

IMPORTANTE:

Este material esta en permanente modificación y mejora.

Puede haber errores.

Importante comparar con las clases del profesor para no inducir a algún durante el estudio para controles.



Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Eléctrica

APUNTES EL55A
Sistemas de Telecomunicaciones

Prof. Néstor Becerra Yoma
Aux. Eduardo Ortega Canario

1 INTRODUCCION

El curso “Sistemas de Telecomunicaciones”, EL 55 A, es el último en el cuál todos los alumnos de la carrera tienen la oportunidad de ver los conceptos y tecnologías de las telecomunicaciones modernas. Por ese motivo, y tal como dice su nombre, el objetivo principal de este curso, que es bastante ambicioso, es dar una visión amplia y lo más completa posible de las tecnologías que están cambiando las sociedades y economías de hoy en día. La variedad y la complejidad de los temas que se abordan requieren varios libros textos y, dada la imposibilidad de tener varios ejemplares de cada uno de ellos en la biblioteca, es que se pensó en la elaboración de apuntes que sirvieran de guía a los alumnos para los asuntos vistos en clase. Sin embargo, estos apuntes no pretenden reemplazar completamente las referencias que se citan, siendo que la profundización en ciertos temas, cuando sugerido en clase, es de responsabilidad de los alumnos.

La variedad y complejidad de las tecnologías que se estudian imponen algunas restricciones con relación al orden en que estos se presentan. Es de vital importancia, por razones didácticas, que haya una secuencia visiblemente coherente. Si esto no ocurre, se corre el riesgo que el curso se vea como un aglomerado de temas independientes. Así, el orden natural utilizado va desde la capa Física hasta la de Aplicación en los modelos TCP/IP y OSI. Esto se observa en los capítulos II a VIII. En el capítulo IX se ve la teoría de tráfico y de colas utilizada en el diseño de sistemas de telefonía fija y celular. En el capítulo X se ven los conceptos básicos de telefonía celular y, finalmente, en el capítulo XI se hace una introducción a redes ópticas WDM de última generación.

2 REDES DE COMUNICACIONES

Unos de los puntos importantes en el desarrollo de la humanidad es la necesidad de comunicación que nace entre personas, grupos o naciones, con el fin último de compartir y conocer ideas, es decir información. Los medios de comunicación han variado a lo largo del desarrollo de la historia del hombre, pero siempre han tendido a una mayor apertura, es decir, han logrado que más personas puedan comunicarse entre sí, con menor demora y sin importar la ubicación geográfica. De aquí es que surgen como una solución a esta necesidad intrínseca las redes de comunicaciones, las cuales nos permiten acceso a información de variada forma (voz, datos, imágenes, etc) cada vez más eficientes y a menor costo. Este capítulo intenta introducir algunos conceptos básicos para el entendimiento de redes de comunicaciones y de datos.

2.1 Arquitectura de redes

Cuando se habla de arquitectura esta hace relación a todos los componentes de diseño tanto de hardware como software que componen una red de comunicación. En esta sección se especificarán alguno de estos componentes.

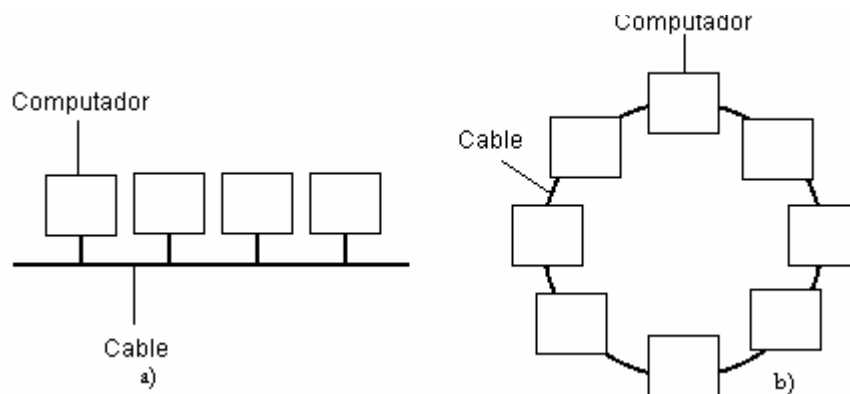


Figura 2.1 Redes de difusión a) Bus b) Anillo

2.1.1 Topología y Hardware

Una de las formas de clasificar las redes son a través de la tecnología de transmisión que utilizan. Estas pueden ser redes de *difusión* o redes *punto a punto*. Las primeras tienen un solo canal de comunicación compartido por todas las máquinas. Los mensajes que envía una máquina son recibidos por todas las demás pero un campo de dirección dentro del frame especifica a quién se dirige. En contraste, las redes punto a punto consisten de muchas conexiones entre pares individuales de máquinas. Para llegar a su destino puede ser que el paquete deba pasar antes por varias máquinas, lo cual hace muy importante el ruteo

que se de a este paquete. En las Figuras 2.1 y 2.2 se muestran posibles topologías para las redes de difusión y punto a punto.

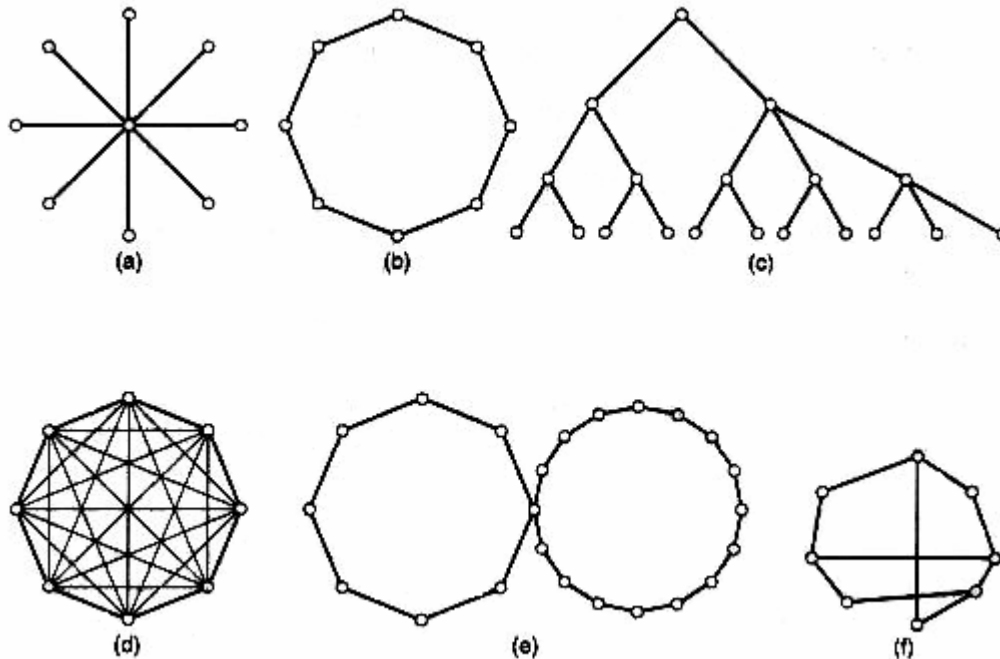


Figura 2.2: Posibles topologías para una subred punto a punto a) Estrella b) Anillo c)Arbol d)Completa e)Intersección de anillos f)Irregular

En función de las dimensiones físicas y del hardware, las redes se pueden clasificar en:

Redes de Área Local (LAN): son redes de propiedad privada dentro de un solo edificio o campus de hasta unos cuantos kilómetros de extensión. Se utilizan para interconectar computadoras dentro de espacios físicos relativamente reducidos (Ej: casa, oficina, campus, etc) con el objeto de compartir recursos. Pueden ser difusión o punto a punto, siendo las más conocidas y utilizadas las primeras, como por ejemplo redes de difusión Ethernet, IEEE 802.3 o IEEE 805.5 (Token Ring).

Redes de Área Metropolitana (MAN): es una versión más grande de una LAN y normalmente se basa en una tecnología similar. Podría abarcar un grupo de oficinas cercanas o una ciudad. Una MAN sólo tiene uno o dos cables y no contiene elementos de conmutación, lo cual simplifica el diseño.

Redes de Área Ampla (WAN): se extiende sobre un área geográfica extensa, a veces un país o continente. Contiene una serie de máquinas de usuario (hosts) los cuales se conectan a una subred de comunicación, la cual se encarga de rutear los paquetes entre hosts. La

subred se compone de líneas de transmisión y elementos de conmutación. Estos últimos conectan dos o más líneas y se encargan de asignar una línea de salida para cada paquete. Como termino genérico se les conoce como *enrutadores (routers)*. En la Figura 2.3 se muestra un ejemplo de subred.

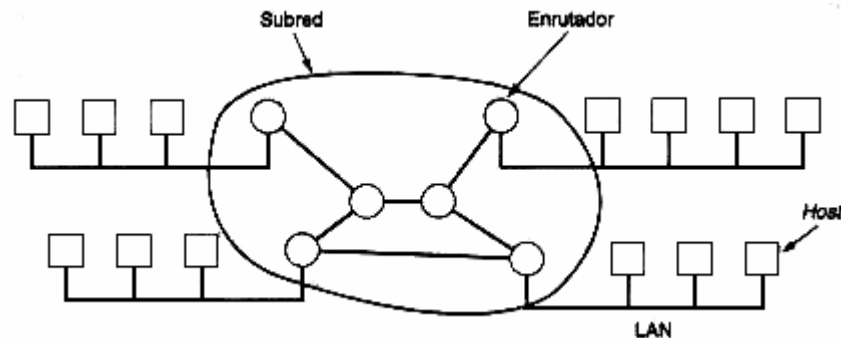


Figura 2.3: Subred y hosts

Pueden existir muchos tipos de redes alrededor del mundo con distintas tecnologías de software y hardware. Un conjunto de redes interconectadas se le llama *internet* (inter.-red). La Internet (con "I" mayúscula) es definida como la red formada por los hosts interconectados con TCP/IP.

2.1.2 Tipos de Conmutación

La comunicación entre máquinas en una red puede describirse desde dos enfoques: *conmutación de circuito* o también conocida como *orientada a conexión*, y *conmutación de paquetes* o *sin conexión*. La primera de ellas, como su nombre lo indica, necesita establecer un canal o circuito físico antes de establecer la comunicación. Una vez establecida, la información se transmite sin pérdida de datos y en orden. El mejor ejemplo de este tipo de conexión es el sistema público de telefonía.

Las redes de conmutación de paquetes funcionan de forma distinta. Estas dividen la información a ser enviada por la red en pequeñas unidades de hasta unos cientos de bytes, llamados paquetes, a los que se agrega identificaciones de destino y origen. El hardware envía estas unidades hacia su destino y cada uno de ellos viaja en forma independiente de su antecesor. El camino a seguir dependerá de las condiciones de la subred. Una vez en el destino el software se encargará de reensamblar los paquetes con la información que estos posean. Una analogía para este tipo de redes se observa en el servicio de correo.

2.1.3 Modelo de Capas

Como es posible deducir, el problema de comunicación entre equipos, los cuales pueden ser de fabricantes diferentes, y por ende, de tecnologías incompatibles puede llegar a ser muy complejo de resolver e implica un sin fin de consideraciones de diseño tanto de hardware como software. Una forma de abordar este desafío es dividir el problema en múltiples etapas, cada una independiente de la otra, y que en conjunto sean capaces de entregar una solución a variados requerimientos. Por lo anterior, muchas redes se

organizan como una serie de capas o niveles cada una construida sobre la inferior. El propósito de cada capa es ofrecer ciertos servicios a las capas superiores. Esto independiza las labores de cada capa pues a la capa superior no le interesa como la capa inferior se encarga de implementar su servicio. Luego, se llega a un nivel de abstracción tal que se puede mirar como si cada capa conversara con la capa adyacente en otra máquina gracias a un *protocolo* definido para cada nivel. Básicamente, un protocolo es un acuerdo entre las partes que se comunican sobre cómo va a proceder la comunicación. En la Figura 2.4 se muestra una red compuesta de cinco capas.

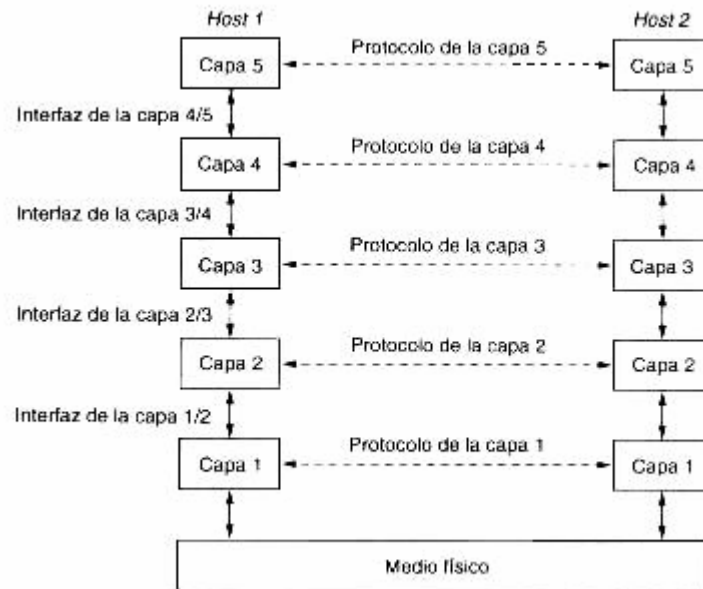


Figura 2.4: Capas, protocolos e interfaces.

De acuerdo a la Figura 2.4, los datos no se transfieren de la capa n de una máquina a la capa n de otra, sino que la capa transfiere los datos e información de control a la capa inferior, hasta llegar a la capa más baja. Luego es el medio físico en donde ocurre la comunicación real. Como se ve en la figura, entre las capas existen las interfaces las cuales definen operaciones y servicios primitivos que ofrece la capa inferior a la superior. La implementación de las funciones que definen las interfaces no se incluye en los protocolos, es decir, es de libre realización. Sólo deben cumplir con asegurar el o los servicios ofrecidos por cada capa.

Un conjunto de capas y protocolos recibe el nombre de *arquitectura de red*. La lista de protocolos empleados por cierto sistema, con un protocolo por capa, se llama *stack de protocolos*. Una analogía, tomada de “*Redes de Computadoras*” de Tanenbaum puede ayudar a explicar la idea de la comunicación multicapas. Imagine a dos filósofos (procesos pares de la capa 3), uno de los cuales es de India y habla indú e inglés, y el otro es de China y habla chino y francés. Ya que no tienen un idioma en común, cada uno contrata un traductor (procesos pares de la capa 2), cada uno de los cuales, a su vez, establece contacto con una secretaria (procesos pares de la capa 1). El filósofo indú desea comunicar a su par su afecto por los conejos. Para hacerlo, pasa a su traductor un mensaje (en inglés), por conducto de la interfaz 2/3, que dice: “I like rabbits”, como se ilustra en la Figura 2.5. Los

traductores acuerdan el uso del holandés como idioma neutro y el mensaje se convierte en *"Ik hou van konijnen"*. El idioma elegido es el protocolo de la capa 2 y la elección corresponde a los procesos pares de la capa 2. A continuación, el traductor entrega el mensaje a su secretaria para que lo transmita, por ejemplo, por fax (el protocolo de la capa 1). Cuando el mensaje llega, se traduce al francés (*"J'aime les lapins"*) y se pasa a través de la interfaz 2/3 al filósofo chino. Observe que cada protocolo es independiente por completo de los otros mientras las interfaces no cambien. Los traductores pueden cambiar a voluntad del holandés, por decir, al finlandés, siempre que ambos lo acuerden y que nada cambie su interfaz ya sea con la capa 1 o con la 3. De manera similar, las secretarias pueden cambiar de fax a correo electrónico o a teléfono sin molestar (o incluso sin informar) a las otras capas. Cada proceso puede añadir algo de información dirigida únicamente a su par. Esta información no se pasa a la capa de arriba.

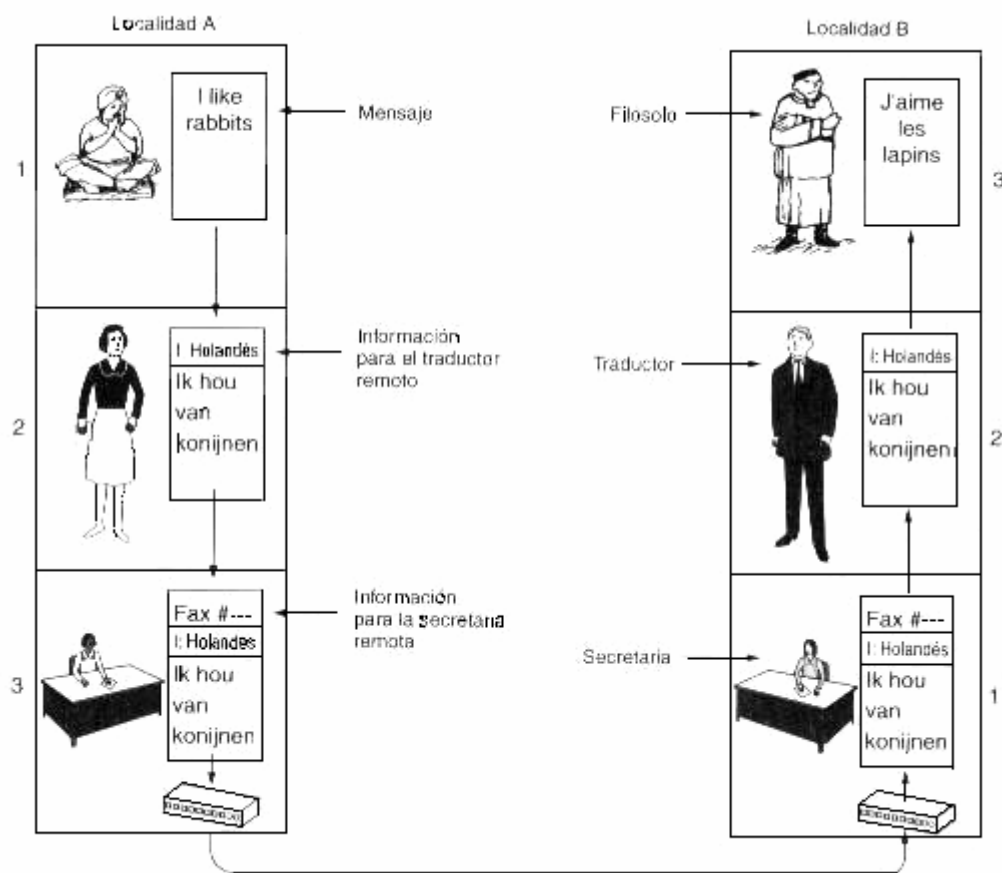


Figura 2.5: Arquitectura filósofo – traductor - secretaria

2.2 Modelo OSI

La Organización para la estandarización (ISO) en 1984 reconoció la necesidad de un modelo de red que ayudara a los productores a crear implementaciones de red interoperables y abiertas. De aquí nace el modelo de referencia **OSI** (*Open Systems Interconnection*) de la ISO, el cual sigue la idea de una arquitectura de red compuesta de 7

capas, las cuales debían tener funciones bien definidas y a la vez seguir estándares internacionales.

Aunque otros modelos de arquitectura han sido creados, la mayoría de los proveedores de redes asocian sus productos de red al modelo OSI cuando ellos quieren informar a los usuarios sobre sus productos. Luego, el modelo OSI es la mejor herramienta disponible para enseñanza de tecnologías de redes y de los sistemas de telecomunicaciones en general. Es importante mencionar que el modelo en sí no es una arquitectura de red pues no especifica los servicios y protocolos exactos que se han de usar en cada capa; sólo menciona lo que debe hacer cada una. Sin embargo, la ISO también ha elaborado estándares para todas las capas, aunque no sean parte del modelo de referencia mismo. La Figura 2.6 muestra la estructura de capas del modelo de referencia OSI.

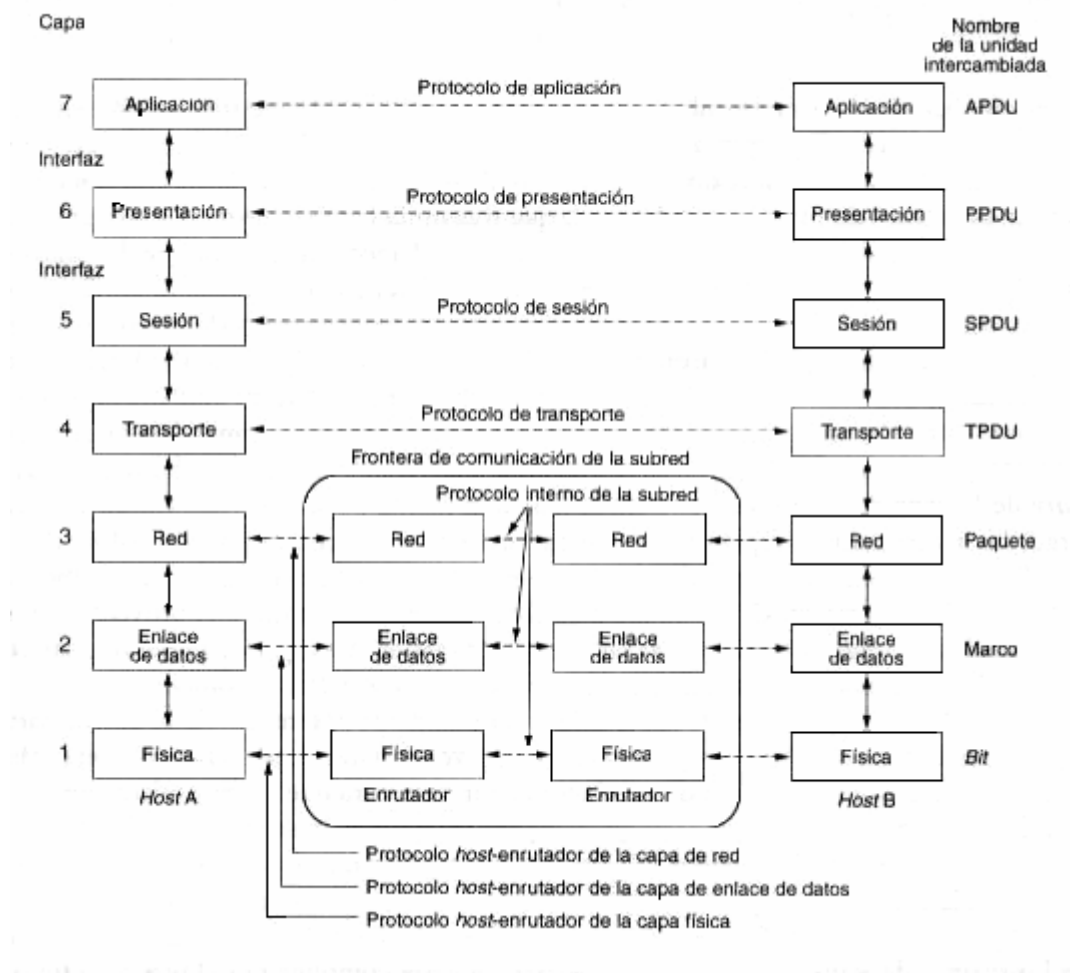


Figura 2.6: Modelo de referencia OSI

2.2.1 Capa Física

Toma en cuenta las interfaces mecánicas y eléctricas que harán posible la comunicación en un medio (el cual puede ser de difusión o punto a punto). Aquí la información es manejada a nivel de bits, debido a esto es necesario definir niveles de voltaje, tiempo de duración de

cada bit, sentido de la comunicación, técnica de modulación, etc. Todas estas consideraciones toman en cuenta el medio físico por el cual se llevará a cabo la comunicación. (Ej: fibra óptica, UTP, Cable coaxial, etc)

2.2.2 Capa de Enlace de datos

Se encarga de agrupar los bits de la capa física en *frames o marcos* y a la vez lograr que los hosts sean capaces de identificar un principio y un fin de frame. Además, tratará de enviar estos frames de forma de evitar errores y efectuará control sobre el envío con tal que un emisor no sature el buffer de recepción del equipo destino. La sofisticación en esta capa dependerá del medio de transmisión. .

2.2.3 Capa de Red

Es la encargada del funcionamiento de la subred de comunicaciones, es decir la forma en que los distintos componentes de esta efectúan el enrutamiento de los paquetes y control de congestión en los nodos. La capa de subred también es la encargada de comunicar entre sí redes que no necesariamente utilizan el mismo protocolo en niveles inferiores.

2.2.4 Capa de Transporte

Esta capa se puede pensar como el límite entre las capas concernientes a la aplicación (sesión, presentación y aplicación inclusive) y las capas que guardan relación con el transporte de los datos. Específica asuntos como la confiabilidad en el transporte sobre una interred, es decir, provee mecanismos para establecer, mantener y terminar circuitos virtuales.

Es posible dar un enfoque de extremo a extremo en esta capa, en cuanto a protocolos se refiere, logrando una comunicación directa entre el emisor y usuario final, a diferencia de las capas inferiores en donde la comunicación se realiza entre máquinas o nodos de subred vecinos. Otro punto importante de mencionar es que en esta capa se realiza la multiplexación de mensajes provenientes de distintas aplicaciones por un mismo canal y viceversa.

2.2.5 Capa de Sesión

Esta capa establece, maneja, y termina sesiones entre aplicaciones. Una sesión consiste en un dialogo entre una o más entidades de presentación (servicio a la capa superior). Ejecuta un manejo de control de dialogo y sincronización para transferencia, por ejemplo, de archivos de larga duración.

2.2.6 Capa de Presentación

Esta capa se asegura que la información enviada por la capa de aplicación sea entendible, al mismo nivel de capas, en el otro extremo. Si es necesario se implementan traducciones a formatos comunes de representación. También verifica la estructura semántica y sintaxis de los datos a enviar.

2.2.7 Capa de Aplicación

Es la capa más cercana al usuario. Contiene varios protocolos los cuales pueden emular terminales virtuales o ser utilizados para transferencia de archivos u otras labores. Esta capa establece la disponibilidad de comunicación con el destino y verifica si existen los suficientes recursos para que la transferencia pueda llevarse a cabo.

La figura 2.7 muestra un ejemplo de cómo los datos son transmitidos usando el modelo OSI. El proceso transmisor tiene datos que enviar al proceso receptor, entonces entrega estos datos a la capa de aplicación, la cual agrega su encabezado AH y se lo entrega a la capa de presentación. Esta a su vez realiza el mismo trabajo anteponiendo su encabezado y entregando el resultado a la siguiente capa.

Este proceso se repite hasta llegar a la capa física en donde es enviado, y al llegar a la máquina receptora esta saca los encabezados uno por uno en las distintas capas hasta que finalmente llega al proceso receptor.

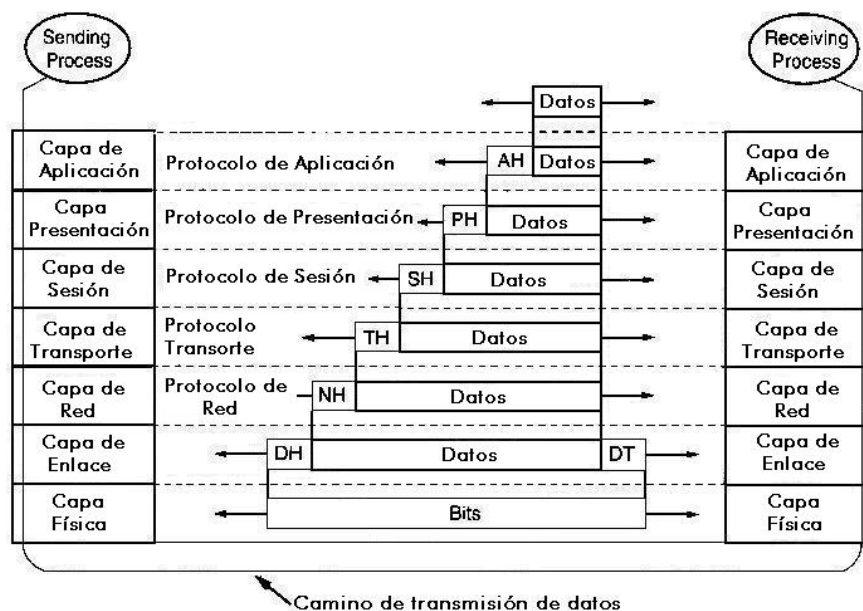


Figura 2.7: Ejemplo de uso del modelo OSI. Algunos de los encabezados pueden ser nulos

2.3 Modelo TCP/IP

Dada las características que podrían poseer las redes de conmutación de paquetes, la Agencia de Investigación de Defensa de los Estados Unidos (ARPA o DARPA) comenzó a interesarse por conformar una red de este tipo a nivel nacional. Esta red debía tener como características principales ser funcional para tecnologías disímiles de red y a la vez ser robusta a la pérdida de uno de sus nodos causada por ataques nucleares.

Con estos objetivos en mente, DARPA financió la investigación para el desarrollo de una serie de protocolos. El resultado de este trabajo fue la creación del grupo de protocolos de TCP/IP, nombre compuesto por las siglas de los dos estándares principales. (*Transmission Control Protocol / Internet Protocol*). En analogía con el modelo OSI, es posible dividir la arquitectura en capas, cada una con interfaces y servicios diferenciados. La Figura 2.8 muestra las distintas capas del modelo TCP/IP en contraste con el modelo OSI.

	OSI	TCP/IP
7	Aplicación	Aplicación
6	Presentación	
5	Sesión	Transporte
4	Transporte	
3	Red	Interred
2	Enlace de Datos	Host de Red
1	Física	

Figura 2.8: TCP/IP

2.3.1 Capa de Interred

Dada las características requeridas para la red se optó por orientar a esta capa hacia un servicio sin conexiones, que permitiera dar mayor flexibilidad a cada uno de los nodos componentes permitiéndoles ingresar paquetes a la red sin importar el destino de estos. Luego esta capa funciona con un filosofía del mejor esfuerzo (*best effort*), no confiable, dejando a las capas superiores la implementación de protocolos más sofisticados para recuperar paquetes perdidos. Luego el trabajo más importante en esta capa es el de llevar estos paquetes por los distintos nodos hacia sus destinos, lo que comúnmente se conoce como ruteo. En la capa de red se trabaja con unidades de largos no fijos, los *paquetes*, y el protocolo oficial para esta capa es el *Internet Protocol* o IP.

2.3.2 Capa de Transporte

Esta capa es similar a la capa del mismo nombre en el modelo OSI, en el sentido que es la que efectúa una comunicación directa entre entidades pares en el emisor y receptor. Para ello se basa en dos protocolos, cada uno de ellos orientado a distinto tipo de servicios. El primero de ellos es TCP (Protocolo de Control de Transmisión) el cual implementa un servicio orientado a conexión y confiable para las capas superiores. A la vez el protocolo TCP se encarga del control de flujo, lo cual evita que un emisor sature la sub-red de comunicación mediante control de congestión. Por otro lado, UDP (Protocolo de

Datagramas de Usuario), a diferencia del anterior, brinda un servicio no orientado a la conexión y no confiable. Es utilizado para aplicaciones que requieren una tasa constante de transmisión en vez de pérdidas iguales a cero. Básicamente UDP aprovecha el protocolo de capa de red para llevar sus mensajes (IP), agregando multiplexación de procesos, la cual permite diferenciar entre aplicaciones que se encuentren en una misma máquina y que estén siendo enviadas a través del mismo canal. Esto se realiza con la utilización de *puertos de protocolo*. Es importante mencionar que TCP también utiliza estos puntos abstractos de acceso llamados puertos, independizando cada uno de los programas o procesos de aplicación que corren en la misma máquina.

2.3.3 Capa de Aplicación

Aquí se encuentran todos los protocolos de alto nivel y que sirven para prestar determinados servicios al usuario. Entre ellos se pueden mencionar: TELNET (terminal virtual), FTP (transferencia de archivos), SMTP (correo electrónico), HTTP, etc. En la Figura 2.9 se observan los distintos protocolos mencionados y otros más que serán descritos en capítulos posteriores. Como se muestra en la figura TCP/IP no especifica protocolo para las capas inferiores a la de interred. Luego este variara y dependerá de la arquitectura de red física a utilizar.

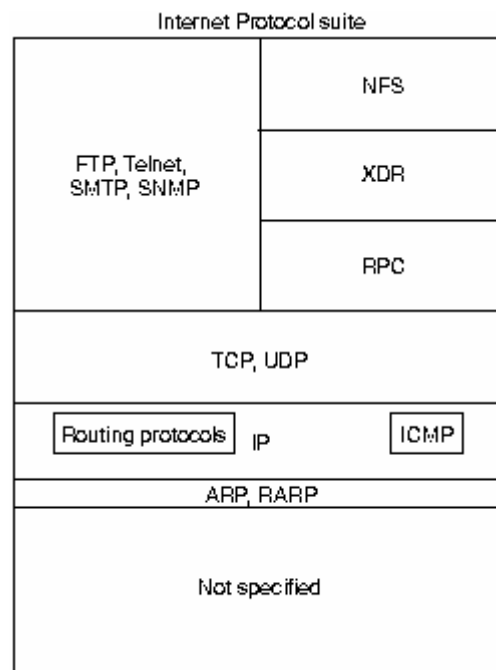


Figura 2.9: Protocolos TCP/IP

2.4 Tipos de Redes

Mientras avanza la técnica e investigación sobre nuevas formas para proveer servicios de comunicación, han surgido una serie de soluciones, las cuales se han ajustado a la problemática del momento llegando algunas de ellas a ser muy populares dadas sus características innovadoras o de menor costo. Algunas de estas tecnologías siguen siendo, en mayor o menor medida, utilizadas por las compañías prestadoras de estos servicios. En esta sección se describirán muy brevemente algunas de estas redes.

2.4.1 Redes X.25

Es un estándar que surge con el objetivo de interconectar terminales con las redes públicas ya existentes en Estados Unidos. Luego X.25 fue desarrollado principalmente por las compañías de teléfonos (portadoras o *carriers*), y fue diseñado para trabajar sin preocuparse del tipo de sistema del usuario. El estándar en cuanto a protocolos fue desarrollado por la ITU-T teniendo como logro resultar en un verdadero estándar mundial. Dado el ambiente de utilización (PDN: *Public Data Networks*) un sistema orientado a conexión, es decir, se necesita establecer una sesión entre el computador emisor y el receptor. Si el receptor acepta la comunicación esta se llevará a cabo en modo full-duplex.

La interacción entre dos puntos (punto a punto) es llevada a cabo entre el *data terminal equipment* (DTE) (en el lugar del usuario) y el *data circuit-terminating equipment* (DCE). A su vez este último se conecta al PSE (*Packet Switching Exchanges*), ambos equipos en las instalaciones del *carrier*, el cual irá conectado un DCE de salida que se conectará a su vez al DTE del receptor final. Esta configuración se muestra en la Figura 2.10.

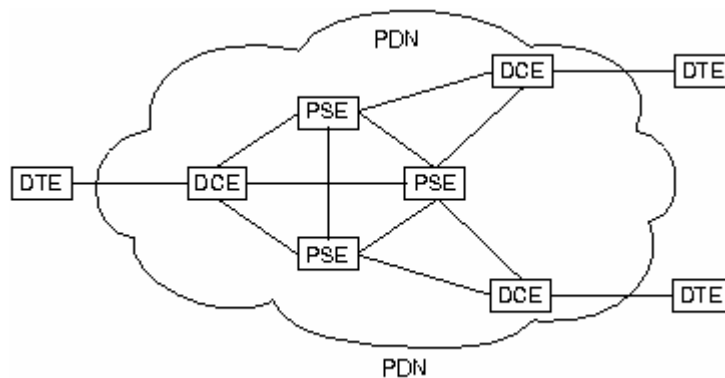


Figura 2.10: X.25

La capa física utiliza el protocolo *X.21 bis*, el cual a grandes rasgos es equivalente al conocido RS-232-C. X.21 bis identifica las características eléctricas y de circuito de intercambio para la interfaz DTE – DCE. Soporta conexión punto a punto, velocidades de hasta 64 Kbps, transmisión síncrona y full – duplex sobre un medio compuesto de cuatro

cables. La distancia máxima entre un DTE y un DCE son 15 metros. La capa de enlace se preocupa de control de transmisión para full – duplex, y manejo de errores en la línea telefónica entre el DTE del usuario y el DCE en la estación del carrier. La capa de red se encarga de los denominados *circuitos virtuales* definiendo identificadores y dependiendo si estos son de carácter permanente o conmutado. También se preocupa del direccionamiento y control de flujo.

2.4.2 Frame Relay

Frame Relay provee una comunicación de conmutación de paquetes, los cuales circulan entre la interfaz de usuario y los equipos de red. Aquí también se distinguen unos de otros como DTE para los equipo en el lado del usuario y DCE para los del lado de red. Este servicio fue pensado para ser más económico y eficiente que sus antecesores, aprovechando los cambio tecnológicos desde redes telefónicas analógicas poco confiables hacia enlaces digitales y menos propenso a errores. Esto permitió olvidarse de algoritmos de corrección de errores solo incluyendo un algoritmo de redundancia cíclico (CRC) que es capaz de detectar bits erróneos sin ejecutan corrección. Todo esto acompañado también con el desarrollo de las computadoras, lo cual permitió delegar a capas superiores trabajos que anteriormente eran ejecutados por las capas más bajas.

Una de las diferencias con X.25 (además de las arriba mencionadas) es que frame relay no incluye explícitamente control de flujo, delegándolo a capas superiores.

Frame Relay hace utilización de los canales virtuales conmutados (SVC), es decir DTE y DCE establecen, terminan y administran dinámicamente un SVC. Este también puede establecer circuitos virtuales permanentes los cuales son configurados por el administrador de la red.

2.4.3 ISDN

Red Digital de Servicios Integrados (RDSI) tiene por principio integrar todos los posibles servicios de voz, datos, texto, gráficos, música y video provistos para el usuario desde un simple terminal sobre la red telefónica pública conmutada.

Los servicios ofrecidos por ISDN son básicamente dos:

- ❑ *Basic Rate Interface (BRI)*: Servicio que ofrece dos canales B y uno D (2B+D). Un canal B trabaja a 64 Kbps y fue pensado para llevar datos. El canal D opera a 16 Kbps y se encarga de llevar información de control y señalización, aunque puede soportar datos bajo ciertas circunstancias.
- ❑ *Primary Rate Interface (PRI)*: servicio que provee se 30 canales B más uno D de 64 Kbps, haciendo un total de 2.048 Mbps (E1). Para EEUU y Japón el servicio consta de 23B + D (1.544 Mbps = T1).

2.4.4 ATM

El Modo de Transmisión Asíncrono (ATM) surge del esfuerzo de la ITU-T por desarrollar un estándar para B-ISDN (Red Digital de Servicios Integrados de Banda Ancha) para permitir transferencia a alta velocidad de voz, video y datos a través de la Red Pública de Datos (PDN).

La idea de ATM es transmitir la información en pequeños trozos de largo fijo llamados *células o celdas*, las cuales se componen de 48 bytes de campo para datos más 5 bytes para encabezado. A nivel de celda ATM, como su nombre lo indica, no necesita de sincronía (no están ligados a un reloj de referencia). Los nodos de una red ATM están compuestos por los llamados switches ATM.

Estas redes son orientadas a la conexión, es decir, es necesario establecer primero una conexión antes de enviar las celdas. Estas conexiones son ejecutadas a una velocidad de 155 Mbps, y trabajan con PVCs o SVCs (Permanent o Switches Virtual Channel) según sea el caso.

El modelo de referencia por capas es más complejo que los anteriormente vistos (TCP y OSI), pues como muestra la Figura 2.11, incluye un tercera dimensión para protocolos de gestión de capas y planos.

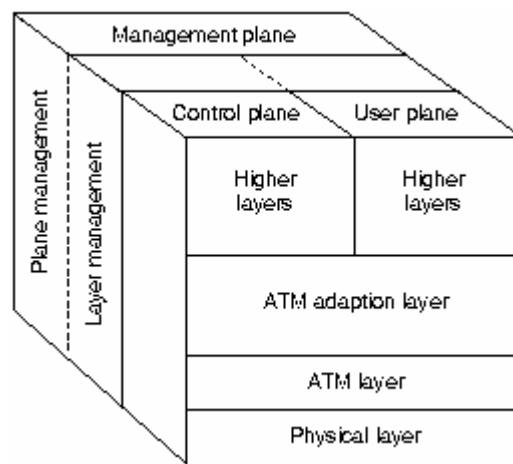


Figura 2.11: ATM

3 TRANSMISIÓN DIGITAL

En este capítulo se abordan los fundamentos básicos en los cuales se sustentan las comunicaciones digitales, a la vez que pretende mostrar las ventajas de la digitalización de la información.

Estas ventajas surgen debido al carácter discreto que adquieren las señales, la cual hace posible utilizar técnicas de procesamiento menos complejas (en los puntos extremos), a la vez que permite a las señales ser más inmunes al ruido (viajando por el mismo canal de comunicación).

Luego las ventajas de la transmisión digital son:

- Inmunidad al ruido
- Procesamiento digital y multicanalización
- Sencillas de medir y evaluar
- Facilidad para medir rendimiento y tasas de error

Y las desventajas:

- El procesamiento digital tiene implícito un retraso
- La conversión A/D y D/A introduce ruido de Cuantización
- Necesita sincronización
- Incompatibilidad con sistemas analógicos.

Como tema relacionado también se tratará un tipo de modulación muy utilizada para transmisión digital, la cual es conocida como PCM (*Modulación de Pulsos Codificados*) y el capítulo concluirá algunos tópicos sobre codificación de voz.

3.1 Ruido

El ruido eléctrico se define como cualquier energía eléctrica no deseada presente en el pasabanda de un circuito de comunicación.

El ruido puede dividirse en dos categorías generales: *Correlacionado* y *No – correlacionado*.

Correlación implica una relación entre la señal y el ruido, mientras que el ruido no correlacionado está presente en la ausencia de cualquier señal. Esto quiere decir que cuando está presente la señal no tiene efecto sobre la magnitud del ruido.

Ruido no correlacionado: Este ruido está presente sin importar si hay señal o no. Puede ser externo o interno.

El ruido externo se debe a ruido atmosférico o estática, ruido del espacio generado por el sol o galaxias. Las frecuencias son de 30 MHz para el ruido atmosférico y entre 8 MHz a 1,5 GHz para ruido del espacio.

El ruido interno es generado dentro de los dispositivos electrónicos y se clasifican en tres tipo: térmico, de disparo (shot – noise) y de tránsito.

Ruido térmico: es el movimiento aleatorio de los electrones por agitación térmica. Se define la densidad de potencia de ruido N_0 .

$$N_o = kT [W / Hz]$$

k = Constante de Boltzmann ($1,38 \times 10^{-23} [J/^{\circ}K]$)

T = Temperatura Absoluta $^{\circ}K$

A temperatura ambiente la densidad de potencia de ruido disponible es:

$$N_o = 1,38 \times 10^{-23} \times 290 = 4 \times 10^{-21} [W/Hz]$$

En dBm:

$$N_o = 10 \log \left[\frac{kT}{10^{-3}} \right] = -174 [dBm]$$

La potencia total de ruido en un ancho de banda B es:

$$N = N_o B = kTB [W]$$

En dBm:

$$N_o = 10 \log \left[\frac{kTB}{10^{-3}} \right] [dBm]$$

Como el ruido térmico es igualmente distribuido en el espectro de frecuencias se le denomina ruido blanco en analogía a la luz blanca que contiene todas las frecuencias de la luz visible.

Ruido de disparo: Es producido por la llegada aleatoria de los portadores a los terminales de un dispositivo semiconductor. Esto se debe al efecto del camino libre medio en el movimiento de los portadores, por lo tanto, es proporcional a la corriente y ancho de banda del sistema de comunicaciones. La potencia de ruido de disparo es aditiva con el ruido térmico.

Ruido de tránsito: Generado en transistores BJT debido al retardo introducido por la base para el viaje de los portadores desde el emisor al colector. Es importante en alta frecuencia.

Ruido correlacionado: Es energía eléctrica no deseada que está presente como un resultado directo de una señal. Ej: distorsión armónica y de intermodulación. Ambas distorsiones son consecuencia de la amplificación no lineal de la señal. La diferencia entre ellas reside en que la primera aparece con sólo 1 señal y la segunda con dos señales.

Distorsión armónica: Es la generación de armónicas no deseadas por la amplificación no lineal de una señal sinusoidal. Ej: amplificador de señales grandes. La distorsión armónica total THD se define como:

$$THD = \frac{\sqrt{V_2^2 + V_3^2 + \dots + V_n^2}}{V_1} \times 100\%$$

Con V en valores efectivos.

Ruido de intermodulación: Son las frecuencias no deseadas del producto cruzado (sumas y diferencias) de dos señales cuando son amplificadas en un dispositivo no lineal. Ejemplo: estaciones radiobase de telefonía celular.

3.2 Teorema Muestreo y Máxima Capacidad del Canal

Para ponernos en el contexto en el cual son aplicables los teoremas a enunciar es necesario hacer una breve descripción de las componentes de un sistema análogo digital. La Figura 3.1 muestra los bloques fundamentales de un sistema de conversión A/D (Análogo a Digital).



Figura 3.1: Diagrama de bloques conversor A/D

En donde:

- $x_a(t)$: señal analógica (t real)
- $x(n)$: señal discreta (n discreto tq. $n=0,1,\dots$)
- $x_q(n)$: señal digital

La primera etapa es el muestreo de la señal análoga. Este muestrea una tasa tal que el proceso pueda ser reversible. Aquí lo que se discretiza es la variable independiente (variable temporal). Esta tasa esta dada por el *Teorema del Muestreo*, el cual se enuncia a continuación.

Teorema del muestreo (Nyquist 1928, Shannon 1949): Si la frecuencia más alta contenida en una señal analógica $x_a(t)$ es $F_{\text{máx}}=B$ y la señal se muestrea a velocidad $F_2 > 2F_{\text{máx}} = 2B$ entonces $x_a(t)$ se puede recuperar totalmente a partir de sus muestras.

Si la tasa de muestreo es menor a $2B$ se produce “*aliasing*” o distorsión de la señal recuperada a esta tasa.

Luego, según el teorema anterior el proceso de muestrear un señal en el tiempo y obtener una representación discreta es reversible siempre y cuando la tasa de muestreo sea el doble de la máxima frecuencia contenida en la señal analógica.

Después de esta etapa se obtiene una señal discreta en tiempo pero que puede tener un rango infinitos de valores dentro de un intervalo (en el cual se define la señal) real. Para obtener una señal digital es necesario, entonces, cuantizar, con tal que el resultado este dentro de un conjunto finito y conocido de valores, y pueda ser posteriormente procesada.

3.3 Error de Cuantización y ruido

El proceso que transforma una señal de amplitud continua discreta en el tiempo en una señal digital al asociar a cada muestra un número de dígitos finito es conocido como *cuantización*. Como se puede deducir, este proceso introduce un error al representar los valores continuos de la señal a niveles discretos, el cual es conocido como *error o ruido de cuantización*.

Con el fin de obtener una expresión que nos sirva para representar el error de cuantización, se seguirá un alcance estadístico como el sugerido en el texto “*Digital Signal Processing: principles, algorithms and applications*” (de los autores John Proakis y Dimitris Manolakis, editorial Prentice Hall).

La cuantización es un proceso no lineal e irreversible, la cual mapea una amplitud dada $x(n) \equiv x(nT)$ en tiempo $t = nT$ dentro de una amplitud x_k , tomado desde un conjunto finito de valores. El procedimiento es ilustrado en la Figura 3.2, donde el rango de amplitud de la señal es dividido en L intervalos con $L+1$ *niveles de decisión* x_1, \dots, x_{L+1} .

$$I_k = \{x_k < x(n) < x_{k+1}\} \quad k = 1, 2, \dots, L \quad (3.0)$$

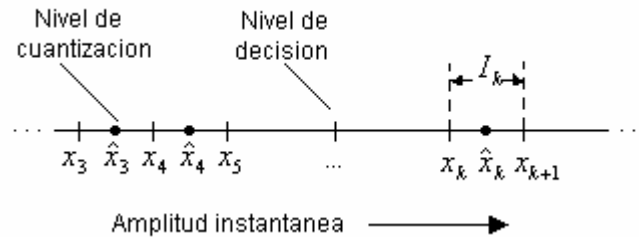


Figura 3.2: Proceso de cuantización

Las posibles salidas del cuantizador son denotadas por $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_L$.

El cuantizador frecuentemente utilizado es el *cuantizador lineal o uniforme*, definido por:

$$\begin{aligned} \hat{x}_{k+1} - \hat{x}_k &= \Delta & k &= 1, 2, \dots, L-1 \\ x_{k+1} - x_k &= \Delta & \text{Par } x_{k+1}, x_k \text{ finito} \end{aligned}$$

donde Δ es el tamaño del paso del cuantizador.

Se puede ver fácilmente que error del cuantizador está siempre en el rango $-\Delta/2$ a $\Delta/2$:

$$-\frac{\Delta}{2} < e_q(n) \leq \frac{\Delta}{2} \quad (3.1)$$

Luego el error de cuantización instantáneo no puede exceder la mitad del paso de cuantización. Si el rango dinámico de la señal, definido como $x_{\max} - x_{\min}$, es más grande que el rango del cuantizador, las muestras que exceden el rango del cuantizador son recortadas, resultando en un error más grande que $\Delta/2$.

El proceso de codificación en un conversor A/D asigna un número binario único a cada nivel de cuantización. Si tenemos L niveles, se necesitarán a lo menos L números binarios distintos. Con una palabra de largo b+1 bits se pueden representar 2^{b+1} números binarios distintos. Entonces podríamos tener $2^{b+1} \geq L$ o, equivalentemente $b+1 \geq \log_2 L$. Entonces el tamaño del paso o la resolución del conversor A/D esta dada por:

$$\Delta = \frac{R}{2^{b+1}} \quad (3.2)$$

Donde R es el rango del cuantizador.

3.3.1 Análisis del error de cuantización

La dependencia del error de cuantización en las características de la señal de entrada y la naturaleza no lineal del cuantizador hacen de análisis determinístico intratable, excepto en casos muy simples.

En el presente análisis se asume que el error de cuantización es de naturaleza aleatorio, y se modela como un ruido que es sumado a la señal original.

Se requieren los siguientes supuestos sobre las propiedades estadísticas de $e_q(n)$:

1. El error $e_q(n)$ está uniformemente distribuido sobre el rango $-\Delta/2 < e_q(n) < \Delta/2$.
2. La secuencia $\{e_q(n)\}$ es una secuencia de ruido blanco estacionario. En otras palabras, el error $e_q(n)$ y el error $e_q(m)$ para $m \neq n$ están no correlacionados.
3. La secuencia de error $\{e_q(n)\}$ está no correlacionada con la secuencia de señal $x(n)$.
4. La secuencia de la señal $x(n)$ es de media cero y estacionaria.

Bajo estas suposiciones (que no siempre se mantienen), el efecto del ruido aditivo sobre la señal puede ser cuantizado al evaluar la relación señal a ruido de cuantización (SQNR), la cual puede ser expresada sobre una escala logarítmica (en dB) como:

$$SQNR = 10 \log_{10} \frac{P_x}{P_n} \quad (3.3)$$

donde $P_x = \sigma_x^2 = E[x^2(n)]$ es la potencia de la señal y $P_n = \sigma_e^2 = E[e_q^2(n)]$ es la potencia del ruido de cuantización.

Si el ruido de cuantización está uniformemente distribuido en el rango $(-\Delta/2, \Delta/2)$ como muestra la Figura 3.3, el valor medio del error es cero y la varianza (la potencia del ruido de cuantización) es:

$$P_n = \sigma_e^2 = \int_{-\Delta/2}^{\Delta/2} e^2 p(e) de = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} e^2 de = \frac{\Delta^2}{12} \quad (3.4)$$

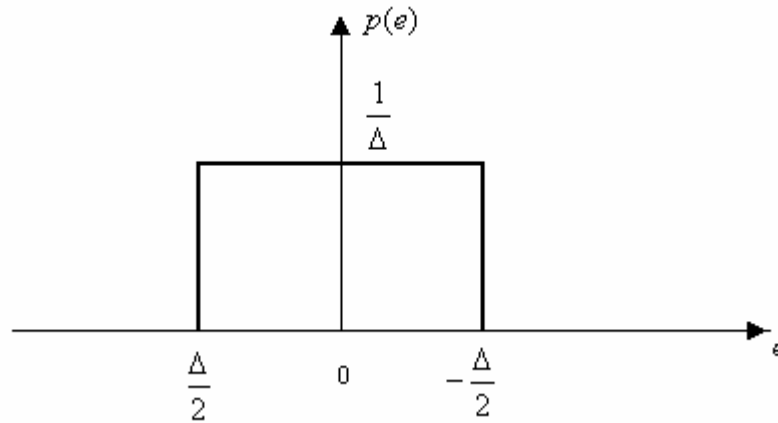


Figura 3.3: Función densidad de probabilidad del error de cuantización

Al combinar la ecuación (3.2) con (3.4) y sustituirla en (3.3), se tiene:

$$SQNR = 6.02b + 16.81 - 20 \log \frac{R}{\sigma_x} \quad (3.5)$$

La cual depende del rango R del conversor A/D y las estadísticas de la señal de entrada. Por ejemplo si asumimos $x(n)$ está distribuida en forma normal y el rango del cuantizador se extiende desde $-3\sigma_x$ a $3\sigma_x$ (i.e. $R = 6\sigma_x$) se tiene:

$$SQNR = 6.02b + 1.25 [dB] \quad (3.6)$$

La formula general para el SQNR es usada frecuentemente para especificar la precisión necesaria en un conversor A/D. Esta simplemente significa que cada bit adicional en el cuantizador aumenta la relación señal ruido en 6 dB.

3.4 PCM (Pulse Code Modulation)

La idea principal de la transmisión digital es lograr convertir la información en pulsos los cuales pueden ser enviados a través del canal de comunicación. Esta información ya puede estar en el formato deseado, o también puede estar en forma analógica. Si es el caso este último, se ha descrito en las secciones anteriores la base de la transformación de la información a formato digital.

Al momento de cuantizar, los niveles continuos de amplitud de la señal son encasillados en niveles discretos los cuales representan una aproximación al valor original. Como último paso necesitamos transformar estos valores discretos en pulsos los cuales serán transmitidos a través del canal de comunicación.

Para ello se utilizan métodos de *modulación de pulsos*. Estos permiten transformar la información deseada en pulsos, siendo los más conocidos:

1. **PWM (*Pulse Width Modulation*)**: El ancho del pulso es proporcional a la amplitud de la señal analógica.
2. **PPM (*Pulse Position Modulation*)**: La posición de un pulso (de ancho constante) dentro de una ranura de tiempo determinada varía proporcionalmente a la amplitud de la señal.
3. **PAM (*Pulse Amplitud Modulation*)**: La amplitud de un pulso fijo (de ancho constante) varía de acuerdo a la amplitud de la señal.
4. **PCM (*Pulse Code Modulation*)**: A cada nivel de cuantización se le es asignado un número binario (código) de largo fijo. El número binario varía su valor de acuerdo a la amplitud de la señal.

La Figura 3.4 muestra un ejemplo para cada tipo de modulación.

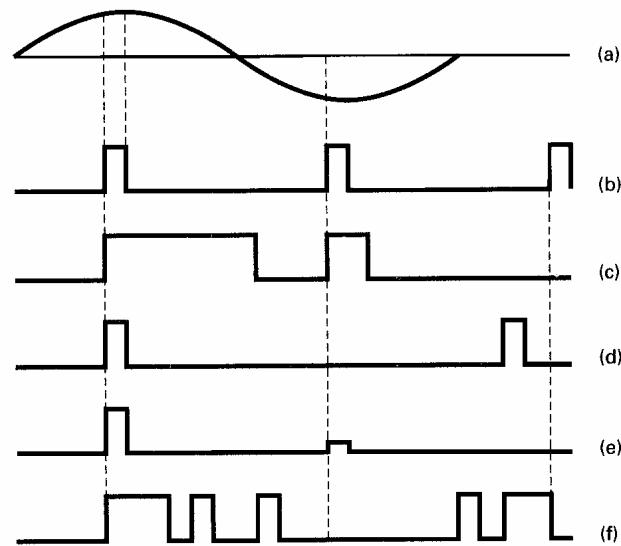


Figura 3.4: a) Señal analógica; b) Pulsos de muestreo; c) PWM; d) PPM; e) PAM; f) PCM

Como se puede notar, la formula obtenida para la relación señal ruido de cuantización (3.6), lleva implícita el hecho de la utilización de un código (de b bits) para cada nivel de cuantización. Luego la relación que se cumple entre los L niveles de cuantización y los b bits del código es:

$$2^b \geq L \quad (3.7)$$

$$\Rightarrow b = \lceil \log_2 L \rceil$$

Esto refleja que el método mayormente utilizado es PCM, y es el que se estudiará en forma breve.

En PCM, tanto la amplitud como la duración de cada pulso son fijos, los cuales son de carácter binario (1 's y 0's). La Figura 3.5 muestra un sistema simplificado de un solo canal PCM.

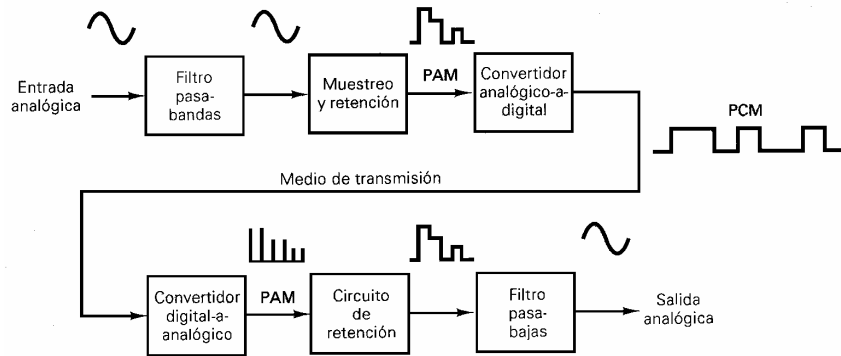


Figura 3.5: Sistema con canal PCM

Debido a la utilización de códigos para representar cada uno de los niveles es necesario el uso de codificadores y decodificadores en la transmisión y recepción respectivamente. Un sistema que realiza la codificación/decodificación es conocido como *codec*.

La eficiencia de un codificador PCM se representa como la razón entre el mínimo número de bits requeridos para lograr un número de niveles específicos y el número real de bits usados.

$$\text{eficiencia del codificador} = \frac{\text{mínimo número de bits} \times 100}{\text{número real de bits (incluyendo bit de signo)}}$$

También es posible distinguir entre códigos PCM *lineales* y *no lineales*. Los primeros son los que asignan en forma uniforme a los niveles de cuantización los códigos correspondientes, y son los que hemos visto hasta el momento. Los no lineales o no uniformes, son aquellos que asignan un mayor número de códigos a ciertos niveles de amplitud. Por ejemplo en codificación de voz es más probable que las señales sean de menor amplitud, por lo tanto se asignan mayor cantidad de códigos, provocando un mayor error de cuantización para señales de mayor amplitud (disminuyendo la relación SQNR).

Compansión

Se refiere al proceso de comprimir una señal (en el lado del transmisor) y luego expandirla (en el lado del receptor). Con PCM este proceso se logra por medio de procesos analógicos o digitales.

La característica de compansión vendrá dada por el tipo de señales a tratar. Por ejemplo, para señales de voz es necesario se necesita un rendimiento SQR relativamente constante, lo que significa que la distorsión debe ser proporcional a la amplitud de la señal para cualquier nivel de señal de entrada[±]. Esto requiere una razón de compresión

[±] Fuente: "Sistemas de comunicaciones electrónicas" Wayne Tomasi.

logarítmica. Existen dos métodos de compresión analógicos que se aproximan a una función logarítmica, y son conocidos como compansión de *ley-μ* y de *ley-A*.

- **Compansión de ley-μ:** Es usada en Estados Unidos y Japón.

$$V_{salida} = \frac{V_{máximo} \times \ln(1 + \mu V_{entrada} / V_{máximo})}{\ln(1 + \mu)} \quad (3.8)$$

en donde: $V_{máximo}$ = máxima amplitud de entrada analógica descomprimida.
 $V_{entrada}$ = amplitud de la señal de entrada en un instante particular del tiempo
 μ = parámetro usado para definir la cantidad de compresión.
 V_{salida} = amplitud de salida comprimida.

- **Compansión de ley-A:** Es utilizada en Europa.

$$V_{salida} = V_{máximo} \frac{A V_{entrada} / V_{máximo}}{1 + \ln A} \quad 0 \leq \frac{V_{entrada}}{V_{máximo}} \leq \frac{1}{A} \quad (3.9)$$

$$V_{salida} = V_{máximo} \frac{1 + \ln(A V_{entrada} / V_{máximo})}{1 + \ln A} \quad \frac{1}{A} \leq \frac{V_{entrada}}{V_{máximo}} \leq 1 \quad (3.10)$$

3.5 Máxima Capacidad de Canal

Una vez revisado los conceptos básicos para la transformación de la señal de información análoga en una señal digital que puede ser transmitida, es necesario describir cual es la capacidad del canal, en cuanto a tasa de bits se refiere.

Nyquist, en 1924, postuló que una señal que ha sido pasada por un filtro de ancho de banda B, y luego de ser cuantizada en L niveles discretos, la máxima tasa de datos a ser transmitidos por un canal sin ruido será:

$$máxima \text{ tasa de datos} = 2B \log_2 L \text{ bits / seg} \quad (3.11)$$

Pero esta relación considera canales sin ruido.

La forma de medir el ruido térmico presente en el canal es con la relación señal ruido S/N. Shannon (1948) postuló que para un canal con una relación señal ruido S/N, ancho de banda B, se tiene una capacidad de canal máxima dada por:

$$C = B \log_2 \left[1 + \frac{S}{N} \right] \quad (3.12)$$

La ecuación (3.12) sólo representa un límite teórico para la capacidad de canal, ya que los sistemas reales raramente lo alcanzan.

3.6 Codificación de Voz

Dado la importancia que ha tenido, sobre todo en comunicaciones móviles, se ha considerado introducir en este capítulo una breve descripción de los fundamentos básicos para la codificación de voz. Muchos de los temas a mencionar ya han sido tratados con anterioridad, así que sólo serán mencionados brevemente.

Gran parte de la información (sino su totalidad) aquí expuesta fue extraída del libro “*Wireless Communications*” (Theodore S. Rappaport), capítulo 7 “*Speech Coding*”.

3.6.1 Introducción

El fin de todo sistema de codificación de voz es transmitir voz con la más alta calidad posible usando la menor capacidad de canal. Estos dos objetivos son posibles manteniendo cierto compromiso entre nivel de complejidad de implementación y retraso en comunicación. Entre más complejo es un algoritmo, más retraso de procesamiento y costo de implementación requiere.

La jerarquía en codificadores de voz se muestra en la Figura 3.6.

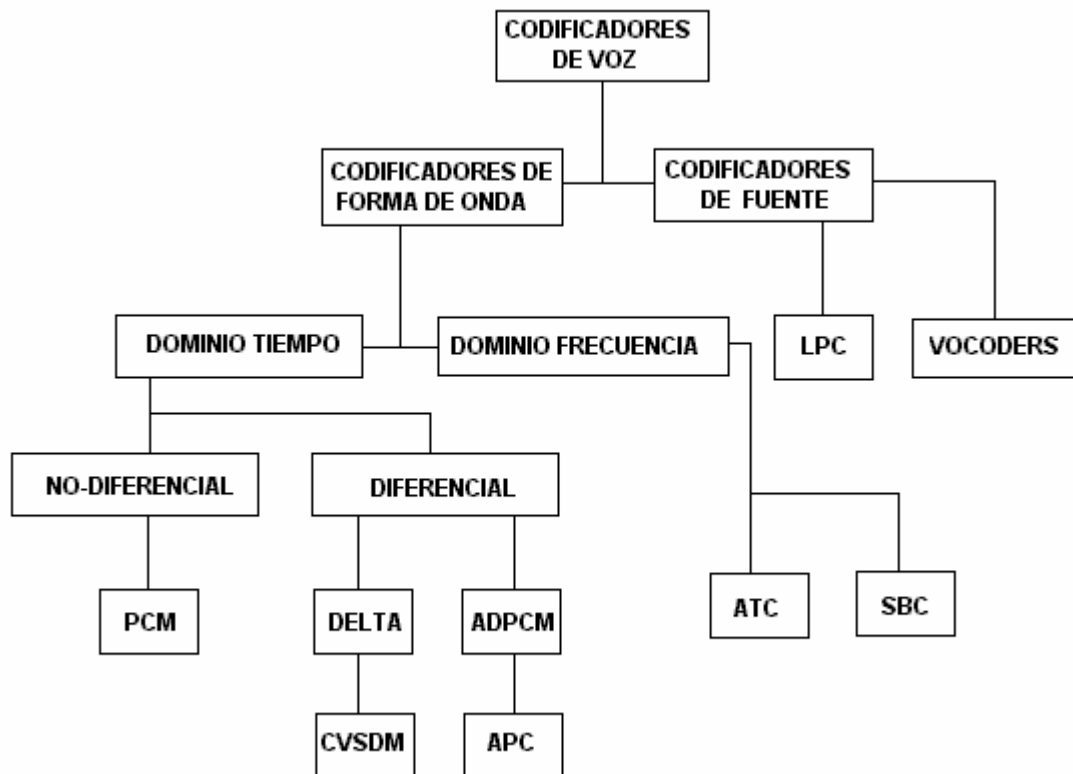


Figura 3.6: Jerarquía Codificadores.

Basados en la forma por la cual logran la compresión, los codificadores de voz son ampliamente clasificados en dos categorías: *Codificadores de forma de onda* y *Vocoders*. Los primeros esencialmente se esfuerzan por reproducir la forma de onda en el tiempo de la señal de voz tan cercanamente como sea posible. En principio fueron diseñados para ser independientes de la fuente y por lo tanto codifican igualmente bien una variedad de señales. Ellos tienen la ventaja de ser robustos para un amplio rango de características de voz y para ambiente ruidoso, con un mínimo de complejidad pero con moderada economía en tasa de bits. Algunos ejemplos de estos codificadores son: PCM, DPCM (PCM diferencial), ADPCM (PCM adaptivo diferencial), modulación delta (DM), modulación delta de gradiente continuamente variable (CVSDM), y código predictivo adaptivo (APC). Por otro lado, los Vocoders alcanzan tasa de bits menores y generalmente son más complejos. Ellos se basan en un conocimiento a priori de la señal a codificar, y por esta razón ellos son específicos para tipos de señales.

3.6.2 Características de la señal de voz

La forma de onda de la señal de voz tiene una serie de características y propiedades que pueden resultar útiles para el diseño de codificadores. Algunas de ellas son: distribución de probabilidad no uniforme de la amplitud de la voz, autocorrelación distinta de cero entre muestras, espectro de voz no plano. La propiedad más básica de la señal de voz que es explotada por todos los codificadores es que es de banda limitada, la cual hace posible el muestreo.

La siguiente propiedad más explotada es la *función de densidad de probabilidad no uniforme* de amplitudes que posee la voz. Esta es caracterizada por una probabilidad muy alta de amplitudes cerca del cero y una función decrecientemente monótona de amplitudes en los extremos. La distribución exacta, sin embargo, depende del ancho de banda de la entrada y condiciones de grabación. La función exponencial dada en la ecuación (3.13) provee una buena aproximación para señales de calidad telefónica.

$$p(x) = \frac{1}{\sqrt{2}\sigma_x} \exp(-\sqrt{2}|x|/\sigma_x) \quad (3.13)$$

La propiedad de las señales de voz la cual indica que existe mucha correlación entre muestras adyacentes de segmentos de voz permite realizar predicción utilizando las muestras previas con un pequeño error aleatorio.

La característica no plana de la densidad espectral hace posible obtener compresión significativa de código de voz en el dominio de la frecuencia.

3.6.3 Técnicas de cuantización

Existen básicamente dos técnicas: Cuantización uniforme (ya analizada) y no uniforme.

La cuantización no uniforme intenta distribuir los niveles de cuantización de forma más eficiente, tratándolos de acuerdo a la fdp de la señal de entrada. Para una señal de entrada se tiene una distorsión de la forma:

$$D = E[(x - f_Q(x))^2] = \int_{-\infty}^{\infty} [x - f_Q(x)]^2 p(x) dx \quad (3.14)$$

Donde:

$x(t)$: señal de voz

f_Q : señal de voz cuantizada

$p(x)$: fdp

Luego se ve que la distorsión se reduce si se reduce el ruido de cuantización en donde fdp es grande.

Una implementación simple de un cuantizador no uniforme es el cuantizador logarítmico. Este usa pasos de cuantización fina para amplitudes bajas que ocurren frecuentemente en voz y usa pasos más gruesos para las menores frecuencias.

Se obtiene cuantización no uniforme al pasar primeramente la señal de voz análoga por un compansor (ley $-A$ o ley $-\mu$) y luego pasar la voz ya comprimida por un cuantizador estándar uniforme.

3.6.4 Vocoders

Los vocoders son una clase de codificadores de voz que analizan la señal de voz, transmitiendo parámetros derivados del análisis. En el receptor se sintetiza la voz usando estos parámetros. Todo Vocoder intenta modelar el proceso de generación de voz como un sistema dinámico tratando de cuantificar ciertas restricciones físicas del sistema. Son en general más complejos que los codificadores de forma de onda y alcanzan más economía en la tasa de transmisión, pero son menos robustos y su ejecución tiende a depender del locutor. El más conocido es el LPC (*Linear Predictive Coder*).

La Figura 3.7 muestra el modelo de generación de voz tradicional que es la base de todo vocoder.

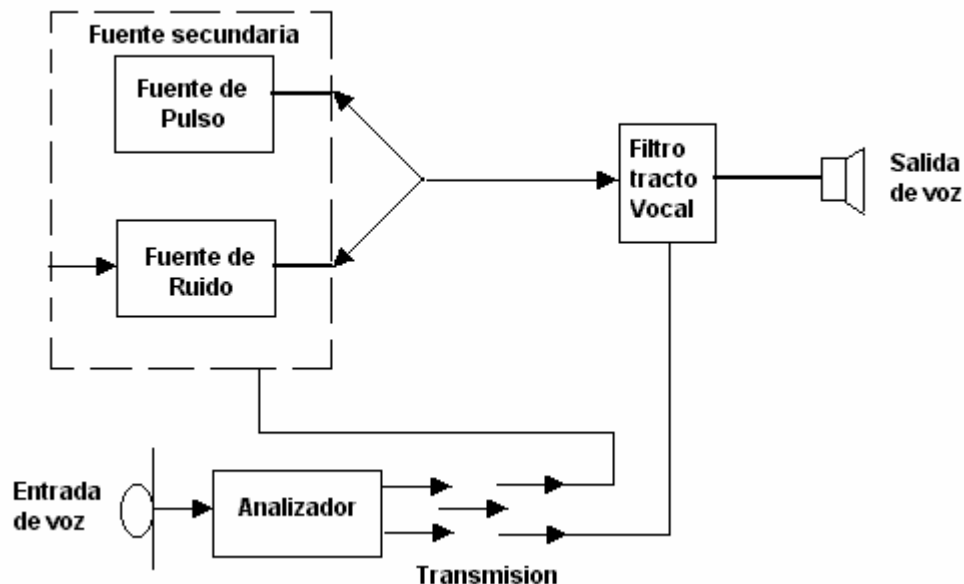


Figura 3.7: Modelo de generación de voz.

4 CAPA FÍSICA

En este capítulo se estudiará una de las capas que forman parte del modelo de referencia OSI, cuya funcionalidad es la proveer el medio de comunicación de más bajo nivel, como lo es la capa física. Se comenzará con una descripción de la forma en que puede ser modulada la información a ser puesta en el canal, para luego tratar los distintos medios físicos que hacen posible esta comunicación.

4.1 Sistemas de Modulación (digitales)

La modulación es el proceso de codificación de la información proveniente de una fuente tal que de esta forma sea más adecuada para el canal de transmisión. Aquí es importante diferenciar dos términos que serán utilizados para esta sección: Señal de *banda base*, la cual corresponde a la señal proveniente de la fuente, y una señal *pasa banda* la cual corresponde a una señal desplazada a frecuencias mucho más altas comparadas con su banda base. La modulación puede llevarse a cabo por variaciones de amplitud (AM), variaciones de fase (PM) o frecuencia (FM). Demodulación es el proceso inverso, el cual extrae la información desde una señal pasa banda (señal modulada).

Existen técnicas de modulación análogas y digitales, siendo las primeras utilizadas para comunicaciones de radio y televisión, y las segundas para comunicaciones inalámbricas y redes de datos.

En este capítulo serán solo tratadas técnicas de modulación digitales, siendo la contraparte temas ya en dominio del lector.

Algunas de las ventajas que ofrece la modulación digital versus la análoga son:

- ❑ Mayor inmunidad al ruido y robustez ante deterioro del canal.
- ❑ Mayor facilidad de multiplexación de varias formas de información.
- ❑ Más seguridad.

La eficiencia de un esquema de modulación puede ser medido (“*Wireless Communications*” (Theodore S. Rappaport), capítulo 5 “*Modulation Techniques for Mobile Radio*”) en términos de su *eficiencia de potencia* y *eficiencia de ancho de banda*. La primera describe la capacidad de un modulador para preservar la fidelidad del mensaje digital a niveles de potencia bajos. En un sistema de comunicación digital para incrementar la inmunidad al ruido es necesario incrementar la potencia de la señal. Sin embargo, la cantidad a la cual la potencia de la señal podría ser incrementada para obtener un cierto nivel de fidelidad depende del tipo particular de modulación empleado. La eficiencia de potencia, η_p de un esquema de modulación digital es una medida de que tan favorable es el intercambio entre fidelidad y potencia de la señal, y es frecuentemente expresado como la razón de la *energía por bit de la señal por la densidad espectral de la potencia del ruido* (E_b/N_0) requerido para recibir entrada para una cierta probabilidad de error.

La *eficiencia de ancho de banda* cuantifica la habilidad de un esquema de modulación para acomodar datos dentro de un ancho de banda limitado. (cubierto en capítulo anterior).

4.1.1 BFSK (Binary Frequency Shift keying)

En este esquema de modulación, la frecuencia de una señal portadora de amplitud constante es conmutada entre dos valores de acuerdo a los posibles estados del mensaje (1 y 0 binario), es similar a la modulación en frecuencia convencional, excepto que la señal modulante posee sólo dos estados. Dependiendo de la forma en que es introducido este cambio en la frecuencia, la señal FSK tendrá o fase continua (CPFSK) o fase discontinua entre bits.

$$S_{FSK}(t) = \sqrt{\frac{2E_b}{T_b}} \cos[(2\pi f_c + 2\pi \Delta f)t] \quad 0 \leq t \leq T_b \quad (\text{binario } 1) \quad (4.1)$$

$$S_{FSK}(t) = \sqrt{\frac{2E_b}{T_b}} \cos[(2\pi f_c - 2\pi \Delta f)t] \quad 0 \leq t \leq T_b \quad (\text{binario } 0) \quad (4.2)$$

Una forma de generar señales FSK es conmutar entre dos osciladores en base a si el bit es 0 o 1. La forma de onda generada no es continua (FSK discontinua). Esto conlleva problemas de desparramamiento espectral, y por lo tanto no es usado en sistemas inalámbricos.

En forma más frecuente se utiliza un tipo de modulación que es similar a un generador análogo de FM, con excepción que la señal modulada $m(t)$ es binaria.

$$S_{FSK}(t) = \sqrt{\frac{2E_b}{T_b}} \cos \left[2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\eta) d\eta \right] \quad (4.3)$$

A pesar que $m(t)$ es discontinua en la transición, la fase es proporcional a la integral de $m(t)$ y es continua.

4.1.2 BPSK (Binary Phase Shift keying)

La fase de una señal portadora de amplitud constante es conmutada entre dos valores de acuerdo a las dos posibles señales correspondientes a 1 y 0 binario. Normalmente las dos fases son separadas por 180° . Si la portadora sinusoidal tiene una amplitud A_c y una energía por bit $E_b = \frac{1}{2} A_c^2 T_b$, entonces la señal transmitida será:

$$S_{BPSK} = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c) \quad 0 \leq t \leq T_b \quad (\text{binario } 1) \quad (4.4)$$

o

$$S_{BPSK} = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c) \quad 0 \leq t \leq T_b \quad (\text{binario } 0) \quad (4.5)$$

4.1.3 QPSK (Quadrature Phase Shift keying)

Tiene el doble de eficiencia de ancho de banda que BPSK, pues 2 bits son transmitidos en un símbolo simple de modulación (00, 01, 10 y 11). La fase de la portadora toma uno de cuatro valores equiespaciados, donde cada uno de los valores de la fase corresponde a un único par de mensajes de bits.

$$S_{QPSK} = \sqrt{\frac{2E_s}{T_s}} \cos\left[2\pi f_c t + (i-1)\frac{\pi}{2}\right] \quad 0 \leq t \leq T_s \quad i = 1, 2, 3, 4 \quad (4.6)$$

Donde T_s es la duración del símbolo y es igual a dos veces T_b

Un dato importante de mencionar es que la probabilidad de error de bit del QPSK es igual a la del BPSK, pero el primero puede enviar el doble de datos por el mismo ancho de banda.

4.1.4 QAM (Quadrature Amplitud Modulation)

En los métodos anteriormente estudiados, se tenía un sistema de modulación el cual entrega una señal de amplitud constante y variante en su fase o frecuencia. La idea de QAM es permitir que junto con la fase, varíe también la amplitud. Para un sistema de modulación general (M-ario) se puede definir QAM como:

$$S_i(t) = \sqrt{\frac{2E_{\min}}{T_s}} a_i \cos(2\pi f_c t) + \sqrt{\frac{2E_{\min}}{T_s}} b_i \sin(2\pi f_c t) \quad 0 \leq t \leq T \quad i = 1, 2, \dots, M \quad (4.10)$$

Donde

E_{\min} es la energía de la señal con la más baja amplitud y a_i y b_i son enteros independientes elegidos de acuerdo a consideraciones métricas para cada punto.

4.1.5 CDMA (Code Division Multiple Access)

En este sistema, el mensaje es multiplicado por una señal de gran ancho de banda denominada señal “dispersa” (spreading), la cual es una secuencia de pseudo ruido que tiene una tasa de símbolo (chip) que es ordenes de magnitud más grande que la tasa de datos del mensaje. Todos los usuarios utilizan la misma frecuencia y pueden transmitir en forma simultanea. Cada usuario posee su código de pseudo ruido aleatorio el cual es aproximadamente ortogonal a los otros códigos. Así todos los otros códigos aparecen como ruido debido a la descorrelación. Para la detección el receptor necesita conocer el código usado por el transmisor.

A continuación se verá con mayor detalle este punto.

Secuencia de pseudo-ruido (PN)

Una secuencia de pseudo ruido o pseudo aleatoria es una secuencia binaria con una autocorrelación que se asemeja a la autocorrelación de ruido blanco limitado en banda. Es una secuencia determinística, pero posee características similares a la de una secuencia aleatoria binaria como tener un número casi igual de unos y ceros, muy baja correlación entre versiones desplazadas de la secuencia, y entre distintas secuencias, etc.

Este pseudo ruido es utilizado en sistemas DS-SS (*direct sequence spread spectrum*), los cuales expanden la banda base de datos al multiplicar los pulsos de datos con este pseudo ruido. Los sistemas DS-SS también son conocidos como sistemas CDMA.

Un pulso simple o símbolo de la secuencia PN es llamado un *chip*. En la Figura 4.0 se muestra un diagrama en bloques de un sistema DS con modulación de fase binaria.

La señal de espectro esparcido para un usuario simple puede ser representada como:

$$S_{ss}(t) = \sqrt{\frac{2E_s}{T_s}} m(t) p(t) \cos(2\pi f_c t + \theta) \quad (4.11)$$

Donde $m(t)$ es la secuencia de datos, $p(t)$ es el PN, f_c es la frecuencia de la portadora y θ es la fase de la portadora en $t=0$.

Cada símbolo en $m(t)$ representa un dato y tiene duración T_s . Cada uno de los pulsos en $p(t)$ representa un chip, y es usualmente rectangular de amplitud igual a 1 y -1, de duración T_c . La transmisión de los símbolos de datos y los chips cumplen con que la razón entre T_s y T_c es un valor entero. Si W_{ss} es el ancho de banda de $S_{ss}(t)$ y B es el de $m(t)\cos(2\pi f_c t)$, se cumple que $W_{ss} \gg B$.

En el receptor, la señal recibida pasa a través de un filtro y luego es multiplicada por una réplica local del pseudo código $p(t)$. Si $p(t)=1$, y entonces $p^2(t)=1$, lo cual lleva a:

$$S_1(t) = \sqrt{\frac{2E_s}{T_s}} m(t) \cos(2\pi f_c t + \theta) \quad (4.12)$$

La cual corresponde a la entrada del demodulador.

En CDMA el piso de ruido es determinado por la potencia de los múltiples usuarios. Lo que ocurre es que la señal más fuerte recibida capturará el demodulador en la estación base, aumentando el piso de ruido para las señales con niveles más débiles, y por lo tanto disminuyendo la probabilidad de que la señal débil sea recibida. Para evitar este efecto se utilizan controles de potencia en cada estación base, asegurando que cada señal recibida tenga el mismo nivel de potencia.

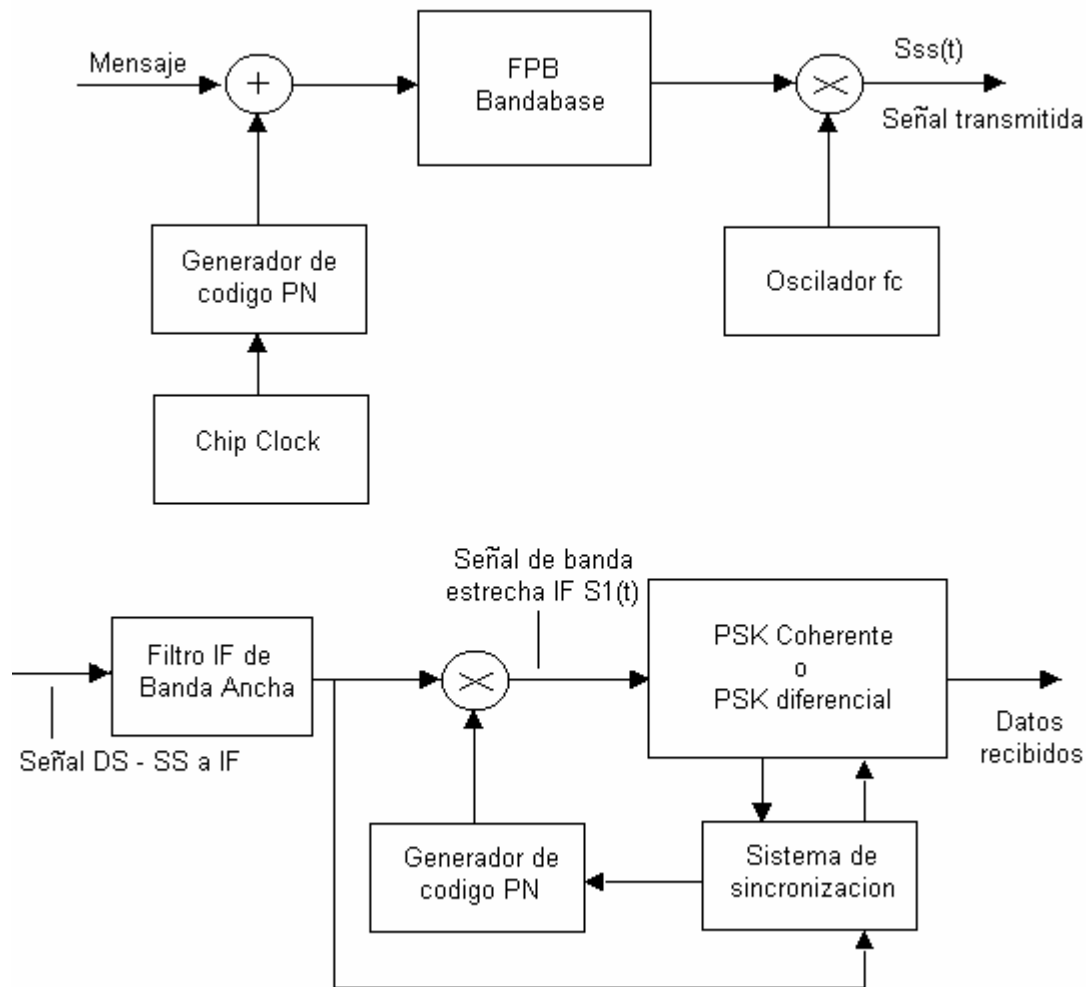


Figura 4.0: Diagrama de un sistema DS - SS

Algunas características del sistema CDMA se describen a continuación:

- Muchos usuarios comparten la misma frecuencia.
- El incremento en el número de usuarios, aumenta el piso de ruido de una manera lineal, luego no hay un número limitado de usuarios, más bien el rendimiento del sistema cae con el número de usuarios.
- El desvanecimiento por camino múltiple es reducido debido a que la señal está esparcida sobre un gran espectro.
- La tasa de datos en el canal es muy alta, con lo cual la duración del chip es muy corta y usualmente mucho menor que el retraso de dispersión del canal.

4.2 Multiplexación y Conmutación

Esta sección será muy breve y estará enfocada a aclarar ciertos aspectos dentro de los sistemas de comunicaciones que son básicos y servirán para entender razonamientos posteriores. Estos dos temas son: Multiplexación y Conmutación.

Estos temas se encuentran estrechamente ligados a los sistemas de telefonía fijos ya que este los utiliza ampliamente.

4.2.1 Multiplexación

Multiplexación intenta enviar información de distintas fuentes hacia varios destinos por el mismo medio (o canal). Existen dos categorías básicas para multiplexación:

- ❑ **FDM (Frequency Division Multiplexing):** el espectro de frecuencia del canal es dividido en bandas de frecuencia, en donde a cada usuario se le es asignada una de ellas. El ejemplo clásico de utilización de este tipo de multiplexación son las bandas de radio AM y FM, en donde a cada estación se le asigna una frecuencia determinada y existe una separación entre estas bandas.

Otra forma de multiplexación en frecuencias sería la técnica de WDM (Wavelength Division Multiplexing) utilizada para canales ópticos. A cada usuario se le asigna una longitud de onda por la cual transmitir información modulada a través de la fibra óptica (recordar la relación entre λ y la frecuencia f)[±]. Algunos de estos temas serán tratados con mayor detalle más adelante.

- ❑ **TDM (Time Division Multiplexing):** cada usuario utiliza un espacio de tiempo determinado para transmitir, es decir toma el ancho completo del canal por un período de tiempo corto. Sólo puede ser usada para datos digitales. Es utilizado ampliamente para las redes de telefónicas fijas y también inalámbricas (TDMA).

El sistema de multiplexación digital utilizado en las redes telefónicas fijas es el sistema de multiportadora digital T1 (para el caso de Estados Unidos) y E1 para el caso de Europa. T1 consta de 24 canales de banda de voz más dos canales de control, lo cual suman un ancho de banda de 1,544 [Mbps]. E1 son 30 canales de voz más 2 de control, totalizando un ancho de banda de 2,048 [Mbps].

4.2.2 Conmutación

La conmutación siempre es asociada al sistema de telefonía pública, y es el mejor ejemplo para entender la idea que existe detrás. Cuando se necesita establecer una comunicación y se disca un número, la central debe en base a este número lograr una conexión permanente entre el transmisor y receptor de la llamada. Para ello lo que se necesita es conectar por medio de pares de cobre los dos extremos de la comunicación estableciendo un camino, lo cual se hace intercambiando conexiones (antiguamente físicas y más recientemente lógicas), logrando así la comunicación directa entre ambos por medio del establecimiento del circuito. Si el circuito se corta en algún punto, la comunicación se pierde y es necesario volver a establecer la llamada. El proceso por el cual se intercambian

[±] $\lambda = c/f$ con $c = 300000$ m/s (velocidad de la luz)

distintos circuitos hasta lograra establecer la llamada se conoce como conmutación (particularmente de circuitos). En esta sección se tratará la conmutación tanto de circuitos como la de paquetes, describiendo las características de cada uno de los enfoques.

Conmutación de circuitos: Como se describió arriba, es la forma usual conocida en que los sistemas de comunicaciones telefónicas fijas establecen una llamada. Son básicamente caminos físicos (conexiones de pares de cobre) entre los dos extremos, gracias a la ayuda de dispositivos de conmutación, los cuales se encargan de establecer el camino (en donde pueden existir enlaces que no necesariamente sean pares de cobre, como microondas o fibra óptica, por las cuales fluye multiplexada información desde distintos orígenes a distintos destinos). Una vez constituido el camino, este es exclusivo para la comunicación que se efectúa, lo que hace limitado el número de comunicaciones posibles de establecer, eventualmente bloqueando las llamadas que entran cuando el sistema se encuentra copado. La comunicación que fluye por este canal se efectúa en tiempo real (sólo existiendo el retardo propio de propagación de la onda electromagnética por el medio).

Luego algunas características de este tipo de conmutación son: canal dedicado, transmisión en tiempo real, probabilidad no nula de bloqueo, transmisión de ancho de banda fija, no existe almacenamiento de información en la red.

Conmutación de paquetes: Otro enfoque para la conmutación sería permitir que exista almacenamiento, por corto periodos de tiempo, de información en la red lo cual permitiría el reenvío de la información hacia el destino específico. Para ello se dividen los datos a enviar en la fuente en segmentos de tamaño reducido conformando frames de datos los cuales serán enviados a través de la red. Estos paquetes pueden tomar trayectorias distintas entre uno y otro, por lo cual estos pueden llegar en desorden a su destino. Por ende no existe una reserva implícita de ancho de banda entre los dos extremos de comunicación, lo que hace que esta conmutación sea más flexible, pero a la vez menos confiable. Los equipos encargados de realizar esta conmutación son los ya conocidos enrutadores.

Algunas características son: no existe ruta establecida, permite almacenamiento de información en la red, uso dinámico de ancho de banda, es posible detectar errores y realizar conversiones de velocidad y código.

El ejemplo más emblemático de la utilización de este tipo de conmutación son las redes basadas en el protocolo IP (Internet).

4.3 Medios de Transmisión

Hasta el momento se han mencionado las formas en que las señales de información viajan por el canal de comunicación, al igual que la forma en que las redes manejan este flujo, pero aún no se ha hecho mención a una de las partes esenciales dentro del canal: la forma física que adopta este. Esta sección intenta presentar algunas de las características más relevantes de los medios físicos de transmisión utilizados en la actualidad.

4.3.1 UTP (Unshielded Twisted Pair)

El par trenzado es el medio más antiguo de establecer una comunicación entre dos puntos en tiempo real, y básicamente consiste de un par de alambres independientemente aislados y trenzados entre si, para evitar la interferencia (efectos de inducción) con pares adyacentes, debido a que dos alambres en paralelo se comportan como una antena.

En la Figura 4.1 se muestra un cable UTP (Par trenzado sin blindaje).

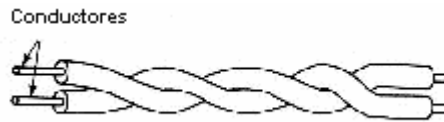


Figura 4.1: Par trenzado

La gran ventaja de este medio es su bajo costo v/s su rendimiento, pudiendo alcanzar varios Mbps durante algunos kilómetros.

Los cables que son utilizados para transportar datos a alta velocidad (10 – 100 Mbps) consiste de cuatro pares agrupados regularmente en una cubierta plástica (o teflón para reducir la diafonía) para su protección. Son normalmente utilizados en redes Ethernet con arquitectura de estrella y en las redes telefónicas privadas.

4.3.2 Cable Coaxial

El gran problema del par trenzado es que a frecuencias relativamente altas su sensibilidad a interferencias y pérdidas por radiación aumentan en forma considerable. Debido a esto lo que se utiliza para estos casos es el cable coaxial, el cual esta constituido por un conductor central rodeado por algún material aislante, como Pirex o poliestireno, el cual a su vez también se encuentra rodeado por otro conductor (trenzado) de tipo cilíndrico. Todo el conjunto se encuentra forrado por una envoltura plástica que sirve de protección. En la Figura 4.2 se aprecia un cable coaxial.

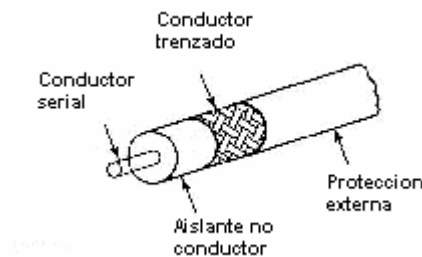


Figura 4.2: Cable coaxial

Estos cables tienen mejor inmunidad a interferencias (mayor blindaje), lo cual conlleva a un mejoramiento en la capacidad, llegando a transportar datos a velocidades de 1 a 2 Gbps en conductores de 1 Km.

Ser dividen en coaxial de banda base y banda ancha, siendo este último empleado para el transporte de señales de televisión. (TV cable)

4.3.3 Fibra Optica

Básicamente son canales cilíndricos muy pequeños contruidos de vidrio o plástico (o combinaciones) por los cuales es posible hacer incidir un haz de luz que se propagará por él hasta de un extremo a otro.

La base de esto sistemas se encuentra en la teoría óptica y en la naturaleza electromagnética de la luz.

Es un hecho físico que cuando un rayo atraviesa desde un material de cierta densidad a otro de distinta se producen refracciones en mayor o menor grado dependiendo de la densidad relativa entre los dos medios (las cuales se reflejan en sus índices de refracción respectivos). Si el ángulo de incidencia del rayo es mayor que un cierto umbral crítico, el rayo rebotará en forma integra, permaneciendo en el medio. Este es el principio básico de funcionamiento de la fibra, la cual esta compuesta de un núcleo, por el cual viaja la luz, y un manto con índice de refracción ligeramente menor, en el cual rebotarán los rayos contenidos en el núcleo. En la Figura 4.3 se muestra un rayo incidiendo sobre la interfaz aire/silicio en un ángulo α_1 y emergiendo en un ángulo β_1 . La cantidad de refracción depende de las propiedades de los dos materiales (índices de refracción). Para ángulos de incidencia sobre un ángulo crítico, la luz es refractada al interior del silicio, sin escapar al aire. Este rayo incidente sobre un ángulo crítico es atrapado en el interior de la fibra y así puede propagarse por muchos kilómetros virtualmente sin pérdidas.

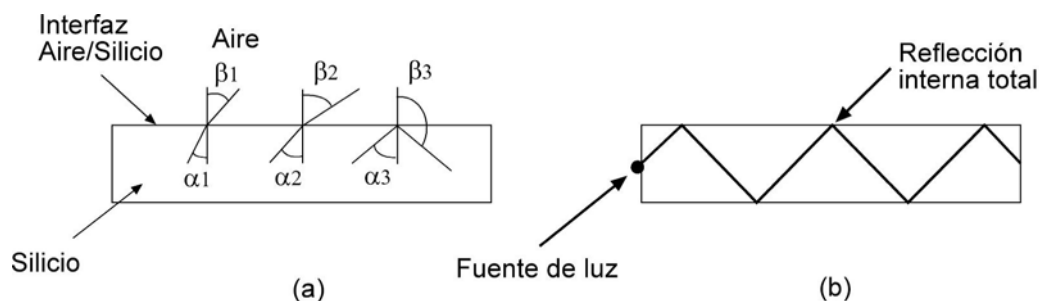


Figura 4.3: (a) Tres ejemplos de rayos incidiendo sobre la interfaz aire/silicio en ángulos distintos; (b) Luz atrapada por reflexión interna total

Las fibras ópticas están hechas de vidrio muy transparente. La atenuación a través de este depende de la longitud de onda de la luz incidente. La atenuación para el tipo de vidrio utilizado en fibras se muestra en la Figura 4.4 en decibels por kilómetro lineal de fibra. Luego, la atenuación esta dada por la formula:

$$\text{Atenuación en decibels} = 10 \log \frac{\text{Potencia transmitida}}{\text{Potencia recibida}} \text{ [dB]} \quad (4.13)$$

La figura 4.4 muestra tres bandas, las cuales son utilizadas en comunicaciones. Están centradas en 850 [nm], 1300 [nm] y 1550 [nm] respectivamente. Las dos últimas presentan menor atenuación que la primera que la primera banda, pero esta presenta la ventaja que tanto lasers como diodo emisores pueden ser contruidos del mismo material.

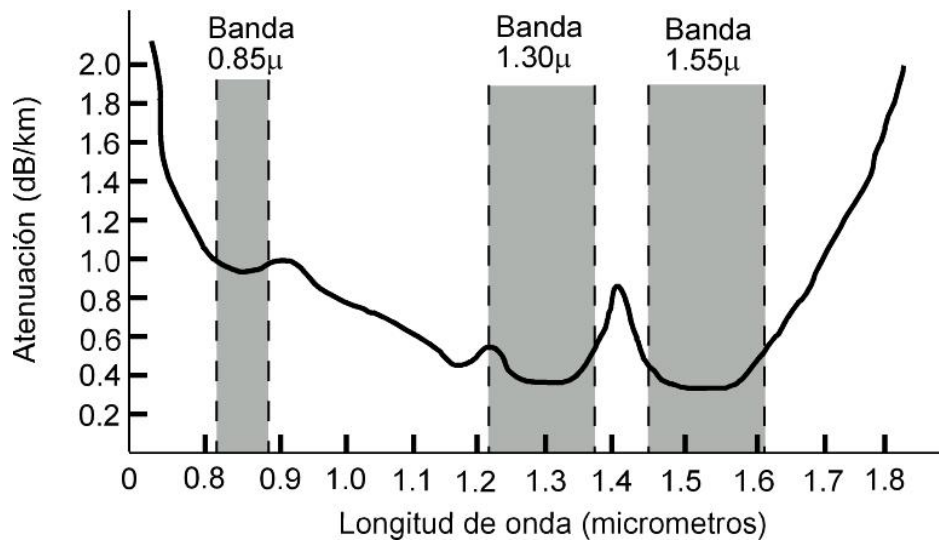


Figura 4.4: Atenuación de la luz a través de la fibra en la región infrarroja

La Figura 4.5 muestra una fibra óptica tipo

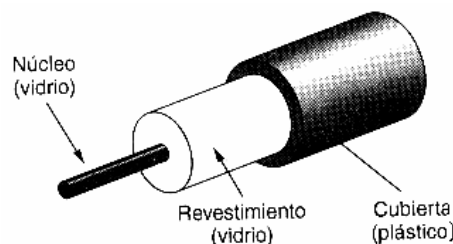


Figura 4.5: Fibra Óptica

Dado que pueden existir varios rayos rebotando con distintos ángulos de incidencia sobre la frontera, se puede hablar de fibras de múltiples modos (Multimodo).

También es posible reducir el tamaño del núcleo con tal que sólo sea posible la propagación de un único modo, y a este tipo de fibra se le denomina fibra Monomodo. Estas fibras son más caras debido al proceso de fabricación y la diferencia de tamaño con las multimodo (Multimodo: $\approx 125 \mu\text{m}$; Monomodo: $\approx 50 \mu\text{m}$). Alcanzan velocidades de datos de varios Gbps a una distancia de 30 Km..

Luego un sistema óptico se encuentra compuesto por una fuente de emisión de luz, la fibra óptica y el receptor de luz. La longitud de onda utilizada en la fuente dependerá de las características físicas de ésta y también (en mayor medida) de las características de la fibra óptica. La información es representada en dos estados (binario), siendo por ejemplo un 0 lógico la ausencia de pulsos de luz, y un 1 lógico la presencia de un pulso de luz. Para tal emisión por lo general se utilizan Diodos especialmente fabricados o LASER[±]. En el lado del receptor se suelen utilizar diodos sensibles a la luz (Fotodiodos).

[±] Light Amplification by Stimulated Emission of Radiation

4.4 Propagación de ondas

La utilización del espacio libre como medio de transmisión es ampliamente utilizado, debido principalmente a las limitaciones físicas que imponen los otros medios (como fibra óptica y cables de cobre) y a las condiciones geográficas existentes entre transmisor y receptor (espacios de agua muy grandes, montañas o terrenos desérticos). Es así como la utilización de este medio ha permitido el desarrollo de la telefonía móvil, enlaces microondas y comunicación satelital.

La propagación en el espacio libre de las ondas electromagnéticas, a menudo se llama *propagación de radiofrecuencia* (RF) o simplemente *propagación de radio*.

Para el análisis de propagación de ondas electromagnéticas en el espacio libre es ampliamente utilizado el concepto de rayos y frentes de onda. Los rayos son líneas dibujadas en el sentido de propagación y sirven para demostrar la dirección relativa de la onda. Sin embargo, un rayo no necesariamente representa la propagación de una sola onda electromagnética. Aquí se utiliza el concepto de frente de onda, el cual representa una superficie de fase constante de una onda.

La figura 4.6 muestra un frente de onda con una superficie que es perpendicular a la dirección de propagación (rectángulo ABCD). Cuando la superficie es plana, su frente de onda es perpendicular a la dirección de propagación.

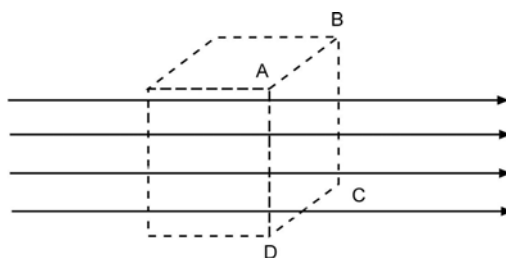


Figura 4.6: Onda Plana

En la figura 4.7 se muestra una fuente puntual, propagando varios rayos a partir de él, y el frente de onda correspondiente. Como se muestra en la figura, el frente de onda generado es una esfera de radio R y con centro en el punto de radiación. A este tipo de fuente, que propaga igualmente sus rayos en todas direcciones se conoce como radiador *isotrópico*.

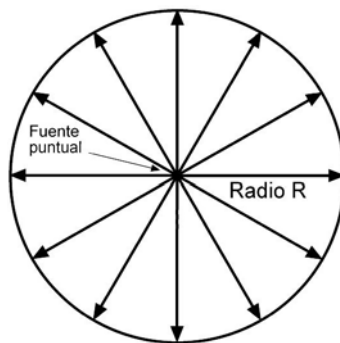


Figura 4.7: Frente de onda, desde una fuente puntual

4.4.1 Densidad de potencia e intensidad de campo

La proporción en la cual la energía cruza por una superficie dada, en el espacio libre, se llama *densidad de potencia*. La *intensidad de campo* es la intensidad de los campos eléctricos y magnéticos de una electromagnética propagándose en el espacio libre. Así la densidad de potencia es:

$$P = EH \quad (4.14)$$

en donde:

P	=	Densidad de potencia (W/m ²)
E	=	Intensidad de campo eléctrico en rms (V/m)
H	=	Intensidad del campo magnético en rms (At/m)

4.4.2 Impedancia característica del espacio libre

Las intensidades de los campos eléctricos y magnéticos de una onda electromagnética, en el espacio libre, se relacionan por la impedancia característica (resistencia) del espacio libre. La impedancia característica del espacio libre es:

$$Z_s = \sqrt{\frac{\mu_0}{\epsilon_0}} \quad (4.15)$$

en donde:

Z _s	=	Impedancia característica del espacio libre (ohms)
μ ₀	=	Permeabilidad magnética del espacio libre (1,26 × 10 ⁻⁶ H/m)
ε ₀	=	Permitividad eléctrica del espacio libre (8,85 × 10 ⁻¹² F/m)

Así, sustituyendo tendríamos que $Z_s = 377 \Omega$.

Por lo tanto, utilizando la ley de Ohm, obtenemos:

$$P = E^2/377 = 377 H^2 W/m^2 \quad (4.16)$$

$$H = E/377 \text{ At/m} \quad (4.17)$$

4.4.3 Frente de onda esférica

Como mencionamos con anterioridad una fuente isotrópica radia potencia a una proporción uniformemente constante, en todas direcciones. Un verdadero radiador isotrópico no existe, pero se aproxima bastante a una antena omnidireccional. Un radiador isotrópico produce un frente de onda esférica de radio R. En cualquier instante de tiempo, la potencia total radiada, P_r watts, esta uniformemente distribuida sobre la superficie de la esfera. Por lo tanto, la densidad de potencia en cualquier punto de la esfera es la potencia total radiada dividida por el área total de la esfera. Esto es

$$P_a = \frac{P_r}{4\pi R^2} \quad (4.18)$$

$$\begin{aligned}
 P_r &= \text{Potencia total radiada (watts)} \\
 R &= \text{Radio de la esfera (el cual es igual a la distancia desde cualquier punto en la superficie de la esfera a la fuente)} \\
 4\pi R^2 &= \text{Área de la esfera}
 \end{aligned}$$

en donde:

Igualando (x.3) con (x.5) se obtiene:

$$E = \frac{\sqrt{30P_r}}{R} \quad (4.19)$$

De la ecuación (x.5) se puede observar que entre más lejos se mueva el frente de onda de la fuente, más pequeña es la densidad de potencia. La relación se llama *ley inversa cuadrática*.

4.4.4 Atenuación

La ley inversa cuadrática describe la reducción de potencia con la distancia de la fuente. Conforme se aleja un frente de onda de la fuente, el campo electromagnético continuo, que es radiado desde esa fuente, se dispersa. Nada de potencia radiada se pierde o se disipa, porque el frente de onda se aleja de la fuente., es decir la onda se disipará sobre un área más grande, disminuyendo la densidad de potencia. Esta reducción de densidad de potencia es equivalente a la pérdida de potencia y comúnmente se llama *atenuación de onda*. La atenuación de la onda se expresa generalmente en términos del logaritmo común de la relación de densidad de potencia (pérdida en dB).

$$\gamma_a = 10 \log \frac{P_1}{P_2} \quad (4.20)$$

4.4.5 Absorción

La reducción en la densidad de potencia debida a la propagación del espacio no libre se llama *absorción*. Esta se produce a que en la atmósfera existen átomos y moléculas, algunos de los cuales son capaces de absorber las ondas electromagnéticas, transfiriéndose la energía desde las ondas a las partículas de la atmósfera. La absorción de onda por la atmósfera es análoga a una pérdida de potencia $I^2 R$.

La absorción de radiofrecuencias en una atmósfera normal depende de la frecuencia y es relativamente insignificante, debajo de aproximadamente los 10 GHz. La Figura 4.8 muestra la absorción en decibels por kilómetro, debido al oxígeno y al vapor de agua para radiofrecuencias superiores a 10 GHz. La atenuación debido a la absorción no depende de la distancia de la fuente radiante, pero si bastante de la distancia total que la onda propaga

por la atmósfera. Además, las condiciones atmosféricas anormales tales como lluvias fuertes o neblina densa absorben más energía que una atmósfera normal.

La absorción atmosférica para una propagación de onda de R_1 a R_2 es $\gamma(R_2 - R_1)$, en donde γ es el coeficiente de absorción.

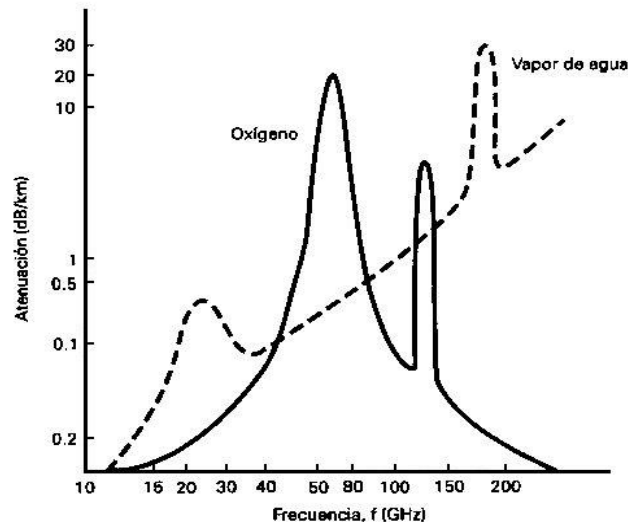


Figura 4.8: Absorción atmosférica de las ondas electromagnéticas

4.4.6 Propiedades ópticas de las ondas de radio

Las ondas que se propagan por el espacio pueden verse afectadas por efectos ópticos como la *refracción*, *reflexión*, *difracción* e *interferencia*. Tiene sentido hablar de propiedades ópticas para propagación de ondas de radio, debido a que las ondas de luz son ondas electromagnéticas de alta frecuencia. A continuación se presentan estos efectos desde una perspectiva de *trazos de rayos geométricos*.

Refracción

La refracción electromagnética es el cambio de dirección de un rayo conforme pasa oblicuamente, de un medio a otro, con diferentes velocidades de propagación. Debido a que la velocidad de propagación es inversamente proporcional a la densidad del medio en que se propaga, la refracción ocurre siempre que una onda pasa de un medio a otro medio de diferente densidad. La Figura 4.9 muestra la refracción de una onda plana en la frontera de dos medios de distinta densidad.

Siempre que un rayo pasa de un medio menos denso a uno más denso, efectivamente se dobla hacia la normal. El ángulo de incidencia es el ángulo formado entre la onda incidente y la normal, y el ángulo de refracción es el ángulo formado entre la onda refractada y la normal.

La cantidad de inclinación depende del índice de refracción de los materiales, el cual es:

$$n = \frac{c}{v} \quad (4.21)$$

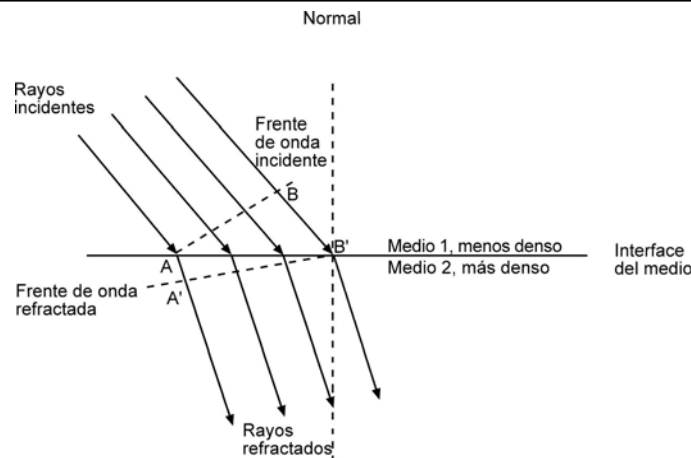


Figura 4.9: Refracción en una frontera plana entre dos medios

n = Índice de refracción (sin unidades)

en donde: c = Velocidad de la luz en el espacio libre (3×10^8 m/s)
 v = Velocidad de la luz en un material dado (m/s)

El índice de refracción es una función de la frecuencia, sin embargo en la mayoría de las aplicaciones es insignificante, aunque como se verá más adelante en este documento, resulta muy útil en aplicaciones con fibra óptica. Cómo reacciona una onda electromagnética, cuando conoce a la interface de dos materiales de transmisión que tienen índices de refracción diferentes, puede explicarse con la *ley de Snell*.

$$n_1 \sen \theta_1 = n_2 \sen \theta_2 \quad (4.22)$$

n_1 = Índice de refracción del material 1

en donde: n_2 = Índice de refracción del material 2
 θ_1 = Ángulo de incidencia (grados)
 θ_2 = Ángulo de refracción (grados)

y debido a que el índice de refracción de un material es igual a la raíz cuadrada de su constante dieléctrica:

$$\frac{\sen \theta_1}{\sen \theta_2} = \sqrt{\frac{\epsilon_{r1}}{\epsilon_{r2}}} \quad (4.23)$$

donde los ϵ_r representan las constantes dieléctricas de los medios correspondientes.

Reflexión

La reflexión electromagnética ocurre cuando una onda incidente choca con una barrera de dos medios y algo, o toda la potencia incidente no entra al segundo material. Las ondas que no penetran al segundo medio se reflejan. La Figura 4.10 muestra la reflexión de una onda electromagnética en una barrera plana entre dos medios.

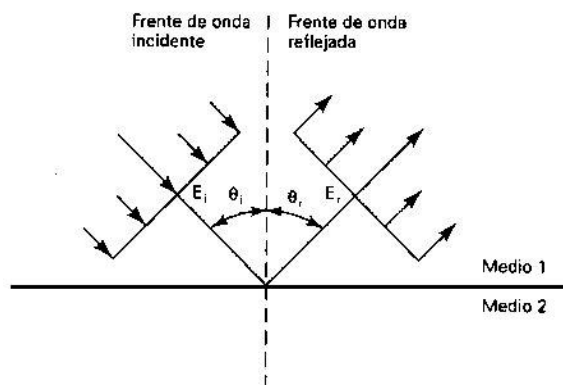


Figura 4.10: Reflexión electromagnética en una frontera plana

Debido a que todas las ondas reflejadas permanecen en el medio 1, las velocidades de las ondas reflejadas e incidentes son iguales, y por lo tanto el ángulo de reflexión es igual al ángulo de incidencia ($\theta_i = \theta_r$). Sin embargo, la intensidad del campo de voltaje reflejado es menor que la intensidad del campo de voltaje incidente. La relación de las intensidades de voltaje reflejado a incidente se llama *coeficiente de reflexión*, Γ . Para un conductor perfecto $\Gamma = 1$. Γ se usa para indicar la amplitud relativa de los campos incidentes y reflejados y, además, el desplazamiento de la fase en el punto de reflexión. Luego, el coeficiente de reflexión es:

$$\Gamma = \frac{E_r e^{j\theta_r}}{E_i e^{j\theta_i}} = \frac{E_r}{E_i} e^{j(\theta_r - \theta_i)} \quad (4.24)$$

Γ = Coeficiente de reflexión (sin unidades)

En donde:

- E_i = Intensidad de voltaje incidente (volts)
- E_r = Intensidad de voltaje reflejado (volts)
- θ_i = Fase incidente (grados)
- θ_r = Fase reflejada (grados)

La porción de voltaje incidente total que no está reflejada se llama *coeficiente de transmisión de potencia* (T). Para un conductor perfecto $T = 0$. La ley de conservación de energía establece que:

$$T + |\Gamma|^2 = 1 \quad (4.25)$$

Si el medio 2 no es un conductor perfecto, algunas de las ondas incidentes lo penetran y se absorben. Las ondas absorbidas establecen corrientes en la resistencia del material y la energía se convierte en calor. La fracción de potencia que penetra al medio 2 se llama el *coeficiente de absorción*.

Difracción

La difracción se define como la modulación o redistribución de energía, dentro de un frente de onda, cuando pasa cerca del extremo de un objeto opaco.

Cuando un frente de onda pasa cerca de un obstáculo o discontinuidad de dimensiones comparables en tamaño a una longitud de onda, no se puede utilizar el simple análisis geométrico para explicar los resultados y el *principio de Huygens* es necesario.

El principio de Huygens indica que cada punto de un frente de onda esférica determinado se puede considerar como una fuente secundaria de puntos de ondas electromagnéticas, desde donde se irradian hacia fuera otras ondas secundarias. Esto se ilustra en la Figura 4.11.

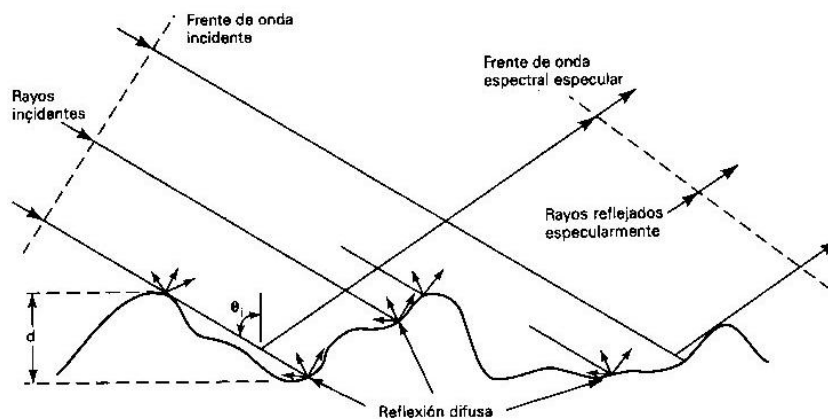


Figura 4.11: La reflexión de una superficie semiáspersa

Cada fuente puntual secundaria irradia energía hacia afuera en todas direcciones. Sin embargo el frente de onda continúa en su dirección original porque la cancelación de ondas secundarias ocurre en todas direcciones excepto de frente.

Cuando se considera un frente de onda plano y finito, la cancelación en direcciones aleatorias es incompleta. En consecuencia, el frente de onda se extiende hacia fuera o se *dispersa*. El efecto de dispersión se llama difracción.

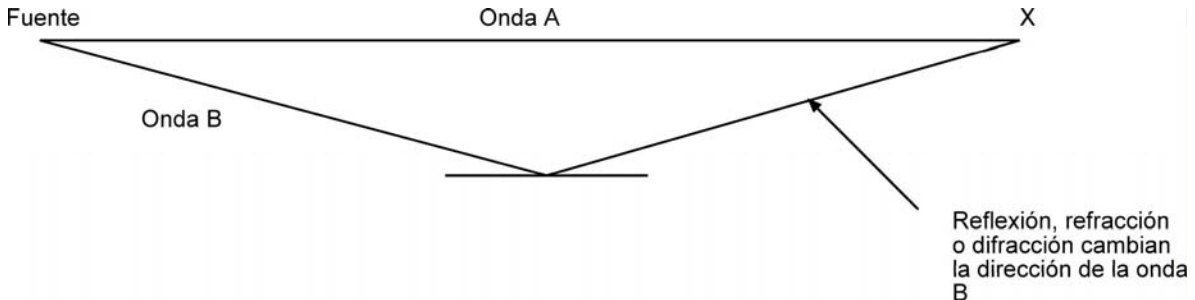
Interferencia

La interferencia de ondas de radio ocurre cuando dos o más ondas electromagnéticas se combinan de tal forma que el funcionamiento del sistema se degrada. La interferencia se sujeta al principio de superposición lineal de ondas electromagnéticas y ocurre cada vez que dos o más ondas ocupan el mismo punto en el espacio. El principio de superposición lineal indica que la intensidad total de voltaje, en un punto determinado del espacio, es la suma de los vectores de ondas individuales.

El voltaje total resultante de la suma no es simplemente la suma de las magnitudes de los dos vectores, sino la suma de las fases de los dos. Luego, dependiendo de los ángulos de fase de los dos vectores puede ocurrir suma o resta.

La Figura 4.12 muestra la interferencia entre dos ondas que viajan por el espacio libre. Puede verse que en el punto X las dos ondas ocupan la misma área de espacio, sin embargo la onda B ha viajado una trayectoria diferente a la onda A y, por lo tanto, puede que sus

ángulos de fase relativos sean distintos. Si la diferencia en la distancia viajada es un múltiplo de un número entero impar de una mitad de longitud de onda, se lleva a cabo el refuerzo. Si la diferencia es un múltiplo de un número entero par de una mitad de longitud de onda, ocurre la cancelación total. Para frecuencias por debajo de VHF, las longitudes de onda relativamente grandes evitan que la interferencia se convierta en un problema



importante.

Figura 4.12: Interferencia de onda electromagnética

4.4.7 Propagación de ondas

Hay tres formas de propagación de ondas electromagnéticas: ondas de tierra, ondas espaciales (que incluyen ondas directas y ondas reflejadas a tierra), y propagación de onda del cielo. La Figura 4.13 muestra los modos normales de propagación entre dos antenas.

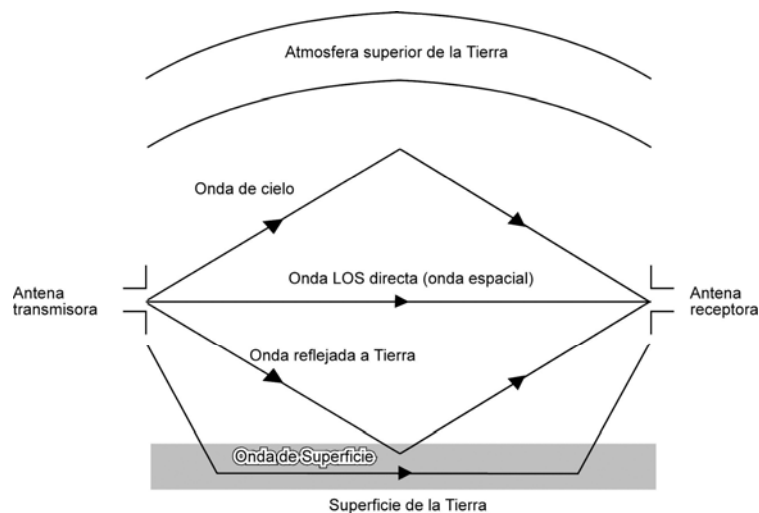


Figura 4.13: Modos normales de la propagación de ondas

Propagación de ondas de tierra

Las ondas viajan por la superficie de la tierra, por lo que suelen llamarse ondas superficiales. Estas ondas deben estar polarizadas verticalmente, pues el campo eléctrico, en una onda polarizada horizontalmente, estaría paralelo a la superficie de la Tierra, y dichas ondas harían cortocircuito por la conductividad de la Tierra. Con las ondas de tierra, el campo eléctrico variante induce voltajes en la superficie de la Tierra, que causa que fluyan corrientes que son muy similares a las de una línea de transmisión. Las pérdidas de

onda de tierra se incrementan rápidamente con la frecuencia, por lo tanto, se limita generalmente a frecuencias por debajo de los 2 MHz.

La propagación de ondas de tierra se utiliza comúnmente para comunicaciones de barco a barco y de barco a tierra (debido a las propiedades conductoras del agua salada), para radio navegación, y para las comunicaciones marítimas móviles.

Las desventajas de la propagación de ondas de tierra son:

1. Requieren de una potencia relativamente alta para la transmisión.
2. Están limitadas a frecuencias muy bajas, bajas y medias (VLF, LF, MF) que requieren de antenas grandes.
3. Las pérdidas por tierra varían considerablemente con el material de la superficie.

Y sus ventajas son:

1. Con suficiente potencia de transmisión, se pueden utilizar para comunicarse entre dos ubicaciones cualesquiera en el mundo.
2. No se ven relativamente afectadas por los cambios en las condiciones atmosféricas.

Propagación de ondas espaciales

Las ondas espaciales viajan por la parte inferior de la atmósfera de la Tierra. Incluyen ondas directas y reflejadas de tierra, como se muestran en la Figura 4.14. Las primeras viajan en línea recta entre las antenas transmisora y receptora. La propagación de ondas espaciales directas se llama comúnmente transmisión de línea de vista (LOS).

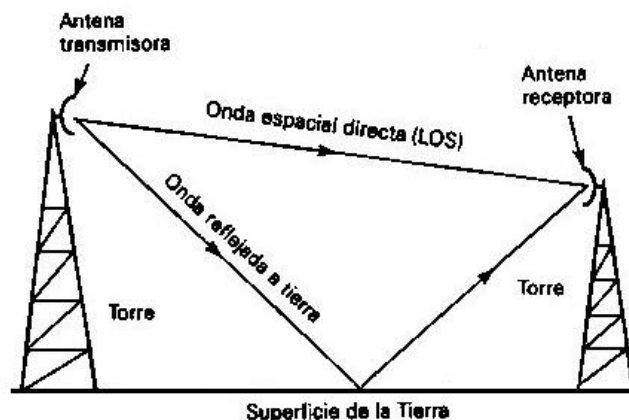


Figura 4.14: Propagación de ondas espaciales

La intensidad de campo en la antena receptora depende de la distancia que hay entre las dos antenas (atenuación y absorción) y si las ondas, directa y reflejada a tierra, están en fase.

La curvatura de la Tierra presenta un horizonte para la propagación de ondas espaciales llamado *radio horizonte*. Debido a la refracción atmosférica, el radio horizonte se extiende más allá del horizonte óptico para la atmósfera estándar común. El radio horizonte es aproximadamente cuatro tercios del horizonte óptico. La troposfera causa la refracción, debido a cambios en su densidad, temperatura, contenido de agua vapor, y conductividad relativa. El radio horizonte puede alejarse simplemente elevando las antenas por arriba de la superficie de la Tierra, con torres o colocándolas arriba de montañas o edificios altos.

La Figura 4.15 muestra el efecto del radio horizonte. El radio horizonte de línea de vista para una sola antena se da como:

$$d = \sqrt{2h} \quad (4.26)$$

d = Distancia a radio horizonte (millas)

en donde:

h = Altura de la antena sobre el nivel del mar (pies)

Por tanto, para una antena transmisora y una receptora, la distancia entre las dos antenas es:

$$d = \sqrt{2h_t} + \sqrt{2h_r} \quad (4.27)$$

d = Distancia total (millas)

en donde: h_t = Altura de la antena transmisora (pies)

h_r = Altura antena receptora (pies)

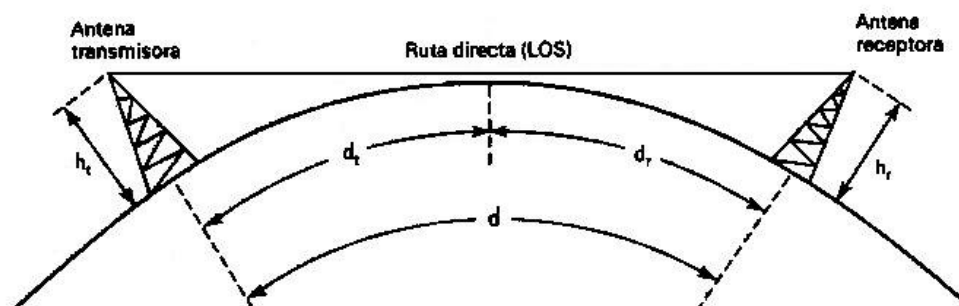


Figura 4.15: Ondas espaciales y radio horizonte

Propagación de ondas de cielo

Las ondas electromagnéticas que se dirigen por encima del nivel del horizonte se llaman ondas de cielo. Las ondas se envían hacia el cielo, donde son reflejadas o refractadas nuevamente a la Tierra por la ionosfera. La ionosfera es la región de espacio localizada aproximadamente de 50 a 400 Km arriba de la superficie de la Tierra y es la porción más alta de la atmósfera, por lo tanto absorbe grandes cantidades de energía radiante del sol, que ioniza las moléculas de aire, creando electrones libres. Cuando una onda de radio pasa a través de la ionosfera, el campo eléctrico de la onda ejerce una fuerza en los electrones libres, haciéndolos vibrar. Esto reduce la corriente, lo que equivale a reducir la constante dieléctrica. A su vez este efecto provócale incremento de la velocidad de propagación y hace que las ondas se doblen alejándose de las regiones de alta densidad de electrones (es decir se incrementa la refracción). Entre más alta la densidad de iones, mayor la refracción. Además, debido a que la composición de la atmósfera no es uniforme y a las variaciones en la temperatura y densidad, se encuentra estratificada, en tres capas: las capas D, E y F. En la Figura 4.16 puede verse que las tres capas varían en ubicación y en densidad de ionización con la hora del día.

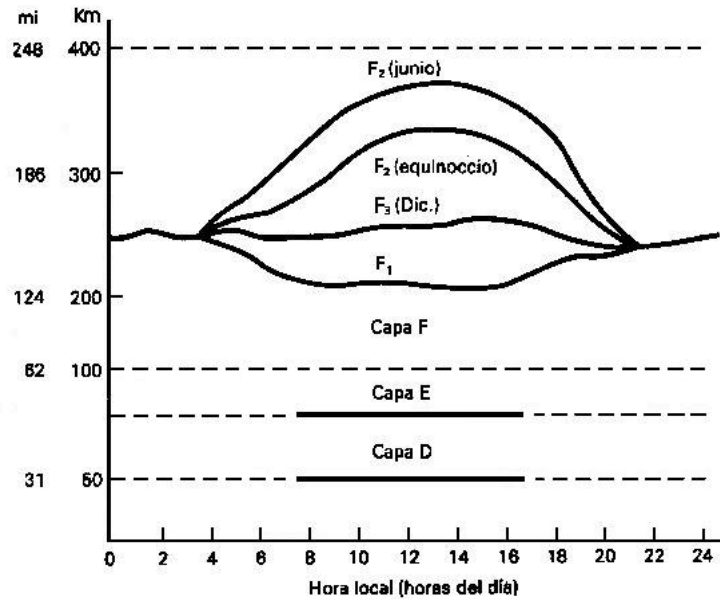


Figura 4.16: Capas ionosféricas

Capa D: es la capa inferior de la ionosfera y se localiza entre 50 a 100 Kilómetros arriba de la superficie de la Tierra. Debido a que es la capa más lejana al sol, existe poca ionización, sin embargo los iones de esta capa pueden absorber cantidades considerables de energía electromagnética. La ionización depende de la altitud del sol en el horizonte, así que desaparece en la noche. La capa D refleja ondas VLF y LF y absorbe ondas MF y HF.

Capa E: se localiza entre 100 y 140 Kilómetros sobre la superficie de la Tierra. Esta capa auxilia la propagación de ondas de superficie MF y refleja ondas HF un poco durante el día. Al igual que la capa anterior, prácticamente desaparece durante la noche.

Capa F: esta capa está compuesta por dos subcapas F_1 y F_2 . Durante el día la primera se ubica entre 140 a 250 Kilómetros y la segunda entre 140 a 300 Kilómetros sobre la superficie de la Tierra. Durante la noche ambas subcapas se combinan y forman una sola. La capa F_1 absorbe y atenúa algunas ondas HF, aunque la mayoría de las ondas pasan a través de F_2 , cuando se refractan nuevamente a la Tierra.

4.5 Antenas

Una antena es un conductor metálico capaz de radiar y recibir ondas electromagnéticas. Se usa como interfaz entre el transmisor y el espacio libre o el espacio libre y el receptor (o más específicamente entre la guía de onda y el espacio libre).

En la Figura 4.17 se muestran los patrones de onda estacionarias de voltaje en una línea de transmisión. La línea termina en un circuito abierto, que representa una discontinuidad abrupta en la onda de voltaje incidente en la forma de una inversión de fase. La inversión de la fase resulta cuando parte del voltaje incidente se irradia, en lugar de ser reflejado de

nuevo a la fuente. La energía radiada se propaga lejos de la antena en forma de ondas transversales.

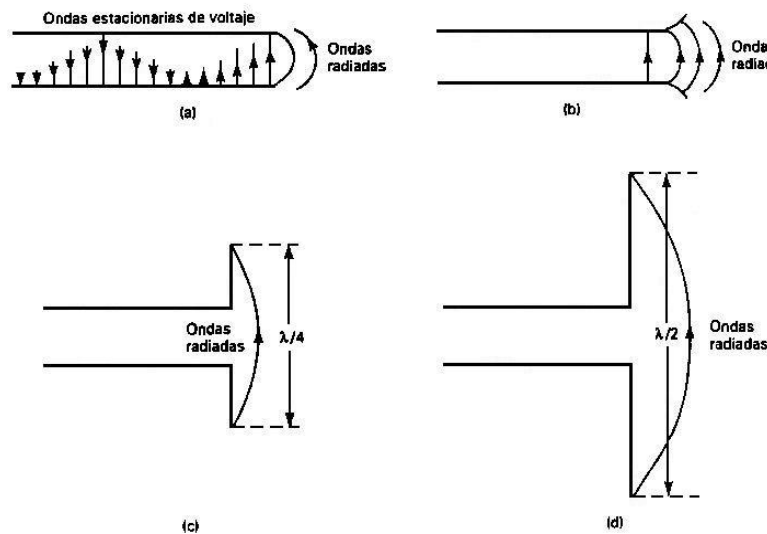


Figura 4.17: (a) Radiación de una línea de transmisión; (b) conductores difusores; (c) antena Marconi; (d) antena Hertz

La eficiencia de radiación es la relación entre la energía radiada y la energía reflejada, y para el caso de una línea de transmisión abierta es en extremo baja. Para radiar más energía es necesario separar más los conductores. Una antena así se llama dipolo y se muestra en la figura anterior.

4.5.1 Patrón de radiación

Un patrón de radiación es un diagrama polar que representa las intensidades de los campos o las densidades de potencia en varias posiciones angulares en relación con una antena. Si el patrón de radiación se traza en relación a la intensidad del campo eléctrico (E) o de la densidad de potencia (P) se llama patrón de radiación absoluto. Si se traza la intensidad de campo o la densidad de potencia en relación al valor en un punto de referencia, se llama patrón de radiación relativo. En la Figura 4.18 se muestra un patrón de radiación absoluto y relativo para una antena no especificada.

En la figura el haz principal se encuentra en una dirección de 90° y se llama *lóbulo principal*. Puede existir más de un lóbulo principal. También hay un haz secundario o lóbulo *menor* en la dirección de -180° . Debido a que el lóbulo principal propaga y recibe la mayor parte de la energía, este lóbulo se llama lóbulo frontal. Los lóbulos adyacentes al frontal se llaman *laterales*, y los que están en dirección opuesta al lóbulo frontal se llaman lóbulos *traseros*. La relación de la potencia del lóbulo frontal con la potencia del lóbulo trasero se llama relación frontal trasera, y la relación del lóbulo frontal con el lateral se llama relación frontal lateral. La línea que divide el lóbulo principal desde el centro de la antena en la dirección de máxima radiación se llama *línea de tiro*.

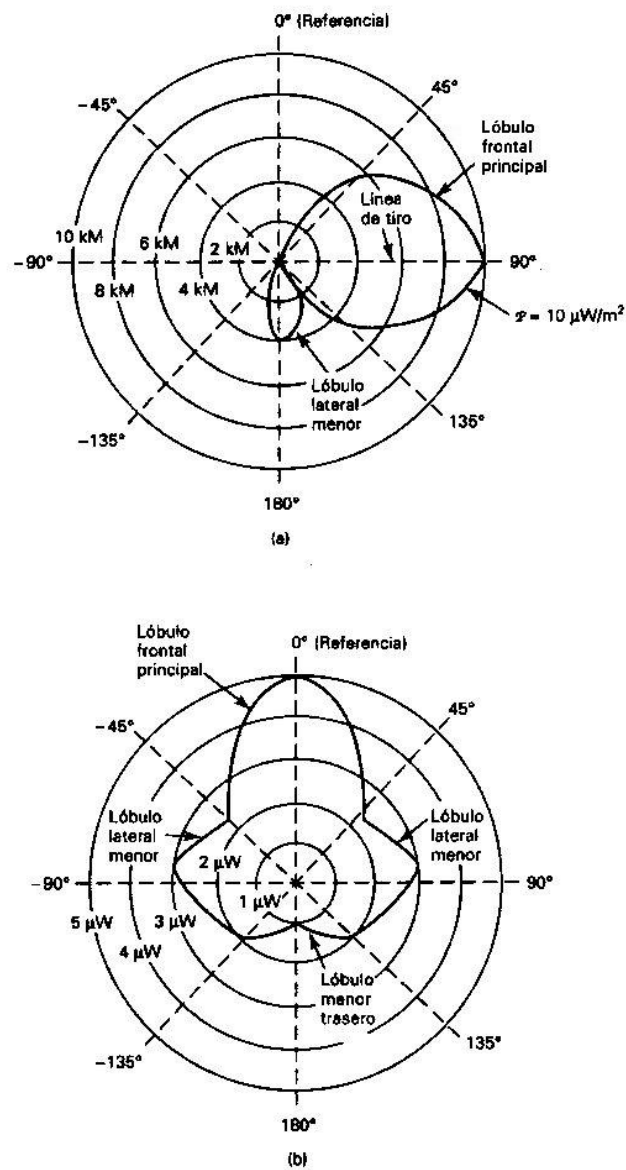


Figura 4.18: (a) Patrón de radiación absoluto; (b) Patrón de radiación relativo

4.5.2 Campos cercanos y lejanos

El término campo cercano se refiere al patrón de campo que esta cerca de la antena, y el término campo lejano se refiere al que esta a gran distancia. La potencia que alcanza el campo lejano continua radiando lejos y nunca regresa (campo de radiación), a diferencia de lo que ocurre en el campo cercano (o de inducción) donde en la segunda mitad del ciclo, la potencia regresa a la fuente. Debido a que importa el campo radiado, los patrones de radiación de la antena se dan para el campo lejano. El campo cercano se define como el

área dentro de una distancia D^2/λ de la antena, en donde λ es la longitud de onda y D el diámetro de la antena en las mismas unidades.

4.5.3 Resistencia de radiación y eficiencia de antena

Parte de la potencia suministrada a la antena se convierte en calor y se disipa. Esto se modela por medio de una *resistencia de radiación* de la antena en ca y es igual a:

$$R_r = \frac{P}{i^2} \quad (4.28)$$

R_r = Resistencia de radiación (ohms)

en donde: P = Potencia radiada por la antena (Watts)
 i = Corriente de la antena en el punto de alimentación (Amperes)

La resistencia de radiación es la resistencia que, si reemplazara la antena, disiparía exactamente la misma cantidad de potencia de la que irradia la antena.

La *eficiencia* de la antena es:

$$\eta = \frac{P_r}{P_r + P_d} \times 100 \quad (4.29)$$

η = Eficiencia de antena (%)

en donde: P_r = Potencia radiada por la antena (Watts)
 P_d = Potencia disipada por la antena (Watts)

En la Figura 4.19 se muestra un circuito equivalente simplificado para una antena. En términos de resistencia y corriente, la eficiencia de la antena es:

$$\eta = \frac{i^2 R_r}{i^2 (R_r + R_e)} = \frac{R_r}{R_r + R_e} \quad (4.30)$$

η = Eficiencia de antena

en donde: R_r = Resistencia de radiación (ohms)
 R_e = Resistencia de la antena efectiva (ohms)
 i = Corriente de antena (amperes)

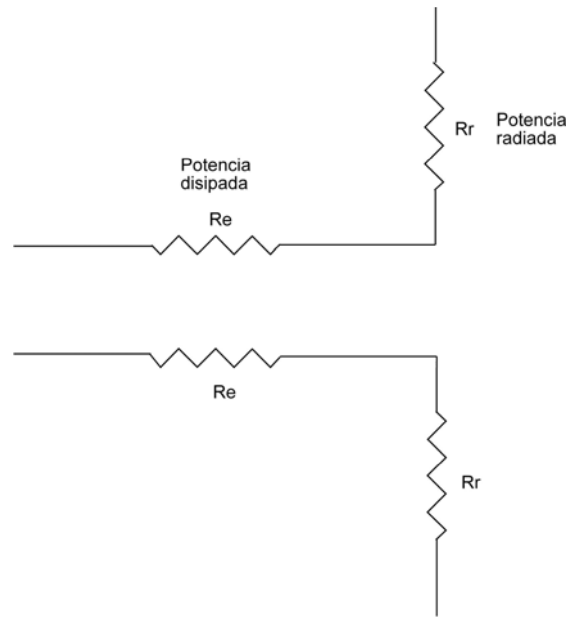


Figura 4.19: Circuito equivalente simplificado de una antena

4.5.4 Ganancia directiva y ganancia de potencia

La ganancia directiva es la relación de la densidad de potencia radiada en una dirección en particular con la densidad de potencia radiada al mismo punto por una antena de referencia, suponiendo que ambas antenas irradian la misma cantidad de potencia. El patrón de densidad de potencia relativa de una antena es realmente un patrón de ganancia directiva si la referencia de la densidad de potencia se toma de una antena de referencia estándar, que por lo general es una antena isotrópica. Luego, la ganancia directiva es:

$$D = \frac{P}{P_r} \quad (4.31)$$

D = Ganancia directiva (sin unidades)

P = Densidad de potencia en algún punto de una antena determinada (W/m^2)

P_{ref} = Densidad de potencia en el mismo punto de una antena de referencia (W/m^2)

donde:

La ganancia de potencia es igual a la ganancia directiva excepto que se utiliza el total de potencia que alimenta la antena (o sea, que se toma en cuenta la eficiencia de la antena). La ganancia de potencia es:

$$A_p = D\eta \quad (4.32)$$

Si una antena no tiene pérdidas, irradia 100% de la potencia de entrada y la ganancia de potencia es igual a la ganancia directiva. En decibels la ganancia de potencia es:

$$A_p = 10 \log \frac{P\eta}{P_{ref}} \quad (4.33)$$

Para una referencia isotrópica, la directividad de un dipolo de media onda es aproximadamente de 1,64 (2,15 dB). Es usual indicar la ganancia de potencia en decibels si se refiere a un dipolo de media onda, sin embargo, si es en referencia a un radiador isotrópico, la cantidad de decibels se indica como dBi. Es importante mencionar que la antena realmente no amplifica la potencia de entrada, sino que solo concentra su potencia radiada en una dirección en particular. Debido a que la antena es un dispositivo recíproco, su patrón de radiación es también su patrón de recepción. Por lo tanto, las antenas receptoras tienen una directividad y una ganancia de potencia exactamente como las antenas transmisoras.

4.5.5 Potencia radiada isotrópica efectiva

La potencia radiada isotrópica efectiva (EIRP) se define como una potencia de transmisión equivalente y se expresa como:

$$\text{EIRP} = P_r A_t \quad \text{Watts} \quad (4.34)$$

Donde P_r es la potencia total radiada y A_t es la ganancia directiva de la antena transmisora. En dBm se tiene:

$$\text{EIRP(dBm)} = 10 \log \frac{P_r}{0,001} + 10 \log A_t \quad (4.35)$$

EIRP o solo ERP es la potencia equivalente que tendría que radiar una antena isotrópica para alcanzar la misma densidad de potencia en la dirección seleccionada en un punto determinado, como otra antena.

Para determinar la densidad de potencia en un punto determinado, se expande la ecuación x.5 para incluir la ganancia de la antena transmisora y se escribe como:

$$P = \frac{P_r A_t}{4\pi R^2} \quad (4.36)$$

Para analizar una antena como receptora, es útil emplear el concepto de área efectiva. Si una antena es puesta en el campo de una onda electromagnética polarizada linealmente, la potencia disponible en los terminales de la antena es igual al área efectiva multiplicada por la potencia por unidad de área llevada por la onda, es decir:

(4.37)

$$C = PA$$

C = Potencia disponible (watts)
 P = Densidad de potencia de la onda (W/m²)

$$A = \text{Área efectiva (m}^2\text{)}$$

en donde:

La definición anterior no considera pérdidas en la antena. El término de densidad de potencia P para la relación de arriba corresponde a la densidad de potencia de la antena transmisora, la cual ya se mostró en la ecuación 4.18.

Una relación muy útil existente entre el área efectiva de una antena y su ganancia esta dada por:

$$A = \frac{\lambda^2 A_r}{4\pi} \quad (4.38)$$

Donde A_r corresponde a la ganancia de la antena receptora. Luego, combinando las ecuaciones 4.36, 4.37 y 4.38 se tiene:

$$C = \frac{P_r A_t A_r \lambda^2}{16\pi^2 R^2} \quad (4.39)$$

Donde C es la potencia disponible en la antena receptora, medida en Watts. De esta última ecuación se obtiene la *razón de transferencia de potencia*, dada por

$$\frac{C}{P_r} = \frac{A_t A_r \lambda^2}{16\pi^2 R^2} \quad (4.40)$$

Esta relación es cierta para campos lejanos, es decir debe existir una distancia mayor o igual a $2D^2/\lambda$.

4.5.6 Polarización de la antena

La polarización de la antena se refiere sólo a la orientación del campo eléctrico radiado desde ésta. Una antena puede polarizarse en forma lineal (polarizada horizontal o verticalmente), en forma elíptica o circular. Así, si la antena irradia una onda electromagnética polarizada verticalmente, la antena se define como polarizada verticalmente.

4.5.7 Ancho de haz de la antena

El ancho de haz de la antena es sólo la separación angular entre los dos puntos de media potencia (- 3 dB) en el lóbulo principal del patrón de radiación del plano de la antena, por lo general tomado en uno de los planos “principales”. En la Figura 4.20 el ancho de haz corresponde al ángulo formado entre los puntos A, X y B. El ancho de haz corresponde al ancho de haz de media potencia.

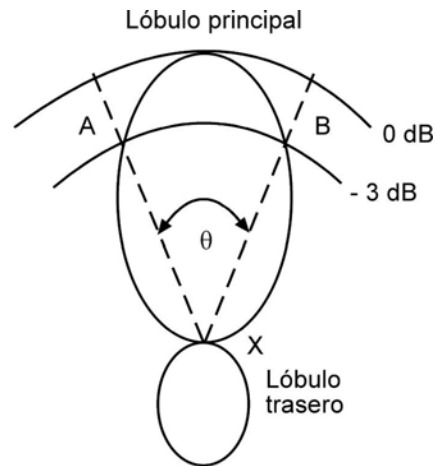


Figura 4.20: Ancho de haz de antena

4.5.8 Antenas básicas

Doblete elemental

Este es un dipolo corto (comparado con la mitad de la longitud de onda) que tiene corriente uniforme en toda su longitud. Sin embargo, la corriente varía en forma sinusoidal en tiempo y en cualquier instante, es decir:

$$i(t) = I \sin(2\pi ft + \theta) \quad (4.41)$$

- $i(t)$ = Corriente instantánea
 I = Amplitud pico de la corriente de RF (amperes)
 F = Frecuencia (hertz)
 t = Tiempo instantáneo (segundos)
 θ = Ángulo de fase (radianes)

en donde:

Su patrón de intensidad de campo eléctrico relativo se muestra en la Figura 4.21.

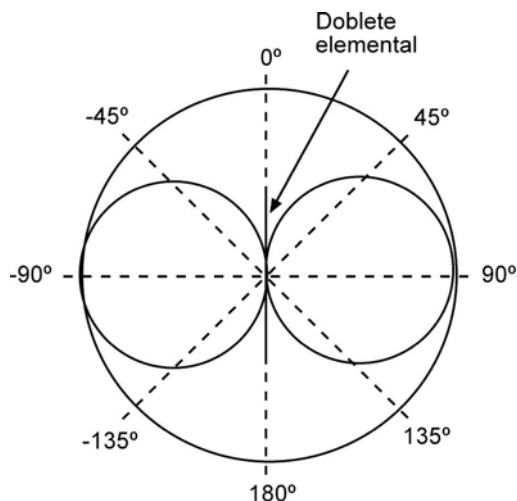


Figura 4.21: Patrón de radiación relativa para un doblete elemental en un plano perpendicular a los dos ejes del dipolo

El patrón de potencia relativa es:

$$P = \frac{30\pi l^2 I^2 \sin^2 \phi}{\lambda^2 R^2} \quad (4.32)$$

Donde ϕ corresponde al ángulo entre el eje de la antena y la dirección de radiación y l la longitud del dipolo.

Dipolo de media onda

Es conocida como *antena de Hertz*. Es una antena resonante, es decir, es múltiplo de un cuarto de longitud de onda de largo y de circuito abierto en el extremo más lejano.

El patrón de radiación de espacio libre para un dipolo de media onda depende de la colocación horizontal o vertical de la antena con relación a la superficie de la Tierra. La Figura 4.22 muestra los distintos patrones posibles para esta antena.

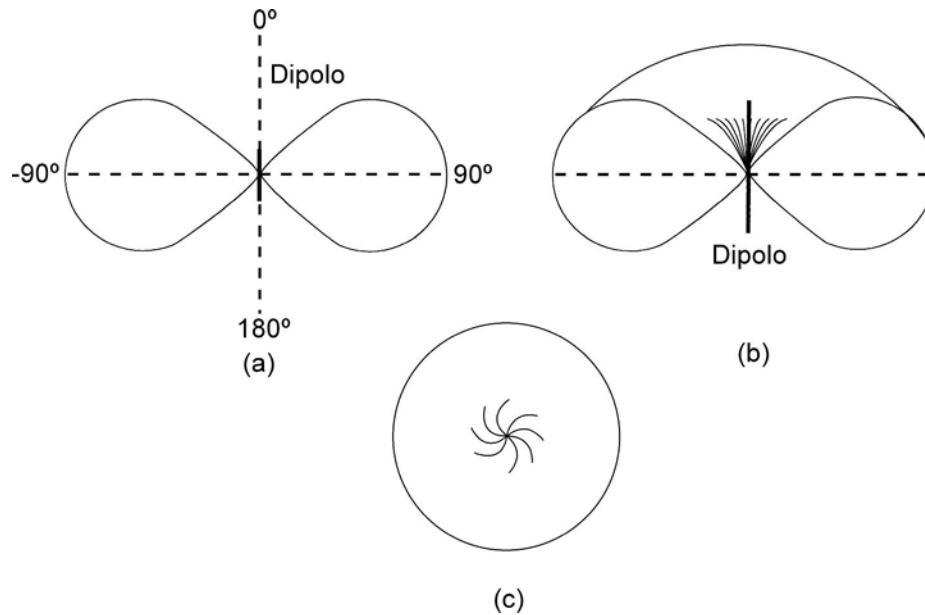


Figura 4.22: (a) vista vertical de un dipolo montado verticalmente; (b) vista de sección transversal; (c) vista horizontal

Antena aterrizada

La antena monopolo de un cuarto de longitud de onda montada verticalmente con el extremo inferior conectado directamente a tierra se llama *antena Marconi*. La corriente máxima ocurre en los extremos aterrizados, que causa un flujo de alta corriente a través de tierra. Para reducir las pérdidas, la tierra deberá ser un buen conductor.

La Figura 4.23 muestra un patrón de radiación para una antena aterrizada de un cuarto de longitud de onda. Puede verse que la mitad inferior de cada lóbulo se cancela con las ondas reflejadas de tierra. Esto por lo general no tiene ninguna consecuencia porque la radiación en dirección horizontal se incrementa, incrementando así la radiación a lo largo de la superficie de la Tierra y mejorando la cobertura de área.



Figura 4.23: Patrones de radiación para una antena aterrizada

5 CAPA DE ENLACE

Cuando se habla de capa de enlace se debe pensar en el canal de comunicación existente entre dos máquinas pertenecientes a una misma red. El canal actúa como un cable por el cual los datos circulan en el mismo orden en que son emitidos.

La capa de enlace se encarga de ofrecer una forma confiable de entregar esta información, a través de la capa física, hacia la máquina que se encuentra en el otro extremo. A pesar de lo sencillo que pudiese parecer este enfoque, surgen una serie de problemas que hacen necesario definir servicios bien específicos que permitan mediante interfaces apropiadas la comunicación con las otras capas del modelo OSI.

Luego, la capa de enlace debe proveer funciones específicas que permitan controlar la comunicación a este nivel de abstracción. Estas funciones incluyen entregar interfaces de servicios bien definidas, determinar de qué forma los bits que serán entregados a la capa física en forma de frames (marcos), manejar los errores de transmisión y regular el flujo para que un transmisor rápido no sature a uno lento.

5.1 Servicios

La capa de enlace puede ser diseñada para ofrecer varios servicios, los cuales dependerán del tipo de sistema y red en las cuales se trabaje. Los servicios más comúnmente provistos son:

- Servicio no confiable y no orientado a la conexión
- Servicio confiable no orientado a la conexión
- Servicio orientado a la conexión y confiable

El primero es el más básico de ellos y sólo se encarga de entregar los marcos (frames) a enviar a la capa física sin importar si el receptor está dispuesto a recibirlos y sin esperar confirmación previa antes de seguir enviando. Esto implica que no se toman medidas en la eventualidad que existiesen pérdidas en el medio producidos por ruido en el canal. Para terminar la comunicación sólo se detiene el envío de frames. Este tipo de comunicación es utilizado cuando la probabilidad de error es muy baja, delegándose este control a alguna capa superior. También es útil para tráfico en tiempo real, en donde no es necesaria la confirmación de los datos recibidos (debido al retardo que estas producen).

El segundo de los servicios continúa con la misma filosofía de envío sin esperar saber si el receptor se encuentra en condiciones de aceptar los frames (sin conexión), pero se implementa acuse de recibo por parte del receptor. Así el transmisor sabe si el frame llegó en forma correcta, y si no es así posibilita el reenvío de este por parte del transmisor. Este tipo de comunicación es menos utilizada que la anterior y se implementa para canales poco confiables.

El último tipo de servicio es orientado a conexión y confiable, es decir, antes de que comience el envío de datos, el transmisor establece una conexión con el receptor. Esto permite que los datos sean transferidos por el canal establecido, a la vez que se asegura

que los frames enviados lleguen en el orden correcto y en su totalidad. Una vez concluida la comunicación, la conexión establecida desaparece liberándose los recursos que esta utilizaba en ambos extremos.

5.2 Enmarcado

La capa física se encarga de enviar y recibir flujos de bits por el canal sin importar el tamaño de este, ni tampoco preocuparse de que estos contengan información con error. Luego el labor de la capa de enlace proveer este tipo de servicio.

Entonces la capa de enlace divide el flujo de bits en segmentos denominados marcos (o frames), a los cuales se les agrega un campo que permita a la capa paralela en la otra máquina detectar posibles errores en los datos. Este campo generalmente es conocido como chequeo de suma (Checksum) pues los algoritmos que utiliza para su calculo se basan en sumas de los datos contenidos en el frame. El método más utilizado para este efecto es el CRC (Cyclic Redundancy Code).

Una consideración que surge al dividir el flujo de bits en frames, es la necesidad de delimitarlo, pudiéndose reconocer en forma clara un principio y un fin de marco, para que pueda ser recibido y chequeado en el receptor. Algunas de las formas para solucionar este problema son: contador de caracteres, caracteres de comienzo y término con carácter de relleno, banderas de comienzo y fin con bit de relleno y violaciones de código de capa física.

El último método es interesante de comentar debido a que es utilizado para las redes Ethernet y 802.3. La violación de código es posible siempre y cuando la forma de los bits que se utilizan incluyan transiciones de estado. Por ejemplo, si para representar un 1 lógico se utiliza una transición desde nivel alto a bajo (5 a 0 Volt) y para el cero una transición de nivel bajo a alto (0 a 5 Volt) entonces en el momento que no exista señal, el canal sólo tendrá valores de 0 Volt, los cuales no representan ningún estado en particular, produciéndose una violación al código. Cuando se detecte transición se entenderá que es el inicio de un nuevo frame. (El código utilizado es el Manchester)

5.3 Manejo de errores

Para poder realizar un debido manejo de errores es necesario poseer un método que permita detectarlos. Luego de esta detección se pueden utilizar dos diferentes enfoques: corrección del o los errores y petición de reenvío de la información.

Como ya se mencionó con anterioridad, para la detección de errores se incluye un campo dentro del frame el cual contiene cierto tipo de información obtenida mediante algún método que permite en el lado del receptor realizar una serie de cálculos con el fin de comprobar la integridad del frame recibido. Generalmente son utilizados métodos de detección de paridad, efectuados por códigos polinomiales (como CRC). Dependiendo del número de detecciones requeridas será el tamaño del campo a incluir en el frame.

Una vez detectado errores, una posibilidad sería intentar corregirlos. El problema que surge con este enfoque es que para efectuar de manera eficiente una corrección es necesario enviar mayor información redundante en el frame. Los métodos utilizados

generalmente se basan en el cálculo de distancias y matrices de corrección. (por ejemplo distancia de Hamming). Mientras mayor capacidad de corrección, mayor es el número de información a enviar, lo cual hace a estos métodos poco viables.

El otro enfoque es sólo pedir retransmisión del frame equivocado, o mejor aún enviar acuses de recibos positivos o negativos dependiendo de la integridad de los datos recibidos. A pesar de sonar simple existen una serie de consideraciones a tener en cuenta en esta implementación.

Si el frame es enviado y por alguna razón desaparece sin nunca llegar a destino, el transmisor quedara esperando la confirmación. Para evitar este problema se utilizan temporizadores, los cuales al espirar provocan la retransmisión automática el último frame que no recibió confirmación. Otro problema surge cuando la confirmación desaparece, pudiendo llegar muchas veces el mismo frame debido a retransmisión o que los frames no lleguen en orden secuencial. Para evitar estos problemas se soporta un número máximo de repeticiones, dando por terminada la comunicación si no se recibe acuse, a la vez que los frames son enumerados para que sean posibles de ordenar en el receptor.

5.4 Control de flujo

El control de flujo trata de dar solución a la situación en que un transmisor de mayor velocidad satura los buffer de recepción del receptor de menor velocidad. Esto se puede dar debido a diferencias en las máquinas (distintas tecnologías y capacidades), lo cual se desea evitar.

Para ello se implementan distintas políticas de control de flujo, siendo una de las más utilizadas el de confirmación de envío. Es decir el receptor autoriza enviar una cantidad determinada de frames, pero una vez cumplida la cuota debe esperar hasta una nueva confirmación con la cantidad de marcos a enviar.

Este problema es tratado en todas las capas del modelo, siendo de particular interés en la capa de transporte para el modelo TCP/IP.

A continuación se presenta un ejemplo de protocolo de capa de enlace, conocido como HDLC.

5.5 HDLC

IBM desarrollo a mediados de la década de los 70's el protocolo de Control de Enlace de Datos Síncrono (SDLC) para ser usado en ambientes de redes SNA (Systems Network Architecture). La particularidad de este protocolo es que fue el primer protocolo de capa de enlace de datos basado en operaciones orientadas al bit, síncronas.

De aquí derivaron un número importantes de protocolos los cuales diferían en detalles de implementación con SDLC, en particular se menciona HDLC (High-level Data Link Control), el cual fue implementado para redes X.25.

HDLC soporta una serie de tipos de enlace y topologías. Puede ser utilizada con enlaces punto a punto, multipunto, medios de transmisión half y full-duplex y redes de conmutación de paquetes y de circuitos.

Aquí se identifican dos tipos de nodos: *primarios* y *secundarios*. Los nodos primarios controlan la operación de las otras estaciones (secundarias). Las primarias preguntan en un determinado, en el cual las secundarias pueden transmitir si tienen datos.

Los nodos primarios y secundarios pueden conectarse de 3 distintas formas:

- ❑ Punto a punto (sólo un primario con un secundario)
- ❑ Multipunto (un primario con varios secundarios)
- ❑ En lazo: los nodos forman un bucle donde el primario se encuentra en primera posición y le siguen los secundarios, los cuales se traspasan entre si la información proveniente del primario.

El formato del frame se muestra en la Figura 5.1.

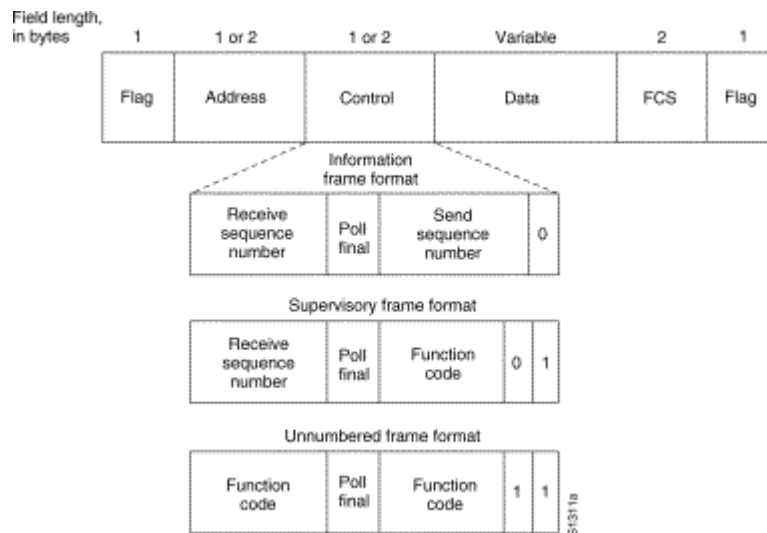


Figura 5.1: Formato de frame SDLC

Se distinguen tres tipos de frame, los cuales varían en el campo de control: Información, Supervisión, y no numerado.

Algunos de los campos son:

Flag: Delimitadores del Frame. (01111110)

Address: Identificador de terminal, utilizada para cuando existen múltiples terminales. Para comunicación punto a punto es a veces utilizada para distinguir comandos, de respuestas.

Control: Se distinguen dependiendo del tipo de frame. Se usa para incluir números de secuenciamiento, y mensajes de control desde los nodos primarios a secundarios y viceversa.

Datos: Campo que contiene la información.

FCS: Frame Check Sequence. Es un código CRC para detección de error.

5.6 Subcapa de Acceso al Medio

En el capítulo 2 se trató el tema de topologías de redes, en donde se presentan las redes de difusión (broadcast) las cuales tienen la característica de utilizar un solo medio físico como canal de comunicación entre varios usuarios conectados. Como se puede vislumbrar, estas redes presentan un problema particular para la implementación de los servicios descritos anteriormente, debido a su naturaleza compartida.

El problema fue resuelto con la agregación de una Subcapa dentro de la capa de enlace, encargada de manejar el acceso de los distintos usuarios al medio compartido la cual es conocida como Subcapa MAC (Media Access Control).

Algunos enfoques que plantean una solución de implementación se mencionan y describen a continuación:

5.6.1 ALOHA

Fue creado por Norman Abramson y sus colegas de la Universidad de Hawai, con la intención de ser utilizado para transmisiones de radio terrestres, la idea es aplicable a cualquier sistema que utilice un medio de transmisión compartido.

Se distinguen dos enfoques:

ALOHA puro: La idea básica es que el que necesite transmitir datos por el medio que lo haga. Si hay dos usuarios utilizando el medio en el mismo instante se provocará una colisión de frames, los cuales serán destruidos debido a ella. Gracias a que es un medio compartido existe realimentación de colisiones, es decir en algún momento el frame destruido llegará hasta el emisor el cual entenderá que debe retransmitirlo. La retransmisión no se hace de forma inmediata, debido a que si así fuese, se volvería a provocar la colisión entre los mismo frames. Entonces se hace esperar un tiempo aleatorio al transmisor antes de poner un nuevo frame en el medio, reduciendo así la probabilidad de colisiones.

Con el fin de determinar la eficiencia de un canal ALOHA, se presenta el siguiente análisis[±]: Se considera un conjunto infinito de usuarios interactivos en la red. Cada usuario pone un frame en el canal y lo verifica para comprobar si llegó con éxito. Si es así el usuario ve la respuesta y continúa enviando. Si no, el frame se retransmite una y otra vez hasta que ocurre un éxito.

Se entiende por tiempo de frame el tiempo necesario para transmitir una frame de longitud fija. La población de usuarios genera frames nuevos según una distribución de Poisson con una media de S frames por tiempo de frame. Para un rendimiento razonable se espera $0 < S < 1$.

Las estaciones también generarán retransmisiones de frames debido a colisiones. Supóngase, además que la probabilidad de k intentos de transmisión por tiempo de frame, viejos y nuevos combinados, también es Poisson, con una media de G por tiempo de frame. ($G \geq S$). Con todas las cargas, el rendimiento es sólo la carga ofrecida, G , por la probabilidad de que una transmisión tenga éxito (es decir, $S = G P_0$, donde P_0 es la probabilidad de que un marco no sufra una colisión).

[±] Extraído desde “Redes de Computadoras”, Andrew S. Tanenbaum, Tercera Edición Prentice – Hall Hispanoamérica, pags. 248 -249

Un frame no colisionará si no hay transmisión durante un intervalo de tiempo de frame desde su inicio, como se muestra en la Figura 5.2. Sea t el tiempo requerido para enviar un marco. Si cualquier otro usuario generó un frame entre el tiempo t_0 y t_0+t , el final de ese frame chocará con el comienzo del frame sombreado. De manera parecida, cualquier otro frame que salga entre t_0+t y t_0+2t chocará con el final del frame sombreado.

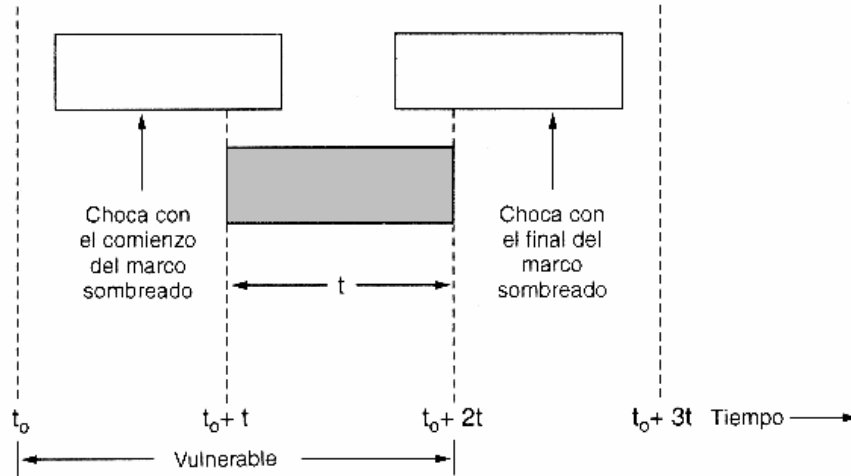


Figura 5.2: Período vulnerable para el frame

La probabilidad de que k frames sean generados durante un tiempo de frame está dado por la distribución de Poisson:

$$P_r[k] = \frac{G^k e^{-G}}{k!} \quad (5.1)$$

En una ranura de dos tiempos de frame de longitud, el número medio de frames es de $2G$. La probabilidad de que no se inicie otro tráfico durante todo el período vulnerable está dada entonces por $P_o = e^{-2G}$. Si $S = GP_o$ se obtiene:

$$S = Ge^{-2G} \quad (5.2)$$

El rendimiento máximo ocurre a $G=0.5$, con $S=1/2e$, que es aproximadamente 18,4%.

ALOHA ranurado: Se propone dividir el tiempo en intervalos discretos, cada uno del tamaño de un frame, todo sincronizado por medio de un terminal que genere un reloj. En comparación con el anterior, ahora se debe esperar por el comienzo de una nueva ranura de tiempo. La probabilidad de que un frame no sufra una colisión es ahora igual a $P_o = e^{-G}$ con lo que se tiene:

$$S = Ge^{-G} \quad (5.3)$$

Luego el máximo uso de canal aumenta a un $\sim 37\%$.

La Figura 5.3 muestra la gráfica para los dos tipos de ALOHA.

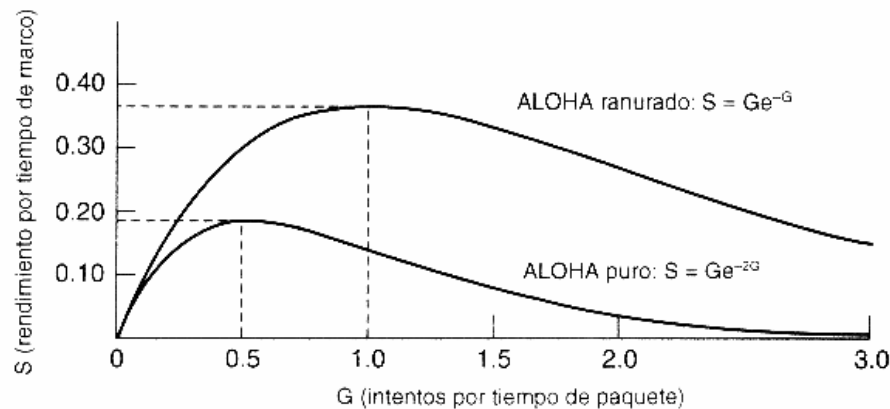


Figura 5.3: Rendimiento de los sistemas ALOHA

5.6.2 CSMA (Carrier Sense Multiple Access)

Este protocolo se encuentra basado en el anterior, pero a diferencia de este, la idea detrás de CSMA es escuchar el medio, y si este se encuentra disponible, transmitir. Para ello se utilizan tres algoritmos distintos:

CSMA persistente-1: Si al momento de escuchar el canal se encuentra ocupado, el transmisor espera hasta que se desocupe. Si logra transmitir y ocurre una colisión, se espera un tiempo aleatorio y se vuelve al inicio. Al momento de transmitir no considera el tiempo que demora la señal en propagarse por el canal de un extremo a otro.

CSMA no persistente: Antes de enviar se escucha el canal, y si esta desocupado pone los datos en el. Si esta ocupado no se queda esperando que el canal se desocupe, sino que vuelve a censar el medio después de un tiempo aleatorio.

CSMA persistente-p: utiliza canales ranurados. Escucha el canal y si este esta desocupado transmite con una probabilidad p . Con probabilidad $q = 1 - p$ no transmite y espera la siguiente ranura de tiempo.

CSMA/CD (CSMA con detección de colisión): si al momento de encontrarse transmitiendo un frame hacia el medio se detecta una colisión, inmediatamente de termina la transmisión, cortando el marco afectado. Esto mejora el ancho de banda al ahorrar el tiempo empleado para propagación por el medio de los frames defectos completos. Este protocolo es ampliamente utilizado por redes LAN Ethernet (IEEE802.3).

6 CAPA DE RED

La capa de red es la encargada de hacer llegar los paquete desde un punto de emisión a otro de recepción. Para lograr este objetivo debe existir un conocimiento de los caminos que debe seguir el mensaje, a través de distintas redes y generalmente varios enrutadores hasta llegar a su destino. Para ello se vale de direcciones lógicas las cuales servirán para establecer políticas de ruteo dentro de la subred de comunicaciones, y en las cuales se basarán los protocolos.

En este capítulo se describen los tipos de servicios brindados por esta red, los algoritmos que hacen posible el establecimiento de rutas y por último la descripción del protocolo utilizado en Internet (IP).

6.1 Consideraciones de diseño

Los servicios proporcionados por la capa de red están principalmente orientados a la capa superior (transporte), la cual solo se preocupa de entregar información a enviar, sin preocuparse de que tipo de red se es parte. Esto quiere decir, que al momento de definir los servicios a especificar para la capa de transporte estos deben ser independientes de la tecnología de red utilizada, pues la capa superior no considera (y no tiene por que saber) en que red se encuentra.

Otro punto importante a considerar es que al momento que la capa de transporte desee enviar datos no considera si al host, al que va dirigido el mensaje, utiliza la misma tecnología de red. Luego, la capa de red ,debe ser capaz, mediante un sistema de direccionamiento universal, de entregar el paquete con la información sin preocuparse de este aspecto.

Además del direccionamiento, existe una serie de factores que se deben tener en cuenta al momento de transportar información entre redes de tecnología distintas. Algunos de ellos son: tamaño de paquete que no exceda el máximo permitido en la red (Ej: en Ethernet el largo máximo se fija en 1500 bytes, incluyendo el encabezado), resolución entre direcciones lógicas (de red) y direcciones físicas (de enlace), pérdida de paquetes por saturación en redes lentas, etc.

En cuanto a los servicios proporcionados por esta capa, ya sabemos que uno de los principales es hacer llegar los paquetes entregados por la capa de transporte a su destino. La discusión surge cuando se plantea de que forma lo hará y que tipo de servicio de entrega proporcionará. Para este último punto existen dos enfoques fundamentales:

- ❑ Confiable orientada a la conexión
- ❑ No confiable sin conexión.

Existen más combinaciones, pero estás son las más utilizadas. La elección surgirá dependiendo del tipo de red en que se trabaje y las aplicaciones que se desean implementar

sobre estas. Hay veces en que es preferible esperar un tiempo más largo pero que asegura una entrega íntegra y correcta (como transmisión de archivos) y otras en que la velocidad es lo importante (voz e imágenes en tiempo real). Ejemplo de estos dos enfoques se dan en redes ATM para el primer caso y en redes IP para el segundo. En este capítulo solo se tratará el segundo de los casos. Un detalle importante de mencionar es que ATM efectivamente es orientada a la conexión, pues se establecen circuitos virtuales entre extremos de comunicación, pero no utiliza métodos de acuse de recibo, lo que no impide que se pierda una celda sin que el transmisor se percate. Si implementa suma de verificación de error, por lo cual se puede estrictamente decir que es un protocolo semiconfiable a este nivel.

6.2 Protocolos de ruteo

Para llevar un paquete de un extremo a otro, es necesario implementar ruteo, es decir, mediante algún criterio, escoger la mejor ruta de salida que permite alcanzar el destino a un paquete entrante a un enrutador. Este ruteo ocurre a nivel de subredes.

Los algoritmos de enrutamiento se clasifican en:

- *Algoritmos no adaptables (estáticos)*: las rutas se calculan por adelantado, no considerando mediciones de tráfico o cambios de topología.
- *Algoritmos adaptables (dinámicos)*: basan su decisión en mediciones de topología y en algunos casos tráfico de la red.

Ahora se presentarán la descripción de algunos de estos algoritmos, partiendo por los estáticos.

Enrutamiento por trayectoria más corta: Se basa en la utilización de grafos, donde cada nodo representa un enrutador los cuales son unidos mediante arcos. Para establecer una salida, primero se establecen los caminos más cortos entre pares de nodos. Para medir esta ruta más corta es necesario especificar la métrica a utilizar, la cual podrá variar dependiendo del criterio de diseño. Algunas de ellas son: distancia geográfica, retardos medios, costo de utilización de enlace, ancho de banda, etc. Uno de los algoritmos utilizados para calcular las rutas más cortas es Dijkstra (1959).

Inundación: La idea detrás este algoritmo es sencilla. El paquete que llega se replica por todas las salidas del enrutador, excepto por la cual arribó. Claramente el problema que surge con este algoritmo es la duplicación del paquete. Además se inunda la red de forma innecesaria contribuyendo a la congestión de nodos con problemas. Una forma de evitar esto es siendo más selectivo, enviando los paquetes por los caminos en donde es más probables que se halle el destino.

Algunos de los algoritmos dinámicos se describen a continuación:

Vector Distancia: es un algoritmo dinámico, en el cual los enrutadores comparten información cada cierto intervalo de tiempo con sus vecinos. Esta información se almacena en forma de tablas (vectores), las cuales contienen la mejor distancia conocida hacia cada destino y por cual de sus interfaces debe enviarse.

Una implementación de este algoritmo es conocida bajo el nombre de RIP (*Routing Information Protocol*), la cual utiliza como métrica la enumeración de saltos. Cada router envía su información a intervalos regulares de tiempo o cuando hay un cambio en la topología, y sólo cambia su tabla si llega información con una mejor ruta (menor número de saltos), manteniendo siempre sólo las más cortas.

Ruteo de Estado de Enlace: este algoritmo nace por la necesidad de reemplazar a RIP (mediados de 1980), el cual era cada vez más incapaz de trabajar con redes heterogéneas, debido a que no incluía consideraciones de ancho de banda y su tiempo de convergencia era pobre. La idea detrás de este algoritmo es lograr obtener información sobre vecinos que se encuentren bajo la misma jerarquía, donde la mayor entidad dentro de esta es conocida como sistema autónomo (se entiende por sistema autónomo una colección de redes bajo un administrador común y que utiliza una estrategia de ruteo común).

El algoritmo se basa en identificar a todos los enrutadores adyacentes mediante algún tipo de mensaje especial, el cual también ayuda con posterioridad a identificar si el terminal aún se encuentra funcionando. Luego de la identificación, se mide el costo de llegar a cada terminal adyacente con algún criterio (el cual puede ser el retraso). Con esta información se arma un paquete especial el cual se envía (por inundación) hacia los otros enrutadores que conforman el sistema autónomo. Con esta información es posible conformar la topología completa del sistema en relación a los enrutadores que la componen. Debido a que son componentes de un mismo sistema todos los routers poseen la misma información. Una vez conformada la topología cada router calcula, mediante algún algoritmo como Dijkstra, las menores distancias hacia los distintos nodos que componen la red con las cuales configura su tabla de ruteo.

Como se puede deducir de los anterior, a diferencia de RIP que sólo tiene una visión de saltos hacia vecinos, este protocolo conoce la topología completa de la red, y en esta se basa para conformar su tabla de ruteo.

Un protocolo ampliamente utilizado y que hace uso de este algoritmo es OSPF (*Open Shortest Path First*).

6.3 Protocolo de Internet (IP)

Como ya se mencionó en el capítulo 2, el Protocolo de Internet (IP) forma parte del paquete o stack de protocolos especificados para la arquitectura TCP/IP. De echo es uno de los protocolos más utilizado en la actualidad para la transmisión de paquetes desde una host a otro ubicados en redes distantes.

IP es el protocolo que especifica la capa de red para el modelo TCP/IP. Este protocolo está especificado para enviar un paquete de información desde un origen a un

destino a través de un sistema de redes interconectadas. Utiliza el criterio de mejor esfuerzo, es decir, no garantiza la llegada del paquete a destino pues no posee mecanismos de control de flujo, ni mecanismos para implementar confiabilidad, ni secuenciamiento de paquetes.

Este protocolo se encarga de implementar dos funciones básicas: direccionamiento y fragmentación.

Las direcciones son utilizadas para identificar a cada uno de los host conectados a la red* y además ayuda al encaminamiento de los paquetes ya que los protocolos de ruteo se basan en las direcciones.

La fragmentación ocurre cuando un paquete de determinado largo debe pasar a través de una red que posee un MTU (Maximun Transmission Unit) menor al largo del paquete. Esta fragmentación es realizada por el enrutador de borde y el paquete no es vuelto a reensamblarse hasta que llega a su destino.

Para un mejor comprensión de estas funciones, la Figura 6.1 muestra el formato del paquete.

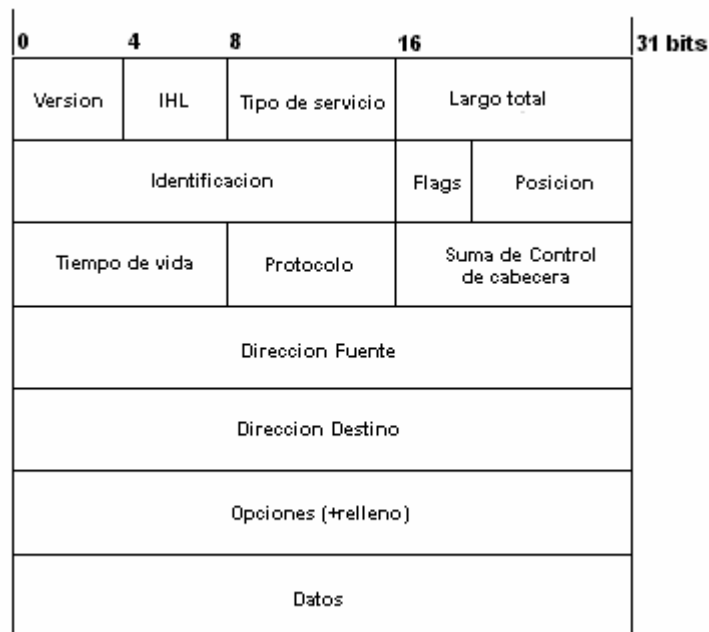


Figura 6.1: Formato paquete IP.

La descripción de cada campo se enuncia a continuación:

- Versión:** Describe el formato de la cabecera Internet (versión 4 para este caso).
- IHL:** Longitud de cabecera de Internet (*Internet Header Length*) en palabras de 32 bits (para un paquete normal es 5).
- Tipo de servicio:** Identificación de los parámetros abstractos de la calidad de servicio. Permite definir prioridad, rendimiento y fiabilidad. No todos los enrutadores toman en cuenta este campo.

* En realidad las direcciones especifican puntos de conexión a red, más que hosts en si mismos.

<i>Largo Total:</i>	Longitud total de paquete medido en número de bytes incluyendo cabecera y datos (largo máximo 65535).
<i>Identificación:</i>	Número que identifica el paquete y que es utilizado como ayuda para el armado en el caso de fragmentación.
<i>Flags:</i>	Indicadores de control de 3 bits, cada uno de ellos utilizado para: 1 ^{er} bit debe ser cero; 2 ^{do} bit si es cero el paquete puede fragmentarse y si es 1 no debe fragmentarse; 3 ^{er} bit indicador de más fragmentos (en el caso que existiese fragmentación) si es cero indica que es el último fragmento, en caso contrario existen más fragmentos.
<i>Posición:</i>	Campo utilizado junto con los flags para controlar la fragmentación. Indica que ocupa el fragmento dentro del paquete original.
<i>Tiempo de vida:</i>	Campo que indica el tiempo máximo que se le está permitido al paquete permanecer en el sistema. Cada vez que pasa por un enrutador este lo decrementa y si llega a cero antes de arribar a su destino es descartado.
<i>Protocolo:</i>	Indica el protocolo de capa de transporte que esta contenido en la parte de datos del paquete.
<i>Suma de control:</i>	Es el complemento a uno de 16 bits de la suma de los complementos a uno de todas las palabras de 16 bits de la cabecera. Como la cabecera cambia al pasar por cada enrutador debe ser recalculada cada vez.
<i>Dirección fuente:</i>	Dirección de origen (32 bits)
<i>Dirección destino:</i>	Dirección de destino (32 bits)
<i>Opciones:</i>	Opciones para compatibilidad futura y de seguridad. Pueden no ser incluidas en el paquete (que es lo general). El relleno se utiliza para asegurar que el paquete sea un múltiplo de 32 bits.
<i>Datos:</i>	Campo en donde se ubican los datos para la capa de red.

A cada host que utiliza el protocolo IP se le es asignada una dirección de 32 bits que lo identifica, la cual esta compuesta de dos partes: una parte que identifica la red y otra parte que identifica el host dentro de esa red. El número de red identifica una red y debe ser asignado por el Centro de Información de Red Internet (InterNIC).

La dirección es dividida en 4 grupos de 8 bits cada uno, los cuales son separados por puntos y representados en formato decimal. La Figura 6.2 ilustra de mejor forma este formato.

Para identificar que parte de la dirección pertenece a la red y que parte pertenece al host se dividen las direcciones en clases, cada una con un número determinado de posibles direcciones de red y de hosts asociados. Esta configuración se muestra en la Tabla 6.1.

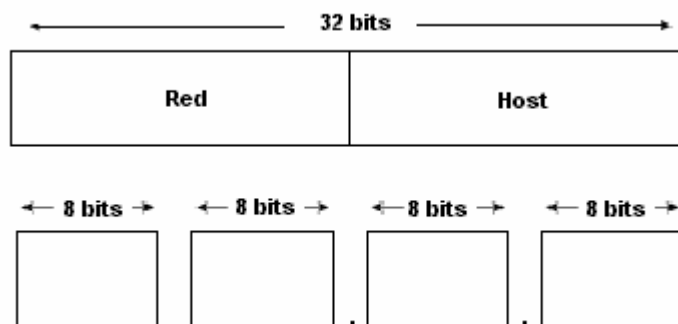


Figura 6.2: Formato de dirección IP.

Tabla 6.1: Clases de direcciones

CLASE	FORMATO	RANGO DE DIRECCIONES
A	8 bits de red, 24 de hosts	1.0.0.0 a 126.0.0.0
B	16 bits de red, 16 de hosts	128.1.0.0 a 191.254.0.0
C	24 bits de red, 8 de hosts	192.0.1.0 a 223.255.254.0
D	Grupos Multicast	224.0.0.0 a 239.255.255.255
E	Experimental	240.0.0.0 a 254.255.255.255

Debido al agotamiento de direcciones de red y a la utilización de subredes en Internet se dio paso a la utilización de máscaras de subred (para mayores detalles ver “Internetworking with TCP/IP” Douglas E. Comer).

6.4 Calidad de Servicio (QoS, Quality of Service)

El concepto de QoS se refiere a “la capacidad de la red de proveer mejor servicio para seleccionar tráfico sobre redes de variadas tecnologías[±]” (Frame Relay, ATM, redes Ethernet, etc), utilizando ruteo IP para todas estas tecnologías subyacentes. Algunos de los factores que influyen en la calidad de servicio son: tiempos de latencia y jitter, ancho de banda dedicado y mejoras en las características de pérdida de la red.

Implementar calidad de servicio no es una tarea fácil, más aún cuando es necesario pasar por varios tipos de redes para llegar al destino deseado. Muchas de las redes no consideran la calidad de servicio como un factor preponderante y sólo realizan el mejor esfuerzo al entregar los paquetes que circulan por ella. Otras, en cambio, si la consideran y definen una serie de factores posibles de configurar al establecer la comunicación (como ATM).

Una de las soluciones posibles es establecer un proceso de negociación entre el emisor y, o el host destino o un elemento de red. Los parámetros importantes a considerar en esta negociación son:

Prioridad : indica que conexiones son más importantes de mantener.

[±] Cisco Product Documentation, www.cisco.com

Rendimiento: cantidad de bytes transferidos por segundo, mediante algún intervalo de tiempo.

Tasa de error de error residual: Cantidad mínima de errores (mensajes perdidos o alterados) como porcentaje del total de datos enviados. (idealmente cero).

Otros factores influyentes son:

Probabilidad de falla al establecer conexión: probabilidad que no se establezca la conexión en un lapso determinado de tiempo.

Retardo de tránsito: tiempo entre mensajes enviados por el usuario hacia el destino y viceversa.

Protección: seguridad otra terceros.

Así se definen tres niveles básicos de calidad de servicio extremo a extremo:

- ❑ Servicio de mejor esfuerzo: conocido como ausente de QoS, define la conectividad sin garantías.
- ❑ Servicio Diferenciado: conocido como soft QoS, algunos tráficos son tratados de mejor forma que otros. (manejo más rápido, mayor ancho de banda promedio, menor tasa de perdidas) Es una preferencia estadística, no garantizada de forma estricta.
- ❑ Servicio Garantizado: reservación de recursos de red para tráfico específico.

Para el caso del protocolo IP, no existe explícitamente una política de calidad de servicio, pues este trabaja con el concepto de mejor esfuerzo. A pesar de ello existe un campo dentro del encabezado IP que permite sugerir ciertos parámetros, este campo es el de Tipo de Servicio el cual se describe a continuación:

0	1	2	3	4	5	6	7
PRECEDENCIA			D	T	R	0	0

Precedencia: Da indicaciones de cómo debe ser tratado el paquete en cuanto a prioridades se trata. Los valores posibles son:

111 – Control de Redes
 110 – Control Entre Redes
 101 – Critico/ECP
 100 – Muy Urgente (Flash Override)
 011 – Urgente (Flash)
 010 – Inmediato
 001 – Prioridad
 000 – Rutina

D: Demora 0 Normal, 1 Baja demora.

T: Rendimiento 0 Normal, 1 Alto rendimiento

R: Fiabilidad 0 Normal, 1 Alta fiabilidad

Como se mencionó anteriormente no todas las redes reconocen estos parámetros, por lo que suelen conservarse con valores normales.

Muchas de las especificaciones de parámetros de calidad de servicio se dejan como variables de decisión para protocolos de capa superior, o son implementados sobre los elementos que componen la subred (como son los enrutadores), permitiendo crear políticas de administración de recursos en redes que están bajo la misma administración.

7 CAPA DE TRANSPORTE

La capa de transporte es la encargada de dar confiabilidad al transporte de datos sobre redes, sin tener consideraciones sobre las capas subyacentes. Se puede decir que dentro de la arquitectura de sistema ésta es la primera capa que verdaderamente realiza una abstracción de niveles inferiores, como si se tratase de un protocolo extremo a extremo. Sobre este nivel de abstracción, esta implementa servicios para las capas superiores, tales como control de flujo, multiplexación, manejo de circuitos virtuales y chequeo y recuperación de errores.

7.1 Concepto de API (Application Program Interface)

Como su nombre lo describe, una API se refiere a la estructura o conjunto de funciones que provee un sistema en particular, tal que permitan dar ciertos servicios a los procesos (usuarios) que se encuentren trabajando en él. La API define como un servicio en particular puede ser invocado mediante algún paquete de software dispuesto para este fin.

Las API's se componen de primitivas de servicio, que son invocaciones al sistema operativo, tal que este se encarga de realizar ciertas funciones especificadas por estas primitivas.

El ejemplo más conocido de API's para protocolos de comunicación es la interfaz *socket* (sistema operativo BSD de UNIX), los cuales son una generalización del mecanismo de acceso a archivos de UNIX que proporciona un punto de acceso para comunicación.

Los socket se basan en un modelo de cliente – servidor. El concepto de servidor se aplica a cualquier proceso dentro de una máquina que ofrece un servicio que se puede obtener mediante una red. El servidor acepta peticiones que llegan por la red, realiza el servicio y devuelve el resultado al solicitante. Luego, se entiende por cliente el proceso (corriendo dentro de la misma máquina o dentro de otra máquina en la red) que envía una petición de servicio y espera respuesta.

La implementación de las características del servicio (control de flujo, confiabilidad, etc) son transparentes para el usuario que las solicita, dejando al sistema encargarse de esos aspectos. La aplicación sólo debe especificar el protocolo de comunicación a utilizar (por ejemplo TCP).

Luego la primitivas definidas y el flujo de información entre un cliente y un servidor, para el caso de una conexión TCP será como se muestra en la Figura 7.1

El significado de cada una de ellas es el siguiente:

- *Socket*: crea un punto de comunicación.
- *Bind*: asocia una dirección local (de la máquina) al socket.
- *Listen*: indica que se comenzará a escuchar conexiones, indicando número máximo de ellas.
- *Accept*: acepta las conexiones provenientes desde los clientes. Se bloquea hasta que llega una petición. (servidor)
- *Connect*: intenta establecer una comunicación. (cliente)

- *Read*: lee, desde el socket conectado, la información que llega a través de él.
- *Write*: escribe por el socket conectado la información que se desea enviar.
- *Close*: cierra el socket, liberando la conexión.

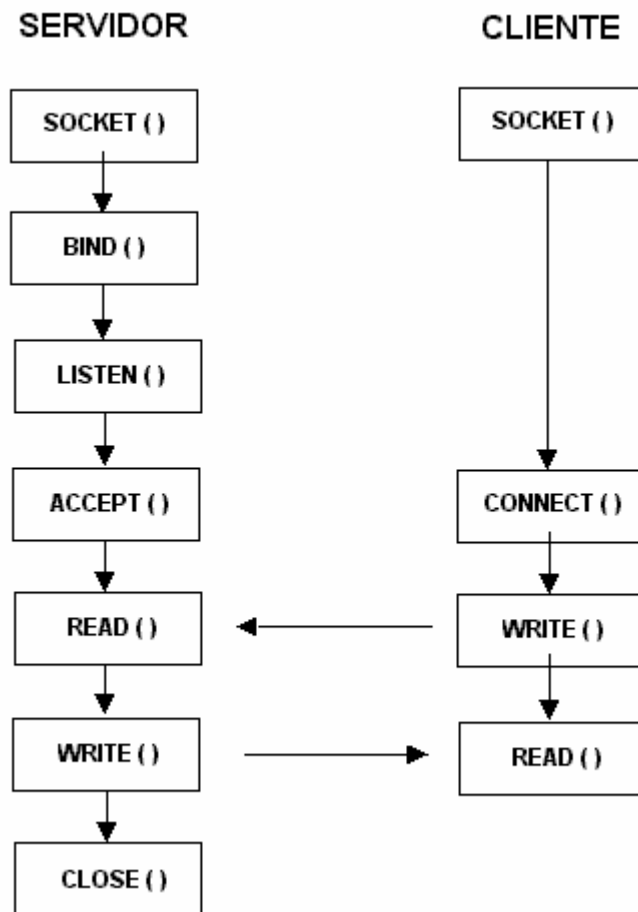


Figura 7.1: Modelo de conexión con socket

Los tipos de conexiones posibles para un protocolo de internet son básicamente dos:

- ❑ No orientado a la conexión y no confiable.
- ❑ Orientado a la conexión y confiable.

Para ilustrar estos dos tipos, se describen a continuación los protocolos dentro de la arquitectura TCP/IP que los implementan.

7.2 UDP (User Datagram Protocol)

Como se vio en el capítulo anterior, para identificar una máquina en una red, IP proporciona direcciones en el interior del encabezado, pero estas direcciones no mencionan nada sobre la aplicación a la que van dirigidos los mensajes. Luego es posible tener varios procesos ejecutándose en una máquina (la cual posee una única dirección IP), entonces ¿cómo saber a cuál de los procesos va dirigido el mensaje contenido en el paquete?. El protocolo más sencillo de capa de transporte que da solución a esta interrogante se conoce como *Protocolo de Datagramas de Usuario* o UDP.

Este protocolo aporta un procedimiento para que los procesos dentro de aplicación puedan enviar mensajes a otros procesos con un mínimo de mecanismo de protocolo. Cada aplicación es identificada por un *puerto de protocolo*, el cual es un número entero contenido dentro del encabezado UDP. El sistema operativo local proporciona un mecanismo de interfaz que los procesos utilizan para especificar o acceder un puerto. A este proceso se le conoce como multiplexado, el cual se grafica en la Figura 7.2.

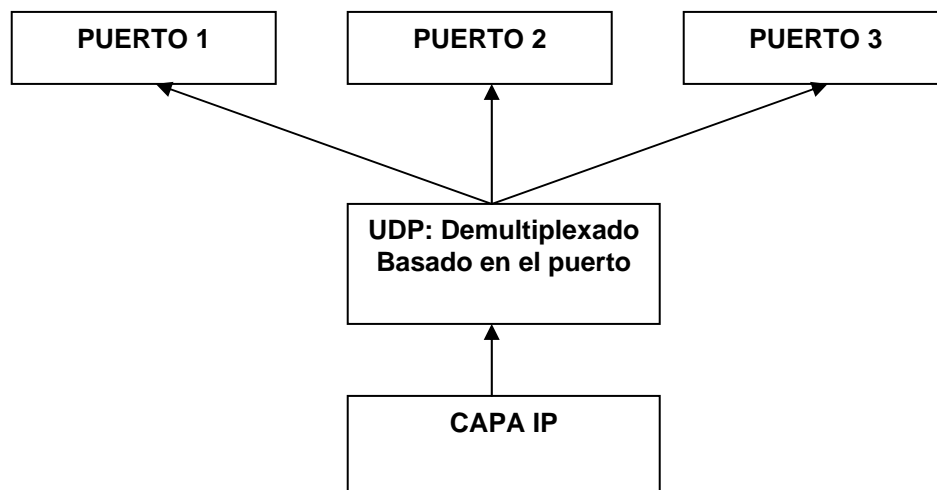


Figura 7.2: Demultiplexado en base a número de puertos

El encabezado UDP, con sus respectivos campos, se muestra a continuación:

0	16	31 bits
Puerto UDP de Origen		Puerto UDP de Destino
Longitud del Mensaje UDP		Suma de Verificacion UDP
Datos		

Figura 7.3: Encabezado UDP

Los campos *Puerto de Origen* y *Puerto de destino* contienen los números que servirán para identificar las distintas aplicaciones, permitiendo el multiplexado de datagramas. El puerto de origen es opcional (puede contener ceros).

El campo de *Longitud* contiene el tamaño del encabezado y datos contenidos en el datagrama UDP en número de bytes.

La *Suma de verificación* es opcional, un valor de cero significa que la suma no se llevó a cabo. Esto se debe a que este protocolo fue pensado para proporcionar un servicio en capa de transporte, no orientado a la conexión y no confiable, más enfocado a aplicaciones en donde el retardo es prioridad. Luego al no realizarse la suma de verificación se ahorra tiempo de procesamiento en el host de origen y destino. Por otro lado, y dado que IP no realiza ninguna operación de verificación sobre el campo de datos, esta instancia en UDP es la única que puede implementar algo de confiabilidad sobre los datos. Esta suma utiliza el mismo algoritmo descrito para IP.

En el campo de datos viajará un paquete IP, como el descrito en el capítulo anterior.

7.3 TCP (Transmission Control Protocol)

El segundo protocolo de capa de transporte para el modelo TCP/IP es el conocido como *Protocolo de Control de Transmisión*, el cual a diferencia del anteriormente estudiado, ofrece un servicio orientado a la conexión y confiable.

Para implementar estos servicios se vale de una serie de métodos y consideraciones, que lo hacen más complejo, y que serán brevemente tratados en esta sección.

Por flujo confiable se entiende la entrega garantizada de los datos enviados de una máquina a otra sin pérdida o duplicación. Luego la transmisión es hecha confiable gracias al uso de número de secuencias y acuses de recibo. La implementación de acuse de recibo no es la que usualmente se pensaría, es decir la que envía un segmento de datos y se espera su confirmación para el envío del siguiente. Este método es ineficiente en el sentido que se desperdicia ancho de banda esperando que llegue una confirmación, y la red estará ociosa en los tiempos efectivos procesamiento de segmento.

El método utilizado se denomina de *ventanas deslizantes* (*Sliding Windows*), el cual es una forma de acuse de recibo positivo y retransmisión, en donde se utiliza el ancho de banda de red de mejor forma debido a que se permite el envío de varios segmentos de datos sin esperar confirmación. Como se muestra en la figura 7.4, la idea consiste en ordenar una secuencia de paquetes que serán enviados por la red, colocando una ventana pequeña de tamaño fijo sobre la secuencia. Luego, todos los segmentos en el interior de la ventana, serán transmitidos en forma consecutiva.

Cuando el transmisor recibe un acuse de recibo del primer paquete mueve la ventana en un segmento y envía el siguiente paquete. Entonces la ventana se moverá a medida que vayan llegando las confirmaciones. El método recuerda y mantiene temporizadores para casos en que no lleguen acuses de recibo, lo cual provocará el reenvío del mismo. (Notar que si el tamaño de ventana es 1, el sistema se transforma en el caso base de confirmación positiva). Siempre es posible elegir un tamaño óptimo tal que los tiempos ociosos se reduzcan al mínimo.

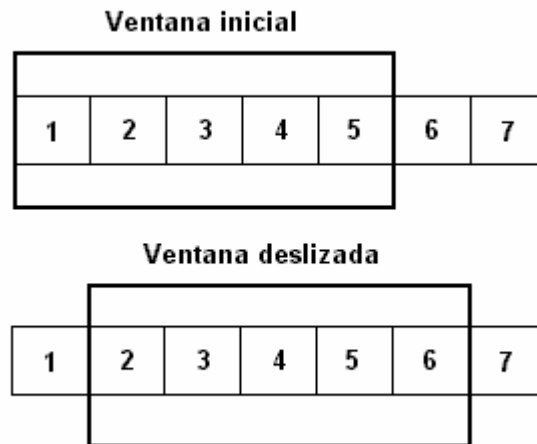


Figura 7.4: Ventana deslizante

Luego TCP implementa este método, pero le agrega una última dimensión, la cual es permitir que el tamaño de la ventana sea variable. Esto es utilizado para realizar control de flujo, es decir, evitar que al enviar muchos paquetes se sature el receptor. Cada acuse de recibo incorpora un aviso de ventana, que especifica los bytes adicionales que esta dispuesto a recibir el receptor. Así el tamaño de la ventana variará en función de la disposición de memoria que tenga el receptor, siendo el caso extremo el tamaño de la ventana igual a cero, lo cual detendría el envío.

Otro aspecto importante de tratar referente a los servicios brindados por TCP, es el establecimiento y fin de la conexión.

El establecimiento se realiza con un saludo de tres etapas, consistentes básicamente en intercambio y acuses de recibo de segmentos TCP con algún campo en particular indicando la acción. Una vez que se establece la conexión, está es Full Duplex. De esta forma se garantiza que ambos lados estén listos para transferir datos (y que sepan que ambos están listos), y permite además acordar un número de secuencia inicial para los segmentos. Este saludo se muestra en la figura 7.5.

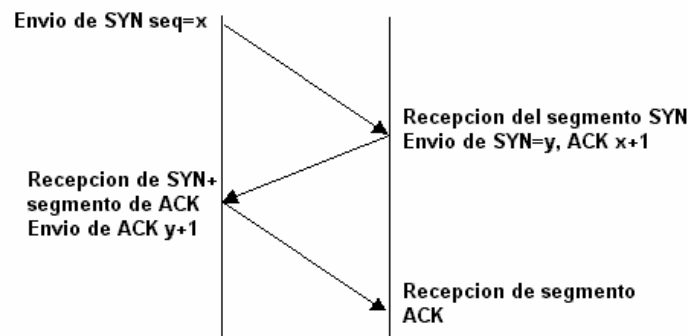


Figura 7.5: Handshake de tres pasos

Para cerrar una sesión se utiliza un procedimiento similar al saludo de tres etapas, pero dado que la comunicación es full duplex, al cerrar un extremo, este envía una trama de fin al otro extremo, lo cual provoca que se avise a la aplicación que concluye la sesión. Mientras esto se hace, ya se envió un segmento de acuse de recibo finalizando las etapas.

TCP utiliza el mismo sistema de multiplexado de UDP, es decir implementa números de puerto para identificar los distintos procesos en una máquina. El encabezado TCP se muestra en la figura 7.6.

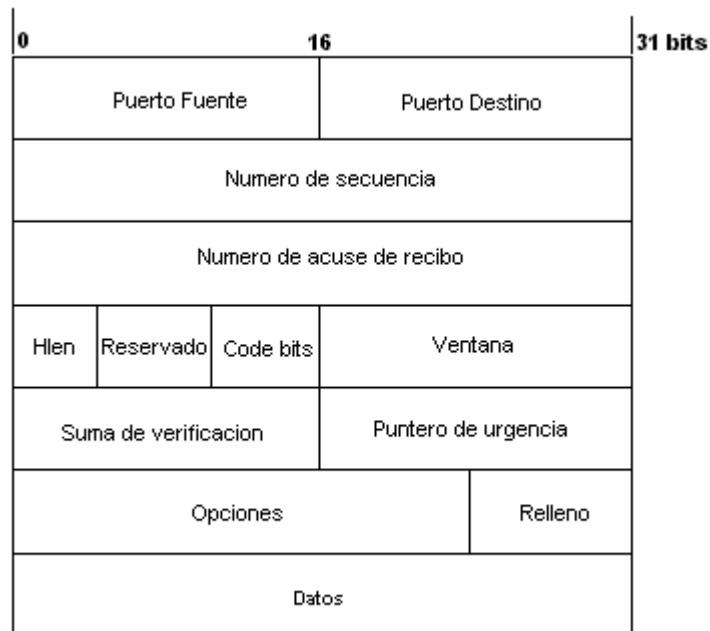


Figura 7.6: Encabezado TCP

8 CAPA DE APLICACIÓN

En este capítulo se estudiarán, en forma resumida, algunos servicios de alto nivel para redes, y algunos protocolos del modelo TCP/IP que las hacen posible. Estos servicios determinan como es que el usuario percibe una red, y la utilidad que esta presta.

Existen variadas aplicaciones, las cuales pueden orientarse a necesidades particulares o como una combinación de servicios, tendencia que va en progresivo aumento, gracias a las distintas ofertas multimediales y al aumento de las prestaciones de redes de área extendida.

8.1 Acceso Remoto

El acceso remoto consiste en autorizar al usuario establecer una sesión hacia una máquina dentro de la red y permitirle entonces ejecutar comandos de su sistema. Para este fin existen softwares servidores de acceso remoto para la mayor parte de los sistemas operativos y programas de aplicación que funcionan como clientes.

El protocolo utilizado para este fin en el modelo TCP/IP es conocido como TELNET, el cual permite establecer una conexión TCP hacia un servidor de acceso remoto. TELNET presta un servicio transparente, ya que simula que el teclado y el monitor del usuario están conectados de manera directa a la máquina remota.

TELNET ofrece tres servicios básicos:

- ❑ Define un terminal virtual de red: proporciona interfaz estándar para los sistemas remotos.
- ❑ Negociación de opciones: permite al cliente y al servidor escoger opciones, al mismo tiempo que ofrece opciones estándares (Ej: ASCII de 7 o 8 bits).
- ❑ Simetría en la conexión: no especifica que la entrada del cliente sea el teclado ni la salida la pantalla, permitiendo que cualquier programa se convierta en cliente.

La tarea de un servidor TELNET es aceptar una conexión TCP del cliente y transmitir los datos entre la conexión y el sistema operativo local. En realidad la labor es un poco más compleja pues debe manejar diversas conexiones concurrentemente. Esto lo logra utilizando un esquema de padre e hijo, donde el padre espera por conexión y una vez establecida la hereda al hijo para que este las atienda, mientras el espera por más.

Una desventaja de este protocolo es su ineficiencia, pues cada pulso de teclado viaja, a través del sistema operativo, hacia el programa cliente, el cual las envía a través de la red hasta la máquina servidor. Una vez allí debe pasar por el sistema operativo del servidor al programa de aplicación del servidor y de este programa, de vuelta al sistema operativo para ser enviado de vuelta por el mismo camino hacia el cliente (eco).

Un problema que surge al trabajar con distintos sistemas, es el adaptar los detalles de cada computador con tal que todos los caracteres y comandos representen la misma acción (Ej: Ctl-C y ESC). Como solución TELNET define el como deben enviarse las secuencias de datos y comandos a través de Internet. Es conocido como *network virtual*

terminal (terminal virtual de red o NTV). La figura 8.1 muestra un ejemplo de esta implementación, donde el cliente traduce las pulsaciones de teclado y los comandos que vienen del terminal de usuario a formato NVT y las envía al servidor. El servidor, a su vez, traduce desde el formato NVT al formato del sistema operativo local.

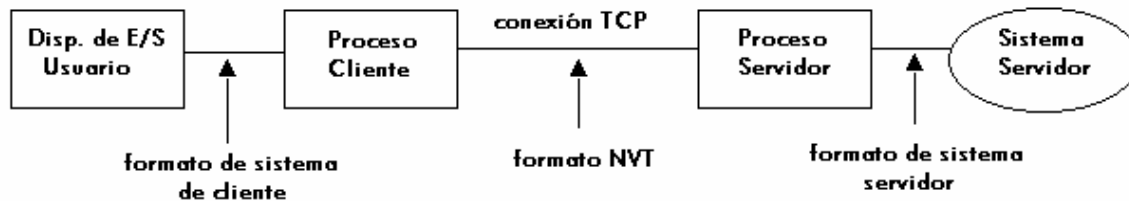


Figura 8.1: TELNET y formato NVT

8.2 Transferencia de Archivos

Para realizar una transferencia de archivos entre máquinas remotas se debe invocar un programa cliente que realice la tarea, especificando la máquina en donde se almacenan los archivos a transferir, y previa autorización esta, se procede al traspaso. Así, el archivo pasará a ser parte del sistema local del usuario, en donde podrá ser modificado sin afectar el archivo original.

El proceso descrito arriba puede ser complejo debido a la no heterogeneidad de las computadoras. Para la transferencia exitosa entonces se debe tener en cuenta una serie de factores como: autorización, nociones de propiedad de archivos y permisos involucrados, formatos de datos, etc.

En esta sección se describirá la forma en que el modelo TCP soluciona estos problemas implementando FTP (*File Transfer Protocol*).

Los servicios ofrecidos por FTP son:

- ❑ Acceso interactivo: permite al usuario interactuar con los servidores remotos (Ej: pedir lista de archivos en directorio).
- ❑ Especificaciones de formato: pueden ser de tipo texto binario, caracteres ASCII o EBCDIC.
- ❑ Control de autenticación: requiere que los clientes sean autorizados, mediante nombre de conexión y clave de acceso al servidor.

Los servidores FTP permiten el acceso concurrente de varios clientes, en forma similar a TELNET, creando un proceso hijo el cual se encarga de atender a los usuarios. Pero a diferencia de TELNET, el proceso hijo no ejecuta todas las acciones necesarias para la transferencia. En realidad lo que ocurre es que al establecerse una sesión FTP se crea un proceso hijo encargado del control de conexión y cada vez que se necesite transferir un archivo este crea otro proceso hijo, el cual se encarga de la transferencia. La conexión de control transporta comandos que indican al servidor que archivo transferir. La Figura 8.2 ilustra el concepto.

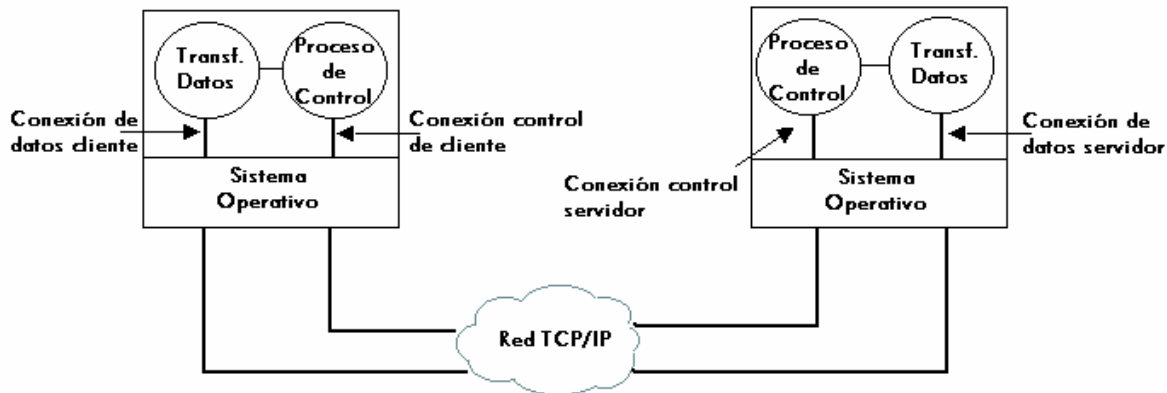


Figura 8.2: Conexión FTP

Mientras dure la sesión se mantendrá la conexión, sin embargo se establece una nueva conexión de transferencia de datos para cada transferencia de archivos.

El proceso de control además de enviar comandos del usuario al servidor, permite coordinar el uso de puertos de protocolo TCP asignándolos dinámicamente y creando procesos de transferencia que utilicen tales puertos.

Se permite que FTP utilice el protocolo de terminal virtual de red de trabajo TELNET para enviar los datos con este formato NVT. FTP no permite negociación de opciones, empleando una definición básica de NVT (a diferencia de TELNET).

8.3 Correo Electrónico

Más conocido como e-mail, es uno de los servicios más utilizados gracias a su sencillez para transferir información a través de la red.

Generalmente no es un método inmediato, debido a la filosofía con que se manejan los mensajes, utilizando una técnica conocida como spooling (o colas de correo). Esta consiste en la ocupar áreas de almacenamiento en donde se colocan los mensajes salientes, junto con la identificación de emisor, destino y hora de depósito. La emisión del mensaje entonces puede ser llevada a cabo, vía TCP como una función secundaria.

Entonces el proceso es el siguiente: el proceso secundario actúa como cliente e intenta establecer una conexión TCP hacia el servidor de correo, si lo logra envía una copia del mensaje saliente al servidor, el cual lo almacena en su área spool. Una vez que ambos están de acuerdo en que la copia se efectuó con éxito, el cliente borra el mensaje de su cola. Ahora, si no se logra la conexión, el proceso cliente almacena la hora del último intento y cierra el proceso. Así es posible mantener los mensajes dentro del área de almacenamiento hasta que, pasado un tiempo, se produzca un nuevo intento. Cada vez que se encuentra un mensaje pendiente el proceso secundario tratará de enviarlo a su destino. Si luego de un tiempo prolongado no es posible encontrar el destino, el software de correo devuelve el mensaje al emisor.

Para facilitar la interacción con otras redes que no necesariamente manejan el mismo protocolo de red, se utilizan mail gateways, con lo cual el emisor no establece contacto directo con la máquina destino, sino que más bien le entrega el mensaje a máquinas intermedias que se encargan del trabajo. Su ventaja ya fue mencionada y es que permite interoperabilidad, pero posee desventajas que provocan que este servicio no sea del todo confiable. Una vez que el emisor transfiere el mensaje a la primera máquina intermedia, este borra la copia local confiando que este será capaz de entregarla. Así, ni el emisor ni el receptor poseen una copia del mensaje en tránsito y si se encuentran fallas intermedias puede producirse la pérdida de la información sin que el receptor ni emisor sean advertidos. Otra desventaja es que introducen retardo, no existiendo control por parte del emisor.

Para hacer más heterogéneo el servicio de correo, TCP/IP lo divide en dos estándares, uno encargado de los formatos del mensaje de correo y el otro de los detalles de intercambio entre dos máquinas.

El estándar para los formatos de mensaje especifica el formato exacto de los encabezados de correo así como el significado. Los encabezados están compuestos por líneas que comienzan con palabras claves seguidas de dos puntos y algún valor. Por ejemplo, una línea puede comenzar con la palabra clave *To:* y el resto debe contener la dirección de correo electrónico del destino. Mantener el formato del mensaje permite que este sea utilizado por un vasto rango de sistemas

El estándar para intercambio de correo entre computadores definido para TCP/IP es conocido con el nombre de SMTP (*Simple Mail Transfer Protocol*) el cual especifica el formato de los mensajes entre cliente y servidor.

Para comenzar, el cliente debe establecer una conexión TCP y espera que el servidor envíe un mensaje 220 READY FOR MAIL. Una vez recibido este mensaje el cliente envía un comando HELO. Una vez que la comunicación se ha establecido, el emisor puede transmitir uno o más mensajes de correo, terminar la conexión (QUIT) o intercambiar funciones de emisor y receptor vía el comando TURN. Para mayor claridad la Figura 8.3 ilustra un ejemplo extraído desde “*Redes globales de información con Internet y TCP/IP*” de Douglas E. Comer, en donde se supone al cliente Smith en Alpha.EDU envía un mensaje hacia el usuario Jones, Green y Brown en Beta.GOV.

```
S: 220 Beta.GOV Simple Mail Transfer Service Ready
C: HELO Alpha.EDU
S: 250 Beta.GOV

C: MAIL FROM:<Smith@Alpha.EDU>
S: 250 OK

C: RCPT TO:<Jones@Beta.GOV>
S: 250 OK

C: RCPT TO:<Green@Beta.GOV>
S: 550 No such user here

C: RCPT TO:<Brown@Beta.GOV>
S: 250 OK

C: DATA
```

```
S: 354 Start mail input; en with <CR><LF>.<CR><LF>
C: ..... cuerpo del mensaje .....
.
.
.
C: <CR><LF>.<CR><LF>
S: 250 OK

C: QUIT
S: 221 Beta.GOV Service closing transmission channel
```

Figura 8.3: Ejemplo de una transferencia SMTP. Las líneas que comienzan con C: son transmitidas por el cliente y las que comienzan con S: por el servidor.

Como se muestra en el cuadro, la máquina Beta.GOV no reconoce al Green, a lo cual responde con un comando 550. También se debe notar que el comando 250 significa todo bien. RCPT identifica a los destinatarios del mensaje.

Por último, es importante mencionar la composición de las direcciones de correo, las cuales, para la Internet tienen la siguiente forma:

local-part@domain-name

Donde *domain name* es el nombre del domino del destino de correo al que el correo debe ser enviado y *local part* es la dirección de buzón de la máquina.

8.4 HTTP (Hypertext Transport Protocol)

El World Wide Web es una arquitectura armada para el acceso de documentos enlazados a través de miles de máquinas sobre toda la Internet. La forma de acceso es a través de los conocidos browsers (buscadores, navegadores) (Ej: Netscape, Internet Explorer, Opera, etc), los cuales utilizan interfaces gráficas muy atractivas, a las cuales se debe la popularidad de la “Web”.

Estos documentos enlazados, conocidos como páginas, pueden contener variada información, como texto, imágenes, enlaces a otros sitios (hyperlinks), etc. Se basan en una conexión confiable vía TCP mediante los mencionados browser, que son programas que se encargan de buscar la información en una dirección determinada, traducirla al formato especificado y desplegarla en pantalla.

Cada servidor Web escucha por conexiones TCP en el puerto 80 y cuando llega una petición de conexión desde un cliente en un browser este responde y se inicia la conexión. El protocolo encargado de definir estas peticiones y réplicas conocido como HTTP.

Para ilustrar el flujo de comunicación entre un cliente y un servidor en la web, se mostrarán los pasos desde un ejemplo extraído desde “*Computer Networks*” de Andrew Tanenbaum.

Supongamos que un usuario desea ver la información de la siguiente URL (Uniform Resource Locator) <http://www.w3.org/hypertext/WWW/TheProject.html>, la cual contiene tres partes: el nombre del protocolo (*http*), el nombre de la máquina donde la página se encuentra localizada (*www.w3.org*) y el nombre del archivo que contiene la página

(*hypertext/WWW/ /TheProject.html*). Los pasos que ocurren entre el clic del usuario y la página siendo desplegada son los siguientes:

1. El browser determina el URL.
2. El browser pregunta al servidor DNS (Domain Name Server) por la dirección IP de *www.w3.org*.
3. DNS responde con 18.23.0.23.
4. El browser realiza una conexión TCP al puerto 80 sobre 18.23.0.23.
5. Entonces envía un comando GET */hypertext/WWW/TheProject.html*.
6. El servidor 18.23.0.23 envía el archivo *TheProject.html*.
7. La conexión TCP es liberada.
8. El browser despliega todo el texto en *TheProject.htm.l*
9. El browser busca y despliega todas las imágenes en *TheProject.html*.

El protocolo HTTP define una serie de métodos que son usados para requerimientos desde el browser al servidor y viceversa. Algunos de estos métodos se muestran en la Tabla 8.1.

Método	Descripción
GET	Pide una página Web para leer
HEAD	Pide un encabezado de página para leer
PUT	Pide almacenar una página Web
POST	Agrega a una página Web
DELETE	Remueve una página Web
LINK	Conecta dos fuentes existentes
UNLINK	Rompe una conexión existente entre dos fuentes

9 TEORÍA DE TRÁFICO

La teoría de tráfico es una herramienta ampliamente utilizada para el análisis del comportamiento de las redes de comunicaciones, las cuales pueden ser de conmutación de circuitos, como las redes telefónicas o de conmutación de paquetes como las redes de datos IP.

En este capítulo el enfoque irá orientado principalmente a las primeras, ya que son la base en donde se sustenta el sistema telefónico fijo.

La conmutación de circuitos consiste el establecimiento de un canal dedicado físico (real), de extremo a extremo, entre los cuales existen elementos de conmutación, que en el caso de la red telefónica se trata de centrales o PBX (Private Branch eXchange), para el caso de empresas.

Los enlaces pueden consistir en ranuras de tiempo en un sistema de multiplexación temporal (TDM) o bandas de frecuencia para el caso de multiplexación en frecuencia (FDM).

Como se verá con mayor detalle, más adelante en el capítulo, las redes telefónicas pueden operar en base a bloqueo de llamadas o a la utilización de colas de espera, siendo la más utilizada la primera en las redes públicas debido a su equidad y eficacia.

El modelamiento para este tipo de comportamiento se realiza con teorías de cola, utilizando distintas notaciones dependiendo de los supuestos y modelos a aplicar para cada proceso. La notación de Kendall para un sistema general de formación de colas es de la forma:

A / B / C:

A representa la distribución de llegada de requerimientos en un conmutador

B la distribución de servicio en un elemento conmutador

C el número de troncales de salida para el caso de una central telefónica

A continuación se hará una breve descripción de los procesos de Poisson para luego deducir formulas que servirán para medir el grado de servicio de los sistemas antes descritos.

9.1 Procesos de Poisson

Se utilizan tres enunciados para básicos para definir los procesos de Poisson. Sea Δt un intervalo de tiempo, entonces se tiene:

1. La probabilidad de una llegada en el intervalo Δt se define como $\lambda \Delta t + O(\Delta t)$, $\lambda \Delta t \ll 1$, siendo λ una constante de proporcionalidad especificada.
2. La probabilidad de cero llegadas en Δt es $1 - \lambda \Delta t + O(\Delta t)$.
3. Las llegadas son procesos sin memoria: cada llegada (evento) en un intervalo de tiempo es independiente de eventos en intervalos previos o futuros.

De acuerdo con 1 y 2 no es posible más de una llegada u ocurrencia de un evento en el intervalo Δt , al menos a $O(\Delta t)$.

Sea un intervalo finito T , entonces la probabilidad $p(k)$ de k llegadas en T está dada por:

$$p(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \quad k = 0, 1, 2, 3, \dots \quad (9.1)$$

A esta se conoce como la distribución de Poisson, en la cual se cumple:

$$E(k) = \sum_{k=0}^{\infty} kp(k) = \sigma_k^2 = \lambda T \quad (9.2)$$

Ahora considerando un intervalo de tiempo mayor, se tendrán una serie de eventos de Poisson, los cuales estarán separados en intervalos. Sea τ el tiempo entre llegadas sucesivas, siendo esta una variable aleatoria. En la estadística de Poisson, τ es una variable aleatoria con distribución exponencial, es decir su función densidad de probabilidad $f_{\tau}(\tau)$ está dada por

$$f_{\tau}(\tau) = \lambda e^{-\lambda \tau} \quad (9.3)$$

A continuación se describen los sistemas utilizados en teoría de tráfico, sección que se basa principalmente en el texto “*Wireless Communications*” de Theodore S. Rappaport, y específicamente en el apéndice A que trata de Teoría de entroncamiento.

Existen dos clases principales de sistemas de entroncamiento:

- ❑ Borrado de llamada pérdida (*Lost Call Cleared o LCC*)
- ❑ Retraso de llamada pérdida (*Lost Call Delayed o LCD*)

En el primer sistema cuando un usuario requiere servicio, existe un tiempo mínimo de configuración, después del cual se le es otorgado el acceso a un canal si este está disponible. En la eventualidad de no existir canal disponible, la llamada es interrumpida sin acceso al sistema, teniendo el usuario la oportunidad de volver a intentar después de un tiempo. Se asume que las llamadas llegan con una distribución de Poisson, y además que existe un número casi infinito de usuarios. La fórmula de Erlang B describe el grado de servicio (GOS) como la probabilidad que un usuario arbitrario experimente un bloqueo de llamada en un sistema LCC. Se asume que todas las llamadas bloqueadas son retornadas instantáneamente a un recipiente de usuarios infinito, y puede volver a llamar en cualquier momento. El tiempo entre llamadas sucesivas para un usuario bloqueado es un proceso aleatorio y es asumido con distribución de Poisson.

En el sistema LCD, se utilizan colas para mantener en espera las llamadas inicialmente bloqueadas. Si un usuario llama y los canales se encuentran ocupados, su requerimiento es retrasado hasta que un canal se desocupe. Entonces, dado que un canal no

esta disponible inicialmente, es necesario conocer la probabilidad de que una llamada sea retrasada, hasta que un canal este disponible para su uso. La probabilidad de que un canal no este inmediatamente disponible en un sistema LCD esta determinada por la formula Erlang C. En LCD el GOS es medido por la probabilidad que la llamada sea retrasada en un tiempo mayor que t segundos. Se asume que existe un número casi infinito de usuarios, que todas las llamadas en la cola son eventualmente servidas, que existe un gran número de canales y un gran número de usuarios con patrones de llamada similares.

A continuación se describirán las formulas de Erlang B en LCC y Erlang C en LCD, los cuales se basan en modelos de cola M/M/m, que implica proceso de llegada de Poisson, estadísticas de servicio con distribución exponencial y m troncales de salida (la M viene de procesos de Markov).

9.2 Erlang B

La formula de Erlang B determina la probabilidad que una llamada sea bloqueada, para sistemas que no utilizan colas de espera (LCC). Está basada en los siguientes supuestos:

- Todos los usuarios, incluso los bloqueados, pueden pedir un canal en cualquier momento (sin memoria).
- Todos los canales libres están disponibles para entregar servicio hasta que todos sean ocupados.
- La probabilidad de utilización de un canal (tiempo de servicio) esta exponencialmente distribuido. Es decir, las llamadas largas tienen menos probabilidad de ocurrencia.
- Hay un número finito de canales disponibles.
- La petición de tráfico esta descrita por una distribución de Poisson, lo cual implica un arribo de llamadas en intervalos de tiempo exponenciales.
- Los intervalos de llegada de peticiones de llamada son independientes unas de otras.
- El número de canales ocupados es igual al número de usuarios ocupados.

Sea un sistema con:

C canales

U usuarios

λ número medio de llegada de llamadas por unidad de tiempo (tasa de llegada)

H duración promedio de una llamada

A carga ofrecida por el sistema

A_U carga promedio ofrecida para cada usuario

λ_1 tasa promedio de llegada de llamada de un usuario

entonces $A_U = \lambda_1 H \Rightarrow A = UA_U = \lambda H$.

La situación se muestra en la Figura 9.1.

La probabilidad que una petición de canal de un usuario sea bloqueada esta dada por:

$$\Pr[\text{Bloqueo}] = \Pr[\text{Ninguno de los } C \text{ canales este libre}] \quad (9.4)$$

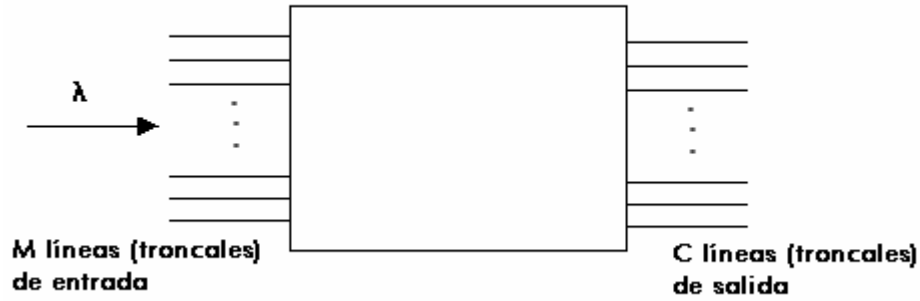


Figura 9.1: Modelo genérico de central de conmutación

Como las llamadas llegan de acuerdo a una distribución de Poisson, como la descrita con anterioridad, de (9.1) se tiene

$$\Pr\{a(t + \tau) - a(t) = n\} = \frac{e^{-\lambda\tau}}{n!} (\lambda\tau)^n \quad \text{para } n = 0, 1, 2, \dots \quad (9.5)$$

Donde $a(t)$ es el número de llegadas o eventos que han ocurrido desde $t=0$ y τ es el intervalo de tiempo entre dos eventos sucesivos.

Como se vio con anterioridad el tiempo de llegada entre eventos es exponencial del tipo descrito en (9.3). Entonces la probabilidad que el tiempo de llegada sea menor que un tiempo s esta dada por:

$$\Pr(\tau_n \leq s) = 1 - e^{-\lambda s} \quad (9.6)$$

El tiempo de servicio es la duración de una llamada particular que ha sido atendida exitosamente en el sistema. El tiempo de servicio se asume exponencial con duración de llamada promedio H , con lo que $\mu = 1/H$ es la tasa de servicio media (número de llamadas por unidad de tiempo). La probabilidad que el tiempo de servicio del n -ésimo usuario sea menor que algún tiempo de duración s esta dada por:

$$\Pr\{S_n < s\} = 1 - e^{-\mu s} \quad s > 0 \quad (9.7)$$

Y la función densidad d probabilidad de tiempo de servicio es

$$p(S_n) = \mu e^{-\mu S_n} \quad (9.8)$$

Donde S_n es el tiempo de servicio del n -ésimo usuario.

Para derivar la formula para Erlang B es necesario utilizar propiedades de las cadenas de Markov. Consideremos un proceso estocástico de tiempo discreto que toma valores desde un conjunto de enteros no negativos, tal que los posibles estados del proceso son $i = 0, 1, 2, \dots$. El proceso es una cadena de Markov si la transición desde el estado presente i al estado próximo $i+1$ depende solo del estado i y no de estados previos. La

operación de sistemas de entroncamiento es de tiempo continuo, pero puede ser analizado en pequeños intervalos δ , donde $\delta > 0$. Si N_k es el número de llamadas (canales ocupados) en el sistema en el tiempo $k\delta$, entonces N_k puede ser representado como

$$N_k = N(k\delta) \quad (9.9)$$

La probabilidad de transición está dada por

$$P_{i,j} = \Pr\{N_{k+1} = j \mid N_k = i\} \quad (9.10)$$

Usando el enunciado básico número 2 de procesos de Poisson y permitiendo que $\delta \rightarrow 0$, se tiene

$$P_{00} = 1 - \lambda\delta + O(\delta) \quad (9.11)$$

$$P_{ii} = 1 - \lambda\delta - \mu\delta + O(\delta) \quad i \geq 1 \quad (9.12)$$

$$P_{i,i+1} = \lambda\delta + O(\delta) \quad i \geq 0 \quad (9.13)$$

$$P_{i,i-1} = \mu\delta + O(\delta) \quad i \geq 1 \quad (9.14)$$

$$P_{i,j} = O(\delta) \quad j \neq i, j \neq i+1, j \neq i-1 \quad (9.15)$$

Estas relaciones quedan mejor graficadas en la figura 9.2.

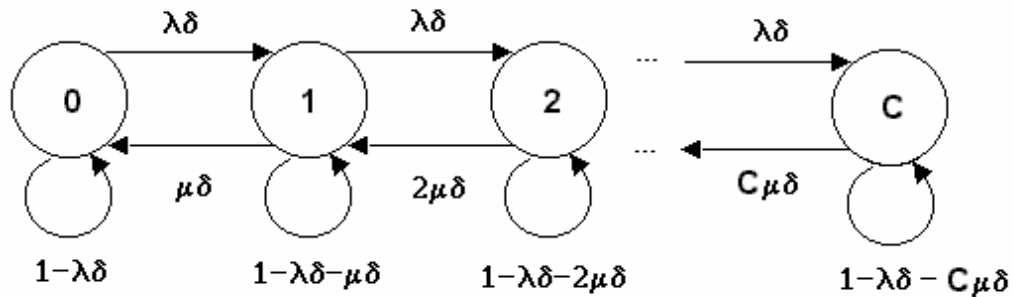


Figura 9.2: Probabilidades de transición representada como una cadena de Markov

Para entender la cadena supongase que al comienzo se tienen 0 canales ocupados, es decir no hay usuarios. Sobre un pequeño intervalo de tiempo, la probabilidad que el sistema continúe sin usuarios es $(1 - \lambda\delta)$. La probabilidad de que haya un cambio desde 0 a 1 usuario esta dada por $\lambda\delta$. En el otro extremo, si un canal esta en uso, la probabilidad de que el sistema pase a 0 canales ocupados esta dada por $\mu\delta$. Similarmente, la probabilidad que el sistema continúe con un canal en uso esta dada por $1 - \lambda\delta - \mu\delta$. Todas las probabilidades de salida para un cierto estado suman 1.

Sobre un gran período de tiempo, el sistema alcanza el estado de régimen permanente y tiene n canales en uso.

Entonces bajo régimen permanente se cumple

$$\lambda \delta P_{n-1} = n\mu \delta P_n \quad n \leq C \quad (9.16)$$

La ecuación (9.16) es conocida como la ecuación general de balance. Además

$$\sum_{n=0}^C P_n = 1 \quad (9.17)$$

Ocupando (9.16) se obtiene

$$P_1 = \frac{\lambda P_0}{\mu} \quad (9.18)$$

Evaluando (9.16) para diferentes valores se obtiene

$$P_n = P_0 \left(\frac{\lambda}{\mu} \right)^n \frac{1}{n!} \quad (9.19)$$

Y

$$P_0 = \left(\frac{\mu}{\lambda} \right)^n P_n n! = 1 - \sum_{i=1}^C P_i \quad (9.20)$$

Sustituyendo (9.19) en (9.20)

$$P_0 = \frac{1}{\sum_{n=0}^C \left(\frac{\lambda}{\mu} \right)^n \frac{1}{n!}} \quad (9.21)$$

En (9.19) la probabilidad de bloqueo para C canales es

$$P_c = P_0 \left(\frac{\lambda}{\mu} \right)^C \frac{1}{C!} \quad (9.22)$$

Sustituyendo (9.21) en (9.22), con $A = \lambda H = \lambda / \mu$ se tiene

$$P_c = \frac{A^C \frac{1}{C!}}{\sum_{n=0}^C A^n \frac{1}{n!}} \quad (9.23)$$

La cual representa la formula para Erlang B.

En la Figura 9.3 se aprecia las curvas características de probabilidad de bloqueo como función del número de canales e intensidad de tráfico en Erlang.

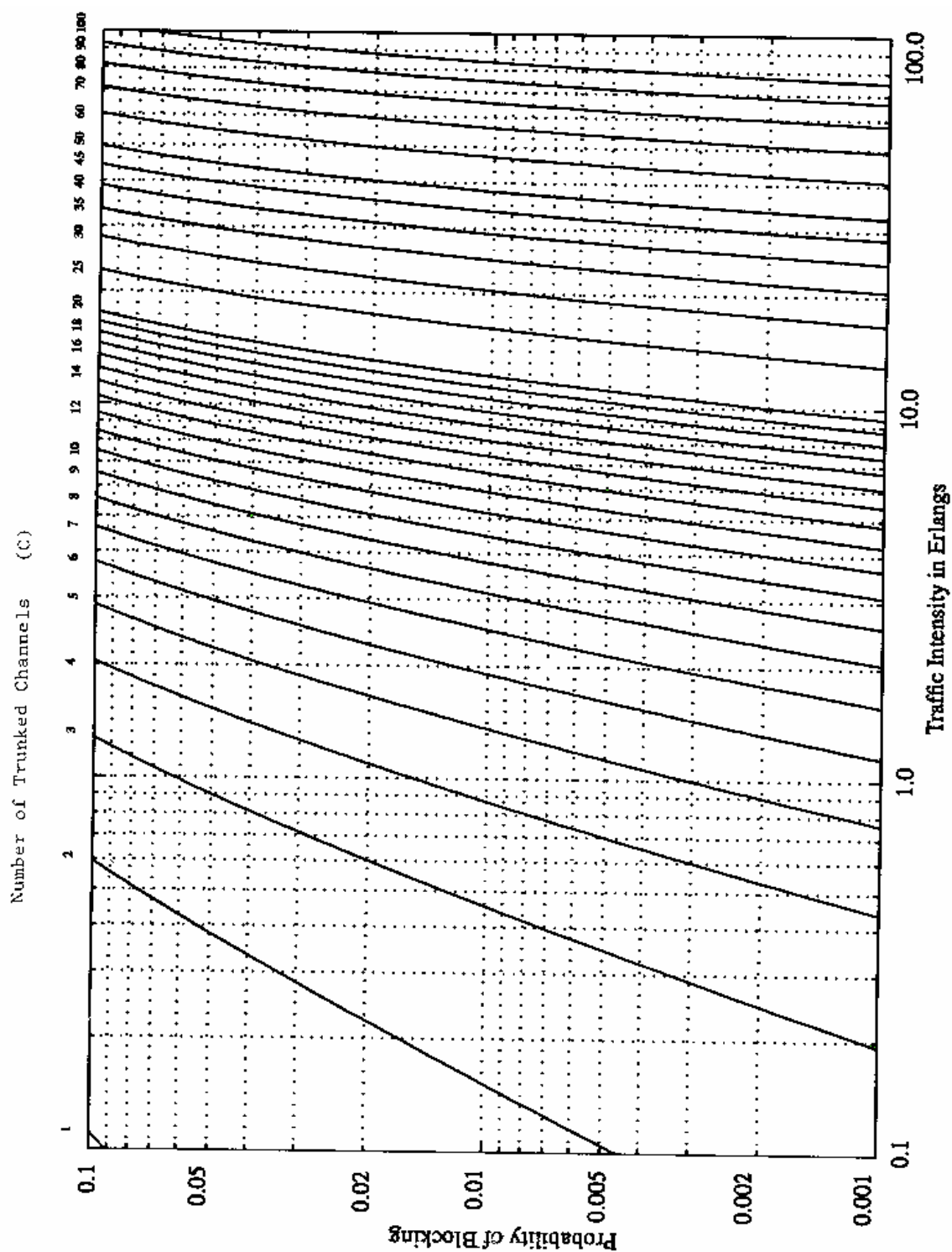


Figura 9.3: Probabilidad de bloqueo como función del número de canales y la intensidad de tráfico en Erlangs. *

* Fuente: "Wireless Communications", T. Rappaport, Capítulo 2

9.3 Erlang C

Sea un sistema con C números de troncales de salida, como se muestra en la figura 9.1. Para derivar la formula de Erlang C se procede de manera similar que en la sección anterior, excepto que ahora se asume que si a una llamada no se le asocia un canal, esta es puesta en una cola (no es bloqueada), la cual tiene un largo infinito. Luego las ecuaciones (9.5), (9.6) y (9.8) siguen siendo validas.

El diagrama de estado para este modelo se muestra en la Figura 9.4.

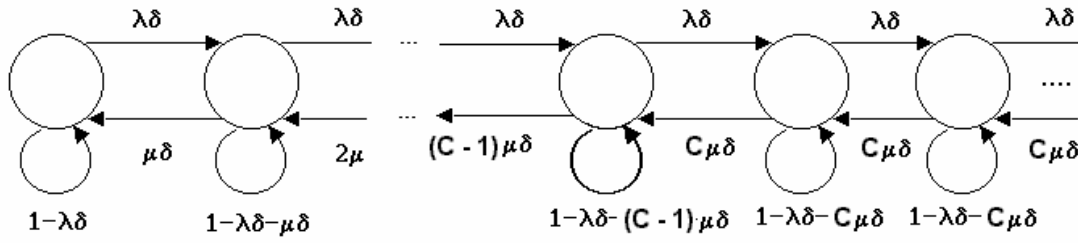


Figura 9.4: Probabilidades de transición como una cadena de Markov

En estado permanente, la probabilidad que el sistema este en estado k y se produzca una transición al estado $k - 1$ en el próximo intervalo de transición es la misma que la probabilidad que el sistema este en estado $k - 1$ y transite hacia el estado k . Entonces desde el diagrama de estado de la figura 9.4

$$\lambda\delta P_{k-1} = k\mu\delta P_k \quad \text{para } k \leq C \quad (9.24)$$

entonces

$$P_k = \left(\frac{\lambda}{\mu}\right) \frac{1}{k} P_{k-1} \quad \text{para } k \leq C \quad (9.25)$$

y

$$\lambda\delta P_{k-1} = C\mu\delta P_k \quad \text{para } k \geq C \quad (9.26)$$

entonces

$$P_k = \left(\frac{\lambda}{\mu}\right) \frac{1}{C} P_{k-1} \quad \text{para } k \geq C \quad (9.27)$$

de lo cual se puede desprender que

$$P_k = \begin{cases} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} P_0 & k \leq C \\ \frac{1}{C!} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{C^{k-C}} P_0 & k \geq C \end{cases} \quad (9.28)$$

Ya que $\sum_{k=0}^{\infty} P_k = 1$ entonces

$$P_0 = \frac{1}{\sum_{k=1}^{C-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \frac{1}{C!} \left(\frac{\lambda}{\mu}\right)^C \frac{1}{\left(1 - \frac{\lambda}{\mu C}\right)}} \quad (9.29)$$

La probabilidad que una llamada llegue cuando todos los C canales estén ocupados y entonces tenga que esperar puede ser determinada usando la ecuación (9.28)

$$\Pr[C \text{ canales esten ocupados}] = \sum_{k=C}^{\infty} P_k = P_0 \frac{1}{C!} \left(\frac{\lambda}{\mu}\right)^C \frac{1}{\left(1 - \frac{\lambda}{\mu C}\right)} \quad (9.30)$$

La cual es válida para $\frac{\lambda}{\mu C} < 1$. Sustituyendo P_0 desde (9.29) y haciendo $A = U\lambda_1 H = \lambda/\mu$ se obtiene:

$$\Pr[C \text{ canales esten ocupados}] = \frac{A^C}{A^C + C! \left(1 - \frac{A}{C}\right) \sum_{k=0}^{C-1} \frac{A^k}{k!}} \quad (9.31)$$

La cual corresponde a la formula de Erlang C.

En la el gráfico de la Figura 9.5 se muestra la probabilidad de llamadas siendo retrasadas como una función del número de canales y la intensidad de tráfico en Erlangs.

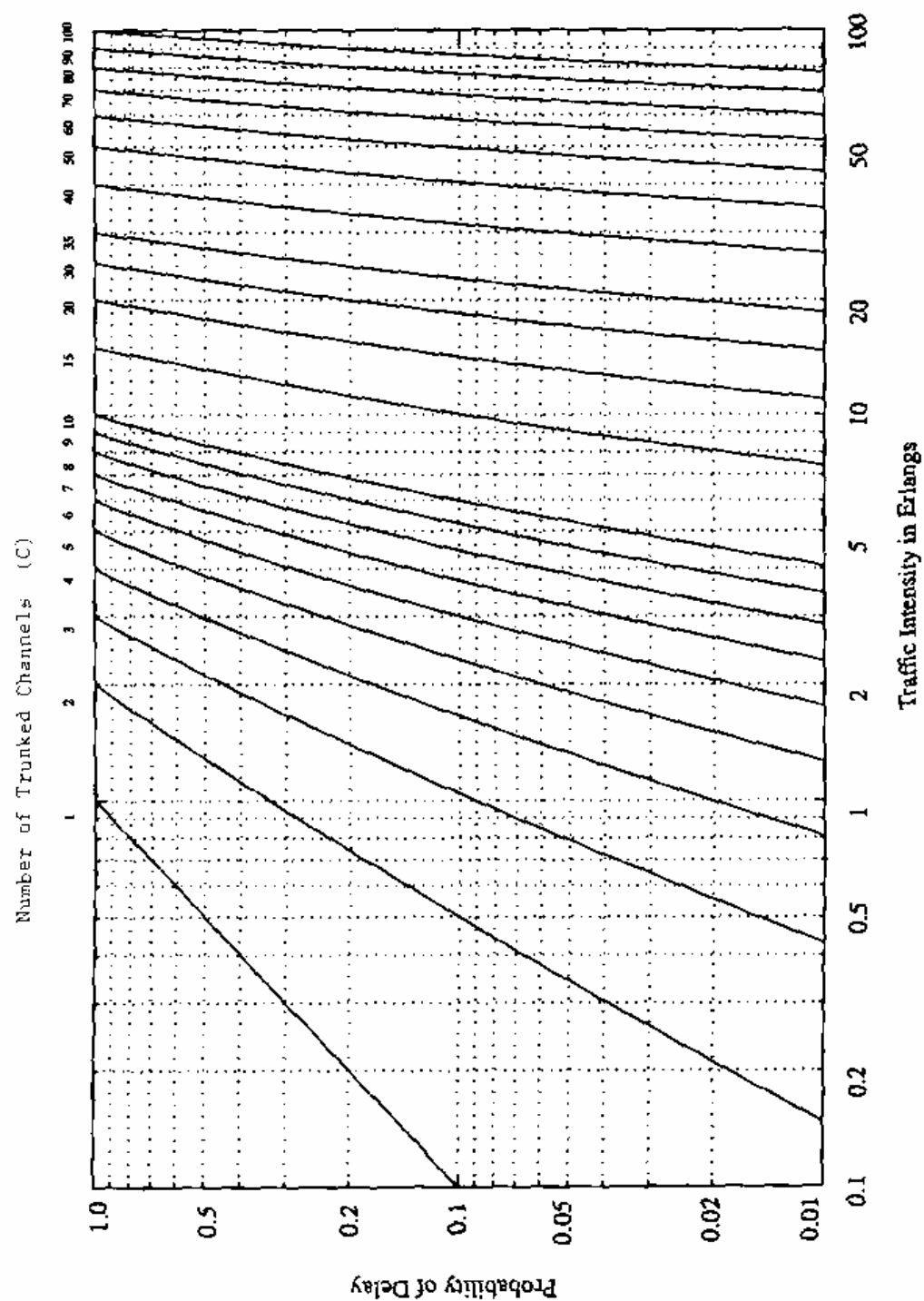


Figura 9.5: Probabilidad que la llamada sea retrasada como una función del número de canales y la intensidad de tráfico en Erlangs. *

* Fuente: “Wireless Communications”, T. Rappaport, Capítulo 2

10 TELEFONÍA CELULAR

La mayoría de la gente se encuentra familiarizada con número de sistemas de comunicación inalámbrico móvil, los cuales pueden ir desde controles remotos, teléfonos inalámbricos, “beepers” y teléfonos celulares. En este capítulo se intentará dar una visión de este último sistema, el cual ha sufrido un crecimiento explosivo en la última década, y ha sido fuente de un gran desarrollo en la industria tecnológica.

El término *móvil* es usado para describir cualquier terminal de radio que pueda ser movido durante su operación, y en forma más específica, se relaciona con equipos que pueden funcionar sobre plataformas de movimiento veloz, como un vehículo.

El primer sistema de telefonía full duplex inalámbrico fue implementado en Estados Unidos bajo el nombre de IMTS (*Improved Mobile Telephone Service*), el cual consistía de transmisores de alta potencia que cubrían áreas de sobre los 50 Km, utilizando modulación FM bajo un sistema analógico FDMA.

El sistema AMPS (*Advance Mobile Phone System*) fue desarrollado por los Laboratorios Bell junto a otras compañías de comunicaciones, los cuales desarrollaron la teoría y técnicas de radiotelefonía celular. Esta implementación consistía en dividir la zona de cobertura en pequeñas celdas, cada una de las cuales reusa una porción del espectro, lo que permite una utilización más eficiente del espectro a expensas de una gran infraestructura (antenas y estaciones base para cada celda). El sistema seguía siendo analógico, utilizando modulación FM y multiplexación en frecuencia (FDMA).

El próximo paso fue la digitalización de los canales de voz, lo que permitió utilizar técnicas de modulación digitales junto con multiplexación en tiempo (TDMA), lo cual incremento la capacidad del sistema desde una a tres comunicaciones simultaneas por cada frecuencia. Para un mayor detalle del desarrollo de los sistemas de comunicaciones celulares ir a “*Wireless Communications*” de Theodore Rappaport.

A continuación se describen brevemente algunos ejemplos de sistemas de comunicación móviles.

Sistema “Paging”

Este sistema consiste en el envío de pequeños mensajes hacia un subcriptor, los cuales pueden ser mensajes numéricos, alfanuméricos o de voz. El mensaje es enviado a través del sistema por medio de un modem o teclado de teléfono, y a este se le conoce como página. El sistema paging, entonces transmite la página a través del área de servicio por medio de una estación base la cual la difunde la página en la portadora de radio.

Un sistema de paging de área amplia consiste de una red de líneas telefónicas, muchas estaciones base transmisoras y torres de radio que simultáneamente difunden una página desde cada una de las estaciones base. Este sistema fue diseñado para proveer comunicación confiable hacia el subcriptor donde quiera que este se encuentre, lo cual implica la utilización de grandes potencias para los transmisores y una baja tasa de datos (un par de miles de bits por segundo).

Teléfono inalámbrico

Es un sistema de comunicación full duplex que usa equipo portátil que mediante ondas de radio se conecta a una estación base fija, la cual a su vez esta conectada a la red pública de telefonía.

Estos teléfonos inalámbricos proveen un rango limitado de movilidad, y típicamente cubren un rango de unos pocos cientos de metros.

Sistema de telefonía celular

Un sistema de comunicación celular acomoda un gran número de usuarios sobre una extensa área geográfica de cobertura, dentro de un limitado espectro de frecuencia.

El área geográfica se encuentra dividida en células (o celdas) las cuales constan de una estación base que tiene asignada una cantidad de frecuencias, que utiliza para llevar a cabo la comunicación. Estas frecuencias (canales) pueden ser reusadas por otra estación base, si esta se ubica lo suficientemente espaciada. Cuando un cliente se mueve desde una celda con una frecuencia hacia otra con otra asignación de frecuencia, se utiliza una técnica de conmutación denominada *handoff* que permite que el subscriptor mantenga su comunicación reasignando la frecuencia.

La figura 10.1 muestra un sistema celular, con sus estaciones bases y un centro de conmutación móvil (MSC).

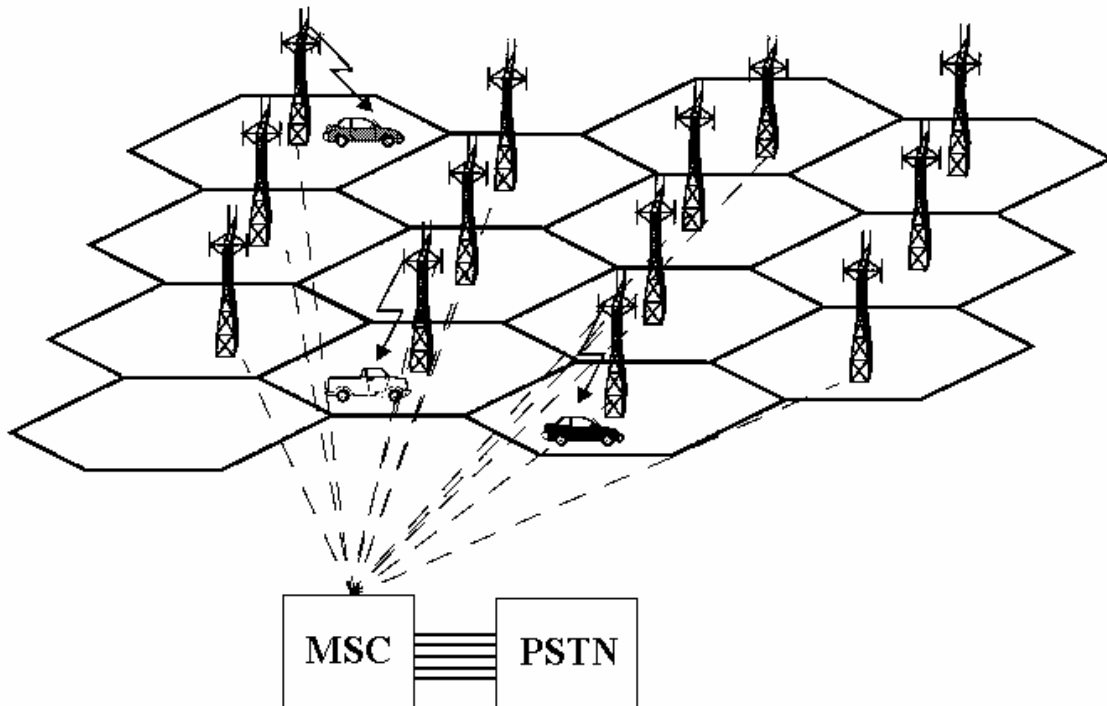


Figura 10.1: Sistema celular, donde las torres representan las estaciones base.

Como muestra la figura, las estaciones base se encuentran conectadas al MSC, cuya función es proveer conexión hacia a la red de telefonía pública conmutada (PSTN) y

coordina la actividad de todas las estaciones base. El subscriptor utiliza un equipo móvil (teléfono celular) para efectuar llamadas, el cual consta de una transceptor, una antena y circuitería de control. La estación base consta de muchas antenas transmisores y receptores que se encargan de manejar las comunicaciones full duplex, y equipo de control.

10.1 Consideraciones de diseño

Reuso de frecuencia

Como se menciono con anterioridad en el capítulo, el área de cobertura para un sistema telefónico móvil se encuentra dividido en regiones de cobertura más pequeños denominados células. A cada una de estas células se le es asignada un grupo de canales que difiere totalmente de sus vecinos, con lo cual se mantiene el nivel de interferencia lo suficientemente bajo, dentro de los límites aceptados. Así es posible volver a asignar el mismo grupo de frecuencia a otra célula que se encuentra en otro sector geográfico. El proceso de diseño de selección y localización de grupos de canales para todas las estaciones base dentro de un sistema es llamado reuso de frecuencia.

Sea un sistema celular con S canales duplex (es decir si un canal simple utiliza 25 kHz, entonces uno duplex utilizará 25×2 kHz) disponibles para el uso. A cada una de las celdas se le asigna un grupo k de canales ($k < S$), y si S es dividido en N celdas con grupos únicos y disjuntos, cada una con el mismo número de canales, el número total de canales disponibles es:

$$S = kN \quad (10.1)$$

Las N celdas que utilizan colectivamente el conjunto completo de frecuencias disponible es llamado *cluster*. Si un cluster es replicado M veces dentro del sistema, el número total de canales duplex está dado por:

$$C = MkN = MS \quad (10.2)$$

El factor N es llamado el tamaño del cluster y sus valores típicos son 4, 7 o 12. El factor de reuso de un sistema celular esta dado por $1/N$. El valor de N es un indicador de que tanta interferencia un teléfono o una estación base es capaz de tolerar mientras mantiene una suficiente calidad de comunicación.

Estrategias de asignación de canal

Básicamente se dividen en: fijas o dinámicas.

Una estrategia fija predetermina un conjunto de frecuencias para cada celda. Cualquier intento de llamada sólo puede ser atendida si existe un canal disponible, y en caso contrario será bloqueada. Una variación de esta es la estrategia de pedir prestado, es decir, la celda puede pedir canales a sus vecinos, si esta posee todos sus canales ocupados. El MSC es el encargado de supervisar este préstamo.

En una asignación dinámica, cada vez que una estación base requiera un canal, esta consulta al MSC, el cual localiza un canal por medio de un algoritmo que toma en cuenta la probabilidad de bloqueo futuro dentro de la célula, la frecuencia de uso del canal

candidato, la distancia de reuso y otras funciones de costo. Esta estrategia reduce la probabilidad de bloqueo, incrementando la capacidad de tráfico del sistema.

Handoff

Se define como el proceso de transferir un móvil desde un canal o estación base a otra. Este proceso debe ser imperceptible para el usuario y por lo general posee mayor prioridad que una secuencia de inicio de llamada, debido que para el usuario es más molesto el corte de la conversación, que el no establecimiento. Para ejecutar este proceso es necesario definir un nivel de señal óptimo de inicio. Para ello se define un nivel mínimo tal que la estación base recibe con una aceptable calidad de voz (normalmente tomado entre -90 dBm y -100dBm), así un nivel levemente más fuerte es usado como umbral para el handoff. Así se define el margen como $\Delta = P_{r \text{ handoff}} - P_{r \text{ mínimo usable}}$, el cual no puede ser ni demasiado pequeño ni demasiado grande. Si es demasiado grande puede ocurrir handoff innecesario y si es demasiado pequeño, el tiempo necesario para completar el handoff puede ser excesivo antes que la llamada se pierda debido a la debilidad de la señal.

Existe el handoff duro (*hard handoff*) y el handoff suave (*soft handoff*). Se diferencian en que el primero al tener un nivel de señal lo suficientemente bajo intenta ubicar frecuencias disponibles en la estación base vecina, en cambio el segundo, utilizado en sistemas CDMA, no intenta cambiar el canal, ya que este es compartido por todas las celdas, y más bien se entiende como el cambio en el manejo de la señal hacia otra estación base.

Interferencia de co-canal

Existen muchas celdas dentro de un sistema que utilizan las mismas frecuencias, luego la interferencia entre señales provenientes de estas celdas es denominada interferencia de co-canal. El aumento en la potencia de la señal portadora incrementa la interferencia, entonces para reducirla es necesario separar las celdas a una distancia mínima, tal que provea la suficiente aislación debido a la propagación.

Cuando se tiene un sistema con tamaño de celdas aproximadamente homogéneo, la interferencia es independiente de la potencia de la portadora, y dependerá del radio de la celda (R), y la distancia al centro más cercano de una celda co-canal (D). El parámetro Q, conocido como radio de reuso de co-canal se encuentra relacionado con el tamaño del cluster de la siguiente manera: (para una geometría de celda hexagonal)

$$Q = \frac{D}{R} = \sqrt{3N} \quad (10.3)$$

Interferencia de canal adyacente

Surge debido a señales que son adyacentes en frecuencia, y se debe a la imperfección en los filtros, los cuales permiten fugas de frecuencias cercanas en el pasabanda.

Puede ser minimizada a través de elección de filtros y asignación inteligente de canales. Esto último se refiere a que no necesariamente en una celda se tienen que establecer frecuencias consecutivas, y puede hacerse de tal modo que la separación entre canales sea máxima.

Grado de servicio y entroncamiento

Entroncamiento (*trunking*) se refiere al hecho de acomodar un gran número de usuarios en un espectro de radio limitado, es decir permitir a un gran número de usuarios compartir un pequeño número de canales en una celda.

Para el diseño se especifica que el sistema puede manejar una cierta capacidad a un específico “grado de servicio”. El grado de servicio (GOS) es una medida de la capacidad de un usuario para acceder a un sistema de trunking durante las horas más ocupadas, y se utiliza para definir el rendimiento deseado de un sistema en particular, especificando la probabilidad de que un usuario obtenga un canal, dado un número de canales disponibles.

El GOS es dado como la probabilidad de que una llamada sea bloqueada (Erlang B) o la probabilidad que experimente un retraso más grande que un cierto tiempo de encolamiento (Erlang C). (ver capítulo anterior)

Para determinar el GOS, generalmente se utilizan los gráficos de las figuras 9.3 para Erlang B y 9.5 para Erlang C, donde un Erlang representa la cantidad de intensidad de tráfico llevado por un canal que está completamente ocupado.

10.2 Técnicas de acceso múltiple

Las técnicas de acceso son las que permiten a múltiples usuarios compartir, de forma simultánea, una cantidad limitada de espectro.

Para establecer una comunicación telefónica es necesario que esta se produzca en dos direcciones a la vez, es decir que la estación base sea capaz de recibir y enviar información en el mismo instante. Este efecto es conocido como *duplexión*, y puede llevarse a cabo utilizando técnicas en el dominio de la frecuencia o en el dominio del tiempo. En frecuencia, esta técnica es conocida duplexación en división de frecuencia (FDD) y utiliza dos bandas distintas, una para recibir y otra para enviar, para cada usuario. La banda de envío provee tráfico desde la estación base al teléfono móvil, y la banda de retorno provee tráfico desde el móvil hacia la estación. Duplexación en división de tiempo (TDD) usa espacios de tiempo para proveer canales de envío y retorno. Si la ranura de tiempo de división entre el canal de envío y el de retorno es pequeño, entonces la transmisión y recepción de datos parece ser simultánea. TDD permite comunicación sobre una canal simple, simplificando el equipo del subscriber.

Las técnicas de acceso múltiples más utilizadas y conocidas son: *acceso múltiple por división en frecuencia (FDMA)*, *acceso múltiple por división de tiempo (TDMA)* y *acceso múltiple por división de código (CDMA)*. Estas técnicas pueden ser agrupadas en sistemas de banda estrecha y sistemas de banda ancha. La técnica de duplexión del sistema de acceso múltiple es frecuentemente especificada junto con el esquema de acceso en particular.

Sistemas de banda estrecha: en estos sistemas el espectro es dividido en un gran número de canales de banda estrecha, operados generalmente con FDD. Para minimizar la interferencia entre canales la división entre las frecuencias es lo más grande posible dentro del espectro de frecuencia. En FDMA de banda estrecha, a cada usuario se le es asignado un canal, el cual no es compartido, y si se utiliza FDD (canal de envío y de retorno) el sistema es llamado FDMA/FDD. Para TDMA de banda estrecha, cada usuario comparte el canal, pero tiene localizado una única ranura de tiempo. Así la técnica de duplexión puede ser FDD o TDD, y el canal es compartido gracias a TDMA.

Sistemas de banda ancha: el ancho de banda de transmisión es mucho más grande que el coherente para un canal. Entonces el desvanecimiento por canales múltiples no afecta mayormente a las señales recibidas. En estos sistemas el usuario se le permite transmitir en una amplia parte del espectro y un gran número de transmisores pueden utilizar un mismo canal. TDMA localiza los slots de tiempo de tal forma que muchos usuarios puedan transmitir en el mismo canal, permitiendo un solo acceso en cada instante de tiempo, mientras que CDMA de espectro disperso permite a todos los transmisores acceder al canal al mismo tiempo. Sistemas TDMA y CDMA pueden usar o FDD o TDD.

En la tabla 10.1 se muestra las diferentes técnicas de acceso en utilización.

Tabla 10.1: Técnicas de acceso múltiple en utilización*

Sistema Celular	Técnica de acceso
Advance Mobile Phone System (AMPS)	FDMA/FDD
Global System for Mobile (GSM)	TDMA/FDD
U. S. Digital Cellular (USDC)	TDMA/FDD
Japanese Digital Cellular (JDC)	TDMA/FDD
CT2 (Teléfonos inalámbricos)	FDMA/TDD
Digital European Cordless Telephone (DECT)	FDMA/TDD
U. S. Narrowband Spread Spectrum (IS-95)	CDMA/FDD

10.2.1 FDMA

Esta técnica de acceso asigna canales individuales para cada usuario. Estos canales son asignados bajo demanda para usuarios que requieran un canal. Durante el período de llamada el canal no puede ser compartido. En sistema FDD a cada usuario se le es asignado un par de frecuencias.

Algunas de las características más importantes de esta técnica son:

- Los canales FDMA llevan sólo un circuito telefónico a la vez.
- Después que un canal de voz es asignado, el teléfono y la estación base transmiten y reciben continua y simultáneamente.
- El ancho de banda de los canales es relativamente estrecho (30 kHz).
- La complejidad de un sistema móvil FDMA es más baja relativa a un sistema TDMA.
- Se necesitan pocos bits para encabezado, debido al esquema de transmisión continua.
- Las celdas FDMA son más costosas que para un sistema TDMA, debido a los canales únicos por portadora y la necesidad de usar costosos filtros pasabanda en las estaciones base.
- FDMA requiere ajustados filtros de RF para minimizar interferencia de canales adyacentes.

10.2.2 TDMA

El sistema TDMA divide el espectro de radio en ranuras (slots) de tiempo, y cada una de las ranuras soporta solo un usuario que envía o recibe. TDMA transmite datos con

* Fuente: Wireless Communications, T. Rappaport

un método de almacenamiento y ráfaga (buffer- and- burst), haciendo que la transmisión para un usuario no sea continua. Esto implica que deben utilizarse señales y técnicas de modulación digitales, a diferencia de FDMA. Las ranuras de tiempo de varios usuarios son puestas en frames cíclicos como muestra la Figura 10.2.

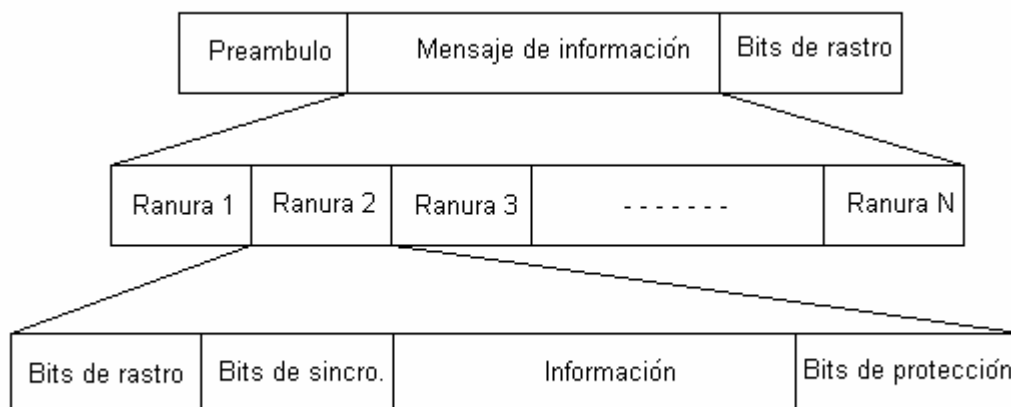


Figura 10.2: Estructura de frame TDMA

En un sistema TDMA/TDD la mitad de las ranuras de tiempo podrán ser usadas para el canal de envío y la otra mitad para el canal de retorno. En un sistema TDMA/FDD se utilizaría el esquema de frame de la figura para dos frecuencias (envío y retorno).

Algunas de las características de TDMA se describen a continuación.

- Muchos usuarios comparten un mismo canal, en diferentes ranuras de tiempo, siendo el tamaño de esta dependiente de la técnica de modulación utilizada, el ancho de banda disponible, etc.
- La transmisión no es continua, pero ocurre en ráfagas.
- Debido a lo anterior el handoff es mucho más simple, ya que le permite escuchar otras estaciones bases en los intervalos desocupados.
- Es necesaria ecualización adaptiva, debido a que las tasas de transmisión son muy altas comparadas con FDMA.
- Se necesita encabezado de sincronización debido a la transmisión en ráfagas. También se requieren bits de protección entre ranuras para separar a los usuarios, lo que lleva a un mayor encabezado comparado con FDMA.

10.2.3 CDMA

La capacidad de sistemas CDMA esta limitada por interferencia, a diferencia de TDMA y FDMA que se encuentran delimitadas por ancho de banda. Por lo tanto, cualquier reducción en la interferencia causará un incremento lineal en la capacidad. CDMA forma parte del estándar IS-95 el cual describe el sistema celular basado en esta técnica. Para mayor detalle referirse a la sección 4.1.5.

11DWDM

(Dense Wavelength Division Multiplexing)

La demanda explosiva por ancho de banda es debida, en mayor medida, al aumento en el tráfico de datos, especialmente el tráfico IP. Esta demanda, además de aumentar, se hace más compleja, pudiendo transitar por una red troncal tráfico basado en circuito (voz y fax TDM), basado en paquetes (IP), o basados en celdas (ATM y Frame Relay). Se suma a esto la porción de datos sensitivos a retraso, como voz sobre IP y video streaming.

En respuesta a este explosivo aumento en la demanda, siendo IP el fundamento común para todos los servicios, las compañías buscan alternativas para los sistemas TDM actuales, que permitan aumentar el ancho de banda sin incurrir en inversiones exageradas, aprovechando la infraestructura disponible.

11.1 Introducción

Las redes troncales de gran capacidad utilizan tecnología basada en fibra óptica, ya que esta permite transmitir un gran ancho de banda a grandes distancias, gracias a que la información viaja en forma de luz, la cual se confina en el núcleo de la fibra. Luego, las posibilidades de aumentar el ancho de banda son básicamente dos: instalar nuevas fibras o incrementar el ancho de banda efectivo de las ya existentes.

Debido a que la primera alternativa necesita de una fuerte inversión, lo que realmente se quiere es aumentar el ancho de banda de las fibras instaladas.

En este capítulo se describirá una técnica que permite aumentar el ancho de banda de una instalación de fibra óptica en un factor de 16 o 32, gracias a la utilización de distintas longitudes de onda sobre una misma fibra, cada una actuando como una canal independiente de las otras longitudes, permitiendo aumentar la capacidad de las redes ya existentes con una relativa baja inversión.

Para aumentar la capacidad de forma efectiva de las fibras existentes, surgen dos alternativas:

- ❑ *Incrementar la tasa de bits de los sistemas existentes*

Usando TDM, los datos pueden ser transmitidos a 2,5 Gbps (OC-48), aumentando hacia los 10 Gbps (OC-192), y considerando los últimos avances se ha logrado llegar a 40 Gbps (OC-768). La circuitería electrónica existe, pero es muy compleja y a la vez cara, además que surgen problemas técnicos como dispersión cromática y por modos de polarización.

- ❑ *Incrementar el número de longitudes de onda sobre la fibra*

Muchas longitudes de onda son combinadas en una sola fibra. Usando *multiplexación por división en longitud de onda (WDM)*, muchos colores de luz multiplexan señales de 2,5 a 40 Gbps sobre una fibra. Sistemas con 128 y 160 longitudes de onda se encuentran en operación, donde los límites de esta tecnología aún no han sido especificados.

11.1.1 TDM

Multiplexación por división en el tiempo fue inventado como una forma de maximizar la cantidad de tráfico de voz que se podía transportar por un medio. Así, usando multiplexación era posible poner a más de una llamada telefónica por enlace.

TDM incrementa la capacidad de transmisión al dividir el tiempo en pequeños intervalos, tal que múltiples entradas de bits pueden ser puestas sobre el mismo medio, utilizando estos intervalos. La Figura 11.1 ejemplifica de mejor forma el concepto.

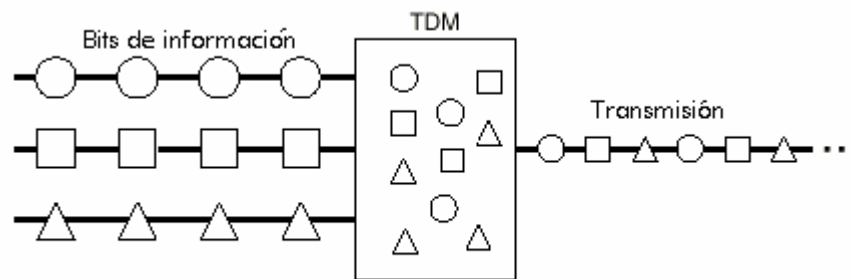


Figura 11.1: Concepto de TDM

Aunque bueno, este método resulta ineficiente, ya que cada ranura de tiempo esta reservada, independientemente si se utiliza.

La industria de las telecomunicaciones adopto como estándares para transporte sobre fibra de datos TDM, las técnicas conocidas como SONET (*Synchronous Optical Network*) en Norteamérica y su contraparte europea SDH (*Synchronous Digital Hierarchy*). Estos estándares especifican parámetros para interfaces, tasas de transmisión, formato de marco (frame), métodos de multiplexación y administración para redes síncronas TDM sobre fibra.

SONET/SDH toma n flujos de bits, los multiplexa, y modula ópticamente la señal, enviándola usando algún componente emisor de luz sobre fibra, con una tasa igual a (tasa de bits entrante) $\times n$. Entonces si sobre un multiplexor SONET llegan 4 entradas a 2,5 Gbps este producida un solo flujo de bits a $4 \times 2,5 \text{ Gbps} = 10 \text{ Gbps}$. La Figura 11.2 ilustra este hecho.

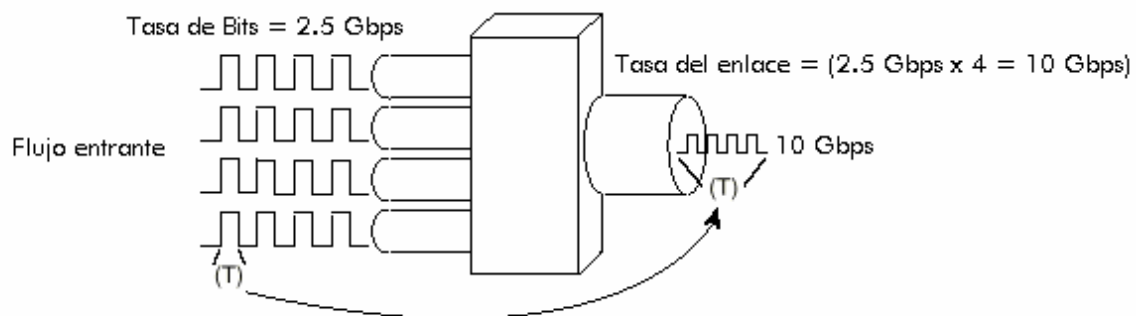


Figura 11.2: Multiplexación TDM en SONET

La unidad básica utilizada para multiplexación, la cual corresponde a una llamada telefónica, es de 64 Kbps, utilizándose en Estados Unidos 24 de estos canales (T1) y en Europa 32 (E1), así cada uno de estos conforman una señal multiplexada con más alta tasa de bits (1,544 Mbps y 2,048 Mbps respectivamente). La jerarquía para multiplexación de llamadas telefónicas se muestra en la Tabla 11.1.

Tabla 11.1 Jerarquía de multiplexación Telco

Señal	Tasa de Bit	Slots de Voz
DS0	64 kbps	1 DS0
DS1	1.544 Mbps	24 DS0s
DS2	6.312 Mbps	96 DS0s
DS3	44.736 Mbps	28 DS1s

Los bloques de construcción básicos para multiplexación en SONET/SDH se muestran en la Tabla 11.2.

Tabla 11.2: Jerarquía de multiplexación SONET/SDH

Portadora Optica	Señal SONET/SDH	Tasa de Bit	Capacidad
OC-1	STS-1	51.84 Mbps	28 DS1s or 1 DS3
OC-3	STS-3/STM-1	155.52 Mbps	84 DS1s or 3 DS3s
OC-12	STS-12/STM-4	622.08 Mbps	336 DS1s or 12 DS3s
OC-48	STS-48/STM-16	2488.32 Mbps	1344 DS1s or 48 DS3s
OC-192	STS-192/STM-64	9953.28 Mbps	5379 DS1s or 192 DS3s

La Figura 11.3 bosqueja esta jerarquía de multiplexación, usando un estándar denominado tributario virtual, los canales de baja velocidad son mapeados en una señal STS-1. Así 28 señales DS1 Pueden ser mapeadas a STS-1 o alternativamente pueden ser multiplexadas a un DS3 y puestas directamente en un STS-1. Es importante notar usando paquetes sobre SONET, es posible alimentar al multiplexor con celdas ATM.

SONET/SDH sufren de algunos problemas, ya con TDM las nociones de prioridad o congestión no existen. Otro problema es que la jerarquía de multiplexación es rígida, es decir, cuando es necesaria más capacidad se debe aumentar el factor de multiplicación, lo que puede significar una inversión en capacidad que no será utilizada. Por ejemplo el siguiente paso incremental de STM-1 (155.52 Mbps) en SDH es STM – 4 (622.08 Mbps).

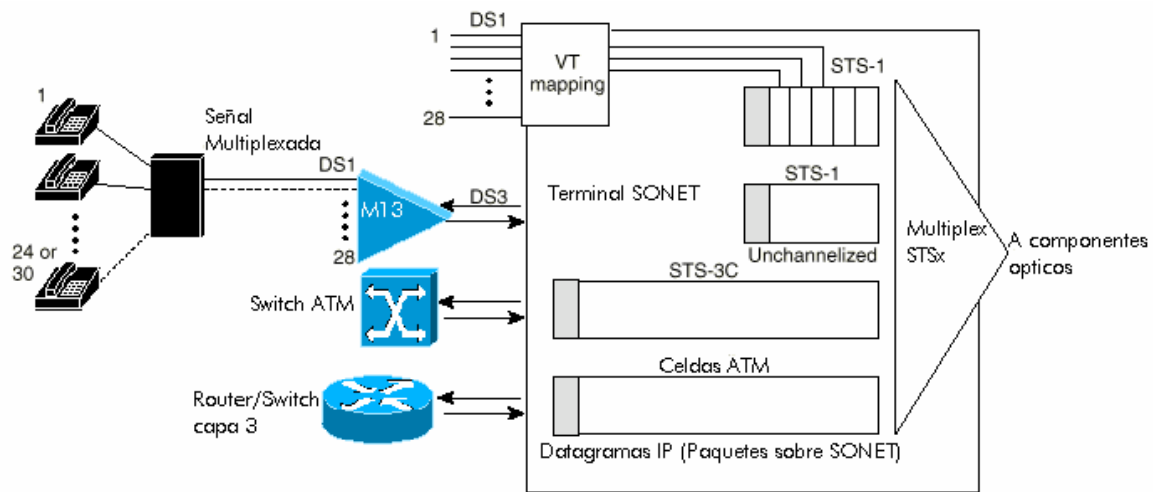


Figura 11.3: Agregación TDM en SONET

11.1.2 WDM (Multiplexación por División en Longitud de Onda)

WDM asigna a cada señal entrante una frecuencia de luz específica dentro de un espectro limitado. Debido a que cada señal se transmite en una frecuencia distinta, no interfieren una con otras. Otra forma de verlo es que cada canal representa un color de luz y un conjunto de señales forman un “arcoiris”. La Figura 11.4 ilustra este método.

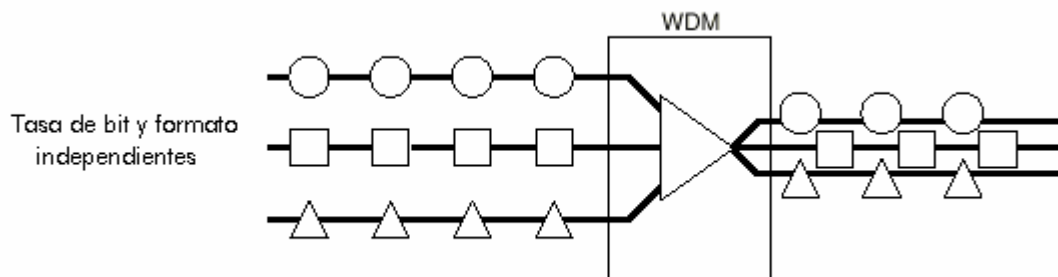


Figura 11.4: WDM

En WDM cada uno de los canales tiene su propio ancho de banda dedicado y todas las señales llegan al mismo tiempo al receptor, siendo demultiplexadas.

En DWDM (*Dense WDM*) las longitudes de onda se encuentran menos espaciadas y por lo tanto alcanza una mayor capacidad, alcanzando hasta el año 2000 128 longitudes de onda en una fibra.

Otro punto importante de mencionar es que tanto WDM como DWDM trabajan sobre fibras monomodo. Esto no debe llevar a confusión, pues las fibras multimodo utilizan diferentes ángulos de reflexión en el núcleo, resultando en distintos “modos” de transmisión.

SONET TDM toma las señales síncronas y asíncronas y las multiplexa en una única señal con gran tasa de bits en una única longitud de onda sobre la fibra. Las señales pueden tener que ser transformadas de eléctricas a ópticas antes de ser multiplexadas. En cambio, en WDM se toman múltiples señales ópticas y cada una de ellas es mapeada en una longitud de onda, y se multiplexan estas longitudes sobre una única fibra. WDM también puede llevar señales de múltiples protocolos sin un formato común, mientras que SONET no. Esta diferencia es ilustrada en la Figura 11.5.

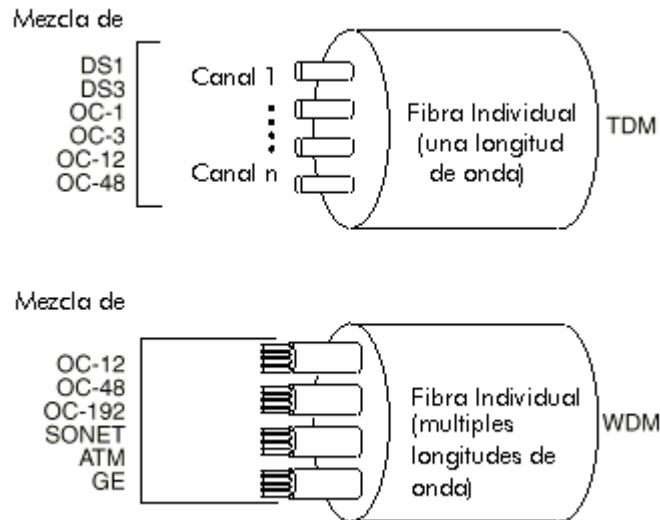


Figura 11.5: Interfaces TDM y WDM

Con WDM es posible transportar TDM, lo que permitiría preservar inversiones en equipos ya existentes para SDH y SONET. Actualmente las nuevas implementaciones pueden eliminar capas de equipo. Por ejemplo, los equipos multiplexores para SONET pueden ser eliminados interfazando conmutadores de paquetes y equipos ATM directamente a equipos WDM, ya que para estos las interfaces OC-48 son comunes. La Figura 11.6 ilustra este hecho.

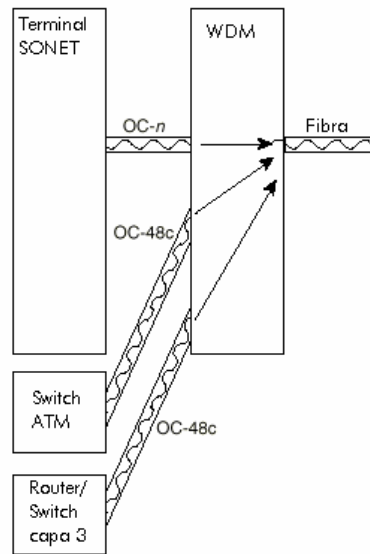


Figura 11.6: Interfaces desde SONET a WDM

11.2 Fundamentos de la tecnología WDM

El primer sistema WDM fue utilizado en la década de los 80's y consistía en el uso de dos longitudes de onda, una en los 1310 nm y la otra en los 1550 nm (o 850 nm y 1310 nm), el cual fue llamado WDM de banda ancha. Para transmitir se utiliza un par de fibras como lo muestra la Figura 11.7.

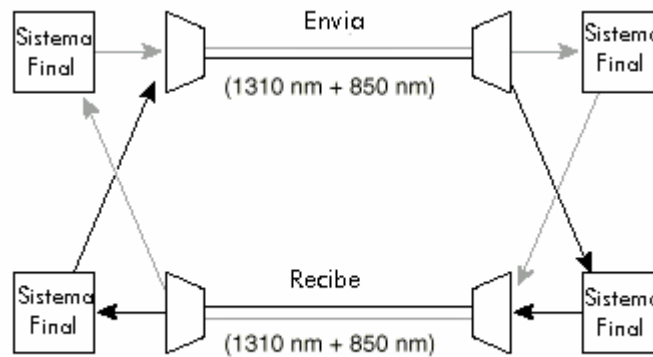


Figura 11.7: WDM de dos canales

En la década de los 90 surgió la segunda generación de sistemas WDM, conocida como WDM de banda estrecha, en la cual los canales podían ir desde dos a ocho. Estos canales fueron espaciados a intervalos de sobre los 400 GHz en la ventana de los 1550 nm. A mediados de los 90's fueron emergiendo los sistemas DWDM con 16 a 40 canales y

espaciamientos de 100 GHz a 200 GHz. A fines de la década, los sistemas DWDM fueron evolucionando hasta tener capacidades de 64 a 160 canales paralelos, densamente empaquetados en espacios de entre 50 a 25 GHz.

En la Figura 11.8 se muestra un sistema WDM de cuatro canales, el cual ilustra las funcionalidades.

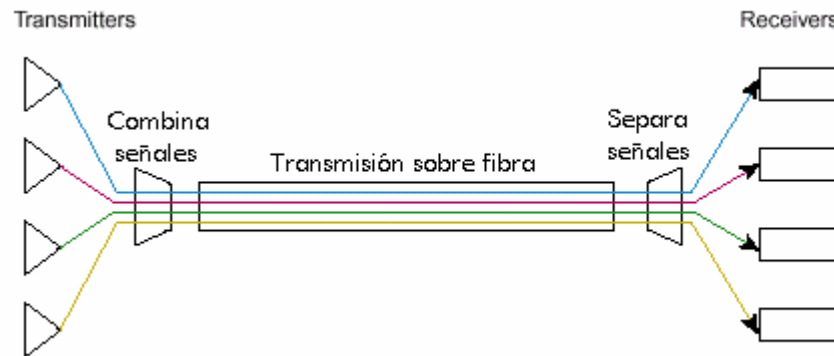


Figura 11.8: Esquema funcional WDM

Así, las principales funciones en un sistema WDM son:

- **Generar la señal:** la fuente (un laser de estado sólido) debe proveer luz estable dentro de un estrecho y específico ancho de banda para transportar los datos digitales, modulándolos como una señal análoga.
- **Combinar las señales:** se emplean multiplexores. Hay pérdidas asociadas a este proceso, las cuales pueden ser reducidas mediante el uso de amplificadores ópticos, los cuales estimulan todas las longitudes de onda a la vez sin realizar conversión optoelectrónica.
- **Transmitir la señal:** Deben tenerse en cuenta los efectos de degradación y pérdida de señal, controlándolos mediante los niveles de potencia de los lasers, y espaciamiento entre longitudes de onda.
- **Separar las señales recibidas:** se deben separar las señales multiplexadas, siendo este uno de los trabajos más difíciles.
- **Recibir la señal:** es efectuado mediante el uso de un fotodetector.

Los componentes esenciales de un sistema DWDM pueden ser clasificados por el lugar que ocupan dentro de él:

- ❑ En el lado del transmisor, lasers con longitudes de onda precisas y estables.
- ❑ Sobre el enlace, fibras ópticas que posean bajas pérdidas y gran rendimiento en el espectro de longitudes de onda utilizado y amplificadores ópticos de ganancia plana que estimulan la señal para tramos de gran distancia.
- ❑ En el lado del receptor, fotodetectores y demultiplexores ópticos que utilicen filtros o elementos difractivos.

A continuación se describirán, en forma breve, algunos de los componentes anteriormente mencionados.

11.2.1 Fibra óptica

El trabajo principal de la fibra es guiar la luz con el mínimo de atenuación. Esta compuesta por un núcleo de vidrio que permite dos tercios de la velocidad de la luz en el vacío. El núcleo se encuentra rodeado por un manto, también de vidrio, pero con un menor índice de refracción que el del núcleo. Así, la transmisión de la luz en la fibra es explicada por el principio de reflexión interna total, en donde el 100 % de la luz que viaja por el núcleo y golpea en la interfaz con el manto es reflectada.

La luz es reflectada o refractada dependiendo del ángulo de incidencia, y la reflexión total interna sucede cuando el haz, que viaja por un material de cierta densidad, llega a una interfaz con un material de menor densidad, y el ángulo de incidencia es menor que el ángulo crítico, el cual es el máximo ángulo de incidencia tal que la luz no es refractada hacia el otro medio. Este fenómeno se ilustra en la Figura 11.9.

Una fibra óptica esta compuesta de dos distintos tipos de vidrio (silicio) de alta pureza, uno para el núcleo y otro para el manto, los cuales son dopados para ajustar sus índices de refracción tal que $n_1 < n_2$.

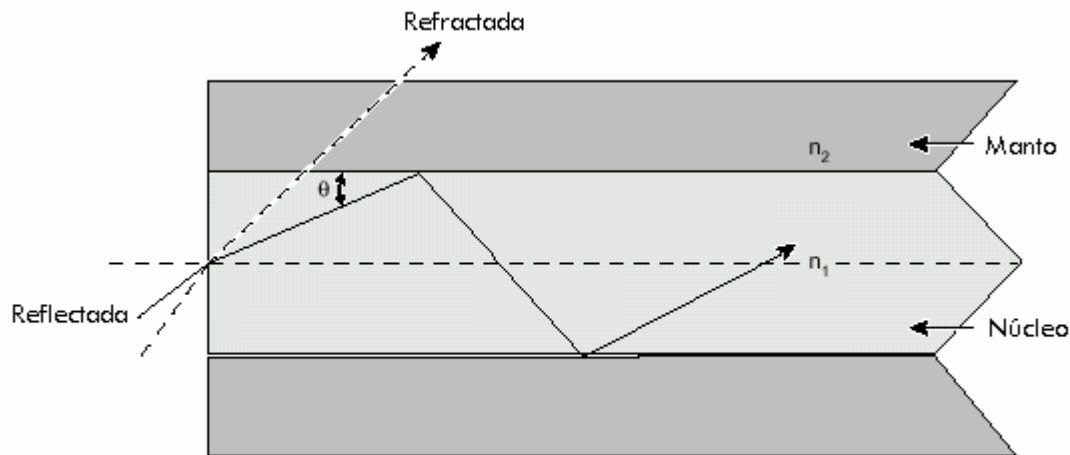


Figura 11.9: Refracción total interna

Existen dos tipos de fibra utilizados hoy en día, la fibra multimodo y la monomodo. La primera posee un mayor núcleo y soporta un mayor número rayos propagándose por el. Es índice de refracción en esta fibra es uniforme en el núcleo, y las distintas señales viajan rebotando con distintos ángulos de incidencia, como muestra la Figura 11.10.

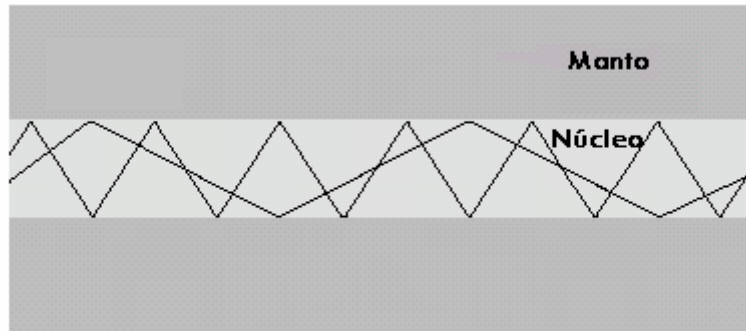


Figura 11.10: Fibra multimodo

Notar que en la figura 11.10, las señales no se demoran lo mismo en llegar a su destino, llamándose a este efecto dispersión modal. Debido a esto es que esta fibra no es utilizada para aplicaciones de área extendida.

El segundo gran tipo de fibra es la monomodo, la cual posee un menor núcleo que la anterior, lo cual permite un solo modo de propagación. Esto provoca una mejor fidelidad de la señal para tramos largos, pues la dispersión modal es altamente reducida. Por su gran capacidad de transportar información y bajas pérdidas intrínsecas, esta fibra es la preferida para aplicaciones de largas distancias y alto ancho de banda, incluyendo DWDM.

Con respecto a las longitudes de transmisión, han sido explotadas principalmente 3 ventanas dentro del espectro infrarrojo. La primera de ellas, cercana a los 850 nm, fue utilizada para tramos cortos en aplicaciones multimodo. Fibras no dispersivas desplazadas, más bien conocidas como fibras monomodo, fueron diseñadas para ser usadas en la segunda ventana, la cual corresponde a los 1310 nm. Así la fibra fue diseñada de tal forma que la dispersión cromática cercana a esta longitud fuera casi zero. La tercera ventana explotada corresponde a la de los 1550 nm, la cual posee menor atenuación y su frecuencia de operación fue la misma que la de los nuevos amplificadores dopados con erbio (EDFAs).

Por último mencionar que la transmisión de la luz por fibra presenta algunos retos clasificados como:

- ❑ Atenuación: decaimiento de la intensidad de la señal, o pérdida de potencia óptica, mientras la señal se propaga por la fibra.
- ❑ Dispersión cromática: ensanchamiento del pulso de luz cuando viaja por la fibra.
- ❑ No linealidades: efectos acumulativos de la interacción de la luz con el material a través del cual esta viaja, resultando en cambios de las ondas de luz e interacciones entre ellas.

Todos estos efectos poseen muchas causas las cuales escapan al propósito de este apunte.

11.2.2 Emisores de luz: LED y Lasers

Los emisores de luz utilizados para transmisión óptica deben ser compactos, monocromáticos, estables y de larga duración.

Son dos los tipos generales de fuentes emisoras, los diodos emisores de luz (LED) y los diodos laser, o semiconductores laser. Los LEDs son componentes confiables para velocidades menores que 1Gbps, y exhiben un espectro relativamente ancho. Son utilizados generalmente para fibras multimodo. Por otro lado, los semiconductores laser poseen mejores características para aplicaciones en fibra monomodo. Algunas características de estos son precisión en longitud de onda y espectro estrecho, lo cual le permite transmitir a mayores tasas (sobre los 10 Gbps).

Dos de los tipos de lasers más ocupados son el laser monolítico Fabry – Perot y el laser de realimentación distribuida (DFB), siendo el último el más confiable para aplicaciones DWDM pues emite luz casi monocromática, soporta grandes velocidades, tiene una favorable razón señal a ruido y tiene linealidad superior. La frecuencia central de estos lasers son alrededor de los 1310 nm, y desde 1520 a 1565 nm.

11.2.3 Detectores de luz

Son dos los tipos de fotodetectores utilizados en comunicaciones ópticas: fotodiodo PIN (positive . intrinsic – negative) y el fotodiodo de avalancha (APD). El PIN trabaja de una forma similar pero inversa que un LED, es decir los fotones son convertidos a electrones en una relación 1:1. Los ADP son similares a los PIN , pero poseen ganancia debido a un proceso de amplificación: un fotón actúa sobre el componente produciendo muchos electrones (efecto avalancha), estos los hace más sensibles y exactos pero disminuye su velocidad de respuesta y son de mayor costo.

11.2.4 Multiplexores y demultiplexores

Ya que DWDM es un sistema que envía señales desde muchas fuentes, sobre una única fibra, ello debe incluir algún medio para combinar tales señales. Este trabajo es realizado por un multiplexor, el cual toma las distintas longitudes de onda provenientes de múltiples fibras y las hace converger en un solo haz. El extremo receptor del sistema debe ser capaz de separar las componentes de luz tal que cada una de ellas pueda ser detectada. El demultiplexor ejecuta esta función, separando el haz en distintas longitudes de onda y acoplándolas a fibras individuales. La demultiplexión es ejecutada antes de la detección.

Como muestra la Figura 11.1 en sistemas unidireccionales, es necesario dos fibras para una comunicación bidireccional.

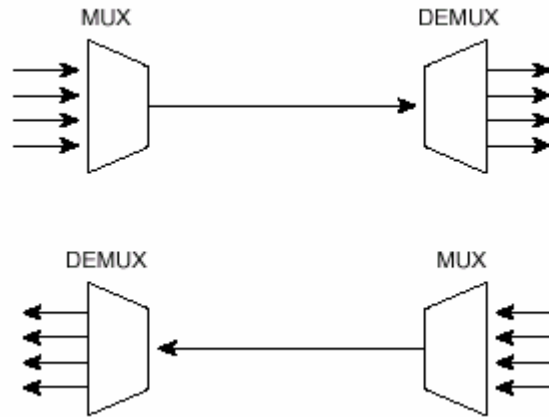


Figura 11.11: Multiplexión y demultiplexión en sistemas unidireccionales

En la figura 11.12 se muestra el caso bidireccional.



Figura 11.12: Multiplexión y demultiplexión en sistemas bidireccionales

Los principales retos de estos equipos son minimizar el cross-talk y maximizar la separación de canal. El cross-talk es la medida de que tan bien separados se encuentran los canales, mientras que la separación de canal se refiere a la capacidad de distinguir entre una longitud de onda y otra.

Una de las técnicas más simple de multiplexación y demultiplexación es la utilización de prismas, como lo muestra la Figura 11.13.

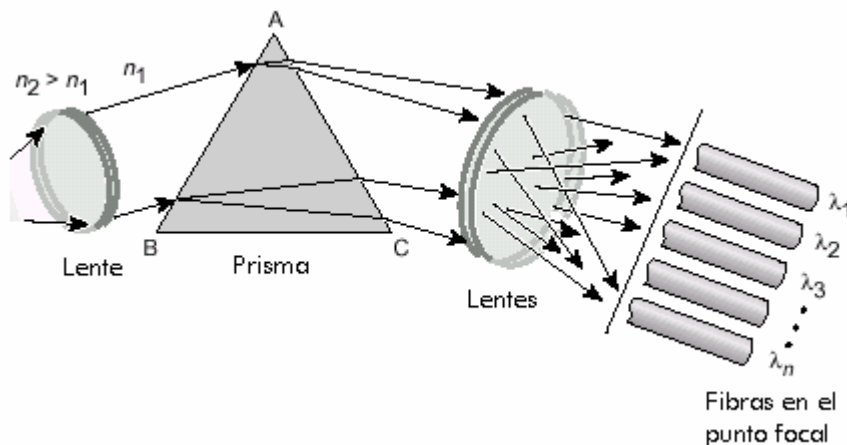


Figura 11.13: Refracción en un prisma como medio de demultiplexión

Otra tecnología utilizada esta basada en el principio de difracción en una interfaz óptica. Cuando una luz policromática choca contra una rejilla de difracción, cada una de las longitudes de onda es difractada en un ángulo distinto y por lo tanto a diferentes puntos del espacio. Con el uso de un lente estas longitudes de onda pueden ser enfocadas dentro de fibras individuales, como lo muestra la Figura 11.14.

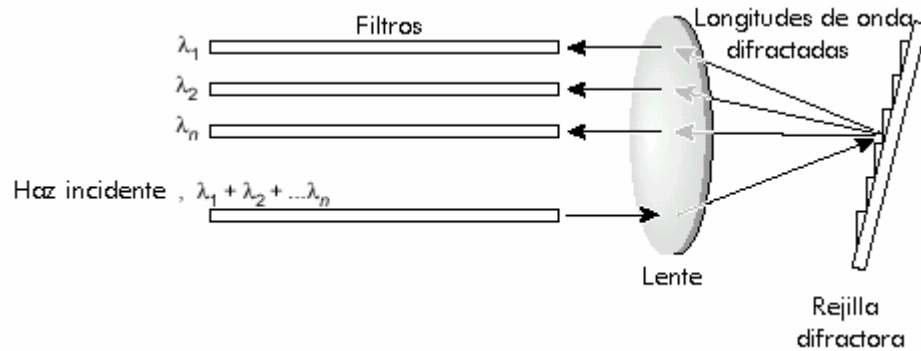


Figura 11.14: Rejilla difractora como guía de onda

Existen más métodos, los cuales se basan básicamente en estos dos principios, por lo cual no serán mencionados en este apunte.

12REFERENCIAS

- *“Redes de Computadoras”*
Andrew Tanenbaum, 3º Ed, Prentice Hall
- Cisco System, Internetworking Technology Overview, Cisco System 1997.
- *“Sistemas de Comunicaciones Electrónicas”*
Wayne Tomasi
- *“Digital Signal Processing: principles, algorithms and applications”*
John Proakis y Dimitris Manolakis, Prentice Hall
- *“Redes Globales de Información con Internet y TCP/IP, principios básicos, protocolos y arquitectura”*
Douglas E. Comer, 3º Ed, Prentice Hall Hispanoamérica, 1996
- *“Redes de Telecomunicaciones: Protocolos, modelado y análisis”*
Schwartz, M., Addison Wesley, 1994
- *“Wireless Communications, principles & practice”*
Theodore S. Rappaport, Prentice Hall, 1996
- *“Introduction to DWDM for Metropolitan Networks”*
Cisco System, Inc., 2000

4 CAPA FÍSICA

En este capítulo se estudiará una de las capas que forman parte del modelo de referencia OSI, cuya funcionalidad es la proveer el medio de comunicación de más bajo nivel, como lo es la capa física. Se comenzará con una descripción de la forma en que puede ser modulada la información a ser puesta en el canal, para luego tratar los distintos medios físicos que hacen posible esta comunicación.

1.1 Sistemas de Modulación (digitales)

La modulación es el proceso de codificación de la información proveniente de una fuente tal que de esta forma sea más adecuada para el canal de transmisión. Aquí es importante diferenciar dos términos que serán utilizados para esta sección: Señal de *banda base*, la cual corresponde a la señal proveniente de la fuente, y una señal *pasa banda* la cual corresponde a una señal desplazada a frecuencias mucho más altas comparadas con su banda base. La modulación puede llevarse a cabo por variaciones de amplitud (AM), variaciones de fase (PM) o frecuencia (FM). Demodulación es el proceso inverso, el cual extrae la información desde una señal pasa banda (señal modulada).

Existen técnicas de modulación análogas y digitales, siendo las primeras utilizadas para comunicaciones de radio y televisión, y las segundas para comunicaciones inalámbricas y redes de datos.

En este capítulo serán solo tratadas técnicas de modulación digitales, siendo la contraparte temas ya en dominio del lector.

Algunas de las ventajas que ofrece la modulación digital versus la análoga son:

- ❑ Mayor inmunidad al ruido y robustez ante deterioro del canal.
- ❑ Mayor facilidad de multiplexación de varias formas de información.
- ❑ Más seguridad.

La eficiencia de un esquema de modulación puede ser medido (“*Wireless Communications*” (Theodore S. Rappaport), capítulo 5 “*Modulation Techniques for Mobile Radio*”) en términos de su *eficiencia de potencia* y *eficiencia de ancho de banda*. La primera describe la capacidad de un modulador para preservar la fidelidad del mensaje digital a niveles de potencia bajos. En un sistema de comunicación digital para incrementar la inmunidad al ruido es necesario incrementar la potencia de la señal. Sin embargo, la cantidad a la cual la potencia de la señal podría ser incrementada para obtener un cierto nivel de fidelidad depende del tipo particular de modulación empleado. La eficiencia de potencia, η_p de un esquema de modulación digital es una medida de que tan favorable es el intercambio entre fidelidad y potencia de la señal, y es frecuentemente expresado como la razón de la *energía por bit de la señal por la densidad espectral de la potencia del ruido* (E_b/N_o) requerido para recibir entrada para una cierta probabilidad de error.

La *eficiencia de ancho de banda* cuantifica la habilidad de un esquema de modulación para acomodar datos dentro de un ancho de banda limitado. (cubierto en capítulo anterior).

1.1.1 BFSK (Binary Frequency Shift keying)

En este esquema de modulación, la frecuencia de una señal portadora de amplitud constante es conmutada entre dos valores de acuerdo a los posibles estados del mensaje (1 y 0 binario), es similar a la modulación en frecuencia convencional, excepto que la señal modulante posee sólo dos estados. Dependiendo de la forma en que es introducido este cambio en la frecuencia, la señal FSK tendrá o fase continua (CPFSK) o fase discontinua entre bits.

$$S_{FSK}(t) = \sqrt{\frac{2E_b}{T_b}} \cos[(2\pi f_c + 2\pi \Delta f)t] \quad 0 \leq t \leq T_b \quad (\text{binario } 1) \quad (4.1)$$

$$S_{FSK}(t) = \sqrt{\frac{2E_b}{T_b}} \cos[(2\pi f_c - 2\pi \Delta f)t] \quad 0 \leq t \leq T_b \quad (\text{binario } 0) \quad (4.2)$$

Una forma de generar señales FSK es conmutar entre dos osciladores en base a si el bit es 0 o 1. La forma de onda generada no es continua (FSK discontinua). Esto conlleva problemas de desparramamiento espectral, y por lo tanto no es usado en sistemas inalámbricos.

En forma más frecuente se utiliza un tipo de modulación que es similar a un generador análogo de FM, con excepción que la señal modulada $m(t)$ es binaria.

$$S_{FSK}(t) = \sqrt{\frac{2E_b}{T_b}} \cos \left[2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\eta) d\eta \right] \quad (4.3)$$

A pesar que $m(t)$ es discontinua en la transición, la fase es proporcional a la integral de $m(t)$ y es continua.

1.1.2 BPSK (Binary Phase Shift keying)

La fase de una señal portadora de amplitud constante es conmutada entre dos valores de acuerdo a las dos posibles señales correspondientes a 1 y 0 binario. Normalmente las dos fases son separadas por 180° . Si la portadora sinusoidal tiene una amplitud A_c y una energía por bit $E_b = \frac{1}{2} A_c^2 T_b$, entonces la señal transmitida será:

$$S_{BPSK} = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c) \quad 0 \leq t \leq T_b \quad (\text{binario } 1) \quad (4.4)$$

o

$$S_{BPSK} = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c) \quad 0 \leq t \leq T_b \quad (\text{binario } 0) \quad (4.5)$$

1.1.3 QPSK (Quadrature Phase Shift keying)

Tiene el doble de eficiencia de ancho de banda que BPSK, pues 2 bits son transmitidos en un símbolo simple de modulación (00, 01, 10 y 11). La fase de la portadora toma uno de cuatro valores equiespaciados, donde cada uno de los valores de la fase corresponde a un único par de mensajes de bits.

$$S_{QPSK} = \sqrt{\frac{2E_s}{T_s}} \cos\left[2\pi f_c t + (i-1)\frac{\pi}{2}\right] \quad 0 \leq t \leq T_s \quad i = 1, 2, 3, 4 \quad (4.6)$$

Donde T_s es la duración del símbolo y es igual a dos veces T_b

Un dato importante de mencionar es que la probabilidad de error de bit del QPSK es igual a la del BPSK, pero el primero puede enviar el doble de datos por el mismo ancho de banda.

1.1.4 QAM (Quadrature Amplitud Modulation)

En los métodos anteriormente estudiados, se tenía un sistema de modulación el cual entrega una señal de amplitud constante y variante en su fase o frecuencia. La idea de QAM es permitir que junto con la fase, varíe también la amplitud. Para un sistema de modulación general (M-ario) se puede definir QAM como:

$$S_i(t) = \sqrt{\frac{2E_{\min}}{T_s}} a_i \cos(2\pi f_c t) + \sqrt{\frac{2E_{\min}}{T_s}} b_i \sin(2\pi f_c t) \quad 0 \leq t \leq T \quad i = 1, 2, \dots, M \quad (4.10)$$

Donde

E_{\min} es la energía de la señal con la más baja amplitud y a_i y b_i son enteros independientes elegidos de acuerdo a consideraciones métricas para cada punto.

1.1.5 CDMA (Code Division Multiple Access)

En este sistema, el mensaje es multiplicado por una señal de gran ancho de banda denominada señal “dispersa” (spreading), la cual es una secuencia de pseudo ruido que tiene una tasa de símbolo (chip) que es ordenes de magnitud más grande que la tasa de datos del mensaje. Todos los usuarios utilizan la misma frecuencia y pueden transmitir en forma simultanea. Cada usuario posee su código de pseudo ruido aleatorio el cual es aproximadamente ortogonal a los otros códigos. Así todos los otros códigos aparecen como ruido debido a la descorrelación. Para la detección el receptor necesita conocer el código usado por el transmisor.

A continuación se verá con mayor detalle este punto.

Secuencia de pseudo-ruido (PN)

Una secuencia de pseudo ruido o pseudo aleatoria es una secuencia binaria con una autocorrelación que se asemeja a la autocorrelación de ruido blanco limitado en banda. Es una secuencia determinística, pero posee características similares a la de una secuencia aleatoria binaria como tener un número casi igual de unos y ceros, muy baja correlación entre versiones desplazadas de la secuencia, y entre distintas secuencias, etc.

Este pseudo ruido es utilizado en sistemas DS-SS (*direct sequence spread spectrum*), los cuales expanden la banda base de datos al multiplicar los pulsos de datos con este pseudo ruido. Los sistemas DS-SS también son conocidos como sistemas CDMA.

Un pulso simple o símbolo de la secuencia PN es llamado un *chip*. En la Figura 4.0 se muestra un diagrama en bloques de un sistema DS con modulación de fase binaria.

La señal de espectro esparcido para un usuario simple puede ser representada como:

$$S_{ss}(t) = \sqrt{\frac{2E_s}{T_s}} m(t) p(t) \cos(2\pi f_c t + \theta) \quad (4.11)$$

Donde $m(t)$ es la secuencia de datos, $p(t)$ es el PN, f_c es la frecuencia de la portadora y θ es la fase de la portadora en $t=0$.

Cada símbolo en $m(t)$ representa un dato y tiene duración T_s . Cada uno de los pulsos en $p(t)$ representa un chip, y es usualmente rectangular de amplitud igual a 1 y -1, de duración T_c . La transmisión de los símbolos de datos y los chips cumplen con que la razón entre T_s y T_c es un valor entero. Si W_{ss} es el ancho de banda de $S_{ss}(t)$ y B es el de $m(t)\cos(2\pi f_c t)$, se cumple que $W_{ss} \gg B$.

En el receptor, la señal recibida pasa a través de un filtro y luego es multiplicada por una réplica local del pseudo código $p(t)$. Si $p(t)=1$, y entonces $p^2(t)=1$, lo cual lleva a:

$$S_1(t) = \sqrt{\frac{2E_s}{T_s}} m(t) \cos(2\pi f_c t + \theta) \quad (4.12)$$

La cual corresponde a la entrada del demodulador.

En CDMA el piso de ruido es determinado por la potencia de los múltiples usuarios. Lo que ocurre es que la señal más fuerte recibida capturará el demodulador en la estación base, aumentando el piso de ruido para las señales con niveles más débiles, y por lo tanto disminuyendo la probabilidad de que la señal débil sea recibida. Para evitar este efecto se utilizan controles de potencia en cada estación base, asegurando que cada señal recibida tenga el mismo nivel de potencia.

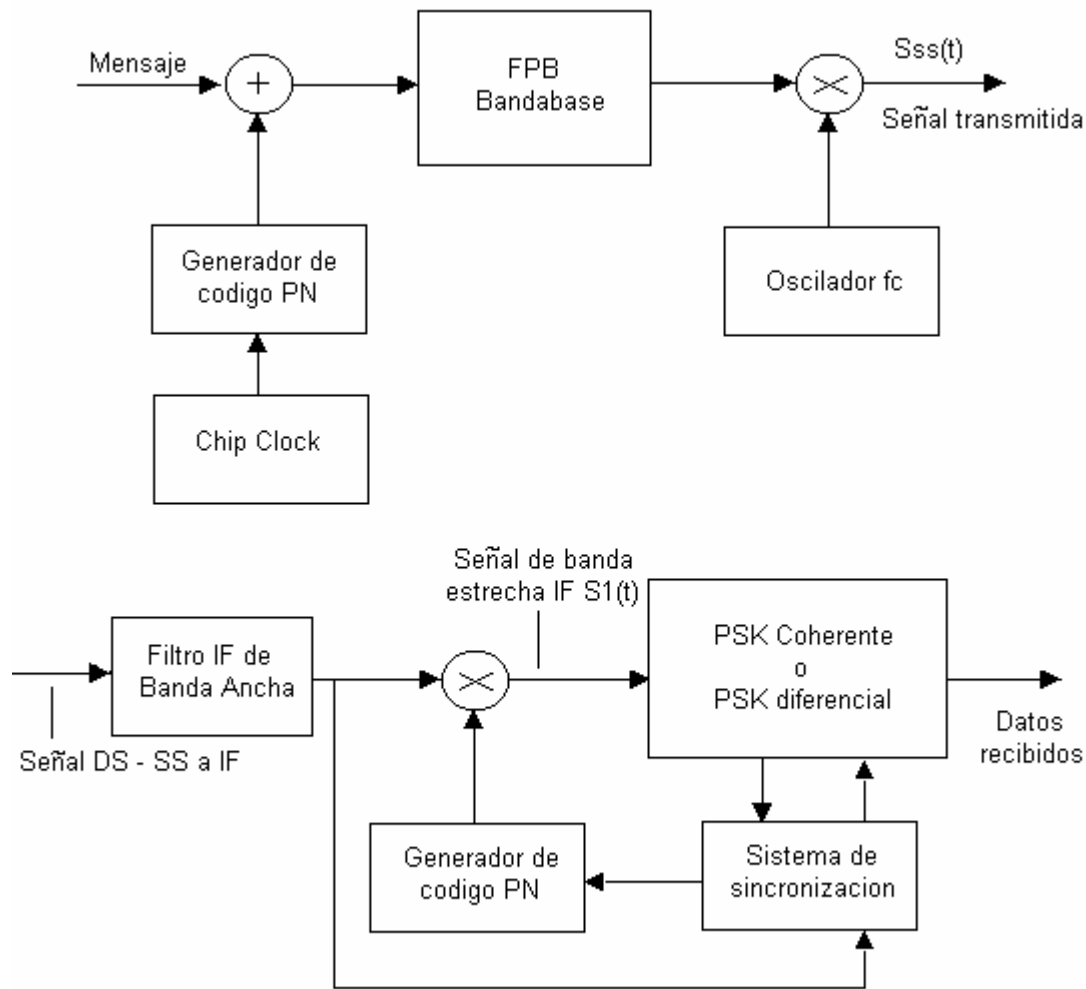


Figura 4.0: Diagrama de un sistema DS - SS

Algunas características del sistema CDMA se describen a continuación:

- Muchos usuarios comparten la misma frecuencia.
- El incremento en el número de usuarios, aumenta el piso de ruido de una manera lineal, luego no hay un número limitado de usuarios, más bien el rendimiento del sistema cae con el número de usuarios.
- El desvanecimiento por camino múltiple es reducido debido a que la señal está esparcida sobre un gran espectro.
- La tasa de datos en el canal es muy alta, con lo cual la duración del chip es muy corta y usualmente mucho menor que el retraso de dispersión del canal.

1.2 Multiplexación y Conmutación

Esta sección será muy breve y estará enfocada a aclarar ciertos aspectos dentro de los sistemas de comunicaciones que son básicos y servirán para entender razonamientos posteriores. Estos dos temas son: Multiplexación y Conmutación.

Estos temas se encuentran estrechamente ligados a los sistemas de telefonía fijos ya que este los utiliza ampliamente.

1.2.1 Multiplexación

Multiplexación intenta enviar información de distintas fuentes hacia varios destinos por el mismo medio (o canal). Existen dos categorías básicas para multiplexación:

- ❑ **FDM (Frequency Division Multiplexing):** el espectro de frecuencia del canal es dividido en bandas de frecuencia, en donde a cada usuario se le es asignada una de ellas. El ejemplo clásico de utilización de este tipo de multiplexación son las bandas de radio AM y FM, en donde a cada estación se le asigna una frecuencia determinada y existe una separación entre estas bandas.

Otra forma de multiplexación en frecuencias sería la técnica de WDM (Wavelength Division Multiplexing) utilizada para canales ópticos. A cada usuario se le asigna una longitud de onda por la cual transmitir información modulada a través de la fibra óptica (recordar la relación entre λ y la frecuencia f)[±]. Algunos de estos temas serán tratados con mayor detalle más adelante.

- ❑ **TDM (Time Division Multiplexing):** cada usuario utiliza un espacio de tiempo determinado para transmitir, es decir toma el ancho completo del canal por un período de tiempo corto. Sólo puede ser usada para datos digitales. Es utilizado ampliamente para las redes de telefónicas fijas y también inalámbricas (TDMA).

El sistema de multiplexación digital utilizado en las redes telefónicas fijas es el sistema de multiprotadora digital T1 (para el caso de Estados Unidos) y E1 para el caso de Europa. T1 consta de 24 canales de banda de voz más dos canales de control, lo cual suman un ancho de banda de 1,544 [Mbps]. E1 son 30 canales de voz más 2 de control, totalizando un ancho de banda de 2,048 [Mbps].

1.2.2 Conmutación

La conmutación siempre es asociada al sistema de telefonía pública, y es el mejor ejemplo para entender la idea que existe detrás. Cuando se necesita establecer una comunicación y se disca un número, la central debe en base a este número lograr una conexión permanente entre el transmisor y receptor de la llamada. Para ello lo que se necesita es conectar por medio de pares de cobre los dos extremos de la comunicación estableciendo un camino, lo cual se hace intercambiando conexiones (antiguamente físicas y más recientemente lógicas), logrando así la comunicación directa entre ambos por medio del establecimiento del circuito. Si el circuito se corta en algún punto, la comunicación se pierde y es necesario volver a establecer la llamada. El proceso por el cual se intercambian

[±] $\lambda=c/f$ con $c = 300000$ m/s (velocidad de la luz)

distintos circuitos hasta lograra establecer la llamada se conoce como conmutación (particularmente de circuitos). En esta sección se tratará la conmutación tanto de circuitos como la de paquetes, describiendo las características de cada uno de los enfoques.

Conmutación de circuitos: Como se describió arriba, es la forma usual conocida en que los sistemas de comunicaciones telefónicas fijas establecen una llamada. Son básicamente caminos físicos (conexiones de pares de cobre) entre los dos extremos, gracias a la ayuda de dispositivos de conmutación, los cuales se encargan de establecer el camino (en donde pueden existir enlaces que no necesariamente sean pares de cobre, como microondas o fibra óptica, por las cuales fluye multiplexada información desde distintos orígenes a distintos destinos). Una vez constituido el camino, este es exclusivo para la comunicación que se efectúa, lo que hace limitado el número de comunicaciones posibles de establecer, eventualmente bloqueando las llamadas que entran cuando el sistema se encuentra copado. La comunicación que fluye por este canal se efectúa en tiempo real (sólo existiendo el retardo propio de propagación de la onda electromagnética por el medio).

Luego algunas características de este tipo de conmutación son: canal dedicado, transmisión en tiempo real, probabilidad no nula de bloqueo, transmisión de ancho de banda fija, no existe almacenamiento de información en la red.

Conmutación de paquetes: Otro enfoque para la conmutación sería permitir que exista almacenamiento, por corto periodos de tiempo, de información en la red lo cual permitiría el reenvío de la información hacia el destino específico. Para ello se dividen los datos a enviar en la fuente en segmentos de tamaño reducido conformando frames de datos los cuales serán enviados a través de la red. Estos paquetes pueden tomar trayectorias distintas entre uno y otro, por lo cual estos pueden llegar en desorden a su destino. Por ende no existe una reserva implícita de ancho de banda entre los dos extremos de comunicación, lo que hace que esta conmutación sea más flexible, pero a la vez menos confiable. Los equipos encargados de realizar esta conmutación son los ya conocidos enrutadores.

Algunas características son: no existe ruta establecida, permite almacenamiento de información en la red, uso dinámico de ancho de banda, es posible detectar errores y realizar conversiones de velocidad y código.

El ejemplo más emblemático de la utilización de este tipo de conmutación son las redes basadas en el protocolo IP (Internet).

1.3 Medios de Transmisión

Hasta el momento se han mencionado las formas en que las señales de información viajan por el canal de comunicación, al igual que la forma en que las redes manejan este flujo, pero aún no se ha hecho mención a una de las partes esenciales dentro del canal: la forma física que adopta este. Esta sección intenta presentar algunas de las características más relevantes de los medios físicos de transmisión utilizados en la actualidad.

1.3.1 UTP (Unshielded Twisted Pair)

El par trenzado es el medio más antiguo de establecer una comunicación entre dos puntos en tiempo real, y básicamente consiste de un par de alambres independientemente aislados y trenzados entre si, para evitar la interferencia (efectos de inducción) con pares adyacentes, debido a que dos alambres en paralelo se comportan como una antena.

En la Figura 4.1 se muestra un cable UTP (Par trenzado sin blindaje).

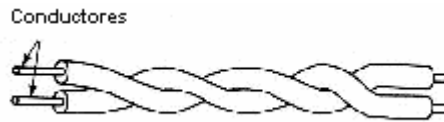


Figura 4.1: Par trenzado

La gran ventaja de este medio es su bajo costo v/s su rendimiento, pudiendo alcanzar varios Mbps durante algunos kilómetros.

Los cables que son utilizados para transportar datos a alta velocidad (10 – 100 Mbps) consiste de cuatro pares agrupados regularmente en una cubierta plástica (o teflón para reducir la diafonía) para su protección. Son normalmente utilizados en redes Ethernet con arquitectura de estrella y en las redes telefónicas privadas.

1.3.2 Cable Coaxial

El gran problema del par trenzado es que a frecuencias relativamente altas su sensibilidad a interferencias y pérdidas por radiación aumentan en forma considerable. Debido a esto lo que se utiliza para estos casos es el cable coaxial, el cual esta constituido por un conductor central rodeado por algún material aislante, como Pirex o poliestireno, el cual a su vez también se encuentra rodeado por otro conductor (trenzado) de tipo cilíndrico. Todo el conjunto se encuentra forrado por una envoltura plástica que sirve de protección. En la Figura 4.2 se aprecia un cable coaxial.

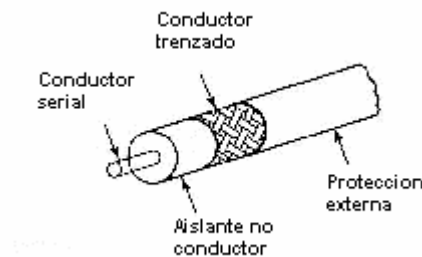


Figura 4.2: Cable coaxial

Estos cables tienen mejor inmunidad a interferencias (mayor blindaje), lo cual conlleva a un mejoramiento en la capacidad, llegando a transportar datos a velocidades de 1 a 2 Gbps en conductores de 1 Km.

Ser dividen en coaxial de banda base y banda ancha, siendo este último empleado para el transporte de señales de televisión. (TV cable)

1.3.3 Fibra Óptica

Básicamente son canales cilíndricos muy pequeños contruidos de vidrio o plástico (o combinaciones) por los cuales es posible hacer incidir un haz de luz que se propagará por él hasta de un extremo a otro.

La base de esto sistemas se encuentra en la teoría óptica y en la naturaleza electromagnética de la luz.

Es un hecho físico que cuando un rayo atraviesa desde un material de cierta densidad a otro de distinta se producen refracciones en mayor o menor grado dependiendo de la densidad relativa entre los dos medios (las cuales se reflejan en sus índices de refracción respectivos). Si el ángulo de incidencia del rayo es mayor que un cierto umbral crítico, el rayo rebotará en forma integra, permaneciendo en el medio. Este es el principio básico de funcionamiento de la fibra, la cual esta compuesta de un núcleo, por el cual viaja la luz, y un manto con índice de refracción ligeramente menor, en el cual rebotarán los rayos contenidos en el núcleo. En la Figura 4.3 se muestra un rayo incidiendo sobre la interfaz aire/silicio en un ángulo α_1 y emergiendo en un ángulo β_1 . La cantidad de refracción depende de las propiedades de los dos materiales (índices de refracción). Para ángulos de incidencia sobre un ángulo crítico, la luz es refractada al interior del silicio, sin escapar al aire. Este rayo incidente sobre un ángulo crítico es atrapado en el interior de la fibra y así puede propagarse por muchos kilómetros virtualmente sin pérdidas.

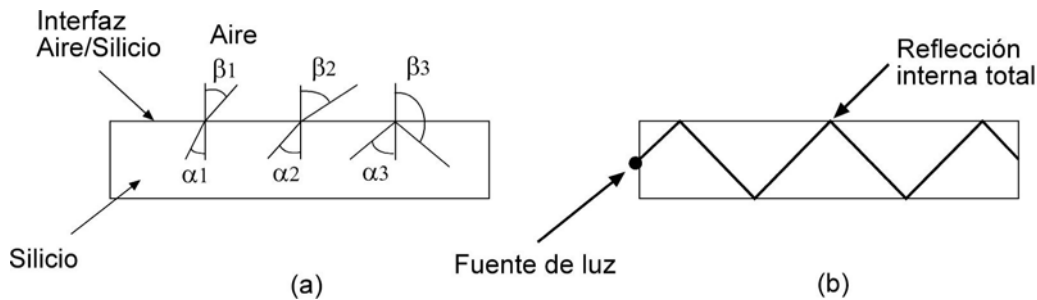


Figura 4.3: (a) Tres ejemplos de rayos incidiendo sobre la interfaz aire/silicio en ángulos distintos; (b) Luz atrapada por reflexión interna total

Las fibras ópticas están hechas de vidrio muy transparente. La atenuación a través de este depende de la longitud de onda de la luz incidente. La atenuación para el tipo de vidrio utilizado en fibras se muestra en la Figura 4.4 en decibels por kilómetro lineal de fibra. Luego, la atenuación esta dada por la formula:

$$\text{Atenuación en decibels} = 10 \log \frac{\text{Potencia transmitida}}{\text{Potencia recibida}} \text{ [dB]} \quad (4.13)$$

La figura 4.4 muestra tres bandas, las cuales son utilizadas en comunicaciones. Están centradas en 850 [nm], 1300 [nm] y 1550 [nm] respectivamente. Las dos últimas presentan menor atenuación que la primera que la primera banda, pero esta presenta la ventaja que tanto lasers como diodo emisores pueden ser contruidos del mismo material.

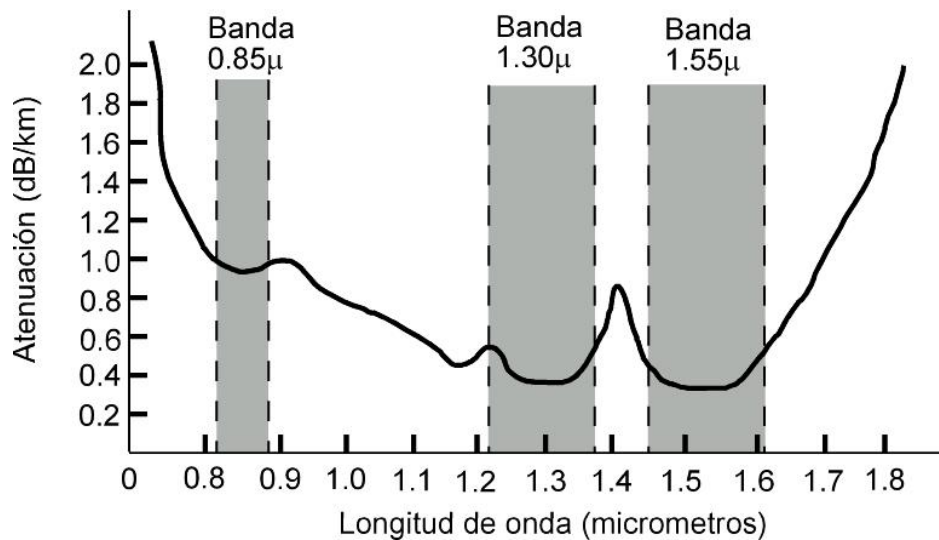


Figura 4.4: Atenuación de la luz a través de la fibra en la región infrarroja

La Figura 4.5 muestra una fibra óptica tipo

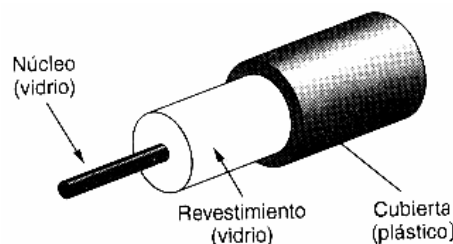


Figura 4.5: Fibra Óptica

Dado que pueden existir varios rayos rebotando con distintos ángulos de incidencia sobre la frontera, se puede hablar de fibras de múltiples modos (Multimodo).

También es posible reducir el tamaño del núcleo con tal que sólo sea posible la propagación de un único modo, y a este tipo de fibra se le denomina fibra Monomodo. Estas fibras son más caras debido al proceso de fabricación y la diferencia de tamaño con las multimodo (Multimodo: $\approx 125 \mu\text{m}$; Monomodo: $\approx 50 \mu\text{m}$). Alcanzan velocidades de datos de varios Gbps a una distancia de 30 Km..

Luego un sistema óptico se encuentra compuesto por una fuente de emisión de luz, la fibra óptica y el receptor de luz. La longitud de onda utilizada en la fuente dependerá de las características físicas de ésta y también (en mayor medida) de las características de la fibra óptica. La información es representada en dos estados (binario), siendo por ejemplo un 0 lógico la ausencia de pulsos de luz, y un 1 lógico la presencia de un pulso de luz. Para tal emisión por lo general se utilizan Diodos especialmente fabricados o LASER[±]. En el lado del receptor se suelen utilizar diodos sensibles a la luz (Fotodiodos).

[±] Light Amplification by Stimulated Emission of Radiation

1.4 Propagación de ondas

La utilización del espacio libre como medio de transmisión es ampliamente utilizado, debido principalmente a las limitaciones físicas que imponen los otros medios (como fibra óptica y cables de cobre) y a las condiciones geográficas existentes entre transmisor y receptor (espacios de agua muy grandes, montañas o terrenos desérticos). Es así como la utilización de este medio ha permitido el desarrollo de la telefonía móvil, enlaces microondas y comunicación satelital.

La propagación en el espacio libre de las ondas electromagnéticas, a menudo se llama *propagación de radiofrecuencia* (RF) o simplemente *propagación de radio*.

Para el análisis de propagación de ondas electromagnéticas en el espacio libre es ampliamente utilizado el concepto de rayos y frentes de onda. Los rayos son líneas dibujadas en el sentido de propagación y sirven para demostrar la dirección relativa de la onda. Sin embargo, un rayo no necesariamente representa la propagación de una sola onda electromagnética. Aquí se utiliza el concepto de frente de onda, el cual representa una superficie de fase constante de una onda.

La figura 4.6 muestra un frente de onda con una superficie que es perpendicular a la dirección de propagación (rectángulo ABCD). Cuando la superficie es plana, su frente de onda es perpendicular a la dirección de propagación.

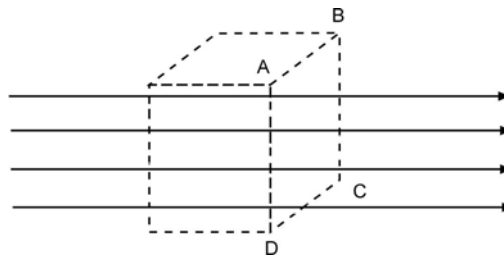


Figura 4.6: Onda Plana

En la figura 4.7 se muestra una fuente puntual, propagando varios rayos a partir de él, y el frente de onda correspondiente. Como se muestra en la figura, el frente de onda generado es una esfera de radio R y con centro en el punto de radiación. A este tipo de fuente, que propaga igualmente sus rayos en todas direcciones se conoce como radiador *isotrópico*.

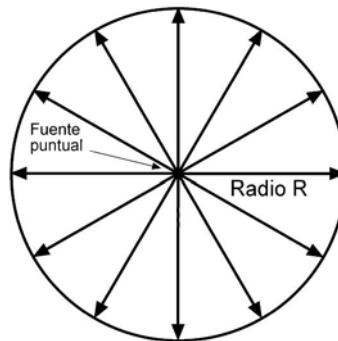


Figura 4.7: Frente de onda, desde una fuente puntual

1.4.1 Densidad de potencia e intensidad de campo

La proporción en la cual la energía cruza por una superficie dada, en el espacio libre, se llama *densidad de potencia*. La *intensidad de campo* es la intensidad de los campos eléctricos y magnéticos de una electromagnética propagándose en el espacio libre. Así la densidad de potencia es:

$$P = EH \quad (4.14)$$

en donde:

P	=	Densidad de potencia (W/m ²)
E	=	Intensidad de campo eléctrico en rms (V/m)
H	=	Intensidad del campo magnético en rms (At/m)

1.4.2 Impedancia característica del espacio libre

Las intensidades de los campos eléctricos y magnéticos de una onda electromagnética, en el espacio libre, se relacionan por la impedancia característica (resistencia) del espacio libre. La impedancia característica del espacio libre es:

$$Z_s = \sqrt{\frac{\mu_0}{\epsilon_0}} \quad (4.15)$$

en donde:

Z _s	=	Impedancia característica del espacio libre (ohms)
μ ₀	=	Permeabilidad magnética del espacio libre (1,26 × 10 ⁻⁶ H/m)
ε ₀	=	Permitividad eléctrica del espacio libre (8,85 × 10 ⁻¹² F/m)

Así, sustituyendo tendríamos que $Z_s = 377 \Omega$.

Por lo tanto, utilizando la ley de Ohm, obtenemos:

$$P = E^2/377 = 377 H^2 W/m^2 \quad (4.16)$$

$$H = E/377 \text{ At/m} \quad (4.17)$$

1.4.3 Frente de onda esférica

Como mencionamos con anterioridad una fuente isotrópica radia potencia a una proporción uniformemente constante, en todas direcciones. Un verdadero radiador isotrópico no existe, pero se aproxima bastante a una antena omnidireccional. Un radiador isotrópico produce un frente de onda esférica de radio R. En cualquier instante de tiempo, la potencia total radiada, P_r watts, esta uniformemente distribuida sobre la superficie de la esfera. Por lo tanto, la densidad de potencia en cualquier punto de la esfera es la potencia total radiada dividida por el área total de la esfera. Esto es

$$P_a = \frac{P_r}{4\pi R^2} \quad (4.18)$$

$$\begin{aligned}
 P_r &= \text{Potencia total radiada (watts)} \\
 R &= \text{Radio de la esfera (el cual es igual a la distancia desde cualquier punto en la superficie de la esfera a la fuente)} \\
 4\pi R^2 &= \text{Área de la esfera}
 \end{aligned}$$

en donde:

Igualando (x.3) con (x.5) se obtiene:

$$E = \frac{\sqrt{30P_r}}{R} \quad (4.19)$$

De la ecuación (x.5) se puede observar que entre más lejos se mueva el frente de onda de la fuente, más pequeña es la densidad de potencia. La relación se llama *ley inversa cuadrática*.

1.4.4 Atenuación

La ley inversa cuadrática describe la reducción de potencia con la distancia de la fuente. Conforme se aleja un frente de onda de la fuente, el campo electromagnético continuo, que es radiado desde esa fuente, se dispersa. Nada de potencia radiada se pierde o se disipa, porque el frente de onda se aleja de la fuente., es decir la onda se disipará sobre un área más grande, disminuyendo la densidad de potencia. Esta reducción de densidad de potencia es equivalente a la pérdida de potencia y comúnmente se llama *atenuación de onda*. La atenuación de la onda se expresa generalmente en términos del logaritmo común de la relación de densidad de potencia (pérdida en dB).

$$\gamma_a = 10 \log \frac{P_1}{P_2} \quad (4.20)$$

1.4.5 Absorción

La reducción en la densidad de potencia debida a la propagación del espacio no libre se llama *absorción*. Esta se produce a que en la atmósfera existen átomos y moléculas, algunos de los cuales son capaces de absorber las ondas electromagnéticas, transfiriéndose la energía desde las ondas a las partículas de la atmósfera. La absorción de onda por la atmósfera es análoga a una pérdida de potencia $I^2 R$.

La absorción de radiofrecuencias en una atmósfera normal depende de la frecuencia y es relativamente insignificante, debajo de aproximadamente los 10 GHz. La Figura 4.8 muestra la absorción en decibels por kilómetro, debido al oxígeno y al vapor de agua para radiofrecuencias superiores a 10 GHz. La atenuación debido a la absorción no depende de la distancia de la fuente radiante, pero si bastante de la distancia total que la onda propaga

por la atmósfera. Además, las condiciones atmosféricas anormales tales como lluvias fuertes o neblina densa absorben más energía que una atmósfera normal.

La absorción atmosférica para una propagación de onda de R_1 a R_2 es $\gamma(R_2 - R_1)$, en donde γ es el coeficiente de absorción.

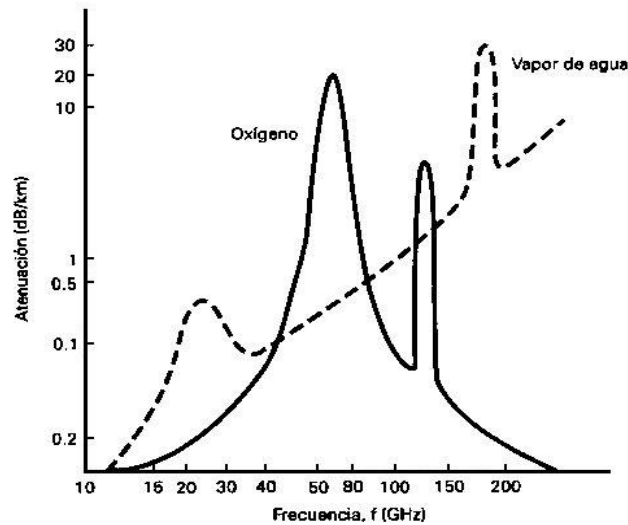


Figura 4.8: Absorción atmosférica de las ondas electromagnéticas

1.4.6 Propiedades ópticas de las ondas de radio

Las ondas que se propagan por el espacio pueden verse afectadas por efectos ópticos como la *refracción*, *reflexión*, *difracción* e *interferencia*. Tiene sentido hablar de propiedades ópticas para propagación de ondas de radio, debido a que las ondas de luz son ondas electromagnéticas de alta frecuencia. A continuación se presentan estos efectos desde una perspectiva de *trazos de rayos geométricos*.

Refracción

La refracción electromagnética es el cambio de dirección de un rayo conforme pasa oblicuamente, de un medio a otro, con diferentes velocidades de propagación. Debido a que la velocidad de propagación es inversamente proporcional a la densidad del medio en que se propaga, la refracción ocurre siempre que una onda pasa de un medio a otro medio de diferente densidad. La Figura 4.9 muestra la refracción de una onda plana en la frontera de dos medios de distinta densidad.

Siempre que un rayo pasa de un medio menos denso a uno más denso, efectivamente se dobla hacia la normal. El ángulo de incidencia es el ángulo formado entre la onda incidente y la normal, y el ángulo de refracción es el ángulo formado entre la onda refractada y la normal.

La cantidad de inclinación depende del índice de refracción de los materiales, el cual es:

$$n = \frac{c}{v} \quad (4.21)$$

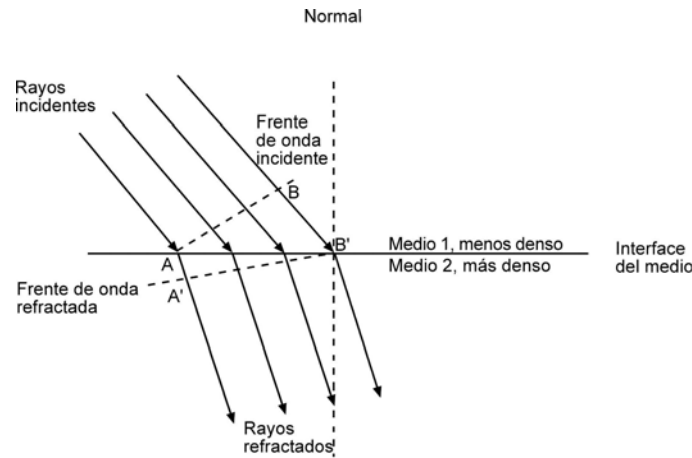


Figura 4.9: Refracción en una frontera plana entre dos medios

n = Índice de refracción (sin unidades)

en donde: c = Velocidad de la luz en el espacio libre (3×10^8 m/s)
 v = Velocidad de la luz en un material dado (m/s)

El índice de refracción es una función de la frecuencia, sin embargo en la mayoría de las aplicaciones es insignificante, aunque como se verá más adelante en este documento, resulta muy útil en aplicaciones con fibra óptica. Cómo reacciona una onda electromagnética, cuando conoce a la interface de dos materiales de transmisión que tienen índices de refracción diferentes, puede explicarse con la *ley de Snell*.

$$n_1 \sen \theta_1 = n_2 \sen \theta_2 \quad (4.22)$$

n_1 = Índice de refracción del material 1

en donde: n_2 = Índice de refracción del material 2
 θ_1 = Ángulo de incidencia (grados)
 θ_2 = Ángulo de refracción (grados)

y debido a que el índice de refracción de un material es igual a la raíz cuadrada de su constante dieléctrica:

$$\frac{\sen \theta_1}{\sen \theta_2} = \sqrt{\frac{\epsilon_{r1}}{\epsilon_{r2}}} \quad (4.23)$$

donde los ϵ_r representan las constantes dieléctricas de los medios correspondientes.

Reflexión

La reflexión electromagnética ocurre cuando una onda incidente choca con una barrera de dos medios y algo, o toda la potencia incidente no entra al segundo material. Las ondas que no penetran al segundo medio se reflejan. La Figura 4.10 muestra la reflexión de una onda electromagnética en una barrera plana entre dos medios.

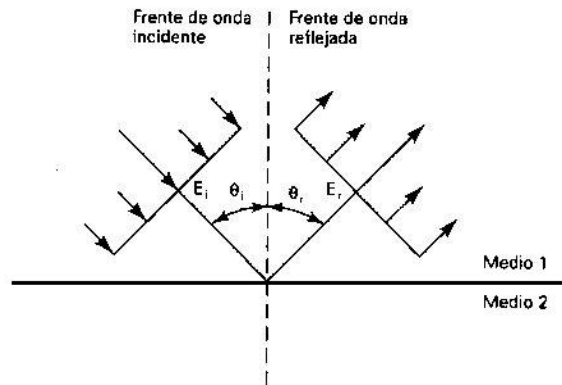


Figura 4.10: Reflexión electromagnética en una frontera plana

Debido a que todas las ondas reflejadas permanecen en el medio 1, las velocidades de las ondas reflejadas e incidentes son iguales, y por lo tanto el ángulo de reflexión es igual al ángulo de incidencia ($\theta_i = \theta_r$). Sin embargo, la intensidad del campo de voltaje reflejado es menor que la intensidad del campo de voltaje incidente. La relación de las intensidades de voltaje reflejado a incidente se llama *coeficiente de reflexión*, Γ . Para un conductor perfecto $\Gamma = 1$. Γ se usa para indicar la amplitud relativa de los campos incidentes y reflejados y, además, el desplazamiento de la fase en el punto de reflexión. Luego, el coeficiente de reflexión es:

$$\Gamma = \frac{E_r e^{j\theta_r}}{E_i e^{j\theta_i}} = \frac{E_r}{E_i} e^{j(\theta_r - \theta_i)} \quad (4.24)$$

Γ = Coeficiente de reflexión (sin unidades)

En donde:

- E_i = Intensidad de voltaje incidente (volts)
- E_r = Intensidad de voltaje reflejado (volts)
- θ_i = Fase incidente (grados)
- θ_r = Fase reflejada (grados)

La porción de voltaje incidente total que no está reflejada se llama *coeficiente de transmisión de potencia* (T). Para un conductor perfecto $T = 0$. La ley de conservación de energía establece que:

$$T + |\Gamma|^2 = 1 \quad (4.25)$$

Si el medio 2 no es un conductor perfecto, algunas de las ondas incidentes lo penetran y se absorben. Las ondas absorbidas establecen corrientes en la resistencia del material y la energía se convierte en calor. La fracción de potencia que penetra al medio 2 se llama el *coeficiente de absorción*.

Difracción

La difracción se define como la modulación o redistribución de energía, dentro de un frente de onda, cuando pasa cerca del extremo de un objeto opaco.

Cuando un frente de onda pasa cerca de un obstáculo o discontinuidad de dimensiones comparables en tamaño a una longitud de onda, no se puede utilizar el simple análisis geométrico para explicar los resultados y el *principio de Huygens* es necesario.

El principio de Huygens indica que cada punto de un frente de onda esférica determinado se puede considerar como una fuente secundaria de puntos de ondas electromagnéticas, desde donde se irradian hacia fuera otras ondas secundarias. Esto se ilustra en la Figura 4.11.

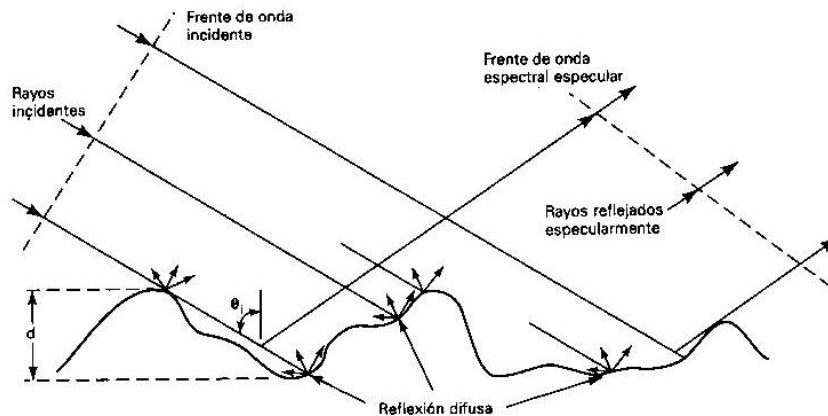


Figura 4.11: La reflexión de una superficie semiáspira

Cada fuente puntual secundaria irradia energía hacia afuera en todas direcciones. Sin embargo el frente de onda continúa en su dirección original porque la cancelación de ondas secundarias ocurre en todas direcciones excepto de frente.

Cuando se considera un frente de onda plano y finito, la cancelación en direcciones aleatorias es incompleta. En consecuencia, el frente de onda se extiende hacia fuera o se *dispersa*. El efecto de dispersión se llama difracción.

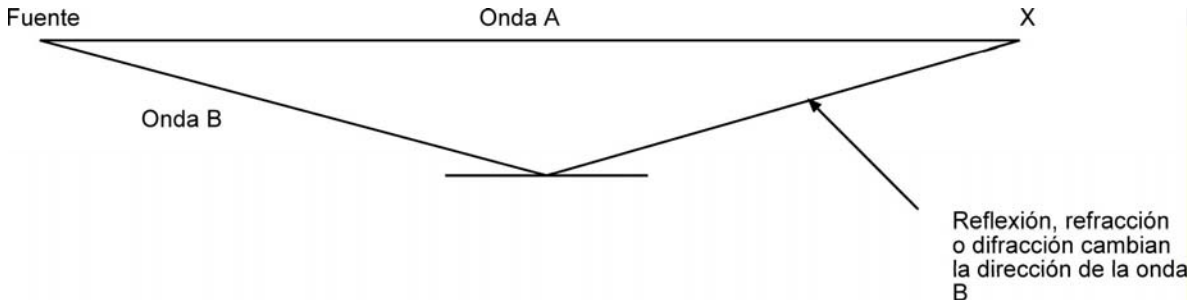
Interferencia

La interferencia de ondas de radio ocurre cuando dos o más ondas electromagnéticas se combinan de tal forma que el funcionamiento del sistema se degrada. La interferencia se sujeta al principio de superposición lineal de ondas electromagnéticas y ocurre cada vez que dos o más ondas ocupan el mismo punto en el espacio. El principio de superposición lineal indica que la intensidad total de voltaje, en un punto determinado del espacio, es la suma de los vectores de ondas individuales.

El voltaje total resultante de la suma no es simplemente la suma de las magnitudes de los dos vectores, sino la suma de las fases de los dos. Luego, dependiendo de los ángulos de fase de los dos vectores puede ocurrir suma o resta.

La Figura 4.12 muestra la interferencia entre dos ondas que viajan por el espacio libre. Puede verse que en el punto X las dos ondas ocupan la misma área de espacio, sin embargo la onda B ha viajado una trayectoria diferente a la onda A y, por lo tanto, puede que sus

ángulos de fase relativos sean distintos. Si la diferencia en la distancia viajada es un múltiplo de un número entero impar de una mitad de longitud de onda, se lleva a cabo el refuerzo. Si la diferencia es un múltiplo de un número entero par de una mitad de longitud de onda, ocurre la cancelación total. Para frecuencias por debajo de VHF, las longitudes de onda relativamente grandes evitan que la interferencia se convierta en un problema



importante.

Figura 4.12: Interferencia de onda electromagnética

1.4.7 Propagación de ondas

Hay tres formas de propagación de ondas electromagnéticas: ondas de tierra, ondas espaciales (que incluyen ondas directas y ondas reflejadas a tierra), y propagación de onda del cielo. La Figura 4.13 muestra los modos normales de propagación entre dos antenas.

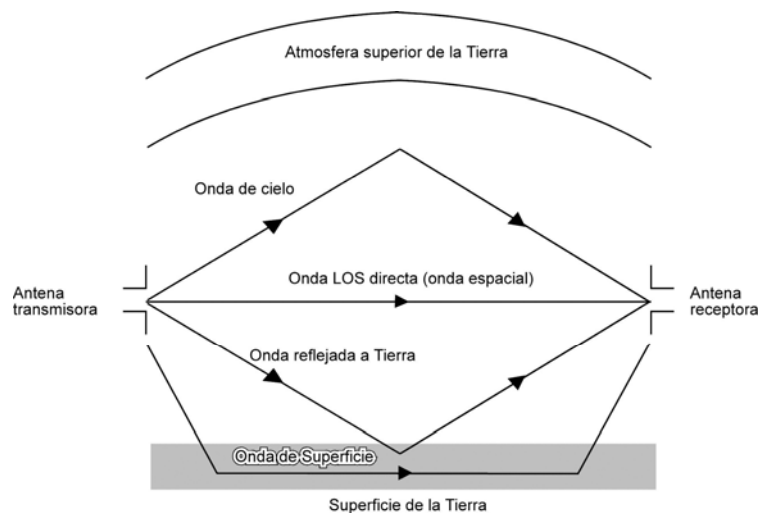


Figura 4.13: Modos normales de la propagación de ondas

Propagación de ondas de tierra

Las ondas viajan por la superficie de la tierra, por lo que suelen llamarse ondas superficiales. Estas ondas deben estar polarizadas verticalmente, pues el campo eléctrico, en una onda polarizada horizontalmente, estaría paralelo a la superficie de la Tierra, y dichas ondas harían cortocircuito por la conductividad de la Tierra. Con las ondas de tierra, el campo eléctrico variante induce voltajes en la superficie de la Tierra, que causa que fluyan corrientes que son muy similares a las de una línea de transmisión. Las pérdidas de

onda de tierra se incrementan rápidamente con la frecuencia, por lo tanto, se limita generalmente a frecuencias por debajo de los 2 MHz.

La propagación de ondas de tierra se utiliza comúnmente para comunicaciones de barco a barco y de barco a tierra (debido a las propiedades conductoras del agua salada), para radio navegación, y para las comunicaciones marítimas móviles.

Las desventajas de la propagación de ondas de tierra son:

1. Requieren de una potencia relativamente alta para la transmisión.
2. Están limitadas a frecuencias muy bajas, bajas y medias (VLF, LF, MF) que requieren de antenas grandes.
3. Las pérdidas por tierra varían considerablemente con el material de la superficie.

Y sus ventajas son:

1. Con suficiente potencia de transmisión, se pueden utilizar para comunicarse entre dos ubicaciones cualesquiera en el mundo.
2. No se ven relativamente afectadas por los cambios en las condiciones atmosféricas.

Propagación de ondas espaciales

Las ondas espaciales viajan por la parte inferior de la atmósfera de la Tierra. Incluyen ondas directas y reflejadas de tierra, como se muestran en la Figura 4.14. Las primeras viajan en línea recta entre las antenas transmisora y receptora. La propagación de ondas espaciales directas se llama comúnmente transmisión de línea de vista (LOS).

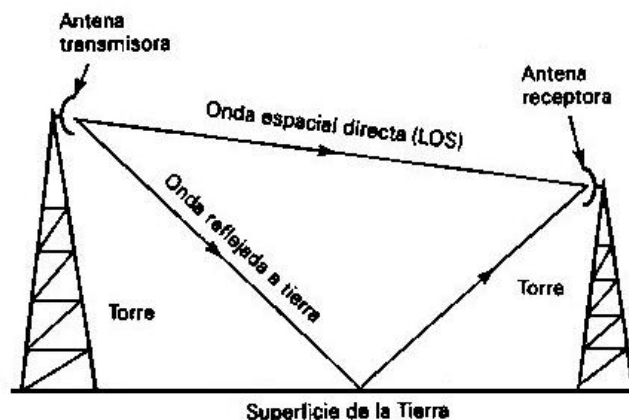


Figura 4.14: Propagación de ondas espaciales

La intensidad de campo en la antena receptora depende de la distancia que hay entre las dos antenas (atenuación y absorción) y si las ondas, directa y reflejada a tierra, están en fase.

La curvatura de la Tierra presenta un horizonte para la propagación de ondas espaciales llamado *radio horizonte*. Debido a la refracción atmosférica, el radio horizonte se extiende más allá del horizonte óptico para la atmósfera estándar común. El radio horizonte es aproximadamente cuatro tercios del horizonte óptico. La troposfera causa la refracción, debido a cambios en su densidad, temperatura, contenido de agua vapor, y conductividad relativa. El radio horizonte puede alejarse simplemente elevando las antenas por arriba de la superficie de la Tierra, con torres o colocándolas arriba de montañas o edificios altos.

La Figura 4.15 muestra el efecto del radio horizonte. El radio horizonte de línea de vista para una sola antena se da como:

$$d = \sqrt{2h} \quad (4.26)$$

d = Distancia a radio horizonte (millas)

en donde:

h = Altura de la antena sobre el nivel del mar (pies)

Por tanto, para una antena transmisora y una receptora, la distancia entre las dos antenas es:

$$d = \sqrt{2h_t} + \sqrt{2h_r} \quad (4.27)$$

d = Distancia total (millas)

en donde: h_t = Altura de la antena transmisora (pies)

h_r = Altura antena receptora (pies)

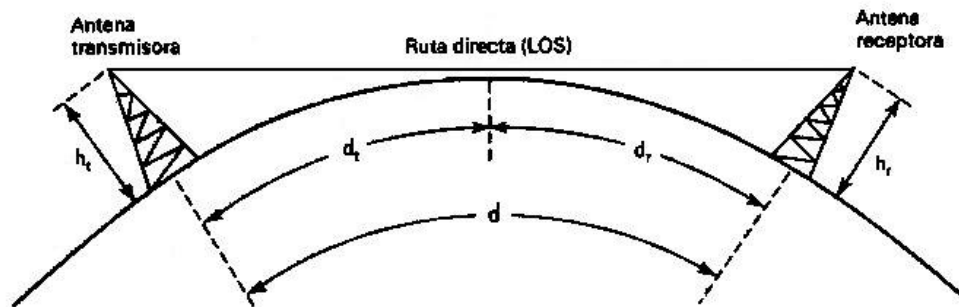


Figura 4.15: Ondas espaciales y radio horizonte

Propagación de ondas de cielo

Las ondas electromagnéticas que se dirigen por encima del nivel del horizonte se llaman ondas de cielo. Las ondas se envían hacia el cielo, donde son reflejadas o refractadas nuevamente a la Tierra por la ionosfera. La ionosfera es la región de espacio localizada aproximadamente de 50 a 400 Km arriba de la superficie de la Tierra y es la porción más alta de la atmósfera, por lo tanto absorbe grandes cantidades de energía radiante del sol, que ioniza las moléculas de aire, creando electrones libres. Cuando una onda de radio pasa a través de la ionosfera, el campo eléctrico de la onda ejerce una fuerza en los electrones libres, haciéndolos vibrar. Esto reduce la corriente, lo que equivale a reducir la constante dieléctrica. A su vez este efecto provócale incremento de la velocidad de propagación y hace que las ondas se doblen alejándose de las regiones de alta densidad de electrones (es decir se incrementa la refracción). Entre más alta la densidad de iones, mayor la refracción. Además, debido a que la composición de la atmósfera no es uniforme y a las variaciones en la temperatura y densidad, se encuentra estratificada, en tres capas: las capas D, E y F. En la Figura 4.16 puede verse que las tres capas varían en ubicación y en densidad de ionización con la hora del día.

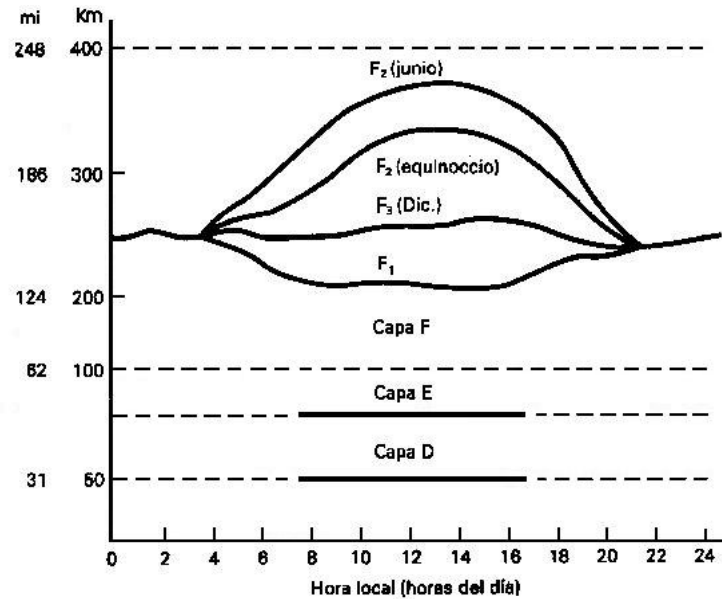


Figura 4.16: Capas ionosféricas

Capa D: es la capa inferior de la ionosfera y se localiza entre 50 a 100 Kilómetros arriba de la superficie de la Tierra. Debido a que es la capa más lejana al sol, existe poca ionización, sin embargo los iones de esta capa pueden absorber cantidades considerables de energía electromagnética. La ionización depende de la altitud del sol en el horizonte, así que desaparece en la noche. La capa D refleja ondas VLF y LF y absorbe ondas MF y HF.

Capa E: se localiza entre 100 y 140 Kilómetros sobre la superficie de la Tierra. Esta capa auxilia la propagación de ondas de superficie MF y refleja ondas HF un poco durante el día. Al igual que la capa anterior, prácticamente desaparece durante la noche.

Capa F: esta capa está compuesta por dos subcapas F_1 y F_2 . Durante el día la primera se ubica entre 140 a 250 Kilómetros y la segunda entre 140 a 300 Kilómetros sobre la superficie de la Tierra. Durante la noche ambas subcapas se combinan y forman una sola. La capa F_1 absorbe y atenúa algunas ondas HF, aunque la mayoría de las ondas pasan a través de F_2 , cuando se refractan nuevamente a la Tierra.

1.5 Antenas

Una antena es un conductor metálico capaz de radiar y recibir ondas electromagnéticas. Se usa como interfaz entre el transmisor y el espacio libre o el espacio libre y el receptor (o más específicamente entre la guía de onda y el espacio libre).

En la Figura 4.17 se muestran los patrones de onda estacionarias de voltaje en una línea de transmisión. La línea termina en un circuito abierto, que representa una discontinuidad abrupta en la onda de voltaje incidente en la forma de una inversión de fase. La inversión de la fase resulta cuando parte del voltaje incidente se irradia, en lugar de ser reflejado de

nuevo a la fuente. La energía radiada se propaga lejos de la antena en forma de ondas transversales.

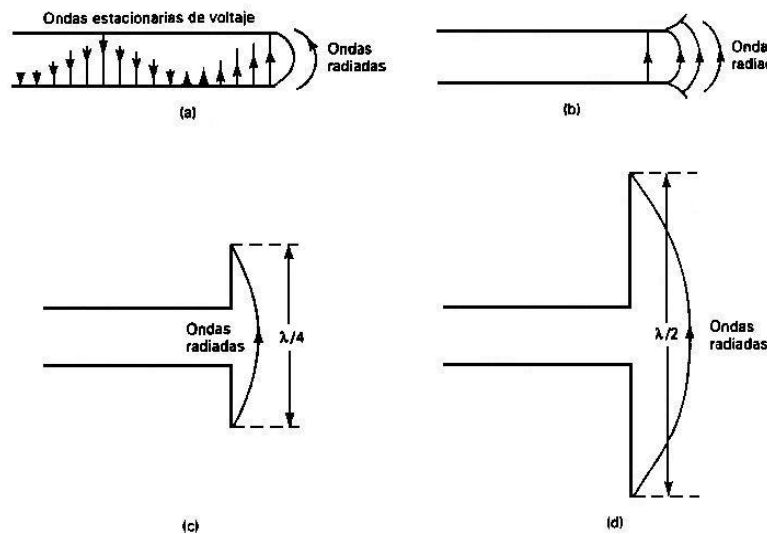


Figura 4.17: (a) Radiación de una línea de transmisión; (b) conductores difusores; (c) antena Marconi; (d) antena Hertz

La eficiencia de radiación es la relación entre la energía radiada y la energía reflejada, y para el caso de una línea de transmisión abierta es en extremo baja. Para radiar más energía es necesario separar más los conductores. Una antena así se llama dipolo y se muestra en la figura anterior.

1.5.1 Patrón de radiación

Un patrón de radiación es un diagrama polar que representa las intensidades de los campos o las densidades de potencia en varias posiciones angulares en relación con una antena. Si el patrón de radiación se traza en relación a la intensidad del campo eléctrico (E) o de la densidad de potencia (P) se llama patrón de radiación absoluto. Si se traza la intensidad de campo o la densidad de potencia en relación al valor en un punto de referencia, se llama patrón de radiación relativo. En la Figura 4.18 se muestra un patrón de radiación absoluto y relativo para una antena no especificada.

En la figura el haz principal se encuentra en una dirección de 90° y se llama *lóbulo principal*. Puede existir más de un lóbulo principal. También hay un haz secundario o lóbulo *menor* en la dirección de -180° . Debido a que el lóbulo principal propaga y recibe la mayor parte de la energía, este lóbulo se llama lóbulo frontal. Los lóbulos adyacentes al frontal se llaman *laterales*, y los que están en dirección opuesta al lóbulo frontal se llaman lóbulos *traseros*. La relación de la potencia del lóbulo frontal con la potencia del lóbulo trasero se llama relación frontal trasera, y la relación del lóbulo frontal con el lateral se llama relación frontal lateral. La línea que divide el lóbulo principal desde el centro de la antena en la dirección de máxima radiación se llama *línea de tiro*.

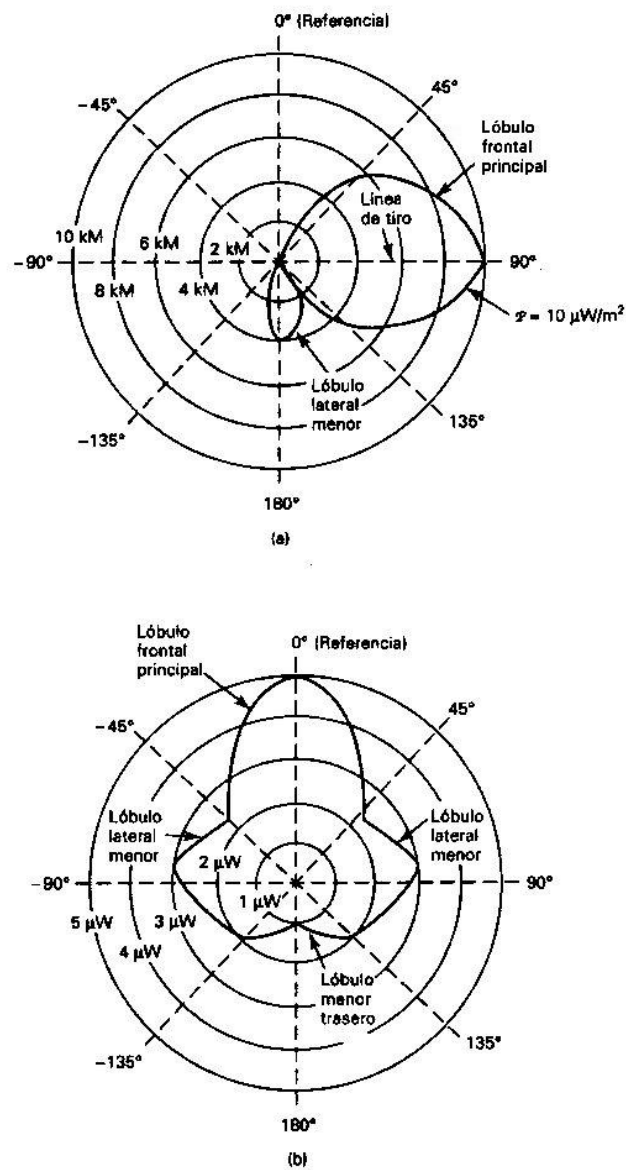


Figura 4.18: (a) Patrón de radiación absoluto; (b) Patrón de radiación relativo

1.5.2 Campos cercanos y lejanos

El término campo cercano se refiere al patrón de campo que esta cerca de la antena, y el término campo lejano se refiere al que esta a gran distancia. La potencia que alcanza el campo lejano continua radiando lejos y nunca regresa (campo de radiación), a diferencia de lo que ocurre en el campo cercano (o de inducción) donde en la segunda mitad del ciclo, la potencia regresa a la fuente. Debido a que importa el campo radiado, los patrones de radiación de la antena se dan para el campo lejano. El campo cercano se define como el

área dentro de una distancia D^2/λ de la antena, en donde λ es la longitud de onda y D el diámetro de la antena en las mismas unidades.

1.5.3 Resistencia de radiación y eficiencia de antena

Parte de la potencia suministrada a la antena se convierte en calor y se disipa. Esto se modela por medio de una *resistencia de radiación* de la antena en ca y es igual a:

$$R_r = \frac{P}{i^2} \quad (4.28)$$

R_r = Resistencia de radiación (ohms)

en donde: P = Potencia radiada por la antena (Watts)
 i = Corriente de la antena en el punto de alimentación (Amperes)

La resistencia de radiación es la resistencia que, si reemplazara la antena, disiparía exactamente la misma cantidad de potencia de la que irradia la antena.

La *eficiencia* de la antena es:

$$\eta = \frac{P_r}{P_r + P_d} \times 100 \quad (4.29)$$

η = Eficiencia de antena (%)

en donde: P_r = Potencia radiada por la antena (Watts)
 P_d = Potencia disipada por la antena (Watts)

En la Figura 4.19 se muestra un circuito equivalente simplificado para una antena. En términos de resistencia y corriente, la eficiencia de la antena es:

$$\eta = \frac{i^2 R_r}{i^2 (R_r + R_e)} = \frac{R_r}{R_r + R_e} \quad (4.30)$$

η = Eficiencia de antena

en donde: R_r = Resistencia de radiación (ohms)
 R_e = Resistencia de la antena efectiva (ohms)
 i = Corriente de antena (amperes)

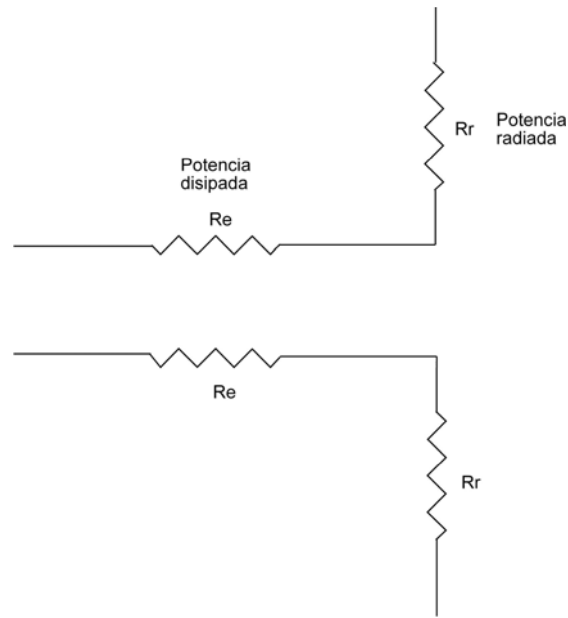


Figura 4.19: Circuito equivalente simplificado de una antena

1.5.4 Ganancia directiva y ganancia de potencia

La ganancia directiva es la relación de la densidad de potencia radiada en una dirección en particular con la densidad de potencia radiada al mismo punto por una antena de referencia, suponiendo que ambas antenas irradian la misma cantidad de potencia. El patrón de densidad de potencia relativa de una antena es realmente un patrón de ganancia directiva si la referencia de la densidad de potencia se toma de una antena de referencia estándar, que por lo general es una antena isotrópica. Luego, la ganancia directiva es:

$$D = \frac{P}{P_r} \quad (4.31)$$

D = Ganancia directiva (sin unidades)

P = Densidad de potencia en algún punto de una antena determinada (W/m^2)

P_{ref} = Densidad de potencia en el mismo punto de una antena de referencia (W/m^2)

donde:

La ganancia de potencia es igual a la ganancia directiva excepto que se utiliza el total de potencia que alimenta la antena (o sea, que se toma en cuenta la eficiencia de la antena). La ganancia de potencia es:

$$A_p = D\eta \quad (4.32)$$

Si una antena no tiene pérdidas, irradia 100% de la potencia de entrada y la ganancia de potencia es igual a la ganancia directiva. En decibels la ganancia de potencia es:

$$A_p = 10 \log \frac{P\eta}{P_{ref}} \quad (4.33)$$

Para una referencia isotrópica, la directividad de un dipolo de media onda es aproximadamente de 1,64 (2,15 dB). Es usual indicar la ganancia de potencia en decibels si se refiere a un dipolo de media onda, sin embargo, si es en referencia a un radiador isotrópico, la cantidad de decibels se indica como dBi. Es importante mencionar que la antena realmente no amplifica la potencia de entrada, sino que solo concentra su potencia radiada en una dirección en particular. Debido a que la antena es un dispositivo recíproco, su patrón de radiación es también su patrón de recepción. Por lo tanto, las antenas receptoras tienen una directividad y una ganancia de potencia exactamente como las antenas transmisoras.

1.5.5 Potencia radiada isotrópica efectiva

La potencia radiada isotrópica efectiva (EIRP) se define como una potencia de transmisión equivalente y se expresa como:

$$\text{EIRP} = P_r A_t \quad \text{Watts} \quad (4.34)$$

Donde P_r es la potencia total radiada y A_t es la ganancia directiva de la antena transmisora. En dBm se tiene:

$$\text{EIRP(dBm)} = 10 \log \frac{P_r}{0,001} + 10 \log A_t \quad (4.35)$$

EIRP o solo ERP es la potencia equivalente que tendría que radiar una antena isotrópica para alcanzar la misma densidad de potencia en la dirección seleccionada en un punto determinado, como otra antena.

Para determinar la densidad de potencia en un punto determinado, se expande la ecuación x.5 para incluir la ganancia de la antena transmisora y se escribe como:

$$P = \frac{P_r A_t}{4\pi R^2} \quad (4.36)$$

Para analizar una antena como receptora, es útil emplear el concepto de área efectiva. Si una antena es puesta en el campo de una onda electromagnética polarizada linealmente, la potencia disponible en los terminales de la antena es igual al área efectiva multiplicada por la potencia por unidad de área llevada por la onda, es decir:

(4.37)

$$C = PA$$

C = Potencia disponible (watts)
 P = Densidad de potencia de la onda (W/m²)

$$A = \text{Área efectiva (m}^2\text{)}$$

en donde:

La definición anterior no considera pérdidas en la antena. El término de densidad de potencia P para la relación de arriba corresponde a la densidad de potencia de la antena transmisora, la cual ya se mostró en la ecuación 4.18.

Una relación muy útil existente entre el área efectiva de una antena y su ganancia esta dada por:

$$A = \frac{\lambda^2 A_r}{4\pi} \quad (4.38)$$

Donde A_r corresponde a la ganancia de la antena receptora. Luego, combinando las ecuaciones 4.36, 4.37 y 4.38 se tiene:

$$C = \frac{P_r A_t A_r \lambda^2}{16\pi^2 R^2} \quad (4.39)$$

Donde C es la potencia disponible en la antena receptora, medida en Watts. De esta última ecuación se obtiene la *razón de transferencia de potencia*, dada por

$$\frac{C}{P_r} = \frac{A_t A_r \lambda^2}{16\pi^2 R^2} \quad (4.40)$$

Esta relación es cierta para campos lejanos, es decir debe existir una distancia mayor o igual a $2D^2/\lambda$.

1.5.6 Polarización de la antena

La polarización de la antena se refiere sólo a la orientación del campo eléctrico radiado desde ésta. Una antena puede polarizarse en forma lineal (polarizada horizontal o verticalmente), en forma elíptica o circular. Así, si la antena irradia una onda electromagnética polarizada verticalmente, la antena se define como polarizada verticalmente.

1.5.7 Ancho de haz de la antena

El ancho de haz de la antena es sólo la separación angular entre los dos puntos de media potencia (- 3 dB) en el lóbulo principal del patrón de radiación del plano de la antena, por lo general tomado en uno de los planos “principales”. En la Figura 4.20 el ancho de haz corresponde al ángulo formado entre los puntos A, X y B. El ancho de haz corresponde al ancho de haz de media potencia.

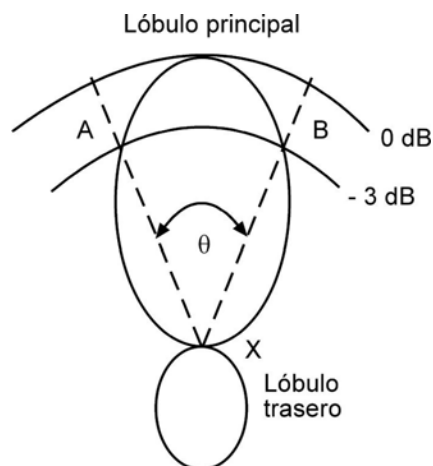


Figura 4.20: Ancho de haz de antena

1.5.8 Antenas básicas

Doblete elemental

Este es un dipolo corto (comparado con la mitad de la longitud de onda) que tiene corriente uniforme en toda su longitud. Sin embargo, la corriente varía en forma sinusoidal en tiempo y en cualquier instante, es decir:

$$i(t) = I \sin(2\pi ft + \theta) \quad (4.41)$$

- $i(t)$ = Corriente instantánea
 I = Amplitud pico de la corriente de RF (amperes)
 F = Frecuencia (hertz)
 t = Tiempo instantáneo (segundos)
 θ = Ángulo de fase (radianes)

en donde:

Su patrón de intensidad de campo eléctrico relativo se muestra en la Figura 4.21.

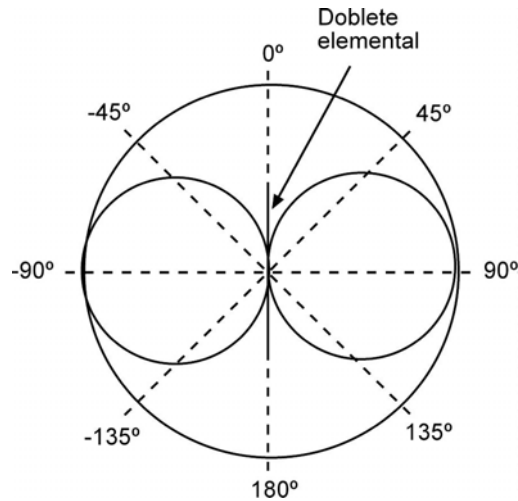


Figura 4.21: Patrón de radiación relativa para un doblete elemental en un plano perpendicular a los dos ejes del dipolo

El patrón de potencia relativa es:

$$P = \frac{30\pi l^2 I^2 \sin^2 \phi}{\lambda^2 R^2} \quad (4.32)$$

Donde ϕ corresponde al ángulo entre el eje de la antena y la dirección de radiación y l la longitud del dipolo.

Dipolo de media onda

Es conocida como *antena de Hertz*. Es una antena resonante, es decir, es múltiplo de un cuarto de longitud de onda de largo y de circuito abierto en el extremo más lejano.

El patrón de radiación de espacio libre para un dipolo de media onda depende de la colocación horizontal o vertical de la antena con relación a la superficie de la Tierra. La Figura 4.22 muestra los distintos patrones posibles para esta antena.

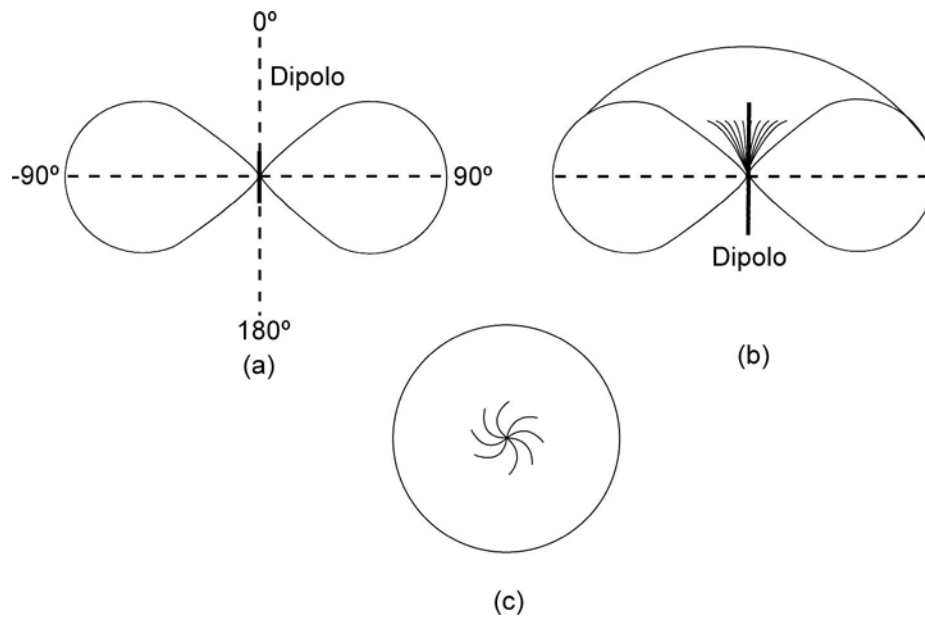


Figura 4.22: (a) vista vertical de un dipolo montado verticalmente; (b) vista de sección transversal; (c) vista horizontal

Antena aterrizada

La antena monopolo de un cuarto de longitud de onda montada verticalmente con el extremo inferior conectado directamente a tierra se llama *antena Marconi*. La corriente máxima ocurre en los extremos aterrizados, que causa un flujo de alta corriente a través de tierra. Para reducir las pérdidas, la tierra deberá ser un buen conductor.

La Figura 4.23 muestra un patrón de radiación para una antena aterrizada de un cuarto de longitud de onda. Puede verse que la mitad inferior de cada lóbulo se cancela con las ondas reflejadas de tierra. Esto por lo general no tiene ninguna consecuencia porque la radiación en dirección horizontal se incrementa, incrementando así la radiación a lo largo de la superficie de la Tierra y mejorando la cobertura de área.



Figura 4.23: Patrones de radiación para una antena aterrizada

5 CAPA DE ENLACE

Cuando se habla de capa de enlace se puede pensar en el canal de comunicación existente entre dos máquinas pertenecientes a una misma red. El canal actúa como un cable por el cual los datos circulan en el mismo orden en que son emitidos. La capa de enlace se encarga de ofrecer una forma confiable de entregar esta información, a través de la capa física, hacia la máquina que se encuentra en el otro extremo. A pesar de lo sencillo que pudiese parecer este enfoque, surgen una serie de problemas que hacen necesario definir servicios bien específicos que permitan mediante interfaces apropiadas la comunicación con las otras capas del modelo OSI. Luego, la capa de enlace debe proveer funciones específicas que controlen la comunicación a este nivel de abstracción. Estas funciones incluyen entregar interfaces de servicios bien definidas, determinar de qué forma los bits que serán entregados a la capa física en forma de *frames* (marcos), manejar los errores de transmisión y regular el flujo para que un transmisor rápido no sature a uno lento.

5.1 Servicios

La capa de enlace puede ser diseñada para ofrecer varios servicios, los cuales dependerán del tipo de sistema y red en las cuales se trabaje. Los servicios más comúnmente provistos son:

- Servicio no confiable y no orientado a la conexión
- Servicio confiable, pero no orientado a la conexión
- Servicio confiable y orientado a la conexión

El primero es el más básico de ellos y sólo se encarga de entregar los marcos (*frames*) a la capa física sin importar si el receptor está dispuesto a recibirlos y sin esperar confirmación previa antes de seguir enviando. Esto implica que no se toman medidas en la eventualidad que existiesen pérdidas en el medio producidos por ruido en el canal. Para terminar la comunicación sólo se detiene el envío de *frames*. Este tipo de comunicación es utilizado cuando la probabilidad de error es muy baja, delegándose este control a alguna capa superior. También es útil para tráfico en tiempo real, en donde no es necesaria la confirmación de los datos recibidos (debido al retardo que estas producen).

El segundo de los servicios continúa con la misma filosofía de envío sin esperar saber si el receptor se encuentra en condiciones de aceptar los *frames* (sin conexión), pero implementa acuse de recibo por parte del receptor. Así el transmisor sabe si el *frame* llegó en forma correcta, y si no es así posibilita el reenvío de este por parte del transmisor. Este tipo de comunicación es menos utilizada que la anterior y se implementa para canales poco confiables.

El último tipo de servicio es orientado a conexión y confiable, es decir, antes de que comience el envío de datos, el transmisor establece una conexión con el receptor. Esto permite que los datos sean transferidos por el canal establecido, a la vez que se asegura que los *frames* enviados lleguen en el orden correcto y en su totalidad. Una vez concluida la

comunicación, la conexión establecida desaparece liberándose los recursos que esta utilizaba en ambos extremos.

5.2 Enmarcado

La capa física se encarga de enviar y recibir flujos de bits por el canal sin importar el tamaño de este, ni tampoco preocuparse de que estos contengan información con error. Luego la labor de la capa de enlace es proveer este tipo de servicio.

Entonces, la capa de enlace divide el flujo de bits en segmentos denominados marcos (o *frames*), a los cuales se les agrega un campo que permita a la capa paralela en la otra máquina detectar posibles errores en los datos. Este campo generalmente es conocido como chequeo de suma (*Checksum*) pues los algoritmos que utiliza para su calculo se basan en sumas de los datos contenidos en el *frame*. El método más utilizado para este efecto es el CRC (*Cyclic Redundancy Code*).

Una consideración que surge al dividir el flujo de bits en *frames*, es la necesidad de delimitarlo, pudiéndose reconocer en forma clara un principio y un fin de marco, para que pueda ser recibido y chequeado en el receptor. Algunas de las formas para solucionar este problema son: contador de caracteres, caracteres de comienzo y término con carácter de relleno, banderas de comienzo y fin con bit de relleno y violaciones de código de capa física.

El último método es interesante de comentar debido a que es utilizado para las redes Ethernet IEEE 802.3. El uso de violación de código es posible siempre y cuando la forma de los bits que se utilizan incluya transiciones de estado. Por ejemplo, si para representar un 1 lógico se utiliza una transición desde nivel alto a bajo (5 a 0 Volt) y para el cero una transición de nivel bajo a alto (0 a 5 Volt) entonces en el momento que no exista señal, el canal sólo tendrá valores de 0 Volt, los cuales no representan ningún estado válido en particular, produciéndose una violación al código. Cuando se detecte transición se entenderá que puede ser el inicio de un nuevo *frame* (El código utilizado es el Manchester).

5.3 Manejo de errores

Para poder realizar un debido manejo de errores es necesario poseer un método que permita detectarlos. Luego de esta detección se pueden utilizar dos diferentes enfoques: corrección del o los errores y petición de reenvío de la información.

Como ya se mencionó con anterioridad, para la detección de errores se incluye un campo dentro del *frame* el cual contiene cierto tipo de información obtenida mediante algún método que permite en el lado del receptor realizar una serie de cálculos con el fin de comprobar la integridad del *frame* recibido. Generalmente son utilizados métodos de detección de paridad, efectuados por códigos polinomiales (como CRC). Dependiendo del número de detecciones requeridas será el tamaño del campo a incluir en el *frame*.

Una vez detectado errores, una posibilidad sería intentar corregirlos. El problema que surge con este enfoque es que para efectuar de manera eficiente una corrección es necesario enviar mayor información redundante en el *frame*. Los métodos utilizados generalmente se basan en el calculo de distancias y matrices de corrección. (por ejemplo

distancia de Hamming). Mientras mayor capacidad de detección y de corrección, mayor es el número de información a enviar, lo cual hace a estos métodos poco viables en ciertos casos.

El otro enfoque es sólo pedir retransmisión del frame equivocado, o mejor aún enviar acuses de recibos positivos o negativos dependiendo de la integridad de los datos recibidos. A pesar de parecer simple, existen una serie de consideraciones a tener en cuenta en esta implementación. Si el frame es enviado y por alguna razón desaparece sin nunca llegar a destino, el transmisor quedara esperando la confirmación. Para evitar este problema se utilizan temporizadores, los cuales al espirar provocan la retransmisión automática del último frame que no recibió confirmación. Otro problema surge cuando la confirmación desaparece, pudiendo llegar muchas veces el mismo frame debido a retransmisión o que los frames no lleguen en orden secuencial. Para evitar estos problemas se soporta un número máximo de repeticiones, dando por terminada la comunicación si no se recibe acuse, a la vez que los frames son enumerados para que sean posibles de ordenar en el receptor.

5.4 Control de flujo

El control de flujo trata de dar solución a la situación en que un transmisor de mayor velocidad satura los buffer de recepción del receptor de menor velocidad. Esto se puede dar debido a diferencias en las máquinas (distintas tecnologías y capacidades), lo cual se desea evitar. Para ello se implementan distintas políticas de control de flujo, siendo una de las más utilizadas el de confirmación de envío. Es decir el receptor autoriza enviar una cantidad determinada de frames, pero una vez cumplida la cuota debe esperar hasta una nueva confirmación con la cantidad de marcos a enviar. Este problema es tratado en otras capas del modelo, siendo de particular interés en la capa de transporte para el modelo TCP/IP. A continuación se presenta un ejemplo de protocolo de capa de enlace, conocido como HDLC.

5.4.1 HDLC

IBM desarrollo a mediados de la década de los 70's el protocolo de Control de Enlace de Datos Síncrono (SDLC) para ser usado en ambientes de redes SNA (Systems Network Architecture). La particularidad de este protocolo es que fue el primer protocolo de capa de enlace de datos basado en operaciones orientadas a bit y síncronas. De aquí derivaron un número importantes de protocolos los cuales diferían en detalles de implementación con SDLC, en particular se menciona HDLC (High-level Data Link Control), el cual fue implementado para redes X.25.

HDLC soporta una serie de tipos de enlace y topologías. Puede ser utilizada con enlaces punto a punto, multipunto, medios de transmisión half y full-duplex y redes de conmutación de paquetes y de circuitos. Aquí se identifican dos tipos de nodos: *primarios* y *secundarios*. Los nodos primarios controlan la operación de las otras estaciones (secundarias). Los nodos primarios preguntan en un determinado instante, en el cual las estaciones secundarias pueden transmitir si tienen datos. Los nodos primarios y secundarios pueden conectarse de 3 distintas formas:

- ❑ Punto a punto (sólo un primario con un secundario)
- ❑ Multipunto (un primario con varios secundarios)
- ❑ En lazo: los nodos forman un bucle donde el primario se encuentra en primera posición y le siguen los secundarios, los cuales se traspasan entre si la información proveniente del primario.

El formato del *frame* se muestra en la Figura 5.1.

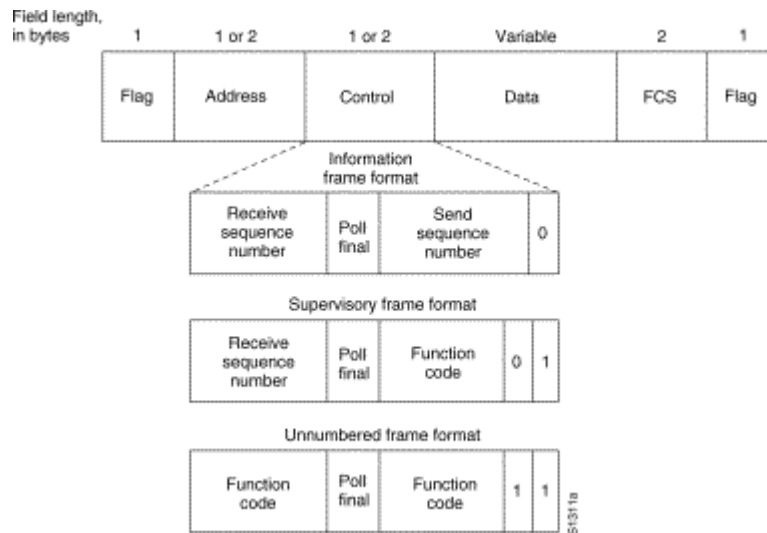


Figura 5.1: Formato de frame SDLC

Se distinguen tres tipos de *frame*, los cuales varían en el campo de control: Información, Supervisión, y no numerado.

Algunos de los campos son:

Flag: Delimitadores del Frame. (01111110)

Address: Identificador de terminal, utilizada para cuando existen múltiples terminales. Para comunicación punto a punto es a veces utilizada para distinguir comandos de respuestas.

Control: Se distinguen dependiendo del tipo de *frame*. Se usa para incluir números de secuenciamiento, y mensajes de control desde los nodos primarios a secundarios y viceversa.

Datos: Campo que contiene la información.

FCS: Frame Check Sequence. Es un código CRC para detección de error.

El protocolo HDLC también utiliza el método de la ventana deslizante explicado a seguir.

5.4.2 Ventana deslizante

La esencia de todos los protocolos de ventana deslizante es que, en cualquier instante, el transmisor mantiene un grupo de números en secuencia que corresponde a los *frames* que tienen permiso para ser enviados. Se dice que estos marcos caen dentro de la ventana transmisora. De manera semejante, el receptor mantiene una ventana receptora

correspondiente al grupo de marcos que tiene permitido aceptar. En general se mantiene una ventana cuyo ancho es igual al tiempo de transmisión de un cuadro más el de transmisión del acuse de recibo. Por ejemplo, suponga que en un enlace satelital transmitiendo a 50 kbps el intervalo de tiempo entre la transmisión de un cuadro y la recepción del acuse de recibo correspondiente es de 500 mseg. Considere también que cada cuadro contiene 1000 bits. Dado este escenario, se pueden analizar dos posibilidades: a) transmitir un cuadro y esperar acuse de recibo antes de enviar el próximo; y, b) enviar todos los cuadros dentro de una ventana de 500 mseg correspondiente a 25 cuadros. ¿Qué ocurre si un cuadro se daña y no es recibido? Se puede utilizar una de estas dos técnicas: Regresa-n; o, Repetición Selectiva. En Regresa-n, el receptor descarta todos los cuadros subsecuentes. En Repetición Selectiva solo se envía el *frame* con problema.

5.5 Subcapa de Acceso al Medio

En el capítulo 2 se trató el tema de topologías de redes, en donde se presentan las redes de difusión (*broadcast*) las cuales tienen la característica de utilizar un solo medio físico como canal de comunicación entre varios usuarios conectados. Como se puede vislumbrar, estas redes presentan un problema particular para la implementación de los servicios descritos anteriormente, debido a su naturaleza compartida. El problema fue resuelto con la agregación de una Subcapa dentro de la capa de enlace, encargada de manejar el acceso de los distintos usuarios al medio compartido la cual es conocida como Subcapa MAC (*Media Access Control*). Algunos enfoques que plantean una solución de implementación se mencionan y describen a continuación:

5.5.1 ALOHA

Fue creado por Norman Abramson y sus colegas de la Universidad de Hawai, con la intención de ser utilizado para transmisiones de radio terrestres. La idea es aplicable a cualquier sistema que utilice un medio de transmisión compartido. Se distinguen dos enfoques:

ALOHA puro: La idea básica es que el que necesite transmitir datos por el medio, que lo haga. Si hay dos usuarios utilizando el medio en el mismo instante se provocará una colisión de *frames*, los cuales serán destruidos. Gracias a que es un medio compartido existe realimentación de colisiones, es decir en algún momento el *frame* destruido llegará hasta el emisor el cual entenderá que debe retransmitirlo. La retransmisión no se hace de forma inmediata, debido a que si así fuese, se volvería a provocar la colisión entre los mismo *frames*. Entonces se hace esperar un tiempo aleatorio al transmisor antes de poner un nuevo cuadro en el medio, reduciendo así la probabilidad de colisiones.

Con el fin de determinar la eficiencia de un canal ALOHA, se presenta el siguiente análisis[±]: Se considera un conjunto infinito de usuarios interactivos en la red. Cada usuario

[±] Extraído desde “Redes de Computadoras”, Andrew S. Tanenbaum, Tercera Edición Prentice – Hall Hispanoamérica, pags. 248 -249

pone un *frame* en el canal y lo verifica para comprobar si llegó con éxito. Si es así el usuario ve la respuesta y continúa enviando. Si no, el cuadro se retransmite una y otra vez hasta que ocurre un éxito. Se entiende por tiempo de *frame* el tiempo necesario para transmitir un cuadro de longitud fija. La población de usuarios genera *frames* nuevos según una distribución de Poisson con una media de N *frames* por tiempo de *frame*. Para un rendimiento razonable se espera $0 < N < 1$.

Las estaciones también generarán retransmisiones de *frames* debido a colisiones. Supóngase, además que la distribución de probabilidad de k intentos de transmisión por tiempo de *frame*, viejos y nuevos combinados, también es Poisson, con una media de G por tiempo de *frame*. ($G \geq N$). El rendimiento S correspondiente a los paquetes transmitidos con éxito es la carga generada, G , veces la probabilidad de que una transmisión tenga éxito (es decir, $S = G \cdot P_0$, donde P_0 es la probabilidad de que un marco no sufra una colisión).

Un cuadro no colisionará si no hay transmisión durante un intervalo de tiempo de *frame* desde su inicio hasta su fin, como se muestra en la Figura 5.2. Sea t el tiempo requerido para enviar un marco. Si cualquier otro usuario generó un *frame* entre el tiempo t_0 y $t_0 + t$, el final de ese *frame* chocará con el comienzo del cuadro sombreado. De manera parecida, cualquier otro *frame* que salga entre $t_0 + t$ y $t_0 + 2t$ chocará con el final del *frame* sombreado.

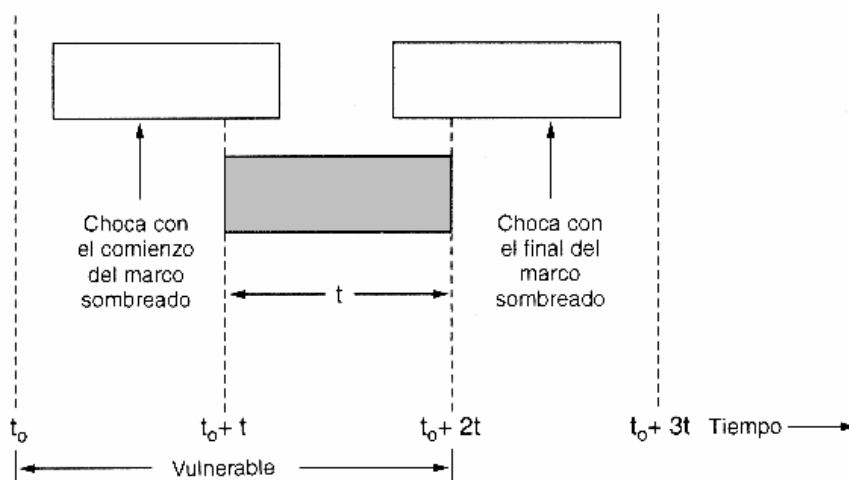


Figura 5.2: Período vulnerable para el frame

La probabilidad de que k frames sean generados durante un tiempo de *frame* está dado por la distribución de Poisson:

$$P_r[k] = \frac{G^k e^{-G}}{k!} \quad (5.1)$$

En una ranura de dos tiempos de *frame* de longitud, el número medio de cuadros es de $2G$. La probabilidad de que no se inicie otro tráfico durante todo el período vulnerable está dada entonces por $P_o = e^{-2G}$. Si $S = GP_o$ se obtiene:

$$S = Ge^{-2G} \quad (5.2)$$

El rendimiento máximo ocurre a $G=0.5$, con $S=1/2e$, que es aproximadamente 18,4%.

ALOHA ranurado: Se propone dividir el tiempo en intervalos discretos, cada uno del tamaño de un frame, siendo todo sincronizado por medio de un reloj generado por un terminal. En comparación con el anterior, ahora se debe esperar por el comienzo de una nueva ranura de tiempo. La probabilidad de que un *frame* no sufra una colisión es ahora igual a $P_o = e^{-G}$ con lo que se tiene:

$$S = Ge^{-G} \quad (5.3)$$

Luego el máximo uso de canal aumenta a un $\sim 37\%$.

La Figura 5.3 muestra la eficiencia de transmisión para los dos tipos de ALOHA.

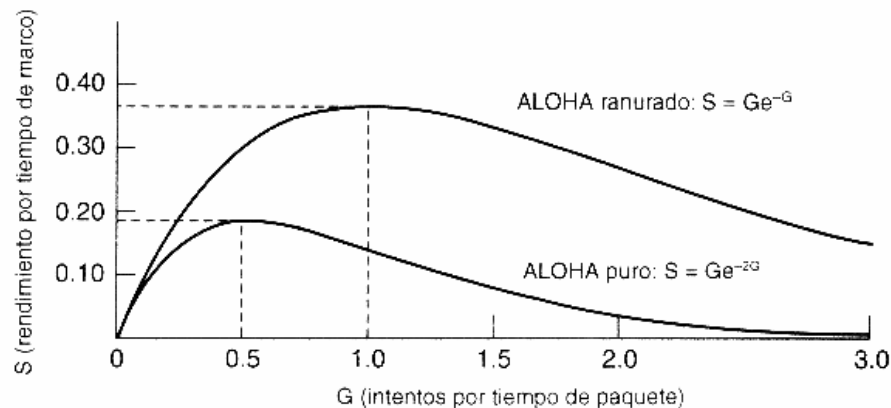


Figura 5.3: Rendimiento de los sistemas ALOHA

5.5.2 CSMA (Carrier Sense Multiple Access)

Este protocolo se encuentra basado en el anterior, pero a diferencia de este, la idea detrás de CSMA es escuchar el medio, y si este se encuentra disponible, transmitir. Para ello se utilizan tres algoritmos distintos:

CSMA persistente-1: Si al momento de escuchar el canal se encuentra ocupado, el transmisor espera hasta que se desocupe. Si logra transmitir y ocurre una colisión, se espera un tiempo aleatorio y se vuelve al inicio. Al momento de transmitir no considera el tiempo que demora la señal en propagarse por el canal de un extremo a otro.

CSMA no persistente: Antes de enviar se escucha el canal, y si esta desocupado pone los datos en el. Si esta ocupado no se queda esperando que el canal se desocupe, sino que vuelve a censar el medio después de un tiempo aleatorio.

CSMA persistente-p: utiliza canales ranurados. Escucha el canal y si este esta desocupado transmite con una probabilidad p . Con probabilidad $q = 1 - p$ no transmite y espera la siguiente ranura de tiempo.

CSMA/CD (CSMA con detección de colisión): si al momento de transmitirse un *frame* por el medio se detecta una colisión, inmediatamente se termina la transmisión, cortando el marco afectado. Esto mejora el ancho de banda al ahorrar el tiempo empleado en la propagación por el medio de todo el cuadro defectuoso. Este protocolo es ampliamente utilizado por redes LAN Ethernet (IEEE802.3).

6 CAPA DE RED

La capa de red es la encargada de hacer llegar los paquete desde un punto de emisión a otro de recepción. Para lograr este objetivo debe existir un conocimiento de los caminos que debe seguir el mensaje, a través de distintas redes y, generalmente, de varios enrutadores hasta llegar a su destino. Para ello se vale de direcciones lógicas las cuales servirán para establecer políticas de ruteo dentro de la subred de comunicaciones, y en las cuales se basarán los protocolos. En este capítulo se describen los tipos de servicios brindados por esta red, los algoritmos que hacen posible el establecimiento de rutas y por último la descripción del protocolo utilizado en Internet (IP).

6.1 Consideraciones de diseño

Los servicios proporcionados por la capa de red están principalmente orientados a la capa superior (transporte), la cual solo se preocupa de entregar información a enviar, sin preocuparse de las características de la sub-red de comunicación. Esto quiere decir que al momento de definir los servicios a especificar para la capa de transporte estos deben ser independientes de la tecnología de red utilizada, pues la capa superior no considera (y no tiene por que saber) en que red se encuentra.

Otro punto importante a considerar es que al momento que la capa de transporte desee enviar datos no considera si al host al que va dirigido el mensaje utiliza la misma tecnología de red. Luego, la capa de red, debe ser capaz, mediante un sistema de direccionamiento universal, de entregar el paquete con la información sin preocuparse de este aspecto.

Además del direccionamiento, existe una serie de factores que se deben tener en cuenta al momento de transportar información entre redes de tecnología distintas. Algunos de ellos son: tamaño de paquete que no exceda el máximo permitido en la red (Ej: en Ethernet el largo máximo se fija en 1500 bytes, incluyendo el encabezado), resolución entre direcciones lógicas (de red) y direcciones físicas (de enlace), pérdida de paquetes por saturación en redes lentas, etc.

En cuanto a los servicios proporcionados por esta capa, ya sabemos que uno de los principales es hacer llegar los paquetes entregados por la capa de transporte a su destino. La discusión surge cuando se plantea de que forma lo hará y que tipo de servicio de entrega proporcionará. Para este último punto existen dos enfoques fundamentales:

- ☐ Confiable orientada a la conexión
- ☐ No confiable sin conexión.

Existen más combinaciones, pero estás son las más utilizadas. La elección surgirá dependiendo del tipo de red en que se trabaje y las aplicaciones que se desean implementar sobre estas. Hay veces en que es preferible esperar un tiempo más largo pero que

asegurando una entrega íntegra y correcta (como transmisión de archivos) y otras en que la velocidad es lo importante (voz e imágenes en tiempo real). Ejemplo de estos dos enfoques se dan en redes ATM para el primer caso y en redes IP para el segundo. En este capítulo solo se tratará el segundo de los casos. Un detalle importante de mencionar es que ATM efectivamente es orientada a la conexión, pues se establecen circuitos virtuales entre extremos de comunicación, pero no utiliza métodos de acuse de recibo, lo que no impide que se pierda una celda sin que el transmisor se percate. Si implementa suma de verificación de error, por lo cual se puede estrictamente decir que es un protocolo semiconfiable a este nivel.

6.2 Protocolos de ruteo

Para llevar un paquete de un extremo a otro, es necesario implementar ruteo, es decir, mediante algún criterio, escoger la mejor ruta de salida que permite alcanzar el destino a un paquete entrante a un enrutador. Este ruteo ocurre a nivel de la sub-red de comunicación. Los algoritmos de enrutamiento se clasifican en:

- *Algoritmos no adaptables (estáticos)*: las rutas se calculan por adelantado, no considerando mediciones de tráfico o cambios de topología.
- *Algoritmos adaptables (dinámicos)*: basan su decisión en mediciones de topología y, en algunos casos, en las condiciones tráfico de la red.

Ahora se presentarán la descripción de algunos de estos algoritmos, partiendo por los estáticos.

Enrutamiento por trayectoria más corta: Se basa en la utilización de grafos, donde cada nodo representa un enrutador los cuales son unidos mediante arcos. Para establecer una salida, primero se establecen los caminos más cortos entre pares de nodos. Para medir esta ruta más corta es necesario especificar la métrica a utilizar, la cual podrá variar dependiendo del criterio de diseño. Algunas de ellas son: distancia geográfica, retardos medios, costo de utilización de enlace, ancho de banda, etc. Uno de los algoritmos utilizados para calcular las rutas más cortas es Dijkstra (1959).

Inundación: La idea detrás de este algoritmo es sencilla. El paquete que llega se replica por todas las salidas del enrutador, excepto por la cual arribó. Claramente el problema que surge con este algoritmo es la duplicación del paquete. Además se inunda la red de forma innecesaria contribuyendo a la congestión de nodos con problemas. Una forma de evitar esto es siendo más selectivo, enviando los paquetes por los caminos en donde es más probables que se halle el destino.

Algunos de los algoritmos dinámicos se describen a continuación:

Vector Distancia: es un algoritmo dinámico, en el cual los enrutadores comparten información cada cierto intervalo de tiempo con sus vecinos. Esta información se almacena en forma de tablas (vectores), las cuales contienen la mejor distancia conocida hacia cada destino y por cual de sus interfaces debe enviarse. Una implementación de este algoritmo es conocida bajo el nombre de RIP (*Routing Information Protocol*), la cual utiliza como métrica la enumeración de saltos. Cada router envía su información a intervalos regulares de tiempo o cuando hay un cambio en la topología, y sólo cambia su tabla si llega información con una mejor ruta (menor número de saltos), manteniendo siempre sólo las más cortas.

Ruteo de Estado de Enlace: este algoritmo nace por la necesidad de reemplazar a RIP (mediados de 1980), el cual quedaba cada vez más obsoleto al trabajar con redes heterogéneas, debido a que no incluía consideraciones de ancho de banda y su tiempo de convergencia era pobre. La idea detrás de este algoritmo es lograr obtener información sobre vecinos que se encuentren bajo la misma jerarquía, donde la mayor entidad dentro de esta es conocida como sistema autónomo (se entiende por sistema autónomo una colección de redes bajo un administrador común y que utiliza una estrategia de ruteo común). El algoritmo se basa en identificar a todos los enrutadores adyacentes mediante algún tipo de mensaje especial, el cual también ayuda con posterioridad a identificar si el terminal aún se encuentra funcionando. Luego de la identificación, se mide el costo de llegar a cada terminal adyacente con algún criterio (el cual puede ser el retraso). Con esta información se arma un paquete especial el cual se envía (por inundación) hacia los otros enrutadores que conforman el sistema autónomo. Con esta información es posible conformar la topología completa del sistema en relación a los enrutadores que la componen. Debido a que son componentes de un mismo sistema todos los routers poseen la misma información. Una vez conformada la topología cada router calcula mediante algún algoritmo, como el de Dijkstra, las menores distancias hacia los distintos nodos que componen la red con las cuales configura su tabla de ruteo.

Como se puede deducir de los anterior, a diferencia de RIP que sólo tiene una visión de saltos hacia vecinos, este protocolo conoce la topología completa de la red, y en esta se basa para conformar su tabla de ruteo. Un protocolo ampliamente utilizado y que hace uso de este algoritmo es OSPF (*Open Shortest Path First*).

6.3 Protocolo IP de la capa de red del modelo TCP/IP

Como ya se mencionó en el capítulo 2, el Protocolo de Internet (IP) forma parte del conjunto o stack de protocolos especificados para la arquitectura TCP/IP. De echo es uno de los protocolos más utilizado en la actualidad para la transmisión de paquetes desde una host a otro ubicados en redes distantes. IP es el protocolo que especifica la capa de red para el modelo TCP/IP. Este protocolo está especificado para enviar un paquete de información desde un origen a un destino a través de un sistema de redes interconectadas. Utiliza el criterio de mejor esfuerzo, es decir, no garantiza la llegada del paquete a destino pues no posee mecanismos de control de flujo, ni mecanismos para implementar confiabilidad, ni

secuenciamiento de paquetes. Este protocolo se encarga de implementar dos funciones básicas: direccionamiento y fragmentación.

Las direcciones son utilizadas para identificar a cada uno de los host conectados a la red* y además ayuda al encaminamiento de los paquetes ya que los protocolos de ruteo se basan en las direcciones.

La fragmentación ocurre cuando un paquete de determinado largo debe pasar a través de una red que posee un MTU (Maximun Transmission Unit) menor al largo del paquete. Esta fragmentación es realizada por el enrutador de borde y el paquete no es vuelto a reensamblarse hasta que llega a su destino.

Para un mejor comprensión de estas funciones, la Figura 6.1 muestra el formato del paquete IP.

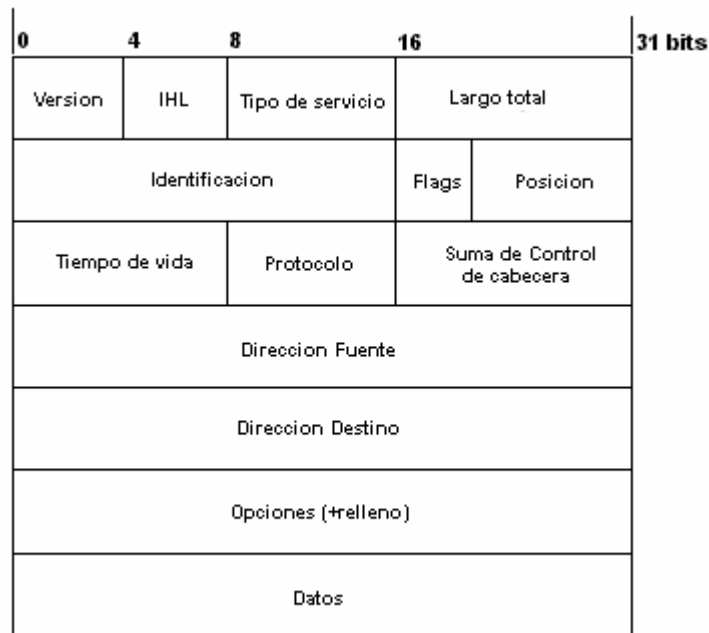


Figura 6.1: Formato paquete IP.

La descripción de cada campo se enuncia a continuación:

- Versión:** Describe el formato de la cabecera Internet (versión 4 para este caso).
- IHL:** Longitud de cabecera de Internet (*Internet Header Length*) en palabras de 32 bits (para un paquete normal es 5).
- Tipo de servicio:** Identificación de los parámetros abstractos de la calidad de servicio. Permite definir prioridad, rendimiento y fiabilidad. No todos los enrutadores toman en cuenta este campo en IPv4.
- Largo Total:** Longitud total de paquete medido en número de bytes incluyendo cabecera y datos (largo máximo 65535).
- Identificación:** Número que identifica el paquete y que es utilizado como ayuda para el armado en el caso de fragmentación.

* En realidad las direcciones especifican puntos de conexión a red, más que hosts en si mismos.

Flags:	Indicadores de control de 3 bits, cada uno de ellos utilizado para: 1 ^{er} bit debe ser cero; 2 ^{do} bit si es cero el paquete puede fragmentarse y si es 1 no debe fragmentarse; 3 ^{er} bit indicador de más fragmentos (en el caso que existiese fragmentación) si es cero indica que es el último fragmento, en caso contrario existen más fragmentos.
Posición:	Campo utilizado junto con los flags para controlar la fragmentación. Indica que ocupa el fragmento dentro del paquete original.
Tiempo de vida:	Campo que indica el tiempo máximo que se le está permitido al paquete permanecer en el sistema. Cada vez que pasa por un enrutador este lo decremента y si llega a cero antes de arribar a su destino es descartado.
Protocolo:	Indica el protocolo de capa de transporte que esta contenido en la parte de datos del paquete.
Suma de control:	Es el complemento a uno de 16 bits de la suma de los complementos a uno de todas las palabras de 16 bits de la cabecera. Como la cabecera cambia al pasar por cada enrutador debe ser recalculada cada vez.
Dirección fuente:	Dirección de origen (32 bits)
Dirección destino:	Dirección de destino (32 bits)
Opciones:	Opciones para compatibilidad futura y de seguridad. Pueden no ser incluidas en el paquete (que es lo general). El relleno se utiliza para asegurar que el paquete sea un múltiplo de 32 bits.
Datos:	Campo en donde se ubican los datos para la capa de red.

A cada host que utiliza el protocolo IP se le es asignada una dirección de 32 bits que lo identifica, la cual esta compuesta de dos partes: una parte que identifica la red y otra parte que identifica el host dentro de esa red. El número de red identifica una red y debe ser asignado por el Centro de Información de Red Internet (InterNIC).

La dirección es dividida en 4 grupos de 8 bits cada uno, los cuales son separados por puntos y representados en formato decimal. La Figura 6.2 ilustra de mejor forma este formato. Para identificar que parte de la dirección pertenece a la red y que parte pertenece al host se dividen las direcciones en clases, cada una con un número determinado de posibles direcciones de red y de hosts asociados. Esta configuración se muestra en la Tabla 6.1.

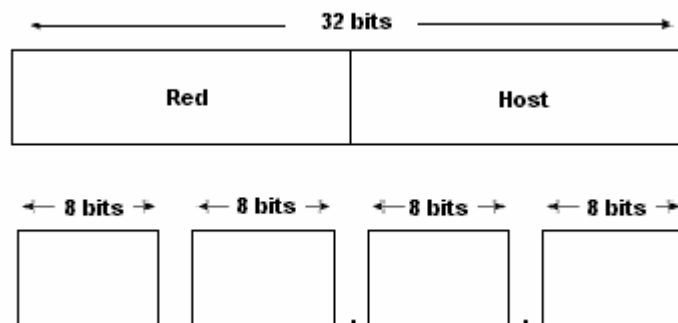


Figura 6.2: Formato de dirección IP.

Tabla 6.1: Clases de direcciones

CLASE	FORMATO	RANGO DE DIRECCIONES
A	8 bits de red, 24 de hosts	1.0.0.0 a 126.0.0.0
B	16 bits de red, 16 de hosts	128.1.0.0 a 191.254.0.0
C	24 bits de red, 8 de hosts	192.0.1.0 a 223.255.254.0
D	Grupos Multicast	224.0.0.0 a 239.255.255.255
E	Experimental	240.0.0.0 a 254.255.255.255

Debido al agotamiento de direcciones de red y a la utilización de subredes en Internet se dio paso a la utilización de máscaras de subred (para mayores detalles ver “Internetworking with TCP/IP” Douglas E. Comer).

6.4 Calidad de Servicio (QoS, Quality of Service)

El concepto de QoS se refiere a “la capacidad de la red de proveer mejor servicio para seleccionar tráfico sobre redes de variadas tecnologías[±]” (Frame Relay, ATM, redes Ethernet, etc), utilizando ruteo IP para todas estas tecnologías subyacentes. Algunos de los factores que influyen en la calidad de servicio son: tiempos de latencia y jitter, ancho de banda dedicado y mejoras en las características de pérdida de la red.

Implementar calidad de servicio no es una tarea fácil, más aún cuando es necesario pasar por varios tipos de redes para llegar al destino deseado. Muchas de las redes no consideran la calidad de servicio como un factor preponderante y sólo realizan el mejor esfuerzo (*best-effort*) al entregar los paquetes que circulan por ella. Otras, en cambio, si la consideran y definen una serie de factores posibles de configurar al establecer la comunicación (como ATM).

Una de las soluciones posibles es establecer un proceso de negociación entre los hosts terminales y los elementos de red. Los parámetros importantes a considerar en esta negociación son:

Prioridad: indica que conexiones son más importantes de mantener.

Rendimiento: cantidad de bytes transferidos por segundo, mediante algún intervalo de tiempo.

Tasa de error de error residual: Cantidad mínima de errores (mensajes perdidos o alterados) como porcentaje del total de datos enviados. (idealmente cero).

Otros factores influyentes son:

Probabilidad de falla al establecer conexión: probabilidad que no se establezca la conexión en un lapso determinado de tiempo.

Retardo de tránsito: tiempo entre mensajes enviados por el usuario hacia el destino y viceversa.

Protección: seguridad contra terceros.

Así se definen tres niveles básicos de calidad de servicio extremo a extremo:

[±] Cisco Product Documentation, www.cisco.com

- ❑ Servicio de mejor esfuerzo: conocido como ausente de QoS, define la conectividad sin garantías.
- ❑ Servicio Diferenciado: conocido como soft QoS, algunos tráficos son tratados de mejor forma que otros. (manejo más rápido, mayor ancho de banda promedio, menor tasa de perdidas). Es una preferencia estadística, no garantizada de forma estricta.
- ❑ Servicio Garantizado: reservación de recursos de red para tráfico específico.

Para el caso del protocolo IP, no existe explícitamente una política de calidad de servicio, pues este trabaja con el concepto de mejor esfuerzo. A pesar de ello existe un campo dentro del encabezado IP que permite sugerir ciertos parámetros, este campo es el de Tipo de Servicio el cual se describe a continuación:

0	1	2	3	4	5	6	7
PRECEDENCIA			D	T	R	0	0

Precedencia: Da indicaciones de cómo debe ser tratado el paquete en cuanto a prioridades se trata. Los valores posibles son:

111 – Control de Redes
 110 – Control Entre Redes
 101 – Critico/ECP
 100 – Muy Urgente (Flash Override)
 011 – Urgente (Flash)
 010 – Inmediato
 001 – Prioridad
 000 – Rutina

D: Demora, 0 Normal, 1 Baja demora.

T: Rendimiento, 0 Normal, 1 Alto rendimiento

R: Fiabilidad, 0 Normal, 1 Alta fiabilidad

Como se mencionó anteriormente no todas las redes reconocen estos parámetros, por lo que suelen conservarse con valores normales.

Muchas de las especificaciones de parámetros de calidad de servicio se dejan como variables de decisión para protocolos de capa superior, o son implementados sobre los elementos que componen la sub-red (como son los enrutadores), permitiendo crear políticas de administración de recursos en redes que están bajo la misma administración.

7 CAPA DE TRANSPORTE

La capa de transporte es la encargada de dar confiabilidad al transporte de datos sobre redes, sin tener consideraciones sobre las capas subyacentes. Se puede decir que dentro de la arquitectura de sistema ésta es la primera capa que verdaderamente realiza una abstracción de los niveles inferiores de la sub-red de comunicación, actuando como un protocolo de extremo-a-extremo. Sobre este nivel de abstracción, la capa de transporte implementa servicios para las capas superiores, tales como control de flujo, multiplexación, manejo de circuitos virtuales y chequeo y recuperación de errores.

7.1 Concepto de API (Application Program Interface)

Como su nombre lo describe, una API se refiere a la estructura o conjunto de funciones que provee un sistema operativo en particular para que los procesos (usuarios) tengan acceso a ciertos servicios.. La API define cómo un servicio en particular puede ser invocado mediante algún paquete de software dispuesto para este fin.

Las API's se componen de primitivas de servicio, que son invocaciones al sistema operativo, tal que este se encarga de realizar ciertas funciones especificadas por estas primitivas.

El ejemplo más conocido de API's para protocolos de comunicación es la interfaz *socket* (sistema operativo BSD de UNIX), los cuales son una generalización del mecanismo de acceso a archivos de UNIX que proporciona un punto de acceso para comunicación.

Los socket se basan en un modelo de cliente – servidor. El concepto de servidor se aplica a cualquier proceso dentro de una máquina que ofrece un servicio que se puede obtener mediante una red. El servidor acepta peticiones que llegan por la red, realiza el servicio y devuelve el resultado al solicitante. Luego, se entiende por cliente el proceso (corriendo dentro de la misma máquina o dentro de otra máquina en la red) que envía una petición de servicio y espera respuesta.

La implementación de las características del servicio (control de flujo, confiabilidad, etc) son transparentes para el usuario que las solicita, dejando al sistema la responsabilidad de encargarse de esos aspectos. La aplicación sólo debe especificar el protocolo de comunicación a utilizar (por ejemplo TCP). Luego las primitivas definidas y el flujo de información entre un cliente y un servidor, para el caso de una conexión TCP, será como se muestra en la Figura 7.1

El significado de cada una de ellas es el siguiente:

- *Socket*: crea un punto de comunicación.
- *Bind*: asocia una dirección local (de la máquina) al socket.
- *Listen*: indica que se comenzará a escuchar conexiones, indicando número máximo de ellas.
- *Accept*: acepta las conexiones provenientes desde los clientes. Se bloquea hasta que llega una petición. (servidor)
- *Connect*: intenta establecer una comunicación. (cliente)

- *Read*: lee, desde el socket conectado, la información que llega a través de él.
- *Write*: escribe por el socket conectado la información que se desea enviar.
- *Close*: cierra el socket, liberando la conexión.

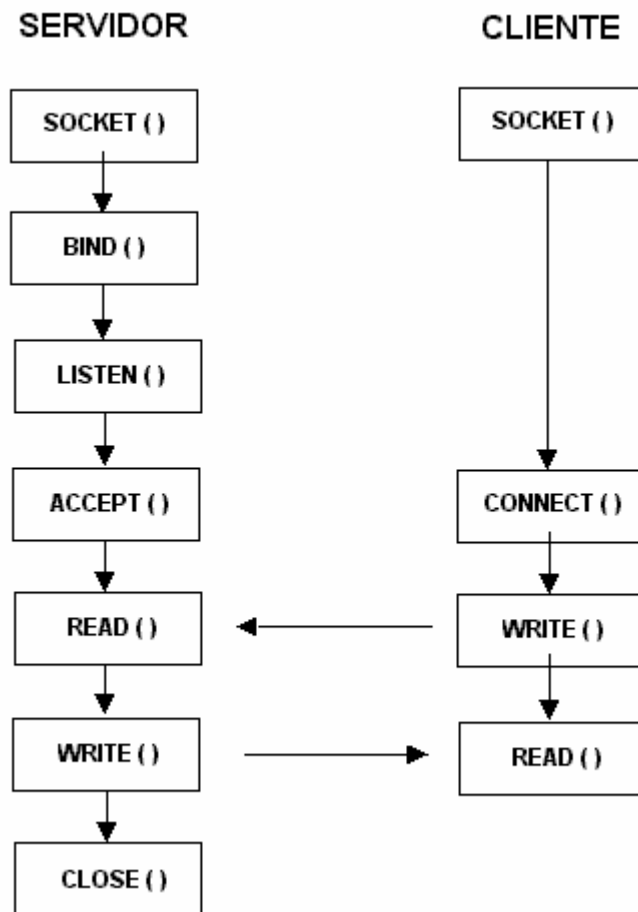


Figura 7.1: Modelo de conexión con socket

Los tipos de conexiones posibles para un protocolo de internet son básicamente dos:

- ❑ No orientado a la conexión y no confiable.
- ❑ Orientado a la conexión y confiable.

Para ilustrar estos dos tipos, se describen a continuación los protocolos dentro de la arquitectura TCP/IP que los implementan.

7.2 UDP (User Datagram Protocol)

Como se vio en el capítulo anterior, para identificar una máquina en una red, IP proporciona direcciones en el interior del encabezado, pero estas direcciones no mencionan nada sobre la aplicación a la que van dirigidos los mensajes. Luego es posible tener varios procesos ejecutándose en una máquina (la cual posee una única dirección IP), entonces ¿cómo saber a cuál de los procesos va dirigido el mensaje contenido en el paquete?. El protocolo más sencillo de capa de transporte que da solución a esta interrogante se conoce como *Protocolo de Datagramas de Usuario* o UDP.

Este protocolo aporta un procedimiento para que los procesos dentro de aplicación puedan enviar mensajes a otros procesos con un mínimo de mecanismo de protocolo. Cada aplicación es identificada por un *puerto de protocolo*, el cual es un número entero contenido dentro del encabezado UDP. El sistema operativo local proporciona un mecanismo de interfaz que los procesos utilizan para especificar o acceder un puerto. A este proceso se le conoce como multiplexado, el cual se grafica en la Figura 7.2.

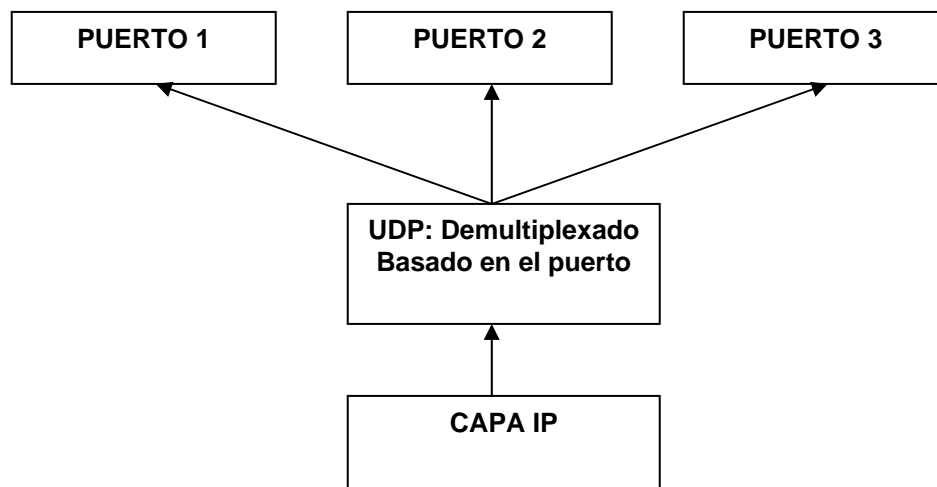


Figura 7.2: Demultiplexado en base a número de puertos

El encabezado UDP, con sus respectivos campos, se muestra a continuación:

0	16	31 bits
Puerto UDP de Origen		Puerto UDP de Destino
Longitud del Mensaje UDP		Suma de Verificacion UDP
Datos		

Figura 7.3: Encabezado UDP

Los campos *Puerto de Origen* y *Puerto de destino* contienen los números que servirán para identificar las distintas aplicaciones, permitiendo el multiplexado de datagramas. El puerto de origen es opcional (puede contener ceros).

El campo de *Longitud* contiene el tamaño del encabezado y datos contenidos en el datagrama UDP en número de bytes.

La *Suma de verificación* es opcional, un valor de cero significa que la suma no se llevó a cabo. Esto se debe a que este protocolo fue pensado para proporcionar un servicio en capa de transporte, no orientado a la conexión y no confiable, más enfocado a aplicaciones en donde el retardo es prioridad. Luego al no realizarse la suma de verificación se ahorra tiempo de procesamiento en el host de origen y destino. Por otro lado, y dado que IP no realiza ninguna operación de verificación sobre el campo de datos, esta instancia en UDP es la única que puede implementar algo de confiabilidad sobre los datos. Esta suma utiliza el mismo algoritmo descrito para IP. En el campo de datos viajará un paquete IP, como el descrito en el capítulo anterior.

7.3 TCP (Transmission Control Protocol)

El segundo protocolo de capa de transporte para el modelo TCP/IP es el conocido como *Protocolo de Control de Transmisión*, el cual a diferencia del anteriormente estudiado, ofrece un servicio orientado a la conexión y confiable. Para implementar estos servicios se vale de una serie de métodos y consideraciones, que lo hacen más complejo, y que serán brevemente tratados en esta sección.

Por flujo confiable se entiende la entrega garantizada de los datos enviados de una máquina a otra sin pérdida o duplicación. Luego la transmisión es hecha confiable gracias al uso de número de secuencias y acuses de recibo. La implementación de acuse de recibo no es la que usualmente se pensaría, es decir, la que envía un segmento de datos y se espera su confirmación para el envío del siguiente. Este método es ineficiente en el sentido que se desperdicia ancho de banda esperando que llegue una confirmación, y la red estará ociosa en los tiempos efectivos de procesamiento de segmento.

El método utilizado se denomina de *ventanas deslizantes (Sliding Windows)*, el cual es una forma de acuse de recibo positivo con retransmisión, en donde se utiliza el ancho de banda de red de mejor forma debido a que se permite el envío de varios segmentos de datos sin esperar confirmación. Como se muestra en la figura 7.4, la idea consiste en ordenar una secuencia de paquetes que serán enviados por la red, colocando una ventana pequeña de tamaño fijo sobre la secuencia. Luego, todos los segmentos en el interior de la ventana, serán transmitidos en forma consecutiva.

Cuando el transmisor recibe un acuse de recibo del primer paquete mueve la ventana en un segmento y envía el siguiente paquete. Entonces la ventana se moverá a medida que vayan llegando las confirmaciones. El método recuerda y mantiene temporizadores para casos en que no lleguen acuses de recibo, lo cual provocará el reenvío del mismo. (Notar que si el tamaño de ventana es 1, el sistema se transforma en el caso base de confirmación positiva). Siempre es posible elegir un tamaño óptimo tal que los tiempos ociosos se reduzcan al mínimo.

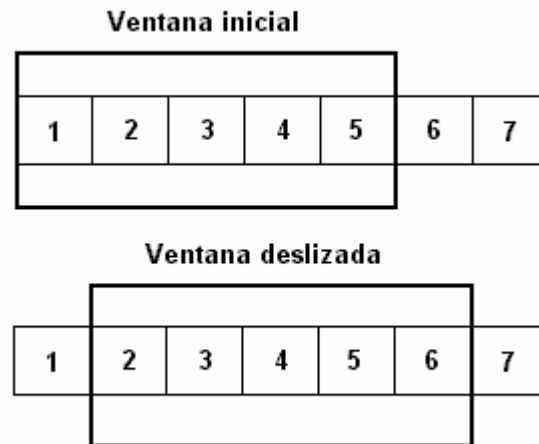


Figura 7.4: Ventana deslizando

Luego TCP implementa este método, pero le agrega una última dimensión, la cual es permitir que el tamaño de la ventana sea variable. Esto es utilizado para realizar control de flujo, es decir, evitar que al enviar muchos paquetes se sature el receptor. Cada acuse de recibo incorpora un aviso de ventana, que especifica los bytes adicionales que esta dispuesto a recibir el receptor. Así el tamaño de la ventana variará en función de la disposición de memoria que tenga el receptor, siendo el caso extremo el tamaño de la ventana igual a cero, lo cual detendría el envío.

Otro aspecto importante de tratar referente a los servicios brindados por TCP, es el establecimiento y fin de la conexión. El establecimiento se realiza con un saludo de tres etapas, consistentes básicamente un intercambio y acuses de recibo de segmentos TCP con algún campo en particular indicando la acción. Una vez que se establece la conexión, está es Full Duplex. De esta forma se garantiza que ambos lados estén listos para transferir datos (y que sepan que ambos están listos), y permite además acordar un número de secuencia inicial para los segmentos. Este saludo (*hand shaking*) se muestra en la figura 7.5.

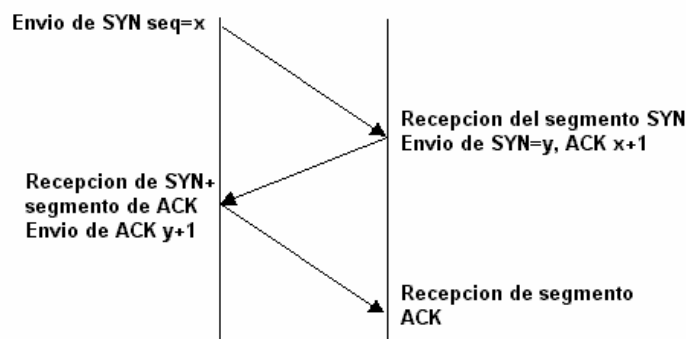


Figura 7.5: Handshake de tres pasos

Para cerrar una sesión se utiliza un procedimiento similar al saludo de tres etapas, pero dado que la comunicación es full duplex, al cerrar un extremo, este envía una trama de fin al otro extremo, lo cual provoca que se avise a la aplicación que concluye la sesión. Mientras esto se hace, ya se envió un segmento de acuse de recibo finalizando las etapas.

TCP utiliza el mismo sistema de multiplexado de UDP, es decir implementa números de puerto para identificar los distintos procesos en una máquina. El encabezado TCP se muestra en la figura 7.6.

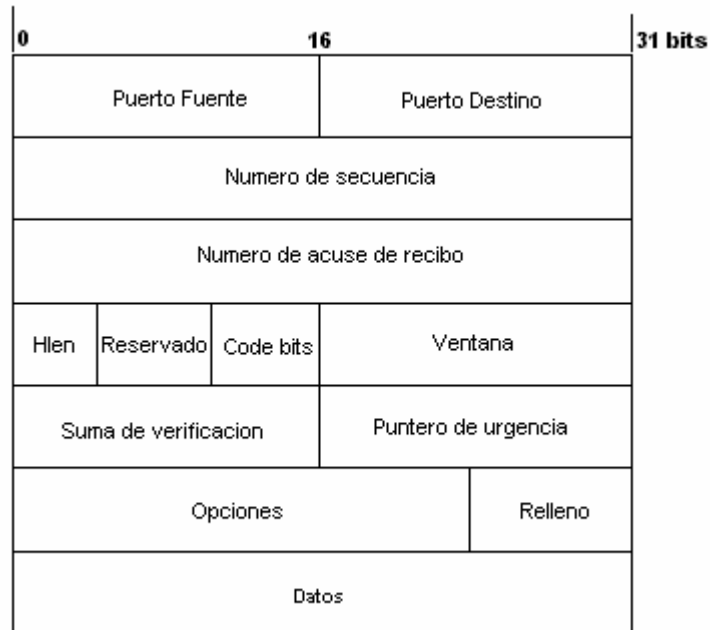


Figura 7.6: Encabezado TCP

8 CAPA DE APLICACIÓN

En este capítulo se estudiarán, en forma resumida, algunos servicios de alto nivel para redes, y algunos protocolos del modelo TCP/IP que las hacen posible. Estos servicios determinan como es que el usuario percibe una red, y la utilidad que esta presta. Existen variadas aplicaciones, las cuales pueden orientarse a necesidades particulares o como una combinación de servicios, tendencia que va en progresivo aumento, gracias a las distintas ofertas multimediales y al aumento de las prestaciones de redes de área extendida.

8.1 Acceso Remoto

El acceso remoto consiste en autorizar al usuario establecer una sesión hacia una máquina dentro de la red y permitirle entonces ejecutar comandos de su sistema. Para este fin existen softwares servidores de acceso remoto para la mayor parte de los sistemas operativos y programas de aplicación que funcionan como clientes.

El protocolo utilizado para este fin en el modelo TCP/IP es conocido como TELNET, el cual permite establecer una conexión TCP hacia un servidor de acceso remoto. TELNET presta un servicio transparente, ya que simula que el teclado y el monitor del usuario están conectados de manera directa a la máquina remota.

TELNET ofrece tres servicios básicos:

- ❑ Define un terminal virtual de red: proporciona interfaz estándar para los sistemas remotos.
- ❑ Negociación de opciones: permite al cliente y al servidor escoger opciones, al mismo tiempo que ofrece opciones estándares (Ej: ASCII de 7 o 8 bits).
- ❑ Simetría en la conexión: no especifica que la entrada del cliente sea el teclado ni la salida la pantalla, permitiendo que cualquier programa se convierta en cliente.

La tarea de un servidor TELNET es aceptar una conexión TCP del cliente y transmitir los datos entre la conexión y el sistema operativo local. En realidad la labor es un poco más compleja pues debe manejar diversas conexiones concurrentemente. Esto lo logra utilizando un esquema de padre e hijo, donde el padre espera por conexión y una vez establecida la hereda al hijo para que este las atienda, mientras el espera por más.

Una desventaja de este protocolo es su ineficiencia, pues cada pulso de teclado viaja, a través del sistema operativo, hacia el programa cliente, el cual las envía a través de la red hasta la máquina servidor. Una vez allí debe pasar por el sistema operativo del servidor al programa de aplicación del servidor y de este programa, de vuelta al sistema operativo para ser enviado de vuelta por el mismo camino hacia el cliente (eco).

Un problema que surge al trabajar con distintos sistemas, es el adaptar los detalles de cada computador con tal que todos los caracteres y comandos representen la misma acción (Ej: Ctl-C y ESC). Como solución TELNET define el como deben enviarse las secuencias de datos y comandos a través de Internet. Es conocido como *network virtual*

terminal (terminal virtual de red o NTV). La figura 8.1 muestra un ejemplo de esta implementación, donde el cliente traduce las pulsaciones de teclado y los comandos que vienen del terminal de usuario a formato NVT y las envía al servidor. El servidor, a su vez, traduce desde el formato NVT al formato del sistema operativo local.

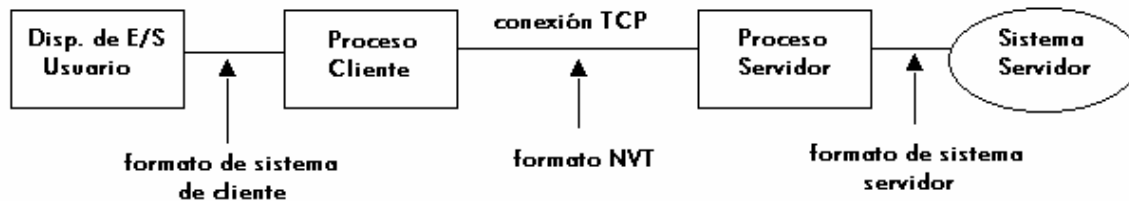


Figura 8.1: TELNET y formato NVT

8.2 Transferencia de Archivos

Para realizar una transferencia de archivos entre máquinas remotas se debe invocar un programa cliente que realice la tarea, especificando la máquina en donde se almacenan los archivos a transferir, y previa autorización esta, se procede al traspaso. Así, el archivo pasará a ser parte del sistema local del usuario, en donde podrá ser modificado sin afectar el archivo original.

El proceso descrito arriba puede ser complejo debido a la no heterogeneidad de las computadoras. Para la transferencia exitosa entonces se debe tener en cuenta una serie de factores como: autorización, nociones de propiedad de archivos y permisos involucrados, formatos de datos, etc.

En esta sección se describirá la forma en que el modelo TCP soluciona estos problemas implementando FTP (*File Transfer Protocol*). Los servicios ofrecidos por FTP son:

- ❑ Acceso interactivo: permite al usuario interactuar con los servidores remotos (Ej: pedir lista de archivos en directorio).
- ❑ Especificaciones de formato: pueden ser de tipo texto binario, caracteres ASCII o EBCDIC.
- ❑ Control de autenticación: requiere que los clientes sean autorizados, mediante nombre de conexión y clave de acceso al servidor.

Los servidores FTP permiten el acceso concurrente de varios clientes, en forma similar a TELNET, creando un proceso hijo el cual se encarga de atender a los usuarios. Pero a diferencia de TELNET, el proceso hijo no ejecuta todas las acciones necesarias para la transferencia. En realidad lo que ocurre es que al establecerse una sesión FTP se crea un proceso hijo encargado del control de conexión y cada vez que se necesite transferir un archivo este crea otro proceso hijo, el cual se encarga de la transferencia. La conexión de control transporta comandos que indican al servidor que archivo transferir. La Figura 8.2 ilustra el concepto.

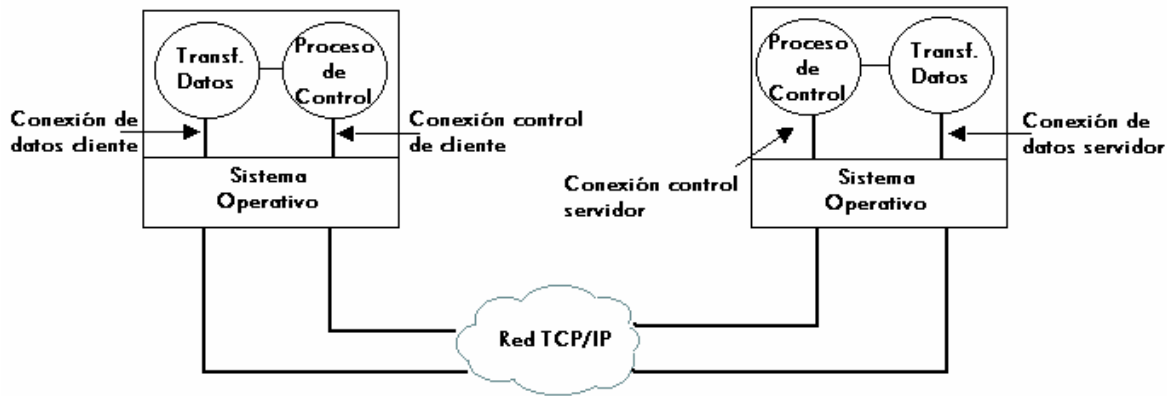


Figura 8.2: Conexión FTP

Mientras dure la sesión se mantendrá la conexión, sin embargo se establece una nueva conexión de transferencia de datos para cada transferencia de archivos. El proceso de control además de enviar comandos del usuario al servidor, permite coordinar el uso de puertos de protocolo TCP asignándolos dinámicamente y creando procesos de transferencia que utilicen tales puertos.

Se permite que FTP utilice el protocolo de terminal virtual de red de trabajo TELNET para enviar los datos con este formato NVT. FTP no permite negociación de opciones, empleando una definición básica de NVT (a diferencia de TELNET).

8.3 Correo Electrónico

Más conocido como e-mail, es uno de los servicios más utilizados gracias a su sencillez para transferir información a través de la red.

Generalmente no es un método inmediato, debido a la filosofía con que se manejan los mensajes, utilizando una técnica conocida como spooling (o colas de correo). Esta consiste en la ocupar áreas de almacenamiento en donde se colocan los mensajes salientes, junto con la identificación de emisor, destino y hora de depósito. La emisión del mensaje entonces puede ser llevada a cabo, vía TCP como una función secundaria.

Entonces el proceso es el siguiente: el proceso secundario actúa como cliente e intenta establecer una conexión TCP hacia el servidor de correo, si lo logra envía una copia del mensaje saliente al servidor, el cual lo almacena en su área spool. Una vez que ambos están de acuerdo en que la copia se ha efectuado con éxito, el cliente borra el mensaje de su cola. Ahora, si no se logra la conexión, el proceso cliente almacena la hora del último intento y cierra el proceso. Así es posible mantener los mensajes dentro del área de almacenamiento hasta que, pasado un tiempo, se produzca un nuevo intento. Cada vez que se encuentra un mensaje pendiente el proceso secundario tratará de enviarlo a su destino. Si luego de un tiempo prolongado no es posible encontrar el destino, el software de correo devuelve el mensaje al emisor.

Para facilitar la interacción con otras redes que no necesariamente manejan el mismo protocolo de red, se utilizan mail gateways, con lo cual el emisor no establece contacto directo con la máquina destino, sino que más bien le entrega el mensaje a máquinas intermedias que se encargan del trabajo. Su ventaja ya fue mencionada y es que permite interoperabilidad, pero posee desventajas que provocan que este servicio no sea del todo confiable. Una vez que el emisor transfiere el mensaje a la primera máquina intermedia, este borra la copia local confiando que este será capaz de entregarla. Así, ni el emisor ni el receptor poseen una copia del mensaje en tránsito y si se encuentran fallas intermedias puede producirse la pérdida de la información sin que el receptor ni emisor sean advertidos. Otra desventaja es que introducen retardo, no existiendo control por parte del emisor.

Para hacer más heterogéneo el servicio de correo, TCP/IP lo divide en dos estándares, uno encargado de los formatos del mensaje de correo y el otro de los detalles de intercambio entre dos máquinas.

El estándar para los formatos de mensaje especifica el formato exacto de los encabezados de correo así como el significado. Los encabezados están compuestos por líneas que comienzan con palabras claves seguidas de dos puntos y algún valor. Por ejemplo, una línea puede comenzar con la palabra clave *To:* y el resto debe contener la dirección de correo electrónico del destino. Mantener el formato del mensaje permite que este sea utilizado por un vasto rango de sistemas

El estándar para intercambio de correo entre computadores definido para TCP/IP es conocido con el nombre de SMTP (*Simple Mail Transfer Protocol*) el cual especifica el formato de los mensajes entre cliente y servidor.

Para comenzar, el cliente debe establecer una conexión TCP y espera que el servidor envíe un mensaje 220 READY FOR MAIL. Una vez recibido este mensaje el cliente envía un comando HELO. Una vez que la comunicación se ha establecido, el emisor puede transmitir uno o más mensajes de correo, terminar la conexión (QUIT) o intercambiar funciones de emisor y receptor vía el comando TURN. Para mayor claridad la Figura 8.3 ilustra un ejemplo extraído desde “*Redes globales de información con Internet y TCP/IP*” de Douglas E. Comer, en donde se supone al cliente Smith en Alpha.EDU envía un mensaje hacia el usuario Jones, Green y Brown en Beta.GOV.

```
S: 220 Beta.GOV Simple Mail Transfer Service Ready
C: HELO Alpha.EDU
S: 250 Beta.GOV

C: MAIL FROM:<Smith@Alpha.EDU>
S: 250 OK

C: RCPT TO:<Jones@Beta.GOV>
S: 250 OK

C: RCPT TO:<Green@Beta.GOV>
S: 550 No such user here

C: RCPT TO:<Brown@Beta.GOV>
S: 250 OK

C: DATA
S: 354 Start mail input; end with <CR><LF>.<CR><LF>
C: ..... cuerpo del mensaje .....
.
.
.
C: <CR><LF>.<CR><LF>
S: 250 OK

C: QUIT
S: 221 Beta.GOV Service closing transmission channel
```

Figura 8.3: Ejemplo de una transferencia SMTP. Las líneas que comienzan con C: son transmitidas por el cliente y las que comienzan con S: por el servidor.

Como se muestra en el cuadro, la máquina Beta.GOV no reconoce al Green, a lo cual responde con un comando 550. También se debe notar que el comando 250 significa todo bien. RCPT identifica a los destinatarios del mensaje.

Por último, es importante mencionar la composición de las direcciones de correo, las cuales, para la Internet tienen la siguiente forma:

local-part@domain-name

Donde *domain name* es el nombre del dominio del destino de correo al que el correo debe ser enviado y *local part* es la dirección de buzón de la máquina.

8.4 HTTP (Hypertext Transport Protocol)

La World Wide Web es una arquitectura armada para el acceso de documentos enlazados a través de miles de máquinas sobre toda la Internet. La forma de acceso es a través de los conocidos *browsers* (buscadores, navegadores) (Ej: Netscape, Internet Explorer, Opera, etc), los cuales utilizan interfaces gráficas muy atractivas, a las cuales se debe la popularidad de la “Web”.

Estos documentos enlazados, conocidos como páginas, pueden contener variada información, como texto, imágenes, enlaces a otros sitios (hyperlinks), etc. Se basan en una conexión confiable vía TCP mediante los mencionados browser, que son programas que se encargan de buscar la información en una dirección determinada, traducirla al formato especificado y desplegarla en pantalla.

Cada servidor Web escucha por conexiones TCP en el puerto 80 y cuando llega una petición de conexión desde un cliente en un browser este responde y se inicia la conexión. El protocolo encargado de definir estas peticiones y réplicas conocido como HTTP. Para ilustrar el flujo de comunicación entre un cliente y un servidor en la web, se mostrarán los pasos desde un ejemplo extraído desde “*Computer Networks*” de Andrew Tanenbaum.

Supongamos que un usuario desea ver la información de la siguiente URL (Uniform Resource Locator) <http://www.w3.org/hypertext/WWW/TheProject.html>, la cual contiene tres partes: el nombre del protocolo (*http*), el nombre de la máquina donde la página se encuentra localizada (*www.w3.org*) y el nombre del archivo que contiene la página (*hypertext/WWW/ /TheProject.html*). Los pasos que ocurren entre el clic del usuario y la página siendo desplegada son los siguientes:

1. El browser determina el URL.
2. El browser pregunta al servidor DNS (Domain Name Server) por la dirección IP de *www.w3.org*.
3. DNS responde con 18.23.0.23.
4. El browser realiza una conexión TCP al puerto 80 sobre 18.23.0.23.
5. Entonces envía un comando GET */hypertext/WWW/TheProject.html*.
6. El servidor 18.23.0.23 envía el archivo *TheProject.html*.
7. La conexión TCP es liberada.
8. El browser despliega todo el texto en *TheProject.htm.l*
9. El browser busca y despliega todas las imágenes en *TheProject.html*.

El protocolo HTTP define una serie de métodos que son usados para requerimientos desde el browser al servidor y viceversa. Algunos de estos métodos se muestran en la Tabla 8.1.

Método	Descripción
GET	Pide una página Web para leer
HEAD	Pide un encabezado de página para leer
PUT	Pide almacenar una página Web
POST	Agrega a una página Web
DELETE	Remueve una página Web
LINK	Conecta dos fuentes existentes
UNLINK	Rompe una conexión existente entre dos fuentes

8.5 APLICACIONES EN TIEMPO REAL

La convergencia entre las redes de datos y las redes de telecomunicaciones ha llevado a la implementación en redes TCP/IP (por ejemplo, Internet) de diversas aplicaciones y servicios que en general requieren de un flujo continuo y constante de paquetes (*streaming traffic*). Estamos hablando de aplicaciones relacionadas con las telecomunicaciones. Entre ellas podemos mencionar VoIP (*Voice over IP*), ToIP (*Telephony over IP*), video conferencia, audio y video de difusión on-line, y, por que no, *video on demand* en un futuro no muy lejano. Estas aplicaciones pueden tolerar en ciertos casos algún nivel de pérdidas de paquetes, pero en general requieren de un flujo más bien sostenidos de paquetes y se implementan sobre el protocolo UDP. Pero como ya se vio, UDP no ejerce ningún control de flujo para evitar la congestión de la red y los siguientes problemas se hacen evidentes: asignación de recursos; garantizar niveles de QoS; y, de proteger las aplicaciones TCP de una degradación extrema.

8.6 INGENIERÍA DE USABILIDAD EN LA OPTIMIZACIÓN DE SISTEMAS HOMBRE-MÁQUINA CONVERSACIONALES

La aceptabilidad por parte del usuario de un sistema de computación es el resultado de la combinación de la aceptabilidad social y de la aceptabilidad práctica. Suponiendo que el sistema es socialmente aceptable, el puede ser analizado bajo la perspectiva de la aceptabilidad práctica, que por su vez se divide en tópicos tradicionales tales como utilidad, precio, compatibilidad con sistemas ya existentes, soporte, confiabilidad e aplicabilidad. Por otro lado, la utilidad depende de la efectividad y de la Usabilidad, donde efectividad se refiere a si la funcionalidad del sistema puede hacer lo que se necesita que haga, y Usabilidad es la cuestión relacionada a cuán bien los usuarios pueden usar tal funcionalidad. El término Usabilidad se aplica así a todos los aspectos de un sistema con los cuales los seres humanos pueden interactuar, incluyendo los procedimientos de instalación y mantenimiento.

Se puede observar que la Usabilidad, así como la aceptabilidad social, es también función de la cultura local y una interface hombre-máquina optimizada para un dado contexto puede no ser apropiada en otra circunstancia (e.g., otro país). Como ejemplo podemos citar que la aceptabilidad de interfaces basadas en control remoto es mucho más elevada en Estados Unidos que en Europa. De manera general, en Estados Unidos hay un consumo mayor que en Europa de sistemas computarizados para diversión (e.g., *video games*) lo que facilita el aprendizaje de las instrucciones de operación de un control remoto. Por otro lado, parece ser que el usuario europeo es, en media, más conservador y tiende a resistir más a paneles de control con varias teclas.

Concluyendo, Ingeniería de Usabilidad se refiere a la cuantificación de todos los parámetros que influyen el uso correcto y eficiente, por parte del usuario, de las funciones para las cuales el sistema fue proyectado.

9 TEORÍA DE TRÁFICO

La teoría de tráfico es una herramienta ampliamente utilizada para el análisis del comportamiento de las redes de comunicaciones, las cuales pueden ser de conmutación de circuitos, como las redes telefónicas, o de conmutación de paquetes como las redes de datos IP. En este capítulo el enfoque irá orientado principalmente a las primeras, ya que son la base en donde se sustenta el sistema telefónico fijo.

La conmutación de circuitos consiste en el establecimiento de un canal dedicado físico (real), de extremo a extremo, entre los cuales existen elementos de conmutación, que en el caso de la red telefónica se trata de centrales públicas (CO – *Central Office*) o PBX (Private Branch eXchange), para el caso de empresas. Los enlaces pueden consistir en ranuras de tiempo en un sistema de multiplexación temporal (TDM) o bandas de frecuencia para el caso de multiplexación en frecuencia (FDM).

Como se verá más adelante con mayor detalle, las redes telefónicas pueden operar en base a bloqueo de llamadas o a la utilización de colas de espera, siendo la más utilizada la primera en las redes públicas debido a su equidad y eficacia. El modelamiento para este tipo de comportamiento se realiza con teorías de cola, utilizando distintas notaciones dependiendo de los supuestos y modelos a aplicar para cada proceso. La notación de Kendall para un sistema general de formación de colas es de la forma:

A / B / C:

A representa la distribución de llegada de requerimientos en un conmutador

B representa la distribución de servicio en un elemento conmutador

C es el número de troncales de salida para el caso de una central telefónica

A continuación se hará una breve descripción de los procesos de Poisson para luego deducir formulas que servirán para medir el grado de servicio de los sistemas antes descritos.

9.1 Procesos de Poisson

Se utilizan tres enunciados básicos para definir los procesos de Poisson. Sea Δt un intervalo de tiempo pequeño ($\Delta t \rightarrow 0$), entonces se tiene:

1. La probabilidad de una llegada en el intervalo Δt se define como $\lambda \Delta t + O(\Delta t)$, $\lambda \Delta t \ll 1$, siendo λ una constante de proporcionalidad especificada.
2. La probabilidad de cero llegadas en Δt es $1 - \lambda \Delta t + O(\Delta t)$.
3. Las llegadas son procesos sin memoria: cada llegada (evento) en un intervalo de tiempo es independiente de eventos en intervalos previos o futuros.

El termino $O(\Delta t)$ denota los elementos $(\Delta t)^n$ con n igual o superior a 2. De acuerdo con 1 y 2, no es posible más de una llegada u ocurrencia de un evento en el intervalo Δt , al menos a $O(\Delta t)$. Sea un intervalo finito T , entonces la probabilidad $p(k)$ de k llegadas en T está dada por:

$$p(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \quad k = 0, 1, 2, 3, \dots \quad (9.1)$$

A (9.1) se le conoce como la distribución de Poisson, en la cual se cumple:

$$E(k) = \sum_{k=0}^{\infty} kp(k) = \sigma_k^2 = \lambda T \quad (9.2)$$

Ahora considerando un intervalo de tiempo mayor, se tendrán una serie de eventos de Poisson, los cuales estarán separados en intervalos. Sea τ el tiempo entre llegadas sucesivas, siendo esta una variable aleatoria. En la estadística de Poisson, τ es una variable aleatoria con distribución exponencial, es decir su función densidad de probabilidad $f_{\tau}(\tau)$ está dada por

$$f_{\tau}(\tau) = \lambda e^{-\lambda \tau} \quad (9.3)$$

A continuación se describen los sistemas utilizados en teoría de tráfico, sección que se basa principalmente en el texto “*Wireless Communications*” de Theodore S. Rappaport, y específicamente en el apéndice A que trata de Teoría de entroncamiento.

Existen dos clases principales de sistemas de entroncamiento:

- ❑ Borrado de llamada pérdida (*Lost Call Cleared o LCC*), sin cola de espera.
- ❑ Retraso de llamada pérdida (*Lost Call Delayed o LCD*), con cola de espera.

En el primer sistema cuando un usuario requiere servicio, existe un tiempo mínimo de configuración, después del cual se le es otorgado el acceso a un canal si este está disponible. En la eventualidad de no existir canal disponible, la llamada es interrumpida sin acceso al sistema, teniendo el usuario la oportunidad de volver a intentar después de un tiempo. Se asume que las llamadas llegan con una distribución de Poisson, y además que existe un número casi infinito de usuarios. La formula de Erlang B describe el grado de servicio (GOS) como la probabilidad que un usuario arbitrario experimente un bloqueo de llamada en un sistema LCC. Se asume que todas las llamadas bloqueadas son retornadas instantáneamente a un recipiente de usuarios infinito, y que cada usuario puede volver a llamar en cualquier momento. El tiempo entre llamadas sucesivas para un usuario bloqueado es un proceso aleatorio y es asumido con distribución de Poisson.

En el sistema LCD, se utilizan colas para mantener en espera las llamadas inicialmente bloqueadas. Si un usuario llama y los canales se encuentran ocupados, su requerimiento es retrasado hasta que un canal se desocupe. Entonces, dado que un canal no está disponible inicialmente, es necesario conocer la probabilidad de que una llamada sea

retrasada, hasta que un canal este disponible para su uso. La probabilidad de que un canal no este inmediatamente disponible en un sistema LCD esta determinada por la fórmula Erlang C. En LCD el GOS es medido por la probabilidad que la llamada sea retrasada en un tiempo mayor que t segundos. Se asume que existe un número infinito de usuarios, y que todas las llamadas en la cola son eventualmente servidas.

A continuación se describirán las formulas de Erlang B en LCC y Erlang C en LCD, los cuales se basan en modelos de cola M/M/C, que implica proceso de llegada de Poisson, estadísticas de servicio con distribución exponencial y C troncales de salida (la M viene de procesos de Markov).

9.2 Erlang B

La formula de Erlang B determina la probabilidad que una llamada sea bloqueada, para sistemas que no utilizan colas de espera (LCC). Está basada en los siguientes supuestos:

- Todos los usuarios, incluso los bloqueados, pueden pedir un canal en cualquier momento (sin memoria).
- Todos los canales libres están disponibles para entregar servicio hasta que todos sean ocupados.
- La probabilidad de utilización de un canal (tiempo de servicio) está exponencialmente distribuido. Es decir, las llamadas largas tienen menos probabilidad de ocurrencia.
- Hay un número finito de canales disponibles.
- La petición de tráfico esta descrita por una distribución de Poisson, lo cual implica un arribo de llamadas en intervalos de tiempo exponenciales.
- Los intervalos de llegada de peticiones de llamada son independientes unas de otras.
- El número de canales ocupados es igual al número de usuarios ocupados.

Sea un sistema con:

C canales

U usuarios

λ número medio de llegada de llamadas por unidad de tiempo (tasa de llegada)

H duración promedio de una llamada

A tráfico total ofrecido por el sistema

A_U tráfico promedio ofrecido para cada usuario

λ_1 tasa promedio de llegada de llamada de un usuario

$$\text{entonces } A_U = \lambda_1 H \Rightarrow A = U A_U = \lambda H.$$

Esta situación se muestra en la Figura 9.1. La probabilidad que una petición de canal de un usuario sea bloqueada esta dada por:

$$\Pr[\text{Bloqueo}] = \Pr[\text{Ninguno de los } C \text{ canales este libre}] \quad (9.4)$$

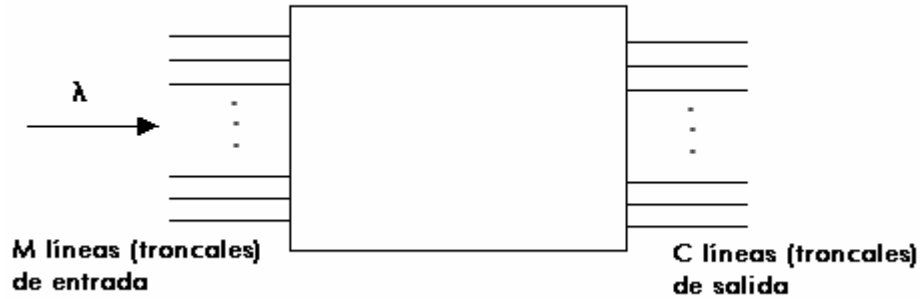


Figura 9.1: Modelo genérico de central de conmutación

Como las llamadas llegan de acuerdo a una distribución de Poisson se tiene

$$\Pr\{a(t + \tau) - a(t) = n\} = \frac{e^{-\lambda\tau}}{n!} (\lambda\tau)^n \quad \text{para } n = 0, 1, 2, \dots \quad (9.5)$$

Donde $a(t)$ es el número de llegadas o eventos que han ocurrido desde $t=0$, y τ es el intervalo de tiempo entre dos eventos sucesivos. Como se vio con anterioridad el tiempo de llegada entre eventos es exponencial del tipo descrito en (9.3). Entonces la probabilidad que el tiempo de llegada sea menor que un tiempo s esta dada por:

$$\Pr(\tau_n \leq s) = 1 - e^{-\lambda s} \quad (9.6)$$

El tiempo de servicio es la duración de una llamada particular que ha sido atendida exitosamente en el sistema. El tiempo de servicio se asume exponencial con duración de llamada promedio H , con lo que $\mu = 1/H$ es la tasa de servicio media (número de llamadas por unidad de tiempo). La probabilidad que el tiempo de servicio del n -ésimo usuario sea menor que algún tiempo de duración s esta dada por:

$$\Pr\{S_n < s\} = 1 - e^{-\mu s} \quad s > 0 \quad (9.7)$$

Donde la función densidad d probabilidad de tiempo de servicio es

$$p(S_n) = \mu e^{-\mu S_n} \quad (9.8)$$

y S_n es el tiempo de servicio del n -ésimo usuario.

Para derivar la formula para Erlang B es necesario utilizar propiedades de las cadenas de Markov. Consideremos un proceso estocástico de tiempo discreto que toma valores desde un conjunto de enteros no negativos, tal que los posibles estados del proceso son $i = 0, 1, 2, \dots, C-1, C$. En otras palabras, cada estado de la cadena de Markov corresponde al número de troncales de salidas siendo utilizados. El proceso es una cadena de Markov si la transición desde el estado presente i al estado próximo $i+1$ depende solo del estado i y no de estados previos. La operación de sistemas de entroncamiento es de

tiempo continuo, pero puede ser analizado en pequeños intervalos δ ($\delta \rightarrow 0$), donde $\delta > 0$. Si N_k es el número de llamadas (canales ocupados) en el instante $k\delta$, entonces N_k puede ser representado como

$$N_k = N(k\delta) \quad (9.9)$$

La probabilidad de transición está dada por

$$P_{i,j} = \Pr\{N_{k+1} = j \mid N_k = i\} \quad (9.10)$$

Usando el enunciado básico número 2 de procesos de Poisson y permitiendo que $\delta \rightarrow 0$, se tiene

$$P_{00} = 1 - \lambda\delta + O(\delta) \quad (9.11)$$

$$P_{ii} = 1 - \lambda\delta - \mu\delta + O(\delta) \quad i \geq 1 \quad (9.12)$$

$$P_{i,i+1} = \lambda\delta + O(\delta) \quad i \geq 0 \quad (9.13)$$

$$P_{i,i-1} = i \cdot \mu \cdot \delta + O(\delta) \quad i \geq 1 \quad (9.14)$$

$$P_{i,j} = O(\delta) \quad j \neq i, j \neq i+1, j \neq i-1 \quad (9.15)$$

Estas relaciones quedan mejor graficadas en la figura 9.2.

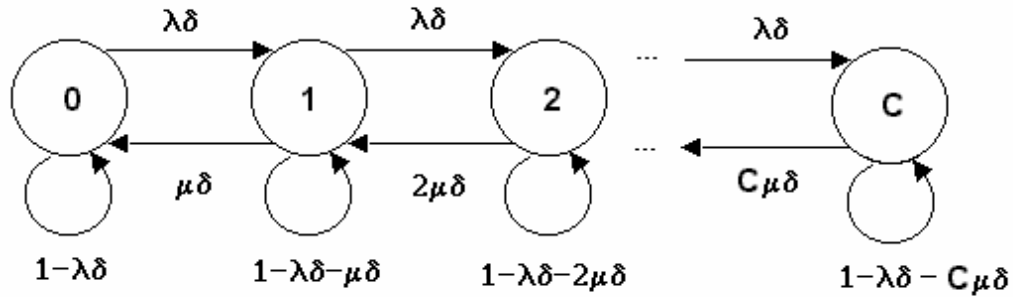


Figura 9.2: Probabilidades de transición representada como una cadena de Markov

Para entender la cadena supongase que al comienzo se tienen 0 canales ocupados, es decir no hay usuarios. Sobre un pequeño intervalo de tiempo, la probabilidad que el sistema continúe sin usuarios es $(1 - \lambda\delta)$. La probabilidad de que haya un cambio desde 0 a 1 usuario esta dada por $\lambda\delta$. En el otro extremo, si un canal esta en uso, la probabilidad de que el sistema pase a 0 canales ocupados esta dada por $\mu\delta$. Similarmente, la probabilidad que el

sistema continué con un canal en uso esta dada por $1 - \lambda\delta - \mu\delta$. Todas las probabilidades de salida para un cierto estado suman 1.

Sobre un gran período de tiempo, el sistema alcanza el estado de régimen permanente y tiene n canales en uso. Entonces bajo régimen permanente se cumple

$$\lambda\delta P_{n-1} = n\mu\delta P_n \quad n \leq C \quad (9.16)$$

La ecuación (9.16) es conocida como la ecuación general de balance. Además

$$\sum_{n=0}^C P_n = 1 \quad (9.17)$$

Ocupando (9.16) se obtiene

$$P_1 = \frac{\lambda P_0}{\mu} \quad (9.18)$$

Evaluando (9.16) para diferentes valores se obtiene

$$P_n = P_0 \left(\frac{\lambda}{\mu} \right)^n \frac{1}{n!} \quad (9.19)$$

Y

$$P_0 = \left(\frac{\mu}{\lambda} \right)^n P_n n! = 1 - \sum_{i=1}^C P_i \quad (9.20)$$

Sustituyendo (9.19) en (9.20)

$$P_0 = \frac{1}{\sum_{n=0}^C \left(\frac{\lambda}{\mu} \right)^n \frac{1}{n!}} \quad (9.21)$$

En (9.19) la probabilidad de bloqueo para C canales es

$$P_c = P_0 \left(\frac{\lambda}{\mu} \right)^C \frac{1}{C!} \quad (9.22)$$

Sustituyendo (9.21) en (9.22), con $A = \lambda H = \lambda/\mu$ se tiene

$$P_c = \frac{A^C \frac{1}{C!}}{\sum_{n=0}^C A^n \frac{1}{n!}} \quad (9.23)$$

La cual representa la formula para Erlang B. En la Figura 9.3 se aprecia las curvas características de probabilidad de bloqueo como función del número de canales e intensidad de tráfico en Erlang.

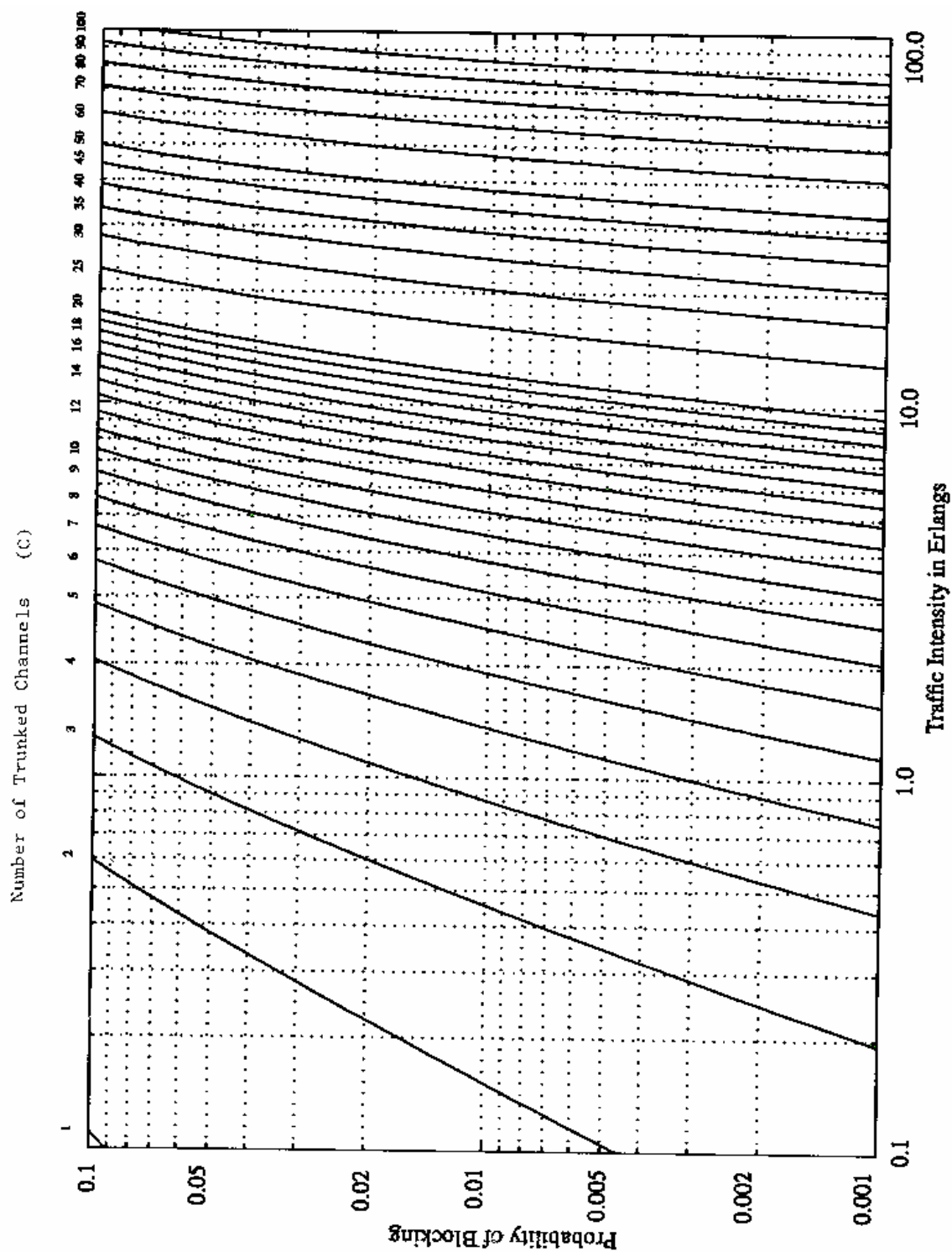


Figura 9.3: Probabilidad de bloqueo como función del número de canales y la intensidad de tráfico en Erlangs. *

* Fuente: "Wireless Communications", T. Rappaport, Capítulo 2

9.3 Erlang C

Sea un sistema con C números de troncales de salida, como se muestra en la figura 9.1. Para derivar la formula de Erlang C se procede de manera similar que en la sección anterior, excepto que ahora se asume que si a una llamada no se le asocia un canal, esta es puesta en una cola (no es bloqueada), la cual tiene un largo infinito. Luego las ecuaciones (9.5), (9.6) y (9.8) siguen siendo validas.

El diagrama de estado para este modelo se muestra en la Figura 9.4.

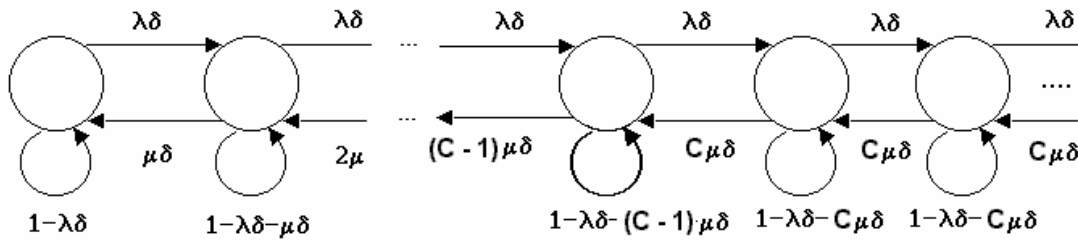


Figura 9.4: Probabilidades de transición como una cadena de Markov

En estado permanente, la probabilidad que el sistema este en estado k y se produzca una transición al estado $k - 1$ en el próximo intervalo de transición es la misma que la probabilidad que el sistema esté en estado $k - 1$ y transite hacia el estado k . Entonces desde el diagrama de estado de la figura 9.4

$$\lambda\delta P_{k-1} = k\mu\delta P_k \quad \text{para } k \leq C \quad (9.24)$$

entonces

$$P_k = \left(\frac{\lambda}{\mu}\right) \frac{1}{k} P_{k-1} \quad \text{para } k \leq C \quad (9.25)$$

y

$$\lambda\delta P_{k-1} = C\mu\delta P_k \quad \text{para } k \geq C \quad (9.26)$$

entonces

$$P_k = \left(\frac{\lambda}{\mu}\right) \frac{1}{C} P_{k-1} \quad \text{para } k \geq C \quad (9.27)$$

de lo cual se puede desprender que

$$P_k = \begin{cases} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} P_0 & k \leq C \\ \frac{1}{C!} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{C^{k-C}} P_0 & k \geq C \end{cases} \quad (9.28)$$

Ya que $\sum_{k=0}^{\infty} P_k = 1$ entonces

$$P_0 = \frac{1}{\sum_{k=1}^{C-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \frac{1}{C!} \left(\frac{\lambda}{\mu}\right)^C \frac{1}{\left(1 - \frac{\lambda}{\mu C}\right)}} \quad (9.29)$$

La probabilidad que una llamada llegue cuando todos los C canales estén ocupados y entonces tenga que esperar puede ser determinada usando la ecuación (9.28)

$$\Pr[C \text{ canales esten ocupados}] = \sum_{k=C}^{\infty} P_k = P_0 \frac{1}{C!} \left(\frac{\lambda}{\mu}\right)^C \frac{1}{\left(1 - \frac{\lambda}{\mu C}\right)} \quad (9.30)$$

La cual es válida para $\frac{\lambda}{\mu C} < 1$. Sustituyendo P_0 desde (9.29) y haciendo $A = U\lambda_1 H = \lambda/\mu$ se obtiene:

$$\Pr[C \text{ canales esten ocupados}] = \frac{A^C}{A^C + C! \left(1 - \frac{A}{C}\right) \sum_{k=0}^{C-1} \frac{A^k}{k!}} \quad (9.31)$$

La cual corresponde a la formula de Erlang C.

En la el gráfico de la Figura 9.5 se muestra la probabilidad de llamadas siendo retrasadas como una función del número de canales y la intensidad de tráfico en Erlangs.

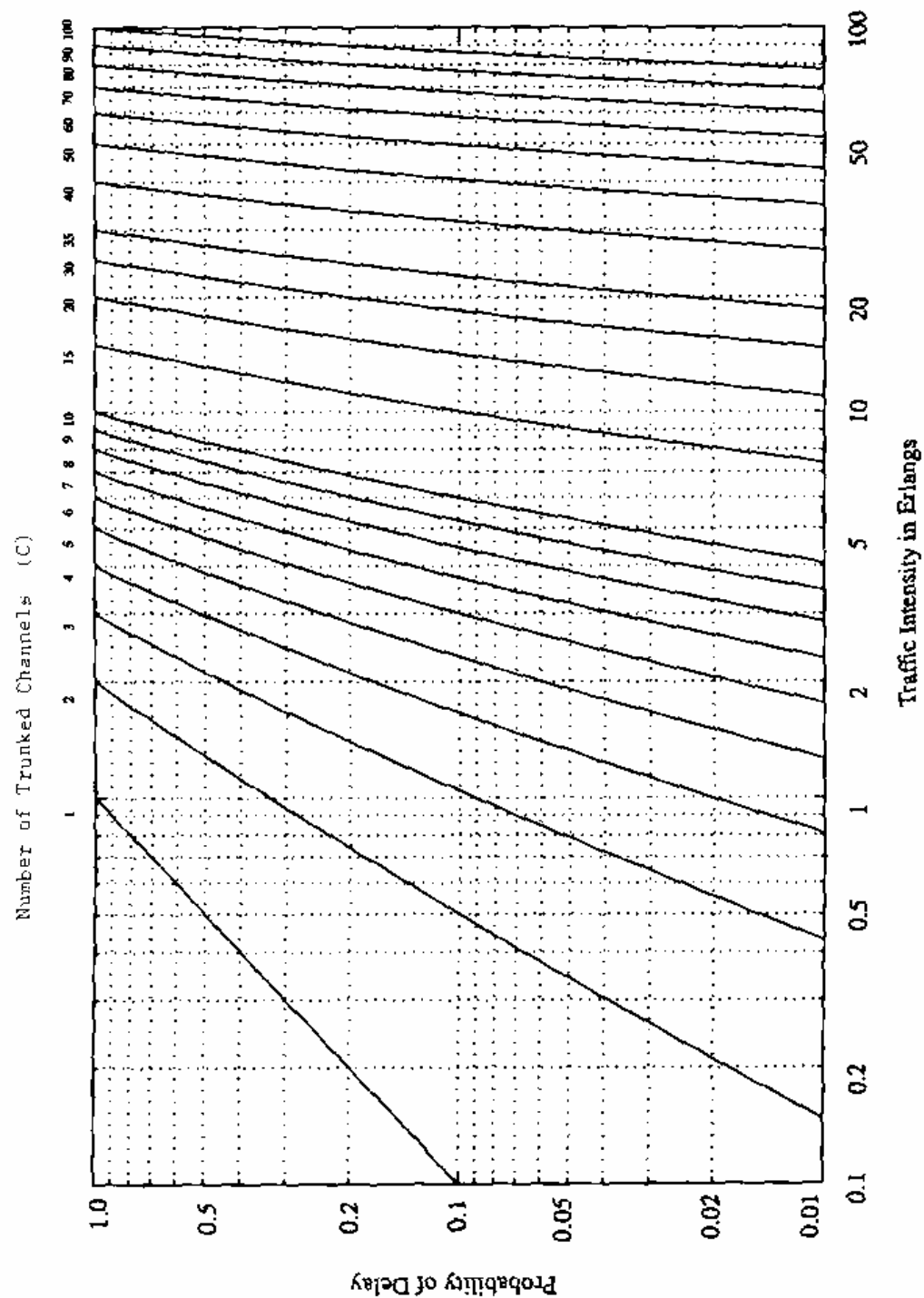


Figura 9.5: Probabilidad que la llamada sea retrasada como una función del número de canales y la intensidad de tráfico en Erlangs. *

* Fuente: “Wireless Communications”, T. Rappaport, Capítulo 2

