

**GUIA DE ESTUDIO
“COMUNICACIONES II PARTE I”**

JOSE LUIS RODRIGUEZ BONFANTE

**UNIVERSIDAD TECNOLOGICA DE BOLIVAR
FACULTAD DE INGENIERIA ELECTRICA Y ELECTRONICA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
CARTAGENA DE INDIAS D. T. Y C.**

2009

**GUIA DE ESTUDIO
“COMUNICACIONES II PARTE I”**

JOSE LUIS RODRIGUEZ BONFANTE

**Trabajo de monografía presentado como requisito para optar al título de
Ingeniero Electrónico**

**DIRECTOR
ING. GONZALO LÓPEZ**

**UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR
FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
CARTAGENA DE INDIAS D. T. Y C.**

2009

NOTA DE ACEPTACIÓN:

Firma del Presidente del Jurado

Firma del Jurado

Firma del Jurado

Cartagena de Indias D. T. y C., Julio de 2009

Señores

COMITÉ CURRICULAR

Facultad de Ingeniería Eléctrica y Electrónica

Universidad Tecnológica de Bolívar

Ciudad

Respetados señores:

De la manera más atenta me dirijo a ustedes, con el propósito de informarles que la monografía titulada “**Guía de Estudio COMUNICACIONES II PARTE I**” ha sido desarrollada conforme a los objetivos establecidos.

Como autor de la monografía considero que el trabajo es satisfactorio y solicito sea estudiado, evaluado y posteriormente aprobado por ustedes.

En espera de los resultados de dicha evaluación.

Atentamente,

José Luís Rodríguez Bonfante

CC N° 73.006.542 de Cartagena

Cartagena de Indias D. T. y C., Julio de 2009

Señores

COMITÉ CURRICULAR

Facultad de Ingeniería Eléctrica y Electrónica

Universidad Tecnológica de Bolívar

Ciudad

Cordial Saludo.

A través de la presente me permito poner en consideración para su respectiva evaluación, la monografía titulada “**Guía de estudio COMUNICACIONES II PARTE I**”, la cual fue realizada por el estudiante JOSE LUIS RODRIGUEZ BONFANTE, a quien asesoré en su ejecución.

Atentamente;

GONZALO LÓPEZ

Asesor de Monografía

AUTORIZACIÓN

Cartagena de Indias D. T. y C., Julio de 2009

Yo JOSE LUIS RODRIGUEZ BONFANTE, identificado con cedula de ciudadanía N° 73.006.542 de Cartagena, autorizo a la UNIVERSIDAD TECNOLOGICA DE BOLIVAR, para hacer uso de mi monografía y publicarlo en el catalogo en línea de la biblioteca.

Atentamente;

José Luís Rodríguez Bonfante

CC N° 73.006.542 de Cartagena

AGRADECIMIENTOS

En la elaboración de la Guía de Estudio COMUNICACIONES II PARTE I, agradezco en primer lugar a Dios y a mi familia por el apoyo y motivación dada durante el desarrollo de este trabajo.

Doy gracias a la Universidad Tecnológica de Bolívar, en especial a sus programas de Ingeniería Electrónica e Ingeniería Eléctrica por trazar un camino, en el cual se ganó muchos conocimientos necesarios para mi formación como profesional.

Agradezco al Ingeniero Gonzalo López Vergara por su compromiso, disposición y colaboración presentada como asesor en el proceso de realización de esta guía, el cual me motivo a trabajar con mucha dedicación sin mirar atrás.

Agradezco a los docentes de los programas de Ingeniería Electrónica e Ingeniería Eléctrica, por aportar sus conocimientos, los cuales me fueron muy útiles en la realización de esta guía, así como en mi formación profesional.

A mis compañeros de clase y amigos, los cuales estuvieron presentes en mi formación y en el desarrollo de esta guía.

GRACIAS!!!

The background of the cover features a semi-transparent, green-tinted image of a modern architectural complex. On the left, a tall palm tree stands in front of a curved building with multiple levels and windows. In the lower center, a large, dome-shaped structure with a grid-like pattern is visible. The overall aesthetic is clean and academic.

GUIA DE ESTUDIO **COMUNICACIONES II PARTE I**

JOSE LUIS RODRIGUEZ BONFANTE

UNIVERSIDAD TECNOLOGICA DE BOLIVAR

2009

INTRODUCCIÓN

Guía de Estudio COMUNICACIONES II PARTE I, es un texto que servirá a los estudiantes de ingeniería electrónica, comprender y hacer énfasis en los conceptos y la matemática más importante de las redes de conmutación, el sistema de señalización SS7 y el tráfico telefónico.

Esta guía se encuentra dividida en tres unidades: la primera de ella consiste en las redes de conmutación en donde se estudiarán las dos principales técnicas de conmutación, las cuales son las técnicas de conmutación de circuitos y la técnica de conmutación de paquetes.

En la segunda unidad se analizan la estructura, protocolos y aspectos más importantes del sistema de señalización SS7, en donde se observa además los diferentes formatos de la unidad de señalización, control de errores y los conceptos de tratamiento y gestión de mensajes.

Por último en la tercera unidad se abarca los principales tópicos de tráfico telefónico, introduciéndose en los modelos básicos de sistemas de colas, así como sus componentes más importantes, los procesos de nacimiento y muerte y los sistemas basados en este concepto.

Al inicio de cada unidad encontrará una introducción referente a los temas vistos en cada unidad, en el transcurso de la unidad se podrán ver ilustraciones así como ejercicios los cuales ayudarán a comprender los temas de la unidad.

Al final de la guía se encontrará las referencias bibliográficas utilizadas para la elaboración de la guía.

Espero que la Guía de Estudio de COMUNICACIONES II PARTE I, sea un apoyo que aporte al desarrollo académico y profesional de los lectores, que en su mayoría serán estudiantes de ingeniería electrónica.

INDICE

	Pag
UNIDAD I: REDES DE CONMUTACIÓN	13
1.1 INTRODUCCIÓN A LAS REDES DE CONMUTACIÓN	15
1.2 CONMUTACIÓN DE CIRCUITOS	15
1.2.1 Tipos de Conmutadores	16
1.2.1.1 Conmutador espacial	17
1.2.1.2 Conmutador temporal	21
1.2.2 Conmutación Digital	23
1.2.2.1 Bloques funcionales en una central digital	24
1.2.2.2 La red de conmutación digital	26
1.2.2.3 Circuitos de línea de abonado	30
1.2.2.4 Interfaz de troncal digital	30
1.3 CONMUTACIÓN DE PAQUETES	32
1.3.1 Técnicas de Conmutación	33
1.3.1.1 Datagramas	33
1.3.1.2 Circuitos virtuales	34
1.3.2 Comparación entre técnicas de datagramas y c. virtuales	36
1.3.3 Algoritmos del Camino más Corto	40
1.3.3.1 Algoritmo de Bellman Ford	40
1.3.3.2 Algoritmo de Dijkstra	42
UNIDAD II: SISTEMA DE SEÑALIZACIÓN SS7	46
2.1 INTRODUCCIÓN AL SISTEMA SS7	48

2.2 ESTRUCTURA DEL SISTEMA SS7	48
2.2.1 Puntos de Señalización	48
2.2.2 Jerarquía	50
2.3 PROTOCOLOS	51
2.3.1 Parte de Transferencia de Mensajes (MTP)	51
2.3.2 Parte de Usuario Telefónico (TUP)	52
2.3.3 Parte de Usuario de Servicios Integrados (ISUP)	52
2.3.4 Parte de Control de Conexión de Señalización (SCCP)	53
2.3.5 Parte de Aplicación de las Capacidades de Transacción (TCAP)	53
2.3.6 Parte de Administración, Mantenimiento y Operación (OMAP)	53
2.4 MENSAJES Y TRANSFERENCIA DE MENSAJES	54
2.5 EJEMPLO BÁSICO DE LLAMADA	55
2.6 PETICIÓN A UNA BASE DE DATOS	58
2.7 FUNCIONES MTP	60
2.7.1 Enlace de Datos de Señalización (MTP 1)	60
2.7.2 Funciones del Enlace de Señalización (MTP 2)	61
2.7.2.1 Estructura MTP 2	61
2.7.2.2 Descripción general de MTP 2	62
2.7.2.3 Unidades de señalización	63
2.7.2.4 Control de errores	69
2.7.3 Funciones de la Red de Señalización	76
2.7.3.1 Tratamiento de mensajes de señalización	77
2.7.3.2 Gestión de la red de señalización	82
2.8 PARTE DE CONTROL DE CONEXIÓN DE SEÑALIZACIÓN (SCCP)	84
2.9 PARTE DE APL. DE LAS CAPACIDADES DE TRANSACCIÓN (TCAP)	85

UNIDAD III: TRÁFICO TELEFÓNICO	87
3.1 INTRODUCCIÓN A LOS SISTEMAS DE COLAS	89
3.2 ELEMENTOS DE LOS SISTEMAS DE COLAS	89
3.2.1 Magnitudes Fundamentales en un S.E	91
3.3 ANALISIS DE RETARDO Y FÓRMULA DE LITTLE	92
3.3.1 Definiciones de Tasas de Llegada y Carga de Tráfico	93
3.3.2 Fórmula de Little	94
3.4 MODELOS BÁSICOS DE COLAS	97
3.4.1 Procesos de Arribo	97
3.4.2 Proceso de Poisson	98
3.4.3 Tiempos de Servicio	100
3.4.4 Clasificación de los Sistemas de Cola	102
3.4.5 Factor de Utilización y Estabilidad	104
3.5 PROCESOS DE NACIMIENTO Y MUERTE	105
3.5.1 Sistemas M/M/1	108
3.5.2 Sistemas M/M/1/K	111
3.6 CONCEPTOS DE TELETRÁFICO	114
3.6.1 Modelo Erlang B	115
BIBLIOGRAFIA	118

UNIDAD I

REDES DE CONMUTACIÓN

INTRODUCCIÓN

En este capítulo se estudiarán y analizarán las redes de conmutación y se hará énfasis en los tipos de conmutación más utilizados, analizando sus aplicaciones, sus estructuras, así como ventajas y desventajas.

Se realizará un apoyo al lector para ayudar a comprender este tema mediante un resumen de este basándose en los tópicos más importantes.

Además se hará énfasis en la conmutación digital en donde se detallará a profundidad los diferentes bloques que la componen.

Se describirán los diferentes tipos de conmutación, los cuales son:

- Conmutación de Circuitos
- Conmutación de Paquetes

Estas dos tipos de técnicas serán analizadas y tratadas en este capítulo viendo sus características, ventajas y aplicaciones.

En las redes de conmutación de circuitos veremos aspectos importantes como sus tipos de conmutadores, los cuales son espaciales y temporales, analizando sus características; por otra parte se hará énfasis en aspectos importantes de la red de conmutación digital, en la cual veremos los bloques de una central digital, los circuitos de línea de abonado y la interfaz de troncal digital.

En las redes de conmutación de paquetes analizaremos sus diferentes técnicas de conmutación, las cuales son datagramas y circuitos virtuales, viendo las ventajas de cada una de ellas, además se hará énfasis en aspectos importantes de las comparaciones de estas dos técnicas de conmutación, así como el concepto de retardo en la transmisión.

Al final de este capítulo se hará un resumen sobre los algoritmos del camino más corto, siendo estos el algoritmo de BELLMAN-FORD y el algoritmo de DIJKSTRA.

1.1 INTRODUCCIÓN A LAS REDES DE CONMUTACIÓN

Una red de conmutación es aquella que permite la transferencia de información entre dos usuarios de una red de telecomunicaciones, la conmutación permite la descongestión entre los usuarios de la red disminuyendo el tráfico y aumentando el ancho de banda. A continuación se verán las técnicas de conmutación utilizadas en una red de telecomunicaciones, las cuales son la técnica de conmutación de circuitos y la técnica de conmutación de paquetes.

1.2 CONMUTACIÓN DE CIRCUITOS

En la conmutación de circuitos los equipos de conmutación deben establecer un camino físico entre los medios de comunicación previa a la conexión entre los usuarios. Este camino permanece activo durante la comunicación entre los usuarios, liberándose al terminar la comunicación. Ejemplo: Red Telefónica Conmutada.

La conmutación de circuitos puede llegar a ser bastante ineficiente. La capacidad del canal se dedica permanentemente a la conexión mientras dura esta, incluso si no se transfieren datos. Aunque no se alcanza el cien por cien, la utilización puede ser bastante alta para una conexión de voz. Por su parte para comunicaciones entre un terminal y un computador es posible que el canal este libre durante la mayor parte de la conexión.

La conmutación de circuitos fue desarrollada para el tráfico de voz, pero en realidad se usa también para el tráfico de datos. El mejor ejemplo conocido de una red de conmutación de circuitos es el de la red telefónica pública (PSTN, Public switching telephone Network, ver figura 1.1), la cual es en la actualidad un conjunto de redes nacionales interconectadas para ofrecer un servicio internacional.

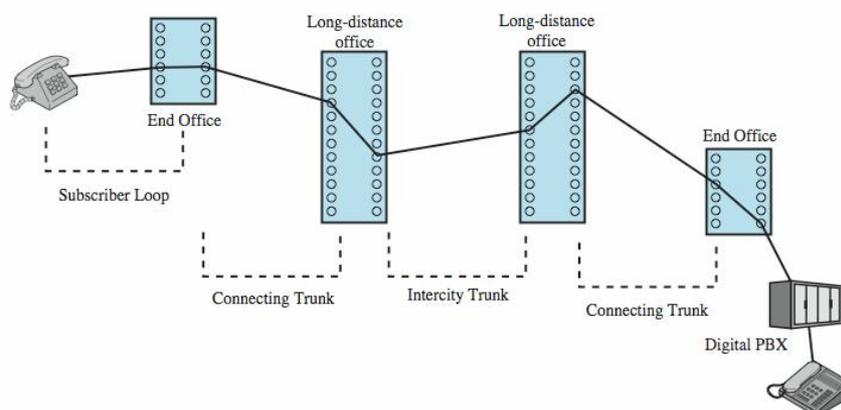


Figura 1.1 Red de conmutación de circuitos (Imagen tomada del libro *Comunicación y Redes de Computadoras*, Williams Stallings, Octava edición).

Una comunicación mediante circuitos conmutados posee tres etapas bien definidas:

Establecimiento del circuito

Cuando un usuario quiere obtener servicios de la red para establecer una comunicación se deberá establecer un circuito entre la estación de origen y la de destino. En esta etapa, dependiendo de la tecnología utilizada, se pueden establecer la capacidad del canal y el tipo de servicio.

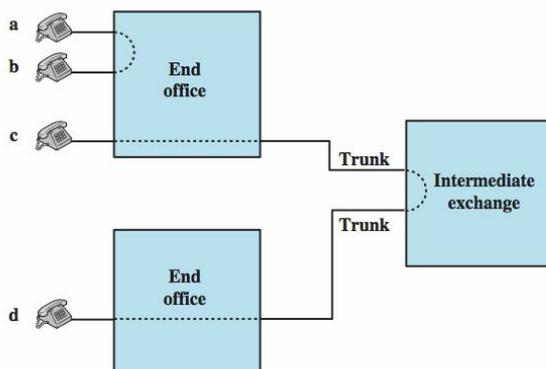


Figura 1.2 Establecimiento de un circuito (Imagen tomada del libro Comunicación y Redes de Computadoras, Williams Stallings, Octava edición).

Transferencia de información

Una vez que se ha establecido un circuito puede comenzar la transmisión de información. Dependiendo del tipo de redes y del tipo de servicio la transmisión será digital o analógica y el sentido de la misma será unidireccional o full duplex.

Liberación de llamada

Una vez que se ha transmitido todos los datos, una de las estaciones comienza la terminación de la sesión y la desconexión del circuito. Una vez liberado los recursos utilizados por el circuito pueden ser usados por otra comunicación.

1.2.1 TIPOS DE CONMUTADORES

Para introducirse en el concepto de la conmutación digital, hay que tener en cuenta que toda información precedente de un abonado llega al conmutador en forma digital. Cada abonado tiene una interfaz de línea de abonado (ILA) que además de realizar la conversión de señales analógicas a digitales en el sentido de transmisor y al contrario en recepción, poseen otras funciones que hacen posible el establecimiento de la conexión.

Una central telefónica digital tal como lo muestra la figura 1.3, esta constituida básicamente por un conmutador digital, un control, un módulo de atención a abonados, un módulo de troncales, y generadores y receptores de tonos digitales.

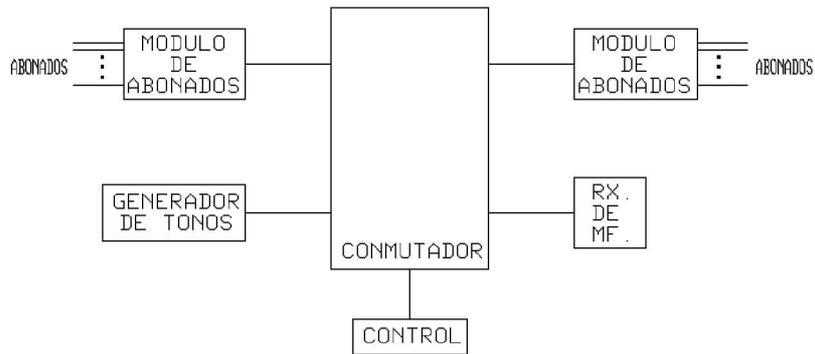


Figura 1.3 Conmutador Digital en una central telefónica

El sistema tiene separado los buses de transmisión y recepción, sobre los cuales se hallan conectadas las interfaces de línea de abonados pertenecientes a la central, los servicios y las troncales tanto entrantes como salientes que permiten que sus abonados se conecten con otros abonados que forman parte de otras centrales.

La conmutación de la información transportada en un intervalo de tiempo de entrada sobre uno de salida es unidireccional, lo que significa que para lograr el intercambio de mensajes en forma bidireccional, como lo exige una conversación, es necesario crear dos enlaces, uno que haga posible que la información de un abonado A le llegue a otro abonado B, y otro enlace en sentido contrario.

Existen dos tipos de conmutadores digitales que se estudiarán a continuación:

- Conmutador Espacial
- Conmutador Temporal

1.2.1.1 Conmutador Espacial

Este conmutador es implementado mediante un juego de compuertas que actúan a la manera de una matriz de puntos de cruce, en el las rutas de señal que se establecen son físicamente independientes entre si (separadas en el espacio). Cada conexión necesita del establecimiento de un camino físico a través del conmutador que se dedique únicamente a la transferencia de señales entre los dos extremos. El bloque básico de un conmutador consiste en una matriz de conexiones o puntos de cruce que una unidad de control puede habilitar o deshabilitar.

En la figura 1.4 se muestra una matriz de conexiones simple con 10 líneas de entrada/salida full-duplex. La matriz tiene 10 entradas y 10 salidas; cada estación se conecta a la matriz a través de una línea de entrada y otra de salida. La conexión entre dos líneas cualesquiera es posible habilitando el punto de cruce correspondiente. Obsérvese que es necesario un total de 100 conexiones. Los conmutadores matriciales presentan varias limitaciones:

- El número de conexiones crece con el cuadrado del número de estaciones conectadas, lo cual resulta costoso para conmutadores grandes.
- La pérdida de un cruce impide la conexión entre los dos dispositivos cuyas líneas se interseccionan en ese punto de cruce.
- Las conexiones se utilizan de forma ineficiente; incluso cuando todos los dispositivos conectados se encuentran activos, solo está ocupada una pequeña fracción de los puntos de cruce.

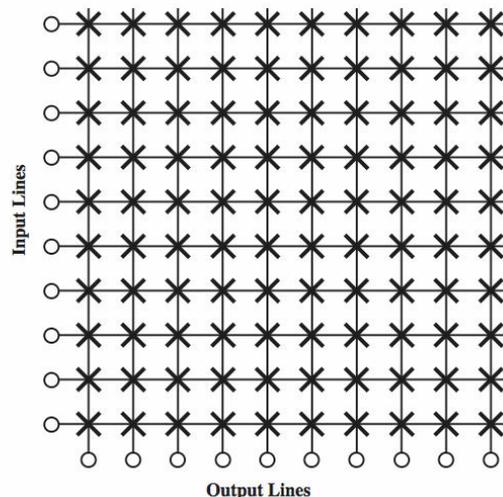


Figura 1.4 Conmutador por división en el espacio (Imagen tomada del libro *Comunicación y Redes de Computadoras*, Williams Stallings, Octava edición).

Como es necesario que se realice una conmutación periódica, si se desea mantener el camino establecido, se requiere de una memoria de escritura y lectura que contenga la información de los puntos que se deben cerrar en un determinado intervalo de tiempo.

Estas memorias de control son de lectura cíclica, mientras que la escritura, que depende del tráfico, es llevada a cabo de manera aleatoria por el procesador.

En las figuras 1.5, 1.6, 1.7 y 1.8 se muestra el funcionamiento de un conmutador espacial, en la figura 1.5 se muestran las memorias de retención, los canales o enlaces de entrada y salida con sus informaciones en tres intervalos de tiempo y la matriz de cruce.

Los puntos de cruce de cada vertical de ésta matriz están controlados por la información que está en la memoria de retención, las cuales en cada instante de tiempo (t_1, t_2, t_3, \dots) seleccionan los puntos de cruces de la matriz para que se interconecten los enlaces de entrada y salida.

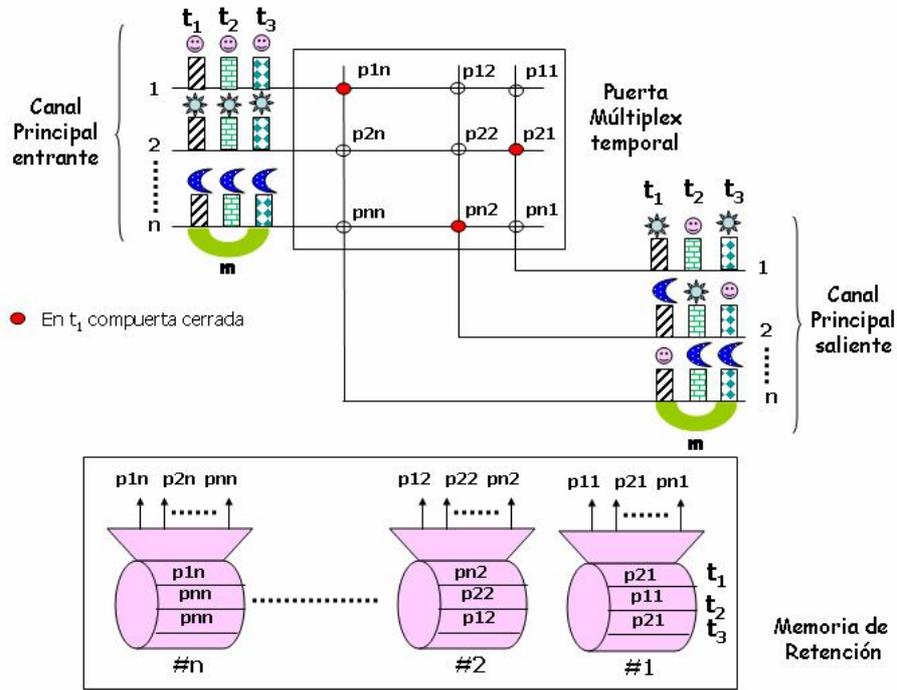


Figura 1.5 Funcionamiento de conmutador espacial

En la figura 1.6 en la memoria de retención #n, #2 y #1 en el tiempo t_1 se lee la informaciónes p_{1n} , p_{n2} , y p_{21} que corresponden respectivamente a la activación de los puntos de cruce de la horizontal 1 y vertical n de la matriz de cruce interconectando la entrada 1 con la salida n, la entrada n con la salida 2 y la entrada 2 con la salida 1.

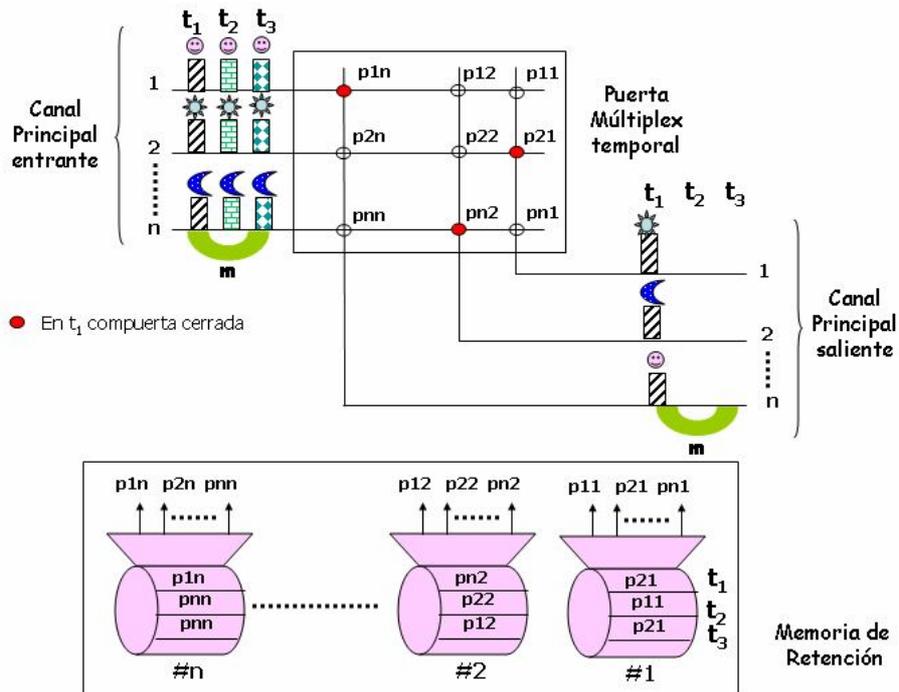


Figura 1.6 Activación de t_1 en memoria de retención del conmutador espacial

En la figura 1.7 y 1.8 con la información en la memoria de retención en los tiempos t_2 y t_3 se producen las conmutaciones mostradas en las graficas

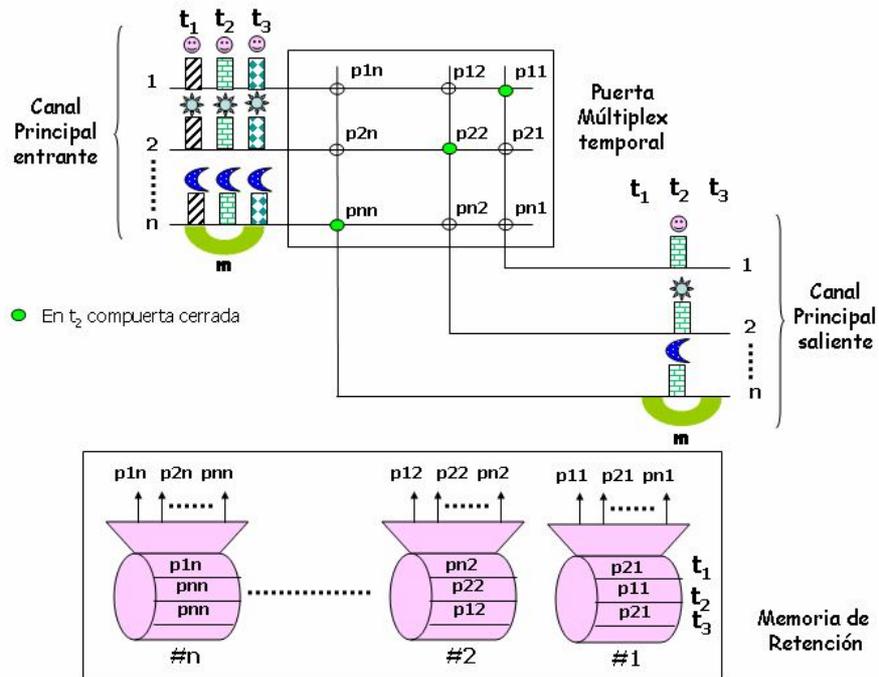


Figura 1.7 Activación de t_2 en memoria de retención del conmutador espacial

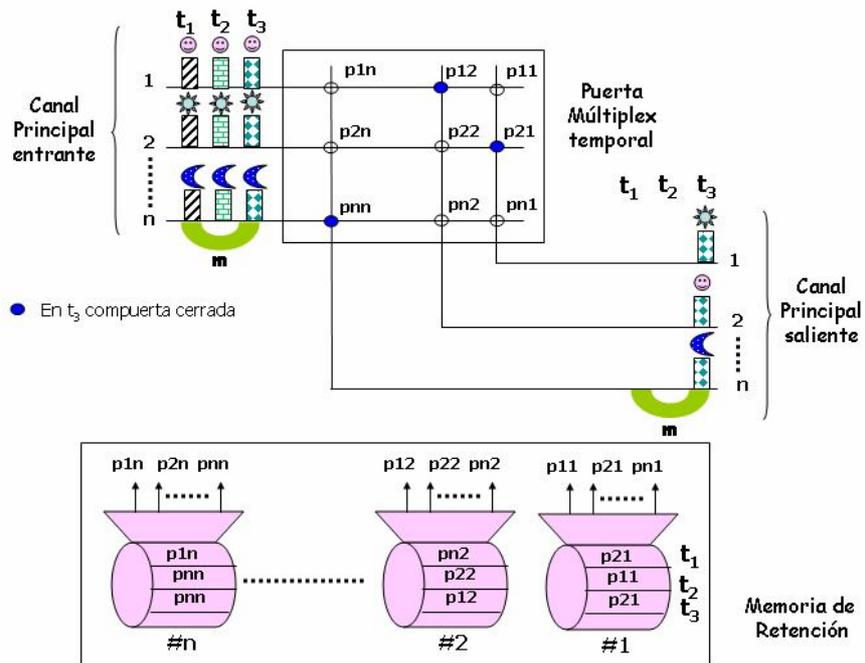


Figura 1.8 Activación de t_3 en memoria de retención del conmutador espacial

Como se puede observar, las conmutaciones se realizan a nivel de enlaces por cada intervalo de tiempo.

1.2.1.2 Conmutador Temporal

El conmutador está constituido básicamente por una memoria intermedia o memoria de conversación y una memoria de control. En la memoria de conversación la lectura y escritura dependen del tipo de control, mientras que en la memoria de control la lectura está comandada por la base de tiempo y la escritura la ordena el procesador dependiendo del número de conexiones nuevas que deban establecerse.

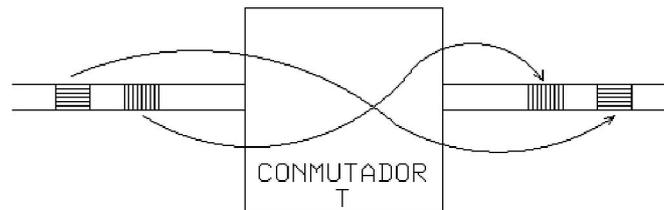


Figura 1.9 Descripción funcional básica del conmutador temporal

El retraso es implementado usando memorias RAM en las cuales se escriben las muestras que van llegando y se leen cuando estas deben ser transferidas a la salida. Para esto se realiza un acceso de escritura a la memoria por cada intervalo de tiempo que ingrese y un acceso de lectura por cada intervalo de tiempo que salga.

El enrutamiento de los intervalos de tiempo de salida se almacena en cada una de las posiciones de la memoria de control.

El efecto del conmutador temporal es el de intercambio de intervalos de tiempo en un mismo enlace

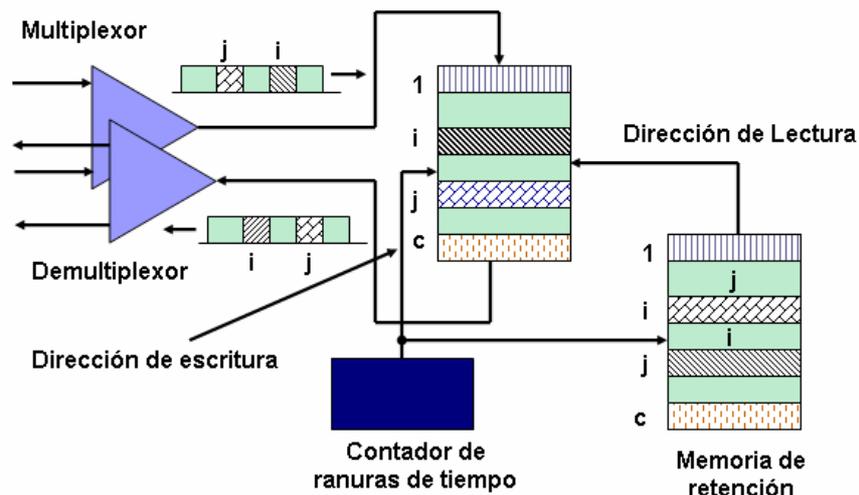


Figura 1.10 Operación funcional básica del conmutador temporal

La arquitectura funcional básica del conmutador temporal se puede ver en la figura 1.10.

Existen básicamente dos maneras en las cuales pueden ser controladas las memorias intermedias:

- Escritura secuencial y lectura aleatoria (control por la salida).
- Escritura aleatoria y lectura secuencial (control por la entrada).

Conmutador Temporal controlado por la Salida (Escritura secuencial y lectura aleatoria)

En este conmutador las muestras van llegando y se van almacenando ordenadamente en la memoria intermedia, mientras que en la memoria de control se encuentran los datos que determinan el camino que debe seguir la muestra, o sea que la lectura de la memoria intermedia es comandada por la memoria de control y la escritura de la memoria de control se realiza de manera cíclica.

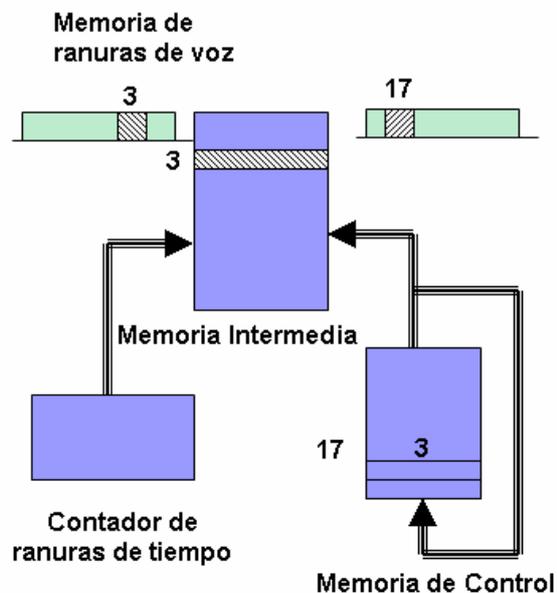


Figura 1.11 Ejemplo de funcionamiento de conmutador temporal controlado por la salida

Del ejemplo de la figura 1.11 podemos decir que la información de la ranura 3 se almacena automáticamente en la tercera posición de la memoria intermedia, A la salida, la información es recobrada desde la memoria de retención que especifica cual dirección se debe acceder para cada ranura de tiempo particular. En el ejemplo, la palabra 17 de la memoria de retención tiene asignado el número 3, lo cual quiere decir que el contenido de la memoria intermedia, en la dirección 3, se transfiere al enlace de salida en la ranura de tiempo 17.

Conmutador Temporal controlado por la Entrada (Escritura aleatoria y lectura secuencial)

Este método es exactamente opuesto al descrito anteriormente. La escritura de la memoria intermedia es comandada por la memoria de control, por lo tanto la lectura de la memoria intermedia es cíclica bajo la supervisión de un contador de intervalos de tiempo a la salida.

Debido a que cuando llega la muestra se decide el camino final que va a tomar, no se permite la difusión de tonos.

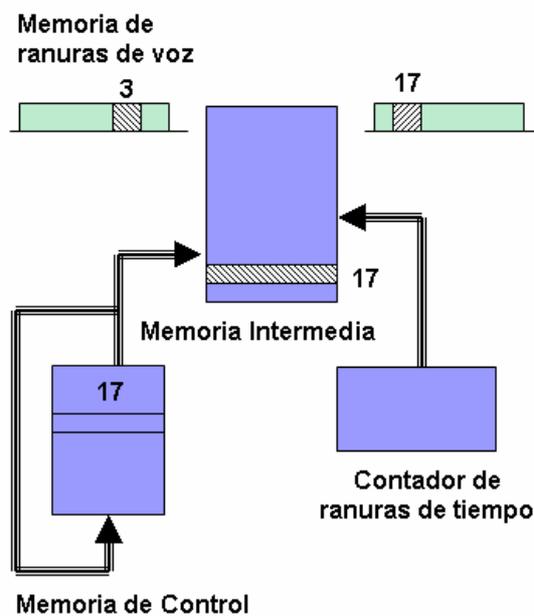


Figura 1.12 Ejemplo de funcionamiento de conmutador temporal controlado por la entrada

Del ejemplo de la figura 1.12 podemos decir que la información contenida en la ranura de tiempo 3, se escribe directamente en dirección 17 de la memoria intermedia ya que en esa dirección fue definida por la memoria de control, de allí se traslada secuencialmente a la ranura numero 17 de la salida.

1.2.2 CONMUTACIÓN DIGITAL

La tecnología digital está constantemente disminuyendo el costo de las soluciones de los problemas de telecomunicaciones. Por razones económicas, se introdujo ampliamente la transmisión MIC (Modulación por impulsos codificados, PCM) durante los años 60.

La transmisión de conversaciones entre las centrales, se logró poniendo voz analógica dentro de paquetes digitales o intervalos de tiempo, siendo

necesaria la conversión nuevamente a analógica con propósitos de conmutación. El desarrollo ha hecho ahora posible y económico conmutar directamente intervalos de tiempo de entrada al intervalo de tiempo de salida requerido.

Con propósitos de enrutamiento, una conversación entrante en un cierto intervalo de tiempo de un sistema MIC necesita conectarse a un circuito de salida, otro determinado intervalo de tiempo en otro sistema MIC. El conmutador digital conmuta una muestra de voz digital en un intervalo de tiempo de entrada al intervalo de tiempo de salida escogido, hacia el siguiente punto en la red telefónica. Ver figura 1.13.

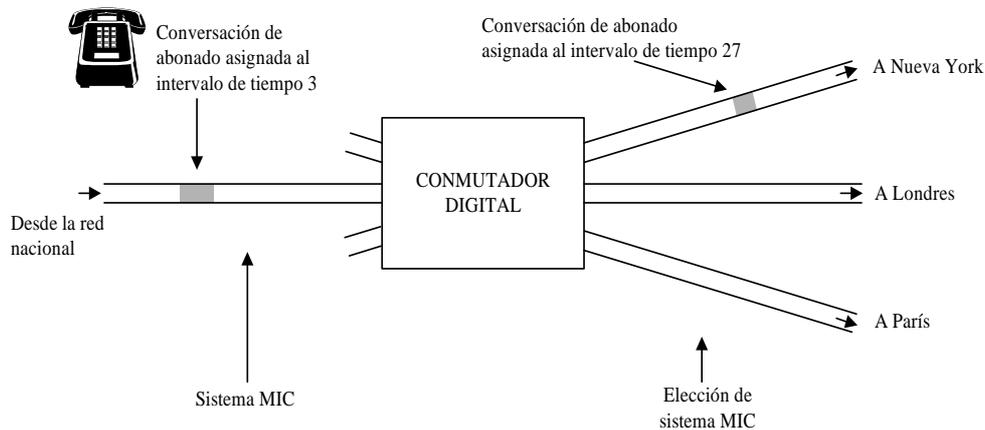


Figura 1.13 El abonado se va a conectar a Nueva York; un intervalo de tiempo libre ha de ser seleccionado en el sistema MIC de salida hacia Nueva York (Por ej. el intervalo de tiempo 27)

1.2.2.1 Bloques Funcionales de una central digital

A continuación se estudiarán los bloques funcionales que hacen parte de una central digital más importantes con sus respectivas funciones.

Conmutador (CX). Es el elemento central del sistema en las implementaciones actuales. En él fluyen los demás módulos, y su función es la de tender las vías digitales para que estos se comuniquen entre sí. La información que fluye a través del conmutador corresponde a voz y datos (pertenecientes a los abonados), señalización (entre la central y los abonados u otras centrales) y control (entre los procesadores que gobiernan los distintos módulos de la central).

Control. Este bloque funcional representa la inteligencia del sistema, la cual, tal como se acaba de sugerir, no está necesariamente localizada en un mismo módulo. La tendencia actual consiste en contribuir el control del sistema hacia los módulos periféricos.

Módulo de Abonados (MA). En el diagrama funcional que ha sido planteado cumple dos funciones básicas: digitalizar la información proveniente de los abonados, tanto voz como señalización, y concentrar su tráfico. Como resultado la conexión con el conmutador se realiza mediante enlaces MIC de 30 o 120 canales, que manejan un alto tráfico. En este módulo se encuentran ubicadas las ILA.

Módulo de Abonados Remoto (MAR). La conexión de MA con el conmutador mediante enlaces MIC, hizo posible la utilización de MAR, que consiste básicamente en un MA, que es sacado del edificio de la central, ubicado en el sector donde están los abonados atendidos, y conectado a la central a través de enlaces MIC. La utilización de MAR significa ahorros para la administración en la adquisición, instalación y mantenimiento de las redes de abonado

Módulo de Troncales Analógicas (MTA). La interconexión de la central digital con otras centrales a través de enlaces analógicos requiere, tal como en el caso de MA, la digitalización de la información proveniente de las troncales. Aquí no hace falta la función de concentración puesto que las líneas troncales ya soportan un alto tráfico.

Módulo de Troncales Digitales (MTD). Cuando la interconexión con otras canales se realiza mediante enlaces también digitales (enlaces MIC), las funciones del módulo que las atiende son muy distintas a la del MTA. En este caso se reducen a la supervisión y sincronización de las señales digitales.

Generador de Tonos (GT). Se tiene un módulo que genera los distintos tonos de señalización de abonado (TIM, TOC, etc.) y de central (MFC). Estos tonos se entregan al conmutador, a través del cual pueden ser conectados hasta el abonado o el enlace que lo requiera.

Receptor de Tonos (RxT). Este módulo es el encargado de recibir e interpretar los diferentes tonos de señalización de abonado (marcación multifrecuencial) y de central (MFC). Las señales provenientes de los abonados o de las troncales llegan a este módulo también a través del conmutador.

A continuación vamos a describir en forma pormenorizada los bloques más importantes:

La red de conmutación digital, los circuitos de línea y la Interfaz de troncal digital.

1.2.2.2 La red de conmutación digital

La red de conmutación digital realiza conmutación entre buses multiplexados en el tiempo. Para que haya conexión entre diferentes intervalos de tiempo en diferentes buses, se requiere tanto conmutación temporal como espacial; conmutación temporal espacial por medio de matrices de puntos cruzados.

Una conexión a través de la red de conmutación implica un intercambio de información entre un canal de entrada y un canal de salida. Este intercambio se logra mediante una cierta secuencia de conmutación temporal y espacial.

Cuando una llamada normal está en progreso durante muchas tramas MIC (en el orden de un millón) esta secuencia deberá repetirse una vez por cada trama MIC durante toda la llamada. Esto requiere un cierto tipo de control cíclico, logrado por memorias de control.

Las diferentes combinaciones de conmutadores espaciales, E, y temporales, T, dan redes de conmutación con diferentes cualidades. Aquí discutiremos la estructura TET (tiempo-espacio-tiempo; Time-space-time, TST) .Para redes muy grandes deben agregarse etapas adicionales, por ejemplo, formando estructuras simétricas EETEE (SSTSS) o TEET (TSST).

Conmutadores TET (TST)

En la figura 1.14 se muestra el modelo de una red TET. Tiene tres buses de entrada y tres buses de salida, cada uno conteniendo a su vez 32 intervalos de tiempo. En la práctica, los números de intervalos de tiempo son mayores, por ejemplo 216 ó 512, logrados después de la multiplexación y de la conversión serial/en paralelo en el terminal de central, pero esto no influye en el principio de funcionamiento ni en el conmutador. Las cifras más altas tienen que ver con la capacidad requerida y la optimización del costo del conmutador.

Para explicar el funcionamiento del conmutador, asumimos, de acuerdo al diagrama, que el sistema de control ordena una conexión entre la entrada A, intervalo de tiempo de entrada asignado No. 2 al bus entrante No. 1, y la salida B, intervalo de tiempo de salida asignado No. 31 al bus saliente No. 3.

Para realizar esto, se necesita encontrar un trayecto libre a través de la red. Esto implica la búsqueda de un intervalo de tiempo interno que esté desocupado, tanto en el lado de entrada A como en el lado de salida B de la matriz espacial. La búsqueda es realizada por el control central. Al hallar el primer intervalo de tiempo vacante, digamos el No. 7, se encuentra este número y las direcciones requeridas, (2, 31 y 1), se envían a las memorias de control MC-A, MC-B, MC-C. Las direcciones se almacenan en la celda No. 7 de estas memorias.

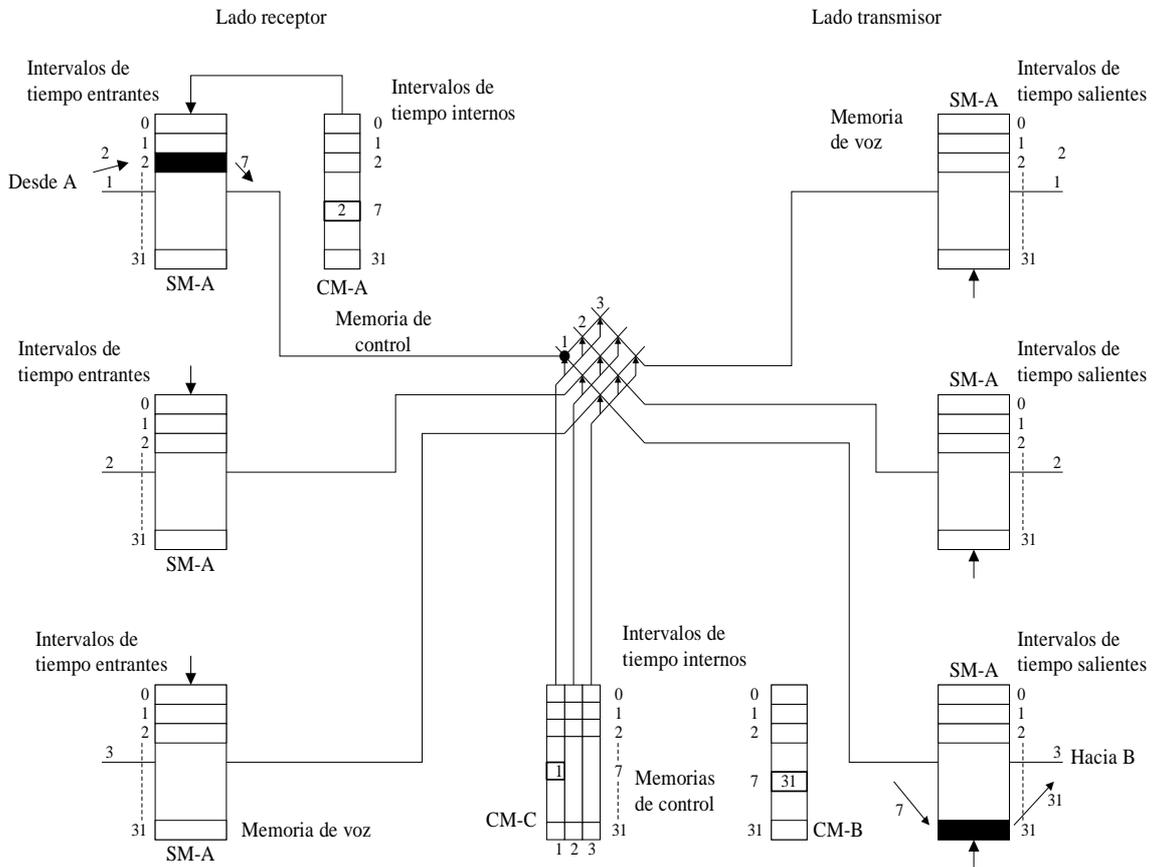


Figura 1.14 Red de conmutación digital, estructura TET

La palabra MIC A se escribe en la celda No. 2 de la MV-A (memoria de voz A) durante el intervalo de tiempo de entrada No. 2. Se almacena allí hasta que llega el intervalo de tiempo interno No. 7. Como los números de intervalos de tiempo de entrada probablemente no coincidan con los números de intervalos de tiempo internos, este tiempo de almacenamiento puede ser cualquiera, desde 0 hasta 31 intervalos de tiempo. Durante el intervalo de tiempo interno No. 7, la dirección hacia la celda No. 2 se envía desde la MC-A y la palabra MIC es leída hacia la matriz espacial. Simultáneamente, la dirección al punto de cruce apropiado (No. 1) se envía desde MC-C y la palabra MIC es conmutada hacia la MV-B requerida. Finalmente, cuando llega el intervalo de tiempo No. 31, la palabra MIC se envía hacia B.

Esta secuencia, que se repite cada trama, crea un trayecto desde A hacia B. Sin embargo, no existe transmisión desde B hacia A, ver figura 1.14. Para lograr esto, se puede hacer uso de dos métodos. Ya sea que el segundo trayecto se establezca completamente independiente del primero o que ambos trayectos se establezcan en coordinación. El primer método tal vez ofrece un sistema más flexible mientras que el segundo método hace posible ahorrar circuitos (hardware) debido al comportamiento simétrico del conmutador. Con el segundo método, la búsqueda de los dos trayectos es una sola, mientras que el primer método requiere dos búsquedas por separado.

Una manera especial de controlar los dos trayectos, el de ida y el de vuelta, es el método antifase. Si se encuentra un trayecto libre de A hacia B, durante un cierto intervalo de tiempo, el trayecto de regreso está garantizado media trama más tarde. Al aplicarlo a nuestro ejemplo, obtenemos el trayecto de ida durante el intervalo de tiempo No. 7 y, consecuentemente el trayecto de regreso durante el intervalo de tiempo $(7+32/2) = 23$. El método se combina con una reducción de la memoria de control. La figura 1.15 ofrece una explicación del método.

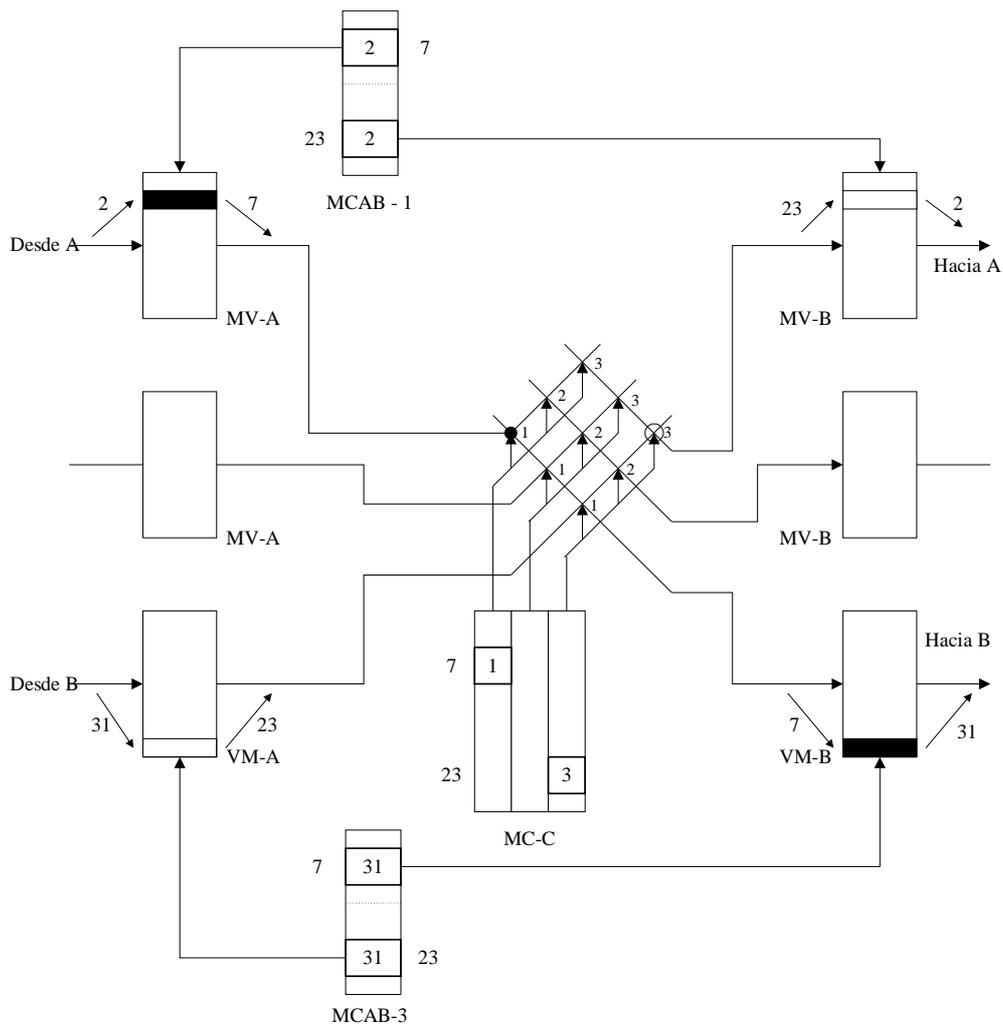
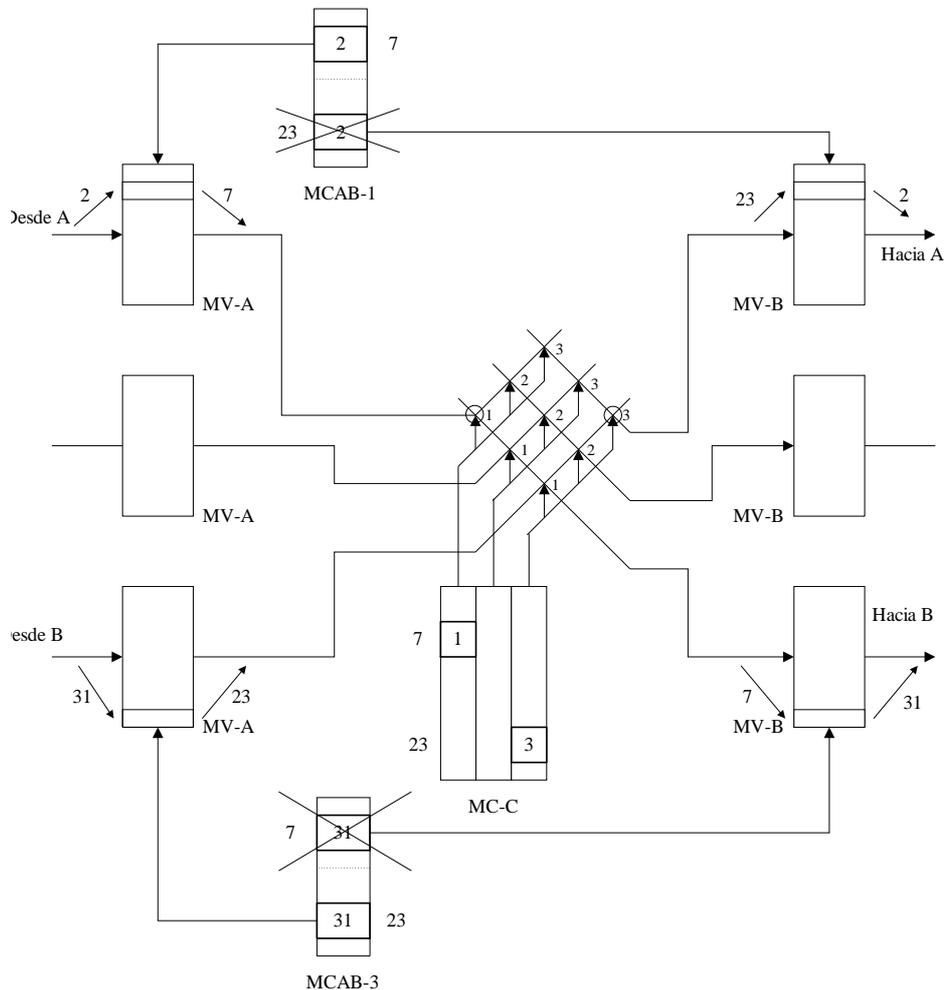


Figura 1.15 Estructura TET con método antifase.

En realidad, la memoria de control no tiene la misma información en dos intervalos de tiempo separados, como en la figura 1.15. Para ahorrar en la memoria de control, el método antifase toma ventaja de la dirección que siempre está desplazada media trama. Ver figura 1.16.



- Dirección A → B (durante el interno TS 7)
- CMAB-1 Leer de SMA en la dirección dada en el interno TS 7 (2).
- CMAB-3 Escribir, vía el conmutador especial, en SMB en la dirección dada en el interno TS $7 + 16 = 23$ (31)
- Dirección B → A (durante el interno TS 23)
- CMAB-3 Leer de SMA en la dirección dada en el interno TS 23 (31)
- CMAB-1 Escribir, vía el conmutador especial, en SMB en la dirección dada en el interno TS $23 + 16 = 7$ (2).

Figura 1.16 Realización TET del método antifase.

Para desconectar la llamada, el control central ordena que se borre la palabra apropiada en las memorias de control.

El bloqueo creado por la red TET depende, obviamente, de que pueda hallarse un par de intervalos de tiempo vacantes para la transmisión entre los dos conmutadores temporales

1.2.2.3 Circuitos de Línea de Abonado

En los sistemas de telecomunicaciones digitales, a las funciones del circuito de abonado o interfaz de línea de abonado se las identifica con la palabra BORSCHT, conformada con las iniciales de los nombres en inglés de cada función así:

Battery:	Alimentación microfónica
Overload:	Protección contra sobrecargas
Ringing:	Conexión de la señal de repique
Supervision:	Supervisión de bucle
Codec:	Codificación y Decodificación MIC
Hybrid:	Conversión de dos a cuatro hilos
Test:	Conexión al equipo de pruebas

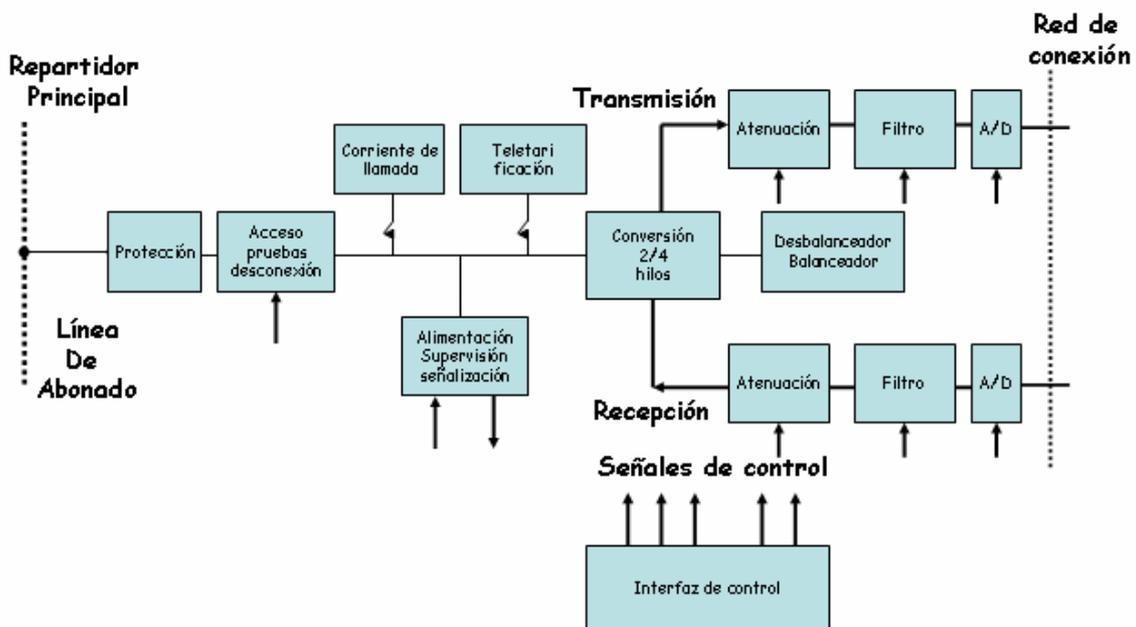


Figura 1.17 Diagrama en bloques de circuito de línea de abonado

1.2.2.4 Interfaz de Troncal Digital

El interfaz de enlace digital es un conjunto de vías de comunicación en las que las señales han sido digitalizadas mediante la técnica PCM y multiplexadas en el tiempo para compartir el mismo medio de transmisión.

En el caso Colombiano se ha adoptado la velocidad básica de 2048 Mbit/s para los enlaces digitales. El funcionamiento y los bloques de la interfaz de troncal digital se pueden ver en la figura 1.18.

En esta memoria, con capacidad para contener varias tramas, se almacena la señal entrante, haciendo uso para esta operación el reloj recuperado de la misma señal (reloj externo).

Simultáneamente, pero con cierto desfase, se extrae la misma señal de la memoria, pero utilizando la lectura del reloj propio de la central.

De esa forma, la memoria absorbe las diferencias de velocidad, obteniéndose una señal digital síncrona con el reloj de la central.

1.3 CONMUTACIÓN DE PAQUETES

Se denomina Conmutación de Paquetes al establecimiento, por parte de una red de comunicaciones, de un intercambio de bloques de información (o “paquetes”) con un tamaño específico entre dos puntos, un emisor y un receptor. En el origen, extremo emisor, la información se divide en “paquetes” a los cuales se les indica la dirección del destinatario. Esto es, cada paquete contiene, además de datos, un encabezado con información de control (prioridad y direcciones de origen y destino).

Los paquetes se transmiten a través de la red y, posteriormente, son reensamblados en el destino obteniendo así el mensaje original. En cada nodo de red, un paquete puede ser almacenado brevemente y encaminado dependiendo de la información de la cabecera. De esta forma, pueden existir múltiples vías o “caminos” de un punto a otro, siendo gestionado por la red el camino óptimo. Las redes basadas en la conmutación de paquetes evitan que mensajes de gran longitud signifiquen grandes intervalos de espera ya que limitan el tamaño de los mensajes transmitidos. La red puede transmitir mensajes de longitud variable pero con una longitud máxima. En la figura 1.19 se puede apreciar una red de conmutación de paquetes.

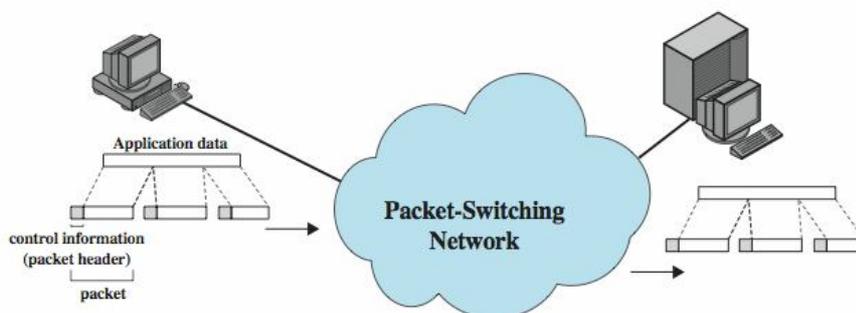


Figura 1.19 Red de conmutación de paquetes (Imagen tomada del libro Comunicación y Redes de Computadoras, Williams Stallings, Octava edición).

La conmutación de paquetes resulta más adecuada para la transmisión de datos comparada con la Conmutación de circuitos.

Su principal ventaja es que únicamente consume recursos del sistema cuando se envía (o se recibe) un paquete, quedando el sistema libre para manejar otros paquetes con otras información o de otros usuarios. Por tanto, la conmutación de paquetes permite inherentemente la compartición de recursos entre usuarios y entre informaciones de tipo y origen distinto. Este es caso de Internet. Su inconveniente reside en las dificultades en el manejo de informaciones de tiempo real, como la voz, es decir, que requieren que los paquetes de datos que la componen lleguen con un retardo apropiado y en el orden requerido. Evidentemente las redes de conmutación de paquetes son capaces de manejar informaciones de tiempo real, pero lo hacen a costa de aumentar su complejidad y sus capacidades.

1.3.1 TÉCNICAS DE CONMUTACIÓN DE PAQUETES

Si una estación tiene que enviar un mensaje de longitud superior a la del tamaño máximo del paquete permitido a través de una red de conmutación de paquetes, fragmenta el mensaje en paquetes y los envía, de uno en uno, hacia la red. La cuestión que surge es como gestiona la red esta secuencia de paquetes para encaminarlos en su seno y entregarlos en el destino deseado. Existen dos aproximaciones usadas en las redes actuales:

- Datagramas
- Circuitos Virtuales

1.3.1.1 Datagramas:

En esta técnica cada paquete se trata de forma independiente, sin referencia alguna a los paquetes anteriores. Esta técnica se muestra en la figura 1.20. Cada nodo elige el siguiente nodo en la ruta del paquete de acuerdo con información recibida de los nodos vecinos acerca de tráfico, fallos en las líneas, etc. De este modo no todos los paquetes, aunque con el mismo destino, seguirán la misma ruta (véase subfigura c), pudiendo recibirse desordenados en el ultimo nodo. En este ejemplo, el nodo final almacena todos los paquetes y los reordena antes de retransmitirlos al destino.

En algunas redes de datagramas es el destino final, en lugar del nodo, el responsable de llevar a cabo la reordenación de los paquetes. También es posible que los paquetes se pierdan en la red; por ejemplo, si un nodo de conmutación de paquetes falla momentáneamente, se perderán todos los paquetes en su cola. De nuevo, será responsabilidad del nodo final o del destino detectar la pérdida de un paquete y decidir como recuperarlo. En esta técnica cada paquete se denomina datagrama y se trata de forma independiente del resto.

Ventajas:

- Considerado el método más sensible.
- No tiene fase de establecimiento de llamada.
- El paso de datos es más seguro.
- No todos los paquetes siguen una misma ruta.
- Los paquetes pueden llegar al destino en desorden debido a que su tratamiento es independiente.
- Un paquete se puede destruir en el camino, cuya recuperación es responsabilidad de la estación de destino (esto da a entender que el resto de paquetes están intactos).

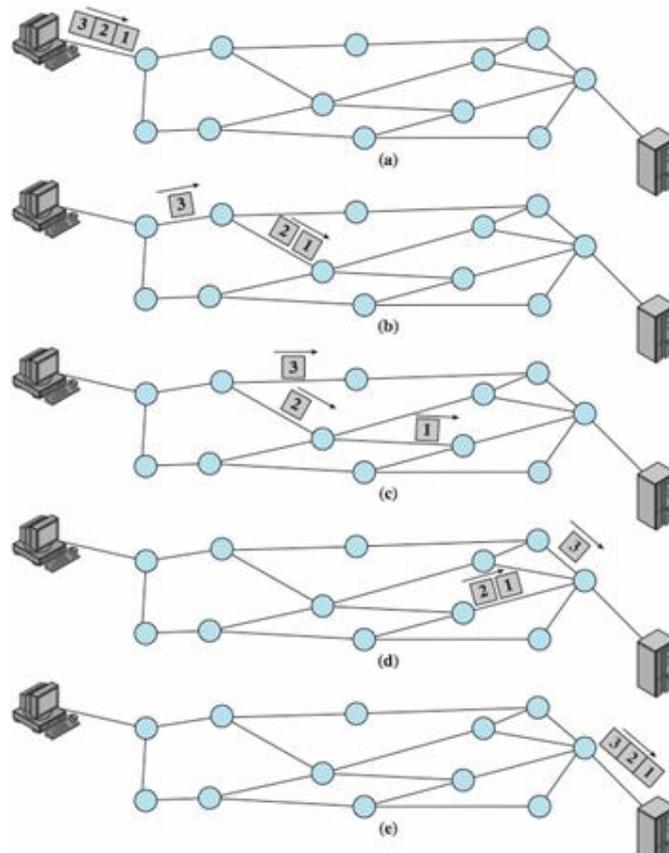


Figura 1.20 Conmutación de paquetes mediante datagramas (Imagen tomada del libro *Comunicación y Redes de Computadoras*, Williams Stallings, Octava edición).

1.3.1.2 Circuitos Virtuales:

En esta técnica se establece una ruta previa al envío de los paquetes. Una vez establecida esta, todos los paquetes intercambiados entre dos partes comunicantes siguen dicho camino a través de la red. Esto se ilustra en la figura 1.21. Dado que el camino es fijo mientras dure la conexión lógica, este es similar a un circuito en redes de conmutación de circuitos, por lo que se le llama circuito virtual. Además de los datos, cada paquete contiene un identificador de circuito virtual. Cada nodo de la ruta preestablecida sabe hacia donde dirigir los paquetes, no precisándose por tanto la toma de decisiones de

encaminamiento. En un instante de tiempo dado, cada estación puede tener más de un circuito virtual hacia otra u otras estaciones. La principal característica de la técnica de circuitos virtuales es que la ruta entre las estaciones se establece antes de la transferencia de los datos. Obsérvese que esto no significa que sea una ruta dedicada como en el caso de la conmutación de circuitos.

Ventajas:

- Son los más usados.
- Su funcionamiento es similar al de redes de conmutación de circuitos.
- Previo a la transmisión se establece la ruta previa a la transmisión de los paquetes por medio de paquetes de petición de llamada (pide una conexión lógica al destino) y de llamada aceptada (en caso de que la estación destino esté apta para la transmisión envía este tipo de paquete); establecida la transmisión, se da el intercambio de datos, y una vez terminado, se presenta el paquete de petición de liberación (aviso de que la red está disponible, es decir que la transmisión ha llegado a su fin).
- Cada paquete tiene un identificador de circuito virtual en lugar de la dirección del destino.
- Los paquetes se recibirán en el mismo orden en que fueron enviados.
- Posee control de errores que asegura que los paquetes no solo lleguen en orden sino que además sean correctos.

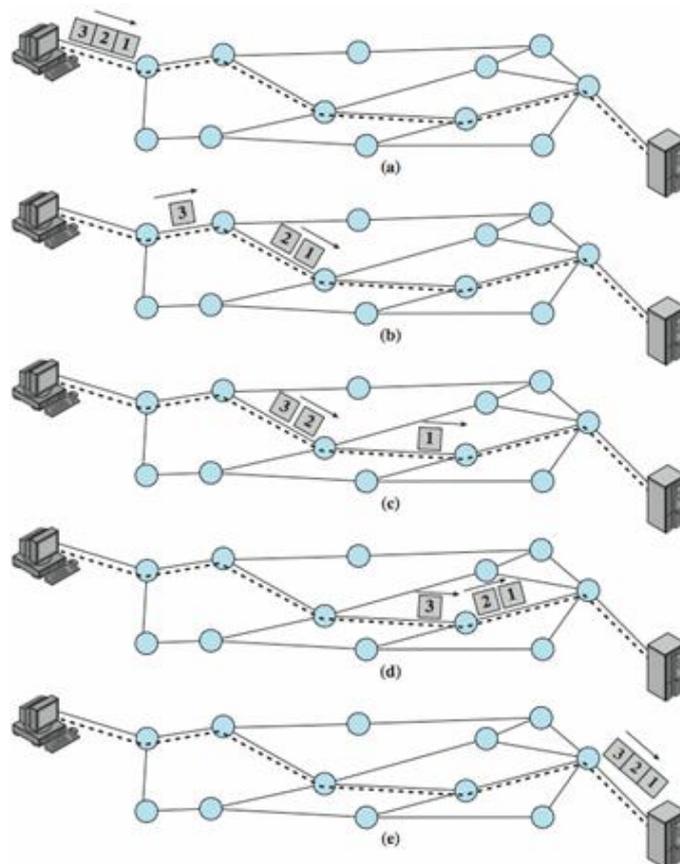


Figura 1.21 Conmutación de paquetes mediante circuitos virtuales (Imagen tomada del libro *Comunicación y Redes de Computadoras*, Williams Stallings, Octava edición).

1.3.2 COMPARACIONES ENTRE LAS TÉCNICAS DE CONMUTACIÓN DE CIRCUITOS Y DE PAQUETES

Una vez visto el funcionamiento interno de la técnica de conmutación de paquetes, a continuación se realizara una comparación de ella con la conmutación de circuitos. Se estudiarán las prestaciones y luego otras características.

En la figura 1.22 se muestra el retardo total sufrido cuando un mensaje dado se transmite por una ruta que consta de dos elementos intermedio. En primer lugar el mensaje debe atravesar el enlace que conecta el origen con el primer conmutador, suponiéndose que el enlace tiene un retardo de propagación de p segundos y que el mensaje conlleva un tiempo de transmisión de T segundos.

Seguidamente el mensaje atraviesa el enlace que conecta los dos conmutadores. Suponiendo por simplicidad que el retardo de propagación y la velocidad de las líneas de transmisión tienen el mismo valor. En consecuencia el retardo extremo a extremo mínimo para el mensaje será $3p+3T$. Obsérvese que este retardo no tiene en cuenta ni los retardos de cola que puedan aparecer en los distintos enlaces debido a la espera previa a la retransmisión de los mensajes, ni el tiempo consumido en la comprobación de los errores ni las retransmisiones asociadas.

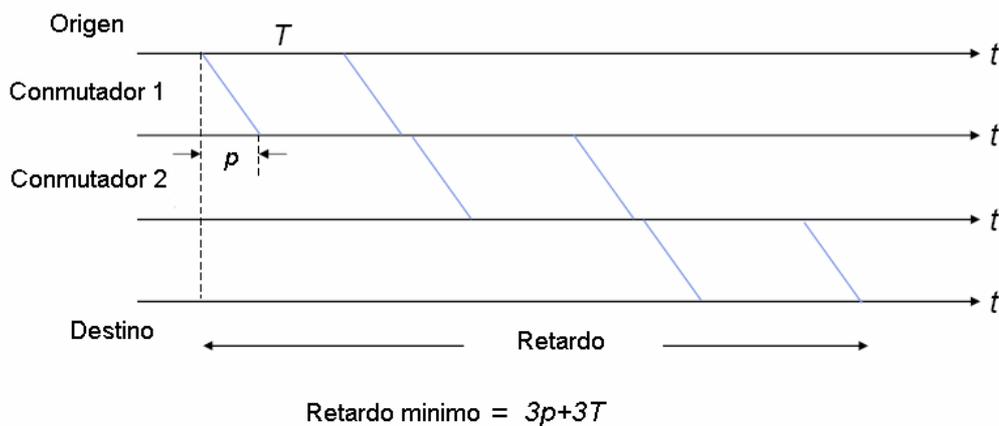


Figura 1.22 Retardos en la conmutación de mensajes

En la figura 1.23 se muestra el retardo sufrido por un mensaje que se ha dividido en tres paquetes distintos. Aquí se ha supuesto que los tres paquetes siguen el mismo camino y que se transmiten en orden secuencial. Se han despreciado además los datos suplementarios correspondientes a las cabeceras y se ha supuesto que cada paquete precisa $P=T/3$ para su transmisión. Los tres paquetes se transmiten sucesivamente desde el origen hasta el primer conmutador de paquetes.

El primer paquete en la figura 1.23 se recibe en el primer conmutador tras $p+P$ segundos. Suponiendo que el paquete se recibe correctamente, se puede

comenzar a transmitir sobre el siguiente salto tras un breve periodo de procesamiento. El primer paquete se recibirá así en el segundo conmutador en el instante de tiempo $2p+2P$. Como antes se supone que el paquete comienza a transmitirse por el ultimo salto tras un breve periodo de tiempo de procesamiento; se reciben por tanto, el primer paquete en el ultimo enlace en el instante de tiempo $3p+3P$.

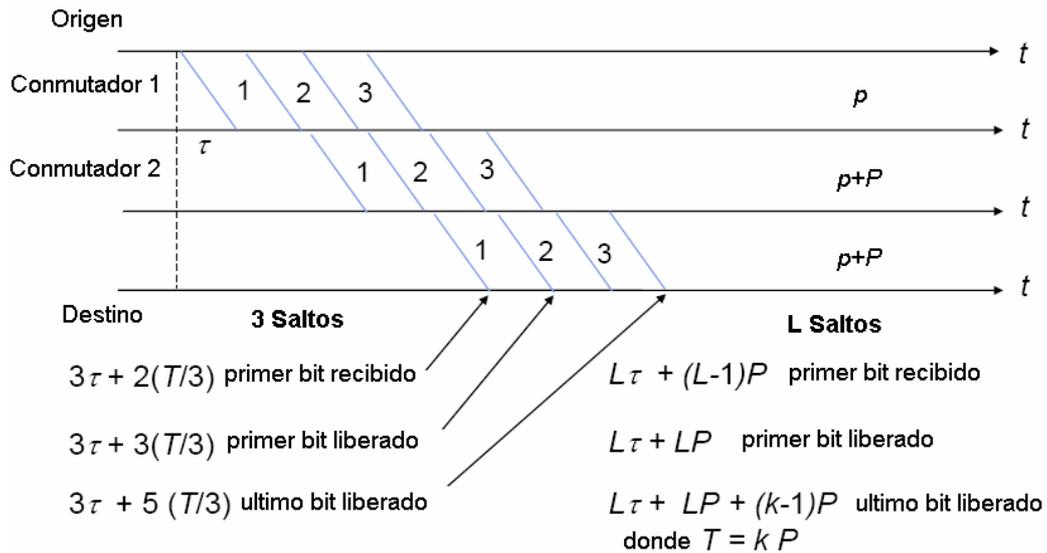


Figura 1.23 Retardos en la conmutación de paquetes

Como se muestra en la figura, los siguientes paquetes atraviesan la red inmediatamente después del primero. Si no se producen errores de transmisión, el ultimo paquete se recibirá en el destino en el instante de tiempo $3p+3P+2P=3p+5P=3p+T+2P$, que es menor que el retardo sufrido en el ejemplo de la conmutación de mensajes mostrado en la figura 1.22. En general, si el camino que sigue la secuencia de paquetes consta de L saltos con iguales valores del retardo de propagación y de la velocidad de transmisión, el retardo total sufrido por un mensaje de k paquetes esta dado por:

$$Lp+LP+(k-1)P$$

Por el contrario, el retardo sufrido cuando se emplea la técnica de conmutación de mensajes es:

$$Lp+LT=Lp+L(kP)$$

Por tanto la conmutación de mensajes implica un retardo adicional igual a $(L-1)(k-1)P$. Obsérvese que en los valores de retardo anteriores se han despreciado los tiempos de puesta en cola y de procesamiento en cada uno de los saltos en la red.

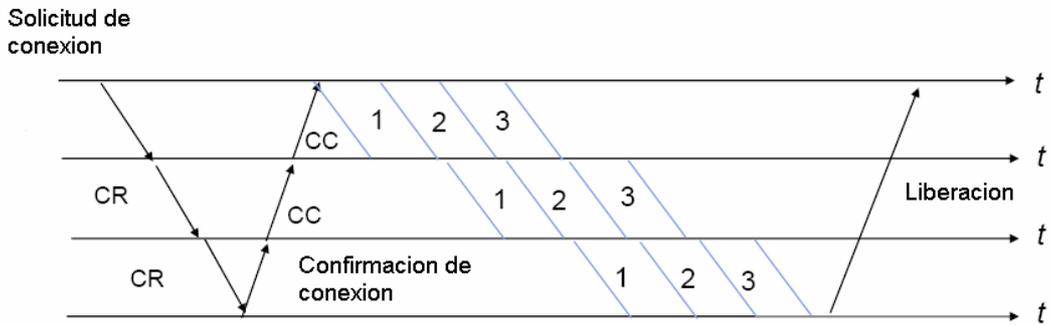


Figura 1.24 Retardos en la conmutación de paquetes mediante circuitos virtuales

En la figura 1.25 se muestra una comparación sencilla entre la conmutación de circuitos y las dos formas de conmutación de paquetes. Esta figura ilustra la transmisión de un mensaje a través de cuatro nodos, desde una estación emisora conectada al nodo 1 hasta una estación de destino conectada al nodo 4. En esta figura se indican tres tipos de retardo:

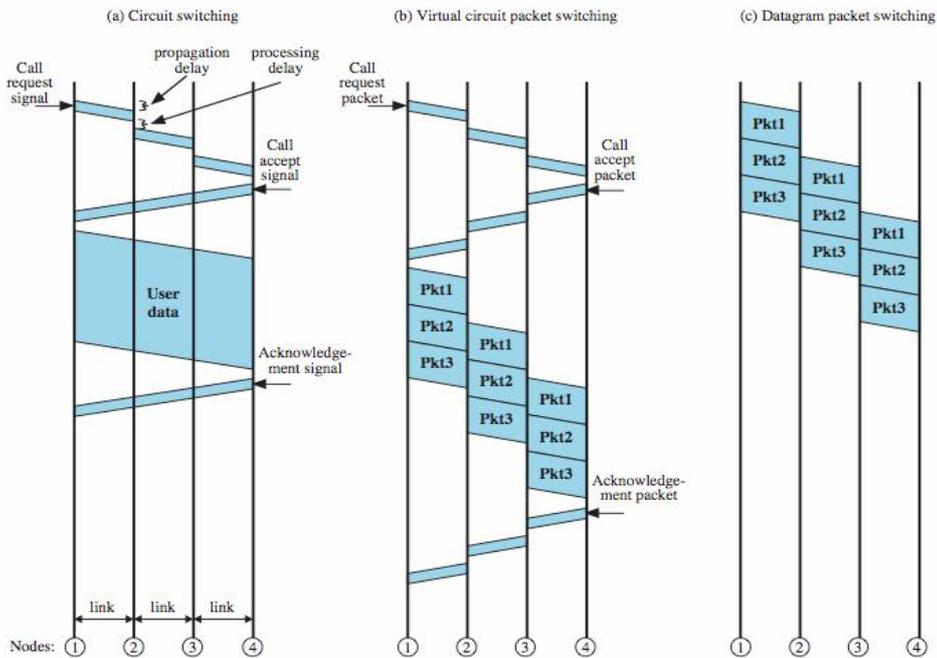


Figura 1.25 Eventos temporales en conmutación de circuitos y en conmutación de paquetes (Imagen tomada del libro *Comunicación y Redes de Computadoras*, Williams Stallings, Octava edición).

- **Retardo de propagación:** tiempo que tarda la señal en propagarse desde un nodo hasta el siguiente. Este tiempo es generalmente despreciable, ya que la velocidad de las señales electromagnéticas a través de un cable, por ejemplo, es generalmente de 2×10^8 m/s.

- **Tiempo de transmisión:** tiempo que tarda un transmisor en enviar un bloque de datos. Por ejemplo en una línea de 10 kbps se tarda 1 segundo en transmitir un bloque de datos de 10.000 bits.
- **Retardo de Nodo:** tiempo que tarda un nodo en realizar los procesos necesarios para la conmutación de los datos.

En conmutación de circuitos existe un cierto retardo antes de que se pueda enviar el mensaje. Primero se envía a través de la red una señal “Petición de Llamada” (Call Request) para establecer una conexión con el destino.

Si la estación de destino no está ocupada, devuelve una señal “Llamada Aceptada” (Call Accepted).

Obsérvese la aparición de un retardo de procesamiento en cada nodo durante la solicitud de la llamada debido a la necesidad de establecer la ruta para la conexión. A la vuelta lo se requiere procesamiento dado que la conexión ya está establecida. Una vez establecida la conexión, el mensaje se envía como un único bloque, sin retardos en los nodos de conmutación.

La técnica de conmutación de paquetes mediante circuitos virtuales parece muy similar a la de conmutación de circuitos. Un circuito virtual se solicita mediante el uso de un paquete “Petición de Llamada” (Call Request), el cual sufre un retardo en cada nodo. El circuito virtual se acepta mediante un paquete “aceptación de Llamada” (Call Accept). Al contrario que en el caso de conmutación de circuitos, la aceptación de llamada también experimenta retardos en los nodos aunque la ruta del circuito virtual se encuentre ya establecida. La razón de ello es que el paquete se pone en cola en cada nodo y debe esperar turno para su transmisión.

La técnica de conmutación de paquetes mediante datagramas no precisa un establecimiento de llamada, de modo que para mensajes cortos resulta más rápida que la conmutación de paquetes mediante circuitos virtuales y, quizá, que la conmutación de circuitos. Sin embargo, dado que cada datagrama individual se encamina de forma independiente, el procesamiento de cada uno de ellos en cada nodo puede llegar a ser superior que en el caso de los circuitos virtuales. Por tanto, para mencionar mensajes grandes, la técnica de circuitos virtuales puede ser mejor.

A partir de la figura 1.25 se pueden comprender aproximadamente las prestaciones relativas de las distintas técnicas. Las prestaciones reales dependen de varios factores como el tamaño de la red, su topología, la carga y las características típicas de intercambios.

1.3.3 ALGORITMOS DEL CAMINO MÁS CORTO

Prácticamente todas las redes de conmutación de paquetes y todas las redes de tipo Internet basan sus decisiones de encaminamiento en algún criterio de mínimo costo. Si el criterio consiste en minimizar el número de saltos, cada enlace tendrá asociado un valor igual a 1. Normalmente, el valor asociado al enlace es inversamente proporcional a su capacidad, proporcional a su carga actual o alguna combinación de ellos. En cualquier caso, el costo de las líneas se emplea como entrada a un algoritmo de encaminamiento de mínimo costo, que establece que:

Dada una red de nodos conectados entre si por enlaces bidireccionales, donde cada enlace tiene un costo asociado en cada sentido, se define el costo asociado de una ruta entre dos nodos como la suma de los costos de los enlaces atravesados. Así, para cada par de nodos se obtiene el mínimo camino de mínimo costo.

Se puede hacer uso de varias métricas para asignar un costo a los enlaces en función del parámetro que se desee optimizar. Algunos ejemplos son:

- **Costo inversamente proporcional a capacidad:** se asignan costos superiores a los enlaces de menor capacidad. El objetivo será así el envío de un paquete dado a través de la ruta con capacidad superior. Si los enlaces tuviesen la misma capacidad, el camino más corto será el que presente menor número de saltos.
- **Costo proporcional a retardo de paquete:** el cual incluye el retardo de cola en la memoria temporal de un conmutador y el retardo de propagación en el enlace. El camino más corto corresponde al más rápido en alcanzar el destino.
- **Costo proporcional a congestión:** un ejemplo es carga de tráfico. En este caso, el camino más corto trata de evitar enlaces congestionados.

La mayor parte de los algoritmos de encaminamiento de costo utilizados en las redes de conmutación de paquetes y en las redes de Internet son variantes de uno de los dos algoritmos comunes: el de DIJKSTRA y el de BELLMAN-FORD los cuales se verán a continuación.

1.3.3.1 ALGORITMO DE BELLMAN-FORD

El algoritmo de BELLMAN-FORD se puede enunciar así: encontrar los caminos más cortos desde un nodo origen dado con la condición de que estos contengan a lo sumo un enlace; a continuación encontrar los caminos más

cortos con la condición de que contengan dos enlaces como máximo, y así sucesivamente. Se fija primero el nodo de destino. Se define D_j como la estimación actual del costo mínimo (o distancia mínima) entre el nodo j y el destino, y sea C_{ij} el costo del enlace entre los nodos i y j . el costo del enlace del nodo i así mismo se fija a 0 ($C_{ii}=0$), y el costo entre los nodos i y k es infinito si estos no se encuentran conectados. Definido el nodo destino el algoritmo de BELLMAN-FORD de puede resumir como sigue:

1. Inicialización:

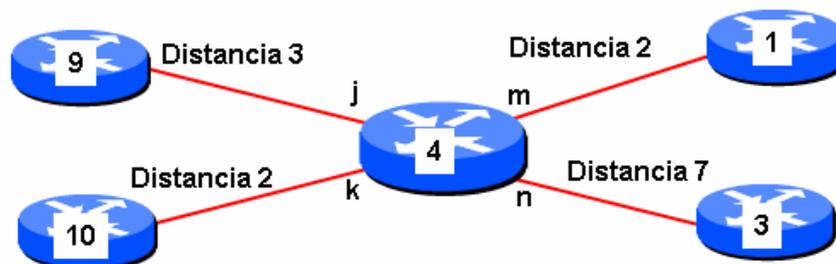
$$D_i = \infty, \forall i \neq d$$

$$D_d = 0$$

2. Actualización: para cada $i \neq d$

$$D_i = \min_j \{C_{ij} + D_j\}, \forall j \neq 1$$

Repetir el paso 2 hasta que no se produzcan mas cambios en la iteración



Destino:	1	2	3	4	5	6	7	8	9	10	11
Recibido de j (+3):	12	3	15	3	12	5	6	18	0	7	15
Recibido de k (+2):	5	8	3	2	10	7	4	20	5	0	15
Recibido de m (+2):	0	5	3	2	19	9	5	22	2	4	7
Recibido de n (+7):	6	2	0	7	8	5	8	12	11	3	2
Distancia mínima:	2	6	5	0	12	8	6	19	3	2	9
Interfaz de salida:	m	j	m	0	k	j	k	n	j	k	n

Figura 1.26 Ejecución del algoritmo de BELLMAN-FORD

En este ejemplo el router 4 recibe de sus cuatro vecinos los correspondientes vectores distancia. Obsérvese que el router 4 no conoce quienes son sus vecinos, aunque puede deducirlo viendo que elemento de cada uno de los vectores recibidos tiene una distancia cero.

Para cada columna (cada router) se calcula la distancia desde el router 4 sumando a la distancia recibida la que el router conoce al siguiente router por esa interfaz. En cada caso se elige la distancia más corta. Por ejemplo las distancias al router 1 son de 15, 7, 2 y 13 a través de j, k, m y n respectivamente.

Por tanto nuestro router elegirá la interfaz m que permite llegar a 1 con una distancia de 2 únicamente (casualmente el router vecino por la interfaz m es 1, pero esto es algo irrelevante desde el punto de vista del algoritmo). De la misma forma procede el router 4 con el resto de columnas, utilizando en cada caso la información recibida de todos sus vecinos y construyendo a partir de ella su propio vector distancia. Una vez termina el cálculo el router distribuye el vector calculado a sus cuatro vecinos.

En el cálculo de rutas óptimas puede producirse empate de métrica entre dos rutas. Este es el caso de la ruta al router 3 en la figura, que se encuentra a una métrica 5 tanto por la interfaz 'm' como por la 'k'. En estos casos ya depende del protocolo de ruteo (no del algoritmo) que se utilicen todas las rutas de igual métrica o solo una de ellas.

1.3.3.2 ALGORITMO DE DIJKSTRA

El algoritmo de DIJKSTRA es un algoritmo alternativo para determinar los caminos más cortos desde un nodo origen hasta todos los otros nodos en una red. Este algoritmo es generalmente más eficiente que el de Bellman-Ford, pero requiere que todos los costos de enlace sean positivos, lo cual, por suerte sucede en las redes de comunicaciones.

La idea principal de este algoritmo consiste en identificar los nodos más cercanos desde el nodo origen en orden creciente del costo del camino, el algoritmo es iterativo. En la primera iteración el algoritmo obtiene el nodo más cercano desde el nodo origen, que debe ser el vecino del nodo origen si los costos de los enlaces son positivos.

En la segunda iteración, el algoritmo obtiene el segundo nodo más cercano desde el nodo origen, este nodo debe ser el vecino o bien del nodo origen o bien del nodo más cercano a este; en caso contrario, debe existir uno más cercano. En la tercera iteración, el tercer nodo más cercano debe ser el vecino de los dos primeros nodos más cercanos, y así sucesivamente. Por tanto en la iteración k el algoritmo habrá obtenido los k , nodos más cercanos desde el origen.

El algoritmo se puede implementar a través del mantenimiento de un conjunto de nodos N nodos permanentemente etiquetados, los cuales corresponden a los de los caminos más cortos. En cada iteración se añade el nodo siguiente

más cercano al conjunto N hasta que todos los nodos se encuentren incluidos en la lista. Para formalizar el algoritmo, definamos D_i como el costo mínimo actual desde el nodo origen (etiquetado como s) al nodo i . El algoritmo de DIJKSTRA se puede describir como sigue:

1. Inicialización:

$$\begin{aligned} N &= \{s\} \\ D_j &= C_{sj}, \forall j \neq s \\ D_s &= 0 \end{aligned}$$

2. Búsqueda del siguiente nodo mas cercano: encontrar $i \notin N$ tal que:

$$D_i = \min_{j \notin N} D_j$$

- Añadir i a N .
- Finalizar si N contiene todos los nodos

3. Actualización de los costos mínimos: para cada nodo $j \notin N$:

$$D_j = \min \{D_j, D_i + C_{ij} \}$$

Ir al paso 2.

N = conjunto de nodos de la red.

s = nodo origen.

T = lista o conjunto de nodos añadidos o incorporados por el algoritmo.

$L(n)$ = costo en curso obtenido por el algoritmo para el camino de mínimo costo del nodo s ; al finalizar el algoritmo, este costo corresponde al del camino de mínimo costo de s a n en el grafo.

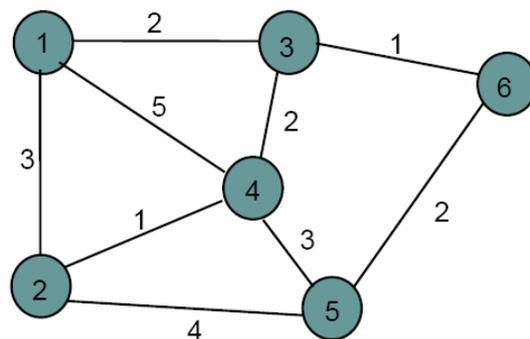


Figura 1.27 Ejemplo de red con los costos de los enlaces asociados

En la tabla 1.1 se muestra la ejecución del algoritmo de DIJKSTRA basado en la figura 1.27 al final de la inicialización y en cada iteración. Para cada iteración se subraya el valor del mínimo costo del siguiente nodo más cercano. En caso de empate, se elige aleatoriamente el nodo más cercano. El costo mínimo para los nodos que no se encuentren etiquetados permanentemente se actualiza secuencialmente. La última fila indica el costo mínimo para cada uno de los nodos.

Si en cada iteración se recuerda el nodo predecesor del siguiente nodo más cercano, se puede obtener, como se muestra en la figura 1.28 un árbol del camino mas corto cuyo nodo raíz es el nodo 1. Cuando el algoritmo finaliza, se conoce el costo mínimo a cada nodo y el nodo siguiente es el camino mas corto.

Iteración	T	L(2) Ruta	L(3) Ruta	L(4) Ruta	L(5) Ruta	L(6) Ruta
1	{1}	3 1-2	2 1-3	5 1-4	∞	∞
2	{1,3}	3 1-2	2 1-3	4 1-3-4	∞	3 1-3-6
3	{1,2,3}	3 1-2	2 1-3	4 1-3-4	7 1-2-5	3 1-3-6
4	{1,2,3,6}	3 1-2	2 1-3	4 1-3-4	5 1-3-6-5	3 1-3-6
5	{1,2,3,4,6}	3 1-2	2 1-3	4 1-3-4	5 1-3-6-5	3 1-3-6
6	{1,2,3,4,5,6}	3 1-2	2 1-3	4 1-3-4	5 1-3-6-5	3 1-3-6

Tabla 1.1 Ejecución del algoritmo de DIJKSTRA

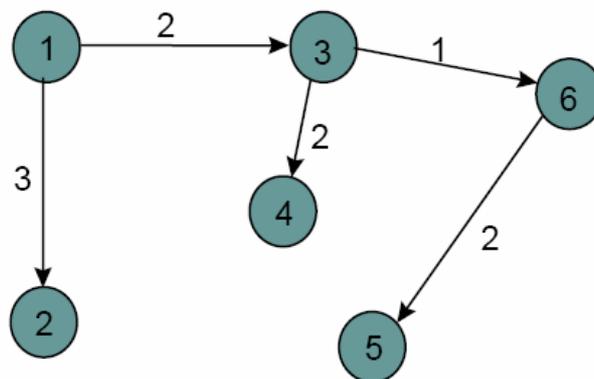
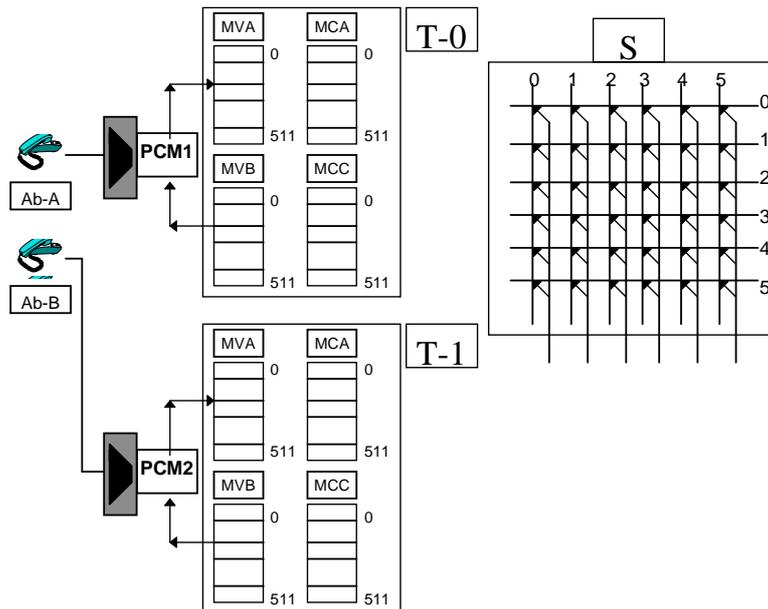


Figura 1.28 Árbol del camino mas corto desde el nodo 1 a otros nodos

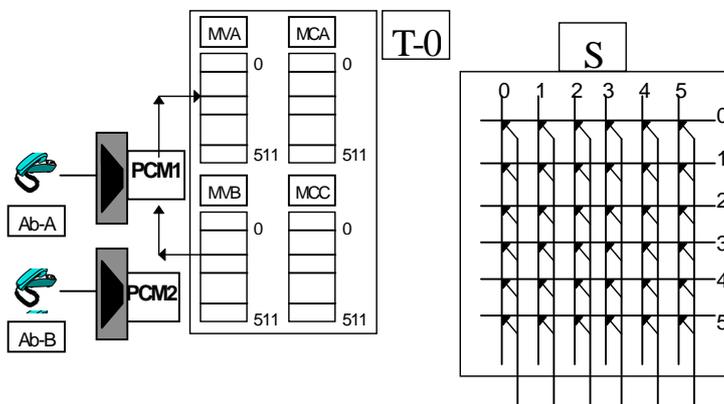
Pero el tránsito en la red es mucho menor. Esta técnica también libera de cálculos para seleccionar el encaminamiento.

EJERCICIOS PROPUESTOS

1. En el conmutador T-S-T de la figura debe conectarse el abonado A, conectado por su concentrador de abonado al time slot 5 del PCM 3 del conmutador T-0, con el abonado B, conectado por su concentrador en el time slot 17 del PCM 9 del conmutador T-1. El sistema asigna el intervalo interno 45 para la dirección $A \Rightarrow B$. Describir el contenido de las memorias de control de cada módulo T y la transferencia de las muestras de voz en ambos sentidos.



2. En el conmutador T-S-T de la figura debe conectarse el abonado A, conectado por su concentrador de abonado al time slot 5 del PCM 3 del conmutador T-0, con el abonado B, conectado por su concentrador en el time slot 9 del PCM 15 del conmutador T-0. El sistema asigna el intervalo interno 37 para la dirección $A \Rightarrow B$. Describir el contenido de las memorias de control de cada módulo T y la transferencia de las muestras de voz en ambos sentidos.



UNIDAD II

SISTEMA DE SEÑALIZACIÓN SS7

INTRODUCCIÓN

En este capítulo se hará énfasis en el sistema de señalización SS7 que se utiliza para implementar protocolos de señalización en redes de telecomunicaciones para soporte de redes fijas y celulares.

Se estudiarán aspectos importantes en su estructura, así como sus puntos de señalización los cuales son los puntos de servicio de conmutación, puntos de transferencia de señal y puntos de control de servicio, y las funciones que cada uno de estos desempeñan.

La jerarquía es otro aspecto importante en las redes de señalización SS7 que se verá en este capítulo, así como los protocolos que hacen parte de la red, en donde haremos énfasis en sus estructuras y papeles que desempeñan dentro de los niveles de la jerarquía SS7.

Se analizarán los diferentes pasos que se llevan a cabo para el establecimiento de una llamada detallando el proceso ocurrido al enlazar dos centrales telefónicas, además de observar el método para realizar una petición a una base de datos por medio del proceso de llamada gratuita.

Se detallará los diferentes niveles del protocolo de la parte de transferencia de mensajes, los cuales poseen diferentes funciones dentro del sistema, por otra parte se observarán los parámetros que componen una unidad de señalización y los diferentes métodos para las correcciones de errores.

Al final de este capítulo se hará un breve resumen de los protocolos de la parte de control de conexión de señalización y parte de aplicación de las capacidades de transacción.

2.1 INTRODUCCIÓN AL SISTEMA DE SEÑALIZACIÓN SS7

Es un estándar definido por la International Telecommunication Union (ITU). Este estándar define los procedimientos y protocolos con los que los elementos de la red de conmutación telefónica pueden intercambiar información en una red digital de señalización para el establecimiento, ruteo y control de llamadas. En este sistema la señalización se da fuera de banda (en canales dedicados) en vez de ir en banda (en los canales de voz).

2.2 ESTRUCTURA DEL SISTEMA SS7

A continuación se verán aspectos importantes de la estructura SS7 como sus puntos de señalización, su jerarquía la cual consiste de los niveles de la red SS7, los protocolos donde se definirán su función y posiciones en la jerarquía.

2.2.1 Puntos de señalización

Las entidades que se conectan a una red SS7 están clasificadas de la manera siguiente:

- Puntos de servicio de conmutación (SSP)
- Puntos de transferencia de señal (STP)
- Puntos de control de servicio (SCP)

La figura 2.1 representa gráficamente esas entidades:



*Figura 2.1 Puntos de señalización SS7
(Imagen tomada del tutorial <http://www.iec.org>).*

2.2.1.1 Puntos de servicio de conmutación (SSP)

Los puntos de conmutación de servicios (SSP) representan conmutadores telefónicos que incorporan la funcionalidad de SS7 y enlaces terminales de señalización. Son estos los que originan, terminan o conmutan las llamadas. Estos procesan tráfico de voz de banda (voz, fax, módem, etc.) que incorpora la señalización SS7. El SSP generalmente está integrado en el mismo conmutador de voz, mediante módulos con tarjetas que incorporan

microprocesadores cargados con el programa SS7. Los SSP también inician las consultas a las bases de datos del proveedor, inicia acciones en SSP remotos y realizan acciones en respuesta a las solicitudes de los SSP y otras entidades de señalización.

2.2.1.2 Puntos de transferencia de señal (STP)

Los puntos de transferencia de señal (STP) son parte importante de la arquitectura SS7, representan los conmutadores de paquetes de la red, son los encargados de recibir y dirigir los mensajes de señalización hacia el destinatario correcto por lo que llevan a cabo funciones de ruteo.

Cuando el STP recibe un mensaje procedente de un SSP, el STP verifica el destino del mensaje y de no ser para el, elige, a partir de sus tablas de ruteo, el punto destinatario de señalización y el enlace a través del cual se enviara el mensaje a este nodo. En conclusión cuando un punto de señalización ejecuta las funciones de un STP nunca envía los mensajes a las capas de aplicación, es decir, no procesa los mensajes en el nivel 4.

Los únicos mensajes que no son transferidos están relacionados con la gestión de la red y funciones de la red Inteligente que veremos más adelante.

2.2.1.3 Punto de control del servicio (SCP)

Un punto de control de servicio actúa como una interfaz entre las bases de datos de telecomunicaciones y la red SS7, ofrecen una lógica complementaria, utilizada para ofrecer servicios adicionales en las denominadas redes inteligentes. Básicamente estos puntos se tratan de base de datos que proveen características avanzadas como por ejemplo, servicios a números gratuitos 1-8000, o en la localización de líneas celulares, etc. Para poder utilizar estos servicios el SSP envía un mensaje al SCP solicitando instrucciones.

Para que la arquitectura SS7 sea robusta, la red deberá diseñarse de tal forma que ofrezca un alto grado de redundancia. De esta forma, cualquier problema que pudiera surgir en alguno de los nodos o en alguno de los enlaces, no provocaría una catástrofe en la red, y en consecuencia, se logra una arquitectura confiable y veloz. La figura 2.2 muestra un sencillo ejemplo de la disposición de los elementos de la red SS7 y la manera en que estos forman dos redes interconectadas. A continuación se hará énfasis en ciertas características importantes a las que se hace referencia en la figura 2.2:

- Los STPs W y X llevan a cabo las mismas funciones, por lo tanto, son parte de la redundancia de la arquitectura, lo mismo sucede para los STPs Y y Z.

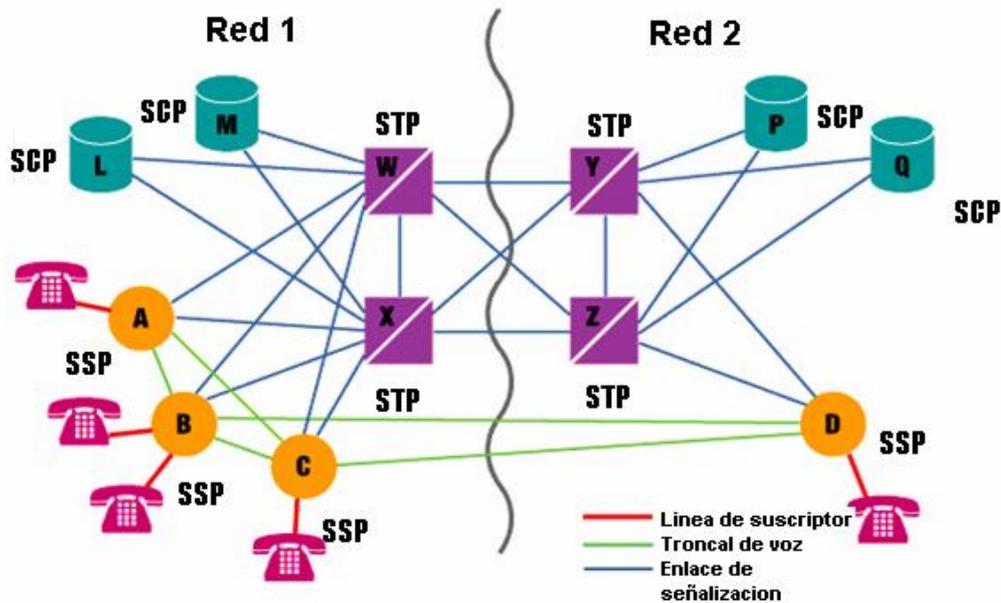


Figura 2.2 Arquitectura de red SS7
(Imagen tomada del tutorial <http://www.iec.org>).

- Cada SSP cuenta con dos enlaces uno para cada STP. La señalización SS7 al resto del mundo se envía a través de estos enlaces, por lo cual hay más redundancia.
- Los STPs que forman un par entre si, se unen mediante un enlace.
- Dos pares de STPs siempre se encuentran interconectados entre si.
- Al igual que los STPs, los SCPs se incrementan en pares. Sin embargo no existen un enlace que una los puntos de control que forman un par.
- Siempre existe una señalización indirecta asociada a los elementos de ambas redes interconectadas.

2.2.2 Jerarquía

Las partes de SS7 son organizadas en una jerarquía de cuatro niveles, el nivel mas alto es un usuario de los servicios proveídos por la parte del nivel mas bajo. Este acuerdo es similar a la estructura del modelo OSI (sistemas de interconexión abiertos) usado como protocolo para comunicación de datos.

Los niveles OSI 1 y 2 corresponden a los niveles 1 y 2 del sistema de señalización SS7, pero depuse empiezan a diferir en las partes altas de la jerarquía.

2.3 PROTOCOLOS

Los protocolos de SS7 y sus niveles se muestran en la figura 2.3, así como su relación con el modelo OSI. La MTP (Parte de transferencia de mensajes) y los bloques ISUP (Parte de usuario de servicios integrados), TUP (Parte de usuario telefónico) y OMAP (Parte de operación, gestión y mantenimiento) constituyen el estándar inicial de SS7. Los bloques TCAP (Parte de capacidades de transacción de la aplicación) y SCCP (Parte de control de la señalización de conexión) adicionales conforman la llamada red inteligente (RI).

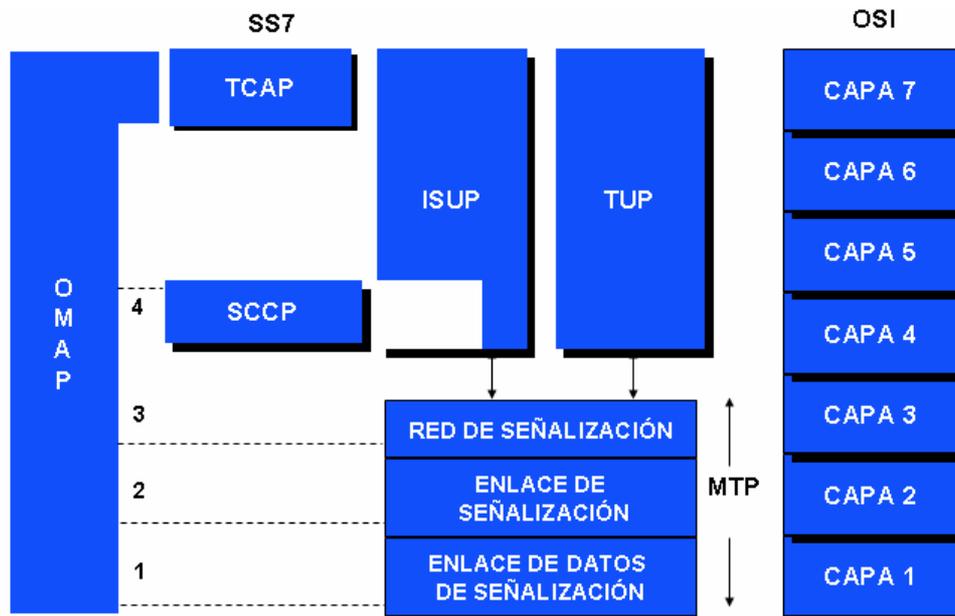


Figura 2.3 Capas del modelo OSI y Capas SS7

2.3.1 Parte de transferencia de mensajes (MTP)

Este protocolo proporciona el servicio de transferencia de mensajes para sus usuarios. Se divide en tres partes denominadas como MTP1, MTP2, y MTP3 que ocupan los niveles 1, 2 y 3 de la jerarquía SS7.

Nivel MTP1 (Capa Física), es equivalente a la capa física del modelo OSI. Aquí se definen las características físicas, eléctricas, y funcionales relacionadas con los enlaces de señalización de la red, sigue la recomendación Q702.

Este nivel está compuesto de una línea de transmisión entre centrales y de un equipo de interfaz de centrales (equipo terminal).

La recomendación de este nivel es que la transferencia de datos posea una velocidad de transferencia de 64 kbps. (Un canal de voz muestreado a la frecuencia de Nyquist, 8000Hz, con 8 bits por muestra).

Nivel MTP2 (Capa de enlace), se ocupa de la transferencia de mensajes de un nodo a otro, y asegura que entre estos dos puntos el intercambio sea confiable. Los mensajes de señalización son enviados por las capas superiores del modelo al segundo nivel de MTP, en donde se forman los paquetes que serán enviados a través del enlace, sigue la recomendación Q703.

Esta etapa también proporciona control de flujo, validación de secuencia de mensajes y verificación de error. Cuando ocurre un error en el enlace de señalización el mensaje se retransmite.

Nivel MTP3 (Capa de red), Este nivel determina el camino el cual los mensajes deberán seguir y es el encargado de la transmisión de mensajes a través de la red de señalización, vista como un todo, sigue la recomendación Q704.

Se encarga de rutear los mensajes entre los puntos de señalización en la red SS7. Rerutea tráfico lejos de enlaces y puntos de señalización con fallas y controla tráfico cuando ocurren congestiones.

2.3.2 Parte del usuario telefónico (TUP)

Comprende los procedimientos funcionales para el intercambio de mensajes de señalización relacionados con la aplicación telefónica. Estos mensajes se intercambian a través de un enlace de señalización y siguen la recomendación Q723.

TUP utiliza los servicios de transferencia de mensajes ofrecidos por la MTP. Cuando una central intenta establecer una conexión de llamada, la TUP de esta central intercambia los mensajes de señalización con la de la central distante. Los mensajes de señalización contienen información relacionada con el establecimiento inicial de llamada y la condición de la vía de conversación.

Estos mensajes se emiten a través del enlace de señalización de la red de señalización por canal común. Con el fin de transmitir un mensaje de señalización a la TUP remota, la TUP, que envía el mensaje, lo transfiere a la función de red de señalización (MTP 3). Luego la MTP selecciona una ruta que vaya a la central distante en la red de señalización por canal común. La TUP de la central distante recibe el mensaje de señalización desde su función de red de señalización. De esta manera las TUP de ambas centrales se comunican entre si y la función de transferencia de mensajes es transparente para ellas.

2.3.3 Parte del usuario de servicios integrados (ISUP)

Es un protocolo de señalización que provee servicios a aplicaciones RDSI. Básicamente se ocupa de la iniciación y terminación de llamadas telefónicas entre SSPs. Los mensajes de configuración de llamada son un ejemplo de mensajes que utilizan el protocolo ISUP.

Entre estos mensajes de configuración se tiene al mensaje de dirección inicial (IAM), que se emplea para iniciar una llamada entre dos SSPs, el mensaje de repuesta (ANM), que indica que una llamada ha sido aceptada por el usuario destino, y el mensaje de liberación de llamada (REL), utilizado para iniciar el proceso de desconexión.

ISUP es un protocolo orientado a conexión, lo cual significa que esta relacionado con el establecimiento de la conexión entre usuarios, es decir, se trata de un servicio, en el que se establece una conexión, se utiliza y se libera.

En conclusión ISUP realiza todas las funciones TUP más otras adicionales y está amparado por la recomendación Q763 y es el estándar utilizado en Colombia.

2.3.4 Parte de control de la conexión de señalización (SCCP)

Es normalmente un protocolo no orientado a conexión. La señalización no orientada a conexión se refiere al intercambio de información si necesidad de establecer una configuración de conexión previa al intercambio de información. De esta forma, la información simplemente es enviada, pudiendo llegar al destino en un orden diferente al cual fue transmitida.

SCCP es el protocolo que permite que los mensajes sean utilizados por aplicaciones independientes dentro de un nodo. A estas aplicaciones se les conoce como subsistemas y entre ellas se puede mencionar el procesamiento de llamadas 1-8000, llamadas con tarjeta y redes inteligentes avanzadas.

Además de esto SCCP cuenta con un mecanismo de direccionamiento avanzado que le permite llevar a cabo la señalización entre dos puntos, aun cuando estos no conozcan la dirección del otro, este método es conocido como direccionamiento de título global.

2.3.5 Parte de aplicación de las capacidades de transacción (TCAP)

Esta capa define los mensajes y el protocolo utilizado en los nodos para comunicarse entre aplicaciones. TCAP se emplea en los servicios de bases de datos y se encuentra definido únicamente para señalización no orientada a conexión, por lo que necesita de la capa SCCP para el transporte.

2.3.6 Parte de administración, mantenimiento y operación (OMAP)

OMAP define mensajes y protocolos diseñados para asistir a los administradores de la red SS7. Hasta la fecha, el más plenamente desarrollado y desplegado de estas capacidades son los procedimientos para validación de las tablas de enrutamiento de red y para problemas de diagnóstico de enlace. OMAP incluye los mensajes que utilizan tanto el MTP y el SSP para ruteo.

2.4 MENSAJES Y TRANSFERENCIA DE MENSAJES

En la figura 2.4 se distinguen varios tipos de mensajes en un punto de señalización. Los mensajes de usuario son mensajes entre el nivel 4 del protocolo y MTP3: son llamados como los mensajes TUP, ISUP y SCCP. La unidad de mensaje de señalización (MSU) son mensajes entre MTP2 y MTP1 de un enlace de señalización en un punto de señalización determinado, y entre el MTP1 y ambos extremos de un enlace de señalización, un MSU contiene un mensaje originado por MTP3 o por un usuario MTP3.

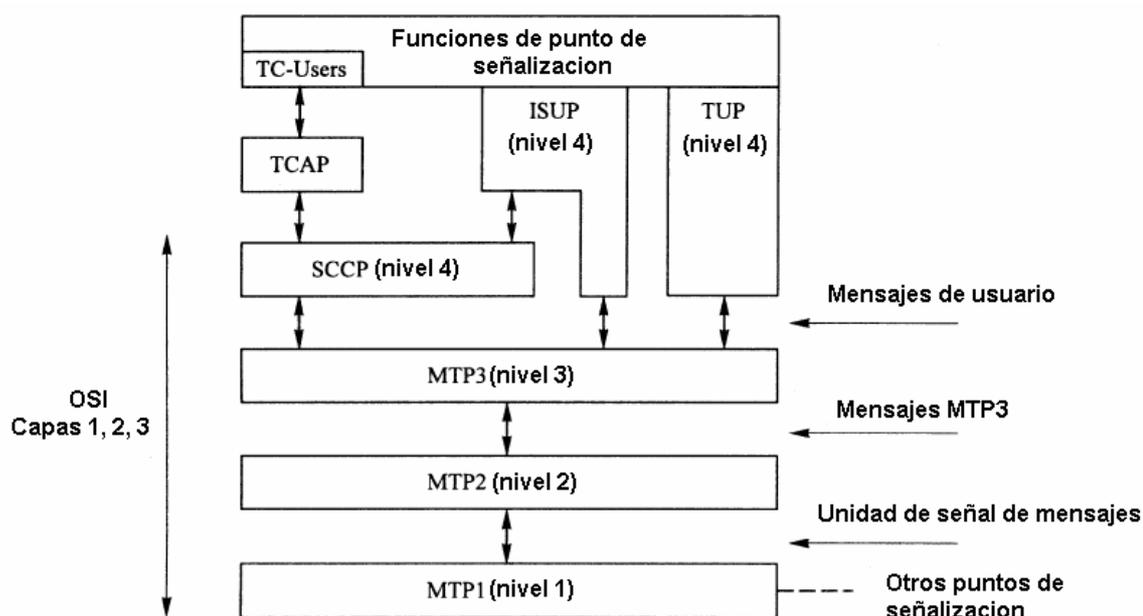


Figura 2.4 Estructura de mensajes en SS7

Los mensajes de usuarios son transferidos entre dos pares de protocolos a nivel cuatro en dos puntos de señalización. La figura 2.5 ilustra la transferencia de un mensaje TUP que proviene de TUP-A hacia TUP-C (en los puntos de señalización A y C, respectivamente) el cual es ruteado por el punto de transferencia de señalización B.

En el punto de señalización A, TUP-A pasa el mensaje TUP hacia abajo hacia su MTP3 el cual lo expande dentro de un mensaje MTP3 y lo pasa al MTP2 del enlace de señalización hacia B. MTP2 expande el mensaje MTP3 dentro de una unidad de señal de mensaje (MSU) y lo pasa a su MTP1.

El MSU atraviesa el enlace de señalización y llega al MTP1 del punto de señalización B, donde MTP2 extrae el mensaje MTP3 y lo pasa a su MTP3. MTP3 transfiere el mensaje MTP3 al MTP2 del enlace de señalización C.

La segunda etapa del mensaje es transferida de manera similar: el mensaje es pasado y expandido hacia abajo en el punto de señalización B, atraviesa el enlace de señalización entre B y C (como un MSU), y es pasado hacia arriba

en el punto de señalización C, finalmente llega como un mensaje TUP a TUP-C.

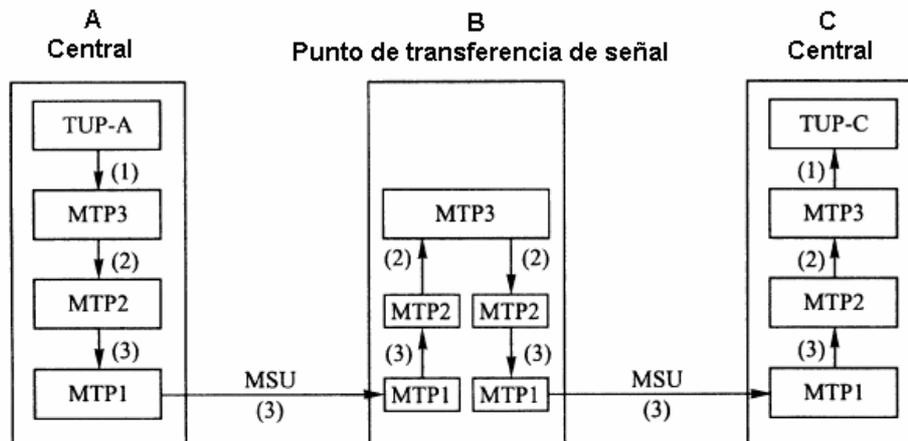


Figura 2.5 Transferencia de un mensaje TUP
 { (1) Mensaje TUP; (2) Mensaje MTP3; (3) Unidad de señal de mensaje }

2.5 EJEMPLO BÁSICO DE LLAMADA

A continuación veremos un ejemplo en donde un suscriptor de la central de conmutación A el cual está conectado a la entidad SSP que se encuentra en la parte inferior izquierda de la figura 2.6 realizará una llamada a un suscriptor de la central de conmutación B, el cual está conectado a la entidad SSP que se encuentra en la parte inferior derecha de la misma gráfica, además se pueden observar los STP W y STP X los cuales serán los encargados de transferir los mensajes, se explicará paso por paso el procedimiento necesario para establecer una llamada.

1. La central de conmutación A analiza los dígitos marcados y determina que será necesario mandar la llamada a la central de conmutación B.
2. La central de conmutación A selecciona una troncal entre ella y B y elabora un IAM (Mensaje Inicial de Llamada), en él se identifica a A como origen, a B como destino, el número que llama, el número a quien se le llama y más información.
3. La central de conmutación A escoge uno de sus enlaces A por medio del STP W y transmite el mensaje IAM hacia la central de conmutación B.
4. El STP W recibe el mensaje, revisa la etiqueta de ruteo y observa que pertenece a la central de conmutación B, por lo que lo transmite a B.

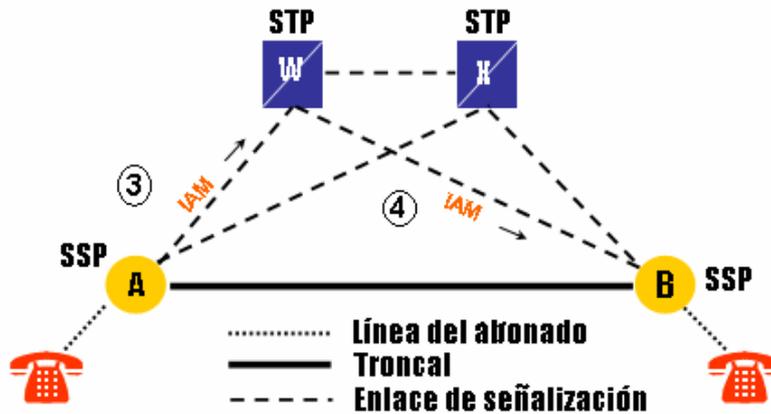


Figura 2.6 Envío de mensaje inicial de llamada de switch A a switch B

5. La central de conmutación B revisa el mensaje y se da cuenta que el número llamado le pertenece y que éste está desocupado.

6. La central de conmutación B elabora un ACM (Mensaje de Dirección Completa), la cual indica que el IAM ha llegado a su destino.

7. La central de conmutación B escoge uno de sus enlaces A por medio del STP X y transmite el ACM, al hacer esto completa la conexión de la llamada hacia atrás; manda tono de llamando hacia A y llama hacia la parte llamada.

8. El STP X recibe el mensaje, revisa su etiqueta de ruteo y ve que debe ser ruteado hacia A.

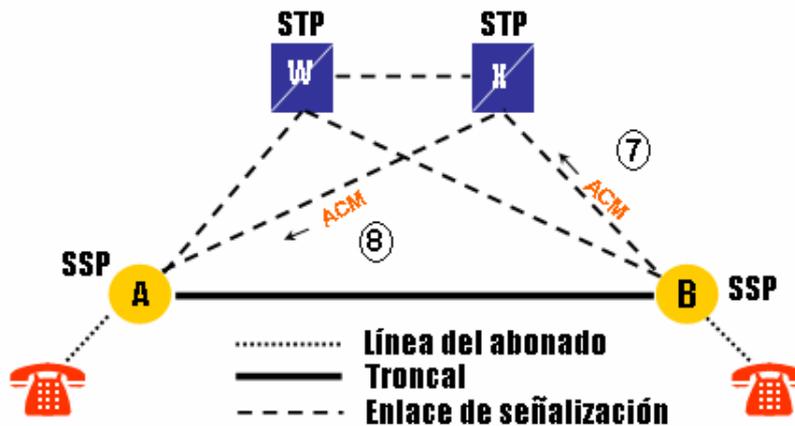


Figura 2.7 Envío de tono de llamada de Switch B a switch A

9. Al recibir el ACM, la central de conmutación A conecta la parte llamante a la troncal de regreso para que pueda escuchar el tono de llamada.

10 Cuando la parte llamada contesta, la central de conmutación B elabora un Mensaje de Respuesta (ANM).

11. La central de conmutación B utiliza el mismo enlace A anterior y transmite el ANM, para esto la troncal ya debe estar conectada a la línea llamada en ambas direcciones para permitir la conversación.

12. El STP X reconoce que el ANM está direccionado a la central de conmutación A y lo manda por el enlace A anterior.

13. La central de conmutación A se asegura que la parte llamante esté conectado a la troncal de salida (en ambas direcciones) para permitir la conversación.

14. Si la parte llamante cuelga primero, A genera un mensaje de Liberación (REL) hacia la central de conmutación B, identificando a la troncal asociada con la llamada.

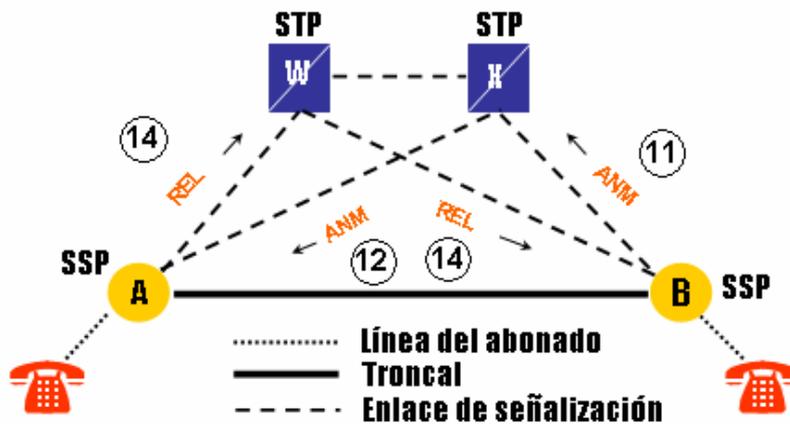


Figura 2.8 Identificación de troncales para conexión de llamada

15. El STP W recibe el mensaje y lo retransmite al switch B.

16. La central de conmutación B recibe el mensaje, desconecta la troncal utilizada, poniéndola en estado desocupado y genera un Mensaje de Liberación Completa (RLC) para la central de conmutación A.

17. El STP X recibe el mensaje y lo direcciona a la central de conmutación A.

18. Al recibir la central de conmutación A el mensaje, esta pone en estado desocupado a la troncal involucrada y completa la llamada.

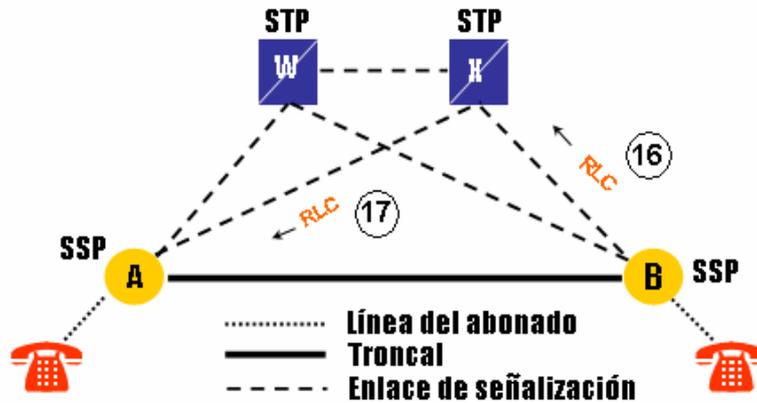


Figura 2.9 Liberación de llamada

2.6 PETICIÓN A UNA BASE DE DATOS

Las personas generalmente están familiarizadas con los números de llamadas gratis 1-8000 pero esos números tienen capacidades adicionales que son posibles gracias a la red SS7.

Aunque ellos son usados para referenciar números telefónicos reales. Ellos no son asignados a la línea suscriptor identificándose como su propio número.

Cuando un suscriptor llama a un número 1-8000, es una señal para que la central telefónica suspenda la llamada y busque más instrucciones de una base de datos. Esta proveerá cualquier número telefónico real al cual pueda ser dirigida la llamada, o identificara otra red a la cual la red será ruteada para más adelante ser procesada. Mientras la respuesta de la base de datos puede ser la misma para cualquier llamada (como por ejemplo un numero personal 1-8000), esto puede ser hecho para variar en función del numero llamado, la hora del día, el día de la semana, o otros factores.

El siguiente ejemplo muestra como una llamada 1-8000 puede ser ruteada (ver figura 2.10:

1. Un suscriptor de la central de conmutación A quiere reservar unos boletos de avión en la sucursal más cercana de la compañía de aviación. Marca el número 1-8000 de la compañía de aviación.
2. Una vez que la central de conmutación A recibe el número completo, esta reconoce que es un número 1-8000 y que requiere ayuda para manejar éste
3. La central de conmutación A elabora un mensaje de petición 1-8000 incluyendo quien llama y el número llamado, lo manda a cualquiera de sus STP por un enlace A.

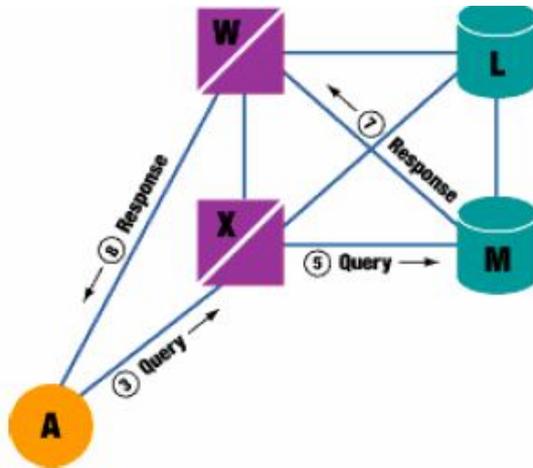


Figura 2.10 Procesamiento de una llamada en una base de datos
 (Imagen tomada del tutorial <http://www.iec.org>).

4. El STP X recibe y reconoce que el mensaje es una petición de número 1-8000 y selecciona una base de datos que pueda responder a la petición.
5. El STP X manda una petición al SCP M por un enlace A, el SCP M recibe la petición, y basándose en sus registros selecciona el número telefónico real o la red a la cual debe ser ruteada la llamada
6. El SCP M elabora un mensaje de respuesta con la información necesaria para procesar la llamada, la direcciona hacia la central de conmutación A, escoge un STP y un enlace A para mandarlo y rutea la respuesta.
7. El STP W recibe el mensaje de respuesta, reconoce que es para la central de conmutación A y lo rutea hacia él.
8. Una vez que la central de conmutación recibe la respuesta utiliza la información para determinar a donde debe ser ruteada la llamada, escoge una troncal hacia ese destino, elabora un IAM y procede de la misma forma que en el ejemplo anterior.

2.7 FUNCIONES DE LA PARTE DE TRANSFERENCIA DE MENSAJES

La parte de transferencia de mensajes (MTP) es un medio de transferencia, entre dos puntos de señalización, de los mensajes de señalización preparados por la parte de usuario en la jerarquía de funciones del sistema de señalización SS7 tal como se muestra en la figura 2.11; sin manipular el contenido de los mensajes.

La MTP esta compuesta del enlace de datos de señalización de nivel 1, la función de enlace de señalización de nivel 2, y las funciones de red de señalización de nivel 3. A continuación estas tres funciones se explicaran con más detalle.

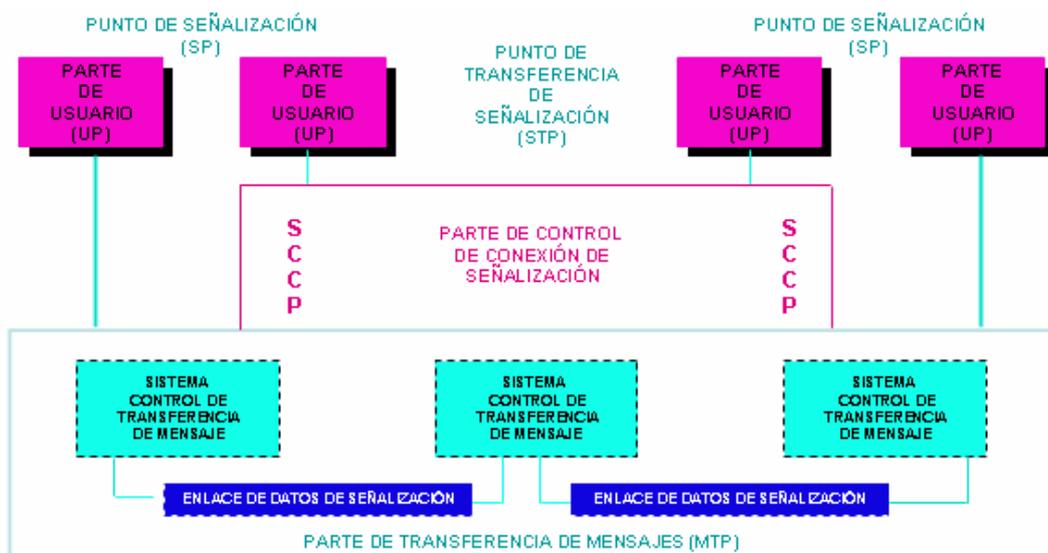


Figura 2.11 Esquema funcional MTP

2.7.1 ENLACE DE DATOS DE SEÑALIZACIÓN (MTP 1)

El enlace de datos de señalización es una trayectoria de transmisión física que permite transportar las unidades de señalización en ambas direcciones, y ocupa el nivel 1 de la jerarquía de funciones de sistema de señalización SS7.

El enlace de datos de señalización está compuesto de dos líneas de transmisión entre centrales de idéntica velocidad de transmisión para la transmisión de datos en ambas direcciones, y dos equipos de interfaz que están directamente conectados entre si a través de estas líneas de transmisión.

El enlace de datos de señalización en la trayectoria de transmisión digital entre centrales se denomina "enlace de datos de señalización analógica", ya la trayectoria de transmisión analógica entre centrales analógicas, se denomina "enlace de datos de señalización analógica". Como equipo de interfaz, se pueden utilizar los conmutadores digitales o el equipo interfaz apropiado, y cada uno de ellos tiene una función de conectar o conmutar las unidades de

señalización del sistema de señalización SS7 por medio de la línea de transmisión entre centrales con el equipo de señalización del nivel 2.

La velocidad máxima y normalizada de transmisión de unidades de señalización es de 64 kbps, la cual es la velocidad normalizada de la red PCM. Sin embargo la transmisión a velocidades inferiores también es posible, por ejemplo se permite un mínimo de 4,8 kbps para la transmisión de unidades de señalización para el proceso de llamadas telefónicas.

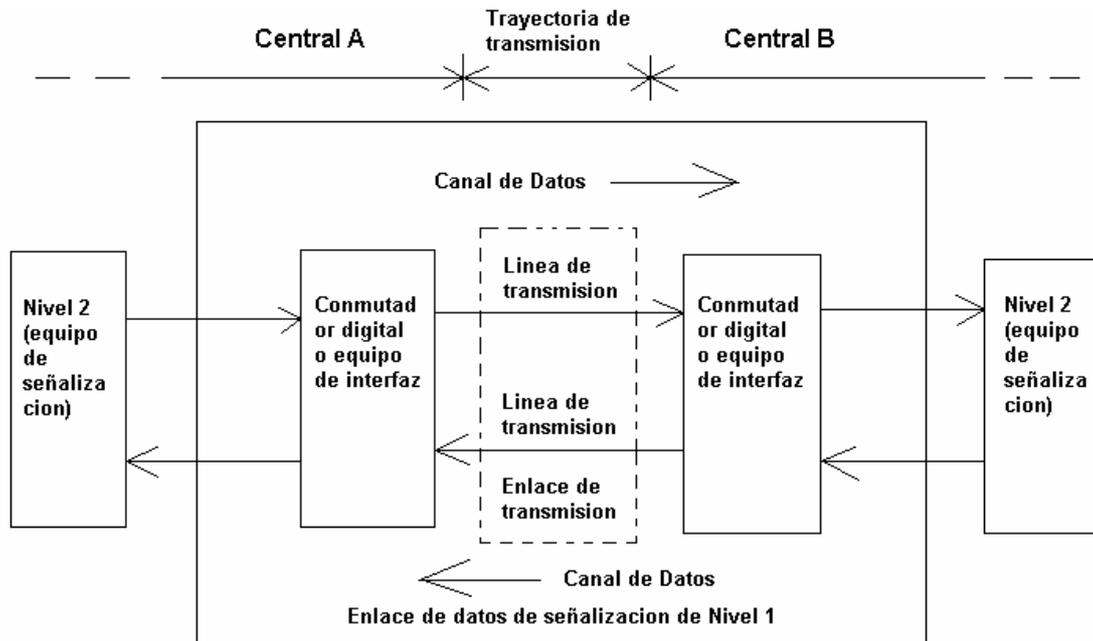


Figura 2.12 Estructura de MTP1

2.7.2 FUNCIONES DEL ENLACE DE SEÑALIZACIÓN (MTP 2)

Las funciones del enlace de señalización que corresponden al nivel 2 dentro de la jerarquía funcional del sistema de señalización SS7, definen la transferencia de mensajes de señalización por un enlace de datos de señalización. Estas funciones están diseñadas para efectuar una transferencia confiable de mensajes de señalización entre dos puntos de señalización conectados directamente. Junto con el enlace de datos de señalización (MTP1), las funciones del enlace de señalización ofrecen un servicio de soporte para los niveles jerárquicos superiores.

2.7.2.1 Estructura MTP 2

Las partes principales de la estructura MTP 2 se pueden observar en la figura 2.13, el enlace de control (LC) controla las otras unidades funcionales de MTP 2. En primer lugar, el enlace de control coordina la transferencia de unidades de señalización. El enlace de control también monitorea la operación del enlace de señalización. Este lo comunica con su MTP3, aceptando los comandos de estado de enlace (C) y reportando información de estados de enlace con

indicaciones (I), finalmente el enlace de control se comunica con el enlace de control del extremo distante del enlace de señalización, usando unidades de señalización del estado de enlace (LSSU).

El MTP3 en un punto de señalización ocupa la salida de los mensajes de MTP3 en el buffer de salida (OB) del enlace de señalización. El buffer de retransmisión (RB) almacena los mensajes que han sido enviados pero que aun no han sido reconocidos positivamente por el MTP2 distante.

Cada mensaje a ser transmitido o retransmitido pasa a través del procesador de salida (OP) y entonces entra al enlace de señalización de datos como un MSU (Unidad de señalización de mensaje). Una unidad de señalización recibida del enlace de señalización de datos es procesada por el procesador de entrada (IP). Los mensajes MTP3 en los MSUs que son aceptados por el procesador de entrada toman lugar en el buffer de entrada (IB) y son recuperados por MTP3.

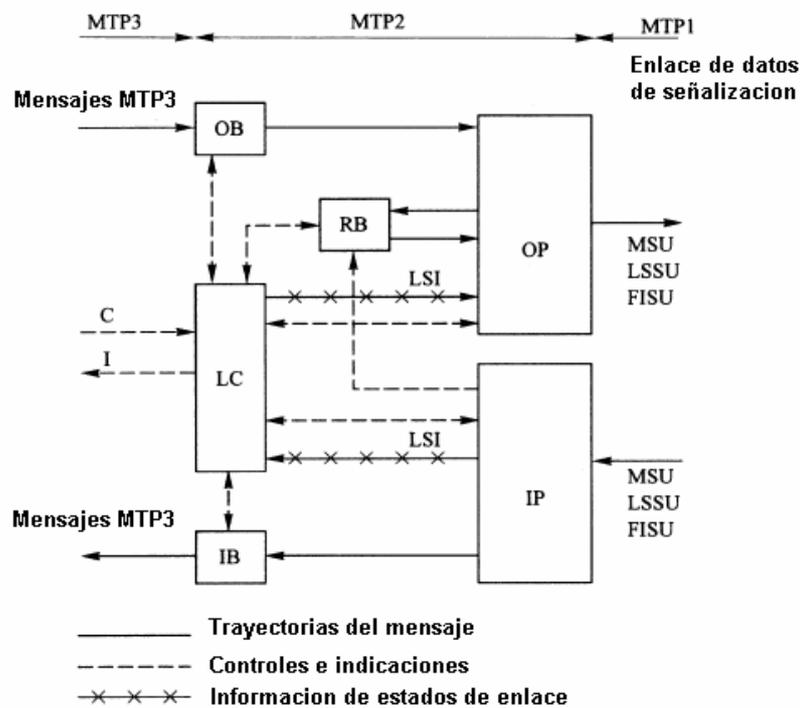


Figura 2.13 Estructura de MTP2

2.7.2.2 Descripción general de las funciones del enlace de señalización

Las funciones del enlace de señalización (MTP2) del sistema de señalización SS7 comprenden los siguientes bloques funcionales;

- Control del estado del enlace: coordina la operación de cada bloque funcional en el nivel 2 y recibe instrucciones de reporte sobre las condiciones del enlace.

- Control de alineación inicial: controla la secuencia para el establecimiento de un enlace de señalización.
- Control de transmisión: controla el proceso secuencial de las unidades de señalización y administra los registros intermedios de transmisión y retransmisión.
- Control de alineación y detección de errores (transmisión): genera 16 bits de comprobación, encierra la unidad de señalización con un patrón de banderas y transmite el caudal de bits por el enlace de datos.
- Control de alineación y detección de errores (recepción): recibe el caudal de bits procedente del enlace de datos, detecta el patrón de banderas y efectúa una comprobación de redundancia cíclica.
- Control de recepción: transfiere los mensajes recibidos al nivel 3, notifica al control del estado de enlace al recibir el LSSU y supervisa la ocupación de registro intermedio de recepción.
- Monitor de la tasa de errores de la unidad de señalización: supervisa la calidad de la transmisión, mientras el enlace esta en el estado de servicio, contando los errores de las unidades de señalización.
- Monitor de la tasa de errores de alineación: supervisa la calidad de transmisión contando el número de errores de las unidades de señalización durante el periodo de prueba de alineación final.
- Control de interrupción del procesador: notifica la situación de interrupción de l procesador local a la central opuesta y notifica al nivel 3 cuando de recibe una indicación de la interrupción del procesador remoto y transmite la FISU.
- Control de congestión: controla los estados de congestión en el nivel 2.

2.7.2.3 Unidades de señalización

La unidad de señalización, es una unidad de información con cierto significado que se transmite por el enlace de señalización. Es de longitud variable en unidades de octetos (8 bits). La unidad de señalización tiene longitud variable para adecuarse a los diversos requerimientos de los usuarios.

Cada unidad de señalización esta encerrada por banderas de 8 bits. Hay tres tipos de unidades de señalización denominadas, unidad de señalización de mensajes (MSU), unidad de señalización del estado de enlace (LSSU) y unidad de señalización de relleno (FISU). La MSU se utiliza para la transmisión de datos desde el nivel 3 o el 4, la LSSU contiene información sobre el estado del enlace de señalización utilizado, y la FISU indica que no hay información de los niveles 3 o 4 disponibles para la transmisión.

En la figura 2.14 se puede observar la estructura de la unidad de señalización en los diferentes formatos, con sus respectivos componentes los cuales serán detallados a continuación:

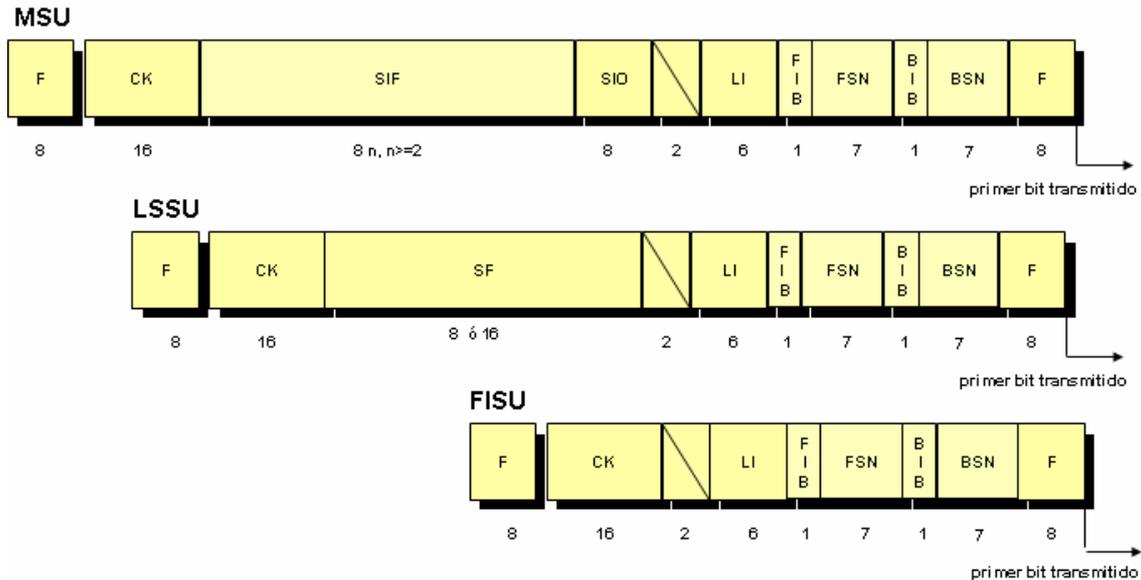


Figura 2.14 Formatos de unidades de señalización

2.7.2.3.1 Campos de la unidad de señalización

A continuación se describirán cada uno de los campos de la unidad de señalización.

- **Bandera (F):** una bandera es un tipo de información de 8 bits que tiene un patrón de bits expresado por “01111110”, que indica el principio o el fin de una unidad de señalización. Se utiliza para alinear la transmisión o recepción de unidades de señalización entre dos centrales.

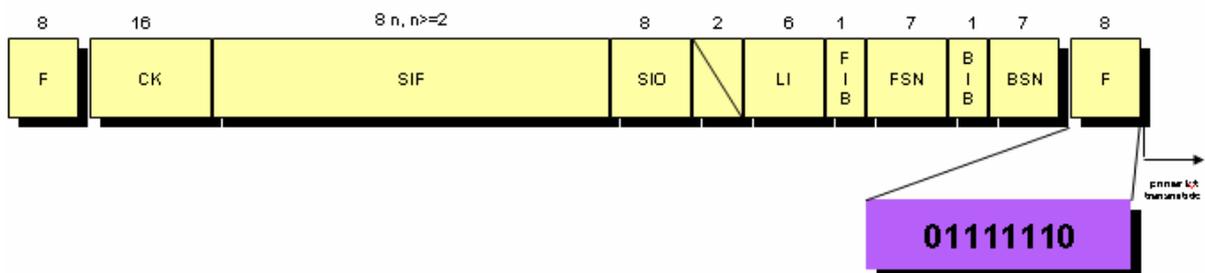


Figura 2.15 Configuración de Bandera

- Número secuencial inverso (BSN): indica el número de secuencia de la última unidad de señalización recibida correctamente. Se utiliza como señal de reconocimiento de todas las unidades de señalización precedentes incluyendo la indicada por el BSN que se han recibido correctamente. Toma valores de 0 a 127, cíclicamente.
- Bit indicador inverso (BIB): es la información de 1 bit utilizada para solicitar la retransmisión de unidades de señalización después de haberse detectado algún error.
- Número secuencial directo (FSN): es la información de 7 bits asignada en el lado de transmisión para indicar la secuencia de las unidades de señalización. Toma valores de 0 a 127 en orden cíclico. Los errores de señalización generados en la línea de transmisión se detectan comprobando los FSN que faltan en las unidades de señalización recibida. Sin embargo esta comprobación, solo se efectúa cuando el FIB recibido es idéntico al último BIB transmitido, esto es, solo cuando la retransmisión no está en progreso.
- Bit indicador directo (FIB): es la información de 1 bit utilizada para indicar que la unidad de señalización transmitida corresponde a la información retransmitida. La central que ha recibido la unidad de señalización con el BIB invertido, envía unidades de señalización con el FIB invertido hacia la central opuesta como indicación de que la retransmisión se ha iniciado.
- Indicador de longitud (LI): es información de 6 bits que indica el número de octetos (Bytes) de los datos que siguen al campo LI y que preceden al campo CK (es decir, SIO y SIF en MSU, o SF en LSSU). Se utiliza para distinguir si la unidad de señalización es MSU; LSSU o FISU.

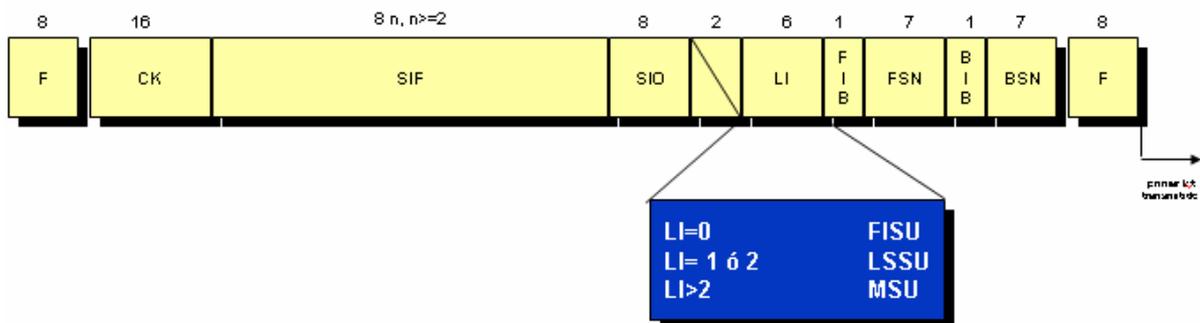


Figura 2.16 Configuración de LI

- Octeto de información de servicio (SIO): es la información de 8 bits que contiene un indicador de servicio de 4 bits (SI) y un campo de subservicio de 4 bits (SSF). El SI es una información que identifica al usuario al cual pertenece el mensaje y el SSF indica si el mensaje se utiliza en una red nacional o internacional.

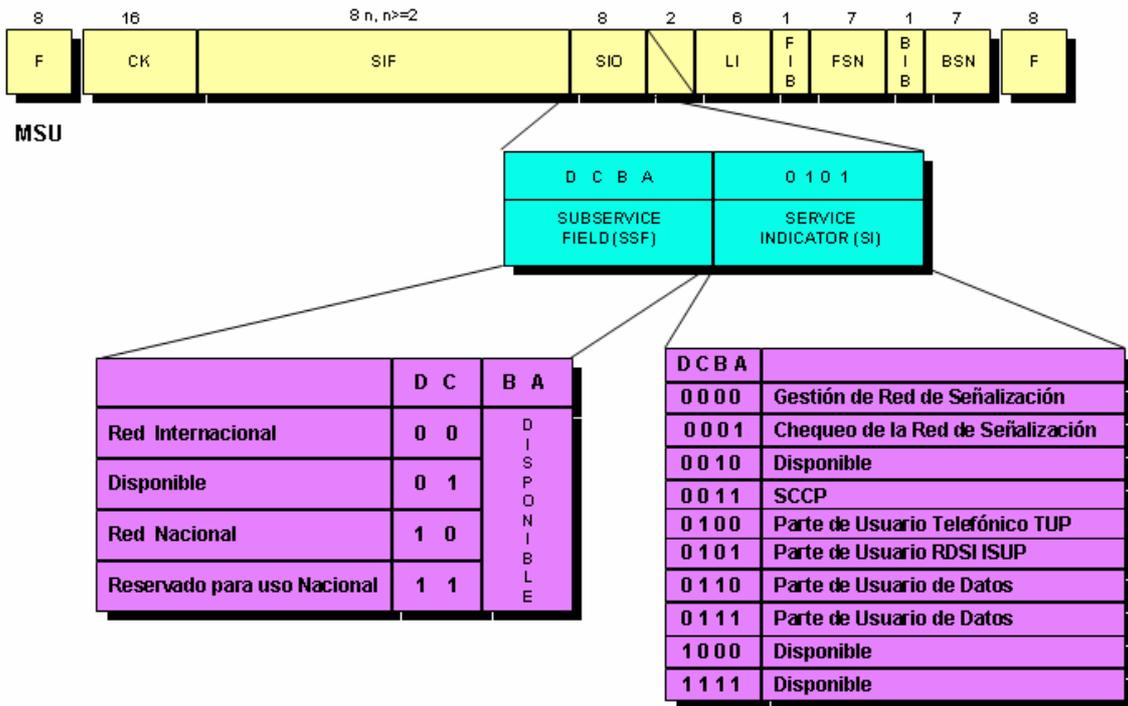


Figura 2.17 Configuración de SIO

- Campo de información de señalización (SIF): contiene los datos a transmitirse o recibirse por cada usuario. Esta información consiste de la etiqueta y de los datos de los usuarios. El SIF tiene una longitud variable en unidades de octetos u su longitud mínima es de 2 octetos (2 Bytes). La etiqueta contiene los datos de enrutamiento del mensaje que se requieren para transferir la unidad de señalización a través de varios puntos de señalización (centrales). El numero de octetos del SIF varia: Red internacional $2 < SIF < 62$ octetos, Red nacional $2 < SIF < 272$ octetos.

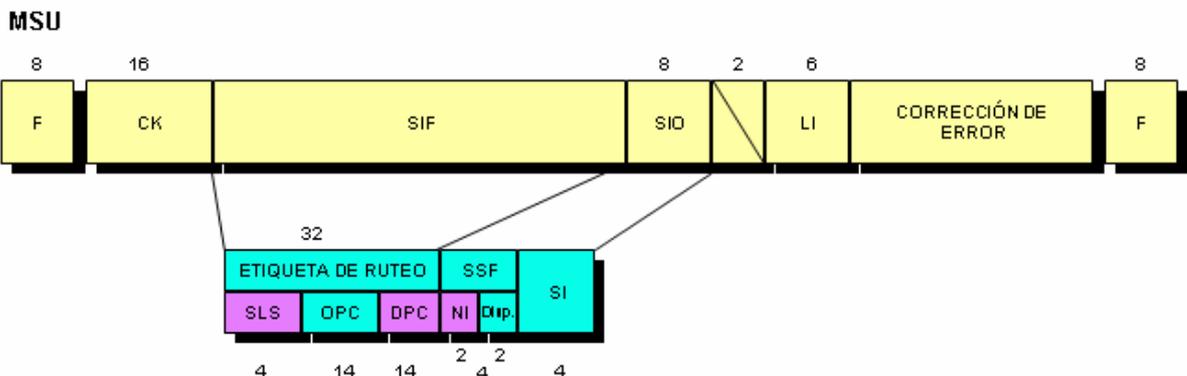


Figura 2.18 Configuración de SIF

- Bits de comprobación de error (CK): son una información de 16 bits utilizada para la detección de errores. Los CK se añaden a cada unidad de señalización antes de la transmisión.

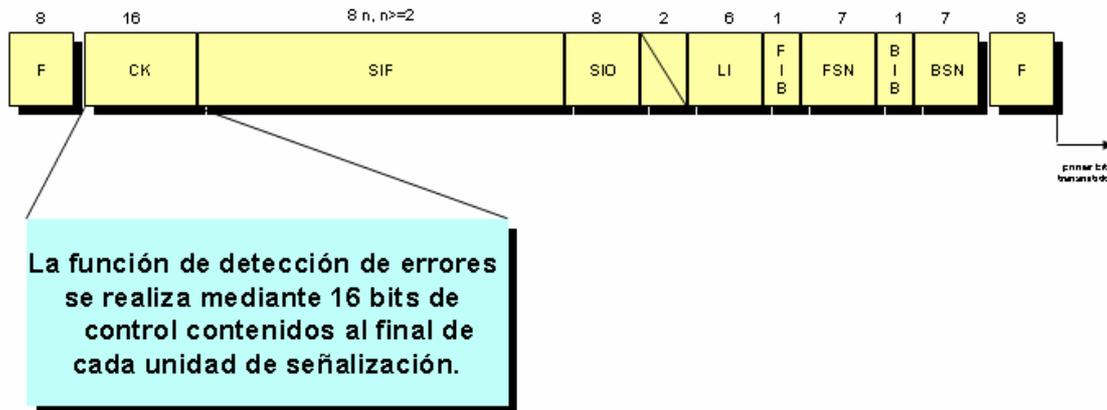


Figura 2.19 Configuración de CK

- Campos de estado (SF) es un campo de 1 o 2 octetos que indica el estado del enlace de señalización. Cuando LI es 1, SF consta de 1 octeto; y cuando LI es 2, SF consta de 2 octetos. Sin embargo, solo los tres primeros bits del primer octeto están definidos por las recomendaciones de UIT.

LSSU

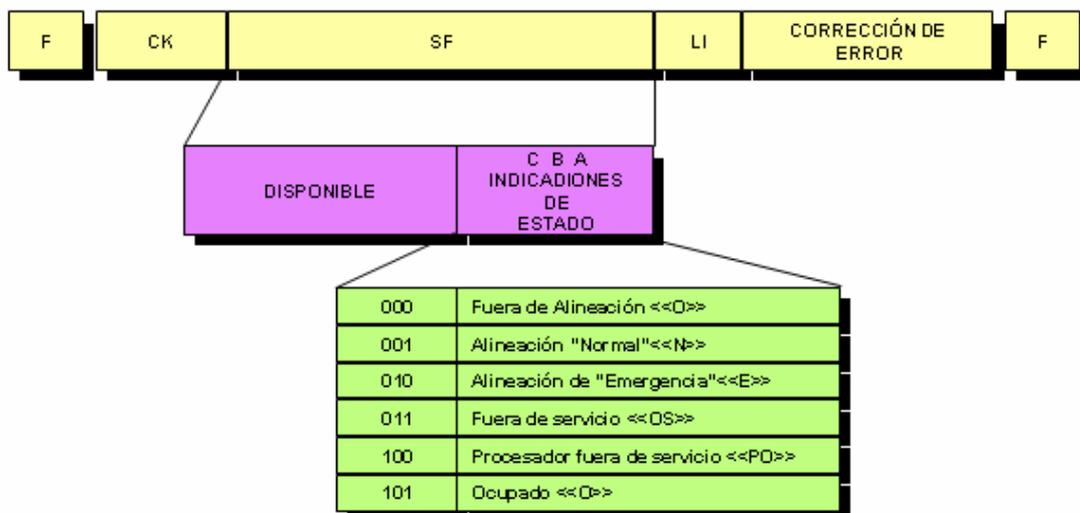


Figura 2.20 Configuración de SF

2.7.2.3.2 Tipos de unidades de señalización

Los tres tipos de unidades de señalización se diferencian mediante el valor del LI que se incluye en todas las unidades de señalización tal como se ha mencionado anteriormente. A continuación se describen brevemente las aplicaciones de las unidades de señalización:

Unidad de señalización de mensaje (MSU):

Se procesa en los niveles 3 o 4. Su información principal que consta de la etiqueta y los datos del usuario están incluidos dentro del SIF.

Los datos típicos del usuario corresponden a las unidades de señalización para la conexión de llamadas telefónicas, unidades de señalización para la gestión de la red de señalización y las unidades de señalización para pruebas y mantenimiento de la red. La parte del usuario telefónico incluye las unidades de señalización generales sin información adicional que se envían al momento de la contestación, desconexión, etc; y las señales de dirección con datos adicionales, como las señales numéricas, etc.

Las unidades de señalización de la red de señalización se utilizan para realizar la gestión de la red de señalización y las unidades de señalización para pruebas y mantenimiento de la red de señalización se utilizan para probar los enlaces de señalización y las rutas de señalización para confirmar su normalidad en el momento del establecimiento de nuevos enlaces de señalización. Si se detecta un error en la MSU, se retransmite tal MSU.

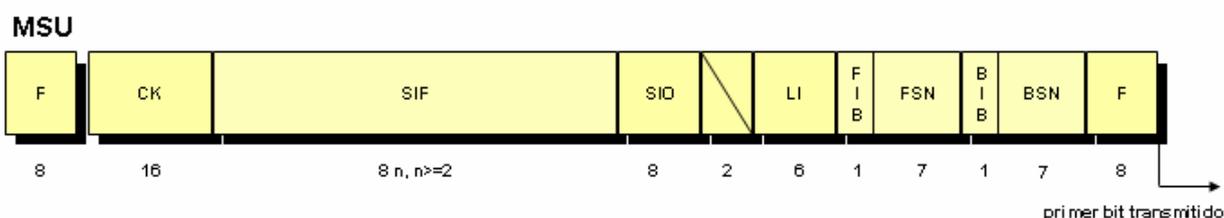


Figura 2.21 Formato unidad de señalización de mensaje MSU

Unidad de Señalización del estado de enlace (LSSU)

La unidad de señalización del estado de enlace (LSSU) sirve para notificar el estado de un enlace de señalización. Se utiliza para indicar el estado de alineación inicial y para probar el enlace de señalización así como para notificar de los problemas a la central del extremo opuesto. Aunque se detecte algún error en la LSSU, tal LSSU no se retransmite.

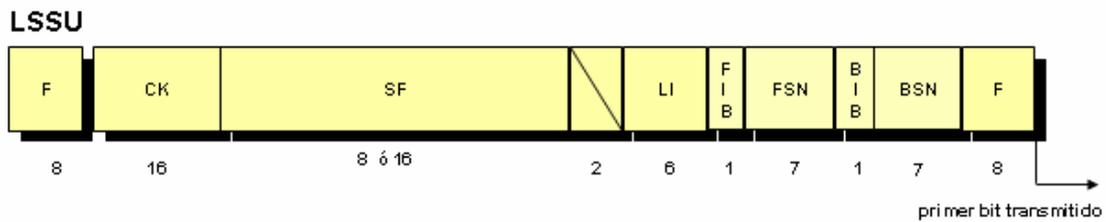


Figura 2.22 Formato unidad de señalización del estado de enlace

Unidad de Señalización de relleno (FISU)

La unidad de señalización de relleno (FISU) se envía cuando no hay datos para transferirse por el enlace de señalización. Contiene la mínima información requerida para el control de transmisión.

La FISU se utiliza como una unidad de señalización ficticia para suprimir el retardo de las unidades de señalización dentro de un valor predeterminado, y como una señal para indicar la conclusión de la prueba de la tasa de errores del enlace de señalización en el momento de la alineación inicial del enlace de señalización. Aunque se detecte un error en una FISU, tal FISU no se retransmite.

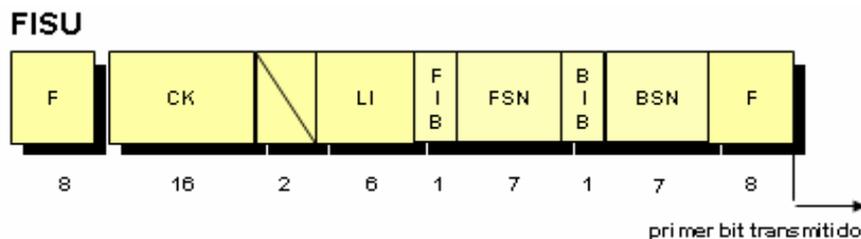


Figura 2.23 Formato unidad de señalización de relleno

2.7.2.4 Control de errores

Cuando los datos se transmiten por una línea de comunicación inevitablemente se generan errores a una tasa constante. Los errores son creados por los efectos del ruido, interferencias, etc.

Los errores pueden provocar un cambio en el estado de los bits (de 0 a 1, o de 1 a 0) o incluso la pérdida de varios bits. La función de control de errores tiene el propósito de detectar estos errores de transmisión de datos en el terminal de señalización de recepción y los corrige en cooperación con el terminal de señalización de transmisión. Los efectos de estos errores son reducidos a un nivel aceptable ya que estos no se pueden detectar y corregir todos estos errores.

2.7.2.4.1 Detección de errores

Se pueden detectar dos tipos de errores, el primer tipo de error es el de los errores de bits de las unidades de señalización, se detectan mediante los 16 bits de comprobación (CK) como se puede ver en la figura 2.19, que se acoplan a cada unidad de señalización mediante el lado de transmisión.

Esta comprobación se denomina comprobación de redundancia cíclica (CRC), y puede detectar errores de hasta 3 bits así como todos los errores de ráfagas de longitud inferior a 16 bits. Un error de ráfaga es un error en el cual se ha producido un error de bit que ocurre consecutivamente en varios bits. Cuando los bits de comprobación detectan un error, se descarta la unidad de señalización que contiene el error en el lado de recepción.

El segundo tipo de errores, es la pérdida de unidades de señalización en el lado de recepción se detectan mediante la comprobación de la consecutividad secuencial de los números de secuencia progresiva (FSN) asignados a todas las unidades de señalización.

La detección de errores utilizando los bits de comprobación es posible solo después de que la unidad de señalización, incluyendo sus bits de comprobación, haya llegado al lado receptor. Sin embargo, si la unidad de señalización se pierde durante la transmisión, la detección de errores utilizando los bits de comprobación no resultara eficaz. Por lo tanto, es muy necesaria la comprobación de la consecutividad secuencial de los FSN.

Se utilizan acuses de recibo positivo para indicar la transferencia correcta de la MSU y acuses de recibo negativos como peticiones explícitas de retransmisión de MSU erróneas recibidas.

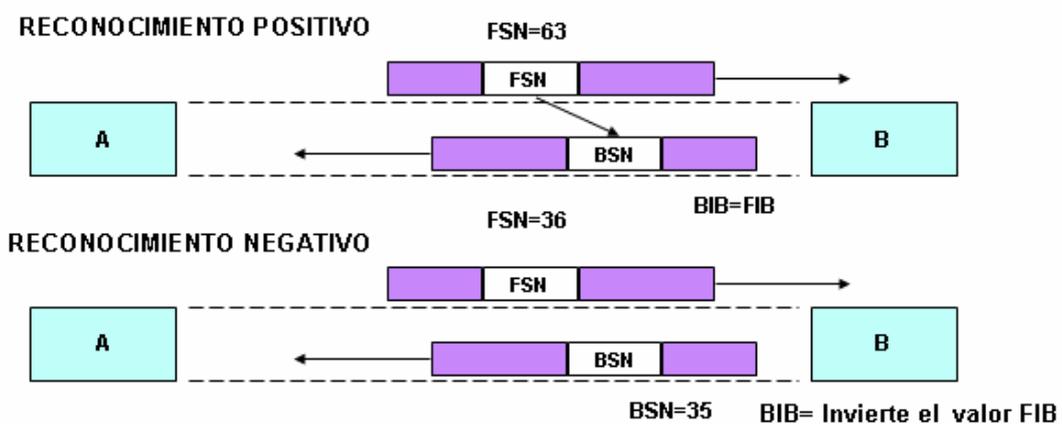


Figura 2.24 Acuses de recibo positivo y negativo

- Cuando el mensaje se recibe correctamente el terminal de señalización envía un RECONOCIMIENTO POSITIVO insertando el FSN del mensaje recibido como un BSN y haciendo el BIB=FIB del mensaje recibido.

- Si el mensaje recibido es incorrecto, el terminal de señalización envía un RECONOCIMIENTO NEGATIVO invirtiendo el BIB y el FSN del último mensaje recibido correctamente se inserta en el campo BSN

2.7.2.4.2 Corrección de errores

Descripción

El método básico de corrección de errores es un método de secuencia no obligada en el que la corrección se efectúa mediante la retransmisión. En este método los acuses de recibo positivos se utilizan para indicar la transferencia correcta de las MSU. Mientras que los acuse de recibo negativos se utilizan como solicitudes explícitas para la retransmisión de las MSU recibidas incorrectamente.

De este modo, durante la retransmisión puesto que no solo se retransmite la unidad de señalización errónea, sino todas las unidades de señalización que la siguen, nunca ocurre la recepción duplicada de las unidades de señalización, si secuencia errónea. Además puesto que se efectúa la solicitud de retransmisión explícita invirtiendo el BIB.

Control de secuencia de la unidad de señalización

Cada unidad de señalización contiene un número de secuencia progresiva (FSN). El FSN únicamente identifica cuando una unidad de señalización se recibe correctamente y en la secuencia correcta. Antes de la transmisión de la MSU, el valor del FSN se obtiene incrementando en 1 el último valor asignado. Después de la transmisión, la MSU se transfiere desde el registro intermedio de transmisión al registro intermedio de retransmisión. Permanece en el registro de intermedio de retransmisión hasta la confirmación de recepción correcta por medio del acuse de recibo.

Una vez recibida la unidad de señalización, se efectúa una comprobación de redundancia cíclica utilizando los bits de comprobación. Si se detecta un error, se descarta la unidad de señalización. De lo contrario, el FSN de la unidad de señalización se compara con el FSN de la última unidad de señalización aceptada. Después de la comparación se toman diferentes medidas dependiendo del tipo de la unidad de señalización.

En el caso de que la unidad de señalización sea una MSU, esta es aceptada y suministrada al nivel intermedio superior cuando el valor de su FSN es mayor en 1 al de la última unidad de señalización aceptada y cuando el FIB esta en el mismo estado que el último BIB transmitido. El acuse de recibo positivo se devuelve a la central transmisora mediante el ajuste del valor BSN de la siguiente unidad de señalización de transmisión regresiva de acuerdo con el valor de FSN de la MSU aceptada.

La unidad de señalización se descarta cuando el valor de FSN no es mayor en 1 comparado con la última unidad de señalización aceptada. Se envía un acuse de recibo negativo, siempre que el FIB no este en el mismo estado que el último FIB transmitido, significa solicitud de retransmisión y que la unidad de señalización recibida no es la retransmitida, por lo tanto la unidad de señalización también será descartada.

Si una FISU recibida tiene un valor FSN igual al de la última MSU aceptada, la unidad de señalización se procesa dentro de la parte de transferencia de mensajes (MTP). Si el valor FSN es diferente al de la última MSU aceptada y el FIB esta en el mismo estado que el último BIB enviado, se enviara un acuse de recibo negativo.

Retransmisión

El valor de BSN de las MSU y FISU recibidas que han satisfecho la comprobación de redundancia cíclica es examinado. El valor de BSN recibido indica que la MSU enviada previamente con un valor de FSN idéntico a este valor de BSN ha sido aceptada así como todas aquellas MSU aun sin acuse de recibo y fueron enviados antes de esta MSU. Por lo tanto, estas MSU serán extraídas del registro intermedio de retransmisión.

El número de MSU que han sido transmitidas pero de las que aun no se tiene acuse de recibo desde el lado de recepción (es decir, el numero de MSU mantenidas en el registro intermedio de retransmisión) esta limitado a 127 aunque el numero máximo de FSN que se pueden asignar es de 128 (de 0 a 127). Cuando la unidad de señalización no es correctamente recibida y descartada, o cuando la unidad de señalización se pierde durante la transmisión, el valor de BSN no se actualiza. El FSN de la última MSU aceptada se establece como BSN de las MSU o FISU enviadas en sentido contrario. Cuando se recibe el mismo valor de BSN, no se toman medidas posteriores.

En el caso de que haya una MISU o FISU recibidas que tengan un valor de BSN que no sea el mismo que el anterior a unos de los valores de FSN de la unidad de señalización en la memoria intermedia de retransmisión, la unidad de señalización será descartada.

Después de emitir una MSU, se activa un temporizador para supervisar el acuse de recibo de esta MSU desde la centra opuesta. El enlace se considera con falla cuando el temporizador expira y no se ha recibido el acuse de recibo. Luego el nivel 3 recibe el informe de que el enlace esta defectuoso. El valor de la temporización normalmente se ajusta a un valor de 1 segundo aproximadamente.

La retransmisión de las MSU se inicia cuando se recibe un acuse de recibo negativo. El acuse de recibo negativo se detecta siempre que el BIB recibido no este en el mismo estado que el último FIB enviado. Todas las MSU del registro intermedio de retransmisión, se retransmiten en su orden original a partir de la MSU que tiene el valor de FSN incrementado en uno con respecto al valor de

BSN recibido. El lado de transmisión notifica al lado de recepción que esta MSU es retransmitida, invirtiendo el FIB de la primera MSU en el inicio de la retransmisión. Este nuevo valor de FIB se mantiene en todas las unidades de señalización transmitidas subsiguientemente hasta que se inicia otra retransmisión. De ahí que en condiciones normales, el BIB incluido en la unidad de señalización transmitida sea igual al valor del BIB de la unidad de señalización recibida. En el caso que se pierda una MSU retransmitida, el lado de recepción la detecta comprobando su FSN y FIB y luego envía de retorno una nueva solicitud de retransmisión invirtiendo el BIB.

Durante la retransmisión, la nueva MSU que debe ser transmitida se mantiene en el registro intermedio de transmisión hasta que la última MSU disponible se mantiene en el registro intermedio de transmisión hasta que la última MSU disponible para la retransmisión haya sido enviada.

Ejemplo #1:

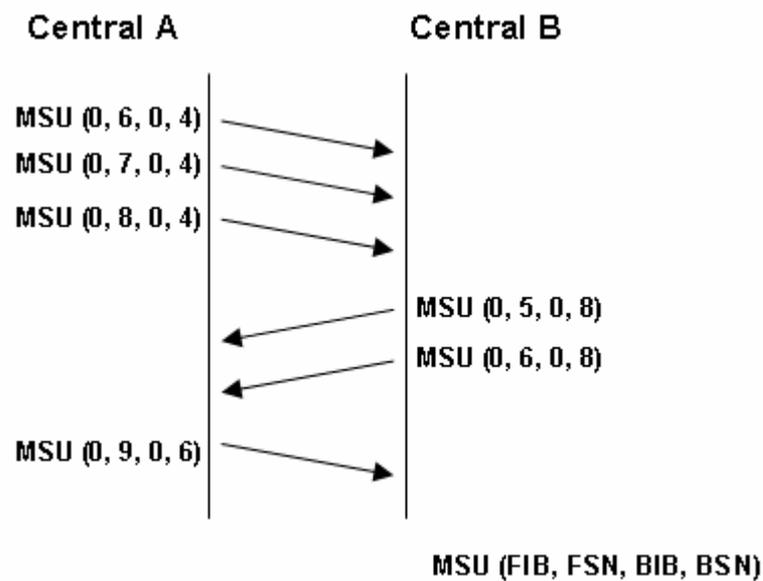


Figura 2.25 Ejemplo de retransmisiones

Según el ejemplo #1 de las figura 2.25 observamos como la central A transmite una MSU con un FSN=6 y un BSN=4 siguen con una retransmisión de los FSN=7 y FSN=8 de esta central, cuando la central B reconoce el mensaje de la FSN=8 de la central A, esta realiza un acuse de recibo positivo por lo cual esta central coloca el valor de BSN=8 como el valor del FSN de la central A que reconoció además coloca el valor de FSN=5 que incrementa en un número el valor de la BSN de la central A el cual es 4, por lo tanto el valor de BIB de la central B es igual al valor FIB del mensaje recibido, esta central a su vez retransmite el MSU hacia la central A, retransmitiendo hasta su FSN=6 y su BSN queda estático en el valor de 8, luego lo recibe la central A y hace también ajuste de recibo positivo por lo cual la FSN de la central B que es igual a 6 se vuelve la BSN de la central A y esta a su vez incrementa su FSN hasta el valor de 9.

Ejemplo #2:

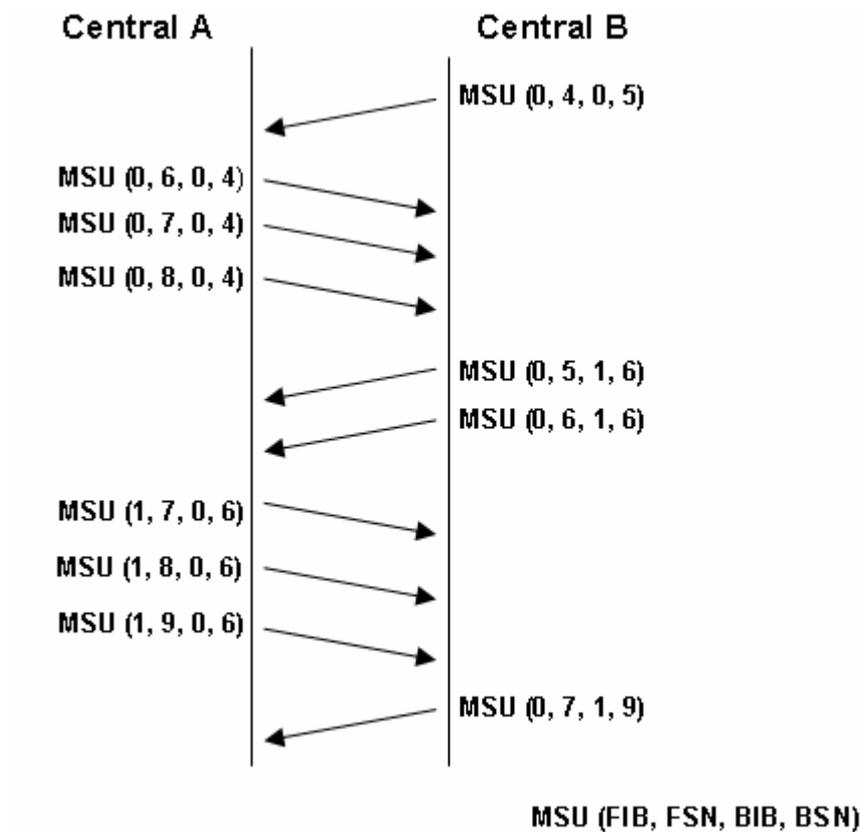


Figura 2.26 Ejemplo de retransmisiones

Según el ejemplo #2 de las figura 2.26 podemos observar como la central B envía un mensaje hacia la central A con FSN=4 y BSN=5 como la central A recibe este mensaje con un ajuste de recibo positivo por lo cual hace que su FSN=6 incrementando en un número el BSN de la central B, además su BSN toma el mismo valor del mensaje recibido de la central B y su BIB toma el mismo valor del FIB de la central B, luego la central B retransmite este mensaje que como no es leído incrementa su FSN hasta tomar el valor de 8, la central B al no recibir ningún mensaje hace un acuse de recibo negativo por lo cual hace su BSN=6 lo cual es el valor de FSN del mensaje el cual estaba esperando por lo cual invierte el BIB para tomar un valor de 1 e incrementa su FSN hasta 6 el cual lo transmite hacia la central A la cual hace un ajuste de recibo positivo por lo cual hace un ajuste recibo positivo por lo cual hace que su FSN=7 incrementando en un número el BSN de la central B y su BIB toma el mismo valor del FIB de la central B y lo retransmite hasta que tome el valor de FSN=9 el cual es recibido por la central B con un ajuste de recibo positivo.

2.7.2.4.3 Método preventivo de retransmisión cíclica

Introducción

El método preventivo de retransmisión cíclica (PCR) es designado para el uso de enlaces de señalización con tiempos largos de propagación (T_p), por ejemplo enlaces de señalización que se encuentran ligados a circuitos satelitales. Como en un método básico de corrección de errores la FSN identifica la posición de una MSU en su secuencia original de transmisión y la BSN siempre identifica la más reciente MSU aceptada.

El método preventivo de retransmisión cíclica (PCR) usa solo ajuste de recibo o conocimiento positivo. Los bits indicadores FIB y BIB son ignorados (ellos permanecen en el valor de 1) y el siguiente proceso simplemente acepta o descarta una MSU libre de errores basado en el valor de su FSN, el cual tiene que exceder el FSN de la MSU mas recientemente aceptada por una unidad.

Retransmisión cíclica preventiva

En el método preventivo de retransmisión cíclica (PCR), las unidades de señalización se retransmiten sin recibir una solicitud de transmisión explícita. La retransmisión se efectúa sin invierte el BIB.

En el lado de recepción acepta las MSU comprobando si los valores de FSN son consecutivos. En el caso de que se detecten errores en la MSU, esta se descarta y como consecuencia el valor de FSN de la siguiente unidad de señalización recibida deja de ser consecutivo. Luego, la central de recepción espera la llegada de la MSU retransmitida. El acuse de recibo negativo no se envía con el método PCR, durante la retransmisión, si no hay una nueva MSU disponible para la transmisión, el ciclo de retransmisión se interrumpe y la nueva MSU se envía con prioridad.

Procedimiento de retransmisión forzada

En el procedimiento de retransmisión forzada, el número de MSU (N1) almacenadas en la memoria intermedia de retransmisión y el numero total de octetos de estas MSU (N2) son supervisados continuamente. Siempre que N1 o N2 llegue a su límite predeterminado, se suspende la retransmisión de nuevas MSU y FISU. Todas las MSU contenidas en la memoria intermedia de retransmisión se retransmiten una vez en su orden de transmisión original. Al finalizar la retransmisión, el procedimiento PCR se repone a la situación normal siempre y cuando N1 y N2 estén por debajo de sus valores límites respectivos. El valor de N1 esta limitado por el número máximo de MSU disponibles para la retransmisión, es decir, si N1 es menor o igual a 127. En ausencia de errores, N2 esta determinado por el retardo del bucle del enlace de señalización (T_L). El valor de N2 debe ser menor o igual a $T_L / T_p + 1$.

2.7.3 FUNCIONES DE LA RED DE SEÑALIZACIÓN (MTP 3)

Mientras que las funciones de los niveles 1 y 2 están relacionadas con el enlace de señalización en si, el nivel 3 realiza las funciones más amplias relacionadas con el tratamiento de los mensajes de señalización y la gestión de la red de señalización.

El servicio básico del nivel 3 consiste en proveer la transferencia transparente de todos los mensajes originados en los niveles 3 y 4 y encaminados a cualquier otra entidad correspondiente a través de la red de señalización de canal común (CSS). La parte de usuario queda libre de toda responsabilidad de esta forma de interconexión de los puntos de señalización. También se supervisa y controla el estado de los enlaces de señalización. El nivel 3 efectúa dos funciones importantes relacionadas con el encaminamiento y la retransmisión.

Antes de continuar definimos algunos términos: cualquier nodo de la red de telecomunicaciones constituye un punto de señalización. En la red de señalización cada punto de señalización se identifica mediante un código que de acuerdo al origen o destino de los mensajes se denomina código de punto de origen (OPC) o código de punto de destino (DPC). Por lo tanto en la etiqueta de encaminamiento de cada mensaje se halla contenido un OPC y un DPC, para identificar únicamente los puntos de origen y destino de los mensajes.

Las funciones de la red de señalización (MTP3) es dividido en los siguientes dos grupos de funciones:

- Tratamiento de mensajes de señalización.
- Gestión de red de señalización.

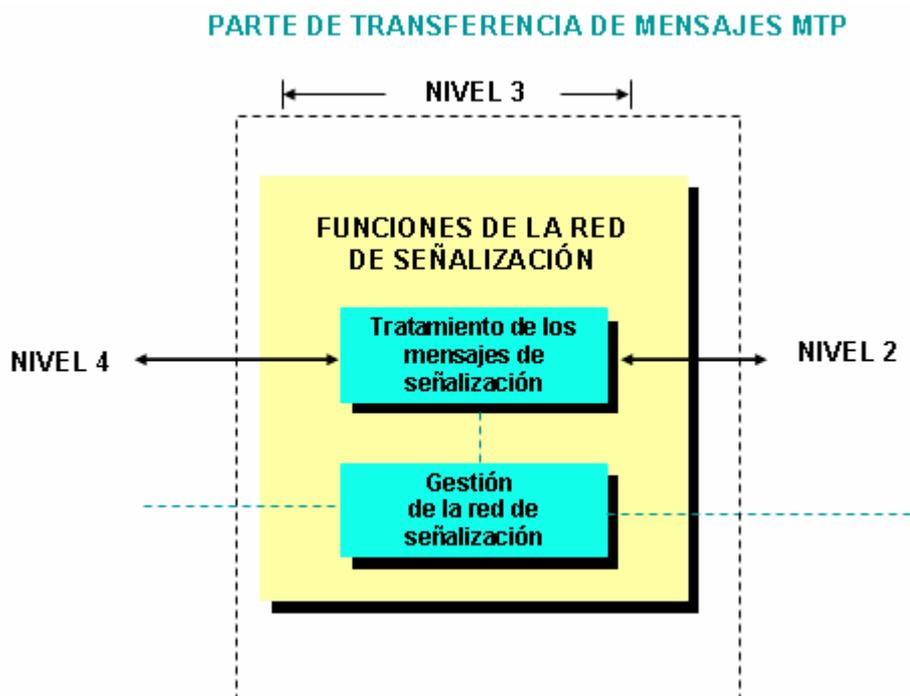


Figura 2.27 Estructura de MTP3

2.7.3.1 TRATAMIENTO DE MENSAJES DE SEÑALIZACIÓN

El tratamiento de mensajes de señalización (SMH) maneja la transferencia de mensajes entre dos pares de usuarios MTP de la parte de usuario del teléfono (TUP), parte de usuario de servicios integrados (ISUP) y la parte de control de conexión de señalización (SCCP).

2.7.3.1.1 Formato de mensajes

El Formato general de mensajes MTP3 es mostrado en la figura 2.28, se distingue el octeto de información de servicio (SIO) y el campo de información de señalización (SIF). A su vez SIF se divide en etiqueta de enrutamiento (RL) y en mensaje de usuario (UM).

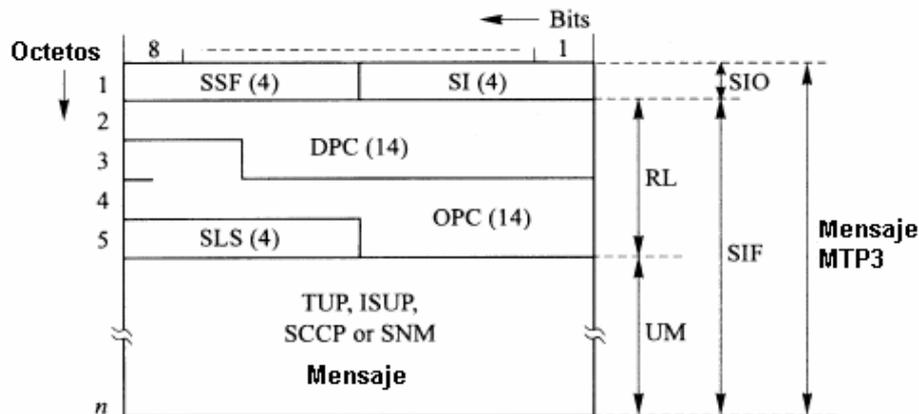


Figura 2.28 Formato de mensajes MTP3

2.7.3.1.2 Parámetros de etiqueta de encaminamiento (RL)

Los parámetros de la etiqueta de encaminamiento (RL) son usados por los MTP3s en los puntos de señalización a lo largo de la ruta del mensaje para determinar el tipo de la ruta de señalización para establecer el destino del mensaje.

- Punto código de origen (OPC) y punto de código de destino (DPC): estos parámetros identifican el código de punto del destino y los puntos de señalización de origen. Los códigos de punto en Colombia tienen 14 bits.
- Selector de enlace de señalización (SLS): este parámetro divide las cargas del mensaje saliente del usuario de MTP en un punto de señalización. SLS es usado para la selección de un enlace de señalización particular en la ruta de señalización para establecer el destino del mensaje.

2.7.3.1.3 Mensajes de usuario (UM)

Se clasifican en mensajes de nivel 3 y de nivel 4, los mensajes de nivel 3, se refieren a los mensajes de gestión (SNM) y su formato se muestra en la figura 2.29.

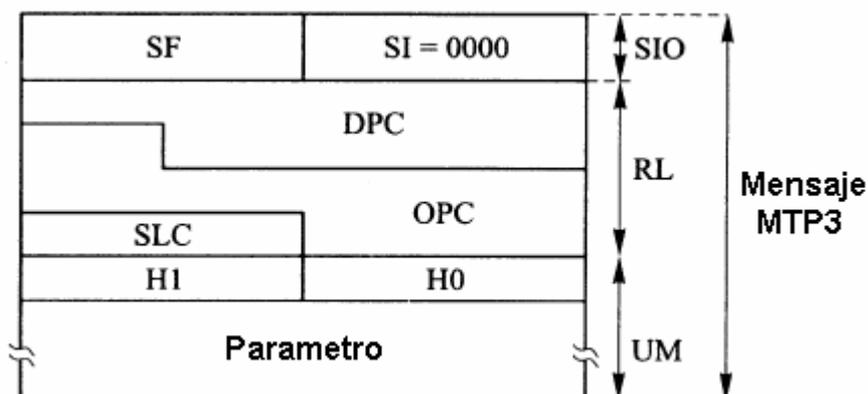


Figura 2.29 Mensaje MTP3 en el campo de usuario de mensaje

Donde H0 y H1 son códigos de cuatro bits que indican el tipo de mensaje, como se detalla en la tabla 2.1.

Nombre de Mensaje	Siglas	H0	H1
Changeover Order	COO	0001	0001
Changeover Acknowledgment	COA	0001	0010
Changeback Declaration	CBD	0001	0101
Changeback Acknowledgment	CBA	0001	0110
Emergency Changeover	ECM	0010	0001
Emergency Changeover Acknowledgment	ECA	0010	0010
Transfer Prohibited	TFP	0100	0001
Transfer Allowed	TFA	0100	0101
Transfer Controlled	TFC	0011	0010

Tabla 2.1 Mensajes para gestión de red de señalización

Por otra parte los mensajes de nivel 4 o de la parte de usuario, presentan el siguiente formato:

Los formatos de mensajes TUP son mostrados en la figura 2.30. El octeto (a) es el octeto de información de servicio (SIO), conformado por el indicador de servicio (SI) y el campo de subservicio (SSF). El valor SI=0100 indica un mensaje TUP.

La etiqueta de encaminamiento (RL) se encuentra entre los octetos (b) y (e) y contiene los puntos de código de origen y destino (OPC,DPC) y el selector de

enlace de señalización (SLS). La troncal por la cual el mensaje es destinado es identificada por la combinación de OPC, DPC y CIC (Código de identificación de circuito).

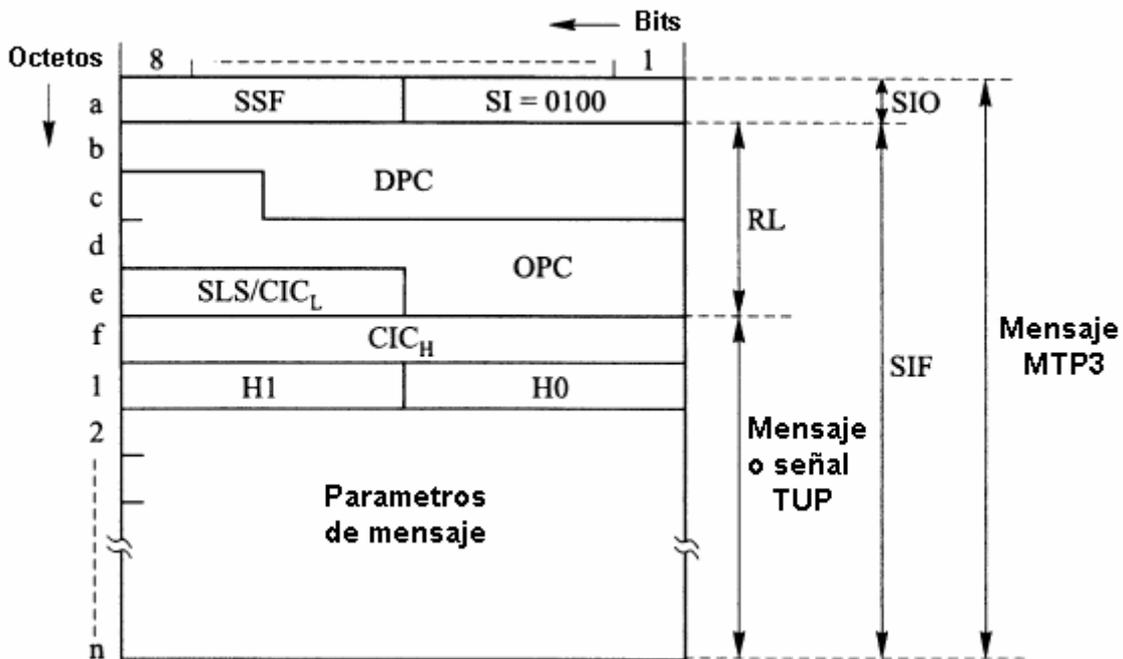


Figura 2.30 Formato general de mensajes TUP y señales

Los bits (8) a (5) tienen una función dual: ellos representan el SLS y los cuatro bits de mas bajo orden (CIC_L) de el código de identificación de circuito. Los bits de orden más alto (CIC_H) están en el octeto (f). Como resultado cada troncal tiene una ruta de señalización asociada, este es uno de los requerimientos para la entrega de mensajes en secuencia.

Sigla	Nombre	Codigo de tipo de mensaje
ACM	Address Complete	0000 0110
ANM	Answer	0000 1001
APM	Application Transport	0100 0001
CPG	Call Progress	0010 1100
CON	Connect	0000 0111
COT	Continuity	0000 0101
FOT	Forward Transfer	0000 1000
IAM	Initial Address	0000 0001
INF	Information	0000 0100
INR	Information Request	0000 0011
PAM	Pass-Along	0010 1000
REL	Release	0000 1100
RES	Resume	0000 1110
RLC	Release Complete	0001 0000
SUS	Suspend	0000 1101

Tabla 2.2 Tipos de mensajes

2.7.3.1.4 Funciones de tratamiento de mensajes de señalización

El tratamiento de los mensajes de señalización (SNM) está principalmente relacionado con los mensajes salientes y entrantes, los mensajes salientes que son proporcionados por las partes de usuario o por el mismo nivel 3 (función de la gestión de la red de señalización) debe encaminarse por los enlaces apropiados de señalización (función de encaminamiento de mensajes)

Los mensajes entrantes procedentes del nivel 2 (es decir, procedentes del terminal de señalización transmisor) deben entregarse a sus correspondientes destinos. Los mensajes pueden o no pertenecer al punto de señalización local en cuestión (discriminación de mensajes).

Si el mensaje pertenece al punto de señalización local, el destino puede ser el mismo nivel 3 (función de la gestión de la red de señalización) o de una de las partes de usuario (distribución de mensajes). Las funciones de encaminamiento, distribución y discriminación de mensajes, se llevan a cabo basándose en el análisis de la etiqueta de encaminamiento de cada uno de los mensajes.

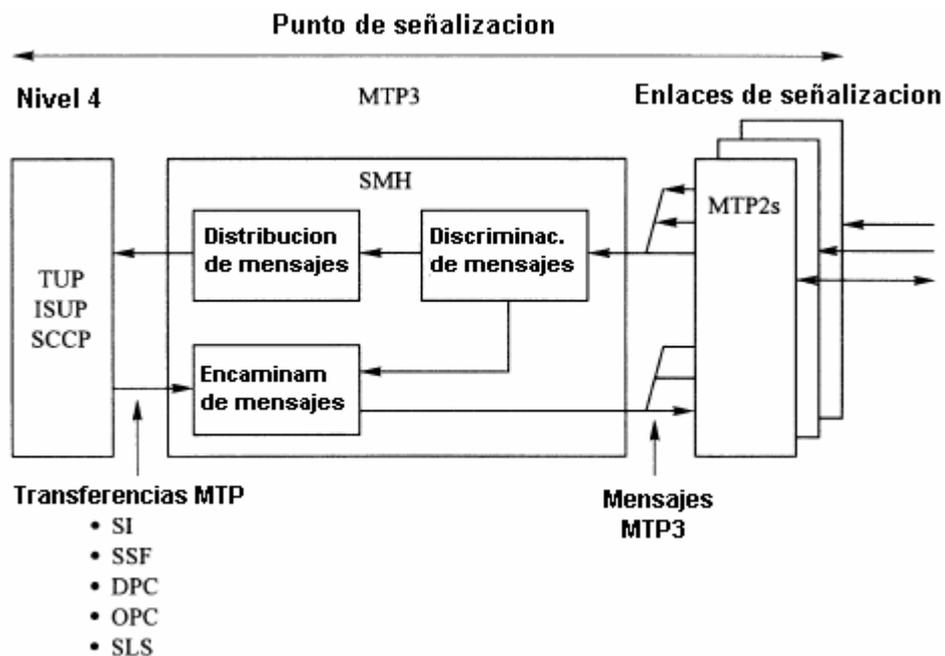


Figura 2.31 Estructura del tratamiento de mensajes de señalización

Las funciones de tratamiento de mensajes de señalización están divididas en tres grupos:

- **Discriminación de mensajes:** la discriminación de mensajes se basa en el análisis del código del punto de destino (DPC) para determinar si un mensaje recibido se destina o no a dicho punto. Cuando un punto de señalización recibe un mensaje que le está destinado, este debe

transferirse a la función de distribución; pero cuando los códigos no se corresponden, es decir cuando el mensaje no está destinado a dicho punto, el mensaje se transfiere a la función de encaminamiento en cuyo caso el punto de señalización actúa como un punto de transferencia de señalización (STP).

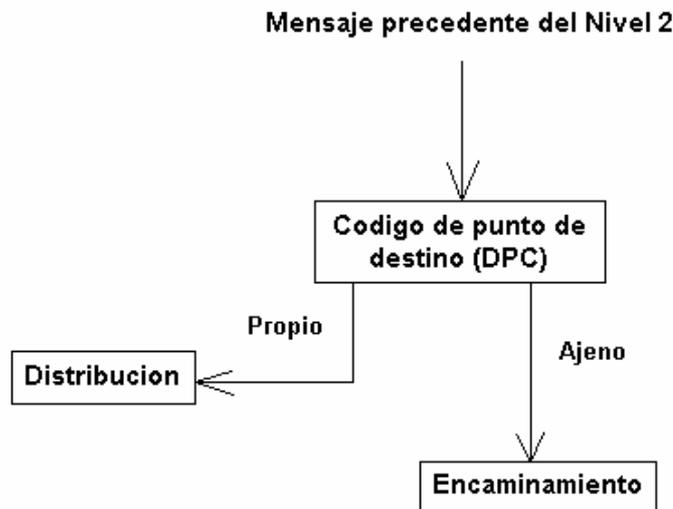


Figura 2.32 Función de discriminación de mensajes

- **Distribución de mensajes:** después de la discriminación el mensaje se distribuye de acuerdo con el indicador de servicio (SI), el cual identifica a las partes del usuario del nivel 4 o de nivel 3 (función de gestión de la red de señalización).

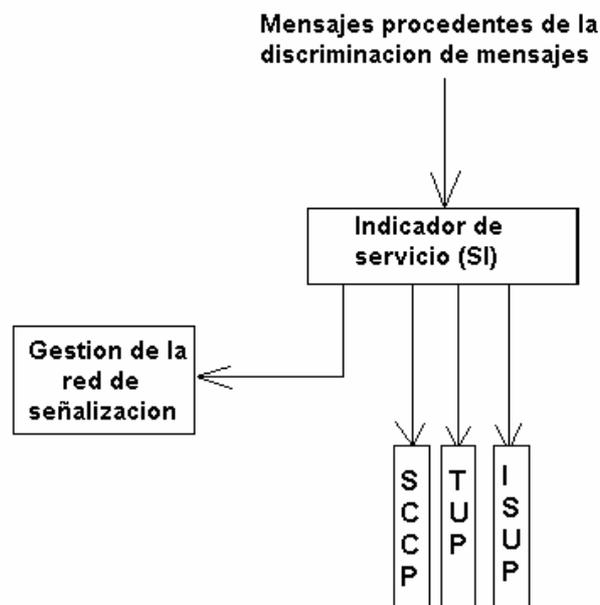


Figura 2.33 Función de distribución de mensajes

- **Encaminamiento de mensajes:** el encaminamiento de mensajes consiste en la selección del enlace de señalización apropiado de salida para la transmisión de un mensaje, de acuerdo con el código de punto de destino (DPC) y con la información sobre la selección del enlace de señalización (SLS), que forman parte de la etiqueta de enrutamiento. El SLS se utiliza para compartición de carga (del tráfico de señalización) entre los diversos enlaces de señalización que interconectan los dos puntos de señalización concernientes.

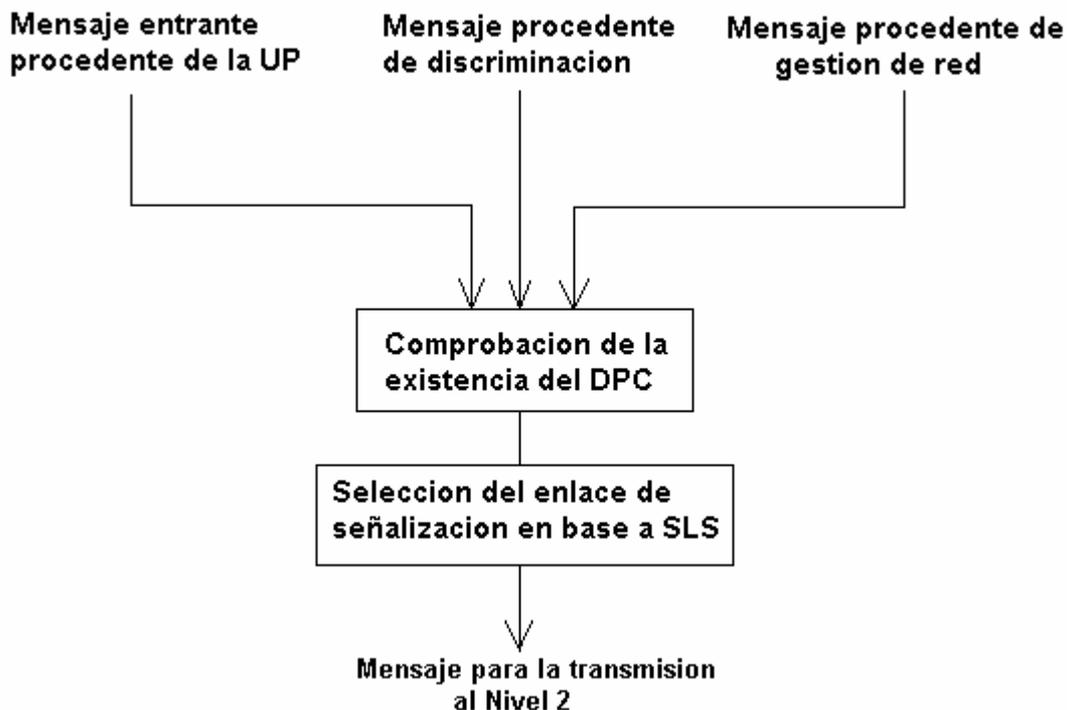


Figura 2.34 Función de encaminamiento de mensajes

2.7.3.2 GESTIÓN DE LA RED DE SEÑALIZACIÓN

El propósito de la gestión de red de señalización es de mantener el tráfico de mensajes de señalización operando fuera de condiciones anormales (congestiones, fallas) en la red de señalización. Algunas de esas condiciones podrían requerir una reducción temporaria o suspensión de mensajes salientes de tráfico a ciertos destinos. En ese caso la gestión de red de señalización alerta a los protocolos de nivel 4 en su punto de señalización.

La gestión de red de señalización (SNM) tiene interfases con los usuarios de MTP (protocolos de nivel 4), los MTP2s de los enlaces de señalización, y la función de encaminamiento de mensajes del tratamiento de mensajes de señalización en sus puntos de señalización. Esta también se comunica con las SNMs de otros puntos de señalización, enviando y recibiendo mensajes SNM. Los caminos pertenecientes a la gestión los podemos clasificar en:

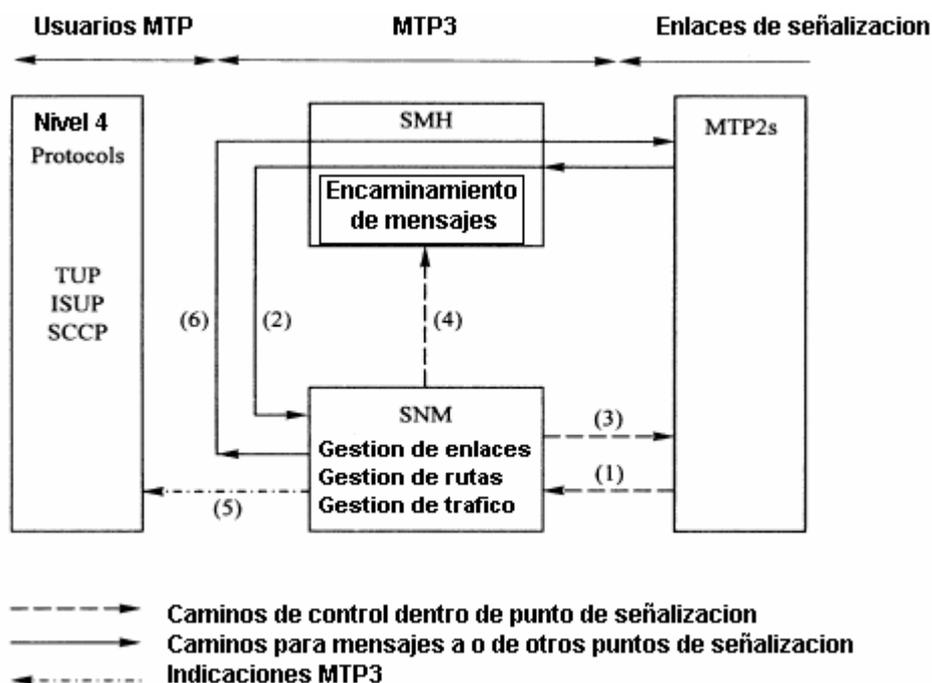


Figura 2.35 Estructura e interfaces de gestión de la red de señalización

Entradas: la gestión de red de señalización (SNM) basa sus acciones en las siguientes entradas (los números corresponden a los de la figura 2.35):

1. Indicaciones de los estados de los enlaces de señalización en sus puntos de señalización, recibidos de los MTP2s.
2. Mensajes SNM recibidos de otras funciones SNM en otros puntos de señalización.

Salidas: Las acciones SNM resultan en las salidas que se ven a continuación:

3. Comando a los MTP2s de los enlaces de señalización
4. Comando a la función de encaminamiento de mensajes del tratamiento de mensajes de señalización en sus puntos de señalización. Por ejemplo, desviar mensajes de señalización a ciertos destinos de su enlace de señalización a un enlace alternativo.
5. Indicaciones de los usuarios MTP en sus puntos de señalización, acerca del estado de rutas de señalización establecidas a sus destinos individuales. El estado, la pausa y la reanudación MTP indica que la ruta de señalización establecida a un destino particular ha de ponerse congestionada, indisponible y disponible otra vez.
6. Mensajes a SNMs de otros puntos de señalización.

2.7.3.2.1 Estructura de la gestión de la red de señalización

La gestión de la red de señalización consta de tres partes como se puede ver en la figura 2.35:

- **Gestión de enlaces de señalización:** su propósito es de establecer y mantener una capacidad de transferencia dada por medio de varios conjuntos de enlaces de señalización entre SP y SPT. Esta gestión monitorea y controla el estado de los MTP2 individuales (enlaces de señalización) del punto de señalización.
- **Gestión de rutas de señalización:** esta se comunica con sus funciones pares en otros puntos de señalización, enviando y recibiendo información denotando el estado de las rutas de señalización a destinos individuales.
- **Gestión del tráfico de señalización:** recibe información de la gestión de enlaces de señalización acerca del estado (disponible, no disponible) de los enlaces de señalización en sus puntos de señalización, e información de la gestión de rutas de señalización acerca de los problemas de las rutas de señalización seleccionada a sus destinos particulares. Cuando es necesario informa al encaminamiento de mensajes del tratamiento de mensajes de señalización y los usuarios MTP en sus puntos de señalización.

2.8 PARTE DE CONTROL DE LA CONEXIÓN DE SEÑALIZACIÓN (SCCP)

Es normalmente un protocolo no orientado a conexión. La señalización no orientada a conexión se refiere al intercambio de información si necesidad de establecer una configuración de conexión previa al intercambio de información. De esta forma, la información simplemente es enviada, pudiendo llegar al destino en un orden diferente al cual fue transmitida.

SCCP es el protocolo que permite que los mensajes sean utilizados por aplicaciones independientes dentro de un nodo. A estas aplicaciones se les conoce como subsistemas y entre ellas se puede mencionar el procesamiento de llamadas 1-8000, llamadas con tarjeta y redes inteligentes avanzadas.

Además de esto SCCP cuenta con un mecanismo de direccionamiento avanzado que le permite llevar a cabo la señalización entre dos puntos, aun cuando estos no conozcan la dirección del otro, este método es conocido como direccionamiento de título global.

La capa parte de control de la conexión de señalización (SCCP) se incluye por encima de la parte de transferencia de mensajes (MTP) de la red SS7 para proporcionar funciones adicionales de servicios de transferencia de información a nivel de red.

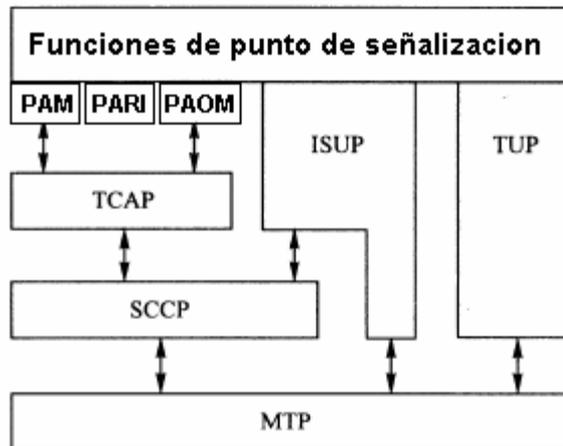


Figura 2.36 Posición de SCCP

Sirve como soporte (añadiendo la capa de capacidades de transacción (TCAP) a determinados usuarios de parte de aplicación, como:

- PAM: Parte de aplicación de móviles.
- PAOM: Parte de aplicación de operación y mantenimiento.
- PARI: Parte de aplicación de red inteligente.

2.9 PARTE DE APLICACIÓN DE LAS CAPACIDADES DE TRANSACCIÓN (TCAP)

Esta capa define los mensajes y el protocolo utilizado en los nodos para comunicarse entre aplicaciones. TCAP se emplea en los servicios de bases de datos y se encuentra definido únicamente para señalización no orientada a conexión, por lo que necesita de la capa SCCP para el transporte.

Las capacidades de transacción proporcionan funciones y protocolos a gran variedad de aplicaciones distribuidas entre centrales y centros especializados de las redes de telecomunicación. El término de "capacidades de transacción" se refiere a un conjunto de funciones de comunicación que proveen interfaz entre las aplicaciones y un servicio de capa de red.

La parte de aplicación de las capacidades de transacción (TCAP), es una interfaz entre aplicaciones de usuario y el servicio de la capa de red de N7. Se añade, por tanto, por encima del SCCP.

Por encima de TCAP se instalan los usuarios TCAP. Los usuarios TCAP dialogan entre sí mediante operaciones y respuestas, que se cursan a través de los servicios TCAP.

Las capacidades de transacción se dividen en dos partes o subcapas funcionales denominadas de componentes y de transacciones. Como puede observarse en la figura 2.37, la capa de componentes provee los servicios al usuario TCAP, mientras la capa de transacciones inter-actúa con el nivel de red (en nuestro caso la SCCP).

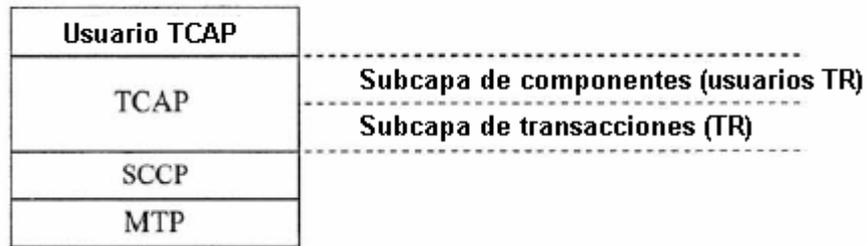


Figura 2.37 Estructura TCAP

Una operación es una petición de un usuario TCAP sobre el usuario TCAP remoto; es decir, una acción que debe llevar a cabo el extremo distante. Esta acción puede requerir una serie de parámetros asociados. La invocación de una operación se identifica mediante una identidad o referencia de invocación. Esta identidad de invocación permite que varias peticiones o invocaciones estén activas simultáneamente entre dos usuarios TCAP, referidas a una misma operación o a operaciones diferentes.

UNIDAD III
TRÁFICO TELEFÓNICO

INTRODUCCIÓN

En este capítulo se hará énfasis en los sistemas de colas, los cuales están compuestos de diferentes elementos, como clientes, servidores, líneas de esperas y otros. Estos elementos se detallarán en este capítulo y se verá como estos influyen en el proceso de una cola.

Se tratará aspectos importantes en un sistema de colas como lo son los procesos de arribo, tiempos de servicio, número de clientes, tiempos de espera, etc. Los cuales son parámetros muy importantes que rigen el funcionamiento o proceso del sistema.

Además se hará énfasis en el proceso de nacimiento y muerte el cual es la base para diferentes modelos de sistemas de colas, en las cuales se explicaran en cada una de ellas sus parámetros para diferenciarlos en cada sistema de cola.

Al final de este capítulo se analizarán conceptos de teletráfico como intensidad de tráfico, tasa de llamadas, tráfico cursado, tráfico recibido, etc. Para después analizar el modelo Erlang B y analizar sus características.

3.1 INTRODUCCIÓN A LOS SISTEMAS DE COLAS

La teoría de colas es el estudio matemático del comportamiento de líneas de espera. Esta se presenta, cuando los clientes llegan a un lugar demandando un servicio a un servidor, el cual tiene una cierta capacidad de atención. Si el servidor no está disponible inmediatamente y el cliente decide esperar, entonces se forma la línea de espera.

Una cola es una línea de espera y la teoría de colas es una colección de modelos matemáticos que describen sistemas de línea de espera particulares o sistemas de colas. Los modelos sirven para encontrar un buen compromiso entre costos del sistema y los tiempos promedio de la línea de espera para un sistema dado.

3.2 ELEMENTOS DE LOS SISTEMAS DE COLAS

Un sistema de colas puede dividirse en sus dos componentes de mayor importancia, la cola y la instalación de servicio. Las llegadas son las unidades que entran en el sistema para recibir el servicio. Siempre se unen primero a la cola; si no hay línea de espera se dice que la cola esta vacía. De la cola, las llegadas van a la instalación de servicio de acuerdo con la disciplina de la cola, que estudiaremos más adelante. Una vez que se completa el servicio, las llegadas se convierten en salidas, en la figura 3.1 se ven los componentes más importantes en un sistema de colas:

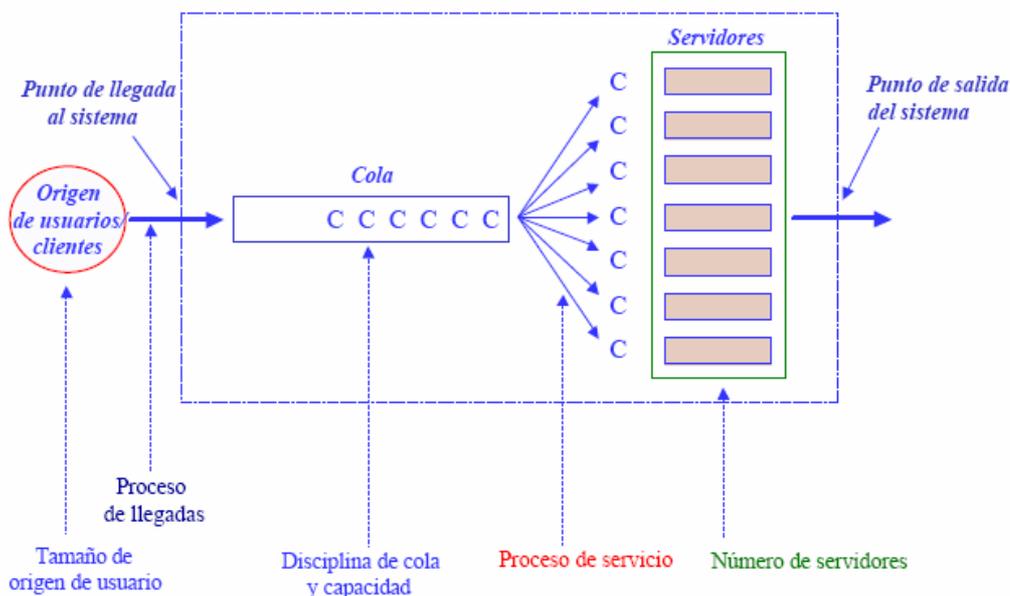


Figura 3.1 Esquema básico de un sistema de colas

Tamaño de la población: es el número total de clientes que pueden requerir servicio en determinado momento, es decir el número total de clientes potenciales distintos (puede suponerse de tamaño finito o infinito).

Cliente: Es todo individuo de la población potencial que solicita servicio. Suponiendo que los tiempos de llegada de clientes consecutivos son $0 < t_1 < t_2 < \dots$, será importante conocer el patrón de probabilidad según el cual la fuente de entrada genera clientes.

Número de servidores: indica el número de servidores de que dispone el sistema para procesar las demandas de servicio, trabajando en paralelo; lo designaremos por la letra K. Por ejemplo, podemos decir que el sistema tiene $K = 2$ servidores.

Capacidad de la cola: Es el máximo número de clientes que pueden estar haciendo cola (antes de comenzar a ser servidos). De nuevo, puede suponerse finita o infinita. Lo más sencillo, a efectos de simplicidad en los cálculos, es suponerla infinita. Aunque es obvio que en la mayor parte de los casos reales la capacidad de la cola es finita, no es una gran restricción el suponerla infinita si es extremadamente improbable que no puedan entrar clientes a la cola por haberse llegado a ese número límite en la misma.

Disciplina de la cola: Es el modo en el que los clientes son seleccionados para ser servidos. Las disciplinas más habituales son:

La disciplina FIFO (first in first out), según la cual se atiende primero al cliente que antes haya llegado, o el primer cliente en llegar es el primer cliente en salir.

La disciplina LIFO (last in first out): consiste en atender primero al cliente que ha llegado de último, o el último cliente en llegar es el primer cliente en salir.

La disciplina SIRO (service in random order), que selecciona a los clientes de forma aleatoria, no posee ningún patrón establecido.

Mecanismo de servicio: Es el procedimiento por el cual se da servicio a los clientes que lo solicitan. Para determinar totalmente el mecanismo de servicio debemos conocer el número de servidores de dicho mecanismo (si dicho número fuese aleatorio, la distribución de probabilidad del mismo) y la distribución de probabilidad del tiempo que le lleva a cada servidor dar un servicio. En caso de que los servidores tengan distinta destreza para dar el servicio, se debe especificar la distribución del tiempo de servicio para cada uno.

La cola: es el conjunto de clientes que hacen espera, es decir los clientes que ya han solicitado el servicio pero que aún no han pasado al mecanismo de servicio.

El sistema de la cola: es el conjunto formado por la cola y el mecanismo de servicio, junto con la disciplina de la cola, que es lo que nos indica el criterio de qué cliente de la cola elegir para pasar al mecanismo de servicio.

Tasa de llegada: indica el número medio de clientes que llegan al sistema (demandando servicio) por unidad de tiempo; se designa por la letra griega λ . Por ejemplo, podemos decir que la tasa de llegada del sistema es de $\lambda = 5$ clientes por minuto.

Tasa de servicio: indica el número medio de clientes que un servidor puede procesar por unidad de tiempo; se designa por la letra griega μ . Por ejemplo, podemos decir que la tasa de servicio de cada servidor es de $\mu = 6$ clientes por minuto.

3.2.1 Magnitudes fundamentales en un sistema de espera

En la figura 3.2 se pueden apreciar las diferentes magnitudes presentes en un sistema de espera en el cual se observa el comportamiento de los clientes en este sistema, en donde secuencialmente ingresan a la cola permaneciendo con determinado tiempo en esta, luego llegan al servidor en donde permanecen otro tiempo para que finalmente abandonen el sistema

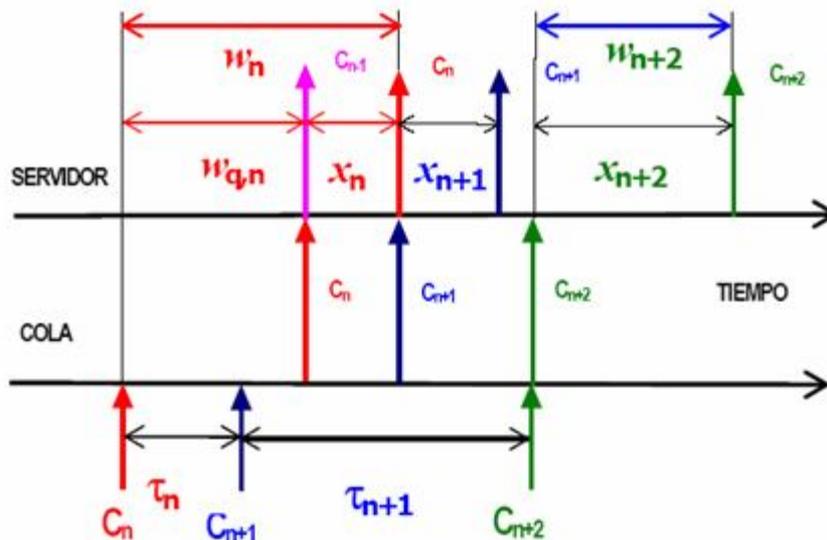


Figura 3.2 Magnitudes de un sistema de colas

Donde tenemos:

- τ_n = tiempo entre llegada del cliente n y el cliente n+1.
- w_n = tiempo de permanencia en el sistema de espera del cliente n.
- $w_{q,n}$ = tiempo de permanencia en cola del cliente n.
- x_n = tiempo de servicio del cliente n.
- C_n = cliente n.

Tiempo medio entre llegadas: El tiempo medio entre las llegadas de dos clientes consecutivos es de $1/\lambda$ unidades de tiempo.

Tiempo medio de servicio: El tiempo medio de servicio de un cliente es de $1/\mu$ unidades de tiempo.

En la figura 3.2 se puede observar como un cliente C_n ingresa en la cola pasa un determinado tiempo en ella w_{q_n} para luego ingresar al servicio en donde transcurre otro tiempo x_n para después abandonar el sistema, al determinado tiempo entre la entrada a la cola del cliente n y el cliente n+1 se denomina τ_n , luego el cliente n+1 (C_{n+1}), realiza los mismo pasos ya mencionados de C_n

3.3 ANÁLISIS DEL RETARDO Y FÓRMULA DE LITTLE

La figura 3.3 muestra un modelo básico para un sistema retardo/pérdida. Los clientes llegan al sistema de acuerdo con algún patrón de llegada, presentando peticiones de conexión mensajes individuales, paquetes o celdas.

El sistema puede ser una línea de transmisión individual, un multiplexor, un conmutador o incluso una red completa. El cliente pasa algún tiempo T en el sistema y después de este tiempo el cliente sale del sistema. Es posible que bajo ciertas condiciones el sistema esté en un estado de bloqueo, por ejemplo, debido a la falta de recursos. Los clientes que llegan al sistema cuando esté en este estado quedan bloqueados o se pierden.



Figura 3.3 Análisis del retardo de red (Imagen tomada del libro Redes de comunicación, García y Widjaja, Segunda edición).

Nos interesan las siguientes medidas de prestaciones:

- Tiempo pasado en el sistema: T
- Número de clientes en el sistema: N(t)
- Fracción de clientes que llegan que se pierden o se bloquean: P_b
- Número medio de mensajes/segundo que pasan a través del sistema: rendimiento

3.3.1 Definiciones de tasas de llegada y carga de tráfico

Sea $A(t)$ el número de llegadas al sistema en el intervalo de tiempo de 0 a t . Sea $B(t)$ el número de clientes bloqueados y $D(t)$ el número de clientes que salen en el mismo intervalo de tiempo. El número de clientes $N(t)$ en el sistema en un instante t viene dado por:

$$N(t) = A(t) - D(t) - B(t)$$

Ya que el número que ha entrado en el sistema hasta el instante t es $A(t) - B(t)$ y ya que $B(t)$ de esos clientes se han ido en el instante t . Se supone que el sistema está vacío en el instante $t=0$.

A largo plazo la velocidad de llegada al sistema viene dada por:

$$\lambda = \lim_{t \rightarrow \infty} \frac{A(t)}{t} \text{ clientes / segundo}$$

El rendimiento del sistema es igual a la velocidad de salida a largo plazo, que viene dada por:

$$\text{Rendimiento} = \lim_{t \rightarrow \infty} \frac{D(t)}{t} \text{ clientes / segundo}$$

El número medio en el sistema viene dado por:

$$E[N] = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t N(t') dt' \text{ clientes}$$

La fracción de clientes bloqueados es por tanto:

$$P_b = \lim_{t \rightarrow \infty} \frac{B(t)}{A(t)}$$

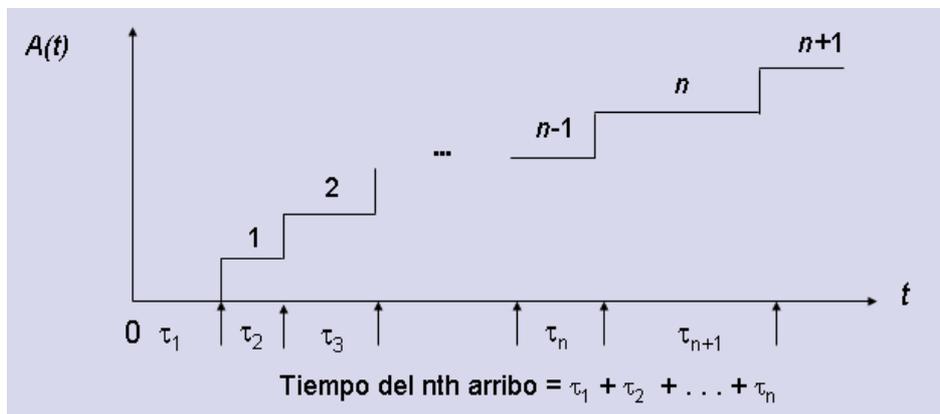


Figura 3.4 Proceso de llegadas de clientes (Imagen tomada del libro *Redes de comunicación*, Garcia y Widjaja, Segunda edición).

La figura 3.4 representa una función muestra típica $A(t)$ que representa el número de llegadas al sistema. Se supone que se empiezan a contar los clientes en el instante $t=0$. El primer cliente llega en el instante τ_1 y $A(t)$ cambia de 0 a 1 en este instante.

La llegada del n -ésimo cliente es en el instante $\tau_1 + \dots + \tau_n$ donde τ_i es el tiempo transcurrido entre la llegada del cliente $i-1$ y el i .

La velocidad de llegada en el instante en el que el n -ésimo cliente llega viene dada por $n/(\tau_1 + \dots + \tau_n)$ clientes /segundo.

Por tanto la velocidad de llegada a largo plazo viene dada por:

$$\lambda = \lim_{t \rightarrow \infty} n/(\tau_1 + \dots + \tau_n) = \lim_{t \rightarrow \infty} 1/(\tau_1 + \dots + \tau_n)/n = 1/E[\tau]$$

Se supone que todos los tiempos entre llegadas son estadísticamente independientes y tienen la misma distribución de probabilidad y que el valor medio o esperado viene dado por $E[t]$; por tanto la velocidad de llegada viene dada por el recíproco del tiempo medio entre llegadas.

3.3.2 Fórmula de little

Relaciona el tiempo medio pasado en el sistema $E[T]$ y la velocidad de llegada λ con el número medio de clientes en el sistema $E[N]$ por medio de la siguiente fórmula:

$$E[N] = \lambda E[T]$$

Suponemos que el sistema no bloquea a ningún cliente, como se ve en la figura 3.4. El número en el sistema $N(t)$ varía según $A(t) - D(t)$. Supóngase que se representan $A(t)$ y $D(t)$ en el mismo gráfico de la figura 3.6.

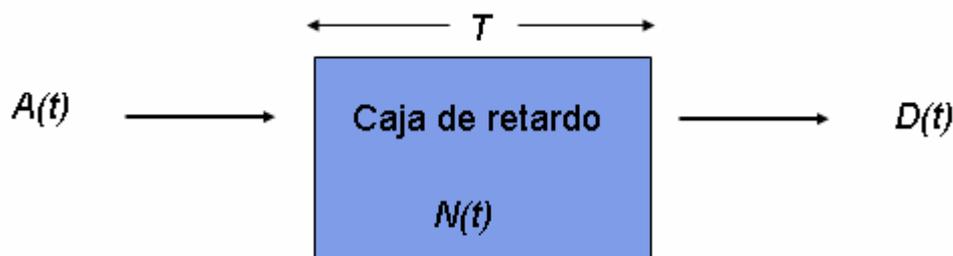


Figura 3.5 Esquema de proceso de colas (Imagen tomada del libro Redes de comunicación, García y Widjaja, Segunda edición).

$A(t)$ aumenta 1 cada vez que un cliente llega, y $D(t)$ aumenta 1 cada vez que se va un cliente. El número de clientes en el sistema $N(t)$ viene dado por diferencia entre $A(t)$ y $D(t)$. El número de salidas nunca puede ser mayor que el número de llegadas, y por tanto $D(t)$ no puede estar por detrás de $A(t)$ como se muestra en la figura 3.6.

Supongamos que los clientes se atienden en forma de FIFO (el primero que entra, es el primero en salir), entonces el tiempo T_1 utilizado por el primer cliente es el tiempo transcurrido entre el instante en el que $A(t)$ pasa de 0 a 1 y en el instante en el que $D(t)$ pasa de 0 a 1, siendo T_1 el área del rectángulo definido por esos dos instantes en la figura 3.6.

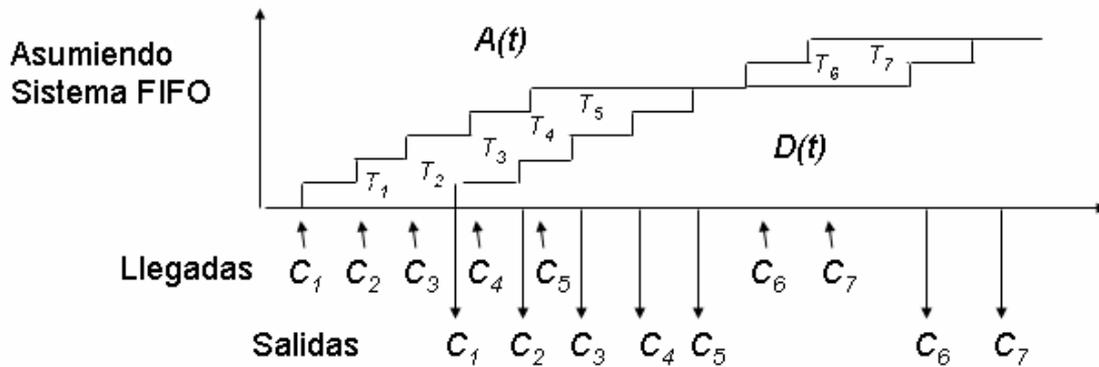


Figura 3.6 Medidas en sistemas de colas (Imagen tomada del libro Redes de comunicación, García y Widjaja, Segunda edición).

Considérese un instante t_0 en el que $D(t)$ alcanza a $A(t)$; es decir $N(t_0) - A(t_0) - D(t_0) = 0$. Nótese que el área entre $A(t)$ y $D(t)$ viene dada por la suma de los tiempos T_0 empleados en el sistema por los primeros $A(t_0)$ clientes.

El tiempo medio del número de clientes en el sistema hasta el instante t_0 es:

$$\frac{1}{t_0} \int_0^{t_0} N(t') dt' = \frac{1}{t_0} \sum_{j=1}^{A(t_0)} T_j$$

Si se multiplica y divide la expresión anterior por $A(t_0)$ se obtiene:

$$\frac{1}{t_0} \int_0^{t_0} N(t') dt' = \frac{A(t_0)}{t_0} \left\{ \frac{1}{A(t_0)} \sum_{j=1}^{A(t_0)} T_j \right\}$$

Esta ecuación quiere decir que hasta el instante t_0 el número medio de clientes en el sistema viene dado por el producto de la velocidad media de llegada $A(t_0)/t_0$ y la media aritmética del tiempo pasado en el sistema por los primeros $A(t_0)$ clientes. La fórmula de Little supone que:

$$E[N] = \lambda E[T]$$

Ahora consideremos un sistema en que los clientes se pueden bloquear. La fórmula anterior implica que si reemplazamos $A(t)$ por $A(t) - B(t)$, tenemos el verdadero número de clientes que entran en el sistema. La verdadera velocidad de llegada al sistema con bloqueo es $\lambda (1 - P_b)$ donde P_b es la fracción de llegadas que están bloqueadas, por tanto la fórmula de Little para un sistema con bloqueo es: $E[N] = \lambda (1 - P_b) E[T]$. Esta fórmula es válida para un sistema que puede ser una línea de transmisión individual, un multiplexor, un conmutador o una red.

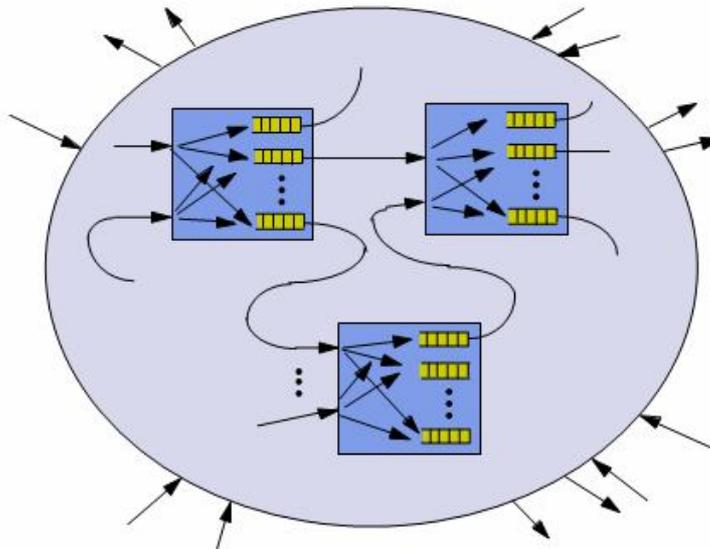


Figura 3.7 Red de conmutación de paquetes (Imagen tomada del libro Redes de comunicación, Garcia y Widjaja, Segunda edición).

En la figura 3.7 se muestra una red de conmutación de paquetes formada por conmutadores de paquetes interconectados. Suponemos que cuando un paquete llega a un conmutador el paquete es encaminado instantáneamente y situado en un multiplexor para esperar su transmisión por una línea de salida. Entonces cada conmutador de paquetes se puede ver como un conjunto de multiplexores. Empezamos a aplicar la fórmula de Little a la red como un todo. Sea N_{red} el número total de paquetes en la red; T_{red} el tiempo que pasa el paquete en la red y λ_{red} la velocidad total de llegada del paquete a la red.

Entonces la fórmula de Little establece $E[N_{red}] = \lambda_{red} E[T_{red}]$

Esta fórmula implica que el retardo medio experimentado por los paquetes al atravesar la red es:

$$E[T_{red}] = E[N_{red}] / \lambda_{red}$$

Podemos refinar la ecuación anterior aplicando la fórmula de Little a cada multiplexor individualmente. Para el multiplexor m la fórmula de Little da:

$E[N_m] = \lambda_m E[T_m]$ donde λ_m es la velocidad de llegada del paquete al multiplexor y $E[T_m]$ es el tiempo medio que ha pasado el paquete en el multiplexor.

El número total de paquetes en la red N_{red} es igual a la suma de los paquetes en todos los multiplexores:

$$E[N_{red}] = \sum_m E[N_m] = \sum \lambda_m E[T_m]$$

Combinando las tres ecuaciones anteriores se obtiene una expresión para el retardo total experimentado por un paquete cuando atraviesa toda la red:

$$E[T_{red}] = E[N_{red}] / \lambda_{red} = \{ \sum_m \lambda_m E[T_m] \} / \lambda_{red}$$

Por tanto el retardo de la red depende de la velocidad total de llegada a la red, de la velocidad de llegada a los multiplexores individuales, y del retardo en cada multiplexor. La velocidad de llegada a cada multiplexor viene dada por el algoritmo de encaminamiento. El retardo en un multiplexor depende de la velocidad de llegada y de la velocidad a la que la línea de transmisión asociada pueda transmitir paquetes.

Entonces la fórmula anterior simplemente incorpora el efecto del encaminamiento así como el efecto de las capacidades de las líneas de transmisión en la red. Por este motivo la expresión anterior se usa frecuentemente en el diseño y gestión de redes de conmutación de paquetes.

3.4 MODELOS BÁSICOS DE COLAS

El primer trabajo de Erlang sobre ingeniería del tráfico en sistemas telefónicos llevó al desarrollo de varios modelos fundamentales para el análisis de los sistemas con recursos compartidos. En una aplicación típica los clientes solicitan recursos en instantes aleatorios y los utilizan durante un tiempo variable. Cuando todos los recursos están en uso, es posible que cuando lleguen clientes formen una línea o cola para esperar a que los recursos estén disponibles. La teoría de colas trata del análisis de estos tipos de sistemas.

3.4.1 Proceso de arribo

El proceso de arribo para modelos de población infinita es visto en términos de tiempo entre arribos de un cliente y el próximo. Los arribos pueden ocurrir en forma aleatoria o ser planificados

El modelo más importante para los arribos aleatorios es el proceso de Poisson. La tasa de arribos es de λ clientes por unidad de tiempo. El número de arribos en un intervalo de tiempo de longitud t , $N(t)$, sigue una distribución de Poisson con media λt .

La figura 3.8 muestra los elementos básicos de un sistema de colas. Los clientes llegan al sistema en instantes de llegada $\tau_1, \tau_2 \dots$; Suponemos que los intervalos de tiempo de llegada son variables aleatorias independientes con la misma distribución. Los resultados del proceso de llegada de la figura 3.4 siguen siendo válidos.

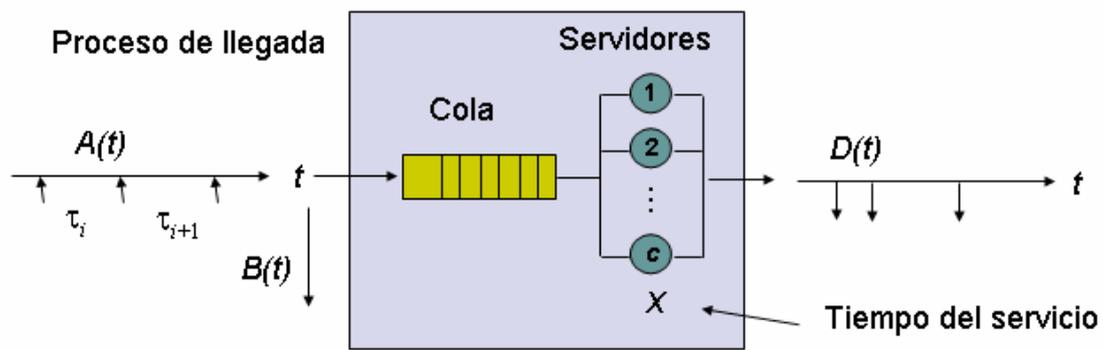


Figura 3.8 Modelo de cola (Imagen tomada del libro Redes de comunicación, Garcia y Widjaja, Segunda edición).

En particular la velocidad de llegada al sistema viene dada por

$$\lambda = \frac{1}{E[\tau]} \text{ cliente / segundo}$$

3.4.2 PROCESO DE POISSON

El proceso de Poisson satisface las siguientes hipótesis, donde $P[X]$ significa la probabilidad de X .

- 1) P [un arribo en el intervalo de tiempo $(t, t+h)$, para $h \rightarrow 0$] = λh , donde λ es una constante.
- 2) P [más de un arribo en el intervalo $(t, t+h)$, para $h \rightarrow 0$] $\rightarrow 0$.
- 3) Las probabilidades ya vistas no dependen de t (propiedad de no memoria= tiempo de independencia= estacionario).

Definimos $p_n(t) = P[n \text{ arribos en el intervalo de tiempo } (0, t)]$. Usando las anteriores hipótesis 1) y 2), es posible expresar la probabilidad $p_n(t+h)$, $h \rightarrow 0$:

$$p_n(t+h) = p_n(t) [1 - \lambda h] + p_{n-1}(t) \lambda h$$

(La probabilidad de estar en el estado n equivale a que estando en el estado n no se presenten llamadas o estando en el estado $n-1$ se presente una llamada).

$$p_0(t+h) = p_0(t)[1 - \lambda h]$$

(no arribos para t , no mas arribos) (1)

La ecuación (1) puede ser escrita de esta manera:

$$\frac{p_n(t+h) - p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad \frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t) \quad (2)$$

Debido a lo pequeño del término h los términos a la izquierda de (2) pueden ser considerados como derivadas:

$$\frac{dp_n(t)}{dt} = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad \frac{dp_0(t)}{dt} = -\lambda p_0(t) \quad (3)$$

Las ecuaciones (3) representan una serie de ecuaciones diferenciales. Al resolverla despejamos para $p_0(t)$:

$p_0(t) = e^{-\lambda t}$ (Verificado esta, calculamos la derivada de $p_0(t)$ y la comparamos con (3))

Usando $p_0(t)$ es posible calcular $p_1(t)$:

$p_1(t) = \lambda t e^{-\lambda t}$ (Verificado por calculo la derivada y reemplazando en (3) con $p_0(t)$)

O generalmente (puede ser verificada por inducción matemática):

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (4)$$

Debido a las hipótesis 3) la formula (4) se aferra a cualquier intervalo $(s, s+t)$, en otras palabras la probabilidad de n arribos durante algún intervalo de tiempo solo depende del largo de este intervalo de tiempo (no del inicio del intervalo de tiempo).

Número de arribos N_t durante algún intervalo de tiempo t

Es una variable discreta al azar a asociada con los procesos de Poisson. Teniendo las probabilidades de valores aleatorias (4), es posible encontrar los parámetros usuales de la variable aleatoria N_t . definiendo $E[x]$ como el valor promedio, $\text{Var}[x]$ la variabilidad, y $\text{Std}[x]$ la desviación estándar de la variable aleatoria x :

$$E[N_t] = \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda t e^{-\lambda t} \sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1}}{(n-1)!} = \lambda t e^{-\lambda t} e^{\lambda t} = \lambda t \quad (5)$$

La ecuación (5) da la interpretación de una constante λ , que es el número promedio de arribos por unidad de tiempo. Con lo que λ es llamada tasa de arribo.

$$\text{Var}[N_t] = \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!} e^{-\lambda t} - (\lambda t)^2 = \dots = \lambda t \quad (6)$$

$$\text{Std}[N_t] = \sqrt{\text{Var}[N_t]} = \sqrt{\lambda t} \quad (7)$$

Otra variable aleatoria asociada con los procesos de Poisson es el intervalo de tiempo entre dos arribos consecutivos. Definimos x como el intervalo cualquiera. A encontrar esta distribución, expresamos la función distribución $F(x)$ así:

$$F(x) = P[\text{intervalo} < x] = P[\text{en al menos un arribo durante el intervalo de tiempo } x] = 1 - p_0(x) = 1 - e^{-\lambda x}$$

Debido a que el intervalo es una variable aleatoria continua, es posible calcular la densidad de probabilidad como una derivada de la función distribución:

$$f(x) = dF(x)/dx = \lambda e^{-\lambda x}$$

Estas dos funciones definen la distribución de probabilidad:

$$F(x) = 1 - e^{-\lambda x}, f(x) = \lambda e^{-\lambda x} \quad (8)$$

La distribución (8) es llamada distribución exponencial o distribución exponencial negativa. Esos parámetros son:

$$E[x] = \int_0^{\infty} t e^{-\lambda t} dt = \frac{1}{\lambda} \quad \text{Var}[x] = E[x^2] - (E[x])^2 = \frac{1}{\lambda^2} \quad \text{Std}[x] = \sqrt{\text{Var}[x]} = \frac{1}{\lambda} \quad (9)$$

La ecuación (9) da otra interpretación de la constante λ . Su valor invertido es el intervalo promedio entre arribos. Como el número de arribos, la distribución de intervalos entre arribos no depende del tiempo. Note el gran valor de la desviación estándar (igual al valor promedio) esto significa, que la distribución exponencial es muy aleatoria. De hecho todos los intervalos de 0 a ∞ son teóricamente posibles. Entonces eso puede pasar, que habrá grupos de arribos muy cerrados separados por intervalos muy grandes. De hecho no todas las distribuciones prácticas pueden ser tratadas como exponenciales. Como dice esta regla general:

“Sin un largo número de eventos independientes fluyen, estos caen juntos, y si un evento en cada cascada de eventos ocurre a muy baja tasa, entonces la fluidez resultante es aproximadamente igual al proceso de Poisson con intervalos exponenciales”.

Esta regla es más o menos satisfecha para arribos en sistemas con grandes poblaciones (cada cliente potencial representa una baja tasa de fluidez)

3.4.3 Tiempos de servicio

Es el tiempo que transcurre desde el inicio del servicio para un cliente hasta su terminación. Los tiempos de servicio pueden ser constantes o aleatorios. Los aleatorios están modelados por distribuciones probabilísticas.

Un sistema de colas consiste de un número de centros de servicio y colas de espera interconectados. Cada centro de servicio consiste de uno o más servidores: un servidor, múltiples servidores, infinitos servidores.

Los recursos se denominan servidores porque su función es servir las peticiones del cliente. El tiempo necesario para servir a un cliente se llama tiempo de servicio y se denota con X . El servidor es usualmente una línea de transmisión y el tiempo de servicio puede ser el tiempo requerido para transmitir un mensaje o la duración de una llamada telefónica.

La velocidad máxima a la que un servidor puede atender a clientes se obtiene cuando el servidor está continuamente ocupado. En este caso el tiempo medio entre las salidas de los clientes es igual al tiempo de servicio medio. La capacidad de procesamiento de un único servidor viene dado por el máximo rendimiento o velocidad de salida. Del razonamiento que condujo a la fórmula de la velocidad de llegada, la capacidad de procesamiento viene claramente dada por:

$$\mu = \frac{1}{E[X]} \text{cliente/segundo}$$

La variabilidad y la desviación estándar pueden ser calculadas por reemplazar λ por μ en la fórmula (9), a diferencia de los arribos, el servicio exponencial es una abstracción que es difícilmente satisfecha por sistemas, debido a que en la mayoría de los casos es muy diferente tener tiempos de servicio muy largos que tiempos de servicio muy cortos. La duración del servicio real será típicamente menos aleatoria que la distribución exponencial teórica.

Otro parámetro muy importante en los sistemas de colas es el ratio ρ del arribo y las tasas de servicio llamado tasa o velocidad de tráfico. (a veces llamado intensidad de tráfico o factor de utilización):

$$\rho = \frac{\lambda}{\mu} \quad (10)$$

La capacidad de procesamiento μ se puede comparar con el flujo máximo que puede soportar una cascada de eventos. El número de servidores c en un sistema de cola puede ser mayor que uno. La capacidad total de procesamiento de un sistema de cola viene dada por $c\mu$ clientes/segundo.

Un sistema ideal de cola es aquel en el que los clientes llegan a intervalos iguales y en el que requieren un tiempo de servicio constante. Mientras el tiempo de servicio sea menos que el tiempo entre llegadas, cada cliente llega a un servidor disponible y no hay tiempo de espera. En general el tiempo entre llegadas y los tiempos de servicio son aleatorios.

La combinación de un tiempo de servicio largo seguido por un tiempo entre llegadas corto puede conducir a una situación en la que el servidor no esté disponible cuando llegue un cliente. Por este motivo en muchas aplicaciones se crea una cola de forma que el cliente puede esperar a que haya un servidor disponible, como se muestra en la figura 3.8.

Cuando un servidor queda disponible, se selecciona al siguiente cliente para recibir el servicio según los criterios del mismo, FIFO, LIFO, según prioridad de la clase y orden aleatorio del servicio. Normalmente se suponen tipos de servicio FIFO.

El número máximo de clientes permitido en un sistema de cola se denota con k . Nótese que k incluye tanto a los clientes en cola como a los que están atendándose. Denotamos el número total de clientes en el servicio con $N(t)$; el número en cola, con $N_c(t)$ y el número en servicio con $N_s(t)$. Cuando un sistema está completo, es decir, $N(t) = k$, las llegadas de nuevos clientes se bloquean o se pierden.

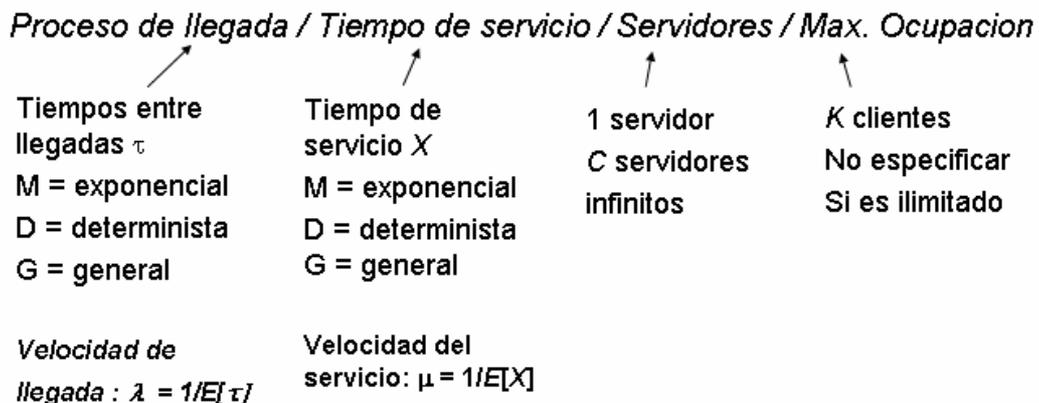
3.4.4 Clasificación de un sistema de cola

Los sistemas de cola se clasifican según las siguientes características:

- Patrón de llegada de un cliente
- Distribución del tiempo de servicio.
- Número de servidores
- Número máximo en el sistema.

Notación KENDALL

En los modelos de colas se emplea esta notación con el propósito de identificar los diversos modelos que se presentan:



Modelos de multiplexor: $M/M/1/K$, $M/M/1$, $M/G/1$, $M/D/1$
 Modelos de líneas principales: $M/M/c/c$, $M/G/c/c$
 Actividad del usuario: $M/M/\infty$, $M/G/\infty$

Un ejemplo, el sistema de cola $M/M/1/K$ corresponde a un sistema en el que los tiempos entre llegadas están distribuidos exponencialmente (M); los tiempos de servicio están exponencialmente distribuidos (M); hay un único servidor (1); y al menos se permiten K clientes en el sistema.

La figura 3.9 muestra los parámetros que se usan para analizar un sistema de cola.

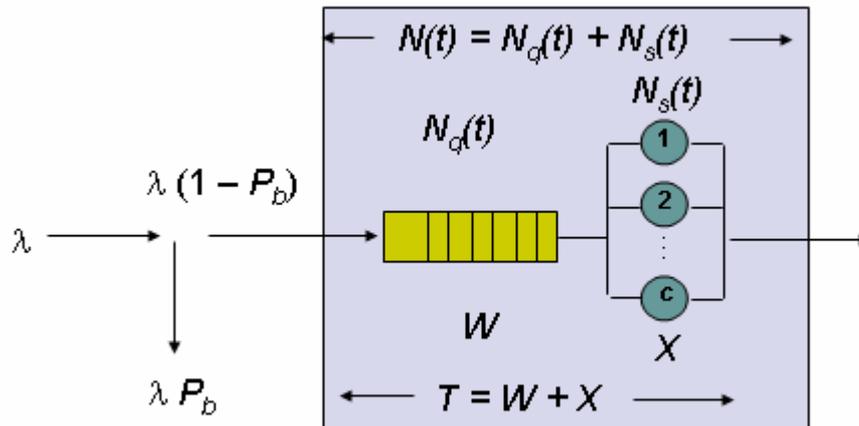


Figura 3.9 Medidas de desempeño (Imagen tomada del libro Redes de comunicación, García y Widjaja, Segunda edición).

$N(t)$ = número en el sistema; $N_q(t)$ = Número en cola; $N_s(t)$ = Número en servicio
 T = retardo total; W = tiempo de espera; X = tiempo de servicio; 1,2, c = servidores

El tiempo total que pasa un cliente en el sistema se denota con T , que representa el tiempo empleado en esperar en la cola W más el tiempo empleado en servicio X .

Cuando un sistema está bloqueado, P_b denota la fracción de clientes que están bloqueados, por tanto, la velocidad de llegada real al sistema viene dada por $\lambda(1 - P_b)$. Este es el valor de la velocidad de llegada que debería emplearse cuando se aplica la fórmula de Little. Por tanto el número medio en el sistema y el retardo medio en el sistema están relacionados según:

$$E[N] = \lambda(1 - P_b)E[T]$$

Si aplicamos la fórmula de Little cuando el “sistema” es precisamente la cola, entonces el número medio de clientes en cola y el tiempo medio de espera están relacionados según:

$$E[N_q] = \lambda(1 - P_b)E[W]$$

Finalmente si el “sistema” se define como un conjunto de servidores, entonces el número medio de clientes en servicio y el tiempo de servicio medio están relacionados según:

$$E[N_s] = \lambda(1 - P_b)E[X]$$

3.4.5 Factor de utilización y estabilidad

Intuitivamente, todos comprendemos, por experiencia personal, que un sistema de colas puede estar más o menos congestionado, y que puede llegar a colapsarse. Los periódicos se hacen eco a menudo de noticias sobre los atascos que se producen en contextos variados, ejemplo: judicial o sanitario.

En sistemas informáticos, también se producen fenómenos de congestión cuando el tráfico de paquetes de bits aumenta de forma excesiva, lo cual causa tiempos de respuesta grandes, o incluso el colapso del sistema. La principal medida del nivel de saturación de un sistema de colas del tipo considerado en este curso es el factor de utilización, que se designa por la letra griega ρ

ρ = Tasa media de llegada de trabajo por unidad de tiempo/ capacidad máxima de procesamiento de trabajo por unidad de tiempo

Consideremos el caso de una cola con tasa de llegada λ , tasa de servicio μ , K servidores en paralelo, y espacio de almacenamiento de clientes ilimitado. Calculemos su factor de utilización, comenzando por el numerador. Sabemos que llegan al sistema λ clientes por unidad de tiempo. Cada cliente trae consigo una carga de trabajo media igual a su tiempo medio de servicio, que es de $1/\mu$ unidades de tiempo. Por tanto, tenemos que

Tasa media de llegada de trabajo por unidad de tiempo = $\lambda \times \frac{1}{\mu}$

Continuemos con el denominador. Como el sistema tiene K servidores en paralelo, y cada uno puede procesar una unidad de trabajo por unidad de tiempo (ya que el trabajo se mide en tiempo), tendremos que:

Capacidad máxima de procesamiento de trabajo por unidad de tiempo = K

Ya que tal tasa máxima se alcanza cuando todos los servidores están ocupados. Por tanto, en tal sistema, podemos calcular el factor de utilización como:

$$\rho = \frac{\lambda}{K\mu}$$

Conocer el valor del factor de utilización nos permite determinar si el sistema de interés está o no saturado/colapsado, o, en terminología de TC, si es o no estable. Intuitivamente, un sistema de colas es estable si puede procesar trabajo más rápido de lo que este llega. En caso contrario, es inestable, es decir, se colapsa al aumentar su congestión sin cesar. Ello motiva la siguiente definición.

3.5 PROCESO DE NACIMIENTO Y MUERTE

La mayor parte de los modelos elementales de colas suponen que las entradas (llegadas de clientes) y las salidas (clientes que se van) del sistema ocurren de acuerdo al proceso de nacimiento y muerte. Lo cual se interpreta así:

Nacimiento: Llegada de un nuevo cliente al sistema de colas.

Muerte: Salida del cliente servido.

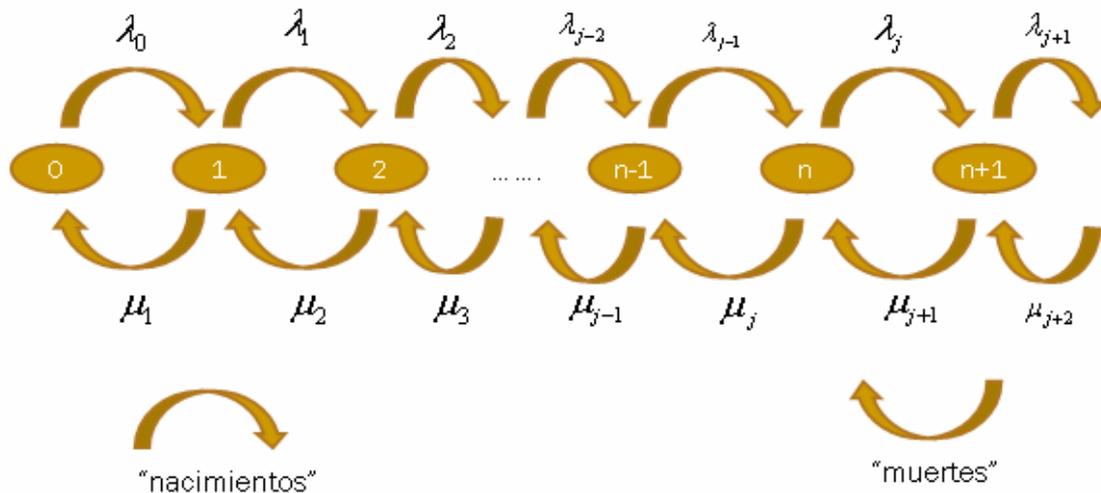


Figura 3.10 Diagramas de transiciones para el proceso de nacimiento y muerte

El estado del sistema en el tiempo t ($t > 0$), denotado por $N(t)$, es el número de clientes que hay en el sistema de colas en el tiempo t . El proceso de nacimiento y muerte describe en términos probabilísticos cómo cambia $N(t)$ al aumentar t . En general, dice que los nacimientos y muertes individuales ocurren aleatoriamente, en donde sus tasas medias de ocurrencia dependen del estado actual del sistema. De manera más precisa, las suposiciones del proceso de nacimiento y muerte son las siguientes:

Suposición 1:

Dado $N(t) = n$, la distribución de probabilidad actual del tiempo que falta para el próximo nacimiento (llegada) es exponencial con parámetro λ_n ($n = 0, 1, 2, \dots$).

Suposición 2:

Dado $N(t) = n$, la distribución de probabilidad actual del tiempo que falta para la próxima muerte (terminación del servicio) es exponencial con parámetro μ_n ($n = 1, 2, \dots$).

Suposición 3:

Las variables aleatorias de los tiempos que faltan para la próxima llegada y para la terminación del servicio son mutuamente independientes.

Transición en el estado del proceso $n \rightarrow n+1$ (un nacimiento) o $n \rightarrow n-1$ (una muerte)

Considere cualquier estado n ($n=0,1,2,\dots$) del sistema suponga que en el tiempo 0 se inicia el conteo del número de veces que el sistema entra a este estado y el número de veces que sale de él, como se ve a continuación:

$E_n(t)$ = Número de veces que el proceso entra al estado n , hasta el tiempo t .

$L_n(t)$ = Número de veces que el proceso sale del estado n , hasta el tiempo t .

Como los dos tipos de eventos (entrar y salir) deben alternarse, estos dos números serán iguales o diferirán en solo 1, es decir,

$$|E_n(t) - L_n(t)| \leq 1$$

Al dividir ambos lados entre t y después hacer que $t \rightarrow \infty$ se obtiene:

$$\left| \frac{E_n(t) - L_n(t)}{t} \right| \leq \frac{1}{t} \quad \text{Así,} \quad \lim_{t \rightarrow \infty} \left| \frac{E_n(t) - L_n(t)}{t} \right| = 0$$

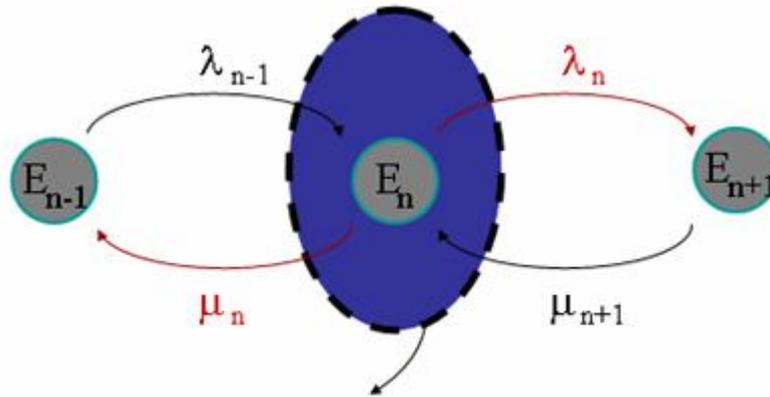
Si se dividen $E_n(t)$ y $L_n(t)$ entre t se obtiene la tasa real (número de eventos por unidad de tiempo) a la que ocurren estos dos tipos de eventos, y cuando $t \rightarrow \infty$ se obtiene la tasa media (número esperado de eventos por unidad de tiempo):

$$\lim_{t \rightarrow \infty} \frac{E_n(t)}{t} = \text{tasa media a la que el proceso entra al estado } n.$$

$$\lim_{t \rightarrow \infty} \frac{L_n(t)}{t} = \text{tasa media a la que el proceso sale del estado } n.$$

Este resultado conduce al principio clave:

Tasa media de entrada = Tasa media de salida, tal como se puede ver en la figura 3.11.



Flujo saliente = Flujo entrante

$$(\lambda_n + \mu_n) \pi_n = \lambda_{n-1} \pi_{n-1} + \mu_{n+1} \pi_{n+1}$$

Figura 3.11 Tasas de entrada y salida en un proceso de nacimiento y muerte

De estas formulas de tasas de entrada y salida se obtienen los diferentes estados que se pueden observar en la tabla 3.1.

Estado	Tasa de entrada = Tasa de salida
0	$\mu_1 P_1 = \lambda_0 P_0$
1	$\mu_0 P_0 + \mu_2 P_2 = (\lambda_1 + \mu_1) P_1$
2	$\mu_1 P_1 + \mu_3 P_3 = (\lambda_2 + \mu_2) P_2$
3	$\mu_2 P_2 + \mu_4 P_4 = (\lambda_3 + \mu_3) P_3$
·	...
n	$\mu_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = (\lambda_n + \mu_n) P_n$
·	...

Tabla 3.1 Ecuaciones de balance

Simplificando estas ecuaciones vistas en la tabla 3.1 obtenemos:

$$C_n = \frac{\lambda_{n-1} \lambda_{n-2} \dots \lambda_0}{\mu_{n-1} \mu_{n-2} \dots \mu_1}, \quad \text{para } n = 1, 2, \dots,$$

Y después se define $C_n = 1$ para $n=0$, así las probabilidades de estado estable son:

$$P_n = C_n P_0, \quad \text{para } n = 0, 1, 2, \dots,$$

El requisito:

$$\sum_{n=0}^{\infty} P_n = 1$$

Implica que:

$$\left(\sum_{n=0}^{\infty} C_n \right) P_0 = 1,$$

asi,

$$P_0 = \frac{1}{\left(\sum_{n=0}^{\infty} C_n \right)}$$

Cuando un modelo de líneas de espera se basa en un proceso de nacimiento y muerte, de manera que el estado del sistema n representa el número de clientes en el sistema de colas, las medidas de desempeño para el sistema se obtienen así:

Numero promedio de clientes en el sistema y en cola:

$$L = \sum_{n=0}^{\infty} n P_n, \quad L_q = \sum_{n=1}^{\infty} (n - s) P_n$$

Tiempo promedio de clientes en el sistema y en cola

$$W = \frac{L}{\lambda}, \quad W_q = \frac{L_q}{\lambda}.$$

En donde λ es la tasa de llegadas promedio. Como λ_n es la tasa media de llegadas cuando el sistema se encuentra en el estado n ($n=0,1,2,\dots$) y P_n es la proporción de tiempo que el sistema esta en este estado:

$$\lambda = \sum_{n=0}^{\infty} \lambda_n P_n$$

3.5.1 SISTEMAS M/M/1

El ejemplo más sencillo de colas es la cola M/M/1. Las distribuciones involucradas son exponenciales. Caracterizamos el estado del sistema con $X(t)$ = número de clientes en el sistema (en cola + servicio). Este es un proceso de Nacimiento y Muerte con $\lambda_n = \lambda$ y $\mu_n = \mu$. Como se puede ver en la grafica

3.12 Usando las ecuaciones de nacimiento y muerte, podemos escribir todas las probabilidades de estado estable como:

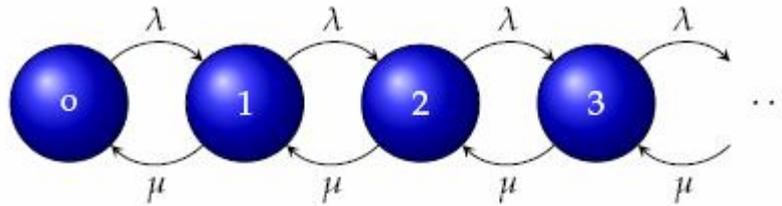


Figura 3.12 Diagrama de transiciones para la cola M/M/1

Parámetros de sistemas M/M/1

Este valor es la fracción promedio de que el sistema este ocupado, también sería el número promedio de unidades que están siendo atendidas en cualquier momento

El factor de Utilización de la cola M/M/1 esta dado por:

$$\rho = \lambda / \mu$$

esto es, por el cociente entre la tasa de llegada y la tasa de servicio. Para que la cola sea estable, se ha de cumplir que dicho factor de utilización ha de ser menor que la unidad:

$$\rho < 1.$$

En otras palabras, la cola M/M/1 es estable si $\lambda < \mu$, esto es, si la tasa de llegada no supera la tasa de servicio.

Además, vimos que se cumple la identidad

$$\rho = \bar{B}$$

que indica que ρ es el numero medio de servidores ocupados. Como solo hay un servidor, en este caso interpretamos que ρ indica la fracción media de tiempo que el servidor está ocupado.

Los factores C_n del proceso de nacimiento y muerte se reducen a:

$$C_n = \left(\frac{\lambda}{\mu} \right)^n = \rho^n \quad \text{Por lo tanto} \quad P_n = \rho^n P_0$$

Donde:

$$P_0 = \frac{1}{\left(\sum_{n=0}^{\infty} \rho^n \right)} = 1 - \rho$$

La primera característica de operación que calculamos es el número promedio de unidades que se encuentran en el sistema, ya sea esperando o siendo atendidas. Denominaremos a este número promedio de unidades promedio, L . Entonces tenemos que:

$$L = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$$

Con estos valores obtenidos podemos calcular el número promedio de unidades que esperan ser atendidas, L_q . Dado que L es el número de unidades que están esperando o están siendo atendidas, y ρ es el número promedio de unidades que están siendo atendidas en algún momento dado entonces:

$$L = L_q + \rho$$

A partir de esto es fácil observar que:

$$L_q = L - \rho$$

O también podríamos decir que:

$$L_q = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

Ahora examinaremos el tiempo de espera. Utilizaremos W para representar el tiempo promedio o esperado que una unidad se encuentra en el sistema. Para encontrar W , observaremos que se L el número esperado de unidades de en el sistema y λ es el número promedio de unidades que llegan para ser atendidas por periodo, entonces el tiempo promedio de cualquier unidad que llega debe estar en el sistema está dado por:

W = tiempo promedio de una unidad en el sistema

$$W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda}$$

De manera similar, el tiempo esperado o promedio que una unidad tiene que esperar antes de ser atendida, W_q , esta dado por:

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)}$$

La probabilidad de una cola vacía P_0 es obtenida de la condición de normalización $P_0 + P_1 + P_2 \dots = 1$, la probabilidad que el servidor este vacío es $1 - \rho$, la probabilidad que el servidor este ocupado es ρ .

3.5.2 SISTEMAS M/M/1/K

Hay una sola cola, cuya disciplina será FIFO, La capacidad del sistema es limitada, de tal modo que sólo puede haber k clientes como máximo en el sistema, Por lo tanto, el número máximo de clientes en la cola es $k-1$, Si un cliente llega y el sistema está lleno, es rechazado y nunca más regresa.

Las llegadas se producen según un proceso de Poisson de razón λ , Los tiempos entre llegadas se distribuirán exponencialmente, $\text{Exp}(\lambda)$.

Los tiempos entre servicios también se distribuirán exponencialmente, $\text{Exp}(\mu)$, de tal manera que μ es el número medio de clientes que el servidor es capaz de atender por unidad de tiempo.

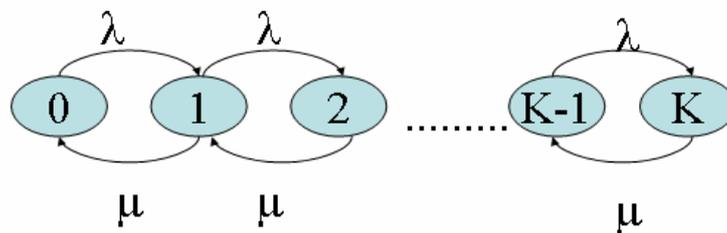


Figura 3.13 Diagrama de transiciones para la cola M/M/1/K

Las ecuaciones de equilibrio son:

$$\begin{aligned} \lambda \pi_0 &= \mu \pi_1 \\ (\lambda + \mu) \pi_n &= \lambda \pi_{n-1} + \mu \pi_{n+1}, \quad 1 \leq n \leq K - 1 \\ \mu \pi_K &= \lambda \pi_{K-1} \\ \sum_{n=0}^K \pi_n &= 1 \\ \pi_n &\geq 0, \quad \forall n = 0, 1, \dots, K \end{aligned}$$

Sumando cada ecuación con la anterior se pasa al sistema $\lambda \pi_{n-1} = \mu \pi_n$, para $n = 1, 2, \dots, K$, que unido a la condición de que la suma de las probabilidades tiene que ser 1, obtenemos:

- Si $\lambda = \mu$,

$$\pi_n = \frac{1}{K+1}, \quad \forall n = 0, 1, \dots, K$$

- Si $\lambda \neq \mu$,

$$\pi_0 = \frac{1}{\sum_{n=0}^K \left(\frac{\lambda}{\mu}\right)^n} = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{K+1}}$$

y

$$\pi_n = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{K+1}} \left(\frac{\lambda}{\mu}\right)^n, \quad n = 0, \dots, K$$

Puesto que el tamaño de la cola es acotado, no puede crecer indefinidamente, por lo que el sistema alcanza el equilibrio para todos los valores de λ y μ y no hace falta introducir la condición $\lambda/\mu < 1$.

Si $\lambda < \mu$, cuando $K \rightarrow \infty$, se tiene la convergencia deseada al sistema M/M/1.

Podemos calcular también el número medio de individuos en el sistema:

- Si $\lambda = \mu$,

$$L = \sum_{n=0}^K n\pi_n = \frac{1 + 2 + \dots + K}{K+1} = \frac{K}{2}$$

- Si $\lambda \neq \mu$,

$$\begin{aligned} L = \sum_{n=0}^K n\pi_n &= \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{K+1}} \sum_{n=0}^K n \left(\frac{\lambda}{\mu}\right)^n = \\ &= \frac{1 - a}{1 - a^{K+1}} \sum_{n=0}^K na^n = \\ &= \frac{a(1 + Ka^{K+1} - (K+1)a^K)}{(1 - a^{K+1})(1 - a)} \end{aligned}$$

ya que

$$\begin{aligned}\sum_{n=0}^K na^n &= a \sum_{n=1}^K na^{n-1} = a \sum_{n=1}^K \frac{da^n}{da} = a \frac{d}{da} \sum_{n=0}^K a^n = \\ &= a \frac{d}{da} \left(\frac{1 - a^{K+1}}{1 - a} \right) = a \frac{-(1-a)(K+1)a^K + (1-a^{K+1})}{(1-a)^2}\end{aligned}$$

Para calcular W hay que tener cierto cuidado con lo que se entiende por clientes en el sistema. Todo el tráfico que llega al sistema no entra ya que si hay K clientes en el sistema no se admite a los clientes que llegan. Así, consideramos que los clientes son aquellos que entran en el sistema y lo hacen con tasa:

$$\lambda_a = \lambda(1 - \pi_K)$$

$$W = \frac{L}{\lambda_a} = \frac{L}{\lambda(1 - \pi_K)}$$

Además, podemos calcular L_s , para ello sea N_s la v.a. número de clientes en el servidor. Entonces:

$$L_s = E(N_s) = E(N_s/N=0)P(N=0) + E(N_s/N>0)P(N>0) = 0 \times \pi_0 + 1 \times (1 - \pi_0) = 1 - \pi_0$$

Así,

$$L_q = L - L_s = L - (1 - \pi_0)$$

y

$$W_q = \frac{L_q}{\lambda(1 - \pi_K)}$$

El uso verdadero del servidor es:

$$\rho = \lambda_a W_s = \lambda(1 - \pi_K) W_s = \frac{\lambda}{\mu} (1 - \pi_K)$$

También pueden derivarse las distribuciones de w y w_q , aunque de forma algo más compleja.

3.6 CONCEPTOS DE TELETRÁFICO

A continuación describiremos algunos conceptos básicos de tráfico telefónico:

Tasa de llamadas (y): es también llamada tasa de arribo, es el número promedio de llamadas iniciadas por unidad de tiempo, cada arribo es independiente de otras llamadas. Se define por la letra y.

Tiempo de espera (h): es el periodo promedio de retención de los circuitos o de los órganos de conmutación por parte de las llamadas que logran tomarlos. Se define por la letra h.

Intensidad de tráfico (A): es el número promedio de ocupaciones simultáneas de un grupo de circuitos.

La intensidad de tráfico en Erlangs se calcula como:

$$A = yh$$

Erlang (E): es una unidad adimensional utilizada en telefonía como una medida estadística del volumen de tráfico. Un Erlang representa teóricamente un circuito ocupado durante una hora.

Tráfico ofrecido (To): es igual a la intensidad de tráfico (A), toma en cuenta todos los intentos de llamada, el cual es igual al tráfico cursado (Tc) más el tráfico perdido (TL).

$$T_o = yh$$

Tráfico cursado (Tc): son las llamadas que han sido aceptadas.

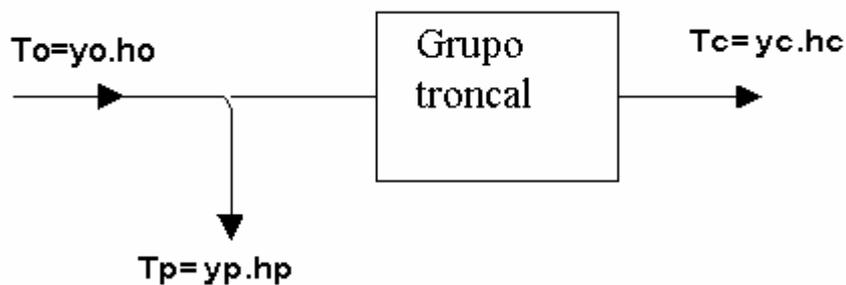


Figura 3.14 Tráficos ofrecidos, cursados y perdidos

Tráfico perdido (TL): son las llamadas que no han sido aceptadas.

Por lo tanto el tráfico ofrecido es:

$$T_o = A = T_c + T_L$$

Además el tráfico perdido es igual a:

$$T_L = T_O \times P_B$$

Grado de servicio: es la relación entre el tráfico perdido y el tráfico ofrecido define que tan eficiente es el sistema.

$$Gos = T_L / T_O$$

Congestión en el tiempo: es la probabilidad de que todos los servidores estén ocupados, esto es, el porcentaje de tiempo en que los servidores permanecerán totalmente ocupados, también se le llama probabilidad de bloqueo P_B .

Congestión en llamadas: es la probabilidad de que una llamada se pierda, también se le llama probabilidad de pérdidas P_P .

3.6.1 Modelo Erlang B

El modelo de cola M/M/c/c (ver figura 3.14) se puede usar para diseñar un sistema que gestiona las peticiones de conexión de una línea troncal por parte de muchos usuarios.

Se suponen peticiones a la línea principal con tiempos entre llegadas exponenciales con velocidad λ peticiones/segundo. Cada línea principal es vista como un servidor y el tiempo de conexión como el tiempo de servicio X . Entonces cada línea principal o servidor tienen una velocidad de servicio $\mu = 1/E[X]$. Supondremos que las peticiones de conexión se bloquean si todas las líneas principales están ocupadas.

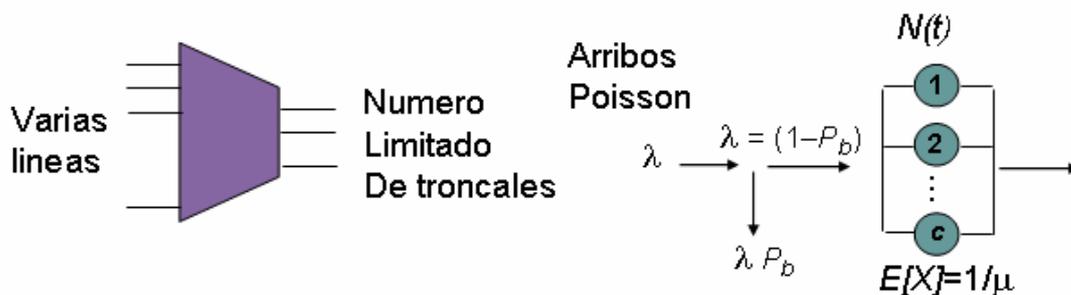


Figura 3.15 Esquema modelo Erlang B (Imagen tomada del libro Redes de comunicación, García y Widjaja, Segunda edición).

Las características de este modelo son:

- Población infinita.
- Régimen de llegadas de Poisson.

- Tiempo de servicio distribuido exponencialmente.
- C servidores.
- No hay sala de espera. Y por lo tanto los clientes se pierden si todos los servicio están ocupados.

Las tasas de nacimiento y muerte vienen dadas por:

$$\lambda_k = \lambda \quad k= 0,1,\dots,C-1$$

$$\mu_k = k\mu \quad k=1,2,\dots,C$$

Diagrama de estado estable:

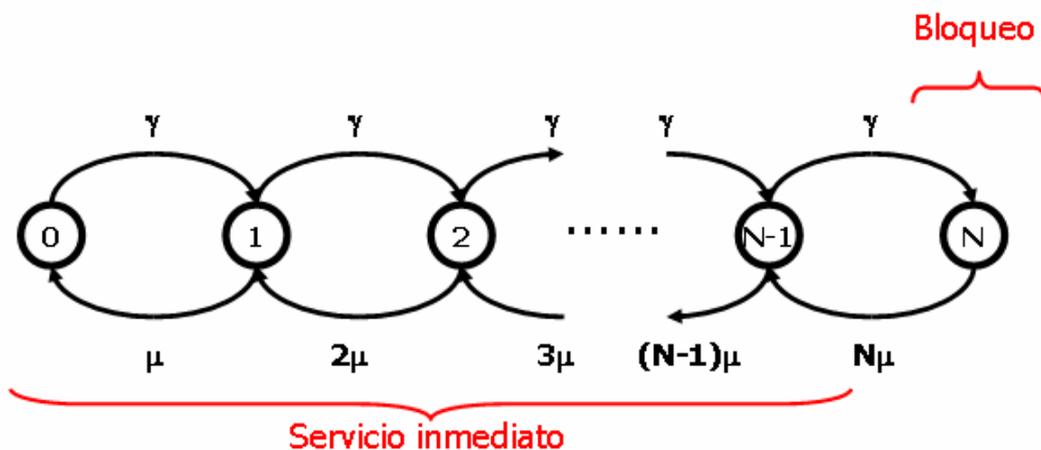


Figura 3.16 Diagrama de transición para Erlang B

En la figura 3.16 se puede apreciar los diagramas de transiciones para Erlang B podemos apreciar que los primeros clientes son atendidos pero los que arriban después, al estar los servidores llenos quedan bloqueados ya que en este modelo no hay sala de espera por lo tanto estos clientes se pierden.

Si la población es grande y su actividad relativamente baja, la tasa de llegada λ_k no se ve apreciablemente afectada por el numero de llamadas en curso, es decir, por el estado del sistema ($\lambda_k = \lambda = \text{constante}$). Por tanto en estos modelos podemos tener resultados muy precisos utilizando este modelo.

Las ecuaciones de balance nos llevan a:

$$\begin{array}{rcl}
\gamma P_0 = \mu P_1 & \longrightarrow & \gamma P_0 = \mu P_1 \longrightarrow P_1 = \frac{\gamma}{\mu} P_0 \\
\mu P_1 + \gamma P_1 = 2\mu P_2 + \gamma P_0 & \longrightarrow & \gamma P_1 = 2\mu P_2 \longrightarrow P_2 = \frac{\gamma}{2\mu} P_1 = \left(\frac{\gamma}{\mu}\right)^2 \frac{P_0}{2} \\
2\mu P_2 + \gamma P_2 = 3\mu P_3 + \gamma P_1 & \longrightarrow & \gamma P_2 = 3\mu P_3 \longrightarrow P_3 = \frac{\gamma}{3\mu} P_2 = \left(\frac{\gamma}{\mu}\right)^3 \frac{P_0}{6} \\
3\mu P_3 + \gamma P_3 = 4\mu P_4 + \gamma P_2 & & \vdots \\
\vdots & & \gamma P_{k-1} = k\mu P_k \\
\vdots & & \vdots \\
(N-1)\mu P_{N-1} + \gamma P_{N-1} = N\mu P_N + \gamma P_{N-2} & & \vdots \\
\vdots & & \vdots \\
N\mu P_N = \gamma P_{N-1} & \longrightarrow & \gamma P_{N-1} = N\mu P_N \longrightarrow P_k = \left(\frac{\gamma}{\mu}\right)^k \frac{P_0}{k!}
\end{array}$$

Regla de probabilidad total:

$$\sum_{i=0}^N P_i = 1 = \sum_{i=0}^N \left(\frac{\gamma}{\mu}\right)^i \frac{P_0}{i!} \longrightarrow P_0 = \frac{1}{\sum_{i=0}^N \left(\frac{\gamma}{\mu}\right)^i \frac{1}{i!}}$$

Con esta ecuación y utilizando:

$$P_k = \left(\frac{\lambda}{\mu}\right)^k \frac{P_0}{k!} \quad \text{y} \quad A = \frac{\lambda}{\mu}$$

Obtenemos:

$$P_k = \frac{\frac{A^k}{k!}}{\sum_{i=0}^N \frac{A^i}{i!}}$$

Para bloqueo en el estado k=N:

$$P(B) = B(N, A) = P_N = \frac{A^N}{N!} / \sum_{i=0}^N \frac{A^i}{i!}$$

" B " = Erlang B
" N " = # Servidores
" A " = Trafico ofrecido

BIBLIOGRAFIA

- GARCIA, Alberto León y WIDJAJA, Indra., *Redes de Comunicación*, Segunda edición.
- STALLING, Williams., *Comunicación y Redes de Computadoras*, Séptima edición.
- KEAGY, Scott., *Integración de Redes de Voz y Datos*, Ed. Cisco Systems.
- FUJITSU, *Descripción del Sistema de Señalización SS7*.
- VAN BOSSE, John., *Signaling in Telecommunication Networks*, Edit. Wiley, Segunda edición.
- MARTINEZ, Jorge., *Conmutadores de Paquetes: Arquitectura y Prestaciones*, Editorial U.P.V.
- HILER and LIBERMAN, *Teoría de Colas*.
- RIAÑO, Germán., *Modelos Probabilísticos con Aplicaciones*, Universidad de los Andes.

Paginas WEB

- <http://www.textoscientificos.com/redes/conmutacion/circuitos>
- <http://www.dednet.net/institucion/itba/cursos/000183/demo/unidad01/conmutaciondecircuitos.htm>
- <http://www.mailxmail.com/curso-redes-transmicion-datos-1>
- <http://www.monografias.com/trabajos16/telefoniasenalizacion/telefoniasenalizacion.shtml>
- <http://www.scribd.com/doc/7353167/10-Sistema-de-SeNalizaciOn-Ss7>
- http://www.nautestech.com/RandD/knowledge/ss7_2.htm