

REDES DE COMPUTADORES PARA INGENIEROS EN INFORMÁTICA

José Ángel Berná Galiano
Manuel Pérez Polo
Luis Miguel Crespo Martínez

REDES DE
COMPUTADORES
PARA INGENIEROS
EN INFORMÁTICA

UNIVERSIDAD DE ALICANTE

© J.A. Berná Galiano, M. Pérez Polo y L.M. Crespo Martínez, 2002

© de la presente edición

Publicaciones de la Universidad de Alicante

Campus de San Vicente s/n

03690 San Vicente del Raspeig

Publicaciones@ua.es

<http://publicaciones.ua.es>

Diseño de portada: Alfredo Candela

Impresión: Compobell S.L.

C/. Palma de Mallorca, 4 - bajo

30009 Murcia

I.S.B.N.: 84-7908-664-5

Depósito Legal: MU-726-2002

Reservados todos los derechos. No se permite reproducir, almacenar en sistemas de recuperación de la información ni transmitir alguna parte de esta publicación, cualquiera que sea el medio empleado —electrónico, mecánico, fotocopia, grabación, etc.—, sin el permiso previo de los titulares de los derechos de la propiedad intelectual.

ÍNDICE

	Página
Capítulo 1. Introducción.....	11
1.1 Las redes de computadores. Conceptos básicos.....	11
1.2 Evolución histórica.....	12
1.3 Tipos de redes.....	14
1.4 Arquitectura de red.....	22
1.5 Modelo de referencia OSI/ISO.....	24
1.6 Modelo de referencia TCP/IP.....	34
1.7 Interconexión de redes.....	38
Capítulo 2. Especificación y validación de protocolos.....	41
2.1 Especificación formal e informal.....	41
2.2 Máquinas de estado finito (MEF)	42
2.3 Redes de Petri.....	48
2.4 Otros métodos formales de especificación de protocolos.....	51
2.5 Verificación y validación de protocolos.....	52
Capítulo 3. Transmisión de señales.....	53
3.1 Introducción.....	53
3.2 Análisis de señales con series de Fourier. Ancho de banda.....	55
3.3 Velocidad de transmisión. Teorema de Nyquist.....	61
3.4 Distorsión en el medio de transmisión.....	65
3.5 Ruido. Tipos. Teorema de Shannon.....	67
3.6 Filtrado de señales.....	68
Capítulo 4. Señalización de la información.....	71
4.1 Señalización en banda base.....	71
4.2 Señalización en banda modulada.....	74
4.3 Multiplexión.....	89
4.4 Modalidades de transmisión.....	92
Capítulo 5. Medios de transmisión.....	95
5.1 Cables eléctricos.....	95
5.2 Fibra óptica.....	100
5.3 Ondas electromagnéticas.....	111
5.4 Técnicas de compartición del medio físico en LAN's.....	120

Capítulo 6. Funciones del nivel de enlace.....	125
6.1 Introducción.....	125
6.2 Servicios y funciones del nivel de enlace.....	126
6.3 Delimitación de tramas.....	128
6.4 Direccionalamiento.....	132
6.5 Detección y corrección de errores.....	132
Capítulo 7. Control del flujo.....	147
7.1 Introducción.....	147
7.2 Protocolo unilateral no restringido.....	147
7.3 Protocolo unilateral de parada y espera.....	148
7.4 Protocolo unilateral de parada y espera. Canal con errores.....	149
7.5 Protocolo bilateral de parada y espera piggyback.....	153
7.6 Protocolos de ventana deslizante.....	154
7.7 Rendimiento de protocolos.....	161
7.8 Cadencia eficaz.....	163
Capítulo 8. Protocolos de nivel de enlace.....	171
8.1 Protocolo HDLC.....	171
8.2 Protocolo PPP.....	175
Capítulo 9. Funciones del nivel de red.....	179
9.1 Introducción.....	179
9.2 Servicios proporcionados a la capa de transporte.....	180
9.3 Organización interna de la capa de red.....	183
Capítulo 10. Algoritmos de encaminamiento y control de la congestión.....	187
10.1 Introducción.....	187
10.2 Algoritmos de encaminamiento.....	187
10.3 Algoritmos de control de la congestión.....	196
10.4 Análisis de congestión. Teoría de colas.....	199
Capítulo 11. Protocolo de nivel de red IP.....	207
11.1 Fundamentos del protocolo IP.....	207
11.2 Interconexión de redes empleando IP.....	212
11.3 Protocolo de encaminamiento RIP.....	213
11.4 Protocolos de encaminamiento OSPF y BGP.....	214
11.5 Tunneling.....	215
11.6 Seguridad en redes a nivel IP.....	216
Bibliografía.....	219

PRÓLOGO

*Este texto es fruto del trabajo de los autores en la elaboración de unos contenidos sobre **Redes de Computadores** para estudiantes de la titulación de ingeniería en informática. Su experiencia acumulada en el mundo profesional de los sistemas de comunicación ha resultado una pieza clave para elaborar un libro que contenga una fundamentación teórica básica que permita afrontar al futuro Ingeniero en Informática los problemas relativos a transmisión de datos entre computadores. Estos problemas han de ser abordados con una metodología que el ingeniero ha de aplicar en la resolución de situaciones reales. Por ello se hace especial hincapié en proporcionar métodos que permitan cuantificar el rendimiento de los sistemas de comunicación por computador y cuáles son las acciones más idóneas para su mejora y optimización.*

El mundo profesional de las comunicaciones por computador está irremediablemente asociado a ingenieros en informática y de telecomunicación. Para establecer unos contenidos más adecuados a Ingenieros en Informática y que les proporcione una mayor competitividad profesional, este libro se centra en el estudio de la operatividad de las redes de computadores y su optimización. En primer lugar se introducen conceptos relativos a la arquitectura de red, elemento clave en la modelización del funcionamiento de las comunicaciones entre computadores, analizando dos modelos en concreto: el modelo OSI de la ISO y el modelo TCP/IP de Internet. En este apartado se relacionan los elementos de la arquitectura de red con los recursos del computador que emplea. La elaboración de algoritmos que solucionan problemas de computación es uno de los pilares de la ingeniería en informática, y por ello se aplicará a la modelización del funcionamiento de protocolos.

El estudio de los niveles inferiores de la arquitectura de red se realiza desde el punto de vista del ingeniero en informática: un análisis cualitativo del nivel físico para proporcionar una base mínima de conocimientos (un campo más propio de los ingenieros de telecomunicación) y un análisis exhaustivo y profundo de los algoritmos de los protocolos empleados en los niveles de enlace y red, donde un ingeniero en informática puede aplicar sus conocimientos de optimización para mejorar el rendimiento de las comunicaciones por computador en los niveles más inferiores de la arquitectura de red. En la actualidad, la modelización y optimización del funcionamiento interno de un router (elemento clave en la interconexión de redes de computadores y en definitiva un computador de propósito específico) es un campo de estudio abierto y promovido por importantes compañías de comunicaciones, y ahí el futuro ingeniero en informática

puede hacer valer su formación y preparación en la teoría de la computación.

*Por último, indicar que esta obra tiene como objetivo proporcionar una guía docente para que los estudiantes de ingeniería en informática puedan aprovechar al máximo el desarrollo de la asignatura **Redes de Computadores** que se imparte en un cuatrimestre (30 horas de clases teóricas) dentro de los estudios de Ingeniería en Informática de la Escuela Politécnica Superior de la Universidad de Alicante. Agradeceríamos a los lectores que nos proporcionaran sus comentarios e impresiones que puedan ayudar a la mejora del libro.*

Los autores, noviembre de 2001

CAPÍTULO 1. INTRODUCCIÓN

1.1 LAS REDES DE COMPUTADORES. CONCEPTOS BÁSICOS

El término **redes de computadores** hace referencia a una *colección de computadores autónomos capaces de transmitir información entre ellos empleando un medio de comunicación*. De especial importancia es el término **computador autónomo**, pues los sistemas maestro - esclavo, en los que un terminal "tonto" se conecta a un mainframe (los computadores del tamaño de habitaciones que en la década de los sesenta se empleaban para el proceso de datos) no se consideran una red de computadores.

Las redes de computadores se engloban dentro del campo de los sistemas de telecomunicación, considerando únicamente los sistemas de transmisión de datos entre computadores. La telecomunicación o comunicación a distancia se inicia en la primera mitad del siglo XIX con la invención del telégrafo. Es el primer sistema que permite transmisión de información a distancia y en ese momento se empieza a desarrollar todo un campo tecnológico que se denomina **telecomunicaciones**. Desde la invención del telégrafo, y sobre todo después, con la invención del teléfono a finales del siglo XIX, los sistemas de comunicación a distancia se orientan a la transmisión de voz. No es sino hasta mediados del siglo XX, con el desarrollo del computador y su aplicación al procesamiento masivo de datos, cuando los sistemas de telecomunicación se orientan, además de a la transmisión de voz a la transmisión de datos. En el momento en que varios computadores autónomos se interconectan entre sí para intercambiar información surgen las **redes de computadores**. El estudio de estas redes engloba varias disciplinas, como son la teoría de transmisión de señales, la electrónica de procesado y el *software* de comunicación. En este libro, dada su orientación a estudiantes de una titulación de Ingeniería en Informática, se analizarán de forma conceptual los aspectos relativos a transmisión de señales y medios físicos de transmisión, orientándolos hacia la arquitectura software de comunicación empleada en las redes de computadores. Se estudiarán aspectos relativos a la arquitectura de red, protocolos y servicios, pilares en el diseño de una red de computadores, pasando al

estudio concreto de dos arquitecturas de red: el modelo **OSI** de la *Organización Internacional de Estándares* y el modelo **TCP/IP** de *Internet*.

1.2 EVOLUCIÓN HISTÓRICA

Los **primeros proyectos de envergadura en la transmisión de datos** entre computadores datan de la **década de 1960**. El proyecto **SAGE** (*service automatic ground environment*), desarrollado por las fuerzas aéreas de los EEUU, realizó la primera interconexión de varios centros de tratamiento de datos. **Posteriormente**, en un entorno experimental y con patrocinio militar surge la red ARPA o **ARPANET**. Patrocinada por el DoD (*department of defense*) de los EEUU en plena guerra fría (1970), pretende construir un sistema de comunicación entre los centros de defensa capaz de resistir un ataque nuclear. Para ello se diseña una arquitectura redundante, con múltiples caminos entre origen y destino, de forma que en caso de que la comunicación entre dos centros se viera interrumpida, la información sería capaz de llegar a su destino buscando caminos alternativos. Esta arquitectura se basa en la técnica denominada **comutación de paquetes** o **mensajes** y que posteriormente se analizará más en detalle.

A **mediados de los setenta** aparecen las primeras arquitecturas de sistemas distribuidos de mano de **empresas informáticas** como son **IBM**, que propuso la arquitectura **SNA** (1974), y **Digital**, cuya propuesta se denominó **DNA** (1976). En ese mismo año, 1976, se normalizan a través del organismo internacional **CCITT** (Comité Consultivo Internacional Telegráfico y Telefónico) las redes de comutación de circuitos, denominadas **X.25**. Este organismo modificará su nombre en 1993, pasándose a **ominar en la actualidad ITU (International Telecommunications Union)** Unión Internacional de Telecomunicaciones. Las primeras redes que empiezan a funcionar bajo estas recomendaciones en 1978 son Transpac en Francia y Datapac en Canadá (X.25) y la red Nómada de comutación de circuitos (X.21). Sin embargo, en **España** ya había un antecedente anterior de red de **comutación de paquetes**, la **R.E.T.D.** (Red Especial de Transmisión de Datos) de Telefónica que comenzó su operación en 1972.

Hasta finales de los setenta no existe una normalización a la hora de desarrollar sistemas de comunicación entre computadores. Los estándares surgen *de facto* a partir de los desarrollos de las grandes compañías informáticas, que son incompatibles con los de sus competidores. Se plantea entonces la necesidad de establecer una arquitectura de **red única** a partir de la cual todos los fabricantes desarrollen sus productos, de forma

que exista compatibilidad entre los diferentes sistemas. Con esta filosofía de trabajo en 1977 la **ISO (International Standards Organization)** - Organización Internacional de Normas) crea el subcomité ISO/TC 97/SC16, que inicia los trabajos para la modelización de arquitecturas y comportamientos normalizados en la comunicación entre computadores. El resultado del trabajo del subcomité fue el desarrollo de una serie de **estándares** denominado Modelo Básico de Referencia para la Interconexión de Sistemas Abiertos (**OSI**, *open systems interconnection*). El documento, publicado en 1984, se denomina **IS/7498** y al mismo tiempo fue aceptado por la entonces CCITT como **X.200**.

La década de los ochenta supondrá una revolución en el desarrollo de las comunicaciones por computador debido en gran medida al espectacular avance en el desarrollo tecnológico. Es en este momento cuando se inician los estudios de un concepto al que en la actualidad se tiende: la **integración de la información**. Surge así la idea de la creación de una Red Digital de Servicios Integrados, **RDSI**, donde toda la información que en ese momento se transmite en las telecomunicaciones; voz, vídeo, datos, fax, se desea integrar en una única comunicación digital. Este concepto tiene su **realización** a partir de principios de la década de 1990, con la aparición en Europa por un lado y en América y Japón por otro, de las primeras redes digitales integradas, hasta la actualidad donde ya existe una RDSI consolidada y que, como veremos más adelante, quizás ya esté desfasada.

Este concepto de integración de los diferentes tipos de información en una sola de tipo digital y basándose en el modelo de referencia **OSI** se convierte en el motor del avance en el campo de las telecomunicaciones (telefonía fija y móvil, fax) hasta mediados de los noventa. Sin embargo, ¿qué ha sucedido en el campo de las comunicaciones por computador? El modelo OSI ha resultado una arquitectura demasiado complicada y cara a la hora de introducirla en las redes de computadores. Los fabricantes de redes (IBM, Novell, Digital,...) se niegan a aceptar este estándar en su totalidad y hacen una 'readaptación' de sus productos al modelo OSI de forma parcial. Y a principios de los noventa un fenómeno surge en EEUU y se propaga con rapidez a Europa: **Internet**. La antigua red de comunicación académica ARPANET, basada en el conjunto de protocolos TCP/IP, proporciona una serie de ventajas que la hacen popular. Se trata de una red flexible a la situación tecnológica de las comunicaciones y proporciona una serie de aplicaciones muy útiles. Comienza a difundirse con rapidez el correo electrónico (**E-mail**), el acceso a servidores de ficheros (**ftp**), el acceso a información en hipertexto (**WWW**), los grupos de discusión de noticias (**news**)... Y el proceso es imparable. La arquitectura de red de Internet, que adopta algunas de las características de OSI se convierte en el estándar de

las redes de computadores, y por otro lado las compañías de telecomunicaciones tradicionales permanecen ancladas al modelo OSI.

Y se alcanza la situación actual, finales de los noventa, principios del nuevo milenio. El desarrollo tecnológico permite la existencia de redes a velocidades de los cientos de Mbps (millones de bits por segundo) como **ATM** (*asynchronous transfer mode*), en contraste con las velocidades de cientos de Kbps (miles de bits por segundo) de principios de la misma década. Las compañías de telecomunicaciones emprenden un giro inesperado: se plantean ofrecer sus servicios tradicionales empleando la arquitectura de red de Internet. El futuro está en ofrecer servicios integrados a través de redes TCP/IP y se ponen en marcha los primeros prototipos. Ya se está experimentando la telefonía en una red sin caminos fijos entre origen y destino (en contraste con la telefonía tradicional de conmutación de circuitos), la denominada **voz bajo IP (VoIP)**.

1.3 TIPOS DE REDES

La clasificación de los diferentes tipos de redes de computadores puede realizarse ateniendo a diversos aspectos, siendo dos los más sobresalientes:

- 1.1 La forma de interconexión de las estaciones en la red.
- 1.2 La escala geográfica de la red.

1.3.1 Redes de difusión y redes punto a punto.

Esta clasificación está basada en las diferentes características que presentan ambos tipos de tecnología de transmisión. En las **redes de difusión** existe un único canal de comunicación compartido por todas las máquinas de la red. Los mensajes transmitidos por una máquina de la red al medio de comunicación son recibidos por todas las demás. Dentro del mensaje existe un campo que indica a qué máquina va dirigida el mensaje, el campo de dirección. La máquina cuya dirección sea la de destino del mensaje enviado será la encargada de procesarlo, descartándolo el resto.

Los sistemas de difusión también suelen ofrecer la posibilidad de enviar un mensaje a todas las máquinas que se encuentran en la red de comunicación. Es lo que se conoce como **difusión (broadcasting)**, y se realiza indicando un código especial en el campo de dirección del mensaje para que todas las máquinas puedan procesarlo. Algunos sistemas contemplan también la posibilidad de enviar un mensaje a un subconjunto de máquinas dentro de la red, proceso denominado **multidifusión**. En este esquema se reserva un bit del campo de dirección en el mensaje para indicar multidifusión, siendo

los $n-1$ bits restantes del campo de dirección empleados para direccionar grupos de máquinas en la red.

En contraste con las redes de difusión, en una red punto a punto se establecen múltiples conexiones entre pares individuales de máquinas. Para realizar la transmisión de un mensaje entre origen y destino, es posible que dicho mensaje deba visitar máquinas intermedias para llegar a su destino. Dado que es posible que en estas redes puedan existir diferentes rutas o caminos para llegar de un origen a un destino, son de especial importancia los algoritmos de encaminamiento, los cuales se analizarán en el capítulo 10. Una de las ventajas que presentan estas redes frente a las redes de difusión es que el fallo del medio de transmisión entre un par de máquinas no implica necesariamente la incomunicación entre ambas máquinas, debido a la existencia de rutas alternativas. Por el contrario las redes punto a punto tienen un coste económico mucho mayor que las redes de difusión al precisar de mayor cableado.

1.3.2 Redes LAN, MAN y WAN.

La escala geográfica de la red de computadores es otro de los aspectos a la hora de realizar una clasificación de las mismas. Las redes de área local (*Local Area Networks, LAN*) son redes privadas que se encuentran en un solo edificio o campus y se extienden hasta distancias de unos pocos kilómetros. Este tipo de redes son las de menor coste económico y están diseñadas para su empleo en la interconexión de ordenadores personales y compartir recursos. Las velocidades de transmisión van desde los 10 a los 100 Mbps (10^6 bits por segundo). Además, las LAN pueden adoptar distintas topologías, siendo las más frecuentes el bus común y el anillo. Como ejemplo de bus común tenemos la red Ethernet (IEEE 802.3), en la que todas las máquinas que conforman la red están conectadas a un mismo bus de comunicación común consistente en un cable coaxial. Esta configuración da lugar a la aparición de colisiones cuando dos máquinas tratan de transmitir información simultáneamente. Empleando la topología en anillo está la red Token Ring (IEEE 802.5) donde las máquinas se encuentran conectadas en serie completando una circunferencia. De esta forma, la transmisión de un mensaje de una máquina a otra se realiza mediante el paso del mismo de máquina a máquina hasta llegar a su destino.

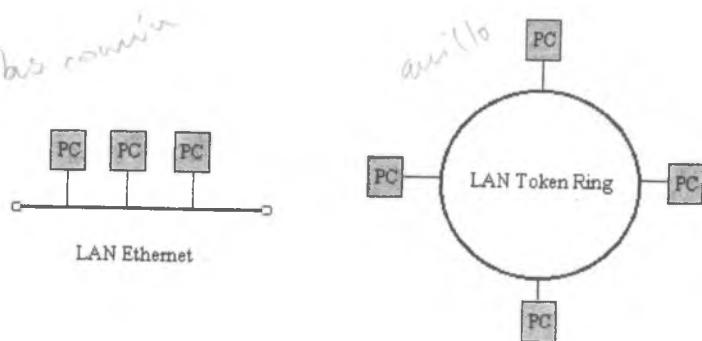


Figura 1.1 Topologías de redes de área local (LAN).

Las **redes de área metropolitana** (*Metropolitan Area Networks, MAN*) son básicamente una versión ampliada en extensión de las **LAN** que emplean una tecnología similar. Puede abarcar un conjunto de edificios e incluso una ciudad y pueden ser tanto de ámbito público como privado. En cuanto a la aplicación, están muy relacionadas con la transmisión de datos, voz y vídeo. Existe un estándar para este tipo de redes denominado **bus dual de cola distribuida** (*distributed queue dual bus, DQDB*), normalizado por la IEEE con el estándar 802.6. El DQDB consiste en un par de buses unidireccionales a los cuales están conectados todas las computadoras, según la siguiente figura.

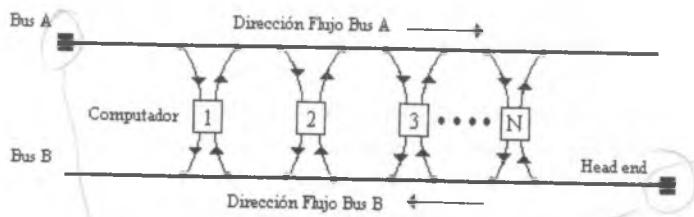


Figura 1.2 Arquitectura de la red de área metropolitana DQDB.

Cada bus tiene una cabeza terminal (*head-end*), un dispositivo que inicia la actividad de transmisión. El tráfico destinado a una computadora situada a la derecha del emisor usa el bus superior. El tráfico hacia la izquierda usa el de abajo.

Las **redes de área extensa** (*Wide Area Networks, WAN*) son redes que se extienden a lo largo de un área geográfica extensa, como puede ser un país o un continente. La arquitectura de este tipo de redes está basada en la interconexión de diferentes estaciones (*hosts*) remotos empleando un conjunto de nodos encaminadores (*routers*) interconectados entre sí empleando líneas punto a punto. Estos nodos encaminadores tienen como función encaminar la información de un *host* de origen a un *host* de destino. A cada uno de estos nodos puede conectarse uno o varios *hosts* remotos o una red de área local LAN.

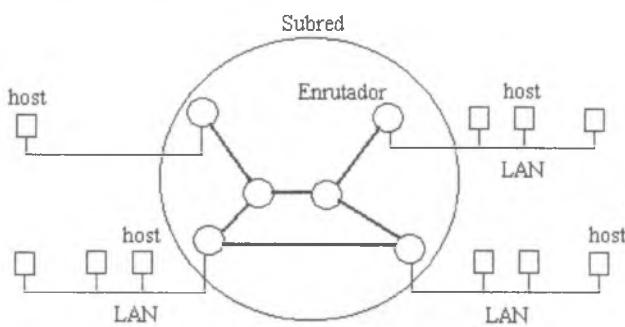


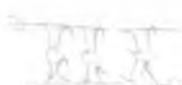
Figura 1.3 Esquema de una red WAN.

La topología de este tipo de redes es variada: en estrella, anillo, árbol, malla, irregular, etc., pero siempre empleando líneas punto a punto entre los nodos. Las velocidades de transmisión en este tipo de redes son más reducidas que en las LAN's, pues las comunicaciones remotas son más susceptibles de sufrir errores, siendo del orden de los kilobits por segundo (kbps). Es posible establecer diferentes clasificaciones dentro de las redes WAN.

5-10 kbps - 20 Mbps

Si se tiene en cuenta el ámbito de los datos que se van a transmitir en la red, pueden clasificarse en,

- a) **Redes públicas de datos**, aquellas cuyo moderador o gestor es un organismo o entidad pública, o aquellas cuya utilización está abierta a un público general. Es el caso de la red telefónica conmutada RTC y la red digital de servicios integrados RDSI.
- b) **Redes privadas de datos**, aquellas cuyo moderador o gestor es una entidad corporativa y la emplea para fines propios. Es el caso de las redes SNA de IBM o DNA de Digital.



Si se tiene en cuenta la forma en que se establece la comunicación en la red, pueden clasificarse en,

- Redes de conmutación de circuitos.**
- Redes de conmutación de paquetes.**

En las **redes de conmutación de circuitos**, la comunicación entre los equipos terminales de datos (*Data Terminal Equipment, DTE*) se establece empleando un camino fijo y dedicado a través de un canal físico de comunicación empleando conmutadores (*switches*). Es el caso de la RTC, en la cual se establecen caminos físicos mediante conmutadores en las centrales telefónicas.

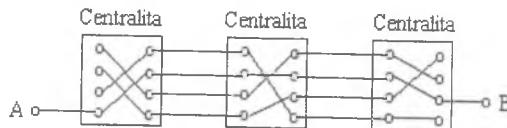


Figura 1.4 Esquema de conmutación de circuitos en la RTC.

Para realizar la transmisión de datos a través de la RTC es necesario emplear un dispositivo especial. Las líneas telefónicas analógicas están diseñadas para la transmisión de voz, por lo que para transmitir la información digital que emplea una computadora (dígitos binarios) es necesario añadir información digital a una señal analógica que se transmite por la RTC. Esta función la realiza el dispositivo denominado **módem** (**modulador-demodulador**), que transmite y recibe información digital empleando un medio de transmisión analógico.

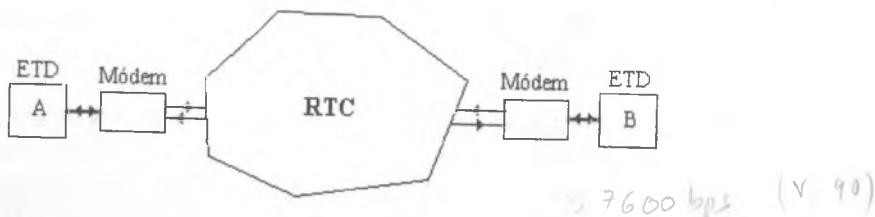


Figura 1.5 Esquema de conexión de dos ETD a través de la RTC.

Las características de este tipo de redes están limitadas por las posibilidades de la RTC empleada. La velocidad de transmisión máxima permitida depende del tipo de módem empleado y del ancho de banda de la

propia RTC. En el caso de España, es posible alcanzar velocidades teóricas de hasta 57600 bps, aunque la velocidad efectiva es menor. Algunos de los inconvenientes que presenta es la existencia de un tiempo inicial para el establecimiento del circuito físico y la falta del aprovechamiento total de la capacidad del canal. Además el control de los errores que se puedan producir durante la transmisión y el control del flujo de datos entre emisor y receptor son funciones de los DTE, siendo la RTC indiferente a estos aspectos.

Las **redes de conmutación de paquetes** presentan la característica de que la **información** a transmitir se fragmenta en unidades de información más pequeñas denominadas **paquetes o tramas**, cada una con un mismo formato común. Generalmente cada uno de estos paquetes consta de un campo de **cabecera**, donde se incorpora información acerca de donde va dirigido el paquete y como ha de ser recomposto el mensaje, y un campo de **datos**, donde se incorpora la información a transmitir en sí.



Figura 1.6 Esquema de un paquete de información.

A la hora de establecer la comunicación entre emisor y receptor las redes de conmutación de paquetes pueden elegir entre dos estrategias, dando lugar a diferencias en el funcionamiento de la **subred**.

La primera de ellas es el establecimiento de comunicación empleando **circuitos virtuales**. Este tipo de comunicación es una **analogía con el sistema telefónico de conmutación de circuitos**. En la transmisión de paquetes de información de origen a destino se establece un **camino virtual** en la red, de forma que todos los paquetes siguen el mismo camino durante la comunicación. Una vez liberada la comunicación, si se produjera otra nueva el **camino virtual** asignado no tiene porque ser el mismo. Este esquema presenta más retardos que la conmutación de circuitos, pues un nodo de una red de conmutación de paquetes es un dispositivo de computación que precisa de recursos software y sufre **congestión**, mientras que un nodo en una red de conmutación de circuitos es un dispositivo de conmutación electrónico que tiene una capacidad física determinada y sufre **saturación**. El ancho de banda del medio se aprovecha mejor, pues el medio físico esta compartido para varias comunicaciones y no dedicado a una sola. La calidad de la comunicación es muy alta, presentando un bajo

índice de errores al realizar una transmisión de los paquetes en orden secuencial. Por último, destacar que los circuitos virtuales pueden ser permanentes o no permanentes. En los permanentes, la tarificación por el uso depende del volumen de información transmitido, mientras que en los no permanentes depende del tiempo empleado en la comunicación.

La otra modalidad de comunicación son los datagramas. En este esquema los paquetes de información contienen información acerca de su destino, pero no del camino a seguir a través de la subred. Es ésta la que se encarga de encaminar los paquetes por un camino u otro dependiendo de los recursos disponibles. Por ello, es posible que los paquetes lleguen a su destino desordenados o incluso que no lleguen, por lo que no es un método de transmisión de datos que se pueda emplear de forma fiable. Más bien, su uso está orientado a la señalización (información que circula en la red notificando eventos) y datos de control de la red. Sin embargo, si los DTE disponen de funciones de recuperación de errores y reorganización de la información pueden emplear este esquema para la transmisión fiable de datos, aunque con unas velocidades de transmisión efectivas inferiores a los circuitos virtuales. Este es el esquema empleado en la Internet mundial.

Como ejemplo de red de conmutación de paquetes puede analizarse el caso de la Red X.25 en España. X.25 es una estándar para una red de conmutación de paquetes de tecnología digital que engloba un conjunto de normas distintas. El origen de X.25 en España hay que buscarlo en la década de los 70, con la aparición de la primera red de datos por conmutación de paquetes, y que se denominó RETD (Red Especial de Transmisión de Datos). Hacia los años 80 se la rebautizó con el nombre de IBERPAC, realizando una evolución que lleva a la aparición de dos redes independientes:

- a) Red IBERPAC RSAN, que emplea un protocolo (conjunto de normas para realizar una comunicación) propietario de Telefónica S.A. y no admite más asociados.
- b) Red IBERPAC X.25, que cumple con la recomendación X.25 de la ITU.

Para conectarse con una computadora a la red X.25 es necesario un dispositivo denominado PAD (ensamblador-desensamblador de paquetes). El esquema de comunicación en una red X.25 así como los protocolos de comunicación empleados se muestran en la siguiente figura.

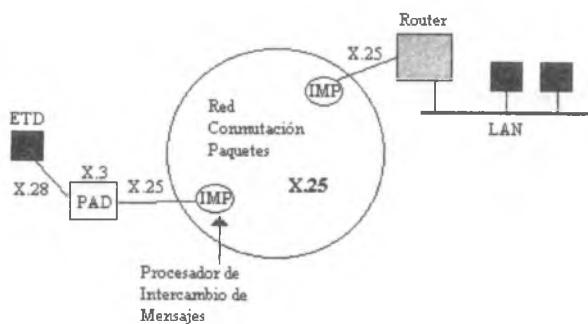


Figura 1.7 Esquema de una red X.25.

Como ejemplo de una red de datos pública se analizará el caso de la RDSI, Red Digital de Servicios Integrados (*integrated services digital network, ISDN*). A finales de los 80, la CCITT establece una normativa para el desarrollo de la RDSI. El objetivo fundamental de esta red es la integración de todos los servicios (datos, voz, imágenes, fax) de las redes de comunicación en una única red de transmisión de información digital. Los servicios disponibles en esta red son:

- Telefonía digital, en la que se realiza una codificación digital de la voz.
- Audioconferencia, conversaciones simultáneas entre varias personas.
- Teletex, correo electrónico de uso doméstico y negocios.
- Fax.
- Videotex, acceso a bases de datos remotas.
- Videotelefonía, llamadas telefónicas con soporte de imágenes.
- Videoconferencia, comunicaciones con soporte de sonido y vídeo entre varias personas.

La RDSI, que tiene una arquitectura física distinta en Europa que en EEUU y Japón, se proporciona al usuario final en un paquete denominado **acceso básico**, que proporciona en el terminal telefónico dos canales **tipo B** de transmisión de datos a 64 Kbps cada uno, y uno **tipo D** de señalización a 16 Kbps. En Europa, para grandes corporaciones se proporciona el **acceso primario** que consta de 30 canales de tipo B y uno de tipo D a 64 Kbps.

En un principio, la RDSI se implanta sobre la actual RTC, quedando para un proyecto futuro la creación de una red RDSI pura. En la actualidad, el auge de las redes TCP/IP y la aparición de tecnologías con mayores

velocidades de transmisión como ATM y ADSL (*asymmetric digital subscriber line*), hacen que el futuro de la tecnología RDSI actual sea incierto, aunque la filosofía de la integración digital se mantiene.

1.3 ARQUITECTURA DE RED

El concepto de **Arquitectura de una Red de comunicaciones** hace referencia a un **conjunto de protocolos** perfectamente definidos e **implementados** que caracterizan cómo se realiza la **transmisión de datos en una red de comunicaciones**. Este concepto se concreta aplicándolo a diferentes redes de comunicaciones: redes de computadores, redes telefónicas fijas y móviles, sistemas de comunicación vía satélite, etc.

La arquitectura de red **se define** en una serie de **capas o niveles** que **interaccionan entre sí**, de forma que se facilita el diseño de la red y la gestión de fallos en el funcionamiento. El objetivo de cada una de esas **capas** es proporcionar un conjunto de servicios a la capa superior. Véase a continuación un ejemplo de arquitectura de red de 3 niveles.

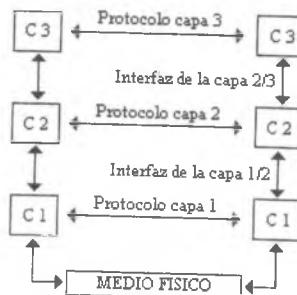


Figura 1.8 Arquitectura de red de 3 niveles o capas.

La capa *n* conversa con la capa *n* del otro extremo de la comunicación empleando unas normas definidas en los **protocolos**. A cada una de estas capas que dialogan a un mismo nivel dentro de la arquitectura se las denomina **entidades pares**. Sin embargo, esta comunicación horizontal entre entidades pares es **virtual**. La comunicación **real** se produce enviando información de datos y control de las capas superiores a las inferiores (comunicación **vertical**) alcanzando el nivel físico, donde la información digital se transmite empleando señales que se propagan por un medio físico. En el *host* del otro extremo de la comunicación se realiza el proceso inverso. La **interfaz entre capas** es el elemento encargado de

definir las operaciones y servicios, y de realizar la comunicación entre capas adyacentes dentro de un mismo *host*.

Resumiendo

- Cada nivel es un **usuario de servicios** ofrecidos por el nivel inferior y **proveedor de servicios** al nivel superior.
- En la especificación del comportamiento de un determinado nivel, es necesario definir:
 - Servicios** que se ofrecen al nivel superior.
 - El protocolo** que se emplea para la comunicación entre entidades pares para proporcionar servicios a la capa superior.

En el siguiente ejemplo se estudia cual es el proceso de envío de un mensaje entre dos *hosts* empleando una arquitectura de red de 4 niveles.

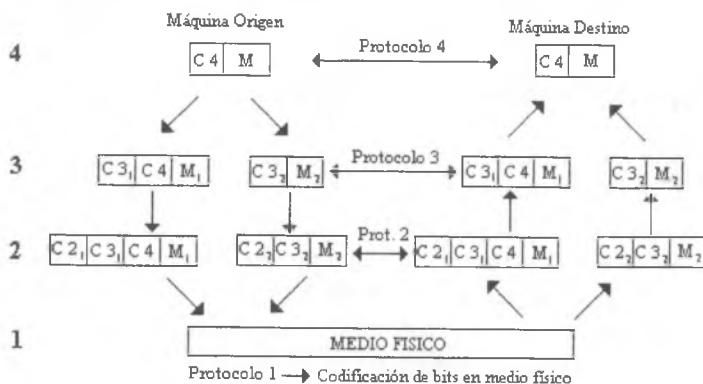


Figura 1.9 Ejemplo de comunicación entre dos máquinas empleando una arquitectura de red de 4 niveles.

En la estación (*host*) de origen, el usuario envía un mensaje M al usuario remoto que se encuentra en el *host* de destino. El nivel 4 de la arquitectura interacciona con el usuario leyendo el mensaje que será transmitido a la entidad par correspondiente empleando unas normas especificadas según el protocolo de nivel 4. Es necesario por tanto añadir al mensaje una información adicional, una **cabecera** de información C4, que especifica como tiene que ser esa comunicación horizontal. La unidad de información que se pasará al nivel inferior será por tanto C4+M, que es la información que se pretende llegue al nivel 4 del *host* destino. En el nivel 3, la

especificación del protocolo de comunicación no permite enviar mensajes largos, por lo que es necesario fragmentar los datos de la capa superior en paquetes más pequeños. Por tanto, el mensaje C4+M se divide en dos: C4+M₁ y M₂. A cada uno de estos fragmentos se les añade las cabeceras correspondientes al protocolo de nivel 3, C₃₁ y C₃₂. Ambos paquetes de información se envían al nivel inferior y, dado que no precisan fragmentación (se supone que el nivel 2 puede procesar el tamaño de los paquetes provenientes del nivel superior) se les añade las cabeceras correspondientes del protocolo C₂₁ y C₂₂. A continuación, se envía cada uno de los paquetes de información al nivel 1, el medio físico, que transmite de forma independiente cada uno de los paquetes empleando una codificación en señales adecuadas al medio físico. En el *host* del otro extremo de la comunicación, en cada nivel se eliminan las cabeceras de protocolo correspondiente y se reagrupan los paquetes en caso de ser necesario, proporcionando al usuario remoto el mensaje original. Nótese que en cada nivel pueden abstraerse los niveles inferiores y considerar la comunicación como si fuera horizontal, permitiendo que no sea necesario conocer el funcionamiento a bajo nivel de la red para poder interaccionar con ella.

1.4 MODELO DE REFERENCIA OSI/ISO

El modelo **OSI** (*Open Systems Interconnection*, Interconexión de Sistemas Abiertos) es una normativa internacional de la **ISO** (*International Standards Organization*, Organización Internacional de Normas). Hasta el año 1984 no existía una normativa internacional en arquitectura de red. Cada fabricante de hardware de red diseñaba la suya propia dando lugar a sistemas propietarios cuya interconexión a otros era cara y compleja. En 1984, la ISO establece una arquitectura de red estándar y abierta basada en el modelo de arquitectura de red **SNA de IBM**. Este modelo tiene **7 capas** y la elección de las mismas se basó en una serie de criterios:

- a) Una capa se define cuando se necesita un **nivel** diferente de **abstracción**.
- b) Cada **capa** tiene una **función** bien definida.
- c) La **función** de cada capa se define con la intención de crear protocolos reconocidos **internacionalmente**.
- d) Los límites entre cada capa se eligen de forma que el **flujo** de información en las **interfaces** sea **mínimo**.
- e) El **número** de **capas** debe ser elevado para que éstas sean lo más **independientes** posible, y pequeño para que sean de fácil manejo.

1.4.1 Niveles del modelo OSI

Véase a continuación el conjunto de los 7 niveles del modelo OSI, del nivel inferior al nivel superior.

1.4.1.1 Capa 1. Nivel físico.

El nivel físico en el modelo OSI describe la forma en que se realiza la transmisión y recepción de secuencias de bits a través de un canal de comunicación (medio de transmisión). A este nivel se definirán:

1. La especificación de las conexiones mecánicas y eléctricas.
2. La especificación de velocidades de transmisión estándar.
3. La especificación de la codificación de los bits en señales adecuadas al medio físico de transmisión.
4. La especificación de la sincronización emisor – receptor a nivel de bits.
5. La especificación de la modalidad de transmisión: simplex, dúplex, semidúplex.
6. La especificación de control de errores a nivel físico.

1.4.1.2 Capa 2. Nivel de Enlace.

El nivel de enlace tiene como función proporcionar al nivel de red (capa 3) una línea de comunicación libre de errores de transmisión. Para conseguir este objetivo, el nivel de enlace realiza una serie de funciones:

1. Los datos a transmitir procedentes del nivel de red se dividen en fragmentos o tramas con información de control.
2. La información se transmite secuencialmente y se numera para reagruparla.
3. Se emplea el reconocimiento de la recepción de la información.
4. Se realiza el reenvío de tramas perdidas.

1.4.1.3 Capa 3. Nivel de Red.

El nivel de red tiene como función controlar la operatividad de la subred (conjunto de nodos o hosts que conectados entre sí conforman la topología de la red) gestionando el flujo de paquetes que se encaminan de origen a destino. Este nivel es el encargado también de resolver los problemas de

interconexión de redes heterogéneas. Genéricamente, los métodos de control del flujo se clasifican en tres tipos:

1. *Control de flujo estático.* En cada nodo de la subred se especifica una tabla con caminos fijos dependiendo de la dirección de destino.
2. *Control de flujo dinámico.* En cada nodo de la red los paquetes de información se encaminan hacia el nodo menos congestionado.
3. *Control de flujo mixto.* Al iniciarse una comunicación se averigua el camino óptimo al destino y se mantiene fijo hasta el fin de la misma.

1.4.1.4 **Capa 4. Nivel de Transporte.**

El nivel de transporte tiene como objetivo proporcionar una **comunicación libre de errores** entre dos máquinas remotas, independientemente de la estructura de subred que exista. El nivel de transporte inicia una conexión con el nivel de red cuando el nivel de sesión inicia una comunicación, y libera la conexión con el nivel de red cuando el nivel de sesión finaliza la comunicación. Es competencia también del nivel de transporte gestionar varias conexiones simultáneas con el *host* remoto, pudiendo así emplear caminos distintos en la subred y agilizar las comunicaciones.

1.4.1.5 **Capa 5. Nivel de Sesión.**

El nivel de sesión permite el **establecimiento de sesiones de comunicación de usuarios entre máquinas remotas**, como puede ser el acceso a un sistema de tiempo compartido, transferencia de ficheros, etc. Uno de los aspectos que se gestiona en el establecimiento de una sesión es la **administración del testigo**, que controla la ejecución simultánea de acciones por parte de usuarios remotos impidiendo la aparición de incongruencias en el sistema.

1.4.1.6 **Capa 6. Nivel de Presentación.**

El nivel de presentación tiene como función resolver el problema de la semántica y sintaxis de la información transmitida. Un ejemplo de ello se tiene en la codificación de los datos, que puede ser a diferentes niveles:

1. Nivel de texto: ASCII, EBCDIC,...
2. Nivel de palabras: codificación de bits,...
3. Nivel de números: C2, coma flotante,...

Otros aspectos son la compresión y descompresión de datos y la seguridad de la transmisión mediante criptografía(cifrado/descifrado).

1.4.1.7 Capa 7. Nivel de Aplicación.

El nivel de aplicación define un conjunto de protocolos que interactúan con las aplicaciones o el usuario final.

Un ejemplo de interacción con una aplicación se tiene en el protocolo de terminal virtual (VT). Cada sistema dispone de sus propios editores, por lo que a la hora de transmitir información de visualización de texto se producen problemas. VT se encarga de definir una conjunto de funciones de terminal virtual (por ejemplo posicionar el cursor en una posición de pantalla) que los terminales reales interpretan y realizan en base a las particularidades propias de cada sistema.

Como ejemplo de aplicaciones de usuario final está la transferencia de archivos o el correo electrónico, que están fuertemente ligadas al software de red existente.

1.5.2 Servicios y Protocolos. Unidades de Transferencia de Información.

1.5.2.1 Servicios

Un **servicio** se define como un conjunto de operaciones, denominadas **primitivas de servicio**, que una capa ofrece a la capa inmediata superior. La **realización** de un servicio conlleva la utilización de 1 o más primitivas de servicio. Los puntos en los que la capa $n+1$ tiene acceso a los servicios de la capa n se denominan **puntos de acceso al servicio** (*Service Access Point, SAP*). Cada SAP tiene una dirección única. Para clarificar un poco este último concepto se puede hacer una analogía con el sistema telefónico. En la red telefónica, el SAP es el terminal de conexión que la compañía telefónica instala en casa y la dirección del SAP es el número de teléfono que se nos asigna.

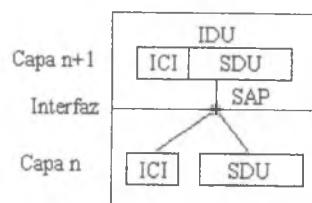


Figura 1.10 Comunicación vertical en un modelo de arquitectura basado en niveles.

Como se aprecia en la figura anterior, la comunicación entre capas para acceder a los servicios disponibles se realiza empleando unas **unidades de transferencia de información**. Cuando la capa $n+1$ quiere acceder a un servicio de la capa n , la capa superior envía a la inferior un paquete de información denominado **IDU**. IDU es la **Unidad de Datos de la Interfaz** y se define como un paquete de información que la capa $n+1$ envía a la capa n a través del **SAP**. La IDU esta compuesta por una **ICI** y una **SDU**. ICI es la **Información de Control de la Interfaz**, que no forma parte de los datos transferidos y ayuda a la comunicación entre capas para proveer un servicio. SDU es la **Unidad de Datos de Servicio**, que consiste en la información transferida entre capas para la realización del servicio.

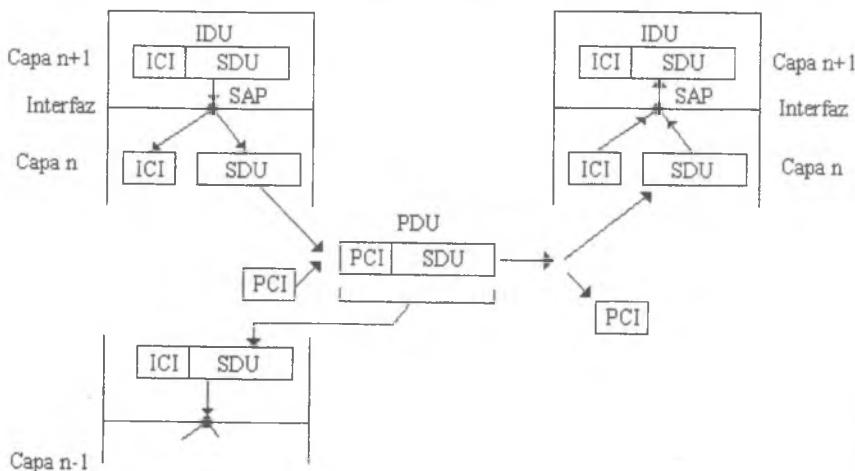


Figura 1.11 Comunicación horizontal en un modelo de arquitectura basado en niveles.

La capa n que recibe una petición de servicio de la capa $n+1$ necesita conversar con su entidad par de la capa n para proveer dicho servicio. Para ello se elimina la ICI de la IDU procedente de la capa $n+1$ y se le añade

una **PCI**. PCI es la **Información de Control del Protocolo** y contiene la información necesaria para poder realizar la comunicación entre las entidades pares de la capa n (numeración de tramas, direccionamiento, etc).

Al conjunto de la PCI y SDU (proveniente de la capa $n+1$) se le denomina **PDU, Unidad de Datos del Protocolo**. Esta PDU será la SDU, que junto a su ICI correspondiente, se enviará a la capa $n-1$ para que pueda ser transmitida a la entidad par correspondiente (recuérdese que la comunicación real es vertical).

En el otro extremo de la comunicación se realiza el proceso inverso, eliminando las cabeceras correspondientes.

A veces, la SDU de la capa $n+1$ que ha de transmitirse a la entidad par de la capa n es demasiado grande para encapsularla en una única PDU, por lo que se produce una **fragmentación** de la SDU en varias PDU, cada una con su correspondiente cabecera PCI. En el otro extremo se realizará la reconstrucción de la SDU original.

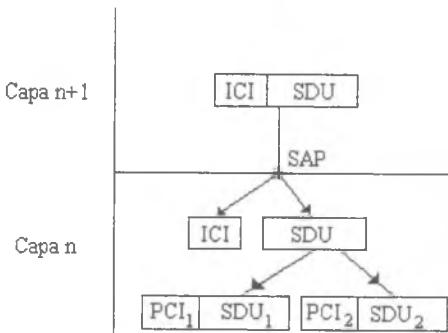


Figura 1.12 Fragmentación de una SDU en varias PDU.

Más aún, es posible realizar multiplexaciones en las conexiones entre entidades pares donde se intercambian PDU. Existen dos modalidades de multiplexación:

- Multiplexación hacia arriba.** Permite unir varias conexiones del nivel $n+1$ en una sola conexión en el nivel $n-1$, lo que permite un mejor aprovechamiento del medio físico de transmisión.

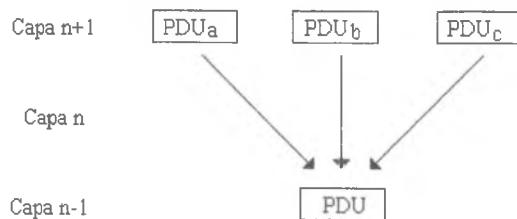


Figura 1.13 Multiplexación hacia arriba.

- Multiplexación hacia abajo.** Permite dividir una conexión del nivel $n+1$ en varias conexiones en el nivel $n-1$, produciendo una incremento en el rendimiento de la transmisión (se emplean distintos caminos en la comunicación).

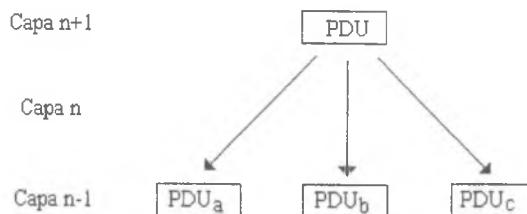


Figura 1.14 Multiplexación hacia abajo.

1.5.2.2 Protocolos.

Un protocolo de nivel n especifica la forma en que dos entidades pares de ese nivel intercambian información en forma de n -PDU (PDU de nivel n de la arquitectura de red) con la finalidad de ofrecer a sus usuarios el servicio de nivel n . Esta especificación consta de:

- Cómo se utilizan los servicios del nivel $n-1$ para realizar el intercambio de n -PDU.
- Qué n -PDU se intercambian, cuál es su misión y cómo están codificadas.

- c) Reglas de comportamiento en el diálogo.

Para conseguir un protocolo cuyo funcionamiento sea fiable y correcto, es preciso que éste cumpla con un conjunto de características determinadas:

- a) **Direccionamiento.** Cada capa necesita un mecanismo para poder identificar emisores y receptores en el proceso de comunicación.
- b) **Transferencia de datos.** Es necesario describir como va a ser la transferencia de datos entre emisor y receptor. Ésta puede ser:
 1. *Comunicación simplex.* La transferencia de datos es en un sólo sentido.
 2. *Comunicación semidúplex.* La transferencia de datos puede ser en los dos sentidos, pero no de forma simultánea.
 3. *Comunicación dúplex o full dúplex.* La transferencia de datos puede ser en los dos sentidos y de forma simultánea.
- c) **Control de errores.** Ambos extremos de la conexión precisan de algoritmos comunes de detección y corrección de errores. El receptor indicará al emisor los errores producidos para proceder con las retransmisiones necesarias.
- d) **Recepción de mensajes.** Es necesario un mecanismo que sea capaz de ordenar mensajes que puedan llegar desordenados debido a la fragmentación. Además ha de controlarse el flujo entre emisor y receptor para no saturar a los receptores lentos.
- e) **Multiplexación, Demultiplexación.** Un protocolo debe permitir que puedan establecerse varias conexiones simultáneas empleando una comunicación entre dos *hosts*.

1.5.2.3 Tipos de servicios.

En forma genérica existen dos tipos de servicios que las capas ofrecen al nivel superior:

1. **Servicio orientado a conexión.** En analogía con el modelo telefónico. El usuario establece una conexión, utiliza el servicio y la libera.
2. **Servicio no orientado a conexión.** En analogía con el modelo del sistema de correos. Cada petición del servicio se realiza cuando se precisa indicando el destino del mismo y sin atender otros aspectos como puede ser el estado de la comunicación.

En cuanto a la calidad del servicio que se ofrece, es posible distinguir dos modalidades:

1. **Servicio confiable.** El receptor realiza un acuse de la recepción del mensaje. Este tipo de servicio produce retardos, por lo que no se tolera para ciertas aplicaciones. Un ejemplo de ello puede ser el servicio de correo electrónico con acuse de recibo.
2. **Servicio no confiable.** No se asegura que el mensaje llegue a su destino y por tanto que el servicio pueda ser llevado a cabo. Un ejemplo sería el envío de correo electrónico ordinario.

Ambas clasificaciones de tipos de servicio pueden combinarse, dando lugar a servicios que son propios para determinadas aplicaciones y algunos de cuyos ejemplos se recogen en la siguiente tabla.

	Confiable	No Confiable
Conexión	Transferencia de un archivo de datos	Transmisión de voz en una red de commutación de paquetes
Sin Conexión	Correo electrónico con acuse de recibo	Correo electrónico ordinario

Tabla 1.1 Ejemplos de combinaciones de tipos de servicios.

1.5.2.4 Primitivas de Servicio.

Las primitivas de servicio son el conjunto de órdenes que definen un servicio. Existen cuatro tipos de primitivas:

1. **Petición o *request*.** Empleada cuando una entidad solicita la realización del servicio.
2. **Indicación o *indication*.** Empleada cuando se avisa a la entidad de un evento en la realización de un servicio.
3. **Respuesta o *response*.** Empleada cuando una entidad responde a un evento en la realización de un servicio.
4. **Confirmación o *confirmation*.** Empleada para confirmar a la entidad solicitante del servicio de la realización del mismo.

Dependiendo del número de primitivas asociadas a un servicio se distinguen dos categorías:

- 1. Servicios confirmados.** Aquellos en los que el usuario solicitante del servicio recibe confirmación de la realización del mismo.



Figura 1.15 Ejemplo de Servicio Confirmado.

- 2. Servicios no confirmados.** Aquellos en los que el usuario solicitante del servicio **no** recibe confirmación de la realización del mismo.

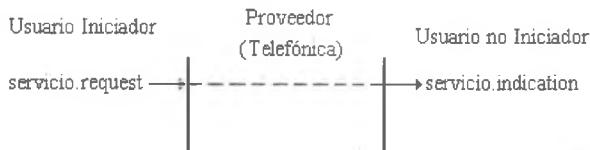


Figura 1.16 Ejemplo de Servicio No Confirmado.

Como ejemplo del funcionamiento de la realización de un servicio empleado primitivas de servicio se analizará el caso de una llamada telefónica de un usuario a otro en la que mantienen una conversación. Para ello se asociará al servicio 8 primitivas distintas que serán:

1. *Connect.request* – Petición para establecer una conexión.
2. *Connect.indication* – Envía una señal a la parte llamada.
3. *Connect.response* – La usa el receptor para aceptar o rechazar llamadas.
4. *Connect.confirm* – Indica al iniciador si se aceptó o no la llamada.
5. *Data.request* – Petición de envío de datos.
6. *Data.indication* – Señal de llegada de los datos.
7. *Disconnect.request* – Petición para liberar una conexión.
8. *Disconnect.indication* – Indica la petición de desconexión.

Al realizar una llamada un usuario a otro y establecer una conversación la secuencia temporal de primitivas que se produciría sería:

1. *Connect.request* – Marcar el número de teléfono del usuario remoto.
2. *Connect.indication* – El teléfono del usuario remoto suena.
3. *Connect.response* – El usuario remoto descuelga.
4. *Connect.confirm* – El usuario iniciador escucha que el teléfono ya no llama.
5. *Data.request* – El usuario iniciador habla.
6. *Data.indication* – El usuario remoto le escucha.
7. *Data.request* – El usuario remoto habla.
8. *Data.indication* – El usuario iniciador le escucha.
9. *Disconnect.request* – El usuario iniciador cuelga el teléfono.
10. *Disconnect.indication* – El usuario remoto escucha y cuelga también.

Este procedimiento puede expresarse también de forma gráfica, visualizando que primitivas se intercambian en cada instante.

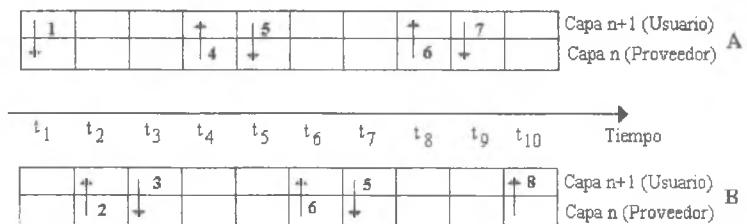


Figura 1.17 Gráfico del funcionamiento de primitivas de servicio.

1.6 MODELO DE REFERENCIA TCP/IP

Se ha visto que el modelo básico de referencia OSI es un estándar internacional para el diseño de redes de comunicaciones. A continuación se describirá uno de los modelos de arquitectura de red que se ha convertido en estándar de facto, el modelo **TCP/IP** de la **Internet** mundial.

El modelo de referencia TCP/IP recibe su nombre de las siglas de dos de los protocolos que lo definen: TCP (*Transmission Control Protocol*) Protocolo de Control de la Transmisión e IP (*Internet Protocol*) Protocolo de Interred. El origen de este modelo se encuentra en la red ARPANET, la red del Departamento de Defensa (DoD) de EEUU. ARPANET surge con el objetivo de crear un sistema de comunicación entre las computadoras de control de los sistemas de defensa nucleares de EEUU para que, en caso de un ataque nuclear, los daños sean los mínimos posibles. Ello se consigue diseñando un protocolo de comunicación que sea tolerante a fallos, buscando caminos alternativos para que la información llegue a su destino.

en caso de que ocurra un fallo en la comunicación entre dos nodos de la red.

El modelo TCP/IP está basado en capas siguiendo el modelo de funcionamiento de la arquitectura OSI. Una de las diferencias entre el modelo OSI y TCP/IP es el número de capas. El modelo TCP/IP sólo posee cuatro, algunas de las cuales engloban a varias del modelo OSI.

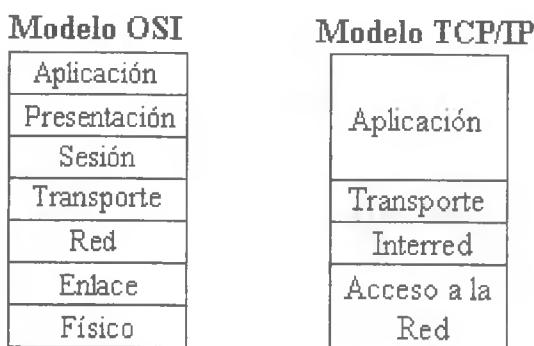


Figura 1.18 Comparación modelos OSI – TCP/IP.

1.6.1 Capa 1. Nivel de Acceso a la Red.

Este es el nivel inferior del modelo TCP/IP y además el más oscuro en cuanto a su normalización. Genéricamente, este nivel hace referencia a como se transmite un paquete del nivel de Interred (capa 2) al medio físico para llegar a su destino libre de errores. No existe una norma específica acerca de cómo debe estar diseñada esta capa, pero engloba a los niveles físico y de enlace del modelo OSI, por lo que se emplean las especificaciones de ese modelo.

1.6.2 Capa 2. Nivel de Interred (*internet*).

Este nivel es el eje de todo el modelo de arquitectura de red. Su objetivo es permitir que los nodos puedan enviar paquetes de información indicando su destino y éstos viajen de forma independiente, sin seguir un itinerario preestablecido. Se define un formato de paquetes y un protocolo asociado: IP. Cada máquina en una red con protocolo IP tiene una dirección única que la identifica dentro de la misma. Esta dirección consta de 4 dígitos separados por puntos que toman valores de 0 a 255 (por ejemplo, un valor válido de dirección sería 194.234.13.5). Una característica de este

protocolo es que permite identificar tanto máquinas dentro de una red como subredes, lo que lo convierte en un protocolo muy útil para interconexión de redes.

1.6.3 Capa 3. Nivel de Transporte.

Este nivel **permite la comunicación extremo a extremo entre origen y destino**, sin tener en cuenta consideraciones acerca de la estructura de la subred. En este nivel TCP/IP define dos protocolos distintos, cada uno con aplicaciones diferentes.

1. **TCP (Transmission Control Protocol)**. Se trata de un protocolo confiable y orientado a conexión, permitiendo la transmisión de una secuencia de bytes de origen a destino sin errores. Precisa por ello de algoritmos de control de errores y del flujo.
2. **UDP (User Datagram Protocol)** Protocolo de Datagramas de Usuario. Se trata de un protocolo no confiable y no orientado a conexión. Se emplea para aplicaciones en las que no es necesario el control de la secuencia ni el flujo, en definitiva en aquellas en que premie la prontitud frente a la fiabilidad (transmisión de voz y vídeo).

El nivel de transporte, tanto en el protocolo TCP como en el UDP, identifica las conexiones entre máquinas empleando un **número de puerto**. Este número de puerto es un número entre 0 y 65535 e identifica un **buffer** de memoria donde se almacena la información contenida en los paquetes del nivel de red (paquetes IP) dirigidos a ese puerto. Este **buffer** sirve de interfaz con el nivel de aplicación, de forma que cada número de puerto está asociado a una aplicación. La comunicación entre dos procesos del nivel de aplicación se establece con una conexión a nivel de transporte que se identifica con un número de **puerto de origen y puerto de destino** para cada uno de los extremos de la comunicación. Los datos contenidos en los paquetes procedentes del nivel de red se almacenan en los **buffers** asociados a cada puerto y cada aplicación lee la información del puerto que tiene asociado.

1.6.4 Capa 4. Nivel de Aplicación.

En este nivel **se definen los protocolos de aplicación** que tanto impacto tienen en la sociedad actual y han convertido a TCP/IP en un modelo de arquitectura aceptado internacionalmente. Entre los más populares podemos destacar:

- TELNET, protocolo de terminal virtual. Permite el acceso a una estación remota con un sistema operativo multiusuario y multitarea para el procesamiento de datos de forma remota.
- FTP (*File Transfer Protocol*), protocolo de transferencia de archivos. Permite el acceso a sistemas de archivos remotos para enviar y recibir ficheros.
- SNMP (*Simple Network Management Protocol*), protocolo simple de gestión de red. Permite la gestión remota de dispositivos de red para conocer su estado y solucionar problemas.
- SMTP (*Simple Mail Transfer Protocol*), protocolo simple de transferencia de correo. Protocolo que se emplea en las redes de comunicaciones con arquitectura TCP/IP para la transferencia de correo electrónico.
- NNTP (*News Network Transfer Protocol*), protocolo de transferencia de noticias en la red. Permite el acceso a variados grupos de discusión donde se intercambian mensajes.
- HTTP (*Hypertext Transfer Protocol*), protocolo de transferencia de hipertexto. Protocolo de aplicación por el que se conoce popularmente a Internet y que permite la transferencia de páginas Web.

Existen más protocolos de nivel de aplicación, así como otros que están en fase de desarrollo e introducción, y que debido a su extensión no se abordan. Se invita al lector a profundizar en la numerosas aplicaciones que emplean TCP/IP como modelo de arquitectura de red consultando la bibliografía.

1.6.5 El modelo OSI versus TCP/IP.

Puede realizarse un análisis comparativo exhaustivo entre ambos modelos de arquitectura de red, pero aquí se analizarán sólo algunas de las diferencias fundamentales.

El modelo OSI se apoya en tres conceptos fundamentales que son los **servicios**, la **interfaz** y los **protocolos**. Además estos conceptos se ajustan bastante bien a las actuales tendencias de programación orientada a objetos. TCP/IP no incluía originalmente estos conceptos, pero al mismo tiempo que se fue convirtiendo en un estándar fue adoptándolos acercándose cada vez más al modelo OSI.

Por un lado OSI se apoya en una comunicación orientada a conexión y sin conexión a nivel de red, siendo a nivel de transporte una comunicación siempre orientada a conexión. Por el contrario, TCP/IP se apoya en una

capa de red no orientada a conexión (permitiendo una mayor tolerancia a fallos en la llegada de los paquetes de información a su destino) y una capa de transporte donde se emplea una comunicación orientada tanto a conexión como a no conexión (TCP, UDP).

Estas diferencias hacen que cada uno de los modelos tengan un campo de aplicación determinado. El modelo OSI es el modelo de referencia en los sistemas de telefonía tanto móvil como fija, por lo que tiene en las compañías telefónicas sus grandes defensores. Por el contrario, el modelo TCP/IP se emplea en Internet, en las redes de computadores donde se está convirtiendo en la panacea de las comunicaciones por computador. En la actualidad, algunas compañías telefónicas están realizando grandes inversiones en proporcionar sus servicios tradicionales de telefonía empleando la arquitectura TCP/IP, por lo que quizás nos encontremos en un futuro no muy lejano con una integración de las telecomunicaciones en Internet.

1.7 INTERCONEXIÓN DE REDES

Uno de los motores del auge de las redes de computadoras en la actualidad ha sido la existencia de la **Internet** mundial. Internet crece día a día con la conexión de más redes a este entramado mundial, denominándose **subredes** a cada una de las diferentes redes que se interconectan entre sí formando una red mayor.

Esta interconexión de redes requiere de dispositivos dedicados que se encuentran íntimamente ligados a los conceptos de **arquitectura de red** y **niveles o capas de red**. Existen diversos dispositivos, cada uno con la función de interconectar una red con otra a diferente **nivel** dentro de la **arquitectura de red**.

- a) **Gateway o pasarela.** El término en inglés *gateway* se emplea frecuentemente para denotar a un dispositivo que se encuentra en una red y que la conecta con otra. Más formalmente, el término pasarela denota a un dispositivo que **interconecta redes con distinta arquitectura de red**, y que por tanto no presenta unos niveles comunes dentro de la arquitectura. Un ejemplo de ello sería la interconexión de una red con una arquitectura del modelo **OSI (Open Systems Interconnection)** con otra que no siga el modelo OSI.

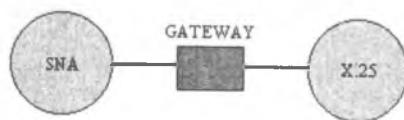


Figura 1.19 Interconexión de redes empleando pasarelas.

- b) **Router o encaminador.** Este dispositivo interconecta generalmente redes que presentan necesidades de encaminamiento de la información, por lo que trabaja en el **nivel de red** de la arquitectura. Es el caso de la interconexión de una red LAN con una WAN.

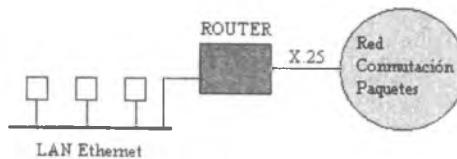


Figura 1.20 Interconexión de redes empleando encaminadores.

- c) **Bridge o puente.** Este dispositivo interconecta generalmente redes teniendo en cuenta las características que presentan a **nivel de enlace**. Es el caso de la interconexión de dos redes de tipo LAN, donde no existe encaminamiento de la información. Por ejemplo la conexión de una red Ethernet a una red Token Ring.

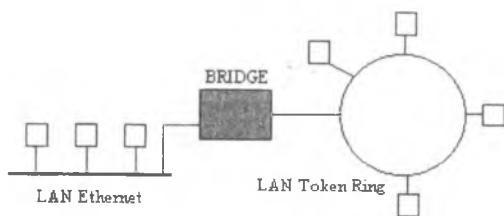


Figura 1.21 Interconexión de redes empleando puentes.

- d) **Repetidor.** Este dispositivo interconecta redes que son semejantes, pues únicamente se encarga de amplificar las señales eléctricas provenientes de una red y que deben llegar a otra. Un ejemplo sería la construcción de una gran LAN Ethernet, que,

debido a las limitaciones en distancia, precisa de varios segmentos conectados entre sí por repetidores.

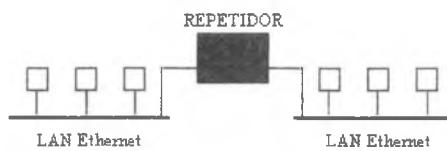


Figura 1.22 Interconexión de redes empleando repetidores.

CAPÍTULO 2. ESPECIFICACIÓN Y VALIDACIÓN DE PROTOCOLOS

2.1 ESPECIFICACIÓN FORMAL E INFORMAL

Por **Especificación de un Protocolo** se entiende la descripción del *conjunto de reglas de utilización de las primitivas de servicio suministradas por el nivel inferior para la comunicación a nivel horizontal.* En situaciones reales el funcionamiento de los protocolos de comunicación entre dos entidades pares de un mismo nivel dentro del modelo de capas puede ser muy complejo. Por ello, en la mayor parte de las ocasiones es necesario recurrir a una herramienta que permita describir su funcionamiento de forma clara y concisa, lo que redundará en una detección de errores en el funcionamiento más sencilla y eficaz.

Es posible establecer dos líneas fundamentales en la especificación de protocolos, la informal y la formal. La especificación informal es la más sencilla y libre de ellas. Consiste en la especificación del funcionamiento del protocolo empleando un lenguaje pseudocódigo cercano al lenguaje natural que explique cómo se realiza la comunicación entre dos entidades pares. Este tipo de especificación resulta útil para protocolos simples, pero presenta grandes desventajas a la hora de validar un protocolo (proceso por el cual se comprueba que el protocolo se adecua a las especificaciones descritas) y de modelar un protocolo complejo.

Puede citarse como ejemplo de especificación informal de protocolos la descripción del protocolo unidireccional de parada y espera con ausencia de errores. El funcionamiento de este protocolo consiste en el envío de paquetes de datos desde un emisor a un receptor, enviando éste último un acuse de recibo para cada paquete recibido.

Entidad que envía el emisor: DATOS.

Entidad que recibe el emisor: ACK.

Entidad que envía el receptor: ACK.

Entidad que recibe el receptor: DATOS.

Algoritmo de Funcionamiento

EMISOR	RECEPTOR
Inicio: Envía DATOS	Inicio: Si recibe DATOS entonces envía ACK
Bucle: Si recibe ACK entonces envía DATOS <i>Ir Bucle</i>	<i>Ir Inicio</i>

En contraste con este tipo de especificación está la especificación formal. La especificación formal emplea una **herramienta matemática** para describir el funcionamiento de un protocolo. Dentro de las herramientas de especificación de protocolos destacan las **Máquinas de Estado Finito** y las **Redes de Petri**, cada una de las cuales se analizarán por separado.

2.2 MÁQUINAS DE ESTADO FINITO (MEF)

Con esta técnica, **cada maquina de protocolo** (transmisor y/o receptor) **siempre está en un estado específico** en cualquier instante. **Su estado consistirá en una combinación de los valores de sus variables.**

Desde el punto de vista formal, una MEF se define como un conjunto de cinco elementos (**X, E, S, FT, FS**), donde:

X , Conjunto finito de estados.

E , Conjunto finito de entradas.

S , Conjunto finito de salidas.

FT , Función de transición φ . $X(t+1) = \varphi(X(t), E)$, donde t es el tiempo y $X(t+1)$ será un elemento de X .

FS , Función de salida ψ . $S(t) = \psi(X(t), E)$, donde t es el tiempo y $s(t)$ será un subconjunto de S .

El **estado** se define como el conjunto de información que describe de forma completa la situación de la MEF en cada instante. En cada instante de tiempo, cada una de las máquinas de protocolo tendrá un estado, que estará determinado por los valores de las variables y el estado del canal. Por tanto, el **número de estados** dependerá del **número de variables**. De hecho si el número de variables es n y pueden tomar valores 0 o 1, el número total posible de estados es 2^n , aunque muchos de ellos no se producirán nunca.

La **transición** se define como el proceso por el cual la MEF cambia de un estado a otro distinto (en ocasiones será al mismo estado). La transición se

produce cuando ocurre un **evento** de entrada a la MEF, produciendo un evento de salida. Los estados se representan por circunferencias donde se indican los valores de variables asociadas, y las transiciones por flechas que unen estados indicando entradas y salidas.

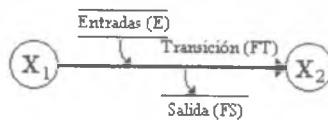


Figura 2.1 Esquema de una transición en una MEF.

En cada estado existen cero o más transiciones a otros estados. El evento de entrada que produce una transición en una máquina de protocolo es muy variado: puede ser la llegada de una trama, la finalización de un temporizador, etc. En toda MEF existe un estado particular que se denomina **estado inicial**. Este estado describe el sistema justo cuando comienza a funcionar. Desde este estado inicial pueden alcanzarse algunos o quizás todos los estados de la MEF mediante las transiciones.

Empleando la **teoría de grafos** es posible estudiar los estados que son alcanzables o no, en definitiva realizar un **análisis de accesibilidad**. De esta forma es posible determinar si un protocolo es correcto o no, y permite detectar errores como:

- Incompletitud**, que se produce cuando existen estados en los que pueden producirse eventos para los que no hay definidas transiciones.
- Bloqueo**, que se produce cuando existen estados para los que no hay salida y no se puede avanzar.
- Transiciones ajenas**, que son transiciones asociadas a eventos que no pueden producirse en el estado origen de la transición.

Véase el siguiente ejemplo de especificación de protocolo empleando una MEF. Se pretende modelar un protocolo de comunicación unidireccional con asentimiento de las tramas enviadas (cada vez que el receptor recibe una trama de datos envía una trama especial al emisor indicando que la ha recibido correctamente), empleando una numeración de las mismas basada en un solo bit. Además el protocolo contempla la pérdida de tramas de datos y de asentimientos en el canal, así como el reenvío de las tramas perdidas en las situaciones anteriores.

Los estados están determinados por tres variables {X,Y,Z}

$X :$

0 → Emisor espera un asentimiento 0
1 → Emisor espera un asentimiento 1

$Y :$

0 → Receptor espera una trama 0
1 → Receptor espera una trama 1

$Z :$

0 → Canal con trama 0
1 → Canal con trama 1
- → Canal sin trama
A0 → Canal con asentimiento 0
A1 → Canal con asentimiento 1

Las entradas o eventos que pueden provocar una transición serán,

- $ACK0_IN \rightarrow$ Receptor recibe asentimiento de trama 0.
- $ACK1_IN \rightarrow$ Receptor recibe asentimiento de trama 1.
- $DATA0_IN \rightarrow$ Receptor recibe trama de datos 0.
- $DATA1_IN \rightarrow$ Receptor recibe trama de datos 1.
- $TEMP \rightarrow$ Vencimiento de temporizador de espera de respuesta en emisor.
- $LOST \rightarrow$ Pérdida de información en el canal.

Las salidas o eventos que se generan en la transición serán,

- $ACK0_OUT \rightarrow$ Receptor envía asentimiento de trama 0.
- $ACK1_OUT \rightarrow$ Receptor envía asentimiento de trama 1.
- $DATA0_OUT \rightarrow$ Emisor envía trama de datos 0.
- $DATA1_OUT \rightarrow$ Emisor envía trama de datos 1.

El grafo de la MEF describe el funcionamiento del protocolo, donde cada transición describe cómo se realiza la comunicación entre emisor y receptor.

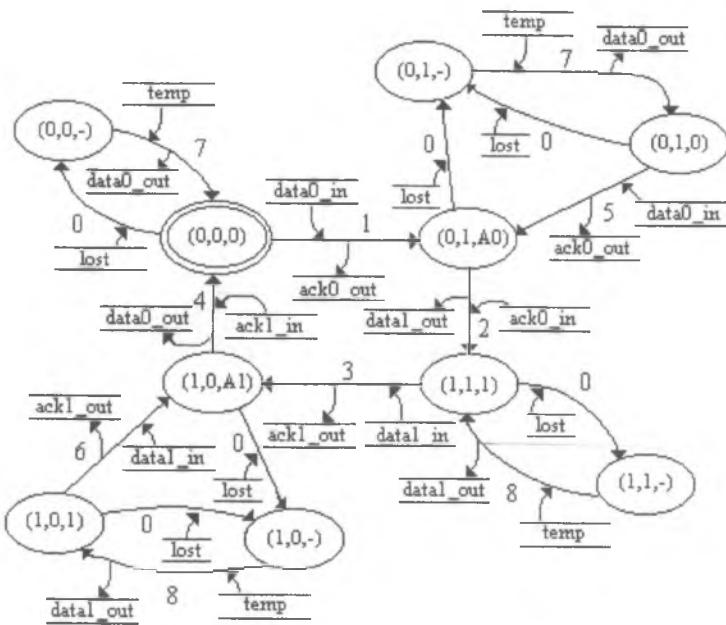


Figura 2.2 Grafo de la MEF correspondiente a un protocolo unidireccional.

Las transiciones en el grafo se corresponden a.

- 0 → Se pierden datos del canal.
- 1 → El receptor recibe una trama de datos 0 y envía el asentimiento correspondiente, cambiando al estado de espera de una trama 1.
- 2 → El emisor recibe el asentimiento de trama 0 y envía una trama 1 al canal, cambiando al estado de espera de un asentimiento de trama 1.
- 3 → El receptor recibe una trama de datos 1 y envía un asentimiento de trama 1, cambiando al estado de espera de trama 0.
- 4 → El emisor recibe el asentimiento de trama 1 y envía una trama 0 al canal, cambiando al estado de espera de un asentimiento de trama 0.
- 5 → El receptor envía un asentimiento de trama 0 después de haber recibido una trama de datos 0 que ha sido retransmitida.

- 6 → El receptor envía un asentimiento de trama 1 después de haber recibido una trama de datos 1 que ha sido retransmitida.
- 7 → En el emisor vence un temporizador al pasar un cierto tiempo sin recibir un asentimiento de trama 0 y se reenvía la trama de datos 0.
- 8 → En el emisor vence un temporizador al pasar un cierto tiempo sin recibir un asentimiento de trama 1 y se reenvía la trama de datos 1.

En el siguiente ejemplo se describe la MEF de un protocolo de comunicación bidireccional con asentimiento de las tramas enviadas, empleando numeración de un solo bit. El protocolo contempla la pérdida de tramas de datos y de asentimientos en el canal, así como el reenvío de las tramas perdidas en las situaciones anteriores. Cada vez que el usuario envíe una solicitud de envío de bloque de datos (RBU), el emisor envía una trama de datos, siguiendo la numeración de un bit. Por otra parte, cada vez que el receptor recibe una trama de datos, la envía al usuario en forma de bloque de datos de usuario (EBU).

Al tratarse de una comunicación bidireccional, cada estación en cada extremo de la misma se considerará que tendrá un emisor y un receptor, por lo que se desarrollarán dos MEF que estarán presentes en cada máquina de protocolo.

Deben describirse dos conjuntos de estados, uno para la MEF emisora de tramas y otra para la receptora,

Estados de la MEF emisora son de la forma {X,Y}

$X :$ EU → Emisor espera bloque de datos del usuario
 EA → Emisor espera trama de asentimiento

$Y :$ 0 → Estado de numeración 0
 1 → Estado de numeración 1

Estados de la MEF receptora son de la forma {Z}

$Z :$ ET0 → Receptor espera trama de datos 0
 ET1 → Receptor espera trama de datos 1

Entradas,

RBU → Se recibe un bloque de datos del usuario.

ACK0_IN → Se recibe un asentimiento de trama 0.

ACK1_IN → Se recibe un asentimiento de trama 1.

DATA0_IN → Se recibe una trama de datos 0.

DATA1_IN → Se recibe una trama de datos 1.

TEMP → Expira el temporizador en el emisor.

Salidas,

EBU → Se envía un bloque de datos al usuario.

ACK0_OUT → Se envía una trama de asentimiento de trama 0.

ACK1_OUT → Se envía una trama de asentimiento de trama 1.

DATA0_OUT → Se envía una trama de datos 0.

DATA1_OUT → Se envía una trama de datos 1.

CT → Conectar temporizador en emisor.

DT → Desconectar temporizador en emisor.

Los grafos de las MEF siguientes describen el funcionamiento del protocolo,

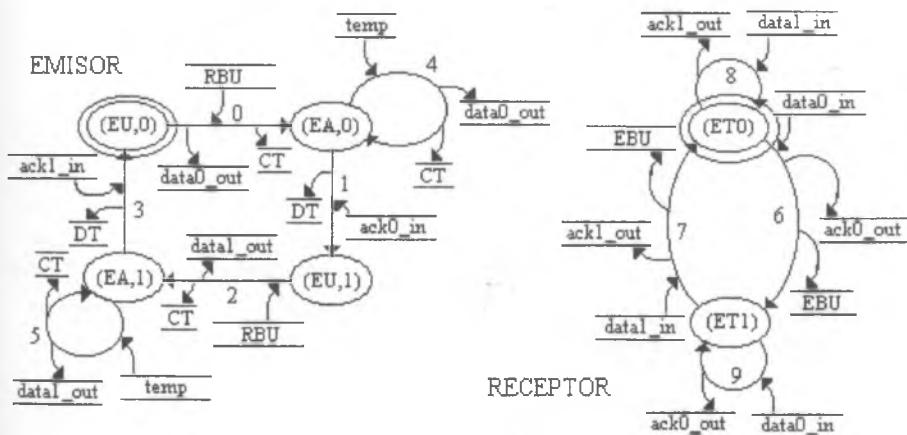


Figura 2.3 Grafo de la MEF correspondiente a un protocolo bidireccional.

Las transiciones en el grafo se corresponden a,

0 → Emisor envía trama de datos 0 al recibir un bloque de usuario y activa temporizador.

- 1 → Emisor recibe trama de asentimiento 0 y desconecta el temporizador.
- 2 → Emisor recibe bloque de datos del usuario y envía trama de datos 1, activando el temporizador.
- 3 → Emisor recibe trama de asentimiento 1 y desconecta el temporizador.
- 4 → Al expirar el temporizador, el emisor reenvía la trama de datos 0 y vuelve a conectar el temporizador.
- 5 → Al expirar el temporizador, el emisor reenvía la trama de datos 1 y vuelve a conectar el temporizador.
- 6 → El receptor recibe una trama de datos 0, la envía al usuario en un bloque de datos y envía el asentimiento de trama 0.
- 7 → El receptor recibe una trama de datos 1, la envía al usuario en un bloque de datos y envía el asentimiento de trama 1.
- 8 → Si se pierde una trama de asentimiento 1, el receptor recibirá la trama de datos 1 repetida, por lo que volverá a enviar el asentimiento e ignorará la trama repetida.
- 9 → Si se pierde una trama de asentimiento 0, el receptor recibirá la trama de datos 0 repetida, por lo que volverá a enviar el asentimiento e ignorará la trama repetida.

2.3 REDES DE PETRI

La técnica de las Redes de Petri, introducida por Danthine en 1980, es otra de las herramientas formales para la especificación de protocolos. Una red de Petri es un gráfico conformado por 4 elementos básicos: lugares, transiciones, arcos y marcas. Un **lugar** representa un estado en el que puede estar el sistema o parte de él. Se representan por círculos. La **transición** se representa mediante una barra horizontal o vertical, la cual posee cero o más **arcos de entrada** y cero o más **arcos de salida**. Las transiciones se establecen entre lugares y representan los posibles cambios en el estado del sistema.

Por último, las **marcas**, representadas por puntos dentro de los lugares, pueden haber una o más, indican el flujo de información o ejecución de la red.

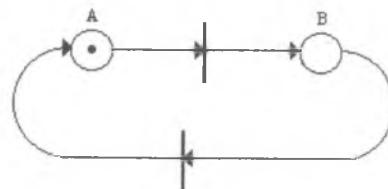


Figura 2.4 Ejemplo de red de Petri.

Una transición se dice que se **habilita** cuando en todos los lugares de entrada (asociados a arcos de entradas de la transición) existe, al menos, una marca. El **disparo** de una transición consume una marca de cada lugar de entrada a la transición y produce una marca en cada lugar de salida. En el caso de que se habiliten simultáneamente dos o más transiciones, la decisión de disparar una u otra transición es indeterminista, es decir, puede realizarse la ejecución tanto de una como de otra, pero no ambas a la vez.

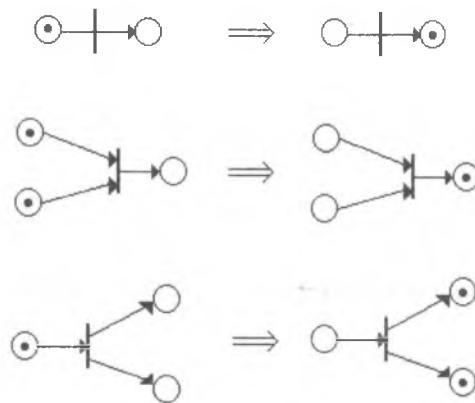


Figura 2.5 Ejemplos de disparo de transiciones.

Inicialmente se especifican unas marcas en ciertos lugares de la red, de forma que a partir de estas **marcas iniciales** el funcionamiento del protocolo es correcto. La evolución del sistema, de la red, se realiza partiendo de estas marcas iniciales, disparando de forma indeterminista las transiciones habilitadas.

Dentro del grafo de la red de Petri existirán unos lugares sin arcos de entrada asociados a las **entradas al protocolo**, y unas transiciones con arcos de salida que no llevan a ningún lugar asociadas a las **salidas del protocolo**. Por otra parte, a diferencia de las MEF, en las redes de Petri no es necesario especificar estados compuestos, por lo que están bien diferenciados los lugares asociados al emisor, al receptor y al canal. Una de las ventajas de esta técnica es que permite detectar fácilmente errores en el protocolo, como es la presencia de **bloqueos**.

Como ejemplo de la aplicación de las redes de Petri a la especificación formal de un protocolo, se muestra a continuación el esquema correspondiente a la red de Petri del protocolo unidireccional visto en el apartado 2.2.

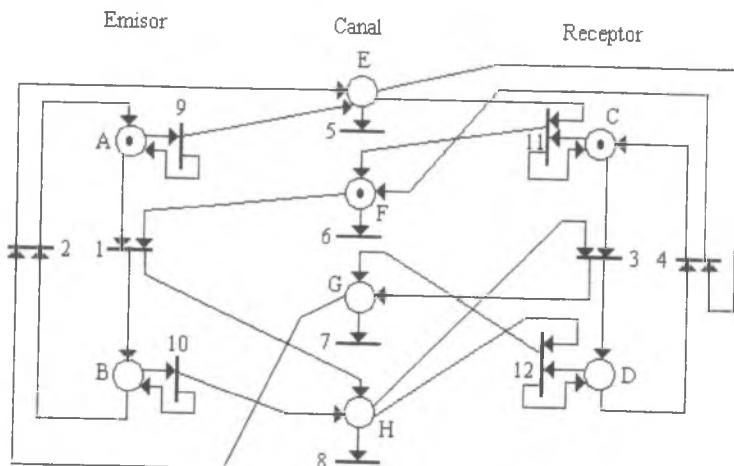


Figura 2.6 Grafo de la red de Petri correspondiente a un protocolo unidireccional.

Los lugares asociados al protocolo son,

- A → Emisor espera asentimiento de datos 0.
- B → Emisor espera asentimiento de datos 1.
- C → Receptor espera trama de datos 1.
- D → Receptor espera trama de datos 0.
- E → Canal con trama de datos 0.
- F → Canal con asentimiento de datos 0.
- G → Canal con asentimiento de datos 1.
- H → Canal con trama de datos 1.

Las transiciones que aparecen en la red se corresponden a,

- 1 → El emisor recibe asentimiento de datos 0 y envía una trama de datos 1.
- 2 → El emisor recibe asentimiento de datos 1 y envía una trama de datos 0.
- 3 → El receptor recibe trama de datos 1 y envía el asentimiento de datos 1.
- 4 → El receptor recibe trama de datos 0 y envía el asentimiento de datos 0.
- 5 → Pérdida en el canal de una trama de datos 0.
- 6 → Pérdida en el canal de una trama de asentimiento de datos 0.
- 7 → Pérdida en el canal de una trama de asentimiento de datos 1
- 8 → Pérdida en el canal de una trama de datos 1.
- 9 → Expiración del temporizador en el emisor y reenvío de la trama de datos 0.
- 10 → Expiración del temporizador en el emisor y reenvío de la trama de datos 1.
- 11 → Recepción de trama de datos 0 fuera de secuencia (debido a un reenvío), por lo que se envía de nuevo el asentimiento de datos 0 y se ignoran los datos recibidos.
- 12 → Recepción de trama de datos 1 fuera de secuencia (debido a un reenvío), por lo que se envía de nuevo el asentimiento de datos 1 y se ignoran los datos recibidos.

2.4 OTROS MÉTODOS FORMALES DE ESPECIFICACIÓN DE PROTOCOLOS

Además de las MEF y las redes de Petri existen otras técnicas de especificación formal de protocolos. Son menos empleadas y por ello se analizarán de forma muy breve.

La primera de ellas son los **Lenguajes de transición de estado extendido**. Esta técnica es una extensión de las MEF para evitar el elevado número de estados que aparecen al analizar protocolos de comunicación complejos. Para ello se introducen variables dentro de los estados, de forma que se realiza una reducción en el número de los mismos al poder agruparlos. Cada proceso (emisor o receptor) se modela empleando lo que se denomina un **módulo**. Un módulo consta de una serie de estados y reglas de transición. En cada regla se indican los eventos que la activan y las acciones que se ejecutan, así como el estado al que evoluciona el sistema al realizar la transición. Para el intercambio de tramas entre los módulos se define una entidad denominada **canal**.

Otra de las técnicas consiste en emplear los **lenguajes estructurados estándar** para la especificación formal de protocolos. Permiten representar fácilmente operaciones sobre mensajes y variables de control, pero presentan problemas con el flujo de control y el estado del protocolo. Algunos de los lenguajes a emplear son los tradicionales como PASCAL, C, C++ y existen algunas aplicaciones empleando sistemas de tiempo real orientados al procesamiento de eventos.

2.5 VERIFICACIÓN Y VALIDACIÓN DE PROTOCOLOS

La **validación** de un protocolo consiste en comprobar que un protocolo cumple con ciertas propiedades. Algunas de las más significativas son,

- a) Ausencia de **bloqueos** o *deadlocks*.
- b) **Viveza** o *liveness* del protocolo, que determina qué partes del protocolo son útiles, eliminando los estados inalcanzables.
- c) Ausencia de **lazos improductivos**. Se produce cuando desde un estado vuelve a alcanzarse a él mismo empleando una serie de transiciones sin aportar nada al funcionamiento del protocolo.
- d) Capacidad de **recuperación** o *selfsynchronization*. Define la capacidad del protocolo de volver desde los estados de funcionamiento anómalo del protocolo (situaciones de error) a los normales.

Por **verificación** de un protocolo se entiende la comprobación de que el protocolo desempeña todas las funciones que se especificaron en el servicio que debe proveer la capa.

CAPÍTULO 3. TRANSMISIÓN DE SEÑALES

3.1 INTRODUCCIÓN

Con este capítulo se inicia una serie en la que se abordan los diferentes niveles de las arquitecturas de red vistas en el capítulo 1. En la línea de la estandarización de redes indicada por el modelo **OSI** de la ISO y el estándar de facto actual de las redes de computadores **TCP/IP**, se analizan diferentes niveles de ambas arquitecturas. En los niveles inferiores, donde TCP/IP no establece normas estrictas, se analizarán el nivel físico (capítulos 3, 4 y 5) y nivel de enlace (capítulos 6, 7 y 8) definidos en el modelo OSI. A continuación se estudia el nivel de red (capítulos 9, 10 y 11) haciendo especial hincapié en las redes de computadores y en Internet.

La definición de **función del nivel físico** en el modelo OSI se establece en *la transmisión de una sucesión de bits a través de un canal de comunicación o medio físico*. En base a este objetivo principal se determina que el nivel físico tiene las funciones siguientes:

- a) **Señalización y Modulación.** Especifica cómo se realiza la modulación de la información a transmitir en señales adecuadas al medio físico.
- b) **Definición de componentes de interconexión con el medio físico.** Define las características de los conectores a nivel mecánico (número de *pins*, tipos de cables, longitudes máximas, etc.) y nivel eléctrico (características de las señales eléctricas a transmitir: niveles de voltaje empleados, velocidades de transmisión, etc.).
- c) **Sincronización.** Define el modo en que emisor y receptor se sincronizan para realizar el muestreo e interpretación correcta de la señal.
- d) **Monitorización de la calidad de señalización de bits.** Define los rangos de niveles de señal que permiten el reconocimiento de los bits de información.

En toda comunicación de datos entre dos estaciones existen unos elementos básicos que son:

- a) **DTE**, *Data Terminal Equipment* o Equipo Terminal de Datos (ETD). Es el dispositivo que se encuentra en cada extremo de la comunicación y desea realizar la transmisión de datos (computador, host, estación de trabajo, ...).
- b) **DCE**, *Data Circuit-terminating Equipment* o Equipo de Terminación de Circuito de Datos (DCE). Es el dispositivo que interacciona con el medio físico y convierte la información proveniente del DTE en señales adecuadas para su transmisión (módem, adaptadores de red, etc.).
- c) **Medio de Transmisión**. Es el medio físico empleado para la transmisión de señales y así establecer la comunicación entre dos puntos distantes. Es posible hacer una clasificación en dos tipos:
 - 1) **Confinados**: Son los medios en los que las señales se transmiten limitadas espacialmente (cables eléctricos, fibra óptica).
 - 2) **No Confinados**: Son los medios en los que las señales se transmiten por el espacio libremente (transmisión de ondas electromagnéticas).
- d) **Regeneradores de Señal**. Estos dispositivos actúan como amplificadores y se emplean cuando las señales deben de transmitirse a distancias elevadas.

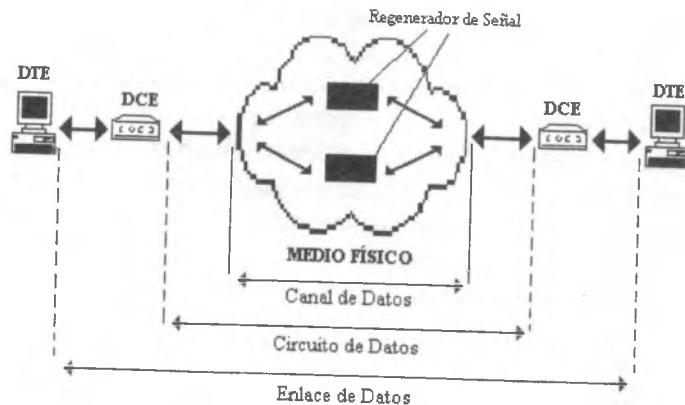


Figura 3.1 Diagrama del esquema de una comunicación entre DTE.

En el esquema de comunicación se definen también diferentes jerarquías en función de los elementos que intercambian información:

- a) **Canal de Datos.** Define una transmisión unidireccional de datos a través del medio físico. Comprende por tanto el medio físico y los regeneradores de señal.
- b) **Círculo de Datos.** Define la comunicación entre dos DCE. Esta comunicación puede comprender más de un canal de datos.
- c) **Enlace de Datos.** Define la comunicación entre dos DTE. En este nivel se consideran funciones adicionales de control de errores que implementan los DTE.

3.2 ANÁLISIS DE SEÑALES CON SERIES DE FOURIER. ANCHO DE BANDA

La transmisión de datos entre dos computadores redonda en última instancia en la propagación de señales a través de un medio físico. Se entiende por señal la variación en el tiempo de una magnitud física. Atendiendo a la naturaleza de la magnitud física, las señales pueden clasificarse en,

- a) **Eléctricas**, cuando la magnitud física es el voltaje o el amperaje.
- b) **Óptica**, cuando la magnitud física es la intensidad luminosa.
- c) **Electromagnéticas**, cuando la magnitud física es la amplitud, frecuencia o fase de ondas electromagnéticas (aunque la luz está incluida dentro de esta clasificación, suelen tratarse por separado dejando este grupo para las ondas de radio).

A su vez las señales pueden clasificarse en dos tipos atendiendo a su naturaleza.

- a) **Señales analógicas**, cuando la magnitud física varía de forma continua a lo largo del tiempo.

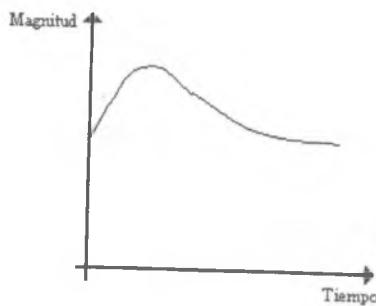


Figura 3.2 Señal analógica.

- b) **Señales digitales**, cuando la magnitud física adquiere determinados valores en determinados instantes de tiempo, no estando definida en otros instantes.

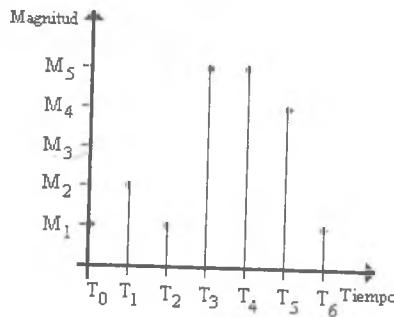


Figura 3.3 Señal digital.

En la transmisión de datos es preciso estudiar como se produce la transmisión de señales por un medio físico para así determinar si la comunicación se realizará de forma correcta. Para ello es necesario una herramienta matemática que permita modelizar las señales y estudiar su propagación. Esta herramienta son las **séries de Fourier**.

Las series de Fourier son un **desarrollo matemático** que permite expresar cualquier función periódica $f(t)$ como un sumatorio infinito de funciones seno y coseno. Su demostración se debe al matemático francés Jean-Baptiste Fourier, quien a principios del siglo XIX, estudiando la solución

a la ecuación en derivadas parciales que rige la transmisión del calor en un medio físico, llegó a la siguiente expresión.

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(2n\pi f_0 t) + \sum_{n=1}^{\infty} b_n \sin(2n\pi f_0 t)$$

$$a_n = \frac{2}{T} \int_0^T f(t) \cos(2n\pi f_0 t) dt \quad n = 0, \dots, \infty$$

$$b_n = \frac{2}{T} \int_0^T f(t) \sin(2n\pi f_0 t) dt \quad n = 1, \dots, \infty$$

Siendo $f_0 = \frac{1}{T}$ la frecuencia de la señal periódica $f(t)$. Si se analiza en detalle este desarrollo se llega a la conclusión de que toda función $f(t)$ periódica puede construirse como una suma de infinitas funciones seno y coseno de frecuencias múltiplo de la frecuencia de la señal $f(t)$. Cada componente a_n , b_n de la expresión representa la amplitud de la señal senoidal o cosenoidal de frecuencia nf_0 . A cada una de estas componentes se la denomina **armónico** de la señal $f(t)$.

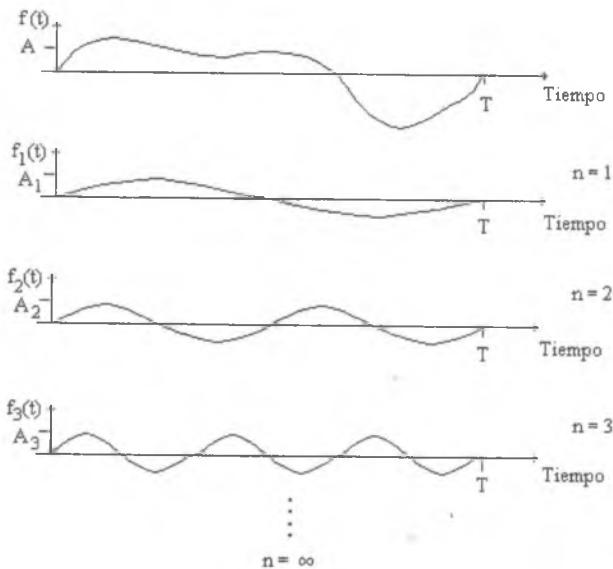


Figura 3.4 Componentes frecuenciales de una señal periódica.

En la transmisión de datos el tipo de señales que aparecen son genéricamente señales analógicas que codifican información binaria y se propagan a través de medios físicos analógicos. La señal analógica

genérica que codifica información binaria es una señal cuadrada (realmente no es cuadrada sino trapezoidal, es decir los flancos de subida y bajada de los pulsos presentan un ángulo muy próximo a 90 grados) donde el 0 lógico corresponde a un voltaje y el 1 lógico a otro distinto.



Figura 3.5 Señal analógica cuadrada que codifica información binaria.

Sin embargo, para poder aplicar los desarrollos de Fourier, además de precisar una función continua o analógica es necesario que la señal sea periódica. En la transmisión de datos la periodicidad de las señales es una situación que apenas se produce, pues la información que se transmite entre dos DTE nunca es la misma y no tiene periodicidad ninguna. Por ello es preciso realizar una aproximación, y es que para estudiar el fenómeno de la transmisión de señales con información binaria, se considerará la transmisión de un octeto de bits (correspondientes a un carácter ASCII) de forma continua, por lo que el periodo de la señal se considerará como el tiempo en transmitir esos ocho bits u ocho pulsos de la señal cuadrada. Basándose en esta suposición se realizará el estudio de la transmisión de este tipo de señales, que si bien no proporcionará información exacta de la situación real, si proporciona unos resultados bastante aproximados.

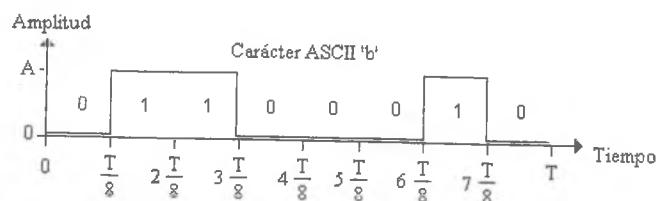


Figura 3.6 Señal periódica asociada a la transmisión del carácter ASCII 'b'.

Si se representan los primeros 10 términos de la Serie de Fourier, o lo que es lo mismo los 10 primeros armónicos, de la señal correspondiente al carácter ASCII 'b', puede apreciarse como la señal reconstruida no es exactamente igual a la original, pero sí que es posible identificar la información binaria de unos y ceros lógicos. Si se calcularan los infinitos armónicos de la serie (cosa que no es posible computacionalmente) se obtendría exactamente la señal original.

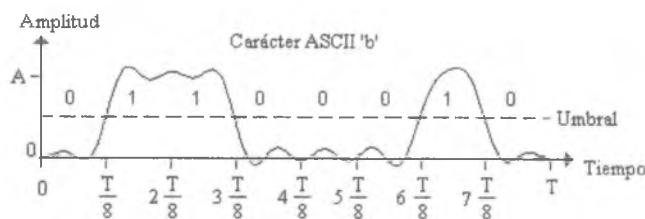


Figura 3.7 Función compuesta por la suma de los 10 primeros armónicos de la señal de la figura 3.6.

Una expresión muy útil para el estudio de la transmisión de señales por un medio es el **Espectro de Potencia**. El espectro de potencia mide la contribución de cada armónico a la reconstrucción de una señal periódica. Esta contribución se visualiza representando la expresión $F_n = \sqrt{a_n^2 + b_n^2}$ frente al orden del armónico n . Para la señal considerada anteriormente es posible obtener el espectro de potencia para los primeros 10 armónicos.

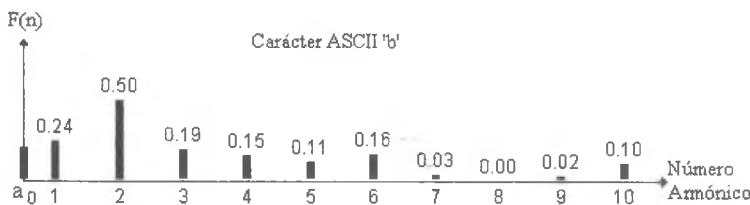


Figura 3.8 Espectro de potencia de la señal asociada al carácter ASCII 'b'.

Como puede apreciarse, los primeros armónicos son los que tienen mayor importancia en la reconstrucción de la señal, pues están asociados a señales senoidales y cosenoidales de mayor amplitud. Conforme aumenta el número del armónico la contribución en amplitud es menor, por lo que su efecto en la reconstrucción de la señal es menor.

En la transmisión de las señales por un medio físico real se producen fenómenos de atenuación. Estas atenuaciones se producen por variaciones en la amplitud de la señal que dependen directamente de la frecuencia de la señal transmitida. Como anteriormente se ha demostrado que toda señal periódica puede considerarse como la composición de infinitas señales armónicas cada una con una determinada frecuencia, la señal periódica se distorsionará debido a la distorsión de cada uno de sus armónicos. Estas atenuaciones son debidas a una propiedad que presentan todos los medios de transmisión y que se denomina **ancho de banda**. El ancho de banda de un medio físico determina cuál es el rango de frecuencias que un medio transmite produciendo un decremento en amplitud que no afecta al reconocimiento de la señal. El valor de la atenuación depende de la frecuencia y se define como la relación entre la amplitud de la señal introducida en el medio físico y la amplitud de la señal en el extremo del medio. Además la atenuación está relacionada con la ganancia en forma inversa.

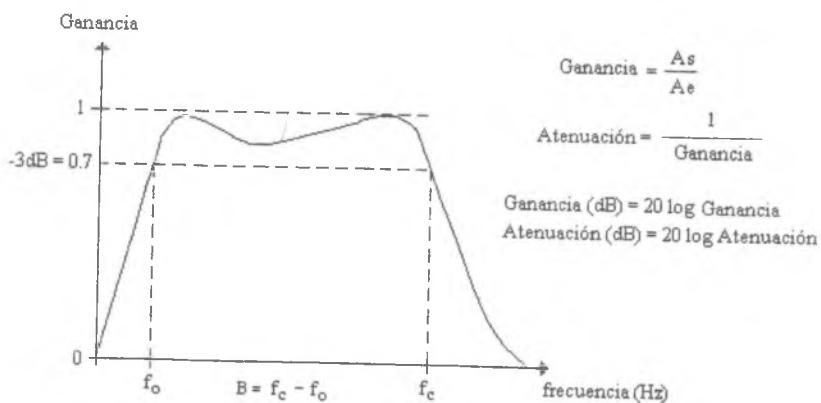


Figura 3.9 Atenuación de un medio frente a la frecuencia.

Existe un rango de frecuencias, desde f_0 hasta f_c , para el cual se considera que la atenuación producida en la señal (atenuación < 3dB) no es suficiente como para no recuperarla de forma correcta al atravesar el medio. A este rango de frecuencias se le denomina **Ancho de Banda del Medio (B o W)**, luego $B = f_c - f_0$ **Hertzios**. Dependiendo del valor del ancho de banda el número de armónicos asociados a la señal que puedan

atravesarlo sin problemas será mayor o menor. En definitiva, el ancho de banda afectará a la reconstrucción de la señal transmitida y, como se demostrará más adelante, afectará a la velocidad máxima de transmisión de información binaria a través del medio.

3.3 VELOCIDAD DE TRANSMISIÓN. TEOREMA DE NYQUIST

Se define la **velocidad de modulación** de una señal digital de pulsos como el número de veces por unidad de tiempo que la magnitud física de la señal puede cambiar su valor. Si la unidad de tiempo es el **segundo**, la unidad de velocidad de modulación se denomina **baudio**.

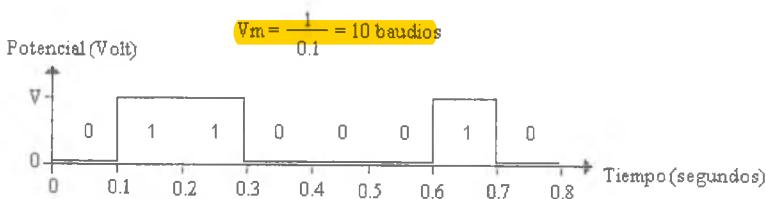


Figura 3.10 Señal binaria transmitida por un medio físico.

En la señal que aparece en la figura 3.10 es posible determinar la velocidad de modulación asociada. Para ello, es necesario determinar en qué tiempo mínimo la señal puede variar su valor, hallando así la velocidad de modulación empleada. Hay que notar que la velocidad de modulación hace referencia al número de veces que la señal *puede* cambiar de valor, pudiendo hacerlo o no. En base a esto, la señal de la figura *puede* variar su valor como mínimo una vez cada 0.1 segundos, existiendo situaciones en que la señal varía su valor cada 0.1 segundos y otras en que el tiempo es mayor (el que la señal cambie o no su valor dependerá de la información binaria a modular). Por tanto la velocidad de modulación de la señal vendrá dada por: $Vm = \frac{1}{0.1s} = 10 \text{ baudios bd.}$

Por otra parte se define la **velocidad de transmisión** como el número de bits transmitidos por un medio por unidad de tiempo. La unidad de velocidad de transmisión comúnmente empleada es el **bit por segundo (bps)** y sus múltiplos asociados: **Kbps** (1,000 bps), **Mbps** (1,000,000 bps) y el **Gbps** (1,000 Mbps). La velocidad de transmisión está relacionada con la velocidad de modulación de forma que, si se conoce el número de bits **b** o el número de niveles **n** que una señal codifica en una de sus variaciones, la velocidad de transmisión viene dada por:

$$V_t = V_m \cdot \log_2 n = V_m \cdot b \text{ bps}$$

Como ejemplo véase el cálculo de la velocidad de transmisión de la señal de la siguiente figura, donde se codifican 4 niveles en la señal.

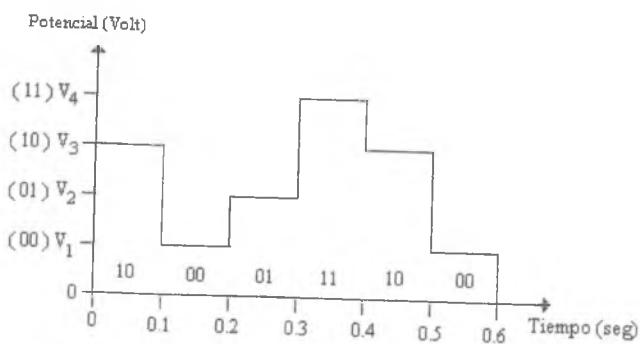


Figura 3.11 Señal de información binaria codificada con 4 niveles de tensión.

La velocidad de modulación para esta señal es de 10 baudios, pues es el número máximo de cambios que puede tener la señal por segundo, y dado que existen cuatro niveles para codificarla información binaria, la velocidad de transmisión viene dada por:

$$V_t = V_m \cdot \log_2 4 = \frac{1}{0.1s} \cdot 2b = 20 \text{ bps}$$

Se estudiará a continuación que relación existe entre la velocidad de transmisión en un medio y el número de armónicos que éste permite transmitir sin atenuaciones considerables. Considérese una señal que transmite de forma periódica un byte de datos (carácter ASCII), siendo T el tiempo de transmisión de esos 8 bits que se repiten. Teniendo en cuenta la definición de velocidad de transmisión, ésta puede expresarse como:

$$V_t = \frac{\text{nº de bits transmitidos}}{\text{tiempo empleado}} = \frac{8}{T}$$

Por otra parte, la inversa del periodo T se denomina f_0 y se corresponde con la frecuencia fundamental de la señal o 1^{er} armónico. Despejando en la definición de Vt se tiene que

$$f_0 = \frac{1}{T} = \frac{V_t}{8} \quad (3.1)$$

Si se dispone de un medio físico con ancho de banda B (Hz), el número de armónicos de la señal que permitirá pasar el medio sin prácticamente atenuación vendrá dado por la relación

$$n \cdot f_0 \leq B \quad (3.2)$$

siendo n el orden o número del armónico. Combinando las ecuaciones (3.1) y (3.2) se llega a la siguiente relación

$$n \cdot \frac{V_t}{8} \leq B \quad (3.3)$$

donde están relacionados el ancho de banda del medio (B), el número de armónicos (n) y la velocidad de transmisión V_t . Si se analiza la condición crítica de igualdad de la inecuación (3.3) fijando algunos de los parámetros, se obtienen conclusiones muy útiles. En primer lugar, si se desea que el medio transmita un determinado número de armónicos para que la señal sea reconstruida correctamente, la velocidad de transmisión máxima es directamente proporcional al ancho de banda del medio. Si, por el contrario, el ancho de banda del medio es fijo (lo que sucede en la realidad) el número de armónicos que pasan adecuadamente a través del medio es inversamente proporcional a la velocidad de transmisión. Es decir, si aumenta la velocidad el número de armónicos disminuye y por tanto la señal se degrada produciéndose errores en la transmisión, mientras que si disminuye la velocidad, el número de armónicos que pueden pasar aumenta y la reconstrucción de la señal es mejor.

A principios del siglo XX, en 1924, H. Nyquist, quien desarrollaba trabajos acerca del muestreo de señales, demostró que si se hace pasar una señal a través de un medio con un ancho de banda B, dicha señal puede reconstruirse a partir de las muestras tomadas con una frecuencia igual a dos veces el ancho de banda del medio ($f_m = 2B$). Para que un dispositivo digital pueda capturar o leer una señal que se transmite por un medio físico, precisa de un proceso denominado **muestreo**. El muestreador, captura el valor de la magnitud de la señal en instantes de tiempo determinados por el **periodo de muestro T_m** ($\frac{1}{f_m} = T_m$) permitiendo la reconstrucción de la señal de forma aproximada.

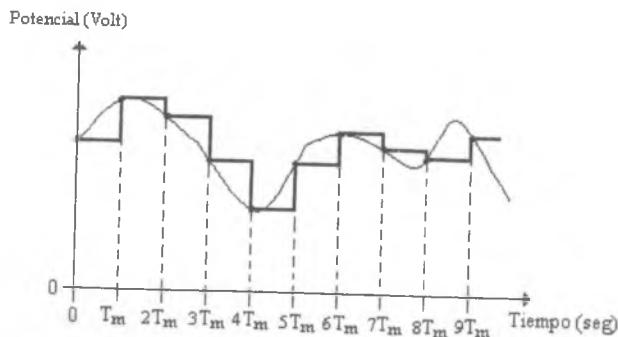


Figura 3.12 Reconstrucción de una señal realizando el muestreo de la misma.

Lógicamente, cuanto menor sea el periodo de muestreo o mayor la frecuencia de muestreo, el sistema será capaz de detectar variaciones temporales de la magnitud de la señal más pequeñas, aumentando así la calidad en la señal de reconstrucción. Sin embargo, existe un límite máximo al aumentar la frecuencia del muestreo a partir del cual ya no se obtiene más información. Dado que la señal se ha transmitido a través de un medio físico, el ancho de banda del mismo habrá limitado la frecuencia máxima en las componentes de la señal. De esta forma, componentes armónicas de frecuencia superior a B Hertzios no existirán. Nyquist llegó a la conclusión que muestreando la señal a $2B$ Hertzios exactamente la señal puede recuperarse en su totalidad, por lo que se habrá transmitido correctamente. Si el sistema muestrea con un periodo igual a $\frac{1}{2B}$ segundos, entonces será capaz de detectar variaciones en la señal de cómo mucho una cada $\frac{1}{2B}$ segundos, siendo por tanto la velocidad de modulación máxima de $V_{máx} = \frac{1}{T_m} = 2B$ baudios. De esta expresión es posible obtener la velocidad máxima de transmisión a través de un medio físico, que vendrá dada por

$$V_{máx} = 2B \log_2 n \text{ bps} \quad \text{Teorema de Nyquist} \quad \text{Teorema 3.1}$$

Supóngase el caso de la línea telefónica comutada en la que el medio físico para la transmisión de datos tiene un ancho de banda de 4000 Hz. Si se emplea una codificación de bits con dos niveles, ¿cuál será la velocidad máxima a la que se podrá transmitir información sin que el medio deteriore la señal?

Aplicando el teorema de Nyquist se obtiene que

$$V_t = 2 \cdot 4000 \text{ Hz} \cdot \log_2 2 = 8000 \text{ bps}$$

Llama la atención que la velocidad máxima de transmisión según el teorema de Nyquist sea bastante inferior a las tasas de 57600 bps que ofrecen las actuales operadoras de telefonía. Ello es debido a que la codificación empleada en la transmisión de datos es de más de dos niveles y se utilizan técnicas de compresión de datos.

3.4 DISTORSIÓN EN EL MEDIO DE TRANSMISIÓN

Hasta ahora se ha analizado la distorsión en la amplitud de las señales que producen los medios de transmisión. Sin embargo ésta no es la única forma en que las señales se distorsionan en un medio. A continuación se analizarán en detalle aquellas que resultan más importantes.

- Atenuación.** La atenuación de una señal cuando se propaga por un medio de transmisión consiste en un decremento en la amplitud de la señal original. Esta atenuación limitará la longitud máxima que se pueda emplear en un medio de transmisión, precisando amplificadores de la señal cuando se precisen distancias mayores. Dado que la atenuación dependerá de la frecuencia (véase figura 3.9) cada componente armónica de la señal transmitida sufrirá una atenuación distinta. Para compensar este efecto existen unos dispositivos denominados **ecualizadores**, que amplifican de forma distinta cada componente armónica de una señal, de forma que la atenuación de la señal en el medio no impide que se recupere correctamente.

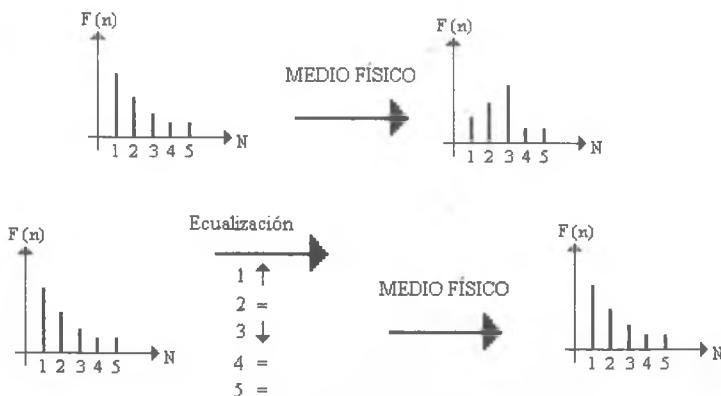


Figura 3.13 Principio de funcionamiento de un ecualizador.

La atenuación de un medio se mide en **decibelios dB** y se define como

$$\text{Atenuación} = 10 \cdot \log_{10} \frac{P_1}{P_2} \text{ dB}$$

siendo P_1 la potencia en **watts** (vatio) de la señal antes de entrar al medio y P_2 la potencia en watts de la señal al salir del medio.

2. Ancho de banda limitado. La limitación en el ancho de banda del medio de transmisión distorsiona las señales al eliminar componentes armónicas de la señal. Esto produce que la señal a la salida del medio físico no sea exactamente la original, al faltar información acerca de los componentes armónicos absorbidos en el medio.

3. Distorsión de retardo. La velocidad de propagación de una señal senoidal pura (armónico) a través de un medio físico varía con su frecuencia. Este efecto produce que cada armónico correspondiente a una señal llegue en diferentes instantes de tiempo al receptor, produciendo una distorsión en la señal. Además, según aumenta la velocidad de transmisión esta distorsión es mayor, por lo que se produce el aumento de la distancia entre dos bits consecutivos y por tanto interferencias en el valor del bit siguiente.

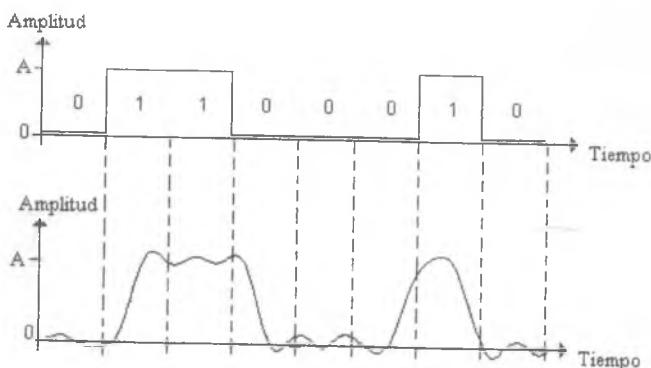


Figura 3.14 Distorsión de retardo.

4. Ruido. El ruido son aquellas perturbaciones aleatorias que están presentes en los medios de transmisión y que producen distorsiones adicionales a las señales que se propagan a través de ellos. Se analizará el ruido con más detalle en el siguiente apartado.

3.5 RUIDO. TIPOS. TEOREMA DE SHANNON

Si en un medio de transmisión no se introduce ninguna señal, cabría esperar que en el extremo de la línea existiera una ausencia de señal. Sin embargo, en los medios de transmisión reales aparecen **perturbaciones aleatorias** en la línea cuando está en estado de **ausencia de señal**. Estas perturbaciones aleatorias reciben el nombre de **ruido de la línea**. Si a esta señal de ruido de la línea se añade la atenuación sufrida al atravesar el medio se obtiene el denominado **ruido de fondo de un medio**.

El ruido presente en un medio se mide mediante la **razón señal a ruido** (*signal to noise ratio*) que **expresa la relación entre la potencia de la señal propagada en el medio y la del ruido**. Esta relación se indica generalmente en decibelios **dB** según la siguiente expresión.

$$\text{Relación S/N (dB)} = 10 \cdot \log_{10} \left(\frac{S}{N} \right)$$

donde S es la potencia de la señal y N la potencia del ruido.



Figura 3.15 Ruido de fondo en un medio de transmisión.

En 1948, Claude Shannon, que ampliaba los estudios realizados por Nyquist, llegó a obtener una expresión que permitía obtener la velocidad máxima de transmisión (bps) en un medio con ancho de banda B y con una determinada relación señal-ruido. Esta velocidad máxima es independiente del tipo de codificación o modulación empleada, estableciendo un límite teórico para la velocidad de transmisión que en la práctica no se alcanza. La relación a la que llegó Shannon se denomina **Teorema de Shannon** y se expresa como

$$V_{\text{máx}} = B \log_2 \left(1 + \frac{S}{N} \right) \text{ bps} \quad \text{Teorema de Shannon} \quad \text{Teorema 3.2}$$

Como ejemplo se realizará el cálculo del límite de Shannon para la línea telefónica de 4000 Hz de ancho de banda con un nivel de ruido de 30dB (valor aproximado en la realidad). Para ello aplicando el Teorema 3.2 se obtiene:

$$V_{\text{máx}} = 4000 \cdot \log_2 \left(1 + \frac{S}{N} \right) = 4000 \cdot \log_2 (1 + 1000) = 4000 \cdot \frac{\log_{10} 1000}{\log_{10} 2} = 39868.91 \text{ bps}$$

$$30 \text{ dB} = 10 \cdot \log_{10} \left(\frac{S}{N} \right) \rightarrow 3 = \log_{10} \left(\frac{S}{N} \right) \rightarrow \left(\frac{S}{N} \right) = 10^3 = 1000$$

Nótese que el límite de Shannon es bastante mayor que el límite de Nyquist. La justificación está en que el límite de Shannon es un límite físico que no es posible superar con ningún tipo de codificación, mientras que el límite de Nyquist se incrementa con el número de niveles. Eso sí, aunque el límite de Nyquist permita obtener velocidades de transmisión superiores al de Shannon, en la práctica esas tasas de velocidad no podrán alcanzarse debido a que las distorsiones producidas por el medio en la señal la hacen irreconocible.

Por otra parte el ruido presente en los medios de transmisión puede clasificarse, atendiendo a su naturaleza, en distintos tipos.

- a) **Ruido cruzado o diafonía** (*crosstalk*). Este ruido se produce por el acoplamiento entre medios de transmisión cercanos (inducción electromagnética).
- b) **Autoacoplamiento**. Se produce cuando una señal de alta intensidad a la salida del DCE induce perturbaciones en una señal débil de entrada al DCE.
- c) **Ruido de impulso**. Debido a la operación de aparatos que generan ruido electromagnético en las cercanías del medio de transmisión.
- d) **Ruido térmico**. Debido a la agitación térmica de los electrones asociados a cada átomo en el material del medio de transmisión.

3.6 FILTRADO DE SEÑALES

Antes de realizar la transmisión de una señal a través de un medio de transmisión se realiza un proceso de **filtrado** de la misma, que consiste en la eliminación de componentes frecuenciales que están fuera del ancho de banda del medio. Como resultado a este filtrado:

- a) La señal a transmitir podrá recomponerse en todas sus frecuencias, de forma que los armónicos transmitidos serán los que se recibirán.
- b) Se evitan las distorsiones que se producen en las componentes de altas frecuencias de la señal, las cuales no son eliminadas por completo por el medio y pueden afectar a la señal recibida.

Eliminando estas frecuencias se evitan posibles errores en la transmisión.

Un filtro se representa gráficamente visualizando una **ganancia** del filtro que amplifica la señal frente a la frecuencia. Un filtro ideal **pasa-baja**, que elimina las componentes frecuenciales de una señal por encima de un determinado valor de frecuencia, se representa como una señal en forma de pulso en el espectro de frecuencias.

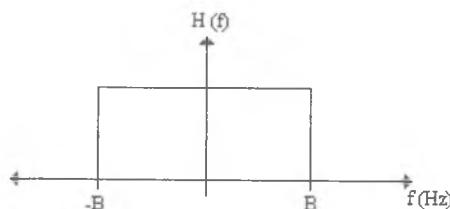


Figura 3.16 Filtro ideal pasa-baja.

Para obtener la señal resultante del proceso de filtrado al hacer pasar la señal a través del filtro se realiza la operación matemática de convolución temporal entre la señal y la función temporal del filtro. Para ello se calcula la **transformada inversa de Fourier** de la función frecuencial del filtro obteniendo su función temporal.

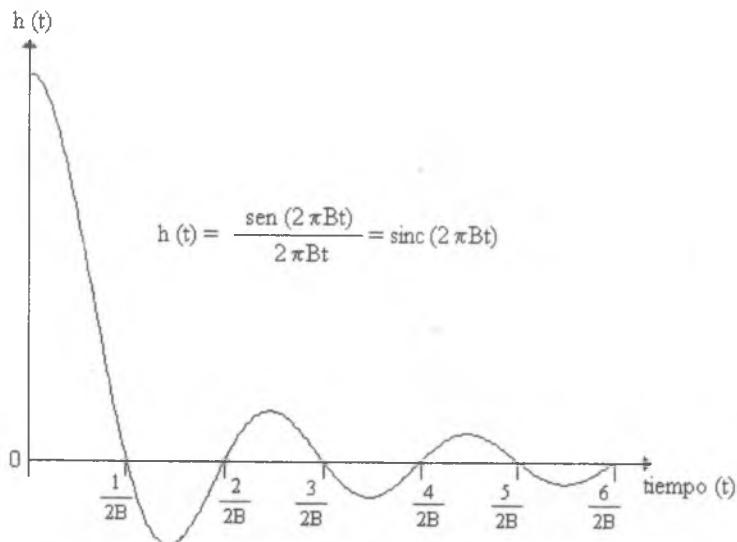


Figura 3.17 Función temporal del filtro ideal.

No existe ningún dispositivo electrónico que permita generar la función temporal asociada al filtro pasa-baja ideal, por lo que en la práctica se emplean otros tipos de filtros que son realizables físicamente y que tienen un efecto similar. Algunos de estos filtros son:

- a) Coseno alzado.

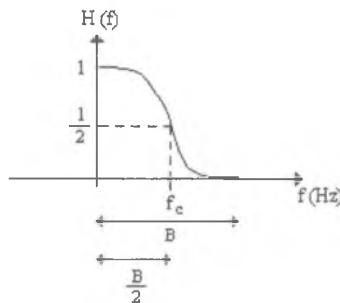


Figura 3.18 Filtro coseno alzado.

- b) Caída senoidal.

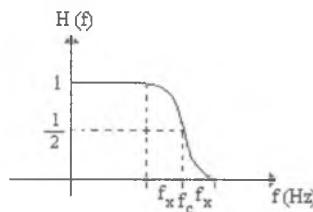


Figura 3.19 Filtro caída senoidal.

- c) Filtro de Butterworth

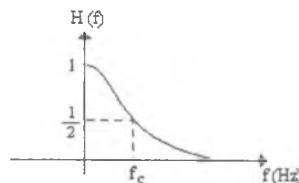


Figura 3.20 Filtro de Butterworth.

CAPÍTULO 4. SEÑALIZACIÓN DE LA INFORMACIÓN

La **señalización** especifica como se indica o señala la información en las señales transmitidas en un medio físico. Existen dos tipos de señalización:

- a) **Señalización en banda base.** La información a transmitir se envía tal cual al medio físico. Dado que en las redes de computadores se transmitirá información digital, es preciso emplear señales digitales. Un ejemplo de señal digital realizable es una señal analógica de pulsos. Se emplea con distancias cortas y velocidades bajas, produciéndose así pequeñas distorsiones en la señal.
- b) **Señalización en banda modulada.** La información a transmitir debe ser adaptada al medio físico antes de ser transmitida. Para ello se realiza un proceso denominado **modulación**, en el que una señal **portadora** que se transmite por el medio modifica alguna de sus características para incorporar la información de una señal **moduladora**.

4.1 SEÑALIZACIÓN EN BANDA BASE

Dentro de la señalización en banda base se abordará el envío de señales de pulsos en un medio físico analógico (medio físico adaptado para el envío de señales analógicas), que es una de las situaciones frecuentes en la transmisión de datos entre computadores. La codificación de las señales digitales en el medio puede ser de varios tipos.

4.1.1 Codificación binaria.

En la codificación binaria a cada valor lógico de la señal digital (cero o uno) se le asigna un nivel de tensión. El inconveniente que presentan este tipo de señales es que la sincronización de la señal entre emisor y receptor

para interpretarla es difícil. Dentro de la codificación binaria existen dos modalidades.

- a) **Codificación binaria sin retorno a cero (NRZ).** En esta codificación el valor de tensión asociado a un nivel lógico (bit) se mantiene constante durante el tiempo que dura el bit. Además la asignación de niveles de tensión puede ser **bipolar** o **unipolar**.



Figura 4.1 Codificación binaria NRZ unipolar.

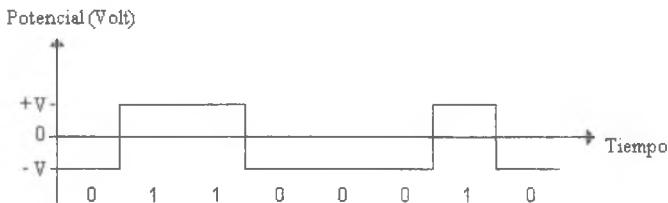


Figura 4.2 Codificación binaria NRZ bipolar.

- b) **Codificación binaria con retorno a cero (RZ).** En esta codificación el valor de tensión asociado a un nivel lógico (bit) se mantiene constante durante la primera mitad del tiempo que dura el bit, tomando el valor de tensión cero durante la segunda mitad. Puede ser además unipolar o bipolar.

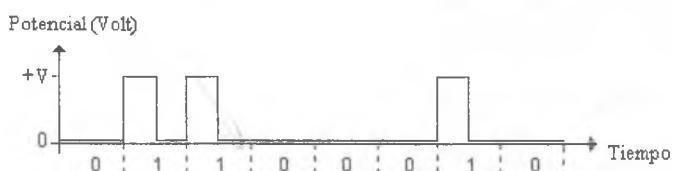


Figura 4.3 Codificación binaria RZ unipolar.

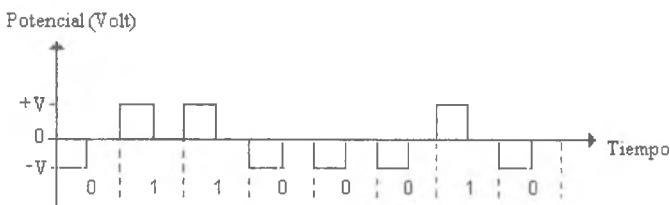


Figura 4.4 Codificación binaria RZ bipolar.

4.1.2 Codificación Manchester.

La codificación Manchester emplea transiciones de nivel, lo que permite una sincronización emisor-receptor más fácil al disponer de una señal de reloj intrínseca en la señal. En la transmisión de un **1 lógico** la primera mitad de la célula de un bit está a nivel bajo (transición $0 \rightarrow 1$), mientras que en la transmisión de un **0 lógico** la segunda mitad de la célula del bit está a nivel bajo (transición $1 \rightarrow 0$).

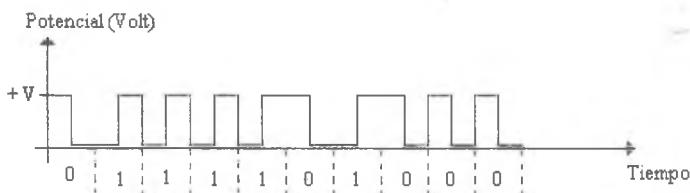


Figura 4.5 Codificación Manchester.

4.1.3 Codificación Manchester diferencial.

La codificación Manchester diferencial es una variación de la Manchester en la que los valores lógicos se asocian a cambios en las transiciones. Inicialmente se asume un tipo de transición ($1 \rightarrow 0$ o $0 \rightarrow 1$) y a partir de este tipo de flanco si se transmite un **0 lógico** el tipo de transición se mantiene y si se transmite un **1 lógico** el tipo de transición cambia.

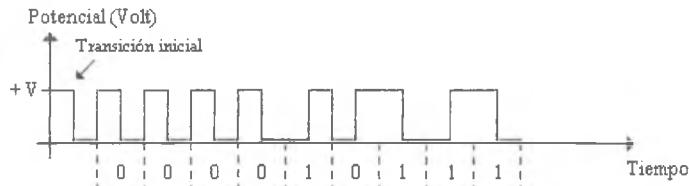


Figura 4.6 Codificación Manchester diferencial.

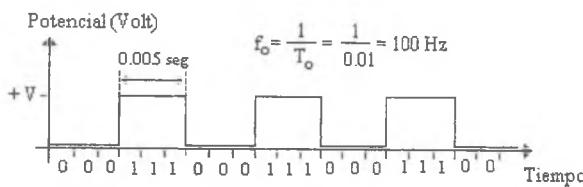
4.2 SEÑALIZACIÓN EN BANDA MODULADA

La señalización en banda modulada consiste en incorporar información de una señal **moduladora** en una señal **portadora** que se transmite de forma adecuada por un medio de transmisión. Puede distinguirse entre dos tipos de modulación atendiendo a la naturaleza de las señales portadora y moduladora. Cuando la señal portadora es analógica y la moduladora digital el proceso de modulación se denomina **modulación analógica**. Si, por el contrario, la portadora es digital y la moduladora analógica la denominación es de **modulación digital**.

4.2.1 Modulación analógica.

Este tipo de modulación se emplea en la transmisión de información digital a través de líneas de transmisión analógicas o que están diseñadas para la transmisión de señales analógicas. En el proceso de modulación la señal portadora modificará la magnitud de alguna de sus características físicas según los valores de la señal digital moduladora, codificando así la información en ella contenida.

Un ejemplo de su uso se encuentra en la transmisión de datos a través de líneas telefónicas empleando los dispositivos **módems** (**modulador-demodulador**) que realizan el proceso de modulación de señales. La línea telefónica posee un ancho de banda de 4000 Hz, siendo el rango de frecuencias transmitidas de 400 Hz a 4400 Hz. Supóngase que se realiza una transmisión empleando señalización en banda base con una codificación binaria NRZ a 600 bps y que se transmiten secuencias consecutivas de 3 1's lógicos y 3 0's lógicos.



Como puede apreciarse en la **figura 4.7** en el medio de transmisión se enviará una señal de pulsos cuya frecuencia fundamental será $\frac{1}{T_0}$. Dado que T_0 , el período de la señal de pulsos es 10^{-2} segundos, la frecuencia fundamental será de 100 Hz. Claramente este valor está fuera del ancho de banda del medio y el 1º y 2º armónico de la señal (los más importantes) no serán transmitidos por el medio.

Por tanto, se hace preciso emplear una señal analógica adecuada al medio, como es una señal senoidal de una frecuencia adecuada al ancho de banda, que modifique el valor de alguno de sus parámetros para codificar la información digital a transmitir. Dependiendo del parámetro que se emplee en la modulación se distinguen tres tipos de modulación.

4.2.1.1 **Modulación por cambio en Amplitud (ASK-Amplitude Shift Keying).**

En la modulación en amplitud, la amplitud de la señal portadora se modifica en función del valor de la señal digital moduladora. Una de las modulaciones en amplitud más sencilla es emplear una señal senoidal de amplitud A, que toma este valor durante el tiempo que dura un bit a 1 y que toma valor de amplitud 0 durante el tiempo que dura un bit a 0.

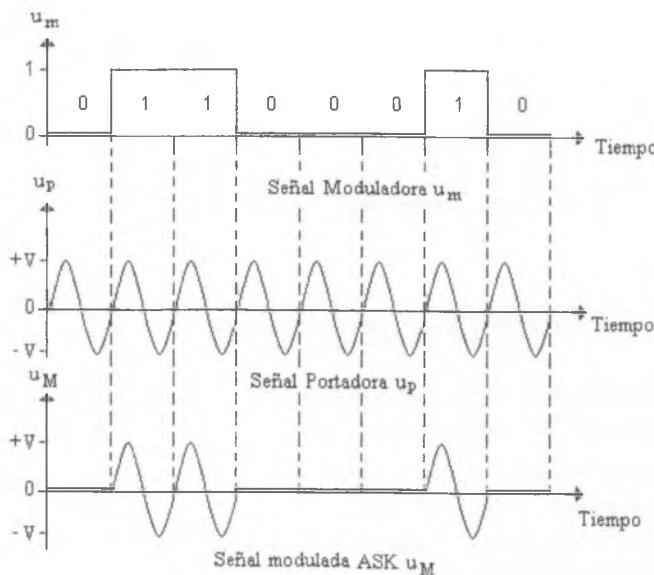


Figura 4.8 Modulación ASK de una portadora analógica.

El proceso de modulación en amplitud puede expresarse analíticamente como

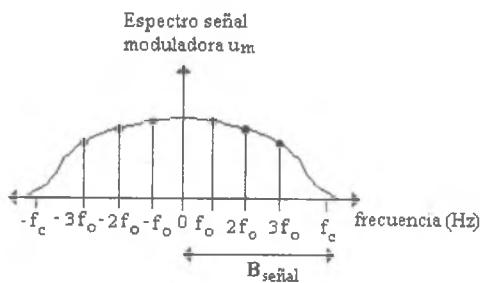
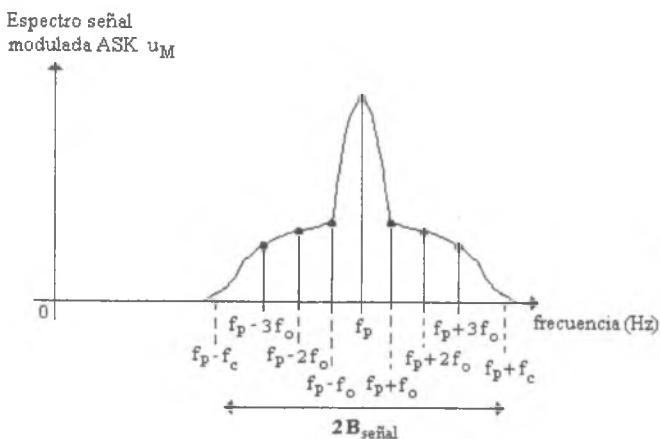
$$u_p(t) = A \operatorname{sen}(2\pi f_p t + \varphi)$$

$$u_m(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(2n\pi f_m t) + \sum_{n=1}^{\infty} b_n \operatorname{sen}(2n\pi f_m t)$$

$$u_M(t) = u_p(t) \cdot u_m(t)$$

donde u_p es la señal portadora, u_m la señal moduladora y u_M la señal modulada.

Si se analiza el rango espectral de las señales moduladora y modulada es posible obtener resultados interesantes.

Figura 4.9 Espectro de frecuencias de la señal moduladora u_m .Figura 4.10 Espectro de frecuencias de la señal modulada u_M .

Puede apreciarse de forma clara que la modulación en amplitud produce un desplazamiento en frecuencia del espectro de la señal moduladora igual a la frecuencia de la señal portadora f_p . Esto produce que se envíe el espectro de la señal moduladora por duplicado en las denominadas **bandas laterales**. Por tanto, el ancho de banda necesario para la señal modulada es el **doble** del ancho de banda de la señal moduladora. Sin embargo, dado que ambas bandas poseen la misma información, es posible eliminar una de las dos bandas y la componente de la portadora empleando un filtro pasa-alta dando lugar a la **transmisión en banda lateral única**. El inconveniente es que entonces la potencia de la señal se reduce en gran medida, por lo que la probabilidad de que se produzca un

error en la transmisión aumenta. Como ventaja se consigue un mejor aprovechamiento del ancho de banda, pues es posible emplear la otra banda lateral para transmitir información adicional. Este tipo de modulación no se empleaba en los primeros módems comerciales, pues las atenuaciones que presentaban las líneas telefónicas antiguas hacían que la comunicación fuera muy susceptible a errores. En la actualidad, con una mayor fiabilidad en las líneas telefónicas, la comunicación entre el usuario y la centralita telefónica se puede realizar empleando modulación ASK.

4.2.1.2 Modulación por cambio en frecuencia (FSK-Frequency Shift Keying).

Este tipo de modulación, que fue la empleada en los primeros modelos de módems, modifica la frecuencia de la señal portadora para incorporar información digital de la señal moduladora. El proceso se realiza en realidad empleando dos portadoras a diferentes frecuencias moduladas en amplitud.

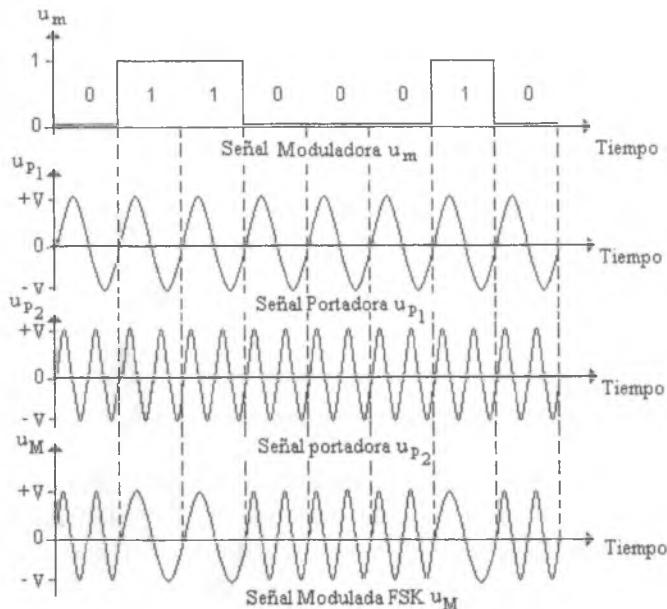


Figura 4. 11 Modulación FSK de portadora analógica.

La forma analítica de la modulación en frecuencia se puede obtener a partir de la de modulación en amplitud de la forma

$$u_{p1}(t) = A_1 \operatorname{sen}(2\pi f_1 t + \varphi)$$

$$u_{p2}(t) = A_2 \operatorname{sen}(2\pi f_2 t + \varphi)$$

$$u_M(t) = u_m(t) \cdot u_{p1}(t) + u'_m(t) \cdot u_{p2}(t)$$

donde $u'_m(t)$ es la señal de pulsos inversa a $u_m(t)$.

El diagrama espectral de la señal modulada, considerando el espectro de la señal moduladora de la **figura 4.9**, proporciona información adicional.

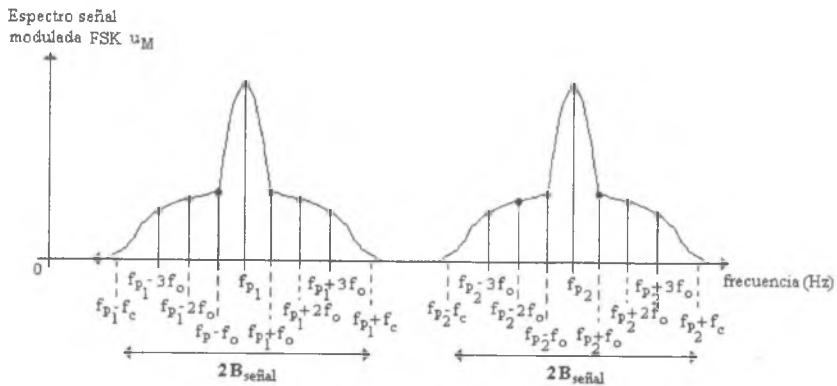


Figura 4.12 Espectro de frecuencias de la señal modulada u_M

Es interesante indicar que las frecuencias de las señales portadoras p_1 y p_2 deben elegirse de forma que no se solapen los espectros de las bandas laterales. Como inconveniente aparece la necesidad de un ancho de banda cuatro veces mayor que el de la señal moduladora, aunque es posible reducirlo a la mitad eliminando las bandas laterales redundantes.

4.2.1.3 Modulación por cambio de fase (PSK-Phase Shift Keying).

La modulación en fase modifica la fase de la señal portadora senoidal en base a la información digital de la señal moduladora. Una de las formas de modulación consiste en asociar un cambio de fase de 0° en la señal modulada cuando se transmite un 0 lógico y un cambio de 180° cuando se transmite un 1 lógico. Este tipo de modulación, en la que los cambios de fase se miden en la propia señal modulada y no respecto de una señal portadora, se denomina modulación PSK diferencial.

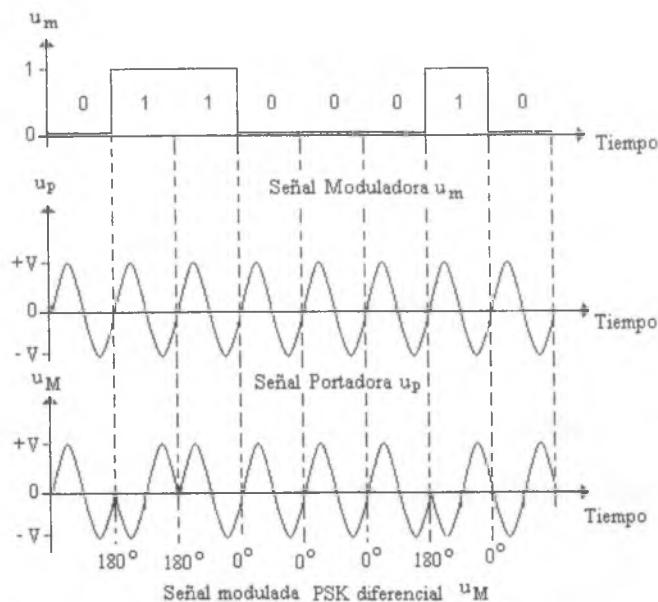


Figura 4.13 Modulación PSK diferencial de portadora analógica.

Si se considera el espectro de la señal moduladora de la figura 4.9, se obtiene el siguiente espectro para la señal modulada PSK.

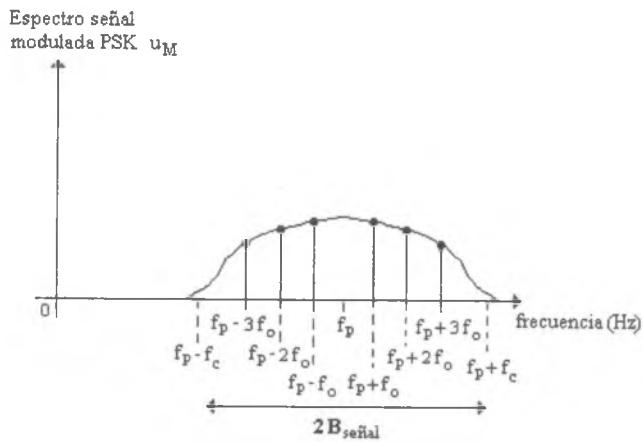


Figura 4.14 Espectro de frecuencias de la señal modulada PSK diferencial.

En el espectro de la señal modulada puede apreciarse como no existe una componente debido a la frecuencia de la portadora, por lo que al eliminar una de las bandas laterales la pérdida de potencia es menor que en la modulación ASK.

4.2.1.4 Métodos de modulación de múltiples niveles.

En las modulaciones anteriores la velocidad de transmisión en bits/seg coincide con la velocidad de modulación o señalización. Es posible conseguir un aumento en la velocidad de transmisión aumentando el número de bits (aumentando el número de niveles) que representa cada variación de la señal.

Si se emplea como modulación base la PSK, el número de niveles puede aumentarse empleando un **diagrama de fase**. En el diagrama de fase se indica la secuencia de bits que codifica cada cambio de fase y amplitud. Cuando se emplea un diagrama de fase es preciso disponer en el receptor de la señal portadora, pues los cambios de la fase se miden respecto de la señal portadora. A este tipo de modulación se denomina modulación PSK de fase coherente.

Como ejemplo véase la modulación en fase definida en el siguiente diagrama, en el que cada bit 0 o 1 tiene asociado un cambio de fase distinto.

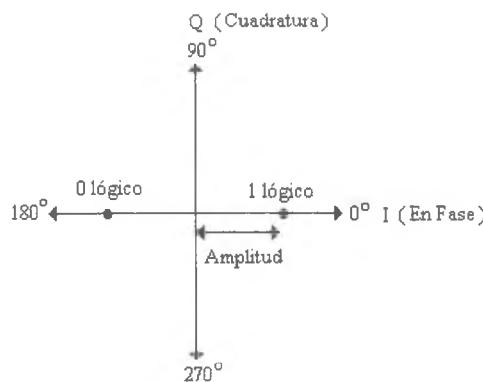


Figura 4.15 Diagrama de fase para una modulación de 1 bit/baudio.

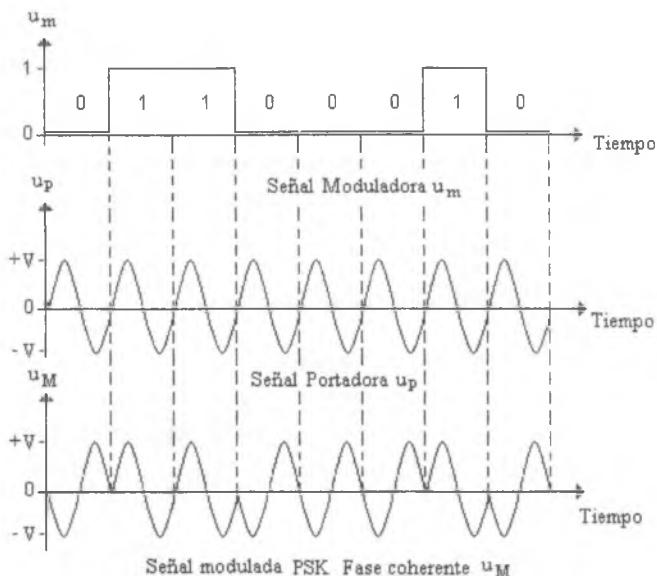


Figura 4.16 Modulación PSK según diagrama de figura 4.15.

Aumentando el número de bits codificados por cambio de fase a 2 se obtiene la denominada **modulación de cambio de fase en cuadratura (QPSK)**.

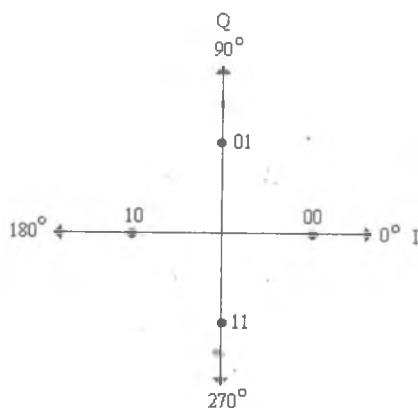


Figura 4.17 Diagrama de fase de la modulación QPSK.

Si además del cambio de fase se tiene en cuenta cambios en la amplitud de la señal portadora se obtiene modulaciones como la **modulación de amplitud en cuadratura** (QAM).

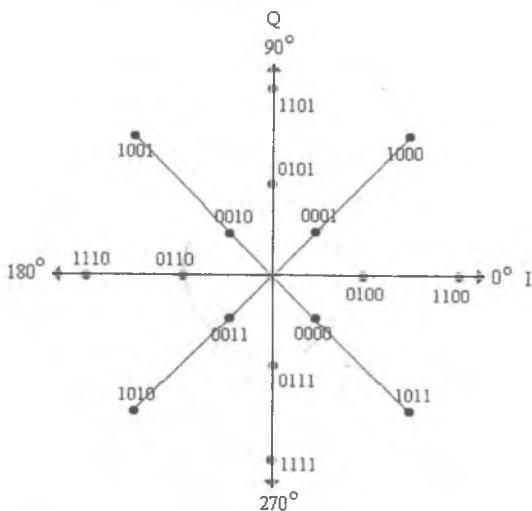


Figura 4.18 Diagrama de fase de la modulación QAM.

De esta forma es posible aumentar la velocidad de transmisión a través de un medio, pero hay que tener en cuenta que cuanto mayor sea el número de niveles más cercanos estarán entre sí. Ello produce un aumento del ancho de banda de la señal modulada y una mayor sensibilidad al ruido en

el medio, por lo que en la práctica no es posible aumentar la velocidad de transmisión cuanto se quiera, como ya lo indicaba Shannon en su teorema.

4.2.2 Modulación Digital.

La modulación digital realiza una modulación de una señal portadora digital en base a la información de una señal moduladora analógica. El origen de este tipo de modulación hay que buscarlo en las aplicaciones en que se precisa la transmisión de señales analógicas a través de un medio de transmisión digital, como es la transmisión de voz a través de la RDSI (Red Digital de Servicios Integrados).

4.2.2.1 Modulación por código de pulsos (PCM).

PCM son las siglas de *Pulse Code Modulation* (Modulación por código de pulsos) y se emplea para la transmisión de información analógica a través de medios de transmisión digital con un elevado ancho de banda. La señal consiste en una secuencia de pulsos de la misma amplitud y duración de pulso situados en determinadas posiciones temporales.

La señal se genera a partir de la señal analógica a modular realizando un proceso de muestreo. El periodo de muestreo empleado se determina a partir del Teorema de Nyquist obteniendo un valor de $f_m = \frac{1}{T_m} = 2B \text{ Hz}$,

siendo B el ancho de banda de la señal analógica. En cada instante de muestreo se genera un pulso de amplitud igual a la amplitud de la señal analógica. Sin embargo, dado que un sistema digital no puede generar un pulso de cualquier valor de amplitud es necesario realizar una cuantización de los valores de amplitud. El pulso sólo podrá tomar determinados valores de amplitud que están limitados por q, el número de niveles del dispositivo conversor analógico-digital empleado. Cuando la señal analógica tenga una amplitud que no se corresponda con ninguno de los niveles q se aproximará al valor más cercano. Esto introduce un error en la reconstrucción de la señal analógica denominado error de cuantización. En la siguiente figura se describe gráficamente como se genera la señal modulada por amplitud de pulsos, también denominada señal PAM.

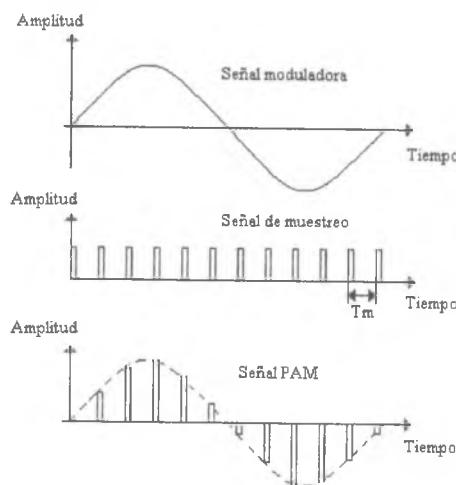


Figura 4.19 Señal PAM generada a partir de una señal analógica.

El siguiente paso en la generación de la señal PCM consiste en la codificación de cada uno de los niveles del conversor A/D y el envío de los valores de cada instante de muestreo. La codificación se realiza empleando un código binario sencillo que precisa n bits, siendo $q = 2^n$. La señal PCM consistirá en el envío de grupos de n pulsos de la misma duración y con amplitud A o cero. Cada grupo de n pulsos codifica un valor de n bits correspondiente a un muestreo donde, si el bit i tiene valor 1 el pulso i tendrá amplitud A, y si el bit i tiene valor 0 el pulso i tendrá amplitud 0. Además cada grupo de n pulsos ha de ser enviado en un tiempo no superior a T_m de forma que en el receptor se pueda recomponer la señal analógica inicial de forma completa y correcta.

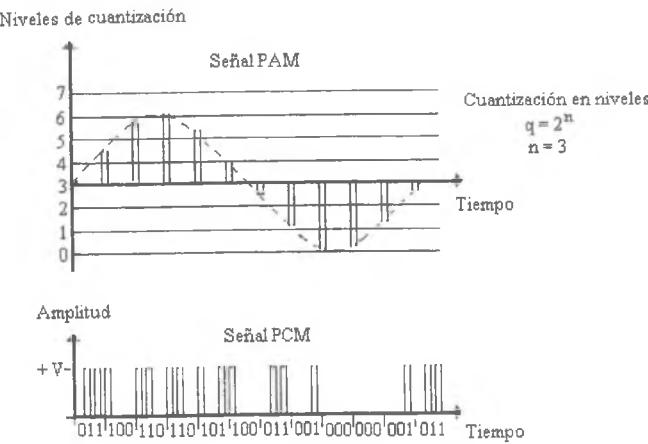


Figura 4.20 Señal PCM generada a partir de la señal PAM de la figura 4.19

Este envío de n pulsos en un tiempo máximo de T_m establece una velocidad mínima de transferencia de los pulsos de la señal PCM. Dado que cada pulso está asociado a un bit, la velocidad de transmisión mínima necesaria en el medio digital vendrá dada por

$$V_{t-med-dig} = \frac{n}{T_m} = n \cdot f_m \text{ bps}$$

Las ventajas más interesantes de este tipo de modulación son la fácil regeneración de la señal original al estar codificada en simples pulsos y la posibilidad de poder realizar multiplexado de varios canales de transmisión empleando división en el tiempo, aspecto que se analizará en la sección 4.3.

Entre las desventajas de la codificación PCM destaca el error debido a los errores de cuantización en el conversor A/D y la limitación en el número máximo de bits que se pueden emplear en la modulación.

Supóngase una señal analógica de ancho de banda $B_{señal}$ que desea ser modulada empleando la modulación PCM y ser enviada a través de un medio digital con un ancho de banda B_{medio} . Aplicando el teorema de Nyquist es conocido que la frecuencia máxima de muestreo con la que se obtiene la mayor cantidad de información de una señal es $f_{m-s} = 2 \cdot B_{señal}$

Hz, siendo el tiempo de muestreo $T_{m-s} = \frac{1}{2 \cdot B_{señal}}$. Dado que se emplea una codificación PCM en cada periodo de muestreo T_{m-s} se envían n bits, por lo que el periodo de muestreo necesario para el medio digital será de $T_{m-m} = \frac{T_{m-s}}{n}$, luego

$$f_{m-m} = \frac{n}{T_{m-s}} = n \cdot 2 \cdot B_{señal} \text{ Hz. (4.1)}$$

A su vez, la frecuencia máxima de muestreo en el medio también está limitada por el teorema de Nyquist, por lo que

$$f_{m-m} = 2 \cdot B_{medio} \text{ Hz (4.2)}$$

Por tanto, combinando las ecuaciones (4.1) y (4.2) se obtiene el número máximo de bits a emplear en la codificación PCM, pues el medio impedirá que se puedan transmitir más.

$$n_{max} = \frac{2 \cdot B_{medio}}{2 \cdot B_{señal}} \quad \text{Teorema 4.1}$$

4.2.2.2 Modulación por código de pulsos diferencial.

La modalidad de PCM diferencial difiere del PCM original en la forma de realizar la cuantización de la señal analógica muestreada. En la cuantización diferencial se determina el valor inicial de la señal analógica y se codifican incrementos y decrementos en el valor original de la señal. La cuantización de estos incrementos se realiza empleando un número de bits inferior al necesario para la modulación PCM, con lo que se consiguen dos ventajas: aumentar la velocidad de transmisión y emplear medios físicos que tienen limitado el número de bits que se pueden emplear en la cuantización.

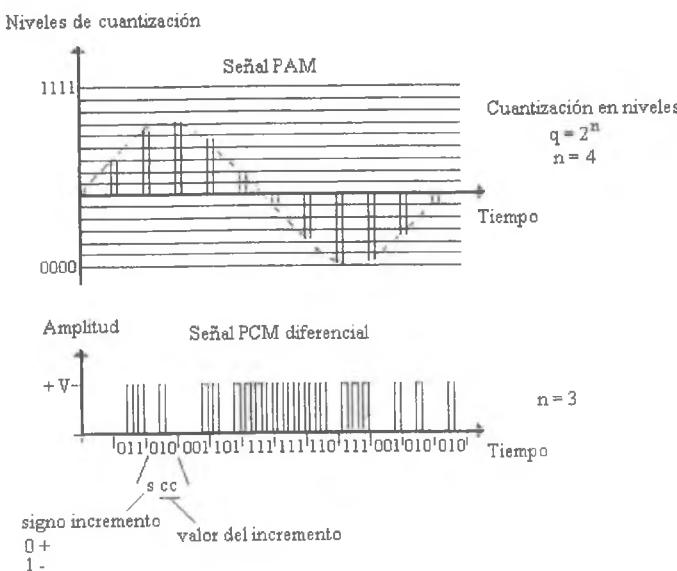


Figura 4.21 Modulación PCM diferencial.

Si estos incrementos y decrementos tienen unos valores predeterminados fijos y se codifican empleando un único bit, la modulación se denomina **modulación Delta**. Un decremento de la señal se codifica como un pulso a 0 y un incremento de la señal como un pulso a 1.

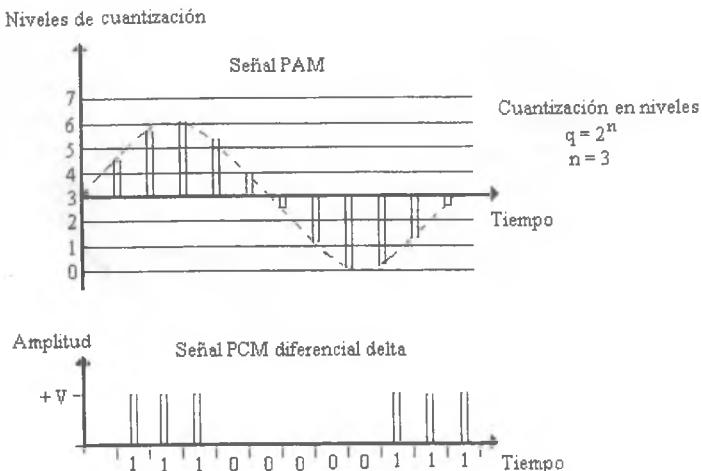


Figura 4.22 Modulación Delta.

La ventaja fundamental de este tipo de modulación es que reduce la cantidad de información a transmitir, por lo que se precisa un ancho de banda menor en el medio digital. Sin embargo presenta varias desventajas, como son la imposibilidad de codificar señales constantes, la limitación en el incremento detectado (si se produce un incremento mayor del establecido como fijo no se detecta) y la falta de precisión, que en última instancia está limitada al valor del incremento seleccionado.

4.2.2.3 Modulación por código de pulsos predictiva.

La modalidad predictiva de la modulación PCM se fundamenta en la interpolación de valores anteriores de la señal para predecir el siguiente valor, transmiéndose la codificación de la diferencia entre el valor real y predicho en cada instante de muestreo.

El funcionamiento de esta modulación precisa la existencia de unidades de predicción estadística de señales tanto en emisor como en receptor para generar las señales de predicción. Inicialmente se negocia un incremento máximo de la señal codificable para las diferencias entre emisor y receptor, cuantizando este incremento en distintos valores dependiendo del número de bits de los conversores A/D.

Esta modulación presenta una mayor precisión en la transmisión de señales analógicas que la PCM diferencial y al emplear menos bits para codificar las diferencias es posible aprovechar mejor el ancho de banda aumentando la velocidad de transmisión.

4.3 MULTIPLEXIÓN

El mecanismo de multiplexión permite establecer varios **canales de datos** en un único **circuito de datos** (véase **figura 3.1**), uniendo varios canales de velocidad moderada en un canal de alta velocidad. Esta multiplexión puede realizarse en dos modalidades: multiplexión por división en frecuencias y por división en el tiempo.

4.3.1 Multiplexión por división de frecuencias (FDM-Frequency Division Multiplexion).

Este tipo de multiplexión **se emplea en líneas de transmisión analógicas**, como son las líneas telefónicas. Su funcionamiento se basa en **repartir el ancho de banda del medio de transmisión en ventanas de frecuencia donde se incorporan los espectros de las señales a transmitir, de forma que pueden enviarse de forma simultánea**. Para ello debe cumplirse que, si

B_{medio} es el ancho de banda del medio físico y n el número de canales a multiplexar,

$$B_{medio} = n \cdot (B_{canal} + \Delta B)$$

donde B_{canal} es el ancho de banda de las señales a multiplexar y ΔB es un factor que evita el solapamiento de los canales en la transmisión.

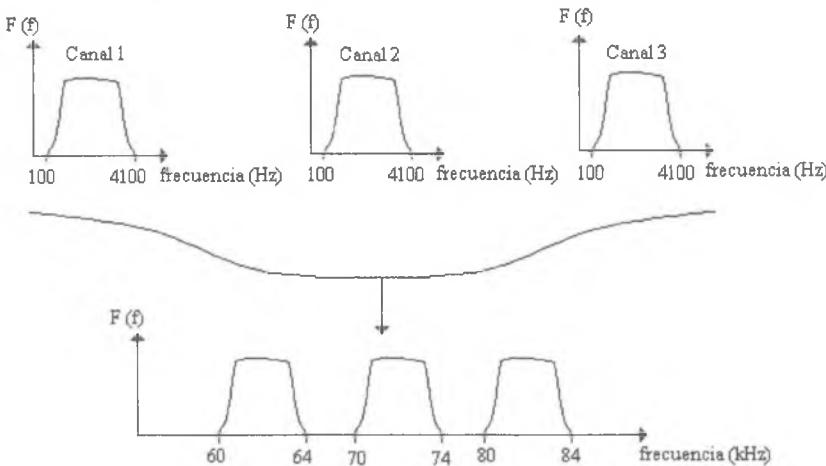


Figura 4.23 Multiplexación por división de frecuencias.

4.3.2 Multiplexión por división en el tiempo (TDM-Time Division Multiplexion).

Este tipo de multiplexión basa su funcionamiento en asignar celdas de tiempo para la transmisión de información de cada canal. Se emplea con frecuencia en medios de transmisión digitales, donde es posible separar la información de cada canal más fácilmente. Existen dos tipos de TDM atendiendo a la estrategia de asignación de la celda temporal a cada canal.

- a) **Multiplexión síncrona.** Los fragmentos de tiempo asociados a cada canal son fijos y asignados antes de iniciar la transmisión.

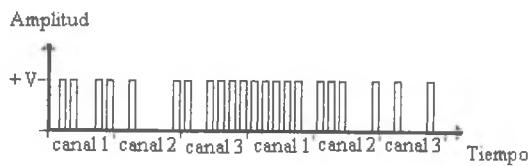


Figura 4.24 Multiplexión TDM síncrona.

- b) **Multiplexión estadística (STDM)**. Los fragmentos de tiempo asociados a cada canal se asignan dinámicamente en base a la demanda, mejorando el rendimiento.

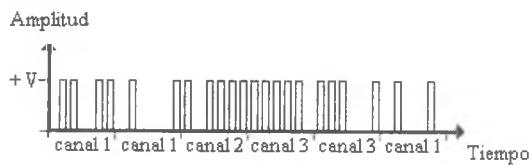


Figura 4.25 Multiplexión TDM estadística.

Un ejemplo de multiplexión temporal síncrona es la empleada en la transmisión de varios canales de voz digitales empleando modulación PCM. La multiplexión de canales de transmisión empleando división en el tiempo en PCM está normalizada en dos estándares: El norteamericano-japonés y el europeo. La normativa de EEUU-Japón se denomina **T1**. En esta normativa cada tiempo de muestreo es una celda de $125\mu s$ (es el tiempo de muestreo aconsejado para el muestreo de señales de voz) donde se multiplexan 24 canales. Cada uno de esos canales consiste en 8 bits de datos correspondientes al valor de una muestra de un canal. Existe una excepción y es que en los canales 6 y 12 se emplea un bit para señalización, por lo que quedan 7 disponibles para datos. Al final de la celda de $125\mu s$ se añade un bit para señalización de la celda, lo que hace un total de 193 bits por celda. A partir de estos datos es fácil calcular la velocidad de transmisión necesaria del medio digital que será de **1.544 Mbps**.

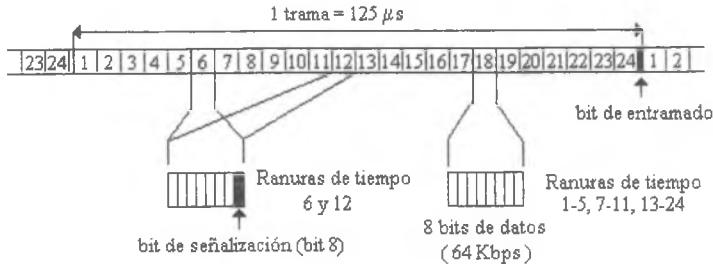


Figura 4.26 Multiplexión T1.

Por otra parte, la norma Europea denominada **E1** emplea celdas con la misma duración temporal pero con 256 bits divididos en 30 canales de datos de 8 bits y 2 canales de 8 bits para sincronización y señalización. Esto supone una velocidad de transmisión para el medio de **2.048 Mbps**.

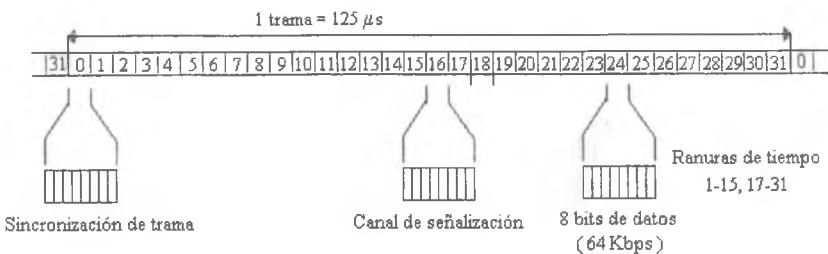


Figura 4.27 Multiplexión E1.

4.4 MODALIDADES DE TRANSMISIÓN

La transmisión de datos digitales puede clasificarse en diferentes tipos atendiendo a diversos factores. Si se tiene en cuenta la **forma de secuenciación de los bits** en la transmisión, es posible distinguir entre:

- Transmisión paralela:** cuando se transmiten secuencias de bits de forma simultánea por varias líneas físicas diferentes. La principal ventaja que se consigue con este tipo de transmisión es el aumento de la velocidad de transmisión, a costa de realizar

una mayor inversión económica en el cableado y la presencia de interferencias entre las líneas. Un ejemplo de este tipo de transmisión se encuentra en el puerto paralelo de un PC.

- b) **Transmisión serie:** cuando se transmiten secuencias de bits seguidos a través de una única línea física. El menor coste en cableado que supone este tipo de transmisión impide alcanzar velocidades muy elevadas si se compara con la transmisión en paralelo. Este tipo de transmisión es empleado en los puertos de comunicación serie de un PC.

Si se tiene en cuenta la **simultaneidad en la comunicación entre dos dispositivos** es posible distinguir tres tipos:

- a) **Comunicación simplex.** Cuando la comunicación entre dos dispositivos es unidireccional, en un solo sentido.
- b) **Comunicación semidúplex.** Cuando la comunicación entre dos dispositivos es bidireccional, en los dos sentidos, pero no simultánea.
- c) **Comunicación duplex o full duplex.** Cuando la comunicación entre dos dispositivos es bidireccional y simultánea.

Por último, es posible establecer una clasificación en función del **sincronismo en la transmisión**, de forma que se distingue entre:

- a) **Transmisión síncrona.** Se emplea cuando han de transmitirse grandes volúmenes de bits a una velocidad elevada. En este tipo de transmisión, cuando se realiza la transmisión de un bloque de datos se envía al principio de éste una secuencia de bits especiales que permiten al receptor sincronizarse para interpretar correctamente los bits. Una vez leídos los bits de sincronización, el receptor lee toda la trama de bits utilizando la información de duración de bit proporcionada por los bits de sincronización. En otras ocasiones, se acompaña a la señal de datos una señal adicional de reloj, que informa al receptor de los instantes en los que debe interpretar la información de la señal de datos.
- b) **Transmisión asíncrona.** Se emplea cuando la tasa de bits a emplear no es conocida o es variable y el volumen de datos es bajo. En este tipo de transmisión la sincronización del receptor se realiza por cada unidad de datos que se quiera transmitir, generalmente un byte. Cada byte de datos dispondrá de un serie de bits de sincronización para que el receptor interprete

correctamente la información. Este tipo de transmisión es la empleada por los puertos de comunicación serie de un PC.

CAPÍTULO 5. MEDIOS DE TRANSMISIÓN

El elemento más importante dentro del nivel físico en la transmisión de datos es sin duda el medio físico de transmisión. De las características que éste presente dependerá en medida la fiabilidad del medio, tasa de velocidad de transmisión, limitaciones geográficas de la comunicación, etc. A continuación se abordarán distintos medios de transmisión empleados en la comunicación entre computadores, analizando sus características y para qué aplicaciones son más adecuados.

5.1 CABLES ELÉCTRICOS

El medio físico más económico y empleado en las redes de comunicaciones es el cable eléctrico. Cualquier cable eléctrico, un hilo de material conductor (generalmente cobre) protegido por un material aislante, puede ser modelado empleando un **modelo de parámetros distribuidos**.

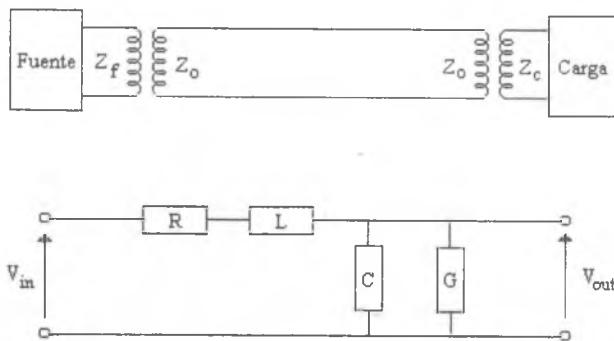


Figura 5.1 Modelo de parámetros distribuidos de un cable eléctrico.

En la figura anterior se indica el esquema de conexión de una fuente y una carga para la transmisión de una señal eléctrica entre la fuente y la carga a través de un par de hilos eléctricos. Cada uno de los elementos (fuente, carga y cable) presenta una **resistencia distinta a las señales senoidales**

denominada **impedancia Z** que se mide en **ohmios**. A su vez, el cable eléctrico posee una resistencia a la corriente continua denominada tradicionalmente como resistencia eléctrica **R**, medida en Ohmios/metro, una autoinductancia debido al campo magnético que genera cada uno de los hilos **L**, medida en Henrios/metro, una capacidad entre ambos hilos **C**, medida en Faradios/metro, y una **admitancia G** entre ambos hilos, medida en Siemens/metro. Esta admitancia mide la resistencia entre los dos hilos y se define como la inversa de la impedancia, Z^{-1} .

Los valores de cada uno de los parámetros del cable Z , R , L , C y G han de escogerse en base a dos criterios que permiten una transmisión de señales eléctricas lo más óptima posible.

- 1) **La señal enviada no debe reflejarse al llegar al otro extremo de la línea.** Para poder permitir esta condición ha de cumplirse que la impedancia de la línea sea igual a la de la carga, es decir

$$Z_c = Z_0$$

- 2) **La atenuación de la señal debe ser mínima e independiente de la frecuencia de la misma.** Para ello debe cumplirse la relación

$$RC = GL$$

Esta restricción no puede cumplirse nunca debido a restricciones físicas, pero es posible conseguir una aproximación que funcione adecuadamente.

5.1.1 Par Paralelo.

El cable par paralelo consta de dos hilos de cobre recubiertos por un material aislante y que se colocan de forma paralela. Se trata éste de un medio poco fiable, pues es muy sensible al ruido. Los hilos actúan como antenas de transmisión y recepción de señales electromagnéticas que inducen ruido en la comunicación y presentan inducción electromagnética entre sí. Dentro de la comunicación de datos, este medio se emplea cuando las distancias que separan los extremos de la comunicación son pequeñas (menos de 50m.) y las velocidades de transmisión bajas (menor de 19.2 Kbps aproximadamente). Las aplicaciones más usuales están en la comunicación de un equipo DCE a un equipo DTE.

5.1.2 Par trenzado.

El par trenzado es una variante del par paralelo, en el que el hilo de masa rodea al hilo de señal, anulándose los efectos de las autoinducciones entre los hilos y reduciendo las señales de ruido electromagnético externas. Sus aplicaciones se encuentran en la telefonía analógica y las redes de área local de baja velocidad (inferior a 1 Mbps). El par trenzado permite transmitir a velocidades de entre 1 y 5 Mbps cuando las distancias son inferiores a los 100 m. Es posible alcanzar distancias más elevadas reduciendo la tasa de velocidad, aunque a partir de distancias de 1 Km es recomendable el uso de amplificadores.

En la actualidad existen diferentes categorías de cables de par trenzado que se adaptan a las características de transmisión de alta velocidad en las redes LAN. Los cables par trenzado se dividen en dos tipos principales: cables de par trenzado no blindado (UTP - *Unshielded Twisted Pair*) y cables de par trenzado blindado (STP - *Shielded Twisted Pair*).

- a) **Cables UTP.** El cable UTP es el típico par de hilos trenzados entre sí. Dependiendo de la torsión del par de hilos sobre sí mismo, el cable presentará mayor o menor diafonía y atenuación. Dentro del cable UTP se establecen diferentes categorías según la torsión del cable sea mayor o menor, destacando las categorías 3 y la 5, que presenta mayor torsión que la categoría 3.

Categoría 3: Se emplea en redes de área local de velocidades de 10 Mbps, permitiendo alcanzar velocidades de 30 Mbps a distancias de 100 metros.

Categoría 5: Se emplea en redes de área local de 100 Mbps, permitiendo alcanzar velocidades de 100 Mbps a distancias de 100 metros.



- b) **Cables STP.** El cable STP consiste en un par de hilos de cobre al que se le añade un apantallamiento similar al del cable coaxial para evitar el ruido externo. Este cable permite alcanzar velocidades de hasta 800 Mbps a distancias de 100 metros.

5.1.3 Cable coaxial.

La estructura del cable coaxial consiste en un conductor eléctrico rodeado por un material dieléctrico o aislante. Alrededor del dieléctrico se coloca

una malla de un material conductor que a su vez está recubierto por un material aislante.

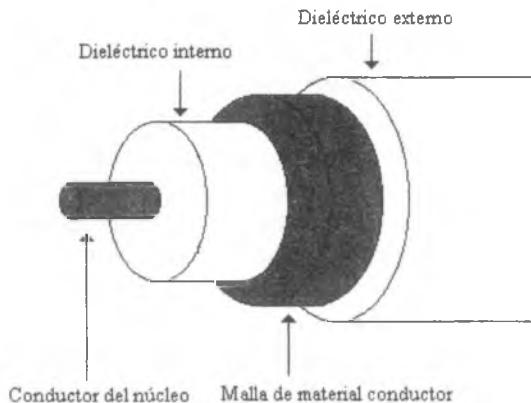


Figura 5.2 Estructura de un cable coaxial.

El conductor interno actúa como portador de la señal eléctrica y la pantalla exterior es la masa. Además el conductor interno y la malla poseen distinta impedancia, lo que hace que el par se denomine par no equilibrado o balanceado.

Este medio presenta gran inmunidad al ruido electromagnético, pues el campo eléctrico en el interior de una superficie conductora (malla) es nulo. Esto permite que el conductor eléctrico presente muy poco ruido en la señal que transmite. Esta gran inmunidad al ruido hace del cable coaxial un medio con un gran ancho de banda, permitiendo alcanzar velocidades de hasta 10 Mbps a varios cientos de metros. Dentro del cable coaxial es posible distinguir entre dos tipos: el cable coaxial de $50\ \Omega$ de impedancia (empleado en la transmisión en banda base digital) y el de $75\ \Omega$ de impedancia (empleado en la transmisión analógica).

Cable Coaxial de $50\ \Omega$.

Este cable presenta una gran adecuación a la transmisión de señales digitales (señales de pulsos) en banda base empleando codificación Manchester y Manchester Diferencial, permitiendo alcanzar velocidades de hasta 10 Mbps a distancias de cientos de metros. Sus aplicaciones más comunes son las redes de área local LAN y la telefonía de larga distancia.

Generalmente las **LAN** emplean el cable coaxial como un medio físico compartido, de forma que todos los ETD están conectados al cable coaxial y envían señales al medio. Si en algún momento dos ETD pretenden transmitir información simultáneamente se producirá una señal de interferencia en el medio perdiéndose la información transmitida. Es lo que se denomina una **colisión**. La conexión del ETD al cable coaxial es posible realizarla de dos formas:

- a) **Conexión en T.** El cable coaxial se corta en un punto determinado y se coloca un conector de derivación en forma de T que permite al ETD enviar y recibir señales del medio. Esta modalidad precisa de la interrupción del funcionamiento de la red.
- b) **Conexión Vampiro.** Este tipo de conexión no requiere la interrupción del funcionamiento de la red. El ETD tendrá acceso al cable coaxial a través de un conector que perfora el cable coaxial alcanzando el núcleo conductor y estableciendo contacto físico para transmitir señales. El inconveniente es que este tipo de conexión es muy delicada y poco fiable.

Cable Coaxial 75Ω .

Este cable presenta una gran adecuación a la **transmisión simultánea de señales analógicas** de distinta frecuencia asociadas a distintos canales de información: **transmisión broadband** (transmisión en banda ancha). Este cable tiene un ancho de **banda de hasta 300 Mhz** y una de sus aplicaciones más importantes es la transmisión de **vídeo a la carta**.

Este servicio, ofrecido por numerosas compañías de **televisión por cable**, permite que el usuario seleccione cuáles son sus canales de vídeo o televisión preferidos y pague por la visualización de los mismos. Por tanto, es preciso una comunicación bidireccional entre usuario y proveedor que puede realizarse de dos formas:

- a) **Hilo doble.** Empleando un cable coaxial para transmisión y otro para recepción. Es una opción muy cara y poco empleada por los proveedores de servicios.
- b) **Hilo único.** Empleando un único cable con dos canales, uno para transmisión y otro para recepción, multiplexados en frecuencia. Para la comunicación usuario → proveedor se emplea un canal de bajo ancho de banda (rango de frecuencias 50Hz - 30 MHz) pues el volumen de información a transmitir es bajo, y para la comunicación proveedor → usuario un canal de alto ancho de

banda (rango de frecuencias 40 MHz - 300 MHz) para proporcionar los diversos canales de vídeo.

Aunque el cable coaxial de $75\ \Omega$ esta adecuado para la transmisión de señales analógicas, es posible emplearlo para la transmisión de datos digitales. Para ello es preciso modular la señal digital antes de transmitirla empleando conversores digital-analógico (D/A) y analógico-digital (A/D) para recuperarla del medio (la transmisión en banda base digital presenta una fuerte atenuación). Una aproximación del consumo de ancho de banda para la transmisión de datos digitales a una determinada tasa de bps es que se precisa un ancho de banda de 1 a 4 Hz por cada bps, lo que permite velocidades máximas de transmisión entre 200 y 75 Mbps.

5.2 FIBRA ÓPTICA

La fibra óptica es un medio de transmisión de la luz que se emplea en la transmisión de datos codificando la información digital mediante pulsos de luz. Una de las principales ventajas de este medio es que al propagar señales luminosas es posible aprovechar la alta frecuencia de las mismas y conseguir así elevadas tasas de velocidad de transmisión en bps. Además, la inmunidad al ruido electromagnético externo, la flexibilidad de las fibras y las enormes distancias de trazado que permite, la convierten en un medio físico muy potente que en la actualidad es un estándar para redes de comunicación que precisan un elevado ancho de banda a grandes distancias.

5.2.1 Estructura.

Una fibra óptica está compuesta, en similitud al cable coaxial, por un núcleo interno de cristal de sílice cuyo diámetro oscila entre 5 y 100 μm rodeado por una recubrimiento de silicona. Alrededor de la silicona se dispone una capa de poliuretano que actúa como protección ante los agentes externos.

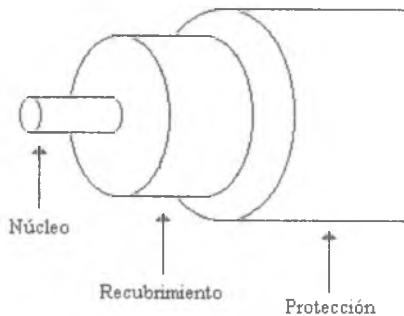


Figura 5. 3 Estructura de una fibra óptica.

Las longitudes de onda empleadas para las señales luminosas se encuentra entre 200 nm y 1 μm. En estas longitudes el efecto de dispersión cromática (dependiendo de la longitud de onda λ de una señal luminosa, el ángulo de refracción al atravesar un medio varía) es prácticamente nulo, por lo que los dispositivos empleados para generar los pulsos luminosos puede emplear señales con cierto ancho espectral (no es posible generar un haz luminoso a una longitud de onda λ exacta, el haz dispondrá de longitudes de onda entre $\lambda - \Delta\lambda$ y $\lambda + \Delta\lambda$. Cuanto menor sea $\Delta\lambda$ más complejo y caro es el dispositivo).

5.2.2 Modelo de propagación.

De especial interés es el fenómeno de propagación de ondas luminosas en una fibra óptica, pues precisamente el hecho de que una señal luminosa pueda confinarse y propagarse en un espacio físico limitado es lo que permite que sea empleada para la transmisión de datos.

En primer lugar ha de introducirse el concepto de **índice de refracción** de un medio **n**, que mide la relación entre la velocidad de propagación de la luz en el vacío ($V_c = 3 \cdot 10^8 \text{ m/s}$) y la velocidad de propagación en ese medio V_n .

$$n = \frac{V_c}{V_n}$$

La trayectoria de un haz de luz incidente sobre el límite entre dos medios con índice de refracción distinto se modifica, pudiéndose producir dos fenómenos:

- Reflexión**, en el que la onda rebota en la superficie de separación de ambos medios. La trayectoria de la onda reflejada tiene entonces el mismo ángulo respecto a la perpendicular de la superficie de separación de ambos medios que la onda incidente.
- Refracción**, en el que la onda atraviesa la superficie de separación de ambos medios modificando el ángulo de salida respecto de la perpendicular. El ángulo de salida cumple entonces la Ley de Snell.

$$n_1 \operatorname{sen} \theta_{inc} = n_2 \operatorname{sen} \theta_{refrac}$$

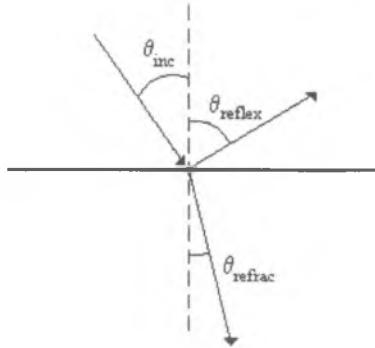


Figura 5.4 Reflexión y Refracción de una onda luminosa.

Si un haz luminoso incide en una fibra óptica, parte del haz es absorbido por la envoltura y el recubrimiento. La parte del haz que incide en el núcleo lo hará con cierto ángulo en la superficie que separa el núcleo de la envoltura, de forma que podrá reflejarse o atravesarlo modificando la trayectoria. En el caso de que el ángulo de incidencia en la superficie de contacto de núcleo y recubrimiento fuera el adecuado, el haz se reflejará en el interior del núcleo de la fibra, y mediante sucesivas reflexiones se propagará al otro extremo de la misma. El ángulo mínimo incidente a partir del cual el haz quedará confinado en el núcleo de la fibra se denomina **ángulo crítico** $\vartheta_{crítico}$ y es posible calcularlo a partir de la Ley de Snell.

$$\vartheta_{crítico} = \arcsen\left(\frac{n_{rec}}{n_{nuc}}\right)$$

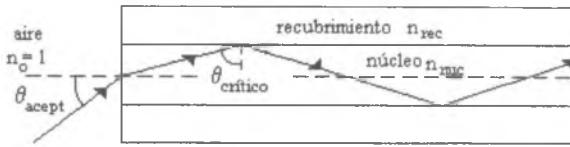


Figura 5.5 Ángulo crítico y ángulo de aceptación de una fibra óptica.

Por otra parte, se define el **ángulo de aceptación** ϑ_{acept} como el ángulo máximo que forma el haz incidente con el eje del núcleo de la fibra a partir del cual no es posible que el haz luminoso quede confinado en la fibra. Éste es el ángulo máximo con el que el dispositivo emisor debe hacer incidir el haz en la fibra para su propagación en la misma. Frecuentemente, en vez de proporcionar el ángulo de aceptación se indica la **apertura numérica** de una fibra, que viene dada por

$$AN = \operatorname{sen} \vartheta_{acept} = \sqrt{n_{nuc}^2 - n_{rec}^2}$$

Como ejemplo puede calcularse la apertura numérica y el ángulo crítico para una fibra de sílice con índice de refracción para el núcleo de 1.62 y de 1.52 para el revestimiento.

$$\vartheta_{crítica} = \arcsen\left(\frac{n_{rec}}{n_{nuc}}\right) = \arcsen\left(\frac{1.52}{1.62}\right) \approx 69.76^\circ$$

$$AN = \sqrt{n_{nuc}^2 - n_{rec}^2} = \sqrt{1.62^2 - 1.52^2} = 0.560$$

5.2.3 Tipos de fibra óptica.

Dependiendo de las características del núcleo de la fibra óptica es posible hacer una clasificación en tres tipos.

- a) **Fibras Multimodo o de salto de índice.** Estas fibras presentan la característica de que el núcleo de la misma tiene un índice de refracción

constante en toda la sección del núcleo. En este caso los distintos haces de luz que inciden en la fibra dentro del ángulo de aceptación se propagan con trayectorias diferentes. Cada una de estas trayectorias recorre un camino óptico diferente (longitud física de trayectoria diferente), por lo que los haces de luz llegan al extremo de la fibra con desfases temporales (**dispersión intermodal**). Este efecto limita la frecuencia a la que se pueden enviar pulsos de luz para evitar el solapamiento de los mismos, de forma que limita la velocidad máxima de transmisión en bps.

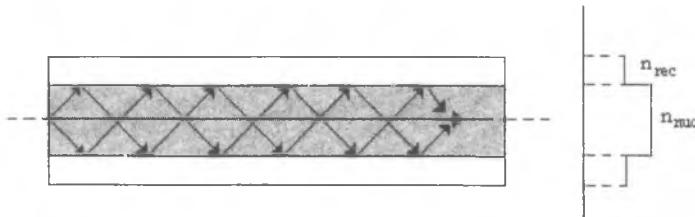


Figura 5.6 Propagación en fibras multimodo.

- b) **Fibras de índice gradual.** Este tipo de fibras se caracteriza por presentar un núcleo con índice de refracción variable en la sección del núcleo. El índice de refracción en el eje de la fibra es de $n_{núcleo}$ y disminuye hacia el exterior hasta alcanzar el valor de n_{rec} en el recubrimiento. Esta estructura consigue que los haces de luz que recorren una mayor distancia óptica lo hagan a mayor velocidad que los de menor recorrido óptico, por lo que se eliminan las desfases temporales entre los haces ya que éstos convergen hacia el eje del núcleo. De esta forma se consigue aumentar la velocidad de transmisión de datos en la fibra.

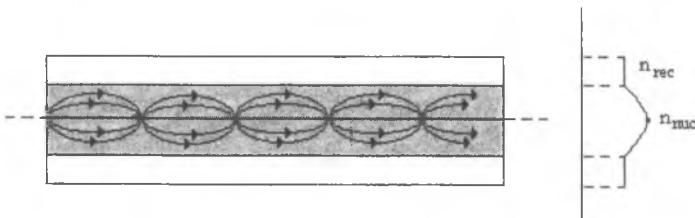


Figura 5.7 Propagación en fibras de índice gradual.

c) **Fibras monomodo.** Este tipo de fibras es un subconjunto dentro de las fibras de índice de salto. Las fibras monomodo son fibras de índice de salto que permiten la propagación de un único haz de luz en el núcleo. Esto es posible debido a un diámetro del **núcleo de la fibra muy pequeño**, alrededor de $5 \mu\text{m}$. La teoría ondulatoria determina un conjunto finito de trayectorias de haces de luz que pueden ser transmitidas por una fibra, separadas entre sí por un determinado ángulo. Este ángulo está relacionado con la longitud de la onda luminosa que se propaga y el diámetro de la fibra. Bajo ciertas condiciones de longitud de onda y diámetro de núcleo, el ángulo de aceptación de haces en la fibra es tal que solo permite que un único haz axial sea propagado por la misma (*por ejemplo, para una fibra de núcleo de silice con $n_{\text{núcleo}}=1.447$ y empleando una longitud de onda $\lambda=1.3 \mu\text{m}$, si se desea que se propague un único haz axial, el núcleo de la fibra ha de ser inferior a $4.86 \mu\text{m}$*). En la propagación de una haz en una fibra monomodo se hace más notable una distorsión (también presente en los otros tipos de fibras) denominada **distorsión intramodal**. Esta distorsión es debida a que en la práctica no es posible generar un haz luminoso de una única longitud de onda, sino que tiene componentes de distinta longitud en un determinado ancho de banda. Dado que cada onda con una longitud se propaga a una velocidad distinta, en el extremo de la fibra aparecerá un pulso de luz distorsionado por el retardo debido a cada onda del haz. Esta distorsión es de menor magnitud que la de la distorsión intermodal, por lo que en las fibras donde está presenta esta distorsión la intramodal no es muy significativa.

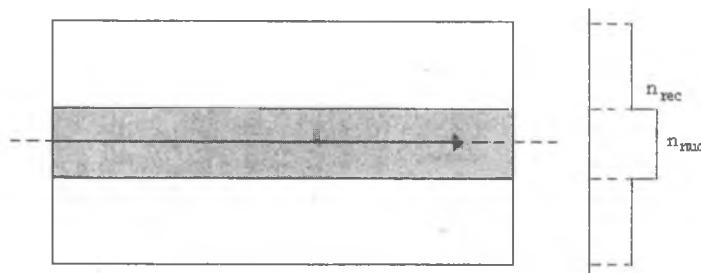


Figura 5.8 Propagación en fibras monomodo.

5.2.4 Velocidad de transmisión.

La velocidad máxima de transmisión de datos en una fibra óptica estará determinada por dos factores: la codificación de los bits en pulsos luminosos y la frecuencia máxima para los pulsos que permita la fibra.

La codificación de los bits de información en pulsos luminosos es muy variada, y como ejemplo se analizará un esquema de codificación bipolar. En este esquema, cada nivel lógico estará codificado en tres niveles de potencia lumínosa del haz de luz, utilizando transiciones diferentes para el 0 y el 1.

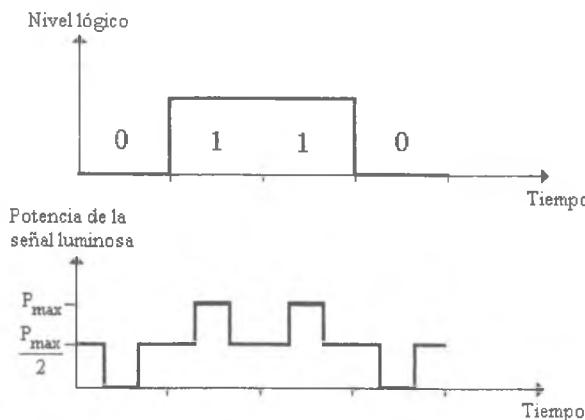


Figura 5.9 Codificación de bits en señales luminosas para fibras ópticas.

La frecuencia máxima de los pulsos que pueden ser enviados por la fibra estará limitada por el efecto de la dispersión intermodal. Las fibras monomodo, en las que la dispersión intermodal es nula, serán las que alcancen mayores frecuencias en los trenes de pulsos, y por tanto mayores velocidades de transmisión. Por otra parte hay que notar que la atenuación en las fibras ópticas es muy pequeña, inferior a **0.2 dB/Km**, por lo que la transmisión puede realizarse a grandes distancias (cientos de kilómetros) sin pérdidas cuantitativas en la potencia de la señal.



TIPO DE FIBRA	ANCHO DE BANDA (Hz/Km)
Multimodo	20 MHz/Km
Índice gradual	500 - 1000 MHz/Km
Monomodo	1 - 10 GHz/Km

Tabla 5.1 Anchos de banda de fibras ópticas.

5.2.5 Dispositivos luminosos. Conexión de fibras ópticas.

Los dispositivos luminosos y detectores son elementos principales en la transmisión de datos por fibras ópticas. Sus características (frecuencias máximas de conmutación, convergencia de los haces generados, etc.) limitarán la velocidad máxima de transmisión de datos en las fibras. Debe hacerse una distinción entre dispositivos emisores de ondas luminosas y receptores.

- a) **Dispositivos emisores.** Existen dos tipos de dispositivos emisores para generar los haces luminosos para las fibras. Los **diodos emisores infrarrojos (IRED)**, que generan ondas luminosas con un ancho de banda de 40 nanómetros y presentan tiempos de conmutación del orden de los 20 nanosegundos. Los **diodos láser** proporcionan mejores prestaciones al generar ondas con un ancho de banda de 2 nanómetros y presentar tiempos de conmutación del orden del nanosegundo.
- b) **Dispositivos receptores.** El dispositivo más empleado es el **fotodiodo semiconductor en avalancha (APD)**, que genera una señal eléctrica proporcional a la luz incidente en el material semiconductor del diodo.

La principal desventaja de la fibra óptica reside en la complejidad de la conectrización. Realizar un empalme entre dos fibras o hacer una derivación de un bus principal es realmente complicado y delicado. Para llevar a cabo este proceso con suficiente precisión (nótese el pequeño diámetro de los núcleos de las fibras ópticas) se emplea un sistema de ajuste en tres ejes utilizando bucles de realimentación que maximizan la cantidad de luz transmitida en la unión.

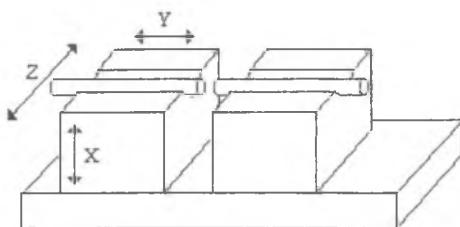


Figura 5.10 Sistema de unión de fibras ópticas en tres ejes.

A pesar de la precisión empleada en la conexión, ésta no suele ser perfecta por lo que las señales sufren atenuaciones al atravesar las uniones.

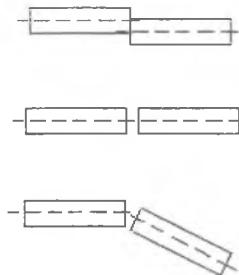


Figura 5.11 Imperfecciones en las uniones de fibras ópticas.

5.2.6 Redes de fibra óptica.

Existen diferentes topologías de redes en las que se emplea la fibra óptica como medio de transmisión. La primera de ellas es la **red en anillo FDDI** (*fiber distributed data interface*) y una de las más populares en el entorno de los pocos cientos de kilómetros (campus universitarios) o el uso con altas velocidades de transmisión en empresas privadas. Presenta dos anillos, cada uno a 100 Mbps, unidireccionales y contrarrotatorios. Uno de ellos, el **primario**, es el que se emplea normalmente en la transmisión de datos mientras que el **secundario** se emplea para subsanar posibles interrupciones del anillo primario o aumentar la tasa de velocidad.

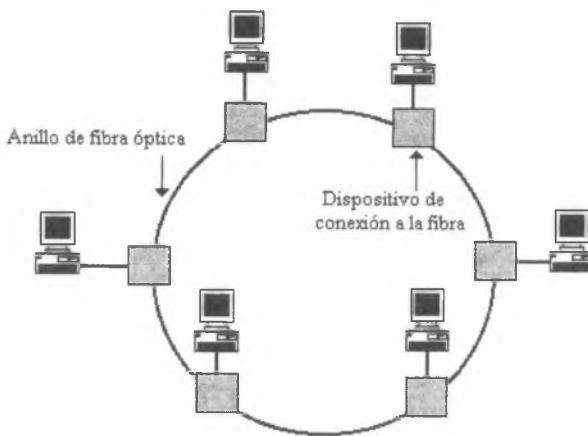


Figura 5.12 Red de anillo FDDI.

Las estaciones se conectan a los anillos de fibra capturando la señal luminosa y regenerándola empleando un detector y un emisor. En este proceso de regeneración de la señal la estación captura la información digital. En este tipo de redes la longitud de onda empleada es de 850 nm, permitiendo distancias entre estaciones de hasta 2 Km. y longitudes de anillo de hasta 200 Km.

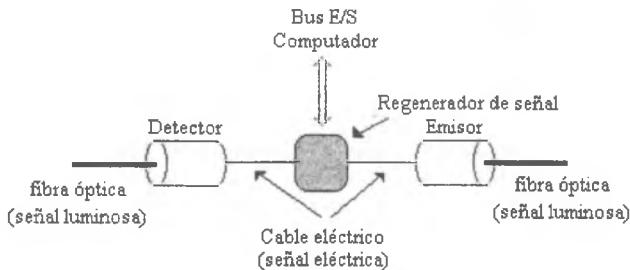


Figura 5.13 Conexión de una estación al anillo FDDI.

Otra topología menos frecuente es la **estrella**. En esta configuración se dispone de un bus de fibra óptica denominado estrella pasiva que recibe la señal luminosa de todas las estaciones y la envía de nuevo a todas las

estaciones. Éstas disponen de un dispositivo emisor y otro receptor de señales luminosas para enviar y recibir información en la red de estrella.

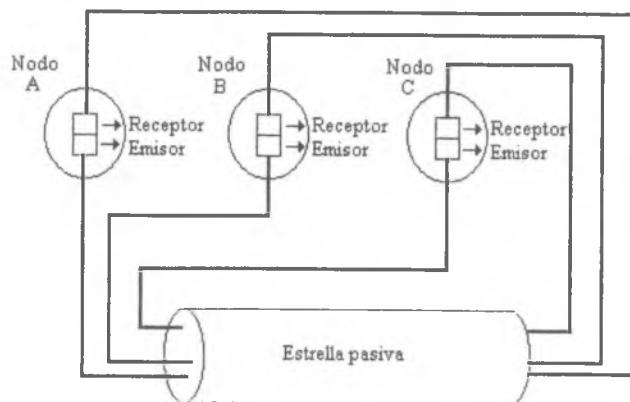


Figura 5.14 Red en estrella de fibra óptica.

Por último, existe un tipo de interconexión con fibra óptica que se emplea frecuentemente. Cuando se desean interconectar dos redes LAN a una gran distancia se emplea una conexión punto a punto entre dos estaciones de las LAN empleando fibra óptica. El efecto que tiene es el de extensor del bus de las dos LAN.

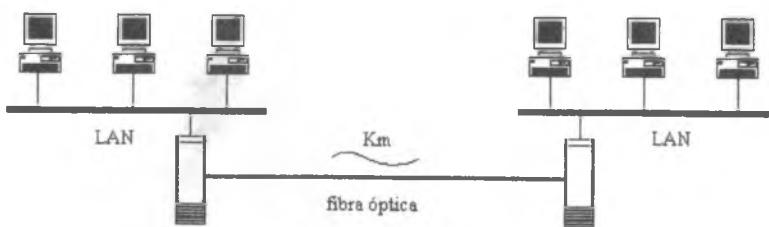


Figura 5.15 Extensión de bus con fibra óptica.

5.3 ONDAS ELECTROMAGNÉTICAS

5.3.1 Introducción. Espectro electromagnético.

Una **onda electromagnética** consiste en la propagación conjunta de un campo eléctrico y un campo magnético perpendiculares entre sí y a la dirección de propagación. Si a una onda electromagnética se le modifican alguno de sus parámetros físicos para codificar información digital es posible emplearla para la transmisión de datos por un medio que puede ser el aire o el vacío.

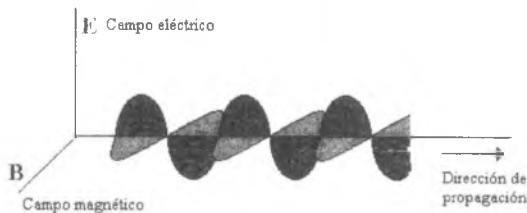


Figura 5.16 Propagación de una onda electromagnética.

Una onda electromagnética se caracteriza por su **frecuencia** f (número de ciclos de onda por unidad de tiempo), que a su vez está relacionada con la **longitud de la onda** λ (longitud en metros de un ciclo de onda) y la **velocidad de transmisión** c de la luz en el vacío.

$$f = \frac{c}{\lambda}$$

Cualquier onda electromagnética se propaga en el vacío a una velocidad constante c de $3 \cdot 10^8 \text{ m/s}$. Cuando una onda electromagnética atraviesa un medio que no es el vacío su velocidad de propagación disminuye, y dado que la frecuencia de la onda (su energía) debe mantenerse constante se produce una variación en la longitud de onda.

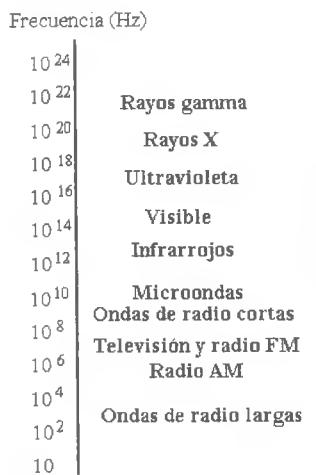


Figura 5.17 Espectro de radiación electromagnética.

Las ondas electromagnéticas engloban todo el espectro de radiación de energía en el universo. Desde las ondas de radio de muy baja frecuencia hasta la nociva radiación gamma. Dentro del espectro electromagnético, si exceptuamos las ondas de "luz" (del infrarrojo cercano al ultravioleta) empleadas en las fibras ópticas, la radiación más empleada para la transmisión de información son las clásicas señales de radio con frecuencias desde 30 KHz a los 30 GHz. Por encima de los 30 GHz se encuentran la radiación infrarroja, entre los 30 y 3000 GHz, que se emplea para transmisión de datos a corta distancia (conexión de ordenadores portátiles) pues requieren contacto visual.

Banda	Frecuencia	Aplicaciones
VLF	< 30 KHz	Audio
LF	30 KHz - 300 KHz	Marítima
MF	300 KHz - 3 MHz	Radio AM
HF	3 MHz - 30 MHz	
VHF	30 MHz - 300 MHz	Radio FM, TV, Radar
UHF	300 MHz - 3 GHz	Radar, TV, Microondas
SHF	3 GHz - 30 GHz	Satélite, Microondas, Radar
EHF	30 GHz - 300 GHz	Radar, Infrarrojo
SEHF	300 GHz - 3000 GHz	Infrarrojo

Tabla 5.2 Espectro de frecuencia empleado en radiocomunicación.

Las ondas electromagnéticas (ya se vio en las ondas luminosas) sufren

diferentes fenómenos en su propagación que han de ser tenidos en cuenta para la transmisión de datos de forma correcta.

- a) **Reflexión.** Este fenómeno se produce cuando la onda que incide con un cierto ángulo sobre una superficie se refleja siguiendo una trayectoria con el mismo ángulo que el incidente.
- b) **Refracción.** Este fenómeno se produce cuando la onda atraviesa un medio con diferente índice de refracción al del medio en el que se propaga, y modifica su trayectoria.
- c) **Difracción.** Este fenómeno se produce cuando una onda electromagnética incide sobre un obstáculo cuyas dimensiones son del orden de la longitud de onda. En esta situación la onda realiza un cambio brusco en su trayectoria distorsionándose, por lo que este fenómeno será de especial importancia en ondas de radio con longitudes de onda de las decenas de metros, ya que obstáculos como los edificios pueden imposibilitar la comunicación.
- d) **Atenuación.** En el proceso de propagación las ondas electromagnéticas sufren atenuaciones

5.3.2 Modelos de propagación.

Los fenómenos mencionados anteriormente condicionarán la distancia alcanzada y la trayectoria descrita por las ondas electromagnéticas al propagarse. Además, estos fenómenos son dependientes de la frecuencia de la onda, por lo que dependiendo de la aplicación que se desee realizar se emplearán ondas en una determinada **banda de transmisión** (rango de frecuencias).

- a) **Ondas de Superficie (VLF - MF 30 KHz - 3 M Hz).** En este tipo de ondas el frente se va curvando al propagarse hasta que se cortocircuita con tierra, por lo que su alcance es limitado. No pueden atravesar obstáculos y precisan de antenas grandes, con tamaños del orden de la longitud de onda λ .

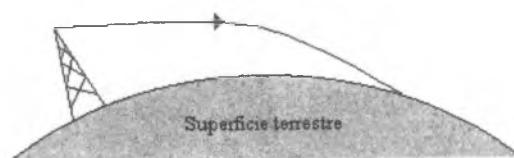


Figura 5.18 Propagación de ondas de superficie.

- b) **Ondas del cielo (HF 3MHz - 30 MHz).** Se emplean para la comunicación de puntos muy distantes entre sí en la superficie terrestre. El frente de ondas se refleja en la ionosfera (capa de la atmósfera a unos 40 Km. de la superficie terrestre) y vuelve a la tierra reflejándose nuevamente hacia la atmósfera. Estas reflexiones sucesivas de las ondas permiten alcanzar puntos muy alejados en el globo terráqueo. Además, es muy importante enviar la onda hacia la ionosfera con un determinado ángulo de incidencia, pues si no está dentro de unos determinados límites no se refleja y escapa hacia el espacio.

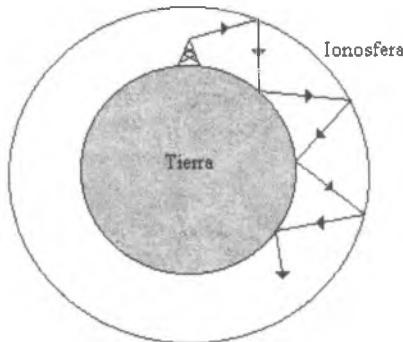


Figura 5.19 Propagación de ondas del cielo.

- c) **Ondas del Espacio (VHF - UHF 30 MHz - 3 GHz).** Estas ondas se propagan en línea recta, por lo que precisan que la **distancia al horizonte** de las antenas emisora y receptora sea tal que coincidan en un punto común. Su alta frecuencia produce que sufran fenómenos de reflexión y refracción con los obstáculos. Un ejemplo de aplicación es la transmisión de TV, y la propagación en línea recta de estas ondas es la explicación de la necesidad de repetidores en lugares elevados.

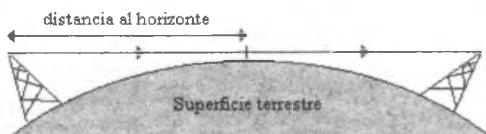


Figura 5.20 Propagación de ondas del espacio.

- d) **Ondas Troposféricas (SHF 3 GHz - 30 GHz).** Estas ondas se propagan sin problemas hacia el espacio exterior atravesando todas las capas de la atmósfera. Por ello se emplean mayoritariamente en comunicaciones vía satélite.

5.3.3 Redes inalámbricas.

Las ondas electromagnéticas tienen su aplicación en la transmisión de datos entre computadores cuando se realizan configuraciones de acceso remoto a redes LAN. Es frecuente que computadores portátiles que no disponen de emplazamiento fijo precisen del acceso a recursos en una red LAN o que el realizar el cableado entre diferentes equipos sea muy complicado o costoso (pequeñas estaciones automáticas de campo). En estas situaciones, donde realizar un cableado de una red no es factible, el empleo de una red de comunicación donde el medio físico de transmisión es el aire se convierte en la única solución. Para paliar esta situación se desarrollaron un tipo de redes, denominadas inalámbricas, que permitían la conexión de equipos portátiles a redes LAN cableadas o la conexión de equipos portátiles entre sí. Estas dos situaciones dan lugar a dos tipos de topologías de redes inalámbricas.

Una de ellas consiste en redes LAN a las que se conecta una **unidad de acceso portátil (PAU - portable access unit)**, que se encarga de gestionar el acceso vía radio de computadores portátiles a la LAN. El alcance de las PAU suele estar entre 50 y 100 metros. Esta topología se denomina **LAN inalámbrica de infraestructura**. El otro tipo de topología consiste en la conexión de diferentes equipos portátiles entre sí empleando enlaces de radio, conformando una LAN propia. Este tipo de redes se denominan **LAN inalámbricas ad hoc**.

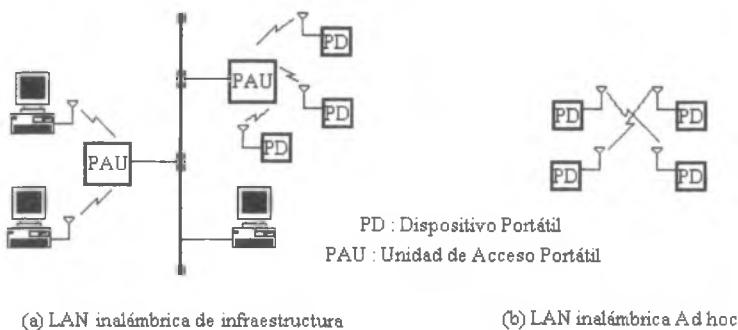


Figura 5.21 Topologías de redes inalámbricas.

5.3.4 Transmisión vía satélite.

El 4 de Octubre de 1957 la extinta Unión Soviética inicia la carrera espacial colocando en órbita el primer satélite artificial de la tierra, el **Sputnik**. Desde entonces el número de ingenios espaciales en la órbita terrestre se ha multiplicado, y una de las aplicaciones más frecuente es la transmisión de datos. Los satélites pueden cubrir grandes zonas geográficas y, empleando un número reducido en determinadas órbitas, es posible tener una cobertura mundial.

Sin embargo, cualquier órbita no es útil para la transmisión de datos. Generalmente, cuando se ofrece un servicio de comunicación de información (ya sea voz, vídeo o datos) se precisa que el servicio sea ininterrumpido las 24 horas del día. Para ello el satélite debe permanecer siempre cubriendo una zona geográfica determinada, lo que se consigue colocando el satélite en una órbita denominada **geoestacionaria**. Una órbita geoestacionaria es aquella en la que el satélite tiene un periodo de traslación alrededor de la tierra igual al tiempo de rotación de la tierra. Las órbitas que cumplen esta condición son aquellas órbitas circulares a una distancia de 35000 Km de la superficie terrestre.

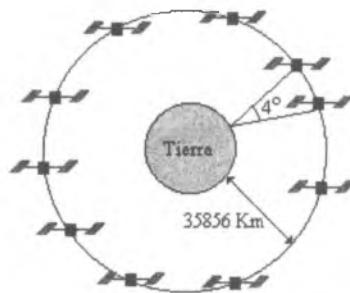


Figura 5.22 Configuración de satélites en órbita geoestacionaria.

Inicialmente, dado que las órbitas geoestacionarias están limitadas, los satélites se disponían a distancias entre sí de 4°, permitiendo hasta 90 satélites en cada una de estas órbitas. En la actualidad, además del problema de órbitas limitadas, hay que tener en cuenta la basura espacial que orbita la tierra. Cuando un satélite deja de ser útil, siempre dispone de una reserva de combustible para colocar el satélite en una órbita denominada de **cementerio**, para no ocupar órbitas útiles. Otra de las opciones es hacerlos caer en la tierra en lugares no habitados, generalmente los océanos. A pesar de ello, siempre existen en órbita pequeños fragmentos de los satélites que pueden producir graves desperfectos en otros satélites.

En comunicaciones vía satélite es muy importante elegir el rango espectral de las señales electromagnéticas portadoras empleadas en la comunicación tierra - satélite. Si la frecuencia de la portadora es demasiado baja el efecto de la refracción en la atmósfera es acusado, y la onda puede desviarse y no llegar a la estación en tierra. Por el contrario, si la frecuencia es demasiado alta, la atenuación de la onda al atravesar la atmósfera se incrementa produciéndose pérdidas y errores en la señal. El rango de frecuencias empleado está entre **4 - 6 GHz**, que en la actualidad está saturado, y se ha ampliado al uso de la banda **12 - 14 GHz**.

Los sistemas de comunicación en los satélites se integran en dispositivos individuales denominados **transponders**. Estos elementos consisten en un sistema emisor/receptor que emplea un ancho de banda y rango de frecuencias determinado, de forma que cada satélite dispone de varios transponders que no se solapan en el espectro frecuencial. Además en cada transponder la frecuencia para la transmisión y la recepción es distinta,

evitando interferencias y permitiendo una comunicación tierra - satélite **full duplex**.

Generalmente un satélite trabaja con una ancho de banda de **500 MHz** empleando una señal portadora con una frecuencia entre 4 - 6 GHz o 12 - 14 GHz. El ancho de banda de 500 MHz se distribuye entre los 12 - 13 transponders que suele tener un satélite de telecomunicaciones, de forma que cada transponder tiene un ancho de banda de aproximadamente **36 MHz**. Con este ancho de banda, un transponder puede transmitir a una velocidad de hasta **50 Mbps** y ésta tasa de velocidad suele dividirse en canales de 64 Kbps.

Para aprovechar al máximo el ancho de banda proporcionado por el satélite y que se optimice el uso del mismo por varias estaciones terrestres, se emplea multiplexación. Esta multiplexación puede ser de varios tipos:

- a) **Multiplexación en tiempo en recepción y emisión.** En este tipo de multiplexación, cada estación terrestre dispone de una celda de tiempo para transmitir información al satélite. En la recepción, el satélite emplea celdas temporales en las que transmite bloques de datos que son recibidos por todas las estaciones terrestres. En cada uno de los bloques el satélite incorpora información para saber a qué estación terrestre va dirigida la información.

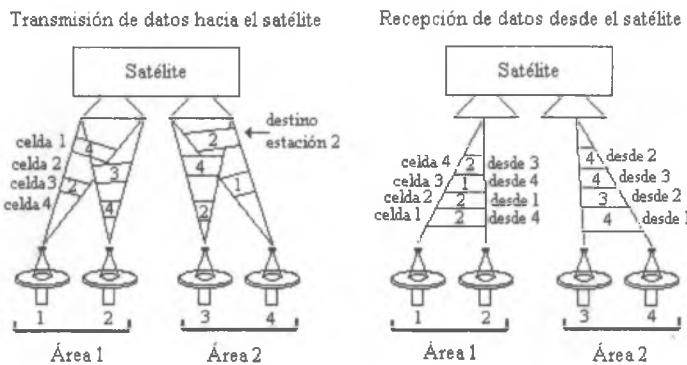


Figura 5.23 Transmisión vía satélite empleando multiplexación temporal.

- b) **Multiplexación por división de frecuencias.** En esta situación, todas las estaciones terrestres transmiten y reciben información simultáneamente, pero empleando rangos de frecuencia distintos para evitar solapamientos.

- c) **Multiplexión en frecuencia para recepción y multiplexado en tiempo para emisión.** Como su nombre indica, este tipo de multiplexación es combinada. En la recepción desde el satélite cada estación terrestre recibe información en un rango de frecuencias determinado y en la transmisión al satélite las estaciones disponen de una celda temporal. De esta forma se consigue una mejor separación entre canales de diferentes estaciones que emplean el satélite.

Uno de los aspectos a tener muy en cuenta en la transmisión vía satélite son los retardos que se producen en la transmisión satélite - estación terrestre. Si se considera la velocidad de propagación de las ondas electromagnéticas en el vacío (en la atmósfera se reduce en un determinado factor, pero despreciable si se tiene en cuenta el reducido grosor de la misma frente a la distancia a la que se encuentran los satélites), que es de 300000 Km/s, y la altura de la superficie de un satélite geoestacionario, que es de unos 35000 Kms, se obtiene un tiempo de propagación de la onda de unos **240 ms**. Este retardo contrasta con los retardos entre enlaces terrestres de cable, que son del orden de **3.5 - 4 μ s/Km**. Debido al elevado retardo que se produce en la transmisión vía satélite, es necesario evaluar la cantidad de información a transmitir, pues muchas veces la transmisión de una cantidad pequeña de datos es más rápida a través de una línea de baja velocidad que a través de un satélite.

Como ejemplo, calcúlese la cantidad de información mínima a partir de la cual la transmisión a través de un enlace vía satélite a 5 Mbps es más rápida que a través de una línea telefónica a 9600 bps. Se supone que el retardo del satélite es de 240 ms y el de la línea telefónica despreciable.

Sea X la cantidad en bits a transmitir, el enlace a través de la línea telefónica empleará $\frac{X}{9600}$ segundos y el enlace a través de satélite $\frac{X}{5 \cdot 10^6} + 0.240$ segundos. Si queremos obtener la cantidad de bits a partir de la cual la transmisión vía satélite es más rápida que la telefónica, ha de resolverse la siguiente inecuación

$$\frac{X}{9600} > \frac{X}{5 \cdot 10^6} + 0.240$$

Despejando la cantidad de bits X, se obtiene la condición $X > 2308 \text{ bits}$. Es decir, si la cantidad de información a transmitir es superior a 2.3 Kbits la transmisión vía satélite es más rápida que la terrestre.

5.4 TÉCNICAS DE COMPARTICIÓN DEL MEDIO FÍSICO EN LAN's

Las redes de área local, LAN, se caracterizan por emplearse en entornos geográficos pequeños (edificios, campus, etc.) y suponer un coste económico pequeño. Una de las características que permite un ahorro económico en las LAN es la compartición del medio físico de transmisión por el conjunto de los ETD conectados a la red. Esto determina la existencia de criterios en cuanto a cómo se determina cuando una estación de la red puede transmitir datos sin colisionar (transmitir simultáneamente) con las demás estaciones en la red.

5.4.1 Clasificación.

Dentro de las técnicas de compartición del medio físico es posible realizar una clasificación en tres grupos:

- a) **Selección.** En este tipo de técnica, cada estación en la red es avisada de cuando llega su turno y toma el control de la misma hasta que acaba de transmitir. La asignación de turnos no ocurre siempre con un mismo intervalo de tiempo, sino que es variable.
- b) **Contienda.** En esta técnica, el acceso al canal para la transmisión se realiza en un proceso de contienda o de lucha por el medio en el que las estaciones que desean transmitir compiten. Los protocolos que emplean esta técnica se denominan protocolos de acceso múltiple y presentan colisiones de los paquetes que se envían al medio.
- c) **Reserva.** En esta técnica cada estación realiza una petición de reserva de uso del medio confirmándosele posteriormente que puede emplear el medio para transmitir. En esta situación sólo se producen colisiones en las peticiones y no en la transmisión. La reserva presenta una mejora frente a la contienda y es que presenta menos retardos en las colisiones, pues la cantidad de información que se envía en una reserva es mucho menor que la que se envía en una trama de información.

5.4.2 Técnicas de contienda.

A continuación se analizarán las técnicas de contienda para compartir un medio físico entre varias máquinas, pues éstas son de las más difundidas dentro de las redes de comunicación de difusión. En primer lugar es posible establecer una división en dos tipos fundamentales,

- a) **Transmisión Sorda (ALOHA).** En la que la transmisión de información se realiza sin tener en cuenta la información que existe en el medio.
- b) **Transmisión con escucha (CSMA - Carrier Sense Medium Access).** En la que previamente a la transmisión se escucha el medio para detectar si ya existe una señal de datos en el mismo.

5.4.2.1 ALOHA

En los años 70, Norman Abramson y sus colegas de la Universidad de Hawái inventaron un método para resolver el problema del reparto del canal. Este sistema, denominado ALOHA por tener sus orígenes en las islas Hawái, se diseñó para radiotransmisión vía satélite, aunque es aplicable a cualquier otro medio.

Puede distinguirse entre dos modalidades de ALOHA,

- a) **ALOHA puro.** El esquema de contienda de ALOHA puro es muy simple. Cada máquina que precisa transmitir información al medio realiza el envío sin más. Ello produce que cuando dos máquinas intenten transmitir al medio simultáneamente o la información de una máquina se solapa en el tiempo con la de otra, se produzca una colisión. Una máquina puede detectar que se ha producido una colisión en el medio escuchando el canal después de realizar la transmisión. Si se ha producido una colisión la señal presentará errores y distorsiones y se procederá a la retransmisión de la información tras esperar un tiempo aleatorio. Esta técnica presenta muy poco rendimiento, pues en cuanto el número de estaciones que intentan acceder al medio es suficientemente elevado, la red queda colapsada al producirse colisiones continuamente. Por término medio, el aprovechamiento del medio para la transmisión es de un 20% aproximadamente.
- b) **ALOHA ranurado.** En 1972 Roberts publicó un método para duplicar la capacidad del método ALOHA. Su propuesta consistió en establecer **ranuras temporales**, de forma que una estación sólo podía transmitir al inicio de cada ranura temporal. Esto convierte el ALOHA puro continuo

en uno discreto. Para permitir esta sincronización es preciso algún mecanismo, que podría ser que una estación de la red transmitiera una señal especial al inicio de cada ranura, como si fuera un reloj. La ventaja es que ahora la probabilidad de que se produzca una colisión en un determinado instante de tiempo es menor que la de que se produzca en cualquier instante. Esto redundaría en una mejora del aprovechamiento del medio, que puede alcanzar hasta un 37%.

5.4.2.2 Acceso al medio con detección de portadora - CSMA.

Esta modalidad de contienda en el acceso a un medio físico compartido se caracteriza por comprobar, antes de realizar la transmisión, de que no existe ninguna señal transmiéndose por el canal. Con esta filosofía sólo se detectarán colisiones cuando dos estaciones detecten simultáneamente que el medio está libre y traten de transmitir a la vez. Existen varias versiones del protocolo CSMA, que fueron analizadas por Kleinrock y Tobagi en la década de los 70 (1975).

- a) **CSMA Persistente-1.** Cuando una estación tiene datos para transmitir escucha el canal para determinar si existe alguna máquina transmitiendo. Si el canal está ocupado, la estación espera hasta que se desocupa. Una vez que el canal está libre realiza la transmisión del marco y en caso de producirse una colisión la estación espera un tiempo aleatorio, pasando de nuevo a escuchar el medio para realizar la retransmisión de la información.

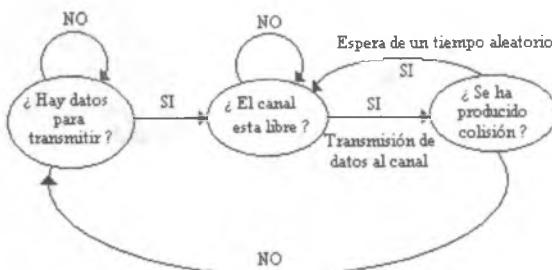


Figura 5.24 Algoritmo del protocolo CSMA Persistente-1.

- b) **CSMA No Persistente.** Esta modalidad introduce una variante con la anterior en la espera de que el medio esté libre. Supone que pueden existir otras estaciones esperando a que el medio quede libre para transmitir, por lo que si el medio está ocupado espera un tiempo aleatorio y vuelve a comprobar que esté libre para transmitir.



Figura 5.25 Algoritmo del protocolo CSMA No Persistente.

- c) **CSMA Persistente- p .** Esta modalidad se aplica a canales con ranuras temporales. Cuando una estación está lista para transmitir datos y se encuentra al inicio de una ranura temporal escucha el canal. Si el canal está libre, transmite los datos con una probabilidad p o espera a la siguiente ranura para transmitir con probabilidad $q = 1 - p$. Si la siguiente ranura se encuentra libre, vuelve a transmitir o esperar con probabilidades p y q nuevamente. Este proceso se repite hasta que los datos han sido transmitidos o se detecta que otra estación ha comenzado a transmitir. En este último caso se considera como si se hubiera producido una colisión, por lo que se espera un tiempo aleatorio y se inicia el procedimiento. Si inicialmente la estación detecta que el canal está ocupado, espera hasta la siguiente ranura y aplica el algoritmo anterior.

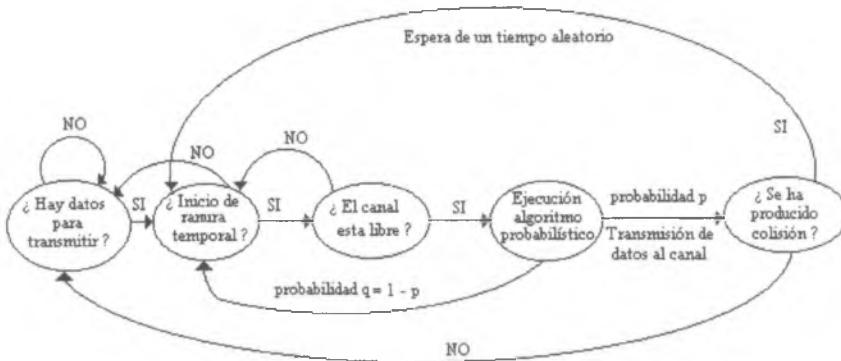


Figura 5.26 Algoritmo del protocolo CSMA p Persistente.

- d) CSMA/CD (*Carrier Sense Medium Access with Collision Detection*).

Esta modalidad presenta la característica de que cuando el canal se encuentra libre para transmitir, las estaciones interrumpen de forma inmediata la transmisión de un marco o trama de datos al detectar una colisión. La colisión puede detectarse observando la potencia de la señal recibida y comparándola con la señal transmitida, por lo que el hardware de acceso al medio debe permitir escuchar el medio al mismo tiempo que transmite. Una vez detectada la colisión y abortada la transmisión, la estación espera un tiempo aleatorio y vuelve a transmitir el marco, suponiendo que ninguna otra estación ha comenzado a transmitir en ese lapso. Esta técnica comprenderá por tanto periodos en que se produce contención (colisiones de dos o más estaciones que tratan de transmitir simultáneamente), periodos de transmisión correcta de datos y periodos de inactividad donde ninguna estación tiene datos para transmitir.

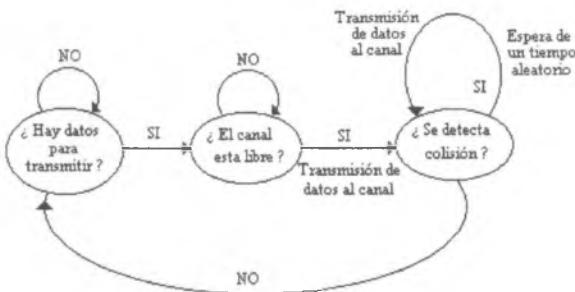


Figura 5.27 Algoritmo del protocolo CSMA/CD.

CAPÍTULO 6. FUNCIONES DEL NIVEL DE ENLACE

6.1 INTRODUCCIÓN

La capa de enlace es el segundo nivel dentro del modelo de arquitectura OSI y su objetivo es, dado un canal físico de comunicación con una cierta tasa de error, **permitir establecer un enlace lógico libre de errores entre entidades de la capa superior**, la capa de red.

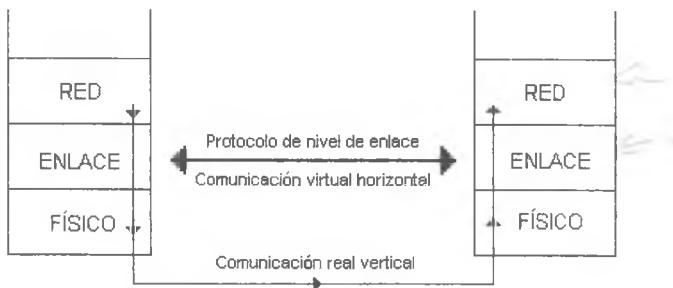


Figura 6.1 Comunicación vertical y horizontal en la capa de enlace.

Los paquetes de bits de información (**tramas**) procedentes del nivel de red y que deben ser transmitidos a la entidad par del otro extremo de la comunicación, son fragmentados en grupos de bits a los que se añaden bits de información de control. Estas tramas se transmitirán de forma secuencial, pudiendo numerar cada una de las tramas para reagrupar la información en el otro extremo.

La entidad de nivel de enlace emisora podrá conocer si las tramas de datos enviados han llegado correctamente a la entidad par del otro extremo manejando el reconocimiento de la información. Cada vez que la entidad receptora recibe una trama de datos procedente de la entidad emisora, le

envía una pequeña trama de información de control (trama de asentimiento) indicando que la trama de datos se recibió correctamente.

Dado que existe una tasa de error en el medio, se producirán pérdidas de tramas de datos y tramas de asentimiento o errores en el contenido de los mismos. Para subsanar este tipo de errores el nivel de enlace realiza el reenvío de las tramas de datos perdidos o erróneos.

Por otra parte, en el intercambio de información entre dos estaciones a través de un medio físico es preciso controlar el flujo de la información. Frecuentemente la comunicación se establecerá entre elementos heterogéneos que serán capaces de procesar la información procedente de la red a diferente velocidad. Para evitar que emisores de información con elevada tasa de velocidad de transmisión de datos saturen a receptores que operan a una tasa inferior, es preciso establecer unos mecanismos de sincronización y control del flujo de datos.

6.2 SERVICIOS Y FUNCIONES DEL NIVEL DE ENLACE

El nivel de enlace proporciona al nivel superior de red diferentes servicios según la calidad que se pretende conseguir en la transmisión de una trama de datos del nivel de red a través de un medio de transmisión.

6.2.1 Servicio sin conexión y sin reconocimiento

En este servicio la estación origen envía tramas de información independientes a la máquina de destino sin pedir que se confirme que han llegado o no. No se establecen conexiones entre la estación origen y destino, pues al enviar tramas de datos independientes se consigue una transmisión más rápida. Este servicio se empleará en líneas con una tasa de errores baja en el medio físico o en aquellas aplicaciones en que los retardos son más perjudiciales que los errores.

Si el nivel de enlace proporciona este servicio, los niveles superiores deben contemplar el control de errores que se produce en la transmisión a nivel de enlace. La gestión del control de errores por capas superiores permite un nivel de enlace rápido, que no precisa de esperar confirmaciones de los datos enviados.

6.2.2 Servicio sin conexión y con reconocimiento

En este servicio, por cada trama de datos independiente que es recibida por una estación receptora, ésta envía una trama de asentimiento **confirmando la recepción de la trama de datos al emisor**. Se emplea, por tanto, **cuando el nivel físico presenta una tasa de error no despreciable** y es preciso una transmisión fiable de la información. Un ejemplo de aplicación se tiene en las redes de comunicaciones inalámbricas, donde los errores debido a interferencias electromagnéticas en las señales son frecuentes.

6.2.3 Servicio con conexión y con reconocimiento

El servicio conectado se caracteriza por la presencia de las primitivas de servicio de establecimiento de conexión, liberación de conexión y envío de las tramas de datos. En el establecimiento de la conexión se reservan los recursos asociados al servicio, como son *buffers*, variables, etc. A continuación se envían cada una de **las tramas de datos numeradas, que serán confirmadas** por el receptor y, **en caso de** que alguna sufra **errores** en la transmisión, se realizará el **reenvío** de la misma. Por último, una vez finalizado el envío de las tramas de datos, se procede con la liberación de la conexión en la que se liberarán los recursos reservados a la misma.

La diferencia entre el servicio conectado y el desconectado con reconocimiento está en el grado de fiabilidad en la transmisión de tramas de datos. **El servicio conectado es más fiable**, pues tiene en cuenta que una determinada trama no sea enviada más de una vez en caso de que sea la trama de confirmación la que no se ha recibido de forma correcta por el emisor, que el receptor no se encuentre disponible para recibir información, etc.

6.2.4 Funciones del nivel de enlace

Para proporcionar el conjunto de servicios mencionados anteriormente, el nivel de enlace tiene un conjunto de funciones asignadas. Estas funciones son principalmente:

- a) **Delimitación de tramas.** Especifica como identificar el inicio y fin de una trama de datos.
- b) **Direccionamiento.** Permite identificar origen y destino en el envío de una trama.
- c) **Control de errores.** Asegura una transmisión de tramas sin errores, producidos por el ruido y atenuaciones del medio físico.

- d) **Control del flujo.** Proporciona un control del flujo de tramas entre emisor y receptor para evitar saturaciones en receptores lentos.

6.3 DELIMITACIÓN DE TRAMAS

El nivel de enlace define un formato de trama de datos con una cabecera inicial y una cola en la parte final, y entre ellas se encuentran los datos del nivel superior. La cabecera inicial indica el inicio de la trama, donde se incorpora información de control como es la numeración de la trama, destino de la trama, etc. A continuación se incorporan los datos procedentes del nivel superior: la trama de datos del nivel de red. En ocasiones la trama de datos del nivel de red es demasiado grande para incorporarla en una única trama del nivel de enlace, por lo que se produce la fragmentación en la forma en que se describe en el capítulo 1 de este libro. Al final de la trama de nivel de enlace aparece la cola, donde se incorpora información para detectar errores en los datos de la trama e indicar el final de la misma.



Figura 6.2 Formato de una trama de nivel de enlace.

La introducción de una cabecera y una cola en la trama de nivel de enlace permite **delimitarla**, y en base a cómo sea la delimitación de la trama se establecen distintas categorías de protocolos de nivel de enlace.

6.3.1 Delimitación temporal.

En un principio se diseñó un sistema de delimitación de tramas basado en la duración temporal de la trama sin emplear cabeceras. Se estimaba el tiempo que el emisor tardaba en transmitir una trama y el receptor leería datos del medio físico durante ese mismo tiempo. Esta estrategia se empleó con muy poco éxito y se descartó rápidamente, pues los retardos en la transmisión de señales por el medio introducían muchos errores.

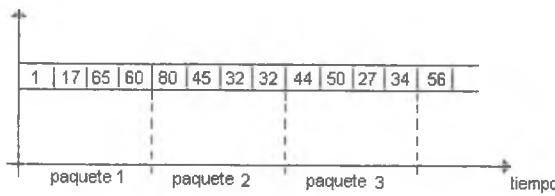


Figura 6.3 Delimitación temporal de tramas.

6.3.2 Delimitación por numeración de caracteres.

Una primera estrategia de delimitación de tramas con resultados satisfactorios se basa en indicar el número de caracteres en la trama. En las primeras comunicaciones por computador donde se transmitían bytes de datos, se introducía al principio de la cabecera uno o más bytes donde se indicaba el número de caracteres que contenía la trama. De esta forma se podían distinguir las tramas de forma unívoca, pero es un método muy sensible al ruido. Si se produce un error en el byte que indica la longitud de la cabecera se pierde la sincronización emisor - receptor y las tramas no se reconocen de forma correcta.

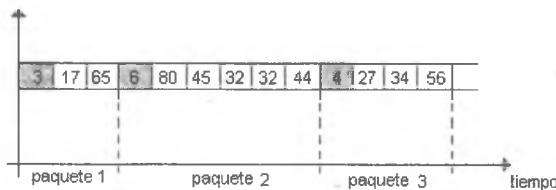


Figura 6.4 Delimitación de tramas por numeración de caracteres.

6.3.3 Delimitación por secuencias de caracteres especiales.

Esta estrategia emplea un conjunto de caracteres ASCII especiales para delimitar la trama de nivel de enlace. Estos caracteres especiales tienen valores de código ASCII que están reservados y no se emplean en la transmisión de datos alfabéticos. Algunos de estos caracteres son: DLE, STX, ETX, ACK, NACK, SYN, etc. El carácter DLE se denomina secuencia de escape y siempre se acompaña a los demás caracteres reservados si han de ser interpretados. La delimitación de un paquete se

realiza empleando la secuencia DLE STX al inicio del paquete y DLE ETX al final del mismo. Este tipo de enmarcado tiene un funcionamiento adecuado cuando los datos transmitidos en el paquete de nivel de enlace son caracteres de texto. De ahí que a los protocolos que emplean esta delimitación se les denomine *protocolos orientados a carácter*. Sin embargo, cuando los datos que se incorporan en el paquete de nivel de enlace son de tipo binario pueden aparecer problemas. Es posible que en los bits de datos se produzcan combinaciones que se correspondan con los códigos ASCII de los bits reservados para el enmarcado y se interprete de forma errónea el fin de un paquete o el inicio de uno nuevo. Para ello, cuando se transmiten datos binarios empleando un protocolo orientado a carácter, es preciso realizar *un relleno de caracteres* en los datos a transmitir. El proceso consiste en incorporar la secuencia de bits correspondiente al carácter DLE delante de las secuencias de bits del carácter DLE que aparezcan en los datos. El receptor por su parte interpretará cada carácter DLE que aparezca en los datos en función del siguiente conjunto de bits asociado a un carácter. Si el siguiente conjunto de bits corresponde a un carácter DLE el anterior se interpreta como de relleno y el leído como dato. De esta forma sólo se interpretan las secuencias DLE + carácter reservado adecuadas. En la siguiente figura se ilustra el proceso anterior.

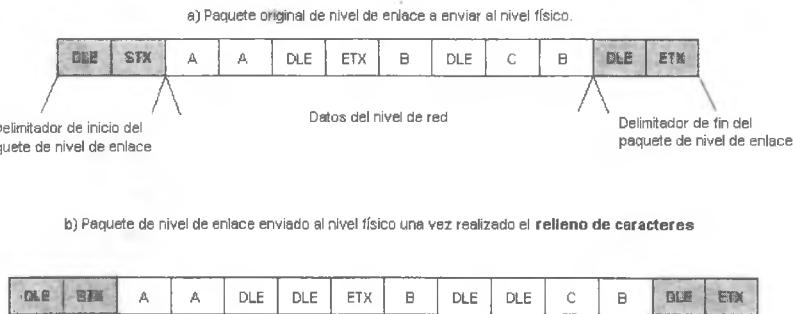


Figura 6.5 Delimitación por secuencias de caracteres especiales.

6.3.4 Delimitación por secuencias de bits especiales.

La delimitación por secuencia de bits especiales soluciona los problemas que presenta la delimitación por secuencias de caracteres cuando los datos del paquete son orientados a bit. Los protocolos que emplean este tipo de enmarcado se denominan por tanto *protocolos orientados a bit*. La delimitación del paquete se realiza empleando una secuencia de bits que es

única y que no se emplea en los datos del paquete. Para impedir que una secuencia determinada pueda aparecer en los datos es necesario realizar también el *relleno* de los bits de datos, pero esta vez se empleará un solo bit en el relleno, por lo que se reduce la información redundante respecto del esquema anterior y se aprovecha más el ancho de banda del medio. En la siguiente figura se ilustra el proceso de *relleno de bits* cuando la secuencia de bits para el inicio y el fin del marco es **01111110**. En el emisor se introduce un bit **0** en los datos cada vez que aparece la secuencia **011111**, mientras que el receptor comprueba el siguiente bit cada vez que recibe la secuencia **011111**. Si ese bit siguiente es **1** interpreta el fin de trama y procesa el paquete, mientras que si es **0** elimina el bit, pues es de relleno, y busca la siguiente secuencia **011111**.

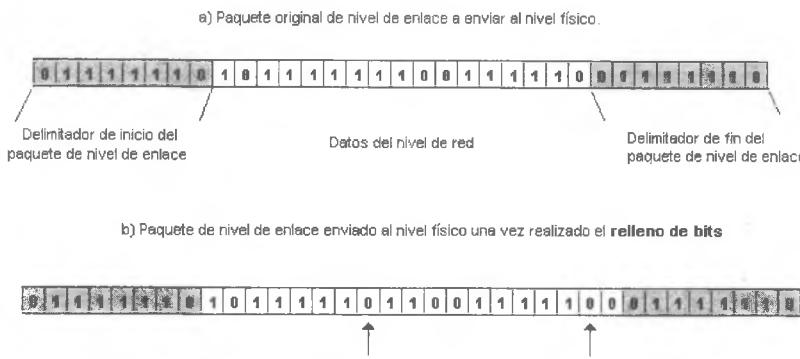


Figura 6.6 Delimitación por secuencias de bits especiales.

6.3.5 Delimitación por violaciones de la codificación de la capa física.

Este tipo de delimitación puede emplearse con varios esquemas de señalización en banda base y banda modulada. Una de ellas es la codificación Manchester, donde cada bit lógico está asociado a un tipo de transición en la señal que se propaga por el medio físico. En concreto, un **1** lógico está asociado a una transición de tipo bajo - alto en la amplitud de la señal durante el tiempo que dura un bit, y un **0** lógico a una transición de tipo alto - bajo. La violación de la codificación consiste en enviar al inicio de la trama una secuencia de la señal que no está asociada a información interpretable. En el caso de la codificación Manchester podrían enviarse varios pulsos de la señal con un ancho igual a la duración temporal de un bit y en los cuales no se produce ninguna transición. De esta forma el receptor no interpreta información pero se encuentra listo para recibir la información que vendrá a continuación codificada de forma correcta.

6.4 DIRECCIONAMIENTO

El direccionamiento de nivel de enlace permite identificar las estaciones que intercambian información en el canal. Existen distintos tipos de direccionamiento atendiendo a la forma en que las estaciones están comunicadas empleando el canal. Entre los diferentes esquemas de direccionamiento destacan:

- a) **Implícito.** Este direccionamiento no precisa de la especificación de la estación origen y destino de los datos transmitidos. Se emplea cuando se establecen conexiones punto a punto entre pares de estaciones.
- b) **Explícito.** En este esquema cada estación conectada al canal de comunicación tiene una dirección única. En la transmisión de datos, cada paquete incorpora en su cabecera la dirección de la estación origen del paquete y la dirección de la estación de destino. Se emplea frecuentemente en redes de difusión y redes punto a punto donde existen enlaces multipunto (una estación conectada a varias empleando líneas punto a punto).
- c) **Preselección.** El direccionamiento por preselección emplea un contador que selecciona secuencialmente cada una de las estaciones de destino accesibles desde una estación. De esta forma una estación no puede transmitir datos a un destino cuando lo precise, sino cuando le sea concedido el turno a ese destino.
- d) **Master único.** Este direccionamiento está basado en la comunicación unidireccional maestro - esclavo, donde existe una estación principal (maestro) que puede transmitir datos a un conjunto de estaciones esclavo, las cuales sólo pueden comunicar a la estación maestro si recibieron los datos o no.
- e) **Master múltiple.** Es una variante del esquema de direccionamiento anterior, en el que las estaciones esclavo pueden además transmitir datos a la estación maestro.

6.5 DETECCIÓN Y CORRECCIÓN DE ERRORES

El nivel de enlace debe proporcionar mecanismos para verificar la integridad de los paquetes transmitidos. Estos paquetes, que contienen la información procedente del nivel de red, al ser transmitidos como señales

por un medio físico son sensibles al ruido presente en el mismo. La superposición de señales aleatorias presentes en el medio sobre las señales asociadas a un paquete de información digital pueden producir que uno o varios bits del paquete sean interpretados de forma incorrecta. El extremo de la comunicación que recibe este paquete con datos erróneos debe ser capaz de detectar esta situación de error y subsanarla.

La detección de errores en un paquete de datos se consigue analizando un pequeño conjunto de datos que suele añadirse en la cola del paquete de nivel de enlace y que se denomina **secuencia de verificación de trama (FCS)**. Dependiendo del tipo de información que aporte esta secuencia se distingue entre:

- a) **Códigos de detección de error.** En los códigos de detección de error, la FCS sólo incorpora información que permita detectar si el paquete de datos es correcto o posee algún error en 1 o más bits. El receptor, por tanto, tendrá que informar al emisor de que el paquete enviado ha llegado de forma incorrecta y ha de ser reenviado. Estos códigos se emplean cuando el tiempo de reenvío de un paquete es menor que el que se necesitaría para determinar los bits erróneos y modificarlos.
- b) **Códigos de corrección de error.** Estos códigos se caracterizan porque la FCS incorpora información que, además de detectar si el paquete de datos presenta errores, determina el conjunto de bits erróneos. Estas técnicas son más complejas y requieren de un tiempo de cómputo para determinar cuáles son los bits erróneos, por lo que se emplearán cuando el reenvío del paquete de información tenga un retardo elevado. Un ejemplo de aplicación son las transmisiones en satélites de exploración del sistema solar, donde los retardos de propagación de las señales son del orden de las horas.

La medida de los errores que presenta un medio de transmisión se cuantifica con la **tasa de error**. Esta tasa mide el número de errores que aparecen por bit transmitido, por lo que su valor siempre es inferior a la unidad, siendo valores típicos los comprendidos entre 10^{-3} y 10^{-6} .

$$T_{error} = \frac{\text{nº de errores}}{\text{nº de bits transmitidos}}$$

Los factores que afecta al valor de la tasa de error son muy variados y algunos de ellos son:

- a) **Tipo de medio de transmisión.** Las características del medio físico, tales como ruido de fondo, ancho de banda, etc. determinan la calidad de las señales transmitidas.
- b) **Entorno del medio de transmisión.** La presencia, en el caso de transmisión de señales eléctricas, de dispositivos electrónicos que generen ondas electromagnéticas en su funcionamiento introducen señales de ruido adicionales en el medio de transmisión.
- c) **Velocidad de transmisión.** La velocidad empleada en la transmisión influirá también en la calidad de la señal y por tanto en los errores que se puedan producir. Conforme se aproxime más a la velocidad de transmisión máxima del medio físico, la probabilidad de error aumentará.
- d) **Calidad de servicio del medio físico.** Atendiendo a la calidad de servicio del proveedor del medio físico la tasa de error podrá ser mayor o menor. Generalmente, las líneas de transmisión de mayor calidad serán las privadas y las de menor calidad las públicas. En un término medio se encuentran las líneas alquiladas, líneas públicas en las que se asegura una determinada calidad de servicio.
- e) **Horario.** Este factor únicamente debería tenerse en consideración en las líneas públicas, donde la calidad del medio no es fija, sino que varía dependiendo del horario. En determinadas redes privadas de área extendida el horario laboral donde mayor número de usuarios emplean la red presenta mayor tasa de error.

6.5.1 **Métodos de detección de errores**

Se analizarán tres técnicas para la detección de errores en paquetes transmitidos por el nivel de enlace al medio físico. Dos de ellas, la detección de paridad y los códigos de redundancia cíclica, incorporan el campo FCS en la trama de nivel de enlace, mientras que la detección por chequeo de lazo no precisa de información adicional en la trama.

PARIDAD DE FILAS Y COLUMNAS

La detección de errores por paridad precisa de añadir a la trama de datos un bit de paridad cuyo valor se determina para que el número total de bits '1' en la trama sea impar (*paridad impar*) o par (*paridad par*). La

paridad sólo permite detectar errores en un número impar de bits, pues al modificar dos bits (o un número par de bits) cualesquiera de una palabra con bit de paridad, la paridad sigue siendo correcta.

Por tanto, la detección de paridad con un sólo bit no proporciona un rendimiento aceptable cuando transmitimos bloques de datos grandes donde pueden producirse varios errores. Sin embargo, podemos aumentar la capacidad de detectar errores añadiendo más de un bit de paridad al paquete de datos. Este aumento de los bits de paridad se consigue distribuyendo los bits de datos a enviar en una matriz de k filas y n columnas, calculando un bit de paridad por cada fila y otro por cada columna.

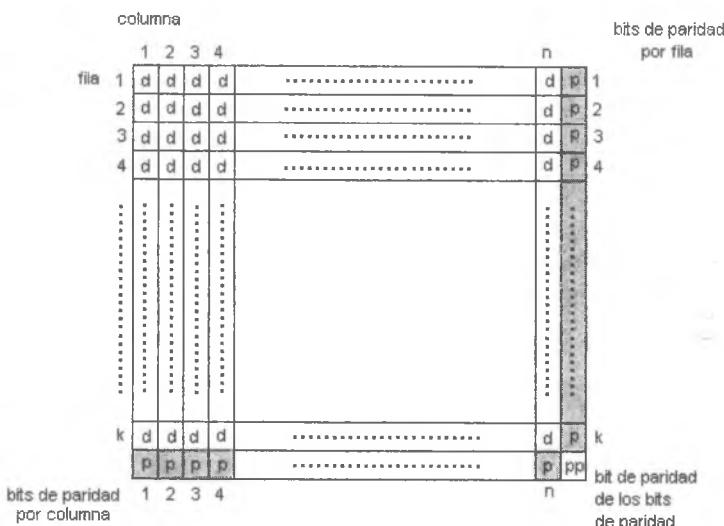


Figura 6.7 Paridad de bloque de datos por filas y columnas.

Por tanto se enviarán $k + n$ bits en el campo FCS de la trama de enlace. Opcionalmente se puede enviar un bit adicional de paridad correspondiente a la paridad de los bits de paridad. Con este esquema siguen detectándose errores en un número impar de bits, pero además puede detectarse errores en un número par de bits en determinadas condiciones. Por ejemplo, si se producen 2 errores en 2 bits situados en cualquier posición de la matriz se detectarán errores de paridad en filas o columnas, por lo que la trama podrá ser rechazada y solicitar su reenvío.

CÓDIGOS DE REDUNDANCIA CÍCLICA (CRC)

Un código de redundancia cíclica consiste en una secuencia de bits que se incorporan en el campo FCS de una trama de nivel de enlace y permiten detectar errores. Este CRC se obtiene considerando los bits de datos como los coeficientes de un polinomio en x y realizando operaciones aritméticas sobre el mismo.

Dada una secuencia de k bits, esta secuencia tiene asociada un polinomio de grado $k-1$. Véase como ejemplo la siguiente secuencia de bits:

$$\begin{aligned} 11101110 \text{ (8 bits)} &\rightarrow 1 \cdot x^7 + 1 \cdot x^6 + 1 \cdot x^5 + 0 \cdot x^4 + 1 \cdot x^3 + 1 \cdot x^2 + 1 \cdot x^1 + 0 \cdot x^0 \\ &= x^7 + x^6 + x^5 + x^3 + x^2 + x \end{aligned}$$

Las operaciones aritméticas de suma y resta que se realizarán con estos polinomios se considerarán en módulo dos, o lo que es lo mismo, emplearemos la operación lógica XOR. Para las multiplicaciones se realizará la multiplicación binaria normal.

La obtención de la FCS pasa por la elección para el emisor y el receptor de un polinomio común denominado **polinomio generador $G(x)$** . Este polinomio tendrá asociado una secuencia de bits que sólo debe cumplir la restricción de que el primer y último bit de la secuencia han de tener valor 1. Existen, por otra parte, un conjunto de polinomios generadores estándares, como son:

$$\begin{aligned} \text{CRC-12} &\rightarrow x^{12} + x^{11} + x^3 + x^1 + 1 \\ \text{CRC-16} &\rightarrow x^{16} + x^{15} + x^2 + 1 \\ \text{CRC-CCITT} &\rightarrow x^{16} + x^{12} + x^5 + 1 \\ \text{CRC-32} &\rightarrow x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1 \end{aligned}$$

La obtención de los bits de la secuencia de verificación de trama se obtiene a partir de las propiedades de la división del polinomio asociado a los datos entre el polinomio generador $G(x)$.

Sea $D(x)$ el polinomio asociado a la secuencia de bits del paquete de nivel de enlace y r el grado del polinomio generador. Se obtendrá el resto de la división de:

$$R(x) = x^r D(x) \% G(x)$$

empleando para ello la aritmética en módulo dos para la resta en la división. Si ahora al dividendo $x^r D(x)$ le restamos el resto de la división $R(x)$, al dividir el resultado nuevamente por el polinomio generador $G(x)$ el resto será cero. Por tanto, si transmitimos la secuencia de bits asociada al polinomio $T(x) = x^r D(x) - R(x)$, el receptor podrá detectar si se han producido errores en la transmisión dividiendo la secuencia de bits recibida por el polinomio generador. Si el resto es distinto de cero se habrán producido errores en la transmisión.

$$T(x) = x^r D(x) - R(x)$$

$$T(x) \% G(X) = 0 \rightarrow \text{No hay errores}$$

Sin embargo, ello supondrá que los datos transmitidos no son los mismos que se pretenden enviar, y esto se subsana empleando la aritmética en módulo 2, de forma que el polinomio $T(x)$ es la secuencia de bits de datos acompañada de un conjunto de r bits correspondientes a $R(x)$ y que será el valor del campo FCS en el paquete de nivel de enlace.

Para apreciar mejor el funcionamiento de este método véase el siguiente ejemplo.

Sea $D(x)$ el polinomio asociado a la secuencia de bits **1101011011** y el polinomio generador $G(x) = x^4 + x + 1$. $x^r D(x)$ será la secuencia de bits **11010110110000**. Si se realiza la división con el polinomio generador se obtiene:

$$\begin{array}{r} \xrightarrow{\quad 8 \quad} \\ (x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1) \cdot x^4 \Rightarrow \\ \xrightarrow{\quad 16 \quad 12 \quad 10 \quad} \\ x^{12} + x^{10} + x^8 + x^7 + x^5 + x^4 \\ \boxed{11010110110000} \end{array}$$

$$\begin{array}{r}
 \overbrace{11010110110000}^{\oplus} \\
 \underline{10111} \\
 010011 \\
 \oplus 10011 \\
 \hline 000001 \\
 \oplus 00000 \\
 \hline 000010 \\
 \oplus 00000 \\
 \hline 000101 \\
 \oplus 00000 \\
 \hline 010110 \\
 \oplus 10011 \\
 \hline 001010 \\
 \oplus 00000 \\
 \hline 010100 \\
 \oplus 10011 \\
 \hline 001110 \\
 \oplus 00000 \\
 \hline \overbrace{01110}^{R(x)}
 \end{array}
 \quad
 \begin{array}{r}
 10011 \\
 \hline 1100001010
 \end{array}$$

Notar que al realizar el cálculo del bit del cociente, únicamente se comprueba el primer bit del término del dividendo y el primero del divisor. Una vez obtenido el resto $R(x)$, $T(x) = x^r D(x) - R(x)$ se calcula como:

$$\begin{array}{r}
 \oplus 11010110110000 \\
 \underline{00000000001110} \\
 \hline 11010110111110
 \end{array}$$

Como puede apreciarse, la trama $T(x)$ transmitida consiste en la secuencia de bits de los datos $D(x)$ más la secuencia de bits del resto de la división $R(x)$. Se invita al lector a comprobar cómo efectivamente el resto de la división de $T(x)$ entre $G(x)$ es cero.

Esta técnica se implementa en circuitos hardware dedicados, debido a su escasa complejidad y la necesidad de realizar la detección de errores con gran rapidez, pues si se detecta un error hay que realizar el reenvío de la trama de enlace.

El empleo de los códigos de redundancia cíclica utilizando alguno de los polinomios generadores normalizados permite detectar errores en un

número impar de bits y en 2 bits. Además se podrán detectar errores en ráfaga de longitud menor que r , el orden del polinomio generador $G(x)$. Los **errores en ráfaga** son una forma frecuente en que los errores se producen en un medio de transmisión. Debido a que los errores se producen generalmente por la presencia puntual de señales aleatorias de ruido, éste produce errores en varios bits seguidos. Si se produce esta situación y el número de bits errores es k , se dice que el medio presenta **errores en ráfaga de longitud k** .

CHEQUEO DE LAZO

El chequeo de lazo es una técnica de detección de errores que no emplea un campo FCS en la trama de nivel de enlace. Se fundamenta en que cada paquete que es enviado por el emisor, el receptor lo reenvía de nuevo al emisor y éste verifica la integridad de los datos. Para ello compara la trama de datos enviada con la trama de datos recibida y si son distintas interpreta que se ha producido un error en la transmisión, volviendo a reenviarla. La información de control para que el receptor pueda determinar si un paquete enviado es un reenvío por un error o no, se incorpora en el campo de control de la cabecera del nivel de enlace.

Esta técnica se emplea frecuentemente en redes de anillo unidireccionales, pues la topología permite que no se desperdicie el ancho de banda del medio. Si se empleara un medio de difusión, donde los paquetes enviados por el emisor y el receptor comparten el mismo medio, el aprovechamiento del medio sería sólo del 50%.

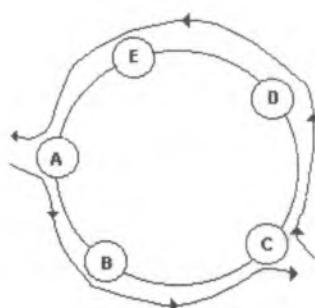


Figura 6.8 Detección de errores empleando el chequeo de lazo. La estación A transmite a C.

6.5.2 Métodos de corrección de errores

Las técnicas de corrección de errores no emplean un campo FCS en la trama de nivel de enlace. La información adicional (información redundante) que permite detectar si hay algún error en la trama de datos, y en caso de que exista determinar de forma unívoca los bits erróneos para corregirlos, está intercalada con los bits de datos. Por tanto, el receptor debe identificar en la información recibida cuáles son los bits de datos y cuáles los redundantes para enviar al nivel de red la información adecuada. Estas técnicas serán por tanto más lentas que las de detección al precisar de un tiempo de computación adicional en la identificación y reorganización de los bits de datos.

Únicamente se estudiará una técnica de corrección de errores que se denomina **Código de Hamming** y para ello es necesario introducir un conjunto de conceptos y una nomenclatura apropiada.

Dado un conjunto de m bits de datos se dispondrá de un conjunto de 2^m **palabras de datos** a transmitir. A cada una de estas palabras se añadirá un conjunto de r bits redundantes que permitirán corregir los errores en los datos. Sea $n=m+r$, se tendrá un conjunto de 2^n palabras a transmitir, cada una de ellas denominada **palabra código**. El esquema de corrección de errores se basa en que no todas las palabras código posibles son correctas. Sólo un subconjunto de ellas son correctas, en concreto 2^m , asociadas a las palabras de datos con sus bits redundantes. El determinar a partir de una palabra código incorrecta, cual es la correcta que se envió originalmente depende de la relación existente entre los bits de datos m y los bits redundantes r . La elección de esta relación debe ser tal que a partir de una palabra código incorrecta sólo exista una correcta asociada. Conseguir esto para un número infinito de errores no es posible, de forma que estas relaciones se definen para poder corregir hasta un número máximo de errores.

El código de Hamming se fundamenta en el concepto de **distancia de Hamming**. La distancia de Hamming entre dos palabras código se define como el **número de bits en que difieren las dos palabras**. Dado un conjunto de palabras código correctas, se define la **distancia de Hamming de un código** como la menor de todas las distancias de Hamming entre las palabras del código.

Supóngase un código de palabras correcta formado por $m=3$ bits de datos y $r=1$ bit redundante, calculado como un bit de paridad par. Las palabras código correctas (**palabras del código**) serán:

0000	1001
0011	1010
0101	1100
0110	1111

Si se determina la distancia de Hamming de este código puede comprobarse que es 2. Supóngase ahora que se produce un error de un bit en cualquiera de las palabras código y en cualquier posición. Puede comprobarse que las palabras código erróneas obtenidas son todas distintas de las palabra código correctas, por lo que este código podrá detectar errores de un bit. Sin embargo, si se producen dos errores en dos bits cualesquiera de cualquier palabra, puede comprobarse como se obtiene una palabra código correcta, por lo que este código no podrá detectar errores en 2 bits. Se generaliza, además, que dado un **código con distancia Hamming d** , este código puede **detectar errores en $d-1$ bits**.

Supóngase ahora el siguiente conjunto de palabras de un código, donde se obvia la relación existente entre los bits redundantes y los bits de datos.

0000000000	1111100000
0000011111	1111111111

Es sencillo comprobar que la distancia de Hamming de ese código es 5 y por tanto podrá detectar errores de hasta 4 bits como mucho (es posible detectar errores en más bits, pero sólo en determinadas condiciones. El valor de $d-1$ asegura que cualesquiera que sean los $d-1$ bits erróneos, siempre se detectará el error). Supóngase que se producen como mucho errores en un bit. Puede comprobarse que dada una palabra cualquiera de código, si se modifica cualquier bit no se obtiene otra palabra del código. Si se supone que como mucho se producen dos errores, nuevamente puede comprobarse que no se obtienen palabras del código. Sin embargo, si se suponen 3 errores como mucho y el receptor recibe la palabra 0001111111, determinar la palabra código original ya no es posible. La palabra 0001111111 puede obtenerse a partir de 2 errores en la palabra código 0000011111 o 3 errores en la palabra código 1111111111. Como en la transmisión de datos no es posible conocer a priori los errores que se producirán, pero sí el número máximo de errores que pueden producirse, es posible afirmar que el código anterior puede corregir errores de hasta 2 bits. Se generaliza además que, dado un **código de distancia de Hamming d** , es posible **corregir un número de errores e tal que $d = 2 \cdot e + 1$** .

CÓDIGO DE HAMMING

El código de Hamming debe su nombre a Hamming, quien introdujo en 1950 un método para generar un código de distancia 3 para cualquier conjunto de m bits de datos, de forma que es posible corregir errores en un bit. La relación entre los bits r y los bits m se establece mediante el cálculo de la paridad (par o impar) de un conjunto de bits.

En primer lugar hay que determinar cuantos bits redundantes son precisos para generar el código de Hamming de un conjunto de m bits de datos. Si se disponen de m bits de datos son posibles 2^m palabras de datos. Cada palabra de datos tendrá asociada una palabra del código de n bits, donde $n=m+r$. Si se desean corregir errores en un bit, es necesario que cada palabra código tenga asociadas n palabras código incorrectas y además todas las palabras código incorrectas asociadas a cada palabra código han de ser diferentes entre sí, para asegurar que cuando se produzca un error en un bit sólo exista una palabra código asociada. Con esta condición es fácil comprobar que el número de palabras código de n bits necesarias son $2^m(n+1)$: por cada palabra código, n palabras código incorrectas. Y además todas estas palabras se tienen que obtener por combinación de n bits, por lo que

$$2^m(n+1) \leq 2^n \quad (6.1)$$

Teniendo en cuenta que $n=m+r$ y sustituyendo en (6.1),

$$m + r + 1 \leq 2^r \quad (6.2)$$

De (6.2) es posible obtener, mediante sucesivas iteraciones, el número de bits r necesarios para obtener el código de Hamming de m bits de datos.

El procedimiento para generar las palabras del código viene dado a continuación.

1. Numerar los n bits de izquierda a derecha y en orden ascendente desde $k=1$.
2. Identificar los bits de datos y bits redundantes. Los bits redundantes será aquellos que se encuentran en posiciones que son potencia de 2, es decir $k=1,2,4,8,\dots,2^{r-1}$. El resto serán bits de datos.
3. Colocar los valores de los bits de datos en las posiciones adecuadas.

4. Calcular el valor de los bits redundantes como bits de paridad (par o impar). El bit redundante k se calculará como la paridad de los bits k y los bits de datos en posiciones que estén compuestos por la potencia del bit k . Por ejemplo, sea el bit redundante $k=4$, la paridad se calculará en base al bit k y los bits de datos donde aparezca la potencia 2^2 , es decir los bits $k=5, 6, 7, 12, \dots$
5. Realizando el procedimiento anterior para cada palabra de datos se obtendrán todas las palabras del código.

Determine el código de Hamming con paridad par asociado a un conjunto de palabras de datos de 2 bits.

En primer lugar hay que determinar el número de bits r necesarios. Para ello se emplea la ecuación 6.2 y se varía r desde $r=1$. De esta forma se obtiene que $r=3$. Por tanto las palabras del código tienen una longitud de $m+r=5$ bits. A continuación, y realizando el procedimiento anterior, se obtiene el siguiente código de Hamming.

2^0	2^1	2^0+2^1	2^2	2^0+2^2
r_0	r_1	m_0	r_2	m_1
1	2	3	4	5
0	0	0	0	0
1	0	0	1	1
1	1	1	0	0
0	1	1	1	1

Puede comprobarse como los conjuntos de bits: $\{r_0, m_0, m_1\}$, $\{r_1, m_0\}$ y $\{r_2, m_1\}$ tienen paridad par.

Por otra parte, es necesario describir un procedimiento para, una vez recibido un conjunto de n bits, determinar si ese conjunto es una palabra código válida, y en caso de que no lo sea corregirla. Este procedimiento sigue los siguientes pasos.

1. Numerar los n bits recibidos de izquierda a derecha y en orden ascendente desde $k=1$.
2. Inicializar una variable $contador=0$.
3. Verificar para cada bit redundante si la paridad es correcta.

4. En caso de que el bit redundante tenga un valor erróneo se incrementará la variable *contador* en el valor de la posición del bit redundante, pasando a verificar en cualquier caso el siguiente bit redundante.
5. Una vez verificados todos los bits redundantes, si la variable *contador* tiene valor cero no se han producido errores en la transmisión y pueden extraerse de la palabra código los bits de datos. Si por el contrario *contador* tiene un valor distinto de cero, este valor es la posición del bit erróneo, por lo que se corregirá cambiando el valor del mismo.

Dado el código de Hamming del ejercicio anterior, verificar si las palabras 01111 y 11101 son correctas y en caso de que no, corregir el error.

a) Para la palabra 01111, el procedimiento sería:

$$\begin{array}{ccccccc} \text{cont} = 0 & \xrightarrow{k=1} & \text{cont} = 0 & \xrightarrow{k=2} & \text{cont} = 0 & \xrightarrow{k=4} & \text{cont} = 0 \\ & & & & & & \rightarrow \text{No se han producido errores.} \end{array}$$

b) Para la palabra 11101, el procedimiento sería:

$$\begin{array}{ccccccc} \text{cont} = 0 & \xrightarrow{k=1} & \text{cont} = 1 & \xrightarrow{k=2} & \text{cont} = 1 & \xrightarrow{k=4} & \text{cont} = 5 \\ & & & & & & \rightarrow \text{Bit } 5 \text{ erróneo. Palabra} \\ & & & & & & \text{correcta: 11100} \end{array}$$

Ya se ha mencionado anteriormente que los errores en la transmisión de datos por medios físicos no suelen producirse de forma aislada sino en ráfagas de longitud k . De esta forma si queremos emplear un código corrector de errores para transmitir datos a través de un medio con errores en ráfaga precisamos de un código con distancia de Hamming $2 \cdot k + 1$, lo que precisa de un proceso complejo de generación del código. Sin embargo, es posible emplear un código de Hamming para transmitir datos que pueden ser corregidos a través de un medio con errores en ráfaga. Para ello simplemente hay que agrupar las palabras código a transmitir en bloques de k filas y transmitir los datos por columnas, de forma que el receptor recomponer el bloque de datos a partir de columnas y puede así corregir hasta k errores. Así se consigue que cuando se produzcan k errores seguidos cada palabra código sólo presenta un error.

	1 2 3 4 5	6 7 8 9 10
	↓ ↓ ↓ ↓	↓ ↓ ↓ ↓
k=4	0 0 0 0 0	1 0 0 1 1
	1 0 0 1 1	1 1 1 0 0
	1 0 0 1 1	0 1 1 1 1
	1 0 0 1 1	1 0 0 1 1

Figura 6.9 Envío de palabras de código de Hamming por columnas.



X

CAPÍTULO 7. CONTROL DEL FLUJO

7.1 INTRODUCCIÓN

El control del flujo es la última función del nivel de enlace que se analizará. Esta función tiene como objetivo que cada paquete del nivel de enlace, procedente de la fragmentación de un paquete del nivel de red, llegue al receptor y allí se recomponga el paquete de nivel de red original que debe llegar al otro extremo. Esta función debe controlar que un receptor lento no sea saturado por el envío demasiado rápido de paquetes desde el emisor. El envío de paquetes debe ser correcto, de forma que no se produzcan duplicaciones de paquetes en el receptor ni que el emisor interprete de forma correcta un paquete que no llegó al receptor. Todos estos aspectos se tienen en cuenta en una variedad de protocolos de control del flujo que se analizarán desde el más simple al más complejo.

7.2 PROTOCOLO UNILATERAL NO RESTRINGIDO *NO*

Éste es el protocolo de control del flujo más sencillo, en el que se considera que precisamente el control del flujo no es necesario. El término **unilateral** hace referencia a que la **transmisión de datos** sólo es **en un sentido**, de forma que el receptor sólo puede enviar información de control al emisor. En el caso de este protocolo el emisor no envía ninguna información de control al emisor, por lo que el medio físico se considera de tipo simplex. No se producirán perdidas de paquetes ni errores en los mismos. Además, el emisor siempre está listo para recibir paquetes del nivel de red y enviar paquetes de nivel de enlace al receptor, y por otro lado el receptor siempre está listo para recibir paquetes. Ello se conseguirá con la existencia de *buffers* de tamaño *infinito*. No se considera la fragmentación de los paquetes del nivel de red. Es evidente que esta situación no se aproxima a la realidad en ningún punto, pero sirve de partida para plantear situaciones reales que complican el protocolo de control del flujo.

El esquema de funcionamiento para emisor y receptor será el siguiente.

Emisor

1. Espera un paquete del nivel de red.
2. Construye un paquete de nivel de enlace.
3. Envía el paquete al medio físico y vuelve a 1.

Receptor

1. Espera un paquete del medio físico.
2. Extrae los datos del paquete de nivel de enlace.
3. Pasa los datos al nivel de red y vuelve a 1.

7.3 PROTOCOLO UNILATERAL DE PARADA Y ESPERA

Una primera aproximación a la situación real en el control del flujo de información es suponer que emisor y receptor no procesan los paquetes a la misma velocidad. En esta situación, el receptor ha de informar al emisor que está listo para recibir el siguiente paquete enviando un paquete de información pequeño denominado **de aceptación (ACK)**. El emisor por su parte no envía ningún paquete de nivel de enlace hasta que no recibe el ACK del anterior. Este mecanismo precisa por tanto de un medio físico semidúplex y los *buffers* del receptor serán de tamaño finito, pues puede producirse saturación en el receptor. No se consideran pérdidas en el medio físico ni errores en los paquetes.

Este mecanismo de parada y espera permite un control sencillo del flujo, modificando el esquema de funcionamiento de emisor y receptor anterior.

Emisor

1. Espera un paquete del nivel de red.
2. Construye un paquete de nivel de enlace.
3. Envía el paquete al medio físico y espera ACK.
4. Vuelve a 1.

Receptor

1. Espera un paquete del medio físico.
2. Extrae los datos del paquete de nivel de enlace.

3. Envía ACK al medio físico y datos al nivel de red.
4. Vuelve a 1.

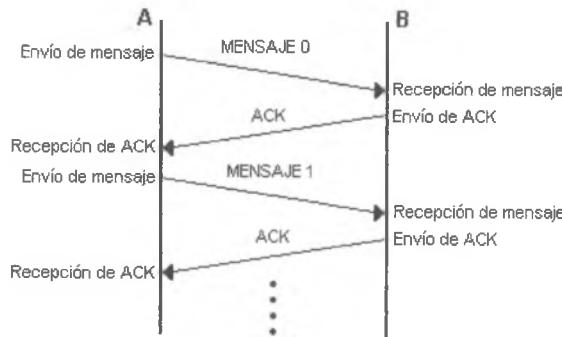


Figura 7.1 Protocolo unidireccional de parada y espera. Canal sin errores.

7.4 PROTOCOLO UNILATERAL DE PARADA Y ESPERA. CANAL CON ERRORES

Si se desea considerar una situación real de comunicación a nivel de enlace es preciso tener en cuenta los errores en el canal de comunicación. Los paquetes transmitidos en un medio físico real pueden sufrir errores en algunos de los bits que los componen o incluso el paquete puede no llegar al receptor. Con estas consideraciones el paquete de enlace dispondrá de un campo FCS que permitirá al receptor detectar errores en las tramas y proceder al reenvío de las mismas por parte del emisor. El mecanismo de comunicación con el emisor para indicarle que reenvíe una trama se basa en un temporizador. El emisor, cada vez que envía un paquete al receptor, inicia un temporizador y pasa a esperar la llegada del asentimiento (ACK). Si el paquete es recibido con errores por el receptor éste no envía el ACK y si el paquete no llega al receptor tampoco se enviará. En esta situación expirará el temporizador del emisor y se procederá con el reenvío del último paquete transmitido.

El esquema del funcionamiento del protocolo queda modificado de la siguiente forma.

Emisor

1. Espera un paquete del nivel de red.
2. Construye un paquete de nivel de enlace.
3. Envía el paquete al medio físico.
4. Inicia temporizador y espera ACK.
5. Si recibe ACK vuelve a 1.
6. Si no expira el temporizador. Reenvía el último paquete enviado.
7. Vuelve a 4.

Receptor

1. Espera un paquete del medio físico.
2. Extrae los datos del paquete de nivel de enlace y comprueba integridad de los datos.
3. Si Datos son Correctos Entonces Envía ACK al medio físico y datos al nivel de red.
4. Sino Descarta Datos.
5. Vuelve a 1.

Sin embargo, este protocolo presenta algunos problemas en su funcionamiento que provocan diferentes situaciones de error. Una de ellas se produce cuando se pierde el paquete de asentimiento en el canal. En ese caso el receptor habrá recibido correctamente el paquete del emisor y lo habrá enviado a su nivel de red. Sin embargo, como el emisor no recibirá el ACK reenviará la trama de datos y el receptor volverá a enviarla al nivel de red produciendo una **duplicación**.

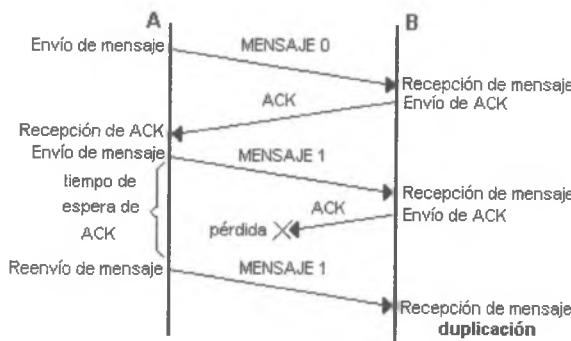


Figura 7.2 Duplicación de datos en el nivel de red por pérdida del ACK.

Otra situación de error en la que existe duplicación de datos en el nivel de red del receptor se produce cuando el receptor presenta un retardo elevado en el envío del ACK, produciéndose además una **pérdida de la sincronización**.

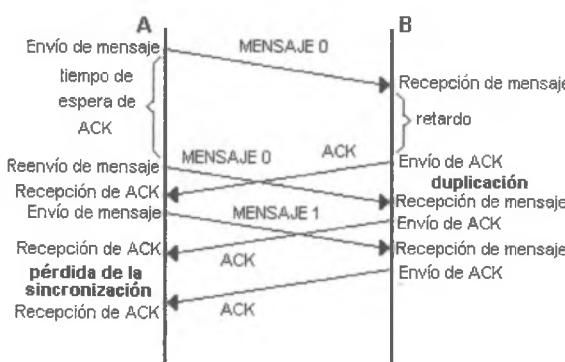
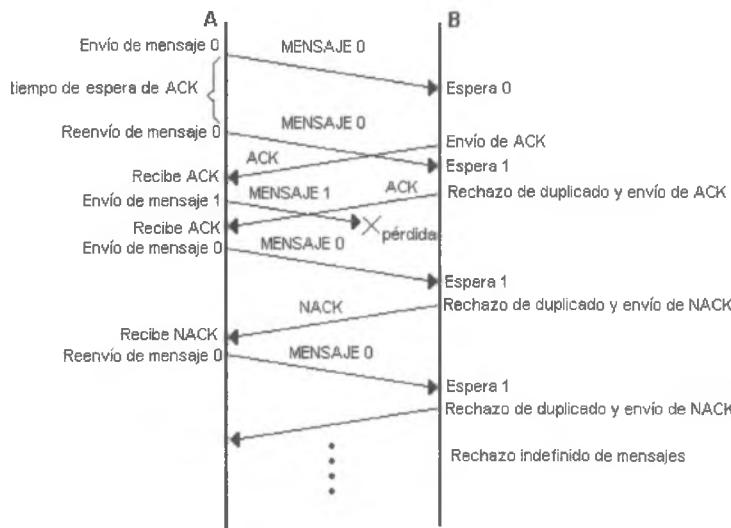


Figura 7.3 Pérdida de sincronización y duplicación por retardo en el envío del ACK.

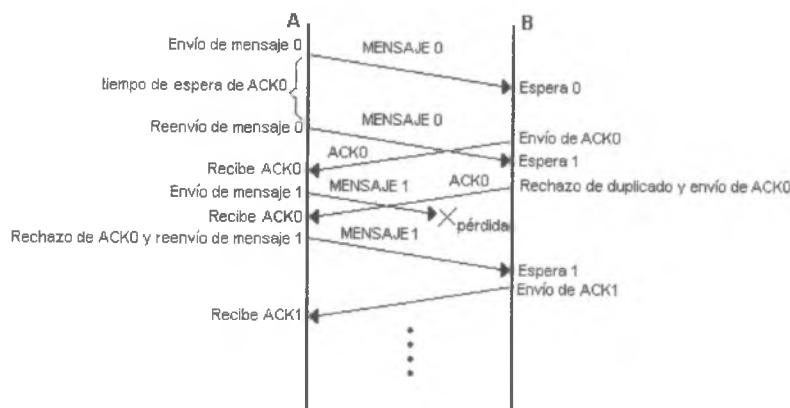
Para solventar el problema de la duplicación se realiza la numeración de las tramas empleando un bit. De esta forma el emisor envía tramas con secuencias 0-1-0-1-0-1..... El receptor, por su parte, sólo aceptará tramas que sigan la numeración 0-1-0-1-0-1.....



c) Numeración de tramas de datos

Figura 7.4 Protocolo de parada y espera con numeración de tramas empleando 1 bit.

La numeración de tramas no es suficiente para evitar los problemas debido a un temporizador en el emisor muy corto o retardos elevados en el envío de ACK en el receptor. Para subsanarlos es preciso numerar también las tramas de asentimientos.



d) Numeración de tramas de datos y asentimientos.

Figura 7.5 Protocolo de parada y espera con numeración de tramas y asentimientos.

7.5 PROTOCOLO BILATERAL DE PARADA Y ESPERA PIGGYBACK

Esta modalidad del protocolo de parada y espera permite la transmisión bidireccional de datos. El término *piggyback* significa **incorporación**, es decir, se aprovecha el envío de los paquetes ACK para incorporar datos del nivel de red que han de ser transmitidos al otro extremo.

El esquema de funcionamiento de este protocolo es el siguiente, donde cada extremo de la comunicación es emisor y receptor. Para simplificar el funcionamiento no se considerará numeración de tramas y asentimientos y se supondrá que el medio físico no presenta errores. Además los dos extremos de la comunicación siempre tienen datos del nivel de red para enviar.

Emisor/Receptor

1. Espera datos del nivel de red.
2. Construye fragmento con datos y envía paquete al medio físico.
3. Espera paquete del nivel de enlace.
4. Si recibe paquete con datos. (*Situación inicial del algoritmo*)
Espera paquete del nivel de red.
Construye fragmento con datos y ACK.

Envía paquete al medio físico.

Vuelve a 3.

5. Si recibe paquete con datos y ACK.

Espera paquete del nivel de red.

Construye fragmento con datos y ACK.

Envía paquete al medio físico.

Vuelve a 3.

7.6 PROTOCOLOS DE VENTANA DESLIZANTE

Los protocolos que emplean el esquema de ventana deslizante para el control del flujo se caracterizan por permitir transmisión de datos bidireccional con incorporación (piggyback). Las tramas de datos y asentimiento son numeradas, empleando para ello n bits, por lo que el rango de numeración es desde 0 hasta $2^n - 1$. El medio físico se considera de tipo Full Duplex con errores en los paquetes y pérdidas de los mismos.

La característica fundamental, y de ahí reciben su denominación, es la existencia de una **lista del emisor** y una **lista del receptor**, que consisten en unas listas con los números de secuencia consecutivos de los paquetes. Para un funcionamiento correcto ambas listas deben contener los mismos números de secuencia. Se define la **ventana del emisor** como el conjunto de números de secuencia de la lista del emisor asociados a paquetes enviados y que no han recibido confirmación. A su vez, se define la **ventana del receptor** como el conjunto de números de secuencia de la lista del receptor asociados a paquetes que el receptor está esperando recibir y para los que enviará aceptación. El **tamaño de la ventana del emisor** se define como el número de secuencias en la ventana del emisor, y de forma similar se define el **tamaño de la ventana del receptor**. El tamaño de estas ventanas no tienen porque ser el mismo para emisor y receptor y además puede ser variable durante el funcionamiento del algoritmo.

El funcionamiento básico de este esquema es el siguiente. El emisor aceptará paquetes de red y construirá paquetes de nivel de enlace aumentando el tamaño de su ventana al introducir los números de secuencia enviados. Este envío de paquetes se produce mientras el tamaño de la ventana del emisor no alcance un tamaño máximo prefijado. Además, cada vez que recibe una aceptación de un paquete enviado, elimina el número de secuencia de su ventana.

Por otra parte, el receptor tiene un tamaño de ventana prefijado k , donde se encuentran las primeras k secuencias de la numeración para ser aceptadas. Al recibir el primer paquete dentro de su ventana envía la aceptación y rota en una posición la misma, permitiendo de nuevo k secuencias.

La evolución del contenido de las ventanas en el funcionamiento de un protocolo de ventana deslizante se indica a continuación. Considerese el tamaño máximo de la ventana del emisor 1 ($W_{emisor} = 1$) y la del receptor 2 ($W_{receptor} = 2$). La numeración de los paquetes y ACK emplea 3 bits, luego el conjunto de secuencias será {0,1,2,...,7}.

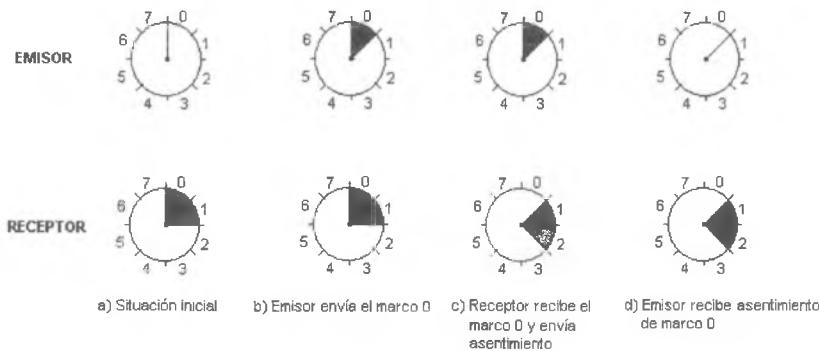


Figura 7.6 Evolución temporal de las ventanas en un protocolo de ventana deslizante.

7.6.1 Protocolo de ventana deslizante de 1 bit

Este protocolo se caracteriza por que la numeración de tramas y ACK se realiza empleando un único bit, por lo que las secuencias enviadas estarán numeradas como 0-1-0-1..... El tamaño de la ventana del emisor y receptor es el mismo y tiene valor 1, por lo que es equivalente al protocolo de parada y espera con incorporación (piggyback).

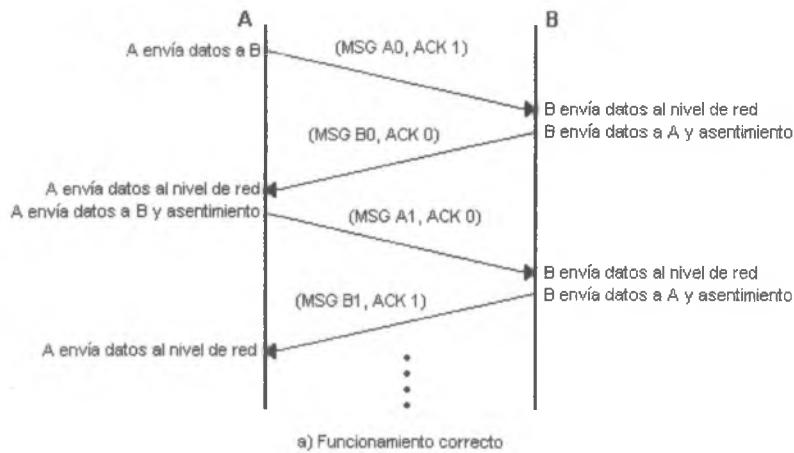


Figura 7.7 Funcionamiento del protocolo de ventana deslizante de 1 bit.

Este protocolo permite la recuperación de errores debido a pérdidas en el medio, que serán tanto de datos como de los asentimientos al emplear la **incorporación**.

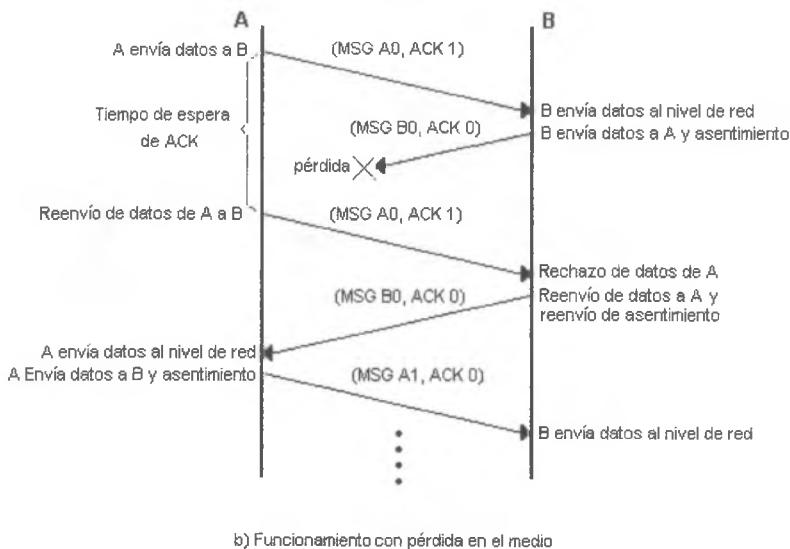


Figura 7.8 Funcionamiento del protocolo con pérdida en el medio.

En cuanto a las situaciones anómalas debido a un temporizador de espera de ACK demasiado corto, el protocolo de ventana deslizante de 1 bit

recupera los errores. Sin embargo, si los dos extremos de la comunicación inician la transmisión de datos de forma simultánea, se producen continuos reenvíos de la información en la transmisión.

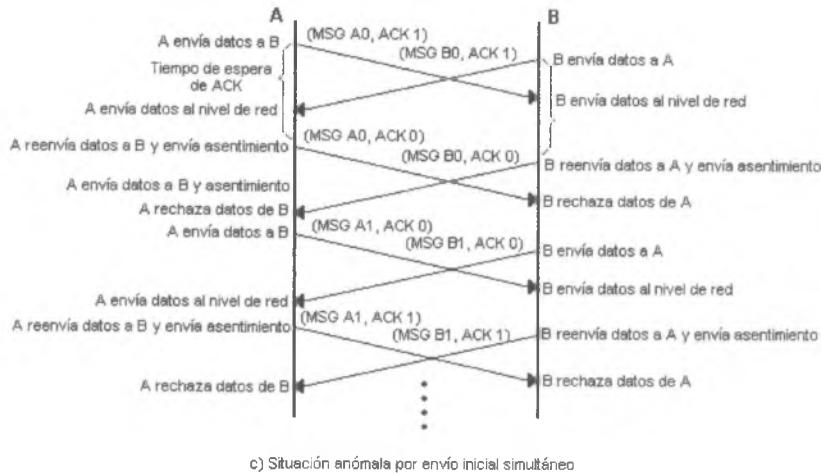
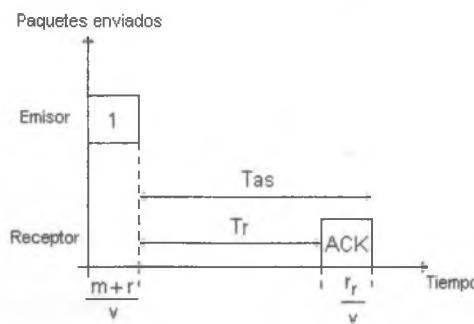


Figura 7.9 Envío simultáneo con temporizador de ACK corto.

7.6.2 Protocolo de ventana deslizante de envío continuo con repetición no selectiva

En el protocolo anterior, el **tamaño de la ventana del emisor limitado a 1** impide un mejor aprovechamiento del medio físico para el envío de datos, que está inactivo hasta la llegada del ACK del receptor. En el siguiente esquema se indica esta situación de inactividad del medio.



Se considerará que el paquete de datos consta de m bits de datos y r bits redundantes debido a las cabeceras del protocolo. V es la velocidad de transmisión del medio en bits/seg, r , son los bits del paquete de asentimiento y t , el tiempo de propagación (ida y vuelta) de un bit en el medio físico. Frecuentemente se indica T_{as} , el tiempo de asentimiento que se define como el tiempo transcurrido desde que el emisor envía el último bit de la trama de datos hasta que interpreta el ACK. De esta forma, se cumple que $T_{as} = t_r + \frac{r}{v}$.

Una medida del aprovechamiento de un medio físico en la transmisión de paquetes de datos es el **coeficiente de utilización** o **eficiencia** de una transmisión, y se define como el cociente entre el tiempo que precisa el envío de un paquete de datos y el tiempo total que el paquete ocupa el medio sin que otros paquetes puedan ser enviados. Para el caso del protocolo de parada y espera será:

$$E(\%) = \frac{T_{trama}}{T_{total}} \cdot 100 = \frac{\frac{m+r}{v}}{\frac{m+r}{v} + T_{as}} \cdot 100 = \frac{m+r}{m+r+vT_{as}} \cdot 100$$

Si se considera una comunicación vía satélite a $v = 50$ Kbps, paquetes de tamaño $m+r = 1000$ bits y un $T_{as} = 500$ ms, se obtiene una eficiencia $E = 3.846\%$. Es decir, el medio sólo se aprovecha en un 3.846% de su capacidad. Para solventar este problema es necesario aumentar la ventana del emisor, de forma que se pueda realizar el envío de tramas de datos aunque no haya llegado la aceptación. Si la ventana del receptor se mantiene a valor 1 se está considerando el denominado **protocolo de ventana deslizante de envío continuo con repetición no selectiva**. El tamaño de la ventana del emisor ha de ser como mínimo un valor tal que permita enviar paquetes sin aceptación hasta que llega el ACK del primer paquete. Este tamaño de ventana se obtiene como:

$$W_{emisor} = \frac{T_{total}}{T_{trama}}$$

De esta forma, cuando llega el primer ACK, la ventana del emisor está llena y puede entonces disminuir su tamaño en una unidad y enviar otro paquete. Debido a que se realizó un envío continuo de paquete de datos, los asentimientos llegarán de forma continua y el aprovechamiento del medio será del 100%. Sin embargo, si se produce un error en algún paquete, el emisor se percatará de ello cuando expire el temporizador para el

asentimiento del paquete. Por otro lado, el receptor, al tener un tamaño de ventana de 1, no aceptará los paquetes de datos enviados después del error y habrán de ser reenviados. Esto produce una disminución del aprovechamiento del medio, que depende de la tasa de error del mismo.



Figura 7.10 Protocolo de ventana deslizante de envío continuo con repetición no selectiva.

7.6.3 Protocolo de ventana deslizante de envío continuo con repetición selectiva

Para mejorar el aprovechamiento del ancho de banda del medio se aumenta el tamaño de la ventana del receptor, denominándose ahora a esta modalidad el **protocolo de ventana deslizante de envío continuo con repetición selectiva**. El funcionamiento es similar al protocolo anterior, pero ahora los paquetes no han de recibirse en orden al aceptar el receptor más de un número de secuencia. Se hace preciso por tanto de la existencia de un *buffer* en el receptor donde las tramas son almacenadas temporalmente y ordenadas antes de ser enviadas al nivel de red. Cuando se produce un error, el receptor almacena tramas dentro de su ventana y espera la secuencia inicial de la misma. Todas las tramas que lleguen a continuación y que no estén en la ventana del receptor serán rechazadas. Cuando llega la trama inicial de la ventana, el receptor envía el ACK de la última trama en el *buffer*. El emisor se percibirá del error producido cuando expire el tiempo de espera del ACK y procederá con el reenvío de la trama perdida, continuando con el envío de paquetes de su ventana. Cuando reciba el ACK del receptor continuará con la siguiente trama indicada en la aceptación. En el siguiente diagrama se ejemplifica el funcionamiento de este protocolo con un tamaño de ventana para el emisor y el receptor de 7, y un conjunto de secuencias para los paquetes suficientemente grande.

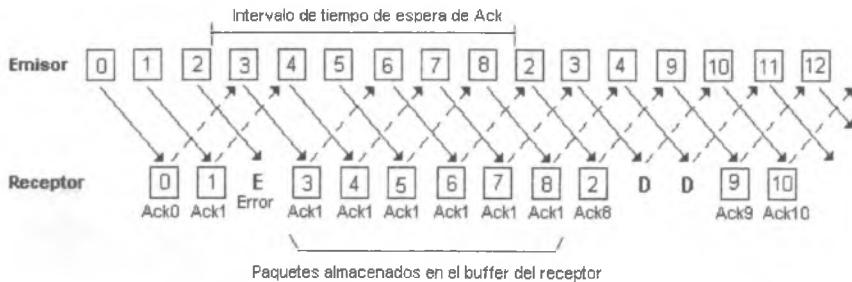


Figura 7.11 Protocolo de ventana deslizante de envío continuo con repetición selectiva.

En el ejemplo anterior, se producirán descartes en el receptor debido a que éste ha rotado su ventana y el emisor ha alcanzado el tamaño de ventana máximo, por lo que procederá con el reenvío de los paquetes de los que no ha recibido el ACK. Esto produce que el receptor reciba paquetes fuera de su ventana y los descarte. El tiempo que se desaprovecha el canal en estos descartes es debido al T_{as} . Si se emplea un tamaño de **ventana** mayor que 7 para el **emisor y receptor** de la situación anterior (de forma general un **tamaño suficientemente grande** limitado por los bits de la numeración), entonces **no se producen descartes de paquetes**.

Puede optimizarse los retardos en el reenvío de paquetes empleando tramas de rechazo NACK. Esta estrategia, denominada de **rechazo selectivo**, consiste en que el receptor envía una trama de no aceptación cuando el paquete llega con errores, de forma que el emisor procede con el reenvío y no hay que esperar a que expire el temporizador. Éste sólo se empleará cuando las tramas lleguen al receptor y no se pierdan los paquetes NACK.

Un aspecto a tener en cuenta en el protocolo de ventana deslizante selectivo es la **elección del tamaño de la ventana** del emisor y receptor en función de los números de secuencia disponibles para que no se produzcan errores.

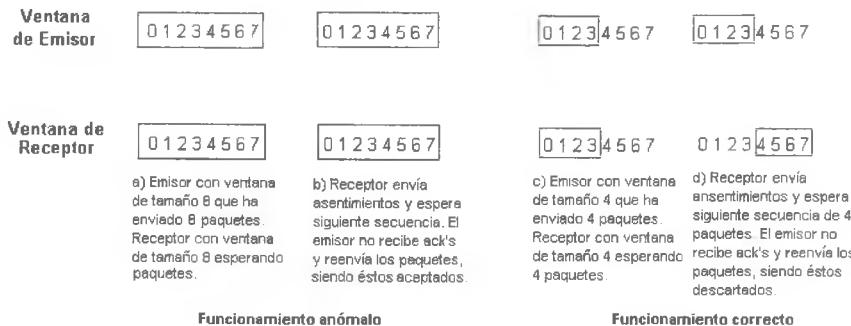


Figura 7.12 Elección del tamaño de las ventanas.

Supóngase una numeración de tramas y ACK empleando 3 bits donde las secuencias serán el conjunto $\{0,1,2,3,4,5,6,7\}$. Si emisor y receptor tienen tamaño de ventana 8, el emisor envía los primeros ocho paquetes y el receptor los está esperando. Si éstos son recibidos de forma correcta, la ventana del receptor rotará esperando de nuevo el conjunto de secuencias $\{0,1,2,3,4,5,6,7\}$ y enviará 8 paquetes ACK. Si se pierden todas las tramas de asentimiento, cuando en el emisor expire el temporizador se reenviarán todas las tramas de nuevo y el receptor volverá a aceptarlas. Por tanto se produce una situación de error de duplicación. Para evitar esto, la ventana del receptor rotada no debe contener números de secuencia de la ventana anterior. Para ello debe cumplirse que si n es el número de bits para la numeración, la ventana del emisor y receptor debe ser tal que:

$$W_{emisor} = W_{receptor} \leq \frac{2^n}{2}$$

7.7 RENDIMIENTO DE PROTOCOLOS

En este apartado se realizará una comparación de los protocolos de ventana deslizante y parada y espera en cuanto a su aprovechamiento del medio físico. Se supondrá que no se producen errores y los paquetes de datos contienen cabeceras de protocolo que suponen el 20% de la misma. De esta forma, el aprovechamiento máximo del medio será del 80% en la mejor situación.

Se introducirá el concepto de **longitud en tramas de un canal** para representar el uso del canal o aprovechamiento frente a esta magnitud. Sea un canal de transmisión que presenta un tiempo de propagación de un bit t_p .

Durante este tiempo, un emisor que esté enviando bits secuencialmente a una velocidad v habrá transmitido un secuencia de $t_p v$ bits en el medio. Esta cantidad de bits se denomina **longitud de un canal en bits**. Si existe un tamaño de trama de l bits, la longitud de un canal en tramas será de $\frac{t_p \cdot v}{l}$.

Si se compara la utilización del canal frente a la longitud del mismo en número de tramas, se aprecia cómo el protocolo de ventana deslizante tiene un aprovechamiento máximo del canal mientras la longitud del mismo sea menor o igual que $\frac{W_{emisor}}{2}$. Esta condición permite que el emisor pueda estar enviando tramas al canal hasta que llegue el ACK. Si la longitud del canal es mayor, entonces, cuando el emisor ha enviado la última trama de su ventana el ACK no ha llegado todavía, por lo que no se aprovecha de forma total el canal. Para el protocolo de ventana y espera el aprovechamiento total del canal se produce cuando la longitud del mismo es nula y por tanto no hay retardo en la interpretación del ACK.

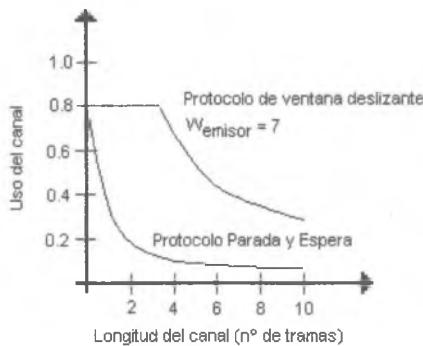


Figura 7.13 Uso del canal frente a la longitud del mismo en número de tramas.

Es posible apreciar cómo afecta el tamaño de la ventana deslizante del emisor al aprovechamiento del canal comparando medios con diferente longitud.

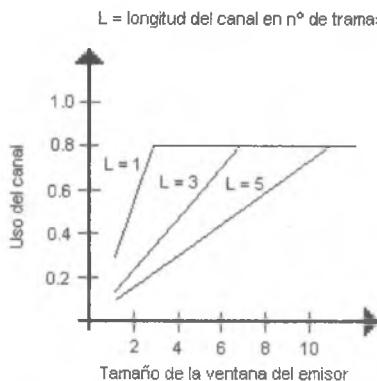


Figura 7.14 Uso del canal frente al tamaño de la ventana deslizante.

Puede comprobarse como efectivamente el aprovechamiento máximo se produce cuando la longitud del canal es menor o igual que $\frac{W_{emisor}}{2}$.

7.8 CADENCIA EFICAZ *No*

Anteriormente se ha visto cómo una de las métricas para el estudio del rendimiento de un protocolo de control del flujo era la **eficiencia**. Sin embargo, para tener además en cuenta la cantidad de información redundante que un protocolo soporta en los paquetes de datos y el efecto de los errores, existe una métrica más real que es la **cadencia eficaz** de una transmisión. La cadencia eficaz determina la tasa de bits de datos por segundo (C_e) que un protocolo proporciona empleando un canal con una tasa de v bits por segundo y teniendo en cuenta los errores en el mismo. Para determinar la cadencia eficaz introduciremos la siguiente notación:

$m \rightarrow$ bits de datos en una trama.

$r \rightarrow$ bits redundantes en una trama.

$n=m+r \rightarrow$ bits totales en una trama.

$v \rightarrow$ velocidad de transmisión del canal (bps).

$r_r \rightarrow$ bits de la trama de asentimiento.

$t_r \rightarrow$ tiempo de retardo de ida y vuelta de un bit en el canal (segundos).

$T_{as} = t_r + \frac{r_r}{v} \rightarrow$ tiempo de asentimiento (segundos).

$p \rightarrow$ probabilidad de recibir una trama errónea. Esta probabilidad es proporcional a la longitud de la trama. Si **BER** (tasa de error de un medio) es la probabilidad de que un bit transmitido en el canal sea erróneo, entonces $p=BER(m+r)$.

$N_t \rightarrow$ número medio de transmisiones de paquetes realizados para el envío de una trama de datos de forma correcta. Este valor puede determinarse de forma experimental determinando el número total de paquetes enviados debido a retransmisiones para enviar un conjunto de paquetes. $N_t = \frac{\text{Tramas totales}}{\text{Tramas útiles}}$. Sin embargo, un valor medio más preciso puede determinarse a partir de la probabilidad de recibir una trama errónea. Sea p_k la probabilidad de que un paquete requiera k transmisiones para ser enviado de forma correcta, entonces se cumple que:

$$N_t = \sum_{k=1}^{\infty} k \cdot p_k \quad (7.1)$$

Por otra parte, p_k se puede determinar como la probabilidad de que se produzcan $k-1$ errores y una transmisión correcta, por lo que $p_k = p^{k-1} \cdot (1-p)$. Sustituyendo en la ecuación (7.1) se obtiene,

$$N_t = \sum_{k=1}^{\infty} k \cdot p^{k-1} \cdot (1-p) = (1-p) \cdot \sum_{k=1}^{\infty} k \cdot p^{k-1} = (1-p) \frac{1}{(1-p)^2} = \frac{1}{1-p}$$

$N_r \rightarrow$ número de retransmisiones medio de paquetes en el envío de una trama de datos de forma correcta. Como se conoce el número medio de paquetes transmitidos (N_t), si no se considera el paquete enviado de forma correcta se obtiene el número de reenvíos. Es decir,

$$N_r = N_t - 1 = \frac{1}{1-p} - 1 = \frac{p}{1-p}$$

Así la cadencia eficaz se expresa como,

$$C_e = \frac{m}{T_0} bps$$

donde T_0 es el tiempo que una trama ocupa el canal hasta que se interpreta su recepción correcta. El valor de la cadencia eficaz dependerá del protocolo de control del flujo que se emplee, pues cada uno presentará un valor diferente de T_0 .

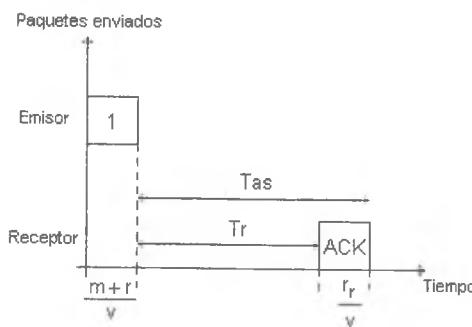
Hay que tener en cuenta que para comparar la cadencia eficaz que presentan diferentes canales con distinta velocidad de transmisión v , el valor absoluto de la misma no es adecuado. Para comparar la cadencia eficaz entre distintos canales hay que obtener su valor respecto de la velocidad del canal en %.

$$C_e (\%) = \frac{C_e}{v} \cdot 100 \%$$

De esta forma, la cadencia eficaz se convierte en una medida del aprovechamiento del canal, siendo del 100% en la situación óptima, cuando la C_e es igual a v . Esta situación no se dará nunca ya que los protocolos reales precisan de información redundante en forma de cabeceras.

Cadencia Eficaz para el protocolo de parada y espera

El esquema de funcionamiento del protocolo de parada y espera es el siguiente.



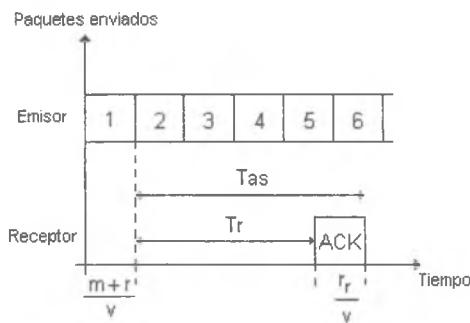
En esta situación de funcionamiento el tiempo total empleado en la transmisión de un paquete de forma correcta será el producto del tiempo asociado a la transmisión de un paquete por el número medio de paquetes necesarios.

$$T_0 = \left(\frac{m+r}{v} + T_{as} \right) \cdot N_t, \text{ luego}$$

$$C_e (\text{parada y espera}) = \frac{m}{T_0} = \frac{m}{\left(\frac{m+r}{v} + T_{as} \right) \cdot N_t} \text{ bps}$$

Cadencia Eficaz para el protocolo de envío continuo con repetición no selectiva

El esquema de funcionamiento de los protocolos de ventana deslizante con envío continuo es el siguiente.



Para el caso de la modalidad de repetición selectiva, se supondrá que el tiempo de expiración del temporizador en el emisor es igual al tiempo de asentimiento T_{as} . De esta forma el tiempo T_o empleando en la transmisión de una trama lo dividiremos en dos términos: t_1 y t_2 . El término t_1 se corresponde con el tiempo que se emplea en la transmisión del número medios de paquetes necesarios, es decir $t_1 = \left(\frac{m+r}{v} \right) N_t$. Por otra parte, el término t_2 referencia el tiempo adicional que se ocupa el canal sin transmitir datos debido a los descartes del receptor cada vez que se produce un reenvío. El tiempo durante el que se producen descartes en cada error será igual al tiempo de expiración del temporizador (ver figura 7.10), es decir T_{as} . Por tanto $t_2 = T_{as} \cdot N_r$, luego

$$T_0 = \left(\frac{m+r}{v} \right) N_t + T_{as} \cdot N_r$$

$$C_e = \frac{m}{T_0} = \frac{m}{\left(\frac{m+r}{v} \right) N_t + T_{as} \cdot N_r} \text{ bps}$$

Cadencia Eficaz para el protocolo de envío continuo con repetición selectiva

Con el mismo esquema de funcionamiento anterior, pero considerando la repetición selectiva y un tamaño de ventanas de receptor y emisor suficientemente grandes, el aprovechamiento del medio es total para el envío de tramas. De esta forma,

$$T_0 = \left(\frac{m+r}{v} \right) N_t$$

$$C_e = \frac{m}{\left(\frac{m+r}{v} \right) N_t} \text{ bps}$$

Como ejemplo se comparará el aprovechamiento de dos medios de transmisión diferentes como son el satélite en órbita geoestacionaria y el cable eléctrico cuando se comunican dos puntos separados una distancia de 15000 Km en la superficie terrestre. Los parámetros de la comunicación son:

	Satélite	Cable
m	1024 bits	1024 bits
r	64 bits	48 bits
r_r	64 bits	48 bits
BER	$2 \cdot 10^{-4}$	$5 \cdot 10^{-5}$
v	50 Kbps	9600 bps
t_r	500 ms	75 ms

En primer lugar hay que determinar el valor de p y T_{as} que no se proporciona en los valores dados.

$$p_{satelite} = BER(m+r) = 2 \cdot 10^{-4} (1024 + 64) = 0.2176$$

$$p_{cable} = BER(m+r) = 5 \cdot 10^{-5} (1024 + 48) = 0.0536$$

$$T_{as-satelite} = t_r + \frac{r_r}{v} = 0.5 + \frac{64}{50 \cdot 10^3} = 0.501 \text{ seg}$$

$$T_{as-cable} = t_r + \frac{r_r}{v} = 0.075 + \frac{48}{9600} = 0.08 \text{ seg}$$

Protocolo de parada y espera

Para el caso del satélite, la cadencia eficaz será:

$$C_{e-\text{satélite}} = \frac{m}{\left(\frac{m+r}{v} + T_{as}\right) \frac{1}{1-p}} = \frac{m \cdot (1-p) \cdot v}{m+r+T_{as} \cdot v} = \frac{1024 \cdot 0.7824 \cdot 50 \cdot 10^3}{1088 + 0.501 \cdot 50 \cdot 10^3} = 1532.59 \text{ bps}$$

$$C_{e-\text{satélite}} (\%) = \frac{C_{e-\text{satélite}}}{v} = \frac{1532.59}{50 \cdot 10^3} = 3.065\%$$

Para el caso del cable, la cadencia eficaz será:

$$C_{e-\text{cable}} = \frac{m \cdot (1-p) \cdot v}{m+r+T_{as} \cdot v} = \frac{1024 \cdot 0.9464 \cdot 9600}{1072 + 0.08 \cdot 9600} = 5056.24 \text{ bps}$$

$$C_{e-\text{cable}} (\%) = \frac{C_{e-\text{cable}}}{v} = \frac{5056.24}{9600} = 52.669\%$$

Protocolo de ventana deslizante con repetición no selectiva

Para el caso del satélite, la cadencia eficaz será:

$$C_{e-\text{satélite}} = \frac{m}{\left(\frac{m+r}{v} \frac{1}{1-p} + T_{as} \frac{p}{1-p}\right)} = \frac{m \cdot (1-p) \cdot v}{m+r+v \cdot p \cdot T_{as}} = \frac{1024 \cdot 0.7824 \cdot 50 \cdot 10^3}{1024 + 64 + 50 \cdot 10^3 \cdot 0.2176 \cdot 0.501} = \\ = 6126.26 \text{ bps}$$

$$C_{e-\text{satélite}} (\%) = \frac{C_{e-\text{satélite}}}{v} = \frac{6126.26}{50 \cdot 10^3} = 12.25\%$$

Para el caso del cable, la cadencia eficaz será:

$$C_{e-\text{cable}} = \frac{m \cdot (1-p) \cdot v}{m+r+v \cdot p \cdot T_{as}} = \frac{1024 \cdot 0.9464 \cdot 9600}{1024 + 48 + 9600 \cdot 0.0536 \cdot 0.08} = 8357.69 \text{ bps}$$

$$C_{e-\text{cable}} (\%) = \frac{C_{e-\text{cable}}}{v} = \frac{8357.69}{9600} = 87.059\%$$

Protocolo de ventana deslizante con repetición selectiva

Para el caso del satélite, la cadencia eficaz será:

$$C_{e-satélite} = \frac{m}{\left(\frac{m+r}{v}\right)1-p} = \frac{m \cdot v \cdot (1-p)}{m+r} = \frac{1024 \cdot 50 \cdot 10^3 \cdot 0.7824}{1024+64} = 36818.823 \text{ bps}$$

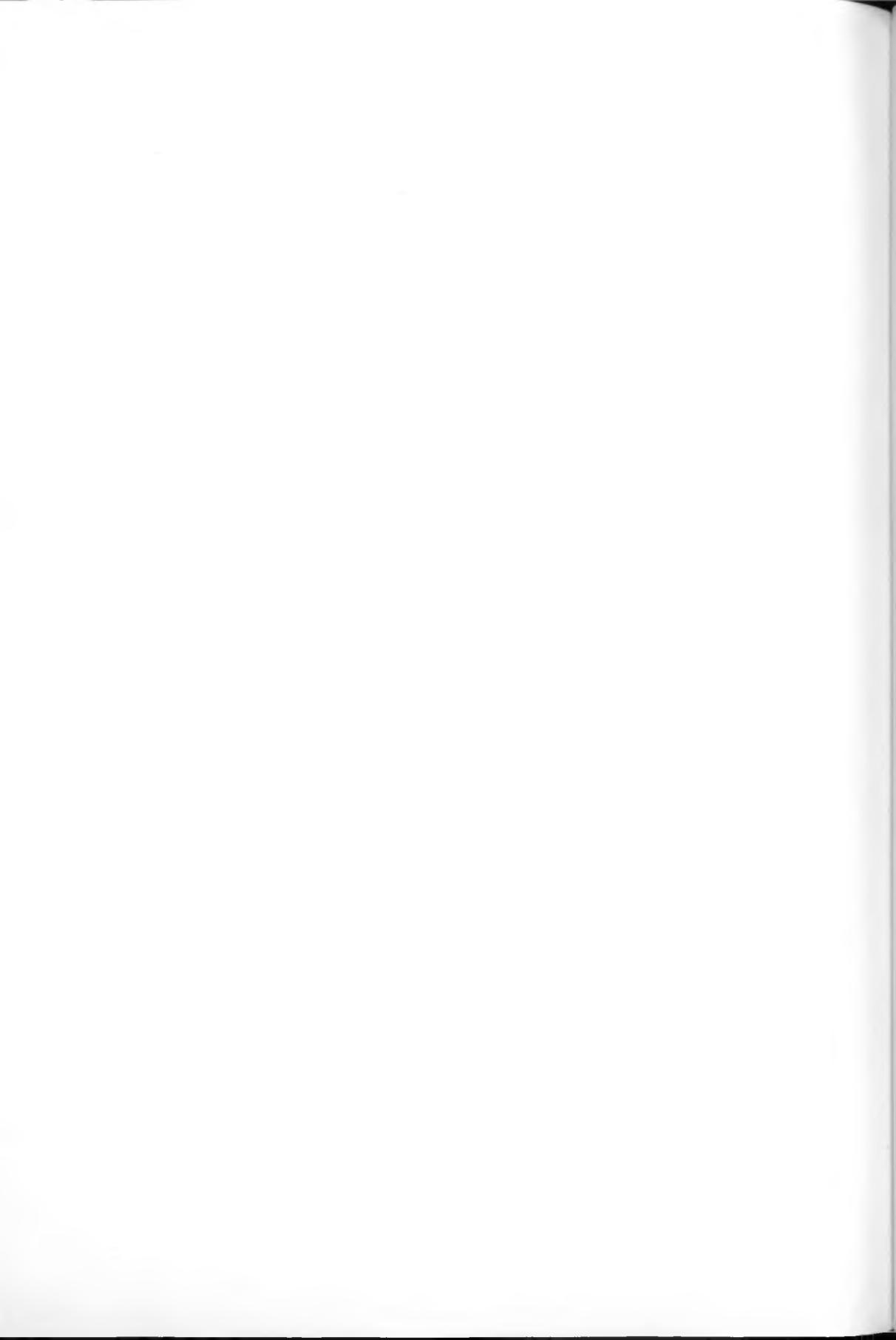
$$C_{e-satélite} (\%) = \frac{C_{e-satélite}}{v} = \frac{36818.823}{50 \cdot 10^3} = 73.637\%$$

Para el caso del cable, la cadencia eficaz será:

$$C_{e-cable} = \frac{m \cdot v \cdot (1-p)}{m+r} = \frac{1024 \cdot 9600 \cdot 0.9464}{1024+48} = 8678.629 \text{ bps}$$

$$C_{e-cable} (\%) = \frac{C_{e-cable}}{v} = \frac{8678.629}{9600} = 90.40\%$$

Como puede apreciarse, las mayores diferencias en la cadencia eficaz (%) entre el cable y el satélite se producen en los protocolos donde influye el tiempo de asentimiento. El elevado valor que presenta la transmisión vía satélite es un factor importante en el aprovechamiento del medio. Sin embargo, cuando el tiempo de asentimiento no influye para el rendimiento, como es el caso de la ventana deslizante con repetición selectiva, el medio con menor tasa de error es el que mejor aprovecha el medio. En este caso es el cable.



CAPÍTULO 8. PROTOCOLOS DE NIVEL DE ENLACE

Con este capítulo se pretende analizar dos ejemplos prácticos de protocolos de nivel de enlace que se emplean en redes de comunicaciones. Se han seleccionado dos por su importancia dentro de las redes de computadores actuales: HDLC, como un protocolo de nivel de enlace normalizado por la ISO y que sirve de base para otros protocolos de enlace empleados en numerosos tipos de arquitectura de red, y PPP como el protocolo de nivel de enlace empleado para conectar máquinas a redes TCP/IP empleando un canal de comunicación punto a punto, como es el acceso telefónico a Internet.

8.1 PROTOCOLO HDLC

HDLC son las siglas de **Control de Alto nivel del Enlace de Datos** (*High-Level Data Link Control*). Este protocolo, normalizado por la ISO (normas ISO 3309, ISO 4335) está basado en el protocolo de control del enlace síncrono, **SDLC**, de la arquitectura de red **SNA** de IBM.

Este protocolo está orientado a bit, empleando un control del flujo basado en ventana deslizante y con numeración de tramas de 3 y 7 bits (secuencias {0-7} y {0-127}).

8 bits	8 bits	8 o 16 bits	longitud variable	16 o 32 bits	8 bits
Delimitador	Dirección	Control	Datos del nivel superior	FCS	Delimitador
01111110				CRC-16 CRC-32	01111110

Figura 8.1 Formato de la trama de enlace HDLC.

El **delimitador** de inicio y fin es la secuencia de 8 bits **01111110**, empleando el relleno con un bit a 0 cuando aparecen 5 unos consecutivos

en los datos del nivel superior. El campo **dirección** consta de 8 bits que permite identificar diferentes estaciones en una topología multipunto o de difusión. El campo **control** indica el tipo de trama HDLC e incorpora información de control de numeración, asentimientos, etc. y puede tener una longitud de 8 o 16 bits. A continuación se disponen los **datos del nivel superior**, que generalmente son múltiplos de 8 bits. Finalmente aparece la **secuencia de verificación de trama** (FCS) para detectar errores en la trama de datos transmitida y que puede ser un código CRC-CCITT de 16 bits o CRC-32.

Los tipos de trama en el protocolo HDLC son tres: de información, supervisión y no numeradas.

Trama de información

Las tramas de información incorporan datos del nivel de red que han de ser transmitidas al otro extremo. Se caracterizan porque el primer bit del campo de control de la trama tiene valor 0.

	1	2	3	4	5	6	7	8
I: Información	0	N(S)		P/F		N(R)		

Figura 8.2 Formato del campo de control de una trama de información.

El campo N(S) indica el número de secuencia de la trama de enlace transmitida y N(R) indica el número de secuencia que el otro extremo espera recibir. El bit P/F permite identificar la última trama de un conjunto asociada a un paquete de nivel de red fragmentado. Todas los paquetes creados en la fragmentación tendrán el bit P/F a 1 (bit P) excepto el último que lo tendrá a 0 (bit F).

Trama de supervisión

Las tramas de supervisión controlan el funcionamiento del protocolo de ventana deslizante. Se distinguen porque los dos primeros bits del campo de control valen **10**.

	1	2	3	4	5	6	7	8
S: Supervisión	1	0	S	P/F		N(R)		

Figura 8.3 Formato del campo de control de una trama de supervisión.

Los bits S se emplean para determinar el tipo de trama de supervisión y en N(R) aparece un número de secuencia al que hace referencia la trama de supervisión. Existen cuatro tipos:

- a) **Tipo 0.** Receptor Listo (RR). Se corresponde con la trama de asentimiento del protocolo de ventana deslizante. Indica que se han recibido correctas hasta la trama N(R).
- b) **Tipo 1.** Rechazo (REJ). Trama de asentimiento negativo debido a un error en la transmisión. Sigue al emisor el reenvío de tramas desde la secuencia N(R).
- c) **Tipo 2.** Receptor no listo (RNR). Indica al emisor que se ha recibido correcta hasta la trama N(R) y que deje de transmitir más paquetes.
- d) **Tipo 3.** Rechazo selectivo (SREJ). Sigue al emisor el reenvío del paquete con secuencia N(R).

Trama no numerada

Las tramas no numeradas se emplean para controlar la conexión del enlace. Se caracterizan por tener los dos primeros bits del campo control a valor **11**.

	1	2	3	4	5	6	7	8
U: No Numeradas	1	1	M	P/F		M		

Figura 8.4 Formato del campo de control de una trama no numerada.

Los bits M codifican el tipo de trama no numerada, algunas de cuyas funciones son: solicitar conexión, solicitar desconexión, fijar el modo de comunicación, solicitud de RESET de la conexión, TEST, etc.

Cuando el campo de control tiene una longitud de 8 bits, únicamente hay disponibles 3 bits para los números de secuencia, es decir el conjunto {0,1,2,3,4,5,6,7}. Sin embargo, para las tramas de información y

supervisión, el campo de control puede tener 16 bits, dando lugar a un formato ampliado.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Información	0			N(s)			P/F									N(R)
Supervisión	1	0	S	0	0	0	0	P/F								N(R)

Figura 8.5 Formato ampliado del campo de control en tramas de información y supervisión.

De esta forma es posible aumentar el número de bits para las secuencias hasta 7, comprendiendo el conjunto {0,1,2,.....,127}. Con ello se consigue aumentar el tamaño de la ventana de emisor y receptor y obtener un rendimiento mayor del protocolo del control del flujo.

Tipos de estaciones HDLC

Cuando se establece un enlace de datos entre dos estaciones empleando el protocolo HDLC, existen 3 tipos distintos de estaciones.

- a) **Primarias.** Estaciones P. Estas estaciones tienen como función el controlar la transmisión de información enviando **comandos** y recibiendo **respuestas**. Estos comandos se envían a estaciones secundarias y se espera su respuesta.
- b) **Secundarias.** Estaciones S. Se comunican sólo con estaciones primarias, enviando respuestas a los comandos recibidos.
- c) **Combinadas.** Estaciones P/S. Estas estaciones se caracterizan porque pueden actuar tanto como primarias como secundarias.

Tipos de enlaces HDLC

La conexión de estaciones primarias, secundarias o combinadas empleando el protocolo HDLC establece tres tipos distintos de enlaces.

- a) **Enlace Balanceado.** Este enlace se establece entre estaciones combinadas, por lo que cualquiera de los extremos puede enviar comandos en cualquier instante y esperar respuestas.

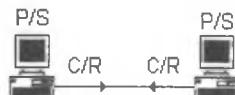


Figura 8.6 Enlace balanceado.

- b) **Enlace no Balanceado.** Se establece entre estaciones Primarias y Secundarias, de forma que la estación P envía comandos y la S responde.

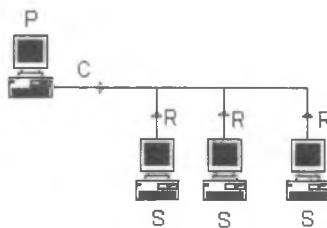


Figura 8.7 Enlace no balanceado.

Dentro del enlace no balanceado se distinguen dos modos:

1. **Modo Normal.** En el que las estaciones S sólo pueden transmitir respuestas a comandos enviados por la estación P.
2. **Modo Asíncrono.** En el que las estaciones S pueden transmitir respuestas a la estación P sin esperar ningún comando de la misma.

8.2 PROTOCOLO PPP

El protocolo PPP, **Protocolo Punto a Punto** (*Point to Point Protocol*), es un protocolo de nivel de enlace especificado para el acceso a redes TCP/IP. Está normalizado por el IETF (*Internet Engineering Task Force*) dependiente del IAB (*Internet Architecture Board*) en el documento **RFC 1661** (*Request For Comments 1661*).

PPP proporciona sobre una línea punto a punto detección de errores, verificación de autenticación en el enlace, reconocimiento de varios protocolos de nivel de red (IP, IPX, OSI CLNP, etc.) y negociación de direcciones de red IP, lo que le convierte en el protocolo de nivel de enlace empleado por los proveedores de acceso a Internet a través de líneas telefónicas.

El formato de trama de PPP se basa en HDLC, empleando un enmarcado con la secuencia de bits **01111110**. El protocolo, a diferencia de HDLC, es orientado a carácter, por lo que el relleno de los datos del nivel de red se realiza añadiendo la secuencia **01111110** cada vez que ésta aparece en los datos. Por tanto, el tamaño de la trama PPP es múltiplo de 8 bits, aunque su tamaño máximo es variable y negociable entre los dos extremos del canal.

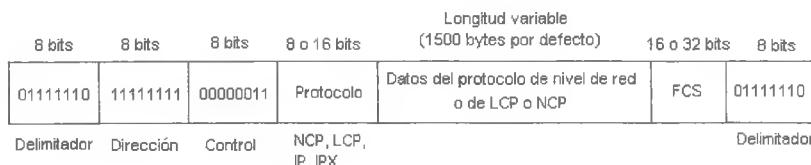


Figura 8.8 Formato del paquete de nivel de enlace PPP.

El campo **dirección** se rellena con los 8 bits a **1** dado que al ser una línea punto a punto el direccionamiento es implícito. El campo de **control** también es fijo a valor **00000011** y el campo **protocolo** especifica el tipo de protocolo al que pertenece la información del campo de datos. Por supuesto, se añade un campo **FCS** para detectar errores en la trama.

PPP también especifica un par de protocolos para el control del enlace punto a punto. El **protocolo de control del enlace LCP** (*Link Control Protocol*) tiene como objetivo activar la líneas punto a punto, probarlas, negociar opciones y desactivarlas cuando no son necesarias. La negociación de las opciones de la capa de red se realiza empleando el **protocolo de control de red NCP** (*Network Control Protocol*), que dependerán del protocolo de nivel de red empleado.

Una de las aplicaciones más importante del protocolo PPP es el establecimiento de un medio libre de errores para el intercambio de paquetes IP entre un usuario y un *router* de un proveedor de acceso a Internet a través de la línea telefónica.

El acceso a Internet con un proveedor se inicia cuando el PC del usuario llama al *router* del proveedor a través de una línea telefónica y un módem. En España, el coste de la llamada al *router* del proveedor es de llamada metropolitana independientemente de en qué lugar del territorio nacional se encuentre el mismo. Esta funcionalidad la proporcionaba **Infovía** hasta hace unos años de forma única; en la actualidad todos los proveedores de acceso a Internet proporcionan tarifas metropolitanas en las llamadas a nodos de acceso a Internet y tarifas planas para los usuarios intensivos. Una vez establecido el camino físico a través de la RTC (Red Telefónica Conmutada), el PC y el *router* se intercambian paquetes **LCP** (encapsulados en paquetes PPP) negociando parámetros de la conexión. Uno de los más importantes es la autenticación del usuario que solicita el acceso empleando un nombre de usuario y una clave. Existen varios protocolos de autenticación, entre los que destacan por su uso **PAP** (*password authentication protocol*) y **CHAP** (*challenge-handshake authentication protocol*).

A continuación se negocian los parámetros del nivel de red empleando paquetes **NCP**. En este caso el nivel de red es IP, y para que el usuario pueda ejecutar el conjunto de protocolos TCP/IP es preciso proveerle de una dirección IP **legal**. El número de direcciones IP legales está limitado, por lo que un proveedor de Internet tiene un conjunto limitado **n**. Cada vez que un usuario se conecta con el *router* del proveedor se le proporciona una dirección IP del conjunto que tenga disponibles y ésta es una medida de la calidad de un proveedor. Frecuentemente el usuario se encuentra con problemas en el acceso en el que la autenticación es incorrecta a pesar de que los datos son correctos. Ello es debido a que el proveedor no cuenta con las suficientes direcciones IP para un acceso simultáneo de todos los clientes, por lo que a algunos se les negará el acceso. Un proveedor de calidad dispondrá de un conjunto de direcciones IP suficientes para soportar un acceso simultáneo elevado de clientes.

A partir de este momento, en el que al usuario se le ha asignado una dirección IP, el PC puede recibir y enviar paquetes IP como si fuera un nodo conectado a Internet. Una vez que se desea finalizar la conexión, el protocolo NCP finaliza la conexión de red y libera la dirección IP. LCP libera la capa de enlace de datos y el módem cierra la conexión física colgando el teléfono.

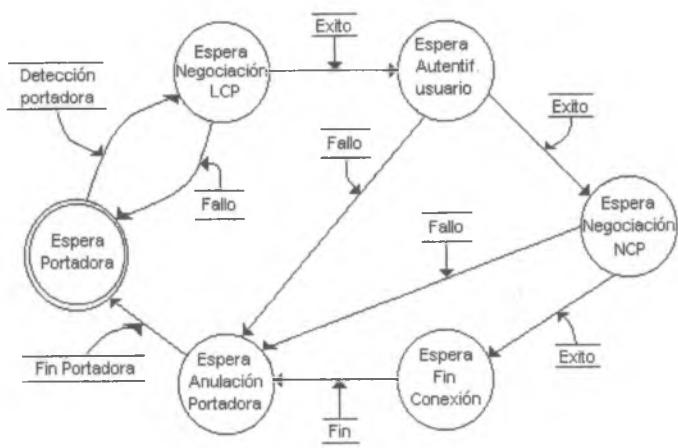


Figura 8.9 MEF simplificada del funcionamiento del protocolo PPP.

CAPÍTULO 9. FUNCIONES DEL NIVEL DE RED

El capítulo 9 inicia el estudio del tercer nivel de la arquitectura de red del modelo OSI, también presente en el modelo TCP/IP y conocida como capa de interred o Internet.

9.1 INTRODUCCIÓN

El objetivo del nivel de red, tanto en el modelo OSI como en el TCP/IP, es realizar el correcto encaminamiento de los paquetes de información que van de un nodo origen a un nodo destino de la subred. Esta funcionalidad de encaminamiento se presenta en las redes WAN (Redes de Área Extendida), pues son éstas las que presentan una topología de nodos (encaminadores o *routers*) interconectados entre sí, en los que se realiza la comutación de paquetes entre un *host* (estación o ETD) origen y un *host* destino. Es función por tanto del nivel de red controlar la topología de la subred, determinando los caminos con menor número de saltos entre origen y destino, coste de los enlaces entre pares de nodos, reducir la congestión en la red, etc.

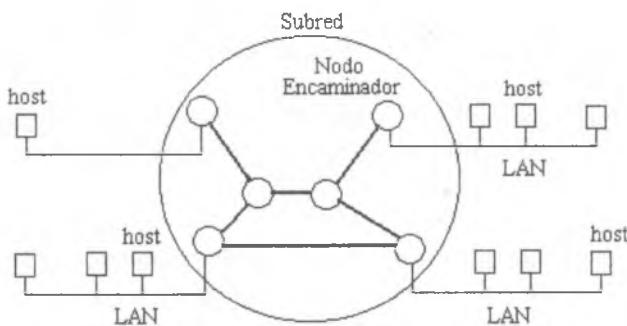


Figura 9.1 Esquema de una red WAN.

En la comunicación de información a través de una red WAN, los nodos de conmutación o encaminamiento de la red tienen implementadas funcionalidades hasta el nivel de red. De esta forma, los ETD de los extremos de la comunicación tienen implementadas las funcionalidades de la capa de transporte. Para esta capa del modelo de arquitectura de red, la comunicación será como si existiera un enlace dedicado entre origen y destino.

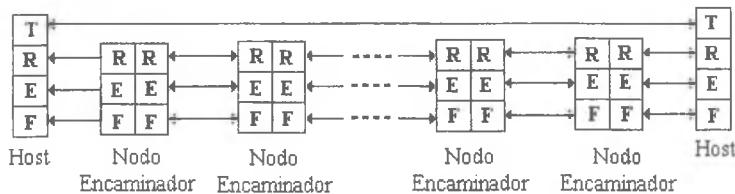


Figura 9.2 Modelo de comunicación en una red WAN.

En este capítulo se estudiarán cuáles son los servicios que el nivel de red proporciona a la capa de transporte, así como las diferentes organizaciones para la misma.

9.2 SERVICIOS PROPORCIONADOS A LA CAPA DE TRANSPORTE

La capa de red proporciona un conjunto de servicios a su capa superior, la capa de transporte, que se diseñaron persiguiendo una serie de objetivos:

- **Los servicios ofrecidos deben ser independientes de la tecnología de la subred.** Los servicios proporcionados a la capa de transporte deben ser independientes de la arquitectura de la red (comutación de circuitos, comutación de paquetes, mallas, etc) permitiendo así que la subred pueda evolucionar en el tiempo proporcionando a la capa superior servicios cada vez más fiables y rápidos.
- **La capa de transporte debe ser transparente a la cantidad, tipo y topología de las subredes presentes.** La capa de red ha de tener en cuenta los aspectos de encaminamiento al atravesar los paquetes subredes de diferentes características, de forma que para la capa de transporte la comunicación entre origen y destino es extremo a extremo, como si estuvieran unidas por un único enlace.

- **Las direcciones de red disponibles para la capa de transporte deben seguir un plan de numeración uniforme, aún a través de varias LAN y WAN.** El nivel de red tiene que proporcionar un sistema de numeración para identificar cada nodo de la red WAN, independientemente del tipo de red donde se encuentra. Esta funcionalidad está asociada a la interconexión de redes, para así poder interconectar redes con topologías y medios físicos distintos.

El servicio que la capa de red proporciona a la capa de transporte se puede clasificar en **orientado a conexión** y **no orientado a conexión**. La elección entre si la capa de red debe proporcionar un servicio orientado a conexión o sin conexiones presenta dos alternativas en las redes de comunicaciones actuales.

El servicio no orientado a conexión es característico de una subred WAN inestable. El principal defensor de este tipo de servicio para la capa de red es la comunidad Internet. Con esta filosofía en el tipo de servicio, la capa superior (transporte) deberá realizar el control del flujo de datos y de los errores. De esta forma, la capa de red proporciona un servicio con dos primitivas básicas: enviar paquete y recibir paquete. Cada uno de ellos circulará por la red de forma independiente, pudiendo llegar desordenados a su destino.

El servicio orientado a conexión tiene su principal defensor en las compañías telefónicas, pues el servicio telefónico mundial se caracteriza por proporcionar un servicio fiable y orientado a conexión en la subred. Este servicio, proporcionado a la capa superior, se caracteriza por:

1. Antes de enviar los datos, un proceso de la capa de red del extremo iniciador de la comunicación establece una conexión con un proceso de la capa de red del extremo no iniciador. A la conexión se le asignará un identificador, que se mantendrá hasta que sea liberada.
2. Al establecer la conexión, se negocian los parámetros de la misma: coste y calidad de servicio.
3. La transmisión de información se realiza de forma secuencial y bidireccional.
4. Se realiza un control del flujo automático por parte de la red para evitar la saturación.

5. La liberación de la conexión finaliza la transmisión de datos y libera el identificador de la conexión.

La controversia entre ambas modalidades está en dónde se implementa la complejidad de la comunicación. Para el servicio orientado a conexión, la mayor complejidad se implementa en la capa de red (en los nodos de la subred), mientras que en el servicio no orientado a conexión, la complejidad se encuentra en la capa de transporte (en los *hosts* de la subred).

Cada uno de los defensores de la modalidad de servicio proporcionan una serie de argumentaciones que justifican el uso del tipo de servicio.

Defensores del servicio no orientado a conexión

- La capacidad de cómputo del usuario se ha vuelto cada vez más barata, por lo que la complejidad se puede implementar en los hosts que interconecta la subred y no en los nodos.
- La subred es una inversión de gran escala que durará décadas, por lo que no se deben incluir en ella técnicas que podrían quedar obsoletas en breve tiempo.
- Aplicaciones tales como la transmisión de voz digitalizada en redes de conmutación de paquetes o la adquisición de datos en tiempo real precisan de una entrega rápida, más que exacta.

Defensores del servicio orientado a conexión

- Los usuarios demandan un servicio confiable y libre de problemas, lo que se consigue empleando conexiones a nivel de red.
- La mayoría de los usuarios no quieren ejecutar complejos protocolos de transporte en sus máquinas, demandan dispositivos sencillo adecuados a la aplicación que se desea de la red (por ejemplo el terminal telefónico).
- Es más rápido proporcionar sonido y vídeo en tiempo real a través de una red orientada a conexiones que una sin conexiones, pero puede producirse congestionamiento más rápidamente.

9.3 ORGANIZACIÓN INTERNA DE LA CAPA DE RED

Independientemente del tipo de servicio proporcionado por la capa de red, el funcionamiento interno de la subred puede ser mediante circuitos virtuales o mediante datagramas. En el estudio de las redes de computadores nos centraremos en estos dos tipos de redes de commutación de paquetes, empleadas más frecuentemente para la transmisión de datos.

La commutación de paquetes (analizada en el **capítulo 1**) se fundamenta en la fragmentación de la información de niveles superiores en paquetes con el mismo formato que son encaminados a través de los nodos de la subred.

9.3.1 Circuitos virtuales

Esta organización interna para la subred se emplea en aquellas cuyo servicio primario está orientado a conexión, como son las **redes X.25**. En este tipo de funcionamiento de commutación de paquetes, cuando se requiere el envío de información entre un nodo origen y un nodo destino es necesario un proceso de conexión. En esta fase inicial de conexión, se determina un camino en la subred por la que circularán secuencialmente los paquetes de información a transmitir. Una vez transmitidos los paquetes se procederá con la liberación de la conexión, y por tanto del camino determinado. Cada vez que se realiza una conexión para transmitir datos a un mismo nodo destino, el camino que se elige para todos los paquetes no tiene por qué ser el mismo, dependiendo de la congestión que exista en los nodos de la subred. La subred puede permitir que un usuario disponga en todas sus conexiones a un mismo destino el mismo camino en la subred: a esta modalidad se la denomina **circuitos virtuales permanentes**.

9.3.2 Datagramas

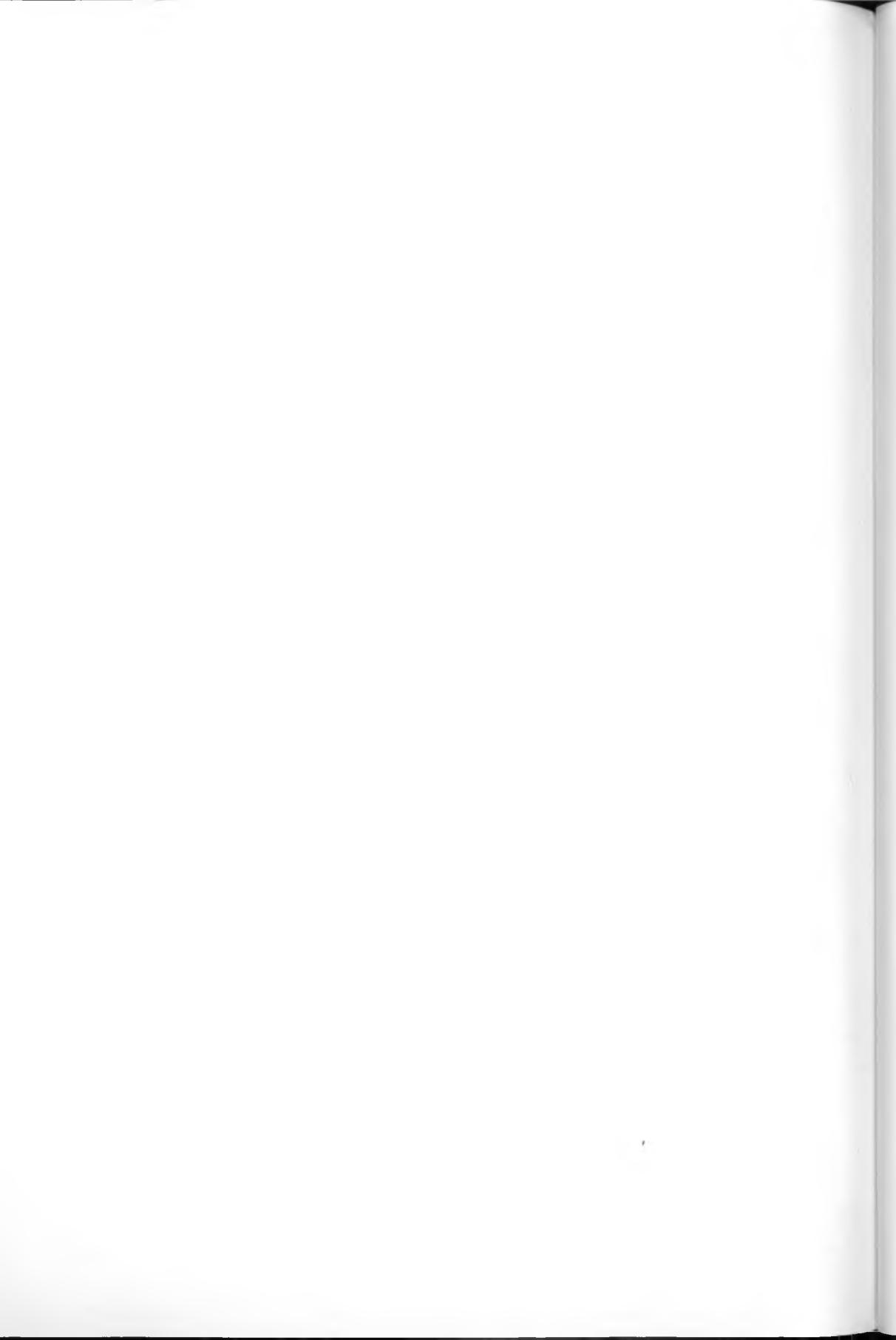
Las redes de commutación de paquetes basadas en datagramas son características de las subredes que proporcionan un servicio básico no orientado a conexión, como son las **redes IP** (Internet). En este tipo de redes, cada paquete de información es encaminado en la subred de forma independiente, de forma que éstos pueden seguir caminos independientes en la misma, pudiendo llegar a su destino desordenados. Los nodos de este tipo de subredes son más complejos y disponen de una mayor carga de cómputo, pues han de realizar un proceso de encaminamiento por cada paquete. Sin embargo, esta característica dota a las redes basadas en datagramas de una mayor tolerancia a fallos, pues los paquetes serán encaminados automáticamente por caminos alternativos si algún nodo falla.

9.3.3 Circuitos Virtuales versus Datagramas

Una vez indicadas las características de cada modalidad de conmutación de paquetes, se procederá a realizar una comparación entre las mismas atendiendo a diversos aspectos.

- **Establecimiento del circuito.** La subred basada en circuitos virtuales precisa de un tiempo inicial para establecer el circuito en la conexión. Si la cantidad de información a transmitir es pequeña es posible que el tiempo de establecimiento sea superior o muy grande frente al tiempo de transmisión de la información en la subred, por lo que es más aconsejable el empleo de redes de datagramas. Por el contrario, las redes de datagramas precisan de un tiempo de procesamiento del encaminamiento cuando cada paquete atraviesa un nodo de la subred.
- **Direccionamiento.** Las subredes basadas en datagramas precisan en cada paquete de información la dirección completa del nodo origen y destino para realizar el proceso de encaminamiento. Por el contrario, las subredes basadas en circuitos virtuales precisan de un identificador de circuito virtual en el paquete de longitud pequeña, por lo que incorporarán en los paquetes menor cantidad de información redundante y por tanto habrá un mejor aprovechamiento del ancho de banda.
- **Información de estado.** Las subredes de circuitos virtuales precisan de recursos para conocer el estado del circuito virtual: volumen de tráfico transmitido, tiempo de duración de la conexión, etc. Este tipo de información no existe en las subredes de datagramas, por lo que se precisan menos recursos en el nodo.
- **Encaminamiento.** En las subredes de datagramas, en cada nodo encaminador de la subred es necesario determinar cuál es el siguiente nodo al que se enviará el paquete, dependiendo de la dirección de destino que aparezca en el mismo. Esta funcionalidad se lleva a cabo consultando en el nodo una **tabla de encaminamiento** o **tabla de rutas**. En las subredes de circuitos virtuales, cada nodo encaminador dispone de una tabla con circuitos virtuales de entrada y salida relacionados entre sí. De esta forma, el proceso de encaminamiento se lleva a cabo buscando en una tabla un identificador de circuito virtual, lo que conlleva un encaminamiento muy rápido de los paquetes frente a las subredes de datagramas.

- **Efecto de fallos en el encaminador.** Si en una subred de circuitos virtuales se produjera un fallo en un nodo encaminador, se interrumpirían todos los circuitos virtuales que atravesaran ese nodo, por lo que se precisaría volver a establecerlos desde el origen. Por el contrario, un fallo en un nodo de una subred de datagramas produce la pérdida de los paquetes de información en tránsito en ese nodo, por lo que sólo habría que retransmitir éstos. El resto alcanzaría su destino por otros caminos en la subred. De este forma, las subredes de datagramas presentan una mayor tolerancia a fallos.
- **Control del congestionamiento.** La gestión del congestionamiento en los nodos de la subred es más fácil y se lleva a cabo más eficientemente en las subredes de circuitos virtuales. Estas subredes permiten tener información de cada nodo acerca del número de circuitos virtuales establecidos, los recursos consumidos, etc. Cuando un nodo se encuentra muy congestionado en una subred de circuitos virtuales no se permitirá el establecimiento de más circuitos virtuales, mientras que no es posible a priori eliminar paquetes en una subred de datagramas sin producir errores.



CAPÍTULO 10. ALGORITMOS DE ENCAMINAMIENTO Y CONTROL DE LA CONGESTIÓN

10.1 INTRODUCCIÓN

Las funcionalidades más importantes del nivel de red hacen referencia al encaminamiento de la información y al control de la congestión en la subred. En este capítulo se analizará la problemática y los algoritmos de encaminamiento más importantes, centrándonos en las redes de conmutación de paquetes, características de las redes de computadores. A continuación, se abordará el problema de congestión en los nodos de redes de conmutación de paquetes, estudiando algunos algoritmos de prevención del congestionamiento. Finalmente se introducirá una herramienta para la modelización de la congestión en redes de conmutación de paquetes, la teoría de colas.

10.2 ALGORITMOS DE ENCAMINAMIENTO

Los procedimientos o algoritmos de encaminamiento se implementan como procesos del nivel de red que determinan por qué líneas han de ser retransmitidos los paquetes del nivel de red para que alcancen su destino. Si la subred de conmutación de paquetes se basa en datagramas la decisión de encaminamiento se tomará por cada paquete que alcanza el nodo, mientras que si se basa en circuitos virtuales la decisión se tomará sólo una vez en cada establecimiento de conexión entre origen y destino.

Un algoritmo de encaminamiento debe presentar un conjunto de características que se indican a continuación:

- 1. Simplicidad.** El algoritmo de encaminamiento debe tener reglas sencillas y costes de computación reducidos, sobre todo en el caso de subredes WAN grandes.

- 2. Fiabilidad.** El algoritmo debe responder de manera adecuada en caso de errores o fallos en la transmisión.
- 3. Calidad.** El algoritmo debe llegar a soluciones globales óptimas.
- 4. Convergencia.** Para un tráfico en la red determinado, el algoritmo debe converger a una situación de régimen permanente sin oscilaciones.
- 5. Adaptación.** El algoritmo debe adaptarse a cambios de tráfico y topología de forma rápida y uniforme para que pueda funcionar en tiempo real.

El objetivo del algoritmo de encaminamiento es determinar una ruta entre origen y destino. En el caso de subredes basadas en circuitos virtuales esa ruta será la misma para todos los paquetes de información transmitidos durante la conexión, y en el caso de subredes de datagramas por cada paquete se determinará una ruta, que podrá ser la misma o no.

Atendiendo a la naturaleza de las rutas en el algoritmo de encaminamiento, es posible realizar una clasificación en distintos tipos.

10.2.1 Tipos de algoritmos

Es posible realizar una distinción entre diferentes tipos de algoritmos basada en la naturaleza de las rutas entre los pares de nodos de la subred.

- 1. Encaminamiento fijo.** En este tipo de algoritmos, en cada nodo se definen unas tablas de encaminamiento (conocida la topología de la red), las cuáles, conocida la dirección destino, informan acerca del siguiente nodo al que ha de ser enviado el paquete. Este tipo de algoritmos presenta una baja tolerancia a fallos en los nodos y congestión de la red, pues si un nodo falla parte de la red dejará de funcionar y en las rutas fijas aumenta la congestión. Los fallos en los nodos pueden solucionarse empleando rutas de respaldo para el caso de que algún nodo en alguna de las rutas deje de funcionar. Una ventaja clara de este tipo de encaminamiento es la sencillez y rapidez del mismo en una red con una carga de tráfico estable.
- 2. Inundación.** En este tipo de encaminamiento cada paquete entrante se envía por todas las rutas o enlaces del nodo excepto por el de llegada. Esto genera un gran número de paquetes duplicados, y si no existe algún mecanismo para detener la retransmisión de paquetes se producirá la

congestión de la red. Para evitar la congestión existen básicamente dos mecanismos limitadores:

- a) **Contador de salto.** Es un valor que se incluye en la cabecera de cada paquete. Cada vez que el paquete pasa por un nodo este valor se decremente, y si alcanza el valor cero se elimina. Este valor se determina de forma que los paquetes que siguen las rutas más largas se eliminan.
- b) **Número de secuencia.** Es un valor que se incluye en la cabecera de cada paquete. Cada nodo fuente asigna un valor distinto para cada paquete que le llega de un terminal conectado al mismo y es enviado a la subred. Los nodos intermedios en la red generan una lista de secuencias de paquetes enviados por cada nodo de la red, de forma que si aparece un paquete con un número de secuencia que ya está en la lista se descarta.

La técnica de inundación no se emplea en la práctica para el encaminamiento de la información, pero presenta dos propiedades interesantes que la hacen útil para determinadas situaciones.

- Se ensayan todas las rutas posibles entre fuente y destino, de forma que si algún nodo no funciona se asegura que el paquete llegará por un camino alternativo (siempre que éste exista).
- Al menos una de las copias del paquete llegará a su destino a través de la ruta de coste mínimo.

Estas propiedades hacen que esta técnica se emplee para el envío de **mensajes urgentes** de alta prioridad a través de la red. Estos mensajes se emplean en aplicaciones de tiempo real.

3. **Encaminamiento adaptativo.** Las estrategias anteriores no reaccionan a las condiciones de la red, pudiendo el operador en ocasiones realizar acciones para adecuarlas. Las estrategias de encaminamiento adaptativas se caracterizan porque se adaptan a las condiciones de cambio de la subred. Con ello se consigue una mejora del rendimiento al realizar una distribución del tráfico en la subred lo más equitativa posible. Sin embargo, también presenta inconvenientes: la decisión de encaminamiento será más compleja y se precisará de una mayor capacidad de proceso en el nodo; en ocasiones la estrategia se basa en informaciones de un punto de la subred que son evaluadas en otro, por lo que aumenta el tráfico en la red; y por último, si la reacción a los

cambios no es lo suficientemente rápida, la congestión en la red aumenta. Dentro del encaminamiento adaptativo es posible distinguir entre 3 tipos distintos.

- A. Centralizado.** En la red existirá un **Centro de Control del Encaminamiento** (CCE), el cual, basándose en la información de los nodos y su conocimiento global de la subred, calculará las rutas óptimas entre todos los nodos y las transmitirá a cada uno de ellos. Sería similar al encaminamiento fijo pero las rutas se determinan dinámicamente y periódicamente. Algunos inconvenientes de esta modalidad es la vulnerabilidad del nodo CCE y la existencia de inconsistencia en las tablas con los cambios de topología.
- B. Aislado.** En esta modalidad las decisiones de encaminamiento sólo se basan en información propia de cada nodo.

Un ejemplo de algoritmo que emplea esta estrategia es el *Algoritmo de la patata caliente de Baran*. Los nodos encaminan los paquetes hacia enlaces de salida con longitud de cola de espera más corta, equilibrando así la carga en los enlaces.

- C. Distribuido.** Esta modalidad es una mezcla de las dos anteriores. La elección de la ruta del paquete al destino se realiza obteniendo la ruta menos congestionada en una determinada vecindad de nodos, de forma que se optimizan las rutas en una determinada vecindad.

La elección de la ruta es uno de los aspectos más importantes en la descripción de un algoritmo de encaminamiento. La elección de la misma se realiza en base a un criterio que permita un alto rendimiento en la red. Atendiendo a la naturaleza del criterio para la elección de la ruta existen diversos tipos de algoritmos que se analizarán a continuación. Dentro del gran número de criterios y algoritmos existentes, se han seleccionado aquellos que se emplean en las redes de comunicaciones actuales más frecuentes.

10.2.2 Algoritmo de enrutamiento por la trayectoria más corta

Este criterio de selección de ruta, la más corta, es el más simple de todos los existentes. La filosofía consiste en determinar cual es la ruta más corta entre un par de nodos origen y destino, es decir determinar el camino en la subred por el que el paquete tardará menos tiempo en llegar a su destino. Una métrica posible de la longitud de la trayectoria puede ser la longitud existente (metros) entre el par de nodos, pero no siempre es una medida correcta. El tiempo que tardará el paquete en llegar a su destino no sólo depende de la distancia entre ambos, sino que también de la velocidad de propagación de la señal en el medio. De esta forma, es posible que dos nodos cercanos físicamente entre sí tengan un retardo en el envío de un paquete mayor que otros nodos más alejados.

Frecuentemente el término 'trayectoria más corta' se sustituye por el de 'trayectoria de menor coste', donde la distancia entre dos nodos se mide evaluando la distancia física, la velocidad de transmisión, el nivel de congestión, etc. entre un par de nodos. Así, en una subred se determinarán los caminos de coste mínimo entre cada par de nodos para obtener el menor retardo posible en la transmisión de información en la subred.

Existen diversos algoritmos que obtienen las rutas más cortas en un grafo asociado a una subred, y en este apartado se estudiará el algoritmo de Dijkstra (1959) para la búsqueda de las trayectorias más cortas en un grafo donde cada enlace entre nodos tiene un coste.

Algoritmo de Dijkstra

Dada una subred formada por nodos conectados por enlaces bidireccionales que tiene asociado un coste en cada sentido, se define el **coste de una ruta** entre nodos como la suma de los costes de los enlaces que los unen. El algoritmo de Dijkstra determina la ruta de menor coste entre cada par de nodos. Por simplicidad, consideraremos el mismo coste en ambos sentidos en cada enlace.

El algoritmo progresó en etapas, de forma que en la etapa k han sido determinadas las rutas de menor coste a los k nodos más próximos. Estos nodos se añaden a un conjunto M . En la etapa $k+1$, el nodo que no está en el conjunto M y tiene el menor coste desde el nodo F (nodo fuente) se añade a M , y así hasta que todos los nodos de la subred estén en el conjunto M .

El algoritmo se formaliza en un conjunto de pasos para los que se empleará la siguiente notación:

N: Conjunto de nodos de la subred.

F: Nodo fuente (uno cualquiera de la red, a partir del que se iniciará el algoritmo).

M: Conjunto de nodos incorporados por el algoritmo en cada una de las etapas.

c(i,j): Coste del enlace del nodo **i** al **j** (este valor será ∞ si no están conectados).

C₁(n): Coste de la ruta de coste mínimo del nodo F al nodo **n** obtenida por el algoritmo en una determinada etapa.

Los pasos del algoritmo se describen a continuación:

1. $M = \{F\}$. Para cada nodo $n \in N - \{F\}$ se inicializará $C_1(n)$ como $c(F,n)$.
2. Hallar $w \in N - M$ tal que $C_1(w)$ sea mínimo e incluirlo en M . A continuación, determinar:

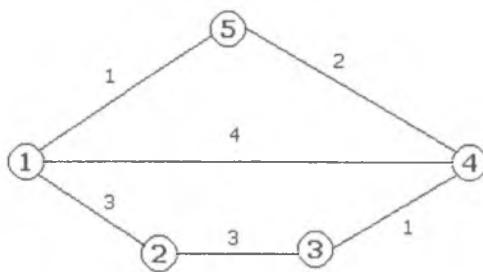
$$C_1(n) = \text{mínimo } \{C_1(n), C_1(w) + c(w,n)\} \text{ para cada nodo } n \in \{N - M\}$$

Si el término $C_1(w) + c(w,n)$ es el mínimo, se actualizará la ruta de F a n, que será ahora la ruta de F a w concatenada con la ruta de w a n.

3. Repetir el **paso 2** hasta que $M = N$.

Se ilustrará la aplicación del algoritmo en el ejemplo que se acompaña a continuación.

Sea una subred con una estructura de nodos y costes entre los mismos en la forma de la siguiente figura.



Tomaremos como $F=1$ y se mostrarán las diversas iteraciones del algoritmo, indicando las rutas de menor coste del nodo F al resto y su coste asociado.

Iteración	M	$C_I(2)$	Ruta	$C_I(3)$	Ruta	$C_I(4)$	Ruta	$C_I(5)$	Ruta
1	{1}	3	1-2	∞	-	4	1-4	1	1-5
2	{1,5}	3	1-2	∞	-	3	1-5-4	1	1-5
3	{1,5,2}	3	1-2	6	1-2-3	3	1-5-4	1	1-5
4	{1,5,2,4}	3	1-2	4	1-5-4-3	3	1-5-4	1	1-5
5	{1,5,2,4,3}	3	1-2	4	1-5-4-3	3	1-5-4	1	1-5

El algoritmo de Dijkstra se emplea frecuentemente en estrategias de encaminamiento centralizado, pero será estable siempre que la topología y los costes de los enlaces sea estática, pues si el coste varía con el tráfico se producirán inestabilidades.

10.2.3 Algoritmo de enrutamiento por vector de distancia

Los algoritmos de encaminamiento o enrutamiento por vector de distancia operan haciendo que cada encaminador mantenga una tabla que da la mejor distancia conocida a cada destino y la línea a usar para llegar allí. Estas tablas se actualizan intercambiando información con los vecinos.

El algoritmo de enrutamiento por vector de distancia a veces recibe otros nombres, destacando el algoritmo de enrutamiento *Bellman-Ford* distribuido y el algoritmo *Ford-Fulkerson*, por los investigadores que lo desarrollaron (Bellman en 1957, y Ford y Fulkerson en 1962). Este fue el algoritmo de encaminamiento original de ARPANET y se emplea en Internet con el nombre de RIP.

En el enrutamiento por vector de distancia, cada encaminador mantiene una tabla de encaminamiento indexada por, y conteniendo un registro de, cada encaminador de la subred. Esta entrada comprende dos partes: la línea preferida de salida hacia ese destino y una estimación del tiempo o distancia a ese destino. La métrica usada suele ser el número de saltos o el retardo de propagación en milisegundos.

Se supone que el encaminador conoce la 'distancia' a cada uno de sus vecinos. Cada cierto tiempo, cada encaminador envía a sus encaminadores vecinos la tabla con la distancia a cada uno de los destinos, de forma que éstos pueden actualizar sus tablas o no. Si en la tabla procedente de otro encaminador existe una ruta a un destino con menor distancia que la actual, se procederá a la actualización. De esta forma se realiza una propagación de la información de encaminamiento a través de los nodos de la subred, alcanzando una solución óptima. Sin embargo, aunque este algoritmo converge, puede hacerlo lentamente y ello produce que puedan darse situaciones de congestión.

10.2.4 Algoritmo de enrutamiento por estado del enlace

Este algoritmo es una variante del algoritmo de enrutamiento por vector de distancia. Además de la métrica de la distancia se considera el ancho de banda de las rutas. Este es un aspecto importante para evitar la congestión en la red. Al considerar el retardo en el envío de un paquete se tiene en cuenta el retardo de propagación de la información. Sin embargo, este factor por si sólo no indica que una línea tenga menor coste que otra.

Supóngase un par de rutas que presentan un retardo similar en la propagación de los bits de un paquete de información (rutas de la misma longitud física y mismo tipo de medio de transmisión). El coste de ambas rutas podría ser el mismo en un principio, pero si las velocidades de transmisión son diferentes esto no es así. La ruta que presenta una mayor velocidad de transmisión puede transmitir paquetes al medio más rápidamente que la otra, por lo que presentará menor longitud de cola de espera de paquetes para salida del nodo y, en definitiva, una menor congestión. Por tanto, esta ruta de mayor velocidad de transmisión debe tener un coste menor que la de menor velocidad, pues se producirá menor congestión en la subred si se selecciona ésta.

El algoritmo de enrutamiento por estado del enlace considera el ancho de banda de los enlaces para determinar su coste y puede describirse en un conjunto de 5 pasos:

1. Descubrir a sus vecinos y conocer sus direcciones de red.
2. Medir el retardo o costo para cada uno de sus vecinos.
3. Construir un paquete que indique todo lo que acaba de aprender.
4. Enviar este paquete a todos los demás enrutadores.
5. Calcular la trayectoria más corta a todos los demás enrutadores.

10.2.5 Algoritmo de enrutamiento jerárquico

A medida que crecen en tamaño las redes, crecen proporcionalmente las tablas de encaminamiento del enrutador. Las tablas que siempre crecen no sólo consumen memoria del enrutador, sino que también se necesita más tiempo de CPU para examinarlas y más ancho de banda para enviar informes de estado entre enrutadores. En cierto momento, la red puede crecer hasta el punto en que ya no es factible que cada encaminador tenga una entrada para cada uno de los demás encaminadores, por lo que el enrutamiento tendrá que hacerse jerárquicamente, como ocurre en la red telefónica.

Al emplearse el enrutamiento jerárquico, los enrutadores se dividen en lo que llamaremos **regiones**, donde cada encaminador conoce todos los detalles de la manera de encaminar o enrutar paquetes a destinos dentro de su propia región, pero no sabe nada de la estructura interna de las otras regiones. Al interconectar diferentes redes, es natural considerar cada una como una región independiente, a fin de liberar a los enrutadores de una red de la necesidad de conocer la estructura topológica de las demás.

En las redes enormes puede ser insuficiente una jerarquía de dos niveles; puede ser necesario agrupar las regiones en cúmulos, los cúmulos en zonas, las zonas en grupos, etc., hasta que se nos agoten los nombres para los agregados.

Desafortunadamente, estas ganancias de espacio en las tablas de encaminamiento no son gratuitas. Se paga un precio, y este precio adopta la forma de rutas de longitud mayores, ya que no se conoce la globalidad de la red y las rutas son óptimas en una región determinada. Al volverse muy grande una sola red surge una pregunta interesante: ¿cuántos niveles debe tener la jerarquía? Por ejemplo, considérese una subred con 720 enrutadores. Si no hay jerarquía, cada encaminador necesita 720 tablas de encaminamiento. Si partimos la subred en 24 regiones de 30 enrutadores cada una, cada enrutador necesitará 30 entradas locales más 23 entradas remotas, en total 53 entradas por encaminador. Kamoun y Kleinrock (1979) demostraron que el número óptimo de niveles para una subred de

enrutadores es de $\ln N$, donde N es el número de nodos enrutadores, requiriéndose un total de $e \cdot \ln N$ entradas por enrutador. También demostraron que el aumento en la longitud media efectiva de ruta causada por el enrutamiento jerárquico es lo bastante pequeña como para ser generalmente aceptable.

10.3 ALGORITMOS DE CONTROL DE LA CONGESTIÓN

Cuando hay demasiados paquetes presentes en la subred (o en una parte de ella), hay una degradación del desempeño. Esta situación se denomina **congestionamiento**. En la **figura 10.1** se muestra este síntoma.

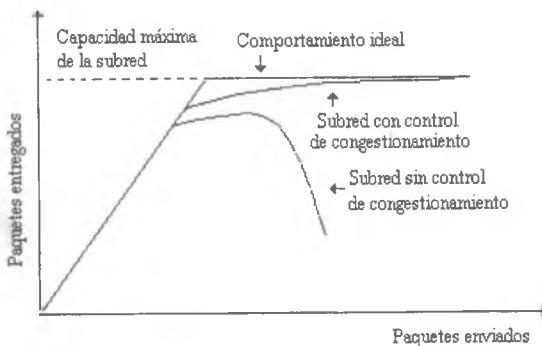


Figura 10.1 Efecto del congestionamiento en el rendimiento de una subred.

Cuando la cantidad de paquetes descargados en la subred por los hosts está dentro de su capacidad de conducción, todos se entregan (excepto unos pocos debido a errores de transmisión) y la cantidad entregada es proporcional al número enviado. Sin embargo, a medida que aumenta el tráfico, los enrutadores ya no pueden manejarlo y comienzan a perder paquetes. Esto tiende a empeorar las cosas. A muy alto tráfico, el rendimiento se desploma por completo y casi no hay entrega de paquetes, produciéndose el bloqueo de la subred.

El congestionamiento puede ocurrir por varias razones, entre las que destacan:

- Si repentinamente comienzan a llegar paquetes por varias líneas de entrada al nodo encaminador y todos necesitan la misma línea de salida, se generará una cola. Si no hay suficiente memoria los paquetes se pierden y tienen que ser reenviados. Incluso si la

memoria es suficiente, es posible que el tiempo de espera de un paquete expire en la cola y sea reenviado, por lo que aumentará el tráfico en la red.

- b) Los procesadores lentos también pueden causar congestionamientos. Si las CPU de los enrutadores son lentas para llevar a cabo las tareas de administración requeridas, pueden alargarse las colas, aun cuando haya un exceso de capacidad de línea.
- c) Otro de los aspectos que afecta al congestionamiento es la capacidad de la línea. La tasa de salida de información del nodo afecta al congestionamiento, de forma que si es baja aumenta el número de paquetes en espera en la cola de salida.

Los diferentes algoritmos de control del congestionamiento pueden clasificarse en dos tipos: de ciclo abierto y ciclo cerrado. Los algoritmos de **ciclo abierto** realizan un control del congestionamiento basándose en un buen diseño de la subred, para asegurarse que no ocurra desde un principio. Los algoritmos de **ciclo cerrado** se basan en el concepto de un ciclo de realimentación, en el que el sistema detecta cuándo y dónde ocurren los congestionamientos, pasan esa información a los lugares donde pueden tomarse acciones y ajustan el funcionamiento del sistema para corregir el problema.

En este capítulo se estudiarán tres estrategias de control del congestionamiento basadas en el concepto de ciclo de realimentación.

10.3.1 Algoritmo de control del flujo

Éste es un mecanismo empleado frecuentemente para el control de la congestión. En esta técnica existen diferentes niveles de control:

1. Control del flujo en el acceso a la red (**CFAR**).
2. Control del flujo local (**CFL**).
3. Control del flujo entre nodo origen y nodo destino (**CFNO-ND**).
4. Control del flujo extremo a extremo, entre transmisor y receptor (**CFEE**).

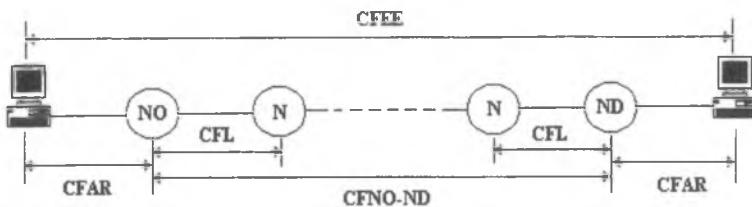


Figura 10.2 Control del congestionamiento empleando control del flujo

El control del flujo en el acceso a la red **CFAR** tiene como objetivo disminuir la entrada de tráfico externo a la subred basándose en las indicaciones de la congestión interna de la subred.

El objetivo del control del flujo local **CFL** es evitar la congestión que se produce entre pares de nodos de conmutación de la red. El procedimiento sólo tiene en cuenta la situación local, sin saber qué ocurre a nivel global en la red.

El control del flujo entre nodo origen y nodo destino **CFNO-ND** tiene como objetivo evitar la congestión que pueda producirse en el nodo de destino de la red cuando los enlaces de salida del origen generan una tasa de tráfico mayor que la que los terminales de destino están absorbiendo.

Por último, el control del flujo de extremo a extremo **CFEE** realiza un control del flujo del tráfico entre el terminal de origen y de destino, apoyándose en el protocolo de nivel de transporte.

10.3.2 Algoritmo de control de admisión

Esta técnica está diseñada para su empleo en subredes de conmutación de paquetes basadas en circuitos virtuales. La técnica pretende evitar que empeoren los congestionamientos que ya han empezado a producirse y la idea es sencilla: una vez que se ha detectado el congestionamiento no se establecen circuitos virtuales nuevos hasta que ha desaparecido el problema. Por tanto, fallan los intentos de establecer conexiones nuevas de la capa de transporte. Permitir el acceso a más gente simplemente empeoraría las cosas. Aunque este enfoque es burdo, su implementación es sencilla. En el sistema telefónico, al sobrecargarse un conmutador también se pone en práctica el control de admisión, al no proporcionarse tonos de llamada.

Un enfoque alternativo es permitir el establecimiento de nuevos circuitos virtuales, pero enrutando los nuevos circuitos por otras rutas que no presenten problemas.

10.3.3 Paquetes de estrangulamiento

Este enfoque puede usarse tanto en circuitos virtuales como en subredes de datagramas. Cada enrutador puede supervisar fácilmente el uso de sus líneas de salida y de otros recursos. Por ejemplo, puede asociar a cada línea de salida una variable real u , que refleja el uso reciente de esa línea. Siempre que u rebasa un cierto valor de umbral, la línea de salida entra en un estado de **advertencia**. Cada paquete nuevo que llega se revisa para ver si su línea de salida está en el estado de advertencia. De ser así, el enrutador envía un **paquete de estrangulamiento** de regreso al enrutador anterior, dándole como dirección de destino la del paquete original. El paquete original se marca (se enciende un bit de cabecera) para que no pueda generar más paquetes de estrangulamiento más adelante en la trayectoria, y se reenvía de la manera normal.

Al recibir el paquete de estrangulamiento el enrutador anterior, está obligado a reducir el tráfico que envía al destino especificado en un determinado porcentaje. Dado que otros paquetes destinados al mismo lugar ya están en camino y generarán nuevos paquetes de estrangulamiento, el enrutador deberá ignorar los paquetes que se refieran a ese destino durante un intervalo de tiempo fijo.

10.4 ANÁLISIS DE CONGESTIÓN DE REDES. TEORÍA DE COLAS

El funcionamiento de cualquier nodo en una red de conmutación de paquetes está basado en la llegada de paquetes por una línea de entrada, su almacenamiento temporal en una cola de entrada, el procesado del mismo para determinar la línea de salida y el envío a través de la misma. Este comportamiento puede modelizarse empleando la teoría de colas.

10.4.1 Modelo probabilístico del comportamiento de un nodo de la red

El tiempo que un paquete permanece en un nodo de la red depende del tiempo de procesamiento, la longitud del paquete, la capacidad de transmisión del enlace y la forma en que llegan los paquetes. Para estudiar

estos procesos en términos probabilísticos se emplean procesos de Poisson, de forma que un enrutador de una subred tiene la estructura siguiente.

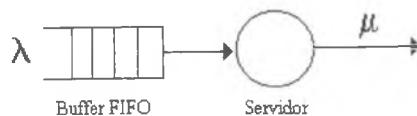


Figura 10.3 Estructura de un nodo de la subred

En este esquema, λ es la tasa de promedio de entrada al nodo en paquetes/seg y μ es la tasa promedio de salida del nodo en paquetes/seg. Los paquetes llegarán al nodo a una velocidad de λ paquetes/seg, serán almacenados en un área temporal y serán atendidos a una velocidad de μ paquetes/seg. Además es posible que exista más de un proceso servidor en el nodo realizando el procesamiento de paquetes.

Para que el nodo encaminador de la subred pueda modelarse empleando procesos de Poisson, deben cumplirse tres condiciones referentes al comportamiento probabilístico del nodo:

- La probabilidad de llegada de un paquete en el intervalo Δt viene dada por:

$$p(1) = \lambda \cdot \Delta t + O(\Delta t)$$

donde $p(1)$ es la probabilidad y $O(\Delta t)$ es un término del orden de Δt , de forma si $\Delta t \rightarrow 0$ entonces $O(\Delta t) \rightarrow 0$.

- La probabilidad de la llegada de ningún paquete en el intervalo Δt viene dada por:

$$p(0) = 1 - \lambda \cdot \Delta t + O(\Delta t)$$

- La llegada de paquetes es un proceso sin memoria, cada llegada en un intervalo de tiempo es independiente de eventos en otros intervalos.

De esta forma, la probabilidad de la llegada de k paquetes en un tiempo t viene dada por la expresión:

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{donde } k \geq 0 \quad \text{y} \quad t \geq 0.$$

A continuación se analizarán los diferentes tipos de colas regidos con la función de probabilidad anterior y que permiten modelar distintos tipos de nodos en una red de conmutación de paquetes. Los tipos de colas se notarán como **A/B/m**, donde **A** y **B** hacen referencia al tipo de función de probabilidad de llegada y salida de paquetes del nodo respectivamente. En nuestro caso se analizarán funciones de probabilidad de tipo Poisson, que se denotarán por **M** (Markov). El término **m** hace referencia al número de servidores que atienden la cola.

10.4.2 La cola M/M/1

Este tipo de cola caracteriza a un nodo que presenta funciones de probabilidad de llegada y salida de paquetes de tipo Poisson y tiene un solo servidor para procesar los paquetes. Se denotará por λ la tasa promedio de llegada de paquetes a la cola del servidor y por μ la tasa promedio de salida de paquetes del servidor.

Se distinguirán dos casos: la cola M/M/1 con tamaño de *buffer* de entrada infinito y finito.

Cola infinita

En el caso del modelo con cola infinita, todos los paquetes que lleguen al nodo serán almacenados en el *buffer* FIFO de entrada. Puede demostrarse que, si la tasa de llegada de paquetes al nodo es λ , la tasa de salida de paquetes del nodo es μ y p_n es la probabilidad de que existan n paquetes en la cola del nodo, se cumple la relación:

$$\lambda \cdot p_n = \mu \cdot p_{n+1} \quad (1)$$

cuando la cola del sistema alcanza el estado estacionario.

A partir de la expresión (1) se puede deducir:

$$p_n = (1 - \rho) \cdot \rho^n$$

donde ρ es la **intensidad de tráfico** y se expresa como $\rho = \frac{\lambda}{\mu}$.

De la expresión (1) se deduce que para que la cola sea estable conforme aumente el número de paquetes ($n \rightarrow \infty$), debe verificarse que $\rho < 1$.

A partir de la expresión (1) también puede obtenerse el valor de **la ocupación promedio de la cola (E(n))**, que indica el número de paquetes que hay presentes en la cola por término medio, y que viene dado por:

$$E(n) = \sum_{n=0}^{\infty} n \cdot p_n = \frac{\rho}{1 - \rho}$$

La ocupación promedio de la cola está relacionada con **el retardo medio en la cola para un paquete (E(T))**, que vendrá dado por:

$$E(T) = \frac{E(n)}{\lambda}$$

Un aspecto interesante en la modelización de la cola es el cálculo del **rendimiento (γ)** de la misma, que se define como **el número de usuarios atendidos por unidad de tiempo**. Podemos distinguir entre el rendimiento en la entrada de la cola y la salida de la cola, por lo que se tendrán las expresiones:

$$\begin{aligned}\gamma_{IN} &= \lambda \\ \gamma_{OUT} &= \mu \cdot (1 - p_0)\end{aligned}$$

Cola finita

Para el modelo de cola M/M/1 con longitud de cola finita, se considerará un *buffer* temporal de almacenamiento FIFO de capacidad N paquetes. En estas condiciones, puede demostrarse que la probabilidad de que haya n paquetes en la cola de entrada vendrá dada por:

$$p_n = \frac{(1 - \rho) \cdot \rho^n}{1 - \rho^{N+1}}$$

A partir de esta expresión se deduce que la **probabilidad de bloqueo de la cola** vendrá dada por la probabilidad de que haya N paquetes en la misma, por lo que:

$$p_B = \frac{(1 - \rho) \cdot \rho^N}{1 - \rho^{N+1}}$$

Las expresiones para el retardo medio de la cola y la ocupación promedio de la cola para el modelo de cola finita son más complejas que el de cola infinita, pero los resultados para la cola finita pueden aproximarse bastante bien para una cola de tamaño suficientemente grande.

El cálculo del rendimiento para la cola finita se realiza como:

$$\gamma_{IN} = \lambda \cdot (1 - p_B)$$

$$\gamma_{OUT} = \mu \cdot (1 - p_0)$$

10.4.3 Modelos de colas dependientes del estado (M/M/m)

Dentro de este tipo de colas, que se caracterizan porque la tasas de llegada y salida de paquetes del nodo no es constante sino que varía y depende del estado del sistema, se analizarán los modelos de la cola M/M/2 , M/M/ ∞ y la cola de desaliento.

Cola M/M/2

Este tipo de cola se caracteriza porque existen dos procesos servidores en el nodo, una cola de entrada y dos líneas de salidas.

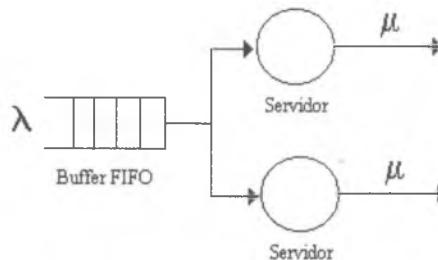


Figura 10.4 Modelo de cola M/M/2

En este caso supondremos por simplicidad que la cola de entrada tiene tamaño infinito, y en ese caso, la intensidad de tráfico puede aproximarse por:

$$\rho = \frac{\lambda}{2\mu}$$

donde λ y μ son las tasas promedios de llegada y salida de paquetes respectivamente.

En estas condiciones, la ocupación promedio de la cola vendrá dada por:

$$E(n) = \frac{2\rho}{1 - \rho^2}$$

Cola M/M/ ∞

Este tipo de cola se caracteriza porque existen suficientes servidores para atender cualquier paquete que llegue a la cola, por lo que el número de líneas de salida del nodo es igual al número de paquetes. De esta forma, la tasa de salida está relacionada con el número de paquetes en la cola en la forma:

$$\mu_n = n \cdot \mu$$

donde μ es la tasa promedio de salida de cada servidor.

Cola de desaliento

La cola de desaliento presenta un único servidor, con tasa promedio de salida μ y tasa promedio de llegada según la relación:

$$\lambda_n = \frac{\lambda}{n+1}$$

donde λ es una tasa de llegada de paquetes fija para el nodo (y será la máxima que presente éste).

En este modelo, conforme aumenta el número de paquetes en la cola disminuye la tasa de llegada de paquetes, debido a que estos son desviados o la cola se bloquea (cuando ésta es finita).

10.4.4 Redes de colas

Algunos sistemas de colas están compuestos por un número de subsistemas ligados en red. En tales sistemas los paquetes reciben servicio en más de un subsistema (o no) antes de ser atendidos completamente. Normalmente, un paquete inicia su peregrinación en un nodo dado, espera su turno en la cola y cuando es atendido va hacia otro nodo para obtener otro servicio más, etc. Estos sistemas son llamados **redes de colas**. De manera informal, una red de colas es un conjunto de nodos interconectados entre sí. Cada nodo posee uno o más servidores y un lugar para que los paquetes formen la(s) cola(s). Existen **redes abiertas** donde los paquetes entran en el sistema y eventualmente salen del sistema; **redes cerradas**, las cuales tienen un número fijo de paquetes circulando por la red sin salir y, desde el punto de vista del sistema, es como si no hubiese ni entrada ni salida de paquetes, y **redes mixtas**, que son abiertas para ciertas clases de paquetes y cerradas para otros.



TEMA 11. PROTOCOLO DE NIVEL DE RED IP

11.1 FUNDAMENTOS DEL PROTOCOLO IP

El protocolo IP es el protocolo de nivel de red empleado en Internet. Internet es la expansión mundial de la originaria ARPANET. Es a principios de los 90 cuando Internet tiene su mayor auge, y debido a que presenta una serie de ventajas que lo propician:

1. Una arquitectura de red que puede emplear cualquier tipo de medio físico para la transmisión de los paquetes de información, pues la arquitectura de Internet (la arquitectura TCP/IP) es independiente de los niveles de enlace y físico, no necesitando de un conjunto de primitivas de nivel inferior determinadas (como si sucede en el modelo OSI).
2. Existencia de un conjunto de protocolos de nivel de aplicación con grandes beneficios para la comunicación de información: servidores Web, correo electrónico, servidores FTP, servidores ARCHIE de búsqueda de ficheros, etc.

Las características anteriores hacen que la migración de las redes LAN a una arquitectura TCP/IP no sea compleja y sea factible la conexión de cada vez más redes a la Internet mundial.

El protocolo IP trabaja sobre una subred de conmutación de paquetes basada en datagramas, lo que permite una tolerancia a fallos debido a que no existen caminos preestablecidos para los paquetes en la red. La información procedente del nivel de transporte es fragmentada en paquetes con un formato determinado (paquetes IP) que llegarán a su destino al ser encaminados por los nodos de la subred.

Todo protocolo de nivel de red debe definir un esquema de direccionamiento de los nodos de una subred para poder realizar el proceso de encaminamiento. El protocolo IP identifica cada nodo o host de Internet por una **dirección IP**, que es única para cada máquina.

Direcciones IP

Las direcciones IP tienen una longitud de 32 bits, organizadas en 4 grupos de 8 bits cada uno. Se dividen fundamentalmente en dos partes: la porción de la red y la porción de la máquina.

La porción de red identifica a un grupo de máquinas que comparten el mismo protocolo de enlace dentro de un medio físico. El campo de máquina hace referencia a todas aquellas estaciones conectadas a la misma red.

El tamaño de cada parte dentro de la dirección IP depende del valor de los bits de mayor peso, tal y como se muestra en la siguiente tabla.

Clase	7bit	24bit	
A	0 Red	Máquina	0.0.0.0 127.255.255.255
B	1 0 Red	Máquina	128.0.0.0 191.255.255.255
C	1 1 0 Red	Máquina	192.0.0.0 223.255.255.255
D	1 1 1 0 Multicast		224.0.0.0 239.255.255.255
E	1 1 1 1 0 Futuras Ampliaciones		240.0.0.0 247.255.255.255

Figura 11.1 Direccionamiento IP y clases de direcciones.

De aquí surge una clasificación en 5 tipos de redes en función del contenido de cada uno de los campos de dirección.

De esta forma, se logra una mayor optimización en las tablas de encaminamiento de los encaminadores o *routers*, puesto que únicamente tienen que localizar la porción de la red a la hora de encaminar un datagrama.

Dentro del direccionamiento IP, existe una **dirección de broadcast** definida con todos los bits a 1 correspondientes a la porción de máquina. Es decir, la dirección 134.215.255.255 sería una dirección de broadcast perteneciente a la red 134.215. Esta dirección tiene el mismo objetivo que la dirección de difusión para redes LAN analizada en el capítulo 1, el envío

de un paquete de datos a todas las máquinas de una red. En el protocolo IP las redes también poseen direcciones, que se obtienen con todos los bits de la porción de máquina a 0. Continuando con el ejemplo anterior, la dirección 134.215.0.0 correspondería a la dirección IP de la red 134.215.

Cada interfaz IP situado dentro de una misma máquina, tiene una dirección propia IP. Significa que si tuviéramos un nodo en la subred con dos adaptadores de red (dos tarjetas de red), éste presentaría dos direcciones IP. Podríamos acceder a él a través de cualquiera de ellas siempre que sus tablas de encaminamiento lo permitiesen.

La máscara de subred

Para que un nodo de Internet actúe como encaminador precisa interconectar dos o más redes IP. Para ello dispondrá de un interfaz por cada red a la que esté conectado, y cada interfaz debe soportar el protocolo IP. Para configurar el interfaz se necesita como mínimo dos parámetros: la dirección IP y su máscara asociada.

La máscara se compone de un valor de 32 bits, donde los n primeros bits están puestos a valor '1' y los $32-n$ bits restantes a '0'. Estos se superponen bit a bit a la dirección IP de tal forma que aquellos cuyo valor es '1', indican que la porción correspondiente a la dirección, es la parte de red. El valor '0', señala la parte de máquina. Lógicamente, existe siempre una máscara por defecto asociada a la dirección IP, en función de la clase.

Por ejemplo, la dirección 10.2.45.1 pertenece a la red 10.0.0.0 de clase A. Su máscara por defecto deberá ser 255.0.0.0 en notación decimal o 11111111.00000000.00000000.00000000 en notación binaria.

En un único segmento de subred (por ejemplo una red Ethernet) el esquema de direccionamiento resulta muy sencillo. Todas las máquinas conectadas llevarían la máscara 255.0.0.0 y se numerarían 10.2.45.1, 10.7.23.124, 10.0.12.253 ,etc., manteniendo la porción de la red siempre igual a 10. Se dispondría por tanto de $2^{24} - 2$ máquinas direccionables: la dirección de broadcast 10.255.255.255 y la dirección de la red 10.0.0.0 no son válidas para numerar máquinas.

Si en vez de disponer de un único segmento de red dispusiéramos de varios a interconectar entre sí, se necesitaría ampliar la máscara de red como mínimo en 2 bits más para poder así direccionar 4 subredes. De este modo se tendría una máscara de 11111111.11000000.00000000.00000000 o 255.192.0.0. En este caso las subredes serían:

00001010.00000000.00000000.00000000 ó 10.0.0.0
 00001010.01000000.00000000.00000000 ó 10.64.0.0
 00001010.10000000.00000000.00000000 ó 10.128.0.0
 00001010.11000000.00000000.00000000 ó 10.192.0.0

El número de máquinas por cada una de estas subredes sería $2^{22} - 2$. Por tanto, cada vez que se amplía la máscara, se pierden 2 direcciones IP en cada subred (dirección de broadcast y subred).

Resumiendo, hemos considerado que una dirección IP está compuesta de dos identificadores, uno para la red y otro para la máquina. El ámbito de cada uno de ellos depende de la clase a la que pertenece esa dirección.

Formato del paquete IP

El formato del datagrama IP queda representado en la figura siguiente. Su tamaño es de 20 bytes, a no ser que presente datos en el campo **opciones**.

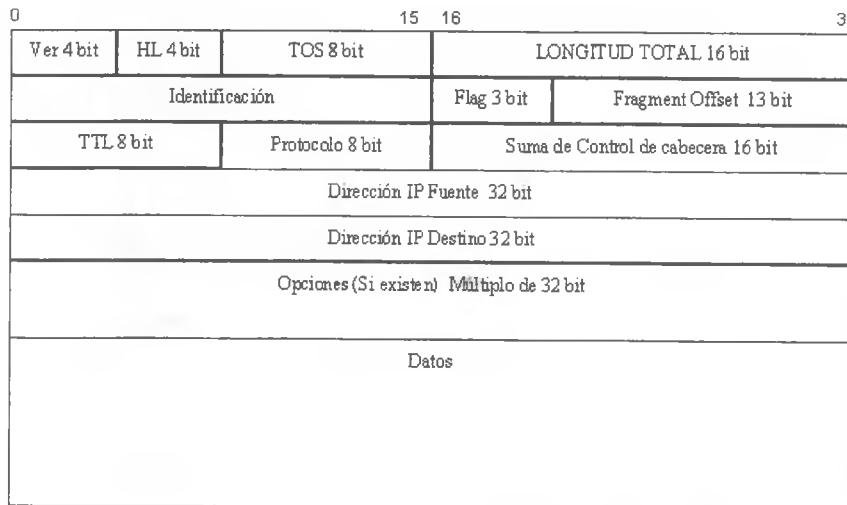


Figura 11.2 Formato de la cabecera del protocolo IP.

El bit más significativo está marcado como 0 en el lado izquierdo, mientras que el menos significativo de la palabra de 32 bits se etiqueta como 31 en el lado derecho. Los octetos de cada palabra de 32 bits se transmiten empezando por el 0 hasta el 31.

La versión en curso actualmente es la 4 también conocida como **IPv4** aunque ya ha comenzado a ser operativa la **IPv6**. El campo **Ver** con 4 bits de longitud transporta esta información.

El campo **HL** (*header length*) indica el número de palabras de 32 bits que componen la cabecera, incluyendo las opciones eventuales. Puesto que su tamaño es de 4 bits, tendremos que: 15×32 bits por palabra son 60 bytes de longitud máxima en la cabecera IP. Este campo incorpora habitualmente el valor 5 (cuando no existen opciones).

TOS (Type of service) indica el tipo de servicio. Actualmente los 3 primeros bits son ignorados, los 4 siguientes identifican el servicio (datos o voz digitalizada) y el último está inutilizado y su valor debe ser siempre 0.

El campo **Longitud Total** contiene el tamaño en octetos del datagrama IP. Gracias a él y al campo **HL** podemos conocer donde empieza y termina la porción de datos. Como utiliza 16 bits, se puede deducir que el tamaño máximo o MTU de un datagrama IP será de 65535 bytes .

El mecanismo de fragmentación utilizado por IP emplea los siguientes 3 campos. El primero, **Identificación**, permite marcar de forma única cada datagrama enviado por una máquina. Se incrementa normalmente en cada nuevo envío. Cuando se produce una fragmentación, este valor es copiado en cada uno de los trozos o fragmentos que componen el datagrama original. El campo **flag** de 3 bits, activa entonces uno de ellos (el número 2) conocido como **More fragments** tomando el valor 1 en todos los trozos excepto en el último. El campo **Frame Offset** contiene el índice del fragmento a partir del datagrama original. Además, el nuevo campo **Longitud Total** de cada fragmento es actualizado a su nuevo valor.

Existe un bit (el número 1) en el campo **flag** conocido como **Don't fragment**. Si está activado a 1, IP no producirá ninguna fragmentación eliminando el datagrama y enviando un mensaje de error a la máquina de origen, empleando mensajes del protocolo ICMP (*internet control message protocol*, protocolo de mensajes de control de interred). El protocolo ICMP se describe en los documentos RFC (*request for comments*) del IETF (*internet engineering task effort*, fuerza de trabajo de ingeniería en Internet). En concreto el documento se denomina RFC 792.

Para evitar que un datagrama quede atrapado en algún bucle dentro de la red (problemas con los protocolos de encaminamiento, por ejemplo) existe un tiempo de vida representado mediante el campo **TTL** (Time to

Live). Se inicializa a un cierto valor por el remitente y se decrementa en una unidad por cada *router* o encaminador que atraviesa. Cuando alcanza el valor 0, el datagrama se elimina y un mensaje ICMP es enviado a la fuente indicando el suceso.

IP identifica el protocolo de nivel superior (TCP, UDP, ICMP ...) al cual debe hacer llegar la información a través del campo **Protocolo**.

La **Suma de Control** abarca únicamente la cabecera IP. Se calcula como una suma sin acarreo sobre 16 bits, de todos los bytes que componen la cabecera IP considerándolos como una secuencia de palabras de 16 bits. Sin embargo, otros protocolos como TCP, UDP, ICMP utilizan códigos de redundancia cíclica (CRC) basados en algoritmos más sofisticados. El motivo es claro: un *router* debe procesar grandes cantidades de paquetes por unidad de tiempo. Generalmente, el único valor que modifica a cada datagrama es el TTL, decrementándolo en una unidad. El cálculo de la nueva suma de control puede ser realizado de forma incremental disminuyendo drásticamente el tiempo de proceso de cada datagrama por los encaminadores intermedios.

Cada datagrama contiene la **dirección IP** de la máquina que envió el paquete y a la que va dirigido, para que los encaminadores puedan hacer llegar el paquete a su destino.

El campo **Opciones** es una lista de longitud variable con información específica del datagrama.

11.2 INTERCONEXIÓN DE REDES EMPLEANDO IP

El esquema de direccionamiento que proporciona el protocolo IP permite realizar la interconexión de máquinas situadas en segmentos de red distintos. A cada segmento físico con el mismo nivel de enlace se le asocia una dirección de red IP determinada y se interconecta con otras redes IP empleando encaminadores o *routers*.

Cada uno de los *routers* tiene acceso a dos o más redes IP empleando un interfaz de red para cada una, por lo que dispondrá de más de una dirección IP. El *router* deberá ser capaz de determinar el interfaz por el que enviar los paquetes que le llegan por uno de ellos. Esta decisión (proceso de encaminamiento) se realiza en base a la dirección IP de destino que figurará en el paquete IP a encaminar. A nivel de red se definen diferentes tipos de protocolos de encaminamiento para la búsqueda de las rutas óptimas. Estas rutas se actualizan en una **tabla de encaminamiento** que

tiene el *router*. Esta tabla consta de una serie de entradas, donde para cada posible dirección de red de destino se especifica el interfaz por el que hay que enviar el paquete. En numerosas ocasiones, el destino del paquete no es alguna de las redes a las que está conectado el *router* directamente, y en ese caso el paquete se reenvía a otro *router* que se encuentre en alguna de ellas. Este *router* recibe el nombre de **gateway** o **puerta de enlace** en la tabla de enrutamiento, aunque no hay que confundir esta nomenclatura con el término formal de *gateway*, que es un dispositivo que interconecta redes con distinta arquitectura.

En una tabla de encaminamiento, y para evitar que ésta sea excesivamente grande y el proceso de encaminamiento sea lento, existe generalmente una entrada denominada **default gateway**. Esta entrada indica un *router* por defecto al que enviar todos aquellos paquetes IP cuyo destino no aparece en ninguna entrada de la tabla de encaminamiento. Además de conseguir que así el tiempo de procesamiento de los paquetes en el encaminamiento sea bajo, el *router* no tiene porque conocer toda la estructura de Internet. Éste sólo conoce su entorno cercano y el resto de la red es una *nube* genérica a donde se envían todos los paquetes para que lleguen a su destino a través de otros *routers*.

11.3 PROTOCOLO DE ENCAMINAMIENTO RIP

RIP (Routing Information Protocol) es un protocolo de encaminamiento de vector de distancia, que en la actualidad está en su versión 2. Está definido en los documentos RFC1721 (análisis), RFC1722 (utilidad) y RFC1723 (descripción), en base a lo ya establecido para el protocolo RIP versión 1, definido en el RFC 1058.

RIP se encarga de mantener actualizadas las tablas de los *routers* a través de mensajes de difusión. Se dice que es un protocolo de ‘Vector de distancia’ ya que emplea el número de saltos (o métrica) para decidir que ruta debe ser aplicada para alcanzar un destino dado. El número de saltos es el número de *routers* que debe atravesar un paquete para llegar al destino. Con RIP el máximo número de saltos se sitúa en 16, y por ello es utilizado en redes con dimensiones reducidas en cuanto a número de *routers*. Una métrica de 16 indica el valor infinito.

En cada ruta propagada a los demás *routers* mediante un mensaje RIP se especifica, además del número de saltos para llegar a la red IP referenciada, la dirección IP del siguiente *router* al que pueden ser enviados los paquetes en vez de utilizar el *router* que genera el mensaje. Permite de esta forma la optimización del encaminamiento en la red.

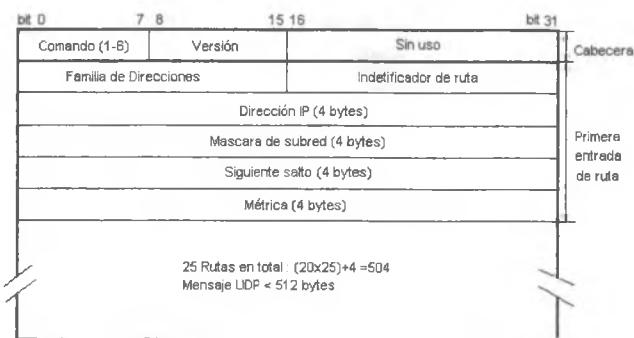


Figura 11.3 Formato del paquete de información del protocolo RIP versión 2.

Por cada ruta dinámica en la tabla de encaminamiento del *router* existe un temporizador asociado. Un sistema con RIP que encuentra una ruta no actualizada desde hace 3 minutos procede a marcarla para su destrucción con el valor infinito (16). La eliminación permanente se retrasa 60 segundos más para asegurarse de que esta acción ha sido notificada al resto de la red con el tiempo suficiente.

11.4 PROTOCOLO DE ENCAMINAMIENTO OSPF Y BGP

Los protocolos **OSPF** (Protocolo de encaminamiento de pasarela interior) y **BGP** (Protocolo de encaminamiento de pasarela exterior) permiten realizar un encaminamiento jerárquico en Internet.

OSPF es un protocolo que permite realizar el encaminamiento de información dentro de lo que se denomina un **sistema autónomo (AS)**: conjunto de redes interconectadas entre sí y gestionadas por un mismo organismo y que está conectado a otros sistemas autónomos. OSPF especifica áreas dentro del AS y permite realizar un encaminamiento entre áreas y dentro de áreas, empleando un encaminamiento dinámico y con diferentes tipos de métricas. OSPF distingue entre:

1. Enrutadores internos que están conectados a una sola área.
2. Enrutadores de borde de área que conectan dos o más áreas.
3. Enrutadores de *backbone* (el área principal dentro del AS) que están en el *backbone* (red troncal).
4. Enrutadores de frontera de AS que intercambian información con los enrutadores de otros AS.

Para realizar el encaminamiento de información entre los distintos sistemas autónomos (AS) se emplea el protocolo BGP. Éste permite

establecer políticas para el encaminamiento de la información entre distintos AS. Mientras que el algoritmo OSPF determina las rutas más eficientes dentro de un AS, el algoritmo BGP determina cual es la política más correcta para el envío de paquetes entre AS, que no tiene porque ser las más rápidas o de mejor rendimiento. Algunos aspectos que se consideran en el protocolo BGP pueden ser consideraciones políticas (no enviar paquetes de información a través de AS de países no afines políticamente), de seguridad o económicas (el tráfico entre dos AS se puede enviar por una ruta más larga pero de menor coste de uso).

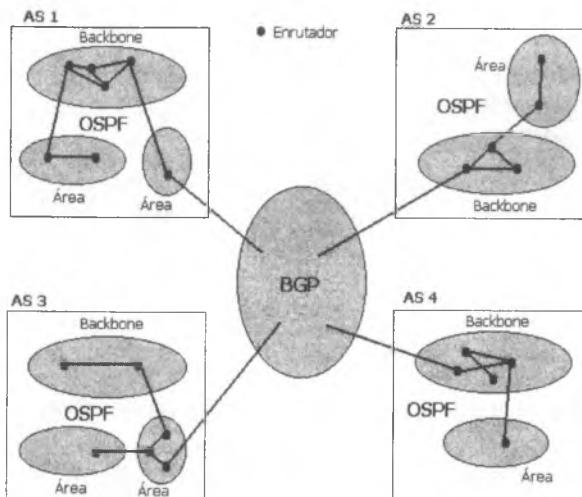


Figura 11.4 Esquema de encaminamiento jerárquico OSPF y BGP.

11.5 TUNNELING

El proceso de tunneling es una situación de intercambio de información a través de redes heterogéneas que cumple con cierta característica. Ésta consiste en que el tipo de subred de la máquina origen del paquete sea el mismo que el de la subred de la máquina de destino. En estas condiciones es posible realizar un proceso mediante el cual el paquete de nivel de red original se encapsula en el paquete de nivel de red de la red intermedia diferente, procediendo a su desencapsulación al llegar a la red de destino: la red intermedia actúa como túnel.

Un ejemplo de túnel muy frecuente se produce cuando se ha de conectar dos redes IP con un determinado esquema de direccionamiento a través de una red WAN con un esquema de direccionamiento distinto. En

esta situación, los paquetes IP de las redes a conectar se encapsulan dentro de paquetes IP de la red intermedia. De esta forma, para las redes extremo el encaminamiento sólo se realiza con un salto, es decir, las cabeceras IP de los protocolos encapsulados sólo tienen en cuenta un salto entre los *routers* extremos de la comunicación. Sin embargo, los cabeceras IP de los paquetes que encapsulan son tenidas en cuenta a lo largo de todo el encaminamiento intermedio entre las dos redes extremo.

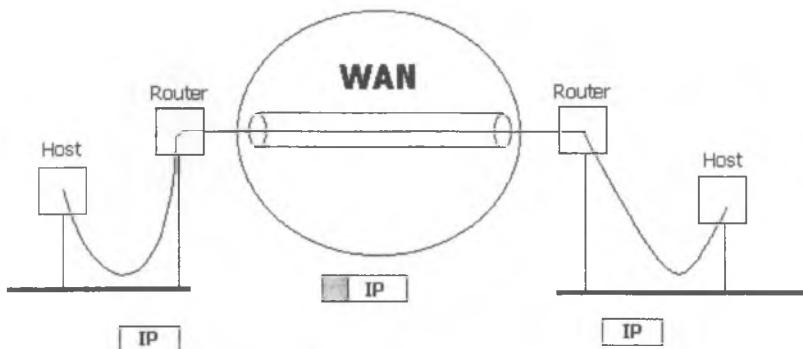


Figura 11.5 Esquema de funcionamiento del tunneling.

De esta forma, para los usuarios de las redes extremo interconectadas con el túnel, el acceso entre las redes remotas es totalmente transparente a la red WAN intermedia.

11.6 SEGURIDAD EN REDES A NIVEL IP

La seguridad en Internet es un aspecto especialmente importante en la actualidad, donde cada vez más el uso de Internet se dedica a la transmisión de información privada. Es posible realizar un estudio exhaustivo y detallado de los sistemas de seguridad que existen en redes TCP/IP, pero que no son objetivo de este libro. Únicamente se indicarán de forma genérica cuales son los esquemas de seguridad que se emplean en el nivel de red de Internet.

Para asegurar la confidencialidad de los datos que circulan en Internet es necesario realizar una encriptación de los datos en los paquetes de información. Esta encriptación de datos puede realizarse a cualquier

nivel dentro de la arquitectura (aplicación, transporte, red, etc.), realizándose ésta frecuentemente en el nivel de red y aplicación.

En el nivel de aplicación se emplean los servidores Web seguros, con el protocolo HTTPS, y los certificados de encriptación para el correo electrónico (Netscape). En el nivel de red mencionaremos un estándar para la encriptación de los paquetes a nivel de red: IPSEC. Este estándar permite la encriptación de los datos de los paquetes IP, siendo necesario que los *routers* que intercambian estos paquetes sean capaces de decodificarlos cuando lleguen a la red de destino. Las técnicas de encriptación son complejas y extensas y escapan del alcance de este libro, por lo que se invita al lector a profundizar en la bibliografía existente en estos aspectos.

Otra funcionalidad de seguridad que proporciona IP es el establecimiento de redes privadas. Existe un conjunto de direcciones de red IP que se encuentran reservadas para el uso de redes privadas y que los encaminadores de Internet rechazan o no encaminan de forma adecuada, por lo que no existirán paquetes con estas direcciones en sus cabeceras IP circulando en Internet. Un esquema de acceso seguro a Internet por parte de una red local, consiste en realizar un direccionamiento de la red local con direcciones IP privadas, de forma que se empleará un *router* a modo de *firewall* (cortafuegos, barrera de fuego, muro de seguridad) para que los paquetes puedan ser modificados y enviados a Internet empleando una dirección IP pública o legal, pero que personas externas a la red LAN (usuarios de Internet) no puedan acceder a la misma. Esta funcionalidad, implementada en la mayor parte de encaminadores comerciales, se denomina **NAT** (*Network Address Translation* - Conversión de direcciones de red).



BIBLIOGRAFÍA

- **Linux Máxima Seguridad.** *Anónimo*. Editorial Prentice Hall. 2000.
- **Redes Locales de Computadoras.** *José Antão Beltrão, Jacques Philippe Sauvé, William Ferreira, José Fábio Marinho*. Editorial McGraw Hill. 1990.
- **Internetworking with TCP/IP. Volume I: Principles, Protocols and Architecture.** *Douglas E. Comer*. Editorial Prentice Hall International, Inc. 1995.
- **Introducción a TCP/IP.** *Luis Miguel Crespo, Francisco Andrés Candelas*. Publicaciones de la Universidad de Alicante. 1998.
- **Fundamentos analíticos de las redes de computadores.** *Francisco Javier Gil*. Publicaciones de la Universidad de Alicante. 1996.
- **Comunicación de Datos, Redes de Computadores y Sistemas Abiertos.** *Fred Halsall*. Editorial Addison-Wesley Iberoamericana. 1998.
- **CCNA Exam Certification Guide.** *Wendell Odom*. Cisco Press. 1999.
- **Comunicaciones y Redes de Computadores.** *William Stallings*. Editorial Prentice Hall. 2000.
- **Redes de Computadoras.** *Andrew S. Tanenbaum*. Editorial Prentice Hall Hispanoamericana S.A. 1997.
- **Sistemas para la Transmisión de Datos.** *Fernando Torres, Francisco Andrés Candelas, Santiago T. Puente*. Publicaciones de la Universidad de Alicante. 1999.
- **Documentos RFC (Request for Comments) IETF.** Disponibles en la dirección URL: www.faqs.org.



