

# Tema 1: Introducción a las redes de ordenadores y protocolos de comunicación

## 1.1 Introducción.

Durante las primeras décadas de existencia, los sistemas de computación eran centralizados. Generalmente existía una computadora u ordenador en un despacho al cual los usuarios llevaban sus trabajos para procesar. Con posterioridad y especialmente en la década de los 90, el modelo de un ordenador atendiendo las necesidades de una organización ha sido reemplazado por un modelo en el cual gran número de ordenadores separados, pero conectados entre sí, hacen el trabajo. Estos sistemas se conocen con el nombre de redes de ordenadores.

Definiremos como *red de ordenadores* a una colección interconectada de ordenadores autónomos de forma que son capaces de intercambiar información entre sí. Entendemos como *ordenador autónomo* al ordenador que no depende de otro para su funcionamiento. Por ejemplo, las terminales remotas conectadas a un gran ordenador no pueden ser consideradas ordenadores autónomos. El medio de conexión de los ordenadores puede ser muy variado, desde un simple cable de cobre, un enlace de fibra óptica, hasta una comunicación mediante microondas, etc.

Conviene explicar en este punto la diferencia existente entre una red de ordenadores y un *sistemas distribuidos*. La principal diferencia entre ambos es que en un sistema distribuido el sistema operativo se encarga de realizar la asignación de trabajos a procesadores, de archivos a discos, etc., de forma transparente al usuario, mientras que en una red de ordenadores es el propio usuario el que elige el procesador (ordenador) sobre el que ejecuta el programa, etc.

## 1.2 Tipos de redes.

Las redes de ordenadores pueden clasificarse según un gran número de aspectos. Sin embargo y de forma general, suele emplearse para su clasificación dos aspectos: la tecnología de transmisión y la escala (dimensión o tamaño) de la red. De acuerdo a la tecnología de transmisión las redes pueden clasificarse en:

1. Redes de difusión (broadcast en inglés).
2. Redes punto a punto.

Y de acuerdo a su escala pueden clasificarse en:

1. Redes de área local (LAN, Local Area Network).
2. Redes de área metropolitana (MAN, Metropolitan Area Network).
3. Redes de área amplia (WAN, Wide Area Network).

Veamos de forma más detenida estas redes.

### 1.2.1. Redes de difusión.

Las redes de difusión tienen un solo canal de comunicación compartido por todos los ordenadores de la red. Los mensajes cortos (llamados *paquetes*) que envía un ordenador son recibidos por todos los demás. La forma de diferenciar a que ordenador se dirige es un trozo de mensaje (*campo*) dentro del paquete que especifica la dirección del ordenador destino del mensaje, si el mensaje va dirigido a otro ordenador sencillamente se ignora. Un ejemplo para su comprensión es un aeropuerto, en el cual la salida de un vuelo se anuncia por los altavoces. Mucha gente oye el mensaje, pero solo los que van a tomar ese vuelo lo procesan y acuden a la sala de embarque.

En una red de difusión, generalmente existe la posibilidad de dirigir un paquete a todos los destinos, colocando para ello un código especial en el campo de dirección. Esta transmisión es recibida y procesada por todos los ordenadores de la red. Eso se conoce como *difusión* (broadcast en inglés).

En una red de difusión la velocidad de transmisión indica la capacidad agregada (sumada) de todas las máquinas conectadas a la red; por ejemplo, la red Ethernet tiene una velocidad de 10 Mbps, lo cual significa que la cantidad máxima de datos de todos los equipos conectados no puede superar este valor. Conviene resaltar aquí que en telemática los prefijos Kilo, Mega, etc., se utilizan con su significado métrico ( $10^3$ ,  $10^6$ , etc.), no con el significado informático ( $2^{10}$ ,  $2^{20}$ , etc.). Así, 1 Kbps significa 1.000 bits por segundo, no 1.024 bits por segundo. Sin embargo, cuando se trata de cantidad de información sin dividir por el tiempo se sigue empleando el concepto informático, así, si el tamaño máximo de un paquete es de 1 Kbytes queremos decir 1.024 bytes y no 1.000 bytes.

Las redes de difusión más conocidas son ethernet (en sus diversos tipos), token ring, FDDI, etc.

### 1.2.2 Redes punto a punto.

Las redes punto a punto se construyen mediante conexiones entre pares de ordenadores, también llamadas líneas, enlaces, circuitos o canales. Una vez un paquete es depositado en la línea, el destino es conocido de forma unívoca y no es preciso en principio que lleve la dirección del destino. Los enlaces que constituyen una red punto a punto pueden ser de tres tipos de acuerdo con el sentido de la transmisión:

1. Simplex: La transmisión sólo puede efectuarse en un sentido.
2. Semi-dúplex o half-dúplex: La transmisión puede hacerse en ambos sentidos, pero no de forma simultánea.
3. Dúplex o full-dúplex: La transmisión puede efectuarse a la vez en ambos sentidos.

En los enlaces semi-dúplex y dúplex, si la velocidad de conexión es la misma en ambos sentidos, se dice que el enlace es simétrico, es el caso de un enlace mediante un módem; en caso contrario se dice que es asimétrico, como por ejemplo una línea ADSL.

La gran mayoría de los enlaces en líneas punto a punto son dúplex. Así, cuando se habla de un enlace de 64 Kbps sin especificar más se quiere decir 64 Kbps en cada sentido, por lo que la capacidad total del enlace es de 128 Kbps.

En contraste a una red de difusión, en la que el camino a seguir de una máquina a otra es único, no existen ordenadores intermedios y el grado de ocupación es el mismo para todos los ordenadores; al unir múltiples máquinas con líneas punto a punto es posible llegar a formar redes de topologías complejas en las que no sea trivial averiguar cual es la ruta óptima a seguir para ir de un punto a otro, ya que puede haber múltiples caminos posibles con distinto número de ordenadores intermedios, con enlaces con diversas velocidades y distintos grados de ocupación.

Cada ordenador que participa en una red de enlaces punto a punto es un *nodo* de la red. Si el nodo tiene un único enlace se dice que es un *nodo terminal*, de lo contrario se dice que es un *nodo intermedio* o *nodo de encaminamiento*. Cada nodo intermedio ha de tomar la decisión sobre por donde debe dirigir los paquetes que reciba, por lo que también se les llama *conmutadores* (*routers* en inglés). Algunos ejemplos de topología de red punto a punto pueden verse en la figura siguiente:

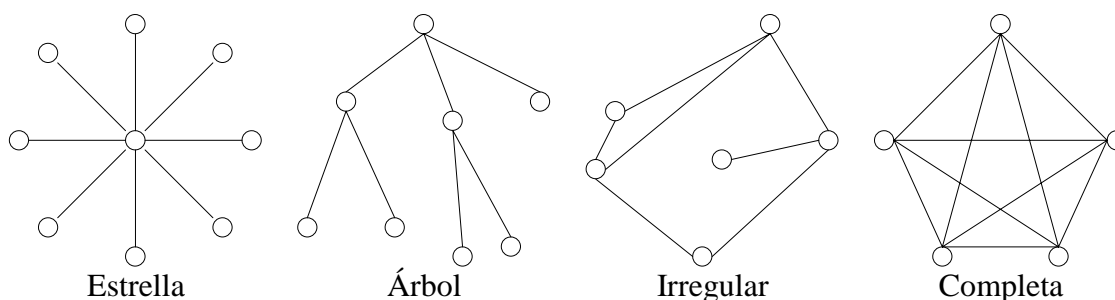


Figura 1.2.2.1: Algunos ejemplos de topología posibles en redes punto a punto.

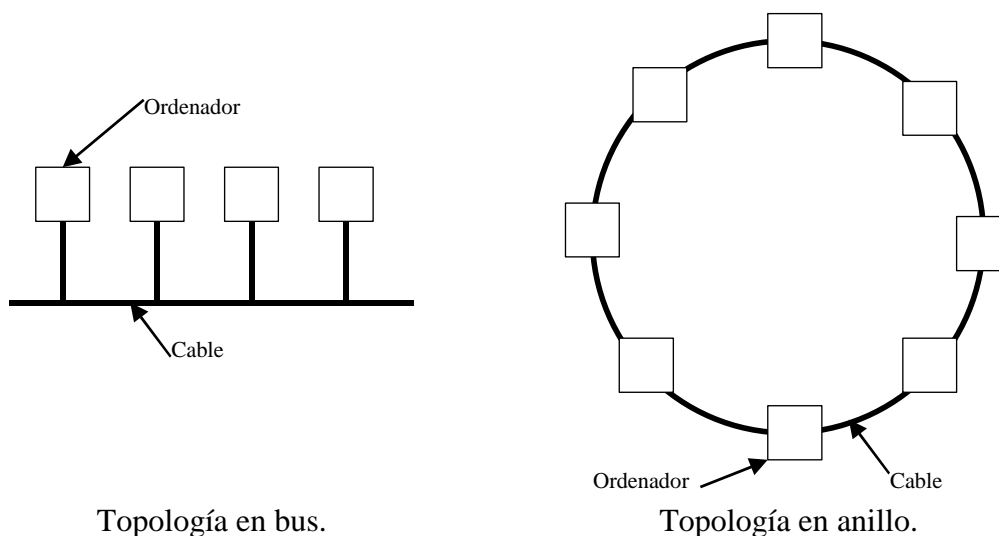
### 1.2.3 Redes de área local.

Las redes de área local (conocidas como Local Area Network) son redes de propiedad privada dentro de un edificio o un campus de hasta unos cuantos kilómetros de extensión. Se usan ampliamente para conectar ordenadores personales y estaciones de trabajo en oficinas y fábricas con el objeto de compartir recursos (impresoras, escáneres, etc.).

Las redes de área local están restringidas en tamaño, lo cual significa que el tiempo de transmisión del peor caso está limitado y se conoce de antemano, lo cual hace posible usar ciertos diseños que de otra manera no serían prácticos y además se simplifica la administración de la red.

Las redes de área local usan una tecnología de transmisión que consiste en un sencillo cable al cual están conectadas todas las máquinas. Operan a velocidades de 10 a 100 Mbps, el tiempo que tardan en transmitir un mensaje es pequeño, por lo que se dice

que tienen bajo retardo (décimas de microsegundos) y experimentan muy pocos errores. Las topologías más usadas en una red de área local son el bus y el anillo (figura 1.2.3.1).



Topología en bus.

Topología en anillo.

*Figura 1.2.3.1: Ejemplos de topologías en redes de área local.*

En una red de bus, en cualquier instante un ordenador es la máquina maestra y puede transmitir, solicitando a las otras máquinas que se abstengan de enviar mensajes mientras dura su transmisión. Es necesario un mecanismo de arbitraje para resolver conflictos cuando dos o más máquinas quieren transmitir simultáneamente. El mecanismo de arbitraje puede ser centralizado (una sola máquina controla el mecanismo) o distribuido (todas las máquinas controlan el mecanismo de arbitraje). Por ejemplo, Ethernet es una red de transmisión basada en bus con control de operación distribuido a 10 o 100 Mbps. Las computadoras de una Ethernet pueden transmitir cuando quieran; si dos o más paquetes chocan (dos computadoras desean transmitir simultáneamente), cada computadora sólo espera un tiempo al azar y lo vuelve a enviar.

En una red en anillo, cada bit se propaga por sí mismo, sin esperar al resto del paquete al cual pertenece. Típicamente, cada bit recorre el anillo entero en el tiempo que toma transmitir unos pocos bits, a veces antes de que el paquete completo se haya transmitido. La regla para arbitrar el acceso simultáneo al anillo más popular es el token ring de IBM, que opera a 4 o 16 Mbps.

#### 1.2.4 Redes de área metropolitana.

Una red de área metropolitana (conocida como Metropolitan Area Network) es básicamente una versión más grande de una red de área local y normalmente se basa en una tecnología similar. Puede abarcar un grupo de oficinas corporativas cercanas o una ciudad y podría ser privada o pública.

La principal razón de distinguir las redes de área metropolitana de las redes de área local es que se ha adoptado un estándar para ellas: el bus dual de cola distribuida (Distributed Queue Dual Bus). El DQDB consiste en dos buses (cables) unidireccionales, a los cuales están conectados todos los ordenadores (figura 1.2.4.1). Cada bus tiene una cabeza terminal (head-end), un dispositivo que inicia la actividad de

transmisión. El tráfico destinado a una computadora situada a la derecha del emisor usa el bus superior. El tráfico hacia la izquierda usa el de abajo.

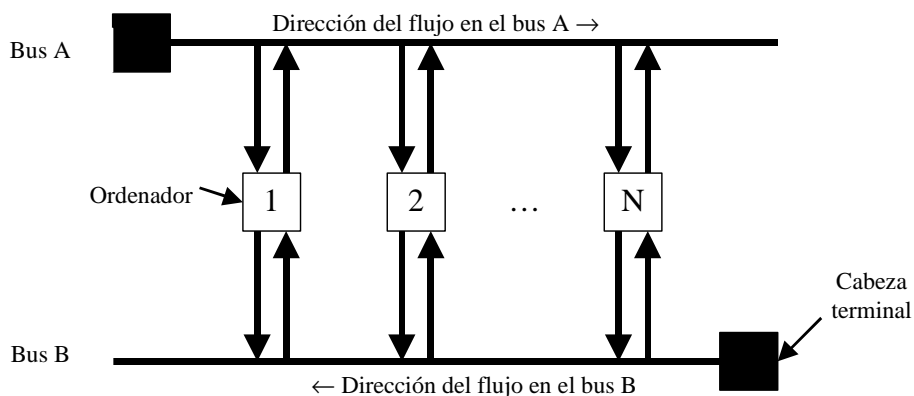


Figura 1.2.4.1: Arquitectura de la red de área metropolitana DQDB.

El aspecto clave de las redes de área metropolitana es que existe un medio de difusión (dos cables en el caso de DQDB), al cual se conectan todos los ordenadores, simplificando el diseño en comparación con otros tipos de redes.

### 1.2.5 Redes de área amplia.

Una red de área amplia (Wide Area Network) se extiende sobre un área geográfica extensa, por ejemplo un país o continente y contiene ordenadores dedicados a ejecutar programas de usuario (aplicaciones). Llamaremos a estas máquinas *hosts*. Los hosts están conectados por una subred de comunicación. El trabajo de esta subred es conducir mensajes de un hosts a otro.

En muchas redes de área amplia, la subred tiene dos componentes distintos: las líneas de transmisión y los elementos de conmutación. Las líneas de transmisión mueven bits de un ordenador a otro. Los elementos de conmutación son ordenadores especializados que conectan dos o más líneas de transmisión. Cuando los datos llegan por una línea de entrada, el elemento de conmutación debe escoger una línea de salida para reenviarlos. Este elemento de conmutación recibe el nombre de *router*.

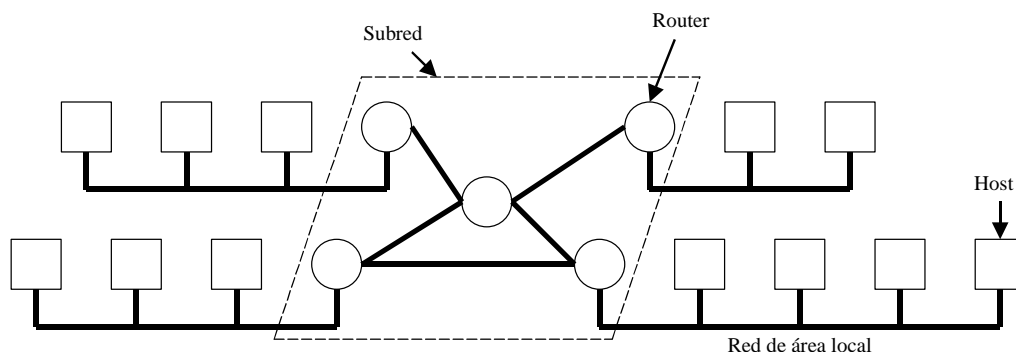


Figura 1.2.5.1: Red de área amplia.

Un ejemplo de red de área amplia puede verse en la figura 1.2.5.1, en ella cada host está conectado a una red de área local en la cual está presente un router, aunque

puede suceder en algunos casos que un host este conectado directamente a un router. La colección de líneas de comunicación y routers (pero no los hosts) forman la subred.

En casi todas las redes de área amplia, la red contiene numerosos cables o líneas telefónicas, cada una conectada a un par de routers. Si dos routers que no comporten un cable desean comunicarse, deberán hacerlo indirectamente, por medio de otros routers. Cuando se envía un paquete de un router a otro a través de uno o más routers intermedios, el paquete se recibe completo en cada router intermedio, se almacena hasta que la línea de salida requerida está libre y a continuación se envía. Una subred basada en este principio se llama de *paquete conmutado*.

### 1.2.6. Interredes.

En el mundo existen muchas redes, a veces con diferente hardware y software. Si deseamos conectar dos redes diferentes frecuentemente incompatibles entre si, es necesario usar máquinas llamadas *pasarelas* para hacer la conexión y la traducción necesaria, tanto hardware como software. Una colección de redes interconectadas se llama interred.

Una forma común de interred es una colección de redes de área local conectadas a una red de área amplia. En efecto, si en la figura 1.2.5.1 reemplazamos la palabra subred por la palabra red de área amplia, nada más en la figura debería cambiar. En este caso, la única distinción entre una subred y una red de área amplia es si están o no presentes los hosts. Si el sistema dentro de la curva cerrada contiene únicamente routers, es una subred, si contiene routers y hosts, es una red de área amplia.

## 1.3 Protocolos de comunicación.

Las primeras redes de computadoras se diseñaron con el hardware como preocupación principal y el software como algo secundario. Sin embargo, con el tiempo, esta estrategia ha sufrido cambios y ahora el software de la red sufre un elaborado proceso de desarrollo, dando lugar a software altamente estructurado.

Para reducir la complejidad del diseño, la mayoría del software de red está organizado como una serie de *capas o niveles*, cada una construida sobre la inferior. El número de capas y el nombre, el contenido y la función de cada una difieren del tipo de red. Sin embargo, en todas las redes el propósito de cada capa es ofrecer ciertos servicios a las capas superiores de modo que no tengan que ocuparse del detalle de la implementación real de los servicios.

La capa N de una máquina lleva a cabo una comunicación con la capa N de otra máquina. Las reglas y convenciones que se siguen en la comunicación se conocen como *protocolo de la capa N*. Básicamente el protocolo es un acuerdo entre las partes que se comunican sobre cómo va a llevarse a cabo la comunicación.

En la figura 1.3.1 puede verse una red de cinco capas. Las entidades que comprenden las capas correspondientes en las diferentes máquinas se denominan *pares*, por lo cual, son los pares los que se comunican usando el protocolo.

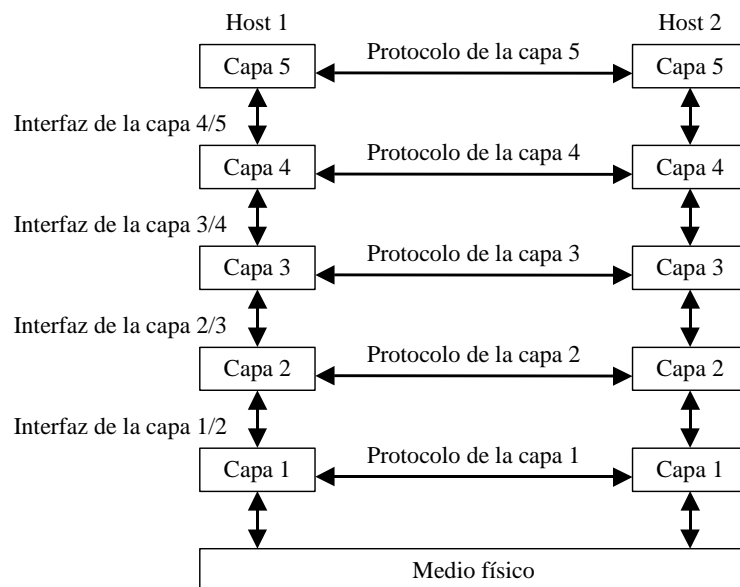


Figura 1.3.1: Capas, protocolos e interfaces de una red.

Los datos no se transfieren directamente de la capa N de una máquina a la capa N de la otra, sino que cada capa pasa datos e información de control a la capa que está inmediatamente debajo de ella, hasta llegar a la capa más baja. Bajo la capa 1 se encuentra el *medio físico* a través del cual ocurre la comunicación. Entre cada par de capas adyacentes se encuentra una *interfaz*. La interfaz define que operaciones y servicios ofrece la capa inferior a la superior. Un conjunto de capas y protocolos recibe el nombre de *arquitectura de red*.

Existen muchas arquitecturas de red implementadas como capas. Veremos a continuación las dos arquitecturas de red más importantes: el modelo de referencia OSI y el modelo de referencia TCP/IP.

#### 1.4 El modelo de referencia OSI.

El modelo de referencia OSI se muestra en la figura 1.4.1. Este modelo se basa en una propuesta que desarrolló la Organización Internacional de Normas (ISO) como paso hacia la estandarización de los protocolos que se usan en las diversas capas. El modelo se llama *modelo de referencia OSI* (Open Systems Interconnection, interconexión de sistemas abiertos) puesto que se ocupa de la conexión de sistemas que están abiertos para la comunicación con otros sistemas. El modelo OSI tiene siete capas, cuyo número se obtuvo de la aplicación de los siguiente principios en el diseño del protocolo:

1. Se debe crear una capa siempre que se necesite un nivel diferente de abstracción.
2. Cada capa debe realizar una función bien definida.
3. La función de cada capa se debe elegir pensando en la definición de protocolos estandarizados internacionalmente.

4. Los límites de las capas deben elegirse a modo de minimizar el flujo de información a través de las interfaces.
5. La cantidad de capas debe ser suficiente para no tener que agrupar funciones distintas en la misma capa y lo bastante pequeña para que la arquitectura no se vuelva inmanejable.

De la aplicación de estas reglas surgió el modelo de protocolo que puede verse en la figura siguiente:

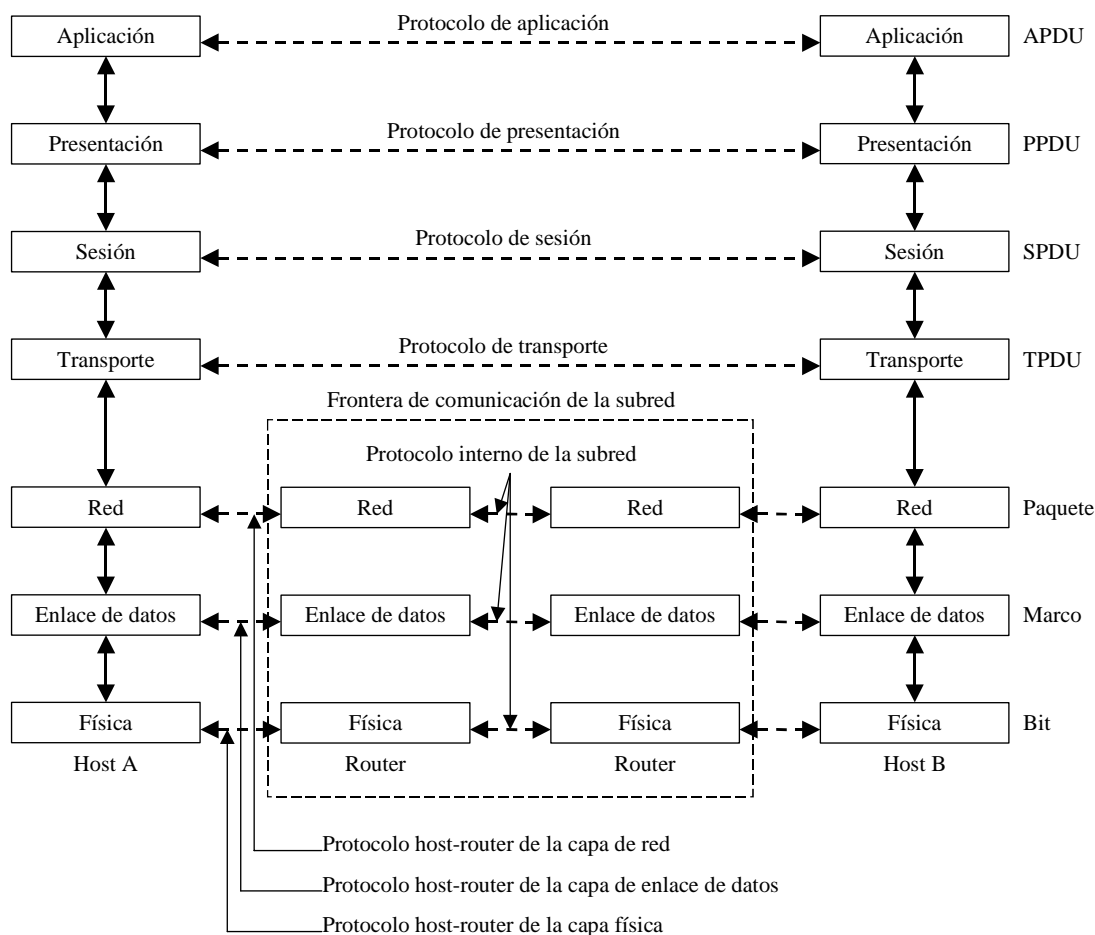


Figura 1.4.1: Modelo de referencia OSI.

Procederemos a continuación a ver, de forma sucinta, cada una de las capas del protocolo OSI.

### 1.4.1 La capa física.

La capa física tiene que ver con la transmisión de bits por un canal de comunicación. Las consideraciones de diseño tiene mucho que ver con las interfaces mecánica, eléctrica, de procedimiento y con el medio de transmisión físico que está bajo la capa física. Estas consideraciones de diseño se pueden resumir en que cuando un lado envíe un bit a 1, el otro lado reciba un bit a 1 e incluyen cuantos voltios deben usarse para representar un 1 y cuantos para un 0; cuantos microsegundos dura un bit; el tipo de transmisión que se puede efectuar (simplex, semi-dúplex o dúplex); el método de



establecimiento de la conexión inicial y como se interrumpe cuando se ha terminado, etc.

#### 1.4.2 La capa de enlace de datos.

La principal tarea de la capa de enlace de datos es tomar un medio de transmisión en bruto y transformarlo en una línea que parezca libre de errores de transmisión no detectados a la capa de red (capa superior). Esta tarea se realiza haciendo que el emisor divida los datos de entrada en *marcos de datos* (unos cientos o miles de bytes), que transmita los marcos de forma secuencial y procese los *marcos de acuse de recibo* que devuelve el receptor.

Para cumplir sus especificaciones, y dado que la capa física solo reconoce una corriente de bits, la capa de enlace de datos debe crear y reconocer los límites de los marcos, añadiendo patrones especiales de bits al principio y al final del marco y teniendo especial cuidado de que estos patrones de bits no puedan interpretarse incorrectamente dentro de los marcos, si aparecen.

Además debe, en caso de la pérdida de un marco por un problema en el medio físico (por ejemplo una ráfaga de ruido), retransmitir el marco. Sin embargo, las transmisiones repetidas de marcos introducen la posibilidad de duplicar marcos, pues se puede enviar un marco duplicado si lo que se pierde no es el marco enviado sino el marco de acuse de recibo que el receptor devuelve al emisor. Corresponde a esta capa resolver el problema de los marcos dañados, perdidos o duplicados.

Por último, la capa de enlace de datos debe implementar un mecanismo de regulación de tráfico para que el transmisor sepa cuánto espacio de almacenamiento temporal (buffer) tiene el receptor libre en ese momento y evitar con ello la saturación del receptor y con ello la pérdida de marcos.

#### 1.4.3 La capa de red.

La capa de red se ocupa de controlar el funcionamiento de la subred. Su principal misión es la determinación de mediante que rutas se encaminan los paquetes del ordenador emisor al ordenador de destino. Las rutas pueden ser tablas estáticas, pueden determinarse al principio de cada conversación o bien pueden determinarse de nuevo con cada paquete para reflejar la carga actual de la red. Además, debe encargarse de controlar la congestión en la subred, de forma que no se encuentren en la misma un excesivo número de paquetes, pues esto produce una disminución en la velocidad de transmisión de la red.

Otro problema que debe resolver la capa de red es la interconexión entre distintos tipos de redes, pues el tipo de direcciones en una red pueden ser distintos en otra red; los tamaños de paquete pueden diferir, etc.

Por último, es misión de la capa de red contar los paquetes, caracteres o bits que envía cada cliente, pues esto permite, en caso de ser necesario, facturar a cada cliente

por el uso que realiza de la red, en concreto, contar los paquetes enviados y recibidos para, después de asignarles un precio, facturar al cliente.

#### 1.4.4 La capa de transporte.

La función básica de la capa de transporte es aceptar datos de la capa de sesión, dividirlos en unidades más pequeñas si es necesario, pasarlos a la capa de red y asegurar que todos los pedazos llegan correctamente. De esta forma, se aísla a las capas superiores de los cambios en la tecnología hardware.

La capa de transporte funciona, de forma general, creando una conexión de red distinta para cada conexión de transporte que requiera la capa de sesión. Sin embargo, si es costoso crear una conexión de red, puede multiplexar varias conexiones de transporte en la misma conexión de red para reducir el costo. En cualquiera de ambos casos, es misión de la capa de transporte que resulte transparente para la capa de sesión. Una vez terminado el uso de una conexión de red, la capa de transporte debe liberar la conexión a través de la red.

La capa de transporte es una verdadera capa de extremo a extremo, del origen al destino. Un programa en el ordenador fuente sostiene una comunicación con un programa en el ordenador destino, mientras que en las capas más bajas los protocolos se usan entre cada máquina y sus vecinas inmediatas, y no entre las máquinas origen y destino.

Por último, la capa de transporte debe regular el flujo de información, a fin de que un nodo rápido no sature a un nodo lento. Este mecanismo se conoce con el nombre de *control de flujo* y desempeña un papel clave en la capa de transporte.

#### 1.4.5 La capa de sesión.

La capa de sesión permite a los usuarios de máquinas diferentes establecer sesiones entre ellos. Una sesión permite el transporte de datos, encargado a la capa de transporte, pero también proporciona servicios que pueden requerirse como el control del diálogo. Además, si la conexión es simplex, la capa de sesión debe manejar el control de los turnos.

Un problema importante cuya solución depende de la capa de sesión es la sincronización entre procesos de las computadoras conectadas. Esta sincronización permite, por ejemplo, que si se desea enviar un archivo entre dos computadoras cuya transferencia dura dos horas, pero la conexión sufre problemas de desconexiones cada hora, la transferencia es imposible. Sin embargo, la capa de sesión puede insertar puntos de sincronización en la corriente de datos, con lo cual, después de cada interrupción sólo se deben enviar los datos que se encuentran después del último punto de sincronización transferido.

#### 1.4.6 La capa de presentación.

La capa de presentación se ocupa de la sintaxis y la semántica de la información que se transmite. Se encarga, por ejemplo, de convertir la representación de las cadenas de caracteres (por ejemplo ASCII, Unicode, etc.), los enteros (big-endian o little-endian), etc., a formatos que hagan posible su comunicación entre distintos tipos de ordenadores. Esto se hace definiendo un formato estándar de red, de forma que los datos son convertidos del formato interno del ordenador al formato de red en la máquina fuente y viceversa en la máquina destino.

#### 1.4.7 La capa de aplicación.

La capa de aplicación contiene los protocolos de aplicación, esto es, aquellas aplicaciones que se utilizan por la red. Su misión es proporcionar dichas aplicaciones y solventar los problemas que puedan surgir por la disparidad de formato existentes entre los computadores conectados a la red. Así, por ejemplo, existen miles de tipos de terminales incompatibles entre sí. La capa de aplicación soluciona el problema definiendo una *terminal virtual de red* abstracta, que los editores y programas pueden manejar. Después, la capa de aplicación debe traducir las ordenes dadas a la terminal virtual de red en las ordenes propias de la terminal real que se este manejando. Otro ejemplo es la transferencia de archivos, pues estos se nombran de forma diferente en cada sistema operativo, tienen distinta forma de representar las líneas de texto, etc. La capa de aplicación debe encargarse de proveer una forma general de solucionar estos problemas.

### **1.5 El modelo de referencia TCP/IP.**

El modelo de referencia TCP/IP es una arquitectura cuyo principal objetivo de diseño es poder conectar entre sí múltiples redes de manera transparente para el usuario. El modelo TCP/IP surgió por el interés del Departamento de Defensa de Estados Unidos en poseer una red con capacidad de mantener las conexiones intactas mientras las máquinas de origen y destino estuvieran funcionando independientemente de si alguna de las máquinas o líneas de transmisión en el trayecto dejara de funcionar de forma repentina.

El modelo de referencia TCP/IP esta formado por cuatro capas, que pueden verse en la figura siguiente (figura 1.5.1), en la cual también se muestra la relación de cada capa con la capa equivalente del modelo de referencia OSI.

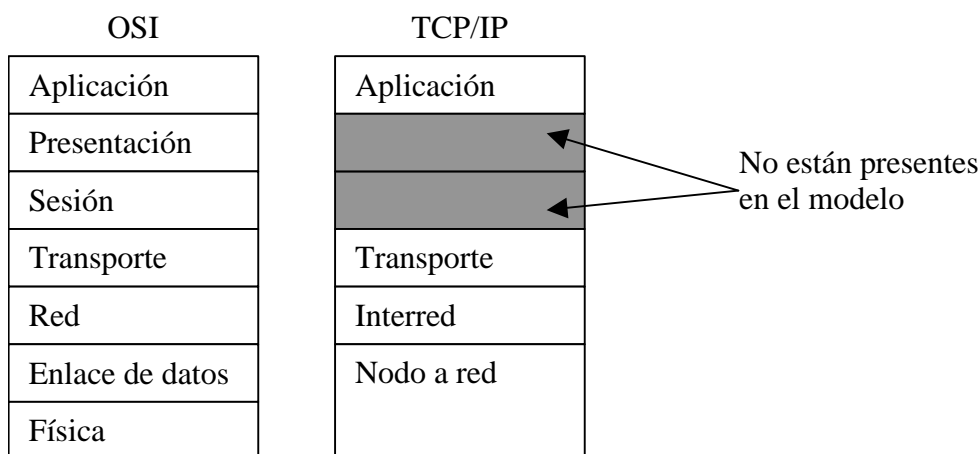


Figura 1.5.1: Modelo de referencia TCP/IP y su relación con el modelo OSI.

De igual forma a como hicimos con las capas del protocolo OSI, expondremos a continuación cada una de las capas del protocolo TCP/IP.

### 1.5.1 La capa del nodo a la red.

El modelo de referencia TCP/IP es muy sucinto en la especificación de esta capa del protocolo. Básicamente, solo indica que el nodo se ha de conectar a la red haciendo uso de algún protocolo que permita el envío de paquetes IP. Este protocolo no está definido y varía de un nodo a otro y de una red a otra red.

### 1.5.2 La capa interred.

La capa interred es el eje que mantiene unida toda la arquitectura. La misión de esta capa es permitir que los nodos inyecten paquetes en cualquier red y los hagan viajar de forma independiente a su destino, que incluso puede estar en una red diferente. Los paquetes enviados por esta capa pueden llegar en un orden diferente a aquel en que se enviaron, siendo misión de las capas superiores ordenarlos, si lo que se desea es una entrega ordenada. Además, la capa interred debe controlar la congestión de la red, evitando el envío de paquetes en caso de congestión en la misma.

Para ello, la capa de interred define un formato de paquete y protocolo oficial llamado IP (Internet Protocol, protocolo de interred). Estudiaremos dicho protocolo de forma detenida en un tema posterior.

### 1.5.3 La capa de transporte.

La capa que está sobre la capa de interred se conoce como capa de transporte. Esta capa se diseñó para permitir que las entidades pares en los nodos origen y destino lleven a cabo una comunicación, cumpliendo el mismo papel que la capa de transporte OSI. En la capa de transporte de TCP/IP se definieron dos protocolos de extremo a extremo, TCP y UDP.

TCP (Transmission Control Protocol, protocolo de control de transmisión) es un protocolo confiable orientado a la conexión que permite que una corriente de bytes originada en una máquina se entregue sin errores en cualquier otra máquina de la red. Este protocolo fragmenta la corriente entrante de bytes en mensajes discretos y pasa cada uno a la capa de interred. En el destino, el proceso TCP receptor reensambla los mensajes recibidos para formar la corriente de salida. El TCP también se encarga del control del flujo para asegurar que un emisor rápido no pueda colapsar a un receptor lento con más mensajes de los que puede manejar.

UDP (User Data Protocol, protocolo de datos de usuario) es un protocolo sin conexión, no confiable, para aplicaciones que no necesitan la asignación de secuencia ni el control de flujo que proporciona TCP, pues o bien no lo desean, o bien desean utilizar los suyos propios. Este protocolo se usa ampliamente para consultas de petición y respuesta de una sola ocasión, del tipo cliente-servidor, y en aplicaciones en las que la entrega rápida es más importante que la entrega precisa del mensaje, tal y como sucede en la transmisión de audio y vídeo.

#### 1.5.4 La capa de aplicación.

Encima de la capa de transporte está directamente la capa de aplicación, pues el modelo TCP/IP no tiene capas de sesión ni de presentación. No se pensó que fueran necesarias, por lo que no se incluyeron. La capa de transporte contiene todos los protocolos de alta nivel. Algunos ejemplos de estos protocolos son el protocolo de terminal virtual (TELNET), el de transferencia de archivos (FTP), el de correo electrónico (SMTP), el de servicio de nombres de dominio (DNS), el protocolo que se usa para recuperar páginas de la World Wide Web (HTTP), etc.

### **1.6 Transmisión de datos en un protocolo de comunicación.**

Veamos ahora como funciona un protocolo de capas. Para ello tomemos como ejemplo el protocolo de cinco capas mostrado en la figura 1.3.1.

Un proceso de aplicación, generalmente una aplicación de usuario, que se ejecuta en la capa 5 desea comunicarse con otro ordenador. Para ello, el proceso produce un mensaje M y se lo entrega a la capa 4 para su transmisión.

La capa 4 coloca un *encabezado* al principio del mensaje para identificarlo y pasa el resultado a la capa 3. El encabezado incluye información de control, como números de secuencia, para que la capa 4 en la máquina de destino pueda entregar los mensajes en el orden correcto si las capas inferiores no mantienen la secuencia.

Generalmente, la capa 3 impone un límite al tamaño de los mensajes que puede procesar, por lo que divide los mensajes que le llegan en unidades más pequeñas, llamadas *paquetes*, anexando un encabezado de la capa 3 a cada paquete. Además decide que línea de salida (si existen varias) será usada y pasa los paquetes a la capa 2.

La capa 2 añade un encabezado y un apéndice al paquete y lo entrega a la capa 1, la cual se encarga de transmitirlo de forma física.

En la máquina receptora el mensaje se mueve hacia arriba de capa a capa, perdiendo los encabezados y apéndices conforme avanza. Ninguno de los encabezados y apéndices para las capas inferiores a la N llega hasta la capa N. En la figura 1.6.1 puede verse de forma esquemática

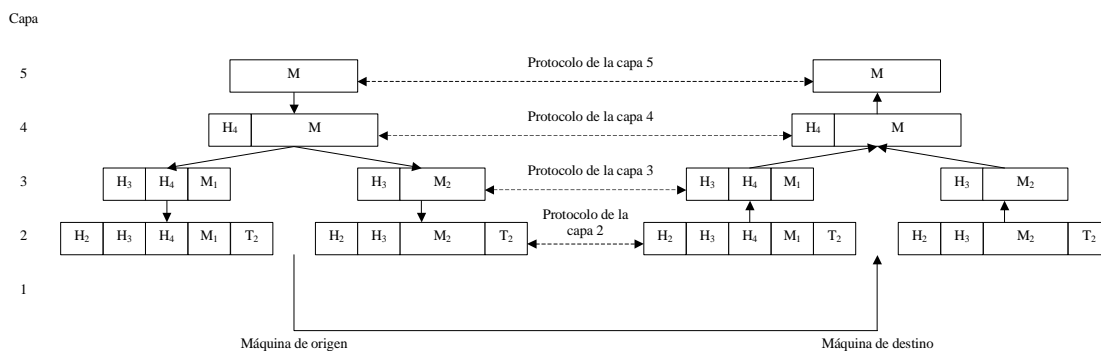


Figura 1.6.1: Ejemplo del flujo de información en un protocolo.

### 1.7 Breve historia de Internet.

No parece adecuado terminar un tema de introducción a las redes de ordenadores sin comentar algo de la corta, pero ajetreada vida de la red que hoy conocemos como Internet.

A mediados de la década de los 60, el Departamento de Defensa de los Estados Unidos quería una red de control que pudiera sobrevivir a una guerra nuclear. Las redes telefónicas tradicionales se consideraban muy vulnerables, pues la pérdida de una línea o un conmutador terminaría toda comunicación que los estuviera usando. Para resolver este problema, el Departamento de Defensa acudió a ARPA (Advanced Research Projects Agency).

ARPA decidió que el Departamento de Defensa necesitaba una red de paquete conmutado, que consistía en una subred y computadoras hosts. La subred consistía en minicomputadoras llamadas IMP (procesadores interfaz de mensajes) conectadas con líneas de transmisión. Para lograr alta fiabilidad, cada IMP se conectaría al menos a otras dos. La subred iba a ser una subred de datagrama, de forma que si algunas líneas de IMP resultaban destruidas, los mensajes se podrían reencaminar de forma automática a través de trayectorias alternativas. ARPANET entró en funciones de forma experimental en diciembre de 1969 con cuatro nodos en UCLA, UCSB, SRI y la Universidad de Utah.

Con posterioridad se produjeron cambios en ARPANET, como el cambio del software de IMP para permitir que las terminales se conectaran de forma directa a un IMP especial llamado TIP (procesador de interfaz de terminal) sin tener que pasar por un hosts. Los cambios siguientes incluyeron el tener múltiples hosts por cada IMP (para ahorrar dinero), hosts que se comunicaban con múltiples IMP (para protegerse de fallos de IMP), etc.

Sin embargo, el cambio más radical fue la invención y desarrollo del modelo y protocolo TCP/IP y su adopción. Para fomentar la adopción, ARPA concedió contratos a BBN y a la Universidad de California en Berkeley para integrar TCP/IP con el UNIX de Berkeley. Los investigadores de Berkeley desarrollaron una interfaz de programa conveniente para la red (sockets) y muchos programas de aplicación, administración, etc., para facilitar el trabajo con redes. Esto hizo que en 1983 ARPANET contará con más de 200 IMP y cientos de hosts. En este momento ARPA cedió el manejo de la red a la Agencia de Comunicaciones de la Defensa (DCA). Lo primero que hizo la DCA fue separar la porción militar de la red en una subred independiente llamada MILNET, con pasarelas estrictas entre MILNET y la red de investigación restante.

En 1990 la red ARPANET había sido rebasada por redes más nuevas que ella misma había generado, de manera que se clausuró y desmanteló, pero su versión militar MILNET aún continúa operando.

A finales de 1970, la NSF (Fundación Nacional de la Ciencia) de Estados Unidos observó el enorme impacto que ARPANET tenía en la investigación universitaria. Sin embargo, para conectarse a ARPANET una universidad debía tener contratos de investigación con el Departamento de Defensa. Esta falta de acceso universal motivó a la NSF a establecer una red virtual, CSNET, centrada en una sola máquina y que permitía tener conexión con ARPANET y otras redes. Mediante CSNET los investigadores podían hacer llamadas y dejar correo electrónico para que otras personas lo recogieran.

En 1984, la NSF empezó a diseñar un sucesor de alta velocidad de ARPANET que se abriría a todos los grupos universitarios de investigación. Para ello conectó sus seis centros de supercomputadoras en San Diego, Boulder, Champaign, Pittsburgh, Ithaca y Princeton formando una red de backbone (tronco o columna vertebral), financiando también algunas redes regionales (cerca de 20), que se conectaron a la red de backbone. Por fin NSFNET se conectó a ARPANET mediante un enlace en el cuarto de máquinas de Carnegie-Mellon.

La NSFNET fue un éxito inmediato y se sobrecargó desde el primer momento. Al continuar el crecimiento, la NSF se dio cuenta que el gobierno no podía seguir financiando eternamente el uso de redes. Más aún, las organizaciones comerciales querían unirse pero los estatutos de la NSF les prohibían usar redes pagadas por la propia fundación. Por ello NSF animó a MERIT, MCI e IBM a formar una corporación no lucrativa, ANS (Advanced Network and Services) como paso hacia la comercialización. En 1990, ANS se hizo cargo de NSFNET y formó ANSNET.

La cantidad de redes, máquinas y usuarios conectados a ARPANET creció con rapidez después de que TCP/IP se convirtió en el único protocolo oficial en 1983. Cuando se conectaron NSFNET y ARPANET el crecimiento se hizo exponencial; se unieron muchas redes regionales y se hicieron conexiones con redes de Canadá, Europa y el Pacífico. Hacia finales de la década de los 80 existían 3000 redes y 200.000 computadoras conectadas, y la gente fue viendo la aglomeración de redes como una interred y más tarde empezó a denominarla INTERNET.

En 1992 se adhirió a INTERNET el host número un millón. Y en la actualidad se calcula que el tamaño se duplica cada año aproximadamente. Esto es debido a que

hasta mediados de la década de los 90, Internet se poblaban en gran medida de investigadores académicos, del gobierno y de la industria. Una aplicación nueva, la WWW cambió esto y atrajo a millones de nuevos usuarios no académicos a la red. Esta aplicación, inventada por el físico del CERN Tim Berners-Lee, no cambió ninguno de los recursos subyacentes, pero los hizo más fáciles de usar.