altavoz en cada extremo pueden ser de un teléfono habilitado para VoIP o la computadora del empleado. Para las empresas ésta es una maravillosa forma de ahorrar en sus cuentas telefónicas.

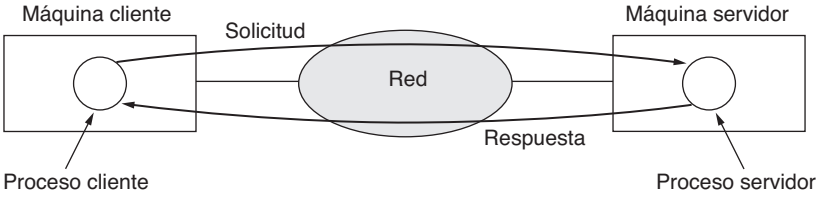


Figura 1-2. El modelo cliente-servidor implica solicitudes y respuestas.

Las redes de computadoras hacen posibles otras formas de comunicación más completas. Se puede agregar video al audio de manera que los empleados en ubicaciones distantes se puedan ver y escuchar mientras sostienen una reunión. Esta técnica es una poderosa herramienta para eliminar el costo y el tiempo dedicados a viajar. Los **escritorios compartidos** permiten a los trabajadores remotos ver una pantalla gráfica de computadora e interactuar con ella. Gracias a ello es posible que dos o más personas que trabajan a distancia lean y escriban en un pizarrón compartido, o escriban juntos un informe. Cuando un empleado realiza una modificación en un documento en línea, los demás pueden ver esa modificación de inmediato, en vez de tener que esperar varios días para recibir una carta. Dicha agilización facilita la cooperación entre los grupos remotos de personas, lo cual antes hubiera sido imposible. Hasta ahora se están empezando a utilizar formas más ambiciosas de coordinación remota como la telemedicina (por ejemplo, el monitoreo remoto de pacientes), lo cual puede tomar aún más importancia en un futuro cercano. En ocasiones se dice que la comunicación y el transporte están en constante competencia, y quien resulte ganador hará que el perdedor se vuelva obsoleto.

Un tercer objetivo para muchas empresas es realizar negocios electrónicamente, en especial con los clientes y proveedores. A este nuevo modelo se le denomina **e-commerce** (**comercio electrónico**) y ha crecido con rapidez en los años recientes. Las aerolíneas, librerías y otros vendedores han descubierto que a muchos clientes les gusta la conveniencia de comprar desde su hogar. En consecuencia, muchas empresas proveen catálogos de sus artículos y servicios en línea, e incluso reciben pedidos en línea. Los fabricantes de automóviles, aeronaves y computadoras entre otros, compran subsistemas de una variedad de proveedores y después ensamblan las piezas. Mediante el uso de redes de computadoras, los fabricantes pueden colocar los pedidos en forma electrónica según sea necesario. Esto reduce la necesidad de tener grandes inventarios y mejora la eficiencia.

1.1.2 Aplicaciones domésticas

En 1977, Ken Olsen era presidente de Digital Equipment Corporation, en ese entonces la segunda empresa distribuidora de computadoras más importante del mundo (después de IBM). Cuando se le preguntó por qué Digital no iba a incursionar a lo grande en el mercado de las computadoras personales, dijo: “No hay motivos para que una persona tenga una computadora en su hogar”. La historia demostró lo contrario y Digital desapareció. En un principio, las personas compraban computadoras para el procesamiento de palabras y para juegos. En los últimos años, probablemente la razón más importante sea acceder a Internet. En la actualidad muchos dispositivos electrónicos para el consumidor, como los decodificadores (*set-top boxes*), las consolas de juegos y los dispositivos de radio reloj, vienen con computadoras y redes integradas, especialmente redes inalámbricas; además las redes domésticas se utilizan ampliamente para actividades de entretenimiento, como escuchar, ver y crear música, fotos y videos.

El acceso a Internet ofrece a los usuarios domésticos **conectividad** a las computadoras remotas. Al igual que en las empresas, los usuarios domésticos pueden acceder a la información, comunicarse con otras personas y comprar productos y servicios mediante el comercio electrónico. Ahora el principal beneficio se obtiene al conectarse fuera del hogar. Bob Metcalfe, el inventor de Ethernet, formuló la hipótesis de que el valor de una red es proporcional al cuadrado del número de usuarios, ya que éste es aproximadamente el número de conexiones distintas que se pueden realizar (Gilder, 1993). Esta hipótesis se conoce como la “ley de Metcalfe” y nos ayuda a explicar cómo es que la enorme popularidad de Internet se debe a su tamaño.

El acceso a la información remota puede ser de varias formas. Podemos navegar en la World Wide Web para buscar información o sólo por diversión. La información disponible puede ser de varios temas, como arte, negocios, cocina, gobierno, salud, historia, ciencia, deportes, viajes y muchos más. Hay muchas maneras de divertirse como para mencionarlas aquí, además de otras que es mejor no mencionar.

Muchos periódicos se han puesto en línea y se pueden personalizar. Por ejemplo, es posible indicarle a un periódico que queremos recibir toda la información sobre políticos corruptos, grandes incendios, celebridades envueltas en escándalos y epidemias, pero nada de fútbol. También es posible hacer que se descarguen los artículos seleccionados en nuestra computadora mientras dormimos. Mientras continúe esta tendencia, cada vez más repartidores de periódicos se quedarán sin empleo, pero a los dueños de los periódicos les gusta la idea debido a que la distribución siempre ha sido el eslabón más débil en toda la cadena de producción. Claro que para que este modelo funcione tendrán primero que averiguar cómo ganar dinero en este nuevo mundo, algo que no es muy obvio dado que los usuarios de Internet esperan que todo sea gratuito.

El siguiente paso más allá de los periódicos (además de las revistas y las publicaciones científicas) es la biblioteca digital en línea. Muchas organizaciones profesionales como la ACM (*www.acm.org*) y la Sociedad de Computación del IEEE (*www.computer.org*) ya tienen todas sus publicaciones y memorias de congresos en línea. Tal vez los lectores de libros electrónicos y las bibliotecas en línea hagan obsoletos los libros impresos. Los escépticos deben tener en cuenta el efecto que tuvo la imprenta sobre el manuscrito ilustrado medieval.

Para acceder a una gran parte de esta información se utiliza el modelo cliente-servidor, aunque hay un modelo distinto y popular para acceder a la información que recibe el nombre de comunicación de **igual a igual** (*peer-to-peer*)(Parameswaran y colaboradores, 2001). En este modelo, los individuos que forman un grupo informal se pueden comunicar con otros miembros del grupo, como se muestra en la figura 1-3. En teoría, toda persona se puede comunicar con una o más personas; no hay una división fija en clientes y servidores.

Muchos sistemas de igual a igual, como BitTorrent (Cohen, 2003) no tienen una base de datos central para el contenido. En su defecto, cada usuario mantiene su propia base de datos en forma local y provee una lista de otras personas cercanas que son miembros del sistema. Así, un nuevo usuario puede ir con cualquier miembro para ver qué información tiene y obtener los nombres de otros miembros para inspeccionar si hay más contenido y más nombres. Este proceso de búsqueda se puede repetir de manera indefinida para crear una amplia base de datos local de lo que hay disponible en la red. Es una actividad que sería tediosa para las personas, pero para las computadoras es muy simple.

La comunicación de igual a igual se utiliza con frecuencia para compartir música y videos. Su mayor impacto fue en 2000 con un servicio de compartición de música llamado Napster, el cual se desmanteló después de lo que tal vez haya sido el caso de infracción de derechos de autor más grande de la historia que se haya documentado (Lam y Tan, 2001; y Macedonia, 2000). También existen aplicaciones legales

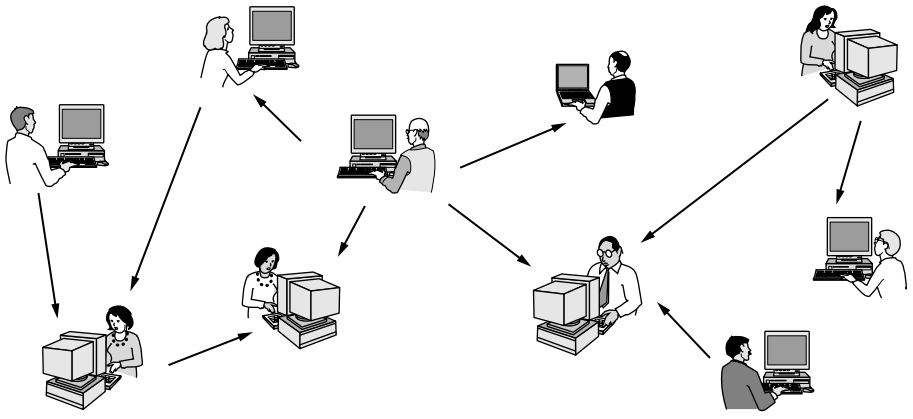


Figura 1-3. En un sistema de igual a igual no hay clientes y servidores fijos.

para la comunicación de igual a igual, como los fanáticos que comparten música del dominio público, las familias que comparten fotografías y películas caseras, y los usuarios que descargan paquetes de software públicos. De hecho, una de las aplicaciones más populares de Internet, el correo electrónico, es sin duda una aplicación de comunicación de igual a igual. Y es probable que esta forma de comunicación crezca de manera considerable en lo futuro.

Todas las aplicaciones antes mencionadas implican interacciones entre una persona y una base de datos remota llena de información. La segunda categoría importante de uso de redes es la comunicación de persona a persona, lo cual es básicamente la respuesta del siglo XXI al teléfono del siglo XIX. En la actualidad hay millones de personas en todo el mundo que utilizan el correo electrónico a diario y su uso se está extendiendo con rapidez. Es muy común que contenga audio y video, así como texto e imágenes. Tal vez la capacidad de oler los correos electrónicos todavía tarde un poco.

Todo adolescente que se precie de serlo es un adicto a la **mensajería instantánea**. Esta herramienta, que se deriva del programa *talk* de UNIX, se utiliza desde la década de 1970 y permite que dos personas se escriban mensajes entre sí en tiempo real. También hay servicios de mensajes multipersonas, como **Twitter**, que permite a las personas enviar mensajes cortos de texto, denominados *tweets*, a su círculo de amigos o cualquier otra audiencia dispuesta a recibirlos.

Las aplicaciones pueden usar Internet para transmitir audio (por ejemplo, las estaciones de radio de Internet) y video (por ejemplo, YouTube). Además de ser una forma económica de llamar a los amigos distantes, estas aplicaciones pueden proveer experiencias enriquecedoras como el teleaprendizaje, con lo cual un estudiante puede asistir a sus clases de las 8:00 a.m. sin tener que levantarse de la cama. A la larga, el uso de las redes para mejorar la comunicación de humano a humano tal vez demuestre ser más importante que cualquier otra aplicación. Quizás en el futuro sea muy importante para que las personas con inconveniencias geográficas, puedan obtener el mismo acceso a los servicios que las personas que viven en medio de una gran ciudad.

Las aplicaciones de **redes sociales** se encuentran entre las comunicaciones de persona a persona y de acceso a la información. Aquí el flujo de información se controla mediante las relaciones que las personas se declaran entre sí. Uno de los sitios de redes sociales más popular es **Facebook**. Este sitio permite a las personas actualizar sus perfiles y compartir las actualizaciones con otros que estén declarados como sus amigos. Otras aplicaciones de redes sociales pueden hacer presentaciones de amigos a través de amigos, enviar mensajes de noticias a éstos como el servicio de Twitter antes mencionado, y mucho más.

Incluso de una manera informal, grupos de personas pueden trabajar en conjunto para crear contenido. Por ejemplo, una **wiki** es un sitio web colaborativo que editan los miembros de una comunidad. La wiki más famosa es **Wikipedia**, una enciclopedia que todo el mundo puede editar, aunque hay miles de wikis más.

Nuestra tercera categoría es el comercio electrónico en el sentido más amplio del término. Las compras desde el hogar ya son populares, además de que permiten a los usuarios inspeccionar los catálogos en línea de miles de empresas. Algunos de estos catálogos son interactivos: muestran productos desde distintos puntos de vista y configuraciones que se pueden personalizar.

Si el cliente compra un producto en forma electrónica pero no puede averiguar cómo usarlo, puede obtener soporte técnico en línea.

Otra área en la cual el comercio electrónico se utiliza ampliamente es para acceder a las instituciones financieras. Muchas personas ya pagan sus facturas, administran sus cuentas bancarias y manejan sus inversiones por medios electrónicos. Es muy probable que esta tendencia continúe a medida que las redes se hagan más seguras.

Una de las áreas que casi nadie pudo prever es la de los "mercados de pulgas" electrónicos (bazares). Las subastas en línea de artículos de segunda mano se han convertido en una industria inmensa. A diferencia del comercio electrónico tradicional que sigue el modelo cliente-servidor, las subastas en línea son de igual a igual en cuanto a que los consumidores pueden actuar como compradores y como vendedores.

Algunas de estas formas de comercio electrónico han adquirido pequeñas e ingeniosas etiquetas debido al hecho de que la palabra "to" y el número "2" en inglés se pronuncian igual. En la figura 1-4 se muestra una lista de las más populares.

Etiqueta	Nombre completo	Ejemplo
B2C	Negocio a consumidor (*Business-to-consumer*)	Pedir libros en línea.
B2B	Negocio a negocio (*Business-to-business*)	Un fabricante de autos que pide los neumáticos al proveedor.
G2C	Gobierno a consumidor (*Government-to-consumer*)	El gobierno que distribuye formatos fiscales vía electrónica.
C2C	Consumidor a consumidor (*Consumer-to-consumer*)	Subastar productos de segunda mano en línea.
P2P	Igual a igual (*Peer-to-peer*)	Compartir música.

Figura 1-4. Algunas formas de comercio electrónico.

El entretenimiento es la cuarta categoría. Éste ha hecho grandes progresos en el hogar en años recientes gracias a la distribución de música, programas de radio y televisión, además de que las películas a través de Internet empiezan a competir con los mecanismos tradicionales. Los usuarios pueden buscar, comprar y descargar canciones en MP3 y películas con calidad de DVD para agregarlas a su colección personal. Los programas de TV ahora llegan a muchos hogares por medio de sistemas **IPTV** (**TeleVisión IP**) basados en la tecnología IP en vez de las transmisiones de radio o TV por cable. Las aplicaciones de flujo continuo de medios (*streaming*) permiten a los usuarios sintonizar estaciones de radio por Internet o ver los episodios recientes de sus programas favoritos de TV. Naturalmente, es posible mover todo este contenido por todo el hogar entre distintos dispositivos, pantallas y bocinas, por lo general a través de una red inalámbrica.

En un futuro cercano tal vez sea posible buscar cualquier película o programa de televisión que se haya realizado en cualquier país y hacer que se despliegue en nuestra pantalla al instante. Las nuevas películas tal vez se hagan interactivas, en donde ocasionalmente se le pida al usuario que decida el curso de la historia (¿debería Macbeth asesinar a Duncan o sólo esperar a que le llegue la hora?) y existan escenarios para todos los posibles casos. Probablemente la televisión en vivo también se vuelva interactiva, de manera que la audiencia pueda participar en los programas de preguntas, elegir de entre varios competidores, etcétera.

Los juegos son otra forma más de entretenimiento. Ya existen los juegos de simulación multipersonas en tiempo real, como jugar a las escondidas en un calabozo virtual, y los simuladores de vuelo en donde los jugadores de un equipo tratan de derribar a los jugadores del equipo contrario. Los mundos virtuales ofrecen un entorno persistente en donde miles de usuarios pueden experimentar una realidad compartida con gráficos tridimensionales.

Nuestra última categoría es la de la **computación ubicua**, en donde la computación está integrada a la vida diaria, como en la visión de Mark Weiser (1991). Muchos hogares ya cuentan con sistemas de seguridad que incluyen sensores de puertas y ventanas, y hay muchos sensores más que se pueden conectar a un monitor inteligente en el hogar, como los sensores de consumo de energía. Los medidores de electricidad, gas y agua podrían reportar su consumo a través de la red. Esto ahorraría dinero, ya que no habría necesidad de enviar personas para tomar la lectura de los medidores. Y nuestros detectores de humo podrían llamar al departamento de bomberos en vez de hacer un ruido ensordecedor (que tiene poco valor si no hay nadie en casa). A medida que disminuye el costo de los sensores y las comunicaciones, se realizarán cada vez más mediciones y reportes por medio de las redes.

Cada vez hay más dispositivos electrónicos conectados en red. Por ejemplo, algunas cámaras de gama alta ya cuentan con capacidad para conectarse a una red inalámbrica para enviar fotografías a una pantalla cercana y verlas. Los fotógrafos de deportes profesionales pueden también enviar fotos a sus editores en tiempo real, primero vía inalámbrica a un punto de acceso y después a través de Internet. Los dispositivos como las televisiones que se conectan a la toma de corriente en la pared pueden usar **redes por el cableado eléctrico** para enviar información por toda la casa a través de los cables que llevan electricidad. Tal vez no sea muy sorprendente tener estos objetos en la red, pero los objetos que no consideramos como computadoras también pueden detectar y comunicar información. Por ejemplo, su regadera podría registrar el consumo de agua, proporcionarle retroalimentación visual mientras se enjabona e informar a una aplicación de monitoreo ambiental en el hogar cuando termine para ayudarle a ahorrar en su factura de agua.

Hay una tecnología conocida como **RFID** (**Identificación por Radio-Frecuencia**, del inglés *Radio Frequency IDentification*) que llevará esta idea aún más lejos. Las etiquetas RFID son chips pasivos (es decir, no tienen batería) del tamaño de estampillas y ya es posible fijarlos a libros, pasaportes, mascotas, tarjetas de crédito y demás artículos en el hogar y fuera de él. Esto permite a los lectores RFID localizar los artículos y comunicarse con ellos a una distancia de hasta varios metros, dependiendo del tipo de RFID. En un principio la RFID se comercializó para reemplazar los códigos de barras. No ha tenido éxito aún debido a que los códigos de barras son gratuitos y las etiquetas RFID cuestan unos cuantos centavos. Desde luego que las etiquetas RFID ofrecen mucho más y su precio está bajando rápidamente. Tal vez conviertan el mundo real en la Internet de cosas (ITU, 2005).

1.1.3 Usuarios móviles

Las computadoras móviles como las laptops y las computadoras de bolsillo son uno de los segmentos de más rápido crecimiento en la industria de las computadoras. Sus ventas ya han sobrepasado a las de las computadoras de escritorio. ¿Por qué querría alguien una de ellas? Con frecuencia las personas que pasan mucho tiempo fuera de su oficina u hogar desean usar sus dispositivos móviles para leer y enviar correos electrónicos, usar Twitter, ver películas, descargar música, jugar o simplemente navegar en la Web para buscar información. Quieren hacer todas las cosas que hacen en su hogar y en su oficina. Por ende, quieren hacerlo desde cualquier lugar, ya sea en tierra, en el mar o incluso en el aire.

Muchos de estos usuarios móviles permiten la **conectividad** a Internet. Como es imposible tener una conexión alámbrica en los autos, botes y aviones, hay mucho interés en las redes móviles. Las redes celulares operadas por las compañías telefónicas son un tipo conocido de red inalámbrica que nos ofrece cobertura para los teléfonos móviles. Los **hotspots** basados en el estándar 802.11 son otro tipo de red inalámbrica para computadoras móviles. Han emergido por todos los puntos de reunión de la gente, que ahora cuenta con cobertura en cafés, hoteles, aeropuertos, trenes y aviones. Cualquiera con una laptop y un módem inalámbrico sólo necesita encender su computadora para estar conectado a Internet por medio de un hotspot, como si la computadora estuviera conectada a una red alámbrica.

Las redes inalámbricas son de gran valor para las flotillas de camiones, taxis, vehículos de reparto y técnicos para mantenerse en contacto con su base. Por ejemplo, en muchas ciudades los taxistas son comerciantes independientes, más que trabajar como empleados de una compañía de taxis. En algunas de estas ciudades los taxis tienen una pantalla que el conductor puede ver. Cuando llama un cliente, un despachador central introduce los puntos de partida y de destino. Esta información se despliega en las pantallas de los conductores y suena un timbre. El primer conductor en oprimir un botón en la pantalla obtiene la llamada.

Las redes inalámbricas también son importantes para los militares. Si de repente usted tiene que pelear una guerra en cualquier parte de la Tierra, probablemente no sea buena idea confiar en que podrá usar la infraestructura de red local. Es mejor que lleve su propia red.

Aunque es común que las redes inalámbricas y la computación móvil estén relacionadas, no son idénticas como lo muestra la figura 1-5. Aquí podemos ver una distinción entre las redes **inalámbricas fijas** y las redes **inalámbricas móviles**. Incluso las computadoras tipo notebook se conectan algunas veces mediante un cable de red. Por ejemplo, si un viajero conecta una computadora notebook al cable de red alámbrica en un cuarto de hotel, obtiene movilidad sin necesidad de una red inalámbrica.

Inalámbrica	Móvil	Aplicaciones comunes
No	No	Computadoras de escritorio en oficinas.
No	Sí	Una computadora notebook que se utiliza en un cuarto de hotel.
Sí	No	Las redes en edificios sin cables.
Sí	Sí	El inventario de la tienda con una computadora de mano.

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

En contraste, algunas computadoras inalámbricas no son móviles. En el hogar y en las oficinas u hoteles que carecen de un cableado adecuado, puede ser más conveniente conectar las computadoras de escritorio o los reproductores de medios en forma inalámbrica en vez de instalar cables. Para instalar una red inalámbrica sólo hay que comprar una pequeña caja con ciertos componentes electrónicos en su interior, desempacarla, conectarla y quizás haya que configurar algunos detalles sencillos en los equipos de cómputo. Esta solución puede ser mucho más económica que contratar trabajadores para que coloquen ductos y cables en el edificio.

Por último, también hay aplicaciones verdaderamente móviles e inalámbricas, como cuando las personas caminan por las tiendas con computadoras de mano registrando el inventario. En muchos aeropuertos concurridos, los empleados de los negocios de renta de autos trabajan en el lote de estacionamiento con computadoras móviles inalámbricas; escanean los códigos de barras o chips RFID de los autos que regresan y su dispositivo móvil, que tiene una impresora integrada, llama a la computadora principal, obtiene la información sobre la renta e imprime la factura en ese instante.

Podemos considerar al teléfono móvil como el impulsor clave de las aplicaciones móviles e inalámbricas. La **mensajería de texto** o Servicio de Mensajes Cortos (SMC) es en extremo popular, ya que permite al usuario de un teléfono móvil escribir un mensaje corto de texto que es entregado a través de la red celular a otro suscriptor móvil. Pocas personas hubieran predicho hace 10 años la gigantesca mina de oro que representa para las compañías telefónicas el hecho de que los adolescentes escriban tediosamente mensajes cortos de texto en teléfonos móviles. Pero el servicio de mensajes cortos es muy rentable, ya que a la compañía de telefonía celular le cuesta una pequeña fracción de un centavo transmitir un mensaje de texto, servicio por el cual cobran mucho más que eso.

Por fin ha llegado la tan esperada convergencia de los teléfonos e Internet; esto acelerará el crecimiento de las aplicaciones móviles. Los **teléfonos inteligentes** (como el popular iPhone) combinan los aspectos de los teléfonos y las computadoras móviles. Las redes celulares (3G y 4G) a las cuales se conectan pueden ofrecer servicios de datos rápidos para usar Internet y manejar a la vez las llamadas telefónicas. Muchos teléfonos avanzados se conectan también a los *hotspots* inalámbricos y cambian de una red a otra en forma automática para elegir la mejor opción disponible para el usuario.

Existen otros dispositivos electrónicos que también pueden usar las redes celulares y los *hotspots* de manera que puedan permanecer conectados con computadoras remotas. Los lectores de libros electrónicos pueden descargar un libro recién comprado o la siguiente edición de una revista o del periódico de hoy, en cualquier lugar en el que se encuentren. Los portarretratos electrónicos pueden actualizar sus pantallas al instante con nuevas imágenes.

Dado que los teléfonos móviles pueden ubicarse gracias a que comúnmente están equipados con receptores **GPS** (**Sistema de Posicionamiento Global**, del inglés *Global Positioning System*), algunos de sus servicios dependen de la ubicación. Los mapas móviles y las indicaciones son el ejemplo más obvio, ya que es probable que su teléfono y automóvil habilitados con GPS tengan mejor capacidad que usted para averiguar dónde está ubicado en un momento dado. Otros ejemplos podrían ser buscar una biblioteca o un restaurante chino que esté cerca, o el pronóstico del clima local. Hay otros servicios que pueden registrar la ubicación, como al incluir en las fotos y videos una anotación del lugar donde se tomaron. A esta anotación se le conoce como "geoetiquetado".

El **comercio-m** (**comercio móvil**) es un área en la que los teléfonos móviles están comenzando a utilizarse (Senn, 2000). Los mensajes cortos de texto del dispositivo móvil se utilizan para autorizar pagos de alimentos en las máquinas expendedoras, boletos del cine y otros artículos pequeños en vez de usar efectivo y tarjetas de crédito. Posteriormente el cargo aparece en la factura del teléfono celular. Cuando el dispositivo móvil está equipado con tecnología **NFC** (**Comunicación de Campo Cercano**, del inglés *Near Field Comunication*), puede actuar como una tarjeta inteligente RFID e interactuar con un lector cercano para realizar un pago. Las fuerzas motrices detrás de este fenómeno son los fabricantes de dispositivos móviles y los operadores de red, que hacen su mejor esfuerzo por tratar de averiguar cómo obtener una rebanada del pastel del comercio electrónico. Desde el punto de vista de la tienda, este esquema les puede ahorrar la mayor parte de la cuota de las compañías de tarjetas de crédito, que puede ser del uno por ciento o mayores. Claro que este plan podría fracasar debido a que los clientes en una tienda podrían usar los lectores de código de barras o RFID en sus dispositivos móviles para verificar los precios de la competencia antes de comprar, y también podrían usarlos para obtener un informe detallado sobre la ubicación y precios de la tienda más cercana.

Una de las grandes ventajas del comercio-m es que los usuarios de teléfonos móviles están acostumbrados a pagar por todo (en contraste a los usuarios de Internet, quienes esperan que todo sea gratuito). Si un sitio web en Internet cobrara una cuota por permitir a sus clientes pagar con tarjeta de crédito, habría muchas quejas por parte de los usuarios. No obstante, si una compañía de telefonía móvil permitiera a sus clientes pagar por los artículos en una tienda con sólo ondear su teléfono frente a la caja registradora y después les cobrara una cuota por ese servicio, probablemente los usuarios lo aceptarían como algo normal. El tiempo nos lo dirá.

Sin duda, el número de usuarios de computadoras móviles e inalámbricas aumentará con rapidez en el futuro a medida que se reduzca el tamaño de éstas, probablemente en formas que nadie puede prever por ahora. Demos un vistazo a algunas posibilidades. Las **redes de sensores** están compuestas por nodos que recopilan y transmiten en forma inalámbrica la información que detectan sobre el estado del mundo físico. Los nodos pueden ser parte de elementos conocidos, como autos o teléfonos, o pueden ser pequeños dispositivos independientes. Por ejemplo, su automóvil podría recopilar la información sobre su ubicación, velocidad, vibración y ahorro de combustible desde su sistema de diagnóstico integrado y enviar esta información a una base de datos (Hull y colaboradores, 2006). Esos datos pueden ayudar a encontrar baches, planear viajes alrededor de caminos congestionados e indicarnos si somos unos "devoradores de gasolina" en comparación con otros conductores en la misma extensión del camino.

Las redes de sensores están revolucionando la ciencia al proveer una gran cantidad de datos sobre el comportamiento, lo cual no era posible observar antes. Como ejemplo podemos mencionar el rastreo individual de cebras durante su migración, al colocar un pequeño sensor en cada animal (Juang y colaboradores, 2002). Los investigadores han logrado empacar una computadora inalámbrica en un cubo de 1 mm de grosor (Warneke y colaboradores, 2001). Con computadoras móviles así de pequeñas podemos rastrear incluso hasta aves pequeñas, roedores e insectos.

A lo anterior le podemos dar incluso usos triviales (como en los parquímetros), ya que se utilizan datos que no estaban disponibles antes. Los parquímetros inalámbricos pueden aceptar pagos con tarjetas de crédito o débito con verificación instantánea a través del enlace inalámbrico. También pueden reportar

cuando estén en uso mediante la red inalámbrica. Esto permitiría a los conductores descargar un mapa de parquímetros reciente en su auto, para que puedan encontrar un lugar disponible con más facilidad. Claro que al expirar, un parquímetro podría también verificar la presencia de un automóvil (al enviar una señal y esperar su rebote) y reportar a las autoridades de tránsito su expiración. Se estima que en Estados Unidos tan sólo los gobiernos municipales podrían recolectar unos $10 mil millones de dólares adicionales de esta forma (Harte y colaboradores, 2000).

Las **computadoras usables** son otra aplicación prometedora. Los relojes inteligentes con radio han formado parte de nuestro espacio mental desde que aparecieron en la tira cómica de Dick Tracy, en 1946, ahora es posible comprarlos. También hay otros dispositivos de este tipo que se pueden implementar, como los marcapasos y las bombas de insulina. Algunos de ellos se pueden controlar a través de una red inalámbrica. Esto permitiría a los doctores probarlos y reconfigurarlos con más facilidad. Incluso podrían surgir graves problemas si los dispositivos fueran tan inseguros como la PC promedio y alguien pudiera intervenirlos fácilmente (Halperin y colaboradores, 2008).

1.1.4 Cuestiones sociales

Al igual que la imprenta hace 500 años, las redes de computadoras permiten a los ciudadanos comunes distribuir y ver el contenido en formas que no hubiera sido posible lograr antes. Pero con lo bueno viene lo malo, y esta posibilidad trae consigo muchas cuestiones sociales, políticas y éticas sin resolver; a continuación mencionaremos brevemente algunas de ellas, ya que para un estudio completo de las mismas se requeriría por lo menos todo un libro.

Las redes sociales, los tableros de mensajes, los sitios de compartición de contenido y varias aplicaciones más permiten a las personas compartir sus opiniones con individuos de pensamientos similares. Mientras que los temas estén restringidos a cuestiones técnicas o aficiones como la jardinería, no surgirán muchas dificultades.

El verdadero problema está en los temas que realmente importan a las personas, como la política, la religión y el sexo. Hay opiniones que si se publican y quedan a la vista de todos pueden ser bastante ofensivas para algunas personas. O peor aún, tal vez no sean políticamente correctas. Lo que es más, las opiniones no necesitan limitarse sólo a texto; es posible compartir fotografías a color de alta resolución y clips de video a través de las redes de computadoras. Algunas personas toman una posición del tipo "vive y deja vivir", pero otras sienten que simplemente es inaceptable publicar cierto material (como ataques verbales a países o religiones específicas, pornografía, etc.) y que es necesario censurar dicho contenido. Cada país tiene diferentes leyes contradictorias sobre este tema. Por ende, el debate se aviva.

En el pasado reciente las personas demandaban a los operadores de red afirmando que eran responsables por el contenido de lo que transmitían, al igual que los periódicos y las revistas. La respuesta inevitable es que una red es como una compañía telefónica o la oficina postal, por lo que no es posible que esté vigilando lo que sus usuarios dicen.

Para estos momentos tal vez le sorprenda un poco saber que algunos operadores de red bloquean contenido por motivos personales. Algunos suspendieron el servicio de red a varios usuarios de aplicaciones de igual a igual debido a que no consideraron rentable transmitir las grandes cantidades de tráfico que envían esas aplicaciones. Probablemente estos mismos operadores traten a las diversas empresas de manera diferente. Si usted es una empresa grande y paga bien, recibe un buen servicio, pero si es un comerciante pequeño recibirá un mal servicio. Los que se oponen a esta práctica argumentan que el contenido de las redes de igual a igual y cualquier otro tipo de contenido debe tratarse de la misma forma, ya que son sólo bits en la red. A este argumento que sostiene que no hay que diferenciar las comunicaciones según su contenido u origen, o con base en quién lo provee, se le conoce como **neutralidad de red** (Wu, 2003). Es muy probable que este debate persista por mucho tiempo.

Hay muchas otras partes involucradas en la lucha sobre el contenido. Por ejemplo, la música y las películas piratas impulsaron el crecimiento masivo de las redes de igual a igual, lo cual no agradó a los dueños de los derechos de autor, quienes han amenazado con tomar (y algunas veces han tomado) acción legal. Ahora hay sistemas automatizados que buscan redes de igual a igual y envían advertencias a los operadores de red y usuarios sospechosos de infringir los derechos de autor. En Estados Unidos a estas advertencias se les conoce como **avisos de DCMA para quitar contenido** según la **Ley de Copyright del Milenio Digital**. Esta búsqueda es una carrera armamentista, ya que es difícil detectar de manera confiable el momento en que se violan los derechos de autor. Incluso hasta su impresora podría ser considerada como culpable (Piatek y colaboradores, 2008).

Las redes de computadoras facilitan considerablemente la comunicación. También ayudan a las personas que operan la red con el proceso de husmear en el tráfico. Esto provoca conflictos sobre cuestiones como los derechos de los empleados frente a los derechos de los patrones. Muchas personas leen y escriben correos electrónicos en su trabajo. Muchos patrones han reclamado el derecho de leer y tal vez censurar los mensajes de los empleados, incluyendo los mensajes enviados desde una computadora en el hogar, después de las horas de trabajo. No todos los empleados están de acuerdo con esto, en especial con lo último.

Otro conflicto se centra alrededor de los derechos del gobierno frente a los derechos de los ciudadanos. El FBI ha instalado sistemas con muchos proveedores de servicios de Internet para analizar todo el correo electrónico entrante y saliente en busca de fragmentos que le interesen. Uno de los primeros sistemas se llamaba originalmente Carnivore, pero la mala publicidad provocó que cambiaran su nombre por el de DCS1000, algo más inocente (Blaze y Bellovin, 2000; Sobel, 2001 y Zacks, 2001). El objetivo de este sistema es espiar a millones de personas con la esperanza de encontrar información sobre actividades ilegales. Por desgracia para los espías, la Cuarta Enmienda a la Constitución de Estados Unidos prohíbe las búsquedas gubernamentales sin una orden de cateo, pero a menudo el gobierno ignora esta regulación.

Claro que el gobierno no es el único que amenaza la privacidad de las personas. El sector privado también participa al crear **perfiles** de los usuarios. Por ejemplo, los pequeños archivos llamados **cookies** que los navegadores web almacenan en las computadoras de los usuarios permiten a las empresas rastrear las actividades de los usuarios en el ciberespacio y también pueden permitir que los números de tarjetas de crédito, de seguro social y demás información confidencial se filtren por todo Internet (Berghel, 2001). Las empresas que proveen servicios basados en web pueden mantener grandes cantidades de información personal sobre sus usuarios para estudiar directamente sus actividades. Por ejemplo, Google puede leer su correo electrónico y mostrarle anuncios basados en sus intereses si utiliza su servicio de correo electrónico **Gmail**.

Un nuevo giro en el ámbito de los dispositivos móviles es la privacidad de la ubicación (Beresford y Stajano, 2003). Como parte del proceso de proveer servicio a los dispositivos móviles, los operadores de red aprenden en dónde se encuentran los usuarios a distintas horas del día. Esto les permite rastrear sus movimientos. Tal vez sepan qué club nocturno frecuenta usted y a cuál centro médico asiste.

Las redes de computadoras también ofrecen el potencial de incrementar la privacidad al enviar mensajes anónimos. En ciertas situaciones, esta capacidad puede ser conveniente. Además de evitar que las empresas conozcan los hábitos de sus clientes, también ofrece, por ejemplo, los medios para que los estudiantes, soldados, empleados y ciudadanos puedan denunciar el comportamiento ilegal por parte de profesores, oficiales, superiores y políticos sin temor a las represalias. Por otra parte, en Estados Unidos, y en la mayoría de otras democracias, la ley permite de manera específica que una persona acusada tenga el derecho de confrontar y desafiar a su acusador en la corte, por lo que no se permite el uso de acusaciones anónimas como evidencia.

Internet hace posible encontrar información rápidamente, pero gran parte de ella se considera de dudosa procedencia, engañosa o en definitiva incorrecta. Ese consejo médico que usted obtuvo de Internet en relación con el dolor en su pecho puede haber provenido de un ganador del Premio Nobel o de un chico sin estudios.

Hay otro tipo de información que por lo general es indeseable. El correo electrónico basura (*spam*) se ha convertido en parte de la vida, ya que los emisores de correo electrónico basura (*spammers*) han recolectado millones de direcciones de correo electrónico y los aspirantes a vendedores pueden enviarles mensajes generados por computadora a un costo muy bajo. La inundación resultante de spam rivaliza con el flujo de mensajes de personas reales. Por fortuna hay software de filtrado capaz de leer y desechar el *spam* generado por otras computadoras, aunque su grado de éxito puede variar en forma considerable.

Existe también contenido destinado al comportamiento criminal. Las páginas web y los mensajes de correo electrónico con contenido activo (en esencia, programas o macros que se ejecutan en la máquina del receptor) pueden contener virus que invadan nuestra computadora. Podrían utilizarlos para robar las contraseñas de nuestras cuentas bancarias o hacer que nuestra computadora envíe *spam* como parte de una red zombie (**botnet**) o grupo de equipos comprometidos.

Los mensajes de **suplantación de identidad** o estafas se enmascaran como si se originaran desde un sitio de confianza (como su banco, por ejemplo) para ver si el receptor les revela información delicada, como los números de sus tarjetas de crédito. El robo de identidad se está convirtiendo en un problema grave, a medida que los ladrones recolectan suficiente información sobre una víctima para obtener tarjetas de crédito y otros documentos a su nombre.

Puede ser difícil evitar que las computadoras se hagan pasar por personas en Internet. Este problema ha originado el desarrollo de **cuadros de captura de texto para verificación** (**CAPTCHAs**), en donde una computadora pide a una persona que resuelva una pequeña tarea de reconocimiento; por ejemplo, escribir las letras que se muestran en una imagen distorsionada para demostrar que son humanos (von Ahn, 2001). Este proceso es una variación de la famosa prueba de Turing, en donde una persona hace preguntas a través de una red para juzgar si la entidad que responde es humana.

Podríamos resolver muchos de estos problemas si la industria de la computación tomara en serio la seguridad de las computadoras. Si se cifraran y autenticaran todos los mensajes, sería más difícil tener dificultades. Dicha tecnología está bien establecida y la estudiaremos con detalle en el capítulo 8. El inconveniente es que los distribuidores de hardware y software saben que es costoso incluir herramientas de seguridad y sus clientes no exigen dichas características. Además, una gran cantidad de los problemas son provocados por el software defectuoso, ya que los distribuidores siguen agregando cada vez más características a sus programas, lo cual se traduce inevitablemente en más código y por ende más errores. Tal vez sería conveniente aplicar un impuesto para las nuevas características, pero no todos estarían convencidos de que sea la mejor solución. También sería agradable que hubiera un reembolso por el software defectuoso pero, de ser así, toda la industria del software quedaría en bancarrota en menos de un año.

Las redes de computadoras generan nuevos problemas legales cuando interactúan con las antiguas leyes. Las apuestas electrónicas son un ejemplo de ello. Si las computadoras han estado simulando cosas por décadas, ¿por qué no simular máquinas tragamonedas, ruletas, repartidores de blackjack y demás equipo para apostar? Bueno, porque es ilegal en muchos lugares. El problema es que las apuestas son legales en otras partes (en Inglaterra, por ejemplo) y los propietarios de casinos de esos lugares han captado el potencial de las apuestas por Internet. Pero, ¿qué ocurriría si el apostador, el casino y el servidor estuvieran todos en distintos países, con leyes contradictorias? Buena pregunta.

1.2 HARDWARE DE RED

Ahora es tiempo de dejar a un lado las aplicaciones y los aspectos sociales de las redes para enfocarnos en las cuestiones técnicas implicadas en su diseño. No existe una clasificación aceptada en la que encajen todas las redes, pero hay dos que sobresalen de manera importante: la tecnología de transmisión y la escala. Examinaremos ahora cada una de ellas por turno.

Hablando en sentido general, existen dos tipos de tecnología de transmisión que se emplean mucho en la actualidad: los enlaces de **difusión** (*broadcast*) y los enlaces de **punto a punto**.

Los enlaces de punto a punto conectan pares individuales de máquinas. Para ir del origen al destino en una red formada por enlaces de punto a punto, los mensajes cortos (conocidos como **paquetes** en ciertos contextos) tal vez tengan primero que visitar una o más máquinas intermedias. A menudo es posible usar varias rutas de distintas longitudes, por lo que es importante encontrar las más adecuadas en las redes de punto a punto. A la transmisión punto a punto en donde sólo hay un emisor y un receptor se le conoce como **unidifusión** (*unicasting*).

Por el contrario, en una red de difusión todas las máquinas en la red comparten el canal de comunicación; los paquetes que envía una máquina son recibidos por todas las demás. Un campo de dirección dentro de cada paquete especifica a quién se dirige. Cuando una máquina recibe un paquete, verifica el campo de dirección. Si el paquete está destinado a la máquina receptora, ésta procesa el paquete; si el paquete está destinado para otra máquina, sólo lo ignora.

Una red inalámbrica es un ejemplo común de un enlace de difusión, en donde la comunicación se comparte a través de una región de cobertura que depende del canal inalámbrico y de la máquina que va a transmitir. Como analogía considere alguien parado en una sala de juntas gritando: "Watson, ven aquí. Te necesito". Aunque muchas personas hayan recibido (escuchado) el paquete, sólo Watson responderá; los otros simplemente lo ignorarán.

Por lo general, los sistemas de difusión también brindan la posibilidad de enviar un paquete a *todos* los destinos mediante el uso de un código especial en el campo de dirección. Cuando se transmite un paquete con este código, todas las máquinas en la red lo reciben y procesan. A este modo de operación se le conoce como **difusión** (*broadcasting*). Algunos sistemas de difusión también soportan la transmisión a un subconjunto de máquinas, lo cual se conoce como **multidifusión** (*multicasting*).

Hay un criterio alternativo para clasificar las redes: por su escala. La distancia es importante como medida de clasificación, ya que las distintas tecnologías se utilizan a diferentes escalas.

En la figura 1-6 clasificamos los sistemas multiprocesadores con base en su tamaño físico. En la parte de arriba están las redes de área personal, las cuales están destinadas a una persona. Después se encuentran redes más grandes. Éstas se pueden dividir en redes de área local, de área metropolitana y de área amplia, cada una con una escala mayor que la anterior. Por último, a la conexión de dos o más redes se le conoce como **interred** (*internetwork*). La Internet de nivel mundial es sin duda el mejor ejemplo (aunque no el único) de una interred. Pronto tendremos interredes aún más grandes con la **Internet interplanetaria** que conecta redes a través del espacio (Burleigh y colaboradores, 2003).

En este libro hablaremos sobre las redes de todas estas escalas. En las siguientes secciones le proporcionaremos una breve introducción al hardware de red con base en la escala.

1.2.1 Redes de área personal

Las **redes de área personal**, generalmente llamadas **PAN** (*Personal Area Network*) permiten a los dispositivos comunicarse dentro del rango de una persona. Un ejemplo común es una red inalámbrica que conecta a una computadora con sus periféricos. Casi todas las computadoras tienen conectado un monitor, un teclado, un ratón y una impresora. Sin la tecnología inalámbrica es necesario realizar esta conexión

Distancia entre procesadores	Procesadores ubicados en el (la) mismo(a)	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
1000 km	Continente	
10000 km	Planeta	Internet

Figura 1-6. Clasificación de los procesadores interconectados con base en la escala.

mediante cables. Hay tantos usuarios nuevos que batallan mucho para encontrar los cables adecuados y conectarlos en los orificios apropiados (aun cuando, por lo general, están codificados por colores), que la mayoría de los distribuidores de computadoras ofrecen la opción de enviar un técnico al hogar del usuario para que se encargue de ello. Para ayudar a estos usuarios, algunas empresas se pusieron de acuerdo para diseñar una red inalámbrica de corto alcance conocida como **Bluetooth** para conectar estos componentes sin necesidad de cables. La idea es que si sus dispositivos tienen Bluetooth, no necesitará cables. Sólo hay que ponerlos en el lugar apropiado, encenderlos y trabajarán en conjunto. Para muchas personas, esta facilidad de operación es una gran ventaja.

En su forma más simple, las redes Bluetooth utilizan el paradigma maestro-esclavo de la figura 1-7. La unidad del sistema (la PC), por lo general es el maestro que trata con el ratón, el teclado, etc., como sus esclavos. El maestro dice a los esclavos qué direcciones usar, cuándo pueden transmitir información, durante cuánto tiempo pueden transmitir, qué frecuencias usar, etcétera.

También podemos usar Bluetooth en otras aplicaciones. A menudo se utiliza para conectar unos audífonos a un teléfono móvil sin cables, además se puede conectar el reproductor musical digital a nuestro automóvil con sólo tenerlo dentro del rango. Una clase completamente distinta de red PAN se forma

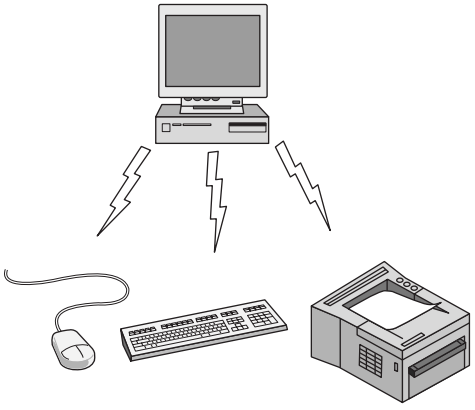


Figura 1-7. Configuración de red PAN con Bluetooth.

cuando un dispositivo médico integrado, como un marcapasos, bomba de insulina o audífono para discapacitados se comunica con un control remoto operado por el usuario. En el capítulo 4 veremos con detalle la tecnología Bluetooth.

Las redes PAN también se pueden construir con otras tecnologías que se comunican dentro de rangos cortos, como RFID en las tarjetas inteligentes y los libros de las bibliotecas. En el capítulo 4 estudiaremos la tecnología RFID.

1.2.2 Redes de área local

Las **redes de área local**, generalmente llamadas **LAN** (*Local Area Networks*), son redes de propiedad privada que operan dentro de un solo edificio, como una casa, oficina o fábrica. Las redes LAN se utilizan ampliamente para conectar computadoras personales y electrodomésticos con el fin de compartir recursos (por ejemplo, impresoras) e intercambiar información. Cuando las empresas utilizan redes LAN se les conoce como **redes empresariales**.

Las redes LAN son muy populares en la actualidad, en especial en los hogares, los edificios de oficinas antiguos, las cafeterías y demás sitios en donde es muy problemático instalar cables. En estos sistemas, cada computadora tiene un módem y una antena que utiliza para comunicarse con otras computadoras. En la mayoría de los casos, cada computadora se comunica con un dispositivo en el techo, como se muestra en la figura 1-8(a). A este dispositivo se le denomina **AP** (**Punto de Acceso**, del inglés *Access Point*), **enrutador inalámbrico** o **estación base**; transmite paquetes entre las computadoras inalámbricas y también entre éstas e Internet. El AP es como el niño popular de la escuela, ya que todos quieren hablar con él. Pero si hay otras computadoras que estén lo bastante cerca una de otra, se pueden comunicar directamente entre sí en una configuración de igual a igual.

Hay un estándar para las redes LAN inalámbricas llamado **IEEE 802.11**, mejor conocido como **WiFi**. Opera a velocidades desde 11 hasta cientos de Mbps (en este libro nos apegaremos a la tradición y mediremos las velocidades de las líneas de transmisión en megabits/segundo, en donde 1 Mbps es 1 000 000 bits/segundo, y en gigabits/segundo, en donde 1 Gbps es 1 000 000 000 bits/segundo). En el capítulo 4 hablaremos sobre el estándar 802.11.

Las redes LAN alámbricas utilizan distintas tecnologías de transmisión. La mayoría utilizan cables de cobre, pero algunas usan fibra óptica. Las redes LAN tienen restricciones en cuanto a su tamaño, lo cual significa que el tiempo de transmisión en el peor de los casos es limitado y se sabe de antemano. Conocer estos límites facilita la tarea del diseño de los protocolos de red. Por lo general las redes LAN alámbricas que operan a velocidades que van de los 100 Mbps hasta un 1 Gbps, tienen retardo bajo (microsegundos

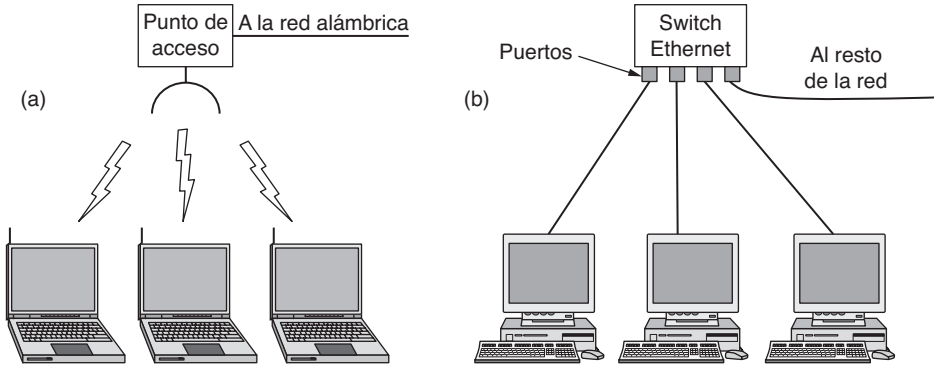


Figura 1-8. Redes inalámbrica y alámbrica. (a) 802.11. (b) Ethernet conmutada.

o nanosegundos) y cometen muy pocos errores. Las redes LAN más recientes pueden operar a una velocidad de hasta 10 Gbps. En comparación con las redes inalámbricas, las redes LAN alámbricas son mucho mejores en cuanto al rendimiento, ya que es más fácil enviar señales a través de un cable o fibra que por el aire.

La topología de muchas redes LAN alámbricas está basada en los enlaces de punto a punto. El estándar IEEE 802.3, comúnmente conocido como **Ethernet**, es hasta ahora el tipo más común de LAN alámbrica. La figura 1-8(b) muestra un ejemplo de topología de **Ethernet conmutada**. Cada computadora se comunica mediante el protocolo Ethernet y se conecta a una caja conocida como **switch** con un enlace de punto a punto. De aquí que tenga ese nombre. Un switch tiene varios **puertos**, cada uno de los cuales se puede conectar a una computadora. El trabajo del switch es transmitir paquetes entre las computadoras conectadas a él, y utiliza la dirección en cada paquete para determinar a qué computadora se lo debe enviar.

Para crear redes LAN más grandes se pueden conectar switches entre sí mediante sus puertos. ¿Qué ocurre si los conectamos en un circuito cerrado? ¿Podrá funcionar la red así? Por fortuna, los diseñadores consideraron este caso. Es responsabilidad del protocolo determinar qué rutas deben recorrer los paquetes para llegar de manera segura a la computadora de destino. En el capítulo 4 veremos cómo funciona esto.

También es posible dividir una gran LAN física en dos redes LAN lógicas más pequeñas. Tal vez se pregunte por qué sería esto útil. En ocasiones la distribución del equipo de red no coincide con la estructura de la organización. Por ejemplo, los departamentos de ingeniería y finanzas de una empresa podrían tener computadoras en la misma LAN física debido a que se encuentran en la misma ala del edificio, pero podría ser más sencillo administrar el sistema si cada departamento tuviera su propia red lógica, denominada **LAN virtual** o **VLAN**. En este diseño cada puerto se identifica con un "color"; por ejemplo, verde para ingeniería y rojo para finanzas. Después el switch reenvía los paquetes de manera que las computadoras conectadas a los puertos verdes estén separadas de las que están conectadas a los puertos rojos. Por ejemplo, los paquetes de difusión que se envíen por un puerto rojo no se recibirán en un puerto verde, tal como si hubiera dos redes LAN distintas. Al final del capítulo 4 veremos los detalles sobre las redes VLAN.

También existen otras topologías de LAN alámbrica. De hecho, la Ethernet conmutada es una versión moderna del diseño original de Ethernet en el que se difundían todos los paquetes a través de un solo cable lineal. Sólo una máquina podía transmitir con éxito en un instante dado, y se utilizaba un mecanismo de arbitraje distribuido para resolver los conflictos. Utilizaba un algoritmo simple: las computadoras podían transmitir siempre que el cable estuviera inactivo. Si ocurría una colisión entre dos o más paquetes, cada computadora esperaba un tiempo aleatorio y volvía a intentar. Llamaremos a esa versión **Ethernet clásica** por cuestión de claridad y, como tal vez se lo imagine, aprenderá sobre ella en el capítulo 4.

Las redes inalámbricas y las alámbricas se pueden dividir en diseños estáticos y dinámicos, dependiendo de la forma en que se asigna el canal. Una asignación estática típica sería dividir el tiempo en intervalos discretos y utilizar un algoritmo por turno rotatorio (*round-robin*), para que cada máquina pueda difundir los datos sólo cuando sea su turno de usar su intervalo. La asignación estática desperdicia la capacidad del canal cuando una máquina no tiene nada que decir durante su intervalo asignado, por lo que la mayoría de los sistemas tratan de asignar el canal en forma dinámica (es decir, bajo demanda).

Los métodos de asignación dinámica para un canal común pueden ser centralizados o descentralizados. En el método de asignación de canal centralizado hay una sola entidad (por ejemplo, la estación base en las redes celulares) que determina el turno de cada quien. Para ello podría aceptar varios paquetes y asignarles prioridades de acuerdo con algún algoritmo interno. En el método de asignación de canal descentralizado no hay una entidad central; cada máquina debe decidir por su cuenta si va a transmitir o no. Tal vez usted piense que esta metodología provoca un caos, pero no es así. Más adelante estudiaremos muchos algoritmos diseñados para poner orden a un potencial caos.

Vale la pena invertir un poco más de tiempo para hablar sobre las redes LAN en el hogar. En lo futuro es probable que todos los dispositivos en el hogar sean capaces de comunicarse con cualquier otro dispositivo, y todos ellos serán accesibles a través de Internet. Tal vez este acontecimiento sea uno de esos conceptos visionarios que nadie solicitó (como los controles remotos de TV o los teléfonos móviles), pero una vez que llegaron nadie se imagina cómo pudo haber vivido sin ellos.

Muchos dispositivos ya son capaces de conectarse en red. Entre ellos tenemos a las computadoras, los dispositivos de entretenimiento como las TV y los DVD, teléfonos y otros dispositivos electrónicos como las cámaras, aparatos como los radios relojes e infraestructura como los medidores de servicios y termostatos. Esta tendencia seguirá avanzando. Por ejemplo, es probable que el hogar promedio tenga una docena de relojes (es decir, en aparatos), los cuales, si estuvieran conectados a Internet, podrían ajustarse de manera automática al horario de verano para ahorrar energía solar. Es muy probable que el monitoreo remoto del hogar sea una aplicación muy popular en el futuro, ya que muchos hijos en edad adulta estarían dispuestos a invertir algo de dinero para ayudar a sus padres envejecidos a vivir con seguridad en sus propios hogares.

Aunque podríamos considerar a la red doméstica como cualquier otra LAN, es muy probable que tenga distintas propiedades. En primer lugar, los dispositivos en red tienen que ser muy fáciles de instalar. Los enrutadores inalámbricos son uno de los artículos que más devuelven los consumidores. Las personas compran uno porque desean una red inalámbrica en su hogar, pero al sacarlo de su caja descubren que no está "listo para usarse"; por lo tanto, prefieren devolverlo en lugar de esperar a ser atendidas en la línea telefónica de asistencia.

En segundo lugar, la red y los dispositivos tienen que operar en un modo a prueba de errores. Los aires acondicionados solían tener una perilla con cuatro posiciones: Apagado, bajo, medio y alto. Ahora tienen manuales de 30 páginas. Una vez que puedan conectarse en red, es probable que tan sólo el capítulo sobre seguridad sea de ese tamaño. Éste es un problema debido a que sólo los usuarios de computadoras están acostumbrados a lidiar con productos que no funcionan; el público que compra autos, televisiones y refrigeradores es menos tolerante. Esperan productos que funcionen al 100% sin tener que contratar a un experto en computadoras.

En tercer lugar, el precio es imprescindible para el éxito. Las personas no pagarán una tarifa de $50 dólares por un termostato con conexión a Internet debido a que pocas personas consideran que sea tan importante monitorear la temperatura de su hogar desde el trabajo. Aunque tal vez por $5 dólares adicionales sí podría llegar a venderse.

En cuarto lugar, debe existir la posibilidad de empezar con uno o dos dispositivos para después expandir el alcance de la red en forma gradual. Esto significa que no debe haber guerras de formatos. Decir a los consumidores que compren periféricos con interfaces IEEE 1394 (*FireWire*) para luego retractarse unos cuantos años después y decir que USB 2.0 es la interfaz del mes, y luego cambiarla por la interfaz 802.11g (¡ups!, no, mejor que sea 802.11n), o quizá mejor 802.16 (distintas redes inalámbricas), son acciones que volverán a los consumidores muy escépticos. La interfaz de red tendrá que permanecer estable por décadas, así como los estándares de transmisión por televisión.

En quinto lugar, la seguridad y la confiabilidad serán de extrema importancia. Perder unos cuantos archivos debido a un virus de correo electrónico es una cosa; que un ladrón desarme nuestro sistema de seguridad desde su computadora móvil y después saquee nuestro hogar es muy distinto.

Una pregunta interesante es si las redes domésticas serán alámbricas o inalámbricas. La conveniencia y el costo favorecen a las redes inalámbricas, ya que no hay cables qué instalar (o peor aún, reinstalar). La seguridad favorece a las redes alámbricas, ya que las ondas de radio que utilizan las redes inalámbricas pueden traspasar las paredes con facilidad. No todos se alegran al saber que los vecinos se están colgando de su conexión a Internet y leyendo su correo electrónico. En el capítulo 8 estudiaremos cómo se puede utilizar el cifrado para proveer seguridad, aunque es más fácil decirlo que hacerlo cuando los usuarios son inexpertos.

Una tercera opción que podría ser interesante es la de reutilizar las redes que ya se encuentren en el hogar. El candidato más obvio es la red formada por los cables eléctricos instalados por toda la casa. Las **redes por el cableado eléctrico** permiten difundir información por toda la casa a los dispositivos que se conectan a los tomacorrientes. De todas formas usted tiene que conectar la TV, y de esta forma puede obtener conectividad a Internet al mismo tiempo. La dificultad está en cómo llevar tanto electricidad como señales de datos al mismo tiempo. Parte de la respuesta es que estas señales utilizan distintas bandas de frecuencia.

En resumen, las redes LAN domésticas ofrecen muchas oportunidades y retos. La mayoría de estos retos se relacionan con la necesidad de que las redes sean fáciles de manejar, confiables y seguras (en especial en manos de los usuarios inexpertos), así como de bajo costo.

1.2.3 Redes de área metropolitana

Una **Red de Área Metropolitana**, o **MAN** (*Metropolitan Area Network*), cubre toda una ciudad. El ejemplo más popular de una MAN es el de las redes de televisión por cable disponibles en muchas ciudades. Estos sistemas surgieron a partir de los primeros sistemas de antenas comunitarias que se utilizaban en áreas donde la recepción de televisión por aire era mala. En esos primeros sistemas se colocaba una gran antena encima de una colina cercana y después se canalizaba una señal a las casas de los suscriptores.

Al principio estos sistemas se diseñaban con fines específicos en forma local. Después, las empresas empezaron a entrar al negocio y consiguieron contratos de los gobiernos locales para cablear ciudades completas. El siguiente paso fue la programación de televisión e incluso canales completos diseñados sólo para cable. A menudo estos canales eran altamente especializados, como canales de sólo noticias, sólo deportes, sólo cocina, sólo jardinería, etc. Pero desde su comienzo hasta finales de la década de 1990, estaban diseñados sólo para la recepción de televisión.

Cuando Internet empezó a atraer una audiencia masiva, los operadores de red de TV por cable empezaron a darse cuenta de que con unos cambios en el sistema, podían proveer servicio de Internet de dos vías en partes no usadas del espectro. En ese momento, el sistema de TV por cable empezó a transformarse, de ser una simple forma de distribuir televisión, para convertirse en una red de área metropolitana. A simple vista, una MAN podría tener la apariencia del sistema que se muestra en la figura 1-9. En esta figura podemos ver que se alimentan señales de televisión y de Internet en un amplificador de cabecera para después distribuirlas a los hogares de las personas. Volveremos a ver este tema con detalle en el capítulo 2.

Cabe mencionar que la televisión por cable no es la única MAN. Los recientes desarrollos en el acceso inalámbrico a Internet de alta velocidad han originado otra, la cual se estandarizó como IEEE 802.16 y se conoce comúnmente como **WiMAX**. Hablaremos sobre ella en el capítulo 4.

1.2.4 Redes de área amplia

Una **Red de Área Amplia**, o **WAN** (*Wide Area Network*), abarca una extensa área geográfica, por lo general un país o continente. Empezaremos nuestra discusión con las redes WAN alámbricas y usaremos el ejemplo de una empresa con sucursales en distintas ciudades.

La WAN en la figura 1-10 es una red que conecta las oficinas en Perth, Melbourne y Brisbane. Cada una de estas oficinas contiene computadoras destinadas a ejecutar programas de usuario (aplicaciones). Seguiremos el uso tradicional y llamaremos a estas máquinas **hosts**. Al resto de la red que conecta estos hosts se le denomina **subred de comunicación**, o para abreviar sólo **subred**. La tarea de la subred es transportar los mensajes de host a host, al igual que el sistema telefónico transporta las palabras (en realidad sólo los sonidos) de la persona que habla a la persona que escucha.

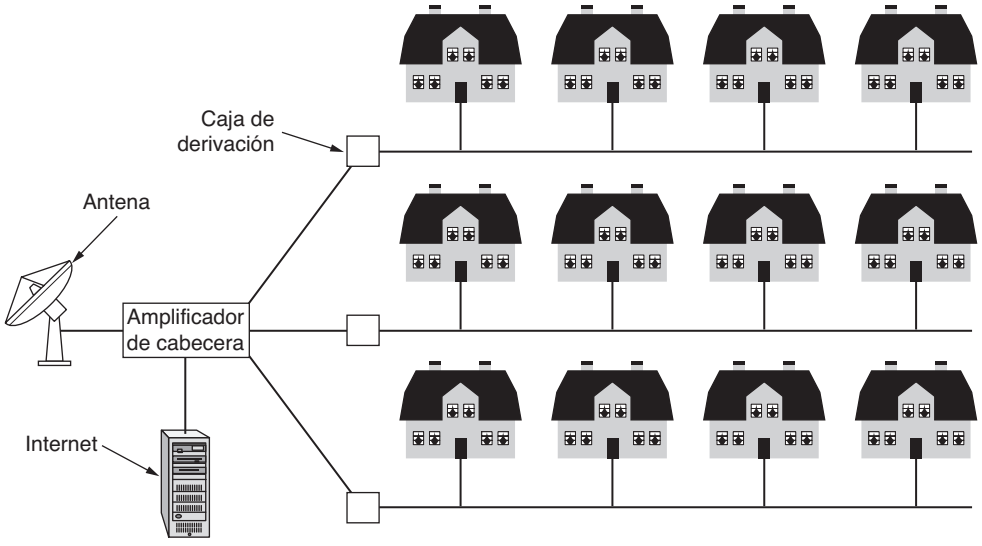


Figura 1-9. Una red de área metropolitana basada en la TV por cable.

En la mayoría de las redes WAN, la subred cuenta con dos componentes distintos: líneas de transmisión y elementos de conmutación. Las **líneas de transmisión** mueven bits entre máquinas. Se pueden fabricar a partir de alambre de cobre, fibra óptica o incluso enlaces de radio. Como la mayoría de las empresas no poseen líneas de transmisión, tienen que rentarlas a una compañía de telecomunicaciones. Los **elementos de conmutación** o **switches** son computadoras especializadas que conectan dos o más líneas de transmisión. Cuando los datos llegan por una línea entrante, el elemento de conmutación debe elegir una línea saliente hacia la cual reenviarlos. En el pasado, estas computadoras de conmutación han recibido varios nombres; ahora se conocen como **enrutador**.

Aprovechemos el momento para hablar un poco sobre el término "subred". En un principio, su **único** significado era el de una colección de enrutadores y líneas de comunicación que transmitían paquetes desde el host de origen hasta el host de destino. Es necesario que nuestros lectores sepan que ha adquirido un segundo significado más reciente en conjunto con el direccionamiento de red. Hablaremos sobre este significado en el capítulo 5 y mientras nos apegaremos al significado original (una colección de líneas y enrutadores).

Según nuestra descripción de la WAN, ésta es muy parecida a una LAN alámbrica extensa, sólo que hay ciertas diferencias importantes que van más allá de los cables extensos. Por lo general, en una WAN los hosts y la subred pertenecen a distintas personas, quienes actúan también como operadores. En nuestro ejemplo, los empleados podrían ser responsables de sus propias computadoras mientras que el departamento de TI de la empresa está a cargo del resto de la red. En los siguientes ejemplos veremos límites más claros, en donde el proveedor de red o compañía telefónica opera la subred. Al separar los aspectos exclusivos de comunicación (la subred) de los aspectos relacionados con la aplicación (los hosts) se simplifica en forma considerable el diseño de la red en general.

Una segunda diferencia es que los enrutadores por lo general conectan distintos tipos de tecnología de red. Por ejemplo, las redes dentro de las oficinas pueden usar la tecnología de Ethernet conmutada mientras que las líneas de transmisión de larga distancia pueden ser enlaces SONET (que veremos en el capítulo 2). Se requiere algún dispositivo para conectarlas. El lector inteligente observará que esto va más allá de nuestra definición de una red. Esto significa que muchas redes WAN serán de hecho **interredes**, o redes compuestas formadas por más de una red. En la siguiente sección veremos más detalles sobre las interredes.

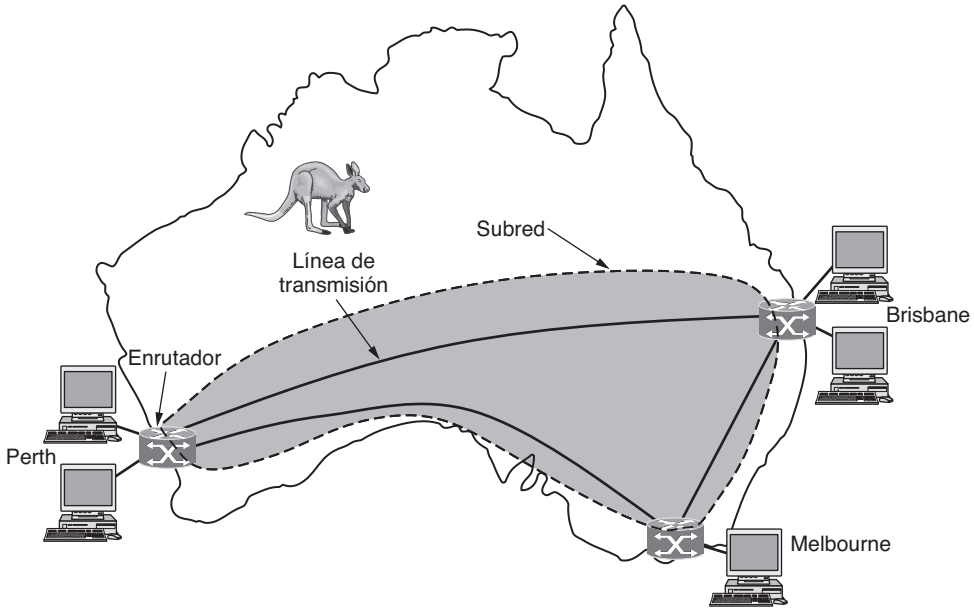


Figura 1-10. Una WAN que conecta tres sucursales en Australia.

Una última diferencia está en lo que se conecta a la subred. Podrían ser computadoras individuales, como en el caso de la conexión a redes LAN, o podrían ser redes LAN completas. Ésta es la forma en que se construyen redes más grandes a partir de otras más pequeñas. En lo que concierne a la subred, ésta hace el mismo trabajo.

Ahora estamos en posición de ver otras dos variedades de redes WAN. En primer lugar, en vez de rentar líneas de transmisión dedicadas, una empresa podría conectar sus oficinas a Internet. Esto le permite hacer conexiones entre las oficinas como enlaces virtuales que utilizan la capacidad subyacente de Internet. A este arreglo, que se muestra en la figura 1-11, se le denomina **VPN** (**Red Privada Virtual**, del inglés *Virtual Private Network*). Si se le compara con un arreglo dedicado, una VPN tiene la ventaja común de la virtualización, lo cual significa que provee flexibilidad en la reutilización de un recurso (conectividad a Internet). Para ver esto, considere lo fácil que sería conectar una cuarta oficina. Una VPN también tiene la desventaja común de la virtualización, lo cual significa que carece de control sobre los recursos subyacentes. Con una línea dedicada, la capacidad está clara. Con una VPN la capacidad puede variar según el servicio de Internet contratado.

La segunda variación es que una empresa distinta puede operar la subred. Al operador de la subred se le conoce como **proveedor de servicios de red** y las oficinas son sus clientes. En la figura 1-12 se muestra esta estructura. El operador de la subred se conecta también con otros clientes, siempre y cuando puedan pagar y les pueda proveer servicio. Como sería un servicio de red decepcionante si los clientes sólo pudieran enviarse paquetes entre sí, el operador de la subred también puede conectarse con otras redes que formen parte de Internet. A dicho operador de subred se le conoce como **ISP** (**Proveedor de Servicios de Internet**, del inglés *Internet Service Provider*) y la subred es una **red ISP**. Los clientes que se conectan al ISP reciben servicio de Internet.

Podemos usar la red ISP para ver por adelantado algunas cuestiones clave que estudiaremos en los capítulos posteriores. En la mayoría de las redes WAN, la red contiene muchas líneas de transmisión, cada una de las cuales conecta a un par de enrutadores. Si dos enrutadores que no comparten una línea de

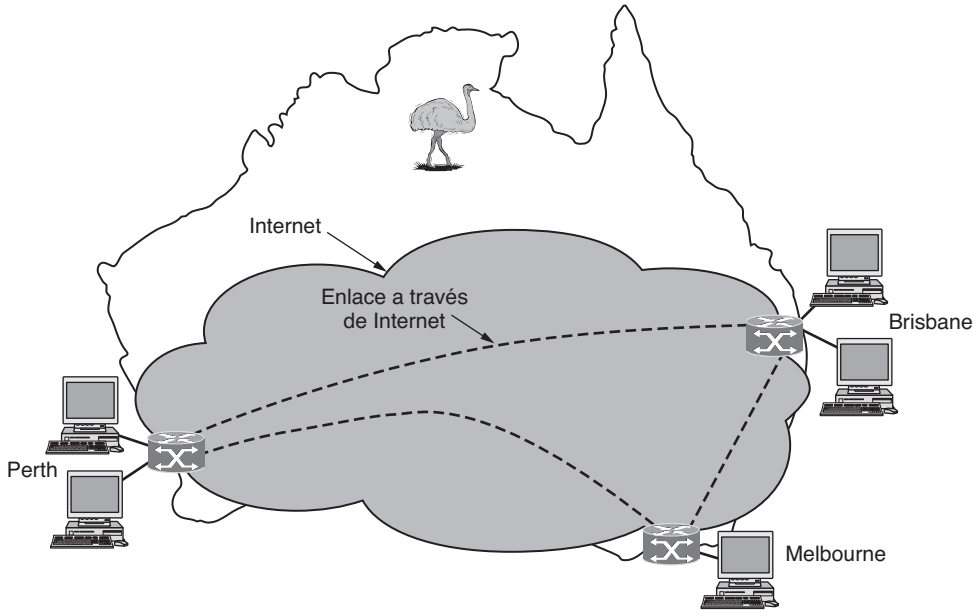


Figura 1-11. Una WAN que utiliza una red privada virtual.

transmisión desean comunicarse, deben hacerlo en forma indirecta a través de otros enrutadores. Puede haber muchas rutas en la red que conecten a estos dos enrutadores. Al proceso por el cual la red decide qué ruta tomar se le conoce como **algoritmo de enrutamiento**. Existen muchos algoritmos de este tipo. La manera en que cada enrutador toma la decisión de hacia dónde debe enviar el siguiente paquete se le denomina **algoritmo de reenvío**. También existen muchos de éstos. En el capítulo 5 estudiaremos ambos tipos de algoritmos con detalle.

Otros tipos de redes WAN utilizan mucho las tecnologías inalámbricas. En los sistemas de satélite, cada computadora en la Tierra tiene una antena a través de la cual es posible enviar y recibir datos de un satélite en órbita. Todas las computadoras pueden escuchar la salida *proveniente* del satélite y, en algunos casos, también pueden escuchar las transmisiones que envían sus computadoras vecinas *hacia* el satélite. Las redes de satélite son de difusión por naturaleza y son más útiles cuando es importante contar con la propiedad de difusión.

La red de telefonía celular es otro ejemplo de una WAN que utiliza tecnología inalámbrica. Este sistema ya pasó por tres generaciones y hay una cuarta por venir. La primera generación fue análoga y sólo para voz. La segunda fue digital y sólo para voz. La tercera generación es digital y se pueden transmitir tanto datos como voz. Cada estación base en un sistema celular cubre una distancia mucho mayor que una LAN inalámbrica, en donde el rango se mide en kilómetros en vez de decenas de metros. Las estaciones base se conectan entre sí mediante una red troncal que por lo general es alámbrica. Las velocidades de datos de las redes celulares se encuentran comúnmente en el orden de 1 Mbps, un valor mucho menor al de una LAN inalámbrica que puede estar en el orden de hasta 100 Mbps. En el capítulo 2 veremos muchos detalles sobre estas redes.

1.2.5 Interredes

Existen muchas redes en el mundo, a menudo con distintos componentes de hardware y software. Por lo general, las personas conectadas a una red se quieren comunicar con las personas conectadas a una red

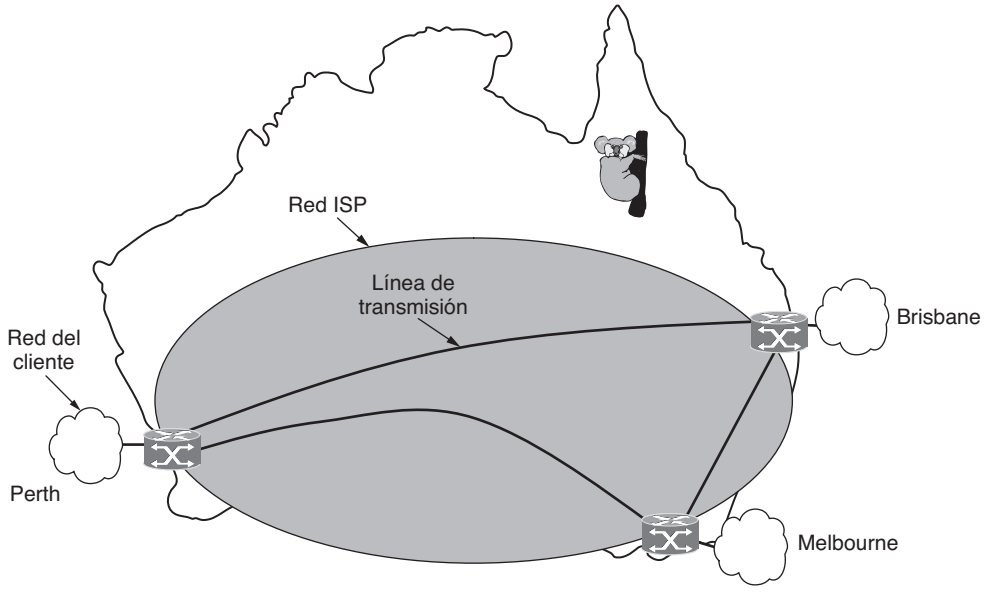


Figura 1-12. Una WAN que utiliza una red de ISP.

distinta; para lograrlo, es necesario conectar redes distintas que con frecuencia son incompatibles. A una colección de redes interconectadas se le conoce como **interred** o **internet**. Utilizaremos estos términos en un sentido genérico, en contraste a la red Internet mundial (que es una internet específica), a la cual nos referiremos siempre con I mayúscula. Internet usa redes de ISP para conectar redes empresariales, domésticas y muchos otros tipos más. Analizaremos la red Internet detalladamente más adelante.

A menudo se confunden las subredes, las redes y las interredes. El término "subred" tiene más sentido en el contexto de una red de área amplia, en donde se refiere a la colección de enrutadores y líneas de comunicación que pertenecen al operador de red. Como analogía, el sistema telefónico está compuesto por oficinas de conmutación telefónica conectadas entre sí mediante líneas de alta velocidad y conectadas a los hogares y negocios mediante líneas de baja velocidad. Estas líneas y equipos, que pertenecen y son administradas por la compañía telefónica, forman la subred del sistema telefónico. Los teléfonos en sí (los hosts en esta analogía) no forman parte de la subred.

Una red se forma al combinar una subred y sus hosts. Sin embargo, la palabra "red" a menudo también se utiliza en un sentido amplio. Podríamos describir una subred como una red, como en el caso de la "red ISP" de la figura 1-12. También podríamos describir una interred como una red, como en el caso de la WAN en la figura 1-10. Continuaremos con una práctica similar y cuando haya que diferenciar una red de otras distribuciones, nos apegaremos a nuestra definición original de una colección de computadoras interconectadas mediante una sola tecnología.

Ahora veamos detalladamente cómo está constituida una interred. Sabemos que una interred se forma cuando hay distintas redes interconectadas. A nuestro parecer, conectar una LAN y una WAN o conectar dos redes LAN es la forma usual de formar una interred, pero la industria no ha llegado a un buen acuerdo en cuanto a la terminología utilizada en esta área. Hay dos reglas prácticas y útiles a este respecto. En primer lugar, si varias organizaciones han pagado para construir distintas partes de la red y cada una se encarga de dar mantenimiento a la parte que le corresponde, entonces tenemos una interred en vez de una sola red. En segundo lugar, si la tecnología subyacente es distinta en diferentes partes (por ejemplo, difusión frente punto a punto y alámbrica frente a inalámbrica), es probable que sea una interred.

Para profundizar en este tema, hablaremos sobre la forma en que se pueden conectar dos redes distintas. El nombre general para una máquina que realiza una conexión entre dos o más redes y provee la traducción necesaria, tanto en términos de hardware como de software, es **puerta de enlace** (*gateway*). Las puertas de enlace se distinguen por la capa en la que operan en la jerarquía de protocolos. En la siguiente sección hablaremos mucho más sobre las capas y las jerarquías de protocolos, pero por ahora basta con imaginar que las capas superiores están más relacionadas con las aplicaciones (como la web), mientras que las capas inferiores están más relacionadas con los enlaces de transmisión (como Ethernet).

Como el beneficio de formar una internet es para conectar computadoras entre distintas redes, no es conveniente usar una puerta de enlace de una capa demasiado baja, ya que no podremos realizar conexiones entre distintos tipos de redes. Tampoco es conveniente usar una puerta de enlace de una capa demasiado alta, o de lo contrario la conexión sólo funcionará para ciertas aplicaciones. A la capa en la parte media que resulta ser la "ideal" se le denomina comúnmente **capa de red**; un enrutador es una puerta de enlace que conmuta paquetes en la capa de red. Así, para detectar una interred o internet hay que buscar una red que tenga enrutadores.

1.3 SOFTWARE DE RED

Las primeras redes de computadoras se diseñaron teniendo en cuenta al hardware como punto principal y al software como secundario. Pero esta estrategia ya no funciona. Ahora el software de red está muy estructurado. En las siguientes secciones examinaremos con cierto detalle la técnica para estructurar el software. La metodología aquí descrita constituye la piedra angular de todo el libro y, por lo tanto, se repetirá en secciones posteriores.

1.3.1 Jerarquías de protocolos

Para reducir la complejidad de su diseño, la mayoría de las redes se organizan como una pila de **capas** o **niveles**, cada una construida a partir de la que está abajo. El número de capas, su nombre, el contenido de cada una y su función difieren de una red a otra. El propósito de cada capa es ofrecer ciertos servicios a las capas superiores, mientras les oculta los detalles relacionados con la forma en que se implementan los servicios ofrecidos. Es decir, cada capa es un tipo de máquina virtual que ofrece ciertos servicios a la capa que está encima de ella.

En realidad este concepto es familiar y se utiliza en muchas áreas de las ciencias computacionales, en donde se le conoce de muchas formas: ocultamiento de información, tipos de datos abstractos, encapsulamiento de datos y programación orientada a objetos. La idea fundamental es que una pieza particular de software (o hardware) provee un servicio a sus usuarios pero mantiene ocultos los detalles de su estado interno y los algoritmos que utiliza.

Cuando la capa *n* en una máquina lleva a cabo una conversación con la capa *n* en otra máquina, a las reglas y convenciones utilizadas en esta conversación se les conoce como el protocolo de la capa *n*. En esencia, un **protocolo** es un acuerdo entre las partes que se comunican para establecer la forma en que se llevará a cabo esa comunicación. Como analogía, cuando a un hombre le presentan una mujer, ella puede elegir si extiende su mano o no. Él a su vez, puede decidir entre estrechar la mano o besarla, dependiendo por ejemplo de si ella es una abogada estadounidense en una reunión de negocios, o una princesa europea en un baile formal. Si se viola el protocolo se hará más difícil la comunicación, si no es que se vuelve imposible.

En la figura 1-13 se ilustra una red de cinco capas. Las entidades que conforman las correspondientes capas en diferentes máquinas se llaman **iguales** (*peers*). Los iguales pueden ser procesos de software,

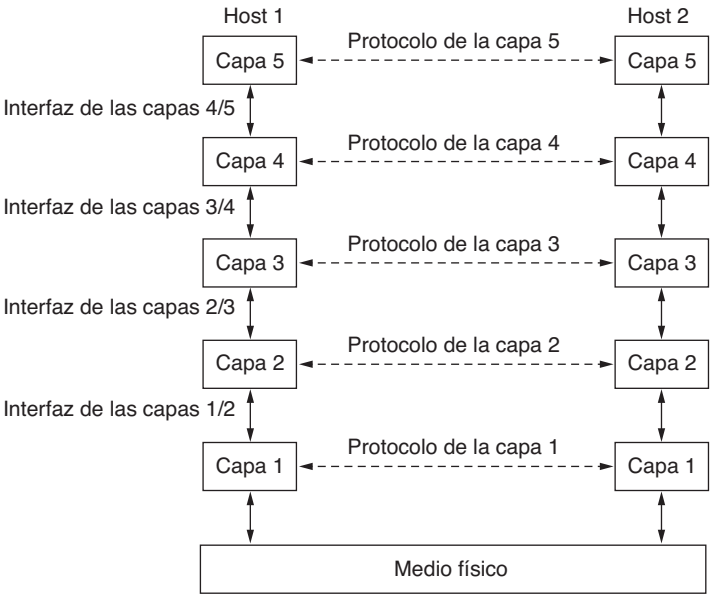


Figura 1-13. Capas, protocolos e interfaces.

dispositivos de hardware o incluso seres humanos. En otras palabras, los iguales son los que se comunican a través del protocolo.

En realidad no se transfieren datos de manera directa desde la capa *n* de una máquina a la capa *n* de otra máquina, sino que cada capa pasa los datos y la información de control a la capa inmediatamente inferior, hasta que se alcanza a 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 se muestra la comunicación virtual con líneas punteadas y la comunicación física con líneas sólidas.

Entre cada par de capas adyacentes hay una **interfaz**. Ésta define las operaciones y servicios primitivos que 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, la consideración más importante es definir interfaces limpias entre las capas. Al hacer esto es necesario 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 también simplifican el reemplazo de una capa con un protocolo o implementación totalmente diferente (por ejemplo, reemplazar todas las líneas telefónicas por canales de satélite), ya que todo lo que se requiere del nuevo protocolo o implementación es que ofrezca exactamente el mismo conjunto de servicios a su vecino de arriba, como lo hacía el protocolo o la implementación anterior. Es común que distintos hosts utilicen diferentes implementaciones del mismo protocolo (a menudo escrito por otras compañías). De hecho, el protocolo en sí puede cambiar en cierta capa sin que las capas superior e inferior lo noten.

A un conjunto de capas y protocolos se le conoce como **arquitectura de red**. La especificación de una arquitectura debe contener suficiente información como para permitir que un programador escriba el programa o construya el hardware para cada capa, de manera que se cumpla correctamente el protocolo apropiado. Ni los detalles de la implementación ni la especificación de las interfaces forman parte de la arquitectura, ya que están ocultas dentro de las máquinas y no se pueden ver desde el exterior. Ni siquiera es necesario que las interfaces en todas las máquinas de una red sean iguales, siempre y cuando cada máquina pueda utilizar todos los protocolos correctamente. La lista de los protocolos utilizados por cierto

sistema, un protocolo por capa, se le conoce como **pila de protocolos**. Las arquitecturas de red, las pilas de protocolos y los protocolos mismos son los temas principales de este libro.

Una analogía podría ayudar a explicar la idea de la comunicación entre múltiples capas. Imagine a dos filósofos (procesos de iguales en la capa 3), uno de los cuales habla urdú e inglés, mientras que el otro habla chino y francés. Como no tienen un lenguaje común, cada uno contrata a un traductor (procesos de iguales en la capa 2) y cada uno de los traductores 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 ello pasa un mensaje (en español) a través de la interfaz de las capas 2-3 a su traductor para decirle: "Me gustan los conejos", como se muestra en la figura 1-14. Los traductores han acordado un idioma neutral conocido por ambos, el holandés, así el mensaje es convertido a "*Ik vind konijnen leuk*". La elección del idioma es el protocolo de la capa 2 y depende de los procesos de iguales de dicha capa.

Después, el traductor pasa el mensaje a una secretaria para que lo transmita, por ejemplo, mediante correo electrónico (el protocolo de la capa 1). Cuando el mensaje llega a la otra secretaria, ésta lo pasa al traductor local, quien lo traduce al francés y lo pasa a través de la interfaz de las capas 2-3 al segundo filósofo 2. Observe que cada protocolo es totalmente independiente de los demás siempre y cuando no cambien las interfaces. Por ejemplo, los traductores pueden cambiar de holandés al finlandés siempre y cuando ambos estén de acuerdo y ninguno cambie su interfaz con las capas 1 o 3. De manera similar, las secretarias pueden cambiar del correo electrónico al teléfono sin molestar (o incluso informar) a las demás capas. Cada proceso puede agregar algo de información destinada sólo a su igual. Esta información no se pasa a la capa superior.

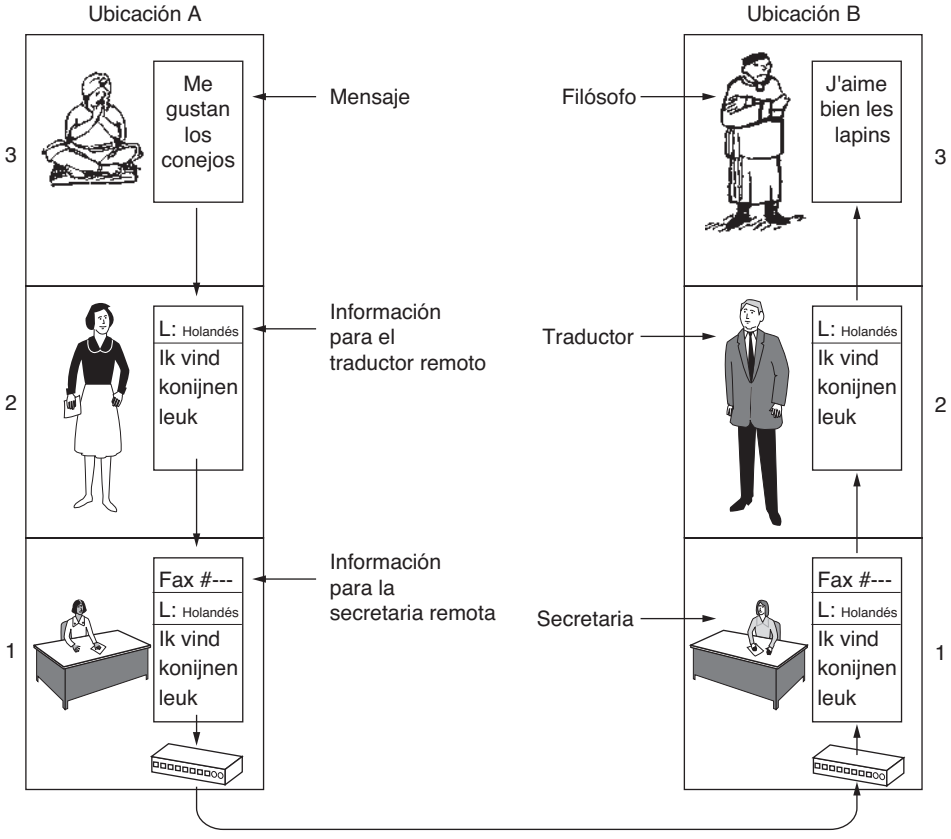


Figura 1-14. La arquitectura filósofo-traductor-secretaria.

Ahora considere un ejemplo más técnico: cómo proveer 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 que lo transmita. La capa 4 coloca un **encabezado** al frente del mensaje para identificarlo y pasa el resultado a la capa 3. El encabezado incluye información de control, como direcciones, para permitir que la capa 4 en la máquina de destino entregue el mensaje. Otros ejemplos de la información de control que se utiliza en algunas capas son los números de secuencia (en caso de que la capa inferior no preserve el orden del mensaje), los tamaños y los tiempos.

En muchas redes no se impone un límite en cuanto al tamaño de los mensajes que se transmiten en el protocolo de la capa 4, pero casi siempre hay un límite impuesto por el protocolo de la capa 3. En consecuencia, la capa 3 debe descomponer los mensajes entrantes en unidades más pequeñas llamadas paquetes, y colocar un encabezado al frente de cada paquete. En este ejemplo, *M* se divide en dos partes: M_1 y M_2, los cuales se transmitirán por separado.

La capa 3 decide cuál de las líneas salientes usar y pasa los paquetes a la capa 2; esta última agrega a cada pieza no sólo un encabezado, sino también un terminador, y pasa la unidad restante a la capa 1 para su transmisión física. En la máquina receptora el mensaje pasa hacia arriba, de capa en capa, y los encabezados se van eliminando a medida que progresa. Ninguno de los encabezados para las capas inferiores a *n* se pasa a la capa *n*.

Lo importante a entender sobre la figura 1-15 es la relación entre la comunicación virtual y real, además de la diferencia entre los protocolos y las interfaces. Por ejemplo, los procesos de iguales en la capa 4 piensan conceptualmente en su comunicación como si fuera "horizontal" y utilizan el protocolo de la capa 4. Es probable que cada uno tenga procedimientos llamados *EnviarAlOtroLado* y *RecibirDelOtroLado*, aun cuando en realidad estos procedimientos se comunican con las capas inferiores a través de la interfaz de las capas 3-4, no con el otro lado.

La abstracción de los procesos de iguales es imprescindible para todo diseño de red. Al usarla, 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.

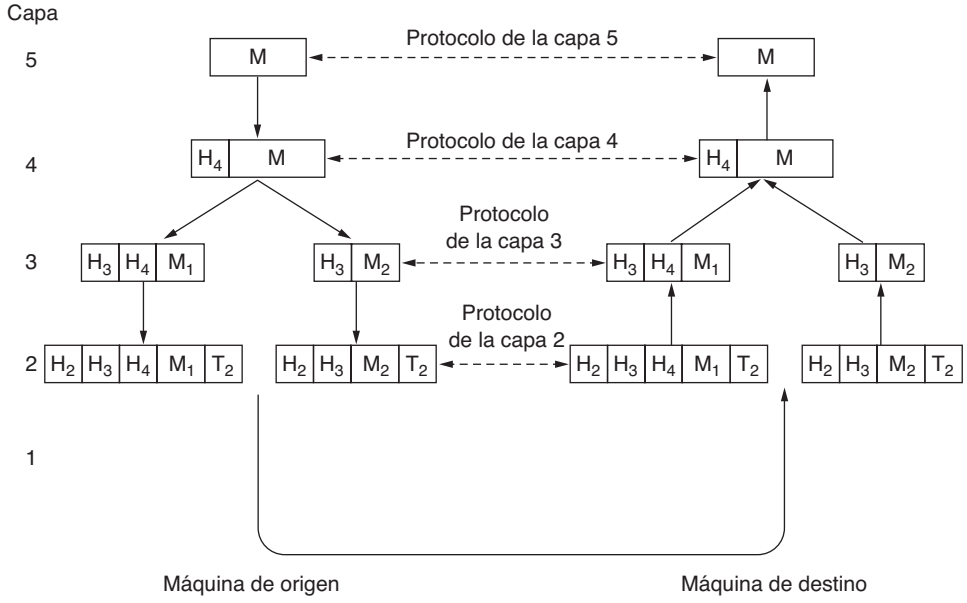


Figura 1-15. Ejemplo de flujo de información que soporta la comunicación virtual en la capa 5.

Aunque la sección 1.3 se llama “Software de red”, vale la pena mencionar que las capas inferiores de una jerarquía de protocolos se implementan con frecuencia en el hardware o firmware. Sin embargo, están implicados los algoritmos de protocolos complejos, incluso aunque estén integrados (en todo o en parte) al hardware.

1.3.2 Aspectos de diseño para las capas

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

La confiabilidad es el aspecto de diseño enfocado en verificar que una red opere correctamente, aun cuando esté formada por una colección de componentes que sean, por sí mismos, poco confiables. Piense en los bits de un paquete que viajan a través de la red. Existe la posibilidad de que algunas de estas piezas se reciban dañadas (invertidas) debido al ruido eléctrico, a las señales aleatorias inalámbricas, a fallas en el hardware, a errores del software, etc. ¿Cómo es posible detectar y corregir estos errores?

Un mecanismo para detectar errores en la información recibida utiliza códigos de **detección de errores**. Así, la información que se recibe de manera incorrecta puede retransmitirse hasta que se reciba de manera correcta. Los códigos más poderosos cuentan con **corrección de errores**, en donde el mensaje correcto se recupera a partir de los bits posiblemente incorrectos que se recibieron originalmente. Ambos mecanismos funcionan añadiendo información redundante. Se utilizan en capas bajas para proteger los paquetes que se envían a través de enlaces individuales, y en capas altas para verificar que el contenido correcto fue recibido.

Otro aspecto de la confiabilidad consiste en encontrar una ruta funcional a través de una red. A menudo hay múltiples rutas entre origen y destino, y en una red extensa puede haber algunos enlaces o enrutadores descompuestos. Suponga que la red está caída en Alemania. Los paquetes que se envían de Londres a Roma a través de Alemania no podrán pasar, pero para evitar esto, podríamos enviar los paquetes de Londres a Roma vía París. La red debería tomar esta decisión de manera automática. A este tema se le conoce como **enrutamiento**.

Un segundo aspecto de diseño se refiere a la evolución de la red. Con el tiempo, las redes aumentan su tamaño y emergen nuevos diseños que necesitan conectarse a la red existente. Recientemente vimos el mecanismo de estructuración clave que se utiliza para soportar el cambio dividiendo el problema general y ocultando los detalles de la implementación: **distribución de protocolos en capas**. También existen muchas otras estrategias.

Como hay muchas computadoras en la red, cada capa necesita un mecanismo para identificar los emisores y receptores involucrados en un mensaje específico. Este mecanismo se conoce como **direccionamiento** o **nombramiento** en las capas altas y bajas, respectivamente.

Un aspecto del crecimiento es que las distintas tecnologías de red a menudo tienen diferentes limitaciones. Por ejemplo, no todos los canales de comunicación preservan el orden de los mensajes que se envían en ellos, por lo cual es necesario idear soluciones para enumerar los mensajes. Otro ejemplo es el de las diferencias en el tamaño máximo de un mensaje que las redes pueden transmitir. Esto provoca la creación de mecanismos para desensamblar, transmitir y después volver a ensamblar los mensajes. A este tema en general se le conoce como **interconexión de redes** (*internetworking*).

Cuando las redes crecen, surgen nuevos problemas. Las ciudades pueden tener problemas de tráfico, escasez de números telefónicos y es fácil perderse. No muchas personas tienen estos problemas en su propio vecindario, pero en toda la ciudad pueden representar un gran problema. Se dice que los diseños que siguen funcionando bien cuando la red aumenta su tamaño son **escalables**.

Un tercer aspecto de diseño radica en la asignación de recursos. Las redes proveen un servicio a los hosts desde sus recursos subyacentes, como la capacidad de las líneas de transmisión. Para hacer bien su

trabajo necesitan mecanismos que dividan sus recursos de manera que un host no interfiera demasiado con otro host.

Muchos diseños comparten el ancho de banda de una red en forma dinámica, de acuerdo con las necesidades a corto plazo de los hosts, en vez de otorgar a cada host una fracción fija del ancho de banda que puede llegar a utilizar o quizás no. A este diseño se le denomina **multiplexado estadístico**, lo cual significa que se comparten los recursos con base en la demanda. Se puede aplicar en capas bajas para un solo enlace o en capas altas para una red, o incluso para aplicaciones que utilizan la red.

Un problema de asignación que ocurre en todas las capas es cómo evitar que un emisor rápido inunde de datos a un receptor lento. Con frecuencia se utiliza retroalimentación del receptor al emisor. A este tema se le denomina **control de flujo**. Algunas veces el problema es que la red sufre un exceso de solicitudes debido a que hay demasiadas computadoras que desean enviar una gran cantidad de información y la red no lo puede entregar todo. A esta sobrecarga de la red se le conoce como **congestión**. Una estrategia es que cada computadora reduzca su demanda cuando experimenta congestión. Esto también se puede usar en todas las capas.

Es interesante observar que la red puede ofrecer más recursos que simplemente el ancho de banda. Para usos como transmitir video en vivo, la puntualidad de la entrega es en extremo importante. La mayoría de las redes deben proveer servicio a las aplicaciones que desean esta entrega en **tiempo real** al mismo tiempo que proveen servicio a las aplicaciones que desean un alto rendimiento. La **calidad del servicio** es el nombre que se da a los mecanismos que reconcilian estas demandas competitivas.

El último aspecto de diseño importante es asegurar la red y defenderla contra distintos tipos de amenazas. Una de las amenazas que mencionamos antes es la de espiar las comunicaciones. Los mecanismos que proveen **confidencialidad** nos defienden contra esta amenaza y se utilizan en múltiples capas. Los mecanismos de **autenticación** evitan que alguien se haga pasar por otra persona. Se pueden usar para diferenciar los sitios web bancarios falsos de los verdaderos, o para permitir que la red celular verifique que una llamada realmente provenga de nuestro teléfono para pagar la cuenta. Otros mecanismos para la **integridad** evitan cambios clandestinos a los mensajes, como cuando se altera el mensaje "cargar $10 a mi cuenta" para convertirlo en "cargar $1000 dólares a mi cuenta". Todos estos diseños se basan en la criptografía que estudiaremos en el capítulo 8.

1.3.3 Comparación entre servicio orientado a conexión y servicio sin conexión

Las capas pueden ofrecer dos tipos distintos de servicio a las capas superiores: orientado a conexión y sin conexión. En esta sección analizaremos estos dos tipos y examinaremos las diferencias entre ellos.

El servicio **orientado a conexión** está modelado a partir del sistema telefónico. Para hablar con alguien levantamos el auricular, marcamos el número, hablamos y después colgamos. De manera similar, para usar un servicio de red orientado a conexión, el usuario del servicio establece primero una conexión, la utiliza y después la libera. El aspecto esencial de una conexión es que funciona como un tubo: el emisor mete objetos (bits) en un extremo y el receptor los toma en el otro extremo. En la mayoría de los casos se conserva el orden de manera que los bits llegan en el orden en el que se enviaron.

En algunos casos al establecer una conexión, el emisor, el receptor y la subred llevan a cabo una **negociación** en cuanto a los parámetros que se van a usar, como el tamaño máximo del mensaje, la calidad requerida del servicio y demás cuestiones relacionadas. Por lo general, uno de los lados hace una propuesta y el otro puede aceptarla, rechazarla o elaborar una contrapropuesta. Un circuito es otro nombre para una conexión con recursos asociados, como un ancho de banda fijo. Esto se remonta a la red telefónica, en la cual un circuito era una ruta sobre alambre que transmitía una conversación telefónica.

En contraste al servicio orientado a la conexión, el servicio **sin conexión** está modelado a partir del sistema postal. Cada mensaje (carta) lleva la dirección de destino completa, y cada uno es enrutado hacia

los nodos intermedios dentro del sistema, en forma independiente a todos los mensajes subsecuentes. Hay distintos nombres para los mensajes en diferentes contextos: un **paquete** es un mensaje en la capa de red. Cuando los nodos intermedios reciben un mensaje completo antes de enviarlo al siguiente nodo, se le llama **conmutación de almacenamiento y envío**. La alternativa en donde la transmisión subsiguiente de un mensaje en un nodo empieza antes de que éste la reciba por completo, se conoce como "conmutación al vuelo". Por lo 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 primero que se envié se retrase de manera que el segundo llegue primero.

Cada tipo de servicio se puede caracterizar con base en su confiabilidad. Algunos servicios son confiables en cuanto a que nunca pierden datos. Por lo general, para implementar un servicio confiable, el receptor tiene que confirmar la recepción de cada mensaje, de manera que el emisor esté seguro de que hayan llegado. El proceso de confirmación de recepción introduce sobrecarga y retardos, que a menudo valen la pena pero algunas veces no son deseables.

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

El servicio confiable orientado a la conexión tiene dos variaciones menores: secuencias de mensajes y flujos de bytes. En la primera variante se conservan los límites de los mensajes. Cuando se envían dos mensajes de 1024 bytes, llegan como dos mensajes distintos de 1024 bytes y nunca como un mensaje de 2048 bytes. En la segunda variante, la conexión es simplemente un flujo de bytes sin límites en los mensajes. Cuando llegan 2048 bytes al receptor, no hay manera de saber si se enviaron como un mensaje de 2048 bytes, como dos mensajes de 1024 bytes o como 2048 mensajes de 1 byte. Si se envían las páginas de un libro a través de una red a una máquina de fotocomposición en forma de mensajes separados, probablemente sea importante preservar los límites de los mensajes. Por otro lado, para descargar una película en DVD, todo lo que se necesita es un flujo de bytes del servidor a la computadora del usuario. Los límites de los mensajes dentro de la película no son relevantes.

En algunas aplicaciones, los retardos de tránsito ocasionados por las confirmaciones de recepción son inaceptables. Una de estas aplicaciones es el tráfico de voz digitalizada o **voz sobre IP**. Es preferible para los usuarios del teléfono escuchar un poco de ruido en la línea de vez en cuando que experimentar un retardo al esperar las confirmaciones de recepción. De manera similar, al transmitir una conferencia de video no hay problema si unos cuantos píxeles están mal, pero es molesto cuando la imagen se sacude mientras el flujo se detiene y avanza para corregir errores.

No todas las aplicaciones requieren conexiones. Por ejemplo, los emisores de correo electrónico basura (*spammers*) envían su correo a muchos destinatarios. Es probable que el emisor no quiera tener que pasar por el problema de establecer y desmantelar una conexión con un destinatario sólo para enviarle un mensaje. Tampoco es esencial una entrega cien por ciento confiable, sobre todo si eso es más costoso. Todo lo que se requiere es una forma de enviar un solo mensaje que tenga una muy alta probabilidad de llegar, aunque sin garantías. Al servicio sin conexión no confiable (que significa sin confirmación de recepción) se le denomina servicio de **datagramas**, en analogía al servicio de telegramas que tampoco devuelve una confirmación de recepción al emisor. A pesar de ser poco confiable, es la forma más dominante en la mayoría de las redes por motivos que veremos más adelante.

En otros casos es conveniente no tener que establecer una conexión para enviar un mensaje, pero la confiabilidad es esencial. En estas aplicaciones se puede utilizar el servicio de **datagramas con confirmación de recepción**. Es como enviar una carta certificada y solicitar una confirmación de recepción. Al regresar la confirmación de recepción el emisor tiene la absoluta certeza de que la carta se entregó al destinatario correcto y que no se perdió en el camino. La mensajería de texto en los teléfonos móviles es un ejemplo.

Hay otro servicio conocido como servicio de **solicitud-respuesta**. En este servicio el emisor transmite un solo datagrama que contiene una solicitud; al receptor envía la respuesta. El servicio de solicitud-respuesta se utiliza mucho para implementar la comunicación en el modelo cliente-servidor; el cliente emite una petición y el servidor le responde. Por ejemplo, el cliente de un teléfono móvil podría enviar una consulta a un servidor de mapas para recuperar los datos del mapa de la ubicación actual. En la figura 1-16 se sintetizan los tipos de servicios antes descritos.

	Servicio	Ejemplo
Orientado a conexión	Flujo de mensajes confiable.	Secuencia de páginas.
	Flujo de bytes confiable.	Descarga de películas.
	Conexión no confiable.	Voz sobre IP.
Sin conexión	Datagrama no confiable.	Correo electrónico basura.
	Datagrama confirmación de recepción.	Mensajería de texto.
	Solicitud-respuesta.	Consulta en una base de datos.

Figura 1-16. Seis tipos distintos de servicios.

Tal vez el concepto de usar una comunicación poco confiable le parezca confuso en un principio. Después de todo, ¿por qué preferiría alguien una comunicación poco confiable en vez de una comunicación confiable? Primero que nada, tal vez la comunicación confiable (en nuestro contexto significa que es con confirmación de recepción) no esté disponible en cierta capa. Por ejemplo, Ethernet no provee una comunicación confiable. Los paquetes se pueden dañar ocasionalmente durante el tránsito. Las capas de protocolos más altas deben tener la capacidad de recuperarse de este problema. En particular, muchos servicios confiables se basan en un servicio de datagramas no confiables. En segundo lugar, los retardos inherentes al proveer un servicio confiable tal vez sean inaceptables, en especial en las aplicaciones de tiempo real como multimedia. Éstas son las razones por las que coexisten la comunicación confiable y la comunicación poco confiable.

1.3.4 Primitivas de servicios

Un servicio se puede especificar de manera formal como un conjunto de **primitivas** (operaciones) disponibles a los procesos de usuario para que accedan al servicio. Estas primitivas le indican al servicio que desarrollen alguna acción o que informen sobre la acción que haya tomado una entidad par. Si la pila de protocolos se encuentra en el sistema operativo, como se da en la mayoría de los casos, por lo general las primitivas son llamadas al sistema. Estas llamadas provocan un salto al modo de kernel, que a su vez devuelve el control de la máquina al sistema operativo para que envíe los paquetes necesarios.

El conjunto de primitivas disponibles depende de la naturaleza del servicio que se va a ofrecer. Las primitivas para el servicio orientado a conexión son distintas de las primitivas para el servicio sin conexión. Como un ejemplo mínimo de las primitivas de servicio que se podrían ofrecer para implementar un flujo de bytes confiable, considere las primitivas que se enlistan en la figura 1-17. Estas primitivas serán familiares para los fanáticos de la interfaz de sockets de Berkeley, ya que son una versión simplificada de esa interfaz.

Primitiva	Significado
LISTEN	Bloquea en espera de una conexión entrante.
CONNECT	Establece una conexión con un igual en espera.
ACCEPT	Acepta una conexión entrante de un igual.
RECEIVE	Bloquea en espera de un mensaje entrante.
SEND	Envía un mensaje al igual.
DISCONNECT	Termina una conexión.

Figura 1-17. Seis primitivas de servicios que proveen un servicio simple orientado a conexión.

Podríamos usar estas primitivas para una interacción petición-respuesta en un entorno cliente-servidor. Para ilustrar esto, vamos a esbozar un protocolo simple que implementa el servicio mediante datagramas con confirmación de recepción.

Primero, el servidor ejecuta LISTEN para indicar que está preparado para aceptar conexiones entrantes. Una forma común de implementar LISTEN es mediante una llamada de bloqueo del sistema. Después de ejecutar la primitiva, el proceso servidor se bloquea hasta que aparezca una petición de conexión.

Después, el proceso cliente ejecuta CONNECT para establecer una conexión con el servidor. La llamada a CONNECT necesita especificar con quién se va a realizar la conexión, por lo que podría incluir un parámetro para proporcionar la dirección del servidor. A continuación, lo más común es que el sistema operativo envíe un paquete al igual para pedirle que se conecte, como se muestra en la sección (1) de la figura 1-18. El proceso cliente se suspende hasta que haya una respuesta.

Cuando el paquete llega al servidor, el sistema operativo ve que el paquete solicita una conexión. Verifica que haya alguien escuchando y, en ese caso, desbloquea al que está escuchando. Ahora el proceso servidor puede establecer la conexión con la llamada a ACCEPT. Esta llamada envía una respuesta (2) de vuelta al proceso cliente para aceptar la conexión. Al llegar esta respuesta se libera el cliente. En este punto, el cliente y el servidor se están ejecutando y tienen una conexión establecida.

La analogía obvia entre este protocolo y la vida real es un cliente que llama al gerente de servicio al cliente de una empresa. Al empezar el día, el gerente de servicio se sienta a un lado de su teléfono en caso de que suene. Después, un cliente hace una llamada. Cuando el gerente levanta el teléfono se establece la conexión.

El siguiente paso es que el servidor ejecute RECEIVE y se prepare para aceptar la primera petición. Por lo general, el servidor hace esto justo después de ser liberado de la primitiva LISTEN, antes de que la confirmación de recepción pueda regresar al cliente. La llamada a RECEIVE bloquea al servidor.

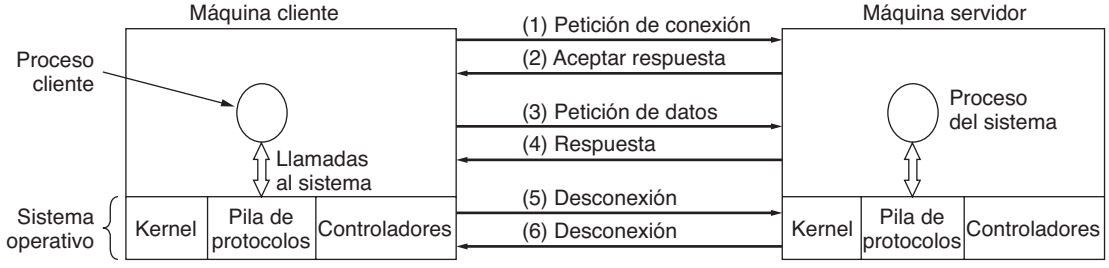


Figura 1-18. Una interacción cliente-servidor simple mediante el uso de datagramas con confirmación de recepción.

Entonces, el cliente ejecuta SEND para transmitir su petición (3) después de ejecutar RECEIVE para obtener la respuesta. La llegada del paquete solicitado a la máquina servidor desbloquea el servidor, de manera que pueda manejar la petición. Después de realizar su trabajo, el servidor usa SEND para devolver la respuesta al cliente (4). Al llegar este paquete se desbloquea el cliente, que ahora puede inspeccionar la respuesta. Si el cliente tiene peticiones adicionales, puede hacerlas ahora.

Cuando el cliente termina, ejecuta DISCONNECT para terminar la conexión (5). Por lo general una primitiva DISCONNECT inicial es una llamada de bloqueo, la cual suspende al cliente y envía un paquete al servidor para indicar que ya no necesita la conexión. Cuando el servidor recibe el paquete también emite una primitiva DISCONNECT por su cuenta, envía una confirmación de recepción al cliente y libera la conexión (6). Cuando el paquete del servidor regresa a la máquina cliente, se libera el proceso cliente y se interrumpe la conexión. En esencia, así es como funciona la comunicación orientada a conexión.

Por desgracia la vida no es tan simple. Aquí pueden salir mal muchas cosas. La sincronización puede estar mal (por ejemplo, que termine CONNECT antes de LISTEN), se pueden perder paquetes, etc. Más adelante analizaremos con mayor detalle estas cuestiones, pero por el momento en la figura 1-18 se resume la forma en que podría trabajar la comunicación cliente-servidor mediante datagramas con confirmación de recepción para poder ignorar los paquetes perdidos.

Dado que se requieren seis paquetes para completar este protocolo, tal vez se pregunte por qué no utilizar mejor un protocolo sin conexión. La respuesta es que en un mundo perfecto podría ser así, en cuyo caso sólo se necesitarían dos paquetes: uno para la petición y otro para la respuesta. Pero cuando hay mensajes extensos en cualquier dirección (por ejemplo, un archivo de un megabyte), errores de transmisión y paquetes perdidos, la situación cambia. Si la respuesta consistiera de cientos de paquetes, algunos de los cuales se pudieran perder durante la transmisión, ¿cómo sabría el cliente que faltan algunas piezas?, ¿cómo sabría si el último paquete que se recibió fue en realidad el último paquete enviado? Suponga que el cliente desea un segundo archivo. ¿Cómo podría diferenciar el paquete 1 del segundo archivo de un paquete 1 perdido del primer archivo que por fin pudo llegar al cliente? En resumen, en el mundo real es inadecuado usar un protocolo simple de petición-respuesta a través de una red poco confiable. En el capítulo 3 estudiaremos con detalle una variedad de protocolos que solucionan éstos y otros problemas. Por el momento basta con decir que algunas veces es conveniente tener un flujo de bytes ordenado y confiable entre procesos.

1.3.5 La relación entre servicios y protocolos

Los servicios y los protocolos son conceptos distintos. Esta distinción es tan importante que la enfatizaremos una vez más. Un *servicio* es un conjunto de primitivas (operaciones) que una capa proporciona a la capa que está encima de ella. El servicio define qué operaciones puede realizar la capa en beneficio de sus usuarios, pero no dice nada sobre cómo se implementan estas operaciones. Un servicio se relaciona con una interfaz entre dos capas, en donde la capa inferior es el proveedor del servicio y la capa superior es el usuario.

En contraste, un *protocolo* es un conjunto de reglas que rigen el formato y el significado de los paquetes o mensajes que intercambian las entidades iguales en una capa. Las entidades utilizan protocolos para implementar sus definiciones de servicios. Pueden cambiar sus protocolos a voluntad, siempre y cuando no cambien el servicio visible para sus usuarios. De esta manera, el servicio y el protocolo no dependen uno del otro. Éste es un concepto clave que cualquier diseñador de red debe comprender bien.

Para repetir este punto importante, los servicios se relacionan con las interfaces entre capas, como se muestra en la figura 1-19. En contraste, los protocolos se relacionan con los paquetes que se envían entre las entidades pares de distintas máquinas. Es muy importante no confundir los dos conceptos.

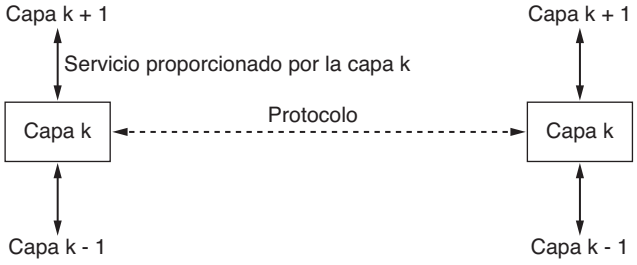


Figura 1-19. La relación entre un servicio y un protocolo.

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

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

1.4 MODELOS DE REFERENCIA

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

1.4.1 El modelo de referencia OSI

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

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

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

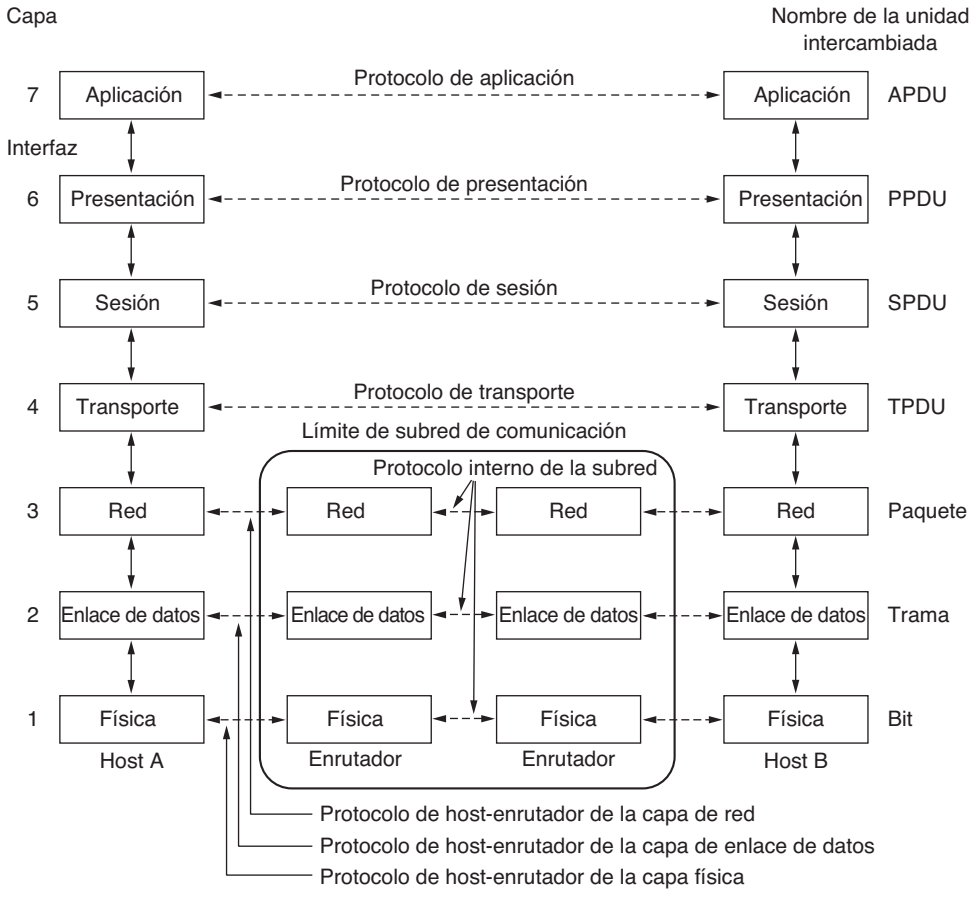


Figura 1-20. El modelo de referencia OSI.

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

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

La capa física

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

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

La capa de enlace de datos

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

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

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

La capa de red

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

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

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

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

La capa de transporte

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

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

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

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

La capa de sesión

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

La capa de presentación

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

La capa de aplicación

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

1.4.2 El modelo de referencia TCP/IP

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

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

La capa de enlace

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

La capa de interred

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

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

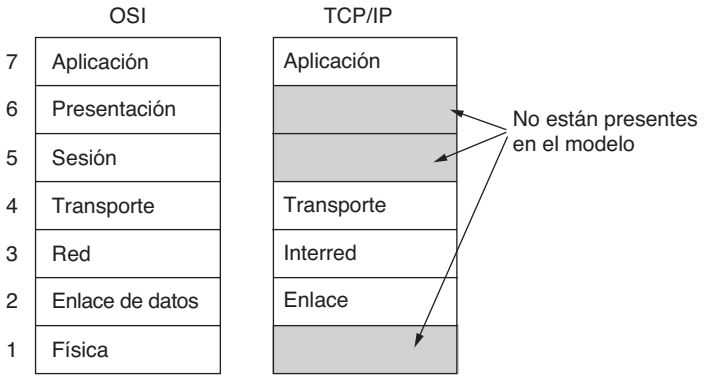


Figura 1-21. El modelo de referencia TCP/IP.

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

La capa de transporte

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

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

La capa de aplicación

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

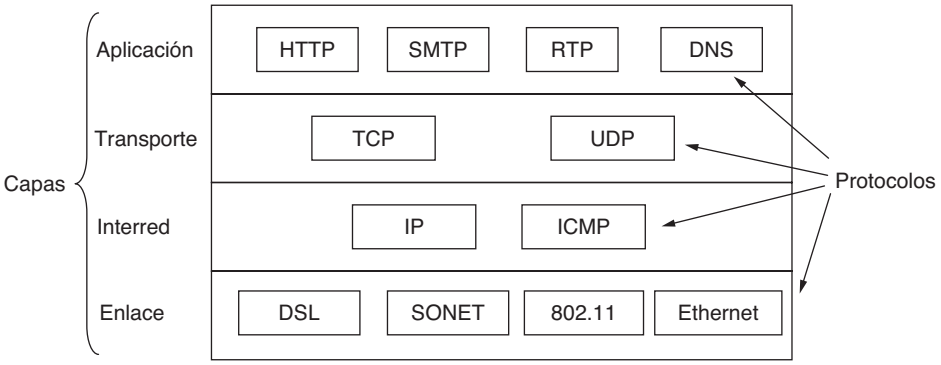


Figura 1-22. El modelo TCP/IP con algunos de los protocolos.

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

1.4.3 El modelo utilizado en este libro

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

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

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

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

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

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

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

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

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

Hay tres conceptos básicos para el modelo OSI:

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

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

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

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

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

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

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

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

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

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

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

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

1.4.5 Una crítica al modelo y los protocolos OSI

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

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

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

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

Mala sincronización

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

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

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

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

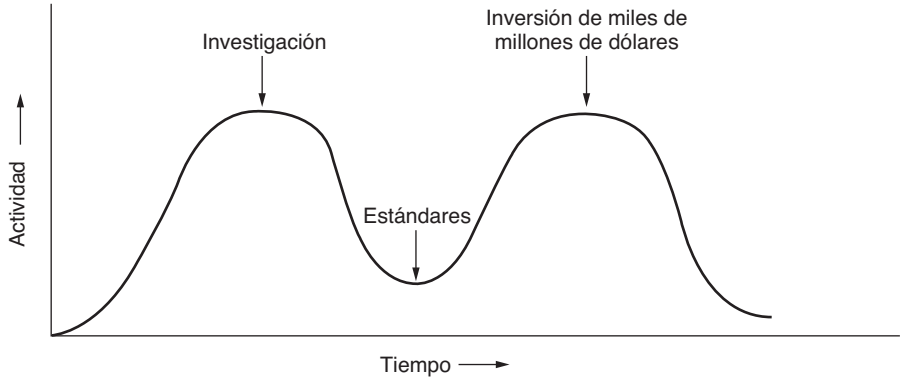


Figura 1-24. El apocalipsis de los dos elefantes.

Mala tecnología

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

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

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

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

Malas implementaciones

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

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

Malas políticas

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

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

1.4.6 Una crítica al modelo de referencia TCP/IP

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

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

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

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

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

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

1.5 REDES DE EJEMPLO

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

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

1.5.1 Internet

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

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

ARPANET

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

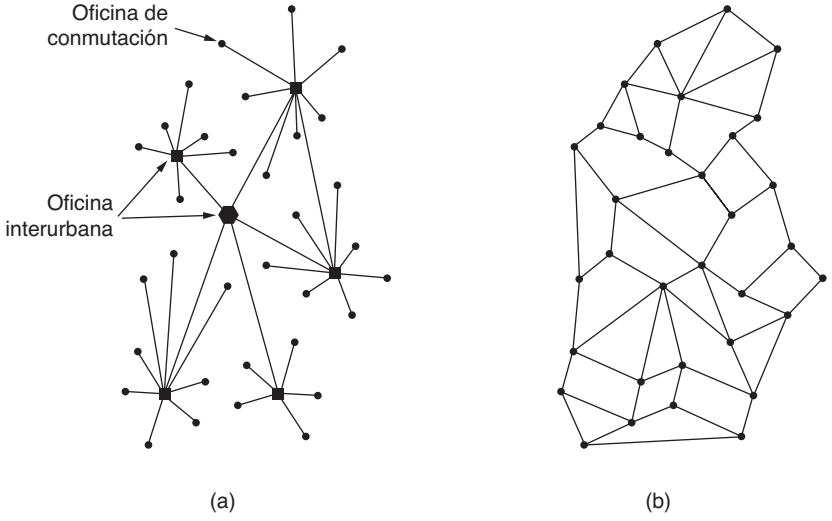


Figura 1-25. (a) Estructura de un sistema telefónico. (b) El sistema de conmutación distribuida propuesto por Baran.

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

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

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

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

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

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

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

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

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

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

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

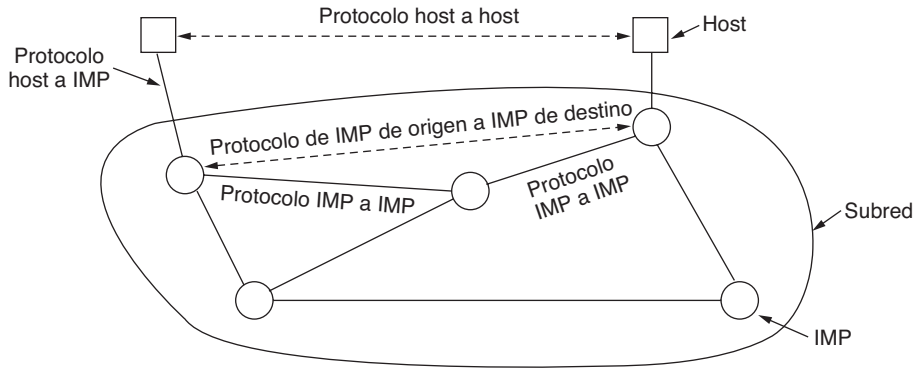


Figura 1-26. Diseño original de ARPANET.

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

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

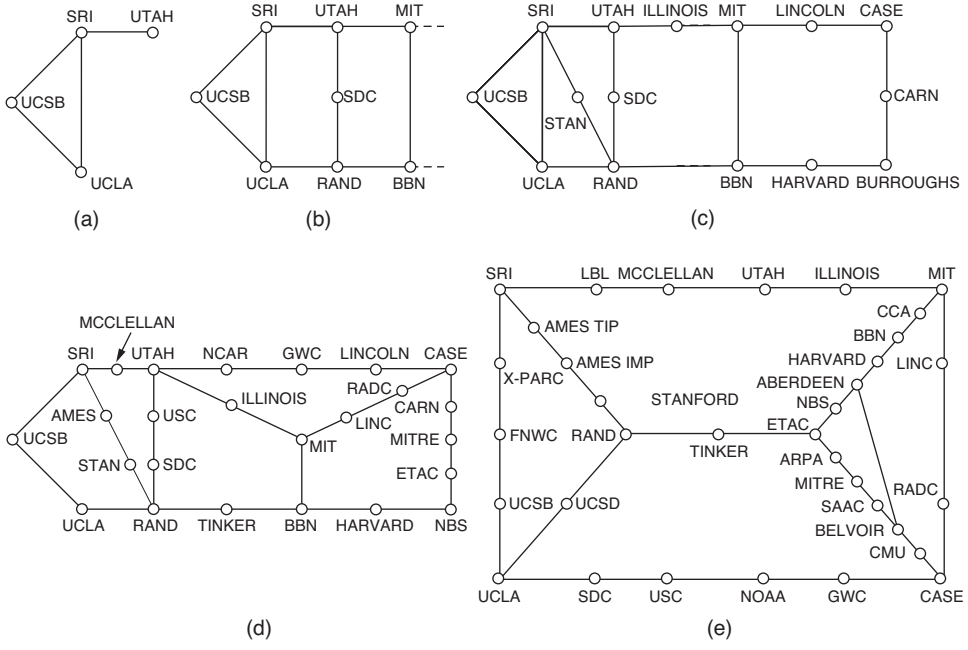


Figura 1-27. Crecimiento de ARPANET. (a) Diciembre de 1969. (b) Julio de 1970. (c) Marzo de 1971. (d) Abril de 1972. (e) Septiembre de 1972.

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

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

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

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

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

NSFNET

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

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

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

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

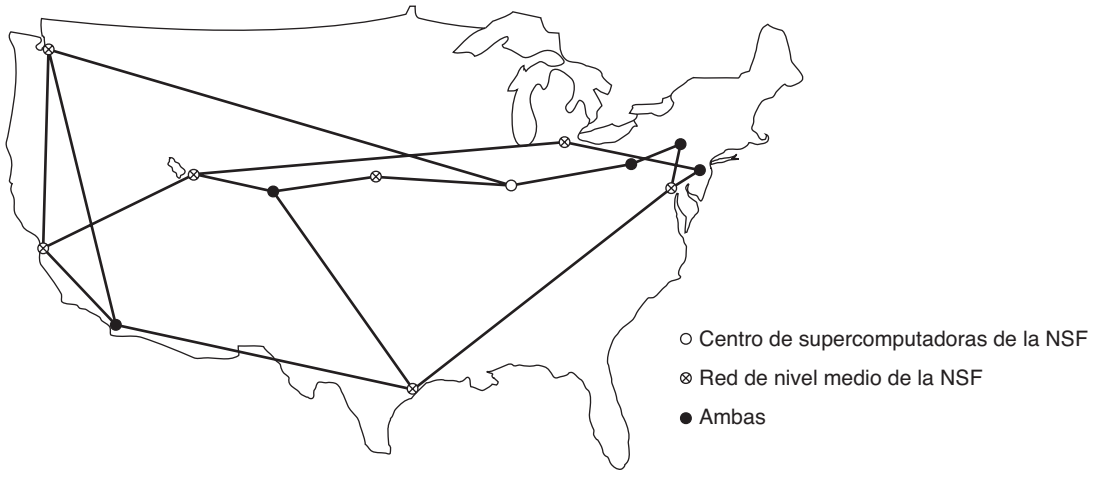


Figura 1-28. La red troncal de NSFNET en 1988.

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

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

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

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

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

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

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

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

Arquitectura de Internet

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

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

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

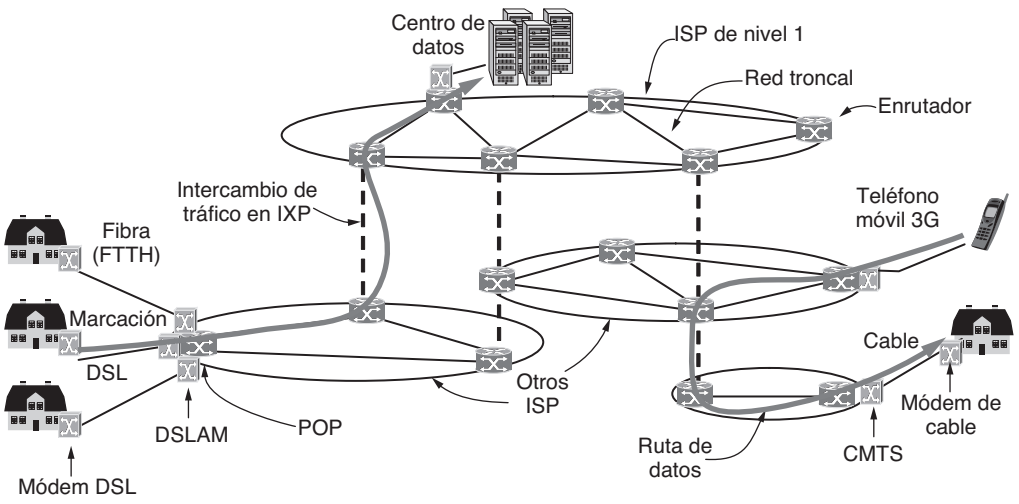


Figura 1-29. Generalidades sobre la arquitectura de Internet.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.5.2 Redes de teléfonos móviles de tercera generación

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

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

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

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

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

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

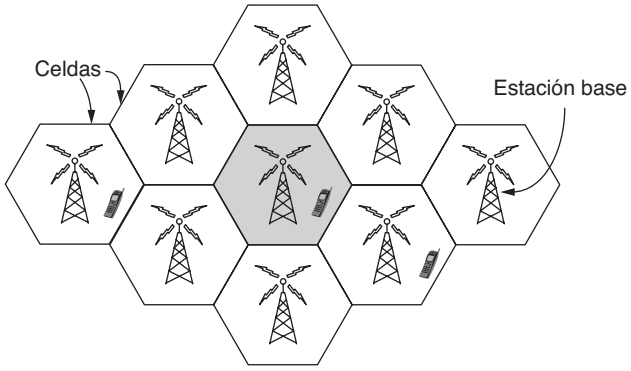


Figura 1-30. Diseño celular de las redes de telefonía móvil.

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

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

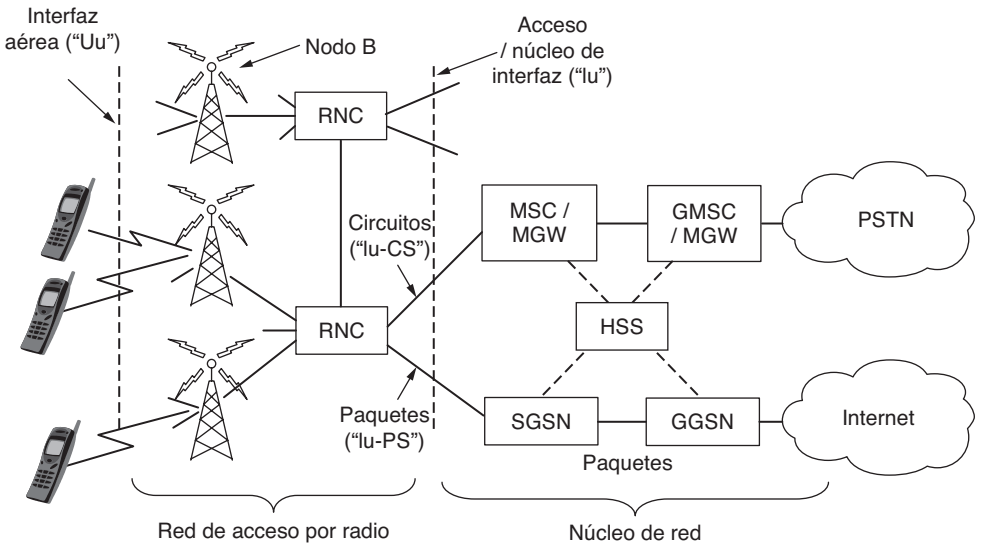


Figura 1-31. Arquitectura de la red de telefonía móvil 3G UTMS.

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

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

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

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

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

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

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

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

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

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

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

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

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

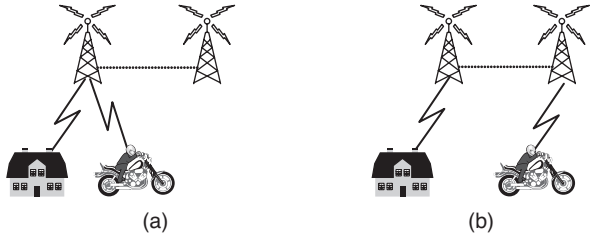


Figura 1-32. Traspaso de telefonía móvil (a) antes, (b) después.

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

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

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

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

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

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

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

1.5.3 Redes LAN inalámbricas: 802.11

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

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

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

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

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

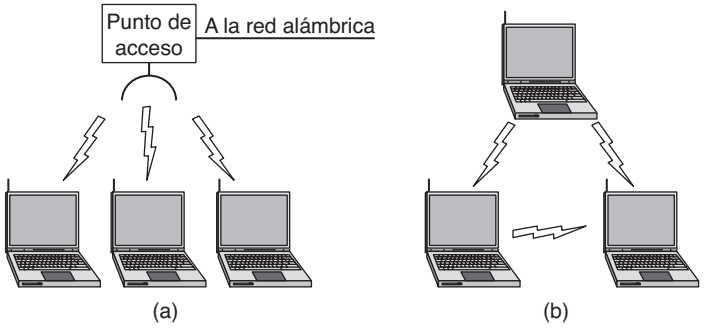


Figura 1-33. (a) Red inalámbrica con un punto de acceso. (b) Red *ad hoc*.

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

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

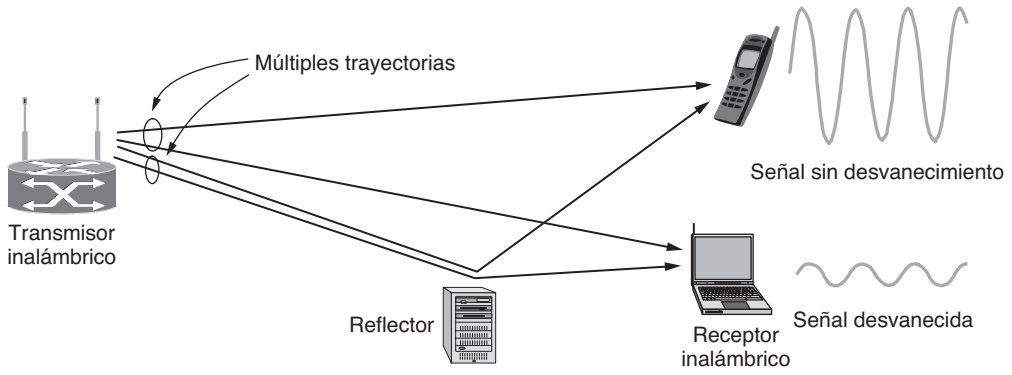


Figura 1-34. Desvanecimiento multitrayectorias.

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

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

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

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

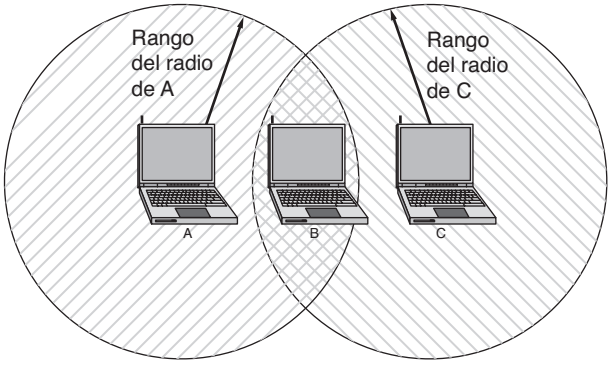


Figura 1-35. El rango de un solo radio tal vez no cubra todo el sistema.

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

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

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

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

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

1.5.3 Redes RFID y de sensores

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

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

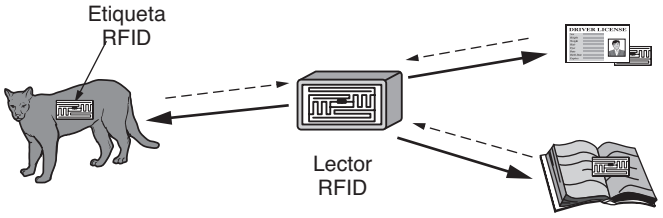


Figura 1-36. La tecnología RFID se utiliza para conectar objetos cotidianos en red.

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

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

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

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

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

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

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

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

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

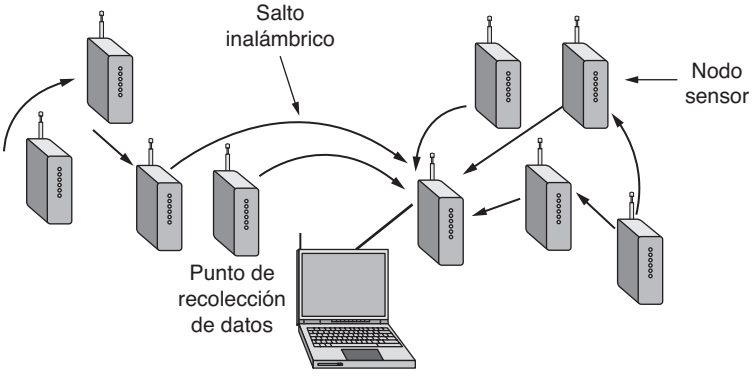


Figura 1-37. Topología multisaltos de una red de sensores.

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

1.6 ESTANDARIZACIÓN DE REDES

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

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

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

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

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

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

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

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

1.6.1 Quién es quién en el mundo de las telecomunicaciones

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

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

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

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

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

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

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

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

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

1.6.2 Quién es quién en el mundo de los estándares internacionales

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

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

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

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

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

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

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

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

1.6.3 Quién es quién en el mundo de estándares de Internet

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

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

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

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

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

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

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

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

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

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

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

1.7 UNIDADES MÉTRICAS

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

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	Zetta
10^{-24}	0.000000000000000000000001	yocto	10^{24}	1 000 000 000 000 000 000 000 000	Yotta

Figura 1-39. Los principales prefijos métricos.

También vale la pena señalar que para medir los tamaños de memoria, disco, archivos y bases de datos, en la práctica común de la industria las unidades tienen significados ligeramente distintos. Así, kilo significa 2^{10} (1 024) en vez de 10^{3} (1 000), ya que las memorias son siempre una potencia de dos. Por ende, una memoria de 1 KB contiene 1 024 bytes, no 1 000 bytes. Observe también que se utiliza una letra "B" mayúscula que significa "bytes" (unidades de ocho bits), en vez de una "b" minúscula que significa "bits". De manera similar, una memoria de 1 MB contiene 2^{20} (1 048 576) bytes, una memoria de 1 GB contiene 2^{30} (1 073 741 824) bytes y una base de datos de 1 TB contiene 2^{40} (1 099 511 627 776) bytes. Sin embargo, una línea de comunicación de 1 kbps transmite 1000 bits por segundo y una red LAN de 10 Mbps opera a 10 000 000 bits/segundo, ya que estas velocidades no son potencias de dos. Por desgracia, muchas personas tienden a mezclar estos dos sistemas, en especial con los tamaños de los discos. Para evitar ambigüedades, en este libro utilizaremos los símbolos KB, MB, GB y TB para 2^{10}, 2^{20}, 2^{30} y 2^{40} bytes, respectivamente, y los símbolos kbps, Mbps, Gbps y Tbps para 10^{3}, 10^{6}, 10^{9} y 10^{12} bits/segundo, respectivamente.

1.8 ESQUEMA DEL RESTO DEL LIBRO

Este libro trata tanto los principios como la práctica de las redes de computadoras. La mayor parte de los capítulos empiezan con una explicación de los principios relevantes, seguida de varios ejemplos que ilustran estos principios. Por lo general estos ejemplos se toman de Internet y de las redes inalámbricas tales como la red de telefonía móvil, ya que ambas son importantes y muy distintas. Donde sea necesario también se dan otros ejemplos.

El libro está estructurado de acuerdo con el modelo híbrido de la figura 1-23. A partir del capítulo 2 comenzaremos a subir por la jerarquía de protocolos, empezando desde los cimientos. Veremos algunos antecedentes en el campo de la comunicación de datos que cubren a los sistemas de transmisión alámbricos e inalámbricos. Este material se enfoca en cómo entregar la información a través de los canales físicos, aunque sólo cubriremos los aspectos de arquitectura, no los de hardware. También veremos varios ejemplos de la capa física, como la red pública de telefonía conmutada, la red de telefonía móvil y la red de televisión por cable.

Los capítulos 3 y 4 tratan sobre la capa de enlace de datos en dos partes. El capítulo 3 analiza el problema de cómo enviar paquetes a través de un enlace, incluyendo la detección y corrección de errores.

Analizaremos la tecnología DSL (que se utiliza para el acceso de banda ancha a Internet sobre líneas telefónicas) como un ejemplo real de un protocolo de enlace de datos.

En el capítulo 4 examinaremos la subcapa de acceso al medio. Ésta es la parte de la capa de enlace de datos que se encarga de cómo compartir un canal entre varias computadoras. Los ejemplos que veremos incluyen redes inalámbricas, como 802.11 y RFID, además de redes LAN alámbricas como Ethernet clásica. Aquí también veremos los switches de la capa de enlace que conectan redes LAN, como Ethernet conmutada.

El capítulo 5 trata sobre la capa de red, en especial el enrutamiento. Veremos muchos algoritmos de enrutamiento, tanto estáticos como dinámicos. Incluso aunque existan buenos algoritmos de enrutamiento, si existe más tráfico del que la red pueda manejar, algunos paquetes se retrasarán o desecharán. Explicaremos esta cuestión, desde cómo evitar la congestión hasta cómo garantizar cierta calidad de servicio. Al conectar redes heterogéneas entre sí para formar interredes también se producen numerosos problemas, de los que hablaremos aquí. Además explicaremos con detalle la capa de red en Internet.

El capítulo 6 trata acerca de la capa de transporte. Daremos mucho énfasis a los protocolos orientados a conexión y la confiabilidad, ya que muchas aplicaciones los necesitan. Explicaremos también con detalle los protocolos de transporte de Internet: UDP y TCP, junto con sus aspectos de rendimiento.

El capítulo 7 se encarga de la capa de aplicación, sus protocolos y aplicaciones. El primer tema es DNS, que es el directorio telefónico de Internet. Después hablaremos sobre el correo electrónico, incluyendo una explicación de sus protocolos. Luego pasaremos a la web, con explicaciones detalladas del contenido estático y dinámico, además de lo que ocurre en los lados cliente y servidor. Más tarde analizaremos la multimedia en red, incluyendo audio y video de flujo continuo. Por último hablaremos sobre las redes de entrega de contenido, incluyendo la tecnología de igual a igual.

El capítulo 8 habla sobre la seguridad en las redes. Este tema tiene aspectos que se relacionan con todas las capas, por lo que es más fácil tratarlo después de haber explicado todas las capas a detalle. El capítulo empieza con una introducción a la criptografía. Después muestra cómo se puede utilizar la criptografía para garantizar la seguridad en las comunicaciones, el correo electrónico y la web. El capítulo termina con una explicación de algunas áreas en las que la seguridad choca con la privacidad, la libertad de expresión, la censura y otras cuestiones sociales.

El capítulo 9 contiene una lista con anotaciones de las lecturas sugeridas, ordenadas por capítulos. El objetivo es ayudar a los lectores que desean llevar más allá su estudio de las redes. Ese capítulo también incluye una bibliografía alfabética de todas las referencias citadas en este libro.

El sitio web de los autores en Pearson tiene una página con vínculos a muchos tutoriales, preguntas frecuentes (FAQ), compañías, consorcios industriales, organizaciones profesionales, organizaciones de estándares, tecnologías, documentos y demás.

1.9 RESUMEN

Las redes de computadoras tienen muchos usos, tanto para empresas como para individuos, en el hogar y en movimiento. Las empresas usan redes de computadoras para compartir la información corporativa, por lo general mediante el modelo cliente-servidor en donde las computadoras de los empleados actúan como clientes que acceden a poderosos servidores en la sala de máquinas. Para los individuos, las redes ofrecen acceso a una variedad de recursos de información y entretenimiento, así como una manera de comprar y vender productos y servicios. Con frecuencia los individuos acceden a Internet por medio de sus

proveedores de teléfono o cable en el hogar, aunque cada vez se utiliza más el acceso inalámbrico para laptops y teléfonos. Los avances tecnológicos permiten nuevos tipos de aplicaciones móviles y redes con computadoras integradas a los electrodomésticos y demás dispositivos para el consumidor. Los mismos avances generan cuestiones sociales tales como las relacionadas con la privacidad.

En términos generales, podemos dividir a las redes en LAN, MAN, WAN e interredes. Por lo general las redes LAN cubren todo un edificio y operan a velocidades altas. Las redes MAN comúnmente cubren toda una ciudad. El sistema de televisión por cable es un ejemplo, ya que ahora muchas personas lo utilizan para acceder a Internet. Las redes WAN pueden cubrir un país o continente. Algunas de las tecnologías utilizadas para construir estas redes son de punto a punto (como un cable), mientras que otras son de difusión (como las redes inalámbricas). Las redes se pueden interconectar con enrutadores para formar interredes, de las cuales Internet es el ejemplo más grande y popular. Las redes inalámbricas, como las redes LAN 802.11 y de telefonía móvil 3G, también se están volviendo muy populares.

El software de red se basa en los protocolos, que son reglas mediante las cuales los procesos se comunican entre sí. La mayoría de las redes soportan jerarquías de protocolos, en donde cada capa proporciona servicios a la capa inmediata superior y los aísla de los detalles sobre los protocolos que se utilizan en las capas inferiores. Por lo general las pilas de protocolos se basan en el modelo OSI o en el modelo TCP/IP. Ambos modelos tienen capas de enlace, red, transporte y aplicación, pero difieren en las otras capas. Los aspectos de diseño incluyen: confiabilidad, asignación de recursos, crecimiento, seguridad, etcétera. Gran parte de este libro se enfoca en los protocolos y su diseño.

Las redes proveen varios servicios a sus usuarios. Estos servicios pueden variar, desde la entrega de paquetes sin conexión de mejor esfuerzo hasta la entrega garantizada orientada a conexión. En algunas redes se proporciona servicio sin conexión en una capa y servicio orientado a conexión en la capa inmediata superior.

Entre las redes más conocidas están: Internet, la red de telefonía móvil 3G y las redes LAN 802.11. Internet evolucionó de ARPANET, a la que se agregaron otras redes para formar una interred. En realidad la Internet de la actualidad es una colección de muchos miles de redes que utilizan la pila de protocolos TCP/IP. La red de telefonía móvil 3G proporciona acceso inalámbrico y móvil a Internet, con velocidades de varios Mbps; además transmite llamadas de voz. Las redes LAN inalámbricas basadas en el estándar IEEE 802.11 se implementan en muchos hogares y cafés; pueden proporcionar conectividad a velocidades en mayores 100 Mbps. También están surgiendo nuevos tipos de redes, como las redes de sensores integradas y las redes basadas en tecnología RFID.

Para permitir que varias computadoras se comuniquen entre sí se requiere una gran cantidad de estandarización, tanto en hardware como en software. Las organizaciones tales como ITU-T, ISO, IEEE e IAB administran distintas partes del proceso de estandarización.

PROBLEMAS

1. Imagine que entrenó a Bernie, su perro San Bernardo, para que transporte una caja de tres cintas de 8 mm en vez de un termo con brandy (cuando se llene su disco, puede considerarlo una emergencia). Cada una de estas cintas contiene 7 gigabytes. El perro puede viajar a donde quiera que vaya, a una velocidad de 18 km/h. ¿Para qué rango de distancias tiene Bernie una velocidad mayor de datos que una línea de transmisión cuya velocidad de datos (sin sobrecarga) es de 150 Mbps? ¿Cómo cambiaría su respuesta si (i) se duplica la velocidad de Bernie; (ii) se duplica la capacidad de cada cinta; (iii) se duplica la velocidad de datos de la línea de transmisión?

2. Una alternativa a una LAN es simplemente un gran sistema de tiempo compartido con terminales para los usuarios. Cite dos ventajas de un sistema cliente-servidor que utiliza una LAN.

3. El rendimiento de un sistema cliente-servidor se ve muy influenciado por dos características principales de las redes: el ancho de banda de la red (es decir, cuántos bits/segundo puede transportar) y la latencia (cuántos segundos tarda el primer bit en viajar del cliente al servidor). Cite un ejemplo de una red que cuente con un ancho de banda alto pero también alta latencia. Después mencione un ejemplo de una red que tenga un ancho de banda bajo y una baja latencia.
4. Además del ancho de banda y la latencia, ¿qué otro parámetro se necesita para tener una buena caracterización de la calidad del servicio ofrecido por una red que se utiliza para:

   (i) tráfico de voz digitalizada?

   (ii) tráfico de video?

   (iii) tráfico de transacciones financieras?
5. Un factor en el retardo de un sistema de conmutación de paquetes de almacenamiento y envío es cuánto tiempo se requiere para almacenar y enviar un paquete a través de un switch. Si el tiempo de conmutación es de 10 μseg, ¿es probable que sea un factor importante en la respuesta de un sistema cliente-servidor en donde el cliente está en Nueva York y el servidor en California? Asuma que la velocidad de propagación en cobre y fibra óptica es de 2/3 la velocidad de la luz en el vacío.
6. Un sistema cliente-servidor utiliza una red satelital, en donde el satélite está 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 todos tengan una terminal casera conectada a una red de computadoras, serán posibles los consultas públicas instantáneas sobre asuntos legislativos pendientes. En algún momento las legislaturas existentes se podrían eliminar para dejar que el deseo del pueblo se exprese de manera directa. Los aspectos positivos de tal democracia directa son bastante obvios; comente sobre algunos de los aspectos negativos.
8. Cinco enrutadores se van a conectar a una subred de punto a punto. Entre cada par de enrutadores, los diseñadores pueden colocar una línea de alta velocidad, una de velocidad media, una de baja velocidad o ninguna línea. Si se requieren 100 ms de tiempo de la computadora para generar e inspeccionar cada topología, ¿cuánto tiempo se requiere para inspeccionarlas todas?
9. Una desventaja de una subred de difusión es la capacidad que se desperdicia cuando varios hosts tratan de acceder al canal al mismo tiempo. Como ejemplo simplista, suponga que el tiempo se divide en porciones (ranuras) discretas y que cada uno de los n hosts trata de usar el canal con una probabilidad de p durante cada porción de tiempo. ¿Qué fracción de las porciones se desperdiciará debido a las colisiones?
10. ¿Cuáles son dos razones para usar protocolos en capas? ¿Cuál es una posible desventaja de usar protocolos en capas?
11. A la presidenta de la empresa Specialty Paint Corp. se le ocurre la idea de trabajar con un fabricante de cerveza local para producir una lata de cerveza invisible (como medida para reducir la basura). La presidenta ordena a los de su departamento legal que investiguen el asunto; ellos a su vez piden ayuda al departamento de ingeniería. Como resultado, el ingeniero en jefe llama a su homólogo en la compañía de cerveza para discutir los aspectos técnicos del proyecto. Después los ingenieros se reportan con sus respectivos departamentos legales, quienes entonces conversan por teléfono para arreglar los aspectos legales. Por último, los dos presidentes corporativos discuten la cuestión financiera del trato. ¿Qué principio de un protocolo multicapas viola este mecanismo de comunicación, en el sentido del modelo OSI?
12. Cada una de dos redes proporciona un servicio confiable orientado a la conexión. Una de ellas ofrece un flujo de bytes confiable y la otra un flujo de mensajes confiable. ¿Son las dos redes idénticas? De ser así, ¿por qué se hace la distinción? Si no es así, mencione un ejemplo de cómo difieren.
13. ¿Qué significa "negociación" al hablar sobre protocolos de red? Cite un ejemplo.
14. En la figura 1-19 se muestra un servicio. ¿Hay algún otro servicio implícito en esta figura? Si es así, ¿en dónde está? En caso contrario, ¿por qué no?
15. En algunas redes, la capa de enlace de datos se encarga los errores de transmisión pidiendo que se retransmitan las tramas dañadas. Si la probabilidad de que una trama se dañe es p, ¿cuál es la cantidad promedio de transmisiones requeridas para enviar una trama? Suponga que las confirmaciones de recepción nunca se pierden.

16. 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?

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

18. La subred de la figura 1-25(b) se diseñó para soportar una guerra nuclear. ¿Cuántas bombas se requerirían para particionar los nodos en dos conjuntos desconectados? Suponga que una bomba destruye a un nodo junto con todos los enlaces conectados a él.

19. Internet duplica su tamaño aproximadamente cada 18 meses. Aunque en realidad nadie lo sabe con certeza, alguien estimó que el número de hosts que incluía era de 600 millones en 2009. Use estos datos para calcular el número esperado de hosts de Internet para 2018. ¿Cree usted esto? Explique por qué si o por qué no.

20. Al transferir un archivo entre dos computadoras, hay dos estrategias de confirmación de recepción posibles. En la primera, el archivo se divide en paquetes y el receptor envía una confirmación de recepción por cada paquete individual, pero no envía una confirmación de recepción para la transferencia del archivo como un todo. En la segunda no se envía una confirmación de recepción para cada paquete individual, sino que se envía una confirmación de recepción de todo el archivo completo cuando llega. Comente sobre las dos estrategias.

21. Los operadores de redes de telefonía móvil necesitan saber en dónde se encuentran los teléfonos móviles (y sus usuarios). Explique por qué esto es malo para los usuarios. Ahora mencione las razones por las que esto es bueno para los usuarios.

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

23. Una imagen tiene $1\,600 \times 1\,200$ píxeles con 3 bytes/píxel. Suponga que no está comprimida. ¿Cuánto tiempo tarda en transmitirse a través de un canal de modem de 56 kbps? ¿A través de un módem de cable de 1 Mbps? ¿A través de una red Ethernet de 10 Mbps? ¿A través de una red Ethernet de 100 Mbps? ¿A través de una red Gigabit Ethernet de 1 gbps?

24. Ethernet y las redes inalámbricas tienen ciertas similitudes y diferencias. Una propiedad de Ethernet es que sólo se puede transmitir una trama a la vez. ¿Comparte la red 802.11 esta propiedad con Ethernet? Explique su respuesta.

25. Mencione dos ventajas y dos desventajas de tener estándares internacionales para los protocolos de red.

26. Cuando un sistema tiene una parte permanente y una removible (como una unidad de CD-ROM y el CD-ROM), es importante que el sistema esté estandarizado de manera que distintas empresas puedan fabricar tanto las partes permanentes como las removibles y que todo pueda funcionar en conjunto. Cite tres ejemplos fuera de la industria de las computadoras en donde existan dichos estándares internacionales. Ahora cite tres áreas fuera de la industria de las computadoras en donde no existan.

27. Suponga que se cambian los algoritmos utilizados para implementar las operaciones en la capa k. ¿Cómo puede afectar esto a las operaciones en las capas $k - 1$ y $k + 1$?

28. Suponga que hay un cambio en el servicio (conjunto de operaciones) ofrecido por la capa k. ¿Cómo afecta esto a los servicios en las capas $k - 1$ y $k + 1$?

29. Proporcione una lista de razones por las que el tiempo de respuesta de un cliente puede ser mayor que el retardo en el mejor de los casos.

30. ¿Cuáles son las desventajas de usar celdas pequeñas de longitud fija en ATM?

31. Haga una lista de actividades que realiza a diario en donde se utilicen redes de computadoras. ¿Cómo se alteraría su vida si de repente se apagaran estas redes?

32. Averigüe qué redes se utilizan en su escuela o lugar de trabajo. Describa los tipos de redes, las topologías y los métodos de conmutación utilizados.

33. El programa *ping* le permite enviar un paquete de prueba a una ubicación dada para ver cuánto tarda en llegar hasta allá y regresar. Pruebe a usar *ping* para ver cuánto tiempo se requiere para ir de su ubicación hasta varias ubicaciones conocidas. Con base en estos datos, trace el tiempo de tránsito de una sola dirección a través de

Internet en función de la distancia. Lo más adecuado es utilizar universidades, ya que la ubicación de sus servidores se conoce con mucha precisión. Por ejemplo, *berkeley.edu* está en Berkeley, California; *mit.edu* está en Cambridge, Massachusetts; *vu.nl* está en Amsterdam, Holanda; *www.usyd.edu.au* está en Sydney, Australia; y *www.uct.ac.za* está en Cape Town, Sudáfrica.

34. Vaya al sitio web del IETF, *www.ietf.org*, para ver lo que están haciendo. Elija el proyecto que desee y escriba un informe de media página sobre el problema y la solución propuesta.

35. Internet está compuesta de una gran cantidad de redes. Su arreglo determina la topología de Internet. Hay una cantidad considerable de información en línea sobre la topología de Internet. Use un motor de búsqueda para averiguar más sobre la topología de Internet y escriba un breve informe con una síntesis de sus hallazgos.

36. Busque en Internet algunos de los puntos de interconexión importantes que se utilizan para encaminar paquetes en Internet en la actualidad.

37. Escriba un programa que implemente el flujo de mensajes desde la capa más alta hasta la capa más baja del modelo de protocolos de siete capas. Su programa debe incluir una función de protocolo separada para cada capa. Los encabezados de los protocolos son secuencias de hasta 64 caracteres. Cada función de protocolo tiene dos parámetros: un mensaje que se pasa del protocolo de la capa superior (un búfer de caracteres) y el tamaño del mensaje. Esta función añade su encabezado al frente del mensaje, imprime el nuevo mensaje en la salida estándar y después invoca a la función de protocolo del protocolo de la capa inferior. La entrada del programa es un mensaje de aplicación (una secuencia de 80 caracteres o menos).