UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES

A R A G O N

LA CALIDAD DE SERVICIO, PARA UNA MEJOR ADMINISTRACION DEL ANCHO DE BANDA

T E S I S

QUE PARA OBTENER EL TITULO DE

INGENIERO EN COMPUTACION

PRESENTA:

RAMON TRUJILLO MACEDO

ASESOR: ING. FERNANDO MARTINEZ ITURBE

MEXICO 2006.

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mis Padres :

Hoy, a un paso de presentar mi examen profesional para conseguir el título de INGENIERO EN COMPUTACIÓN, quisiera decirles cuanto los amo, cuanto los extraño, quisiera agradecerles por todos los esfuerzos que ustedes hicieron a lo largo de mi formación, abrazarlos y fuertemente darles un beso, quisiera decirles que todos esos sacrificios no fueron en vano y que ahora me toca compensarlos, que después de mi examen cual fuera el resultado saber que ustedes están allí conmigo.

Desgraciadamente ya no están presentes, pero sé que en cada momento han estado a mi lado, en esencia, por ese gran recuerdo que conservo de ustedes, por todo el apoyo que en vida siempre mostraron, quiero hacer de estas líneas un homenaje y de esta forma simplemente decirles,

Con todo mi Amor, Admiración y Respeto...

MUCHAS GRACIAS!!!

† **Cayetana Macedo Cristóbal**

1936 – 1991

† **Nabor Trujillo Peña**

1936 – 2000

A mi Familia :

A cada uno de ustedes quiero también agradecer por todo el apoyo que me han brindado, aprovecho hoy para hacerles presentes en esta ocasión tan especial para mí, objetivo que ustedes no hubiese logrado.

Oguilbie, Mary, Jesús, Ale, Edna, Lalo, Mine, Lalito, Benito, Aydee, Eles, Edith, Richie, Brenda, Dany, Mago, Carlos, Carlitos, Victor, Erick, Lila, Mario, Cintya, Cata; Ustedes forman parte de este triunfo, de esta meta y quiero que sepan que este logro también es suyo porque en algún momento de mi vida vi en ustedes un apoyo, y una razón por la cual seguir y llegar hasta el final.

Por todo ese apoyo incondicional...

Gracias!!

ÍNDICE

INTRODUCCIÓN

La capacidad y complejidad de los sistemas de cómputo ha aumentado notablemente en los últimos años. Sistemas operativos multiproceso, redes de computadoras, mecanismos para optimización de recursos (por ej. ancho de banda dinámico), etc., han generado entornos de ejecución altamente dinámicos y poco predecibles. Aunque para muchas aplicaciones esta situación no tiene relevancia, hay algunas para las cuales puede ser un grave problema. Por ejemplo, para aplicaciones en tiempo real (*Voz sobre IP, Videoconferencia, Transacciones en línea, etc.*) y aplicaciones multimedia, para los cuales las propiedades dinámicas de los sistemas que los soportan son fundamentales.

En este tipo de aplicaciones la predicción del comportamiento del entorno de ejecución es deseable y a veces obligatoria, sin embargo lograr un dominio total de éste, es prácticamente imposible, generando un problema difícil de resolver. Aunque los aspectos funcionales de las aplicaciones se pueden asegurar por medio de técnicas y/o mecanismos de control, son más bien los aspectos no funcionales los más afectados.

Para enfrentar este problema se plantea un modelo de objeto activo adaptable, el cual permite reducir los efectos negativos de los cambios impredecibles del entorno o contexto de ejecución. Equipados con mecanismos de decisión, los objetos adaptan su comportamiento en forma dinámica buscando mejorar la calidad de servicio.

En este proyecto se estudiarán algunas de las diferentes técnicas y estrategias utilizadas para la obtención de calidad de servicio, realizando una comparación entre éstas, seleccionando la opción más conveniente para cada caso.

Por lo anterior, el objetivo principal del proyecto es, realizar un estudio de los diferentes mecanismos que se pueden implementar, para buscar un comportamiento estable del entorno donde se trabajará con las aplicaciones de carácter crítico, enfocándome primordialmente en la tecnología de Telefonía IP, logrando así, emplear los principios básicos que conllevan a la calidad de servicio para esta aplicación.

En el primer capítulo se presenta una introducción a la calidad de servicio (*QoS, Quality of Service*), obteniendo así una visión más amplia sobre el tema, además de conocer algunos parámetros y algoritmos implicados en su manejo, permitiendo a su vez, comprender las ventajas de su implementación.

En el capítulo 2, se presentan los mecanismos y herramientas más comunes para la obtención de QoS. Dentro de éstas, se encuentran las técnicas de control, la prioridad del tráfico y mecanismos de señalización, para generar una funcionalidad reflejada en un mejor uso del ancho de banda, dado que éste en frecuentes ocasiones no se utiliza eficientemente ya que se desconoce el uso de sus capacidades y herramientas aplicables para su mejor explotación.

El capítulo 3, describe la importancia de gestionar los recursos de la red para la obtención de la Calidad de Servicio; esto también es conocido como gestión de políticas, sobre todo para fines prácticos en los que ésta es indispensable, dadas las características del tipo de tráfico del que se trate.

En el capítulo 4, se muestran los protocolos y arquitecturas directamente implicadas en la obtención de QoS, es aquí donde términos como RSVP, Diffserv, MPLS serán aclarados.

En el capítulo 5, de una manera genérica se presenta la forma de obtener la Calidad de Servicio en tecnologías de transporte a nivel WAN como lo son ATM y Frame Relay

En el capítulo 6 se aborda un ejemplo, donde se maneja Calidad de Servicio, como lo es Telefonía IP (IP Telephony), explicando algunos aspectos técnicos para su implementación, así como las mejoras obtenidas después de haber realizado la migración a esta nueva tecnología.

Finalmente la bibliografía, apéndices y anexos, consisten en la documentación necesaria para la aclaración de algunas referencias citadas a lo largo del proyecto, y para permitir un mejor entendimiento sobre las ventajas de QoS en la práctica.

1. ANTECEDENTES TÉCNICOS DE QoS

La Internet actual tiene sus raíces en el ARPANET, una red de datos experimental consolidada por la Agencia de Investigaciones Avanzadas de la Defensa de los Estados Unidos (DARPA por sus siglas en inglés), en la década de los 60s. Una meta importante fue construir una red robusta que permitiera sobrevivir a los ataques del ejército activo como los bombardeos. Para lograr esto, el ARPANET se construyó en el modelo de datagramas dónde cada paquete individual se remite independientemente a su destino. La red de datagramas tiene la fuerza de simplicidad y la habilidad para adaptar automáticamente a los cambios en la topología de la red.

Durante muchos años, Internet se usó principalmente para investigaciones científicas, que por medio de una red de computadoras se lograba el intercambio de información. Acceso remoto, transferencia de archivos, y correo electrónico estaban entre las aplicaciones más populares, y para éstas, el modelo de datagramas trabajó eficientemente. El Web, sin embargo, ha cambiado fundamentalmente al Internet, siendo ahora la red pública más grande del mundo, nuevas aplicaciones tales como la videoconferencia, buscadores, medios electrónicos, mesas de discusión y telefonía por Internet están desarrollándose a una velocidad inaudita, así como el comercio electrónico (E-commerce) está revolucionando el camino de los negocios. Ingresando al siglo veintiuno, la Internet se destina para volverse la infraestructura global de comunicaciones (administrativas, comerciales, productivas y hasta políticas).

El éxito fenomenal de Internet ha creado nuevos desafíos. Muchas nuevas aplicaciones tienen los requisitos muy diferentes de aquéllos para los que la Internet fue diseñada originalmente. El modelo de datagramas en que Internet está estructurada, tiene poca capacidad de recursos de dirección dentro de la red y por consiguiente, no puede proporcionar ningún recurso garantizado a los usuarios.

Cuando se intenta ingresar a un sitio Web o hacer una llamada telefónica por Internet, algunas partes de la red pueden estar tan ocupadas que los paquetes no pueden llegar a su destino. Más aplicaciones de tiempo real como la videoconferencia, también exigen algún nivel mínimo de recursos para operar eficazmente. Cuando la Internet se vuelve indispensable en nuestra vida y trabajo, la falta de actuación predecible, es ciertamente un problema que necesita dirigirse, y tratar de encontrar alguna solución práctica a fin de incrementar la eficiencia en la transmisión de algún tráfico en especial.

Otro problema es la diferencia de servicios, debido a que en Internet se trata a todos los paquetes de la misma manera, ofreciendo un sólo nivel de servicio (Best Effort, mejor esfuerzo). Las aplicaciones, sin embargo, tienen diversos requisitos. Las aplicaciones interactivas como la telefonía de Internet son sensibles a la latencia y pérdidas del paquete. Cuando la latencia o la proporción de pérdida exceden ciertos niveles, estas aplicaciones se vuelven inutilizables. En contraste, una transferencia de archivos puede tolerar una cantidad justa de retraso y pérdida de datos, sin mucha degradación percibida en el funcionamiento. Los requisitos del cliente también varían, dependiendo para qué se utilice Internet. Por ejemplo, organizaciones que usan Internet para transacciones bancarias o para el mando de equipo industrial están probablemente dispuestas a realizar una mayor inversión para recibir una mayor prioridad en su tráfico. Muchos proveedores de servicios, proporcionan múltiples niveles de los mismos para reunir diferentes requisitos y demandas del cliente, siendo vital para el éxito de su negocio. *La capacidad para proporcionar convicción del recurso y diferenciación de servicio en una red es a menudo llamada calidad de servicio* (QoS).

Internet se volverá una red multiservicio sólo cuando pueda soportar la diferenciación de servicios. Llevar a cabo las capacidades de QoS en Internet ha sido uno de los desafíos más difíciles en su evolución, tocando casi todos los aspectos de tecnologías de Internet y los cambios requeridos a la arquitectura

básica de la misma, con su trascendencia e implicaciones en un proceso computacional específico.

Por más de una década la comunidad de Internet ha hecho esfuerzos continuos para dirigir el problema y desarrollar nuevas tecnologías para reforzarla con las capacidades de QoS, y solucionar el problema de utilizar adecuadamente el ancho de banda.

1.1 DEFINICIÓN

Para establecer una correcta definición del término Calidad de Servicio (QoS), se debe acudir en primera instancia a estudiar la asignada por el diccionario de la lengua española, el cual, la define como:

CALIDAD == Conjunto de propiedades o atributos que configuran la naturaleza de una persona o cosa.
== Lo que tiene más valor o está en lo más alto.
== Valor intrínseco de una cosa y el valor relativo resultante de compararlas con otras de su misma categoría.

SERVICIO == Acción y efecto de servir
== Prestación o favor que se hace a alguien.
== Estar hecho para algo concreto.

Ambas definiciones llevan contenidas de forma inherente la propiedad de comparación, que se refleja en el valor y su finalidad de servicio, por lo tanto, para determinar si un servicio ofrece mayor o menor calidad, será necesario establecer una comparación con el resto de servicios de ese nivel, finalidades y objetivos de uso.

Son varios los acrónimos terminados en “oS” que hacen referencia a la obtención de calidad de servicio en redes, llevando en ocasiones a situaciones equivocas por el mal uso de los mismos, si bien, QoS es el único que refiere completamente a la Calidad de Servicio, englobando todas la técnicas que se encuentran en torno a ella, mientras que CoS (clase de servicio) y ToS (tipo de servicio), son dos de las técnicas utilizadas para su obtención.

1.1.1 CALIDAD DE SERVICIO (QoS)

Calidad de Servicio (QoS) se refiere a la capacidad de una red (bien una aplicación, un servidor, un router, un switch, etc.), para proveer mejor servicio a tráfico de red seleccionado sobre varias tecnologías, incluyendo Frame Relay, ATM, Ethernet, SONET y redes IP que puedan utilizar cualquiera de estas tecnologías subyacentes.

La meta principal de QoS es proporcionar prioridad incluyendo ancho de banda dedicado, jitter y latencia controlados (requeridas por tráfico interactivo en tiempo real), además de mejorar la pérdida de características. También es importante asegurar proveer prioridad para uno o más flujos, sin hacer que otros fallen.

Adoptarla, requiere además la cooperación de todas las capas de la red, así como de cada elemento de la misma. Desde este punto de vista, la calidad de servicio también suele ser definida como un conjunto de tecnologías que permiten a los administradores de red manejar los efectos de la congestión del tráfico usando óptimamente los diferentes recursos de la red, en lugar de ir aumentando continuamente la capacidad.

La QoS tiene, básicamente, cuatro variantes estrechamente relacionadas:

- La que el usuario desea,
- La que el proveedor ofrece,
- La que el proveedor consigue realmente
- La que finalmente, percibe el usuario.

Habiendo sido definida, la Calidad de Servicio recoge varios parámetros o atributos que describen un servicio, tales como:

- Reserva de ancho de banda
- Retardo extremo a extremo
- Jitter
- Tasa de error

Un ejemplo de tecnología existente que utiliza QoS es RSVP, estudiada mas adelante en el capitulo de protocolos y arquitecturas de calidad de servicio.

1.1.2 CLASE DE SERVICIO (CoS)

Este término implica dos procedimientos: en primer lugar la priorización de los distintos tipos de tráfico claramente definidos a través de la red, en segundo lugar, la definición de un pequeño número de clases de servicio a las cuales aplicarla.

Priorizar es importante en los puntos de congestión de la red, donde las decisiones de priorización pueden ser realizadas por switches y routers.

Las aplicaciones que requieren distinguir clases de servicio incluyen procesos transaccionales, como el vídeo y cualquier otro tráfico sensible al tiempo.

No se debe confundir CoS con QoS, pues, a diferencia de QoS, CoS no garantiza ancho de banda o latencia, en cambio permite a los administradores de red solicitar prioridad para el tráfico basándose en la importancia de éste.

Independientemente de la diferencia, tanto CoS como QoS categorizan el tráfico para asegurar que el considerado como crítico siempre fluya por la red, a pesar del ancho de banda demandado o de las aplicaciones de menor importancia.

Existen muchas posibles definiciones de tipos de calidad de servicio, pero la mayoría de las empresas definen las clases de tráfico por tipo de aplicación, tipo de dispositivo o por tipo de usuario.

1.1.3 TIPO DE SERVICIO (TOS)

En el ToS, se reserva ancho de banda con anticipación y después se asigna el tráfico que necesite preferencia, como el de voz o un CoS con prioridad, de modo que este tráfico pueda utilizar el ancho de banda reservado. ToS no implica, por lo tanto, ningún tipo de garantías.

ToS está incluido como uno de los campos en la tecnología de QoS denominada Diffserv (servicios diferenciados), dónde también es conocido como DiffServ codepoint (DSCP o punto de código Diffserv). Es un campo de 8 bits, estando los dos últimos reservados. Con los otros 6 bits restantes es posible obtener 64 combinaciones o 'codepoint', de ellas, 48 son utilizadas para direccionar el espacio global y 16 son para uso local.

VERS		Tipo de Servicio	Longitud	
IDENTIFICACION			FLAGS	OFFSET del fragmento
Tiempo de vida		PROTOCOLO	CHECKSUM DE LA CABECERA	
Dirección IP fuente				
Dirección IP destino				
Opciones IP				PADDING
DATOS				

Figura 1.1 Paquete Ipv4

1.2 PARÁMETROS

Son muchos los términos manejados en el estudio de la calidad de servicio, que a su vez, son aplicables no sólo a esta área, sino a otros ámbitos de las telecomunicaciones y de la informática, por lo que en este apartado se explicarán aquellos considerados claves para el completo entendimiento de este tema.

Tráfico de red Son aquellos datos que atraviesan la red, generados por diferentes aplicaciones, de esta manera se podría establecer una diferencia del tráfico.

Según el tipo de aplicación Se Obtendrán: tráfico habitual, multimedia, tiempo real, broadcast, multicast, etc.

Según la sensibilidad al retardo En este caso se tendrá:

- Tráfico algo sensible al retardo. Este tipo de aplicaciones requieren retardos de un segundo, o incluso menos. Retardos mayores supondrían hacer esperar a los usuarios por la contestación a sus mensajes antes de que puedan continuar

trabajando, disminuyendo así la productividad de los negocios. Ejemplos son los procesos de transacción on-line, la entrada de datos remota, etc.

- Tráfico muy sensible al retardo. El tráfico en tiempo real de este tipo, es la videoconferencia y la multimedia. Estas aplicaciones requieren un retraso de tránsito muy pequeño (comúnmente menos de una décima de segundo en un sentido, incluyendo el procesamiento en las estaciones finales), y un nivel mínimo de variación (jitter).
- Trafico muy sensible a pérdidas. (Datos tradicionales).
- Tráfico nada sensible. (Servicios de noticias)

Para cada uno de estos tipos de tráfico, podríamos establecer un tipo de QoS según la clasificación realizada, y en consecuencia, la asignación de un nivel de prioridad como se muestra en la figura 1.2.

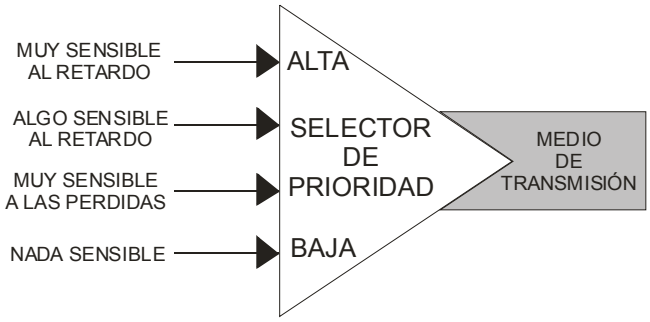


Figura 1.2. Asignación de niveles de prioridad

Retardo Este parámetro, indica la variación temporal y/o retraso en la llegada de los datos a su destino. Es una característica que se hace muy evidente en aplicaciones como la videoconferencia, en donde el retraso existe entre la señal de voz y la señal de video.

Latencia Es el tiempo entre el envío de un mensaje y la recepción del mismo en otra parte de la red. Abarca los retardos sufridos durante el propio camino o en los dispositivos por los que pasa.

Jitter Es la inestabilidad o variabilidad en el retardo. Esto ocurre cuando los paquetes transmitidos en una red no llegan a su destino en debido orden o en la base de tiempo determinada, es decir, varía en latencia. Algo semejante a la distorsión de una señal. Una solución ante el jitter es la utilización de buffers en el receptor, pero esta es una medida poco eficaz, dado que sería necesario un gran tamaño para los buffers, lo que implica un costo económico mayor en los equipos.

Ancho de banda Es la medida de la capacidad de transmisión de datos, expresada generalmente en Kilobits por segundo (kbps) o en Megabits por segundo (Mbps). Indica la capacidad teórica de una conexión, pero esta capacidad se ve disminuida por factores negativos tales como el retardo de transmisión, que puede causar un deterioro en la calidad. Aumentar el ancho de banda significa mayor poder para transmitir más datos, pero también implica un incremento económico, y en ocasiones, resulta imposible su aplicación sin cambiar de tecnología de red.

Pérdida de paquetes Indica el número de paquetes perdidos durante la transmisión. Normalmente se mide en tanto por ciento.

Disponibilidad Indica la utilización de los diferentes recursos. Suele especificarse en tanto por ciento.

Rendimiento Mide el rendimiento de la red en relación a los servicios acordados. El rendimiento es definido también como la velocidad teórica de transmisión de los paquetes por la red. Este depende directamente del ancho de banda y su variación de las posibles situaciones de congestión de la red.

Priorización Priorizar consiste en la asignación de un determinado nivel de QoS al tráfico que circula por una red, asegurando así que las aplicaciones de mayor importancia sean atendidas con anterioridad a las de menor importancia, estando o no ante una situación de congestión. Es necesaria únicamente cuando la red no proporciona la suficiente capacidad para atender todo el tráfico presente en ella.

Encolado El encolado consiste en dividir y organizar el tráfico ante un determinado dispositivo de red para su posterior retransmisión por la misma, según determinado algoritmo que define a la cola y que permite que determinados paquetes sean retransmitidos antes que otros. Es una de las herramientas más utilizadas por la QoS. El sistema de colas no garantiza que los datos importantes lleguen a su destino a tiempo cuando se produce congestión, lo único que aseguran es que los paquetes de alta prioridad llegarán antes que los de baja prioridad.

Planificación Es el proceso de decidir qué paquetes enviar primero en un sistema de múltiples colas.

Flujo Es el conjunto de datos pertenecientes a una misma secuencia que debido a su gran tamaño, han de ser enviados mediante distintos paquetes. Tienen la misma dirección IP fuente y destino, el mismo puerto destino y el mismo protocolo. El flujo necesita llegar secuencialmente a su destino con una frecuencia constante. Por lo que el parámetro más importante para caracterizar un flujo, será su frecuencia constante de bit (constant bit rate, CBR), que proporcionará la frecuencia a la que deberá ser transmitido cada bit de datos.

Acuerdos de niveles de servicio Un Service Level Agreement (SLA, por sus siglas en inglés), es un contrato de servicios entre un proveedor de servicios y su cliente, el cual define las responsabilidades del proveedor en términos del nivel de funcionamiento de la red (rendimiento, tasa de pérdidas,

retrasos, variaciones) y la disponibilidad temporal. Las consecuencias cuando los niveles de servicio no se consiguen o si los niveles de tráfico definidos son superados por el cliente, el SLA puede incluir reglas de condicionamiento del tráfico.

1.3 CLASIFICACIÓN

Es posible realizar una clasificación de QoS bajo distintas especificaciones, así sería factible diferenciarla, según el tipo de tráfico, donde aplicarla, la reserva de recursos de la red y otros parámetros, como se muestra a continuación.

Según la sensibilidad del tráfico.

Teniendo en cuenta la variedad de tráfico existente y los requerimientos de retardo, latencia y ancho de banda para cada tipo, se conceptualizan los siguientes tipos:

QoS muy sensible al retardo. En este caso es necesario garantizar la disponibilidad de una determinada y gran cantidad de ancho de banda para este tráfico y un valor de retardo mínimo que asegure la correcta transmisión del mismo. Para conseguirlo será necesario utilizar mecanismos de prioridad, definidos posteriormente en el capítulo de protocolos y arquitecturas, así como encolar adecuadamente los flujos de datos.

QoS algo sensible al retardo. Al igual que en el caso anterior se garantiza hasta un cierto nivel de ancho de banda, aunque de menor valor. De la misma manera, será necesario asignar propiedades para la transmisión de los datos.

QoS muy sensible a pérdidas. Si se garantiza un nivel de pérdidas de valor cero entonces nunca se descartaran paquetes ni se desbordaran los buffers

de almacenamiento del flujo, lo que facilitará el control de transmisión, por otra parte, esta garantía se hace a nivel de acceso al medio (MAC) o en capas superiores, pero nunca a nivel físico.

QoS nada sensible. La filosofía de este tipo de QoS es usar cualquier oportunidad de transmisión restante y asumir que la capacidad de los buffers posteriores, es suficiente para llevarla a cabo, asignándole a este tipo de tráfico la prioridad mas baja. A este tipo responden los algoritmos Best Effort (mejor esfuerzo), utilizado en Internet.

En la figura 1.3, es posible diferenciar en forma gráfica los tipos de tráfico y sus exigencias de ancho de banda y de sensibilidad a la latencia.

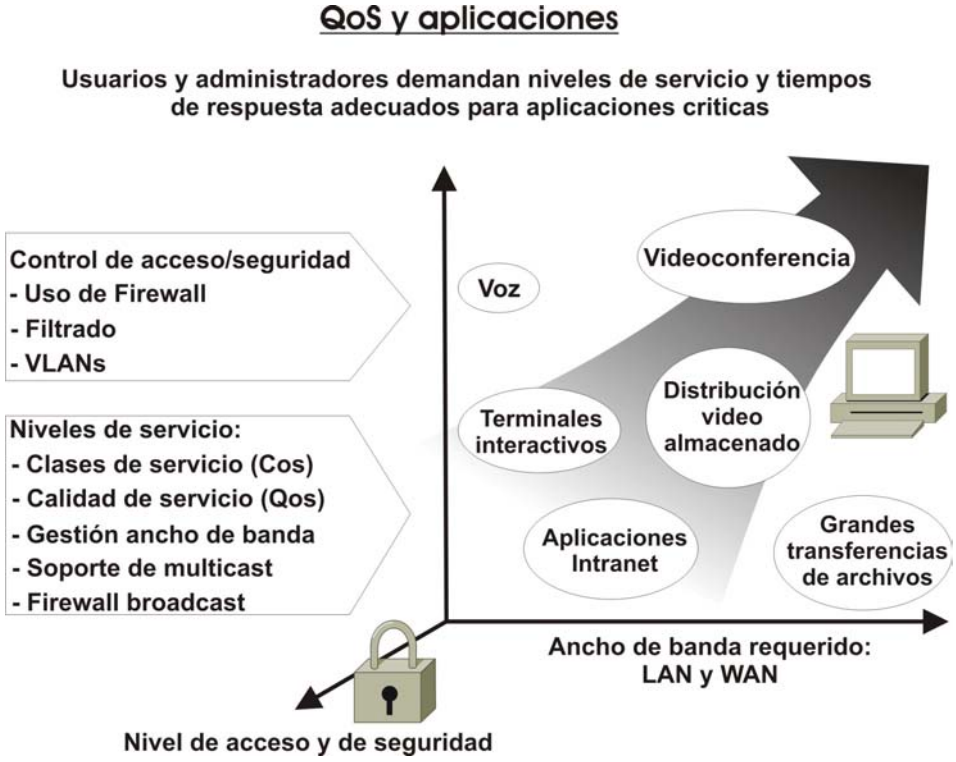


Figura 1.3. QoS y aplicaciones

Según quien solicite el nivel de calidad de servicio.

Teniendo en cuenta que la petición de QoS puede ser realizada por el usuario final o por los conmutadores de la red, me encuentro con:

QoS implícita. En este caso el router o conmutador asigna automáticamente los niveles de calidad en función del criterio especificado por el administrador, dependiendo del tipo de aplicación, protocolo o dirección de origen.

QoS explícita. Este tipo de QoS permite al usuario o aplicación solicitar directamente un determinado nivel de servicio que han de respetar los conmutadores y routers.

Según las garantías.

En esta clasificación se va a tener en cuenta la reserva de recursos del sistema para proporcionar los servicios.

QoS garantizada (Hard QoS). También conocida como Hard QoS la calidad de servicio garantizada es aquella en la que se produce una reserva absoluta de los recursos de la red para un tráfico determinado, asegurándose así, niveles máximos de garantía para este tráfico.

QoS no garantizada (Lack of QoS). Es una calidad de servicio sin garantías, el tráfico es transmitido por la red a expensas de lo que en ella pueda sucederle. Este tipo de QoS pertenece a los servicios Best Effort (mejor esfuerzo).

QoS servicios diferenciados (Soft QoS). También conocida como Soft QoS es el punto medio entre los dos tipos anteriores. Para esta garantía se realiza una diferenciación de tráfico, siendo tratados algunos mejor que el resto

(expedición más rápida, mas ancho de banda promedio, menos tasa de error promedio). Es el utilizado como se explicara más adelante, por DiffServ.

Según el lugar de aplicación.

Es posible aplicar calidad de servicio en los extremos y en los bordes de la red por lo tanto, se obtiene, lo siguiente:

QoS extremo a extremo (End-to-End). Es la aplicación de las políticas de calidad de servicio entre los extremos de la red, pero está menos extendida que la QoS en los bordes de la red (Edge-to-Edge). También se le conoce como la QoS absoluta.

QoS borde a borde (Edge-to-Edge). Es la aplicación de las políticas de calidad de servicio entre dos puntos de la red, a este tipo también se le conoce como calidad de servicio relativa.

1.4 ALGORITMOS

Una vez mencionadas las principales características del término Calidad de Servicio, es necesario exponer el tipo de algoritmos utilizados en la transmisión de paquetes para comprobar cómo estos realizan un control de la congestión y a qué nivel son capaces de proporcionar calidad.

Teniendo en cuenta la clase de servicio que son capaces de ofrecer los algoritmos de transmisión de paquetes, se pueden identificar tres divisiones principales:

Algoritmos de mejor esfuerzo (Best Effort) En este tipo de algoritmos se encuentran los tradicionales, que no ofrecen ningún tipo de garantías de

transmisión, por lo que podría decirse que el nivel de QoS ofrecido es nulo. Un ejemplo es el FIFO (First In First Out).

El principal problema de este tipo de algoritmos es que, si se tienen varios flujos de datos, una ráfaga de paquetes en uno de ellos va a afectar a todos los demás flujos, retardando su transmisión, es decir, que el tiempo de llegada de los paquetes de un flujo puede verse afectado por otros flujos. Cuando esto sucede se dice que el algoritmo utilizado no es capaz de aislar flujos.

Algoritmos deterministas Son aquellos en los que para evitar la posible congestión, antes de aceptar la transmisión de un flujo, se asegura que podrá transmitirse sin problemas, incluso en las peores condiciones. Esto se hace reservando ancho de banda. El reservado es el equivalente a lo que supondría un pico de una transmisión en ráfaga de ese flujo, con lo que se asegura que el mismo nunca se va a salir de su ancho de banda reservado. Si se tiene este comportamiento en cada uno de los flujos de la red, se puede ver que la congestión es imposible, puesto que en el caso de que todos los flujos presentaran un pico al mismo tiempo, tendrían reservado el suficiente ancho de banda para evitar congestiones.

En caso de que por límites físicos de la red, no pudiera asegurarse ese ancho de banda, el algoritmo rechazaría la transmisión del flujo.

Estos tipos de algoritmos fueron los primeros en aparecer cuando surgió la necesidad de asegurar las velocidades de transmisión. Es obvio que consiguen su objetivo, pero lo hacen a un precio muy elevado, puesto que son muy ineficientes respecto al uso de la red. Como ya se ha explicado antes, las situaciones de ráfaga en un flujo son poco frecuentes y de muy corta duración, con lo que en la mayoría de los casos las necesidades de ancho de banda del flujo son mucho menores. Al reservar el equivalente al peor caso, la mayor parte del tiempo se está reservando una capacidad de transmisión que no se usa, y si esto se ejecuta con

varios flujos, el resultado es que los algoritmos rechazan flujos por no poder darles la reserva adecuada cuando en realidad la red presenta un uso muy por debajo de sus posibilidades.

Algoritmos intermedios Son los algoritmos cuyo objetivo es ofrecer calidad de servicio y al mismo tiempo hacer un uso eficiente de los recursos. Entre estos, se pueden diferenciar entre los que ofrecen servicios estadísticos, de degradación limitada y predictivos. Estos algoritmos no aseguran una QoS tan estricta como los deterministas, pero en la mayoría de los casos consiguen un buen comportamiento y aprovechan mucho más los recursos disponibles. Como consecuencia, en estos algoritmos sí que es posible el retraso ocasional de algún paquete, con lo que, si el algoritmo en cuestión se da cuenta de que un paquete ha superado su tiempo de expiración puede descartarlo directamente.

Servicios estadísticos Este tipo de servicios trabaja estadísticamente, asegurando una QoS con una probabilidad determinada. Para ello, antes de aceptar la transmisión de un flujo, obtienen los parámetros que lo modelan. Una vez obtenidos éstos, se calcula el porcentaje de QoS que se le puede asignar, y si es mayor o igual al porcentaje requerido, se acepta el flujo.

Para entender sus ventajas se puede suponer por ejemplo una red con un ancho de banda de 10 Mbps y flujos que requieren 1Mbps de frecuencia constante y 2 Mbps en ráfagas. En un algoritmo determinista es posible transmitir como máximo, cinco flujos. En cambio en uno estadístico es factible transmitir hasta nueve con una probabilidad bastante alta, puesto que presentarían un comportamiento correcto exceptuando los casos en los que dos o más flujos transmitieran en ráfaga al mismo tiempo.

No obstante cabe mencionar que el tener una probabilidad, no implica que tenga que cumplirse necesariamente. Este es el principal inconveniente de esta técnica, que garantiza una probabilidad y no un resultado. Aún así, gran cantidad

de algoritmos de este tipo han resultado ser bastante exactos y usados con probabilidades de fallos del orden 10^{-5} presentan un comportamiento casi determinista aprovechando mucho más la capacidad de la red.

Servicios de degradación limitada Una característica de los flujos es que puede permitirnos la pérdida de algunos datos. Los algoritmos de degradación limitada aprovechan este hecho en la gestión de los paquetes consiguiendo una capacidad de decisión más alta. Por ejemplo, en una aplicación de comunicación por voz, es posible perder algunos paquetes, teniendo en cuenta que estas pérdidas están limitadas por la aplicación, puesto que una pérdida excesiva provocaría que la voz fuera ininteligible.

Con este método, cuando un flujo entra en la red divide sus paquetes en varios tipos, cada uno con una prioridad distinta y con un retardo máximo diferente. Así, en caso de congestión, los paquetes importantes tendrán mayor prioridad.

Servicios predictivos Se caracterizan por utilizar datos obtenidos midiendo las características de los flujos. En la admisión del flujo es necesario confiar en la información que nos da el servidor del flujo, pero una vez dentro de la red se calculan dinámicamente sus parámetros. Con esto se asegura una información fiable y real, puesto que proviene del compartimiento actual del flujo. Este hecho ayuda a tomar decisiones más precisas sobre las necesidades del flujo y, por tanto, conlleva a un funcionamiento bastante correcto con un uso elevado de los recursos.

Otra característica que tienen es organizar los flujos en grupos con necesidades similares. La mayor ventaja reside en que es posible aplicar políticas distintas en cada grupo. Así, se pueden establecer prioridades entre grupos o limitar el uso de los recursos dependiendo del grupo al que pertenezcan. Añaden con mucha facilidad comunicaciones sin calidad de servicio simplemente

añadiendo un grupo con prioridad mínima y sin reserva de ancho de banda. Esto hace que la utilización de la red sea más alta.

1.5 BENEFICIOS

Los beneficios al agregar calidad de servicio a nuestras comunicaciones son de toda índole:

Ahorro de costos.

A la hora de decidir instalar cualquier sistema, el aspecto económico debe considerarse prioritariamente.

En cuanto al objeto de esta ponencia existe un cálculo relativamente sencillo a realizar, que consiste en estimar la cantidad de ancho de banda adicional que habría de contratar para conseguir las calidades requeridas mediante el simple aumento del canal de comunicación sin un sistema de gestión inteligente.

Sólo con estas "cuentas" se produce el retorno de la inversión en unos pocos meses, dependiendo la cifra exacta del tipo de aplicaciones usadas en cada caso, calidades límite requeridas, costos de las comunicaciones impuesta por los proveedores de ancho de banda (cabe mencionar que los precios de estas tarifas son elevados), etc.

Además, si se considera el efecto del aumento de productividad, asociado al uso correcto de los recursos por parte de los distintos elementos de la empresa, se podrá ver que este retorno disminuye a un tiempo mínimo.

Si no se dispone de elementos gestores del ancho de banda, nuestras comunicaciones serán como un sistema de tráfico vehicular donde no existieran semáforos, señalizaciones, ni obligación de ceder el paso a los servicios de urgencia.

Aunque estos servicios de control tengan un costo mas elevado, ¿cree qué es más rentable ahorrarse el dinero de estos elementos que usarlos?

Incremento de producción.

Este es el aspecto fundamental, conseguir una calidad de servicio hasta ahora imposible de obtener, abre el paso a un uso generalizado de Internet y en general de las redes IP en una parte del mercado empresarial que hasta ahora no podía arriesgarse a encaminar sus más críticos procesos de negocio por unos canales con pocas o ninguna garantía de calidad, esto posibilita una empresa mucho más interconectada en sus distintas facetas (clientes, proveedores, etc.) y por tanto mucho más productiva.

Y por supuesto esta necesidad generará un muy importante negocio tanto para los fabricantes de dispositivos y aplicaciones, como para los proveedores de servicios de conectividad, que ya consideran su primer "campo de batalla" el de los servicios añadidos sobre el mero transporte, y entre estos servicios destaca como fundamental, la diferencia cualitativa garantizada del tráfico. Esta diferenciación es una de las bases de las distintas ofertas en cuanto redes IP.

Las estructuras de redes privadas virtuales (VPN) sobre las muchas redes IP públicas de las que es posible disponer (Internet seguirá siendo la mayor aunque también segmentada en función de la calidad), habrán de verse sustentadas por igual tanto en la seguridad (obviamente), como en la calidad de la comunicación, pues de nada sirve reducir costos en la factura del proveedor de canal de datos al usar una red pública en lugar de costosos enlaces dedicados, si no se permite confiar en que se pueda obtener una calidad "profesional" en esta comunicación.

RESUMEN

En este capítulo se muestra inicialmente lo que es la Calidad de Servicio, de dónde nace la necesidad de implementarla, como lo es el enorme crecimiento del Internet, además de algunos aspectos técnicos, cuyo conocimiento es necesario para la mejor comprensión del tema, como los es el Jitter, el ancho de banda, el retardo, la latencia, etc. Pero QoS (Calidad de Servicio), va más allá de los parámetros empleados para su medición; para la implementación de una buena QoS, es necesario hacer buen manejo del tráfico seleccionado, y para ello es que existen mecanismos que nos permiten hacer diferenciación en el tráfico de red, y de diferentes maneras lograr la prioridad de éste.

El siguiente capítulo abarca los mecanismos y herramientas más comunes para la obtención de QoS, mostrando cómo cada mecanismo hace uso del tráfico seleccionado para darle prioridad sobre los demás paquetes de la red.

2. MECANISMOS Y HERRAMIENTAS PARA LA OBTENCIÓN DE QoS.

Hoy en día existen mecanismos suficientes para posibilitar la implantación con garantías de aplicaciones multimedia o aplicaciones críticas. El problema reside en seleccionar el equipamiento adecuado que permita los mecanismos más óptimos para el tipo de tráfico a manejar. Por otro lado, es necesario que todos estos mecanismos sean administrables y configurables de una manera sencilla y centralizada. De otro modo, sería necesario actuar uno por uno sobre todos los dispositivos de la red para dar de alta una política determinada, con el consiguiente peligro para la continuidad del servicio que ello supone. Por último, habilitar todos estos mecanismos no debe implicar disminución del rendimiento en los equipos, ya que podría ocurrir que éstos se colapsaran en un momento dado.

En este capítulo, se estudiará el funcionamiento de diferentes mecanismos y/o herramientas para obtener la Calidad de Servicio, dentro de los cuales se pueden mencionar, las herramientas para prevención y control de congestión, clasificación y marcado de paquetes, mecanismos de prioridad y gestión, etc.

2.1. CONTROL DE ADMISIÓN (CAC)

Este mecanismo, es un filtro que acepta o rechaza la incorporación a la red de una nueva conexión. Las nuevas conexiones se aceptan, sólo si la red puede mantener su desempeño general sin afectar la QoS de las conexiones existentes. La decisión se hace basándose en el contrato de tráfico de la conexión.

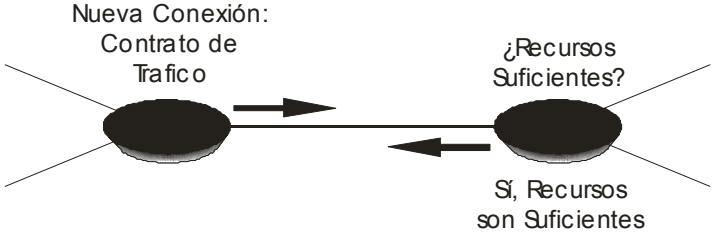


Figura 2.1 Control de Admisión.

2.2. CONFORMADO DEL TRÁFICO GENÉRICO (Generic Traffic Shaping)

El conformado de tráfico es un mecanismo de control de flujo en una interfase determinada, reduciendo la circulación de salida para evitar la congestión, obligando a determinado tráfico a una tasa de bit particular, mientras se encolan las ráfagas de dicho tráfico. Así, el tráfico adherido a una topología puede ser tratado para configurarlo según los requerimientos del tráfico saliente, eliminando cuellos de botella en topologías con tasas de datos desiguales.

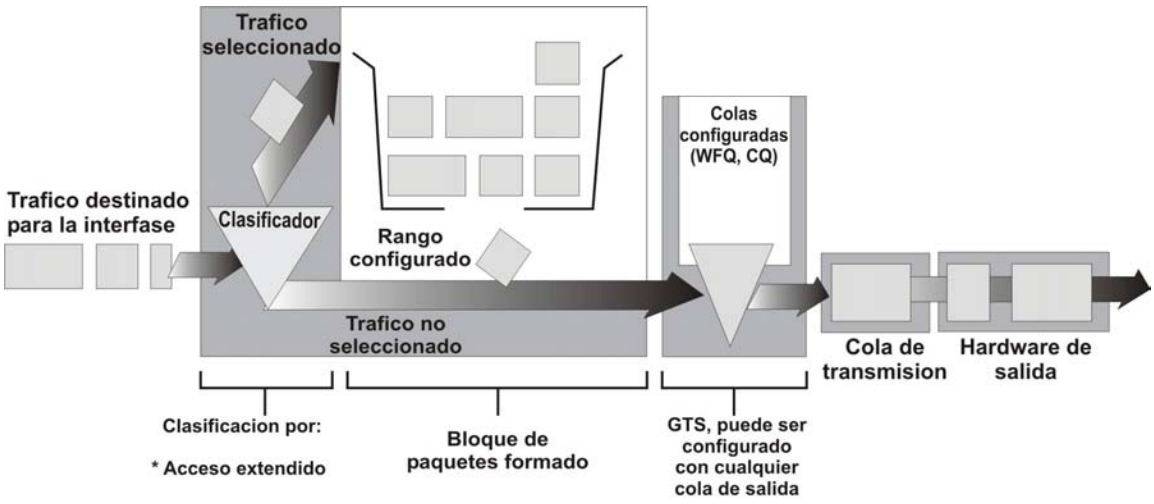


Figura 2.2 Generic Traffic Shaping.

GTS se aplica en una interfase, puede usar listas de acceso para seleccionar el tráfico a formar, y trabaja con una variedad de tecnologías de capa 2, incluyendo Frame Relay, ATM y Ethernet.

2.3. MARCADO Y CLASIFICACIÓN DE PAQUETES

La técnica de clasificación es la encargada de identificar qué aplicaciones han generado qué paquetes. Sin clasificación, la red no puede determinar que hacer con un paquete determinado. Tras su identificación, el paquete se “marca” de modo que otros dispositivos de la red puedan a su vez identificarlo fácilmente.

Todas las aplicaciones dejan huellas sobre los paquetes que pueden ser utilizadas para identificar la aplicación fuente. El proceso de clasificación, examina estas huellas y discierne qué aplicación ha generado el paquete.

Los cuatro métodos de clasificación son:

Protocolo. Las aplicaciones pueden ser identificadas por su EtherType. Por ejemplo Apple Talk utiliza 0x809B e IPX utiliza 0x8137. La priorización basada en este mecanismo representa una buena manera de controlar protocolos que originan retardos de tráfico.

TCP y UDP Socket Number. Muchas aplicaciones utilizan ciertos sockets UDP para comunicarse. Por ejemplo, http utiliza el puerto 80. Examinando el número de socket del paquete IP, la red determina qué tipo de aplicación ha generado el paquete. Está función es conocida como conmutación de nivel 4 debido a que TCP y UDP pertenecen a la capa 4 del modelo OSI.

Source IP Address. Muchas aplicaciones son identificadas por su dirección Source IP (fuente IP). Como a veces algunos servidores están dedicados exclusivamente a soportar una sola aplicación, correo electrónico por ejemplo, el análisis de la dirección IP fuente de un paquete permite identificar la aplicación que lo ha generado.

Physical Port Number. Como las Source IP Address, el Physical Port Number (número de puerto físico) puede indicar que servidor esta enviando los datos.

Una vez identificada la aplicación, el paquete debe ser marcado para asegurar que los equipos de la red sean capaces de darle prioridad, para esto se pueden utilizar alguno de los siguientes métodos.

IEEE 802.1p. Asigna a cada paquete un nivel de prioridad entre 0 y 7. Aunque es el método de priorización más utilizado en el entorno LAN, cuenta con varios inconvenientes, como el requerimiento de una etiqueta adicional de 4 bytes. Esta etiqueta viene definida en el estándar IEEE802.1Q, pero es opcional en redes Ethernet. Además, sólo puede ser soportado en una LAN, ya que las etiquetas 802.1Q se eliminan cuando los paquetes pasan a través de un router.

Differential Services Code Point (DSCP). DSCP es un esquema de marcación de nivel 3 que usa la cabecera IP para almacenar la prioridad del paquete. Las principales ventajas de DSCP sobre IEEE802.1p son que no se precisan etiquetas extras, puesto que el paquete usa la cabecera IP y que la prioridad queda preservada a través de Internet. DSCP utiliza 64 valores para definir distintos niveles de servicio en función del usuario.

IP TOS (Type of Service). Opera introduciendo tres bits en un subcampo de la cabecera IP para marcar la calidad, que podrá ser preservada en la WAN.

2.4. MECANISMOS DE PRIORIDAD Y GESTIÓN

Para satisfacer las necesidades de QoS, los nodos de la red deben aplicar mecanismos de prioridad y gestión.

La prioridad hace referencia normalmente a la capacidad de proporcionar diferentes tratamientos al retardo, por ejemplo, los paquetes de mayor prioridad son servidos siempre antes que los de menor prioridad, en el contexto de dar salida a los paquetes. Los nodos también implementan diferentes técnicas para sufrir menor pérdida con los paquetes de mayor prioridad.

Por otro lado, los nodos también necesitan utilizar algún mecanismo de gestión para asegurarse de que algunas conexiones obtengan los recursos prometidos (en procesamiento y en ancho de banda). Este mecanismo además asegura que cualquier capacidad "**de repuesto**" esté distribuida de la manera más justa. Algunos ejemplos son Generalized Processor Sharing (GPS), Weighted Round Robin (WRR), Weighted Fair Queueing (WFQ) y Class Based Queueing (CBQ), comentados algunos, más adelante.

2.5. PROTOCOLOS DE SEÑALIZACIÓN

Para obtener la QoS requerida por una red, los sistemas extremos necesitan indicárselo a la red, para ello se usan los protocolos de señalización. Un ejemplo de protocolo de señalizaciones RSVP (Resource Reservation Protocol), LDP (Label Distribution Protocol) e IP Precedente. Su escalabilidad y las capacidades de la señalización es un tema que ha estado y estará bajo estudio.

Hay que idealizar que la señalización de QoS es una forma que tiene la red de comunicarse, proporcionando una manera de que cada elemento de la red pueda pedir algo a un vecino. Por ejemplo, una red IP puede usar la parte de la cabecera IP para solicitar un manejo especial para cada prioridad o del tráfico sensible al tiempo. La señalización es útil para coordinar el tráfico que se ocupa de cualquiera de las herramientas de QoS empleadas para conseguir de manera exitosa la QoS extremo a extremo.

La verdadera QoS extremo a extremo requiere que cada elemento en el camino del tráfico por la red (conmutadores, routers, firewalls, host, usuarios, etc), entreguen su parte de QoS y todo ello debe ser coordinado mediante técnicas de señalización.

2.6. EFICIENCIA DEL ENLACE

Existen mecanismos que manejan el encolado y el conformado de tráfico para mejorar la eficiencia y predicción de los niveles de servicio de aplicación, tales como:

LFI: Fragmenting and Interleaving IP Traffic

Fragmentado y separado del trafico IP

El tráfico interactivo como Telnet y/o VoIP (Voz sobre IP), son susceptibles a aumentos de latencia y jitter cuando la red tiene que procesar paquetes grandes (por ejemplo un paquete de LAN a LAN vía FTP atravesando un enlace WAN), sobre todo si necesitan ser encolados en un enlace de menor velocidad. LFI reduce el retardo y el jitter en los enlaces de menor velocidad, partiendo los paquetes grandes y entrelazando los paquetes de menor retardo, obteniendo así paquetes más pequeños.

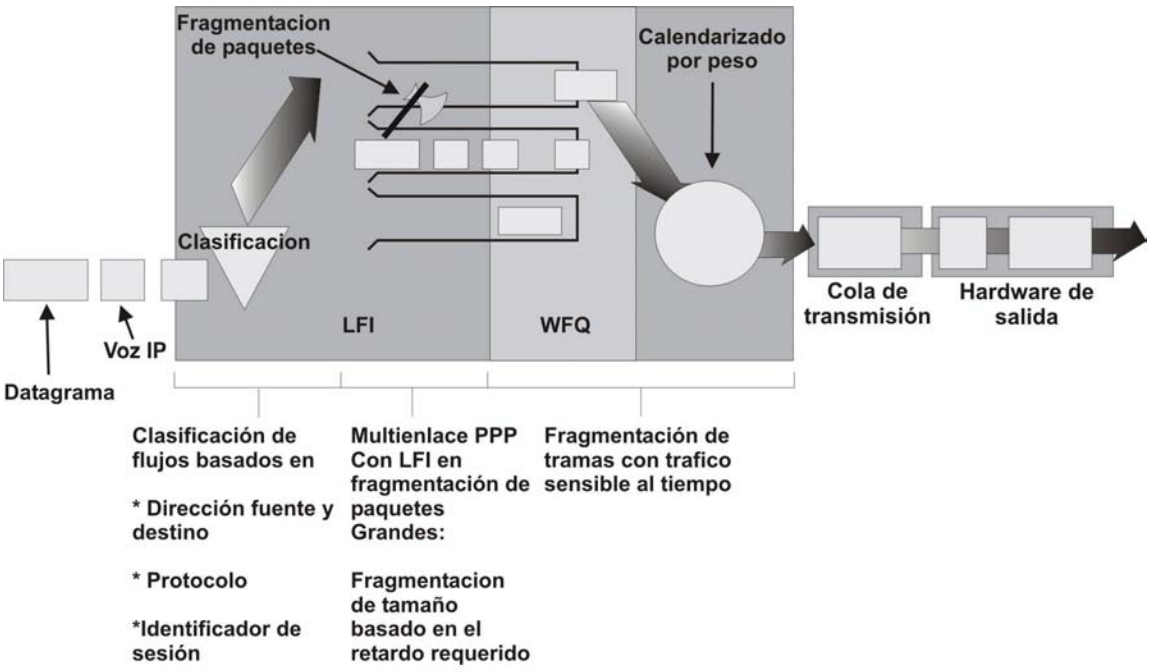


Figura 2.3 Fragmentación y separado del tráfico IP.

Esta herramienta fue diseñada especialmente para enlaces de baja velocidad en que el retardo al serializar es significativo.

RTP (Real-Time Transport Protocol) Header Compression

Incrementando la eficiencia de tráfico en tiempo real

El protocolo de transporte en tiempo real es un protocolo host-to-host usado para transportar el tráfico de nuevas aplicaciones multimedia, incluyendo paquetes de audio y video sobre redes IP, proporcionando funciones de transporte de red extremo a extremo, pensando en aplicaciones con requerimientos en tiempo real, (como es audio, video o simulación de datos sobre servicios de red multicast o unicast). Real-Time Transport Protocol header compression incrementa la eficiencia de muchas nuevas aplicaciones multimedia o de voz sobre IP (VoIP), toman ventajas con dicho protocolo de transporte especialmente en enlaces de baja velocidad.

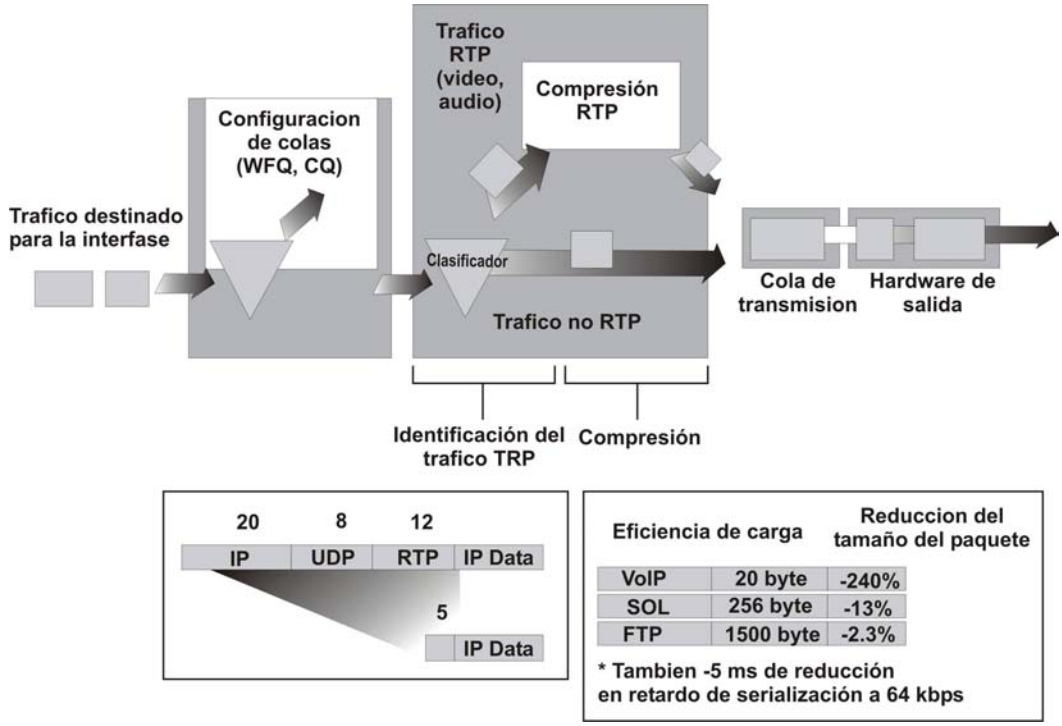


Figura 2.4 Eficiencia del tráfico en tiempo real.

2.7. HERRAMIENTAS DE CONTROL DE CONGESTION

Los elementos de la red deben manejar grandes tasas de tráfico de llegada, para ello usan algoritmos de encolamiento que clasifiquen el tráfico y aplicar después algún método de priorización para su expedición.
Algunos algoritmos de gestión de colas son los siguientes:

- First-in, First-out (FIFO) Primero en entrar, primero en salir de la cola
- Priority Queuing (PQ) Encolado de Prioridad
- Custom Queuing (CQ) Por costumbre
- Weighted fair queuing (WFQ) Por peso

Cada algoritmo de encolamiento ha sido diseñado para resolver un problema especifico del tráfico de la red, obteniendo así un determinado efecto en el funcionamiento de la red, tal y como se describe en cada uno de ellos.

Los algoritmos de encolamiento toman efecto cuando el congestionamiento es experimentado. Por definición, si en enlace no esta congestionado, entonces no es necesario encolar los paquetes. En ausencia de la congestión, todos los paquetes son liberados directamente a la interfase.

FIFO (First-in, First-out) Capacidades básicas de almacenamiento y envío.

En esta forma más simple, el algoritmo FIFO implica almacenar paquetes cuando se congestiona la red y reenviarlos, teniendo en cuenta el orden de llegada, cuando la red no está tan congestionada. FIFO es el algoritmo por defecto, por lo que no requiere ninguna configuración, sin embargo tiene varias limitantes. La más importante es que no toma decisiones sobre la prioridad de los paquetes, es el orden de llegada el que determina el ancho de banda y asigna el buffer. No proporciona protección contra aplicaciones (fuentes) corruptas. El tráfico a ráfagas puede causar grandes retardos en la entrega del tráfico basado

en aplicaciones sensibles al tiempo, así como al control de la red y mensajes de señalización que circulan por la misma. FIFO fue, en definitiva, un primer paso necesario para el control del tráfico de red, pero hoy en día las redes inteligentes necesitan algoritmos más sofisticados.

PQ (Priority Queuing) Priorizando el tráfico.

PQ asegura que el tráfico importante sea administrado más rápidamente en cada punto en donde se utilice. Fue diseñado para dar una mayor prioridad al tráfico importante. Este algoritmo puede dar la prioridad de forma flexible según el protocolo de red utilizado (IP, IPX, o AppleTalk), interfase entrante, tamaño del paquete, la dirección fuente y/o destino, y mucho más. En PQ cada paquete es colocado en una de las cuatro colas de prioridad **- alta – media – normal – baja –** basándose en la prioridad ya asignada, asignando prioridad normal para aquellos paquetes que no tengan ninguna prioridad asignada. Durante la transmisión el algoritmo concede un tratamiento preferencial a las colas de mayor prioridad sobre las de menor prioridad.

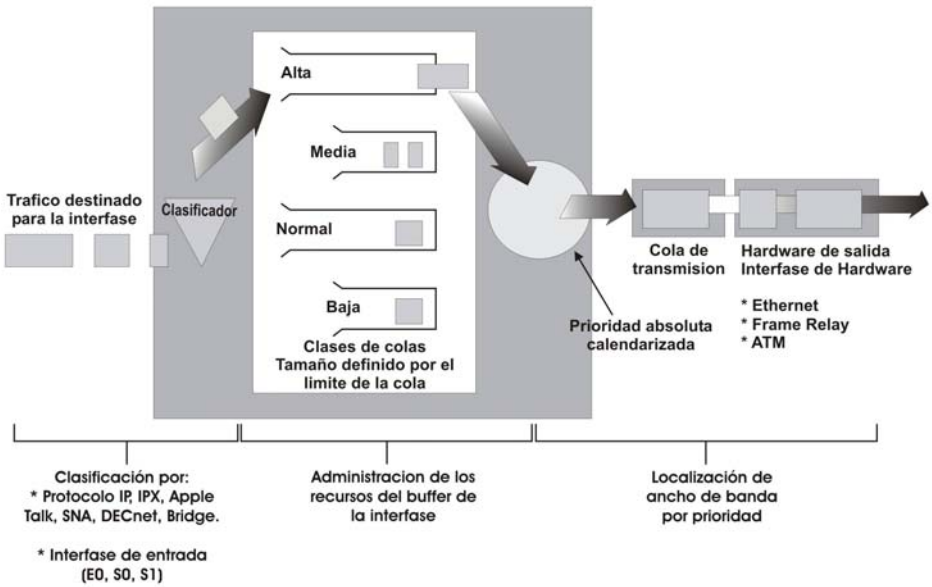


Figura 2.5 Priorización del tráfico.

PQ es útil para asegurarse ese tráfico de misión-crítica que cruza varios enlaces WAN consiguiendo el tratamiento de prioridad. Actualmente, este algoritmo usa una configuración estática, no adaptándose a los requisitos cambiantes de las redes.

CQ (Custom Queuing) Ancho de banda garantizado

CQ, fue diseñado para permitir que varias aplicaciones u organizaciones compartan la red, entre aquellas aplicaciones que necesiten ancho de banda o requisitos de latencia mínimos. En estos entornos debe compartirse proporcionalmente el ancho de banda entre las aplicaciones y los usuarios. CQ puede utilizarse para proporcionar ancho de banda garantizado en aquellos puntos en donde se produzca congestión, asegurando a un tráfico específico una porción fija de ancho de banda disponible y dejando el tráfico restante para cualquier otro tipo de tráfico. CQ maneja el tráfico asignando una cantidad específica del espacio de la cola a cada clase de paquete, aplicando posteriormente el sistema Round-Robin.

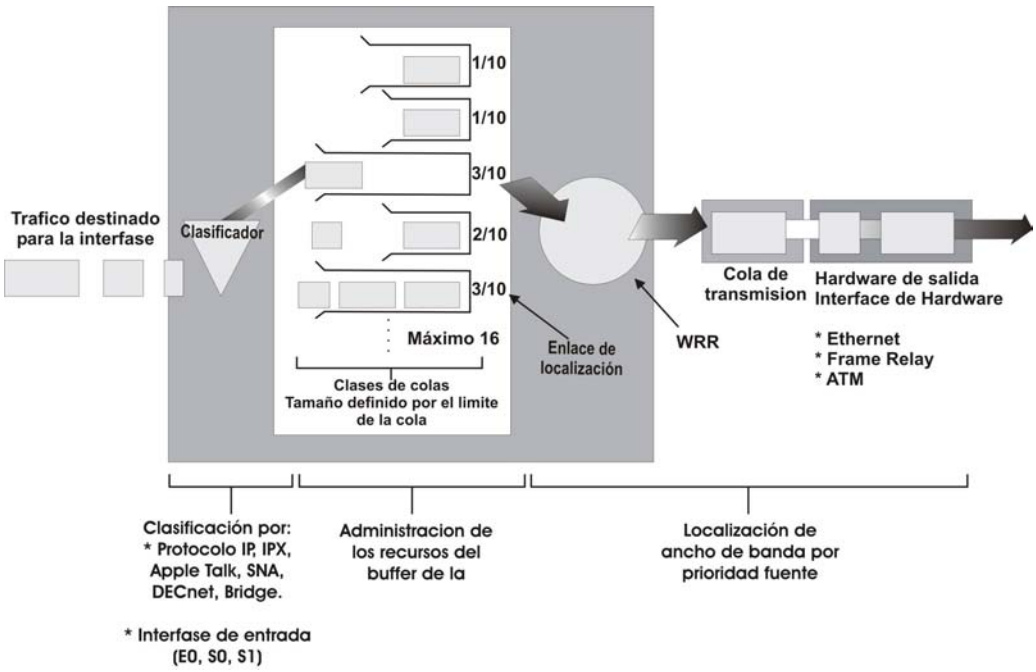


Figura 2.6 Ancho de Banda Garantizado.

El algoritmo de encolamiento coloca los mensajes en una de las 17 colas (la cola 0 es utilizada para almacenar mensajes de sistema como mantenimiento, señalización, etc.) y se vacía con prioridad pesada. Los servicios de enrutamiento van desde la cola 1 a la 16 según el orden asignado por Round-Robin, desencolando un byte de cada cola por cada ciclo. Esto asegura que ninguna aplicación (o grupo de aplicaciones) logre una mayor porción de capacidad global cuando la línea se encuentre bajo presión. Como PQ, CQ se configura estáticamente y no se adapta automáticamente a las condiciones cambiantes de las redes.

WFQ (Weigthted Fair Queuing) Algoritmo para redes inteligentes

Para situaciones en las que es deseable proporcionar un tiempo de respuesta consistente a cualquier tipo de usuarios sin necesidad de aumentar el ancho de banda de forma excesiva, entonces la solución es WFQ. Es un algoritmo de encolamiento basado en el flujo que realiza dos tareas simultáneamente: sitúa el tráfico interactivo a principio de la cola para reducir el tiempo de respuesta y permite así compartir el resto del ancho de banda entre flujos que requieran gran ancho de banda.

WFQ asegura que las colas no se quedarán sin ancho de banda, proporcionando a ese tráfico un servicio predecible. Así mismo, las ráfagas de tráfico de bajo volumen (la mayoría del tráfico) reciben tráfico preferencial, transmitiéndolas rápidamente. Las ráfagas de tráfico de gran volúmen compartirán la capacidad restante de forma proporcional entre ellos, como se observa en la figura 2.7.

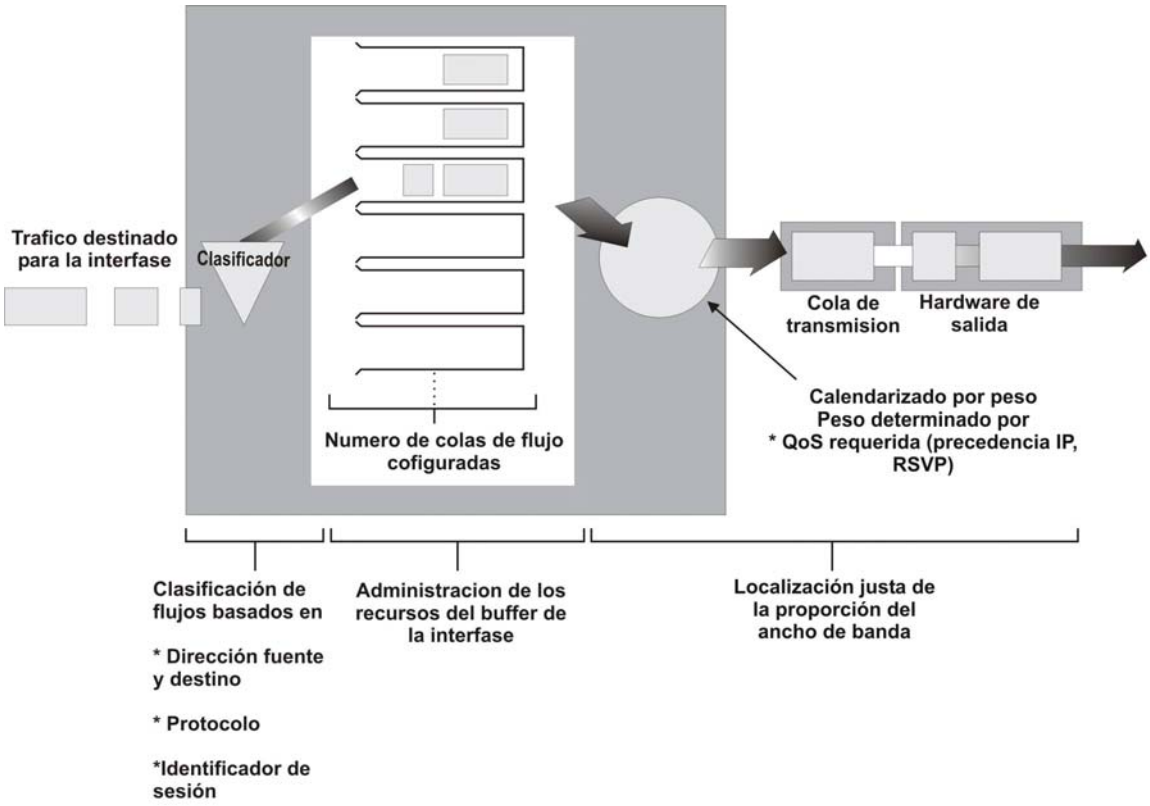


Figura 2.7 Encolamiento por peso del tráfico.

WFQ, ha sido diseñado para minimizar en esfuerzos al configurar, adaptándose automáticamente a las condiciones cambiantes del tráfico de red.

WFQ es eficaz, pues permite que se pueda asignar cualquier cantidad de ancho de banda para flujos de tráfico de baja prioridad si no está presente ningún flujo de alta prioridad. Esto es diferente del multiplexado por división de tiempo (TDM) que simplemente mide el ancho de banda y permite que éste no se utilice si no está presente un determinado tipo de tráfico. WFQ, además, trabaja con las técnicas IP Precedente y RSVP para proporcionar QoS diferenciada así como servicios garantizados.

Este algoritmo también trata el problema de la variabilidad del atraso durante la transmisión. Si hay múltiples conversaciones de alto volumen activas, sus tasas de transferencia, así como sus períodos de llegada se hacen más

predecibles. WFQ refuerza algoritmos como el Control de Enlace Lógico (LLC) y el control de congestión del Protocolo de Control de Transmisión (TCP), obteniendo como resultado una expedición y un tiempo de respuesta más predecible para cada uno de los flujos activos, tal como se muestra en figura 2.8.

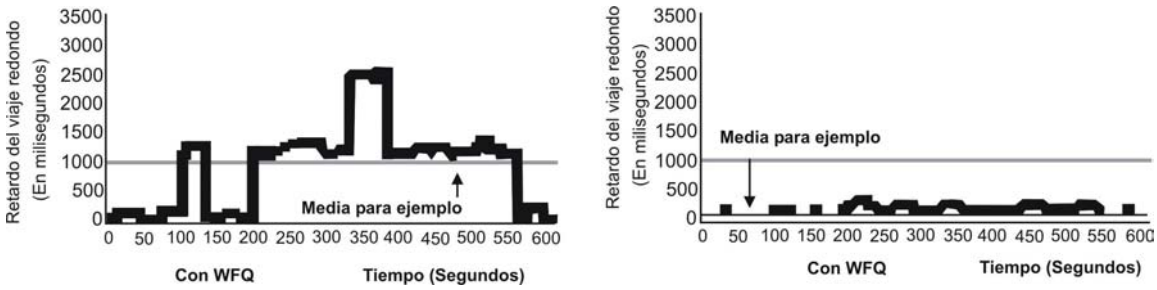


Figura 2.8 Tiempos de respuesta con WFQ y sin ella.

2.8. HERRAMIENTAS DE PREVENCIÓN DE CONGESTION

Porque el tamaño de las colas no es infinito, estas pueden llenarse y desbordarse. Cuando las colas están llenas, cualquier paquete que ya no entre en las colas, será descartado. El problema con estos paquetes es que el router no puede prevenir a dicho paquete para no descartarlo (incluso si es un paquete de alta prioridad). Así que, se necesita que un mecanismo haga dos cosas:

1. Intente asegurarse que la cola no se llene, para que haya espacio para paquetes de alta prioridad.
2. Que contenga algún criterio para descartar primero los paquetes de prioridad mas baja, antes de descartar los paquetes de alta prioridad.

La prevención de la congestión es una forma de administración de las colas. Las técnicas de prevención de congestión monitorean el tráfico de red, en un esfuerzo para anticiparse y evitar cuellos de botella en la red, como opuesto a técnicas que operan para controlar la congestión después de que esta ocurre.

RED (Random Early Detection)

Detección Temprana Aleatoria: Evitar la congestión

Los algoritmos de detección temprana al azar son diseñados para evitar la congestión entre redes, antes de que está se vuelva un problema. RED, supervisa la carga de tráfico en diferentes puntos de la red y descarta paquetes de forma estocástica si aumenta el nivel de congestión. El resultado es que la fuente detecta está situación, retardando su transmisión. RED se ha diseñado para trabajar con TCP en entornos IP.

WRED (Weighted Random Early Detection)

Detección Temprana Aleatoria Pesada:

Cooperación con tecnologías de señalización QoS

WRED combina las capacidades de RED con IP precedente. Esta combinación mantiene tráfico preferencial que maneja como paquetes de prioridad más altos. Puede selectivamente desechar el tráfico de menor prioridad cuando la interfase empieza a congestionarse y proporciona características de gestión distintas para las diferentes clases de servicio. Pero WRED también RSVP, ofreciendo servicios integrados de QoS de carga controlada.

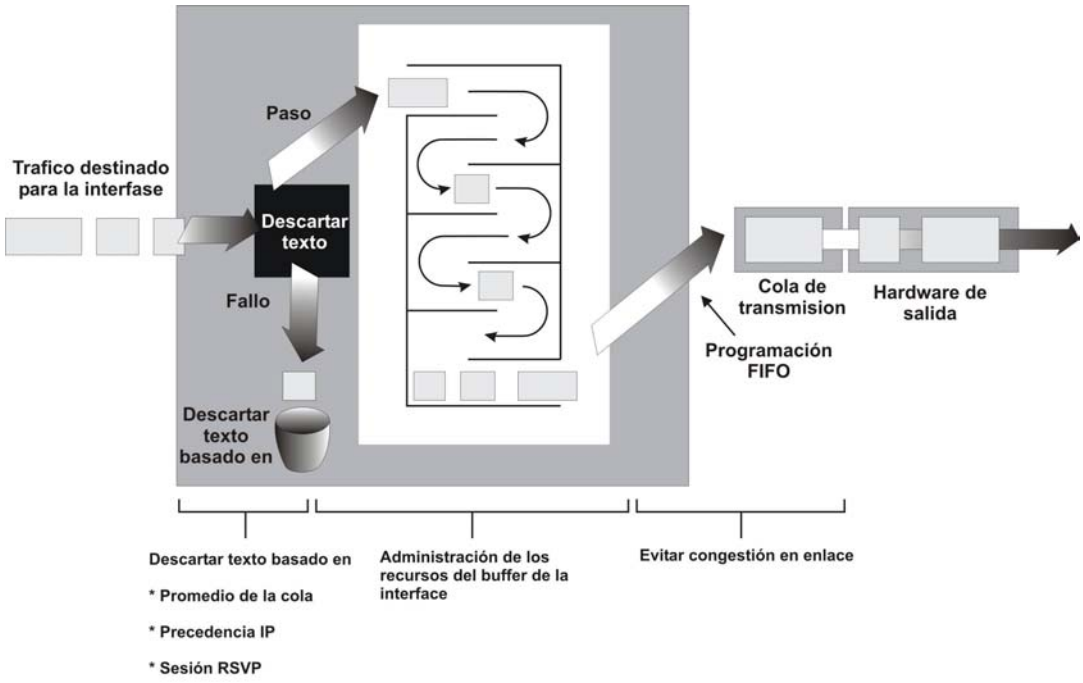


Figura 2.9 Detección temprana aleatoria.

Dentro de cada cola, un número finito de paquetes puede alojarse. Una cola llena causa desbordamientos. Se descartan paquetes que no podrían encajar en la cola porque la misma estaba llena. Esto es indeseable porque el paquete desechado puede haber sido prioritario y el router no tenía una oportunidad para encolarlo. Si la cola no está llena, el router puede mirar la prioridad de todos los paquetes que llegan y descartar los paquetes de baja prioridad, permitiendo los paquetes prioritarios en la cola. A través de manejar la profundidad de la cola (el número de paquetes en la cola) descartando varios paquetes, el router puede asegurarse que la cola no se llene y no se experimente un desbordamiento. Esto permite al router tomar la decisión de descartar paquetes cuando sea necesario. WRED también ayuda a prevenir la congestión global en una red. WRED usa un umbral mínimo para cada nivel de precedencia de IP y determinar cuando un paquete puede descartarse. (El umbral mínimo debe excederse para WRED considere un paquete como un candidato para ser descartado.)

RESUMEN

En el anterior capítulo, se observan las herramientas tanto para controlar la congestión en caso de que esta se presente, como lo es el manejo de colas de diferente prioridad, permitiendo de esta forma el flujo del tráfico de mayor importancia según las necesidades del administrador, así como diferentes técnicas de identificación de tráfico, las cuales pueden estar basadas en el destino, el puerto a conectarse o algún otro parámetro que sea viable para la distinción del flujo en cuestión.

3. GESTIÓN DE POLÍTICAS

Como se ha mostrado en los capítulos anteriores, existen dos formas fundamentales de proporcionar Calidad de Servicio (QoS) en las redes. La más sencilla consiste en incrementar el ancho de banda bruto de las infraestructuras (sobreingeniería de la red). La otra alternativa, sin embargo, es la que se basa en el empleo de funciones como priorización de datos, colas, eliminación de la congestión y modelado de tráfico. Ambas soluciones implican la realización de una buena gestión de los recursos de las redes. Es aquí, donde aparece la QoS basada en políticas, permitiendo al administrador de red asignar anchos de banda más funcionales y priorizar tráficos en la red, en función de un conjunto de políticas administrativas y patrones de uso, optimizando el uso de los recursos disponibles.

En definitiva, la QoS basada en políticas consiste en la identificación de un grupo de tráfico y en el establecimiento de un perfil de calidad de servicio para el mismo. Estos grupos de tráfico pueden estar basados en criterios de topología o de grupos de usuarios, estaciones individuales o sesiones de aplicaciones, es decir una arquitectura sustentada en un aprovechamiento óptimo del ancho de banda.

Los protocolos de Calidad de Servicio (QoS) proveen los mecanismos para diferenciar el tráfico y las políticas definen como será utilizado.

Las políticas definen las reglas que determinarán específicamente cómo, cuándo y dónde será aplicada la Calidad de Servicio para distintos tráficos de red, y funciones de redefinición del modelado del tráfico con un análisis más ágil y preciso que permita opciones para los protocolos de QoS.

Primero, se describirá porqué son necesarias las políticas para QoS, previendo una definición básica de qué son las políticas y describiendo las políticas de QoS desarrolladas en la IETF (Internet Engineering Task Force), y así proveer algunos detalles de cómo serán almacenadas y accesadas.

3.1. LAS NECESIDADES PARA POLÍTICAS

Como ya se ha visto, la Calidad de Servicio se refiere a la clasificación de paquetes con el propósito de tratar ciertas clases o flujos de paquetes de manera particular, comparándolo con otro paquete. Lo que significa que ahora el servicio de entrega de datos que era impredecible “Best Effort”, en este caso será predecible.

Para habilitar QoS se requiere del uso de protocolos como RSVP que provee la señalización requerida para la integración de servicios y DiffServ para diferenciar servicios. RSVP afecta la asignación de recursos de red en base a uno por flujo cuantificando los requerimientos por aplicación, considerando que DiffServ provee las señales dentro de las cabeceras de paquetes IP para permitir ordenar adecuadamente flujos agregados, si éstos se presentan, hecho que ocurre frecuentemente.

Hay circunstancias variables en las que los dueños del tráfico (usuarios finales, aplicaciones, host de internet, entidades de la compañía, etc.), requieren niveles de servicio, allí se necesitan las reglas, que requieren de “policías” para dar fuerza a esas reglas, lo cual sustenta la presencia de “jueces” que decidan dónde serán aplicadas. Las reglas, los policías y los jueces componen un sistema de políticas que es un componente esencial en una red con QoS habilitada.

3.2. QUE ES UNA POLÍTICA

En el sentido más elemental, una política es una o más reglas que describen una o varias acciones que ocurren cuando existe una condición específica, es decir, a toda acción, derivará necesariamente de una característica dada y plenamente establecida.

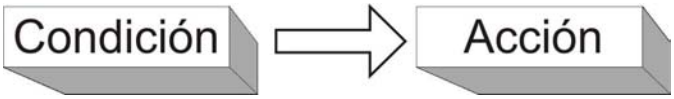


Figura 3.1 Política, toda condición deriva en una acción.

Una regla de la política comprenderá a menudo otras reglas, en efecto “las políticas pueden contener políticas”. Esta noción de jerarquía es crucial, esto es, construir políticas complejas de un juego de políticas más simples, y reducir su administración.

Las redes son fundamentalmente no determinísticas y esto es verdad especialmente para el estándar “Best Effort”, frecuentemente no se dedican recursos a manera de circuito virtual, como lo hacen las redes telefónicas. Los requerimientos para políticas de redes con QoS habilitada, son para evitar la creación de mayores problemas. Las reglas de la política (condiciones y acciones) deben ser inequívocas y comprobables, es decir verificables en términos de aplicabilidad práctica. Debe haber sólo una regla correcta apropiada para cualquier juego específico de condiciones. En otras palabras, las políticas deben definir estados de máquina finitos.

Además, el criterio de decisión de política, debe ser en cierto modo definido, para que pueda entenderse y aplicarse propiamente por los límites administrativos. En otras palabras los procesos para decidir qué condiciones son activas y qué acciones deben resultar, deben estar basado en algunos algoritmos de evaluación de criterio globales que no cambiarán entre los dominios de la

política. Esto es, un requerimiento esencial para interacción de sistemas de políticas, y una aplicación de las mismas, que no manifiesten dudas en ningún momento.

3.3. INFRAESTRUCTURA Y ARQUITECTURA DE POLÍTICAS

El grupo de trabajo RAP (RSVP Admisión Policy) en la IETF, fué originalmente caracterizado para establecer un modelo de control de políticas escalables para la Reservación de protocolos (RSVP) y los servicios integrados que habilita (IntServ). La infraestructura de políticas que ellos diseñaron, fue de inmediato reconocida como inicialmente aplicable a otras tecnologías de QoS, como los servicios diferenciados (DiffServ). De hecho, se ha reconocido subsecuentemente como un modelo generalmente útil para otras tecnologías que necesitan el apoyo de las políticas, como la seguridad en redes (para firewalls, sistemas de acceso, seguridad IP, redes privadas virtuales VPN, etc.)

La infraestructura es en esencia muy simple, identifica dos componentes primarios por su funcionalidad. La infraestructura está comprendida de un punto de entrada en vigor de política (Policy Enforcement Point, PEP) y un punto de decisión de política (Policy Decision Point, PDP), qué es en cierto sentido el “policía” y el “juez”. El PEP es el punto de inicio, donde se forza a la política, y el PDP toma decisiones basadas en políticas permitiendo o no la continuación de la regla, como ejemplo se encuentran entidades tales como servidores de autentificación

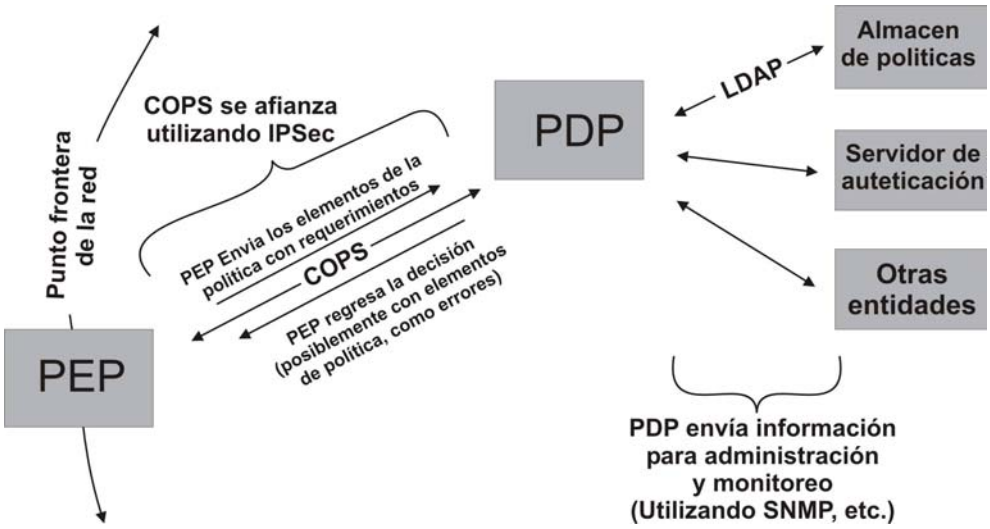


Figura 3.2 La estructura de la política, identifica los elementos funcionales requeridos para soportar las políticas de QoS en la red.

La separación de PEP y PDP, es lógica basada en funcionalidad y no necesariamente en una separación física. En otras palabras, la separación de PEP y PDP no imposibilita la posibilidad de tenerlos construidos en el mismo dispositivo. Actualmente, las especificaciones de infraestructura de políticas describen un componente del PEP llamado “Local PDP” (LPDP) que habilita algunas decisiones de políticas para realizarse en PEP. Sin embargo, para evitar aperturas en la seguridad, un PEP es siempre requerido para enviar una respuesta al PDP para una decisión final de política.

3.4. FUNCIONES DE POLÍTICAS

Además de las funciones de decisión y de la aplicación de la política se ha definido ya el anuncio asignado al PDP y al PEP, otra función importante es la medición. Se muestra a continuación, una descripción de estas tres funciones primarias de sistemas de políticas:

- **Toma de decisiones**: La función del PDP, implica la interpretación de la política, detección de conflictos de políticas, recepción de descripciones de interfaz, elementos de la petición de política y de la decisión de política de PEPs (condiciones), determinando qué política es relevante, aplicando las políticas y regresando los resultados. También implica el enviar elementos de la política al PEP asincrónicamente, de acuerdo con actualizaciones de políticas o peticiones de entidades externas. Los mensajes de los COPS, fueron creados para este propósito, y enviar una instancia codificada de política llamada PIB para ser instalada en el PEP.
- **Aplicación**: Esto implica al PEP, que aplica las acciones según decisiones de PDP y basadas en políticas relevantes y condiciones actuales de la red (el cuál puede ser estático, por ejemplo direcciones o puertos fuente y destino, y/o dinámico, tal como disponibilidad del ancho de banda).
- **Medición**: Esta es una examinación activa o pasiva en curso de la red y de sus dispositivos constitutivos que comprueban el buen funcionamiento de la misma, si las políticas están siendo satisfechas, y si los clientes están tomando la ventaja injusta de los servicios de red. La toma de decisión de PDP utiliza los datos estáticos y/o dinámicos para determinar si un tipo de política está siendo satisfecha y, si no, qué pasos son necesarios para satisfacerla. La medición por PEP es la revisión de la conformidad de la política para verificar que los consumidores de la política pongan en ejecución correctamente las políticas. Esto también se caracteriza como limpiado, aunque ese término se utiliza más específicamente referente al conformado de tráfico activo o descartar paquetes.

3.5. ARQUITECTURA DE LAS POLÍTICAS

La arquitectura de las políticas se estructura sobre su infraestructura y describe como las herramienta de administración de políticas se relacionan, con los otros componentes de la infraestructura de la política.

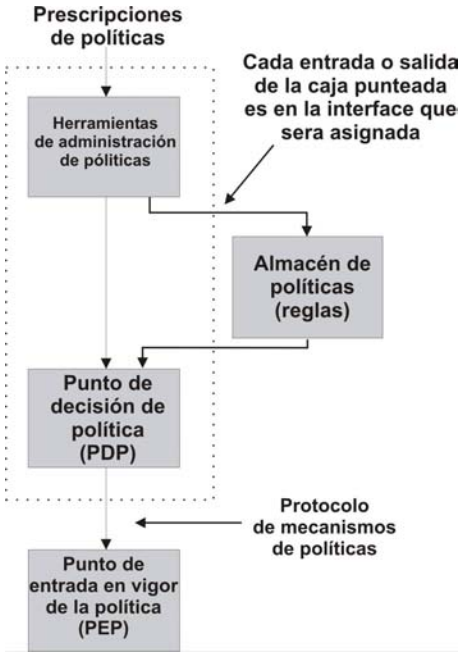


Figura 3.3 La arquitectura de la política muestra otra representación de la estructura de las políticas, que incluyen herramientas de administración y prescripción de políticas.

3.5.1. POLICY WORKING GROUP

El policy working group de la Internet Engineering Task Force ([IETF] es utilizado para definir una estructura escalable y segura para la definición y administración de políticas. La meta inmediata es soportar QoS, aunque en muchas otras tecnologías y servicios que requieren control de acceso también se beneficiarán.

3.5.2. ESQUEMA DEL NUCLEO DE POLÍTICA

Las políticas representan metas comerciales y objetivos. Una traducción debe hacerse entre estas metas, objetivos y su realización en la red. Un ejemplo de esto podrían ser los Servicios de Nivel Acordado (SLA, por sus siglas en ingles Service Level Agreement), y sus objetivos y métrica (Servicio de Niveles Objetivos, o SLO's, por sus siglas en inglés Service Level Objectives), eso se usa para especificar servicios que la red quiere mantener a un cliente dado. El SLA normalmente escribirá la terminología en negocios de alto nivel. Los SLO's direccionan la métrica más específica en apoyo del SLA.

El esquema del núcleo de políticas consiste en un total de 13 clases. Este incluye 6 clases generales y 2 clases específicas del fabricante, las cuales provienen del CIM (Common Information Model, Modelo Común de Información): **policy, policyGroup, policyRule, policyCondition, policyTimePeriodCondition, policyAction, vendorPolicyCondition,** y **vendorPolicyAction.**

El esquema contiene 2 clases auxiliares: **policyGroupContainmentAuxClass** y **policyRuleContainmentAuxClass**, además de 2 clases que son definidas para optimización: **policySubtreesPtrAuxClass** y **policyElement**. Y finalmente, una clase para atar clases auxiliares representando condiciones de políticas y acciones de políticas: **policyInstance**.

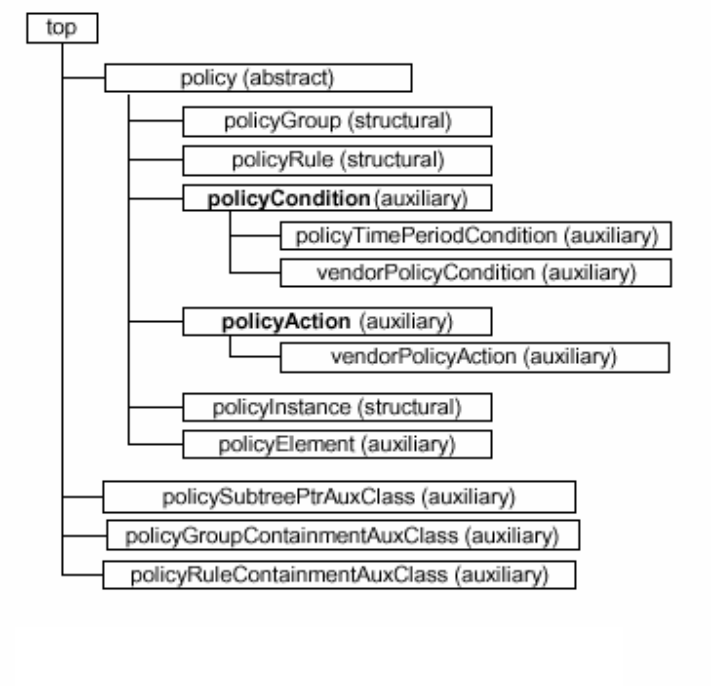


Figura 3.4 Jerarquías para la clase del esquema del núcleo de políticas.

3.5.3. CLASES DE CONDICIONES DE POLÍTICAS

El esquema del núcleo de políticas define las clases generales para la definición de las mismas, pero los usos de la política como QoS tienen diversos requisitos que requieren subclases especializadas derivadas de policyCondition.

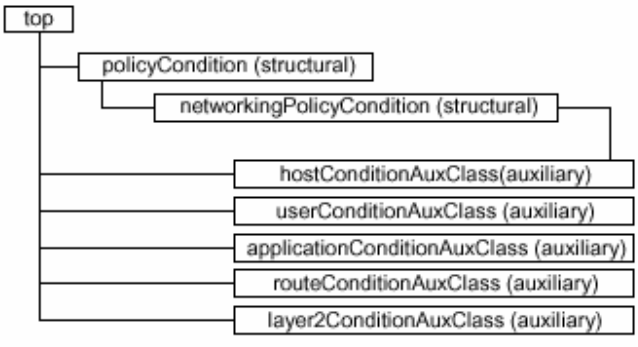


Figura 3.5 Manejo de subclases en una política.

En la figura 3.5, la subclase **networkingPolicyConditions** se observan 5 clases auxiliares diseñadas para categorizar los criterios que un administrador de red comúnmente utiliza para controlar el acceso a los recursos y servicios de red (Condiciones de Política):

Host: Dirección fuente y/o destino ó Host Ids.

User: Tipos de identificador de usuario y valores para enviar y/o recibir.

Aplicación: Número de fuente y/o destino, protocolo de trasnporte y/o recepción del valor del campo IP TOS.

Ruteo: Interfase y dirección de trafico.

Capa 2: Dirección MAC fuente y/o destino, Ethertype, identificador de VLAN, valor de cabecera SNAP, valores SSAP o DSAP.

Estas clases auxiliares se pueden mezclar o emparejar o unir a networkPolicyConditions, según sea requerido por el fabricante o administrador.

Las metas del diseño para estas clases eran de proporcionar la expresión apropiada de la política de una manera que simplificara la administración, pero todavía proporciona extensibilidad y maximiza reutilidad. Con este diseño, "si un fabricante desea ampliar el uso de la categoría, un segundo fabricante desea solamente representar a usuarios en mayor detalle, y un tercero desea hacer ambos, entonces ellos no tuvieran networkPolicyCondition del grado de maneras levemente diversas. Además, el auxiliar clasifica hostPolicyAuxClass, etc., puede ser asociado a otras subclases del policyCondition, DHCPPolicyCondition por caso, de una manera selectiva".

3.6. DEFINICIÓN DE POLÍTICAS

Como se ha descrito con anterioridad, las políticas se comprenden de reglas, y las reglas se componen de condiciones y acciones. Las políticas pueden referirse a otras políticas en una manera jerarquizada, o a la referencia dos o más reglas. Los "grupos de la política" son agregados de las reglas de la política o agregados de los grupos de la política (pero no de ambos). Esta relación del jerárquica se representa en la figura, y describe las clases: **policyGroup, policyGroupContainmentAuxClass** y **policyRuleContainmentAuxClass.**

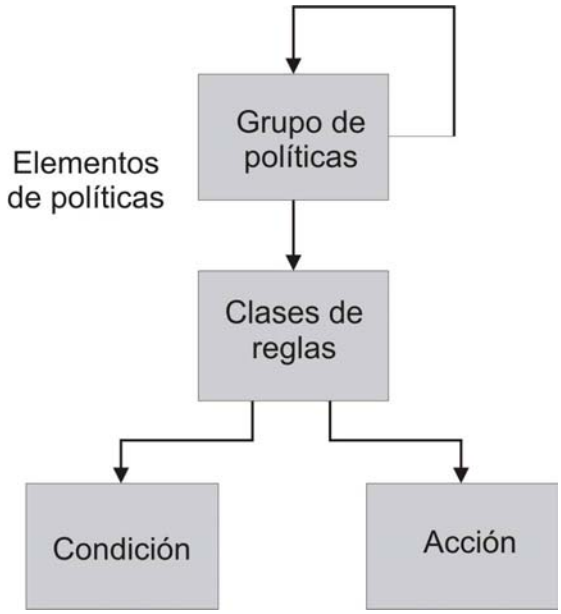


Figura 3.6 El grupo de reglas puede ser formado por la combinación de objetos de políticas (mostrados aquí como “Clases de reglas”), y los grupos de políticas pueden ser anidados.

Las reglas de políticas pueden ser simples o complejas. En una regla de política simple, las condiciones y/o acciones serán agregadas a la misma regla. En una regla de política compleja, las condiciones y/o acciones serán agregadas a las instancias de las clase estructural **policyInstance.** Las políticas simples proveen recuperación eficiente, mientras que las reglas complejas proveen eficiencia de datos.

3.7. EJEMPLO DE POLÍTICAS

Esta sección proporcionará un ejemplo del uso canónico de políticas, reglas de políticas, condiciones y acciones de las mismas.

Asuma que la siguiente regla comercial será llevada a cabo como una política:

Proveer el servicio de video "JitterFreeMPG2", para los usuarios autorizados, entre los puntos autorizados, pero solo en tiempos acordados.

Esta regla puede traducirse someramente como:

SI el usuario EN ApprovedUser
Y repara EN VideoSources
Y destino EN VideoDestinations
Y tiempo EN ApprovedTImePeriods
ENTONCES proporcione JitterFreeMPG2

La condición de la política se traduce inicialmente como:

SI el usuario es un miembro de un grupo aprobado (ApprovedUser que es autorizado para tener este servicio),
Y el servicio pedido es soportado (VideoServicesgroup),

Y la fuente de la demanda es aceptada (en el grupo de VideoSources o se ha autenticado),

Y el destino es aceptado (el grupo del de en de VideoDestinations el o se ha autenticado),

Y el tiempo pedido es ok (en approvedTimePeriods)

Aquí, los tipos de condición de política son: usuario, la fuente, destino, y tiempo, y los elementos de las condiciones de políticas son: ApprovedUsers, VideoServices, VideoSources, VideoDestinations, y ApprovedTimePeriods, que son todas las instancias predefinidas de grupos de objetos.

Note que esta política podría requerir sub-politicas en orden para que pueda ser llevado a cabo. Por ejemplo, RSVP podrían usarse las demandas a la condición previa al camino entre VideoSources y VideoDestinations.

RESUMEN

En este capítulo se presenta la forma en la que será manejado el tráfico que previamente se había seleccionado, esto se logra mediante el manejo de políticas, las cuales se encargan de regular el flujo mediante la validación de dichas condiciones.

Las políticas se comprenden de reglas, y las reglas se componen de condiciones y acciones, las cuales estarán enfocadas a validar o permitir el paso inicialmente del tráfico que tiene mayor prioridad, la cual previamente se había identificado mediante los diferentes mecanismos de identificación.

4.- PROTOCOLOS Y ARQUITECTURAS

“Los protocolos de Calidad de Servicio utilizan una variedad de mecanismos complementarios para permitir la entrega determinista”.

Las redes basadas en el protocolo estándar IP, por default proveen la entrega de datos por “Best Effort” (mejor esfuerzo). El best effort de IP permite que la complejidad permanezca en los host terminales, de esta manera, la red permanece relativamente simple. Esto es bueno, tal como lo muestra la capacidad de Internet para soportar grandes crecimientos. Como más hosts están conectados, la demanda de servicios de red eventualmente excede la capacidad, pero los servicios no son negados. Aunque los resultados varíen en entregas retrasadas (Jitter) y los paquetes perdidos no adversamente afecten a las aplicaciones típicas de Internet (e-mail, transferencia de archivos y aplicaciones Web), otras aplicaciones no pueden adaptarse a los niveles inconsistentes de los servicios.

Los retardos en la entrega, causan problemas en aplicaciones con requerimientos en tiempo real, como son multimedia y los más demandados como son aplicaciones de dos caminos, como la telefonía.

Incrementar el ancho de banda, es un primer paso necesario para acomodar estas aplicaciones en tiempo real, pero esto no es suficiente para evitar el jitter durante el estallido de tráfico. Incluso en una red IP relativamente sin carga, los retrasos de la entrega pueden variar bastante para continuar adversamente afectando a las aplicaciones en tiempo real. Para proporcionar un servicio adecuado (algunos niveles determinantes de calidad y cantidad), los servicios IP deben complementarse, estos requieren adición de algunos mecanismos en la red para distinguir al tráfico, con estrictos requerimientos cronometrados y estos puedan tolerar retrasos, jitter y pérdidas. Para hacer esto,

es que los protocolos de Calidad de Servicio (QoS) están diseñados. La Calidad de Servicio no crea ancho de banda, pero administra y éste es utilizado más efectivamente, reuniendo un rango amplio para los requisitos de las aplicaciones. La meta de QoS es proveer algunos niveles de predictabilidad y control más allá de la corriente de servicios “best effort” de IP.

Un número de protocolos de QoS ha evolucionado para satisfacer la variedad de necesidades de las aplicaciones. En este capítulo, se describirán estos protocolos individualmente, mostrando cómo incorporaron en varias arquitecturas, con el principio extremo a extremo en mente. El desafío de estas tecnologías QoS de IP es proporcionar servicios de entrega diferenciados para flujos individuales o agregados sin romper la red en el proceso. Agregando mecanismos a la red y mejorando en los servicios “best effort”, representa un cambio fundamental al plan que hizo un éxito al Internet.

Para evitar esos problemas potenciales es que los protocolos de QoS se aplican a la red, el principio extremo a extremo es todavía el principal enfoque de los arquitectos de QoS. Como resultado, el principio fundamental de *“Dejar la complejidad en los extremos y guardar el núcleo de la red simple”* es el tema central entre los planes de arquitectura de QoS. Esto no es tanto un enfoque para los protocolos de QoS individuales, pero si, en cómo ellos se complementan para habilitar QoS extremo a extremo.

4.1. PROTOCOLOS DE QoS

Hay más de una manera de caracterizar la Calidad de Servicio (QoS). Generalmente hablando, QoS, es la habilidad de un elemento de red (una aplicación, un host o un router), para proveer algunos niveles de convicción para la entrega consistente de datos de red. Algunas aplicaciones son más severas acerca de sus requerimientos de QoS que otras, y por esta razón (entre otras), se tienen dos tipos básicos disponibles de Calidad de Servicio:

- **Reservación de recursos** (Servicios integrados): Los recursos de la red son proporcionados conforme a la demanda de QoS de las aplicaciones, y sujetos a las políticas de administración de ancho de banda.
- **Priorización** (Servicios diferenciados): El tráfico de red es clasificado y proporcionado a los recursos de red de acuerdo a los criterios de políticas de administración de ancho de banda. Para habilitar QoS, los elementos de la red toman trato preferente a clasificaciones identificadas, como tener los requisitos más exigentes.

Estos tipos de QoS pueden ser aplicados individualmente (flujos) o para flujos agregados, hay otras dos maneras de caracterizar los tipos de Calidad de Servicio.

- **Por flujo**: Un flujo es definido por un individual, unidireccional cadena de datos entre dos aplicaciones (transmisor y receptor), singularmente identificado por 5 parámetros (protocolo de transporte, dirección fuente, número de puerto fuente, dirección destino, número de puerto destino).
- **Por agregado**: Un agregado es simplemente dos o más flujos. Típicamente los flujos tendrán algunas características en común (uno o más de los 5 parámetros mencionados, una etiqueta o número de prioridad, o quizás alguna información de autenticidad).

Las aplicaciones, la topología de red y las políticas dictan qué tipo de QoS es más apropiada para flujos individuales o agregados. Para acomodar las necesidades de estos diferentes tipos de QoS, hay un número de diferentes protocolos y algoritmos de Calidad de Servicio, entre estos se encuentran los siguientes:

- **Protocolo de Reservación (ReSerVation Protocol) RSVP**: Provee la señalización para habilitar la reservación de recursos de red (conocido como Servicios Integrados). Aunque se usa típicamente para un sólo

flujo, RSVP también se utiliza para flujos agregados. Se habla de flujos agregados cuando circule más de un flujo por la red.

- **Servicios Diferenciados (DiffServ)**: Proporciona una manera simple para categorizar y priorizar al tráfico agregado de red (flujos).
- **Multi Protocol Labeling Switching (MPLS)**: Proporciona la posibilidad de administrar el ancho de banda de la red a través de etiquetas en las cabeceras de los paquetes (encapsulamiento) y de routers específicos capaces de reconocerlas.
- **Subnet Bandwidth Management (SBM)**: Es un protocolo de señalización que permite la comunicación y coordinación entre los distintos nodos de red, definiendo cómo relacionar los distintos protocolos de QoS superiores con la diferentes tecnologías de capa 2 (la capa de enlacen el modelo OSI).

En la tabla 4.1, se presenta una comparación de los protocolos de QoS en términos del nivel de calidad que proporcionan y dónde es implementado es servicio y control, si es en la aplicación (App) o en la red (Net).

QoS	***Net***	***App***	***Descripción***
Mayor	**X**		Proporciona recursos extremo a extremo (redes privas y de poco tráfico).
	X	**X**	Servicio RSVP para IntServ garantizado (proporciona retroalimentación a la aplicación).
	X	**X**	RSVP para servicios de carga controlada (proporciona realimentación a la aplicación).
	X		MPLS.
	X	**X**	DiffServ aplicado al ingreso del núcleo de la red para reservación del nivel de servicio RSVP apropiado para ese flujo. Priorización utilizando SBM aplicado en LAN's.
	X	**X**	DiffServ o SBM aplicados en base por flujo de la aplicación origen.
	X		DiffServ aplicado al ingreso del núcleo de la red.
	X		El encolado por los elementos de la red (CFQ, WFQ, RED)
Menor			Servicio Best Effort

Tabla 4.1 Muestra los diferentes algoritmos y protocolos de gestión del ancho de banda, sus relativos niveles de QoS y cuando son activados por elementos de la red (Net) o por aplicaciones (App) o por ambos.

Los protocolos de QoS aquí señalados son diferentes pero no se excluyen unos a otros, sino todo lo contrario, en realidad, se complementan al momento de su aplicación para obtener los niveles de calidad requeridos en una red determinada (convergencia de redes). Esta complementación integra a una gran variedad de arquitecturas en las que los protocolos trabajan conjuntamente para proporcionar Calidad de Servicio extremo a extremo a través de múltiples proveedores de servicio.

4.2. RSVP (Resource reservation)

El protocolo de reservación de recursos [RFC2205, Versión 1 Functional Specifications], es un protocolo de señalización, que proporciona un arreglo de reservación y control para habilitar los servicios integrados [IntServ], orientados fundamentalmente a redes IP. RSVP es la más compleja de todas las tecnologías de QoS, para aplicaciones (host) y para elementos de red (Routers y Switches), Como resultado, también esto representa el mayor estándar creado desde el servicio “best effort” de IP, proporcionando el mayor nivel de QoS en términos de servicios garantizados, granularidad de localización de recursos y el mayor detalle sobre la forma de actuación de aplicaciones y usuarios que proporcionan QoS.

RSVP, fue creado en 1990 por IETF, definiendo un modelo de asignación de QoS en el que cada receptor (para una sesión) fuese responsable de elegir su propio nivel de reserva de recursos, iniciando la reserva y manteniéndola activa tanto tiempo como desee. Consistiendo en una solución distribuida que permite a múltiples receptores heterogéneos efectuar reservas específicamente dimensionadas según sus propias necesidades. Además, para mantener el control, el receptor puede enviar sus especificaciones a la fuente encargada de solicitar las reservas de la red. En definitiva, **RSVP permite que las aplicaciones soliciten una calidad de servicio específica a la red.** Más que un protocolo de encaminamiento es más bien un protocolo de control de Internet. Su tarea consiste

en establecer y mantener las reservas de recursos en un árbol de distribución, con independencia de cómo se hayan creado.

El grupo de trabajo *Integrated Services* del IETF ha considerado la existencia de varias clases de QoS, si bien, actualmente sólo dos de éstas han sido formalmente especificadas para ser utilizadas con RSVP.

- **Servicios Garantizados** (Guaranteed Service) [RFC2211]:
  Este servicio proporciona un nivel de ancho de banda y un límite en el retardo, garantizando la no existencia de pérdidas en colas. Está pensado para aplicaciones con requerimientos en tiempo real, tales como ciertas aplicaciones de audio y video. Cada router, caracteriza el Servicio Garantizado para un flujo específico asignando un ancho de banda y un espacio en buffer.

- **Servicio de Carga Controlada** (Controlled Load Service) [RFC2212]:
  A diferencia del Servicio Garantizado, este servicio no ofrece garantías en la entrega de paquetes. Así, será adecuado para aquellas aplicaciones que toleren una cierta cantidad de pérdidas y un retardo mantenidos en un nivel razonable. Los routers que implementen este servicio deben verificar que el tráfico recibido siga las especificaciones dadas por el Tspec, y cualquier tráfico que no las cumpla será reenviado por la red, como tráfico best effort.

TOMA DE DESICIONES

Para la toma de decisiones de QoS asociadas a los paquetes de una aplicación, RSVP interactúa con las entidades denominadas *packet classifier* (clasificador de paquetes) y *packet scheduler* (programador de paquetes) instaladas en el host. Primero consulta a los módulos las decisiones locales para saber si la QoS deseada puede ser provista (ya sea mediante decisiones basadas en recursos o bien mediante decisiones basadas en políticas), y en consecuencia,

establece los parámetros requeridos en el clasificador y en el programador del paquete.

La clasificación de paquetes, determina la ruta de este y el programador toma las decisiones de envío para alcanzar la QoS deseada, negociando si es necesario, con aquellos host que tengan capacidad propia de gestión de QoS, para proporcionar la QoS solicitada por RSVP.

Algunos aspectos fundamentales en el RSVP son:

- *Merging*: En los diferentes nodos en los que se va atravesando en la red por el camino de datos, se va realizando un proceso de concentración de los diferentes mensajes de petición de reservas.
- *Estado de reserva en cada nodo*: El estado RSVP se crea y refresca periódicamente por mensajes Path y Resv. Permitiendo observar el estado en que se encuentran los recursos.
- *Estilo de reserva*: Una petición de reserva incluye un conjunto de opciones que se conocen como el estilo de reserva. Las distintas combinaciones de estas opciones conforman los tres estilos de reserva en uso, Wildcar-Filter (Filtro Libre), Fixed-Filter (Filtro Fijo), y Shared-Explicit (Explicito Compartido).

TIPOS DE MENSAJES

Existen dos tipos de mensajes RSVP fundamentales, **Resv** y **Path.** Una aplicación solicita participar en una sesión RSVP como **emisor**, enviando un mensaje Path en el mismo sentido que el flujo de datos, por las rutas uni/multicast proporcionadas por el protocolo de ruteo. A la recepción de este mensaje, el **receptor** transmite un mensaje Resv, dirigido hacia el emisor de los datos, siguiendo exactamente el camino inverso al de los mismos, en el cual se especifica el tipo de reserva a realizar en todo el camino.

En la figura 4.2, se pueden visualizar los mensajes descritos así como su intercambio.

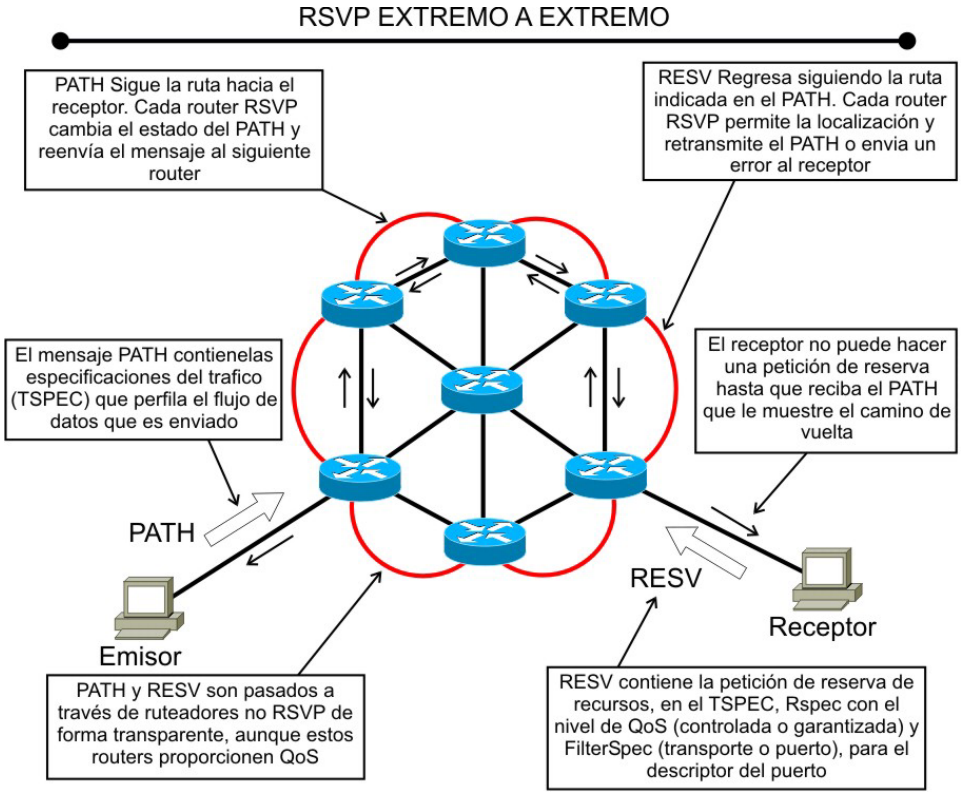


Figura 4.2 Intercambio de mensajes en RSVP

MENSAJE PATH

En general, sin especificar tipos de QoS un mensaje Path, contiene:

- Sender Template: Parámetro por el cual, se describe el formato de los paquetes que el emisor generara.
- Sender Tspec: Describe el tráfico que la aplicación estima que generará.
- Adspec: Información sobre la QoS y propiedades de la aplicación.
- Dirección del PHOP: Necesaria para poder encaminar los mensajes Resv.

De entre estos, para la gestión de Calidad de Servicio, se utilizan los objetos:

- Sender Tspec: Lista los servicios de control de QoS ofrecidos por el remitente y el ancho de banda que requieren. Los ruteadores intermedios anotan esta opción y reexpiden el mensaje sin modificarlo.
- Adspec: Contiene información tal como la disponibilidad de un determinado servicio de control de QoS en el ruteador y los recursos disponibles para cada uno de los servicios de control. Cada router intermedio introduce modificaciones en este objeto, para reflejar sus posibilidades, cuando el mensaje PATH llega a un receptor, el ADSPEC contiene un resumen de los servicios QoS disponibles en la ruta de datos.

  En cada elemento de red, el ADSPEC se pasará al modulo de control de trafico, el cual determinará si el servicio QoS especificado está implementado en el nodo. Por defecto, se generará un objeto que soportará todos los servicios de QoS que admite el emisor. Cuando el mensaje Path llega a un receptor, los datos del SENDER-TSPEC y ADSPEC se pasan a través de la API (Aplication Programming Interface) a la aplicación, la cual interpretará estos datos y los utilizará para seleccionar parámetros de reserva de recursos.

MENSAJE RESV

En el caso de mensajes RESV es posible encontrarse, con:

- Especificaciones de flujo o FLOWSPEC: Los flowspec especifican los recursos a reservar para la sesión (los requisitos de ancho de banda y los retardos que se precisan), estableciendo los parámetros en el *packet scheduler* del nodo.

- Especificaciones de filtro o FILTERSPEC: Los filterspec que indican qué subconjunto de paquetes de la sesión han de recibir la QoS definida por los flowspecs, para ello, utilizan cualquier campo de las cabeceras de protocolo, o de las cabeceras de aplicaciones (discriminar por dirección de origen, por aplicación, por el puerto, por el protocolo, etc.).

  Existen distintos tipos de filtros:

  - Filtro comodín: El receptor reserva recursos que son comunes a todos los emisores del grupo multienvío (usado por ejemplo en audio conferencias, en las que sólo hay un emisor a la vez, los recursos se reservan para todo el grupo, sea cual sea el emisor).
  - Filtro fijo: El receptor especifica una fuente concreta o una lista de fuentes precisas (usado cuando el receptor se suscribe a una película o a un canal de TV de Internet. Los recursos se reservan solo para ese canal).
  - Filtro dinámico: El receptor especifica un conjunto de canales a los que se aplica una reserva (usado en videoconferencias, donde sólo hay un número máximo de cámaras activas a la vez, por ejemplo, si sólo están activas la cámara del que pregunta y la del que responde, se reservarían recursos para dos canales).

Aún cuando las peticiones de reserva RSVP se originan en el receptor final, es en la estación del emisor donde se produce el control QoS.

IMPLEMENTACIONES RSVP/QoS

A partir de las especificaciones publicadas en los diferentes RFCs (2205, 2206, 2207, 2208, 2209), más de 30 empresas del mundo de la Informática y las Telecomunicaciones han decido realizar diferentes implementaciones del

protocolo, tanto en su comportamiento como router como en el host, junto con la realización de diferentes herramientas de aplicación.

Analizando el estado actual de dichas implementaciones para aquellas empresas más destacadas del sector, teniendo en cuenta el sistema operativo utilizado, la tecnología de red, la capacidad de QoS, las aplicaciones, las características no soportadas, la interoperabilidad y la disponibilidad del producto, obtenemos:

- Los sistemas operativos utilizados están en función del sistema que cada compañía utiliza en sus equipos, siendo los sistemas más utilizados: Solaris, Linux, Windows, cada uno de ellos en sus versiones actuales.
- La tecnología de red utilizada es prácticamente común en casi todas ellas: ATM, Frame Relay, Ethernet, FDDI.
- Respecto a la capacidad de QoS, todos los productos cumplen con las especificaciones de RSVP y de Integrated Services ofreciendo servicios de carga controlada. El servicio garantizado sólo está disponible en algunas implementaciones. La opción de servicios diferenciados, es minoritaria encontrándose en proceso de implantación en algunos casos.
- Los productos realizados se ofrecen para aplicaciones de telefonía y videoconferencia esencialmente, y en algunos casos se proponen para ser utilizados en asignación de ancho de banda para usuarios preferentes y Virtual Private Networks.
- La disponibilidad está dividida entre productos en venta y productos gratuitos disponibles al público. El resto son de uso interno, gratis solo para organizaciones o bien se encuentran incluidos en otros productos del mismo fabricante.

En la tabla 4.3, se presenta el estado actual de las implementaciones para distintas empresas:

COMPAÑIA	PLATAFORMA	SISTEMA OPERATIVO	TECNOLOGIA DE RED	DISPONIBILIDAD	QoS	APLICACIONES	NO SOPORTA	INTEROPERABILIDAD
CISCO	Router	IOS	ATM, Frame Relay, FDDI	Venta	CL	Telefonía, Videoconferencia	Ipv6, IPSEC	ISI, Intel, Microsoft
DEC	Host	Digital Unix	Ethernet, FDDI	Publico y Gratis	CL	Telefonía, Videoconferencia	UDP, Encapsulamiento, IPSEC	Sun Solaris, Microsoft
HP	Host	HP-UX	Ethernet	Gratis Empresas	CL	----------	GS, IPSEC, MIB	Cisco Router
IBM	Router	IBM	ATM, FR, Ethernet, FDDI	Productos IBM	CL	VDN / VPN	Blockade State, MIB	Intel, Microsoft
INTEL	Router, Host	Microsoft	----------	Venta	CL	----------	----------	Cisco, Sun Solaris
MICROSOFT	Host	Win95, Win98, WinNT	ATM, Ethernet, FDDI	Con Win98, WinNT	CL, GS	Telefonía, Videoconferencia	----------	Cisco, Intel, Sun Host
SUN	Host, Toolkit	Solaris	----------	Venta	CL	----------	MIB	----------

Tabla 4.3 QoS para las empresas.

Es interesante conocer cuales son las características del protocolo que todavía no están implementadas y que cada fabricante, en función al grado de desarrollo que tenga su producto, indica como futuras realizaciones. Así, la compatibilidad con el protocolo IPv6, el servicio garantizado y el IPSEC son las referencias de no implementación más señaladas. Además, algunos productos no contemplan el encapsulamiento UDP, realización de túneles para el paso por redes no RSVP, mensajes de diagnóstico y autenticación.

4.3. DIFFSERV

Differentiated Services (Servicios Diferenciados), es un protocolo de QoS propuesto por la IETF [RFC 2475 y RFC 2474], que permite distinguir diferentes clases de servicio marcando los paquetes. Permite a los proveedores de servicios de Internet y a usuarios de grandes redes IP corporativas desplegar rápidamente diferentes niveles de Calidad de Servicio. A diferencia de RSVP no especifica un sistema de señalización, y éste, consiste en un método para marcar o etiquetar

paquetes, permitiendo a los routers modificar su comportamiento de envío. Cada tipo de etiqueta representa un determinado tipo de QoS y el tráfico con la misma que, se trata de la misma manera.

Para proporcionar los diferentes niveles de servicio, utiliza el campo *type of service* (ToS) o DiffServ Codepoint (DSCP) de la cabecera del estándar Ipv4 e Ipv6. Este es un campo de 8 bits, estando los 2 últimos reservados. Con los 6 restantes se consiguen 64 combinaciones: 48 para el espacio global y 16 para uso local.

<table>
<tr><td>VERS</td><td></td><td>Tipo de Servicio</td><td colspan="2">Longitud</td></tr>
<tr><td colspan="3">IDENTIFICACION</td><td>FLAGS</td><td>OFFSET del fragmento</td></tr>
<tr><td>Tiempo de vida</td><td colspan="2">PROTOCOLO</td><td colspan="2">CHECKSUM DE LA CABECERA</td></tr>
<tr><td colspan="5">Dirección IP fuente</td></tr>
<tr><td colspan="5">Dirección IP destino</td></tr>
<tr><td colspan="4">Opciones IP</td><td>PADDING</td></tr>
<tr><td colspan="5">DATOS</td></tr>
</table>

Figura 4.4 Paquete Ipv4.

Este tipo de funcionamiento de QoS se vé sustentado con los Service Level Agreement (SLA) o acuerdos de nivel de servicio entre el cliente y su proveedor de servicio (por ejemplo, Internet Service Provider (ISP)), básicamente un SLA especifica las clases de servicios soportadas y la cantidad de tráfico permitida en cada clase. Los SLA pueden ser estáticos o dinámicos, según las necesidades del cliente, en caso que se utilice de forma dinámica, el cliente debe utilizar protocolos de señalización como RSVP.

A continuación, se muestran los componentes que forman parte del protocolo, así como su funcionamiento.

TIPO DE MARCAS

Para clasificar el tráfico mediante DiffServ, se muestran tres opciones básicas de marcas:

- None (ninguna): Se ofrece el servicio de Best Effort convencional.
- Assured and in profile (asegurado y definida dentro del perfil): Está definida en el SLA entre el cliente y el proveedor del servicio.
- Assured and out of profile (asegurado y fuera del perfil): No cumple con lo definido en el SLA entre el cliente y el proveedor del servicio

CLASE DE SERVICIO

El protocolo DiffServ, tiene definido dos tipos de clases de servicios: el servicio Premium y el del Asegurado, además de soportar el convencional servicio Best Effort.

- Premium Service: Proporciona bajo retardo y bajo nivel de jitter para clientes que generen grandes picos de tráfico. Este nivel de servicio está especificado en el SLA que el cliente contrata con el ISP. El SLA especifica la velocidad pico deseada y ofrecida, además del ancho de banda proporcionado. El ISP debe responsabilizarse de proporcionarla y el cliente de no superar esta tasa. Este tipo de servicio es el apropiado para la telefonía por Internet, la videoconferencia o para la creación de líneas virtuales en redes privadas virtuales (VPN's).
- Assured Service: Es solicitado por clientes que necesitan un cierto nivel de fiabilidad de sus ISP's, incluso si existe congestión. Sus especificaciones también vienen determinadas en los SLA's. En donde se indica la cantidad de ancho banda disponible para el cliente, pero es él, quien define cómo compartirán sus aplicaciones en el ancho de

banda. Las aplicaciones apropiadas son las mismas que utilizarían el servicio Best Effort.

ENVÍO DE PAQUETES

El proceso de envío de los paquetes modificados por DiffServ desde un router, se conocen como PHB, Per Hop Behavior policies o comportamiento por salto. El PHB indica qué tratamiento han recibido los paquetes a lo largo de su transmisión para entregar DiffServ, tales como tipo de políticas aplicadas, conformación del tráfico, posibles remarcados en el campo DiffServ, encolamientos y gestión del tráfico.

Existen varios tipos de PHB's:

- Expedited Forwarding: Este proceso, tiene un sólo valor de DiffServ (Codepoint). Minimiza el retardo, el jitter y asegura baja pérdida de paquetes, proporcionando el mayor nivel de QoS.
- Assured Forwarding: Define cuatro clases de probabilidad de tráfico, con 3 variaciones en cada una. La probabilidad de entrega no es tan alta como en el caso anterior, provocándose retardos.
- Default: Funciona como el servicio Best Effort tradicional.

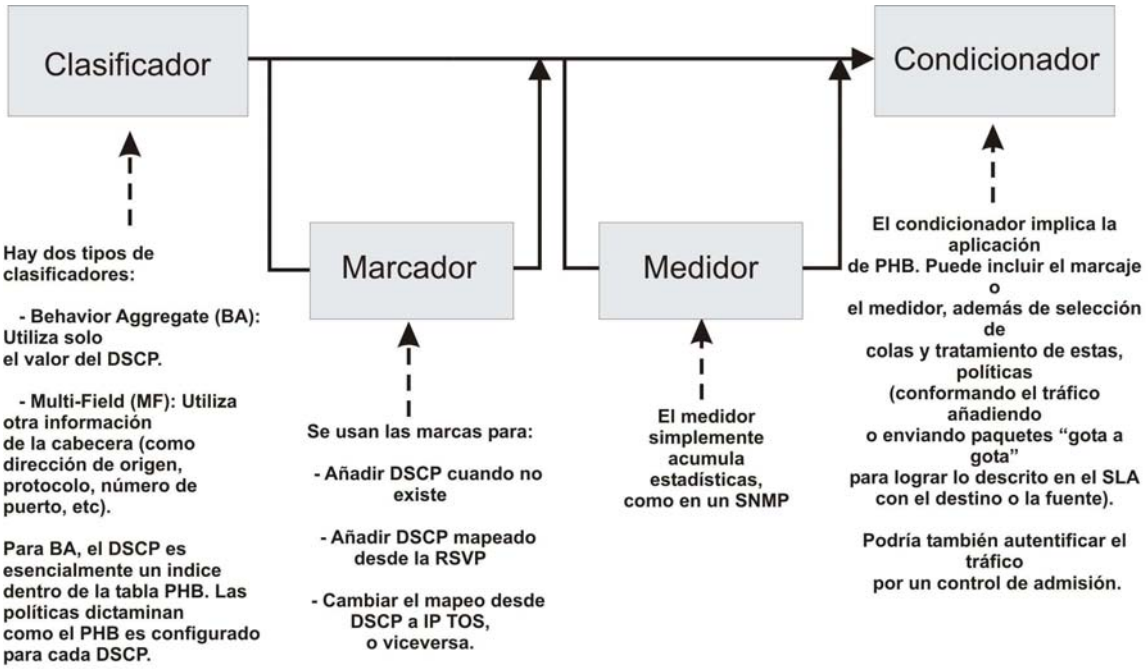


Figura 4.5 Arquitectura de DiffServ.

El borrador de DiffServ, define además la implementación del protocolo en dos tipos de routers: condicionadores de tráfico y capacitados para DiffServ.

- Clasificador de tráfico (capacitado para DiffServ): El clasificador de tráfico selecciona los paquetes basándose en uno o varios campos de la cabecera. Aportan, por lo tanto, funciones de programación y deben de modificar su comportamiento de envío en función de las marcas o etiquetas. Esta diferencia se realiza utilizando el campo DiffServ (ToS), proporcionando un máximo de 64 clases de servicio. Cada router ordena los paquetes en colas basándose en el citado campo, aplicando diferentes políticas de priorización a las mismas.
- Condicionadores de tráfico: Este tipo de router altera los paquetes para que cumplan las reglas de los servicios. Realizando funciones sofisticadas de etiquetado, modelado y monitorización.

4.4. MPLS (Multiprotocol Label Switching)

MPLS (*MultiProtocol Label Switching*), es una de las soluciones propuestas por el *Internet Engeenering Task Force* (IETF) con el objetivo de proporcionar Calidad de Servicio (*QoS*) a una red de datos, éste, trata de un estándar de arquitectura multinivel, capaz de soportar cualquier tipo de tráfico, independiente del nivel de transporte de datos sobre el que se apoya, capaz de ofrecer una gran eficiencia a la hora de realizar la transmisión de paquetes de un extremo a otro de la red MPLS gracias a la combinación de la flexibilidad del nivel de red, con los beneficios propios de un modelo de red orientado a conexión.

En una red IP tradicional, un router conmuta los paquetes de una interfaz de entrada a una interfaz de salida; además, actualiza la información de enrutamiento. Para enviar los paquetes, debe examinar la cabecera del paquete IP para cada uno. Estas dos funciones, envío y enrutamiento, tienen lugar en cada salto que realiza un paquete para cada uno de los que atraviesan la red. Lo que se busca con MPLS a este respecto, es llevar las funciones de enrutamiento únicamente a los equipos exteriores del dominio MPLS, de forma que en el interior de dicho dominio no sea necesario realizar labores de enrutamiento, sino sólo de conmutación mediante la consulta de unas etiquetas añadidas a cada paquete en el momento de entrada al dominio.

ELEMENTOS BÁSICOS DE LA ARQUITECTURA MPLS

LSR (Label Switched Router): Es el tipo de router que permite MPLS. El LSR puede ser interior o extremo. Los LSR extremo añaden o eliminan etiquetas. Los anteriores sustituyen unas etiquetas por otras.

Etiqueta: es un identificador corto, de longitud fija y con significado local empleado para identificar un FEC. Un paquete puede tener una o más etiquetas apiladas (jerarquía). Cuando un paquete atraviesa dominios interiores a otros dominios, es

cuando se produce el apilamiento de etiquetas. El LSR siempre consultará la etiqueta de nivel superior. La etiqueta se añade, de forma general, entre las cabeceras de nivel 2 y 3 (Ethernet, PPP...).

FEC (Forwarding Equivalence Class): Agrupación de paquetes que comparten los mismos atributos (dirección destino, VPN...) y/o requieren el mismo servicio (multicast, QoS...). El FEC se asigna en el momento en que el paquete entra a la red. Todos los paquetes que forman parte del FEC, siguen un mismo LSP.

LSP (Label Switched Path): Es la ruta a través de uno o más LSRs en un nivel de jerarquía que sigue un paquete de un FEC en particular. Esta ruta puede establecerse tanto mediante protocolos de enrutamiento como manualmente.

INTERCAMBIO DE ETIQUETAS

El algoritmo de intercambio de etiquetas requiere la clasificación de los paquetes a la entrada del dominio MPLS para poder hacer la asignación por el LSR de cabecera.

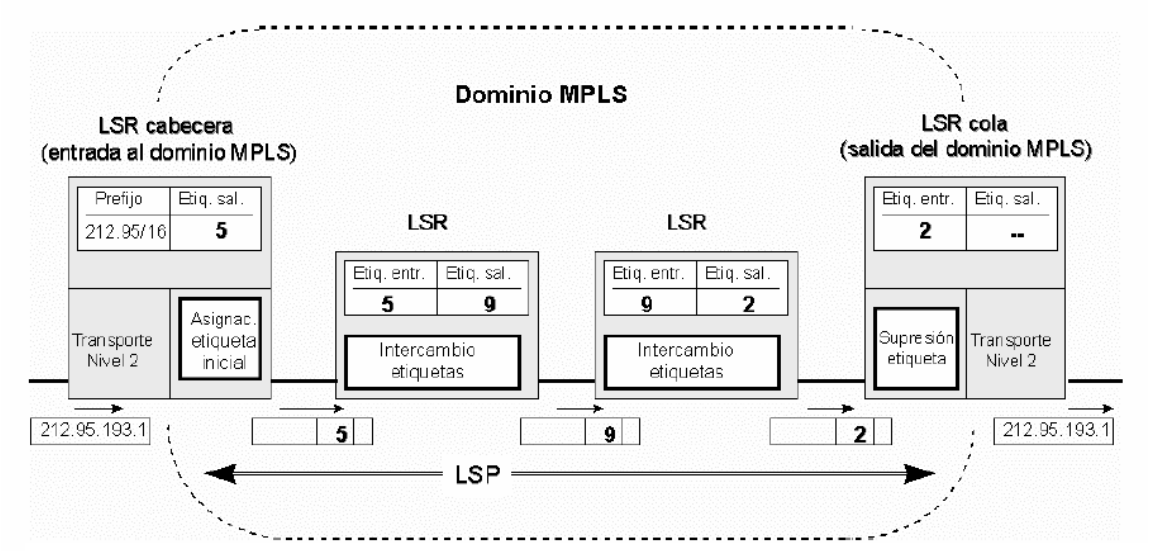


Figura 4.6 Ejemplo de envío de un paquete por un LSP.

En la figura 4.6, el LSR de entrada recibe un paquete normal (sin etiquetar) cuya dirección de destino es 212.95.193.1. El LSR consulta la tabla de enrutamiento y asigna el paquete a la clase FEC definida por el grupo 212.95/16. Asimismo, este LSR le asigna una etiqueta (con valor 5 en el ejemplo) y envía el paquete al siguiente LSR del LSP. Dentro del dominio MPLS los LSR ignoran la cabecera IP; solamente analizan la etiqueta de entrada, consultan la tabla correspondiente (tabla de conmutación de etiquetas) y la reemplazan por otra nueva, de acuerdo con el algoritmo de intercambio de etiquetas. Al llegar el paquete al LSR de cola (salida), ve que el siguiente salto lo saca de la red MPLS; al consultar ahora la tabla de conmutación de etiquetas quita ésta y envía el paquete por *routing* convencional. Como se ve, la identidad del paquete original IP queda enmascarada durante el transporte por la red MPLS, que no "mira" sino las etiquetas que necesita para su envío por los diferentes saltos LSR que configuran los caminos LSP. Las etiquetas se insertan en cabeceras MPLS, entre los niveles 2 y 3. Según las especificaciones del IETF, MPLS debía funcionar sobre cualquier tipo de transporte: PPP, LAN, ATM, Frame Relay, etc. Por ello, si el protocolo de transporte de datos contiene ya un campo para etiquetas (como ocurre con los campos VPI/VCI de ATM y DLCI de Frame Relay), se utilizan esos campos nativos para las etiquetas. Sin embargo, si la tecnología de nivel 2 empleada no soporta un campo para etiquetas (p. ej. enlaces PPP o LAN), entonces se emplea una cabecera genérica MPLS de 4 octetos, que contiene un campo específico para la etiqueta y que se inserta entre la cabecera del nivel 2 y la del paquete (nivel 3).

En la figura 4.7, se representa el esquema de los campos de la cabecera genérica MPLS y su relación con las cabeceras de los otros niveles. Según se muestra en la figura 4.7, los 32 bits de la cabecera MPLS se reparten en: 20 bits para la etiqueta MPLS, 3 bits para identificar la clase de servicio en el campo EXP (experimental, anteriormente llamado CoS), 1 bit de *stack* para poder apilar etiquetas de forma jerárquica (S) y 8 bits para indicar el TTL (*time-to-live*) que sustenta la funcionalidad estándar TTL de las redes IP. De este modo, las cabeceras MPLS permiten cualquier tecnología o combinación de tecnologías de

transporte, con la flexibilidad que esto supone para un proveedor IP a la hora de extender su red.

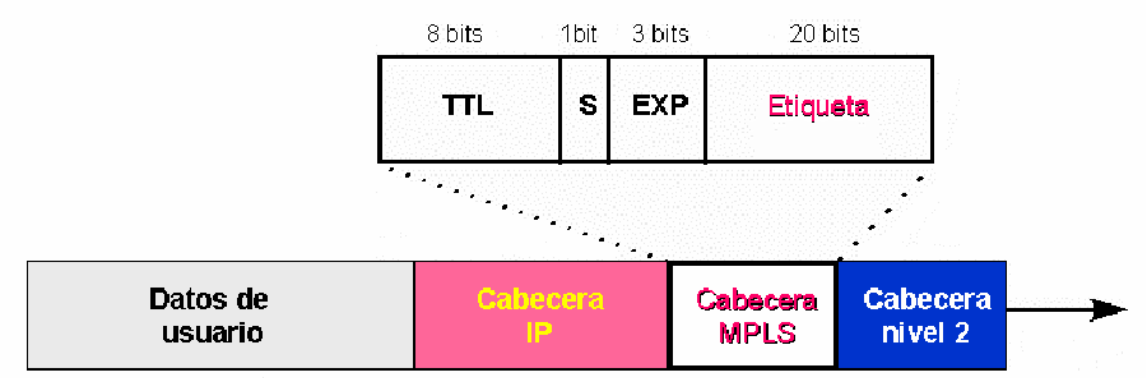


Figura 4.7 Estructura de la cabecera genérica MPLS.

CONTROL DE LA INFORMACIÓN EN MPLS

Hasta ahora se ha visto el mecanismo básico de envío de paquetes a través de los LSPs mediante el procedimiento de intercambio de etiquetas según las tablas de los LSRs. Pero queda por ver dos aspectos fundamentales:

1. Cómo se generan las tablas de envío que establecen los LSPs.
2. Cómo se distribuye la información sobre las etiquetas a los LSRs.

El primero de ellos está relacionado con la información que se tiene sobre la red: topología, patrón de tráfico, características de los enlaces, etc. Es la información de control típica de los algoritmos de encaminamiento. MPLS necesita esta información de *routing* para establecer los caminos virtuales LSPs. Lo más lógico es utilizar la propia información de encaminamiento que manejan los protocolos internos IGP (OSPF, IS-IS, RIP...) para construir las tablas de encaminamiento (cabe recordar que los LSR son *routers* con funcionalidad añadida). Esto es lo que hace MPLS precisamente: para cada "ruta IP" en la red se crea un "camino de etiquetas" a base de concatenar las de entrada/salida en cada tabla de los LSRs; el protocolo interno correspondiente se encarga de pasar la información necesaria.

El segundo aspecto, se refiere a la información de "señalización", pero siempre que se quiera establecer un circuito virtual se necesita algún tipo de señalización para marcar el camino, es decir, para la distribución de etiquetas entre los nodos. Sin embargo, la arquitectura MPLS, no asume un único protocolo de distribución de etiquetas; de hecho se están estandarizando algunos existentes con las correspondientes extensiones; unos de ellos es el protocolo RSVP del Modelo de Servicios Integrados del IETF (recuérdese que ése era uno de los requisitos). Pero, además, en el IETF se están definiendo otros nuevos, específicos para la distribución de etiquetas, cual es el caso del *Label Distribution Protocol* (LDP).

APLICACIONES DE MPLS

Las principales aplicaciones que hoy en día tiene MPLS son:

- **Ingeniería de tráfico.**
- **Diferenciación de niveles de servicio mediante clases (CoS).**
- **Servicio de redes privadas virtuales (VPN).**

A continuación, se describen brevemente las características de estas aplicaciones y las ventajas que MPLS supone para ello frente a otras soluciones tradicionales.

INGENIERÍA DE TRÁFICO

El objetivo básico de la Ingeniería de tráfico es adaptar los flujos de tráfico a los recursos físicos de la red. La idea es, equilibrar de forma óptima la utilización de esos recursos, de manera que no haya algunos que estén suprautilizados, con posibles cuellos de botella, mientras otros puedan estar infrautilizados.

Al inicio de la ultima década del siglo pasado, los esquemas para adaptar de forma efectiva los flujos de tráfico a la topología física de las redes IP eran bastante rudimentarios. Los flujos de tráfico siguen el camino más corto calculado por el algoritmo IGP correspondiente. En casos de congestión de algunos enlaces, el problema se resolvía a base de añadir más capacidad a los mismos. La Ingeniería de tráfico, consiste en trasladar determinados flujos seleccionados por el algoritmo IGP sobre enlaces más congestionados, a otros enlaces más descargados, aunque estén fuera de la ruta más corta (con menos saltos).

En figura 4.8, se comparan estos dos tipos de rutas para el mismo par de nodos origen-destino.

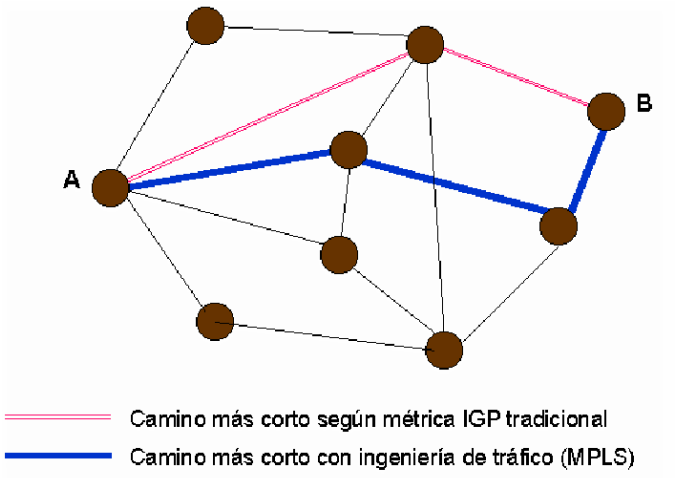


Figura 4.8 Comparación entre el camino más corto IGP con ingeniería de tráfico.

El camino más corto entre A y B según la métrica normal IGP es el que tiene sólo dos saltos, pero puede que el exceso de tráfico sobre esos enlaces o el esfuerzo de los routers correspondientes, haga aconsejable el uso del camino alternativo indicado con un salto más. MPLS es una herramienta efectiva para esta aplicación en grandes *backbones*, ya que:

- Permite al administrador de la red el establecimiento de rutas explícitas, especificando el camino físico exacto de un LSP.
- Permite obtener estadísticas de uso LSP, que se pueden utilizar en la planificación de la red y como herramientas de análisis de cuellos de botella y carga de los enlaces, lo que resulta bastante útil para planes de expansión futura.
- Permite hacer "encaminamiento restringido" (*Constraint-based Routing*, CBR), de modo que el administrador de la red pueda seleccionar determinadas rutas para servicios especiales (distintos niveles de calidad). Por ejemplo, con garantías explícitas de retardo, ancho de banda, fluctuación, pérdida de paquetes, etc.

La ventaja de la Ingeniería de tráfico MPLS es que se puede hacer directamente sobre una red IP, al margen de que haya o no una infraestructura ATM por debajo, todo ello de manera más flexible y con menores costos de planeación y gestión para el administrador, y con mayor calidad de servicio para los clientes.

CLASES DE SERVICIO (COS)

MPLS está diseñado para poder cursar servicios diferenciados, según el Modelo DiffServ del IETF. Este modelo, define una variedad de mecanismos para poder clasificar el tráfico en un reducido número de clases de servicio, con diferentes prioridades. Según los requisitos de los usuarios, DiffServ permite diferenciar servicios tradicionales tales como el WWW, el correo electrónico o la

transferencia de archivos (para los que el retardo no es crítico), de otras aplicaciones mucho más dependientes del retardo y de la variación del mismo, como son las de video y voz interactiva. Para ello, se emplea el campo ToS (*Type of Service*), rebautizado en DiffServ como el octeto DS. Esta es la técnica QoS de marcar los paquetes que se envían a la red.

MPLS se adapta perfectamente a ese modelo, ya que las etiquetas MPLS tienen el campo EXP para poder propagar la clase de servicio CoS en el correspondiente LSP. De este modo, una red MPLS puede transportar distintas clases de tráfico, ya que:

- El tráfico que fluye a través de un determinado LSP se puede asignar a diferentes colas de salida en los diferentes saltos LSR, de acuerdo con la información contenida en los bits del campo EXP.
- Entre cada par de LSR exteriores se pueden provisionar múltiples LSPs, cada uno de ellos con distintas prestaciones y con diferentes garantías de ancho de banda. Por ejemplo, un LSP puede ser para tráfico de máxima prioridad, otro para una prioridad media y un tercero para tráfico *best-effort*, tres niveles de servicio, primera, preferente y turista, que, lógicamente, tendrán distintos precios.

REDES PRIVADAS VIRTUALES (VPNS)

Una red privada virtual (VPN) se construye basándose en conexiones realizadas sobre una infraestructura compartida, con funcionalidades de red y de seguridad equivalentes a las que se obtienen con una red privada. El objetivo de las VPNs es el soporte de aplicaciones intra/extranet, integrando aplicaciones multimedia de voz, datos y video sobre infraestructuras de comunicaciones eficaces y rentables. La seguridad supone aislamiento, y "privada" indica que el usuario "cree" que posee los enlaces. Las IP VPNs son soluciones de comunicación VPN basada en el protocolo de red IP de la Internet. En este

apartado, se van a describir brevemente las ventajas que MPLS ofrece para este tipo de redes frente a otras soluciones tradicionales.

Las VPNs tradicionales se han venido construyendo sobre infraestructuras de transmisión compartidas con características implícitas de seguridad y respuesta predeterminada. Tal es el caso de las redes de datos Frame Relay, que permiten establecer PCVs entre los diversos nodos que conforman la VPN. La seguridad y las garantías las proporcionan la separación de tráficos por PVC y el caudal asegurado (CIR). Algo similar se puede hacer con ATM, con diversas clases de garantías. Los inconvenientes de este tipo de solución es que la configuración de las rutas se basa en procedimientos más bien artesanales, al tener que establecer cada PVC entre nodos, con la complejidad que esto supone al proveedor en la gestión (y los mayores costos asociados). Si se quiere tener conectados a todos con todos, en una topología lógica totalmente mallada, añadir un nuevo emplazamiento supone retocar todos los CPEs del cliente y restablecer todos los PVCs. Además, la popularización de las aplicaciones TCP/IP, así como la expansión de las redes de los NSPs, ha llevado a tratar de utilizar estas infraestructuras IP para el soporte de VPNs, tratando de conseguir una mayor flexibilidad en el diseño e implantación y unos menores costos de gestión y provisión de servicio. La forma de utilizar las infraestructuras IP para servicio VPN (IP VPN) ha sido la de construir túneles IP de diversos modos.

El objetivo de un túnel sobre IP es crear una asociación permanente entre dos extremos, de modo que funcionalmente aparezcan conectados. Lo que se hace es utilizar una estructura no conectiva como IP para simular esas conexiones: una especie de tuberías privadas por las que no puede entrar nadie que no sea miembro de esa IP VPN. “No es el objetivo de esta sección una exposición completa de IP VPNs sobre túneles; se pretende tan sólo resumir sus características para poder apreciar luego las ventajas que ofrece MPLS frente a esas soluciones”.

Los túneles IP en conexiones se pueden establecer de dos formas:

- · En el nivel 3, mediante el protocolo IPSec10 del IETF.
- · En el nivel 2, mediante el encapsulamiento de paquetes privados (IP u otros) sobre una red IP pública de un NSP.

En las VPNs basadas en túneles IPSec, la seguridad requerida se garantiza mediante el cifrado de la información de los datos y de la cabecera de los paquetes IP, que se encapsulan con una nueva cabecera IP para su transporte por la red del proveedor. Es relativamente sencillo de implementar, bien sea en dispositivos especializados, tales como cortafuegos, como en los propios *routers* de acceso del NSP. Además, como es un estándar, IPSec permite crear VPNs a través de redes de distintos NSPs que sigan el estándar IPSec. Pero como el cifrado IPSec oculta las cabeceras de los paquetes originales, las opciones QoS son bastante limitadas, ya que la red no puede distinguir flujos por aplicaciones para asignarles diferentes niveles de servicio. Además, sólo vale para paquetes IP nativos, IPSec no admite otros protocolos.

En los túneles de nivel 2 se encapsulan paquetes multiprotocolo (no necesariamente IP), sobre los datagramas IP de la red del NSP. De este modo, la red del proveedor no pierde la visibilidad IP, por lo que hay mayores posibilidades de QoS para priorizar el tráfico por tipo de aplicación IP. Los clientes VPN pueden mantener su esquema privado de direcciones, estableciendo grupos cerrados de usuarios, si así lo desean. (Además de encapsular los paquetes, se puede cifrar la información por mayor seguridad, pero en este caso limitando las opciones QoS). A diferencia del proceso anterior, la operación de túneles de nivel 2 está condicionada a un único proveedor.

A pesar de las ventajas de los túneles IP sobre los PVCs, ambos enfoques tienen unas características comunes que las hacen menos eficientes frente a la solución MPLS:

- Están basadas en conexiones punto a punto (PVCs o túneles).
- La configuración es manual.
- La provisión y gestión son complicadas; una nueva conexión supone alterar todas las configuraciones.
- Plantean problemas de crecimiento al añadir nuevos túneles o circuitos virtuales.
- La gestión de QoS es posible en cierta medida, pero no se puede mantener extremo a extremo a lo largo de la red, ya que no existen mecanismos que sustenten los parámetros de calidad durante el transporte.

Realmente, el problema que plantean estas IP VPNs es que están basadas en un *modelo topológico superpuesto* sobre la topología física existente, basados en túneles extremos a extremo (o circuitos virtuales) entre cada par de *routers* de cliente en cada VPN. De ahí las desventajas en cuanto a la poca flexibilidad en la provisión y gestión del servicio, así como en el crecimiento cuando se quieren añadir nuevos emplazamientos. Con una arquitectura MPLS se obvian estos inconvenientes ya que el modelo topológico no se superpone sino que se acopla a la red del proveedor. En el modelo acoplado MPLS, en lugar de conexiones extremo a extremo entre los distintos emplazamientos de una VPN, lo que hay son conexiones IP a una "nube común" en las que solamente pueden entrar los miembros de la misma VPN. Las "nubes" que representan las distintas VPNs se implementan mediante los caminos LSPs creados por el mecanismo de intercambio de etiquetas MPLS. Los LSPs son similares a los túneles en cuanto a que la red transporta los paquetes del usuario (incluyendo las cabeceras) sin examinar el contenido, a base de encapsularlos sobre otro protocolo.

Aquí está la diferencia: en los túneles se utiliza el encaminamiento convencional IP para transportar la información del usuario, mientras que en MPLS esta información se transporta sobre el mecanismo de intercambio de etiquetas, que no ve para nada el proceso de *routing* IP. Sin embargo, sí se

mantiene en todo momento la visibilidad IP hacia el usuario, que no sabe nada de rutas MPLS sino que ve una Internet privada (Intranet) entre los miembros de su VPN. De este modo, se pueden aplicar técnicas QoS basadas en el examen de la cabecera IP, que la red MPLS podrá propagar hasta el destino, pudiendo así reservar ancho de banda, priorizar aplicaciones, establecer CoS y optimizar los recursos de la red con técnicas de Ingeniería de tráfico.

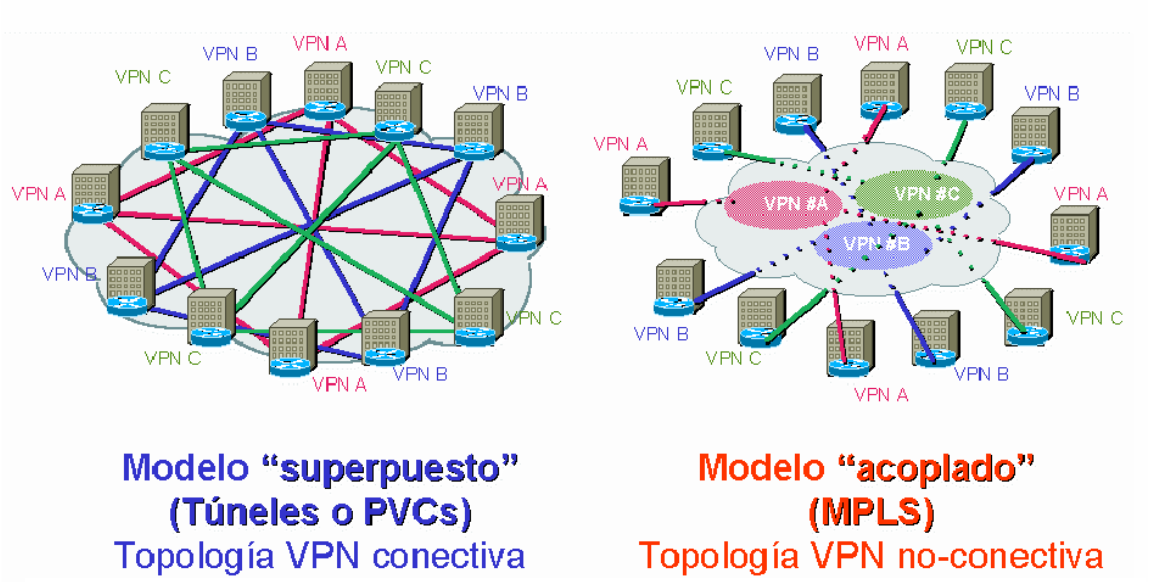


Figura 4.9 Modelo "superpuesto" (Túneles PVC's) vs. Modelo "acoplado" (MPLS).

En la figura 4.9, se representa una comparación entre ambos modelos. La diferencia entre los túneles IP convencionales (o los circuitos virtuales) y los "túneles MPLS" (LSPs) está en que éstos se crean *dentro de la red*, basados en LSPs, y no de extremo a extremo *a través de la red.*

4.5. SBM (Subnet Bandwidth Management)

Hasta ahora se ha visto cómo obtener QoS extremo a extremo entre el emisor y el receptor, esto significa que cada router a lo largo de la ruta debe soportar la tecnología de QoS que se esté utilizando, pero también hay que tener

en cuenta la posibilidad de conseguir QoS en los nodos finales (top-to-bottom). Para ello es necesario que:

- Los **host emisor** y **receptor** permiten la obtención de QoS, siendo necesario que las aplicaciones lo acepten explícitamente o, en su nombre, que lo sustente el sistema implícitamente. Cada capa OSI, desde la de aplicación a las inferiores, deben utilizar también QoS para asegurar que las peticiones de alta prioridad sean tratadas desde el host, emisor y receptor.
- Suponiendo que los sistemas finales se conecten a una red de área local (LAN), éstas, deben permitir QoS, de forma que las tramas de alta prioridad sean tratadas primeramente mientras circulan por la red (host a host, host a router, router a router). De esta forma se está proporcionando QoS en la capa 2 del modelo OSI, capa de enlace, mientras que los protocolos anteriores ofrecían QoS en otras capas: Diffserv en la capa 3 y RSVP y MPLS en capas superiores.

Existen algunas tecnologías creadas para proporcionar QoS en la capa de enlace, como ATM, pero ésta es una tecnología imposible de implementar por algunas empresas, debido a su costo económico y a su complejidad. Todas estas empresas, por el contrario, utilizan otras tecnologías más comunes para sus LANs, tales como Ethernet, que originalmente no fueron diseñadas para ofrecer QoS. Ethernet proporciona, simplemente, un servicio análogo al prestado por IP, el servicio Best Effort, en el que existe la posibilidad de que se produzcan los retardos y variaciones (jitter) que pueden afectar a aplicaciones de tiempo real. Por todos estos eventos, IEEE ha redefinido el estándar Ethernet y otras tecnologías de la capa de enlace para proporcionar QoS, mediante la diferencia de tráfico.

Los estándares de IEEE 802.1p, 802.1q y 802.1D definen como un switch Ethernet puede clasificar las tramas para poder entregar en primer lugar el tráfico

considerado crítico. El grupo de trabajo ISSLL (Integrated Services over Specific Link Layer) del IETF (Internet Engineering Task Force), se encarga de definir cómo relacionar los distintos protocolos de QoS de capas superiores con las diferentes tecnologías de la capa 2, como Ethernet. Entre otros aspectos, el ISSLL ha desarrollado el protocolo Subnet Bandwidth Manager (SBM), o "Gestión del ancho de Banda de la Subred", para aplicarlo con LAN 802. SBM, es un protocolo de señalización que permite la comunicación y coordinación entre nodos de la red y su relación con protocolos de QoS de capas superiores, siendo este concepto fundamental para el desarrollo de esta tesis.

Un requisito fundamental en SBM es que todo el tráfico debe pasar por lo menos por un switch que utilice SBM.

COMPONENTES DE SBM

Los principales componentes de SBM son:

- **Bandwdith Allocator (BA,** Distribuidor de Ancho de Banda**)**: Gestiona la asignación de los recursos y realiza el control de admisión de acuerdo a su disponibilidad y al resto de criterios definidos en las políticas del servicio.
- **Requestor Module (RM,** Modulo del Cliente**)**: Se ubica en cada estación final y no en algún switch. La relación entre el RM y los parámetros de protocolos de QoS superiores son definidas de acuerdo a una política determinada.

En la figura 4.10, es posible observar cómo la localización del BA determina el tipo de configuración de SBM en uso: centralizado o distribuido. Además, cuando existe más de un BA por segmento de red, uno de ellos será elegido como DSBM (Designated SBM).

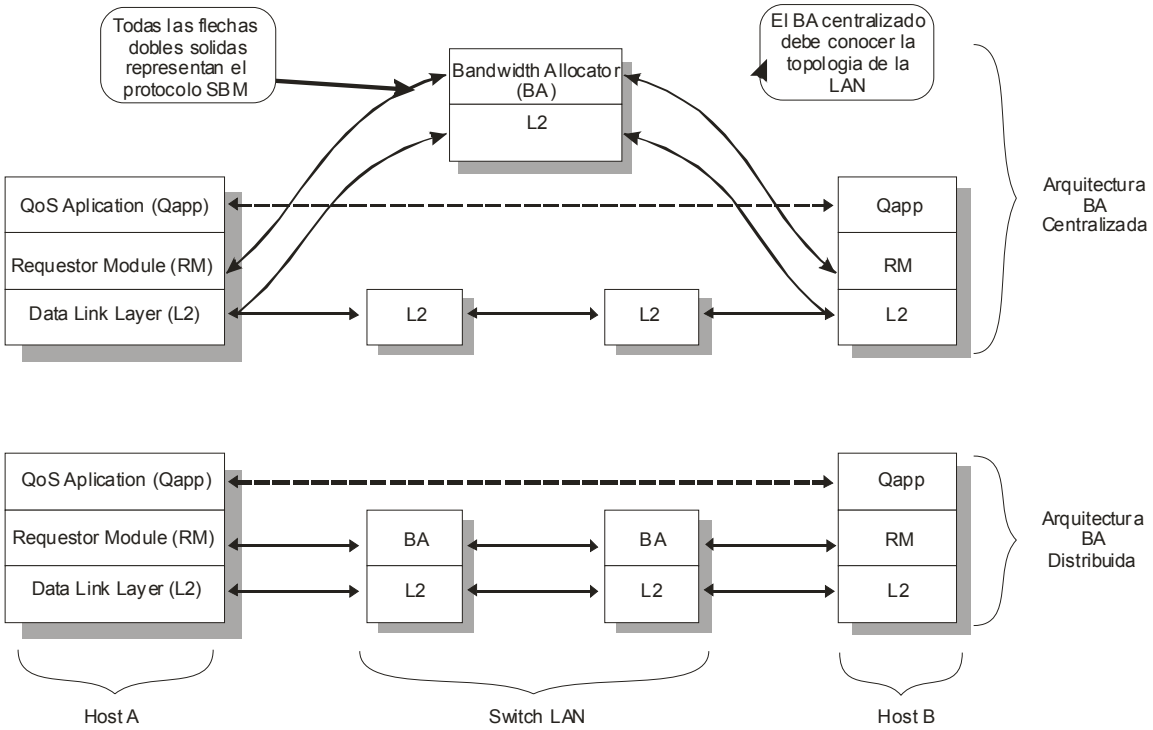


Figura 4.10 Configuración de BA (Centralizado o Distribuido).

FUNCIONAMIENTO

Este protocolo, utiliza un mecanismo de señalización entre RM (Requestor Module) y BA (Bandwith Allocator) para iniciar las reservas, consultar al BA los recursos disponibles y cambiar las reservas. Este mecanismo suele ser RSVP. Para comprobarlo, se tomará como ejemplo cómo se realiza (en forma genérica) el procedimiento del control de admisión en SBM.

1. El DSBM inicializa: consigue la disponibilidad de los recursos.
2. El DSBM (cualquier host o router RSVP) busca el DSBM en el segmento agregado por cada interfase. (Esta tarea está monitorizada con el campo “AllSBMAddress”).
3. El cliente envía un mensaje PATH con el campo “DSBMLogicalAddress”.
4. Una vez recibido el mensaje PATH, el DSBM indica su estado en el switch, almacenando la dirección de origen de capa 2 y capa 3 y la

pone en el mensaje, encaminándolo al siguiente salto (el cual puede ser otro DSBM en el siguiente segmento de red).

5. Cuando se envía un mensaje RESV de RSVP, un host lo envía al primer salto, como lo es el DSBM en este caso.
6. DSBM evalúa los requisitos y si los suficientes recursos están disponibles, lo reenvía al siguiente salto (de lo contrario retorna un error).

Como se ha visto, es un proceso muy parecido al ocurrido en los routers RSVP. Por otro lado, cualquier DSBM puede añadir un objeto denominado TCLASS a los mensajes RESV o PATH del protocolo RSVP. Este objeto contiene información de prioridad basada en la norma 802.1p.

4.6. ARQUITECTURAS DE QoS

Excepto para el caso de SBM en que se utiliza RSVP para la señalización, para el resto de protocolos, se mostró cómo actuaban de manera independiente de extremo a extremo, sin embargo, en la realidad se utilizan varios de estos protocolos para la obtención de QoS extremo a extremo y en los nodos finales, conformándose varias arquitecturas de QoS, de las cuales, se van a describir las más utilizadas.

En la figura 4.11, se muestra una visión general de cómo es posible mezclar los diferentes protocolos.

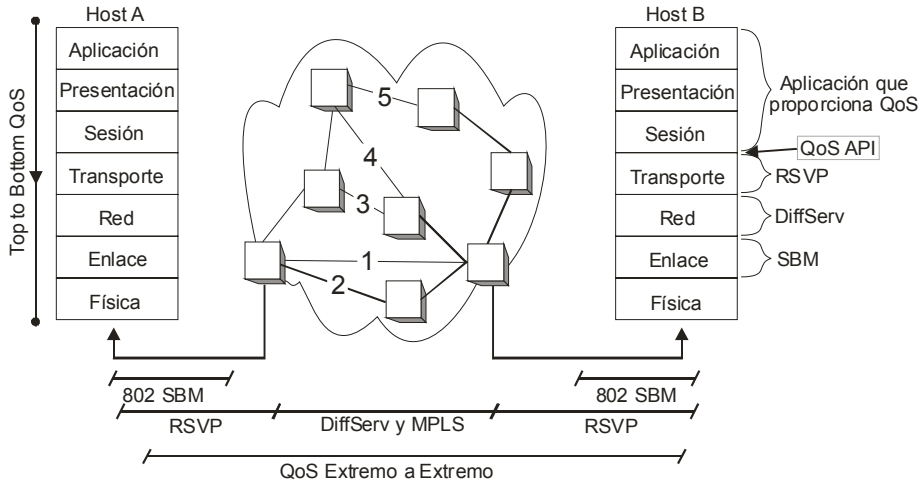


Figura 4.11 Utilización de diferentes protocolos para asegurar QoS extremo a extremo.

RSVP Y DIFFSERV EXTREMO A EXTREMO

RSVP, proporciona recursos para el tráfico de la red, mientras que DiffServ marca y prioriza el tráfico. RSVP es más complejo y demanda más actividad a los routers que DiffServ, esto puede impactar negativamente en los routers del backbone, por eso, normalmente se utiliza DiffServ en el backbone.

DiffServ, es un perfecto complemento para RSVP para habilitar QoS extremo a extremo. Los host finales pueden utilizar peticiones RSVP con alta definición (ancho de banda, umbral de jitter, etc.). Los routers frontera, situados en los puntos de ingreso del backbone pueden asociar esas reservas RSVP a una determinada clase de servcio, indicada por un byte DS y acordada en los niveles de servicio (SLAs).

En la arquitectura representada en la figura 4.12, se muestra a RSVP en los extremos de la red, y DiffServ en el núcleo, con lo que adquiere velocidad y apoyo.

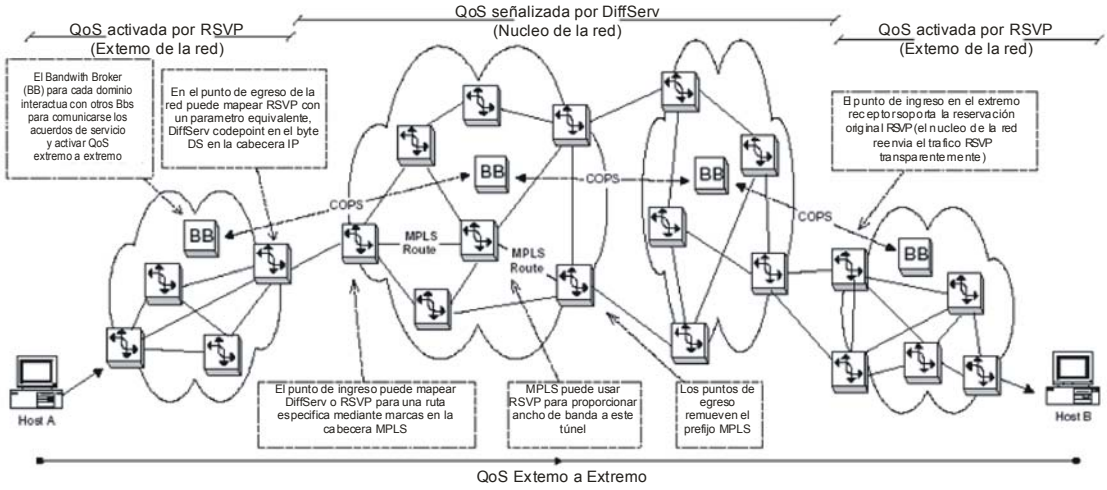


Figura 4.12 Uso de diferentes tecnologías de QoS trabajando conjuntamente para proporcionar QoS extremo a extremo.

RSVP PARA LOS AGREGADOS

Al hablar de clasificación de flujos de tráfico, DiffServ y MPLS se crean eficientemente tuberías para esos agregados. Para que esas tuberías puedan proporcionar cualquier calidad de servicio mejor que el estándar “best effort”, el tráfico en esos tubos virtuales no debe exceder su capacidad. El problema es que ni DiffServ ni MPLS tienen las mecánicas protocolares para detectar cuánto ancho de banda requieren esos flujos, y asignar entonces, los recursos necesarios para el uso especializado. Sólo RSVP está diseñado para hacer esto.

Aunque RSVP fue originalmente diseñado para asignar el ancho de banda de flujos de aplicaciones individuales, también es muy importante para asignar el ancho de banda a las necesidades de flujos agregados. Sin embargo, para los ingenieros de red que utilizan DiffServ o MPLS tratando de conocer las demandas de ancho de banda por anticipado, para hacer una reserva apropiada a la demanda de recursos. Adicionalmente, el emisor y el receptor en ambos extremos

de los tubos virtuales deben hacer apropiadamente esta reserva de recursos para que los mensajes PATH y RESV puedan ser enviados en situaciones factibles.

MPLS PARA RSVP

Existe una propuesta del IETF de usar un objeto en RSVP, denominado, EXPLICIT_ROUTE, para determinar caminos que puedan ser utilizados por flujos de RSVP. Estos flujos, usan tuberías virtuales establecidas a través de routers MPLS. Incluso sin el citado objeto, es posible para MPLS asignar etiquetas de acuerdo al campo *flowspecs* de RSVP.

MPLS PARA DIFFSERV

Al ser DiffServ y MPLS similares, asociar el tráfico DiffServ sobre tuberías MPLS (LSPs), es bastante sencillo.

Para soportar el modelo de DiffServ, un operador de red MPLS necesita asignar una serie de recursos para cada clase de DiffServ transmitida en cada router MPLS y asignar, a su vez, etiquetas.

RESUMEN

Al implementar los mecanismos y políticas para QoS, ya estamos hablando de la implementación de los protocolos y arquitecturas de QoS, lo que nos permite la implementación de una solución integra manejando Calidad de Servicio, ya sea extremo a extremo para una aplicación especifica o en un tramo de la red, para no propiciar la sobre utilización de algún enlace (Ingeniería de tráfico).

En este capítulo se muestran los protocolos y arquitecturas más conocidas de Calidad de Servicio, como es MPLS para creación de VPN, que en la actualidad ha tomado bastante auge para interconectar empresas mediante una VPN sobre Internet.

5. CALIDAD DE SERVICIO EN ATM Y FRAME RELAY.

INTRODUCCIÓN

Durante los últimos años, se han desarrollado numerosos tipos de redes: Redes de Área Local, Redes de Área Extensa, Redes de Área Metropolitana, con tecnologías de transferencia diversa, sea TDM síncrona (como RDSI-BE) o asíncrona (como X.25, Frame Relay o las Redes de Área Local). Unas están diseñadas específicamente para la transferencia de tráfico isócrono y otras son inadecuadas para esta función.

En la actualidad, la tecnología digital hace posible la distribución en red de aplicaciones como multimedia, videoconferencia, etc., que requieren una integración de los servicios de datos, audio e imagen estática y animada. Existen numerosos servicios que demandan esta integración y un elevado uso del ancho de banda, como pueden señalarse, entre otros, telemedicina, conferencias y correo multimedia, enseñanza, servidores de video, además de los nuevos servicios “en línea” como telemarketing, museos y bibliotecas virtuales, etc., que se están ofertando a través de Internet.

La proliferación de redes y servicios hace necesario plantearse un sistema integrado, aplicable a todos, que evite la problemática derivada de la diversificación actual y permita aplicar una economía de escala que proporcione precios accesibles.

Para ello, se requiere un sistema multipropósito que debe funcionar con todo tipo de servicios, tráfico y demanda, opere sobre todas las distancias y alcance grandes velocidades, hasta Gbps (Giga bits por segundo).

5.1. MODO DE TRANSFERENCIA ASINCRONA (ATM).

ATM surge cuando las compañías telefónicas decidieron construir redes para él siglo XXI, donde él trafico de las voces es suave y necesita un ancho de banda bajo pero constante, mientras que el trafico de datos es explosivo, no necesita por lo general un ancho de banda (cuando no hay trafico), pero a veces necesita gran cantidad de recursos durante periodos muy breves.

Después de un análisis exhaustivo, surge una forma híbrida con bloque de tamaño fijo sobre circuitos virtuales, con un acuerdo que proporcionaba un rendimiento razonable para ambos tipos de tráfico. A este esquema se le llamó **Modo de Transferencia Asíncrona (ATM {Asynchronous Transfer Mode})**.

El modelo ATM consiste en que un emisor primero establece la conexión (es decir, un circuito virtual) con el o los receptores. Durante el establecimiento de la conexión se determina una ruta desde el emisor hasta los receptores y se guarda la información del ruteo en los switches a lo largo del camino. Mediante esta conexión se pueden enviar los paquetes, pero el hardware separa a estos en pequeñas partes de tamaño fijo llamadas celdas. Todas las celdas de un circuito virtual siguen la misma ruta guardada en los switches, cuando ya no se necesita la conexión, esta se libera y se purga la información de ruteo en los switches.

Las ventajas que tiene este esquema sobre la comunicación en paquetes y circuitos tradicional, es que puede utilizar una red para transportar un arreglo de voz, datos, televisión, videocintas, radio, etc., debido a que en todos los casos, lo que ve la red son sólo celdas y no le interesa lo que haya en ellas.

Las características más significativas de las redes ATM son: su capacidad de integración de diversos tipos de tráfico; la asignación dinámica y flexible del ancho de banda; la optimización del compromiso entre caudal y latencia; y la

ganancia estadística, es decir, su capacidad de optimizar la relación entre la suma de velocidades pico de las fuentes, y la velocidad del enlace. Por estas razones la tecnología ATM, es recomendada en la actualidad como solución universal para redes de banda ancha por los más importantes organismos de las industrias de comunicaciones y computadoras, como la UIT (Unión Internacional de Telecomunicaciones), el ATM Forum o el IETF (Internet Engineering Task Force).

Las características básicas de ATM son, en esencia, muy simples:

- Operación por conmutación de paquetes, si bien se utilizan paquetes de longitud fija (48 octetos de información y 5 octetos de control), denominados celdas. Esta opción de celdas de tamaño fijo permite el uso de nodos de conmutación a velocidades muy altas.
- Orientado a conexión al nivel más bajo. La información se transfiere por canales virtuales asignados durante la duración de la conexión.
- La asignación del ancho de banda se realiza en función de la demanda de envío de tráfico.
- No se realiza control de errores en el campo de datos, ni control de flujo en la red ATM. Con ello se maximiza la eficiencia.
- Proporciona transparencia temporal, es decir, pequeñas variaciones de retardo entre las señales de la fuente y el destino. Por ello permite la transferencia de señales isócronas.
- Las celdas se transmiten a intervalos regulares; si no hay información se transmiten celdas no asignadas.
- Se garantiza que las celdas llegan a su destino en el mismo orden en el que fueron transmitidas.

Como se ha expuesto, las celdas constan de un campo de información de 48 octetos y una cabecera de 5 octetos, la cual, contiene un conjunto de informaciones de control, como identificadores, que se utilizan para identificación de las conexiones y enrutamiento entre otros fines.

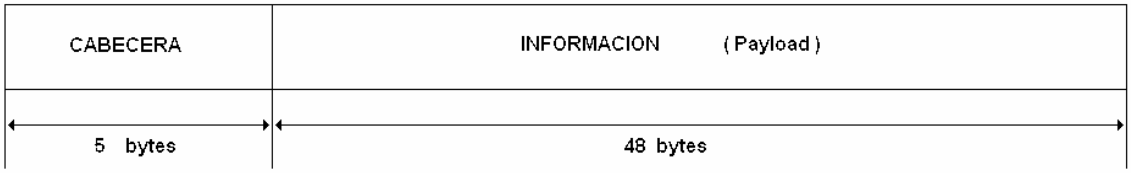


Figura 5.1 Celda ATM

El tamaño de la celda de 48 octetos se deriva de un compromiso entre una serie de características deseables para cada tipo de tráfico. Por una parte, por razones de eficiencia de transmisión es conveniente que las celdas sean de tamaño razonablemente grande. Desde el punto de vista de la transmisión de datos, también es aconsejable que las celdas sean grandes para evitar una excesiva segmentación. Sin embargo, para las aplicaciones sensibles al retardo o a la variación del mismo, es aconsejable que las celdas sean de la menor longitud posible. Con las anteriores, se realizaron varias propuestas, desde 32 octetos, adecuada para transmisiones telefónicas, hasta 64 octetos como tamaño mínimo razonable para transferencia de datos. Es obvio que 48 octetos es un claro compromiso derivado de la media aritmética de las anteriores celdas.

ATM, es considerado un modo de transferencia orientado a conexión, basado en la multiplexación asíncrona por división en el tiempo y el uso de celdas de longitud fija. Cada celda contiene un campo de información y un encabezado. Este, es usado principalmente para identificar celdas pertenecientes al mismo canal virtual dentro de la multiplexación asíncrona por división en el tiempo, y para realizar el direccionamiento apropiado. La integración en la secuencia de las celdas es conservada para cada canal virtual.

El campo de información de las celdas ATM es llevado en forma transparente a través de la red. No se realiza ningún procesamiento, tal como control de errores, sobre este campo dentro de la red. Todos los servicios (voz,

video, datos…) pueden ser transportados vía ATM, incluyendo los servicios no orientados a conexión.

Direccionamiento

Cada celda ATM enviada a través de la red contiene una información de direccionamiento, la cual le permite establecer una conexión virtual desde el origen hasta el destino de esta manera todas las celdas pertenecientes a esta conexión son enviadas en secuencia a través de esta conexión virtual. ATM provee conexiones virtuales tanto permanentes (Permanent Virtual Conexions, PVCs) como conmutadas (Switched Virtual Conexions, SVCs).

Recursos

Como ATM es orientado a conexión, las conexiones son establecidas, ya sea, en forma permanente o por la duración total de la conexión, en el caso de servicios conmutados. Este establecimiento incluye la asignación de un Identificador de Circuito Virtual (VCI Virtual Circuit Identifier) y/o un Identificador de Trayectoria Virtual (VPI Virtual Path Identifier), pero también, la asignación de los recursos empleados en el acceso de usuario y dentro de la red. Estos recursos determinarán el throughput (régimen binario) y Calidad de Servicio (QoS). Pueden ser negociados entre el usuario y la red a través del interfaz de usuario (UNI), durante la fase de inicio de la llamada y posiblemente durante la misma.

Identificadores de Celdas ATM.

Los identificadores de las celdas ATM, tales como el de Trayectoria Virtual (VPI), Circuito Virtual (VCI) y también el del tipo de Carga Útil, identifican una celda ATM en un medio físico de transmisión. El reconocimiento de la celda es la base para todas las demás operaciones. Tanto el VCI como el VPI son únicos

para celdas pertenecientes a una misma conexión virtual en un medio de transmisión compartido.

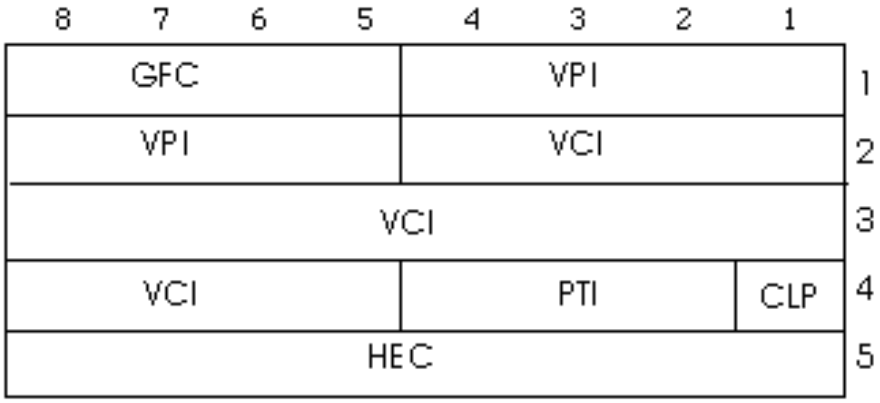


Figura 5.2 Estructura del encabezado ATM en la UNI (User Network Interface).

Dentro de un circuito virtual particular, las celdas pueden ser distinguidas además por su PTI, que depende del tipo de carga útil transportada por la celda. Este campo indica si la celda transporta información del usuario que deba ser entregada en forma transparente a través de la red, o información especial de la red. En caso de que el campo indique información del usuario, una parte del campo de información indica el tipo de control de la red en el cuál la parte restante del campo de información puede ser procesada dentro de la red.

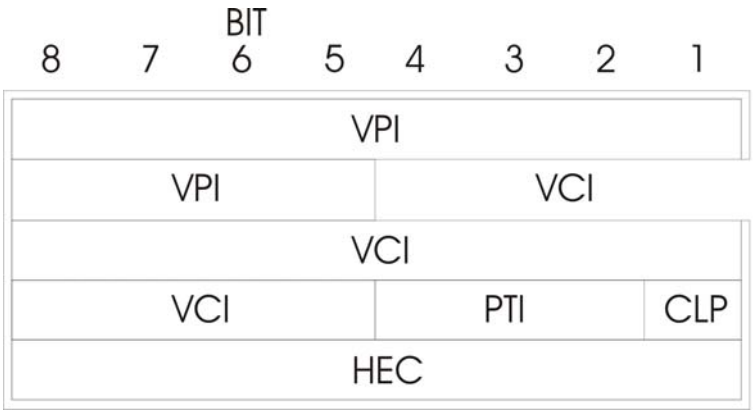


Figura 5.3 Estructura del encabezado ATM en el NNI (Network Node Interface).

Los campos de las celdas ATM son los siguientes:

- Campo GFC (Generic Flow Control [en la UNI]): consta de 4 bits
- Campo VPI / VCI (Virtual Path Identifier / Virtual Channel Identifier): tiene 24 bits en la UNI (8 para la VPI y 16 para la VCI y 28 bits en la NNI (12 para la VPI y 16 para la VCI). Los 4 bits de diferencia se deben al campo GFC de la UNI.
- Campo de tipo de carga útil (PTI, Payload Type Identifier). Está constituido por 3 bits. Indica el contenido de carga útil (datos de usuario, información de gestión, información OAM), así como la situación de congestión en algún punto de la red.
- Campo de Prioridad de Pérdidas de Celdas (CLP, Cell Loss Priority). Tiene un bit de longitud. Las celdas con este bit a 1 son las primeras en ser descartadas en caso de congestión.
- Campo de Control de Error de Cabecera (HEC, Header Error Control). Consta de 8 bits. Es procesado por el nivel físico para detectar errores en la cabecera. El código utilizado permite la corrección de errores simples o detección de errores múltiples.

PRINCIPIO DE OPERACIÓN

Al ser ATM una técnica orientada a conexión, tiene que establecerse una conexión virtual entre usuarios finales antes de que se comience a transmitir la información. Las conexiones pueden establecerse mediante procedimientos de señalización o pueden ser permanentes o semipermanentes.

En la figura 5.4, se representa un esquema simplificado de red con switches ATM.

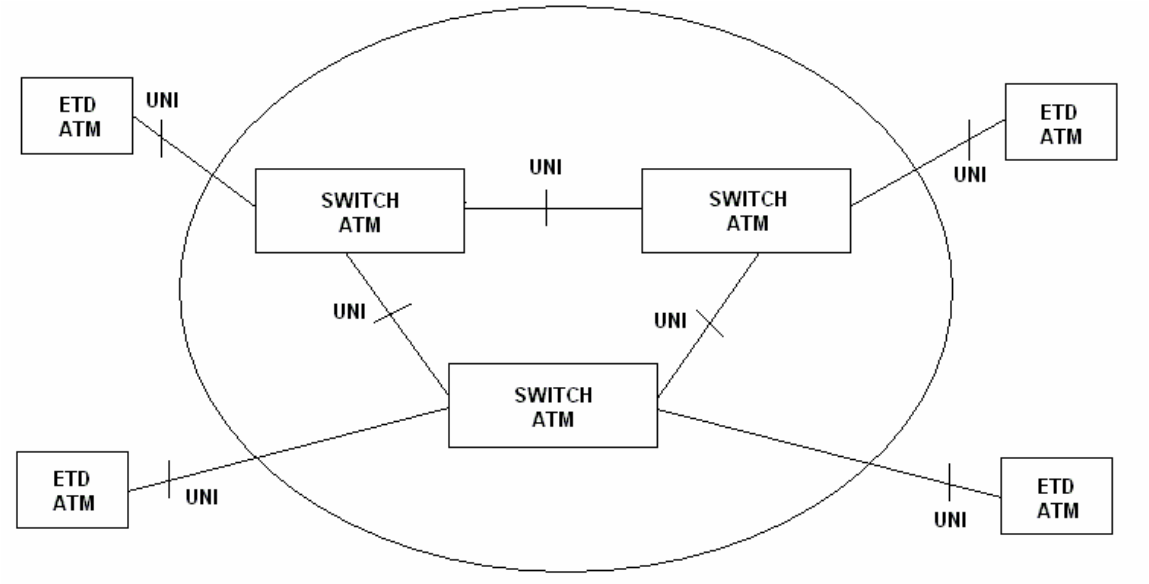


Figura 5.4 Esquema red ATM.

Una red ATM consiste de la interconexión de una serie de switches ATM por medio de enlaces punto a punto o interfaces ATM. Los switches ATM, permiten dos clases de interfaces:

- UNI (User-Network Interface)
- NNI (Network-Node Interface, también conocida como Network-Network Interface)

La UNI proporciona la conexión a la red ATM desde un equipo terminal ATM o bien, desde un sistema intermedio, tal como hub, switch, router, que a su vez controla equipos de usuario final.

La NNI define la interfaz entre los nodos ATM; cuando la NNI conecta nodos pertenecientes a distintas redes se denomina NNI-ICI (NNI-Inter Carrier Interface).

Circuitos Virtuales (VC Virtual Circuits)

ATM utiliza circuitos virtuales para crear enlaces individuales de la red para transportar celdas de longitud fija de 53 bytes entre los nodos de la red. Las celdas de diferentes nodos pueden compartir un circuito virtual cuando viajan hacia un mismo destino. Estos circuitos virtuales llevan todas las transmisiones de datos entre los nodos y mantienen la secuencia correcta de celdas y calidad de servicio a lo largo de toda la transmisión. Un circuito virtual puede atravesar más de un switch ATM.

Existen dos tipos de circuitos virtuales, Circuitos Virtuales Permanentes (Permanent Virtual Circuit, PVCs) y Circuitos Virtuales Conmutados (Switched Virtual Circuit, SVCs).

- Circuitos Virtuales Permanentes (PVC), son conexiones "permanentes" entre dos nodos de la red y operan como una línea física dedicada. En una implementación de PVC, la conectividad de red entre dos nodos es configurada estáticamente en los switches, y el Identificador de Circuito Virtual (VCI Virtual Circuit Identifier) para cada nodo remoto es configurado en cada estación extremo.

- Circuitos Virtuales Conmutados (SVC), son creados dinámicamente para cada transmisión. Son similares a la red telefónica de voz, en donde las conexiones entre dos puntos extremos de la red son creados dinámicamente para cada transmisión. Los SVC pretenden determinar la ruta disponible más corta desde la fuente hasta el destino.

Funcionalidad del encabezado.

Conexiones Virtuales

Una de las características básicas de ATM es la funcionalidad limitada en el encabezado, por lo que no se necesitan las funciones de dirección de origen y destino ni número de secuencia, que son necesarias en las redes no orientadas a conexión. Cada conexión virtual se va a identificar por un número (identificador), que tiene solamente un significado local para cada enlace, en la conexión virtual.

Por lo tanto la función básica que queda en el encabezado es la identificación de la conexión virtual. Esta función, es realizada por dos subcampos del encabezado: VCI (Virtual Channel Identifier) y el VPI (Virtual Path Identifier). El campo VCI ubica dinámicamente conexiones; el campo VPI identifica estáticamente las conexiones.

Canales Virtuales (VC Virtual Channels)

Esta función se realiza por medio del subcampo del encabezado llamado VCI y consiste en utilizar un identificador que se asigna en el establecimiento de la conexión. El VCI tiene solamente un significado local sobre el enlace entre los nodos ATM y se traducirá en cada nodo como se explica en la figura siguiente. Cuando se libera la conexión, los valores del VCI en los enlaces abarcados también serán liberados y podrán ser reutilizados por otras conexiones.

Trayectorias Virtuales (VP Virtual Paths)

Los Trayectorias Virtuales son conexiones semi-permanentes entre dos puntos extremo que tienen que transportar un número muy grande de conexiones simultaneas. Este concepto se conoce también como red virtual. Es aquí donde los recursos disponibles de la red son asignados en forma semi-permanente para

permitir que su manejo sea simple y eficiente (ver figura 5.5). Para establecer esta red virtual, se define otro campo en el encabezado, llamado VPI (Virtual Path Identifier).

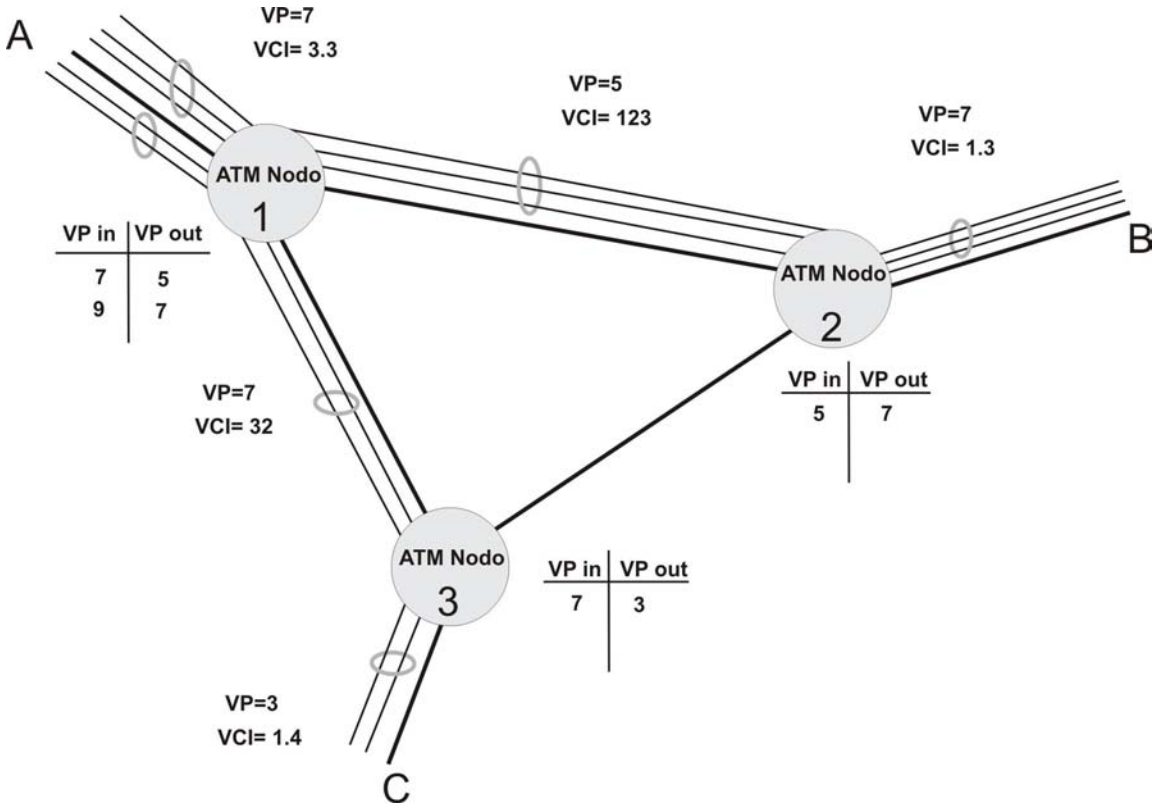


Figura 5.5 Uso de VPI dentro de una red ATM

En el ejemplo de la figura 5.5, se puede observar que se ha establecido una trayectoria virtual (VP) entre el suscriptor A y el suscriptor C, transportando dos conexiones individuales, cada una con un VCI separado. Nótese que los VCI utilizados (3 y 4 en el ejemplo) no son traducidos en los nodos, los cuales solo conmutan el campo VPI.

Adicionalmente, se ha establecido en forma semi-permanente una trayectoria virtual entre los nodos A y B, usando los valores 1,2 y 3 para el VCI. Además en el enlace entre A y el nodo 1, se ha usado dos veces el valor de 3 para el VCI. Esto no crea problemas, puesto que al utilizar dos valores diferentes de VPI permite que dos puntos extremo (A y el nodo 1) discriminen entre dos conexiones virtuales.

Figura 5.6 Un VP puede englobar a varios VC's.

Calidad de Servicio (QOS).

La calidad de servicio de una conexión, se refiere a la pérdida de celdas, el retardo y la variación del mismo en que incurren las celdas pertenecientes a esa conexión en una red ATM. Para ATM, la calidad de Servicio de una conexión está estrechamente ligada al ancho de banda que ésta usa.

Cabe destacar aquí el uso de la función policía, la cual está definida por un conjunto de acciones tomadas por la red para monitorear y controlar tráfico en una conexión ATM en términos del volumen de tráfico de celdas y validación del enrutamiento de las celdas.

El propósito fundamental de la función policía, es forzar a que cada conexión ATM cumpla con el contrato de su tráfico negociado. Todo esto se hace a través del UPC/UNC (Usage Parameter and Network Parameter Control) en la UNI y en la NNI.

Un algoritmo UPC/NPC, debería exhibir las siguientes funciones principales:

- Capacidad de detectar cualquier situación ilegal de tráfico.
- Tiempos de respuesta rápidos en caso de violaciones de parámetros
- Simplicidad de implementación.

Algunos servicios se pueden beneficiar de una indicación de Prioridad de Pérdida de Celdas (CLP, Cell Loss Priority) explicita en cada celda, llevada en un bit específico dentro del encabezado, como un medio de gestionar la pérdida de celdas durante periodos de congestionamiento de la red.

Esto permite al usuario el elegir entre dos regímenes de pérdida de celdas en una sola conexión virtual: prioridad alta para celdas que llevan información básica, y celdas con prioridad baja cuando estén sujetas a ser descartadas, dependiendo de las condiciones de la red. Sin embargo, si este indicador se usa, será necesario indicar durante la fase del inicio de la llamada la incidencia esperada de este indicador. Esto facilitará la asignación apropiada de los recursos de la red y la aplicación de los parámetros de control de la red.

Parámetros de Control de Utilización.

A diferencia del ambiente del Modo de Transferencia Síncrona, en ATM no hay limitaciones físicas en el régimen de acceso de usuario al medio de transmisión, aparte del régimen físico de celdas propio del medio. Por el contrario, el equipo de multiplexación hará lo posible por evitar la pérdida de celdas, para ofrecer el máximo throughput posible que el usuario elija para enviar.

Sin embargo, como las conexiones virtuales comparten recursos físicos, medio de transmisión y espacio en buffer, el uso excesivo de recursos por un usuario desequilibra el tráfico para otros usuarios. Por lo tanto, el throughput debe ser monitoreado (vigilado) en la interfaz de red del usuario por medio de la función

Parámetro de Control de Utilización de la red, para asegurar que el contrato negociado por cada VCI o VPI entre la red y el suscriptor es respetado por cada uno de estos. Los parámetros de tráfico podrían describir el throughput deseado y el QOS en el contrato en forma no ambigua.

En este contexto, es muy importante que los parámetros de tráfico que se han seleccionado para este propósito, puedan ser monitoreados en la recepción de cada celda.

MECANISMOS DE CONTROL

Control de admisión de conexión (CAC)

CAC representa una serie de acciones tomadas por la red durante la fase de establecimiento de un SVC, durante el establecimiento de un PVC o en la fase de renegociación de alguno de éstos, con la finalidad de:

1. aceptar o rechazar una nueva conexión VCC (PVC o SVC), o
2. cambiar la categoría de servicio o descriptores de tráfico de un VCC existente.

La aceptación de la solicitud de una conexión para una nueva llamada se realiza sólo cuando están disponibles suficientes recursos para llevar esta conexión a través de toda la red con la calidad de servicio (QoS) solicitada por la misma y cuando al mismo tiempo es posible mantener las QoS de las conexiones ya existentes en la red (esto se aplica también durante la renegociación de los parámetros dentro de una llamada).

Si no es posible que la conexión reciba estos recursos, la red ATM no aceptará dicha conexión. CAC debe tomar en cuenta cada recurso compartido de todos los componentes de la red, en donde un recurso compartido puede ser, por

ejemplo, la cola de celdas (cell queue) y su enlace de transmisión correspondiente. En otras palabras, CAC al rechazar conexiones o cambios en configuraciones que pueden producir congestión, asegura que la garantía de la pérdida de celdas (CLR) y el retardo de celdas (CTD) fijados para una conexión sean cumplidos.

Durante el establecimiento de la llamada una serie de informaciones, las cuales se encuentran en un contrato de tráfico, deben ser negociadas y acordadas entre el usuario y la red para permitir que el CAC tome las decisiones adecuadas en cuanto a la aceptación/rechazo de la conexión.

Para cada solicitud, (request) de un VCC, CAC emplea la siguiente información para poder tomar la decisión de admisión/rechazo:

- la QoS solicitada
- los valores de los parámetros en el descriptor de tráfico del VCC
- la definición de conformidad UPC solicitada
- las rutas
- el factor de reservación (booking), de existir o ser

La función CAC, es responsable de la asignación de recursos de red para una conexión dada. Según los estándares, estos esquemas son específicos del operador de la red.

Esta, asigna un ancho de banda a cada conexión y limita la asignación del ancho de banda total a la capacidad de cada recurso (por ejemplo a la velocidad del enlace).

La cantidad de recursos de red que una conexión requiere puede ser representada en términos de un Ancho de Banda Virtual V_{bw} (también denominado ancho de banda equivalente) mediante el empleo de los descriptores

de tráfico específicos de la conexión. El ancho de banda virtual, representa la cantidad de ancho de banda empleada por la conexión una vez incluidos los requerimientos de la QoS (principalmente CLR) y el tamaño del buffer.

La función CAC más simple, asigna un ancho de banda virtual equivalente a la PCR para cada conexión. En el caso VBR, puede existir una asignación estadística del ancho de banda en el caso en que el parámetro SCR sea tomado en consideración.

El cálculo del ancho de banda virtual es muy complejo y requiere de grandes suposiciones del comportamiento del tráfico. Continuas investigaciones son realizadas en este aspecto con la finalidad de encontrar el modelo con la mayor precisión que refleje los requerimientos de ancho de banda de cada tipo de comportamiento de tráfico. En la Tabla 5.7, se muestra una estrategia comúnmente empleada para la asignación del ancho de banda para cada una de las categorías de servicio.

Categoría de servicio	**Ancho de banda asignado**
Constant Bit Rate (CBR)	PCR ≤ V_{bw} ≤ Link Rate
Variable Bit Rate (VBR)	SCR ≤ V_{bw} ≤ PCR
Available Bit Rate (ABR)	MCR ≤ V_{bw} ≤ PCR
Unspecified Bit Rate (UBR)	No se le asigna V_{bw}

Tabla 5.7 Asignación del ancho de banda según la categoría de servicio.

Conformado del tráfico (Traffic shaping)

Es un mecanismo que altera las características de tráfico del flujo de celdas de una conexión para alcanzar una mejor eficiencia en la red mientras se mantienen los objetivos QoS o con la finalidad de asegurar que el flujo de celdas sea conforme con los parámetros de tráfico de acuerdo con la configuración del algoritmo leaky bucket del contrato de tráfico. El traffic shaping puede ser

empleado, por ejemplo, para reducir la velocidad pico, limitar la longitud de la ráfaga o reducir la CDV por medio del espaciamiento adecuado de las celdas en el tiempo. El uso y ubicación de esta función es específica de la red.

Control de prioridad de pérdida de celda

Cuando las celdas son conmutadas a través de una red ATM, se van formando colas como una consecuencia natural de los retardos de propagación (demasiadas celdas a ser transmitidas por un enlace de salida) y de los retardos de procesamiento en los nodos de la red. Las celdas en las colas, deben ser colocadas en buffers hasta que puedan ser atendidas. Las redes ATM, bajo cualquier tipo de condición, deben poseer mecanismos adecuados para el servicio de los buffers en los nodos ATM. Bajo condiciones de congestión (es decir, demasiadas celdas en la red), debe existir un mecanismo de prioridad que permita remediar la situación de congestión, como por ejemplo el descarte de ciertas celdas, con la finalidad de poder servir al resto de las mismas con los adecuados parámetros QoS. Para esto, se requiere de un método que permita a los nodos ATM identificar rápidamente celdas que puedan ser descartadas de aquellas que no pueden ser descartadas, a excepción de condiciones extremas de congestión. El control de prioridad, es realizado a través de un bit localizado en el encabezado de la celda denominado bit de prioridad de la celda (CLP)

5.2. RETRANSMISIÓN DE TRAMAS (FRAME RELAY).

Las primeras redes de conmutación de paquetes de los años sesentas/setentas utilizaban las infraestructuras de las redes analógicas. Se trataba de medios de transmisión de baja calidad con una alta tasa de errores. De hecho, justificaba los abundantes controles para la detección de errores de X.25, sus reiterados mecanismos de control de flujo o el pequeño tamaño de los paquetes, más pensados para facilitar las retransmisiones para lograr la máxima eficacia. El resultado, es una comunicación segura entre usuarios, pero lenta e ineficaz debido a la carga de procesamiento que la red debe soportar.

Sin embargo, hoy en día, el entorno donde se diseñan las nuevas redes de comunicaciones es muy diferente: se dispone de nuevas infraestructuras de alta calidad que reducen la probabilidad de error y los usuarios utilizan dispositivos terminales más modernos, con gran capacidad para analizar y manipular flujos de información. Estas dos circunstancias, han promovido una gran estrategia generalizada para vincular los dispositivos de los usuarios con las redes de comunicaciones haciéndolos coparticipes en el proceso de transferencia de información. Esta estrategia, asumida por las redes Frame Relay y ATM, consiste fundamentalmente en delegar el control de flujo y el control de errores a las terminales, mientras que la red es únicamente responsable de la transmisión y conmutación de datos. Si ocurre un error o se saturan los nodos de la red han de ser las terminales de los usuarios las que gestionen estas situaciones reenviando las tramas erróneas o bien reduciendo la velocidad de transmisión para evitar la congestión, mientras que la red se limitará a dar simples indicaciones del estado de sus recursos. En cualquier caso, la red realiza su mejor esfuerzo para entregar las tramas sin errores y controlar la congestión.

Frame Relay originalmente fue estandarizado por la UIT-T para optimizar el uso de los canales de RDSI en Banda Estrecha. Sin embargo, el posterior

desarrollo de los acontecimientos la han convertido en una tecnología de red independiente de RDSI y paradójicamente, muy pocas redes de RDSI han llegado a implementarla. En la actualidad Frame Relay, es también un estándar ANSI.

CONCEPTOS BÁSICOS DE FRAME RELAY

La técnica tradicional de conmutación de paquetes tiene como características básicas:

- Señalización dentro de la banda. Los paquetes de control de llamada, utilizados para establecer y terminar los circuitos virtuales, se transmiten por el mismo canal y el circuito virtual que los paquetes de datos.
- Multiplexacion de los circuitos virtuales a nivel de red.
- Control de flujo y control de errores tanto a nivel 2 como a nivel 3 de OSI.

Estas características suponen una gran carga para el sistema. En el inciso “a” de la figura 5.8, se muestra el flujo de tramas necesarias para la transmisión de un único paquete de datos, con su correspondiente paquete de reconocimiento desde el sistema final-origen, hasta el sistema final destino. Para cada salto en la red, el protocolo de control de enlace de datos necesita el intercambio de una trama de datos y una, de acuse de recibo.

Para cada nodo intermedio es necesario mantener tablas de estado por cada circuito virtual que administren la gestión de llamadas y los aspectos de control de errores y de flujo del protocolo X.25. Para simplificar se supone que el tamaño de la ventana es 1.

Toda esta carga puede estar justificada si existe una posibilidad importante de aparición de errores en cualquiera de los enlaces de la red, pero este no es el caso de la mayoría de las redes actuales.

Frame Relay está diseñada para eliminar en lo posible, todos aquellos procesos no necesarios hoy en día de X.25 y que generan una importante carga en el sistema. Los puntos principales en los que Frame Relay se diferencia de un servicio de conmutación de paquetes convencional es X.25 son:

- Control de llamadas fuera de banda. La señalización del control de llamada se realiza en una conexión lógica separada de la conexión para la transmisión de los datos de usuario.
- La multiplexacion y conmutación de conexiones lógicas, tiene lugar a nivel 2 en vez de nivel 3, eliminando de esta manera un nivel entero de procesamiento.
- La red deja de preocuparse del control de errores y del control de flujo. Estos, si se emplean, pasan a ser responsabilidad del nivel superior y se realizan extremo a extremo.

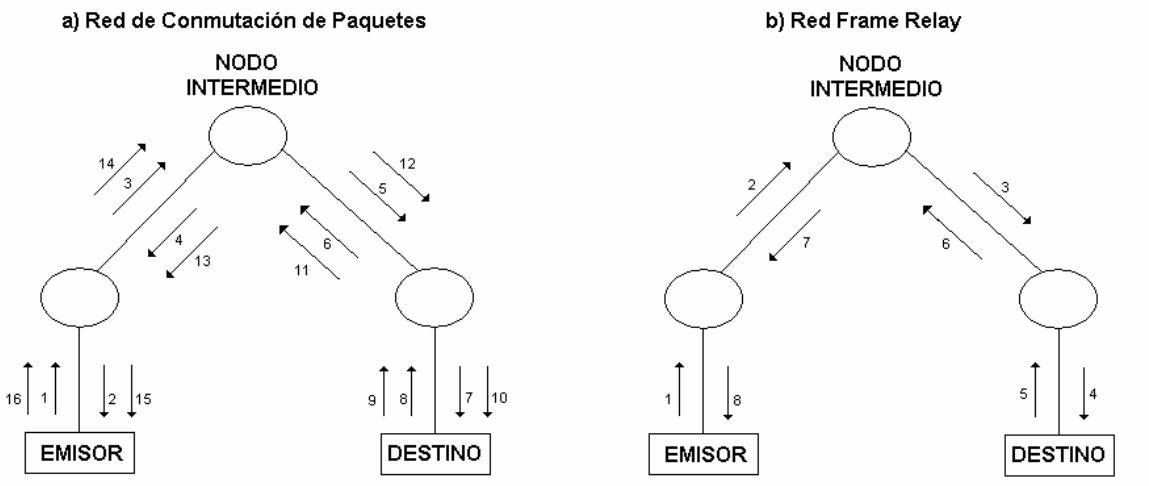


Figura 5.8 Comparación de transmisiones X.25 y Frame Relay.

En la figura 5.8 en el inciso b, se muestra la operación de Frame Relay, en la que se manda una única trama de datos del origen al destino y se genera un acuse de recibo en el nivel superior, transmitido de regreso en otra trama.

La conmutación de tramas o frame switching opera también a nivel 2; sin embargo, realiza las funciones de control de errores y control de flujo de este nivel.

A continuación, se analizan las ventajas e inconvenientes del uso de Frame Relay frente a X.25:

Inconvenientes:

- El inconveniente principal de Frame Relay frente a X.25 es que se pierde la capacidad de realizar el control de flujo y el control de errores en cada uno de los enlaces de la red, pero esta funcionalidad puede ser proporcionada, extremo a extremo, por el nivel superior.
- Es necesaria la disponibilidad de líneas de alta calidad.
- No existe un estándar para la interconexión de servicios Frame Relay, como X.75 para las redes X.25.

Ventajas:

- La mayor ventaja de Frame Relay es que hace más eficiente el proceso de comunicación. La funcionalidad del protocolo requerida en la interfaz usuario-red se reduce, así como el procesamiento interno de la red. Esto conlleva un menor retardo y un mayor rendimiento. El tiempo de proceso de la trama es del orden de la décima parte que en X.25.
- La velocidad de acceso puede alcanzar normalmente los 2 Mbps, frente a los 64 Kbps de X.25.
- La interfaz de usuario es sencilla y conlleva una relativamente simple migración de X.25.

FORMATO DE LA TRAMA DE FRAME RELAY

El formato de la trama, mostrado en la figura 5.9, es similar al de los otros protocolos de nivel 2 como LAP-D y LAP-B, con una diferencia fundamental, es este caso no hay campo de control.

Flag	Address	Information	FCS	Flag
1	2 - 4	Variable	2	1
octeto	octetos	octetos	octetos	octeto

Figura 5.9 Trama Frame Relay.

Esto supone que:

- Sólo existe un tipo de trama, utilizada para transmitir información de usuario.
- No se puede utilizar señalización dentro de banda; una conexión lógica sólo puede transmitir datos de usuario.
- Tampoco existen tramas que permitan a la red ejecutar control de flujo, enviar ACK's o pedir retranmisiones, ya que no hay un numero de secuencia.

Todas estas funciones deben ser implementadas en los equipos terminales tales como routers, bridges, o controladores de comunicaciones, que deberán dispone de los mecanismos necesarios para el secuenciamiento, el control de flujo, el envío de reconocimientos y la recuperación de errores, que permitan garantizar la integridad de los datos transmitidos.

- La red detecta pero no recupera errores; los nodos de la red tienen capacidad de detectar errores y en determinados casos de eliminar tramas, pero nunca recuperarlos.

A continuación, serán descritos uno a uno los campos que componen la trama:

- *Delimitador* (Flag): Este campo funcional igual que en los protocolos LAP-D y LAP-B. Todas las tramas comienzan y terminan con la secuencia de bits 01111110. Para garantizar la transparencia de la información, el nivel de enlace que va a transmitir la trama Frame Relay debe encargarse de comprobar el contenido de la trama entre el delimitador de apertura y cierre e insertar un bit 0 cada vez que aparezca una secuencia de cinco bits 1 consecutivos. Por su parte el nivel de enlace de la entidad receptora se encargará de eliminar dichos bits una vez que contenga los datos de la trama comprendidos entre ambos delimitadores.
- *Dirección:* El campo de dirección está formado por defecto de dos octetos, pero puede extenderse hasta tres o cuatro. Los posibles formatos de este campo se muestran a continuación en la figura 5.10.

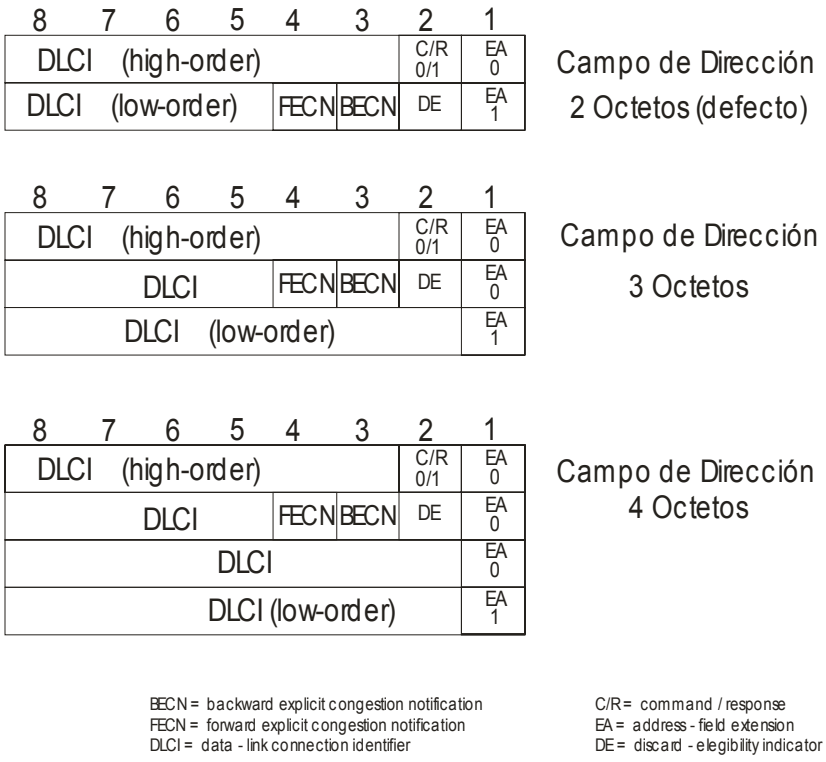


Figura 5.10 Formatos de trama FR.

La longitud del campo de dirección, y por lo tanto del DLCI, está definida por el campo EA (Extended Address), que indica si el campo de dirección continúa en el siguiente octeto (1) o ha terminado (0). El campo C/R es de uso específico en cada aplicación y el protocolo estándar de Frame Relay no lo utiliza. El resto de los bits de este campo están relacionados con el control de congestión, mismo que se tratará mas adelante.

- *Información:* El campo de información transmite datos de nivel superior. Si el usuario elige implementar funciones adicionales de control de nivel de enlace extremo a extremo, entonces en este campo se encuentra una trama de enlace de datos.
- *FCS (Frame*-Check Sequence): Es una secuencia de 16 bits que permite verificar la correcta transmisión de la trama y la futura recuperación de posibles errores en la misma.

FUNCIONAMIENTO DE LA RED

La función de retransmisión de tramas realizadas por Frame Relay consiste en el encaminamiento de las tramas, antes descritas, de acuerdo a los valores de sus DLCI.

Por lo general, el encaminamiento es controlado mediante las entradas de una tabla de conexión que utiliza el DLCI. El manejador conmuta las tramas de un canal de entrada y otro de salida mediante la apropiada entrada de la tabla de conexión y traduce el DLCI de las tramas antes de la transmisión.

Como parte de la función de retransmisión de tramas, se verifica el campo FCS de cada trama. Si se detecta un error, la trama simplemente se descarta, siendo responsabilidad de los usuarios finales la recuperación de este error.

CONTROL DE CONGESTION

Conceptos básicos de control de congestión

Una red Frame Relay es una red de conmutación de paquetes en la que los "paquetes" son tramas de nivel 2. Como en cualquier red de conmutación de paquetes, una de las áreas clave en el diseño de una red Frame Relay es el control de congestión. Para entender algunos términos relacionados con el control de congestión, debemos acudir a algunos resultados de la teoría de colas. Básicamente, una red Frame Relay, es una red de colas. En cada manejador, hay una cola de tramas por cada enlace de salida. Si la velocidad de llegada de las tramas excede la velocidad de transmisión de las mismas, el tamaño de la cola crece sin límite y el retraso sufrido por una trama tiende a infinito. Incluso si la velocidad de llegada de las tramas es menor que la velocidad de transmisión, la longitud de la cola crecerá rápidamente a medida que la velocidad de llegada se aproxime a la velocidad de retransmisión.

En la figura 5.11, se representa la situación de las colas en un administrador de tramas o nodo Frame Relay.

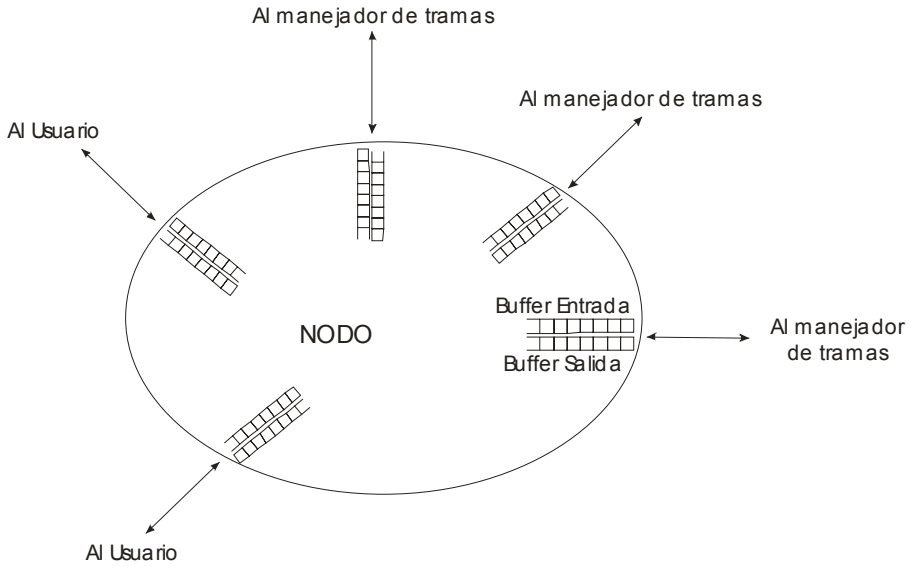


Figura 5.11 Colas en un nodo F R.

Cualquier manejador tiene conectado un determinado número de enlaces de transmisión a otros manejadores y directamente a usuarios finales. En cada enlace las tramas entran y salen. Puede considerarse que hay dos buffers en cada enlace: uno que recibe las tramas que llegan y otro que guarda las tramas que están esperando ser transmitidas. Es factible imaginar cada enlace como dos buffers de tamaño variable, con la única limitación de que la suma de sus tamaños debe ser, siempre constante.

De cualquier manera, cuando llega una trama, se almacena en el buffer de entrada del enlace correspondiente. El manejador examina cada trama de entrada para tomar la decisión de encaminamiento y entonces mueve dicha trama al buffer de salida más apropiado. Las tramas encoladas para salir se transmiten tan rápidamente como sea posible. Pero si las tramas llegan muy rapido al manejador para que éste pueda procesarlas, o llegan más rápido de lo que parten las tramas de los buffers de salida, entonces habrá un momento en el que no se dispondrá de memoria para las nuevas tramas de entrada.

Cuando se alcanza este punto de saturación, se pueden adoptar dos estrategias. La primera consiste simplemente en descartar cualquier trama de entrada para la que no haya espacio en el buffer. Pero este método no es aconsejable, ya que las tramas descartadas deben ser retransmitidas, aumentando de este modo la congestión de la red. La otra alternativa es utilizar algún mecanismo que limite la velocidad a la que las nuevas tramas entran en la red. Este procedimiento es realmente conocido como control de congestión.

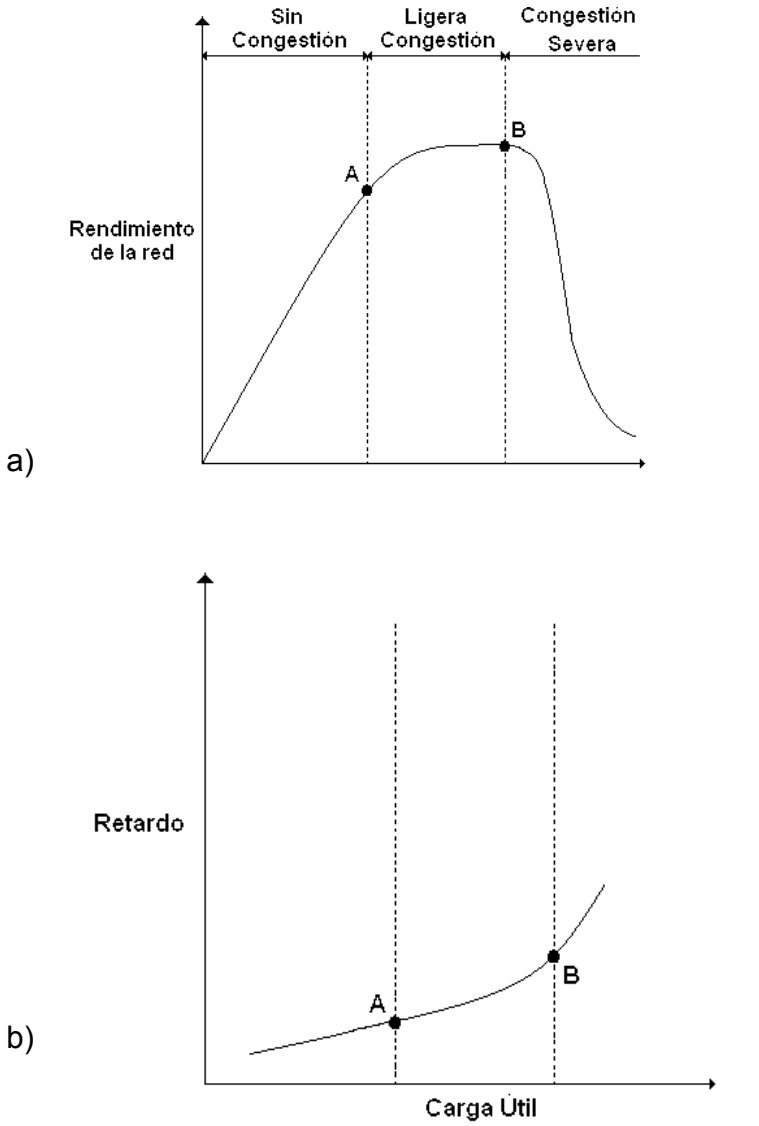


Figura 5.11 Efectos de la congestión.

Las figura 5.11 muestran los efectos de la congestión en términos generales. La primera muestra el rendimiento de una red (número de tramas transmitidas a la estación de destino por unidad de tiempo) frente a la carga ofrecida (número de tramas transmitidas por todos los abonados); mientras que la segunda presenta el retraso medio a través de la red, desde la entrada a la salida. Con poca carga el rendimiento aumenta proporcionalmente al aumento de la carga ofrecida. A medida que la carga va creciendo, se alcanza un punto (punto en el gráfico) a partir del cual el rendimiento de la red crece más lentamente que el

crecimiento de la carga ofrecida. Esto, es debido a que la red está entrando en un estado de congestión ligera. En esta región, la red continúa encargándose de toda la carga aunque con retardos mayores.

A medida que la carga de la red aumenta, la longitud de las colas de los manejadores crece y se alcanza un punto (punto B en el gráfico) más allá del cual el rendimiento disminuye a medida que aumenta la carga ofrecida. Esto es debido a que los buffers de cada manejador son de tamaño finito y cuando se llenan deben descartar tramas. Estas deben ser retransmitidas por el origen, sumándose a las nuevas que entran en la red. Lo único que consigue este hecho es empeorar la situación: a medida que se retransmiten más tramas, la carga del sistema crece y se saturan más buffers. Incluso las tramas que se mandan con éxito tienen que ser retransmitidas, porque el mensaje ACK tarda tanto tiempo en llegar que el origen, asume que la trama no ha llegado al destino. Bajo estas circunstancias, la capacidad efectiva del sistema es virtualmente cero.

Es evidente que es necesario evitar este tipo de situaciones, y es precisamente ésa la misión del control de congestión. El objetivo de todas las técnicas de control de congestión es limitar la longitud de las colas en los manejadores de tramas para evitar el colapso del rendimiento.

CONTROL DE CONGESTION

La UIT, define los objetivos del control de congestión en Frame Relay de la siguiente manera:

- Minimizar el descarte de tramas.
- Mantener, con una probabilidad alta y mínima variación, la calidad de servicio acordada.

- Minimizar la posibilidad de que un usuario monopolice los recursos de la red a expensas de otros usuarios.
- Ser fácil de implementar y suponer poca carga para los usuarios finales de la red.
- Crear el menor tráfico adicional posible en la red.
- Distribuir los recursos de la red equivalente entre los usuarios.
- Limitar la transmisión de la congestión a otras redes del tráfico, en cualquier dirección entre los usuarios finales.
- Tener la mínima interacción con, o impacto sobre, otros sistemas en la red Frame Relay.
- Minimizar la variación de la calidad del servicio debida a las conexiones Frame Relay individuales durante la congestión.

El control de congestión es especialmente en este tipo de redes. El protocolo Frame Relay está orientado a conseguir el máximo rendimiento y eficiencia. Esto tiene como consecuencia que los manejadores de tramas no puedan controlar el flujo de las mismas que llegan de un abonado o de un manejador adyacente mediante el típico protocolo de ventana deslizante.

El control de congestión es una responsabilidad compartida entre la red y los usuarios finales. La red es la que mejor puede monopolizar el grado de congestión limitando el tráfico. Teniendo esto en cuenta, es posible considerar dos estrategias generales para el control de congestión.

- Los procedimientos para evitar la congestión, se utilizan cuando estas se inicia, a fin minimizar sus efectos sobre la red. Estos procedimientos se inician antes, o en el punto A de la figura 5.11, inciso a), para evitar el tratamiento de la congestión que se produce en el punto B. Cerca de punto A, es difícil para el usuario final advertir que la congestión se está incrementando, por lo que debe existir un mecanismo de señalización explicito en la red que dispare estos procedimientos.

- Los procedimientos de recuperación de la congestión se utilizan para prevenir el colapso de la red en la fase de congestión severa. Se inicia generalmente cuando la red empieza a eliminar tramas debido a la congestión, Estas, sirven como un mecanismo de señalización implícito.

UIT-T y ANSI consideran a estas dos estrategias como formas complementarias de control de congestión en el servicio portador de retransmisión de tramas.

Calidad de Servicio, QoS

Es posible contratar para cada conexión, una calidad de servicio distinta. Dicha calidad está definida mediante ciertos parámetros:

- **CIR (***Committed Information Rate*) (bits/s): Es la tasa de información comprometida, es decir, el caudal medio garantizado que la red se compremete a dar en una conexión durante un intervalo de tiempo definido (Tc). Es un parámetro asociado a cada sentido de la transmisión de cada circuito virtual.

Se define como una relación entre el tiempo real y el volumen de información transferida:

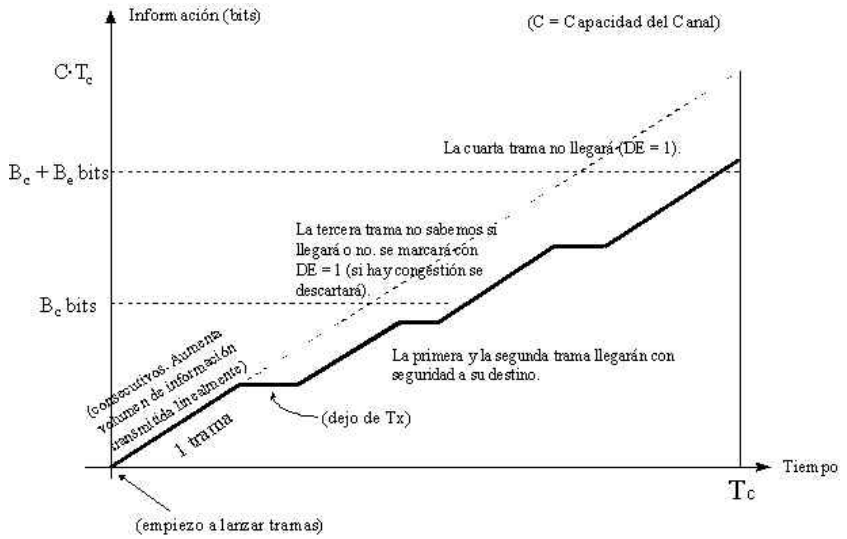


Figura 5.12 Gráfica de transmisión de tramas.

- **T_c** (*Commited rate measurement interval*): Intervalo de observación *(es el tiempo hasta el cual ha sido representado la gráfica anterior)*. Parámetro del algoritmo para calcular el CIR).

- **$C \cdot T_c$**: Máximo volumen de información que se podría cursar en T_c (es lo que posibilita el canal).

El caudal físico (C) de la línea de acceso también se contrata. Así el operador dimensiona la red en función de los parámetros contratados por sus abonados.

En el interfaz usuario-red se controla, para cada circuito virtual, que los usuarios se ajusten a los parámetros Bc, y Be que han negociado. Si la red está bien diseñada no debe perder datos que no superen el tráfico comprometido.
Se definen dos zonas en el diagrama:

- **B_c** (*Committed burst size*): Es el volumen de información comprometida: durante el intervalo T_c la compañía se **compromete** a transmitir un volumen B_c.

- **B_e**: Volumen de información en exceso: la información cursada durante el intervalo T_c que exceda de B_c + B_e **no se sabe si llegará o no** a su destino (la compañía no lo garantiza). El volumen de información que exceda de B_c + B_e seguro que **no llegará**.

  Este método, se aplicará para cada circuito virtual de ingreso a la red.

Existe un bit en la trama (bit DE) que es activado por la red en tramas que superen Bc (es decir aquellas que pertenezcan a Be) para indicar que esas tramas deberían ser descartadas en preferencia a otras, si es necesario. El servicio permite que el propio usuario también pueda marcar este bit para indicar la importancia relativa de una trama respecto a otras (en este caso, estas tramas no se contabilizan como pertenecientes a la zona bajo B_c, sino como perteneciente a la zona sobre B_c y bajo B_c + B_e, no contando para el CIR).

El parámetro $C \cdot T_c$ está asociado a la capacidad física de las líneas, y es lo primero que contrata el abonado. Luego, sobre esa línea física, se definen mallas de circuitos virtuales, cada uno con su CIR asociado.

Su formula es:

$$\mathbf{B_c = CIR \cdot T_c}$$

El CIR no es la capacidad física a la que se transmite. Esa velocidad es la de la capacidad del canal. El CIR sólo es el caudal medio (estadístico).

Si el T_c se toma grande, existe la posibilidad de transmitir grandes picos de información en algunos momentos y nada de información en otros. Por tanto, un T_c pequeño nos garantiza el que la transmisión sea más homogénea (esto interesa a la empresa, ya que así se evita sobredimensionar las redes).

RESUMEN

En este capítulo se presenta la forma en que la calidad de servicio (QoS), se maneja en dos tipos de tecnologías a nivel WAN, como lo son ATM y Frame Relay, haciendo hincapié en la forma que cada una de éstas mediante su operación proporciona la QoS.

6. QoS PARA TELEFONÍA IP (Ejemplo de Aplicación)

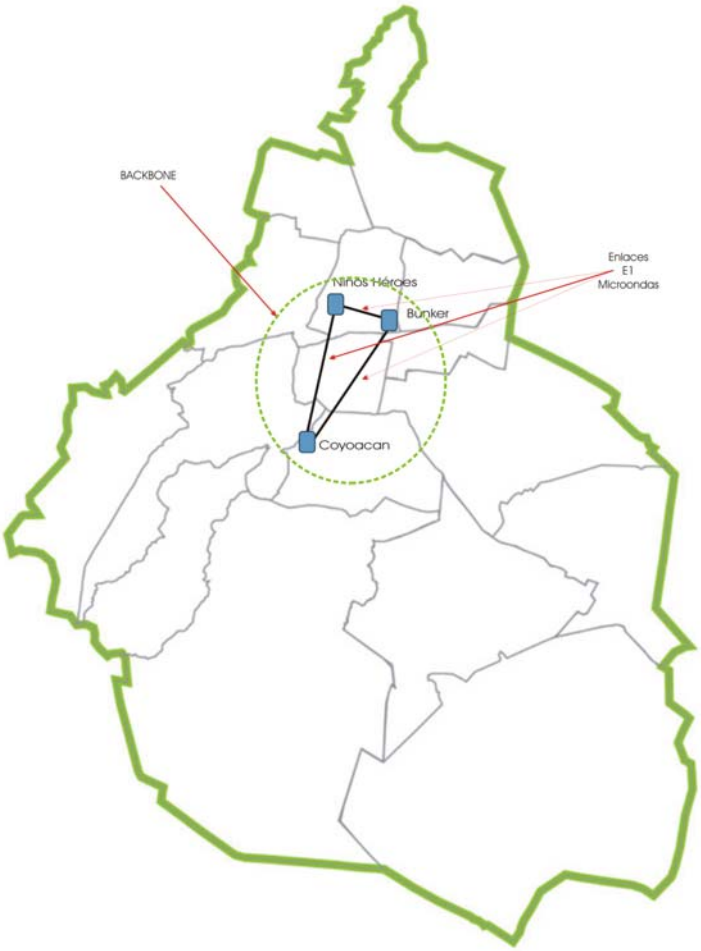
En el presente capitulo se muestra un ejemplo de la aplicación de la calidad de servicio para el manejo de telefonía IP, contando con la explicación de la red en la cual fueron habilitados dichos servicios de voz, además del análisis comparativo entre nodos de similares características (uno contando con los servicios de Telefonía IP y el otro conservando la tecnología con la que trabaja el resto de la red).

Dentro de las principales causas que conllevaron la migración a la telefonía IP, se tienen las siguientes: *La necesidad de homogenizar el uso de sistemas para la atención ciudadana, esto trajo consigo el aumento en el número de usuarios en red, así como el tráfico en la misma*. Y por otro lado, el rezago tecnológico se hizo presente puesto que los servicios y refacciones para esta tecnología se hicieron cada vez más escasos y en consecuencia con mayor costo.

Dentro de los objetivos a cubrir en este capítulo, básicamente, es dar a conocer la red en donde se implementarán los servicio de telefonía IP, haciendo mención de las tecnologías involucradas para el funcionamiento de la misma, así como la implementación de la Calidad de Servicio para el mejor desempeño de esta solución. Para facilitar el manejo de la información dentro de este capítulo, no se profundizará en la mención de las tecnologías, ya que al hacerlo, se puede envolver al lector en tanta tecnología y terminología empleada, sin embargo se hará hincapié en las características más relevantes para mejor comprensión del tema en cuestión.

6.1. ANTECEDENTES TÉCNICOS

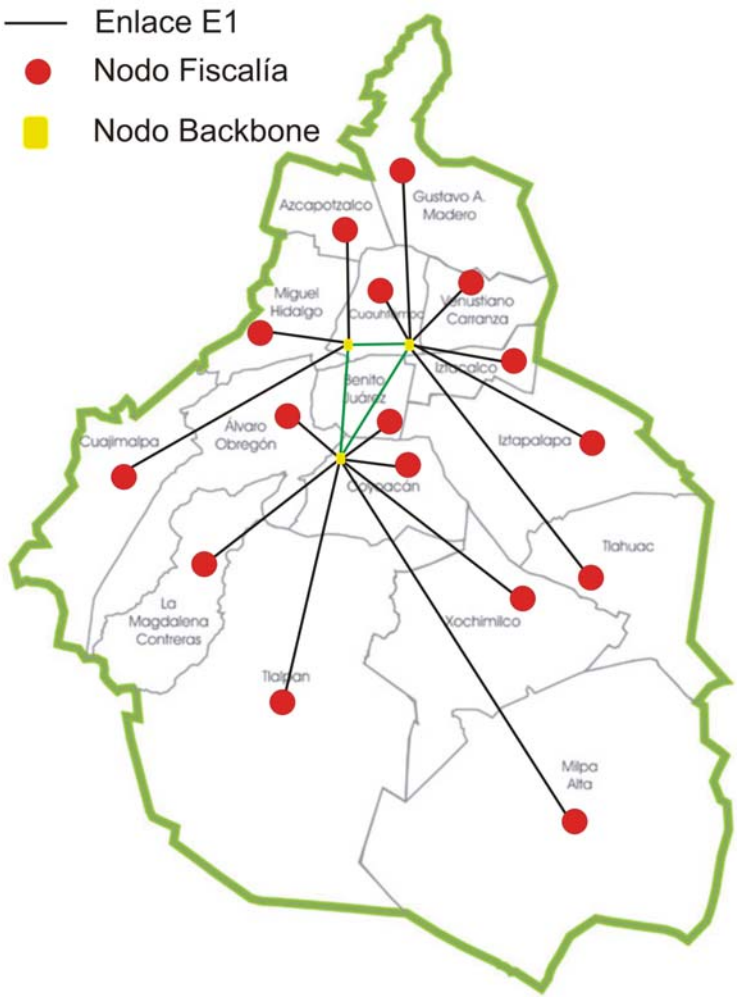
Se trata de una red implementada en la Procuraduría General de Justicia del Distrito Federal, la cual cuenta con 3 nodos principales, mismos que conforman el Backbone o núcleo de la red. Estos, se encuentran en diferentes puntos de la ciudad. Los cuales se interconectan con enlaces E1 de microondas, propios de la institución, empleando una topología en anillo.



6.1. Distribución del Backbone de la P.G.J.D.F.

Posteriormente, en cada una de las delegaciones políticas se ubican las Fiscalías, las cuales son denominadas como nodos de Distribución (esto denominado conforme al modelo de 3 capas propuesto por Cisco, Anexo C). Las Fiscalías se encuentran conectadas a los diferentes nodos de Backbone,

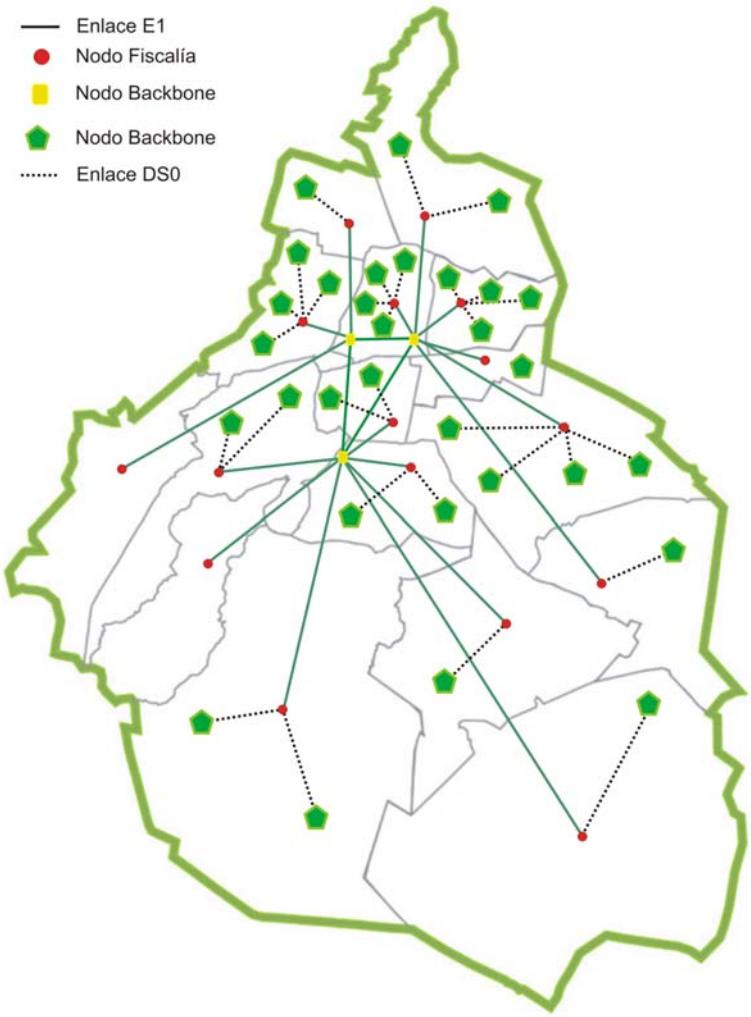
dependiendo de la ubicación territorial correspondiente. De igual manera se interconectan con enlaces de microondas mediante una topología en estrella.



6.2. Distribución de las Fiscalías en la red de la P.G.J.D.F.

Dado que en cada demarcación, son necesarios los servicios que proporciona la institución en cuestión y que la concentración de dichos servicios en un solo punto de la delegación sería insuficiente, se cuenta con diferentes Coordinaciones Territoriales (anteriormente denominadas Agencias del Ministerio Público), distribuidas en puntos estratégicos de dicha delegación, a éstas Coordinaciones Territoriales, se le da el nombre de nodos de Acceso, haciendo nuevamente referencia al modelo de Cisco. La conexión de los nodos de Acceso

hacia los nodos de Distribución se da mediante enlaces DS0, proporcionados por Telmex.



6.3. Distribución de las Coordinaciones Territoriales en la red de la P.G.J.D.F.

ENLACES

Los enlaces utilizados para las interconexiones de cada uno de los niveles. Para la interconexión del backbone se utilizan enlaces E1* Punto a Punto provistos por equipos de microondas, contando con respaldos de la misma capacidad proporcionados por un proveedor de servicios local (Telmex) y en otros casos, se cuenta con enlaces de Gigabit Ethernet, que al igual que los servicios de microondas son infraestructuras de la institución.

Para la conexión del backbone hacia el nivel de distribución se manejan enlaces dedicados E1* Punto a Punto proporcionados de igual manera por equipos de microondas, propios de la Institución.

En la conexión del nivel de distribución hacia el nivel de acceso, se manejan enlaces E1* Punto Multipunto, provisto por el proveedor de servicios locales, en donde cada enlace remoto consta de 2 DS0´s* y en algunos casos, se llegan a manejar hasta 4 DS0´s*.

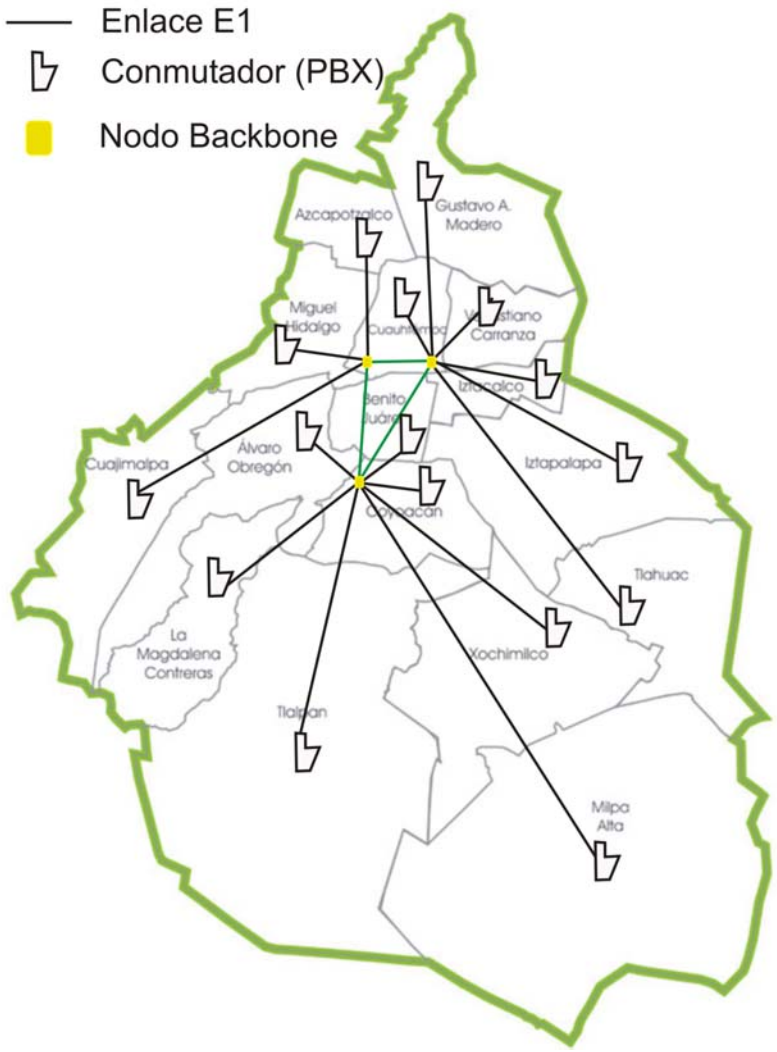
*** Los enlaces E1 cuentan con la capacidad de 2 Mbps y los DS0 con 64 Kbps.**

TELEFONÍA

Por otro lado, se encuentra la red telefónica institucional, la cual esta implementada con el objetivo de contar con servicios telefónicos suficientes, en cada uno de sitios de la Procuraduría, de manera tal que haya comunicación telefónica interinstitucional, así como acceso a la red pública de telefonía.

El equipo encargado de proporcionar las extensiones telefónicas, son conmutadores de la marca Ericsson, habiendo uno en cada nodo de la capa de distribución, siendo éste el que proporcione las extensiones a ser transportadas hacia los diferentes nodos de la capa de acceso dependientes de él. Además de contar con varios equipos en cada nodo de la capa de Backbone para proporcionar el servicio en cada nodo principal.

Cabe mencionar que estos equipos se encuentran interconectados, por medio de enlaces dedicados de microondas con capacidades de 2,048 kbps (E1), y cuentan con una consola de administración, de manera que al igual que el resto de los equipos, se puede acceder a ellos de manera remota y centralizada, haciendo mucho más fácil la operación y administración de cada uno de ellos.



6.4. Red telefónica de la P.G.J.D.F.

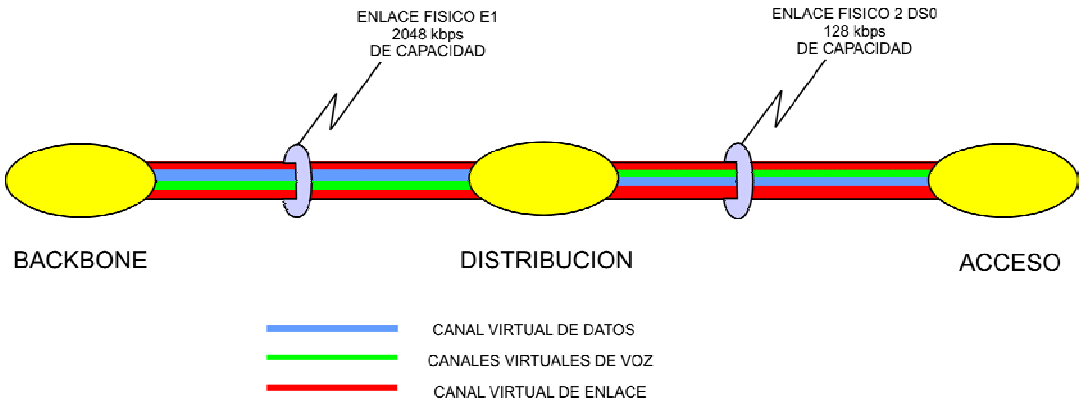
6.2. OPERACIÓN ACTUAL DE LA RED CON MULTIPLEXORES

Como se mencionó anteriormente, la red institucional maneja el transporte de voz y datos con multiplexores, misma que se describe a continuación.

El funcionamiento de esta tecnología se basa en el transporte de aplicaciones en canales lógicos separados definiendo anchos de banda preestablecidos, fijos para cada aplicación, lo que implica destinar esos anchos de banda para cada aplicación de manera permanente sean o no utilizados.

De esta manera, se crea un canal lógico de ancho de banda fijo de la posición “x” (posición de la tarjeta en el multiplexor) en el nodo origen, a una posición “y” (posición de la tarjeta en el multiplexor) en el nodo destino pero transportados todos por un mismo enlace físico.

Inicialmente hay que configurar los circuítos de enlace, mismos que se encargarán de transportar a los circuítos de aplicación (voz y datos) de un nodo a otro, como se muestra en la figura 6.5.

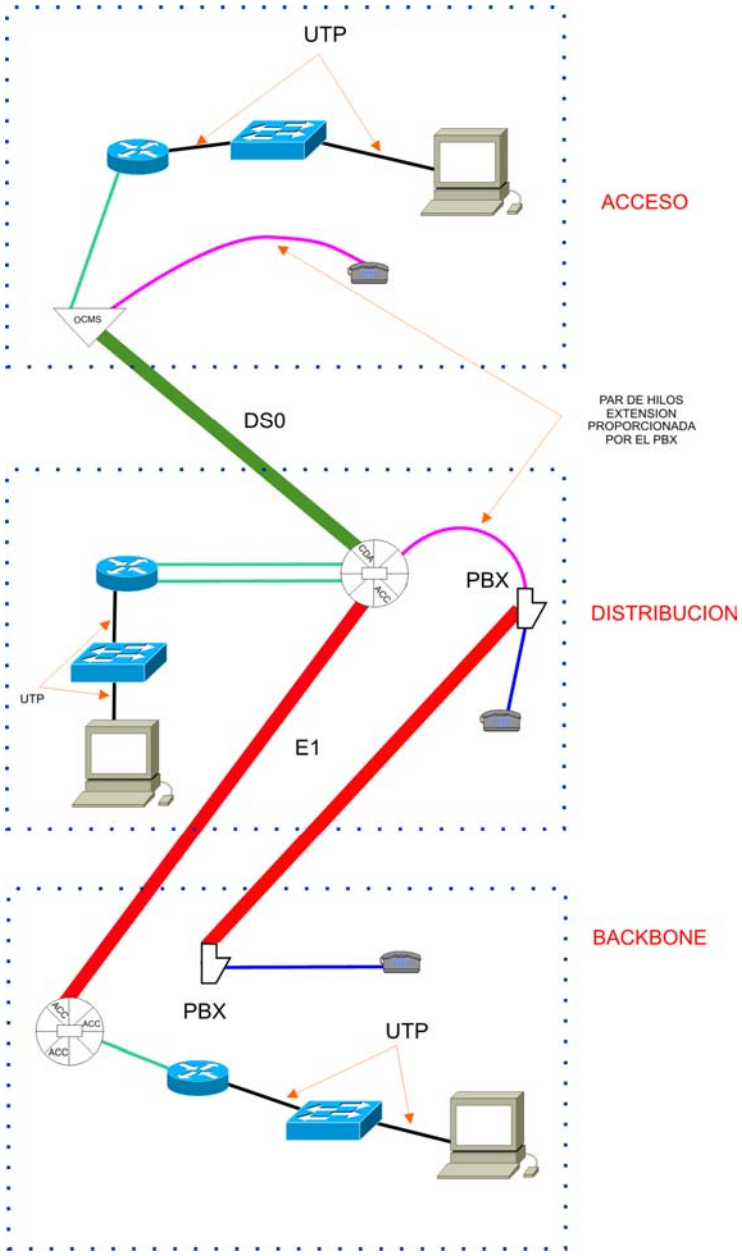


6.5. Canales lógicos de voz, datos y enlace.

Obteniendo como resultado el transporte de las aplicaciones de voz y datos entre diferentes puntos de la red, transportados a través de un enlace principal.

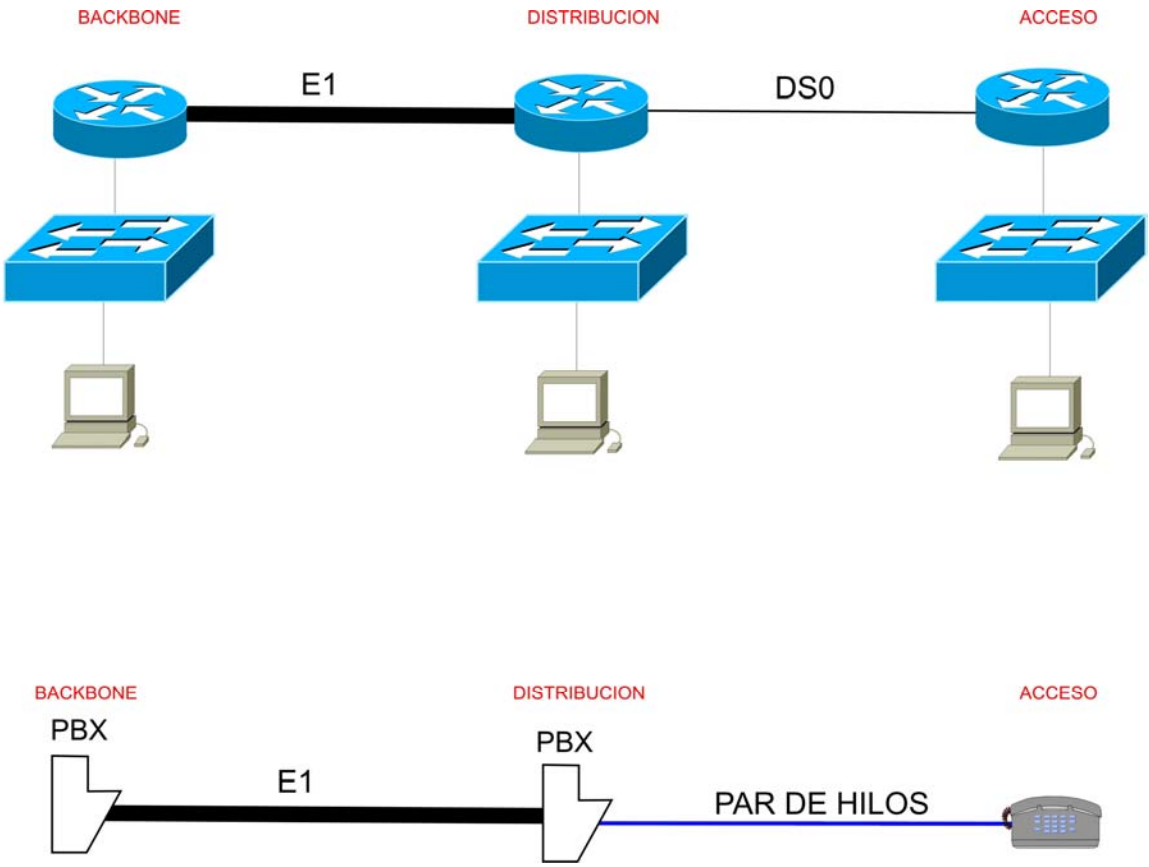
Para la interconexión física de los equipos se utilizan los enlaces descritos con anterioridad (DS0, E1 punto a punto).

En la figura 6.6, se muestra la conexión de un nodo central (Backbone), con un nodo de Distribución y su respectivo nodo de Acceso, implementando las aplicaciones de voz y datos.



6.6. Distribución de aplicaciones de voz y datos en la red.

Como lo muestra la figura 6.6 cada tarjeta tiene su correspondiente en el nodo opuesto, formando así un canal invisible a la aplicación, obteniendo de esta manera el siguiente esquema 6.7.



6.7. Esquema independiente de aplicaciones (voz y datos).

En el caso de datos, los ruteadores, simplemente verán sus routers vecinos, sin conocer qué o quién es el encargado de proporcionar el transporte de los datos.

Para el caso de voz, la conexión entre conmutadores (PBX), se muestra de igual manera invisible, siendo la conexión entre PBX de Backbone y PBX de Distribución un E1 (2.048 Mbps), mientras que entre Distribución y Acceso (PBX - teléfono), se conectan por medio de un par de hilos, obteniendo una extensión remota, la cual ha sido programada en el PBX de Distribución y conectada en el puerto correspondiente hacia el nodo de Acceso, misma que será entregada en el multiplexor de Acceso sin sufrir alteraciones.

Lo más importante a mencionar de esta tecnología es como ya se indicó anteriormente la designación de anchos de banda fijos para cada canal de comunicación creado, esto garantiza que cada aplicación ya sea de voz o datos tenga considerado su canal de comunicación garantizado a un ancho de banda definido siempre que lo requiera utilizar, por lo que en el peor de los casos todas las aplicaciones consideradas funcionarán a cierta velocidad definida evitando de esta manera la saturación por demanda de servicios de los canales y enlaces que se tengan, pero con una desventaja en comparación con las nuevas tecnologías emergentes, y que finalmente empezó a desplazar a ésta, y es que la característica principal de las nuevas tecnologías es que manejan de forma dinámica esos anchos de banda de cada aplicación, esto es de forma resumida, *si una aplicación no utiliza el canal el ancho de banda no utilizado es usado por alguna otra que lo requiera*, adicionalmente, se maneja como se ha venido mencionando mecanismos de priorización de tráfico, lo que redunda en un manejo óptimo de los anchos de banda para las aplicaciones.

6.3. MIGRACION A NUEVA TECNOLOGIA

Con las diferentes propuestas tecnológicas emergentes se logró reforzar el Backbone con la instalación de switches WAN con tecnología Frame Relay, adicionalmente, se presentó la propuesta de implementación de Telefonía IP y siendo ésta, inicialmente instalada en los nodos de acceso de un nodo de distribución, para posteriormente hacerla extensiva a otros nodos fuera del nodo de distribución, ya que las capacidades de los equipos permiten el manejo y la administración de una cantidad suficiente de teléfonos, sin que esto llegue a afectar su operación.

Antes de describir la migración hacia esta nueva tecnología, es necesario describir el alcance al cual fue programado el proyecto de migración.

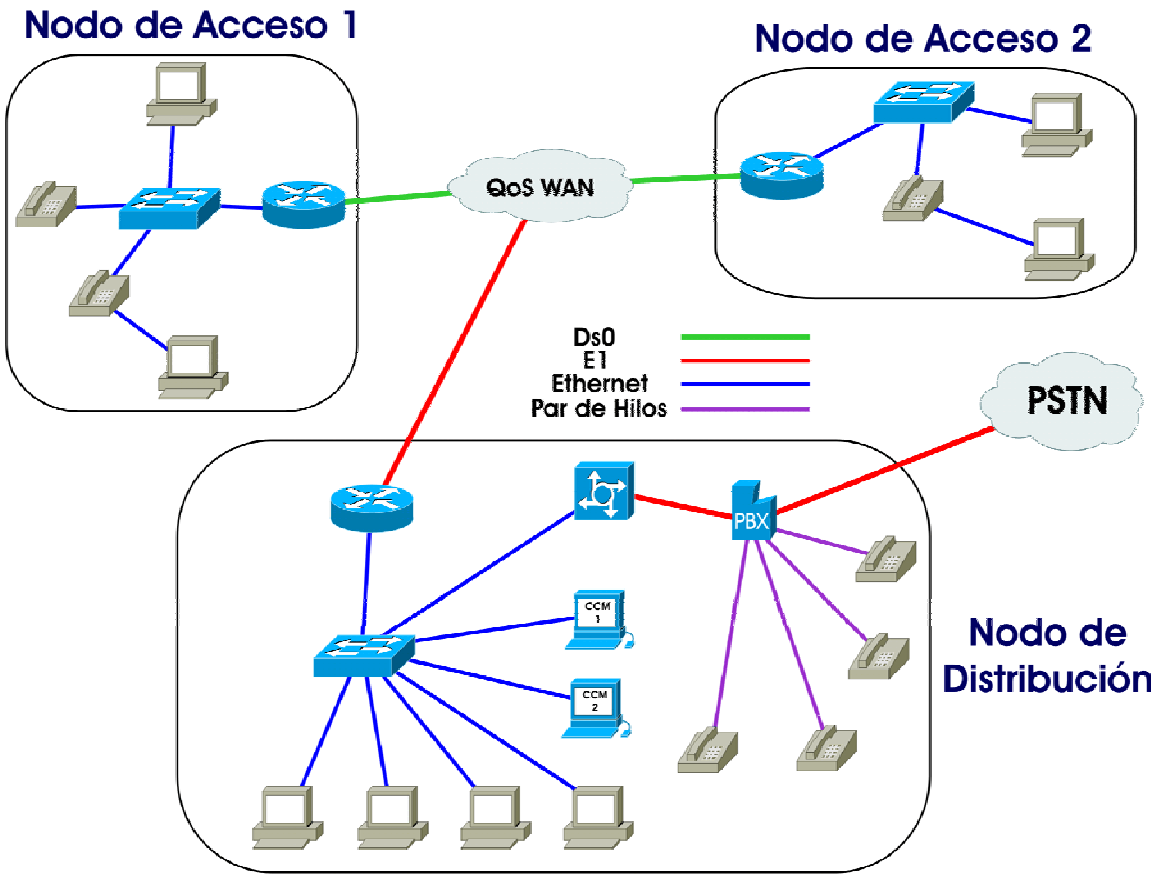
Inicialmente, se contempló la implantación de telefonía IP en los nodos de acceso de un sólo nodo de distribución, así mismo, el equipamiento necesario en el nodo de acceso. De esta manera, se contemplaron 7 teléfonos IP como máximo en cada nodo de acceso, con la restricción de 5 llamadas simultáneas debido a la capacidad del ancho de banda (128 kbps), ya que cada llamada ocuparía aproximadamente 18 kbps, proporcionando la misma calidad de voz que una línea convencional a 64 kbps. Previendo una cantidad especifica mínima para el transporte de datos.

Para el manejo de los equipos de telefonía IP en el nodo de distribución, se implemento el manejo de Redes de Área Local Virtuales (VLAN's, por sus siglas en inglés), de esta forma, se logra manejar por separado el tráfico de voz IP y el tráfico de las aplicaciones de dicho nodo, mientras que la telefonía en el nodo en cuestión (de Distribución), seguiría siendo proporcionada por el equipo Ericsson, al igual que en el resto de la red.

Ahora, para el manejo del tráfico de los nodos de acceso, hubo la necesidad de cambiar el modelo de los ruteadores, debido a que los que se encontraban anteriormente, no tenían la capacidad de manejar QoS, además de las capacidades de memoria, con este criterio, se cambiaron ruteadores de la familia 2500 de Cisco por unos de la familia 2600 de la misma marca. También, fúe necesario instalar al menos un switch Catalyst 3500 XL de la misma marca, ya que este equipo ofrece la característica de ser “In-line power”, esto significa que por el mismo puerto de datos, puede proporcionar la alimentación eléctrica para el teléfono IP, evitando con esto aumentar físicamente un cable más en la conexión de nuestro teléfono y ahorrando una toma de CA.

Dentro de las ventajas que se manifestaron en la implementación de la telefonía IP, además del manejo de tecnologías como ancho de banda dinámico, protocolo H.323, fue el manejo de Calidad de Servicio para la priorización de los paquetes de voz, para lograr así una mejor calidad de voz en las extensiones IP.

En el diagrama 6.8, muestra la forma de interconexión de los equipos antes mencionados para el manejo de telefonía IP.



6.8. Interconexión de aplicaciones implementando Telefonía IP.

Con la sustitución del equipo anteriormente mencionado, se logró mejorar el desempeño de las aplicaciones en red, reduciendo el dominio de colisiones y aumentando la velocidad en el entorno LAN.

En el nodo de distribución, además de cambiarse el router y los hubs por equipo de mayor capacidad, se instaló un Access Server (Voice Gateway), para hacer la interconexión con el PBX, así como dos Administradores de voz (Call Manager), en los cuales se encuentra la base de datos de los teléfonos instalados en cada nodo de acceso, además de grupos de llamadas (para llamadas locales, a celulares, de larga distancia nacional, internacional o mundial). Cabe recordar que en este nodo, el servicio telefónico siguió brindándose con el PBX Ericsson.

Para habilitar los teléfonos IP en los nodos de acceso, después de haber instalado los nuevos equipos (router y switch), sólo fúe necesario configurarlos de acuerdo al direccionamiento IP correspondiente al sitio en cuestión, utilizando un rango específico asignado previamente para ellos.

Cabe mencionar que los teléfonos cuentan con 2 puertos ethernet, uno para conectar al nodo de datos y el segundo proporciona la conexión para otro dispositivo con puerto ethernet que deseé conectarse (computadora personal, impresora), en caso que sólo se tenga un nodo de datos en el lugar donde se instalará el teléfono.

El teléfono IP al ser un elemento de red requiere para su operación dentro de este ambiente, los parámetros necesarios (dirección IP, máscara de red y puerta de enlace) para poder interactuar con los elementos de red Switches, Routers y demás equipos mencionados anteriormente para logra la comunicación con el administrador de llamadas (Call Manager), de esta manera, el dispositivo logra descargar su configuración desde el call manager para posteriormente hacer llamadas dentro del entorno IP y hacia el exterior del mismo (llamadas dentro de la institución pero fuera del entorno IP, llamadas locales, a celulares, de larga distancia nacional, internacional o mundial), todas controladas por dicho dispositivo.

El equipo que se encarga de la interacción entre el entorno IP y el resto de la red o en caso de llamadas al exterior (PSTN), es el Access Server, quien esta directamente conectado al PBX, y al generarse una llamada evalúa si esta dentro del entorno IP o es necesario pasarlo al PBX para que éste se encargue de darle el tratamiento de acuerdo al tipo de llamada.

El código de configuración del puerto del Access Server donde se recibe este enlace es de la siguiente manera.

```
controller E1 0
clock source line primary
line-termination 75-ohm
pri-group timeslots 1-31
description ENLACE ISDN AL PBX ERICSSON MD110
```

Además, en este equipo se encuentra la configuración de los diferentes destinos, además de definir el tipo de codec de audio que estará utilizando para el tráfico de voz. A continuación se presenta el texto de la configuración de los diferentes destinos de extensiones IP.

```
dial-peer voice 30 voip
preference 1
destination-pattern 87..
dtmf-relay h245-alphanumeric
codec g729br8 bytes 50
session target ipv4:172.33.160.30
!
dial-peer voice 50 voip
destination-pattern 87..
dtmf-relay h245-alphanumeric
codec g729br8 bytes 50
session target ipv4:172.33.160.31
!
dial-peer voice 40 voip
preference 1
destination-pattern 165..
dtmf-relay h245-alphanumeric
codec g729br8 bytes 50
session target ipv4:172.33.160.30
!
```

```
dial-peer voice 31 voip
 preference 1
 destination-pattern 88..
 dtmf-relay h245-alphanumeric
 codec g729br8 bytes 50
 session target ipv4:172.33.160.30
!
dial-peer voice 51 voip
 destination-pattern 88..
 dtmf-relay h245-alphanumeric
 codec g729br8 bytes 50
 session target ipv4:172.33.160.31
!
dial-peer voice 32 voip
 preference 1
 destination-pattern 85..
 dtmf-relay h245-alphanumeric
 codec g729br8 bytes 50
 session target ipv4:172.33.160.30
!
```

```
dial-peer voice 60 voip
destination-pattern 165..
dtmf-relay h245-alphanumeric
codec g729br8 bytes 50
session target ipv4:172.33.160.31
```

```
dial-peer voice 52 voip
 destination-pattern 85..
 dtmf-relay h245-alphanumeric
 codec g729br8 bytes 50
 session target ipv4:172.33.160.31
```

En la configuración anterior se puede observar que se repite la configuración para el *destination-pattern*, esto debido a que se hace la primer declaración (*dial-peer voice*) para la dirección IP de Call Manager principal, y la segunda declaración va configurada para el segundo Call Manager, esto es necesario, para que cuando el principal no este presente, pueda consultar en el secundario.

Estos equipos (CALL MANAGER), se encuentran configurados de manera que se pueda acceder a ellos de manera remota, ya sea vía Web Browser (acceso a la aplicación), ó utilizando la utilería *Net Meeting* (acceso remoto al equipo).

Ahora bien, para el manejo de la calidad de servicio en nuestra red, inicia básicamente en el switch, manejando clasificación de paquetes en capa 2 (del modelo OSI), configurando la Clase de Servicio (CoS), agregando en cada puerto la siguiente configuración:

```
interface FastEthernet0/2
switchport priority extend cos 7
spanning-tree portfast
```

La instrucción anterior indica que a la interfase Fast Ethernet 0/2 se le dará prioridad en el entorno LAN, lo que significa que en caso de haber una gran cantidad de tráfico hacia el router, los paquetes provenientes de ese puerto tendrán prioridad sobre los demás.

Posteriormente, en los routers, es necesario configurar cómo será tratado el tráfico previamente seleccionado, de manera que la configuración de éste método se encuentra de la siguiente manera:

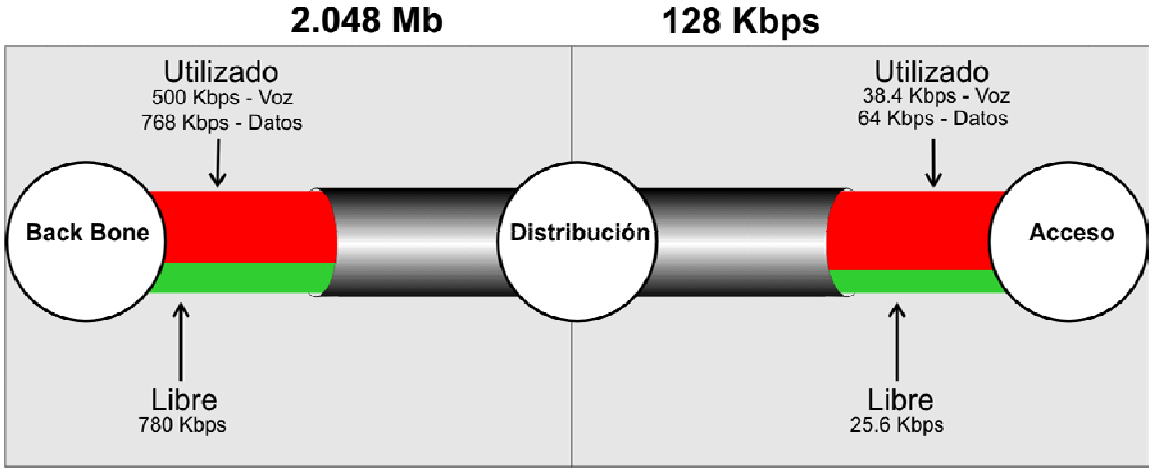
```
class-map match-all VoIP-Control        *declara una clase llamada VoIP-Control
  match access-group 101                *en la clase se llama a una lista de acceso
class-map match-all VoIP-RTP            *declara una clase llamada VoIP-RTP
  match access-group 100                *en la clase se llama a una lista de acceso
!
policy-map QoS-Policy-128k              *declara una política llamada QoS-Policy-128k
  class VoIP-RTP                        *llama a la clase VoIP-RTP
    priority 70                         *define la mas alta prioridad
  class VoIP-Control                    *llama a la clase VoIP-Control
   bandwidth 8
  class class-default
   fair-queue                           *define el tipo de encolamiento a utilizar
!
call rsvp-sync
!
interface Serial0/0
 description ENLACE A DEL. CUAUHTEMOC
 no ip address
 encapsulation frame-relay              *define el tipo de encapsulamiento
 no ip mroute-cache
 load-interval 30
 no arp frame-relay
 frame-relay traffic-shaping            *define el mecanismo para obtener QoS
!
interface Serial0/0.1 point-to-point
 description LINK-TO-CUAUHTEMOC
 bandwidth 128
 ip address 172.18.250.14 255.255.255.252   *define el direccionamiento a utilizar
 no ip mroute-cache
 frame-relay interface-dlci 116             *identifica el dlci a utilizar
  class VoIP-128kbs                         *llama a la clase VoIP-128kbs
 frame-relay ip tcp header-compression      *identifica compresion de cabeceras IP
 frame-relay ip rtp header-compression      * identifica compresion de cabeceras RTP
!
ip classless
no ip http server
!
map-class frame-relay VoIP-128kbs       *define a la clase VoIP-128kbs
 no frame-relay adaptive-shaping
 frame-relay cir 128000
 frame-relay bc 1280
```

```
 frame-relay be 0
 frame-relay mincir 128000
 service-policy output QoS-Policy-128k
 frame-relay fragment 154
access-list 100 permit ip any any precedence critical
access-list 100 permit ip any any dscp ef
access-list 101 permit ip any any precedence flash
access-list 101 permit ip any any dscp af31
!
```

6.4. ANALISIS COMPARATIVO DE RENDIMIENTO

Como se ha venido mencionando, al trabajar la red con los multiplexores, se manejan anchos de banda estáticos para el transporte de voz y datos, teniendo como límite la capacidad total del enlace.

Para ejemplificar de una mejor manera, se muestra el esquema 6.9.



6.9. Utilización del ancho de banda empleando canales separados para cada aplicación.

Como se observa en la figura 6.9, entre el nodo de Backbone y el nodo de Distribución, se tiene que se utilizan 768 kbps para el transporte de datos y 500 kbps para el de voz, además de 12 kbps para la administración, quedando un remanente de 768 kb de un total de 2048 kbps (E1), que es el ancho de banda total del enlace.

Además, del nodo de Distribución al de Acceso, contamos con un enlace de 128 kbps, tomando 64 kbps para el transporte de datos, y 5 canales de 9.6 kbps cada uno para transporte de extensiones remotas, además de 4.8 kbps para la administración, teniendo un remanente de 11.2 kb.

Obteniendo de esta forma las siguientes cifras, la utilización del enlace Backbone – Distribución es del 62 %, mientras que el enlace Distribución – Acceso, tiene una utilización de 91.2 %, esto contemplando que los circuítos de voz y datos se encuentren trabajando al 100 %.

Debido a que éste es el caso más crítico, se tendría una saturación de circuitos, pero en la realidad, los servicios de voz se utilizan aproximadamente en un 60 % en horas pico (de 10:00 a 16:00 hrs), debido a que el personal en sitio no se encuentra permanentemente utilizando la línea telefónica, en cambio en el circuíto de datos, ha habido casos en los que se presenta saturación del medio, provocando que los sistemas utilizados presenten lentitud y en ocasiones no haya conexión con el servidor remoto (esto se presenta principalmente en el enlace Distribución - Acceso), debido a alguna actualización de antivirus, o transferencias de información entre bases de datos (replica de bases de datos).

Haciendo más operaciones, aunado al 38 % de ancho de banda del enlace Backbone – Distribución (aunque se presente este porcentaje no utilizado del enlace, no afecta en la operación de los sistemas debido a la cantidad de ancho de banda dedicada para datos), se suma el 9 % de ancho de banda no utilizado del enlace de Distribución – Acceso, además del 40 % de no utilización de los circuitos de voz (19 kbps aproximadamente), representamos una gran desventaja en la utilización de anchos de banda fijos, característicos de la tecnología TDM.

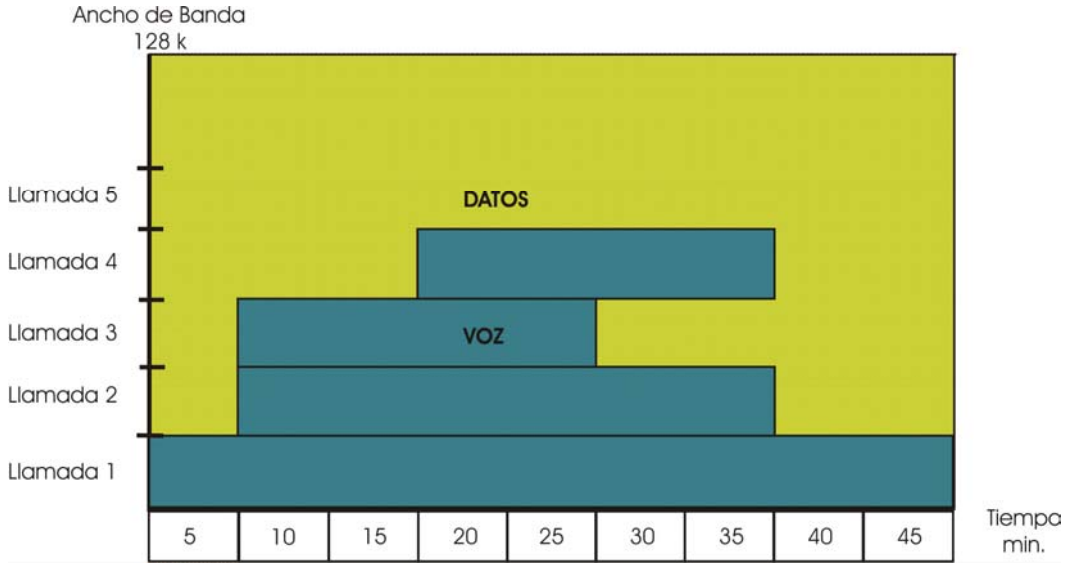
Ahora bien, con la implementación de la Telefonía IP, se logra contar con nuevas tecnologías, como lo es: Manejo de routers con capacidades para QoS,

switch LAN (lo que ayudo a disminuir el numero de hub instalados, obteniendo un mejor desempeño de las aplicaciones en red), teléfonos IP, administrador para telefonía IP (Call Manager).

Con esto, se buscó eficientar al 100 % los anchos de banda, con la misma cantidad de servicios instalados en el sitio en cuestión como mínimo.

Para lograr este beneficio, se utilizaron nuevos conceptos, tales como: manejo de anchos de banda dinámicos, manejo de CODECS para la compresión de cabeceras IP, priorización de paquetes, obteniendo como resultado que cada llamada IP, ocupara aproximadamente 18 kbps, mismos que serán utilizados durante la duración de la llamada y liberando ese ancho de banda ocupado al finalizar la misma.

Si bien es cierto, con la implementación de esta nueva tecnología, cada llamada maneja aproximadamente el doble del ancho de banda que utilizaba anteriormente, ahora el canal para la transmisión es de la capacidad total del enlace, ocupándose solamente durante la duración de la llamada. En cuanto la llamada termine, se liberará el ancho de banda quedando disponible para datos, como se muestra en la figura 6.10.



6.10. Utilización del ancho de banda con implementación de Telefonía IP

RESUMEN

El presente capítulo describe la forma de operación de una red WAN, mencionando algunas de las principales características de ésta, así como la implementación de una nueva tecnología en una parte de ella, logrando manejar en forma eficiente telefonía IP, la cual es una de las nuevas áreas en el ámbito de telecomunicaciones.

Esta tecnología, descrita a lo largo del proyecto de tesis, permite un mejor aprovechamiento de los recursos para el buen desempeño de las nuevas aplicaciones, sin descuidar las que ya se venían manejando.

CONCLUSIONES

Para finalizar el proyecto se tiene que las mejoras que se obtuvieron al implementar IP Telephony y consecuentemente QoS, se vieron reflejadas principalmente en la utilización del ancho de banda, ya que como se ha mencionado en reiteradas ocasiones, no se tiene desperdicio de éste, puesto que ahora se dispone del 100 % del enlace para el tráfico de voz y/o datos.

Teniendo en cuenta que el tráfico de datos en ocasiones se presenta en ráfagas que llegan a alcanzar grandes cantidades de utilización del medio, con la implementación del ancho de banda dinámico, al generarse una llamada IP, el tráfico de datos será reducido para darle paso al tráfico de voz sin que la ráfaga de datos llegue a afectar al de voz, y posteriormente al terminarse la llamada IP, nuevamente el total del ancho de banda estará disponible para el tráfico de datos, logrando con esto una mejor utilización del enlace (ancho de banda).

Además del mayor desempeño de las aplicaciones, se logró mejorar en cuanto a la tecnología, que como se había señalado anteriormente, al adquirir tecnología más reciente, se recuperaron refacciones para los equipos que se encuentran en operación en el resto de la red, así como equipos para habilitar mas servicios de voz y datos en nuevos nodos agregados.

Con estos puntos aclarados, se puede concluir que el objetivo propuesto al inicio de este proyecto se cumplió satisfactoriamente, ya que efectivamente al implementarse la Calidad de Servicio (QoS), específicamente para el manejo de Telefonía IP se permite tener una mejor administración del ancho de banda, reflejándose en un mejor desempeño de las aplicaciones de red y consecuentemente en una mejor atención al público, ya que la institución en donde se encuentra operando este tipo de tecnología es abocada a la atención ciudadana.

OTROS ASPECTOS ENTORNO A QoS

Dentro de las principales tendencias en las que los proveedores de servicios y fabricantes de dispositivos han puesto los ojos, es la convergencia de redes (integración de servicios [voz datos e imágenes]), con ello hacer posible el manejo de servicios multimedia, aplicando Calidad de Servicio, mediante diversos mecanismos de priorización, algoritmos de encolamiento, obteniendo con ello mejor desempeño de la aplicación a transportar por la red, tales como videoconferencia, telefonía IP, voz sobre IP, logrando emplear estas tecnologías para fines más prácticos, como lo es la telemedicina, asistencia remota, comercio electrónico (e-commerce), etc; otro importante avance entorno a la QoS, es el de contemplar la migración a Ipv6,cuyo principal objetivo es el de aportar mayor número de direcciones IP, entre otros, manejando direcciones de 128 bits y que al igual que en Ipv4, tiene un campo en la cabecera destinado para identificar el tipo de tráfico al cual sea necesario asignarle diferentes niveles de prioridad, dicho campo en este caso, es llamado Clase de Tráfico y en Ipv4 es el campo ToS.

En el caso de tecnologías de redes convergentes en los últimos años, se ubica principalmente que las inalámbricas (wireless) se presentan como una solución rápida al momento de instalar una red LAN que está rápidamente ganando mercado, por lo que es importante considerar el papel que juegan este tipo de tecnologías dentro de la integración de servicios considerando un factor muy importante de las mismas: su velocidad de transmisión (los servicios de LAN inalámbrica, manejan hasta 54 Mbps, en tanto los entornos LAN alambrados ya se operan en los 100 Mbps, y se siguen buscando mejoras, por otro lado es posible ubicar la transmisión por fibra óptica, misma que maneja 1 Gbps). Al ser los servicios wireless más bien enfocados al entorno LAN, es necesario mencionar que tambien bridan QoS, mediante la diferenciación de tráfico (CoS).

Con esto, no se concluye que el cableado estructurado esté saliendo del mercado, ya que por su parte los fabricantes de cable, se abocan a hacer mejoras en las características del cobre para soportar mayores velocidades de transmisión.

Quienes se encuentran desarrollando entornos IP a velocidades más grandes, son las empresas de carriers, muchas de ellas ya se encuentran brindando servicios de telefonía inalámbrica a precios accesibles para empresas y muy probablemente más adelante los objetivos de estas empresas sea el mercado público. Hacia el 2001, los ingresos obtenidos por las ventas de gateways se estimaron en mil 800 millones de dólares; y se calcularon, para este mismo año, que la cantidad de minutos de telefonía sobre IP podría llegar a los 12 mil 500 millones. Hacia el 2010 se estima que un 25 por ciento de las llamadas telefónicas en todo el mundo será efectuado sobre redes basadas en IP.

Por otro lado, se desarrollan más y mejores herramientas de seguridad, mediante la encriptación de datos, desarrollo de políticas, detección temprana de virus, lo que permite que cada vez sea menos probable las intromisiones inesperadas de individuos que se divierten intentando acceder a nuestra red, ya sea con fines destructivos o simplemente para comprobar qué tan vulnerable puede ser la misma.

También, se ha venido manejando la implementación de canales lógicos a través de Internet denominadas VPN (Virtual Private Networks), haciendo permisible la interacción de dos o más sitios empleando la conexión a Internet, emulando una red WAN (dentro de estos entornos se pueden ubicar protocolos como MPLS y L2TP, entre otros).

Haciendo uso de la conjunción de estas tecnologías, se puede obtener el manejo de telefonía IP dentro de una empresa trasnacional o a nivel mundial, obteniendo grandes ahorros en telefonía de larga distancia. También se logra hacer uso de las aplicaciones como correo electrónico, manejo de extensión

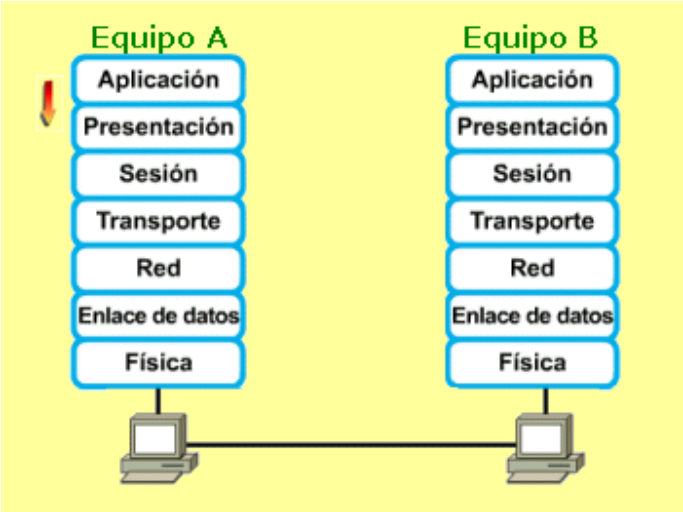
telefónica (utilizando softphone), y de esta manera se puede acceder a nuestros recursos de red desde cualquier parte del mundo, simplemente contando con el acceso a Internet.

Ahora bien el desarrollo tecnológico ya no se vé enfocado solamente a la industria, ya se comienza a ver a los distribuidores de televisión por paga brindar acceso a Internet, además de su acostumbrado servicio, también se logra contar ya con servicio de Internet inalámbrico, manejando servicios con anchos de banda lejos del acostumbrado acceso a 56 kbps (en el mejor de los casos) brindado por los MODEM y la línea telefónica, con esta apertura de mercado también poco a poco se deja ver el servicio telefónico de larga distancia mundial a costos exageradamente bajos, mismos que dejan buen dividendo a quienes se disponen a manejar el servicio telefónico por medio de Internet.

Con esto, se puede concluir que poco a poco la tecnología va llegando hasta nuestro hogar, y las escenas de películas de hace 5 o 10 años, en donde ya se vislumbra una videoconferencia, o de aquel individuo que habla por teléfono desde su reloj de pulsera, etc., ya se podrían vivir día con día.

ANEXO A: MODELO DE 7 CAPAS

Para poder simplificar el estudio y la implementación de la arquitectura necesaria, la ISO dividió el modelo de referencia OSI en capas, entendiéndose por "capa" una entidad que realiza de por sí una función específica. Cada capa, define los procedimientos y las reglas (protocolos normalizados) que los subsistemas de comunicaciones deben seguir, para poder comunicarse con sus procesos correspondientes de los otros sistemas. Esto permite que un proceso que se ejecuta en una computadora, pueda comunicarse con un proceso similar en otra computadora, si tienen implementados los mismos protocolos de comunicaciones de capas OSI.



Los criterios que llevaron a este modelo de referencia fueron:

- deberá crearse una nueva capa siempre que se precise un nuevo grado de abstracción.
- a cada capa deberá asignarse un número bien definido de funciones propias.
- la funcionalidad de cada capa deberá tener en cuenta la posibilidad de definir protocolos normalizados a nivel internacional.

- la frontera de las capas será tal que se minimice el flujo de información a través de la interfaz (especie de pasarela de comunicación entre ellas).
- el número de capas será lo suficientemente grande como para no reunir en un nivel funcionalidades distintas y lo suficientemente pequeño para que el resultado final sea manejable en la práctica.

En el modelo de referencia OSI, hay siete capas numeradas, cada una de las cuales ilustra una función de red particular. La división de la red en siete capas presenta las siguientes ventajas:

- divide la comunicación de red en partes más pequeñas y sencillas.
- normaliza los componentes de red para permitir el desarrollo y el soporte de los productos de diferentes fabricantes.
- permite a los distintos tipos de hardware y software de red comunicarse entre sí de una forma totalmente definida.
- impide que los cambios en una capa puedan afectar las demás capas, de manera que se puedan desarrollar con más rapidez.
- divide la comunicación de red en partes más pequeñas para simplificar el aprendizaje.

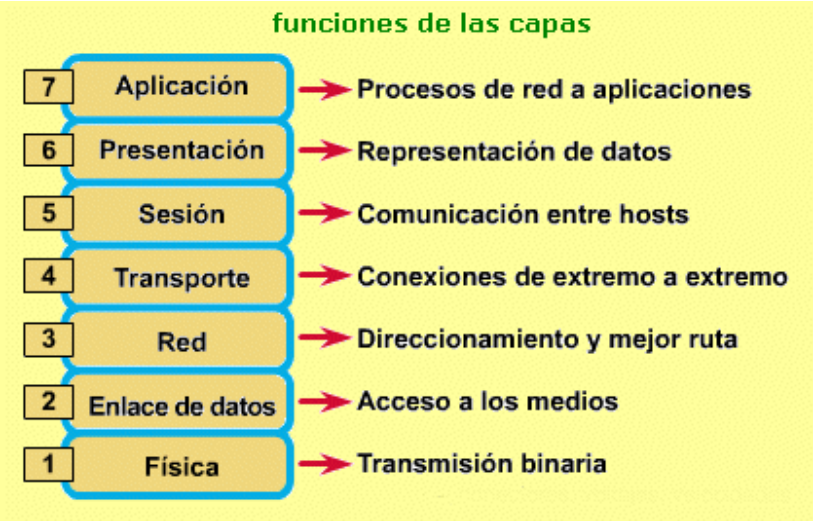
Las siete capas que plantea el modelo de referencia OSI, son las siguientes:

Cada capa individual del modelo OSI tiene un conjunto de funciones que debe realizar para que los paquetes de datos puedan viajar en la red desde el origen hasta el destino.

Las cuatro capas inferiores (Física, de Enlace de Datos, de Red y de Transporte) se van a encargar de la transmisión de los datos (segmentación, empaquetamiento, enrutamiento, verificación y transmisión por los medios físicos), sin importarles el tipo de datos que se transmiten ni la aplicación que los envía o recibe.

Por su parte, las tres capas superiores (de Sesión, de Presentación y de Aplicación) se encargan del establecimiento de sesiones de comunicación entre aplicaciones, del formateo, cifrado y compresión de datos y de suministrar los mismos a las aplicaciones de usuario de forma adecuada.

A continuación, se describen las funciones de cada capa en el modelo OSI.



Capa 7: ***La capa de aplicación***: Es la capa del modelo OSI más cercana al usuario, y está relacionada con las funciones de más alto nivel que proporcionan soporte a las aplicaciones o actividades del sistema, suministrando servicios de red a las aplicaciones del usuario y definiendo los protocolos usados por las aplicaciones individuales. Es el medio por el cual, los procesos de aplicación de usuario acceden al entorno OSI.

Su función principal es proporcionar los procedimientos precisos que permitan a los usuarios ejecutar los comandos relativos a sus propias aplicaciones.

Los procesos de las aplicaciones se comunican entre sí por medio de las entidades de aplicación asociadas, estando éstas controladas por protocolos de aplicación, y utilizando los servicios del nivel de presentación.

Difiere de las demás capas debido a que no proporciona servicios a ninguna otra capa OSI, sino solamente a aplicaciones que se encuentran fuera del modelo OSI. La capa de aplicación, establece la disponibilidad de los diversos elementos que deben participar en la comunicación, sincroniza las aplicaciones que cooperan entre sí y establece acuerdos sobre los procedimientos de recuperación de errores y control de la integridad de los datos.

Algunos ejemplos de procesos de aplicación son:

- programas de hojas de cálculo.
- programas de procesamiento de texto.
- transferencia de archivos (ftp).
- login remoto (rlogin, telnet).
- correo electrónico (mail - smtp).
- páginas web (http).

Capa 6: ***La capa de presentación:*** Proporciona sus servicios a la capa de aplicación, garantizando que la información que envía la capa de aplicación de un sistema pueda ser entendida y utilizada por la capa de aplicación de otro, estableciendo el contexto sintáctico del diálogo. Su tarea principal es aislar a las capas inferiores del formato de los datos de la aplicación, transformando los formatos particulares (ASCII, EBCDIC, etc.) en un formato común de red.

Es también las responsable de la obtención y de la liberalización de la conexión de sesión cuando existan varias alternativas disponibles.

Por ello, de ser necesario, la capa de presentación realiza las siguientes operaciones:

- traducir entre varios formatos de datos utilizando un formato común, estableciendo la sintaxis y la semántica de la información transmitida. Para ello convierte los datos desde el formato local al estándar de red y viceversa.
- definir la estructura de los datos a transmitir. Por ejemplo, en el caso de un acceso a base de datos, definir el orden de transmisión y la estructura de los registros.
- definir el código a usar para representar una cadena de caracteres (ASCII, EBCDIC, etc).
- dar formato a la información para visualizarla o imprimirla.
- comprimir los datos si es necesario.
- aplicar a los datos procesos criptográficos.

Capa 5: ***La capa de sesión:*** Esta capa, proporciona sus servicios a la capa de presentación, proporcionando el medio requerido para que las entidades de presentación en cooperación organicen y sincronicen su diálogo y procedan al intercambio de datos.

Sus principales funciones son:

- establece, administra y finaliza las sesiones entre dos hosts que se están comunicando.
- si por algún motivo una sesión falla por cualquier causa ajena al usuario, esta capa restaura la sesión a partir de un punto seguro y sin perdida de datos o si esto no es posible termina la sesión de una manera ordenada checando y recuperando todas sus funciones, evitando problemas en sistemas transaccionales.

- sincroniza el diálogo entre las capas de presentación de los dos hosts y administra su intercambio de datos, estableciendo las reglas o protocolos para el dialogo entre maquinas y así poder regular quien habla y por cuanto tiempo o si hablan en forma alterna, es decir, las reglas del dialogo que son acordadas.
- ofrece disposiciones para una eficiente transferencia de datos, clase de servicio y un registro de excepciones acerca de los problemas de la capa de sesión, presentación y aplicación.
- manejar tokens. Los tokens son objetos abstractos y únicos que se usan para controlar las acciones de los participantes en la comunicación.
- hacer checkpoints, que son puntos de recuerdo en la transferencia de datos.

Capa 4: ***La capa de transporte:*** Esta capa proporciona sus servicios a la capa de sesión, efectuando la transferencia de datos entre dos entidades de sesión. Para ello segmenta los datos originados en el host emisor y los reensambla en una corriente de datos dentro del sistema del host receptor.

El límite entre la capa de sesión y la capa de transporte puede imaginarse como el límite entre los protocolos de capa de medios y los protocolos de capa de host. Mientras que las capas de aplicación, presentación y sesión están relacionadas con aspectos de las aplicaciones, las tres capas inferiores se encargan del transporte de datos. Además, esta capa es la primera que se comunica directamente con su par de destino, ya que la comunicación de las capas anteriores es de tipo máquina a máquina.

La capa de transporte intenta suministrar un servicio de transporte de datos que aísla las capas superiores de los detalles de implementación del transporte, liberándolas de luchar por conseguir una transferencia de datos segura y económica. Específicamente, temas como la confiabilidad del transporte entre dos hosts es responsabilidad de la capa de transporte. Al proporcionar un servicio de comunicaciones, la capa de transporte establece, mantiene y termina adecuadamente los circuitos virtuales. Al proporcionar un servicio confiable, se utilizan dispositivos de detección y recuperación de errores de transporte.

Se conocen con el nombre de circuitos virtuales a las conexiones que se establecen dentro de una subred, y en ellos no hay la necesidad de tener que elegir una ruta nueva para cada paquete, ya que cuando se inicia la conexión se determina una ruta de la fuente al destino, ruta que es usada para todo el tráfico posterior.

Por último, es posible resumir las funciones de la capa de transporte, en los siguientes puntos:

- controla la interacción entre procesos usuarios.
- incluye controles de integración entre usuarios de la red para prevenir perdidas o doble procesamiento de transmisiones.
- controla el flujo de transacciones y direccionamiento de máquinas a procesos de usuario.
- asegura que se reciban todos los datos y en el orden adecuado, realizando un control de extremo a extremo.
- acepta los datos del nivel de sesión, fragmentándolos en unidades más pequeñas, llamadas segmentos, en caso necesario y los pasa al nivel de red.
- realiza funciones de control y numeración de unidades de información, fragmentación y reensamblaje de mensajes.

- se encarga de garantizar la transferencia de información a través de la subred.

Capa 3: ***La capa de red:*** La capa de red proporciona sus servicios a la capa de transporte, siendo una capa compleja que proporciona conectividad y selección de ruta entre dos sistemas de hosts que pueden estar ubicados en redes geográficamente distintas. También se ocupa de aspectos de contabilidad de paquetes.

Es la responsable de las funciones de conmutación y encaminamiento de la información, proporcionando los procedimientos precisos necesarios para el intercambio de datos entre el origen y el destino, por lo que es necesario que conozca la topología de la red, con objeto de determinar la ruta más adecuada.

Se pueden resumir las funciones de la capa de red en los siguientes puntos:

- divide los mensajes de la capa de transporte en unidades más complejas, denominadas paquetes, y los ensambla al final.
- debe conocer la topología de la subred y manejar el caso en que la fuente y el destino están en redes distintas.
- para ello, se encarga de encaminar la información a través de la subred, mirando las direcciones del paquete para determinar los métodos de conmutación y enrutamiento, y rutea los paquetes de la fuente al destino a través de ruteadores intermedios.
- envía los paquetes de nodo a nodo usando ya sea un circuito virtual o como datagramas.
- debe controlar la congestión de la subred.

En esta capa es donde trabajan los routers.

Capa 2: ***La capa de enlace de datos:*** La capa de enlace proporciona sus servicios a la capa de red, suministrando un tránsito de datos confiable a través de un enlace físico. Al hacerlo, la capa de enlace de datos se ocupa del direccionamiento físico (comparado con el lógico), la topología de red, el acceso a la red, la notificación de errores, formación y entrega ordenada de tramas y control de flujo. Por lo tanto, su principal misión es convertir el medio de transmisión en un medio libre de errores de cualquier tipo.

Sus principales funciones, son:

- establece los medios necesarios para una comunicación confiable y eficiente entre dos máquinas en red.
- agrega una secuencia especial de bits al principio y al final del flujo inicial de bits de los paquetes, estructurando este flujo bajo un formato predefinido llamado trama o marco. Suelen ser de unos cientos de bytes.
- sincroniza el envío de las tramas, transfiriéndolas de una forma confiable libre de errores. Para detectar y controlar los errores se añaden bits de paridad, se usan CRC (Códigos Cíclicos Redundantes) y envío de acuses de recibos positivos y negativos, y para evitar tramas repetidas se usan números de secuencia en ellas.
- envía los paquetes de nodo a nodo usando ya sea un circuito virtual o como datagramas.
- controla la congestión de la red.
- regula la velocidad de tráfico de datos.
- controla el flujo de tramas mediante protocolos que prohíben que el remitente envíe tramas sin la autorización explícita del receptor, sincronizando así su emisión y recepción.
- se encarga de la de secuencia, de enlace lógico y de acceso al medio (soportes físicos de la red).

Capa 1: ***La capa física:*** La misión principal de esta capa, es transmitir bits por un canal de comunicación, de manera que cuanto envíe el emisor llegue sin alteración al receptor.

La capa física proporciona sus servicios a la capa de enlace de datos, definiendo las especificaciones eléctricas, mecánicas, de procedimiento y funcionales para activar, mantener y desactivar el enlace físico entre sistemas finales, relacionando la agrupación de circuitos físicos a través de los cuales, los bits son movidos.

Las características tales como niveles de voltaje, temporización de cambios de voltaje, velocidad de datos físicos, distancias de transmisión máximas, conectores físicos y otros atributos similares se definen a través de las especificaciones de la capa física.

Sus principales funciones se pueden resumir en:

- Definir las características físicas (componentes y conectores mecánicos) y eléctricas (niveles de tensión).
- Definir las características funcionales de la interfaz (establecimiento, mantenimiento y liberación del enlace físico).
- Transmitir el flujo de bits a través del medio. No existe estructura alguna.
- Maneja voltajes y pulsos eléctricos.
- Especificar cables, conectores y componentes de interfaz con el medio de transmisión, polos en un enchufe, etc.
- Garantizar la conexión, pero no la fiabilidad de ésta.

Esta capa, solamente reconoce bits individuales, no reconoce caracteres ni tramas multicaracter.

ANEXO B: MULTIPLEXAJE POR DIVISION EN EL TIEMPO (TDM / MDT)

Es un método, para la transmisión de señales. Para entender éste concepto se partirá del significado de multiplexación:

Multiplexación es un proceso que permite enviar por un mismo medio de transmisión (Ej. pares de cables de cobre) muchas informaciones de diferentes orígenes. (Ej. Voz, transmisión de datos, telegrafía), por División en el Tiempo significa que el medio de transmisión es compartido o multiplexado en el tiempo es decir, que en unas porciones iguales de tiempo se introduce en cada una de ellas señales de diferentes orígenes con el objeto de optimizar el uso del medio de transmisión.

HISTORIA:

Para tener una visión del origen de PCM/MIC es necesario remontarse a finales del siglo pasado, cuando se establecieron las primeras raíces de dos técnicas que se pueden considerar como las bases que más tarde hicieron posible el desarrollo de PCM/ MIC. Una de ellas es la técnica de Múltiple por División en el Tiempo (TDM /MDT) 1853.

Esta técnica, fúe aplicada en telegrafía luego en 1903 W.N.Miner la aplico en telefonía, y los resultados fueron satisfactorios para la transmisión de señales de voz muestreadas a una velocidad de 3500 - 4300 muestras /segundo, pero a velocidades menores y mayores que éstas, no se logró ningún resultado aceptable.

La otra técnica es la Transmisión de Señales Digitales, la cual ha sido ampliamente aplicada en el campo de las telecomunicaciones, inicialmente con la incursión del primer telégrafo hacia mediados del siglo dieciocho, el cual enviaba señales digitales, en diferentes códigos, ej. El código Morse. Sin embargo para esa época todavía era una incógnita el uso de señales digitales para la transmisión de conversaciones en la red telefónica.

A finales de la década de 1930 ya se conocía El Teorema de Muestreo y la técnica de Multiplexación por División en el Tiempo pero el interés primordial era aplicar técnicas digitales para la transmisión de señales análogas por la ventaja de inmunidad al ruido y las descritas para señales digitales.

Casualmente, por ésta misma época un grupo de investigadores en París (Francia) estaba tratando de hallar métodos de modulación adecuados para los enlaces de microondas y así tratar de solucionar un problema latente de ruido y distorsión en éste tipo de transmisión.

Uno de los resultados obtenidos por éste grupo fue la invención de la Modulación por Impulsos Codificados (PCM / MIC) por ALEC H. REEVES EN 1937, la patente francesa se registro en 1938.

Tecnológicamente, era entonces demasiado temprano para usar PCM en la práctica debido a que no se contaban con instrumentos y equipos adecuados para ser aplicada esta nueva técnica, pero con la invención del transistor fue factible colocar en servicio una cantidad creciente de sistemas PCM en la red telefónica a comienzo de los años sesenta.

Con REEVES, se establecieron claramente los tres procesos básicos que permite la conversión de señales análogas en digitales (PCM/MIC).

MUESTREO: Generación de una señal PAM

CUANTIFICACION: Determina el valor de las muestras PAM

CODIFICACION: Representación de la muestras mediante un código numérico generalmente binario.

La gran ventaja de la técnica PCM/MIC radica en la generación de "unos" y "ceros" iguales a los de la información original en cada punto de regeneración a lo largo de la red de transmisión.

Dada la importancia de la técnica PCM/MIC, los organismos internacionales competentes han establecido normas de aplicación de amplia aceptación para el desarrollo de equipos y sistemas basados en PCM/ MIC.

ANEXO C: MODELO JERÁRQUICO DE CISCO

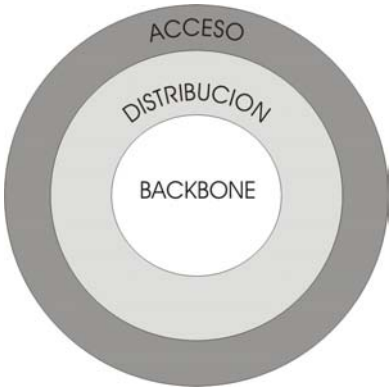
Este modelo se presenta en 3 capas, y es conocido como Nivel Jerárquico de 3 Capas, propuesto por Cisco y que a su vez, es aplicable conforme la distribución y estructura de la institución. De esta manera, se cuenta con los siguientes niveles:

- **Nivel Backbone**.- éste es el núcleo de la red, y es encargado de transportar grandes cantidades de tráfico con rapidez y confiabilidad, con una topología en anillo doblemente ligado.

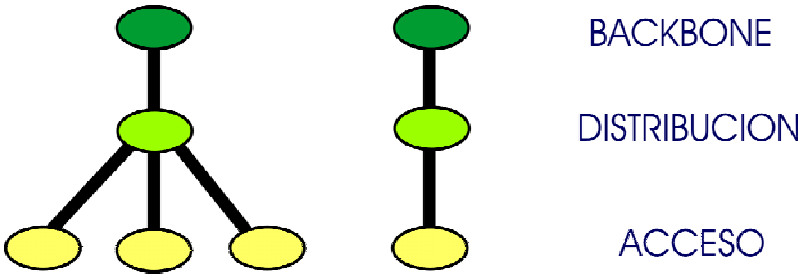
- **Nivel de Distribución**.- es el punto de comunicación entre los niveles de backbone y acceso, y la principal función de este nivel es proveer el ruteo, filtrado y determinar qué paquetes podrán acceder al backbone, en caso de ser necesario. A este nivel se usa una topología en estrella.

- **Nivel de acceso**.- en este nivel, se controla el acceso a los recursos de la red y usuarios. Una de las funciones de éste es la creación de dominios separados de colisiones, mejor conocido como *segmentación*. Utilizando en este último nivel una topología en árbol extendido.

  En la siguiente figura, se representa el Nivel Jerárquico de 3 Capas.



Haciendo la representación esquemática de este modelo, queda de la siguiente forma.



En el nivel de Backbone se ubican 3 edificios principales, los cuales albergan gran parte de los servicios y a su vez un gran número de usuarios.

En el de Distribución se ubican las Fiscalías, éstas, se encuentran en cada una de las delegaciones políticas del Distrito Federal. Y en el nivel de Acceso se encuentran Coordinaciones Territoriales, las cuales se localizan en distintas zonas, dentro de la demarcación correspondiente.

ANEXO D: EQUIPOS DE DATOS LAN Y WAN

Los equipos utilizados en la infraestructura de red de datos son:

- *Routers*: marca Cisco, diferentes modelos, dependiendo de las necesidades del sitio.
- *Switches*: marcas Cisco, 3Com y Allied Telesyn. diferentes modelos dependiendo de las necesidades del sitio.
- *Hubs*: diferentes modelos de las marcas 3Com y Allied Telesyn.

HUBS

Es un dispositivo que provee un punto común de conexión física para los dispositivos de red. Los hubs, son repetidores multipuertos.

Un hub recibe una señal digital y la amplifica o regenera después de lo cual envía la señal a todos los puertos activos excepto al puerto del cual la señal, es recibida. Esto significa que todos los dispositivos (host) conectados al hub están en el mismo *dominio de colisión* y en el mismo *dominio de broadcast*.

Los hubs no "ven" o analizan ningún tráfico que entra y que transmiten. De aquí su principal desventaja, no pueden filtrar tráfico de la red. El filtrado, normalmente se refiere a un proceso o a un dispositivo que protege el tráfico de red con ciertas características, como el equipo de orígen, el equipo de destino y tampoco determina si acelera o desecha la información basándose en criterios establecidos.

SWITCHES LAN

Los switches LAN o simplemente switches representan una tecnología que alivia la congestión en las redes LAN ethernet segmentándolas en varios dominios de colisión. Estos suelen reemplazar a los hubs en los medios compartidos, y trabajan con las infraestructuras de cableado existentes para asegurar que están instaladas con una alteración mínima de las redes instaladas.

Al switch, al igual que al hub se le llama repetidor multipuerto. La diferencia entre el hub y el switch es que los switches toman decisiones basándose en las direcciones MAC y los hubs, no toman decisiones, por ningún concepto.

Gracias a las decisiones que toman los switches, las LAN son mucho más eficientes esto, lo consiguen "conmutando" los datos fuera del puerto al que el propio host está conectado. Por su parte un hub envía los datos a todos sus puertos para que todos los host tengan que ver y procesar (aceptar o rechazar) todos los datos.

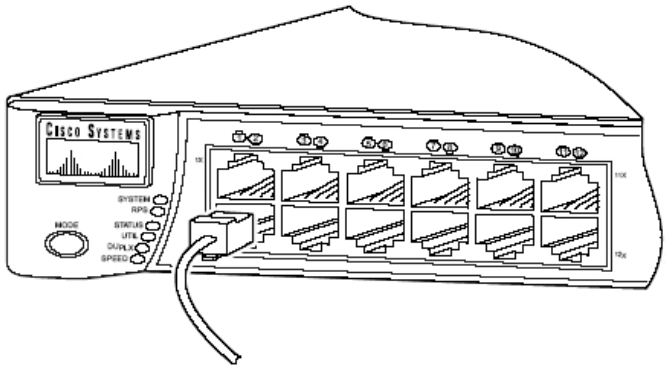
La figura siguiente, muestra el símbolo de un switch. Las flechas de la parte superior representan los datos de rutas separadas que puede haber en un switch.

Símbolo de un switch.

Los switches, usan una tabla de direcciones MAC para decidir el segmento que van a necesitar para transmitir una *trama* y reducen el tráfico de cada

segmento. Los switches operan, a velocidades mucho más altas que los hubs y pueden soportar nuevas funcionalidades como las LAN virtuales.

Aunque un switch LAN reduce el tamaño del dominio de colisión, todos los host conectados con el switch están todavía en el mismo dominio de difusión o broadcast. Por tanto, una difusión desde un nodo sería todavía visible por los otros nodos conectados mediante el switch LAN.



Estos equipos, pertenecen a la capa de enlace del modelo OSI, descrito en el Anexo A.

ROUTERS

La función principal de una red de comunicaciones es la de llevar la información desde el origen hasta el destino. En redes complejas llegar desde el origen hasta el destino puede requerir elegir entre múltiples en ocasiones grandes cantidades de trayectorias. Para lograr este objetivo se debe contar con un dispositivo que se va a encargar de escoger las trayectorias adecuadas a través de la red, el cuál se va a encargar también de seleccionar las rutas a modo de evitar la carga extra de algunas de las líneas de comunicación.

A esta función realizada por un dispositivo de comunicaciones, se le llama enrutamiento y al dispositivo que la realiza se le llama enrutador o router y son equipos que hacen uso de un algoritmo de enrutamiento definido como el software

encargado de decidir la línea de salida por la que se transmitirá la información de entrada.

Con frecuencia, se piensa que los routers son dispositivos de red de área amplia (WAN), pueden ser igualmente útiles en las redes de área local (LAN). Además, pueden implementar una gran variedad de valores añadidos que ayudan a mejorar la relación efectividad-costo de la red. Estas funciones, incluyen el tráfico secuenciado basado en la prioridad y el filtrado del tráfico. Los routers proporcionan interfaces para una amplia gama de enlaces y subredes con un amplio abanico de velocidades. Administran la red proporcionando control dinámico sobre los recursos, y dando respuesta a las tareas y objetivos de la red: conectividad, rendimiento fiable, control administrativo y flexibilidad

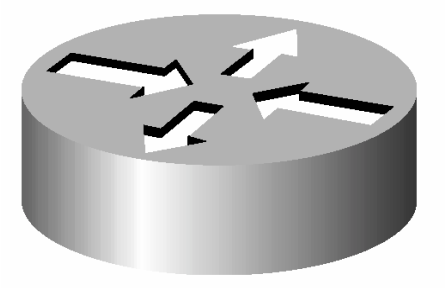
El propósito de un router es examinar los paquetes entrantes, elegir la mejor ruta para ellos a través de la red, y después conmutarlos al mejor puerto de salida.

Los routers, son el dispositivo regulador de tráfico más importante en las redes grandes como la Internet, permitiendo que cualquier tipo de computadora se comunique con otra en cualquier parte del mundo.

Los routers utilizan un esquema de direcciones diferente a las direcciones MAC para tomar decisiones y remitir los datos. Utilizan direcciones de un protocolo llamado IP (Internet Protocol / Protocolo de Internet) o direcciones lógicas. Almacenan la información de los segmentos de red que “aprenden” en una tabla de enrutamiento. Equiparan la información de la tabla de enrutamiento con las direcciones IP de destino de los datos, y envían los datos entrantes hacia la subred y host correctos. Como las direcciones IP se implementan en el software y hacen referencia a la red en la que está ubicado el dispositivo las direcciones, se llaman direcciones de red lógicas. Las direcciones MAC normalmente las asigna el fabricante de la NIC (Network Interface Card) y están codificadas en el interior

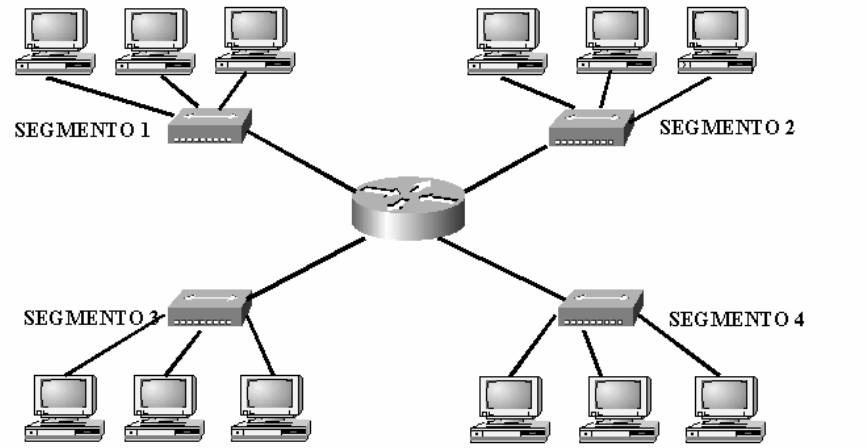
de dicha NIC, las direcciones de red o direcciones IP las asigna normalmente el administrador de la red.

El símbolo de un router, mostrado en la figura, sugiere sus dos propósitos principales: selección de rutas y conmutación de paquetes a la que sea mejor. Un router puede tener muchos tipos diferentes de puertos de interfaz.



Símbolo de un router.

Los routers crean un mayor nivel de segmentación debido a su capacidad para tomar decisiones exactas sobre donde debe enviar los paquetes de datos como consecuencia operan a una mayor tasa de latencia. La figura siguiente muestra un router que se emplea para segmentar una LAN.



Segmentación utilizando un router.

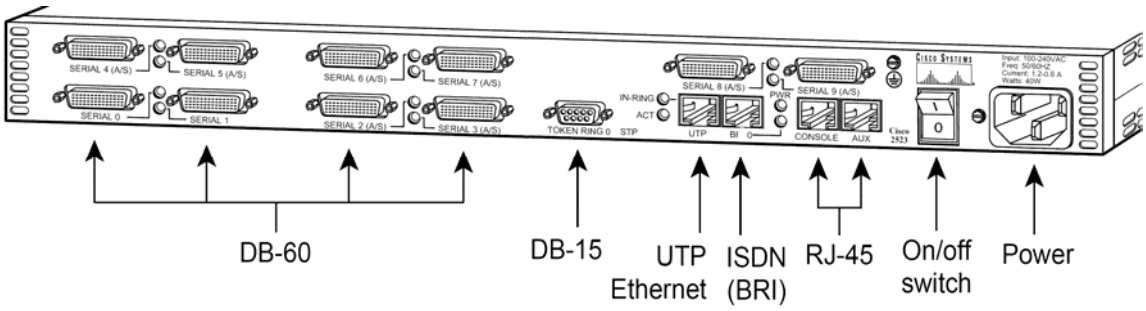
Se distinguen algunas categorías de routers:

- De protocolo simple o multiprotocolo.- Los router´s de protocolo simple son aquellos que únicamente manejan un tipo de protocolo para realizar su función, así un router que maneja paquetes de información en formato IP no pueden manejar paquetes con formato IPX. Estos router´s, fueron muy comunes por muchos años. En el curso natural de la evolución tecnológica, los router´s expandieron sus capacidades con respecto a los protocolos soportados, así un router en la actualidad generalmente maneja mas de un protocolo. Se debe considerar que el costo de incrementar la capacidad trae como consecuencia la pósibilidad de decrementar el ancho de banda en el tráfico de un protocolo en particular. Esto es si un router necesita procesar por ejemplo paquetes con los protocolos IP e IPX necesitará compartir su tiempo disponible y capacidad entre estos dos protocolos.

- Central o periférico.- Un router, puede servir como punto de transferencia para múltiples redes. Por ejemplo, cada red puede ser conectada a una diferente tarjeta en un servidor o un hub. Estos routers centrales están hasta arriba en el precio y rango de capacidades y usualmente son routers multiprotocolo. En contraste, un router periférico sirve primordialmente para conectar una red hacia una gran red. Estos routers, son mas baratos y de menor desempeño. Un router periférico puede limitarse a un solo protocolo.

- Router LAN o Router WAN.- Otro punto común de la evolución tecnológica es extender las capacidades de un dispositivo a través de grandes distancias. En este contexto, los routers WAN, de los cuales, su trabajo es encontrar trayectorias a través de redes ampliamente distribuidas son extensiones de los routers LAN los

cuáles conectan LAN´s que están distribuidas a través de áreas pequeñas y que permiten conexiones sin requerir líneas telefónicas. Los routers LAN y WAN hacen la misma tarea, pero los detalles de cómo realizan esta tarea varía considerablemente. Un router WAN necesita soportar protocolos adecuados para acceso y servicio a grandes distancias. En parte el desarrollo de los routers WAN ha sido esperado por las líneas de comunicaciones con suficiente ancho de banda para hacer factible dicho ruteo.

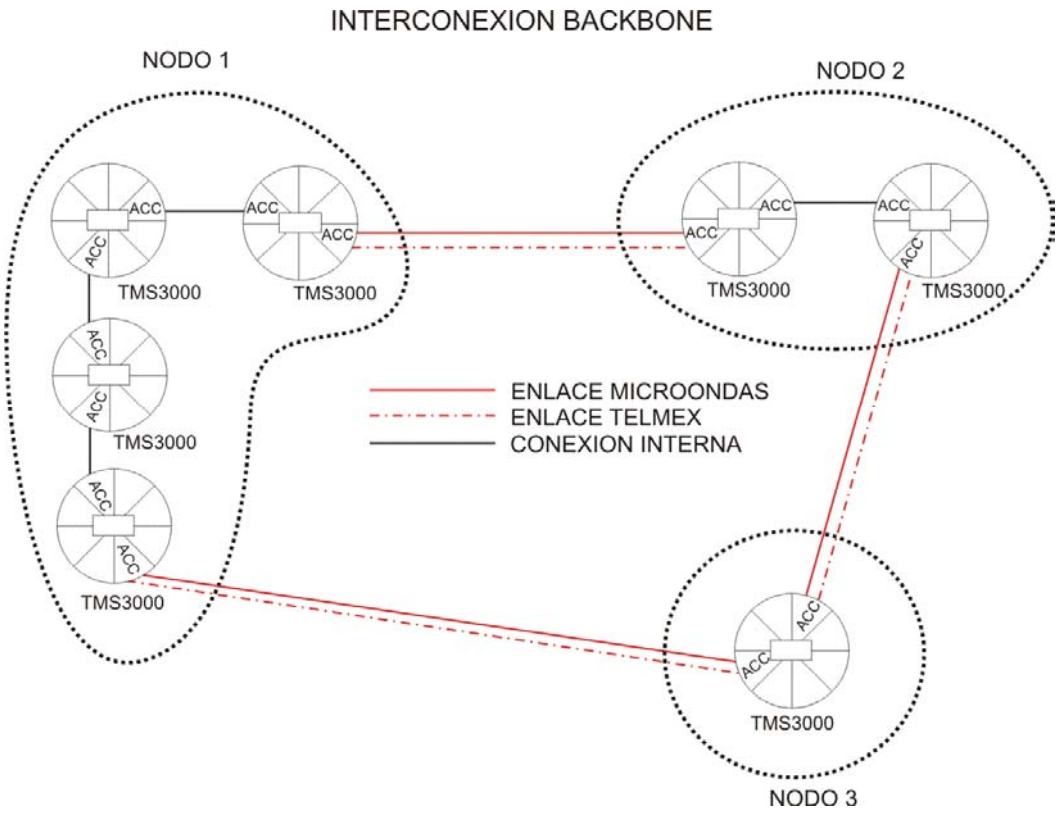


Estos equipos trabajan en la capa 3 del modelo OSI descrito en el Anexo A.

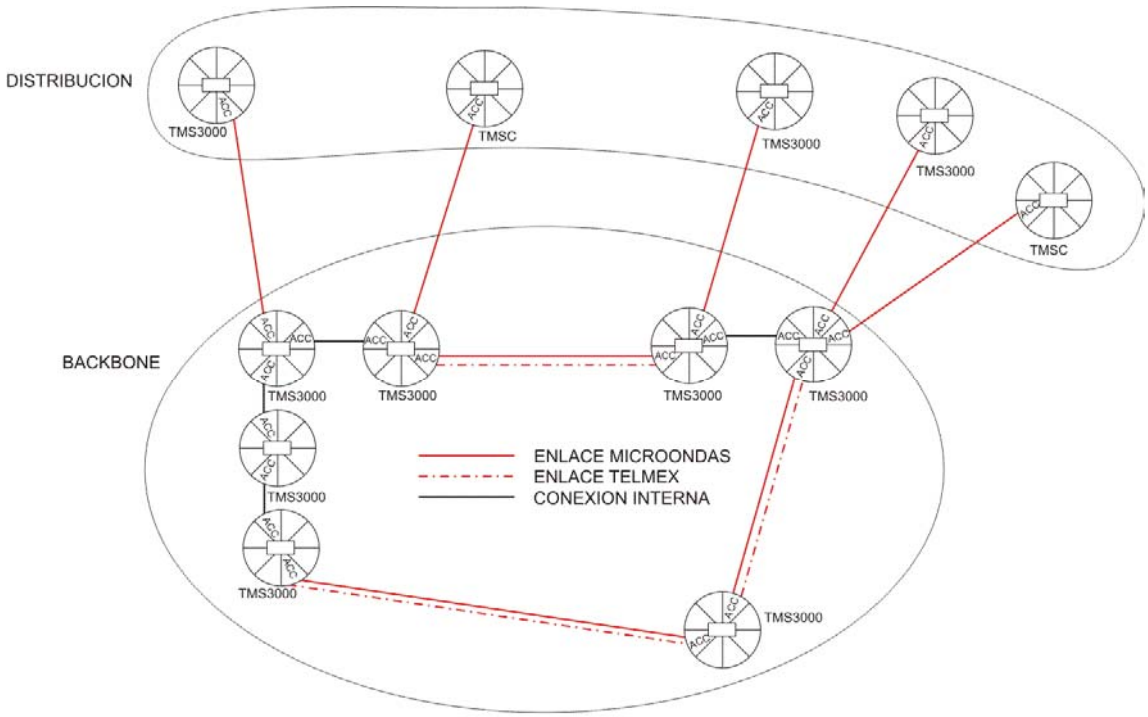
ANEXO E: OPERACIÓN CON TDM

Para abordar la descripción del modo de operación de la red con la tecnología de transporte TDM (Time Division Multiplexing), se hará de la capa de Backbone hacia la capa de Acceso, describiendo la simbología utilizada.

Inicialmente, se utilizan multiplexores, el cual se interconecta a otro del mismo modelo, de esta forma se puede identificar uno o varios multiplexores en un nodo del Backbone, y posteriormente, dependiendo de la cantidad de enlaces hacia la capa de distribución, determinará la cantidad de equipamiento destinado para la conexión de nodos de Distribución. En el siguiente esquema se muestra la interconexión de los multiplexores en la capa de backbone.

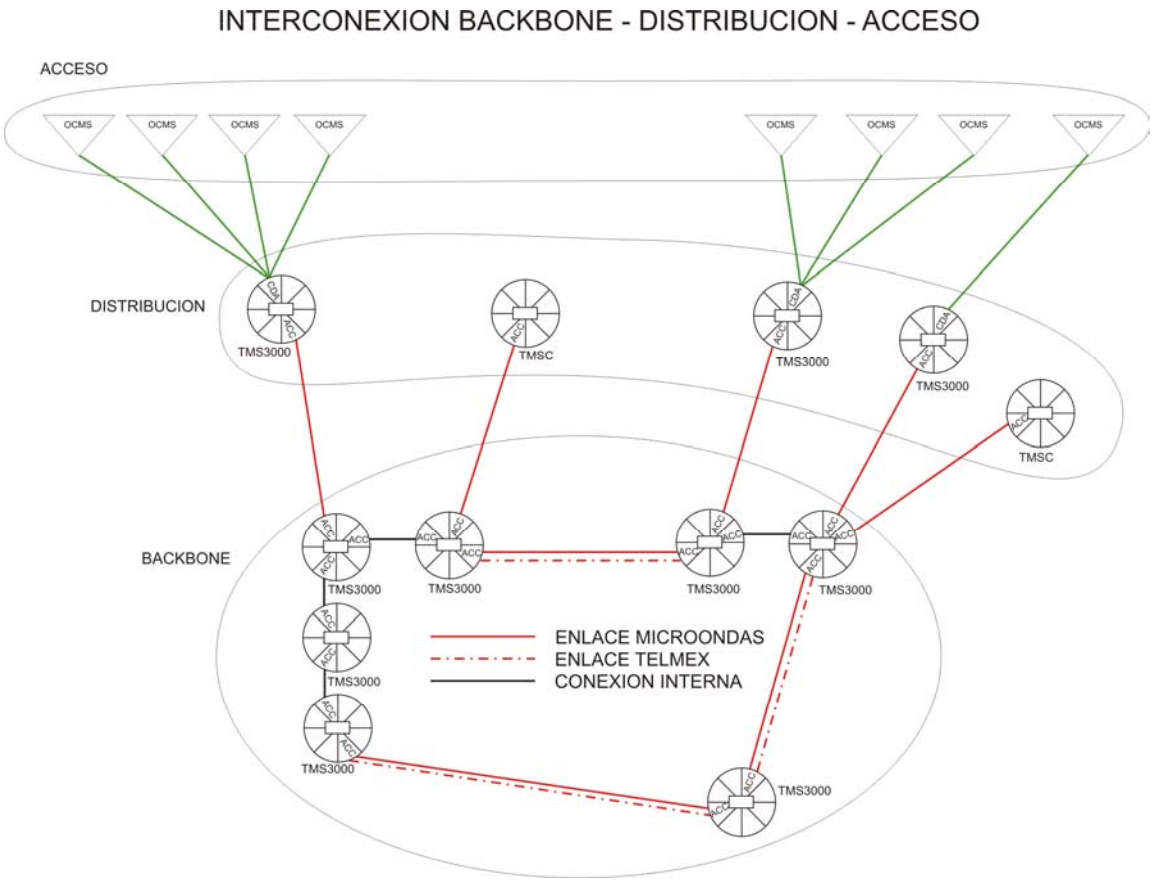


En los nodos de Distribución, también se encuentran instalados multiplexores, el cual cuenta con un equipamiento de características similares a las del nodo de Backbone, para recibir el enlace, a su vez cuenta con un modulo de expansión, en el cual, se proveen las conexiones físicas, con la función de transportar las extensiones telefónicas que se utilizarán en los diferentes nodos de Acceso.



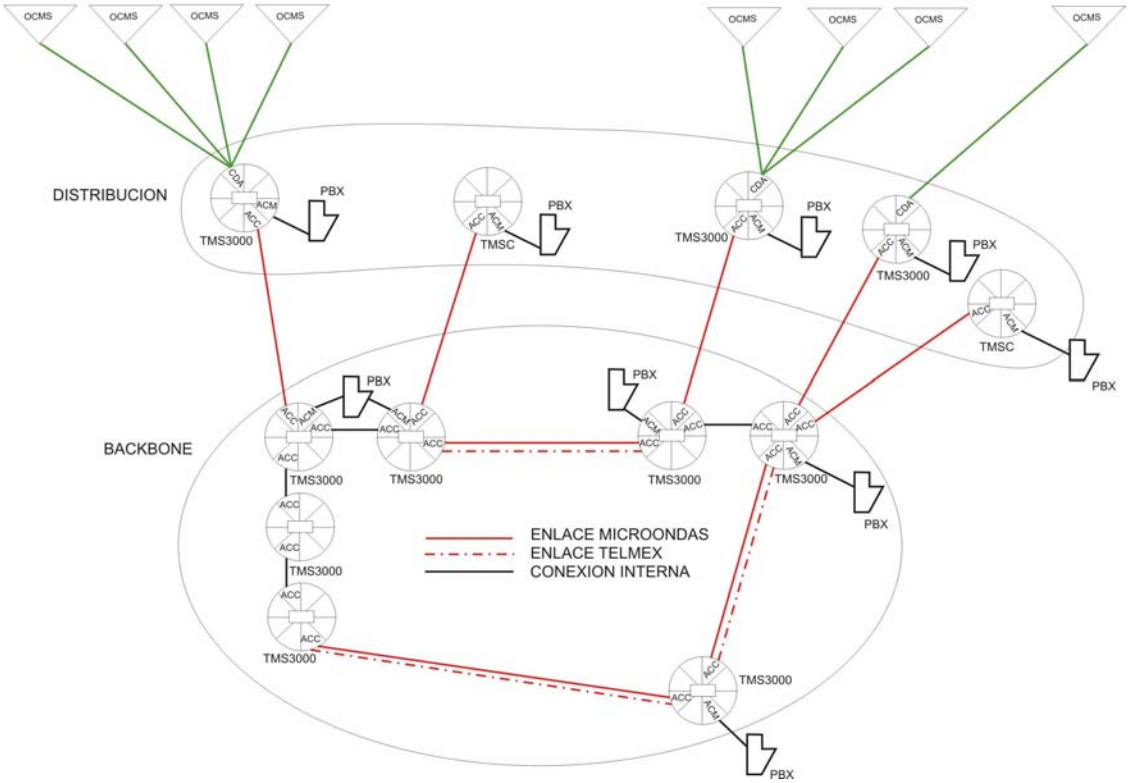
INTERCONEXION BACKBONE - DISTRIBUCION

Dentro del modulo de expansión, se encuentran también los puertos de conexión de datos, en donde se conectaran los puertos del router, y cabe mencionar que habrá tantos puertos de datos como nodos en la capa de Acceso tenga este nodo de Distribución.



Anteriormente, se contaba con un puerto específico para transportar el enlace de los conmutadores (PBX's), y con las diferentes etapas de reestructuración de la red de transporte y datos, se ha migrado la interconexión de los PBX de manera independiente con enlaces de microondas propios de la institución.

INTERCONEXION PBX



GLOSARIO

802.1P	Es un mecanismo de control del tráfico de acumulación apropiado para el uso en muchas redes de área local (LAN).
802.1D	Utilizado en el protocolo de árbol extensible, el estándar IEEE 802.1d es compatible con el puente de MAC para evitar bucles de red.
802.1Q	Define el funcionamiento de los puentes VLAN que permite definir, hacer funcionar y administrar VLAN dentro de las infraestructuras de LAN con puente
ANCHO DE BANDA	Es la medida de la capacidad de transmisión de datos, expresada generalmente en Kilobits por segundo (kbps) o en Megabits por segundo (Mbps). Indica la capacidad teórica de una conexión.
APPLE TALK	Protocolo diseñado para computadoras Apple.
BACKBONE	Es el núcleo de la red, y es encargado de transportar grandes cantidades de tráfico con rapidez y confiabilidad
BANDWIDTH	Véase ancho de Banda
BEST EFFORT	Mejor Esfuerzo; Es un algoritmos que no ofrecen ningún tipo de garantías de transmisión, por lo que podría decirse que el nivel de QoS ofrecido es nulo
BROADCAST	En inglés, difusión. Tipo de comunicación en la que un solo emisor llega a múltiples receptores.

BUFFERS	Área que recibe y almacena los datos esporádicamente, mientras se presenta alguna congestión.
CAC	Connection Admisión Control. Es la secuencia de acciones ejecutadas por un switch ATM para verificar si una petición de conexión es aceptada o rechazada, según los recursos disponibles.
CIR	Committed Information Rate. Es la tasa de información comprometida, es decir, el caudal medio garantizado que la red se compremete a dar en una conexión durante un intervalo de tiempo definido
CODEC DE AUDIO	Codifica la señal desde el equipo de audio para su transmisión y descodifica el código de audio entrante. Los codecs utilizados para voz son G.711, G.722, G.723, G.728, G.729.
COPS	Common Open Policy Service; Protocolo descrito en la RFC 2748, define un modelo cliente/servidor sencillo para proporcionar control de políticas a protocolos de señalización de calidad de servicio.
DATAGRAMA	Es un fragmento de paquete que es enviado con la suficiente información como para que la red pueda simplemente encaminar el fragmento hacia el ordenador receptor, de manera independiente a los fragmentos restantes.
DIRECCIONES MAC	**Dirección física de la tarjeta de red (NIC):** es un "número de serie" único asignado a cada dispositivo de red Ethernet para identificarse en la red. Esta dirección es única e irrepetible y es asignada al momento que de la fabricación del dispositivo, por el fabricante.

DLCI	Data Link Connection Identifier (Identificador de Conexión de enlace de Datos), Es utilizado para identificar circuitos virtuales en una red Frame Relay.
DS0	Digital Signal – 0, Canal de velocidad estándar (64 kbps, definidas para E1), para transporte de voz y datos.
E1	Es un enlace compuesto de 32 canales dúplex completo, constando cada uno de ellos de 64 kbps, que producen un total de 2,048 Mbps.
ETHERNET	Ethernet se estandariza como IEEE 802.3. Ethernet es el estándar de LAN implementado más común. Admite velocidades de transferencia de datos de Mpbs, compatibles con velocidades de 10, 100 ó 1000 Mbps.
FDDI	Fiber Distributed Data Interfase. Es un estándar LAN que maneja velocidades mayores a 200 Mbps en cables de fibra óptica.
FEC	Forwarding Equivalence Class. *):* Agrupación de paquetes que comparten los mismos atributos (dirección destino, VPN...) y/o requieren el mismo servicio.
FIFO (TEORIA DE COLAS)	*First Input First Output*: Los paquetes son transmitidos en el orden en el cual arribaron. Esto en una sola cola para todos los paquetes. Los paquetes son almacenados en la cola, cuando la red presente congestión y enviados cuando sea eliminada dicha congestión. Si la cola esta llena, los paquetes serán descartados.
FIREWALLS	Es una barrera entre una red privada y la red pública, encargada de permitir o negar el acceso a la red privada mediante listas de acceso y otros mecanismos para mantener la seguridad de la red privada.

FRAME RELAY	Es un protocolo de conmutación de paquetes de bajo costo para conectar dispositivos en redes de área amplia.
G.729	Codificador de voz de 8 Kbps utilizando Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP).
GIGABIT ETHERNET	Gigabit Ethernet transmite a 1000 Mbps y es compatible con los estándares Ethernet 10/100 Mbps existentes
H.323	Es una especificación ITU-T para transmitir audio, vídeo y datos a través de una red de Protocolo Internet (IP), incluida la propia Internet.
HOST	Es una máquina perteneciente a algún usuario que publica contenido accesible a través de Internet. También es conocido como servidor y existen máquinas dedicadas a los servicios de mail, noticias o http.
IEEE	Institute of Electrical and electronics Engineers. Es una organización profesional, que entre otras actividades, define estándares.
IP ADDRESS	También llamada Internet Address, es una dirección única que identifica cualquier dispositivo en Internet.
IPSEC	IP Security. Es un conjunto de protocolos desarrollados por la IETF para dar soporte al intercambio seguro de paquetes del lado del IP. El IPsec se implementó ampliamente en las Redes Privadas Virtuales (VPNs).
JITTER	Es la inestabilidad o variabilidad en el retardo.

L2TP L2TP es una extensión del protocolo Punto a Punto (PPP), el cual es utilizado en redes privadas virtuales (VPN's). L2TP soporta múltiples protocolos y administración de direcciones IP sobre Internet.

LATENCIA Es el tiempo entre el envió de un mensaje y la recepción del mismo en otra parte de la red. Abarca los retardos sufridos durante el propio camino o en los dispositivos por los que pasa.

LSP Label Switched Path. Es la ruta a través de uno o más LSRs en un nivel de jerarquía que sigue un paquete de un FEC en particular.

LSR Label Switched Router. Es el tipo de router que permite MPLS

MENSAJE PATH Contiene las especificaciones del tráfico (TSPEC), que perfila el flujo de datos que es enviado.

MENSAJE RESV Contiene la petición de reserva de recursos en el TSPEC, RSPEC con el nivel de QoS y FILTERSPEC para el descriptor del puerto.

MPLS MultiProtocol Label Switching. es una de las soluciones propuestas por el IETF, con el objetivo de proporcionar *QoS*.

MULTICAST Es el envío de la información a múltiples destinos simultáneamente usando la estrategia más eficiente para el envío del mensaje sobre cada enlace de la red únicamente una vez y creando copias cuando los enlaces en los destinos se dividen.

OSPF Open Shortest Path First, protocolo de encaminamiento -o enrutamiento- que abre primero el camino más corto a la hora de enviar paquetes.

PBX Private Branch Exchange. Es un switch telefonico conectado a la red telefonica pero operando del lado del usuario. Un PBX provee el acceso a gran numero de lineas internas (extensiones) y a pocas lineas externas (troncales). En donde, las llamadas salientes son marcadas directamente y las llamadas entrantes, son tomadas por un operador o switcheadas automaticamente por el software del PBX.

PRIORIZACIÓN Consiste en la asignación de un determinado nivel de QoS al tráfico que circula por una red, asegurando así que las aplicaciones de mayor importancia sean atendidas con anterioridad a las de menor importancia, estando o no ante una situación de congestión.

RETARDO Parámetro indica la variación temporal y/o retraso en la llegada de los datos a su destino

RFC Acrónimo inglés de Request For Comments. Conjunto de archivos de caracter técnico donde se describen los estándares o recomendaciones de cualquier cosa. Entre otros los de la propia Internet.

SLA	Service Level Agreement es un contrato de servicios entre un proveedor de servicios y su cliente, el cual define las responsabilidades del proveedor en términos del nivel del funcionamiento de la red (rendimiento, tasa de pérdidas, retrasos, variaciones) y la disponibilidad temporal.
SONET	Es una Red Óptica Síncrona, a través de los medios de fibra óptica.
THROUGHPUT	Es la cantidad de datos que se mueven en una cierta cantidad de tiempo. Normalmente es medido en kilobytes por segundo (Kbps) o megabytes por segundo (Mbsp)
TRAFFIC SHAPING	Conformado De Tráfico, es un mecanismo de control de flujo en una interfase determinada, reduciendo la circulación de salida para evitar la congestión
TSPEC	Describe el tráfico que la aplicación estima generar.
X.25	Es un protocolo de empaquetamiento conmutado, definido por Comité Consultivo de ITT y adoptado luego por ISO.

ACRONIMOS

API	Aplication Programming Interface.
ATM	Asynchronous Transfer Mode (Modo de Transferencia Asincrona**)**
CALL MANAGER	Administrador de Llamadas.
CBR	Constant Bit Rate
CIM	Common Information Model, Modelo Común de Información.
CoS	Class of Service
CQ	Custom Queueing. Mecanismo de control de Congestión.
DIFFSERV	Differentiated Services
DSBM	Designated SBM.
DSCP	DiffServ Codepoint.
E-COMMERCE	Comercio Electrónico
E-MAIL	Correo Electrónico
FTP	File Transfer Protocol

GBPS	Gigabit per Second
http	HiperText Transport Protocol
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
INTSERV	servicios integrados
IP	Internet Protocol
IPV6	IP version 6
IPX	Internetwork Packet Exchange
ISP	Internet Service Provider (Proveedor de Servicios de Internet)
KBPS	Kilobits per Second
LAN	Local Area Network
LDP	Label Distribution Protocol.
LLC	Logia Link Control.
LPDP	Local PDP.

MBPS	Megabits per Second
NIC	Network Interfase Card
OSI	Open Systems Interconnection.
PCV	Permanent Virtual Conexions
PDP	Policy Decisión Point.
PEP	Policy Enforcement Point.
PHB	Per Hop Behavior,
PPP	Point to Point Protocol
PQ	Priority Queuing. Mecanismo de control de Congestión
PSTN	Public Switched Telephone Network
QoS	Quality of Service (Calidad de Servicio)
RDSI	Red Digital de Servicios Integrados
RED	Random Early Dropping. Mecanismo de Prevención de Congestión
RSVP	Resource Reservation Protocol.

RTP	Real-Time Transport Protocol.
SBM	Subnet Bandwidth Manager.
SLO	Service Level Objectives, Servicio de Niveles Objetivos.
SVC	Switched Virtual Conection
TCP	Transmisión Control Protocol
TDM	Time Division Multiplexing (Multiplexación por División de Tiempo)
ToS	Type of Service (Tipo de Servicio)
TTL	Time To Live
UDP	User Datagram Protocol
UIT	Unión Internacional de Telecomunicaciones
UTP	Unshielded Twisted Pair
VCI	Virtual Circuit Identifier (Identificador de Circuito Virtual)
VLAN	Virtual LAN (Red de Area Local Virtual)
VOIP	Voice over Internet Protocol (Voz sobre IP)

VPI	Virtual Path Identifier (Identificador de Ruta Virtual)
VPN	Virtual Private Network (Red Privada Virtual)
WAN	Wide Area Network (Red de Area Amplia)
WFQ	Weighted Fair Queueing. Mecanismo de control de Congestión
WRED	Weighted Random Early Dropping. Mecanismo de prevención de Congestión
WRR	Weigthed Round Robin. Mecanismo de control de Congestión.
WWW	World Wide Web

BIBLIOGRAFÍA

- Wang, Zheng
  Internet QoS: Architectures and Mechanism for Quality of Service
  Morgan Kaufmann Publishers
  ISBN 1-55860-608-4
  2000

- García Tomás Jesús, Raya Cabrera J. Luis, Rodrigo Raya Víctor
  Alta Velocidad y Calidad de Servicio en Redes IP
  AlfaOmega – RAMA
  ISBN 84-7897-503-9
  2002

- Lammle, Todd
  Cisco
  Certified Network Associate, Study Guide
  SYBEX
  ISBN 0-7821-2647-2
  2000

- Black, Uyless
  Tecnologías emergentes para Redes de Computadoras
  Prentice Hall
  ISBN 970-17-0268-9
  1999

Davidson Jonathan, Peter James
Fundamentos de Voz sobre IP
Cisco Press
ISBN 84-205-3190-1
2001

Guijarro Coloma, Luís
Redes ATM, principios de Interconexión y su aplicación
AlfaOmega – RAMA
ISBN 970-15-0538-7
2000

Student Guide
Deploying Cisco QoS for Enterprises Networks
Curso DQOS
Vol. 1, Vol. 2
2001

Student Guide
Cisco IP Telephony
Curso CIPT
Vol. 1, Vol. 2
2001

REFERENCIAS WEB

- http://www.cisco.com
- http://www.ietf.org
- http://www.stardust.com
- http://www.avaya.com
- http://www.nortel.com
- http://www.alcatel.com
- http://www.idg.es
- http://www.microsoft.com
- http://.www.wikipedia.com
- http://www.opalsoft.com
- http://www.red.com.mx
- http://www.babilon.com