

**QUALITY OF SERVICE: CONCEPTO, HISTORIA, EVOLUCIÓN Y
TENDENCIAS**

HÉCTOR GONZALO ANZOLA GARCÍA

ÁLVARO DAVID MENDOZA GARCÍA

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA DE SISTEMAS

CARTAGENA DE INDIAS

2004

**QUALITY OF SERVICE: CONCEPTO, HISTORIA, EVOLUCIÓN Y
TENDENCIAS**

**ANZOLA G, HÉCTOR
MENDOZA G, ÁLVARO**

**Monografía, presentada para optar al título de
Ingeniero de Sistemas**

**Director
GONZALO GARZÓN
Ingeniero de Sistemas**

**UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
CARTAGENA DE INDIAS**

2004

Nota de aceptación

Presidente del Jurado

Jurado

Jurado

Cartagena, Junio del 2004

ARTICULO 107

La institución reservada el derecho de propiedad intelectual de todos los trabajos de grupo aprobados, los cuales no pueden ser explotados comercialmente sin su autorización. Esta observación debe quedar impresa en parte visible del proyecto.

Cartagena, mayo del 2004

Señores

Comité Curricular

Universidad Tecnológica de Bolívar

Ciudad

Apreciados Señores.

Cordialmente me permito informarles que he llevado a cabo la dirección del trabajo de grado de los estudiantes **HÉCTOR GONZALO ANZOLA GARCÍA** y **ÁLVARO DAVID MENDOZA GARCÍA**, titulado: **“QUALITY OF SERVICE: CONCEPTO, HISTORIA, EVOLUCIÓN Y TENDENCIAS”**

Cordialmente,

GONZALO GARZÓN

Cartagena, mayo del 2004

Señores

Comité Curricular

Universidad Tecnológica de Bolívar

Ciudad

De manera muy atenta nos permitimos presentar a su consideración y aprobación el trabajo de grado titulado **“QUALITY OF SERVICE: CONCEPTO, HISTORIA, EVOLUCIÓN Y TENDENCIAS”**. Elaborado por **HÉCTOR GONZALO ANZOLA GARCÍA** y **ÁLVARO DAVID MENDOZA GARCÍA**.

Esperamos que el presente trabajo se ajuste a las expectativas y criterios de la universidad en cuanto a trabajos de grado se refiere.

Cordialmente,

HÉCTOR GONZALO ANZOLA GARCÍA

ÁLVARO DAVID MENDOZA GARCÍA

AUTORIZACIÓN

Cartagena de Indias, D.T.C.H

Yo **HÉCTOR GONZALO ANZOLA GARCÍA** identificado con numero de cedula 15'170.488 de **VALLEDUPAR**, autorizo a la **UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR** para hacer uso de mi trabajo de grado y publicarlo en el catalogo online de la biblioteca.

FIRMA

Debe registrarse esta autorización ante notario público

AUTORIZACIÓN

Cartagena de Indias, D.T.C.H

Yo **ÁLVARO DAVID MENDOZA GARCÍA** identificado con numero de cedula 9'236.779 de **CARTAGENA**, autorizo a la **UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR** para hacer uso de mi trabajo de grado y publicarlo en el catalogo online de la biblioteca.

FIRMA

Debe registrarse esta autorización ante notario público

DEDICATORIA

A Dios por brindarme sabiduría y entendimiento para poder sacar adelante este trabajo.

A dos de los pilares más importantes de mi vida, mi madre DELIA GARCÍA MORENO y mi padre JAIRO ANZOLA VÁSQUEZ, ya que sin sus sabios consejos y su continuo apoyo no habría podido llegar hasta ésta etapa de mi vida.

A mis dos hermanos mayores SANDRA PATRICIA y JAIRO ENRIQUE por estar tan pendientes de mi durante el transcurso de mi carrera.

A mi sobrina ANA LUCIA ANZOLA TORRES por haberse convertido en mi razón de ser para salir adelante en todas las cosas que me he propuesto.

DEDICATORIA

A GLORIA GARCÍA DE MENDOZA, mi madre que todo el tiempo me apoyo para llegar hasta donde estoy.

A ÁLVARO MENDOZA DE LA ESPRIELLA, mi padre que con gran esfuerzo me ha colaborado en el transcurso de mi vida.

A LILIANA MENDOZA, DANIEL MENDOZA y MELISSA MENDOZA, mis hermanos que están siempre pendientes de mi.

A JOHN JAIRO STEER VARELA (q.e.p.d.), mi hermano, compañero y gran amigo que está en el cielo y sé que se encuentra orgulloso de mi.

AGRADECIMIENTOS

Al ingeniero de sistemas GONZALO GARZÓN, por habernos servido de guía en la realización de este trabajo.

Al ingeniero de sistemas ISAAC ZUÑIGA, por sus muchas recomendaciones y excelentes consejos brindados.

A todos los profesores que hicieron parte del MINOR en COMUNICACIONES Y REDES, así como a todos los profesores de la TECNOLÓGICA, por habernos regalado un poquito de todo lo que saben.

A todas las personas que de una u otra manera ayudaron a que nuestro sueño de ser profesionales se hiciera realidad.

CONTENIDO

	Pág.
LISTA DE TABLAS	15
LISTA DE FIGURAS	16
RESUMEN	17
INTRODUCCIÓN	20
OBJETIVO	22
1. CALIDAD DE SERVICIO (QoS).	23
1.1 CONCEPTO DE CALIDAD DE SERVICIO (QoS).	23
1.2 CLASIFICACIÓN DE LA CALIDAD DE SERVICIO (QoS).	27
1.2.1 De acuerdo con el tráfico de la red.	27
1.2.2 De acuerdo a quien requiera el nivel de calidad de servicio (QoS).	30
1.2.3 Según las garantías.	32
1.2.4 Según el lugar de aplicación.	32
1.3 PARÁMETROS DE LA CALIDAD DE SERVICIO (QoS).	34
2. HACIENDO CALIDAD DE SERVICIO (QoS).	38
2.1 CONTROL DE CONGESTIÓN.	38
2.1.1 Principios generales del control de congestión.	39
2.1.2 Factores que pueden influir en la creación de situaciones de	

congestión.	40
2.2 PERFILES DE TRÁFICO Y VIGILANCIA DEL TRÁFICO.	40
2.3 ESPECIFICACIONES DE FLUJO.	41
2.4 CONTROL DE ADMISIÓN.	42
2.5 PAQUETES DE ASFÍXIA.	42
2.6 DERRAMAMIENTO DE LA CARGA.	44
2.7 CONTROL DE JITTER.	45
3. CALIDAD DE SERVICIO (QoS) EN ATM.	46
3.1 CARACTERÍSTICAS DE ATM.	46
3.1.1 Direccionamiento.	47
3.1.2 Multiplexación.	47
3.1.3 Conmutación.	48
3.2 LA CELDA ATM.	50
3.2.1 Descripción de los campos de encabezamiento de una celda atm.	52
3.3 FUNCIONAMIENTO DE ATM.	54
3.4 ATM Y CALIDAD DE SERVICIO.	56
3.5 TIPOS DE TRÁFICO.	56
3.5.1 Clases de servicio en ATM.	57
3.6 POLÍTICAS DE TRÁFICO.	62
3.6.1 Control del retardo.	62

3.6.2 Parámetros de tráfico.	66
3.6.3 Parámetros de calidad de servicio para ATM.	67
3.6.4 Calidad de servicio y descriptores de tráfico.	68
3.7 MECANISMOS DE CONTROL.	72
3.7.1 Control de admisión de Conexión (CAC).	72
3.7.2 Adecuación del tráfico.	76
3.7.3 Control de prioridad de pérdida de celda.	79
3.7.4 Control de congestión.	80
3.8 CONFORMACIÓN Y VIGILANCIA DEL TRÁFICO.	81
4. CALIDAD DE SERVICIO (QoS) EN REDES IP.	84
4.1 CALIDAD DE SERVICIO DE EXTREMO A EXTREMO.	87
4.1.1 Servicios de Best Effort.	87
4.1.2 MPLS aided DiffServ.	88
4.2 MANEJO DE CONGESTION Y TRÁFICO.	89
4.3 PRIORIZACION DE TRÁFICO.	93
4.4 IP QOS SOBRE ATM.	97
4.4.1 IntServ sobre ATM.	98
4.4.2 DiffServ sobre ATM.	105
CONCLUSIONES	108
REFERENCIAS	112

LISTA DE TABLAS

	Pág.
Tabla 1. Valores del campo PT en la celda ATM.	53
Tabla 2. Tipos de tráfico y sus exigencias de calidad.	60
Tabla 3. Resumen de aplicación de los servicios de la red ATM.	61
Tabla 4. Resumen de los parámetros empleados por cada servicio ATM.	68
Tabla 5. Asignación de ancho de banda según la categoría de servicio.	75
Tabla 6. Control de congestión en el buffer de datos.	90
Tabla 7. Control de tráfico.	91
Tabla 8. Incremento de la eficiencia y señalización.	92
Tabla 9. Campos para manejo de Tipo de Servicio en IEEE 802.3.	94
Tabla 10. Mapeos entre ATM e IIS.	101

LISTA DE FIGURAS

	Pág.
Figura 1. Comparación entre sistemas con y sin QoS.	25
Figura 2. QoS y Aplicaciones.	29
Figura 3. Ejemplo de visualización de QoS implícita y explícita.	31
Figura 4. Encolado de servicios.	37
Figura 5. Celda ATM.	51
Figura 6. Retardo en una red ATM.	63
Figura 7. Funciones asociadas con el control del tráfico.	71
Figura 8. Los paquetes ip IntServ atraviesan una nube intermedia ATM.	99

RESUMEN

Actualmente nuestro mundo gira en torno al gran monstruo de la tecnología; podría decirse que casi la mayoría de las cosas que existen actualmente, si no son todas, vienen con tecnología implícita en alguna etapa de su desarrollo o de su uso. Es por eso que se hace necesario que día a día se generen nuevas formas de hacer tecnología, y lo mas importante, nuevas formas de mejorar lo que ya existe; sin perder eso si, la esencia en el proceso, esa esencia que ha logrado llevar al hombre a la luna, desarrollar maquinas súper poderosas e incluso clonar a seres vivos.

En el caso de la comunicación de datos y las redes de computadoras, ocurre exactamente lo mismo. Hoy en día el gran problema para los administradores de sistemas de comunicaciones y redes de computadores, es como mejorar el desempeño de todos los procesos asociados con su trabajo, como lograr hacer mucho con poco. Debido a esta situación aparece lo que en el ámbito computacional, en todo lo relacionado con comunicaciones y redes, se conoce como QoS o calidad de servicio. La calidad de servicio en un sistema, una red de computadores por ejemplo, tiene que ver con la forma como ella misma, la red, puede optimizar sus procesos en pro de la eficiencia; todo esto para satisfacer al usuario final.

Como es bien sabido una red maneja un sinnúmero de aplicaciones (voz, datos y video), que hacen necesario el uso de calidad de servicio para afrontar y dar eficiencia a las mismas. La calidad de servicio se puede clasificar de acuerdo al tráfico de la red, o de acuerdo a quien requiera el nivel de calidad de servicio, o según las garantías o según el lugar de aplicación. Además posee una variedad de parámetros que son de gran importancia para su correcta aplicación en las redes existentes, tales como el retardo, la latencia, el ancho de banda, el tráfico de red, rendimiento, disponibilidad, entre otros. Todos estos parámetros se deben tener presente a la hora de aplicar calidad de servicio.

Utilizar calidad de servicio es un beneficio del cual van a gozar, no solo el administrador de la red sino también el dueño de la misma, las aplicaciones que hacen uso de ella, los proveedores de servicios de red y los más importantes, los usuarios finales de ésta.

Qos, calidad de servicio puede ser implementada en la mayoría de las actuales tecnologías como por ejemplo, ATM, FRAME RELAY, ETHERNET y redes locales IEEE 802. Para ello es necesario manejar una serie de aspectos, los cuales son los que permitirán su correcta aplicación.

La tecnología llamada **Asynchronous Transfer Mode** (ATM) Modo de Transferencia Asíncrona es el corazón de los servicios digitales integrados que ofrecen las nuevas redes digitales de servicios integrados de Banda Ancha (B-

ISDN), donde los usuarios obtienen acceso a los recursos que necesitan y el operador de la red provee las rutas de conexión y asigna el ancho de banda necesario a fuentes de tráfico muy diferentes (voz, datos, vídeo). Esta tecnología puede proveer a los usuarios una calidad de servicio garantizada, dando a conocer a sus usuarios la clase de tráfico esperada que será transmitido en la conexión y el tipo de calidad de servicio que la conexión requiere, para esto se utilizan descriptores de tráfico y categorías de servicio.

Con el creciente uso de Internet, especialmente su proliferación en los negocios, existe una necesidad urgente por parte de los ISPs (Proveedores de Servicio de Internet) para proporcionar QoS. Esto significa proporcionar diferente tratamiento a clientes diferentes y ser capaz para garantizar ciertos niveles de calidad de servicio según las necesidades de los usuarios y a costos razonables. IP es actualmente un servicio de **Best Effort**¹, y la Calidad de Servicio QoS no es uno de los puntos fuertes de IP. Si ATM está disponible, entonces IP puede usar las provisiones de Calidad de Servicio de ATM. Existe un gran esfuerzo para incorporar Calidad de Servicio en IP. Es conveniente beneficiarse de cualquier apoyo inherente para QoS en una capa subyacente (si la capa subyacente es ATM). Muchos Proveedores de Servicio ISP ya están utilizando ATM en su Backbone. Ese es el punto: Calidad de servicio IP sobre ATM al cual se orienta este documento.

¹ Modelo en cuanto a calidad de servicio extremo a extremo se refiere.

INTRODUCCION

Anteriormente, debido a la baja capacidad de las redes, realizar estudios o trabajos relacionados con información multimedia como video-conferencia, audio-conferencia, video bajo demanda o sistemas cooperativos (pizarras compartidas, teletrabajo, telemedicina), así como aplicaciones tradicionales (transferencia de ficheros, base de datos, WWW), era casi imposible; situación que en la actualidad es totalmente lo contrario, hoy en día el avance en el área de tecnologías de redes, aumenta cada vez mas; el mayor de los avances que tenemos hasta hoy ha sido el auge de Internet y la facilidad de poder conectarse desde cualquier lugar utilizando solamente un computador, un módem y una línea telefónica. Por otra parte cada día se emplean mas las tecnologías de fibra óptica que brindan un mayor ancho de banda requerido por las aplicaciones mencionadas, las cuales necesitan una eficaz y eficiente gestión, es decir, necesitan un adecuado uso, garantizando la calidad de éste. Esto es lo que se conoce como calidad de servicio (QoS).

A continuación presentamos una explicación completa, precisa y globalizada de la definición del término calidad de servicio y como algunas de las tecnologías de redes mas conocidas actúan para alcanzarla. Para esto se hará un estudio de todas las medidas y parámetros que se requieren para conseguir calidad de

servicio en dichas tecnologías, de igual manera se describirán los métodos usados para poder conseguir la calidad de servicio en cualquier red, tales como mecanismos de vigilancia y prevención para su buen funcionamiento y de esta forma evitar la congestión en las redes.

OBJETIVO

Realizar un estudio que describa completamente como las tecnologías de redes más usadas actúan para garantizar calidad de servicio y de esta manera ayudar a los administradores de redes a tomar decisiones tendientes a optimizar el desempeño de sus redes.

1. CALIDAD DE SERVICIO (QoS)

1.1 CONCEPTO DE CALIDAD DE SERVICIO (QoS).

La calidad de servicio (QoS) puede definirse como el rendimiento de los servicios observados por el usuario final. Una red debe garantizar que puede ofrecer un cierto nivel de calidad de servicio para un nivel de tráfico que sigue un conjunto especificado de parámetros. En su conjunto, esas condiciones forman un contrato de tráfico entre el usuario y la red. Las siguientes definiciones son importantes para comprender cuando se habla de calidad de servicio:

- La clase de servicio (CS) define un conjunto preciso de parámetros cuando se ofrece un servicio
- El nivel acordado de servicios (**Service Level Agreement: SLA**) establece la calidad de servicio pactada mediante un contrato.

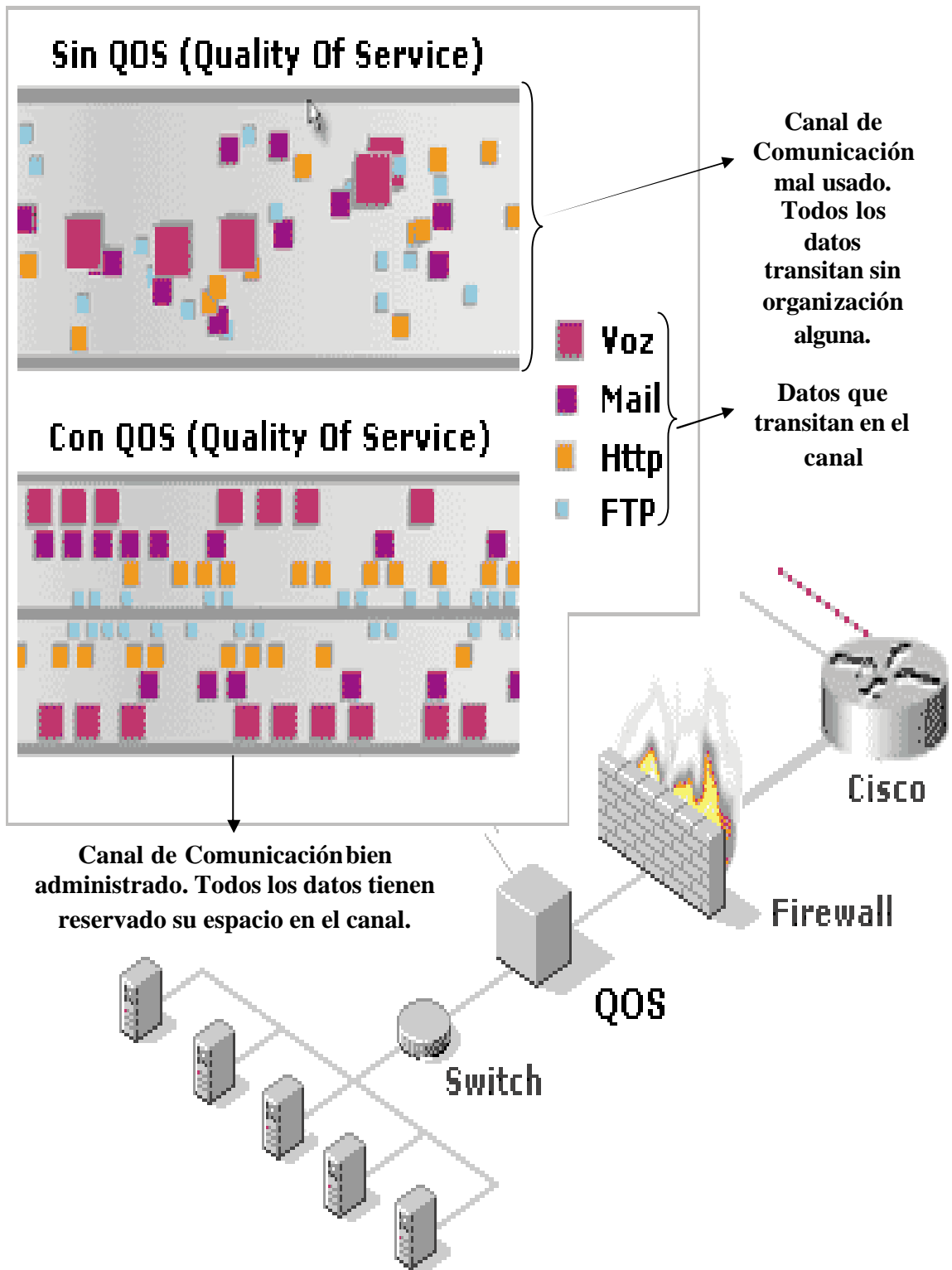
El concepto de calidad de servicio se originó en las técnicas y estándares de redes, pero también puede extenderse al Web, las aplicaciones y los servidores de contenido para administrar las clases de servicio a lo largo de todos los recursos de transmisión y procesamiento que forman la infraestructura de Internet. Las computadoras locales y remotas también pueden administrarse con calidad de

servicio para optimizar las tareas de procesamiento de aplicaciones en Internet (siguiendo el nivel acordado de servicios bajo una verdadera arquitectura distribuida). Deben establecerse protocolos QoS eficientes entre las redes y los servidores para administrar la red y las computadoras, según lo establece el nivel acordado de servicios, adaptándose a las condiciones reales de la red y los servidores, que cambian a cada momento.

Las redes pueden introducir retardos, pérdida de paquetes o errores debido a problemas de multiplexaje, conmutación o transmisión en nodos congestionados, impactando entonces la calidad del servicio. Cuando se maneja asignación de recursos de voz, datos y video, las condiciones de la red se tensan al máximo para poder garantizar el desempeño de los servicios múltiples. La calidad de servicio se definió inicialmente en los protocolos de comunicaciones ATM y luego evolucionó al protocolo IP para disponer de las herramientas para manejar la infraestructura de las redes de nueva generación (NGN).

Las computadoras existentes en Internet pueden presentar problemas de sobrecarga en el procesador, la memoria y los dispositivos de E/S, lo cual disminuye la calidad de servicio. Las aplicaciones deben considerar redundancia, balanceo de cargas y prioridades de asignación de recursos a lo largo de todos los elementos de cómputo disponibles. Las tecnologías QoS son bastante nuevas en la industria de las computadoras, especialmente cuando se establecen protocolos entre redes y procesadores distribuidos.

Figura 1. Comparación entre sistemas con y sin QoS.



El acceso y la seguridad son dos componentes centrales de las tecnologías aplicadas a los servicios corporativos distribuidos mediante Internet, así como dos tecnologías paralelas incluidas en una solución QoS. De hecho, el acceso, la seguridad y los protocolos qos deben interoperar en forma natural con los entornos privados virtuales (VPE) para conformar aplicaciones tipo QoS a lo largo de la red y los servidores. Los recursos de las computadoras y la red pueden programarse para diferenciar los servicios por usuario y por aplicación. Se pueden definir clases de servicio para cada usuario y aplicación que se ejecute en Internet (hasta llegar al nivel de las funciones específicas o los componentes de cada aplicación). Promediando los parámetros de la clase de servicio empleados por los usuarios y sus aplicaciones, cada paquete IP puede ser procesado en forma diferente, de modo tal que se puede disponer de los recursos de la infraestructura de acuerdo con el nivel de servicios pactado.

Aun cuando los costos del ancho de banda se reducen en algunos países, las aplicaciones de nueva generación (NGA) requieren cada vez más recursos de Internet. A medida que la infraestructura cambia con rapidez en todo el mundo, las empresas requieren la calidad de servicio no sólo para garantizar la entrega eficiente y económica de las aplicaciones en Internet, sino también para programar la asignación de los recursos de cómputo y red y aplicar las políticas corporativas.

Cuando los recursos de Internet se administran mediante la calidad de servicio, la frase "la computadora es la red" se vuelve real. Internet se ha hecho tan grande

que requiere de mejores herramientas para administrarla. Las tecnologías QoS permiten recuperar el mismo control que existía en los entornos centrados en las aplicaciones, como en los **mainframes** ² (como si todo se estuviera procesando en una sola computadora).

1.2 CLASIFICACIÓN DE LA CALIDAD DE SERVICIO (QoS)

La calidad de servicio esta clasificada siguiendo algunos patrones, tales como, tipo de trafico, campo de aplicación, reserva de recursos de la red y algunos otros parámetros.

1.2.1 De acuerdo con el tráfico de la red.

- QoS demasiado sensible al retardo. Para este caso, el ejemplo más práctico que se puede mostrar es el tráfico de video comprimido. Se debe garantizar que exista un gran ancho de banda disponible para este tráfico en la red y además un valor de retardo el cual permita la transmisión de dicho trafico, por lo que se requerirá emplear métodos de prioridad y organizar los flujos de datos.
- QoS un poco sensible al retardo. Un caso práctico que ejemplifica esta clasificación de calidad de servicio es la aplicación de la emulación de

² Sistema informático de gran envergadura.

círculo. También se debe garantizar que exista un ancho de banda disponible para el tráfico en la red aunque un poco menor que en el caso anterior, y también se requerirá establecer la priorización para el tráfico de información.

- QoS demasiado sensible a pérdidas. Este caso se puede reflejar en el tráfico de información en la red tradicional. En esta clasificación se puede deducir que si no se tienen pérdidas no se podrá descartar paquetes, ni se desbordarán los **buffers**³ de almacenamiento del flujo; esto proporcionará un mejor y mayor control en el tráfico de información.
- QoS insensible. En esta clase de calidad de servicio se utiliza la posibilidad de transmisión que exista y tomar la capacidad que poseen los **buffers** existentes como necesaria para el tráfico de información con la prioridad más baja. Un ejemplo de este tipo de QoS es el tráfico de servicios de noticias.

En la figura 2 es posible diferenciar de forma gráfica los tipos de tráfico y sus exigencias de ancho de banda y de sensibilidad a la latencia.

³ Porción de memoria utilizada en una tarea determinada.

Figura 2. QoS y Aplicaciones.

QoS y Aplicaciones

Usuarios y administradores demandan niveles de servicio y tiempos de respuesta adecuados para aplicaciones críticas.

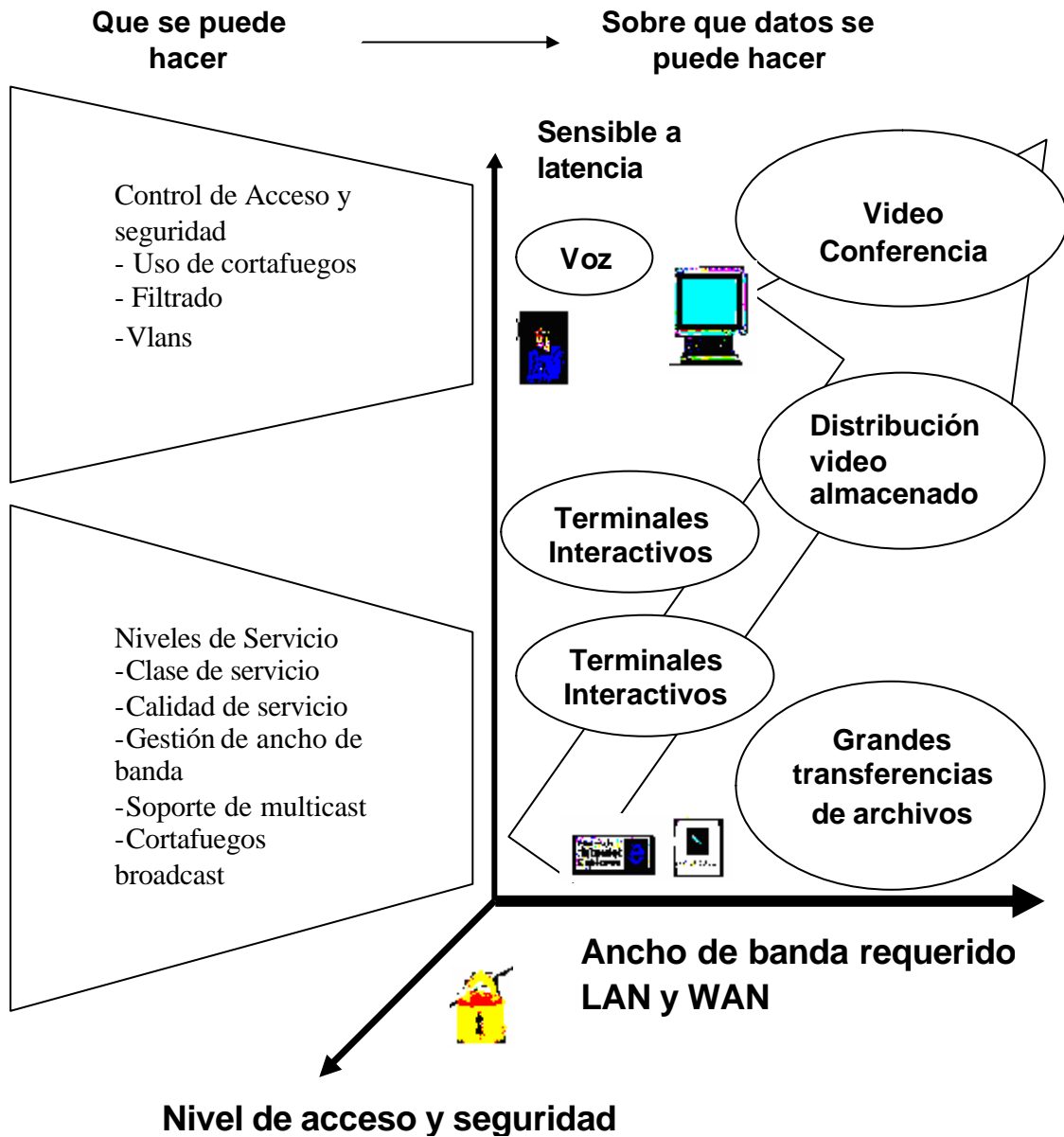


Figura 2. Podemos apreciar algunos servicios, tipos de datos que transitan en la red, sensibilidad a la latencia y ancho de banda requerido.

1.2.2 De acuerdo a quien requiera el nivel de calidad de servicio (QoS). En este caso la calidad de servicio puede ser necesitada y realizada por el usuario final o por los **routers / switches** de la red:

- QoS implícita. El (**router / switch**)⁴ establece los niveles de calidad de servicio, siguiendo lo especificado o estipulado por el administrador (protocolos, dirección de origen). Esta clase de QoS la brindan todos los **routers** y algunos **switches**, en el que las estaciones finales transmiten los paquetes, los cuales llegan al **router** o **switch**, en el que se analiza la información entrante y se prioriza, luego se reparte en distintas colas dependiendo de la prioridad de estas; la información se vuelve a transmitir al próximo **switch** o **router**, en el cual se realiza el mismo proceso.
- QoS explícita. En esta categoría los **routers / switches** reciben un servicio con un nivel escogido previamente por el usuario, en la cual las estaciones finales transfieren una petición RSVP, que si es aceptada, los paquetes son transmitidos; por lo tanto, la información que accede al **switch** o **router** es priorizada dependiendo de las instrucciones del nodo destino, y de esta forma llega al siguiente **switch** o **router**, en el que se repetirá el proceso.

⁴ Equipos de conectividad para redes.

Figura 3. Ejemplo de visualización de QoS implícita y explícita.

Estaciones Finales

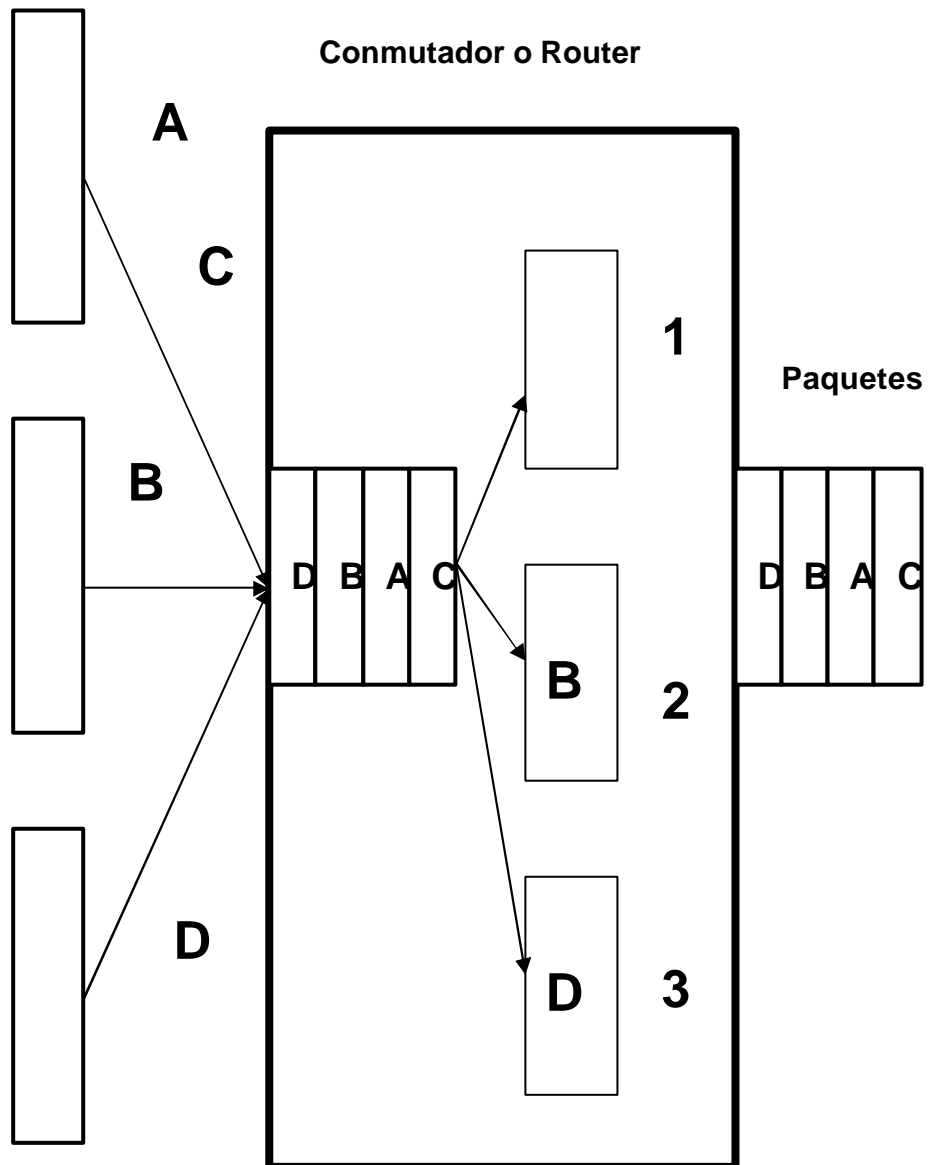


Figura 3. Podemos darnos cuenta de cómo el equipo (Router) se encarga de la administración y salida de los datos.

1.2.3 Según las garantías. Teniendo en cuenta la reserva de recursos del sistema para proporcionar los servicios, se tienen las siguientes clases de QoS:

- QoS garantizada / **Hard QoS**. En este caso se genera una reserva absoluta de los recursos de la red para un tráfico estipulado, garantizando de esta forma niveles máximos para el tráfico de información en una red.
- QoS no garantizada / **Lack of QoS**. Es el tipo de QoS correspondiente a los servicios **Best Effort** (Mejor Esfuerzo). En esta categoría no se tiene ninguna clase de garantía; el tráfico de información está dado por la red, el cual depende de los acontecimientos que en ella se den.
- QoS servicios diferenciados / **Soft QoS**. Es el punto medio entre las dos categorías anteriores. En este caso se hace una diferenciación de tráfico, los cuales son analizados, solamente algunas de ellos, teniendo en cuenta la mejoría referenciada a sus características (expedición más rápida, mas ancho de banda promedio, menos tasa de error promedio).

1.2.4 Según el lugar de aplicación.

- QoS extremo a extremo de la red (**end-to-end**). En este caso la calidad de servicio se aplica de extremo a extremo de la red, se conoce también como calidad de servicio absoluta. Es factible aunque menos utilizada que la calidad de servicio entre dos bordes de la red. Por otro lado, con la

aplicabilidad de este tipo de QoS se reducen los **switches** que se limitan a observar la marca de los paquetes (en el caso de 802.1 p), sin calcular la clase de servicio de cada paquete reducido. Además las aplicaciones podrían escoger dinámicamente el nivel de QoS almacenado temporalmente en los directorios de red o en los **switches** una información estática de clases de servicio.

- QoS borde a borde (**edge-to-edge**). La calidad de servicio se aplica, en este caso, entre dos puntos cualesquiera de la red. Por otro lado, no necesita que los administradores de red toquen los extremos, esto es una ventaja para el caso de las empresas en las que la organización responsable de la infraestructura de red esta separada del grupo de los servidores y del resto de los puestos de trabajo. Además son menos los dispositivos que tienen que ser manejados para la obtención de la QoS; y la accesibilidad por parte de un usuario cualquiera de la red o de un **hacker** ⁵ para cambiar las especificaciones de QoS es mucho menor. Por ultimo al aplicar este tipo de calidad de servicio no es necesario conocer como poner en práctica las reglas de QoS de cada uno de los posibles sistemas operativos que podrían tener los servidores en el caso de aplicar QoS extremo-a-extremo.

⁵ Persona dedicada a labores computacionales de baja reputación.

1.3 PARÁMETROS DE LA CALIDAD DE SERVICIO (QoS)

Al aplicar la calidad de servicio en las redes se manejan diversos conceptos que son de gran importancia para su buen funcionamiento, algunos de estos son:

- **Trafico de red.** Son los datos que se transportan dentro de la red, los cuales dependen de la forma de aplicación que transita en la red. Este se clasifica según su tipo de aplicación (tráfico habitual, multimedia, **multicast**⁶, **broadcast**⁷, tiempo real, etc.), y la sensibilidad al retardo (en este caso puede ser algo sensible, demasiado sensible al retardo, muy sensible a las perdidas o nada sensible).
- **Retardo.** Indica la variación temporal y/o retraso en la llegada de los flujos de datos a su destino. Característica evidente en aplicaciones como videoconferencia. Teniendo en cuenta hacia que tipo de aplicaciones se están orientando las telecomunicaciones, es necesario que en las políticas de QoS definidas para nuestra red este parámetro sea reducido al mínimo.
- **Latencia.** Es el tiempo entre el envío de un mensaje por parte de un nodo y la recepción del mensaje por otro nodo. Abarca los retardos que suceden durante el envío de dicho mensaje, ya sea en el propio camino o en los dispositivos por los que pasa.

⁶ Mensaje enviado a un conjunto de receptores.

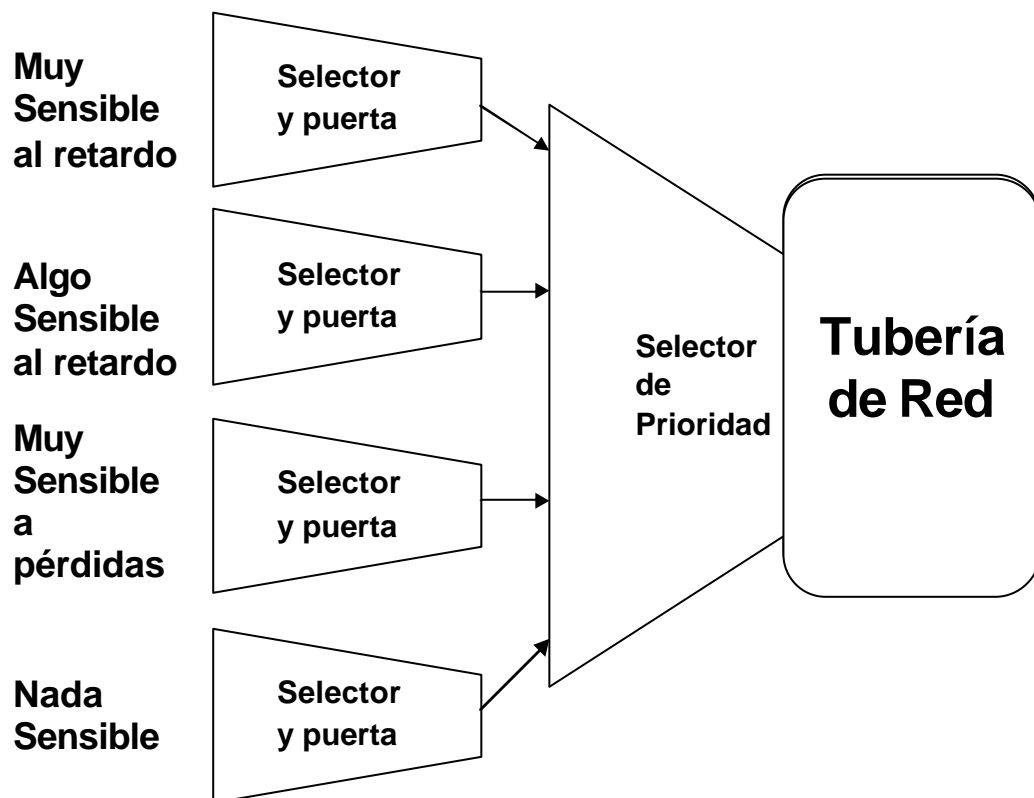
⁷ Mensaje enviado a un conjunto de receptores.

- **Jitter** (Inestabilidad o variabilidad en el retardo). Es lo que ocurre cuando los paquetes transmitidos en una red no llegan a su destino en debido orden o en el tiempo determinado, es decir, varían en latencia.
- **Ancho de banda.** Medida de la capacidad de transmisión de datos, expresada generalmente en Kilobits por segundo (kbps) o en Megabits por segundo (Mbps) y últimamente en Gigabits por segundo (Gbps). Indica la capacidad máxima teórica de una conexión, la cual se ve disminuida por factores tales como el retardo de transmisión, que pueden causar un deterioro en la calidad. El aumento del ancho de banda significa poder transmitir más datos, pero también implica un incremento económico.
- **Perdida de paquetes.** Indica el número de paquete 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 (SLAs o acuerdos de nivel de servicio). El rendimiento es definido también por algunos profesionales como la velocidad teórica de transmisión de los paquetes por la red. Esta depende directamente del ancho de la banda y su variación de las posibles situaciones de congestión de la red.

- **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. Es necesaria únicamente cuando la red no proporciona la suficiente capacidad para atender todo el tráfico presente en la misma.
- **Encolado.** Consiste en dividir y organizar el tráfico ante un determinado dispositivo de red para su posterior retransmisión por la misma según un determinado algoritmo que define la cola y que permite que determinados paquetes sean reexpedidos antes que otros. Es una de las herramientas más utilizadas por la QoS. La idea es ofrecer un mejor servicio al tráfico de alta prioridad al mismo tiempo que se asegura, en diferentes grados, el servicio para los paquetes de menor prioridad. Pero no se asegura la llegada a tiempo de los paquetes.
- **Planificación.** Proceso de decidir que paquetes enviar primero en un sistema de múltiples colas.
- **Flujo.** 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 de destino y el mismo protocolo. En el caso de redes TCP / IP el flujo necesita llegar secuencialmente a su destino con una frecuencia constante.

- Acuerdos de niveles de servicio (SLA). 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, el método de medida, las consecuencias cuando los niveles de servicio no se consiguen o si los niveles de tráfico definidos son superados por el cliente, así como el precio de todos estos servicios.

Figura 4. Encolado de servicios.



2. HACIENDO CALIDAD DE SERVICIO

Para hacer un buen uso de la calidad de servicio, en una tecnología de red, es necesario conocer una serie de mecanismos que QoS utiliza para hacer realidad la calidad de servicio. A continuación entraremos un poco en detalles en cada uno de esos mecanismos que hacen posible la implementación de QoS.

2.1 CONTROL DE CONGESTIÓN

Se entiende por congestión la circunstancia en la que el rendimiento de la red (o una parte de ella) se degrada debido a la presencia de demasiados paquetes / tramas. Normalmente la congestión se produce porque se pretende pasar mas trafico por una línea de lo que esta es capaz de absorber. Por ejemplo un **router / switch** recibe mucho trafico de varias líneas de 2 Mb/s todo dirigido a una sola línea también de 2 Mb/s. Generalmente el cuello de botella es la línea de transmisión, pero en algunos casos la congestión puede ser causada por un **router / switch** sobrecargado o de poca capacidad para el trafico que soporta.

Inicialmente los **buffers** intentan salvar la situación, pero si la situación dura bastante tiempo los **buffers** se llenan y los **routers / switches** empiezan a descartar paquetes.

Tanto en control de congestión como en control de flujo suelen utilizarse mecanismos de notificación al emisor para que baje el ritmo. La congestión es mas compleja, ya que generalmente el emisor es un puro intermediario que se limita a reenviar hacia atrás al verdadero causante del problema el aviso recibido. En líneas de alta velocidad y elevado retardo (gran distancia) este problema se acentúa.

En muchos casos la solución a un problema de congestión se traduce en un control de flujo en el **host**⁸ que genera el tráfico.

2.1.1 Principios generales del control de congestión. Entre los parámetros que permiten detectar la presencia de congestión se encuentran por ejemplo:

- Porcentaje de paquetes / tramas descartados.
- Longitud media de las colas en las interfaces de los **routers / switches**.
- Numero de paquetes / tramas que se retransmiten.
- Retardo medio por paquete / trama.
- Desviación media del retardo por paquete / trama.

Para resolver la congestión solo hay dos posibles medidas:

- Reducir el tráfico informando al emisor para que pare de enviar, o buscando

⁸ Sistema principal o central.

rutas alternativas.

- Aumentar la capacidad.

2.1.2 Factores que pueden influir en la creación de situaciones de congestión. Entre los factores a nivel de enlace que pueden influir en el nivel de congestión se encuentran:

- El intervalo de **timeout** (Tiempo fuera). Si es pequeño originara retransmisiones innecesarias.
- El tamaño de la ventana. Si es grande es más fácil que se produzca congestión.
- El uso de retroceso o repetición selectiva. El retroceso genera mas trafico.

2.2 PERFILES DE TRÁFICO Y VIGILANCIA DEL TRÁFICO (TRAFFIC SHAPING Y TRAFFIC POLICING)

El tráfico a ráfagas es la principal causa de congestión. Si todos los ordenadores transmitieran siempre un flujo constante seria muy fácil evitar las congestiones.

Los perfiles de trafico (**traffic shaping**) establecen unos márgenes máximos al trafico a ráfagas. Suelen utilizarse para fijar una QoS entre el operador y el usuario, entre tanto el usuario respete lo establecido el operador se compromete a

no descartar paquetes (actúa como una especie de contrato).

Se denomina vigilancia del tráfico (**traffic policing**) a la labor de monitorización o seguimiento del tráfico introducido por el usuario en la red para verificar que no excede el perfil pactado. En este proyecto se describirá el algoritmo del Pozal agujereado (**Leaky Bucket**⁹) el cual es uno de los sistemas más utilizados para establecer perfiles de tráfico.

2.3 ESPECIFICACIONES DE FLUJO

Se conocen con este nombre el conjunto de parámetros que especifican el ancho de banda y calidad de servicio esperados en una transferencia de datos entre dos entidades para poder realizar de forma eficiente el perfil de tráfico y la vigilancia de éste.

Estos parámetros fueron descritos en el capítulo 1 y aquí solo se mencionaran algunos de ellos:

- Máximo tamaño de paquete (MTU, **Maximun Transfer Unit**)
- r y C si se utiliza pozal agujereado.
- r , C y M si se utiliza poza agujereado con créditos.
- Tasa de pérdidas.

⁹ Algoritmo utilizado para establecer perfiles de tráfico.

- Cantidad máxima de paquetes consecutivos que pueden descartarse.
- Retardo máximo.
- Variación máxima en el retardo o **jitter**.

2.4 CONTROL DE ADMISIÓN (REDES DE CIRCUITOS VIRTUALES)

El control de admisión consiste en impedir el establecimiento de nuevos circuitos virtuales (para redes ATM y **Frame Relay** ¹⁰) que pasen por la zona congestionada. Si se conoce la capacidad que necesita cada circuito virtual es posible controlar el acceso de forma que nunca se produzca congestión. En general esta técnica no usa los recursos de manera eficiente, por lo que es normal prever un cierto grado de sobre suscripción u **overbooking**.

2.5 PAQUETES DE ASFIXIA (CHOOKE PACKETS)

Se puede aplicar tanto en redes de circuitos virtuales (ATM) como de datagramas (Ethernet).

En esta técnica el **router / switch** sigue de cerca la situación de cada una de sus líneas, monitorizando por ejemplo el grado de utilización, la longitud de la cola o la ocupación del **buffer** correspondiente. Cuando el parámetro inspeccionado supera un determinado valor considerado umbral de peligro se envía un paquete de

¹⁰ Tecnología de red.

asfixia (**choke packet**) al **host** considerado culpable para que reduzca el ritmo.

La evaluación del parámetro a monitorizar se suele hacer con una formula del tipo:

$$U_n = au_{n-1} + (1-a)f$$

Donde f es el valor instantáneo medido del parámetro (utilización, tamaño de la cola u ocupación del **buffer**), U_n el valor medio en la n ésima iteración, y a una constante que permite regular el peso que se le quiere dar a los valores anteriores, o dicho de otro modo con que rapidez se quiere reaccionar a situaciones cambiantes.

Los paquetes de asfixia se envían al **host**, ya que es este y no el **router / switch** el verdadero origen de la congestión. Los **host** cuando reciben estos paquetes suelen reducir a la mitad la velocidad con la que envían datos a la red. Esto lo pueden hacer reduciendo por ejemplo el tamaño de ventana (del protocolo a nivel de transporte) o cambiando los parámetros del pozo agujereado o pozo con créditos si utiliza alguno de estos algoritmos para controlar su flujo. En una situación de congestión normalmente serán muchos los **host** que recibirán este tipo de paquetes.

En ocasiones interesa que los paquetes de asfixia tengan un efecto inmediato en cada uno de los **routers / switches** que atraviesan en su camino hacia el **host** origen del trafico; esto ocurre sobre todo cuando se trata de una conexión de alta

velocidad y gran retardo. Se trata así de resolver la congestión de forma inmediata distribuyendo el tráfico en curso entre los **buffers** de los **routers / switches** que hay por el camino, entretanto el mensaje de alerta llega al verdadero causante de los problemas.

Por desgracia la obediencia a los paquetes de asfixia es completamente voluntaria, si un **host** decide hacer caso omiso de las advertencias no se le puede obligar. Si un **host** obedece las indicaciones y reduce su ritmo mientras otros no lo hacen el primero saldrá perjudicado pues obtendrá una parte aun menor de la ya escasa capacidad disponible.

2.6 DERRAMAMIENTO DE LA CARGA

La solución extrema para resolver la congestión es descartar paquetes. En ocasiones los paquetes llevan alguna indicación de su grado de importancia, en cuyo caso los **routers / switches** intentan descartar los menos importantes primero. Por ejemplo, sería bastante grave si un **router / switch** para resolver una situación de congestión descartara paquetes de asfixia enviados por otro **router / switch**.

A veces el nivel de aplicación puede dar información sobre la prioridad de descarte de los paquetes. Por ejemplo en aplicaciones isócronas (audio y video en tiempo real) suele ser preferible descartar el paquete viejo al nuevo ya que el viejo

seguramente es inútil, mientras que en transferencia de ficheros es al contrario pues en muchos casos el receptor no aceptara el paquete nuevo mientras no reciba el viejo.

En los ficheros MPEG (formato estándar de video digital) algunos fotogramas son completos y otros son diferencias respecto a los fotogramas contiguos, descartar un paquete de un foto grama completo es mas perjudicial. En ATM existen soluciones en las que se le permite al usuario inyectar en la red un ancho de banda superior al contratado, pero a condición de que dicho tráfico excedente sea susceptible de ser descartado en caso de congestión; para esto los **switches** ATM marcan de cierta forma los paquetes que entran en la red con estas condiciones. En todos los casos es preciso que los paquetes poco importantes estén claramente identificados para que los **routers / switches** sepan descartarlos de manera selectiva en caso necesario.

2.7 CONTROL DE JITTER (FLUCTUACIÓN)

En aplicaciones isócronas la fluctuación en el retardo o **jitter**, es tan importante o mas que el retardo mismo. Para intentar minimizar el **jitter** en la red es preciso que cada **router / switch** analice si el paquete va adelantado o atrasado respecto a su horario normal; si va adelantado lo pondrá al final de su cola, y si va atrasado lo pondrá al principio. De esta forma se reduce el **jitter** para todos los paquetes.

3. CALIDAD DE SERVICIO (QoS) EN ATM

Las redes de modo de transferencia asíncrona (**Asynchronous Transfer Mode**, ATM), surgen como idea de un grupo de ingenieros de compañías telefónicas, como respuesta a una demanda de redes más rápidas y de mayor ancho de banda, para acomodarse a los crecimientos en los tamaños de ficheros y aplicaciones experimentados en el sector. Fundamentalmente ATM es una tecnología que simultáneamente transmite tráfico de datos, voz y video sobre circuitos de alto ancho de banda, generalmente cientos de mega bits por segundo (Mbps) en 1997 y Giga bits por segundo (Gbps) en la actualidad. La plataforma de **hardware** y **software** de ATM crea una arquitectura de comunicaciones basada en **switching** (conmutación) y transmisión de pequeñas unidades de información, llamadas celdas (**cells**), algunos también las llaman células; estas celdas son de longitud fija, en concreto 53 bytes. Los bytes son enviados a la red uno a uno, en secuencia, y el propietario de la celda se determina por la información existente en la cabecera de la propia celda.

3.1 CARACTERÍSTICAS DE ATM

Para definir las funcionalidades del Modo de Transferencia Asíncrono se requiere estudiar el funcionamiento de esta tecnología y los servicios que ofrece.

3.1.1 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 Connections, PVCs**) como conmutadas (**Switched Virtual Connections, SVCs**).

3.1.2 Multiplexación. ATM posee la habilidad de multiplexar diversos tipos de servicios, como voz, video y datos, en una única red física no canalizada. Este método de Multiplexación de las celdas ATM define el concepto de Modo de Transferencia Asíncrono, en donde el término asíncrono se refiere a la capacidad de la red ATM de enviar datos asociados a una conexión sólo cuando realmente hay datos que transmitir. En contraste con las redes canalizadas, en donde aún cuando el canal está libre, es necesario enviar un patrón especial en cada ranura de tiempo (**time slot**) que representa a un canal libre. De lo contrario, el receptor no sería capaz de recuperar la información presente en el resto de las ranuras o **time slots**.

Esta es la esencia de las redes que emplean Modos de Transferencia Síncronos, en donde, a cada fuente se le asigna un ancho de banda fijo basada en una posición, por ejemplo, una banda de frecuencia en FDM (**Frequency División Multiplexing**) o un *time slot* en TDM (**Time División Multiplexing**).

Por el contrario, el tráfico en ATM no está basado en una posición fija en el flujo de datos, ya que un encabezamiento en la información ATM identifica hacia donde debe ser dirigido el tráfico; adicionalmente, se presenta bajo demanda: si no hay tráfico no se emplea el ancho de banda, por tanto, el ancho de banda disponible puede ser empleado para otras conexiones.

3.1.3 Conmutación. Una red ATM consiste de la interconexión de una serie de Conmutadores (Switches) ATM, por medio de enlaces punto a punto o interfaces ATM. Los Switches ATM permiten dos clases de interfaces:

- UNI (**User-Network Interface** / Usuario- Interface de red).
- NNI (**Network-Node Interface**, *también conocida como Network-Network Interface* / Red-Interface de red)

La interfaz UNI conecta sistemas finales ATM (**Hosts, Routers**) a un Switch ATM, mientras que la NNI puede ser definida como la Interfaz que conecta a dos Switches ATM que intercambian el protocolo NNI; La conexión entre un Switch ATM privado y un Switch ATM público es una UNI (conocida como UNI Pública), debido a que estos Switches típicamente no intercambian información NNI.

El ATM Forum prevé que la mayoría de los servicios ATM sobre un área amplia

serán provistos por una red ATM pública, entendiéndose por red pública aquella red que posee la habilidad de conectar cualquier usuario con cualquier otro usuario.

En redes atm se observa el empleo de dos formas distintas de la ATM UNI:

- UNI Pública: la cual es típicamente usada para interconectar a un usuario ATM con un Switch ATM ubicado en la red de un Proveedor de Servicio Público, o para conectar a un Switch ATM ubicado en una red privada con un Switch ATM ubicado en una red pública.
- UNI Privada: la cual se usa típicamente para interconectar a un usuario ATM con un Switch ATM que es controlado por la misma red corporativa que es responsable del dispositivo del usuario.

La principal diferencia entre ambas UNI es el alcance físico. Existen también algunas diferencias funcionales entre las mismas debido a los requerimientos asociados con cada una de estas interfaces. Ambas UNI comparten la especificación de la capa ATM, pero pueden emplear diferentes medios físicos.

Dentro de la red pública o privada ATM, los Switches intercambian celdas e información de control por medio de la interfaz Pública NNI o Privada NNI,

respectivamente. En el lado del cliente, algunos equipos del usuario realizan la adaptación ATM, en otras palabras, por un lado hacen interfaz con equipos no basados en celdas y por el otro lado hacen interfaz con el nodo de la red local ATM mediante la UNI.

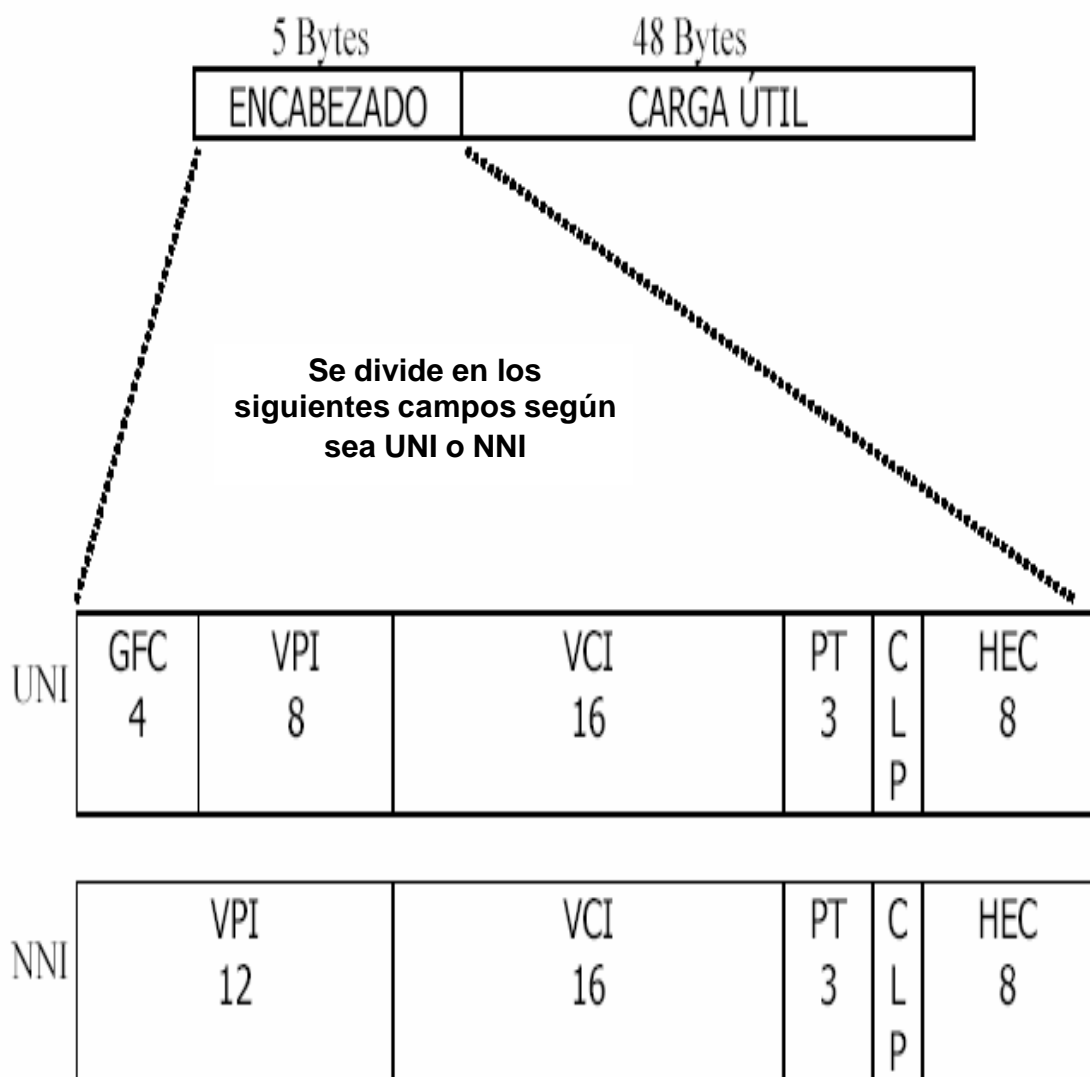
Las redes públicas basadas en ATM pertenecientes a diferentes operadores también deben ser interconectadas con la finalidad de facilitar servicios ATM de extremo a extremo, tanto en el entorno nacional como internacional. Para lograr estos objetivos se requieren métodos adicionales que soporten la multiplexación de servicios entre los operadores de una forma eficiente y controlable. La conexión entre dos diferentes Proveedores de redes públicas puede ser realizada mediante la interfaz B-ICI (*BISDN Inter Carrier Interface*).

3.2 LA CELDA ATM

Como ya se ha mencionado ATM define una celda de tamaño fijo con una longitud de 53 bytes. Consta de dos partes: la carga útil o **payload** de 48 **bytes** que transporta la información generada por un emisor o transmisor, y el encabezamiento o **header** de 5 **bytes** que contiene la información necesaria para la transferencia de la celda. Las celdas son enviadas sobre una estructura de transmisión física, como por ejemplo el DS1, DS3 o SONET de norte América; el E1, E3, E4 o STM de Europa.

Existen dos formatos de cabecera según se trate de la interfaz UNI (**Interface** usuario/red) o NNI (**Interface** nodo/red). La interfaz UNI conecta sistemas finales ATM (**hosts, routers,** etc) a un **switch** ATM, mientras que la NNI puede ser definida como la interfaz que conecta a dos **switches** ATM.

Figura 5. Celda ATM.



3.2.1 Descripción de los campos de encabezamiento de una celda ATM

- Control de flujo genérico (GFC, **Generic Flow Control**): Este campo consta de 4 bits que corresponden a los más significativos del primer byte. Este campo se puede utilizar desde la interfaz del usuario, para asegurar el acceso apropiado de varios terminales sobre un medio compartido (bus o anillo).
- Identificador de Camino Virtual (VPI, **Virtual Path Identifier**) y de Canal Virtual (VCI, **Virtual Circuit Identifier**): Representa una dirección lógica que identifica el circuito virtual al cual la celda está conectada. Los campos de identificación VCI y VPI son esenciales para el enrutamiento y la multiplexación.
- Tipo de Carga Útil (PT, **Type Payload**): Este campo de información permite que el nodo ATM pueda distinguir en una conexión, si el contenido de las celdas corresponde a un usuario o es información de control y gestión de la red. El campo PT contiene 3 bits, en la tabla 2 se describen cada uno de los ocho valores posibles de este campo.

Tabla 1. Valores del campo PT en la celda ATM

VALOR PT	SIGNIFICADO
000	Celda de usuario tipo 0; no se detecta congestión
001	Celda de usuario tipo 1; no se detecta congestión
010	Celda de usuario tipo 0; se detecta congestión
011	Celda de usuario tipo 1; se detecta congestión
100	Información de mantenimiento entre conmutadores vecinos
101	Información de mantenimiento entre conmutadores de origen y destino
110	Celda de gestión de recursos (utilizada para control de congestión ABR)
111	Reservado

- **Prioridad de Pérdida de la Celda (CPL, Cell Loss Priority).** Este campo se emplea para indicarle a la red cuales celdas, dentro de una conexión determinada, son más sensibles a una pérdida que otra. Aquellas celdas que contienen el bit CLP en 1 tienen prioridad más baja y por tal motivo serán descartadas de la red en primera opción frente a una posible congestión de tráfico.

- Control de Errores de Encabezamiento (HEC, **Header Error Check**): Proporciona un CRC de los primeros 40 bits que detecta todos los errores simples y la mayoría de los errores múltiples. En el encabezamiento de las celdas, el transmisor calcula el valor HEC para la totalidad del encabezamiento de la celda, excluido el campo HEC.

3.3 FUNCIONAMIENTO DE ATM

Circuitos virtuales (VC, **Virtual Circuit**). ATM utiliza circuitos virtuales para crear enlaces (conexiones) individuales y transportar las celdas 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. Los circuitos virtuales pueden ser permanentes o conmutados.

- Circuitos virtuales permanentes (PVC, **Permanent Virtual Circuit**). Los circuitos virtuales permanentes se configuran de manera estática en los **switches** y están presentes siempre que la red esta operativa, estos circuitos 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, **Switched Virtual Circuit**). Los circuitos virtuales conmutados, son creados dinámicamente para cada transmisión y son similares a la red telefónica de voz, en donde las conexiones entre dos puntos extremos de la red son creadas dinámicamente para cada transmisión. Los SVC pretenden determinar la ruta disponible mas corta desde la fuente hasta el destino.

Tanto para un PVC como para un SVC, antes de los que datos fluyan, los recursos requeridos tienen que ser asignados y luego que esto ocurra, permanecen asignados durante el circuito. Después de que el circuito ha sido establecido hay una política de seguimiento para que el tráfico cumpla con las condiciones acordadas. Cada **switch** a lo largo de la conexión valida que los recursos requeridos estén disponibles.

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.

3.4 ATM Y CALIDAD DE SERVICIO

A continuación describiremos como la red de modo de transferencia asíncrona brinda calidad de servicio.

Debido a que ATM fue diseñada con un enfoque hacia la calidad de servicio, provee de alta velocidad, bajo retardo, bajo **jitter** (variación del retardo) y garantía de pérdida de celdas, características estas que definen la calidad de servicio para una conexión en particular, además de soportar prácticamente todos los tipos de tráfico.

Para conseguir las características mencionadas ATM realiza dos funciones muy importantes, primero que todo necesita hacer una diferenciación del tipo de tráfico soportado por esta red para lo cual define cuatro clases de servicio donde agrupa el tráfico con exigencias de garantía similares; la segunda función es determinar unas políticas que permitan la mezcla de los distintos tipos de tráfico y que este fluya sin problemas.

3.5 TIPOS DE TRÁFICO

Con el propósito de soportar varios tipos de tráfico se han definido las clases de servicios, cada una asociada a un grupo de parámetros de QoS. Las clases o categorías de servicios con las que cuenta ATM se describen a continuación,

estas difieren en el nivel de garantía que dan al usuario respecto de la disponibilidad de los recursos de red solicitados.

3.5.1 Clases de servicio en ATM

- Servicio CBR (**Constant Bit Rate**, Tasa de Bits Constante). Garantiza una capacidad determinada y constante, independientemente de la utilización que haga de la red este u otros usuarios. Este servicio es el más sencillo de implementar y el más seguro de todos, ya que la red reserva la capacidad solicitada en todo el trayecto de forma estática. No se realiza ningún tipo de control de congestión, ya que se supone que esta no puede ocurrir. Es equivalente a una línea dedicada punto a punto.

La categoría de servicio CBR soporta aplicaciones en tiempo real, requiriendo una cantidad fija de ancho de banda. CBT soporta ajustadamente los parámetros MCTD (Máximo retardo de transferencia de una celda) y CDVT (Tolerancia en la variación del retardo de una celda), los cuales se verán con más detalle a continuación. CBR es perfecto para aplicaciones que no pueden tolerar variaciones en la demora, como aquellas que manejan voz y video en forma constante.

- Servicio VBR (**Variable Bit Rate**, Tasa de Bit Variable): Esta pensado para cuando se prevea una elevada cantidad de tráfico de forma continuada.

Tiene dos modalidades: RT-VBR (**Real Time** VBR), con requerimientos de bajo retardo y **jitter** para cuando se trata de aplicaciones en tiempo real (video conferencia, video bajo demanda, etc.), Y NRT-VBR (**Non Real Time** VBR) para cuando se trata de aplicaciones de trafico elevado pero donde el retardo no es tan importante, por ejemplo correo multimedia o transmisión de ficheros MPEG por la red, que son vistos luego por el usuario locamente de forma asíncrona en su computador. En VBR el usuario especifica un ancho de banda medio pero, en función de sus necesidades y del estado de la red, podrá en muchas ocasiones utilizar anchos de banda superiores, lo cual da mayor flexibilidad y permite al usuario ajustar mas este recurso a sus necesidades medias reales. En algunos servicios VBR el tráfico excedente sale marcado con el bit CLP. Desde el punto de vista de la red VBR tiene una complejidad superior a CBR.

- Servicio ABR (**Available Bit Rat**, Tasa de Bit disponible). ABR esta pensado para tráfico a ráfagas, se supone que habrá instantes de gran demanda de capacidad seguidos de otros de total inactividad. La meta de este servicio es el de proveer dinámicamente el ancho de banda que actualmente no esta en uso por otras categorías de servicios a usuarios que pueden ajustar sus transmisiones a esa tasa. ABR permite establecer un mínimo garantizado en el ancho de banda, y fijar un máximo orientativo. ABR es la única categoría de servicio ATM en la que se pronostica que la red suministre control de flujo al emisor para que reduzca el ritmo en caso

de congestión; esta circunstancia hace de ABR la categoría de servicio mas apropiada para trafico de datos, por ejemplo para enviar datagramas IP cuando no se utilicen aplicaciones isócronas, ABR también es recomendado en las interconexiones del tipo Lan, transferencias de archivos de alta prestación, archivos de bases de datos y navegadores **Web**. Sin embargo debido a su funcionalidad ABR es la categoría de servicio más compleja de implementar.

- Servicio UBR (**Unspecified Bit Rate**, Tasa de Bit no especificada). Se puede considerar el de más baja calidad. No existe ningún tipo de garantías en cuanto al retardo o ancho de banda, y tampoco se informa al emisor en caso de congestión. UBR utiliza la capacidad sobrante en la red de las demás categorías de servicio. Puede utilizarse para emulación de LAN, IP sobre ATM y tráfico de misión no critica.
- Servicio GFR (**Guaranteed Frame Rate** – Tasa garantizada de trama): Ha sido creado para mejorar el tráfico **Best Effort** (mejor servicio) en un mínimo de garantías de rendimiento. Se ha demostrado que este tipo de cola no es suficiente para proporcionar unas mínimas garantías, por lo que el encolado vía circuito virtual con GFR es necesario. El servicio GFR, ha sido creado para aplicaciones que no son a tiempo real y pueden requerir ancho de banda adicional que se libera dinámicamente.

La tabla 2 resume los distintos tipos de tráfico con sus exigencias para cada parámetro de calidad y la tabla 3 resume la aplicabilidad de cada uno de los servicios en las aplicaciones de mayor uso en ATM.

Tabla 2. Tipos de tráfico y sus exigencias de calidad.

	RETARDO	VARIACIÓN DEL RETARDO	RELIABILITY
CORREO	BAJO	BAJO	ALTO
TRANSFERENCIA DE ARCHIVOS	BAJO	BAJO	ALTO
BASE DE DATOS	BAJO / MEDIO	BAJO	ALTO
VIDEOS (MPEG)	ALTO	MEDIO	BAJO
TELEFONÍA	ALTO	ALTO	BAJO
VIDEOCONFERENCIA	ALTO	ALTO	BAJO

Tabla 3. Resumen de aplicación de los servicios de la red ATM.

APLICACIÓN	CBR	RT - VBR	NRT - VBR	ABR	UBR
Datos Críticos	Bueno	Aceptable	Excelente	Aceptable	NO
Interconexión de LAN's	Aceptable	Aceptable	Bueno	Excelente	Bueno
Trasporte de WAN	Aceptable	Aceptable	Bueno	Excelente	Bueno
Emulación de Circuitos	Excelente	Bueno	NO	NO	NO
Telefonía y Videoconferencia	Excelente	Bueno	NO	NO	NO
Audio comprimido	Aceptable	Excelente	Bueno	Bueno	Aceptable
Distribución de video	Excelente	Bueno	Aceptable	NO	NO
Multimedia Interactivo	Excelente	Excelente	Bueno	Bueno	Aceptable

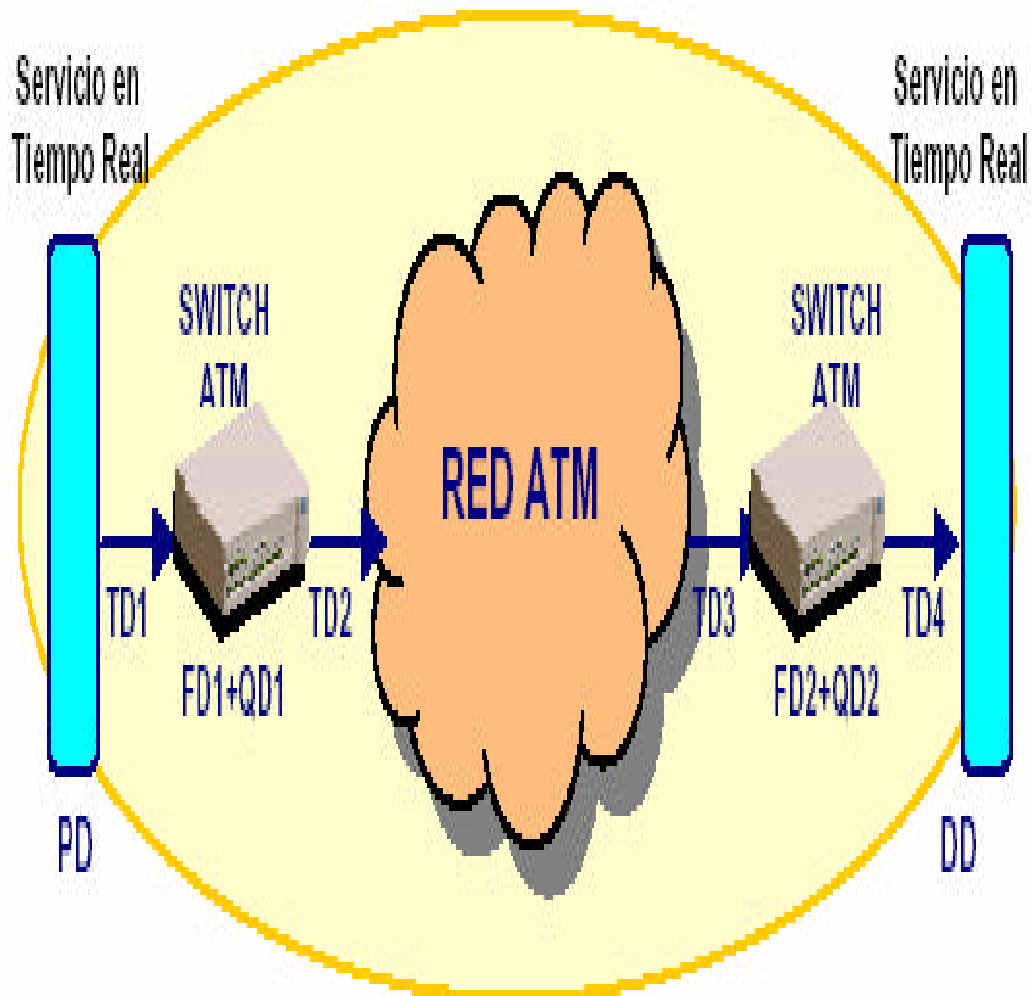
3.6 POLÍTICAS DE TRÁFICO

Los flujos en ATM se asocian a un circuito virtual (VC), los cuales pueden ser permanentes (PVC) o conmutados (SVC). Tanto para un PVC como para un SAC, antes de que los datos fluyan, los recursos requeridos tienen que ser asignados y permanecen asignados durante el circuito. Cada **switch** valida que los recursos requeridos estén disponibles. Después de que el circuito ha sido establecido hay una política de seguimiento para que el tráfico cumpla con las condiciones acordadas.

La designación de estas políticas es, precisamente, la segunda función realizada por ATM para garantizar la calidad de servicio, se determinan normas de tráfico para controlar que un usuario no exceda los límites de ancho de banda pactados en su contrata de servicio, esto se logra evaluando cada una de las celdas que entran en el **switch**. La implementación de las políticas de tráfico se lleva a cabo a través del algoritmo **Leaky Bucket** descrito anteriormente.

3.6.1 Control del retardo. El retardo en una red ATM es determinado por diferentes partes en la red, las cuales individualmente contribuyen al retardo total. En la Figura 26 se muestra una red pura ATM.

Figura 6. Retardo en una red ATM.



En donde: PD: **Packetization Delay.**

TD: **Transmisión Delay.**

QD: **Queuing Delay.**

FD: **Fixed Switching Delay.**

DD: **Depacketization Delay.**

La información es ensamblada en celdas, en el equipo terminal Emisor y es desensamblada en el equipo terminal Receptor. Internamente en la red, sólo existen celdas. Los parámetros que contribuyen al retardo total en la red son:

- Retardo de Transmisión (TD), denominado comúnmente retardo de propagación. Este retardo depende de la distancia entre los dos puntos y de la velocidad de propagación. Dependiendo del medio de transmisión empleado, el TD varía típicamente entre 4 y 5 ms por Km. Este retardo es independiente del tipo de tecnología o modo de transferencia empleado.
- Retardo de Paquetización (PD). Este retardo es introducido cada vez que un servicio en tiempo real (tal como voz y video) es convertido en celdas y depende de la longitud del paquete y de la velocidad a la cual la fuente genera los bits.
- Retardo de Conmutación. En un Switch ATM, el retardo de conmutación está compuesto de dos partes:
 - Retardo de Conmutación Fijo (*FD*, **Fixed Switching Delay**), que como su nombre lo indica, es un retardo fijo. Este retardo es dependiente de la implementación, y es determinado por la transferencia interna de la celda a través del hardware del Switch. Este retardo es el encontrado cuando el Switch se encuentra con

cero cargas.

- o Retardo de la Cola (*QD, Queuing Delay*), el cual es una parte variable determinada por las colas en el Switch. Debido a que los Switches ATM realizan la conmutación y el multiplexaje estadísticamente, es necesario el empleo de colas con la finalidad de evitar una excesiva pérdida de celdas.

Dependiendo de la arquitectura del Switch es posible que estas colas estén distribuidas en varias partes del Switch. Estas colas introducen una fluctuación de fase en las conexiones que la atraviesan, ya que el tiempo de travesía de las celdas (duración de espera, más tiempo de servicio) es aleatorio y depende de la ocupación de la cola (ésta es a su vez, dependiente del tráfico en la red). Este retardo es determinado por medio del comportamiento de las colas, el cual es caracterizado por una función PDF (**Probability Density Function**) de la longitud de la cola.

Las aplicaciones interactivas en tiempo real (por ejemplo, la videoconferencia) son sensitivas a un retardo acumulativo, denominado de latencia (**latency**). Por ejemplo, las redes telefónicas han sido construidas de tal manera que introduzcan hasta 400 ms de latencia en una ida y vuelta (**round-trip**) cuando se emplean canceladores de eco (sin canceladores de eco, es menos de 24 ms). Debido a esto, las redes que proveen aplicaciones multimedia que soportan audio y

videoconferencia, deben ser también diseñadas con los mismos valores de demora. Cuando la red provee un retardo variable para diferentes paquetes o celdas, entonces se habla de *jitter*.

3.6.2 Parámetros de tráfico. Como se menciono anteriormente cada clase de servicio tiene asociada un grupo de parámetros que definen los niveles mínimos de calidad que se debe ofrecer al usuario de la red para cada categoría de servicio. Estos parámetros podemos clasificarlos en dos grupos, los parámetros de tráfico y los parámetros de calidad de servicio.

- PCR (**Peak Cell Rate**, Tasa de Celdas Máxima): Máxima velocidad con la que una conexión ATM puede enviar celdas sin que se produzcan pérdidas.
- SCR (**Sustained Cell Rate**, Tasa de Celdas Sostenida): Velocidad media con la que una conexión ATM puede enviar celdas.
- MCR (**Minimum Cell Rate**, Tasa de Celdas Mínima): Velocidad mínima que el usuario o la aplicación considera aceptable para efectuar la conexión.
- MBS (**Maximum Burst Size**, Tamaño Máximo de la Ráfaga): Máxima ráfaga que puede enviarse por una conexión sin que se produzcan pérdidas.

3.6.3 Parámetros de calidad de servicio para ATM.

- **MCTD (Maximum Cell Transfer Delay, Máximo Retardo de Transferencia de una Celda):** Es el retardo o latencia máximo permitido, es decir, el tiempo máximo que puede tardar la red en transmitir una celda de un extremo a otro de la conexión.
- **CDVT (Cell Delay Variation tolerante, Tolerancia en la Variación del Retardo de una Celda):** Es el **jitter** o máxima variación que se podrá producir en el retardo de las celdas.
- **CLR (Cell Loss Ratio, Tasa de Perdida de Celdas):** Porcentaje máximo aceptable de celdas que la red puede descartar debido a congestión. Cuando una celda es entregada en el destino con un retardo superior a MCTD se considera perdida.

No todos estos parámetros tienen sentido en todas las clases de servicios, en la tabla 4 se muestran aquellos que normalmente son empleados por cada servicio.

Tabla 4. Resumen de los parámetros empleados por cada servicio ATM.

	CBR	VBR - RT	VBR -NRT	ABR	UBR	UBR +
PCR	SI	SI	SI	SI	NO	NO
SCR	NO	SI	SI	NO	NO	NO
MCR	NO	NO	NO	SI	NO	SI
MBS	NO	SI	SI	NO	NO	NO
MCTD	SI	SI	SI	SI	NO	NO
CDVT	SI	SI	NO	NO	NO	NO
CLR	SI	SI	SI	SI	NO	NO

3.6.4 Calidad de Servicio y Descriptores de tráfico. Uno de los principales beneficios de las redes ATM es que pueden proveer a los usuarios con una Calidad de Servicio (QoS) garantizada. Para poder realizar esto, el usuario debe informar a la red, durante el establecimiento de la conexión, de la clase de tráfico esperada que será transmitida en la conexión y del tipo de Calidad de Servicio que la conexión requiere. La clase esperada de tráfico es descrita por medio de una serie de parámetros de tráfico, mientras que la Calidad de Servicio de la conexión es especificada por una serie de parámetros QoS. El nodo origen

debe informar a la red, durante el establecimiento de la conexión, de los parámetros de tráfico y de la deseada QoS para cada dirección de la conexión solicitada, ya que los mismos pueden ser diferentes en cada dirección de la conexión.

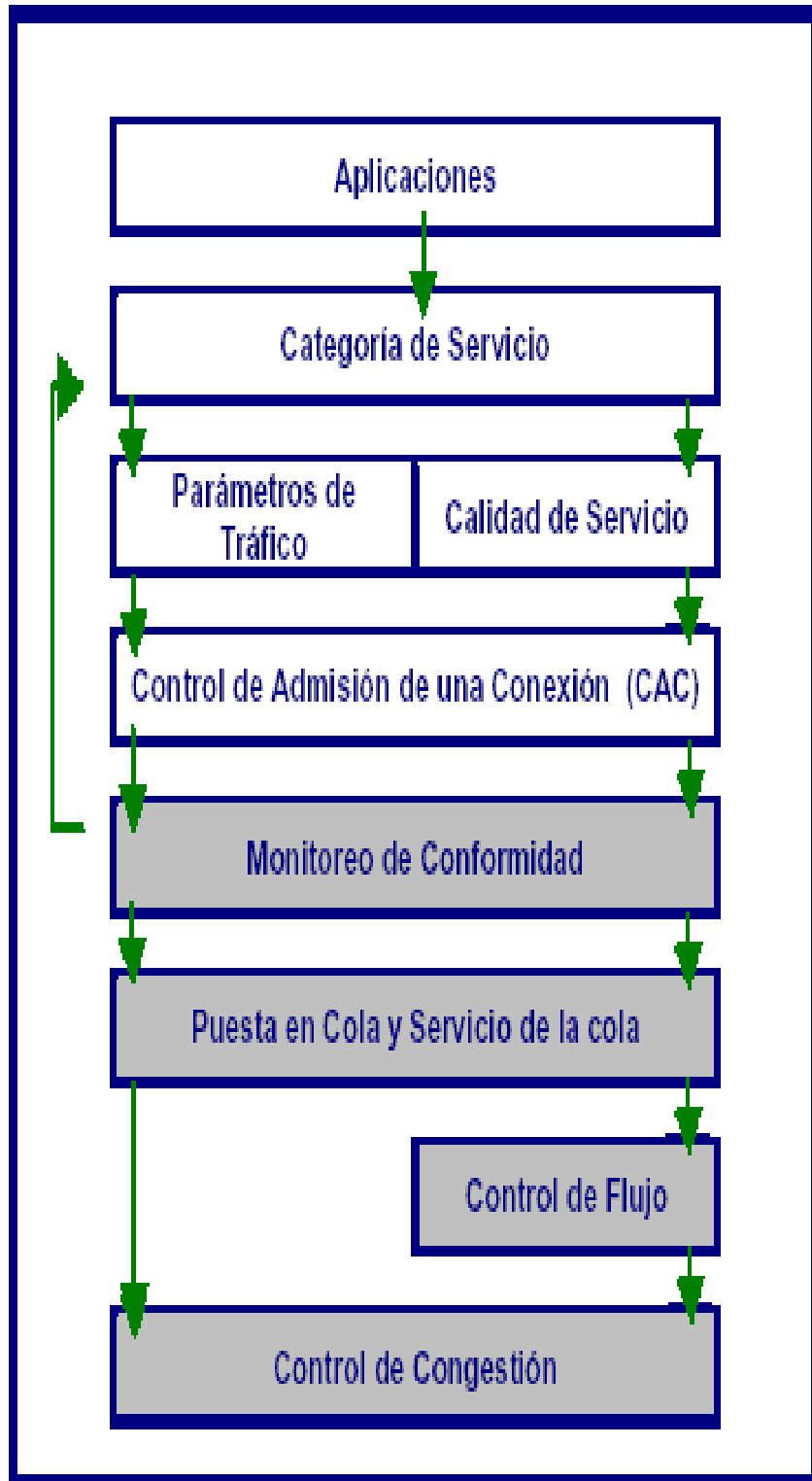
Para una conexión dada, los parámetros de tráfico de una fuente son agrupados en lo que se denomina Descriptor del Tráfico de la fuente, el cual a su vez es un componente del Descriptor de Tráfico de una conexión. Las redes ATM ofrecen una serie específica de categorías de servicio. El usuario debe solicitar a la red, durante el establecimiento de la conexión, una Clase de Servicio (**Class of Service, CoS**) para esta conexión. Las Categorías de Servicio son empleadas para diferenciar entre diferentes tipos específicos de conexiones, en donde cada una de las mismas posee unas características de tráfico y de parámetros QoS particulares. Como resultado, se obtienen las características negociadas de una conexión, la cual constituye el contrato de tráfico.

Un Descriptor del Tráfico de una fuente es una agrupación de parámetros de tráfico para una conexión dada pertenecientes a una fuente **ATM**. Este descriptor es empleado durante el establecimiento de una conexión para capturar las características de tráfico intrínsecas de la conexión solicitada por la fuente particular. El Descriptor de Tráfico de una conexión, especifica las características de tráfico de una conexión.

Este descriptor incluye el Descriptor del Tráfico de una fuente, la Tolerancia a la Variación del Retardo de la Celda CDVT (**Cell Delay Variation Tolerance**), y la definición de conformidad, la cual es empleada para especificar las celdas conformes de una conexión. El Descriptor del Tráfico de la conexión contiene la información necesaria requerida para las pruebas de conformidad de las celdas de la conexión en la UNI.

En la figura siguiente, se presenta el proceso de control del tráfico junto con la serie de funciones asociada a éste.

Figura 7. Funciones asociadas con el control del tráfico.



Durante el establecimiento de la conexión, el nodo solicitante informa a la red del tipo de servicio requerido, de los parámetros de tráfico del flujo de datos en cada dirección, la CDVT (**Cell Delay Variation Tolerance**) o Tolerancia de Variación del retardo de Celdas y de la QoS solicitados para cada dirección. Definición de conformidad, basada en una o más aplicaciones del algoritmo de tasa de células genéricas o *GCRA* (**Generic Cell Rate Algorithm**).

3.7 MECANISMOS DE CONTROL.

Para proporcionar un sistema de calidad, es necesario utilizar una serie de mecanismos que permitan controlar la red.

3.7.1 Control de Admisión de Conexión (CAC). El Control de Admisión de Conexión, CAC, representa una serie de acciones tomadas por la red durante la fase de establecimiento (**Call Set-up**) de un SVC, durante el establecimiento de un PVC o durante la fase de renegociación de alguno de éstos con la finalidad de:

- Aceptar o rechazar una nueva conexión VCC (PVC o SVC).
- 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, establecidas 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 o rechazo de la conexión. Para cada solicitud (**Request**) de un VCC, CAC emplea la siguiente información para poder tomar la decisión de admisión o 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.

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. La función CAC 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). CAC considera los recursos del Switch que están disponibles para los diferentes servicios.

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 incluido 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. Para servicios VBR el enlace se estaría desperdiciando si se asigna un ancho de banda pico, por tanto, 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.

La asignación estadística asigna el ancho de banda teniendo como base PCR, SCR y MBS. 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 siguiente se muestra una estrategia comúnmente empleada para la asignación del ancho de banda para cada una de las categorías de servicio.

Tabla 5. Asignación de ancho de banda según la categoría de servicio.

Categoría de servicio	Ancho de banda asignado
Constant Bit Rate (CBR).	$PCR \leq V_{bw} \leq \text{Link Rate}$
Variable Bit Rate (VBR).	$SCR \leq V_{bw} \leq PCR$
Available Bit Rate (ABR).	$MCR \leq V_{bw} \leq PCR$
Unspecified Bit Rate (UBR).	No se asigna V_{bw}

3.7.2 Adecuación del tráfico (Traffic Shaping). El tráfico se adecua para eliminar picos de ráfagas entregando un perfil de tráfico predecible más manejable con menos celdas perdidas y menos estrés en los recursos de la red. ATM requiere más que convertir paquetes o frames en Celdas, requiere adecuar el tráfico para asegurarse que el intervalo entre las celdas conformará un pico inferior al total del ancho de banda disponible.

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 se adecue 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.

Nota: el retardo pequeño de las Celdas permite una variación de retardo pequeña, CDV. Esto es útil para el transporte isócrono (sensible al retardo), los cuales necesitan enviar datos a intervalos fijos (por ejemplo, la voz). Pequeñas variaciones de retardos permiten transmisiones isócronas virtuales. Para adecuar el tráfico en un flujo isócrono, con intervalos regulares de tiempo a la salida, se usan esquemas de adecuación de tráfico (Traffic Shaping). El **Leaky Bucket** (Estado Continuo de Recipientes con Goteo) es un esquema de

adecuación de tráfico que utiliza un buffer de un tamaño finito en donde se coloca el tráfico entrante. Los datos almacenados se van sacando y enviando a la red a una tasa ρ . El exceso de datos que no cabe en el buffer se desecha o descarga. El algoritmo **Leaky Bucket** tiene el efecto de adecuar el tráfico de ráfaga en un flujo de celdas igualmente espaciadas, cada una de ellas emitidas $1/\rho$ unidades de tiempo después de la celda previa. El tamaño del buffer limita el retardo de la celda. De aquí que, para limitar el CDV, se requiere de un buffer pequeño.

Cada Celda que pasa por la cubeta con orificio, es controlada permitiéndosele el paso. La cubeta es vaciada en un valor específico, el cual se ilustra por el tamaño del orificio. Las Celdas que llegan a una velocidad mayor al de la capacidad del orificio, son aceptadas siempre y cuando no rebase la capacidad de la cubeta (el tamaño de la cubeta define la tolerancia al **jitter**). Mientras más capacidad de la cubeta, se puede enviar ráfagas de datos a una mayor velocidad.

GCRA ($I; L$), con I , que corresponde al rango de vaciado de la cubeta (tamaño del orificio), y L , que corresponde a la capacidad de la cubeta.

El algoritmo genérico de la tasa de celdas GCRA (**Generic Cell Rate Algorithm**) se usa para definir la relación entre la tasa pico de la celdas PCR y la tolerancia a la variación de retardo de las celdas CDV. Se definen dos algoritmos para el GCRA:

- Algoritmo de Programación Virtual (**Virtual Scheduling**)
- Algoritmo de Estado Continuo de Recipientes con Goteo (**Leaky Bucket**).

El Algoritmo de Programación Virtual, emplea un intervalo entre celdas I que se usa para determinar un tiempo teórico de arribo TAT (**Teoretical Arrival Time**) para una conexión virtual. El intervalo es entre celdas igualmente espaciadas para una tasa de celdas dada.

- Si el tiempo de arribo es después de TAT, entonces la celda es ajustada y el siguiente TAT será el tiempo de arribo de la celda más el intervalo I .
- Si es antes que el TAT y después que el TAT menos el valor tolerancia, entonces la celda es ajustada, pero el siguiente TAT será el TAT actual más I .
- Si es después que el TAT menos el valor de tolerancia, entonces la celda no será ajustada y puede ser etiquetada o desechada. EL siguiente TAT será el TAT actual más el intervalo I .

El contrato de tráfico entre el cliente y la red, especifica las características negociadas de una conexión. Este contrato puede consistir de:

- El Descriptor de Tráfico de una conexión.

- Una serie de parámetros QoS para cada dirección de la conexión.
- La definición de cumplimiento de la conexión.

La definición exacta del cumplimiento no está definida en los estándares, es decir, es dependiente de la red. La red, basada en acciones de la función UPC, puede decidir si una conexión es conforme o no. En el caso de que la conexión sea conforme, la red debe soportar la QoS para todas las conexiones conformes.

3.7.3 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 (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 celdas con los adecuados parámetros de QoS. Para esto se requiere de un método que permita a los nodos ATM identificar rápidamente celdas que puedan ser descartadas de aquellas celdas 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 encabezamiento de la celda denominado bit de prioridad de la celda (CLP).

3.7.4 Control de Congestión. La congestión se define como un estado de los elementos de la red en el cual, debido a la sobrecarga de tráfico, la red no está en condiciones para garantizar un QoS para conexiones ya establecidas y para solicitudes de nuevas conexiones. El control de congestión trata de minimizar los efectos de la congestión y evitar la difusión del problema. El control de congestión podría reducir la tasa pico de bits ofrecida a un usuario.

Existen tres mecanismos basados en tasas de bits para el control de flujo: EFICI (**Explicit Forward Congestion Indicator**), ER (**Explicit Rate**), y GFC (**Generic Flow Control**). Cuando se tiene necesidad de transporte de información de varios puertos de entrada a un puerto de salida, el resultado puede ser congestión. Esta congestión puede producirse en el Switch y en el puerto de salida, pero si la congestión no se resuelve, puede producir sobrecarga al Switch.

El descarte de una celda puede ocasionar la pérdida de un paquete completo, por lo que tendría que hacerse una recuperación de errores punto a punto, además de requerirse la retransmisión de la información por un protocolo de más alto nivel. Para prevenir lo anterior, un Switch ATM puede descartar celdas con base en frames usando Descarte de Celdas Primeras (**EPD, Early Packet Discard**) y Descarte Parcial de Celdas (**PPD, Partial Packet Discard**).

EPD descarta el paquete completo después de detectar que la congestión sigue. Si usar EPD no es suficiente para disminuir el evento de congestión, PPD descarta todas las celdas remanentes con excepción de la celda indicadora de fin de paquete, del paquete que ha sido parcialmente descartado. EPD y PPD juntos pueden reducir significativamente la pérdida de paquetes y mejorar el **throughput**, especialmente en ambientes TCP/IP usando UBR (es mejor si se hace el descarte por circuito virtual).

3.8 CONFORMACIÓN Y VIGILANCIA DEL TRÁFICO

Puesto que ATM esta concebida para ofrecer calidad de servicio a sus usuarios es importante que dicho usuario no exceda los límites de ancho de banda pactados en su contrato de servicio. Las tareas dedicadas a este fin, que pueden ser desarrolladas tanto por el usuario o por la red, se denominan conformación de tráfico o **traffic shaping**. Además, la red ha de verificar que efectivamente el usuario no excede sus límites, y aplicar medidas correctoras en caso contrario; a esta labor de vigilancia se la denomina **traffic policing**, o vigilancia de tráfico, ambas tareas, conformación y vigilancia, están íntimamente relacionadas.

Los mecanismos de control de tráfico dependen mucho de la categoría de tráfico que se este tratando. En el caso de tráfico CBR el usuario dispone de una capacidad asignada a él de forma absolutamente estática, por lo que la labor de conformación y vigilancia es muy simple.

Ejemplo: Un usuario contrata un circuito CBR con un PCR de 10000 celdas/s (equivalente a 4,24 Mb/s), en este caso el usuario estará autorizado a introducir una celda en la red cada 100ms, si su conexión física con la red ATM es un enlace OC3 (155.52 Mb/s) el usuario tardara 2.7 ms en enviar una celda, por lo que después de cada envío deberá estar como mínimo 97.3 ms sin enviar otra, ya que de lo contrario excedería su PCR y el conmutador descartara todas las celdas que llegaran antes del plazo previsto.

El otro extremo es el de tráfico UBR. Aquí el usuario no especifica ningún parámetro de calidad de servicio, por lo que recibe un servicio **best effort** (Mejor esfuerzo); normalmente esto se traduce en que los circuitos UBR utilizan la capacidad sobrante al resto de tráfico en la red, con lo que un usuario UBR puede llegar a emplear su enlace físico al máximo de su capacidad sin que se le apliquen medidas correctoras, pero este tráfico de mínima prioridad será desplazado en cuantos otros usuarios tengan alguna necesidad. El tráfico ABR fija una capacidad mínima requerida (MCR) y una máxima prevista (PCR), en caso de ráfagas la red asegura la disponibilidad de MCR, pero no promete nada respecto a PCR, que estará sujeto a disponibilidad.

Dado que el usuario recibe realimentación de la red en caso de congestión, se supone que si inyecta un tráfico superior a MCR y la red no puede soportarlo, recibirá mensajes de congestión que le harán reducir el ancho de banda. No se establece la duración de las ráfagas (parámetro MBS) ya que estas pueden tener

una duración considerable en ratos de baja carga en la red.

El control de tráfico VBR es el más complejo. Los parámetros especificados, SCR, PCR y MBS permiten especificar un algoritmo **Leaky Bucket** (Pozal Agujereado).

En el caso más sencillo se utilizaría SCR como ancho de banda r , MBS sería el tamaño del **buffer** C , y PCR el ancho de banda con que el **host** envía los datos al pozal (del algoritmo **Leaky Bucket**). Pero cabe aclarar que el ancho de banda medio de un tráfico VBR no debería nunca superar el SCR.

4. CALIDAD DE SERVICIO (QoS) EN REDES IP

Internet ofrece Calidad de Servicio QoS de manera muy sencilla: la entrega de datos de mejor esfuerzo (**Best Effort**) punto a punto. Además, debe brindar las garantías de confiabilidad y desempeño que las empresas exigen y a las que están acostumbradas en sus redes privadas. Antes de que las aplicaciones multicast de IP y las aplicaciones en tiempo real (como videoconferencia) puedan trabajarse ampliamente, la infraestructura de Internet debe modificarse para apoyar diferentes niveles de servicios. Las empresas no colocarán datos, voz y aplicaciones de misión crítica hacia las redes de IP públicas, hasta que reciban servicios confiables, medibles y garantizados.

QoS se refiere a la habilidad de un elemento de la red (por ejemplo, una aplicación, un **host** o un Enrutador), para tener algún nivel de seguridad de que sus requerimientos de tráfico y servicios se pueden satisfacer. Habilitar Calidad de Servicio QoS requiere la cooperación de todas las capas de la red, así como, de cada elemento de los extremos de la red.

Los desarrollos de la industria, específicamente en la gestión de redes para tecnología ATM y **Frame Relay**, ha producido muchos de los mecanismos requeridos para llevar a cabo QoS en redes IP. El IETF se ha involucrado en la

implementación de ATM desde que ATM posee implicaciones significativas como protocolo de **backbone**¹¹ para redes IP que forman la base para la Internet y la World Wide Web. El IETF ha desarrollado dos normas que proporcionan capacidades de QoS: Servicios Integrados (IntServ, las aplicaciones tienen la habilidad de escoger entre el múltiples y controlados niveles de QoS para sus paquetes de datos) y Servicios Diferenciados (DiffServ).

El servicio tiempo real requiere servicios garantizados, y esas garantías no pueden lograrse sin las reservaciones. El protocolo de Reservación de Recursos, RSVP (RFC 2205), asume que los recursos son reservados para cada flujo que requiere QoS a cada salto de Enrutador en la ruta entre el receptor y transmisor usando señalización de extremo a extremo. Esto, a su vez, requiere un estado específico de flujo en los Enrutadores, lo cual representa un cambio importante y fundamental al modelo de Internet. También proporciona una manera de comunicar los requerimientos de la aplicación a los elementos de la red a lo largo de la ruta, y para llevar información de administración de QoS entre los elementos de la red y la aplicación.

La recomendación de ITU-T define Calidad de Servicio como *“El efecto colectivo del rendimiento del servicio el cual determina el grado de satisfacción del usuario del servicio”*. Este grado de satisfacción se expresa mediante las siguientes medidas cuantitativas:

¹¹ Parte principal de la red.

- Retraso de extremo a extremo, es el tiempo tomado por un paquete en viajar a través de la red de un extremo a otro.
- Variación del retraso o **jitter**, o sea la variación del retraso encontrada por paquetes similares que siguieron la misma ruta a través de la red.
- **Throughput** o Caudal.

Otras métricas adicionales:

- Disponibilidad del servicio, que es la confiabilidad de la conexión del usuario al dispositivo de Internet.
- Tasa de pérdida de paquetes, tasa a la cual los paquetes van saliendo de la red o se pierden o se corrompen (que es cuando unos bits se cambian dentro del paquete durante el tiempo que están viajando a través de la red).

ATM fue diseñado para .Calidad de Servicio.; permite circuitos virtuales punto a punto y punto a multipunto para que se establezcan con parámetros de Calidad de Servicio previamente especificados. ATM provee una amplia variedad de mecanismos de Calidad de Servicio con una amplia variedad de Categorías de Servicio, los cuales ofrecen un excelente control de los parámetros de tráfico solicitados y administrados. ATM ofrece un sofisticado mecanismo de

señalización que puede utilizarse eficazmente por cualquier agente que necesite encargarse del aprovisionamiento de la Calidad del Servicio.

4.1 CALIDAD DE SERVICIO DE EXTREMO A EXTREMO

El objetivo de la construcción de redes con arquitectura de Calidad de Servicio es acercar los dispositivos de los extremos de manera que incrementen el rendimiento y reduzcan la demora que subyace en la red. Para lograrlo, la red debe mejorar los modelos de servicio. Se han propuesto tres modelos de servicio que se han implementado hasta la fecha. Del primero se dará una explicación a continuación, el segundo y el tercero serán tocados mas adelante.

1. Los Servicios de Best Effort o el mejor esfuerzo.
2. Los Servicios Integrados.
3. Los Servicios Diferenciados.

4.1.1 Servicios de Best Effort. En este modelo, una aplicación envía datos siempre que lo desee y tantas veces lo desee sin requerir permiso de nadie. Los elementos de la red tratan de estar a su mejor nivel para enviar los paquetes a su destino sin ninguna limitación en cuanto a demora, latencia o jitter. Pero si existen inconvenientes en el proceso de envío, los elementos de red dejarán de realizarlo y no informarán ni al emisor ni al receptor (por ejemplo, si no reciben un reconocimiento de recibo o **acknowledgement**, después de intentar enviar un

paquete o aún después de haber transmitido un cierto número de veces). Por lo que la responsabilidad de asegurar que el paquete llegó a su destino corresponde realmente al sistema final. Un ejemplo de este servicio es el envío actual de las redes IP.

Los mecanismos de Calidad de Servicio que se han propuesto para IP son:

- **IntServ**, Servicios Integrados RSVP.
- **Diffserv**, Servicios Diferenciados MPLS.

4.1.2 MPLS aided DiffServ. A continuación se considerará cómo los Servicios Diferenciados pueden suministrarse cuando los Enrutadores utilizan mecanismos de Conmutación de Etiqueta para el reenvío de paquetes. El FEC Forwarding Equivalence Class, es un concepto de MPLS que agrupa paquetes que necesitan el mismo desempeño de reenvío en ellos. Similarmente el BA o Agregados de Comportamiento y OA Agregados de Ordenamiento son conceptos en DiffServ que se refieren a los paquetes que necesitan se les dé el mismo tratamiento y la misma clasificación respectivamente. Por ejemplo, Los Agregados clasificados forman un conjunto de Agregados de Conducta.

Ahora cuando paquetes que han sido marcados por Diffserv codifican que los puntos llegan a una red de MPLS, allí necesita transferir información proporcionada por los puntos del código hacia la etiqueta MPLS. Esto sería

necesario hacerlo si MPLS tuviera la capacidad de tomar decisiones con respecto a los requerimientos de Servicios Diferenciales, para los cuales, los paquetes han sido marcados, en otras palabras, esto haría que MPLS fuera capaz de ejecutar Servicios Diferenciales.

4.2 MANEJO DE CONGESTION Y TRÁFICO

En la Tabla 6 se relacionan los distintos tipos de herramientas que se disponen para asegurar una QoS dentro de una red IP. Se trata de mecanismos que previenen o manejan una congestión, distribuyen el tráfico o incrementan la eficiencia de la red. Los protocolos involucrados en asegurar la calidad de servicio son los indicados en la misma Tabla; a los mismos se refiere como mecanismos de señalización. En el ítem siguiente se analizan con detalle a los mismos.

Tabla 6. Control de congestión en el buffer de datos.

Herramienta	Uso
-FIFO	<p>-(First In, First Out). El primer mensaje en entrar es el primero en salir. Este es el mecanismo de QoS por Default en las redes IP. Es válido solo en redes con mínima congestión. No provee protección, no analiza el ancho de banda ni la posición en la cola de espera.</p>
-PQ	<p>-(Priority Queuing). Este mecanismo de control de congestión se basa en la prioridad de tráfico de varios niveles que puede aportar el encabezado del datagrama IP (ToS Type of Service). Se trata de 3 bits disponibles en el Byte 2 del encabezado de IPv4 (bits de precedencia).</p>
-CQ	<p>-(Custom Queuing). Este mecanismo se basa en garantizar el ancho de banda mediante una cola de espera programada. El operador reserva un espacio de buffer y una asignación temporal a cada tipo de servicio. Es una reservación estática.</p>

Tabla 7. Control de tráfico.

Herramienta	Uso
-WRED	<p>-(<i>Weighted Random Early Detection</i>). Trabaja monitoreando la carga de tráfico en algunas partes de las redes y descarta paquetes en forma random si la congestión aumenta. Está diseñada para aplicaciones TCP debido a la posibilidad de retransmisión. Esta pérdida en la red obliga a TCP a un control de flujo reduciendo la ventana e incrementándola luego en forma paulatina. Un proceso de descarte generalizado, en cambio, produce la retransmisión en "olas" y reduce la eficiencia de la red. La versión ponderada WRED realiza el drop de paquetes de forma que no afecta al tráfico de tipo RSVP. Una versión superior debería considerar el tráfico de aplicación.</p>
-GTS	<p>-(<i>Generic Traffic Shaping</i>). Provee un mecanismo para el control del flujo de tráfico en una interfaz en particular. Trabaja reduciendo el tráfico saliente limitando el ancho de banda de cada tráfico específico y enviándolo a una cola de espera. De esta forma permite una mejor performance en topologías con tasa de bit diferentes. Este control de tráfico se relaciona con CAR.</p>

Tabla 8. Incremento de la eficiencia y señalización.

Herramienta	Uso
-LFI	<p>-(Link Fragmentation and Interleaving).</p> <p>El tráfico interactivo como Telnet y VoIP es susceptible de sufrir latencia y jitter con grandes paquetes en la red o largas colas en enlaces de baja velocidad. Se basa en la fragmentación de datagramas y el intercalado de los paquetes de tráfico.</p>
-RSVP	<p>-(Resource Reservation Protocol). Se trata de implementar el concepto de Señalización. Se dispone de dos tipos de señalización: en-banda (por ejemplo los bits de precedencia para ToS) y fuera-de-banda (mediante un protocolo de comunicación como el RSVP). Este protocolo permite que un host o un router aseguren la reservación de ancho de banda a lo largo de la red IP.</p>
-RTP-HC	<p>-(Real-Time Protocol-Header Compression). El protocolo de tiempo real RTP es estudiado por separado más adelante. La compresión del encabezado permite mejorar la eficiencia del enlace en paquetes de corta carga útil. Se trata de reducir los 40 bytes de RTP/UDP/IP a una fracción de 2 a 5 bytes, eliminando aquellos que se repiten en todos los datagramas.</p>

No todas las herramientas disponibles son usadas en los mismos **routers**. Por ejemplo, la clasificación de paquetes, el control de admisión y el manejo de la configuración se usan en los router de borde (**edge**), en tanto que en los centrales (**backbone**) se gestiona la congestión. El tratamiento de la congestión se fundamenta en el manejo de las colas en buffer mediante diferentes técnicas. El buffer es la primera línea de defensa frente a la congestión. El manejo correcto (mediante políticas de calidad de servicio) del mismo permite determinar el servicio de calidad diferenciada. Una segunda defensa es el control de flujo. El problema del control de flujo en TCP es que se ha planea de extremo-a-extremo y no considera pasos intermedios. En TCP cada paquete de reconocimiento (**Acknowledgment**) lleva un crédito (**Window**) con el tamaño del buffer disponible por el receptor. Un sobreflujo de datos en los routers de la red se reporta mediante el mensaje **Source Quench** en el protocolo ICMP. Estos mecanismos son ineficientes y causan severos retardos en la conexión.

4.3 PRIORIZACION DE TRÁFICO

-ToS/IEEE 802.1Q. El standard IEEE 802.10 y el 802.1Q fueron propuestos para el manejo de las redes VLAN; este último es el utilizado con regularidad. En el standard 802.1Q se define el VLAN **Tagging**¹² **Switch** que permite una identificación de la VLAN y la posibilidad de priorización del servicio. La trama del paquete en capa MAC incluye 4 Bytes adicionales al IEEE 802.3 que se colocan

¹² Etiquetado

luego de las direcciones MAC y antes del **Type/Length**¹³. Los 4 Bytes son indicados a continuación. Obsérvese la presencia de 3 bits para prioridad de tráfico y 12 bits para identificación de VLAN.

Tabla 9. Campos para manejo de Tipo de Servicio en IEEE 802.3

Campo	Explicación
-TPID	(Tag Protocol Identifier) . 2 Bytes. Usados para identificación del protocolo. En Ethernet es hexa=8100.
-TCI	(Tag Control Information) . 2 Bytes usados para las siguientes funciones:
.UP	3 bits para prioridad del usuario (User Priority). Se trata de CoS desde 0 a 7. Esta información permite poner en práctica la CoS definida en IEEE 802.1p.
.CFI	(Canonical Format Indicator) . 1 bit para ser usado por Token Ring.
.VLANI	(VLAN Identifier) . Este campo de 12 bits permite identificar la VLAN (válido desde 1 a 1005). Permite la interoperación entre diferentes productores.

¹³ Tipo/Longitud.

El mecanismo que se define para la CoS (clase 0 a 7 desde alta a baja prioridad) se compone de las colas de recepción y transmisión. El umbral para extraer los paquetes de la cola de recepción son:

- Clase de servicio CoS 0/1: umbral del 50% (máxima prioridad).
- CoS 2/3: umbral al 60%.
- CoS 4/5: umbral al 80%.
- CoS 6/7: umbral al 100% (mínima prioridad).

En transmisión existen dos colas la de alta y baja prioridad. Su relación con la CoS es la siguiente:

- Cola de baja prioridad (corresponde al umbral del 80%) y CoS 0/1: umbral al 40%; con CoS 2/3: umbral al 100%.
- Cola de alta prioridad (corresponde al umbral del 20%) y CoS 4/5: umbral al 40%; con CoS 6/7: umbral al 100%.

Por ejemplo, una puerta del switch que no fue configurada para CoS tiene un valor por **default** ¹⁴ de umbral del 100%. Un servicio clase CoS=2/3 en el buffer de recepción (entrada al switch) tiene un umbral al 60% para la extracción de paquetes, mientras que en el de transmisión se coloca en alta prioridad (umbral al

¹⁴ Defecto.

20%) y con CoS=2/3 tiene una prioridad adicional del 80%.

-QoS/RFC. El campo de precedencia en el encabezado de IPv4 permite definir varios tipos de servicio ToS. Se trata de 3 bits que por razones históricas tienen diferentes nombres (**routing, priority, etc.**) y que pueden ser usados para signar prioridad. Se aplica un control de acceso extendido EACL para definir la política de la red en términos de congestión. En redes heterogéneas se debe mapear este tipo de servicio en equivalentes (tag **switch, Frame Relay** y ATM).

Con Los bits de precedencia se pueden realizar 3 tipos de acciones: routing basado en políticas PBR (**Policy-Based Routing**) (por ejemplo direcciones IP, port de TCP, protocolo, tamaño de paquetes, etc.); propagar la política de QoS mediante el protocolo de routing BGP-4 y desarrollar una política de tasa de acceso contratada CAR. La CAR (**Committed Access Rate**) se ofrece especificando políticas de tráfico y ancho de banda. El umbral de CAR se aplica a la puerta de acceso para cada puerta IP o por flujo de aplicación individual. Una técnica disponible para manejar el CAR es el **netflow switch** que se comenta más adelante. Algunas opciones de política de CAR son:

- Política de prioridad:
 - CAR máximo (el exceso de ancho de banda es descartado);
 - CAR **premium** (el exceso es señalado con un nivel de preferencia

más bajo);

- CAR **best effort** (por encima de un umbral se cambia la preferencia y sobre otro los paquetes son eliminados);

- Política de asignación:

- CAR por aplicación (diferentes políticas son usadas en distintas aplicaciones; por ejemplo bajo nivel para HTTP).
- CAR por puerta (los paquetes que ingresan por un port son clasificados con alto nivel de prioridad).
- CAR por dirección (puede diferenciarse entre la dirección IP de origen y destino y asignar la prioridad en cada caso).

4.4 IP QOS SOBRE ATM

IP y ATM pretenden ser complementarios en vez de ser competidores. IP ha ganado una amplia aceptación, pero la simpleza de su arquitectura que la ha llevado a su lugar de privilegio se convierte en una debilidad, a la cual los grupos de investigación están tratando de apuntar. La Calidad del servicio es una de las debilidades importantes de IP. Sin embargo ATM presume de poseer una arquitectura la cual tiene métodos relativamente sofisticados de proveer Calidad de Servicio QoS. Sin embargo, debido a razones de costo, inercia y complejidad, ATM no ha logrado su status ideal. Así, el segundo paso es obvio para unir las fortalezas complementarias de IP y ATM para lograr IP QoS sobre ATM.

Mientras ATM asegura QoS, no puede hacer nada sobre la capa 2, lo cual significa que todos los flujos de capa 3 que se han agregado no pueden ser diferenciados por ATM, y así, todos ellos finalmente terminan compitiendo unos contra otros por los mismos recursos. Por tanto, existe necesidad de implementar una manera de una granularidad más fina en cuanto al control del flujo de tráfico y esto se hace mejor en la capa 3. Ahí es donde los Servicios Integrados IntServ RSVP y Servicios Diferenciados entran a funcionar. Esta sección revisará el estado actual de los asuntos en la interoperación simbiótica de ATM y las arquitecturas IP en QoS. Primero se consideran Servicios Integrados sobre ATM y luego Servicios Diferenciados sobre ATM. Posteriormente se tratará el MPLS. Muchas de estas arquitecturas están basadas en mayor o menor medida sobre señalizaciones RSVP.

4.4.1 IntServ sobre ATM. La traducción de dirección requerida entre IP y ATM la realizan los dispositivos de frontera mediante el protocolo ATM **Address Resolution Protocol (ATMARP)**, Protocolo de Resolución de Direcciones ATM. El tráfico que necesita cruzar una frontera de la subred lógica IP, necesariamente tiene que pasar a través de un Enrutador IP, el cual es un dispositivo de frontera ATM, cuya red subyacente ATM ha sido configurada para pertenecer a la subred lógica IP y que posiblemente está sirviendo a muchas otras subredes lógicas IP. Esto se muestra esquemáticamente en la siguiente figura.

Figura 8. Los paquetes ip IntServ atraviesan una nube intermedia ATM

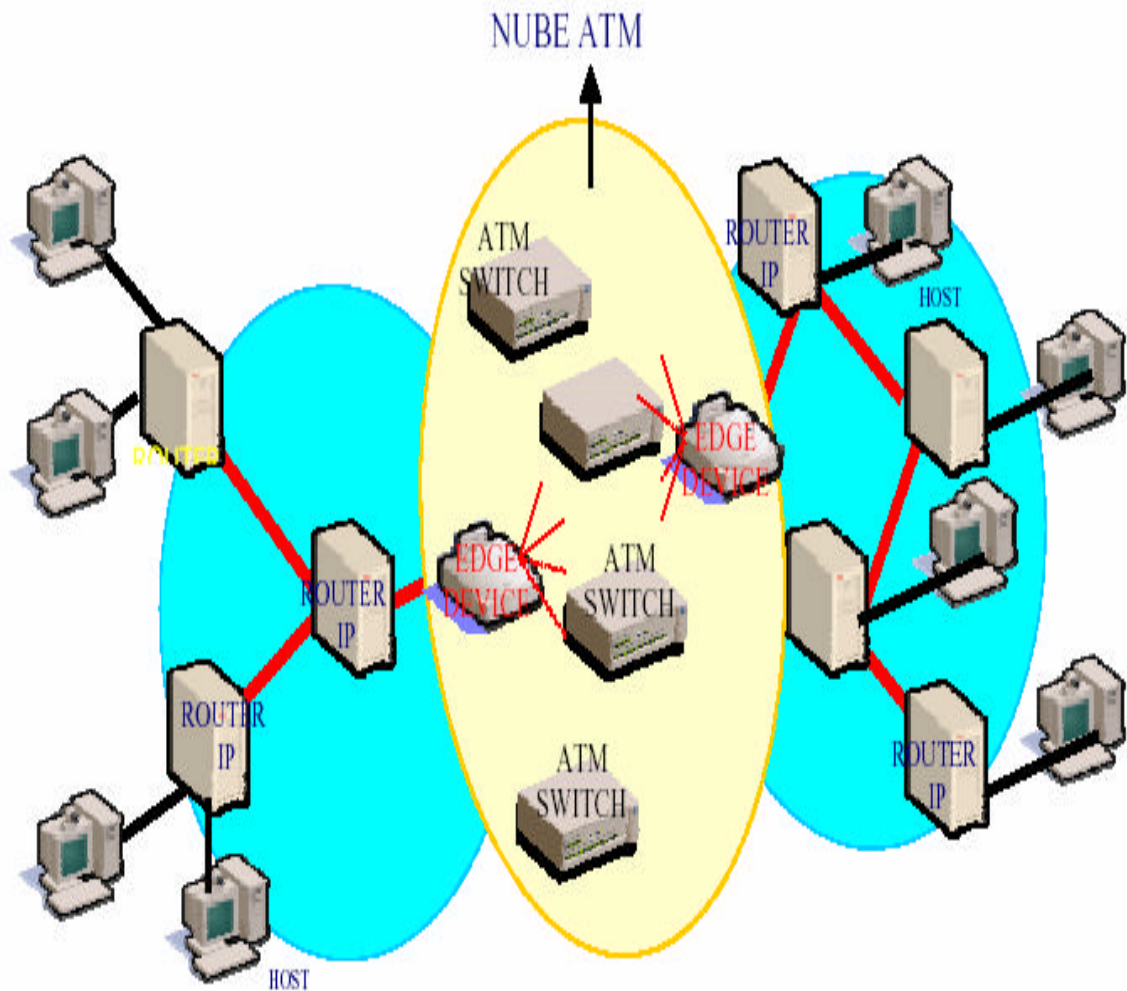


Figura 8. A pesar de la unión de dos tecnologías diferentes, el IntServ que presta ATM con el (Edge Device)¹⁵, permite que los paquetes fluyan sin ningún contratiempo.

El dispositivo de frontera se encarga de la traducción de la dirección requerida y habla ambos lenguajes. Intserv, como se ha visto antes, usa RSVP como un protocolo de señalización para proporcionar QoS.

¹⁵ Dispositivo de orilla.

RSVP no es un protocolo de enrutamiento. En vez de esto utiliza un camino establecido por los protocolos de enrutamiento estándares como el OSPF o el RIP para determinar su próximo salto. RSVP es un protocolo de la capa de transporte si se sigue el modelo de siete capas de OSI. Por lo tanto con los cambios en su enrutamiento debido a los enlaces o a fallas en los nodos, RSVP necesita actualizar el estado de los caminos a los varios enlaces. Esto se hace mediante la emisión por parte del emisor de rutas del mensaje y la respuesta por parte del receptor de mensajes RESV periódicamente.

Los Servicios Integrados utilizan RSVP como protocolo de señalización para suministrar Calidad de Servicio para tráfico IP. Por tanto, cuando la pila de Protocolos IP se implementa sobre ATM y tal conformidad se asegura con los modelos de Servicios Integrados, hay necesidad de mapear la señalización RSVP con la señalización de ATM. Tal mapeado consiste en dos problemas: uno que involucra los parámetros de Calidad de Servicio que se utilizan y el otro la administración de los circuitos virtuales VC. La especificación QoS en el modelo de Servicios Integrados no es idéntica a la de ATM. Se requiere una traducción de QoS para IP QoS sobre ATM. Se agrega tráfico de IP en flujos. ATM, en la capa dos, asigna Circuitos Virtuales a estos agregados. Allí se crea el problema de cuál flujo va en cual circuito virtual VC y cuántos VC se deben crear.

A continuación se consideran algunos de los problemas que están trabajándose para Intserv/RSVP sobre ATM.

- **Control y rutas de mensaje de datos:** En RSVP la ruta del mensaje instala el camino que los datos deben tomar. Esto significa que la ruta del mensaje RSVP debe seguir la misma ruta de los datos. En el caso de una nube ATM subyacente, los puntos de ingreso y egreso debe ser el mismo para los mensajes de control RSVP como para los datos.
- **Mapeo de servicios entre Internet y ATM:** Los Servicios Integrados para IP consisten esencialmente en dos tipos de servicios, además del servicio de Best Effort, que son: Servicio de Carga Controlada y Servicio Garantizado. Los servicios ATM incluyen: CBR Constant Bit Rate, rt-VBR real time Variable Bit Rate, o nrt-VBR non real time Variable Bit Rate, UBR Unspecified Bit Rate y ABR Available Bit Rate. Los siguientes mapeos se han propuesto entre ATM e Integrated Service para IP (IIS IP).

Tabla 10. Mapeos entre ATM e IIS.

ATM	IIS, SERVICIOS INTEGRADOS INTERNET
CBR o rt-VBR (Tasa de bit constante o Tasa de bit Variable en tiempo real)	SERVICIO GARANTIZADO
Nrt-VBR o ABR (mínima tasa de celdas)	CARGA CONTROLADA
UBR o ABR	BEST EFFORT

- **Conexiones QoS para RSVP sobre ATM:** La conexión PVC tiene asignado un ancho de banda fijo que entrega datos entre los puntos terminales especificados, y es el dispositivo del extremo quien debe proporcionar los servicios para asignar varios flujos sobre un VC dado. Es importante tener especial cuidado con los PVCs para evitar la subutilización de los recursos debido a PVCs desocupados. Los PVCs frecuentemente son configurados de manera que el ancho de banda disponible sea menor de lo que se requiere, si todos los PVCs se utilizan simultáneamente. Sin embargo, mientras se use Intserv/RSVP sobre ATM, se necesita asegurar que toda la QoS que han sido garantizadas se satisfagan.
- **Flexibilidad de Circuitos Virtuales Conmutados, SVCs:** La metodología SVC crea un circuito virtual cuando se necesita y, cuando termina la transferencia, el ancho de banda del circuito se devuelve al conjunto de recursos global. Se deben tomar en cuenta los costos y el tiempo necesarios para configurar los SVCs antes de decidir utilizarlos para enrutar tráfico QoS. ATM se basa en modelado estadístico de varios tipos de datos. Los datos tienen que llegar en el orden enviado, lo que significa que si se pierde una celda, habrá que realizar una retransmisión. Esto hace que la congestión sea un punto de preocupación para los usuarios de ATM. Mientras se desarrolla un sistema fiable para el manejo de la congestión, se empleará el método PVC.

- **Soporte IP multicast:** Se requiere aprovechar los VCs (Circuitos Virtuales) punto a multipunto. Sin embargo existe una dificultad: ATM sólo permite una Clase de Servicio para todos los destinatarios de VCs. Para permitirles a los destinatarios requerimientos diferentes de QoS, es necesario tener VCs separados para cada destinatario. Esto da lugar a que se incrementen los problemas.
- **Ruta corta usada por el Protocolo de Resolución del Siguiete Salto (NHRP):** Cuando una nube ATM consiste en varias subredes lógicas IP, es posible usar el protocolo NHRP **Next Hop Resolution Protocol**, para ir directamente de un Enrutador en una subred a un Enrutador de otra subred. Pero esto significaría almacenamiento de información adicional del mapeo de direcciones. El análisis del costo beneficio aún es un problema abierto.
- **Servicios Simultáneos de Best Effort:** El modelo de Servicios Integrados, no permite retirar los paquetes que no adecuados. Esto es para evitar ataques a la seguridad a la denegación del servicio. La implementación de este servicio sobre ATM es un asunto que aún está por resolver.
- **Diversificar VCs:** Se ha propuesto que ATM incorpore VCs punto a multipunto en donde diferentes VCs tienen requerimientos de QoS diferentes. El problema es la preocupación de cómo las celdas se van a retirar cuando el tráfico va de un segmento con gran ancho de banda y

con QoS, ingrese a un segmento con un ancho de banda más pequeño.

- **Renegociación Dinámica de QoS:** RSVP permite cambios dinámicos de parámetros de reservación.
- **Mapeo de grupos de direcciones:** Sería necesario para proporcionar comunicación de muchos a muchos, disponible en IP.
- **Enrutamiento QoS:** Se ha probado que RSVP al mismo tiempo que está solicitando el siguiente salto para una dirección de destino dada, podría también proveer información acerca de los requerimientos de QoS al protocolo de enrutamiento, para que luego el protocolo de enrutamiento tenga en cuenta esta información para llegar a una decisión de enrutamiento. El Enrutamiento PNNI también ha sido considerado para proveer enrutamiento de Calidad de Servicio.
- **Mapeo de parámetros QoS:** Los parámetros QoS necesitan ser mapeados del modelo de ATM al modelo de Intserv/RSVP. Pueden presentarse complicaciones debido a las diferencias en los métodos posibles de implementación, políticas desiguales y problemas de costo. Hay una necesidad de proponer una matriz de posibles mapeos para varias redes. Hay necesidad de establecer una matriz de posibles mapeos para varios tipos de redes.

- **Políticas (Supervisión):** Los mecanismos de CAC encajaron bien con los mecanismos de políticas IP. Algunas políticas únicas para IP sobre ATM podrían no permitir cambio dinámico de QoS sobre ciertos VCs.

Inicialmente se considera la arquitectura general de IP sobre redes ATM con apoyo del modelo de Servicios Integrados. Se ha descrito el papel de los Enrutadores de extremo. Se propuso un mapeo entre ATM QoS e IP QoS. La siguiente sección tratará con proporcionar Diffserv sobre ATM.

4.4.2 DiffServ sobre ATM. Esta sección tratará los asuntos relacionados cuando Diffserv corre sobre una red ATM o cuando los paquetes de Diffserv necesitan atravesar una red intermedia ATM. Diffserv especifica los parámetros de tráfico según Comportamiento por Salto y no dice nada sobre los servicios que se proporcionan. Los servicios específicos quedan a disposición de los Proveedores de Servicio, basado en las Comportamientos por Salto disponible. Sin embargo, ATM especifica el servicio disponible a la Interfase de Red del usuario y le permite alterar los parámetros de tráfico según acuerdos. El mapeo de los parámetros de ATM QoS a los servicios Diferenciados IP no es una tarea fácil.

Dos importantes comportamientos por Salto han sido definidos por IETF para Diffserv: **Expedited Forwarding (EF)** PHB [EF] asegura la disponibilidad del ancho de banda independientemente de que otro tráfico comparta el enlace. Este PHB proporciona las facilidades para las especificaciones de tasa pico y las

políticas de tráfico. Este servicio es similar al proporcionado en ATM CBR o rt-VBR. **Assured Forwarding (AF)** [AF] asegura una cierta cantidad mínima de ancho de banda disponible. **Assured Forwarding** consiste en cuatro subclases de tráfico. AF PHB es entonces similar a ABR y nrt-VBR de ATM. Sin embargo el mapeo, tomado como tal, no es idéntico. ATM proporciona ancho de banda garantizado de extremo a extremo. Ninguno hace que EF o AF tengan control de admisión de la conexión. Diffserv se diseña para el control de flujo a bajo nivel para impartir escalabilidad en la cantidad de trabajo que el Enrutador central necesita, o la cantidad de información que necesita guardar.

Diffserv sobre ATM hace necesario un mapeo de la cabecera IP a un VC, el cual puede leer límites de la trama. También, un mapeo entre el **DS Code Point (DSCP)** y los PHBs necesitan ser identificados. Algunos de los problemas en proporcionar interoperabilidad entre ATM y Diffserv se consideran a continuación.

- **VPs permanente con QoS.** Es sencillo de implementar pero puede llevar a un gasto innecesario si el enlace está subutilizado.
- **Mapeo con SVCs.** Cada nueva conexión IPDS es transportada por un ATM SVC separado. Sin embargo este método no cuesta en términos de ancho de banda sin usar, sin embargo, incrementa el tiempo de configuración y podría llevar a una explosión de VC. RSVP puede utilizarse para implementar este método.

- **El papel de la topología del Enrutador en decidir la simplicidad de Control de Admisión de la conexión (CAC).** La complejidad de CAC aumenta con la topología de red del Enrutador.
- **El ancho de banda se puede utilizar más económicamente.** Se logra permitiendo que los tráficos Best Effort o UBR consuman los recursos no utilizados que fueron ocupados por la Ruta Virtual ATM original.
- **Mapeo del direccionamiento.** Las conexiones permanentes de ATM se mapean a Circuitos Virtuales Individuales. Las rutas virtuales VP también podrían utilizarse entre los usuarios finales para mapear las direcciones IP agregadas, mientras que las direcciones individuales DSCP de los flujos son mapeadas sobre las rutas virtuales correspondientes de ese **Path Virtual**. Los recursos garantizados para los circuitos virtuales individuales, son limitados por los recursos apropiados a los circuitos o rutas virtuales a los cuales pertenecen.

Las acciones que se necesitan que tomen lugar en el Switch ATM para los DiffServ sobre ATM, se clasifican como: aprovisionamiento de ancho de banda, localización relativa de recursos y control de prioridad.

CONCLUSIONES

- ✓ **A lo largo de la historia los ordenadores (o las computadoras) nos han ayudado a realizar muchas aplicaciones y trabajos, el hombre no satisfecho con esto, buscó mas progreso, logrando implantar comunicaciones entre varias computadoras, o mejor dicho: "implantar Redes en las computadoras"; hoy en día la llamada Internet es dueña de las redes, en cualquier parte del mundo una computadora se comunica, comparte datos, realiza transacciones en segundos, gracias a las redes. En los Bancos, las agencias de alquiler de vehículos, las líneas aéreas, y casi todas las empresas tienen como núcleo principal de la comunicación a una RED. Gracias a la denominada INTERNET, familias, empresas, y personas de todo el mundo, se comunican, rápida y económicamente. Las redes agilizaron en un paso gigante al mundo, porque grandes cantidades de información se trasladan de un sitio a otro sin peligro de extraviarse en el camino. Internet es uno de los ejemplos claros de Redes WAN, que tratamos en este Trabajo.**
- ✓ **El deseo de los usuarios de que las redes sean simples, desde su punto de vista, y que se limiten a transmitir y conmutar elementos de información con la menor manipulación posible.**

- ✓ **QoS proporciona utilidades en la entrega de datos críticos de cualquier empresa en periodos establecidos y con unas garantías determinadas, además ofrece a las empresas beneficios brindándoles un servicio de red con mayor calidad y confiabilidad, y asegurando la puntualidad en la entrega de la información transmitida en la red. Por ultimo le permitirá a los proveedores de servicio brindar mayores prestaciones, para un mejor manejo y gestión de la red.**

- ✓ **Para determinadas aplicaciones de calidad de servicio en las redes, los bajos valores de retardo y de variabilidad en el mismo aseguran un tiempo adecuado de llegada de los paquetes al usuario final, buscando de esta forma el aumento de la productividad del negocio, pero en el caso de que estos valores sean muy altos, será imposible obtener un correcto nivel de calidad para estas aplicaciones.**

- ✓ **QoS posee entre sus virtudes mas importantes, la sencillez de implantación, la reducción del trafico de señalización y de control necesarios para su funcionamiento, la facilidad de establecer criterios variables de prioridad, así como los buenos resultados obtenidos en cuanto al aprovechamiento de los recursos del sistema.**

- ✓ **QoS es fundamental para afrontar y dar eficiencia a diferentes tipos de aplicaciones: voz, datos y video, el control de redes complejas y un**

servicio predecible de aplicaciones en red y de tipos de tráfico. El ancho de banda necesario, la priorización por tipo de tráfico (Http, FTP, SMTP, UDP etc.), el retardo, la fluctuación de fase y la pérdida de paquetes pueden controlarse de una manera eficaz. Al garantizar los resultados deseados, las características de QoS hacen posible servicios eficientes y previsibles para su organización.

- ✓ **La red de transporte ATM combina diferentes métodos de gestión y control de tráfico aplicando los más adecuados según la clase de tráfico que se desea controlar.**

- ✓ **ATM promete ser la tecnología de red empresarial virtual del futuro, un término que refleja tanto la evolución del modelo empresarial global y el énfasis en la conectividad lógica, donde los usuarios obtienen acceso a los recursos que necesitan y el operador de la red provee las rutas de conexión y asigna el ancho de banda necesario a fuentes de tráfico muy diferentes (voz, datos, vídeo). Aquellos que construyen y operan redes deben volver los ojos a las capacidades de la tecnología ATM, ya que aspiran a la mágica combinación: interconectividad global - escalabilidad de tecnologías y satisfacción del cliente local.**

- ✓ **A lo largo del documento hemos analizado diferentes mecanismos de control de tráfico en las redes tradicionales. Cuando a esos métodos se**

les aumenta la velocidad, la distancia o ambas cosas, según el caso, hemos analizado las limitaciones. A veces el problema es el propio protocolo, otras la topología de la red, o bien el formato de los paquetes o tramas.

- ✓ Los mecanismos de control por ventana, que tan bien regulan el flujo de tráfico en los nodos de muchas redes de conmutación de paquetes, pierden eficacia en redes de alta velocidad. La razón es que la reacción ya es inútil debido a la gran cantidad de tráfico que se encuentra circulando por la red.

REFERENCIAS

- http://www.cft.gob.mx/html/la_era/info_tel/it0.html
 - http://www.lania.mx/biblioteca/newsletters/1999-otono-invierno/redes_si.html
 - http://www.goslabs.com/definiciones/calidad_d_servicio_gos.html
 - <http://www.eveliux.com/telecom/evolredes.html>
 - <http://www.internet.uson.mx/webpers/hcota/evolution.htm>
 - http://portalqsm.com/zona_banda_ancha_extendida/90_0_6_0_C/
 - <http://www.ist-mobydick.org/publications/cita2002>
 - http://eia.udg.es/~marzo/doctorat/ctav_v00.pdf
 - <http://www.rediris.es/rediris/boletin/46-47/ponencia11.html>
-
- **BOSSIO NIETO, Belky y CUADRADO GONZALEZ, Carlos. Cartagena, 2003, 115 h. Trabajo de grado (Ingenieros de Sistemas). Universidad Tecnológica de Bolívar. Facultad de Ingeniería de Sistemas.**
 - **RIVERA RUEDA, Jaime y RINCON SERRANO, Claudia. Bucaramanga, 2003, 41 h. Trabajo de doctorado (Ingeniería de Telecomunicaciones). Universidad Politécnica de Valencia. Facultad de Ingeniería de Telecomunicaciones.**

