Apuntes

de

Redes de Ordenadores

Tema 1

Introducción

Uploaded by

IngTeleco

http://ingteleco.iespana.es
ingtelecoweb@hotmail.com

La dirección URL puede sufrir modificaciones en el futuro. Si no funciona contacta por email

TEMA 1: INTRODUCCIÓN

1.1.- INTRODUCCIÓN

En tan solo unos años las redes de ordenadores han pasado de ser algo inusual, sólo conocido y utilizado por unos pocos a ocupar un primer plano en cualquier medio informativo de carácter general. Quizá el protagonismo que actualmente se da a términos como “Internet”, “Intranet” o “e-business” sea en parte fruto de una labor de marketing muy bien conducida, pero no cabe duda que dichos términos (o al menos las ideas que representan) tendrán un interés decisivo en los próximos años cambiando radicalmente la actividad del mundo que hoy conocemos.

Podemos hacer un cierto paralelismo entre la explosión de la Telemática en la década de los noventa y el auge de la Informática personal en los ochenta; sin embargo a pesar de su importancia la aparición del PC no parece comparable a la revolución que está protagonizando la Telemática; la razón estriba en que, a pesar de todo, el PC aislado es hasta cierto punto un producto minoritario, mientras que el sistema multimedia de los noventa conectado a las redes se convertirá en una fuente de información y de entretenimiento para el público en general y de negocio para las empresas.

Los precios de la informática vienen sufriendo desde hace bastantes años una disminución exponencial. El precio del espacio en disco (pesetas/Mb) se reduce a la mitad aproximadamente cada 4.5 años, el de la potencia de procesador (pesetas/MIP) cada 2.3 años, y el de la memoria RAM (pesetas/Mb) cada 1.8 años. Como comparación el precio de la transmisión de datos (medido en pesetas/Mbps/Km) se reduce a la mitad cada 1.5 años aproximadamente, es decir, esta teniendo una disminución aun mayor que las tecnologías informáticas. Las investigaciones y desarrollos en materia de transmisión de datos hacen prever que dicha tendencia se mantendrá en el futuro.

Además de los factores tecnológicos en los precios de los servicios telemáticos influyen aspectos legales que en ocasiones alteran la situación de manera importante. Por ejemplo en España, como en otros países de Europa, la decisión de liberalizar las telecomunicaciones en 1998 está produciendo un abaratamiento de los precios gracias a la libre competencia, que de forma transitoria hará aun mayor la reducción que cabría esperar de los factores puramente tecnológicos.

1.1.1.- Telecomunicaciones y Telemática

Comenzaremos por diferenciar estos dos términos fundamentales.

Entendemos por *telecomunicaciones* el conjunto de medios técnicos que permiten la comunicación a distancia. Normalmente se trata de transmitir información sonora (voz, música) o visual (imágenes estáticas o en movimiento) por ondas electromagnéticas a través de diversos medios (aire, vacío, cable de cobre, fibra óptica, etc.). La información se puede transmitir de forma analógica, digital o mixta, pero en cualquier caso las conversiones, si las hay, siempre se realizan de forma transparente al usuario, el cual maneja la información de forma analógica exclusivamente.

El término *telemática* (fusión de *tele*comunicaciones e infor*mática*) trata del uso de las telecomunicaciones para enriquecer las posibilidades de la informática (y no al revés), es decir, del uso de medios de comunicación a distancia para conexiones informáticas (ordenador-ordenador u ordenador-periférico). La información puede transmitirse de forma analógica, digital o mixta, pero esto es transparente al usuario, que la maneja de forma digital únicamente.

Todos los sistemas habituales de telecomunicaciones transmiten la información por medio de ondas electromagnéticas a través de diversos medios: aire, vacío, cable de cobre, fibra óptica, etc.

1.1.2.- Redes de ordenadores y sistemas distribuidos

La expresión *redes de ordenadores* (o simplemente *redes*) se utiliza cuando, por medio de la telemática, se realiza la comunicación entre dos o más ordenadores. Queda excluida aquí la comunicación entre un ordenador y un periférico (terminal, impresora, etc.) independientemente de la distancia a la que dicha comunicación se produzca o el tipo de medios utilizados para ella. Dicho de otro modo, en redes de ordenadores se considera únicamente la comunicación entre elementos que pueden hablar de igual a igual (“peer to peer”), sin tomar en consideración la comunicación asimétrica maestro-esclavo.

Un caso particular de las redes de ordenadores son los *sistemas distribuidos,* en los que se intenta conectar varios ordenadores mediante una red y crear un entorno de utilización tal que el usuario no perciba la existencia de múltiples sistemas, sino que los maneje como un único sistema virtual de forma transparente; para esto se utilizan normalmente protocolos o aplicaciones específicos. Evidentemente si el medio de comunicación es de baja velocidad el usuario percibirá un retraso cuando acceda a un nodo remoto, por lo que generalmente los sistemas distribuidos sólo se implementan en redes de alta velocidad (redes locales por ejemplo). Un ejemplo de protocolo de sistemas distribuidos podría ser el NFS (Network File System) que permite acceso a ficheros remotos de forma transparente.

1.2.- ALGUNOS USOS DE LAS REDES DE ORDENADORES

Podemos diferenciar claramente dos tipos de usos o usuarios de las redes de ordenadores: el profesional, que se da normalmente en la empresa, y el particular, que generalmente tiene lugar en la residencia habitual del usuario.

1.2.1.- Uso de las redes en empresas

Prácticamente cualquier empresa que tenga varios ordenadores hoy en día tiene una red local que los interconecta. Si la empresa dispone de varias sedes u oficinas dispersas dispondrá típicamente de una red local (LAN, Local Area Network) en cada una de ellas y de un medio de interconexión de dichas redes locales a través de enlaces telefónicos (también llamados accesos WAN, Wide Area Network). La red o redes permiten acceder a información importante y actualizada de manera rápida, por ejemplo una base de datos que contenga toda la información comercial de la compañía (productos, stocks, precios, plazos de entrega, etc.). A menudo estas bases de datos están en uno o unos pocos ordenadores de la red, ya que la existencia de múltiples copias complica las actualizaciones.

Antiguamente las aplicaciones se diseñaban para que los usuarios accedieran desde terminales "tontos" al ordenador central en el que se mantenía la base de datos y en el cual se procesaba la transacción en su totalidad, pero la aparición de redes de ordenadores donde el terminal se ha convertido en un PC ha llevado a un nuevo modelo de desarrollo de las aplicaciones llamado *cliente-servidor,* consistente en descargar en el PC (cliente) una parte del proceso de la transacción (por ejemplo toda la labor de validación de los datos introducidos), y dejar para el ordenador central (servidor) únicamente la parte que no es posible hacer en el cliente, como por ejemplo la inclusión del nuevo registro en la base de datos. El modelo cliente-servidor reduce así de forma considerable los recursos necesarios en el ordenador central, y permite aprovechar el PC que el usuario tiene en su mesa (y que muy probablemente tendría de todas formas). Además así la aplicación se integra de forma más amigable en el ordenador del usuario final (mediante el uso de ventanas, interfaces gráficas, ratón, etc.). Así el uso del modelo cliente-servidor, y por tanto de las redes de ordenadores puede llegar a suponer en la práctica un *ahorro* en los gastos informáticos de la empresa, además de una mayor *productividad* de sus empleados.

Por otro lado, la existencia de redes de ordenadores permite a la empresa tener duplicado su servidor de base de datos, o cualquier otra información vital, de forma que en caso de fallo del software, hardware, o destrucción física del servidor la información no se vea afectada al poder los clientes seguir funcionando con el servidor de reserva. Esto se traduce en una mayor *fiabilidad* del sistema, aspecto imprescindible en algunas empresas (por ejemplo bancos, hospitales, cadenas de montaje de fábricas, etc.). Por supuesto para que el sistema en su conjunto sea altamente fiable es preciso duplicar no sólo el servidor de base de datos, sino la propia red (elementos de conmutación, conexión, cables, etc.) de forma que no haya ningún elemento importante susceptible de fallo cuya funcionalidad no este duplicada.

La red en las empresas permite *compartir* recursos, tales como periféricos de elevado costo (impresoras láser, scanners, plotters, filmadoras, etc.), o programas (siempre y cuando la licencia que se posee permita su uso en red) con el consiguiente ahorro de espacio en disco y sencillez de actualización.

Otra utilidad importante de la red en las empresas es como *medio de comunicación* entre sus empleados; el correo electrónico es el servicio básico, pero otros mas avanzados se están implantando, tales como la videoconferencia o las aplicaciones que permiten compartir un documento entre varios usuarios trabajando desde ordenadores distintos. Este tipo de aplicaciones se conoce como CSCW (Computer Supported Cooperative Work) y también como "groupware".

Hasta aquí hemos discutido aplicaciones orientadas fundamentalmente al uso de la red dentro de la propia empresa (lo que actualmente se suele denominar la "Intranet"). Dicha red puede conectarse al exterior, bien directamente o a través de un cortafuego o "firewall", es decir, una pasarela intermedia que permita controlar el acceso (entrante y/o saliente) para evitar problemas de seguridad. Cuando la red de la empresa se conecta al exterior (normalmente a la Internet) aparecen una serie de nuevas aplicaciones que le dan aun mayor utilidad, entre las que cabe destacar las siguientes:

Las actividades de *marketing*; por ejemplo se puede poner el catálogo de productos de la empresa en la red para su consulta por los clientes, con información detallada de características, precio, referencias, etc.; también es posible tramitar pedidos recibidos a través de la red.

Actividades de *soporte en línea*; se puede responder a preguntas de los usuarios a través de la red, tanto por correo electrónico como por listas de distribución o grupos de news. En el caso de empresas de software es frecuente ofrecer a través de la red nuevas versiones de programas, sistemas operativos, parches para la resolución de problemas, etc.

Las herramientas de *comunicación* antes mencionadas (correo electrónico, videoconferencia, CSCW, etc.) adquieren una relevancia mucho mayor cuando su uso no se limita al interior de la empresa.

Algunas empresas encuentran en Internet una manera económica de interconectar sus oficinas remotas, evitando así la contratación de líneas propias de larga distancia.

El empleado puede acceder a una enorme cantidad de *información externa* a su empresa útil para su trabajo, por ejemplo información de suministradores, competidores, clientes, foros de discusión sobre

temas relacionados con su trabajo (especialmente cuando éste es de carácter técnico), etc. Curiosamente esta ventaja conlleva un problema, que es la imposibilidad de evitar que el empleado utilice la conexión al exterior para acceder a información no relacionada con su trabajo (por ejemplo sobre su hobby favorito), perdiendo en ello a veces una parte importante de su jornada laboral. Es prácticamente imposible impedir por medios técnicos que esto suceda, aunque se pueden adoptar algunas medidas protectoras. Este problema ha hecho a algunas empresas cuestionarse la conveniencia de dar acceso Internet a sus empleados.

1.2.2.- Uso de las redes por particulares

El uso de las redes de ordenadores por particulares tiene tres objetivos fundamentales:

- Acceso a información
- Comunicación
- Entretenimiento

El acceso a información actualmente se centra en el acceso a Internet y sobre todo a servidores Web. En torno a esto han aparecido multitud de servicios derivados del uso de la telemática para diversos fines, tales como teletrabajo, telecompra, teleenseñanza, telemedicina, etc.

La comunicación tiene lugar tanto a nivel individual (correo electrónico) como en grupos (listas de distribución, grupos de news, etc.). Esto incluye no solo información textual, sino también multimedia: sonido, imagen y vídeo. Además de estas aplicaciones asíncronas, en las que los participantes no han de coincidir en el tiempo, existen otras (llamadas isócronas) en las que si han de coincidir, como las que permiten utilizar el ordenador como un teléfono, para hablar con un usuario remoto a través de la Internet; esto supone un ahorro importante en algunos casos ya que se puede hacer una llamada a un lugar remoto pagando tarifa local (lo cual ha motivado serias críticas y discusiones con las compañías telefónicas, especialmente en Estados Unidos). También está el servicio de videoconferencia, aunque poco extendido a nivel particular debido a su escasa difusión y a sus requerimientos de capacidad, difíciles de satisfacer con un módem telefónico.

El uso con fines de entretenimiento será la gran aplicación de las redes de ordenadores en el futuro, pero actualmente el reto tecnológico es tan grande que para abordarlo es preciso disponer de potentes y costosos equipos, con lo que la rentabilidad es cuando menos dudosa. Se han hecho ya algunas experiencias de *vídeo bajo demanda* en Estados Unidos, pero las necesidades de red y de servidores para un número elevado de usuarios son tan grandes que los servicios comerciales que actualmente se ofrecen se basan generalmente en el *vídeo casi bajo demanda* (NVOD, Near Vídeo On Demand) donde cada transmisión es vista por un conjunto de usuarios simultáneamente.

1.2.3.- Aspectos sociales

La Internet es noticia casi diaria en los medios de comunicación, y no siempre en sentido positivo. Algunos ejemplos de temas polémicos son los siguientes:

- Distribución de pornografía. En muchos países es ilegal distribuir pornografía a menores, por lo que la disponibilidad de estos materiales en la red limita a veces el acceso a la Internet en colegios. También es polémica la distribución de pornografía infantil a adultos.

- Distribución de información “peligrosa”: por ejemplo se han dado casos de personas que han aprendido a sintetizar drogas a partir de información obtenida en la Internet; o sería posible distribuir información sobre como fabricar explosivos.

- Distribución de publicidad no deseada. Es tremendamente fácil recopilar una enorme lista de direcciones de correo electrónico para distribuir a muy bajo costo cualquier propaganda de tipo comercial, político, religioso, etc. a nivel mundial. Alguna gente recomienda en estos casos utilizar la técnica del “ladrillo a portes pagados”, es decir, devolver al remitente un mensaje con unos cuantos Megabytes de información inútil. Esta acción tomada por un número elevado de usuarios inutiliza el buzón y el servidor desde los que se distribuye la propaganda.

- Discrepancias legales. Es posible que una información distribuida por la red desde un país sea ilegal en otro; por ejemplo, ETA puso un servidor Web en Suiza con información que en España se considera apología del terrorismo. También es posible comprar en el extranjero bienes de consumo sin pagar los impuestos correspondientes a nuestro país.

- Acceso a la Internet desde el puesto de trabajo para fines personales. Este tema, que ya hemos comentado, ha llevado a algunas empresas a "censurar" lo que sus empleados pueden consultar por la red.

- Derecho a la privacidad. La única forma de obtener privacidad en la red es encriptando la información; sin embargo, algunos países (Estados Unidos y Francia, por ejemplo) tienen regulaciones muy severas en ese sentido, al punto de prohibir a los ciudadanos encriptar, salvo si el encriptado es lo bastante "suave" como para poder descifrarlo en caso necesario. Dicho de otro modo: el Estado siempre debe poder descifrar un mensaje si lo considera necesario.

- Anónimos: las redes de ordenadores permiten enviar mensajes anónimos. Si bien esto tiene sus ventajas, se plantean problemas legales cuando se utiliza un anónimo por ejemplo para acusar a una persona.

1.3.- TIPOS DE REDES

De acuerdo con su tecnología de transmisión las redes se clasifican en:

- Redes "broadcast" (difusión).
- Redes punto a punto.

Según su escala también se suelen clasificar en:

- Redes de área local (LAN, Local Area Network)
- Redes de área extensa (WAN, Wide Area Network)

En esta última clasificación también se distingue a veces una categoría intermedia, la formada por las redes de área metropolitana (MAN, Metropolitan Area Network).

La combinación de estos dos criterios nos permite crear una matriz con cuatro categorías posibles; en la práctica existen redes en cada una de estas cuatro categorías, si bien la mayoría encajan en dos de ellas:

	LAN	**WAN**
Broadcast	La mayoría de las LANs (Ethernet, FDDI, Token Ring, etc.), Fibre Channel	Redes de transmisión vía satélite
Punto a punto	HIPPI, Fibre Channel, LANs Conmutadas	La mayoría de las WANs (todas las basadas en enlaces telefónicos, X.25, Frame Relay, RDSI, ATM, etc.)

Tabla 1.1

1.3.1.- Redes broadcast

En las redes "broadcast" el medio de transmisión es compartido por todos los ordenadores interconectados. Normalmente cada mensaje transmitido va dirigido a un único destinatario, cuya dirección aparece en el mensaje, pero para saberlo cada máquina de la red ha de recibir o "escuchar" cada mensaje, analizar la dirección de destino y averiguar si va o no dirigido a ella; las normas de buena educación "telemática" establecen que un ordenador debe descartar sin más análisis todo mensaje que no vaya dirigido a él; sin embargo, algunos programas llamados "sniffers" se dedican a

capturar todo lo que pasa por el cable, independientemente de quien sea su destinatario; con un "sniffer" es muy fácil capturar cualquier cosa, por ejemplo los caracteres que viajan por la red en un proceso de conexión averiguando así de manera rápida el "userid" y la "password" de un usuario cualquiera. La única protección efectiva en las redes "broadcast" es el encriptado de la información.

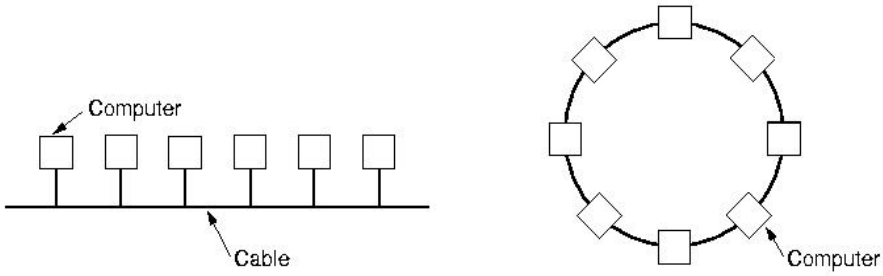

Figura 1.1

A veces en una red "broadcast" lo que se quiere es precisamente enviar un mensaje a todas las máquinas conectadas. Esto se llama un envío *broadcast*. Asimismo es posible enviar un mensaje dirigido a un subconjunto de todas las máquinas de la red (subconjunto que ha de estar definido previamente); esto se conoce como envío *multicast* (y el subconjunto se denomina grupo multicast). En algunos contextos cuando se habla de "broadcast" o multicast el caso en el que el mensaje va dirigido a una máquina concreta se denomina envío *unicast*.

Como ejemplos de redes "broadcast" podemos citar casi todas las tecnologías de red local: Ethernet (en sus diversos tipos), Token Ring, FDDI, etc. También son redes "broadcast" las basadas en transmisión vía satélite. En una red "broadcast" la capacidad o velocidad de transmisión indica la capacidad agregada de todas las máquinas conectadas a la red; por ejemplo, la red conocida como Ethernet tiene una velocidad de 10 Mbps, lo cual significa que la cantidad máxima de tráfico agregado de todos los equipos conectados no puede superar este valor.

Conviene mencionar en este punto que en Telemática siempre que se especifican *capacidades de transmisión* de la información, a menudo referidas erróneamente como *velocidades* de transmisión o *anchos de banda*, 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 Kbp corresponde a 1.000 bits/s, no 1.024 bits/s; análogamente 1 Mbps significa 1.000.000 bits/s, no 1.048.576 bits/s;. Sin embargo cuando no se trata de cantidad de información (sin dividir por el tiempo) el significado sigue siendo el habitual, así por ejemplo si decimos que un determinado protocolo utiliza un tamaño máximo de paquete de 64 Kbytes queremos decir que el paquete puede contener hasta 65535 Bytes; si decimos que hemos transmitido un fichero de 1 MByte, queremos decir que el fichero contiene 1.048.576 Bytes. Normalmente las velocidades o, mas correctamente, las capacidades de transmisión se miden en bits/segundo (*bps*), mientras que el tamaño de una trama, de un paquete o de un fichero se expresa en Bytes.

1.3.2.- Redes punto a punto

Las redes punto a punto se construyen por medio de *conexiones* entre pares de ordenadores, también llamadas *líneas*, *enlaces*, *circuitos* o *canales* ("*lines*", "*links*", "*circuits*", "*channels*" o "*trunks*"). 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 de destino.

Los enlaces que constituyen una red punto a punto pueden ser de tres tipos de acuerdo con el sentido de la transmisión:

- Simplex: la transmisión sólo puede efectuarse en un sentido
- Semi-dúplex o "half-duplex": la transmisión puede hacerse en ambos sentidos, pero no simultáneamente
- Dúplex o "full-duplex": la transmisión puede efectuarse en ambos sentidos a la vez.

En los enlaces semi-dúplex y dúplex la velocidad de conexión es generalmente la misma en ambos sentidos, en cuyo caso se dice que el enlace es simétrico; en caso contrario se dice que es asimétrico.

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

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 de diversas velocidades y distintos grados de ocupación. Como contraste, en una red "broadcast" el camino a seguir de una máquina a otra es único , no existen ordenadores intermedios y el grado de ocupación es el mismo para todas ellas.

Cada uno de los ordenadores 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* o "*end node"*, de lo contrario se dice que es un *nodo intermedio,* de *encaminamiento* o *"routing node"*. Cada nodo intermedio ha de tomar una serie de decisiones respecto a por donde debe dirigir los paquetes que reciba, por lo que también se les llama *nodos de conmutación de paquetes*, *nodos de conmutación*, *conmutadores* o *encaminadores* (los términos equivalentes son respectivamente *packet switching nodes*, *switching nodes*, *switches* y *routers*). Dependiendo del tipo de red que se trate nosotros utilizaremos las denominaciones router o conmutador.

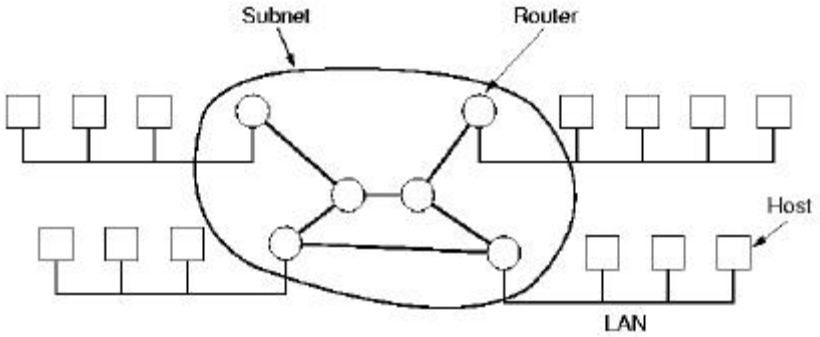

Figura 1.2

Cualquier ordenador (por ejemplo una estación de trabajo UNIX, o incluso un PC con MS/DOS), puede actuar como un router en una red si dispone del programa apropiado; sin embargo, se prefiere normalmente utilizar para este fin ordenadores dedicados, con sistemas operativos en tiempo real y software específico, dejando los ordenadores de propósito general para las aplicaciones del usuario; esto da normalmente mayor rendimiento y fiabilidad. Tradicionalmente al ordenador de propósito general que se conecta a la red como nodo terminal mediante un router se le denomina *host*, anfitrión (aunque esta denominación no se utiliza nunca en este contexto). El conjunto de líneas de comunicación y routers que interconectan a los hosts forman lo que se conoce como la *subred de comunicaciones*, o simplemente *subred.* Obsérvese que los hosts o nodos terminales no forman parte de la subred. Si hacemos la analogía con la red telefónica diríamos que la subred es el conjunto de cables y centralitas telefónicas, incluido el aplique de la pared donde conectamos el teléfono, pero no formaría parte de la subred nuestro teléfono, que enchufamos al aplique.

Para llegar de un nodo a otro en una red se ha de atravesar uno o varios enlaces; el número de enlaces se denomina saltos ("hops"), y depende de la trayectoria seguida y de la topología de la red. Cuando dos nodos no vecinos (es decir a mas de un "hop" de distancia) desean intercambiar información lo han de hacer a través de uno o varios nodos intermedios. Cuando un paquete se envía de un nodo al siguiente normalmente el paquete es transmitido en su totalidad y almacenado; solo entonces el nodo receptor intenta enviar el paquete al siguiente nodo de la red. Esto es lo que se conoce como una red de *almacenamiento - reenvío* ("*store-and-forward"*) o red de *conmutación de paquetes* (*"packet - switched"*). Esta forma de proceder permite una elevada fiabilidad incluso en entornos hostiles donde el número de errores puede ser elevado.

Dado que en una red punto a punto cada enlace puede tener una velocidad distinta, no podemos caracterizar la red con un único dato de forma tan sencilla como en una red "broadcast"; sería preciso adjuntar un esquema de la topología indicando el tipo de cada enlace (símplex, semi-dúplex o dúplex) y su velocidad (en cada sentido si fuera asimétrico).

1.3.3.- Redes de área local

Las redes de área local tienen generalmente las siguientes características:

- Tecnología broadcast: medio compartido
- Cableado específico, instalado normalmente a propósito
- Velocidad de 1 a 1000 Mbps
- Extensión máxima de unos 3 KM (si bien FDDI llega a 200 Km)

Las LANs mas conocidas y extendidas son la Ethernet a 10 Mbps, la IEEE 802.5 o Token Ring a 4 y 16 Mbps, y la FDDI a 100 Mbps. Estos tres tipos de LAN han permanecido prácticamente sin cambios desde finales de los ochenta, por lo que a menudo se les referencia en la literatura como “LANs tradicionales (“legacy LANs”) para distinguirlas de otras más modernas aparecidas en los 90, tales como la Fast Ethernet (100 Mbps) o Gigabit Ethernet (1000 Mbps)

A menudo las LANs requieren un tipo de cableado específico (de cobre o de fibra); esto no suele ser un problema ya que al instalarse en una fábrica, campus o similar, se tiene un control completo sobre el entorno y las condiciones de instalación.

El alcance limitado de las LANs permite saber el tiempo máximo que un paquete tardará en llegar de un extremo a otro de la red, lo cual permite aplicar diseños que de otro modo no serían posibles, y simplifica la gestión de la red.

Como consecuencia del alcance limitado y del control en su cableado, las redes locales suelen tener un retardo muy bajo en las transmisiones (decenas de microsegundos) y una tasa de errores muy baja.

La topología básica de las redes locales suele ser de bus (Ethernet) o de anillo (Token Ring o FDDI). Sin embargo, pueden hacerse topologías más complejas utilizando elementos adicionales, tales como repetidores, puentes, conmutadores, etc., como veremos más adelante.

En los últimos años se ha popularizado una técnica para aumentar el rendimiento de las redes locales, que consiste en dividir una LAN en varias más pequeñas, con lo que el ancho de banda disponible para cada uno es mayor; las diversas LANs así formadas se interconectan en un equipo especial denominado conmutador LAN (o LAN switch); en casos extremos se puede llegar a dedicar una red por equipo, disponiendo así de todo el ancho de banda para él.

También, recientemente se ha empezado a utilizar una tecnología de redes telefónicas, y por tanto típicamente de redes WAN, para la construcción de redes locales: ATM (Asynchronous Transfer Mode).

1.3.4.- Redes MAN

En principio se considera que una MAN abarca una distancia de unas pocas decenas de kilómetros, que es lo que normalmente se entiende como área metropolitana. Existe solamente una red característica de las MANs, la conocida como IEEE 802.6 o DQDB (Distributed Queue Dual Bus), que puede funcionar a diversas velocidades entre 34 y 155 Mbps con una distancia máxima de unos 160 Km. En realidad la distinción de MANs en base a la distancia es un tanto arbitraria, ya que FDDI puede llegar a 200 Km pero raramente se la clasifica como MAN, al no ser un servicio ofrecido por las compañías telefónicas, cosa que sí ocurre con DQDB en algunos países.

La tecnología DQDB ha tenido escasa difusión. Su mayor mérito ha sido servir como predecesora de ATM en algunos aspectos. En el futuro es de esperar que la red DQDB caiga en desuso o desaparezca ya que su espacio ha sido ocupado por completo por las redes basadas en ATM.

Un caso de redes especialmente difíciles de clasificar son las formadas por empresas de televisión por cable. Desde el punto de vista técnico estas redes se podrían considerar tipo LAN; sin embargo el hecho de que sean gestionadas por empresas especializadas y ofrecidas como un servicio contratable por los usuarios les da unas características de WAN desde el punto de vista legal. Estas circunstancias unidas a su alcance máximo (entre 160 y 200 Km) hacen que las podamos considerar en cierto modo como redes MAN.

El término MAN suele utilizarse también en ocasiones para denominar una interconexión de LANs ubicadas en diferentes recintos geográficos (por ejemplo diferentes campus) cuando se dan las siguientes circunstancias:

- La interconexión hace uso de enlaces telefónicos de alta o muy alta velocidad (comparable a la de las propias LANs interconectadas).
- La interconexión se efectúa de forma transparente al usuario, que aprecia el conjunto como una única LAN por lo que se refiere a servicios, protocolos y velocidades de transmisión.
- Existe una gestión unificada de toda la red

1.3.5.- Redes WAN

Las redes de amplio alcance se utilizan cuando no es factible tender redes locales, bien porque la distancia no lo permite por el costo de la infraestructura o simplemente porque es preciso atravesar terrenos públicos en los que no es posible tender infraestructura propia. En todos estos casos lo normal es utilizar para la transmisión de los datos los servicios de una empresa portadora. Hasta hace poco este tipo de servicios eran ofrecidos en régimen de monopolio por las compañías telefónicas en la mayoría de los países de Europa. Afortunadamente esto esta cambiando rápidamente siendo posible por ejemplo en España contratar hoy en día servicios portadores de datos con Retevisión, Euskaltel, BT, Jazztel y en breve con diversas compañías de televisión por cable, si bien en muchos casos la mayor penetración de Telefónica hace que todavía resulte un monopolio de facto.

En la literatura especializada es frecuente referirse a las compañías telefónicas europeas genéricamente con la denominación PTT, abreviatura de Post, Telegraph and Telephone. Esto se debe a que en muchos países de Europa la empresa que se encargaba tradicionalmente de las transmisiones telefónicas era la misma que ofrecía el servicio de correos y telégrafos, todo esto en régimen de monopolio. Con la liberalización del servicio de telefonía y la aparición de diversas compañías competidoras la denominación PTT se esta sustituyendo por la de *operador*; la costumbre hace que en muchos casos se siga aun utilizando el término PTT.

Las redes WAN se han implementado, casi siempre, haciendo uso de enlaces telefónicos que han sido diseñados principalmente para transmitir la voz humana, ya que este es el principal negocio de las compañías telefónicas. Normalmente la infraestructura está fuera del control del usuario, estando supeditado el servicio disponible a la zona geográfica de que se trate. Conseguir capacidad en redes WAN suele ser caro, por lo que generalmente se solicita el mínimo imprescindible.

Hasta tiempos recientes las conexiones WAN se caracterizaban por su lentitud, costo y tasa de errores relativamente elevada. Con la paulatina introducción de fibras ópticas y líneas digitales en las infraestructuras de las compañías portadoras las líneas WAN han reducido apreciablemente su tasa de errores; también se han mejorado las capacidades y reducido los costos. A pesar del inconveniente que en ocasiones pueda suponer el uso de líneas telefónicas tienen la gran virtud de llegar prácticamente a todas partes, que no es poco.

Con la excepción de los enlaces vía satélite, que utilizan transmisión broadcast, las redes WAN se implementan casi siempre con enlaces punto a punto, por lo que prácticamente todo lo que hemos dicho en el apartado de redes punto a punto es aplicable a las redes WAN.

1.3.6.- Redes Inalámbricas y movilidad.

En los últimos años ha habido un auge inesperado de los sistemas de telefonía inalámbrica. Algunos usuarios requieren facilidades para conectar por radioenlaces sus ordenadores personales desde cualquier lugar o mientras se encuentran viajando en tren, autobús, etc. El sistema de telefonía inalámbrica digital GSM (Global System for Mobile Communications), muy extendido en Europa, utiliza un canal digital para transmitir la voz, por lo que es posible conectar un ordenador portátil mediante un teléfono GSM, sin necesidad de módem. Ya pueden realizarse conexiones inalámbricas a 64 Kbps utilizando una versión modificada del GSM, si bien la limitación de la velocidad hace que su interés resulte muy limitado.

La conexión de ordenadores con total movilidad es importante en aplicaciones tales como flotas de taxis, camiones, autobuses, servicios de emergencia, fines militares, etc. En estos casos se emplean, además de los ya familiares ordenadores portátiles conocidos como “laptops”, otros aún más pequeños que se conocen como “palmtop”, asistente digital personal o PDA (Personal Digital Assistant), y que son algo intermedio entre un ordenador portátil y una agenda electrónica.

Las redes inalámbricas también tienen utilidad en algunos casos donde no se requiere movilidad, como en las LANs inalámbricas. Por ejemplo, una empresa que desea establecer una nueva oficina y por rapidez, provisionalidad de la ubicación o simples razones estéticas no desea cablear el edificio puede utilizar una LAN inalámbrica, consistente en una serie de equipos transmisores-receptores. Las LAN inalámbricas son generalmente más lentas que las normales (1-2 Mbps) y tienen una mayor tasa de errores, pero para muchas aplicaciones pueden ser adecuadas.

La movilidad es importante también en casos en que no hay involucradas conexiones inalámbricas. Por ejemplo un representante que desee conectar con su oficina desde su ordenador portátil cuando se encuentra de viaje puede optar por llamar a su oficina directamente, pagando posiblemente una costosa llamada de larga distancia, o bien puede llamar al punto de presencia (POP, Point Of Presence) más próximo de algún proveedor de servicios de comunicación, y a través de este acceder a su oficina por una infraestructura compartida que le resulte más barata (por ejemplo la Internet); en este último caso se dan una serie de problemas de solución no trivial en cuanto a la seguridad y el correcto encaminamiento del tráfico.

1.3.7.- Internetworking

Si bien las clasificaciones de redes antes estudiadas tienen interés como medio de sistematizar su estudio, es obvio que en la realidad casi nunca se da uno de esos tipos en estado puro. Por ejemplo, una LAN (que normalmente será una red de tipo broadcast) casi siempre dispondrá de un router que la interconecte a una WAN (que generalmente consistirá en un conjunto de enlaces punto a punto). Esta interconexión de tecnologías diferentes se conoce como “internetworking” (que podríamos intentar traducir como “interredes”). El router que interconecta redes diferentes está físicamente conectado a todas las redes que se desean interconectar.

Además de la combinación de medios físicos diversos es posible encontrarse con necesidades de “internetworking” en un mismo medio físico; este es el caso cuando coexisten protocolos de comunicación diferentes; por ejemplo, en una misma red ethernet puede haber unos ordenadores utilizando el protocolo TCP/IP y otros utilizando DECNET (protocolo típico de la marca de ordenadores Digital). Al ser protocolos diferentes son completamente independientes y no se pueden hablar entre sí, por lo que un usuario de un ordenador TCP/IP no podría por ejemplo enviar un mensaje de correo electrónico a uno de un ordenador DECNET. Sin embargo, es posible instalar en un ordenador ambos protocolos, y un programa de conversión de correo electrónico, de forma que los usuarios de ambas redes puedan intercambiar mensajes. A la máquina que interconecta el correo electrónico de los dos protocolos se la denomina *pasarela* (“*gateway*”). Generalmente las pasarelas han de implementarse a nivel de aplicación; así disponer en nuestro ejemplo de una pasarela para el correo electrónico no significa que podamos transferir ficheros entre máquinas TCP/IP y DECNET, ya que para esto haría falta una pasarela del servicio de transferencia de ficheros. Una misma máquina puede actuar como pasarela para varios servicios. Haciendo una analogía podemos decir que los protocolos son como idiomas y las pasarelas equivalen a servicios de traducción que permiten entenderse a personas que hablan diferentes lenguas.

Cuando una red esta formada por la interconexión de varias redes se le denomina *internet*. A principios de los setenta se creó en los Estados Unidos una internet mediante la unión de varias redes que utilizando medios de transmisión diversos empleaban un conjunto común de protocolos en el nivel de red y superiores, denominados TCP/IP. Con el tiempo la denominación Internet (con I mayúscula) terminó convirtiéndose en el nombre propio de dicha red, muy conocida en nuestros días.

1.4.- ARQUITECTURA DE REDES

En los inicios de la informática el diseño de un ordenador resultaba en sí mismo una tarea tan compleja que no se tomaba en consideración la compatibilidad con otros modelos de ordenadores; la preocupación fundamental era que el diseño fuera correcto y eficiente. Como consecuencia de esto era preciso crear para cada nuevo modelo de ordenador un nuevo sistema operativo y conjunto de compiladores. Los programas escritos en lenguaje máquina o en ensamblador (que entonces eran la mayoría) tenían que ser prácticamente reescritos para cada nuevo modelo de ordenador.

En 1964 IBM anunció un nuevo ordenador denominado *Sistema/360*. Se trataba en realidad de una familia formada por varios modelos que compartían una *arquitectura* común (era la primera vez que se utilizaba este término referido a ordenadores). La arquitectura establecía unas especificaciones comunes que hacían compatibles a todos los modelos de la familia (conjunto de instrucciones, forma de representar los datos, etc.), pudiendo así ejecutar los mismos programas, utilizar el mismo sistema operativo, compiladores, etc. en toda la familia, que comprendía una gama de ordenadores de potencias y precios diversos. El nombre 360 se eligió en base a la década en que se creó (los 60) y a la idea de que era una arquitectura polivalente, que pretendía servir para aplicaciones de todo tipo (360º, o sea que puede ir en todas direcciones). La arquitectura 360 ha ido evolucionando hasta desembocar en nuestros días en la arquitectura ESA/390, utilizada en los grandes ordenadores IBM (mainframes) actuales, que son aún la base de las aplicaciones críticas en grandes empresas (bancos, líneas aéreas, etc.). Todos los fabricantes de ordenadores actuales utilizan una o varias arquitecturas como base para el diseño de sus equipos.

Las primeras redes de ordenadores tuvieron unos inicios muy similares a los primeros ordenadores: Las redes y los protocolos se diseñaban pensando en el hardware a utilizar en cada momento, sin tener en cuenta la evolución previsible, ni por supuesto la interconexión y compatibilidad con equipos de otros fabricantes (seguramente muchos creían que bastante trabajo suponía conseguir que las cosas funcionaran como para perder el tiempo con florituras). A medida que la tecnología avanzaba y se mejoraba la red se vivieron experiencias parecidas a las de los primeros ordenadores: los programas de comunicaciones, que habían costado enormes esfuerzos de desarrollo, tenían que ser reescritos para utilizarlos con el nuevo hardware, y debido a la poca modularidad prácticamente nada del código era aprovechable.

El problema se resolvió de forma análoga a lo que se había hecho con los ordenadores. Cada fabricante elaboró su propia *arquitectura de red,* que permitía independizar las funciones y el software del hardware concreto utilizado. De esta forma cuando se quería cambiar algún componente sólo la función o el módulo afectado tenía que ser sustituido. La primera arquitectura de redes fue anunciada por IBM en 1974, justo diez años después de anunciar la arquitectura S/360, y se denominó SNA (Systems Network Architecture). La arquitectura SNA se basa en la definición de siete niveles o capas, cada una de las cuales ofrece una serie de servicios a la siguiente, la cual se apoya en esta para implementar los suyos, y así sucesivamente. Cada capa puede implementarse en hardware, software o una combinación de ambos. El módulo (hardware y/o software) que implementa una capa en un determinado elemento de la red debe poder sustituirse sin afectar al resto de la misma, siempre y cuando el protocolo utilizado se mantenga inalterado. Dicho en otras palabras, SNA es una arquitectura altamente modular y estructurada. No vamos a entrar en mas detalles sobre la arquitectura SNA, ya que cae fuera de los objetivos del presente curso, pero sí diremos que el modelo de capas que utiliza ha sido la base de todas las arquitecturas de redes actualmente en uso, incluidas las basadas en el modelo OSI (Open Systems Interconnection) y el TCP/IP.

Las ideas básicas del modelo de capas son las siguientes:

- La capa *n* ofrece una serie de servicios a la capa *n+1.*
- La capa *n* solo “ve” los servicios que le ofrece la capa *n-1.*

- La capa *n* en un determinado sistema solo se comunica con su homóloga en el sistema remoto (comunicación de igual a igual o "peer-to-peer"). Esa "conversación" se efectúa de acuerdo con una serie de reglas conocidas como *protocolo de la capa n.*
- La comunicación entre dos capas adyacentes en un mismo sistema se realiza de acuerdo con una *interfaz.* La interfaz es una forma concreta de implementar un servicio y no forma parte de la arquitectura de la red.

La arquitectura de una red queda perfectamente especificada cuando se describen las capas que la componen, su funcionalidad, los servicios que implementan y los protocolos que utilizan para hablar con sus "iguales". El conjunto de protocolos que utiliza una determinada arquitectura en todas sus capas se denomina *pila de protocolos* ("protocol stack"); así es frecuente oír hablar de la pila de protocolos OSI, SNA, TCP/IP o DECNET, por ejemplo.

Para mejor comprender como funciona el modelo de arquitectura de redes basado en capas hagamos una analogía. Supongamos que un ejecutivo de la empresa A desea enviar de forma urgente un importante informe a un colega suyo en la empresa B. Para esto hablará con aquél notificándole el envío y a continuación pasará a su secretaria el informe con las instrucciones correspondientes. La secretaria llamará a la secretaria de B para averiguar la dirección exacta, pondrá el informe en un sobre y llamará a un servicio de mensajería, que enviará a un motorista para que recoja el paquete y lo lleve al aeropuerto. Cuando el paquete llega al aeropuerto de destino es recogido allí por otro motorista que lo lleva a la oficina de la empresa B y lo entrega a la secretaria; ésta se ocupará de los trámites administrativos (pagar al mensajero, abrir el paquete, comprobar su contenido, acusar recibo, etc.) y lo pasará después a su jefe, quien una vez estudiado el informe llamará al ejecutivo de A.

Obsérvese que en el proceso anterior existen diferentes niveles claramente diferenciados: los ejecutivos, las secretarias, los motoristas, y por último la empresa de líneas aéreas que se ocupa del transporte físico de la mercancía. En todos los niveles (menos probablemente el más bajo) hay dos entidades, la transmisora (A) y la receptora (B). Si todo ocurre según lo previsto cada entidad sólo hablará con su correspondiente en el otro lado, y con sus entidades vecinas, es decir, el jefe de A sólo habla con el jefe de B y con su secretaria, la secretaria habla con su jefe, con el motorista y con la otra secretaria para confirmar el envío, etc. En ningún caso se contempla que la secretaria de A hable con el ejecutivo de B. Si por ejemplo la secretaria de A es sustituida por enfermedad por otra persona los procedimientos seguirán funcionando, siempre y cuando la secretaria suplente desarrolle la misma función. Las variaciones de carácter interno sólo han de ser conocidas por las entidades contiguas, por ejemplo, el motorista de B podría ser reemplazado por una furgoneta de reparto, y este hecho solo ha de ser conocido por la secretaria de B y por la persona que entrega los paquetes en el aeropuerto. Esto es lo que denominamos una interfaz. Obsérvese que el modelo de capas simplifica considerablemente la tarea de cada una de las entidades, que sólo tiene que preocuparse de una pequeña parte de todo el mecanismo. En esencia se trata de aplicar a la resolución de problemas la vieja fórmula de divide y vencerás.

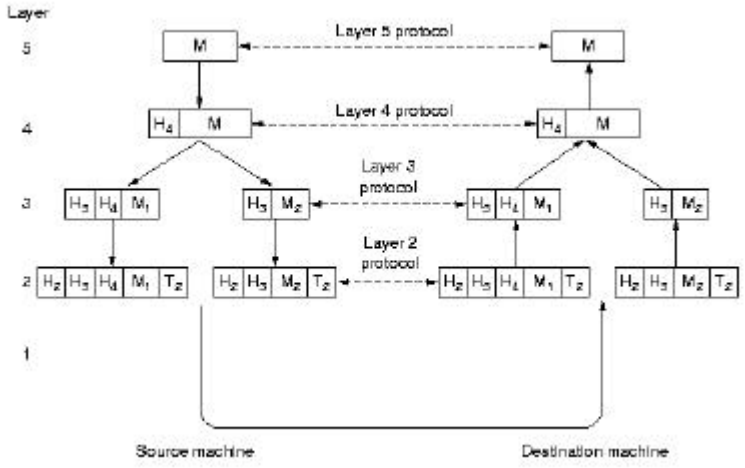

Figura 1.3

Cuando un sistema desea enviar un mensaje a un sistema remoto normalmente la información se genera en el nivel más alto; conforme va descendiendo se producen diversas transformaciones, por ejemplo adición de cabeceras, de colas, de información de control, la fragmentación en paquetes mas pequeños si es muy grande (o mas raramente la fusión con otros si es demasiado pequeño), etc. Todas estas operaciones se invierten en el sistema remoto en las capas correspondientes, llegando en cada caso a la capa correspondiente en el destino un mensaje igual al original.

1.4.1.- Decisiones en el diseño de arquitecturas de redes.

Cuando se diseña una arquitectura de red hay una serie de aspectos y decisiones fundamentales que condicionan todo el proceso. Entre estos cabe mencionar los siguientes:

- *Direccionamiento*: cada capa debe poder identificar los mensajes que envía y recibe. En ocasiones un mismo ordenador puede tener varias instancias de una misma capa, por lo que la sola identificación del ordenador puede no ser suficiente.

- Normalmente cualquier protocolo admite comunicación en ambos sentidos (dúplex); pero no siempre se permite que esta ocurra de forma *simultánea* (full-dúplex). también se debe determinar si se definirán *prioridades*, y cuáles serán éstas.

- En cualquier comunicación es preciso establecer un *control de errores*, ya que los canales de comunicación no son totalmente fiables. Es preciso decidir que código de detección y/o corrección de errores se va a utilizar, y en que capa o capas se va a llevar a cabo. Generalmente a medida que los medios de transmisión mejoran y las tasas de errores disminuyen la detección/corrección se va suprimiendo de las capas inferiores y dejando al cuidado de las más altas, ya que es un proceso costoso que puede llegar a ralentizar apreciablemente la transmisión.

- En algunos casos se debe tener en cuenta la posibilidad de que los paquetes lleguen a su destino en *orden diferente* al de envío.

- Debe contemplarse la posibilidad de que el receptor no sea capaz de “digerir” la información enviada por el transmisor. Para esto es conveniente disponer de algún mecanismo de control de flujo y notificación para indicar la *congestión*.

- Normalmente los equipos funcionan de forma óptima cuando el tamaño de los mensajes que se envían esta dentro de un cierto rango. Para evitar los problemas que puede producir el envío de mensajes muy grandes o muy pequeños se suelen contemplar mecanismos de *fragmentación* y reagrupamiento. Es importante que estos mecanismos estén claramente especificados para evitar la destrucción del mensaje en tránsito.

1.4.2.- Interfaces y servicios

Debido a su importancia vamos a estudiar con mas detalle que es un servicio. Empezaremos con algunas definiciones.

Llamaremos *entidad* a los elementos activos en cada capa. Una entidad puede ser un proceso, un componente hardware, o una combinación de ambos. Un ordenador puede tener una o varias entidades en cada capa (por ejemplo un ordenador con dos tarjetas de conexión a LAN).

Llamaremos *entidades iguales* o *entidades pares* (“*peer entities”*) a dos entidades homólogas, es decir entidades diferentes de la misma capa (generalmente estarán en diferentes máquinas, pero podrían estar en la misma).

Las entidades de la capa *n* implementan los servicios que utiliza la capa *n+1*. En este caso la capa *n* actúa como el *proveedor del servicio* y la capa *n+1* es el *usuario del servicio.* El uso que la capa *n* haga de los servicios de la capa *n-1* es algo que no afecta ni incumbe a la capa *n+1*.

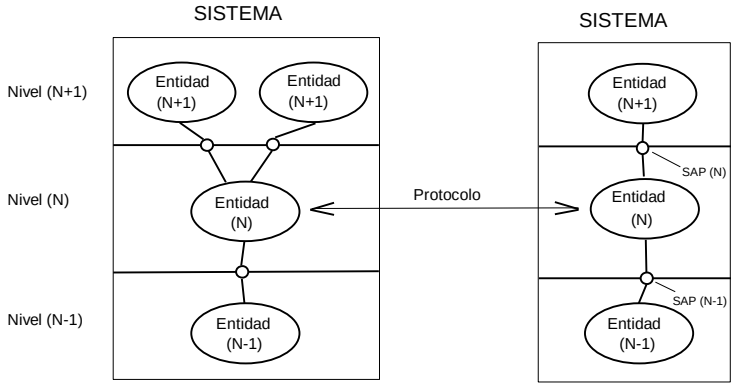

Figura 1.4

Los servicios están disponibles en los *SAPs* (Service Access Points). Los SAPs de la capa *n* son los puntos donde la capa *n+1* puede acceder a los servicios ofertados. Cada SAP de cada entidad de la capa *n* tiene una dirección que le identifica de forma única en toda la red.

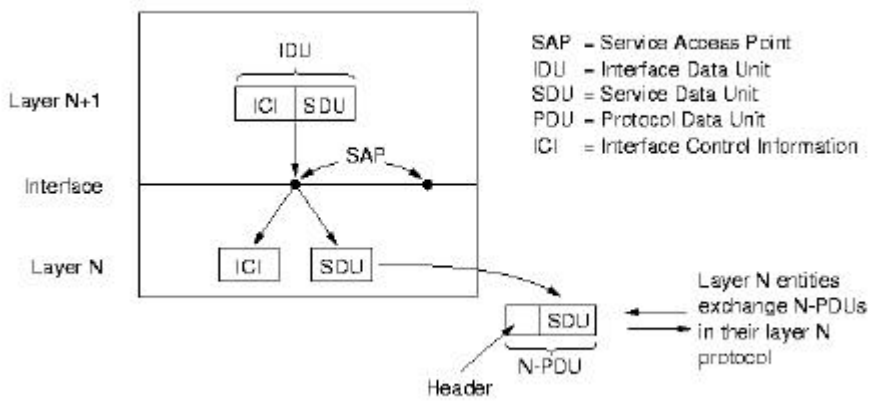

Figura 1.5

Denominamos *interfaz* al conjunto de reglas que gobiernan el intercambio de información entre capas. En una comunicación la entidad de la capa n+1 intercambia una *IDU* (Interface Data Unit) con la entidad de la capa *n* a través del SAP. La IDU esta formada por una SDU (Service Data Unit) e información de control. La SDU es la información que se transmite a la entidad equivalente (peer) en el lado contrario, y de allí a la capa *n+1* a través de su SAP. La información de control es necesaria como su nombre indica para que la capa n haga correctamente su trabajo, pero no es parte de los datos mismos. En la especificación de una arquitectura solo es necesario describir la estructura de la SDU, pero no la de la IDU; ésta se describe en la interfaz, que puede ser distinta para cada implementación.

Para transferir la SDU (Service Data Unit) la entidad de la capa n puede tener que fragmentarla en varias PDUs (Protocol Data Units). Cada PDU llevará una cabecera que permitirá a la entidad de la capa *n* en el otro lado ensamblar de nuevo la SDU correctamente.

1.4.3.- Servicios orientados y no orientados a conexión

En una arquitectura de redes cada capa utiliza los servicios de la capa inmediatamente inferior para comunicar con la correspondiente del otro extremo. En función de como se establezca esa comunicación suelen distinguirse dos tipos de servicios: orientados a conexión y no orientados a conexión.

En el *servicio orientado a conexión*, también llamado CONS (Connection Oriented Network Service), primero se establece el canal de comunicación, después se transmiten los datos, y por último se termina la conexión. Dicha "conexión" se denomina *circuito virtual* (VC, virtual circuit). Una vez establecido el VC el camino físico que van a seguir los datos está determinado; los paquetes deben ir todos por él desde el origen al destino, y llegar en el mismo orden con el que han salido. Dado que el VC establece de forma clara el destino, los paquetes no necesitan contener su dirección. Generalmente se distinguen dos tipos de circuitos virtuales: *conmutados*, también llamados SVCs (Switched Virtual Circuits), y *permanentes*, conocidos también como PVCs (Permanent Virtual Circuits). Los SVCs se establecen y terminan a petición del usuario, normalmente cuando hay paquetes que se quieren transmitir. Los PVCs están establecidos todo el tiempo que la red está operativa (o al menos eso es lo que se pretende). Al hablar de circuitos utilizaremos las denominaciones "establecer" y "terminar" en vez de abrir y cerrar, ya que estos términos tienen un significado completamente opuesto según se trate de ingenieros informáticos o electrónicos (para un ingeniero electrónico un circuito esta abierto cuando esta interrumpido, es decir cuando no puede viajar por el ninguna señal).

En el servicio *no orientado a conexión*, llamado también CLNS (ConnectionLess Network Service) la comunicación se establece de manera menos formal. Cuando una entidad tiene información que transmitir sencillamente la envía en forma de paquetes, confiando que estos llegaran a su destino más pronto o más tarde. No se establece previamente un VC ni otro tipo de canal de comunicación extremo a extremo; los paquetes pueden ir por caminos físicos diversos, y deben incluir cada uno la dirección de destino. Los paquetes pueden ser almacenados por nodos intermedios de la red, y reenviados mas tarde. Aunque lo normal es que lleguen en el mismo orden con que han salido, esto no esta garantizado como ocurría en el servicio orientado a conexión debido al almacenamiento en nodos intermedios y a la diversidad de caminos físicos posibles. A los paquetes enviados en un servicio no orientado a conexión se les denomina *datagramas*, ya que cada paquete viaja hacia su destino de forma completamente independiente de los demás como si fuera un telegrama..

Generalmente se suelen explicar los modelos orientado y no orientado a conexión con dos analogías: el sistema telefónico y el sistema postal. El sistema telefónico es un ejemplo de servicio orientado a conexión, mientras que el sistema postal es un servicio no orientado a conexión. La analogía es bastante exacta salvo por el hecho de que en redes telemáticas la diferencia en el tiempo de entrega del mensaje entre servicios CONS y CLNS no es tan grande como la anterior comparación podría hacer pensar.

En cualquiera de los dos tipos de servicio antes mencionados es posible que se produzca pérdida de información; también puede ocurrir que el tiempo de envío del paquete, también llamado retardo o latencia ("delay" y "latency" respectivamente) sea demasiado grande o fluctúe dentro de un amplio rango debido a la carga o congestión en la red (el término usado para denominar dicha fluctuación es *jitter*, que literalmente significa mieditis, temblar de miedo). En algunos casos se requiere una entrega fiable, es decir que se garantice la entrega de los paquetes, o un retardo y/o jitter garantizados, o sea no superiores a un determinado valor. Por ejemplo si transferimos un fichero, normalmente dividiéndolo en múltiples paquetes, necesitaremos un servicio fiable en la entrega, pero podemos tolerar un retardo o jitter mas o menos grande; por el contrario la voz, o el vídeo (imagen en movimiento) toleran un pequeño porcentaje de pérdidas, pero requieren un retardo y un jitter reducidos y constantes. Cuando al establecer una comunicación se solicita un nivel mínimo para alguno de éstos parámetros se dice que se requiere una *calidad de servicio* (llamada QoS, Quality of Service). La calidad de servicio estipula unos mínimos que la red ha de satisfacer para efectuar la conexión, por ejemplo "transmisión fiable con un retardo no superior a 100 ms"; es posible que la red no sea capaz de satisfacer la calidad solicitada, en cuyo caso podría hacer una propuesta alternativa, a modo de regateo (por ejemplo, "no puedo asegurar 100 ms de retardo, lo mínimo es 250ms, ¿estás conforme?") Una vez pactadas las condiciones de la conexión éstas actúan a modo de contrato que obliga a la red a dar la calidad de servicio prometida al usuario. No todos los protocolos o redes ofrecen la posibilidad de negociar calidades de servicio; en estos casos el protocolo simplemente aprovecha los medios disponibles lo mejor que puede, intentando evitar las congestiones y situaciones críticas en lo posible, y repartir los recursos entre los usuarios de manera mas o menos equilibrada; esta estrategia se denomina del "mejor esfuerzo" (o también "best effort"). Como ejemplos de redes con QoS podemos citar ATM, como ejemplos de redes "best effort" podemos mencionar TCP/IP (la Internet) y Ethernet.

1.4.4.- Primitivas de servicio

Recordemos que, en el modelo de capas, cada capa ofrece sus servicios a la siguiente. El servicio se define por un conjunto de operaciones u órdenes que la capa superior puede mandar a la capa inferior. Dicho conjunto de operaciones se denomina *primitivas.*

Dichas primitivas son el nombre de una función que ordena a un determinado servicio ejecutar determinada acción. Las primitivas pueden ser de cuatro tipos:

- Petición (Request)
- Indicación (Indication)
- Respuesta (Response)
- Confirmación (Confirm)

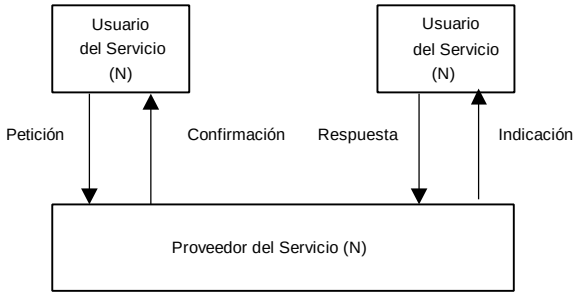

Figura 1.6

Vamos a analizar en detalle las primitivas que participan en el establecimiento y terminación de una conexión entre la capa n de dos sistemas llamados A y B. La entidad A.n (es decir, la capa n del sistema A) inicia la conexión emitiendo la primitiva *CONNECT.request,* que provoca la transferencia de una IDU (Interface Data Unit) a través del SAP (Service Access Point) a la entidad A.n-1; ésta extrae la información de control y la interpreta creando la SDU (Service data Unit), que convierte en una o varias PDUs (Protocol Data Units); las PDUs son transferidas a B.n-1, que regenera a partir de ello la SDU, luego la información de control correspondiente y con ambos la IDU; una vez dispone de la IDU la transmite a B.n a través del SAP mediante la primitiva *CONNECT.indication*, que le indica a B.n que alguna entidad desea establecer conexión con ella. La entidad B.n emite entonces la primitiva *CONNECT.response* para indicar si acepta o rechaza la conexión (las primitivas pueden llevar parámetros y sería aquí donde se indicaría esto). La respuesta se traduce en un paquete que B.n-1 envía a A.n-1, el cual informa a A.n de la situación mediante la primitiva *CONNECT.confirm.*

Obsérvese que el mismo evento origina diferentes primitivas en cada lado. Una CONNECT.request produce una CONNECT.indication en el lado contrario, y la CONNECT.response se convierte en CONNECT.confirm. Existe una cierta simetría entre las primitivas, ya que a una CONNECT.request siempre le corresponderá una CONNECT.indication en el lado opuesto (salvo que falle la comunicación).

En este ejemplo hemos hecho un servicio *confirmado*, es decir, hemos verificado que la conexión se establecía, para lo cual ha tenido que enviarse un paquete en cada sentido. Se podría haber hecho una conexión no confirmada, para lo cual sencillamente se habría emitido la CONNECT.request y la CONNECT.indication.

Una vez establecida la conexión lo normal sería transferir datos, y finalmente terminar la conexión. Un ejemplo del conjunto de primitivas que se emitirían a lo largo de una conexión podría ser el siguiente:

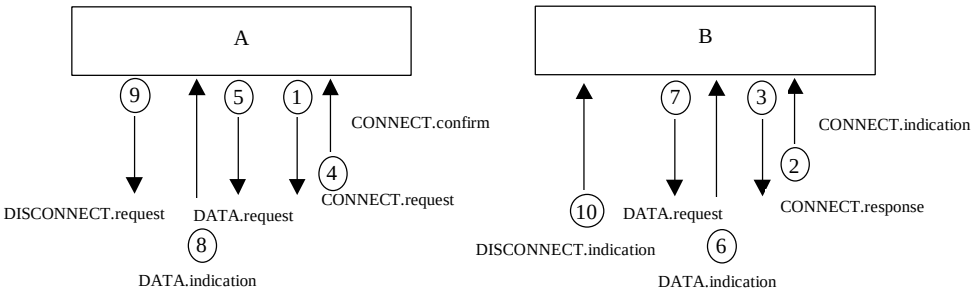

Figura 1.7

Aquí hemos introducido cuatro nuevas primitivas para poder transferir datos y terminar la conexión. Obsérvese que como antes una primitiva *request* va seguida siempre de una *indication* en el lado contrario. En el ejemplo hemos supuesto que se intercambiaban únicamente dos paquetes de datos.

Como ya hemos dicho las primitivas pueden llevar parámetros, y de hecho casi siempre los llevan. Por ejemplo una CONNECT.request llevará la máquina con la que se desea conectar, una CONNECT.indication dirá la máquina que quiere conectar con nosotros, etc. La descripción detallada de estos argumentos, su significado, etc., no es parte de la especificación de las primitivas (y por tanto del servicio) sino del protocolo. El protocolo puede modificarse sin necesidad de cambiar las primitivas. Por ejemplo, un protocolo puede establecer que el servicio de establecimiento de conexión sea confirmado y otro que no lo sea, y ambos pueden utilizar el mismo conjunto de primitivas antes descrito.

Una vez mas diremos que la interfaz no forma parte del protocolo. Por ejemplo imaginemos en el caso anterior que las entidades A.n y A.n-1 acuerdan que la SDU estará codificada en EBCDIC, mientras que B.n y B.n-1 acuerdan utilizar ASCII. Si el protocolo de la capa n-1 establece que la PDU estará en ASCII, entonces A.n-1 sabe que deberá realizar la conversión de códigos cada vez que construya una PDU a partir de una SDU, o viceversa.

1.5.- MODELOS DE REFERENCIA

Hasta aquí hemos hablado del modelo de capas en un sentido genérico. Vamos a hablar ahora con cierto detalle de las dos arquitecturas de redes mas importantes en la actualidad, correspondientes a los protocolos OSI (Open Systems Interconnection) y TCP/IP (Transmission Control Protocol/Internet Protocol). Conviene destacar que la arquitectura es una entidad abstracta, mas general que los protocolos o las implementaciones concretas en que luego se materializan éstos. Típicamente para cada capa de una arquitectura existirán uno o varios protocolos, y para cada protocolo habrá múltiples implementaciones. Las implementaciones cambian continuamente; los protocolos ocasionalmente se modifican o aparecen otros nuevos que coexisten con los anteriores o los dejan anticuados; sin embargo una vez definida una arquitectura ésta permanece esencialmente intacta y muy raramente se modifica.

1.5.1.- El modelo de referencia OSI

Después de la especificación de SNA por parte de IBM cada fabricante importante definió su propia arquitectura de redes; así la evolución de los productos de comunicaciones estaba garantizada, pero no se había resuelto el problema de la interoperabilidad entre diferentes fabricantes. Debido a la posición de hegemonía que IBM disfrutaba en los años 70 y principios de los ochenta la compatibilidad con IBM era un requisito necesario, por lo que la mayoría de los fabricantes tenían implementaciones de los protocolos SNA para sus productos, o estas estaban disponibles a través de terceros. Así, la forma más sencilla de interconectar dos equipos cualesquiera era conseguir que ambos hablaran SNA.

En 1977 la ISO (International Organization for Standardization) consideró que esta situación no era la más conveniente, por lo que entre 1977 y 1983 definió la arquitectura de redes OSI con el fin de promover la creación de una serie de estándares que especificaran un conjunto de protocolos independientes de cualquier fabricante. Se pretendía con ello no favorecer a ninguno a la hora de desarrollar implementaciones de los protocolos correspondientes, cosa que inevitablemente habría ocurrido si se hubiera adoptado alguna de las arquitecturas existentes, como la SNA de IBM o la DNA (Digital Network Architecture) de Digital. Se esperaba llegar a convertir los protocolos OSI en el auténtico *Esperanto* de las redes telemáticas. Por diversas razones que veremos luego el éxito de los protocolos OSI en la práctica ha sido mucho menor de lo inicialmente previsto (cosa que por cierto también le ha ocurrido al Esperanto, aparentemente).

Seguramente la aportación más importante de la iniciativa OSI ha sido precisamente su arquitectura. Ésta ha servido como marco de referencia para describir multitud de redes correspondientes a diversas arquitecturas, ya que la arquitectura OSI es bien conocida en entornos de redes, y su generalidad y no dependencia de ningún fabricante en particular le hacen especialmente adecuada para estos fines. Por este motivo generalmente a la arquitectura OSI se la denomina *Modelo de Referencia OSI,* o también *OSIRM* (OSI Reference Model). Por extensión hoy en día se utiliza a menudo el término *modelo de referencia* para referirse a una arquitectura de red; así oímos hablar del Modelo de Referencia TCP/IP, el Modelo de Referencia ATM, etc.

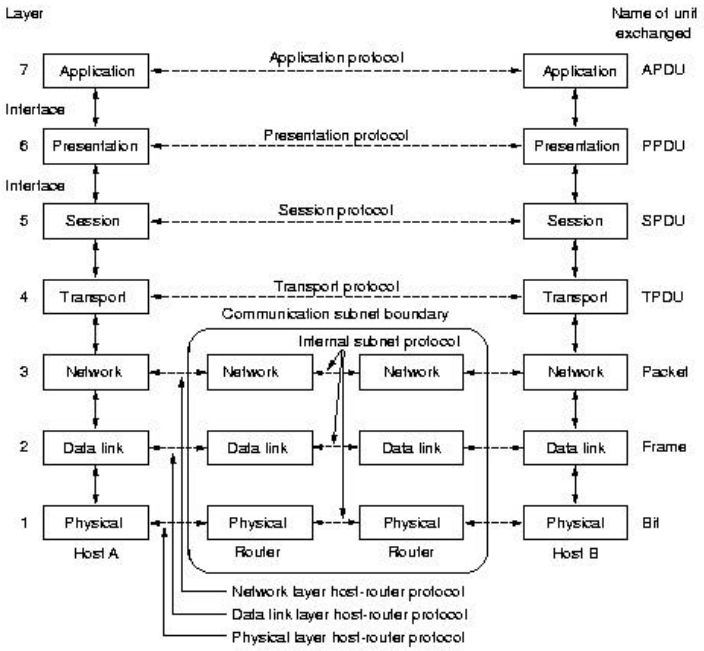

Figura 1.8

El modelo OSI define siete capas, curiosamente como en la arquitectura SNA si bien la funcionalidad es diferente. Las capas son las siguientes:

- Física
- Enlace
- Red
- Transporte
- Sesión
- Presentación
- Aplicación

La ISO ha especificado protocolos para todas las capas, aunque algunos son poco utilizados. En función del tipo de necesidades del usuario no siempre se utilizan todas ellas.

Pasaremos a describir brevemente las funciones desarrolladas por cada una de las capas.

1.5.1.1.- La Capa Física

El nivel 1 o capa física del modelo de referencia OSI es el responsable de llevar a cabo la transmisión de los bits de datos a través del canal de comunicación asegurando que cuanto envíe el transmisor llegue sin alteración alguna al receptor, es decir, cuando un extremo del canal emite un 1, éste debe ser recibido por el otro extremo como un 1 y no como un 0. Esto requiere la especificación de cuestiones tales como: los voltajes precisos para representar los estados lógicos 0 y 1; la velocidad de transmisión empleada; si la transmisión puede ocurrir simultáneamente en ambas direcciones o no (simplex, duplex o semiduplex), cómo se establece y libera la conexión física, cuántos pines posee el conector y para qué sirven, etc.

La transmisión se realiza entre dos entidades (nodos) directamente conectadas; puede tratarse de un enlace punto a punto o de una conexión multipunto (p.e. Ethernet). La comunicación puede ser dúplex, semi-dúplex o simplex.

De esta forma, en el nivel físico se definen las características mecánicas, eléctricas, funcionales y de procedimiento necesarias para la activación, mantenimiento y desactivación de enlaces físicos entre sistemas a través de un determinado medio de transmisión.

- La características **mecánicas** hacen referencia a las propiedades físicas del interfaz entre el sistema y el medio de transmisión: dimensión del interfaz, tipo de conector, configuración, etc.
- Las características **eléctricas** definen los niveles de voltaje, impedancia, y en general todo el conjunto de propiedades eléctricas del interfaz físico.
- Las características **funcionales** especifican las funciones que cada elemento del interfaz, entre el sistema y el medio de transmisión debe de llevar a cabo. Así por ejemplo, especifican el significado de la presencia de determinados niveles de voltaje sobre determinados pines del conector.
- Las características **de procedimiento** determinan el protocolo mediante el cual se realiza el intercambio de secuencias de bits entre entidades de niveles físico pares a través del medio físico.

El nivel físico difiere en gran medida del resto de niveles OSI en cuanto a que mientras el resto de niveles hacen uso de los niveles inferiores para transmitir sus PDUs, el nivel físico debe hacer uso de un medio de transmisión cuyas características no son parte del modelo de referencia OSI. No existe estructura de PDU de nivel físico ya que no existe cabecera de nivel físico con información de control como ocurre en el resto de niveles. Este nivel únicamente trata con secuencias de bits procedentes o dirigidas al nivel de enlace de datos.

FUNCIONES Y SERVICIOS DEL NIVEL FISICO

La funcionalidad básica del nivel físico consiste en la activación, mantenimiento y desactivación de conexiones físicas entre entidades de enlace de datos para la transferencia de secuencias de bits entre los extremos de cada conexión física.

Mediante el control de estas conexiones físicas, el nivel físico proporciona al nivel de enlace de datos los siguientes servicios:

- Cada conexión de nivel físico establece un 'circuito de datos' entre los extremos de la comunicación que proporciona un medio de transferencia de datos transparente entre entidades de enlace de datos.
- El nivel físico ofrece un servicio de secuenciamiento, entregando al nivel de enlace de datos receptor las secuencias de bits transmitidas en el mismo orden en el que el nivel de enlace de datos emisor los emitió.

El circuito de datos que establece una conexión física puede ser directo entre ambos extremos de la comunicación, o indirecto, a través de uno o varios elementos de retransmisión intermedios.

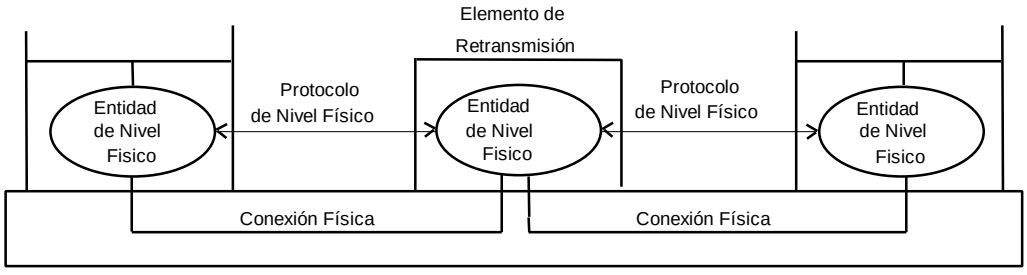

Figura 1.9

Como ejemplos de la capa física podemos mencionar las norma EIA RS-232-C, utilizada por las puertas COM de los ordenadores personales, la EIA-RS-449, CCITT X.21/X.21bis, CCITT V.35. Las normas de redes locales incluyen en sus especificaciones la capa física (IEEE 802.3 o Ethernet, IEEE 802.5 o Token Ring, ISO 9314 o FDDI, etc.)

Muchas de las normas que existen en la capa física se refieren a la interfaz utilizada para conectar un ordenador con un módem o dispositivo equivalente, que a través de una línea telefónica conecta con otro módem y ordenador en el extremo opuesto. Este es el caso por ejemplo de las normas EIA RS-232-C, EIA-RS-449, CCITT X.21/X.21bis y CCITT V.35 antes mencionadas. En estos el conector del

ordenador y el módem son de diferente “sexo” (macho o hembra). En este contexto se suele utilizar la denominación *DTE (Data Terminal Equipment)* para referirse al ordenador y *DCE (Data Circuit-Terminating Equipment)* para referirse al módem. El “módem” en ocasiones no es más que un adaptador, ya que por ejemplo la norma X.21 se utiliza para líneas digitales. En sentido general al equipo que actúa como adaptador entre el ordenador y el medio de transmisión se le denomina CSU/DSU (Channel Service Unit/ Data Service Unit).

1.5.1.2.- La capa de Enlace de Datos (data link)

La principal función de la capa de enlace es ofrecer un servicio de comunicación fiable a partir de los servicios que recibe de la capa física, también entre dos entidades contiguas de la red.

La existencia de una conexión física entre dos sistemas que se comunican asegura la transferencia y posterior entrega de las secuencias de bits exactamente en el mismo orden en el que fueron emitidas. Pero esto no es suficiente, ocasionalmente se producen errores en los circuitos de comunicación, existe un retardo en la propagación de los bits, los sistemas emisor y receptor pueden funcionar a distintas velocidades, etc. Los protocolos de comunicación que tienen en cuenta todos estos factores e intentan paliarlos de alguna manera son los protocolos del nivel de enlace de datos.

Así pues, el nivel de enlace de datos añade ciertas capacidades de comunicación adicionales a las proporcionadas por el nivel físico. La posibilidad de que se generen errores en el nivel físico hace que se realice detección y posiblemente corrección de errores. A diferencia de la capa física, que transmitía los bits de manera continua, la capa de enlace transmite los bits en grupos denominados *tramas* (*frames*) cuyo tamaño es típicamente de unos pocos cientos a unos pocos miles de bytes. Si el paquete recibido de la capa superior es mayor que el tamaño máximo de trama la capa física debe encargarse de fragmentarlo, enviarlo y recomponerlo en el lado opuesto. En caso de que una trama no haya sido transmitida correctamente se deberá enviar de nuevo; también debe haber mecanismos para reconocer cuando una trama se recibe duplicada. Generalmente se utiliza algún mecanismo de control de flujo, para evitar que un transmisor rápido pueda “abrumar” a un receptor lento.

FUNCIONES Y SERVICIOS DEL NIVEL DE ENLACE DE DATOS

Las funciones básicas llevadas a cabo por el nivel de enlace de datos son las siguientes:

- Proporciona los medios funcionales y de procedimiento necesarios para establecer, mantener y liberar conexiones de enlaces de datos entre entidades de red.
- Construye una PDU de nivel de enlace de datos, denominada comúnmente trama, por cada SDU recibida del nivel de red.
- Control de secuencia. El nivel de enlace de datos debe mantener el orden de la secuencia de tramas transmitidas sobre la conexión de enlace de datos.
- Detección de errores. El nivel de enlace de datos debe detectar errores de tipo operacional, de transmisión y de formato que puedan producirse debido a posibles deficiencias del medio físico.
- Control de flujo entre entidades de enlace de datos.

Los servicios que el nivel de enlace de datos proporciona al nivel de red son:

- Establecimiento de conexiones de enlace de datos punto a punto entre entidades de red en sistemas diferentes.
- Información a cerca de los posibles errores irrecuperables detectados.
- Control dinámico de la velocidad de transferencia de las unidades de información mediante técnicas de control de flujo.

Las redes “broadcast” utilizan funciones especiales de la capa de enlace para controlar el acceso al medio de transmisión, ya que éste es compartido por todos los nodos de la red. Esto añade una complejidad a la capa de enlace que no está presente en las redes basadas en líneas punto a punto, razón por la cual en las redes “broadcast” la capa de enlace se subdivide en dos subcapas: la inferior, denominada subcapa MAC (Media Access Control) se ocupa de resolver el problema de acceso al medio, y la superior, subcapa LLC (Logical Link Control) cumple una función equivalente a la capa de enlace en las líneas punto a punto.

Ejemplos de protocolos de la capa de enlace incluyen ISO 7776, la capa de enlace de CCITT X.25, RDSI, LAP-D, ISO HDLC. Como ejemplos de protocolos de la subcapa MAC podemos citar los de IEEE 802.3 (Ethernet), IEEE 802.5 (Token Ring), ISO 9314 (FDDI). El protocolo de subcapa LLC de todas las redes “broadcast” es el IEEE 802.2.

1.5.1.3.- La capa de Red

La capa de red se ocupa del control de la subred. Esta es la capa que tiene “conciencia” de la topología de la red, y se ocupa de decidir por que ruta va a ser enviada la información; la decisión de la ruta a seguir puede hacerse de forma estática, o de forma dinámica en base a información obtenida de otros nodos sobre el estado de la red.

De forma análoga a la capa de enlace la capa de red maneja los bits en grupos discretos que aquí reciben el nombre de *paquetes*; motivo por el cual a veces se la llama la capa de paquete. Los paquetes tienen tamaños variables, pudiendo llegar a ser muy elevados, sobre todo en protocolos recientes, para poder aprovechar eficientemente la elevada velocidad de los nuevos medios de transmisión (fibra óptica, ATM, etc.). Por ejemplo en TCP/IP el tamaño máximo de paquete es de 64 KBytes, pero en el nuevo estándar, llamado IPv6, el tamaño máximo puede llegar a ser de 4 GBytes (4.294.967.296 Bytes).

FUNCIONES Y SERVICIOS DEL NIVEL DE RED

En general, el conjunto de funciones realizadas por el nivel de red son las siguientes:

- Establecimiento, mantenimiento y liberación de conexiones de red entre entidades de transporte a través de las facilidades de comunicación existentes.
- Encaminamiento y reenvío. Puesto que en una conexión de red entre las entidades de red de los sistemas finales pueden estar involucrados uno o varios sistemas intermedios, el nivel de red debe incorporar funciones de reenvío de paquetes entre sistemas intermedios. Los algoritmos de enrutamiento, por otro lado, determinan la ruta más apropiada entre las entidades direccionadas.
- Segmentación y/o Reensamblado. El nivel de red debe poseer la capacidad de segmentar y/o reensamblar las SDUs recibidas del nivel de transporte para facilitar su transferencia.
- Detección de errores. En el nivel de red se utilizan facilidades de detección de errores para comprobar si la calidad de servicio se mantiene en una conexión de red. Esta facilidad de detección de errores del nivel de red hace uso del servicio de notificación de errores proporcionado por el nivel de enlace de datos.
- Secuenciamiento. El nivel de red es capaz de llevar a cabo el secuenciamiento de las unidades de los paquetes en una conexión de datos, de tal forma que sean entregados en su destino en el mismo orden en el que emitieron.
- Control de flujo entre entidades de red.
- Control de congestión para evitar "atascos" en la red.
- En el caso de ofrecer servicios con QoS el nivel de red debe ocuparse de reservar los recursos necesarios para poder ofrecer el servicio prometido con garantías.
- También debe ser capaz de efectuar labores de contabilidad del tráfico en caso necesario (por ejemplo si el servicio se factura en base a la cantidad de datos transmitidos).

Por otro lado, los servicios básicos que el nivel de red es capaz de proporcionar al nivel de transporte son los siguientes:

- Conexiones de red. El nivel de red provee al nivel de transporte de conexiones de red entre entidades de entidades de transporte, haciendo uso de las conexiones de enlace de datos proporcionadas por el nivel de enlace de datos.
- Direcciones de red, necesarias para identificar entidades de transporte.
- Reporte de los errores irrecuperables detectados.

En la capa de red es donde con mas intensidad se observa la distinción entre servicios orientados y no orientados a conexión (CONS vs CLNS). En el curso veremos en detalle las redes ATM, que en el nivel de red dan un servicio de tipo CONS, y las redes TCP/IP, que en el nivel de red dan un servicio

de tipo CLNS.

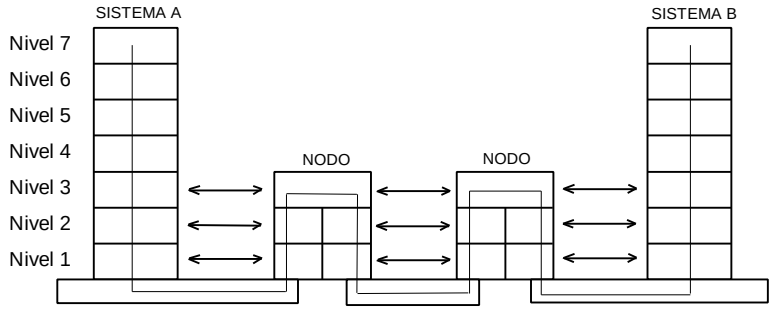

Figura 1.10

Los protocolos de nivel 1 y 2 son protocolos de ámbito local entre cada estación y el nodo de conmutación local al que se encuentra conectado. El protocolo de nivel de red, por el contrario, es el encargado de llevar a cabo las funciones de encaminamiento y reenvío de la información entre los diferentes sistemas intermedios hasta alcanzar su destino final. Se establece un primer diálogo entre la estación emisora y el nodo local al que se conecta, en el cual la estación entregará al nodo las unidades de datos de nivel de red, denominadas paquetes, con la información de direccionamiento que éste requiere para su entrega a través de la red. El nivel de red del nodo local, a partir de la información proporcionada por el nivel de red de la estación emisora, deberá ser capaz de encaminar y conmutar los paquetes a través de la red de conmutación de paquetes hasta alcanzar la estación destino.

Los nodos o sistemas de retransmisión intermedios, normalmente, únicamente incorporan los tres primeros niveles del modelo OSI, ya que su funcionalidad se limita al encaminamiento de la información a través de la red no siendo necesario para ello la utilización de los protocolos de alto nivel que se definen a continuación.

La capa de red es la mas importante en redes de conmutación de paquetes (tales como X.25 o TCP/IP). Algunos ejemplos de protocolos utilizados en la capa de red son los protocolos de nivel de paquete y nivel de pasarela CCITT X.25 y X.75, el IP (Internet Protocol), CCITT/ITU-T Q.931, Q.933, Q.2931, y el OSI CLNP (ConnectionLess Network Protocol).

En las redes de tipo “broadcast” el nivel de red es casi inexistente, ya que desde un punto de vista topológico podemos considerar que en una red “broadcast” los nodos estan interconectados todos con todos, por lo que no se toman decisiones de encaminamiento. Sin embargo veremos que la unión de redes “broadcast” mediante puentes suscita en algunos casos la necesidad de efectuar tareas propias del nivel de red en el nivel de enlace.

1.5.1.4.- La capa de Transporte

La capa de transporte es la primera que se ocupa de comunicar directamente nodos terminales, utilizando la subred como un medio de transporte transparente gracias a los servicios obtenidos de la capa de red. Por esta razón se la ha llamado históricamente la capa host-host. También se suele decir que es la primera capa extremo a extremo.

La capa de transporte actúa de intermediario entre los niveles superiores o niveles orientados a las aplicaciones y los niveles inferiores o niveles dependientes de la red. Su misión básica es la de optimizar los servicios del nivel de red y corregir las posibles deficiencias en la calidad de servicio de los niveles inferiores. De esta forma, el principal propósito del nivel de transporte consiste en proporcionar al nivel de sesión un servicio de transferencia de datos independiente de la red de comunicaciones subyacente.

La principal función de la capa de transporte es fragmentar de forma adecuada los datos recibidos de la capa superior (sesión) para transferirlos a la capa de red, y asegurar que los fragmentos llegan y son recompuestos correctamente en su destino.

En condiciones normales la capa de transporte solicita a la capa de red una conexión diferente por cada solicitud recibida de la capa de sesión, pero puede haber razones de costo que aconsejen multiplexar diferentes conexiones en la capa de sesión sobre una sola conexión en la capa de red o, inversamente, razones de rendimiento pueden requerir que una conexión solicitada por la capa de sesión sea atendida por varias conexiones en la capa de red; en ambos casos la capa de transporte se ocupará de hacer la multiplexación mas adecuada de forma transparente a la capa de sesión.

SERVICIOS PROPORCIONADOS AL NIVEL DE SESION

La capa de transporte establece el tipo de servicio que recibe la capa de sesión, y en último extremo los usuarios. Éste podría ser por ejemplo un servicio libre de errores que entrega los mensajes en el mismo orden en que se envían; también podría ser un servicio de datagramas, es decir, mensajes independientes sin garantía en cuanto al orden de entrega ni confirmación de la misma, o un servicio “broadcast” o multicast en que los paquetes se distribuyen a múltiples destinos simultáneamente.

El control de flujo, que ha aparecido en capas anteriores, es necesario también en la capa de transporte para asegurar que un host rápido no satura a uno lento. La capa de transporte realiza también su propio control de errores, que resulta ahora esencial pues algunos protocolos modernos como Frame Relay o ATM han reducido o suprimido totalmente el control de errores de las capas inferiores, ya que con las mejoras en la tecnología de transmisión de datos éstos son menos frecuentes y se considera mas adecuado realizar esta tarea en el nivel de transporte.

Salvo el caso de transmisiones multicast o “broadcast” el nivel de transporte se ocupa siempre de una comunicación entre dos entidades, lo cual le asemeja en cierto sentido al nivel de enlace; por ello existen similitudes entre ambas capas en cuestiones tales como control de errores o control de flujo.

Tanto los dos tipos de servicios que el nivel de transporte es capaz de proporcionar como las facilidades de direccionamiento y de control de flujo incorporadas en este nivel son muy similares a las proporcionadas por el nivel de red subyacente. A simple vista podría considerarse innecesaria la existencia de uno de estos niveles, ya que ambos parecen proporcionar los mismos servicios. Pero existe una razón crucial que justifica la existencia de ambos. El nivel de red forma parte de la red de comunicaciones y como tal, el servicio que este nivel proporciona depende totalmente del proveedor de la red. Así por ejemplo, si un nivel de red se muestra poco fiable o simplemente el tipo de servicio que proporciona se considera poco eficiente para soportar un determinado tipo de aplicación, los usuarios de dicha red no tienen capacidad para mejorar este servicio, ya que no tienen acceso a los niveles dependientes de la red. La única posibilidad de resolver el problema es definir un nuevo nivel por encima del nivel de red que mejore la calidad del servicio proporcionada por éste. En general la existencia de un nivel de transporte por encima de un nivel de red asegura un servicio de transporte de información más fiable.

Por otra parte, las primitivas de nivel de transporte pueden ser diseñadas de tal forma que sean totalmente independientes de las primitivas del nivel de red, las cuales pueden variar considerablemente de una red a otra. De esta forma es posible la creación de programas de aplicación que corran sobre diferentes tipos de redes, sin tener que preocuparse a cerca de la existencia de interfaces incompatibles debido a la utilización de diferentes primitivas de servicio.

Debido a la capacidad del nivel de transporte de aislar a los niveles superiores de la tecnología, diseño y posibles imperfecciones de la red de comunicaciones, los niveles 1, 2, 3 y 4 del modelo OSI se han definido como los 'proveedores del servicio de transporte', mientras que los niveles 5, 6 y 7 constituirían los 'usuarios del servicio de transporte'.

CALIDAD DE SERVICIO

Como ya hemos visto, el nivel de transporte debe incorporar la funcionalidad necesaria para mejorar la calidad de servicio proporcionada por el nivel de red. Si el servicio que proporciona el nivel de red se considera impecable, la funcionalidad del nivel de transporte puede considerarse casi nula. Si por el contrario el servicio proporcionado por el nivel de red es pobre o insuficiente, el nivel de transporte deberá cubrir las diferencias entre el servicio que los usuarios desean obtener y el servicio que la red es capaz de proporcionar.

La calidad de servicio se caracteriza por un conjunto específico de parámetros. El servicio de transporte OSI permite a los usuarios especificar valores deseables, aceptables e inaceptables de estos parámetros en el momento de establecer la conexión. Algunos de estos parámetros también son aplicables a un servicio de transporte sin conexión. Es responsabilidad del nivel de transporte analizar estos parámetros y, dependiendo del tipo de servicio de red disponible, determinar si es posible proporcionar el servicio requerido. Algunos de los parámetros de Calidad de Servicio más característicos que pueden especificarse son los siguientes:

- Retardo en el establecimiento de la conexión;
- Probabilidad de fallo el establecimiento y liberación de la conexión;
- Rendimiento;
- Probabilidad de fallo en la transferencia;
- Probabilidad de fallo en la liberación de la conexión;
- Prioridad de una conexión.

Cuando un usuario del nivel de transporte solicita el establecimiento de una conexión con unos determinados valores en los parámetros de calidad de servicio, y el nivel de transporte se da cuenta de que todos o algunos de ellos son inalcanzables, informará inmediatamente al usuario que emitió la solicitud del fallo en el establecimiento de la conexión, sin necesidad de establecer ningún tipo de contacto con la parte llamada.

En caso de que no se puedan alcanzar los valores deseados pero sí unos valores aceptables, el nivel de transporte intentará negociar la conexión con el nivel de transporte del extremo llamado. Si éste acepta los valores propuestos se establece la conexión, en caso contrario se rechaza.

CLASES DE PROTOCOLOS DE TRANSPORTE

Cuanto peor es el servicio de red ofrecido, más complicado debe ser el protocolo de transporte. Por este motivo se han definido cinco clases de protocolos de transporte en función de la fiabilidad y otras características del nivel de red subyacente.

- **Clase 0: Protocolos Simples.**

Establecen una conexión de red por cada conexión de transporte solicitada y asumen que el nivel de red está libre de errores. No incorporan mecanismos de secuenciamiento o de control de flujo. Unicamente proporcionan mecanismos para el establecimiento y liberación de conexiones de transporte. Este tipo de protocolos de transporte apenas añaden capacidad alguna al servicio proporcionado por el nivel de red, y son utilizados cuando el error residual es muy bajo, es decir, la proporción del tráfico total enviado que se deteriora, pierde o duplica es muy bajo.

- **Clase 1: Protocolos de recuperación básica de errores.**

Incorporan la misma funcionalidad que los protocolos simples, con la única excepción de que éstos son utilizados en aquellos casos en los que la red es capaz de informar al servicio de transporte de que se ha restablecimiento una conexión. La entidades de transporte involucradas en la conexión restablecida deberán resincronizarse y continuar desde el punto en el que se abandonó. Esta clase de protocolos no incorporan ningún mecanismo de control de errores o de control de flujo adicional al proporcionado por el nivel de red.

- **Clase 2: Protocolos de multiplexación.**

Esta clase de protocolos está formada por protocolos de clase 0 mejorados. Entre las mejoras efectuadas se incluye la posibilidad de multiplexar dos o más conexiones de transporte sobre una misma conexión de red y facilidades individuales de control de flujo para cada conexión de transporte. La capacidad de multiplexación puede resultar muy útil en aquellos casos en los que existen múltiples conexiones de transporte activas y cada una de ellas con un índice de tráfico relativamente bajo.

- **Clase 3: Protocolos de recuperación de errores y de multiplexación.**

Este tipo de protocolos combina las características de los de las clases 1 y 2. Permiten la multiplexación y son capaces de recuperarse de una situación de reestablecimiento de la conexión de red. Adicionalmente incorporan algún mecanismo de control de flujo.

- **Clase 4: Protocolos de detección y recuperación de errores.**

Esta clase de protocolos ha sido definida para su utilización en redes no fiables en las que es posible que se produzca la duplicación o pérdida de paquetes. Deben incorporar mecanismos de detección y corrección de errores, como por ejemplo la inclusión de su propio checksum o número de secuencia. Este tipo de protocolo es el más complejo y puede utilizarse para servicios de red no orientados a la conexión.

Los protocolos de clase 0 y 2 se han diseñado para redes fiables y libres de errores en las que no puede darse el reestablecimiento de las conexiones de red.

Los protocolos de clase 1 y 3 son aplicables a redes con alto grado de fiabilidad pero en las que se puede reestablecer una conexión de red.

Los protocolos de clase 4, por último, están pensados para redes no fiables y en las cuales pueden producirse situaciones de reestablecimiento y de pérdida o duplicación de paquetes.

Ejemplos de protocolos de transporte incluyen el CCITT X.224, también llamado protocolo de transporte OSI TP4 (Transport Protocol 4). En Internet existen dos protocolos de transporte: TCP y UDP.

1.5.1.5.- La capa de Sesión

La capa de sesión es la primera que es accesible al usuario, y es su interfaz más básica con la red. Por ejemplo, mediante los servicios de la capa de sesión un usuario podría establecer una conexión como terminal remoto de otro ordenador. En un sistema multiusuario la capa de sesión se ocupa de ofrecer un SAP a cada usuario para acceder al nivel de transporte.

El nivel de sesión proporciona una funcionalidad novedosa en el modelo de referencia OSI. Ninguna arquitectura de red previa poseía un nivel de sesión o similar. El nivel de sesión es un nivel con una funcionalidad mínima comparada con la de los niveles por debajo de él.

El principal propósito del nivel de sesión es proporcionar a entidades de nivel de presentación un medio de cooperación en el cual organizar y sincronizar su diálogo y gestionar el intercambio de datos. Para ello, el nivel de sesión establece una conexión de sesión en la que impone una estructura para llevar a cabo la interacción y diálogo entre usuarios del nivel de sesión.

El nivel de sesión gestiona el diálogo entre dos procesos de aplicación que se comunican a través de una red de comunicaciones para llevar a cabo alguna actividad, como por ejemplo un login remoto desde un terminal a un ordenador, una transferencia de ficheros, etc.

Aunque en realidad el nivel de sesión proporciona su servicio directamente al nivel de presentación inmediatamente por encima, puesto que la funcionalidad asociada a este nivel de presentación, como veremos más adelante, se limita a ofrecer la información en diferentes formatos, el nivel de sesión puede considerarse íntimamente relacionado con los requerimientos de temporización y diálogo del nivel de aplicación.

Cada conexión del nivel de sesión, denominada simplemente sesión, es mapeada en una conexión de transporte en relación de una a una. No existe capacidad de multiplexación en este nivel.

Las características más importantes del nivel sesión son: el intercambio de datos, la gestión del diálogo, la sincronización, la gestión de actividades y el informe a cerca de excepciones.

· **Intercambio de Datos.**

Una sesión, al igual que una conexión de transporte consta de tres fases: establecimiento, utilización y liberación de la sesión. En la mayoría de los casos, cuando los usuarios del nivel de sesión solicitan el establecimiento de una sesión, el nivel de sesión se limita a invocar el establecimiento de una conexión de transporte sobre la que establecer la sesión.

Una vez establecida una sesión, el intercambio de datos y posterior liberación de dicha sesión deben realizarse de forma ordenada y sin posibilidad de pérdida de información. Así como una conexión de transporte podía finalizar de una forma abrupta y con pérdida de información, una sesión debe finalizar ordenadamente y asegurándose de que no existirá pérdida de datos.

· **Gestión del Diálogo.**

El intercambio de datos en una sesión puede realizarse en ambos sentidos simultáneamente (full duplex) o en ambos sentidos alternativamente (half duplex). La opción elegida depende en gran medida de la aplicación.

Como ejemplo, consideremos un sistema de gestión de bases de datos al que es posible acceder desde terminales remotos. El modo de operación clásico es que un usuario emita una petición al sistema de bases de datos y que espere hasta recibir la correspondiente respuesta. Si previamente a la recepción de la respuesta se permite la emisión de otras peticiones, podrían producirse situaciones indeseadas. Este tipo de aplicaciones debe operar en modo half duplex, de tal forma que en cada momento bien el usuario, o bien el sistema tengan permiso para transmitir, pero nunca ambos al mismo tiempo.

El mecanismo encargado de controlar de quién es el turno para transmitir es lo que se denomina 'gestión del diálogo' y es controlado por medio de un 'testigo'. Cuando se establece una sesión, y se selecciona el modo de operación half duplex, la negociación inicial determina qué lado posee el testigo inicialmente. Unicamente el usuario que posee el testigo puede transmitir, el otro debe permanecer en espera. Cuando el poseedor del testigo a finalizado su transmisión, entregará el testigo a su entidad par para otorgarle el derecho a transmitir.

Si el usuario que no posee el testigo desea transmitir, puede solicitar que se lo entreguen. El poseedor del testigo puede aceptar o negar su entrega.

Si al establecerse la sesión se selecciona un modo de operación full duplex, no es necesaria la utilización de testigos, ya que ambos extremos pueden transmitir cuando lo deseen.

· **Sincronización.**

La sincronización es otro de los servicios proporcionados por el nivel de sesión mediante el cual, es posible hacer que una entidad de sesión retorne a un estado previo debido a la ocurrencia de un error o desacuerdo.

El nivel de transporte ha sido diseñado para hacer frente a errores en la comunicación, pero no contempla la recuperación ante errores producidos en los niveles superiores. La solución a los posibles problemas que puedan surgir en los niveles altos del modelo OSI es responsabilidad del nivel de sesión.

Los usuarios del nivel de sesión pueden, opcionalmente, insertar puntos de sincronización cada cierto tiempo en la secuencia de mensajes entregados al nivel de sesión. En caso de producirse algún problema se deberá retornar el estado de la sesión al punto de sincronización previo a la ocurrencia del problema y continuar a partir de ahí. Para que esta 'resincronización' se realice con éxito, es necesario que el usuario del nivel de sesión sea capaz de mantener la información durante el tiempo necesario hasta su transmisión.

- **Gestión de Actividades.**

Muy relacionado con el servicio de sincronización, el nivel de sesión proporciona el servicio de gestión de actividades. Este servicio permite a los usuarios subdividir la secuencia de mensajes a transmitir en unidades lógicas denominadas 'actividades'. Cada una de estas actividades se considera totalmente independiente del resto de actividades anteriores y posteriores a ella.

Es responsabilidad del usuario determinar el significado específico de una actividad.

Vamos a tomar como ejemplo una sesión que ha sido establecida para la transferencia de varios ficheros entre dos ordenadores. Es necesario disponer de algún sistema que permita delimitar dónde finaliza un fichero y dónde comienza el siguiente. La utilización de un separador ASCII no resulta eficaz ya que dicho separador podría aparecer en la secuencia de información del fichero. Lo que realmente se necesita es alguna forma de insertar una marca en la secuencia de mensajes que sea perfectamente distinguible del resto de la información. La forma más eficiente de realizar esto es considerar cada transferencia de ficheros una actividad independiente.

- **Informe de Excepciones.**

El nivel de sesión incorpora un mecanismo de propósito general que consiste en informar a los usuarios del nivel de sesión a cerca de determinadas situaciones de excepción que hayan podido producirse tanto en el funcionamiento interno del propio nivel de sesión, como en la aplicación de usuario remota.

1.5.1.6.- La capa de Presentación

Hasta aquí nos hemos preocupado únicamente de intercambiar bits (o bytes) entre dos usuarios ubicados en dos ordenadores diferentes. Lo hemos hecho de manera fiable y entregando los datos a la sesión, es decir al usuario, pero sin tomar en cuenta el significado de los bits transportados. El nivel de presentación está relacionado con la forma de representar la información de usuario y cómo esta es codificada para su transmisión. Por ejemplo, si se envía información alfanumérica de un ordenador ASCII a uno EBCDIC será preciso efectuar una conversión, o de lo contrario los datos no serán interpretados correctamente. Lo mismo podríamos decir de la transferencia de datos enteros, flotantes, etc. cuando la representación de los datos difiere en los ordenadores utilizados.

El nivel de presentación se encuentra íntimamente ligado al nivel de sesión debido, principalmente a que el nivel de presentación únicamente proporciona un conjunto muy limitado y específico de servicios al nivel de aplicación, los cuales además no aportan ninguna funcionalidad adicional a los servicios de transferencia de datos proporcionados por los niveles inferiores. Por este motivo el nivel de presentación debe permitir al de aplicación un acceso detallado a los servicios proporcionados por el nivel de sesión.

Cada ordenador posee su forma propia de representar los datos internamente. Por tanto será necesario establecer algún procedimiento que permita el entendimiento entre ordenadores diferentes que se comunican. Es responsabilidad del nivel de presentación preservar el significado de la información independiente de su representación sintáctica en su transcurso a través del medio de comunicación.

De esta forma el nivel de presentación trata con la representación de los datos en cada sistema y con su codificación a la hora de transmitirlos.

REPRESENTACION DE LOS DATOS

Las principales características en la representación de los datos dentro de un sistema son:

- el código de caracteres utilizado: EBCDIC, ASCII, etc.;
- la longitud o número de bits de que consta cada caracter;
- cómo se representan los números: complemento a uno o complemento a dos;
- la forma en que se ordenan los bytes en los números enteros y reales;

- la forma de almacenar los caracteres de un string; etc.

No siempre es posible cambiar la posición de los bytes para conseguir la representación utilizada en otro sistema.

En una comunicación punto a punto, la información se representa de tres formas diferentes: una en cada sistema final, y otra la que adopta durante su transferencia.

Para llevar a cabo el intercambio de un determinado tipo de información, los usuarios deben acordar la utilización de una determinada **Sintaxis Abstracta** que describa los requerimientos de la información de forma genérica.

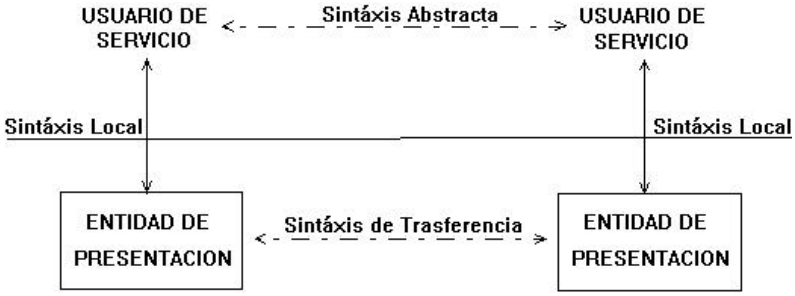

Figura 1.11

La información a transmitir debe ser codificada en una secuencia de bits estructurada según una Sintaxis de Transferencia. Así por ejemplo una sintaxis de transferencia podría establecer que una cadena de caracteres debe representarse mediante el código hexadecimal '06' seguido de un octeto que contenga el número de caracteres de que consta la cadena, y seguido de cada caracter de la cadena codificado en ASCII sin paridad.

Para poder llevar a cabo una transferencia de datos con éxito, debe definirse una relación entre la sintaxis abstracta y la sintaxis de transferencia, relación denominada 'contexto de presentación'. Una aplicación puede requerir la utilización de diversos contextos de presentación para soportar todos los tipos de información que desea transmitir. Todo el conjunto de contextos de presentación utilizados por una aplicación en un momento determinado se denomina 'conjunto de contexto'.

En resumen, el nivel de presentación negocia las sintaxis de transferencia a petición del usuario, y lleva a cabo la transformación de los datos de usuario en y desde esta sintaxis de transferencia.

CODIFICACION DE LOS DATOS

Existen dos conceptos importantes dentro del procedimiento de codificación de los datos: la compresión y la encriptación.

La **compresión** de los datos es una técnica que intenta reducir la cantidad de información transmitida con el principal objetivo de ahorrar costes y espacio. Así por ejemplo para transmitir un número entero de 32 bits, es posible simplemente codificarlo en cuatro octetos según una forma de representación determinada y transmitirlo de esta forma directamente. Sin embargo si podemos deducir que la gran mayoría de los enteros transmitidos toman valores entre 0 y 255, podría resultar más económico transmitir cada entero codificado en un solo octeto sin signo, y utilizar el código 255 para identificar que a continuación se transmite un entero de 32 bits completo. A pesar de que cada cierto tiempo se necesitan cinco octetos para transmitir un entero, la utilización de un solo octeto para cada entero la mayor parte del tiempo resulta ventajosa.

La **encriptación** de los datos, por otro lado, es una técnica de seguridad y privacidad de la información transmitida. En teoría la encriptación de la información transmitida podría realizarse en cualquiera de los niveles OSI, pero finalmente se consideró oportuna su inclusión dentro del nivel de presentación.

Existen muchos y muy variados métodos de encriptación de la información, cuya utilización permiten al nivel de presentación proporcionar una serie de servicios de seguridad tales como:

- Protección de los datos contra su lectura por personal no autorizado;
- Prevención de inserción y borrado de información por personal no autorizado;
- Verificación del emisor de cada mensaje;
- Permitir el envío de mensajes electrónicos firmados por el usuario.

1.5.1.7.- La capa de Aplicación

La capa de aplicación comprende los servicios que el usuario final está acostumbrado a utilizar en una red telemática, por lo que a menudo los protocolos de esta capa se denominan *servicios*.

El nivel de aplicación es el nivel más alto del modelo de referencia OSI. Es en este nivel donde se ubican los programas de usuario o aplicaciones que, haciendo uso de todo el conjunto de servicios que los niveles por debajo de él le proporcionan, se comunican con otras aplicaciones remotas.

Cada usuario puede definir sus propias aplicaciones sobre el resto de los niveles OSI, pero también en este nivel se han definido estándares que además de favorecer el concepto de interconexión total perseguido por OSI, evita que aplicaciones de uso muy común sean elaboradas por diferentes usuarios dando lugar a versiones incompatibles de una misma aplicación.

Dado que se crean continuamente nuevos servicios, existen muchos protocolos para la capa de aplicación, uno o mas por cada tipo de servicio. Ejemplos de protocolos estándar de la capa de aplicación son el CCITT X.400, X.420, X.500, FTAM. SMTP, FTP, HTTP, etc.

1.5.2.- Transmisión de datos en el modelo OSI

La transmisión de datos en el modelo OSI se realiza de forma análoga a lo ya descrito para el modelo de capas. La capa de aplicación recibe los datos del usuario y les añade una cabecera (que denominamos cabecera de aplicación), constituyendo así la PDU (Protocol Data Unit) de la capa de aplicación. La cabecera contiene información de control propia del protocolo en cuestión. La PDU es transferida a la capa de aplicación en el nodo de destino, la cual recibe la PDU y elimina la cabecera entregando los datos al usuario. En realidad la PDU no es entregada directamente a la capa de aplicación en el nodo de destino, sino que es transferida a la capa de presentación en el nodo local a través de la interfaz; esto es una cuestión secundaria para la capa de aplicación, que ve a la capa de presentación como el instrumento que le permite hablar con su homóloga en el otro lado.

A su vez la capa de presentación recibe la PDU de la capa de aplicación y le añade una cabecera propia, (cabecera de presentación) creando la PDU de la capa de presentación Esta PDU es transferida a la capa de presentación en el nodo remoto usando a la capa de sesión como instrumento para la comunicación, de manera análoga a lo ya descrito para la capa de aplicación.

En el caso mas general cada capa añade una cabecera propia a los datos recibidos de la capa superior, y construye así su PDU. La capa homóloga del nodo de destino se ocupará de extraer dicha cabecera, interpretarla, y entregar la PDU correspondiente a la capa superior. En algunos casos la cabecera puede no existir. En el caso particular de la capa de enlace además de la cabecera añade una cola al construir la PDU (trama) que entrega a la capa física.

Volviendo por un momento a nuestra analogía de los dos ejecutivos que intercambian un documento, vemos que a medida que vamos descendiendo capas en el envío (jefe, secretaria, motorista, líneas aéreas) el documento va recibiendo nuevos envoltorios que contienen a los anteriores. A la llegada el paquete es procesado de forma *simétrica*, es decir se le va quitando en cada capa el envoltorio correspondiente antes de pasarlo a la siguiente.

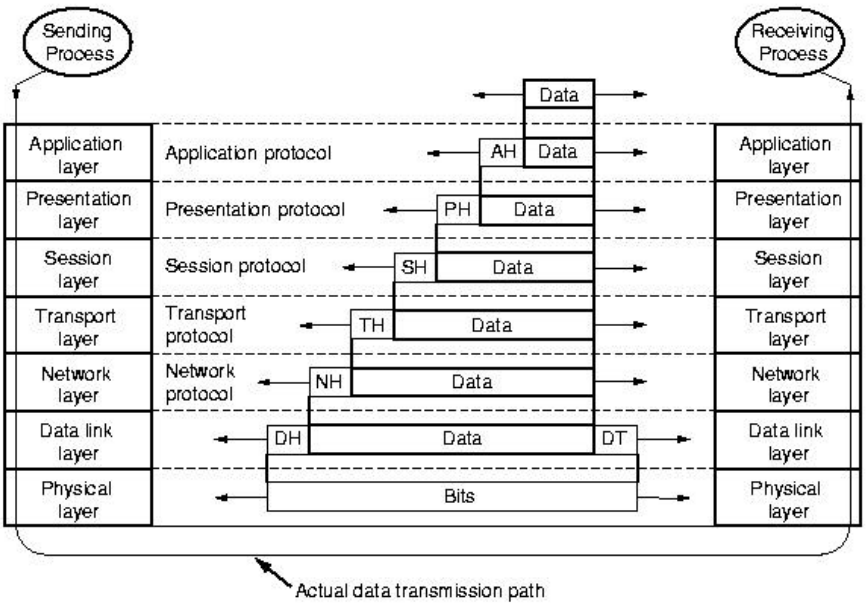

Figura 1.12

1.5.3.- El modelo de referencia TCP/IP

En 1969 la agencia ARPA (Advanced Research Projects Agency) del Departamento de Defensa (DoD, Department of Defense) de los Estados Unidos inició un proyecto de interconexión de ordenadores mediante redes telefónicas. Al ser un proyecto desarrollado por militares en plena guerra fría un principio básico de diseño era que la red debía poder resistir la destrucción de parte de su infraestructura (por ejemplo a causa de un ataque nuclear), de forma que dos nodos cualesquiera pudieran seguir comunicados siempre que hubiera alguna ruta que los uniera. Esto se consiguió en 1972 creando una red de conmutación de paquetes denominada ARPAnet, la primera de este tipo que operó en el mundo. La conmutación de paquetes unida al uso de topologías malladas mediante múltiples líneas punto a punto dio como resultado una red altamente fiable y robusta.

La ARPAnet fue creciendo paulatinamente, y pronto se hicieron experimentos utilizando otros medios de transmisión de datos, en particular enlaces por radio y vía satélite; los protocolos existentes tuvieron problemas para interoperar con estas redes, por lo que se diseñó un nuevo conjunto o pila de protocolos, y con ellos una arquitectura. Este nuevo conjunto se denominó TCP/IP (Transmission Control Protocol/Internet Protocol) nombre que provenía de los dos protocolos mas importantes que componían la pila; la nueva arquitectura se llamó sencillamente *modelo TCP/IP*, los nuevos protocolos fueron especificados por vez primera por Cerf y Kahn en un artículo publicado en 1974. A la nueva red, que se creó como consecuencia de la fusión de ARPAnet con las redes basadas en otras tecnologías de transmisión, se la denominó Internet.

La aproximación adoptada por los diseñadores del TCP/IP fue mucho más pragmática que la de los autores del modelo OSI. Mientras que en el caso de OSI se emplearon varios años en definir con sumo cuidado una arquitectura de capas donde la función y servicios de cada una estaban perfectamente definidas, y solo después se planteó desarrollar los protocolos para cada una de ellas, en el caso de TCP/IP la operación fue a la inversa; primero se especificaron los protocolos, y luego se definió el modelo como una simple descripción de los protocolos ya existentes. Por este motivo el modelo TCP/IP es mucho más simple que el OSI. También por este motivo el modelo OSI se utiliza a menudo para describir otras arquitecturas, como por ejemplo la TCP/IP, mientras que el modelo TCP/IP nunca suele emplearse para describir otras arquitecturas que no sean la suya propia.

En el modelo TCP/IP se pueden distinguir cuatro capas:

1. La capa host-red
2. La capa internet
3. La capa de transporte
4. La capa de aplicación

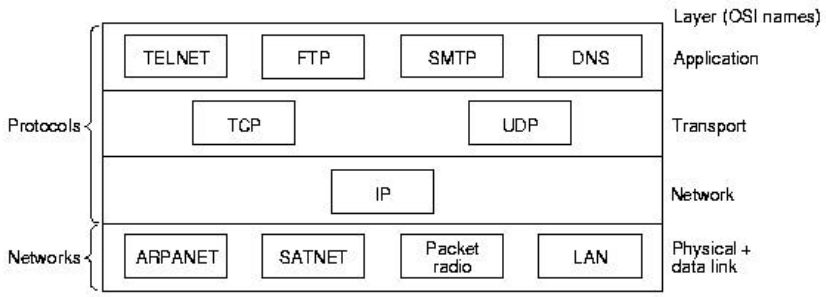

Figura 1.12

1.5.3.1.- La capa host-red (Física + Enlace de Datos)

Esta capa engloba realmente las funciones de la capa física y la capa de enlace del modelo OSI. El modelo TCP/IP no dice gran cosa respecto a ella, salvo que debe ser capaz de conectar el host a la red por medio de algún protocolo que permita enviar paquetes IP. Podríamos decir que para el modelo TCP/IP esta capa se comporta como una “caja negra”. Cuando surge una nueva tecnología de red (por ejemplo ATM) una de las primeras cosas que aparece es un estándar que especifica de que forma se pueden enviar sobre ella paquetes IP; a partir de ahí la capa internet ya puede utilizar esa tecnología de manera transparente.

1.5.3.2.- La capa Internet (Red)

Esta capa es el “corazón” de la red. Su papel equivale al desempeñado por la capa de red en el modelo OSI, es decir, se ocupa de encaminar los paquetes de la forma mas conveniente para que lleguen a su destino, y de evitar que se produzcan situaciones de congestión en los nodos intermedios. Debido a los requisitos de robustez impuestos en el diseño, la capa internet da únicamente un servicio de conmutación de paquetes no orientado a conexión. Los paquetes pueden llegar desordenados a su destino, en cuyo caso es responsabilidad de las capas superiores en el nodo receptor la reordenación para que sean presentados al usuario de forma adecuada.

A diferencia de lo que ocurre en el modelo OSI, donde los protocolos para nada intervienen en la descripción del modelo, la capa internet define aquí un formato de paquete y un protocolo, llamado IP (Internet Protocol), que se considera el protocolo “oficial” de la arquitectura.

1.5.3.3.- La capa de Transporte

Esta capa recibe el mismo nombre y desarrolla la misma función que la cuarta capa del modelo OSI, consistente en permitir la comunicación extremo a extremo (host a host) en la red. Aquí se definen dos protocolos: el TCP (Transmission Control Protocol) ofrece un servicio CONS fiable, con lo que los paquetes (aquí llamados mensajes) llegan ordenados y sin errores. TCP se ocupa también del control de flujo extremo a extremo, para evitar que por ejemplo un host rápido sature a un receptor mas lento. Ejemplos de protocolos de aplicación que utilizan TCP son el SMTP (Simple Mail Transfer Program, correo electrónico) y el FTP (File Transfer Program).

El otro protocolo de transporte es UDP (User Datagram Protocol) que da un servicio CLNS, no fiable. UDP no realiza control de errores ni de flujo. Una aplicación típica donde se utiliza UDP es la transmisión de voz y vídeo en tiempo real; aquí el retardo que introduciría el control de errores

produciría mas daño que beneficio: es preferible perder algún paquete que retransmitirlo fuera de tiempo. Otro ejemplo de aplicación que utiliza UDP es el NFS (Network File System); aquí el control de errores y de flujo se realiza en la capa de aplicación.

1.5.3.4.- La capa de Aplicación

Esta capa desarrolla las funciones de las capas de sesión, presentación y aplicación del modelo OSI. La experiencia ha demostrado que las capas de sesión y presentación son de poca utilidad, debido a su escaso contenido, por ello, la aproximación adoptada por el modelo TCP/IP parece mas acertada.

La capa de aplicación contiene todos los protocolos de alto nivel que se utilizan para ofrecer servicios a los usuarios. Entre estos podemos mencionar tanto los “tradicionales”, que existen desde que se creó el TCP/IP: terminal virtual (TelNet), transferencia de ficheros (FTP), correo electrónico (SMTP) y servidor de nombres (DNS), como los mas recientes, como el servicio de news (NNTP), el Web (HTTP), el Gopher, etc.

1.5.4.- Comparación de los modelos OSI y TCP/IP

Como ya hemos comentado, la génesis del modelo OSI y TCP/IP fue muy diferente. En el caso de OSI primero fue el modelo y después los protocolos, mientras que en TCP/IP el orden fue inverso. Como consecuencia de esto el modelo OSI es mas elegante y esta menos condicionado por ningún protocolo en particular, y se utiliza profusamente como modelo de referencia para explicar todo tipo de redes. El modelo OSI hace una distinción muy clara entre servicios, interfaces y protocolos, conceptos que a menudo se confunden en el modelo TCP/IP. Podríamos decir que la arquitectura (o el modelo) OSI es mas modular y académico que el TCP/IP.

Pero este mayor nivel de abstracción también tiene sus inconvenientes. Los diseñadores del modelo OSI no tenían experiencia práctica aplicando su modelo para desarrollar protocolos y olvidaron algunas funcionalidades importantes. Por ejemplo, las redes “broadcast” no fueron previstas inicialmente en la capa de enlace, por lo que se tuvo que insertar a la fuerza la subcapa MAC para incluirlas. Otro problema era que no se había previsto la interconexión de redes diferentes, cosa que fue como ya hemos visto el *alma mater* del modelo TCP/IP.

El modelo OSI tiene siete capas, mientras que el modelo TCP/IP sólo tiene cuatro. Aunque es desafortunada la fusión de la capa física y la de enlace en una oscura capa host-red, la fusión de las capas de sesión, presentación y aplicación en una sola en el modelo TCP/IP es claramente mas lógica que la del modelo OSI.

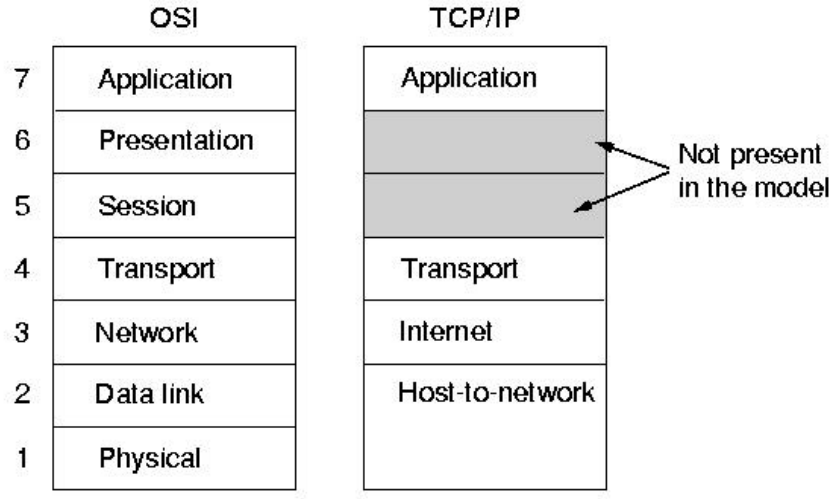

Figura 1.13

Otra diferencia fundamental estriba en los servicios orientados a conexión (CONS) o no orientados a conexión (CLNS). El modelo OSI soporta ambos modos en la capa de red, pero sólo el modo CONS en la capa de transporte, que es la que percibe el usuario. El modelo TCP/IP en cambio soporta solo CLNS en la capa de red, pero ambos en la de transporte. Quizá un sutil detalle pueda explicar esta diferencia: el servicio CONS a nivel de red hace mucho mas sencillo facturar por tiempo de conexión, cosa a la que están muy acostumbradas las compañías telefónicas, que son las que han participado activamente en los comités técnicos de ISO que diseñaron el modelo OSI.

En la práctica los protocolos basados en las normas estándar OSI definidas por la ISO nunca llegaron a tener gran relevancia a nivel mundial, a pesar de que la mayoría de los grandes fabricantes de ordenadores y compañías telefónicas impulsaron su utilización ofreciendo productos y servicios basados en ellos. Las razones principales que motivaron este fenómeno las podemos resumir en los siguientes puntos:

Momento inadecuado: Para cuando estaban disponibles productos comerciales basados en protocolos OSI (finales de los ochenta) ya estaban ampliamente difundidos los productos basados en los protocolos TCP/IP; esto era especialmente cierto en entornos académicos (universidades y centros de investigación), que aunque económicamente no eran los mejor dotados sí tenían las mayores redes a nivel mundial.

Tecnología inapropiada: como ya hemos comentado la elección del modelo de siete capas para el protocolo OSI era algo forzada. Una de las razones que llevaron a elegir este número de capas era que coincidía con el del modelo SNA de IBM, que dominaba el mercado de la informática por aquel entonces; los autores del modelo OSI creían que aproximándose a SNA tenían mayores posibilidades de éxito. La complejidad de la arquitectura OSI (análogamente a la SNA) es considerable, y en muchos aspectos difícil de traducir en programas.

Implementaciones inadecuadas: en parte como consecuencia de su complejidad, los productos comerciales que aparecían basados en los protocolos OSI eran muy caros y poco fiables. Esto creó un círculo vicioso, ya que al ser caros los usuarios no los compraban, y al no usarse en condiciones reales los nuevos productos no se depuraban; además, las empresas fabricantes tenían que mantener un alto precio del software OSI para compensar los elevados costos de desarrollo y mantenimiento. Como contraste una de las primeras implementaciones de TCP/IP formaba parte del UNIX de Berkeley, era muy buena y además se distribuía gratuitamente. No es extraño pues que rápidamente se asociara OSI con baja calidad, complejidad y costos elevados.

Mala política: el desarrollo de OSI era patrocinado principalmente por la ISO, las PTTs europeas, la Comunidad Europea y los gobiernos de sus países miembros; las decisiones eran fruto de multitud de reuniones de los diversos comités y grupos de trabajo, y en ocasiones se tomaban en consideración no sólo aspectos técnicos sino también políticos, buscando el compromiso entre sus miembros. Por el contrario el desarrollo de TCP/IP seguía un curso mucho más improvisado e informal, cualquier persona podía (y puede) proponer un nuevo protocolo para su estandarización independientemente de su nacionalidad, prestigio o situación laboral. Haciendo una simplificación podríamos decir que OSI funcionaba como una “democracia parlamentaria” (similar a un gobierno moderno), mientras que TCP/IP era más similar a una ONG, o a un movimiento alternativo; esto se reflejaba incluso en la indumentaria utilizada por uno y otro colectivo. No es de extrañar que en entornos académicos (de nuevo recordemos los más avanzados en redes globales) se viera con mucha más simpatía el mecanismo de estandarización del TCP/IP que el de OSI.

Aunque por la exposición anterior pueda parecer lo contrario, también existen aspectos negativos en los protocolos TCP/IP. Por un lado no se distinguen claramente los conceptos de servicio, interfaz y protocolo. En segundo lugar, el “modelo” TCP/IP fue diseñado con posterioridad al protocolo, intentando imitar la labor de síntesis que se había hecho en el modelo OSI (podríamos decir que es como si se hubieran cortado los patrones después de cosido el traje). En tercero esta la “caja negra” que hemos llamado capa host-red y que en el modelo TCP/IP es mas bien una interfaz que una capa, ya que lo único que se especifica de ella es que ha de ser capaz de transmitir paquetes IP. Como consecuencia de esto el modelo TCP/IP no distingue entre la capa física y la de enlace, ya que ambas entran en la “capa” host-red.

Por otro lado, aun cuando los protocolos IP y TCP fueron diseñados concienzudamente y bien implementados, algunos protocolos, especialmente del nivel de aplicación, fueron el resultado de una improvisación para resolver un problema concreto; como las implementaciones se distribuían después de forma gratuita se extendían con rapidez por lo que resultaban difíciles de sustituir; un ejemplo de esto lo tenemos en el protocolo TelNet que se utiliza ampliamente a pesar de no tener soporte para interfaz gráfica, ratón, etc.

Durante la década de los ochenta en Europa las redes académicas de la mayoría de los países (incluido España) utilizaban protocolos OSI por imposición de los respectivos gobiernos y de la Comunidad Europea; a la vista de los problemas ya mencionados de los productos OSI, y la extensión y buen resultado de los protocolos TCP/IP, se empezaron a ofrecer en 1991 servicios basados en TCP/IP, lo cual provocó su inmediata difusión por toda Europa y el estancamiento y casi desaparición de los servicios basados en protocolos OSI.

En la tabla siguiente hacemos un resumen del modelo y los principales protocolos de cada capa.

Capa	Protocolo
Aplicación	TCP/IP (DNS, SNMP, SNMP, NNTP, HTTP)
Transporte	TCP/IP (TCP, UDP) ATM (1L1, 1L2, 1L3/4, 1L5)
Red	TCP/IP (IP, ICMP, ARP, RARP, OSPF, BGP, IPv6), ATM (Q2931)
Enlace	ISO(HDLC), TCP/IP (SLIP, PPP), ATM, LANs
Física	N-ISDN, B-ISDN (ATM), GSM, SONET/SDH, LANs Cable coaxial, cable UTP, fibra óptica, microondas, radioenlaces, satélite

Tabla 1.2

1.6.- SERVICIOS DE TRANSMISIÓN DE DATOS DE ÁREA EXTENSA

Dado que cualquier usuario puede solicitar un acceso a las redes que operan las compañías telefónicas, a éstas se las denomina redes públicas de datos (PDN, Public Data Networks). Cuando se desea interconectar ordenadores o redes locales ubicadas a cierta distancia es preciso normalmente utilizar los servicios de alguna de esas redes públicas. Dichos servicios pueden clasificarse de acuerdo con el tipo de conexión que ofrecen, permanente o temporal, y con el tipo de circuito, real o virtual. Esquemáticamente sería:

Tipo de circuito	Tipo de conexión	
	Permanente	Temporal
Real	Líneas dedicadas	Redes de conmutación de circuitos (RTB, RDSI, GSM)
Virtual	Redes de conmutación con PVCs (X.25, Frame Relay, ATM)	Redes de conmutación con SVCs (X.25, Frame Relay, ATM)

Tabla 1.3

En la práctica suele utilizarse en cada caso el servicio mas conveniente por sus prestaciones y precio, por lo que las redes suelen mezclar varios de los servicios que hemos mencionado. Vamos a dar una pequeña descripción de cada uno de ellos.

1.6.1.- Líneas dedicadas

La solución mas simple para una red es el circuito real permanente, constituido por lo que se conoce como *líneas dedicadas o líneas alquiladas (leased line*s); está formado por un enlace punto a punto abierto de forma permanente entre los ordenadores o routers que se desean unir. Una línea dedicada es únicamente un medio de transmisión de datos a nivel físico, todos los protocolos de niveles superiores han de ser suministrados por el usuario.

La red ARPAnet que hemos visto anteriormente se constituyó mediante líneas dedicadas. La Internet incorpora actualmente todos los servicios que hemos mencionado.

Normalmente no es posible contratar una línea dedicada de una velocidad arbitraria, existen unas velocidades prefijadas que son las que suelen ofrecer las compañías telefónicas y que tienen su origen en la propia naturaleza del sistema telefónico, como veremos más adelante. Por ejemplo Telefónica de España ofrece líneas dedicadas de las siguientes velocidades: 9.6, 64, 128, 192, 256, 512 y 2.048 Kbps. El precio de una línea dedicada es una cuota fija mensual que depende de la velocidad y de la distancia entre los dos puntos que se unen.

En las líneas dedicadas la capacidad contratada está reservada de forma permanente en todo el trayecto. Su costo es elevado y por tanto su instalación generalmente sólo se justifica cuando el uso es elevado (al menos tres o cuatro horas al día). Por este motivo las líneas dedicadas no suelen utilizarse en casos en que se necesita una conexión de forma esporádica, por ejemplo una oficina que requiere conectarse unos minutos al final del día para transferir unos ficheros, o un usuario doméstico que se conecta a Internet en los ratos de ocio.

Para mostrar el elevado consumo de recursos que representan las líneas dedicadas pondremos un ejemplo: supongamos que la empresa X con sede central en Bilbao ha abierto treinta sucursales en distintos puntos de España, y necesita que los ordenadores de las sucursales comuniquen con la sede central todos los días durante treinta minutos cada uno para transferir 2 MBytes de información. Para esto la empresa solicita 30 líneas dedicadas de 64 Kbps a la compañía telefónica, y constituye una red con topología de estrella. Aunque cada línea se utiliza únicamente el 2% del tiempo con una eficiencia del 14% el ancho de banda está reservado en su totalidad de forma permanente. Además, se requieren treinta interfaces físicos en el servidor, lo cual lo encarece y complica bastante.

1.6.2.- Conmutación de circuitos

La *conmutación de circuitos* supone una utilización mejor de los recursos que las líneas dedicadas, ya que la conexión extremo a extremo sólo se establece durante el tiempo necesario. Para la transmisión de datos mediante conmutación de circuitos se utiliza la misma red que para la transmisión de la voz, mediante módems o adaptadores apropiados. Genéricamente se la denomina Red Telefónica Conmutada (RTC) o PSTN (Public Switched Telephone Network) y comprende en realidad tres redes diferentes:

- La Red de Telefonía Básica (RTB) también llamada POTS (Plain Old Telephone Service); Está formada por las líneas analógicas tradicionales y por tanto requiere el uso de módems; la máxima velocidad que puede obtenerse en este tipo de enlaces es de 33.6 Kbps (recientemente han aparecido en el mercado módems capaces de comunicar a 56 Kbps por líneas analógicas si se dan ciertas condiciones).

- La Red Digital de Servicios Integrados (RDSI) también llamada ISDN (Integrated Services Digital Network). Está formada por enlaces digitales hasta el bucle de abonado, por lo que el circuito se constituye de forma digital extremo a extremo. La velocidad por circuito es de 64 Kbps, pudiendo con relativa facilidad agregarse varios circuitos (llamados canales) en una misma comunicación para obtener mayor ancho de banda.

- La Red GSM (Global System for Mobile communications). Se trata de conexiones digitales, como en el caso de la RDSI, pero por radioenlaces. La capacidad máxima de un circuito GSM cuando se transmiten datos es de 9.6 Kbps.

La RDSI apareció en España hacia 1994, y la red GSM hacia 1995. Dado que hasta fechas recientes el único sistema de RTC era la RTB a menudo se utilizan ambos términos indistintamente para indicar la red telefónica analógica. Para evitar confusiones conviene usar sólo el término RTB al referirse a la red telefónica analógica, y reservar el término RTC para referirnos al conjunto de todas las redes conmutadas existentes, ahora o en el futuro.

En el caso de la RTC los equipos se conectan a la red pública y en principio cualquier equipo puede comunicar con cualquier otro, siempre que conozca su dirección (número de teléfono). Podemos ver la RTC como una gran nube a la que se conectan multitud de usuarios. Una vez establecido un circuito en RTC la función que éste desempeña para los protocolos de nivel superior es equivalente a la de una línea dedicada.

Telefónica dispone de los tres tipos de RTC (RTB, RDSI y GSM), con tarificación por tiempo de conexión. En el caso de RTB y RDSI se aplica una tarificación con cuatro ámbitos: metropolitano, provincial, nacional e internacional (éste último depende del país). Con la liberalización han aparecido nuevos operadores de telefonía, que en algunos casos disponen de red propia (Retevisión, Euskaltel, ...) y en otros utilizan la redes de dichos operadores para encaminar el tráfico de los usuarios; esto ha supuesto, además una dinamización de la oferta existente haciéndola tan rica que resulta imposible recogerla resumidamente. En el caso de la red GSM (conocida como MoviStar) hay sólo dos ámbitos: nacional e internacional. Existen otras redes GSM operadas por Airtel, Retevisión, y Euskaltel.

Es posible la interconexión entre ordenadores de redes diferentes (RDSI, RTB o GSM); en cuyo caso la velocidad de transmisión será igual a la mas lenta de las conexiones implicadas; en algunos casos puede ser necesario disponer de equipos o contratar servicios especiales.

Siguiendo con nuestro ejemplo anterior de la empresa X, en vez de líneas dedicadas se podría haber utilizado la red telefónica conmutada (por ejemplo la RDSI). En este caso el costo de cada conexión es normalmente menor, ya que sólo se paga por el tiempo que se esta utilizando. Además, la sede central podría contratar menos de treinta enlaces si se planifica un horario escalonado de conexión de las sucursales, o si simplemente se considera que la probabilidad de que todas llamen a la vez es muy reducida. Esto se conoce como *sobresuscripción* ("oversubscription") o *sobrereserva* (“overbooking") y es algo muy normal en redes cuando el número de usuarios es razonablemente elevado y se puede jugar con el factor estadístico. Por ejemplo, supongamos que inicialmente la sede central contrata diez accesos y observa que solo durante el 0,1% del tiempo están todos utilizados; entonces se puede afirmar que el servicio tiene una disponibilidad del 99,9%, es decir, el 99,9% del tiempo hay líneas libres para recibir llamadas de las sucursales; a la vista de esto la empresa puede decidir si aumenta o reduce el número de accesos, según la disponibilidad que se quiera tener y el costo de cada acceso (aquí además del costo de la compañía telefónica se deberá tener en cuenta el de las interfaces , módems y equipo auxiliar).

1.6.3.- Conmutación de paquetes

Con la conmutación de circuitos hemos avanzado en el aprovechamiento de la infraestructura. Sin embargo nos encontramos aún con tres inconvenientes:

- En ocasiones no podremos establecer la conexión por no haber circuitos libres, salvo que contratemos un número de circuitos igual al máximo número posible de conexiones simultáneas, lo cual sería muy costoso.

- Que un circuito se esté utilizando no garantiza que se esté aprovechando el ancho de banda que tiene asignado; en nuestro ejemplo cada sucursal está conectada 30 minutos para enviar 2 MBytes de información, que cual supone un aprovechamiento del 14% suponiendo que se trata de conexiones de 64 Kbps.

- El servidor ha de tener una conexión física por cada circuito, aun cuando la ocupación media sea reducida.

Para evitar estos inconvenientes se crearon redes en las que el usuario puede mantener una única conexión física a la red, y sobre ella varios *circuitos virtuales* con equipos remotos. De esta forma podemos dotar a nuestro ordenador central de treinta circuitos virtuales, con lo que las sucursales siempre van a encontrar un circuito libre sobre el cual establecer la conexión. Al mantener un solo enlace físico el costo de las interfaces, módems, etc., es fijo e independiente del número de circuitos virtuales utilizados. Lógicamente al tener el ordenador central que atender a todas las conexiones por el mismo enlace físico sería conveniente (aunque no necesario) incrementar la velocidad de este; en nuestro ejemplo con conexiones el 2% del tiempo y con un tráfico medio del 14%; para las 30 oficinas agregadas nos daría una ocupación media del 8,4% (0.02x0.14x30) suponiendo un reparto homogéneo (cosa poco probable); como previsiblemente muchas oficinas querrán conectar mas o menos a la misma hora sería conveniente ampliar el enlace del servidor a 128 o 256 Kbps para evitar congestión en horas punta.

Para poder definir circuitos virtuales es preciso disponer de equipos inteligentes en la red que puedan hacer la distribución de los paquetes en función de su destino. Por esto a las redes que permiten crear circuitos virtuales se las denomina redes de *conmutación de paquetes*, y en cierto sentido podemos considerarlas como la evolución lógica de las redes de conmutación de circuitos. En realidad existen dos tipos de redes de conmutación de paquetes, según ofrezcan servicios orientados a conexión o no orientados a conexión (envío de datagramas). La primera red de conmutación de paquetes que existió fue como ya hemos visto ARPAnet, pero como no era orientada a conexión no se adaptaba bien a un servicio de compañía telefónica. Para facilitar la facturación las redes públicas de conmutación de paquetes suelen ofrecer servicios orientados a conexión en el nivel de red. Actualmente hay tres tipos de redes públicas de conmutación de paquetes: X.25, Frame Relay y ATM, y todos ofrecen servicios orientados a conexión. Las tres representan implementaciones bastante completas de los tres primeros niveles del Modelo de Referencia OSI, y tienen muchos puntos en común, según veremos a continuación.

La subred de una red de conmutación de paquetes se constituye mediante conmutadores unidos entre sí por líneas dedicadas. La distribución de los conmutadores y la forma como éstos se unen entre sí (es decir la topología de la red) es algo que decide el proveedor del servicio y que fija la carga máxima que la red podrá soportar en lo que se refiere a tráfico entre conmutadores; la topología fija también la fiabilidad de la red, es decir cuan resistente será a fallos de los enlaces (por ejemplo una red muy mallada será muy resistente). Cuando un usuario desea conectar un equipo a la red el acceso se hace normalmente mediante una línea dedicada entre el equipo a conectar y el conmutador mas próximo del proveedor de servicio (normalmente la Compañía Telefónica). La velocidad de la conexión entre el equipo y el conmutador establece de entrada un máximo a las prestaciones que ese usuario podrá obtener de la red. Puede haber además otras limitaciones impuestas por la capacidad de la red, por saturación o porque se hayan impuesto limitaciones de acuerdo con lo contratado por el usuario con el proveedor del servicio.

Aunque estamos considerando el caso en que la red de conmutación de paquetes la gestiona una compañía Telefónica (con lo que tenemos una red pública de conmutación de paquetes), también es posible que una organización o conjunto de organizaciones (por ejemplo una gran empresa, una administración o un conjunto de universidades) establezcan una red privada basada en X.25, Frame Relay o ATM. En este caso normalmente la gestión de la red se asigna a algún grupo especializado (por ejemplo el departamento de comunicaciones en el caso de la empresa) que se ocupa de diseñar topología, solicitar los enlaces correspondientes, instalar los conmutadores, etc. Si se desea que la red privada esté interconectada con la red pública es preciso prever que al menos uno de los conmutadores de la red privada esté conectado con la red pública. Desde el punto de vista técnico ambas redes son equivalentes en su funcionamiento, salvo que normalmente en una red privada o no se tarifica la utilización, por lo que el control no es tan crítico.

En X.25, Frame Relay y ATM existe el concepto de circuito virtual (VC), que puede ser de dos tipos: conmutado o SVC (Switched Virtual Circuit) y permanente o PVC (Permanent Virtual Circuit). El conmutado se establece y termina a petición del usuario, mientras que el permanente tiene que ser definido por el proveedor del servicio, mediante configuración en los conmutadores a los que se conectan los equipos implicados, normalmente mediante modificación contractual con el cliente. En cierto modo es como si los PVCs fueran “líneas dedicadas virtuales” mientras que los SVCs son como conexiones RTC “virtuales”.

1.6.3.1.- X.25

X.25 fue el primer protocolo estándar de red de datos pública. Se definió por primera vez en 1976 por la CCITT (Comité Consultatif International Télégraphique and Téléphonique). Aunque el protocolo ha sido revisado múltiples veces (la última en 1993) ya se ha quedado algo anticuado y no es en la actualidad un servicio interesante, salvo en algunos casos, debido a su baja eficiencia y velocidad; normalmente no supera los 64 Kbps, aunque se pueden contratar conexiones de hasta 2.048 Kbps. A pesar de estas desventajas conviene conocer los aspectos básicos de X.25 pues aun existe una gran cantidad de usuarios de este tipo de redes. Además, en el protocolo X.25 se definieron por primera vez muchos de los conceptos en que se basa frame relay y ATM, que podemos considerar en cierto sentido como sus descendientes. El conjunto de estándares que definen X.25 ha sido adoptado como parte del modelo OSI para los tres primeros niveles.

A nivel físico se definen en X.25 dos interfaces, la X.21 cuando se usa señalización digital (cosa poco habitual) y la X.21bis (un subconjunto de la EIA-232D/V.24) cuando es analógica.

A nivel de enlace se utiliza un protocolo llamado LAP-B (Link Access Procedure-Balanced) que es una versión modificada del estándar ISO HDLC (High-level Data Link Control), que veremos en detalle al estudiar la capa de enlace.

El protocolo utilizado a nivel de red se conoce como X.25 PLP (Packet Layer Protocol). En este nivel se realizan todas las funciones de control de flujo, confirmación y direccionamiento. Cada NSAP (Network Services Access Point) en una red X.25 viene representado por una interfaz de un conmutador X.25, y tiene una dirección única. Las direcciones son numéricas y típicamente pueden tener entre nueve y quince dígitos. Las redes X.25 públicas de muchos países están interconectadas, como ocurre con las redes telefónicas. Para facilitar su direccionamiento la CCITT ha establecido un sistema jerárquico análogo al sistema telefónico en la recomendación X.121; así es posible por ejemplo llamar desde Iberpac (la red X.25 pública española) a una dirección de Transpac (la red pública X.25 francesa), sin más que añadir el prefijo correspondiente a dicha red en la dirección de destino.

X.25 es un servicio fiable orientado a conexión; los paquetes llegan en el mismo orden con que han salido. Una vez establecido un circuito entre dos NSAPs la información se transfiere en paquetes que pueden ser de hasta 128 bytes (aunque en muchas redes se permiten tamaños de hasta 4 KB). En la red los paquetes son transferidos de cada conmutador al siguiente (almacenamiento y reenvío), y solo borrados cuando se recibe la notificación de recepción. Un mismo NSAP puede tener establecidos varios VCs (PVCs y/o SVCs) hacia el mismo o diferentes destinos.

Los ordenadores que se conectan a un conmutador X.25 necesitan tener la capacidad suficiente para procesar los complejos protocolos X.25. Cuando se definió el estándar X.25 los ordenadores personales eran caros y poco potentes; muchos usuarios que tenían necesidad de conectarse a redes X.25 no disponían de un ordenador adecuado. Para estos casos se diseñó un equipo capaz de conectar un terminal asíncrono, que trabaja en modo carácter (es decir, un paquete por carácter) a una red X.25. A dicho equipo se le denominó PAD (Packet Assembler Disassembler) ya que se ocupaba de ensamblar y desensamblar los paquetes X.25 que recibía. A través de un PAD un usuario de un PC, o incluso de un terminal “tonto”, podía conectarse a un host en una red X.25 y trabajar como un terminal remoto de aquel. La CCITT publicó tres documentos para especificar todo lo relacionado con el funcionamiento de un PAD: el X.3 describe las funciones propias del PAD, el X.28 define el protocolo de comunicación entre el PAD y el terminal asíncrono, y el X.29 define el protocolo entre el PAD y la red X.25. El uso conjunto de estos tres protocolos permite iniciar una sesión interactiva desde un terminal conectado a un PAD con un ordenador remoto, por lo que se le conoce como el logon remoto XXX. Cuando un usuario en un ordenador conectado a X.25 desea establecer una conexión como terminal remoto de otro ordenador a través de una red X.25 lo hace mediante un programa en su ordenador que emula el comportamiento de un PAD (PAD Emulation). El logon remoto XXX ofrece en redes X.25 un servicio equivalente al de Telnet en TCP/IP. Para el caso de usuarios que no dispongan de un PAD propio muchas compañías telefónicas ponen a su disposición un servicio de acceso a PADs por RTC (normalmente RTB). Este servicio se denomina normalmente X.28, por ser este estándar el que define el protocolo de comunicaciones entre el terminal de usuario y el PAD.

El rendimiento que se obtiene de un VC X.25 depende de muchos factores: velocidad de los accesos físicos implicados, número de VC simultáneos, tráfico en cada uno de ellos, carga de la red, infraestructura, etc.

En España Telefónica inició un servicio de red pública de conmutación de paquetes en 1971 con la red RSAN, basada en unos protocolos propios, no estándar. Esta red hoy desaparecida fue la segunda red de conmutación de paquetes del mundo (después de ARPAnet que empezó en 1969), y la primera establecida por un operador de telefonía. En 1984 Telefónica inició la red Iberpac, que ya obedecía a los estándares X.25. A través de Iberpac es posible acceder a mas de 200 redes similares en todo el mundo. Las velocidades de acceso a Iberpac pueden ser de 2,4 a 2.048 Kbps. Es posible contratar PVCs, aunque lo normal es utilizar SVCs. La tarificación se hace por tres conceptos: en primer lugar una cuota fija mensual según la velocidad de la línea de acceso, en segundo por el tiempo que dura cada llamada (o lo que es lo mismo, el tiempo que esta establecido cada SVC), y en tercer lugar por el número de paquetes transferidos por llamada. Para los dos últimos conceptos existen tres ámbitos de tarificación: nacional, europeo e internacional (en X.25 cuesta lo mismo transferir datos entre dos oficinas vecinas que entre Bilbao y La Coruña). Telefónica dispone también de un servicio de acceso X.28 a su red Iberpac, conocido como Datex28.

Los protocolos X.25 se diseñaron pensando en los medios de transmisión de los años setenta, líneas de baja velocidad con tasa de errores elevada. El objetivo era aprovechar lo mejor posible las lentas líneas de transmisión existentes, aun a costa de hacer un protocolo de proceso pesado. Por si esto fuera poco, las redes X.25 casi siempre se utilizan para *encapsular* tráfico correspondiente a otros protocolos, por ejemplo TCP/IP, SNA o DECNET (podríamos decir que los paquetes de estos protocolos viajan “disfrazados” en paquetes X.25); cuando se encapsula un protocolo como TCP/IP en X.25 se realizan de forma redundante las tareas de la capa de red, con lo que el resultado es aún mas ineficiente. Para resolver este tipo de problemas a partir de 1990 se empezaron a crear redes basadas en frame relay.

1.6.3.2.- Frame Relay

Frame Relay (que significa *retransmisión de tramas)* nació a partir de los trabajos de estandarización del servicio RDSI, como un intento de crear una versión “light” de X.25, que permitiera aprovechar las ventajas de poder definir circuitos virtuales pero sin la baja eficiencia que tenían los protocolos excesivamente “desconfiados” de X.25. Mientras que en X.25 la capa de enlace y la capa de red eran sumamente complejas en frame relay ambas se intentaron reducir a su mínima expresión, dejando en manos de los equipos finales toda la labor de acuse de recibo, retransmisión de tramas erróneas y control de flujo; de esta forma frame relay se convertía en el complemento perfecto a otros protocolos, tales como TCP/IP. En muchos casos se considera que frame relay no es un protocolo a nivel de red sino a nivel de enlace (de ahí su nombre), y aun visto como nivel de enlace resulta bastante ligero.

El servicio que suministra frame relay consiste básicamente en identificar el principio y final de cada trama, y detectar errores de transmisión. Si se recibe una trama errónea simplemente se descarta, confiando en que el protocolo de nivel superior de los equipos finales averigüe por sí mismo que se ha perdido una trama y decida si quiere recuperarla, y como. A diferencia de X.25, frame relay no tiene control de flujo ni genera acuse de recibo de los paquetes (estas tareas también se dejan a los niveles superiores en los equipos finales). El tamaño máximo de los paquetes varía según las implementaciones entre 1 KB y 8 KB. La velocidad de acceso a la red típicamente esta entre 64 y 2.048 Kbps, aunque ya se baraja la estandarización de velocidades del orden de 34 Mbps.

Una novedad importante de Frame Relay estriba en que se define un ancho de banda “asegurado” para cada circuito virtual mediante un parámetro conocido como CIR (Committed Information Rate). Un segundo parámetro, conocido como EIR (Excess Information Rate) define el margen de tolerancia que se dará al usuario, es decir, cuanto se le va a dejar “pasarse” del CIR contratado. Por ejemplo, supongamos que un ordenador se conecta a una red frame relay mediante una línea de acceso al conmutador de 1.984 Kbps, y tiene dos circuitos establecidos con otros dos ordenadores, cada uno de ellos con un CIR de 256 Kbps y un EIR de 256 Kbps; en este caso cada circuito tendrá asegurado un ancho de banda de 256 Kbps como mínimo, y si la red no está saturada podrá llegar a 512 Kbps; si un circuito intenta utilizar mas de 512 Kbps el conmutador frame relay empezará a descartar tramas. Obsérvese que en este caso la línea de acceso nunca llegaría a saturarse, ya que como mucho podrían enviarse 512 Kbps por cada circuito. La especificación del CIR para un circuito virtual

se hace de forma independiente para cada sentido de la transmisión, y puede hacerse asimétrica, es decir dar un valor distinto del CIR para cada sentido.

Cuando un usuario hace uso del EIR (es decir, genera un tráfico superior al CIR contratado en un circuito virtual) el conmutador frame relay pone a 1 en las tramas excedentes un bit especial denominado DE (Discard Elegibility). Si se produce congestión en algún punto de la red el conmutador en apuros descartará en primera instancia las tramas con el bit DE marcado, intentando resolver así el problema. Este mecanismo permite a un usuario aprovechar la capacidad sobrante en la red en horas valle sin perjudicar la calidad de servicio a otros usuarios en horas punta, ya que entonces se verá limitado a su CIR. En realidad el CIR tampoco está garantizado, ya que si la congestión no se resuelve descartando las tramas DE el conmutador empezará a descartar tramas normales (no marcadas como DE) que pertenecen a usuarios que no han superado su CIR. Afortunadamente las redes frame relay se suelen dimensionar de forma que el CIR de cada usuario esté prácticamente garantizado en cada momento. En cierto modo podemos imaginar el bit DE como un sistema de “reserva de asiento” en un billete de tren (el bit a 0 significaría tener hecha reserva).

Una red Frame Relay podría utilizarse en vez de líneas dedicadas para interconectar conmutadores X.25; a la inversa sería mucho más difícil ya que al ser X.25 una red mas lenta los retardos introducidos serían apreciados por los usuarios de Frame Relay.

En ocasiones se utilizan redes Frame Relay para transmitir voz digitalizada; esto no es posible con X.25 debido a la lentitud del protocolo, que introduciría unos retardos excesivos; el envío de voz por una red tiene unos requerimientos especialmente severos en cuanto a retardos para que la transmisión se efectúe correctamente.

La red pública Frame Relay de Telefónica se denomina Red Uno, y esta operativa desde 1992. Aunque Telefónica anunció la disponibilidad de SVCs en Frame Relay para 1997, parece que estos aun no estan disponibles y el único servicio contratable es el de PVCs. La tarificación se realiza por dos conceptos: el primero es una cuota fija mensual en función de la velocidad de acceso a la red; el segundo es una cuota fija al mes por cada circuito según el valor de CIR que se tenga contratado; en ambos casos la tarifa depende de la distancia. El EIR no se especifica en el contrato, y por tanto no se paga, pero tampoco se compromete su valor por parte de Telefónica; habitualmente Telefónica pone un EIR que es 256 Kbps superior al CIR contratado. La velocidad del acceso físico puede tener valores comprendidos entre 64 y 1.984 Kbps. El CIR puede ser de 0 a 1.984 Kbps. Al no existir circuitos conmutados la Red Uno no es una red abierta como lo son Iberpac o la RTC. Es posible la conexión internacional con muchas otras redes frame relay gracias a acuerdos suscritos con diversos operadores.

1.6.3.3.- ATM y B-ISDN

Casi todos los servicios de comunicación que hemos visto hasta ahora fueron diseñados para la transmisión de voz o datos, pero no ambos. La RTB y la red GSM, pensadas para la voz, pueden transmitir datos, pero sólo a bajas velocidades. Las líneas dedicadas y redes Frame Relay, pensadas para datos, pueden transmitir voz si se utilizan equipos apropiados y se respetan ciertas restricciones.

El único servicio de los que hemos visto hasta ahora que se diseñó pensando en voz y datos es la RDSI (de ahí el nombre de Servicios Integrados). Pero la RDSI tiene dos inconvenientes importantes:

- Al ser una red de conmutación de circuitos *reales* la reserva del ancho de banda se realiza durante todo el tiempo que esta establecida la comunicación, independientemente de que se estén transfiriendo datos o no (o en el caso de transmitir voz independientemente de que se este hablando o se este callado).

- El estándar RDSI se empezó a definir en 1984. En aquel entonces las líneas dedicadas eran de 9.6 Kbps en el mejor de los casos y hablar de enlaces a 64 Kbps parecía algo realmente avanzado; sin embargo el proceso de estandarización tardó mas de lo previsto (cosa que ocurre a menudo) y cuando aparecieron los primeros servicios RDSI diez años más tarde la red “avanzada” resultaba interesante sólo en entornos domésticos y de pequeñas oficinas; se había quedado corta para nuevas aplicaciones.

Hasta aquí sólo hemos hablado de la transmisión de voz o datos, pero las redes de comunicaciones permiten transmitir también otros tipos de información como imágenes en movimiento (videoconferencia o vídeo), que tienen unos requerimientos distintos. De una forma muy concisa resumimos en la siguiente tabla las características esenciales de cada tipo de tráfico:

Tipo de información	Capacidad	Pérdida tolerable	Retardo	Jittter
Datos	Variable	Muy baja	Alto	Alto
Audio en tiempo real, monólogo	Baja (64 Kbps)	Baja	Bajo	Muy bajo
Audio en tiempo real, diálogo	Baja (64 Kbps)	Baja	Muy bajo	Muy bajo
Vídeo en tiempo real	Alta (2 Mbps)	Media	Bajo	Bajo

Tabla 1.4

Cuando una red está preparada para transmitir tanto audio y vídeo como datos informáticos decimos que es una red multimedia. Generalmente el tráfico multimedia tiene unas necesidades muy variables de ancho de banda, se dice que es un tráfico a *ráfagas* (“bursty traffic”).

Cuando se tiene tráfico a ráfagas resulta especialmente útil disponer de una red de conmutación de paquetes con circuitos virtuales, ya que así unos usuarios pueden aprovechar en un determinado instante el ancho de banda sobrante de otros. Sin embargo las redes de este tipo que hemos visto hasta ahora (X.25 y frame relay) no son apropiadas para tráfico multimedia porque el retardo y el jitter son impredecibles cuando la red esta cargada, y en general son demasiado lentas (especialmente X.25).

Las compañías telefónicas vienen trabajando desde hace bastante tiempo en el diseño de una red adecuada al tráfico multimedia que permita aprovechar las ventajas de la conmutación de paquetes, para así utilizar de forma mas eficiente las infraestructuras y ofrecer servicios nuevos, tales como la videoconferencia o el vídeo bajo demanda. La tecnología que permite todo esto se denomina ATM (Asynchronous Transfer Mode) y sus orígenes se remontan nada menos que a 1968, cuando se concibió en los laboratorios Bell el primer sistema de transmisión de *celdas.*. En esencia lo que se intenta con esta nueva tecnología es integrar todos los servicios en una única red digital, lo mismo que pretendía la RDSI (aunque como hemos visto llegó demasiado tarde). Por este motivo ATM también se denomina a veces RDSI de banda ancha o RDSI-BA (B-ISDN, Broadband-ISDN); por contraste a la “antigua” RDSI se la denomina en ocasiones RDSI de banda estrecha o RDSI-BE (N-ISDN, Narrowband-ISDN). Podríamos decir que la RDSI de banda ancha es lo más parecido a las “autopistas de la información”.

En 1986 la CCITT definió el concepto de RDSI-BA y eligió ATM como la tecnología sobre la que se basarían los futuros estándares. En aquel entonces ATM era una tecnología que interesaba exclusivamente a las compañías telefónicas. Gradualmente los fabricantes de ordenadores se fueron percatando de las posibilidades y futuro de dicha tecnología; para acelerar el proceso de estandarización se creó en 1991 el *ATM forum,* en el que participaban compañías telefónicas y fabricantes de ordenadores. A partir de ese momento se ha producido un avance impresionante en las normas y equipos ATM, especialmente en lo que se refiere a redes de datos. El primer conmutador ATM comercial apareció en 1991.

En cierto sentido ATM puede verse como una evolución de frame relay. La principal diferencia es que los paquetes ATM tienen una longitud fija de 53 bytes (5 de cabecera y 48 de datos) frente al tamaño variable y mucho mayor de las tramas frame relay. Debido a su tamaño pequeño y constante los paquetes ATM se denominan *celdas*, y por esto en ocasiones a ATM se le denomina cell relay (retransmisión de celdas). Manejar celdas de un tamaño tan reducido tiene la ventaja de que permite responder con mucha rapidez a tráfico de alta prioridad que pueda llegar inesperadamente mientras se están transmitiendo otro menos urgente, algo muy importante en tráfico multimedia. El hecho de que todas las celdas sean del mismo tamaño simplifica el proceso en los nodos intermedios, cuestión esencial cuando se quiere que dicho proceso sea lo más rápido posible. En el lado negativo está el hecho de que la eficiencia de una conexión ATM nunca puede superar el 90% (48/53) debido a la información de cabecera que viaja en cada celda.

Al igual que en X.25 o frame relay, una red ATM se constituye mediante conmutadores ATM normalmente interconectados por líneas dedicadas, y equipos de usuario conectados a los conmutadores. Mientras que en X.25 o frame relay se utilizan velocidades de 64 Kbps a 2 Mbps, en ATM las velocidades pueden llegar a 155,52, 622,08 o incluso superiores. La elección de precisamente estos valores se debe a que son los que se utilizan en el nuevo sistema de transmisión sobre fibra óptica en redes WAN denominado SONET/SDH (Synchronous Optical NETwork/Synchronous Digital Hierarchy), que es el que están utilizando las compañías telefónicas actualmente en las infraestructuras. ATM también puede utilizarse a velocidades inferiores, 34 Mbps e incluso 2 Mbps.

Dos equipos conectados a una red ATM pueden establecer entre si un circuito virtual, permanente o conmutado, y transmitir por él información digital de cualquier tipo. ATM da al usuario muchas mas facilidades que X.25 o frame relay para controlar las características de su circuito virtual: se puede fijar un ancho de banda máximo permitido, un margen de tolerancia sobre dicho máximo, un ancho de banda mínimo garantizado, un ancho de banda asimétrico, un perfil horario de forma que el ancho de banda fluctúe con la hora del día de una forma preestablecida, etc. Además es posible definir prioridades y distintos tipos de tráfico, de forma que se prefiera fiabilidad o rapidez, tráfico constante o a ráfagas, etc.

El modelo de referencia ATM

ATM tiene su propio modelo de referencia, constituido por tres capas denominadas *capa física*, *capa ATM* y *capa de adaptación ATM,* o *capa 1L (ATM Adaptation Layer).*

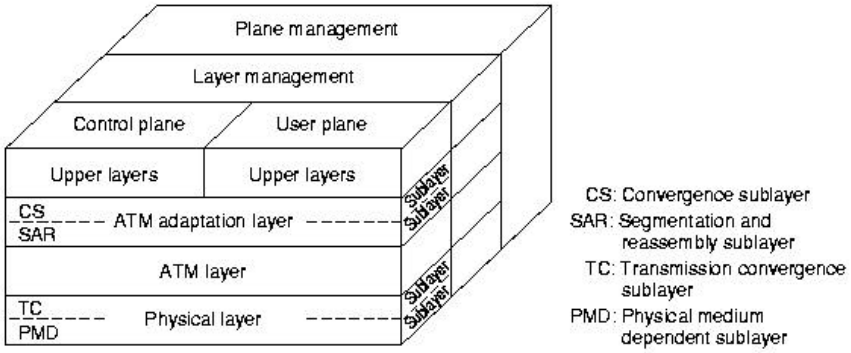

Figura 1.14

La capa física está formada por dos subcapas: la PMD (Physical Media Dependent) y la TC (Transmission Convergence). La subcapa PMD describe la interfaz física con el medio de transmisión, y equivale a la capa física del modelo OSI. La subcapa TC se ocupa de “deshacer” las celdas en bits para pasarlos a la subcapa PMD en el envío, y de recibir los bits de la subcapa PMD para reconstruir las celdas en la recepción. Si consideramos la celda como equivalente a la trama en el modelo OSI esta subcapa haría la función de la capa de enlace. Por este motivo nosotros estudiaremos la subcapa TC en la capa de enlace.

La capa ATM trata de la estructura de las celdas y su transporte. También realiza las tareas de señalización, es decir establece y termina los circuitos virtuales, y realiza el control de congestión. Sus funciones son una mezcla de la capa de enlace y la capa de red en el modelo OSI.

La capa de adaptación ATM (capa 1L) se divide también en dos subcapas; la inferior se denomina subcapa SAR (Segmentation And Reassembly) se ocupa de fragmentar el paquete que recibe desde arriba (normalmente mayor de 48 bytes) en celdas para su envío, y de reensamblarlo en la recepción cuando se lo pasa la capa ATM. La subcapa CS (Convergence Sublayer) se ocupa de suministrar distintos tipos de servicio adecuados al tipo de tráfico (vídeo, audio, datos. etc.). La capa 1L corresponde en sus funciones a la capa de transporte del modelo OSI.

Obsérvese que en el modelo de referencia ATM no se habla de aplicaciones. En realidad el modelo contempla la existencia de capas por encima de la capa 1L, pero no se especifican sus funciones ni características. El modelo deja total libertad a los implementadores sobre como diseñar las aplicaciones que funcionen sobre ATM. Actualmente el principal uso de ATM es como medio de transporte para otros protocolos; hay muy pocas aplicaciones que hayan sido diseñadas para funcionar de manera *nativa*, es decir, directamente sobre la capa 1L.

Futuro de ATM

Probablemente ATM es el acrónimo que está mas de moda en el mundo de las telecomuncaciones actualmente. Prácticamente cualquier revista del área incluye uno o varios artículos sobre algún aspecto de ATM o tema relacionado. En los últimos años ha habido bastante debate sobre si ATM se consolidaría o no como la tecnología de red del futuro, cosa que hoy en día ya pocos ponen en duda. Una de las razones que ha hecho prosperar a ATM es que las compañías telefónicas han visto en esta red su oportunidad para competir con los servicios que ofrecen las compañías de televisión por cable.

Aunque seguirán durante mucho tiempo ofreciéndose todos los tipos de servicios que hemos visto anteriormente, es muy probable que en unos años la infraestructura básica de las compañías telefónicas esté formada por conmutadores ATM unidos mediante enlaces SONET/SDH; y todos los demás servicios utilicen esta red como medio de transporte (de forma análoga a como X.25 puede utilizar frame relay). Por ejemplo Telefónica está evolucionando hacia una red ATM para la interconexión de sus grandes centros como infraestructura básica sobre la que discurrirá el tráfico de todas las otras redes.

Existen aun muy pocas experiencias de servicios públicos ATM; una de las mayores se ha puesto en marcha en Finlandia, país que se encuentra a la cabeza de Europa en muchos aspectos de telecomunicaciones. En España Telefónica inició en 1996 dos servicios de red ATM denominados servicio Gigacom y Servicio Cinco (Comunicaciones Integrales Corporativas). Estos servicios están orientados a clientes con grandes necesidades de transmisión de datos mutimedia; solo se permite la constitución de PVCs; las velocidades de acceso van de 512 Kbps a 155 Mbps. Este servicio puede ser una alternativa interesante a las líneas dedicadas de alta velocidad, ya que además de su precio más interesante permiten contratar servicios de acuerdo a horarios preestablecidos, por ejemplo un periódico que necesita transmitir todos los días un circuito de 4 Mbps entre sus dos oficinas principales de 1 a 2 de la mañana para transmitir la edición del día siguiente.

Curiosamente uno de los campos donde ATM ha encontrado más éxito es como base para constituir redes locales de alta velocidad. Existen hoy en día equipos en el mercado que permiten interconectar redes locales tradicionales (Ethernet, Token Ring, FDDI) a través de conmutadores ATM, pudiendo conectar directamente a ATM a 155 Mbps los servidores mas importantes; de esta forma se consiguen prestaciones y funcionalidades mejores que las de cualquier red local actual. Esta por ver si las redes locales de alta velocidad del futuro se basarán en ATM, pero no hay duda de que esta tecnología WAN tiene un papel que jugar también en la LAN.

1.7.- EJEMPLOS DE REDES

Las redes que hay en el mundo difieren en cuanto a sus medios físicos de transmisión, protocolos, tipos de máquinas que conectan, forma como se crearon, modo de administración, etc. A continuación describimos algunas redes con las que como usuarios de esta Universidad tenemos un contacto más frecuente.

1.7.1.- ARPANET, NSFNET y La Internet

Como ya hemos comentado ARPANET tuvo sus orígenes en un proyecto del Departamento de Defensa de los Estados Unidos que se desarrolló en la segunda mitad de la década de los sesenta, y que pretendía crear una red de ordenadores resistente a ataques militares. Para esto se optó por crear una red de conmutación de paquetes formada por ordenadores especializados denominados IMPs (Interface Message Processors) que se interconectaban por líneas telefónicas punto a punto de

56 Kbps. Para aumentar la fiabilidad de la red se diseñó una topología mallada en la que cada IMP estaba conectado al menos a otros dos. Los IMPs eran los routers de la ARPANET. Como detalle curioso diremos que eran miniordenadores Honeywell con 24 KBytes de memoria RAM especialmente modificados para desarrollar esta tarea; un router pequeño de hoy en día tiene típicamente de 2 a 4 MBytes de memoria RAM.

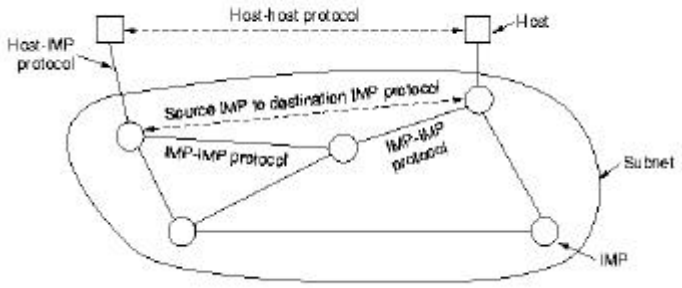

Figura 1.15

Dado que en aquellas fechas aun no existían redes locales, cada IMP se conectaba en forma local a un solo ordenador (host) ubicado en la misma habitación; era este ordenador, que daba servicio a un número razonablemente elevado de usuarios, el que permitía a éstos utilizar la red a través de sus aplicaciones. En la ARPANET la subred estaba formada por los IMPs y las líneas punto a punto que los unían.

La ARPANET empezó a funcionar en diciembre de 1969 con cuatro nodos. En septiembre de 1972 tenía 34 nodos repartidos por todo Estados Unidos.

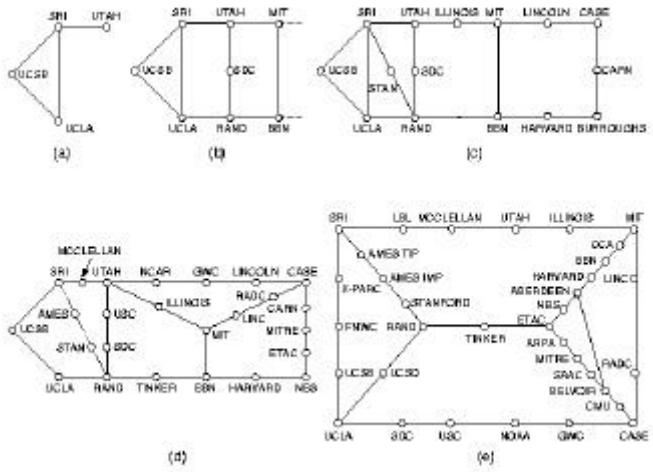

Figura 1.16

Paulatinamente se fueron ampliando las posibilidades de los IMP, permitiendo conexiones de varios hosts en cada IMP, de hosts remotos, o incluso de simples terminales. También se hicieron pruebas con medios de transmisión distintos de las líneas telefónicas, tales como transmisiones por radio o vía satélite. En estas pruebas se puso de manifiesto que los protocolos existentes no eran aptos para interconectar redes con diferentes tecnologías, por lo que con este fin se diseñaron unos nuevos, que son los que conocemos como TCP/IP, y que fueron los oficiales en ARPANET desde principios de 1983. Para fomentar su uso ARPA adjudicó diversos contratos a la Universidad de California en Berkeley y a BBN (empresa encargada de gestionar la ARPANET) para que integraran los nuevos protocolos en el UNIX de Berkeley. El momento fue muy propicio y muchas universidades de Estados Unidos se conectaron a la nueva red. El principal problema que tenía la ARPANET era que sólo las universidades o centros de investigación que tuvieran contratos con el DoD podían conectarse a ella, y muchas no los tenían.

En vista del éxito que tenía ARPANET la NSF (National Science Foundation) diseñó en 1984 una red paralela que permitiría a las universidades conectarse sin las restricciones que imponía ARPANET. Para esto se empezó de una forma muy similar a la ARPANET, interconectando seis superodenadores que la NSF tenía repartidos por Estados Unidos mediante microordenadores adaptados para actuar como routers; el superordenador iba conectado al router y éste a una o dos líneas. En la NSFNET también se utilizaban los protocolos TCP/IP. La Universidad Carnegie-Mellon, que estaba conectada a NSFNET y ARPANET, hacía de "puente" entre ambas redes para permitir el intercambio de paquetes.

El crecimiento de la NSFNET fue espectacular. Multitud de universidades, centros de investigación, museos y bibliotecas se conectaron a ella. En poco tiempo las líneas iniciales de 56 Kbps que constituían la espina dorsal o *backbone* de la red tuvieron que ser aumentadas a 448 Kbps, y después a 1.5 Mbps. A medida que la red se popularizaba las empresas comerciales querían entrar en ella, pero esto no era posible con una red financiada por fondos públicos. En 1990 IBM, MCI y MERIT, impulsados por la NSF, crearon una empresa sin ánimo de lucro denominada ANS (Advanced Networks and Services) que se ocupó de sustituir a NSFNET y crear ANSNET, aumentando la capacidad del backbone a 45 Mbps. También en 1990 desapareció por completo ARPANET, que había ido disminuyendo en importancia a medida que NSFNET y otras redes fueron ocupando su terreno. En 1995 ANSNET fue vendido a America Online; en ese momento existían ya numerosas compañías que operaban redes IP comerciales en los Estados Unidos.

Después de que en 1984 se interconectaran ARPANET y NSFNET también lo hicieron muchas redes regionales y de otros países, aprovechando la ventaja que suponía tener un protocolo común. Así se constituyó poco a poco una superred o *interred* que desde el punto de vista del usuario era una única red. De manera coloquial los usuarios empezaron a llamar a esta superred la *internet*, hasta el punto de que este llegó a ser su nombre propio; no existe una fecha concreta en la que podamos fijar el origen de *La Internet*. Sin duda ARPANET y NSFNET han sido sus dos principales impulsores.

En Europa (y en España) el desarrollo de La Internet estuvo frenado durante la década de los ochenta por razones políticas, ya que los gobiernos y la Comunidad Europea, que financiaban las redes de investigación, fomentaban el uso de protocolos OSI e impedían el uso de TCP/IP (la excepción eran los países nórdicos, que sí estaban integrados en La Internet). La comunicación de los investigadores europeos con sus colegas americanos sólo podía efectuarse a través de *pasarelas,* generalmente limitadas al correo electrónico en modo texto. En 1991 la actitud de los gobiernos y las redes europeas de investigación empezó a cambiar, se permitió la coexistencia de los protocolos TCP/IP con los OSI, y la interconexión a La Internet; de forma inmediata se produjo una explosión en el número de usuarios de la Internet en Europa, que superó incluso a la ocurrida en Estados Unidos.

Actualmente la Internet crece sin parar, a un ritmo que aproximadamente duplica su tamaño cada año. Es de esperar que en unos años este ritmo disminuya. El crecimiento tan elevado, unido a la privatización de la infraestructura básica, ha traído como consecuencia indeseable la saturación de la red en muchas áreas. En octubre de 1996 se puso en marcha una iniciativa, liderada por 34 universidades de Estados Unidos, denominada Internet II. El objetivo de Internet II es rescatar el espíritu original de la NSFNET, es decir, crear un entorno de red específicamente orientado al ámbito académico e investigador que permita garantizar al usuario una calidad de servicio tal que permita experimentar con nuevos servicios que requieran una elevada velocidad de transmisión. Aunque nace como una iniciativa de los Estados Unidos, no sería raro que en un breve plazo se unan a ella instituciones europeas.

El mecanismo informal que solía predominar en los viejos tiempos de La Internet ya no es factible hoy en día dada la gran cantidad de usuarios e intereses comerciales en juego. Por esto se creó en 1992 una asociación denominada *Internet Society (ISOC)* que se ocupa de coordinar, promover y organizar el futuro desarrollo de La Internet, incluida la aceptación de nuevos estándares dentro del conjunto de protocolos.

1.7.2.- La red nacional de I+D, RedIRIS

Como en muchos otros países, en España el desarrollo de las redes de ordenadores se produjo inicialmente por la necesidad de comunicación de las universidades y centros de investigación. Las primeras conexiones se llevaron a cabo en 1984 cuando tres universidades en Madrid y Barcelona se conectaron a la red EARN (European Academic and Research Network) que, patrocinada en sus inicios por IBM, utilizaba unos protocolos de este fabricante llamados NJE (Network Job Entry) anteriores incluso a SNA. Un año más tarde otros centros establecieron otra red basada en ordenadores VAX/VMS de Digital Equipment Corporation (DEC) que utilizaba protocolos DECNET. Ambas redes disponían de enlaces internacionales, pero no estaban interconectadas entre sí a nivel nacional, por lo que la comunicación entre ellas sólo podía establecerse a través de una pasarela en otro país, y sólo para correo electrónico. Las velocidades de interconexión típicas eran de 2.4 a 9.6 Kbps.

En 1988 el PNID (Plan Nacional de Investigación y Desarrollo), órgano dependiente de la CICYT (Comisión Interministerial de Ciencia y Tecnología), en un intento por armonizar estas iniciativas y para extender el uso de las redes a toda la comunidad universitaria española, inició el Programa IRIS (mas tarde RedIRIS), delegando su gestión en Fundesco. Siguiendo directrices entonces normales en todas las redes de investigación europeas, el Programa IRIS fomentaba el uso de protocolos OSI, lo cual supuso que en la práctica el avance de la red se viera limitado por la falta de software adecuado para muchas plataformas, quedando el correo electrónico prácticamente como el único servicio disponible. Para la transmisión de los datos se utilizaba el servicio Iberpac de Telefónica, con velocidades de acceso de 2.4 a 9.6 Kbps. En Madrid, Barcelona y Sevilla se instalaron conmutadores X.25 que se interconectaron mediante líneas dedicadas de 64 Kbps, creando así la primera fase de una red X.25 privada ARTIX (ARTeria de Interconexión X.25).

En 1991, coincidiendo con el cambio a la denominación RedIRIS, se empezó a ofrecer un nuevo servicio de interconexión de redes de área local mediante protocolos IP denominado SIDERAL, con lo que el uso de los protocolos TCP/IP, ya presentes por aquel entonces en las redes locales de muchas universidades, se extendió rápidamente a la WAN. Simultáneamente se fue extendiendo la red ARTIX de forma gradual instalando un conmutador X.25 y una línea dedicada en cada comunidad autónoma, con una topología arborescente con algunos enlaces redundantes. La red ARTIX actuaba como una red multiprotocolo, ya que viajaban sobre ella paquetes de diversos protocolos: OSI (los “propios” de RedIRIS), TCP/IP (el nuevo servicio SIDERAL), DECNET (la red FAENET) y NJE/SNA (la red EARN).

En 1994 la CICYT decidió traspasar la gestión de la red de Fundesco al CSIC. También en 1994 se decidió, a la vista de que el tráfico IP era mayoritario, convertir ARTIX en una red IP nativa para evitar la pérdida de rendimiento que suponía el encapsular paquetes IP en X.25; para esto se sustituyeron los conmutadores X.25 por routers IP, creando una red similar a lo que había sido NSFNET diez años antes. El tráfico remanente de otros protocolos se encapsulaba ahora en paquetes IP para su envío por la red.

En 1995 la CICYT firmó un acuerdo de cooperación tecnológica con Telefónica, gracias al cual RedIRIS podía utilizar los servicios avanzados de transmisión de datos. Actualmente, la red troncal o backbone que soporta los servicios de comunicaciones de RedIRIS, está formada por un conjunto de nodos convenientemente distribuidos por el territorio nacional, conectados entre sí por medio circuitos ATM sobre accesos ATM de 34/155 Mbps, utilizando como mecanismo de backup accesos primarios de RDSI. En la actualidad son 17 los nodos existentes, uno en cada Comunidad Autónoma.

Desde Madrid se establecen PVCs con cada una de las comunidades autónomas; en cada PVCs hay un caudal garantizado por separado para cada sentido, con valores que oscilan entre 2 y 12 Mbps según las necesidades de cada Comunidad Autónoma (por ejemplo en el caso del País Vasco hay 8 Mbps en sentido entrante y 4 Mbps en sentido saliente).

RedIRIS participa en el Proyecto TEN-155 que constituye una red IP paneuropea de 155 Mbps que nos interconecta con las distintas redes académicas y de investigación europeas. La velocidad de acceso de RedIRIS al TEN-155 a Europa y USA es de 45 Mbps y 17Mbps respectiavmente

Para el tráfico con Estados Unidos se dispone de una conexión ATM de 19,2 Mbps (aproximadamente 16 Mbps a nivel de IP) vía MCI.

RedIRIS ha respaldado y participado en la creación, a principios de 1997, de un punto neutro de interconexión para el intercambio de tráfico IP entre los proveedores de tránsito internacional a Internet existentes en España (ESPANIX).

Desde principios de 1996 está operativa una conexión entre RedIRIS e Ibernet, que permite el intercambio directo de tráfico IP entre las redes conectadas por ambos proveedores. En estos momentos la capacidad de esta conexión es de aproximadamente 18 Mbps IP y a través de ella también se intercambia tráfico con otros proveedores comerciales de Internet.

1.7.4.- La red de la Universidad de Deusto

La Universidad de Deusto dispone de dos campus (Bilbao y San Sebastián). Cada uno de los campus dispone de una red local de tecnología Ethernet conmutada (con segmentos de 10 y 100 Mbps) que interconecta sus equipos, utilizando par trenzado para las conexiones de las estaciones de trabajo y figra óptica para el backbone dentro de los edificios y entre éstos. Ambos campus están unidas mediante una línea dedicada de 128 Kbps. La conxión a Internet se realiza a través de RedIris mediante una conexión ATM con el nodo de RedIRIS en la Comunidad Autónoma Vasca (en el campus de Lejona de la Universidad del País Vasco) con un tráfico de 8 Mbps. El protocolo estándar de la red es el TCP/IP, si bien todavía queda tráfico residual SNA de los equipos de IBM de biblioteca que se encapsula sobre IP.

1.8.- ESTÁNDARES

En nuestra vida diaria estamos rodeados de estándares, incluso para las cosas más triviales como los pasos de rosca o el tamaño de las hojas de papel. En algunos casos el estándar hace la vida más cómoda (por ejemplo el formato A4 permite una manipulación cómoda de documentos), en otros es necesario para asegurar la *interoperabilidad* (roscar una tuerca en un tornillo, por ejemplo). Los estándares en materia de telecomunicaciones pertenecen al segundo tipo, es decir, son esenciales para asegurar la interoperabilidad entre diversos fabricantes, cosa esencial si se quieren hacer redes abiertas. Los estándares pueden ser de ámbito regional, nacional o internacional; por ejemplo en Estados Unidos el formato habitual de papel no es el A4 sino tamaño carta (un poco mas pequeño) que constituye un estándar nacional. Las telecomunicaciones son probablemente la primera actividad humana en la que se reconoció la necesidad de definir estándares internacionales; ya en 1865 representantes de muchos países europeos se reunieron para crear una organización que se ocupara de estandarizar las comunicaciones por telégrafo, acordando cosas tales como el código a utilizar; dicha organización fue la predecesora de la actual ITU.

Conviene destacar que la pertenencia de un país a una determinada organización no asegura su adhesión a los estándares emanados de la misma. Por ejemplo, el tamaño de papel A4 es parte de un estándar de la ISO (International Organization for Standardization) que es seguido por prácticamente todos los países del mundo excepto Estados Unidos que utiliza en su lugar el tamaño carta, a pesar de que también es miembro de la ISO.

Generalmente se suele distinguir dos tipos de estándares: *de facto* y *de jure*. Los estándares de facto (del latín “del hecho”) ocurren cuando un determinado producto o modo de comportamiento se extiende en una comunidad determinada sin una planificación previa, hasta el punto de que ese producto o modo de comportamiento se considera “normal” dentro de esa comunidad. Los estándares de facto ocurren de forma natural y progresiva, sin una planificación previa ni un proceso formal que los refrende. Por ejemplo en aplicaciones ofimáticas es un estándar de facto el PC compatible IBM con software de Microsoft; en entornos universitarios de docencia e investigación en informática es un estándar de facto el uso de sistemas operativos UNIX. Los estándares de facto también se llaman a veces “estándares de la industria” (industry standards).

Los estándares de jure (del latín “por ley”) son fruto de un acuerdo formal entre las partes implicadas, después de un proceso de discusión, consenso y generalmente votación. Se adoptan en el seno de una organización que normalmente está dedicada a la definición de estándares; si dicha organización tiene ámbito internacional el estándar definido es internacional. Existen dos clases de organizaciones internacionales: las “oficiales” que son fruto de tratados internacionales y que se crean por acuerdo entre los gobiernos de las naciones participantes, y las “extraoficiales”, que existen gracias al esfuerzo voluntario de sus miembros, sin participación directa de los gobiernos de sus países.

En el mundo de las redes de ordenadores existen hoy en día como hemos visto dos conjuntos de protocolos estándar, el OSI y el TCP/IP, pero ambos son relativamente recientes. En los años setenta y ochenta en que no había protocolos estándar disponibles la forma más sencilla de constituir una red multifabricante era utilizando los protocolos de IBM: SNA o su predecesor el NJE, Network Job Entry; como los equipos IBM eran los más extendidos casi todos los fabricantes disponían de productos que implementaban estos protocolos en sus equipos, con lo que jugaban el papel de protocolos "estándar"; además, en muchos casos una buena parte de los ordenadores a conectar era IBM por lo que el software necesario estaba allí de todos modos. Podemos decir que en aquellos años los protocolos SNA y NJE era hasta cierto punto un estándar de facto.

Pasaremos ahora a describir con mas detalle las principales organizaciones que tienen alguna relación con los estándares del campo de la telemática.

1.8.1.- La ISO

Muchos países tienen organizaciones nacionales de estándares donde expertos de la industria y las universidades desarrollan estándares de todo tipo. Entre ellas se encuentran por ejemplo:

ANSI	American National Standards Institute	Estados Unidos
DIN	Deutsches Institut fuer Normung	Alemania
BSI	British Standards Institution	Reino Unido
AFNOR	Association Francaise de Normalisation	Francia
UNI	Ente Nazionale Italiano de Unificatione	Italia
NNI	Nederlands Normalisatie-Instituut	Países Bajos
S1	Standards Australia	Australia
SANZ	Standards Association of New Zealand	Nueva Zelanda
NSF	Norges Standardiseringsforbund	Noruega
DS	Dansk Standard	Dinamarca
AENOR	Asociación Española de Normalización	España

La ISO (International Organization for Standardization) es una organización voluntaria (es decir, no es fruto de tratados internacionales) creada en 1946 con sede en Ginebra, Suiza. Sus miembros son las organizaciones nacionales de estándares de los 89 países miembros. A menudo un estándar de uno de sus miembros es adoptado por ISO como estándar internacional; esto ocurre especialmente con las más importantes, ANSI, DIN, BSI y AFNOR.

ISO emite estándares sobre todo tipo de asuntos, como por ejemplo: el sistema métrico de unidades de medida, tamaños de papel, sobres de oficina, tornillos y tuercas, reglas para dibujo técnico, conectores eléctricos, regulaciones de seguridad, componentes de bicicleta, números ISBN (International Standard Book Number), lenguajes de programación, protocolos de comunicación, etc. Hasta la fecha se han publicado unos 10.000 estándares ISO que afectan a prácticamente cualquier actividad de la vida moderna

Para realizar esta ingente labor ISO se organiza en cerca de 200 comités técnicos (TC, Technical Committee) numerados según su creación. El TC97 trata de ordenadores y proceso de la información. Cada comité tiene subcomités (SCs) que a su vez se dividen en grupos de trabajo (WG, Working Groups).

El proceso de creación de un estándar ISO es como sigue. Uno de sus miembros (una organización nacional de estándares) propone la creación de un estándar internacional en un área concreta. Entonces ISO constituye un grupo de trabajo que produce un primer documento denominado CD (Committee Draft, borrador del comité). El CD se distribuye a todos los miembros de ISO, que disponen de un plazo de seis meses para exponer críticas. El documento, modificado de acuerdo con las críticas, se somete entonces a votación y si se aprueba por mayoría se convierte en un DIS (Draft International Standard) que se difunde para recibir comentarios, se modifica y se vota nuevamente. En base a los resultados de esta votación se prepara, aprueba y publica el texto final del IS (International Standard). En áreas muy polémicas un CD o un DIS ha de superar varias versiones antes de conseguir votos suficientes, y el proceso entero puede llevar años.

ISO ha generado multitud de estándares en telemática, y en tecnologías de la información en general, siendo OSI su ejemplo más significativo. Además, ha adoptado estándares producidos por sus organizaciones miembros y por otras organizaciones relacionadas.

1.8.2.- La ITU-T

La ITU (International Telecommunication Union) fue creada en 1934, y con la creación de la ONU se vinculó a ésta en 1947. La ITU tiene tres sectores de los cuales solo nos interesa el que se dedica a la estandarización de las telecomunicaciones, que se conoce como *ITU-T.* Desde 1956 a 1993 la ITU-T se conoció con el nombre CCITT, acrónimo del nombre francés Comité Consultatif International Télégraphique et Téléphonique. En 1993 la CCITT fue reorganizada y se cambió el nombre a ITU-T; estrictamente hablando el cambio de nombre tiene efectos retroactivos, es decir, los documentos vigentes, aun si fueron producidos antes de 1993, son hoy documentos de la ITU-T y no de la CCITT.

Los miembros de la ITU-T son de cinco clases:

- Administraciones (PTTs nacionales).
- Operadores privados reconocidos (por ej. British Telecom, Global One, AT&T).
- Organizaciones regionales de telecomunicaciones (p. ej. el ETSI).
- Empresas que comercializan productos de telecomunicaciones y organizaciones científicas
- Otras organizaciones interesadas (bancos, líneas aéreas, etc.)

Entre los miembros hay unas 200 administraciones, unos cien operadores privados y varios cientos de miembros de las otras clases. Sólo las administraciones tienen derecho a voto, pero todos los miembros pueden participar en el trabajo. Cuando un país no tiene un monopolio de comunicaciones, como Estados Unidos, no existe PTT y la representación recae en algún organismo del gobierno relacionado (esto será posiblemente lo que ocurra ahora en la mayoría de los países de Europa).

Para desarrollar su trabajo la ITU-T se organiza en Grupos de Estudio, que pueden estar formados por hasta 400 personas. Los Grupos de Estudio se dividen en Equipos de Trabajo (Working Parties), que a su vez se dividen en Equipos de Expertos (Expert Teams).

Las tareas de la ITU-T comprenden la realización de recomendaciones sobre interfaces de teléfono, telégrafo y comunicaciones de datos. A menudo estas recomendaciones se convierten en estándares reconocidos internacionalmente, por ejemplo la norma V.24 (también conocida como EIA RS-232) que especifica la posición y el significado de las señales en el conector utilizado en muchos terminales asíncronos.

La ITU-T denomina a sus estándares “recomendaciones”; con esto se quiere indicar que los países tienen libertad de seguirlas o ignorarlas; aunque ignorarlas puede suponer quedar aislado del resto del mundo, por lo que en la práctica a menudo las recomendaciones se traducen en obligaciones.

Entre las recomendaciones mas relevantes de la ITU-T en el campo de la telemática podemos destacar la serie V sobre módems (p. ej. V.32, V.42), la serie X sobre redes de datos y OSI (X.25, X.400,...), las series I y Q que definen la RDSI, la serie H sobre codificación digital de sonido y vídeo, etc.

1.8.3.- La Internet Society

Cuando se puso en marcha la ARPANET el DoD creó un comité que supervisaba su evolución. En 1983 dicho comité recibió el nombre de IAB (Internet Activities Board), nombre que luego se cambió a Internet Architecture Board. Este comité estaba constituido por diez miembros. Dada la naturaleza de las organizaciones que constituían la ARPANET (y después la NSFNET) los miembros del IAB representaban básicamente a universidades y centros de investigación..

El IAB informaba al DoD y a la NSF (que eran entonces los que pagaban la Internet) de la evolución de la red y las posibles mejoras a realizar. El IAB también se ocupaba de detectar -después de intensas discusiones- donde era necesario o conveniente especificar un nuevo protocolo; entonces se anunciaba dicha necesidad en la red y normalmente siempre surgían voluntarios que lo

implementaban. La información circulaba en forma de documentos técnicos denominados RFCs (Request For Comments). El nombre da una idea del talante abierto y democrático que tienen todas las actividades de la Internet. Los RFCs se mantienen en la red y cualquiera que lo desee puede consultarlos, redistribuirlos, etc. (como comparación diremos que los documentos de la ITU y la ISO solo pueden obtenerse comprándolos a la oficina correspondiente); actualmente hay mas de 2000 RFCs y su número crece continuamente.

En 1989 el IAB fue reorganizado de nuevo para acomodarse a la evolución sufrida por la red. Su composición fue modificada para que representara a un rango más amplio de intereses ya que la anterior resultaba muy académica, y tenía un procedimiento de nombramiento no democrático (los miembros salientes nombraban a sus sucesores). Además se crearon dos subcomités dependientes del IAB, el IRTF (Internet Research Task Force) y el IETF (Internet Engineering Task Force); el IRTF se concentraría en los problemas a largo plazo, mientras que el IETF debía resolver las cuestiones de ingeniería más inmediatas.

En 1991 se creó la Internet Society (ISOC), una asociación internacional para la promoción de la tecnología Internet y sus servicios. Cualquier persona física u organización que lo desee puede ser miembro de la ISOC sin más que pagar su cuota anual. La ISOC está gobernada por un consejo de administración (Board of Trustees) cuyos miembros son elegidos por votación de los miembros de la ISOC entre una serie de candidatos propuestos. La ISOC absorbió en su seno el IAB con sus dos subcomités, pero cambió radicalmente el mecanismo de elección; estos son ahora nombrados por el consejo de administración de la ISOC.

Dentro de la compleja estructura que es la ISOC el grupo mas importante en lo que a elaboración de estándares se refiere es sin lugar a dudas el IETF. Inicialmente éste se dividió en grupos de trabajo, cada uno con un problema concreto a resolver. Los presidentes de dichos grupos de trabajo se reunían regularmente constituyendo lo que se llamaba el Comité Director. A medida que fueron apareciendo problemas nuevos se fueron creando grupos de trabajo, llegando a haber mas de 70 con lo que se tuvieron que agrupar en ocho áreas; el Comité Director está formado ahora por los ocho presidentes de área.

Paralelamente a la modificación de las estructuras organizativas se modificaron también los procedimientos de estandarización, que antes eran muy informales. Una propuesta de nuevo estándar debe explicarse con todo detalle en un RFC y tener el interés suficiente en la comunidad Internet para que sea tomada en cuenta; en ese momento se convierte en un *Estándar Propuesto (Proposed Standard).* Para avanzar a la etapa de *Borrador de Estándar (Draft Standard)* debe haber una implementación operativa que haya sido probada de forma exhaustiva por dos instalaciones independientes al menos durante cuatro meses. Si el IAB se convence de que la idea es buena y el software funciona declarará el RFC como un *Estándar Internet (Internet Standard)*. El hecho de exigir implementaciones operativas probadas antes de declarar un estándar oficial pone de manifiesto la filosofía pragmática que siempre ha caracterizado a Internet, radicalmente opuesta a ISO e ITU-T.

1.8.4.- Foros industriales

El proceso de definición de estándares de los organismos internacionales “tradicionales” (ITU-T e ISO) siempre se ha caracterizado por una gran lentitud, debida quizá a la necesidad de llegar a un consenso entre muchos participantes y a procedimientos excesivamente complejos y burocratizados. Ya hemos visto que la lentitud en crear los estándares OSI fue uno de los factores que influyó en su rechazo. El caso de RDSI es extremo: la ITU-T empezó a elaborar el estándar en 1972, y lo finalizó en 1984; los servicios comerciales aparecieron hacia 1994, 22 años después de iniciado el proceso; este retraso provocó que lo que se diseñó como un servicio avanzado para su tiempo (accesos digitales a 64 Kbps) resultara cuando se puso en marcha aprovechable sólo en entornos domésticos y de pequeñas oficinas.

Estos retrasos producían grandes pérdidas a los fabricantes de equipos, que no estaban dispuestos a repetir el error. Por ello a principios de los noventa empezó a surgir un nuevo mecanismo para acelerar la creación de estándares, consistente en la creación de grupos independientes formados por fabricantes, usuarios y expertos de la industria con un interés común en desarrollar una tecnología concreta de forma que se garantice la interoperabilidad de los productos de diversos fabricantes. Esto es lo que se conoce como foros industriales.

Los foros no pretenden competir con las organizaciones internacionales de estándares, sino cooperar con ellas y ayudarlas a acelerar su proceso, especialmente en la parte mas difícil, la que corresponde a la traducción de los documentos en implementaciones que funcionen en la práctica. Generalmente los foros trabajan en los mismos estándares intentando aclarar ambigüedades y definir subconjuntos de funciones que permitan hacer una implementación sencilla en un plazo de tiempo mas corto y comprobar la viabilidad y la interoperabilidad entre diversos fabricantes; así los organismos de estandarización pueden disponer de prototipos reales del estándar que se está definiendo. En cierto modo es como traer a la ISO e ITU-T el estilo de funcionamiento de la IETF.

Otra característica de los foros es que se establecen fechas límite para la producción de estándares, cosa que no hacen los organismos oficiales; de esta manera los fabricantes pueden planificar la comercialización de sus productos de antemano, ya que saben para que fecha estarán fijados los estándares necesarios.

Entre las tecnologías que se han estandarizado o se están estandarizando por este procedimiento están frame relay, ATM, ADSL (Asymmetric Digital Subscriber Loop) y algunas variantes de Ethernet de alta velocidad, como el *gigabit Ethernet forum* que está especificando las características de una versión de Ethernet a 1 Gb/s. El *ATM forum* , creado en 1991 por Northern Telecom, Sprint, Sun Microsystems, y Digital Equipment Corporation (DEC), cuenta en la actualidad con mas de 500 miembros

1.8.5.- Otras organizaciones

El IEEE (Institute of Electrical and Electronics Engineers) es una asociación profesional de ámbito internacional. Aparte de otras muchas tareas el IEEE (también llamado IE cubo) tiene un grupo sobre estandarización que desarrolla estándares en el área de ingeniería eléctrica e informática. Entre estos se encuentran los estándares 802 que cubren prácticamente todos los aspectos relacionados con la mayoría de los sistemas habituales de red local. Los estándares 802 han sido adoptados por ISO con el número 8802.

El ANSI es como ya hemos dicho la organización de estándares de los Estados Unidos. La única razón de mencionarlo aquí es porque a menudo sus estándares son adoptados por ISO como estándares internacionales.

El NIST (National Institute of Standards and Technology) es una agencia del Departamento de Comercio de los Estados Unidos., antes conocido como el NBS (National Bureau of Standards). Define estándares para la administración de los Estados Unidos.

El ETSI (European Telecommunications Standards Institute) es una organización internacional dedicada principalmente a la estandarización de las telecomunicaciones europeas. Es miembro de la ITU-T. Entre sus misiones está elaborar especificaciones detalladas de los estándares internacionales adaptadas a la situación de Europa en los aspectos históricos, técnicos y regulatorios.

La EIA (Electrical Industries Association) es una organización internacional que agrupa a la industria informática y que también participa en aspectos de la elaboración de estándares.

La ECMA (European Computer Manufacturers Association), creada en 1961, es un foro de ámbito europeo donde expertos en proceso de datos se ponen de acuerdo y elevan propuestas para estandarización a ISO, ITU-T y otras organizaciones.

La CEPT (Conference European of Post and Telecommunications) es una organización de las PTTs europeas que participa en la implantación de estándares de telecomunicaciones en Europa. Sus documentos se denominan Norme Europeene de Telecommunication (NET). La CEPT esta avalada por la Comunidad Europea.