

**MECANISMO DE MANEJO DE COLA EN REDES IP.**

**INTEGRANTES:**

**JUAN BECERRA GUZMÁN**

**DAVID FERNANDO ARQUEZ BELTRÁN**

**DIRECTOR:**

**GONZALO LÓPEZ VERGARA**

**UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR**

**FACULTAD DE ING. ELECTRÓNICA**

**CARTAGENA**

**2008.**

**MECANISMO DE MANEJO DE COLA EN REDES IP.**

**INTEGRANTES:**

**JUAN BECERRA GUZMÁN**

**DAVID FERNANDO ARQUEZ BELTRÁN**

**MONOGRAFÍA PARA OBTENER TÍTULO EN MINOR DE  
TELECOMUNICACIONES.**

**DIRECTOR:**

**GONZALO LÓPEZ VERGARA**

**UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR**

**FACULTAD DE ING. ELECTRÓNICA**

**CARTAGENA**

**2008**

## CONTENIDO.

	PAG.
• Introducción	9
• Factor de Calidad QoS	11
• Limitaciones Tecnológicas	11
• Ancho de banda o Caudal eficaz.	11
• Pérdida de paquetes	11
• Retardo	12
• Variación del retardo (Jitter)	13
• Modelos de Calidad de Servicio	14
• Servicio de mejor esfuerzo	14
• Servicio diferenciado	14
• Servicio garantizado	14
• Herramientas Para Implementar una Buena Calidad de Servicio	15
• Clasificación y marcado de paquetes	16
• Técnicas de acondicionamiento de tráfico	17
• Técnicas para evitar la congestión	17
• Técnicas para administrar la congestión	19
• Estrategia de Métodos de Encolamiento	20
• FIFO First in – First out	23
• PQ Priority Queuing	25
• CQ Custom Queuing	31
• Encolamiento WFQ	35

- CBWFQ Encolamiento WFQ basado en Clases 39
- Encolamiento mixto PQ/WFQ (IP RTP PRIORITY) 50
- LLQ Low Latency Queuing 56
- CONCLUSIONES 60

## CONTENIDO GRAFICO.

• <i>Diagrama de Bloque Qos</i>	11
• <i>Efectos de la congestión en el tiempo de servicio y rendimiento. Modelamientos de Factor de Calidad</i>	15
• <i>Traffic policing</i>	18
• <i>Traffic shaping</i>	19
• <i>RED Random Early Detection</i>	21
• <i>Estructura de Encolamiento en Interfaz de Salida.</i>	21
• <i>Encolamiento en un Enrutador</i>	22
• <i>Estrategia de Encolamiento FIFO.</i>	24
• <i>Encolamiento Con Prioridad De Cola</i>	26
• <i>Encolamiento Con Prioridad De Cola</i>	26
• <i>PQ Priority Queuing.</i>	27
• <i>Programación en el tiempo de las colas con prioridad</i>	27
• <i>Encolamiento Custom Queing ( colas definidas por el usuario)</i>	32
• <i>Itinerarios en la estrategia CQ</i>	34
• <i>Escenario de caso en el itinerario de Cq</i>	35
• <i>Encolamiento Custom Queing.</i>	36
• <i>Estrategia De Encolamiento WFQ ( Weighted Fair Queuing)</i>	39
• <i>Esquema de descarte de WFQ</i>	42
• <i>Estrategia De Encolamiento WFQ.</i>	46

- ***Weighted Fair Queuing*** 46
- ***CBWFQ Encolamiento WFQ basado en Clases*** 53
- ***LLQ Low Latency Queuing*** 61

## **CONTENIDO DE TABLAS.**

- ***Qos para Aplicaciones*** **7**
- ***Cifras probabilísticas de ocurrencia de 2 flujos terminando en la misma Cola para 3 tamaños de colas en las estrategias WFQ*** **36**
- ***Tiempo de respuesta en Mecanismo de encolamiento*** **45**
- ***Parámetros de respuesta en Mecanismo de encolamiento*** **45**

## GLOSARIO.

- **Algoritmo de Encolamiento:** Algoritmo para suministrar un nivel alto de calidad de servicio priorizando los datos.
- **Buffer:** Memoria temporal de información.
- **Codec:** Abreviación de codificador – decodificador, donde se describe una especificación desarrollada en software, hardware o una combinación de ambos.
- **Custom Queuing (CQ):** Mecanismo de manejo de colas, el cual es un servicio garantizado en ancho de banda hasta para 16 clases.
- **Distributed ToS-based WFQ:** servicio garantizado en ancho de banda hasta para 4 clases.
- **Distributed QoS-group-based WFQ:** servicio garantizado en ancho de banda hasta para 100 clases.
- **Factor de Calidad (Qos):** Parámetro usado para comparar la calidad de un sistema.
- **FIFO:** Mecanismo de manejo de colas, consiste en que el primero que entra es el primero en salir, no hay diferenciación de paquetes (igualitaria pero garantiza QoS).
- **Enrutador:** Dispositivo de Hardware para realizar interconexión de Red.
- **Pérdidas de Paquetes:** Evento como su nombre lo dice, en donde en las redes se saturan, ya que ocurre las ráfagas son colocadas en espera por los routers por no disponer la capacidad suficiente para conmutarlos.



- **Priority Queuing (PQ):** Mecanismo de manejo de colas, consiste en una estricta priorización de paquetes.
- **Jitter:** Variación en el retardo de una fase, se debe en muchos casos al retardo en los encolamientos.
- **Trafico expeditivo:** Es el tráfico que exige un tratamiento preferente derivado de las limitaciones de retardo y jitter.
- **Traffic Policing:** Técnica que trata de garantizar que no excedan el ancho de banda en servicios (SLA).
- **Traffic Shaping:** Técnica que pretende garantizar la no violación del ancho de banda. Intenta limitar el regreso de paquetes a través del mecanismo de encolado para lograr un flujo uniforme y suave del tráfico por debajo del límite de diseño.
- **Latencia:** Se denomina a la suma de retardos temporales dentro de una red.
- **Modified Déficit Round-robin (MDRR):** Servicio garantizado en ancho de banda hasta para 8 clases; bajo retardo garantizado si se usa Strict o Alternate Priority.
- **IP RTP Prioritization:** bajo retardo garantizado para paquetes de voz y video.
- **SLA:** Con sus siglas en ingles (Service level Agree), el cual es un nivel de servicio acordado con el cliente.
- **Tail Drop:** Respuesta por defecto de todo sistema a la congestión, descartar paquetes sin tomar en cuenta el tipo de tráfico.
- **RTP:** Real Time Transport Protocol.
- **Round robin:** Algoritmo para solucionar todos los elementos de manera equitativa y en un orden racional, normalmente empezando por el primero hasta el último para hacer nuevamente el proceso.

- **UDP:** Protocolo de nivel de transporte basado en el intercambio del datagrama.
- **Weighted Fair Queuing (WFQ):** servicio garantizado en ancho de banda para flujos individuales.

## INTRODUCCION.

Las Redes con toda seguridad se desarrollan basándose en el protocolo IP, este protocolo por su naturaleza está enfocado hacia el mundo de los datos y entra en contraposición con los requisitos de calidad de servicio que exige el mundo de las nuevas aplicaciones (voz, video y datos). Una de las tareas más difíciles que enfrenta las redes es proporcionar la capacidad y la flexibilidad necesaria para manejar diferentes niveles de QoS y al mismo tiempo optimizar su transporte bajo la filosofía de diferentes niveles de QoS donde se debe optimizar su transporte con el protocolo IP.

Para abordar la integración de redes de voz y datos a través de una infraestructura de paquetes, hay que considerar varias limitantes que son propias del mundo IP y que son primordiales para tráfico en tiempo real. Aspectos como capacidad de ancho de banda, niveles de retardo, jitter y pérdida de paquetes, son medidas de suma importancia a considerar en las redes, estos conceptos traen a relación el concepto de **FACTOR DE CALIDAD**.

En esta trabajo investigamos sobre los mecanismos usados para tomar los paquetes de las colas de salida de cada interfaz y encapsularlos en el nivel 2 para transmitirlos a su destino se denominan mecanismo de encolamiento o de distribución de los paquetes hacia la salida o simplemente encolamiento (queuing), solamente con el fin de evitar la congestión en las redes.

Son procesos manejados por software en los enrutadores, la selección de que tipo de mecanismo se usa para cada interfaz se hace por configuración con parámetros en el enrutador.

Las colas de espera tienen un papel muy importante ya que entre más

tiempo pasen los paquetes en estas colas, mayor es el tiempo total de la comunicación, existen diferentes mecanismos de encolamiento basados en algoritmos que varían desde lo más simple hasta lo más complejo.

Los recursos finitos de los nodos (tales como la capacidad de los buffer de entrada y salida, capacidad de procesamiento de paquetes por segundo) deben tener un mecanismo de información por el cual se le de al tráfico crítico, un tratamiento diferencial que vaya acorde con la calidad de servicio pactado para cada uno de ellos. Esta opción gana mayor validez si consideramos que en una red multiservicios conviven varios tipos de tráficos y varios de ellos con estrictos requisitos de calidad de servicio, como son las aplicaciones de voz y multimedia.

Estas técnicas de encolamiento son fundamentales para proporcionar Factor de Calidad (**QoS**) en redes convergentes; las mismas entran en acción cuando se ha producido una congestión y es necesario informar para prever el tránsito acelerado del tráfico prioritario a través de los nodos congestionados. En definitiva, los mecanismos de congestión son técnicas correctivas de congestión, que operan en función de una disciplina de servicio que favorece el tratamiento diferencial del tráfico a través del control directo de las colas de los nodos. Para viabilizar esta disciplina de servicio, las diferentes técnicas de encolado ordenan los paquetes en colas bajo diferentes enfoques de prioridad, esto genera en cada nodo una selección del orden y cantidad de bytes que se tomarán de cada cola a transmitir en el proceso de transmisión hacia el siguiente nodo.

## 1. FACTOR DE CALIDAD QoS.

Se refiere a la habilidad de la red de ofrecer prioridad a unos determinados tipos de tráfico o a la calidad de servicio o nivel de servicio que la red ofrece a las aplicaciones o usuarios según algún tipo de especificaciones o parámetros.

La implementación de provisiones de factor de calidad en un router de una red IP presenta dos fases, la cual muestra detalladamente el diagrama de bloques de la *Figura 1*.

- **Distinción de paquetes:** En esta fase se clasifica, se marca y se monitorea el paquete como muestra la figura a continuación.
- **Tratamiento Diferenciado:** En esta fase se planifican los recursos como el ancho de Banda, como muestra la figura a continuación.

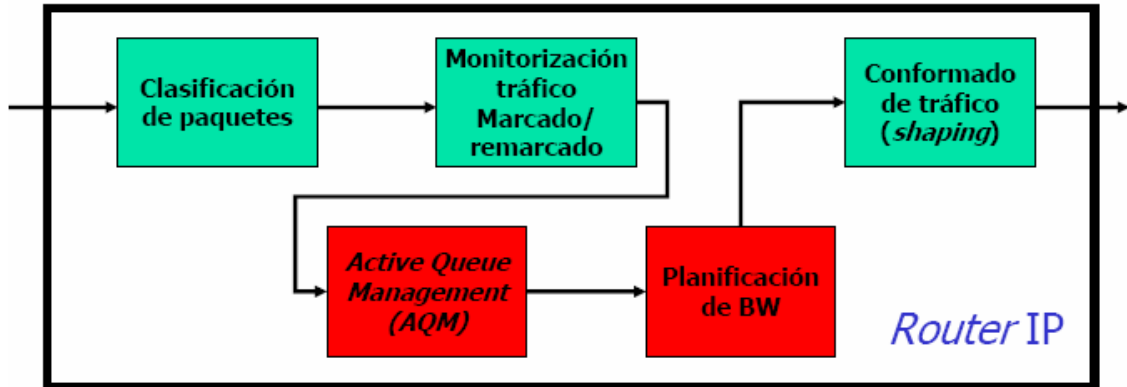


FIG.1 Diagrama de Bloques – Tomado de Diseño e Implementación de Qos.

### 1.1 LIMITACIONES TECNOLOGICAS EN UNA RED.

Los factores principales que afectan la calidad de la información en una red que presenta las aplicaciones necesarias para las redes actuales son:

- Ancho de banda o caudal eficaz
- pérdidas de paquetes
- retardo

- variación del retardo.

#### **1.1.1 Ancho de banda o Caudal eficaz.**

El uso adecuado del ancho de banda, es necesario en las redes actuales para la transmisión de las nuevas aplicaciones, como por ejemplo la señal de voz, la cual es función directa del CODEC utilizado, en donde una red de voz convencional se utiliza 64 Kbps dedicados para transportar voz digitalizada, su calidad es muy buena pero su uso muy ineficiente dada la habitual característica semiduplex del habla humana. En el mundo IP el cual es caracterizado por un uso eficiente del ancho de banda se da preferencia a métodos de codificación y compresión de voz más eficientes que con apoyo de técnicas adicionales de supresión de silencios y generación de ruido de confort, se ha logrado grandes ahorros de ancho de banda, llegando incluso a valores de 5.3 Kbps.

#### **1.1.2 Pérdida de paquetes**

Las redes de paquetes son redes de mejor esfuerzo, en estas redes los paquetes son enviados a la red sin una conexión previa entre emisor y receptor, los paquetes procedentes de varias fuentes que viajan a ráfagas los cuales son colocadas en las colas de espera de los routers, en donde al no disponer de la capacidad suficiente para conmutarlos se saturan, procediendo a descartar los paquetes en la medida que llegan, una vez que se ha congestionado el mismo. Para un servicio de mejor esfuerzo estas pérdidas no representan un gran problema, las políticas de retransmisión de paquetes solucionan este problema, sin embargo en una red que maneja las nuevas aplicaciones y las características de tiempo real, donde toleran mínimas pérdidas antes de degenerar el servicio, en aplicaciones de voz los paquetes son pequeños y es posible minimizar el efecto de esta pérdida en el lado receptor. Si el grado de congestión se vuelve crítico y las políticas de priorización del tráfico de voz no alcanzan a suplir esta carencia temporal de ancho de banda, la calidad de la voz se degrada hasta resultar inaceptable. Cualquier pérdida de paquetes que excedan el cinco por ciento tiene la probabilidad de resultar en

conversación entrecortada.

### **1.1.3 Retardo**

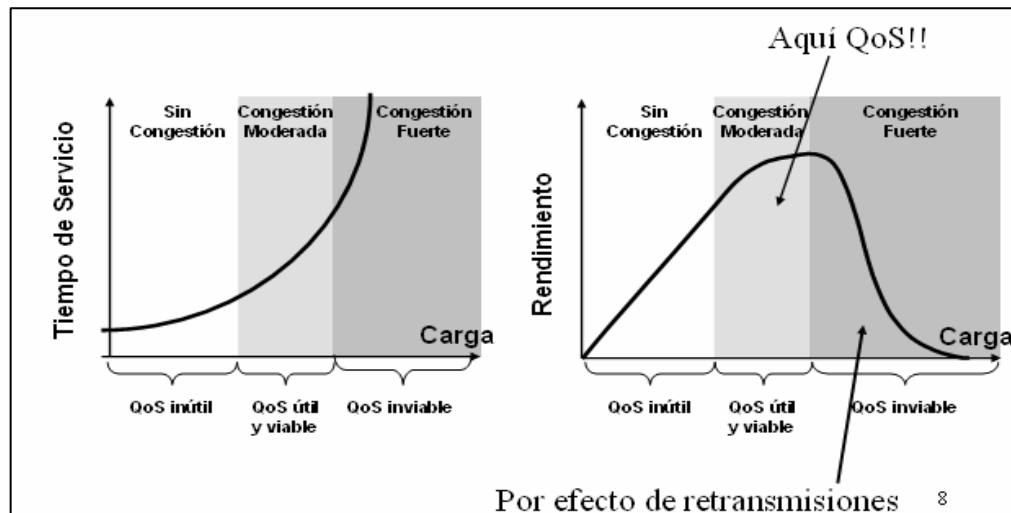
Una de las características más importantes de la información es su temporalidad, en las redes son muy importantes, por ejemplo en aplicaciones como la voz, esta medida es una de las más importantes a considerar en el diseño de servicios de voz a través de una red convergente. Es necesario e imprescindible minimizar este retardo en las redes no solo para evitar las colisiones de las conversaciones, sino también para mitigar los reflejos no deseados de la señal de audio (eco). Los valores de latencia que se admiten para diseño se los sitúa bajo los 250 ms para tráfico unidireccional, valores mayores a este referencial crean ya dificultades en la conversación. La dificultad del retardo mayor provoca que la conversación de voz, que normalmente es unidireccional, se vuelva dúplex completa con las correspondientes colisiones que vuelven impráctica la conversación. Para prevenir esta deficiencia, el estándar de la UIT limita la latencia a una máxima de 150 ms en sentido unidireccional.

### **1.1.4 Variación del retardo (Jitter)**

La variación del retardo o fluctuación de fase, es un tema más serio para las redes de voz sobre IP que el retardo absoluto, las causas principales del jitter son las variaciones del retardo de encolado debido a los cambios dinámicos que sufre la carga de tráfico a través de la red y toda la gama de factores de retardo variables enumerados arriba. Estas variaciones de arribo de los paquetes a un determinado destino no son tan perjudiciales para tráficos no sensibles a esta métrica, tal como las redes de datos, pero es demasiado crítico para tráficos sensibles como la VoIP, los servicios multimedia e interactivos. La falta de sincronismo de llegada de los paquetes a determinado destino, obliga a todos los sistemas de paquetes de voz (de audio en general) a utilizar buffers de reproducción en las pasarelas con el objeto de minimizar este efecto, pero tienen el

inconveniente de aumentar el retardo total sufrido por la comunicación. Sin embargo es mejor un retraso global alto que disponer un flujo de audio con fluctuación de fase.

Continuación se muestra una grafica que representa los efectos de congestión en el tiempo y el rendimiento que se presenta. Ver fig.2



**FIG.2 Efectos de la congestión en el tiempo de servicio y rendimiento – Tomado de AURA, Ganz, Multimedia Wires Network; Standard and Qos**

Con relación a la grafica anterior concluimos que sería mas fácil dar Calidad de Servicio si las redes nunca se congestionaran, para ello habría que sobredimensionar todos los enlaces, cosa no siempre posible o deseable. Para dar QoS con congestión es preciso tener mecanismos que permitan dar un trato distinto al tráfico preferente y cumplir el SLA (Service Level Agreement) ; Acuerdo en nivel de servicio en contratación.

## 1.2 MODELOS DE CALIDAD DE SERVICIO.

Hay tres modelos de **QoS** en donde encajan la gran mayoría de herramientas y tecnologías que controlan la calidad de servicio en las redes, estos son: mejor esfuerzo, diferenciado y garantizado. Estos se muestran a continuación en la fig.





**FIG. 3 Modelos de Factor de Calidad - Tomado de AURA, Ganz, Multimedia Wires Network; Standard and Qos**

**1.2.1 Servicio de mejor esfuerzo:** Es un modelo de servicio donde las aplicaciones envían datos sin límite de cantidad y sin un permiso previo de parte de la red, en respuesta la red cursa el tráfico sin ninguna garantía en cuanto a fiabilidad, retardo y jitter. Este servicio resulta adecuado para aplicaciones insensibles al retardo o jitter, como por ejemplo en aplicaciones de datos, como muestra *la Figura 3*. es llamado también Best Effort la cual no es nada estricto para las aplicaciones y la congestión.

**1.2.2 Servicio diferenciado:** Es un modelo adecuado para aplicaciones multiservicio que requieren servicios de red con calidad diferenciada basados en prioridades. Separa el tráfico en clases de servicio (SLA) cuya información va insertada en los paquetes y es analizada por la red y tramitada con base en niveles de prioridad establecidos previamente en el borde. Al tratarse de un sistema que basa su garantía en factores estadísticos, no ofrece ninguna garantía de calidad de servicio. Pertenece a este servicio el protocolo de Servicios Diferenciados, como muestra la *Figura 3*, también llamado DiffServ, el cual es menos estricto que el Intserv.

**1.2.3 Servicio garantizado:** Es un modelo multiservicio que reserva recursos previo al envío de los datos. Para ello, el emisor debe solicitar a la red la

reserva de recursos entre extremos y lo debe hacer a través de algún mecanismo de señalización explícito que informe a los dispositivos sobre los requerimientos de la aplicación en cuanto a ancho de banda, retardo y jitter. De los tres enumerados, este último es el modelo que ofrece mayor garantía de **QoS**. Pertenece a este servicio el protocolo de reserva de recursos (RSVP), este mecanismo también llamado Intserv, es el más estricto de los tres mecanismos de Factor de calidad como se muestra la *Figura 3*.

### **1.3 HERRAMIENTAS PARA IMPLEMENTAR UNA BUENA CALIDAD DE SERVICIO.**

Las herramientas básicas de diseño de **QoS** tienen como objetivo fundamental acelerar el tráfico sensible al retardo a lo largo de toda ruta por donde fluye este tráfico, pudiendo involucrar las redes del cliente y del operador, sea cual sea el entorno donde se requiera técnicas de QoS, las herramientas que se dispone en la actualidad y que son de aplicación para ofrecer multiservicios sobre la red que maneja varias aplicaciones, son las siguientes:

- Clasificación y marcado de paquetes
- técnicas de acondicionamiento de tráfico
- técnicas para evitar la congestión
- técnicas para administrar la congestión

#### **1.3.1 Clasificación y marcado de paquetes**

Para ofrecer servicios diferenciados, los diferentes tráficos que abordan la red deben ser primero clasificados en grupos de paquetes específicos, luego marcados en base a prioridades y finalmente enviados a la red, como por ejemplo los siguientes estándares:

**1.3.1.1 Estándar IEEE 802.1p:** El estándar 802.1p define como los dispositivos de capa dos deben procesar las prioridades de las tramas basados en

clases de tráfico. Los dispositivos de capa dos que tengan implementado este estándar están en capacidad de manejar tráfico expeditivo (*Traffic expeditivo, es el tráfico que exige un tratamiento preferente derivado de las limitaciones de retardo y jitter*) a través de sus puertos, manejando y gestionando colas en base a prioridades.

**1.3.1.2 Servicios Diferenciados (DiffServ):** Es una primera aproximación a brindar QoS en redes de paquetes IP, su objetivo principal es ofrecer QoS basado en categorías de servicio que son traducidos a acuerdos de servicios (SLA- Service Level Agreement) entre el cliente y el operador de la red. Este basa su operatividad en el estándar RFC 2474 TRAFIC CLASS IPv6 y tiene compatibilidad con el protocolo IP PRECEDENCE (IPv4 TOS BYTE RFC 791), pero a diferencia de este último extiende de 3 a 6 bits el campo de diferenciación DSCP (Differentiated Services Code Point), favoreciendo con ello una mayor granularidad de categorías de servicio con respecto a IP PRECEDENCE.

### **1.3.2 Técnicas de acondicionamiento de tráfico**

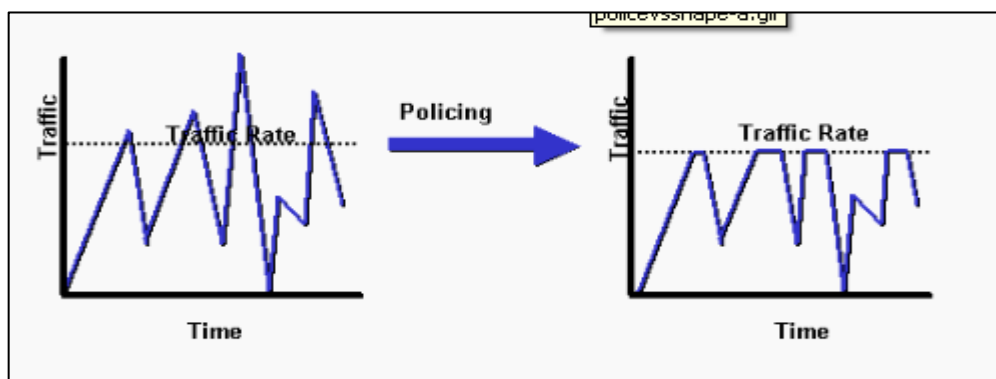
**1.3.2.1 Fragmentación e Intercalado:** El retardo de los paquetes no es exclusivo en los buffers de los nodos. El retardo que sufren los paquetes debido a la serialización es un tema muy importante a considerar, sobre todo cuando los nodos usan interfaces de baja velocidad y tramitan tráfico sensible al retardo. Si el tamaño de los paquetes de datos no es analizado y acondicionado con anterioridad a su ingreso a los buffers, estos inevitablemente retrasarán al resto de tráfico que presionan por salir a través de las interfaces seriales, y se vuelve una condición muy crítica cuando de por medio hay tráfico que no admite retardos como la voz y la multimedia. La fragmentación e intercalado es una técnica usada en enlaces de baja velocidad para asegurar que los paquetes de voz no sufran un retardo excesivo

**1.3.2.2 Traffic Policing y Traffic Shaping:** Estas técnicas responden a las violaciones del ancho de banda por parte de ciertos tráficos que han superado la capacidad previamente contratada y acordada entre el proveedor y sus clientes (SLA).

Cuando las aplicaciones intentan consumir más ancho de banda del que les corresponde, estas técnicas entran en acción con medidas preventivas que aseguran la disponibilidad de la red para tráficos críticos. La diferencia entre una y otra técnica estriba en la forma de responder a dichas violaciones.

**1.3.2.3 Traffic Policing:** Técnica que trata de garantizar que no excedan el ancho de banda en servicios (SLA).

Como muestra la *Figura 4*, estos marcan los paquetes que exceden el límite como descartable, el nodo receptor se encarga de esta gestión. Esta técnica no introduce retardos, pero puede ocasionar excesivas retransmisiones TCP y en el caso de paquetes de voz compromete la calidad de servicio si no se hace selectivo. En definitiva, un policer de tráfico simplemente se encarga de recortar todo pico de tráfico que excede el límite de diseño.

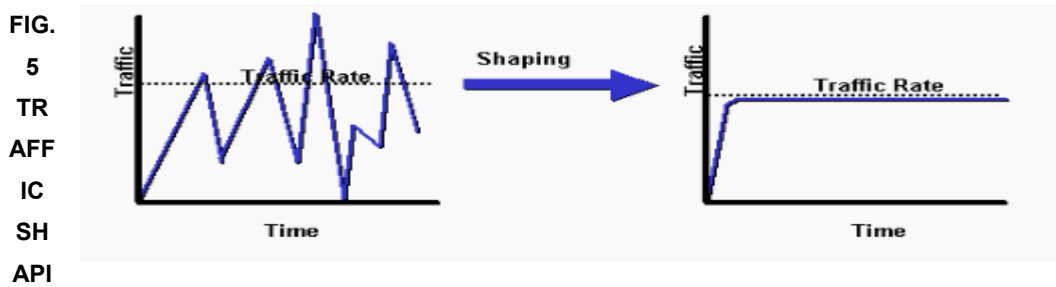


**FIG4. TRAFFIC POLICING - Tomado de Diseño e Implementación de Qos.**

**1.3.2.4 Traffic Shaping.** Técnica que pretende garantizar la no violación del ancho de banda. Intenta limitar el regreso de paquetes a través del

mecanismo de encolado para lograr un flujo uniforme y suave del tráfico por debajo del límite de diseño.

Como muestra la *Figura 5*, intenta limitar el regreso de paquetes a través del mecanismo de encolado para lograr un flujo uniforme y suave del tráfico por debajo del límite de diseño. En otras palabras, incrementa el retardo de los paquetes de salida hasta la deseada tasa de transferencia, con este método las ráfagas son suavizadas y hacia abajo se evita la congestión de otros nodos receptores. Entre los mecanismos de abajo se evita la congestión de otros nodos receptores. Entre los mecanismos de usuario y la red, se tienen a CAR (Committed Access Rate), GTS (Generic Traffic Shaping) y FRTS (Frame Relay Traffic Shaping).



NG - Tomado de Diseño e Implementación de Qos.

### 1.3.3 Técnicas para evitar la congestión

Son técnicas de prevención de la congestión, para ello monitorean la carga de tráfico en un esfuerzo por anticiparse y prevenir posibles focos de congestión que pudieran darse en los cuellos de botella de la red. El método consiste básicamente en descartar paquetes antes de que se llegue a una congestión del tráfico, situación conocida como administración activa de la profundidad de la cola de la interfaz. Hay dos formas de descartar el tráfico:

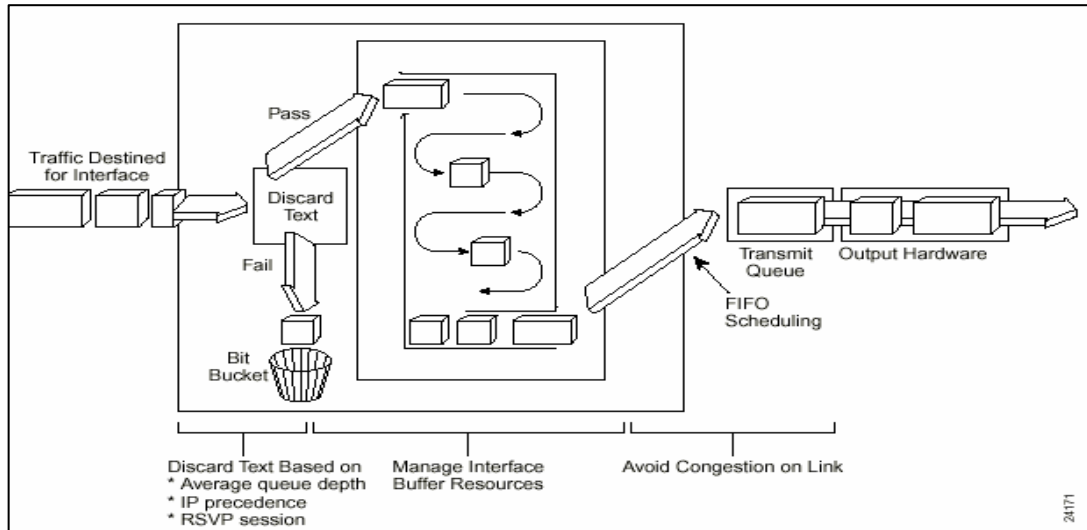
- Tail Drop
- Random Early Detección

**1.3.3.1 Tail Drop.** Es la respuesta por defecto de todo sistema a la congestión, descartar paquetes sin tomar en cuenta el tipo de tráfico.

**1.3.3.2 Random Early Detección..** Es una forma de administración activa de la cola. En lugar de esperar pasivamente hasta que los buffers descarten el tráfico, RED descarta aleatoriamente y en forma progresiva paquetes en cuanto detecta una tendencia a la congestión (detección anticipada aleatoria). La probabilidad y velocidad de descarte de los paquetes aumenta con el grado de ocupación de la cola (profundidad de cola). En respuesta a este aumento de descarte de paquetes por parte del receptor, el emisor reduce la tasa de envío hasta tanto se recupera la profundidad de la cola.

**1.3.3.3 WRED Weighted Random Early Detection:** Weighted RED asocia la probabilidad de descarte al tipo de tráfico, evitando de esta manera que se descarte paquetes de mayor prioridad. El objetivo básico es que los paquetes de mayor prioridad tengan la menor probabilidad de descarte, combina RED con IP Precedente y DSCP y permite el descarte de paquetes en forma selectiva con base en clases de servicio. El tráfico con una reserva RSVP es tratado por encima del valor más alto de prioridad de IP. WRED al permitir la clasificación de tráfico en clases de servicio y administra la profundidad de cola de todos los tipos de tráficos, proporciona una velocidad de descarte mayor para el tráfico menos prioritario. Es útil principalmente cuando la mayoría de tráfico es TCP. WRED por su capacidad de clasificar tráfico en orden de prioridad es usado principalmente en el core de transporte. Los dispositivos de acceso marcan a los paquetes con IP Precedence que es usado en el backbone para decidir que tratamiento de descarte tendrá cada tráfico.

La *Figura 6* muestra como el tráfico deja pasar o descarta la información, dependiendo de la aplicación, en donde seguidamente la información es requerida por los Buffer para luego retirarla.

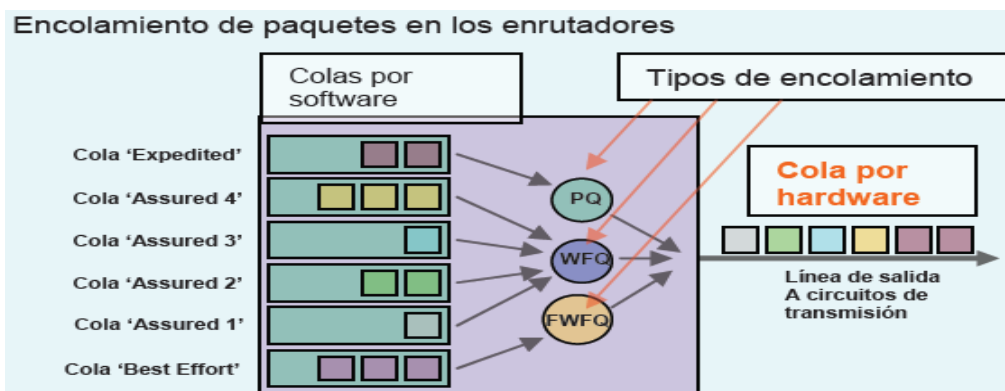


**FIG.6 RED Random Early Detection – Tomado de AURA, Ganz, Multimedia Wires Network; Standard and Qos**

#### 1.4 LA ESTRUCTURA DE ENCOLAMIENTO DEL INTERFAZ DE SALIDA

El encolamiento en los enrutadores es necesario para acomodar ráfagas de paquetes que se presentan cuando la rata de llegada es mayor que la rata de salida de paquetes por una de las siguientes razones:

1. El interfaz de entrada es más rápido que el interfaz de salida
2. La interfaz de salida está recibiendo paquetes de múltiples interfaces de entrada simultáneamente.



**FIG. 7. Encolamiento En Un Enrutador - Tomado de Diseño e Implementación de Qos.**

En la *Figura 7*. se muestra como es el proceso de las ráfagas de paquetes, las cuales son determinadas por el tipo de encolamiento para luego pasar a la siguiente capa.

1.4.1 **Mecanismos de encolamiento.** Más complejos fueron introducidos cuando se necesitaron suplir requerimientos especiales en los enrutadores para diferenciar entre paquetes de diferente importancia (Servicios diferenciados), el encolamiento se dividió en 2 partes (ver figura 8):

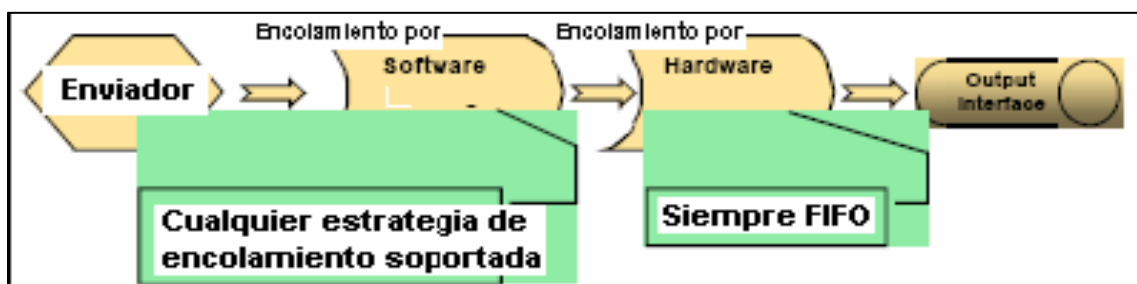


FIG. 8 Estructura de Encolamiento en Interfaz de Salida - Tomado de Diseño e Implementación de QoS.

- La cola en **hardware** la cual usa todavía la estrategia FIFO, la cual es necesaria en razón a que los manejadores (drivers) de la interfaz solo puede manejar los paquetes uno por uno, a la cola de hardware se le denomina también cola de transmisión o **TxQ**.
- La cola de **software** la cual programa en el tiempo el paso de los paquetes en el hardware basada en los requisitos de encolamiento del QoS y los servicios diferenciados.

El enrutador evitara la salida de la cola de software únicamente en 2 casos:

- Cuando la cola por software esta vacía y por tanto no hay nada que evacuar de ella.



- Cuando la cola por hardware esta llena en cuyo caso dará prioridad a la evacuación de la cola por hardware antes de continuar evacuando la cola por software.

Concluimos entonces que la cola por software se adiciona en un paquete cuando la cola por hardware se llena a su máxima capacidad, mientras la cola por hardware tenga espacio los paquetes no van a la cola por software.

Evidentemente mientras la cola por HW no se llene no hay oportunidad de hacer ningún procedimiento QoS, lo cual es natural, pues si la cola por hardware esta con cupo no hay ningún problema de congestión y no hay necesidad de QoS.

Algunas de las estrategias de encolamiento y sus metas son:

### **1.5 Bajo PBR (Policy Based Routing)**

- FIFO: no hay diferenciación de paquetes (igualitaria pero garantiza QoS).
- Priority Queuing (PQ): estricta prioritización de paquetes.
- Custom Queuing (CQ): servicio garantizado en ancho de banda hasta para 16 clases.
- Weighted Fair Queuing (WFQ) y Distributed WFQ: servicio garantizado en ancho de banda para flujos individuales.
- Distributed ToS-based WFQ: servicio garantizado en ancho de banda hasta para 4 clases
- Distributed QoS-group-based WFQ: servicio garantizado en ancho de banda hasta para 100 clases.
- Modified Deficit Round-robin (MDRR): servicio garantizado en ancho de banda hasta para 8 clases; bajo retardo garantizado si se usa Strict o Alternate Priority. ancho de banda para flujos individuales .

- IP RTP Prioritization: bajo retardo garantizado para paquetes de voz y video.

### 1.5.1 FIFO First in – First out

La estrategia de encolamiento FIFO no posee clasificación porque todos los paquetes pertenecen a una sola clase.

Es la técnica más básica de colas. Una cola FIFO es un buffer sencillo que retiene los paquetes salientes hasta que la interfaz los retire en el mismo orden en el que llegaron al buffer, como se muestra en la Figura 9.

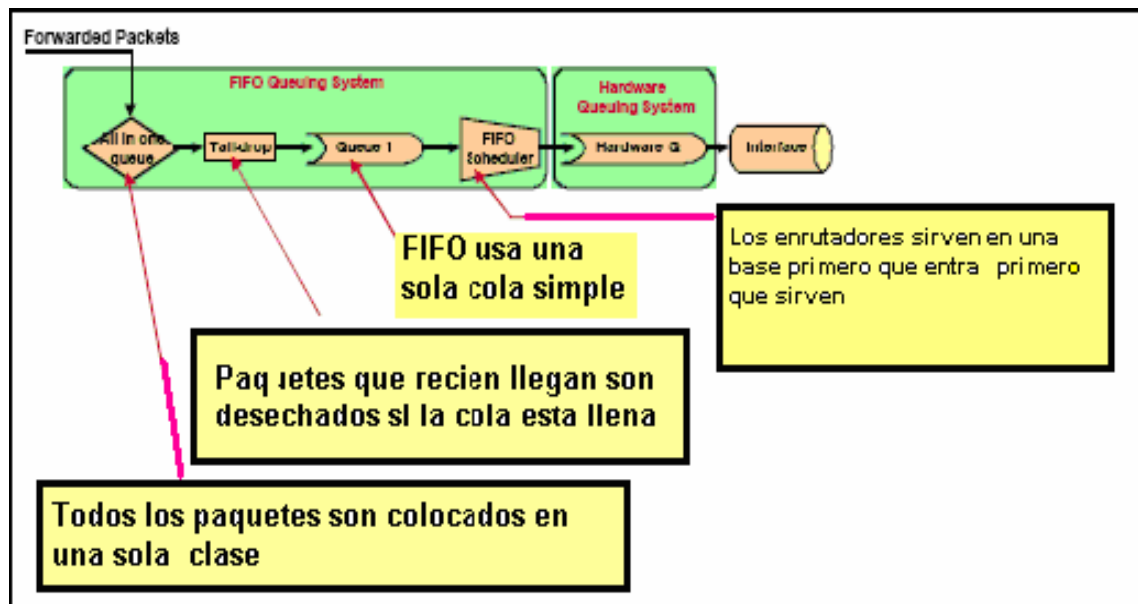


FIG. 9 Estrategia de Encolamiento FIFO - Tomado de Diseño e Implementación de Qos.

Esta técnica de encolamiento no proporciona ninguna QoS para el flujo, ni distribuye un ancho de banda equitativo entre flujos que compartan un enlace. Durante los periodos de congestión, se llama el buffer y los paquetes se descartan sin importar el tipo de paquete o las solicitudes de la aplicación asociada, esto es muy malo para VoIP. Las colas FIFO no tratan por igual a todo el flujo de tráfico, aunque descarten paquetes sin importar sus características.

Suponga que N paquetes se descartan del final de una cola durante un periodo de tiempo. El problema es que algún flujo puede contener solo N paquetes en esa ventana de tiempo, por lo que el flujo completo se podría destruir, puede que N paquetes representen el 1% de otro flujo durante esa misma ventana de tiempo, por eso, las colas FIFO están predispuestas contra flujos con pequeñas u ocasionales peticiones de ancho de banda.

Si la cola esta llena los paquetes se desechan, el programador sirve los paquetes en el orden que llegan. La estrategia de encolamiento **FIFO** es una extensión de la cola de hardware.

Encolamiento **FIFO** puede ser el más igualitario y legal pero tiene una gran lista de problemas:

FIFO no asigna el ancho de banda de una manera igual entre múltiples flujos. Los flujos que usan paquetes de mayor tamaño promedio reciben más ancho de banda que los flujos que usan paquetes de menor tamaño promedio, o lo caudales que generan mas tráfico reciben mayor proporción del ancho de anda.

FIFO es extremadamente desigual cuando un flujo agresivo es compitiendo con un flujo frágil. Un flujo agresivo esta siempre enviando una gran cantidad de paquetes muchos de los cuales son desechados. Un flujo frágil envía una modesta cantidad y la mayoría de ellos son desechados porque la cola esta siempre llena por el caudal agresivo. Este tipo de comportamiento se denomina *starvation (inanición)*.

### 1.5.2 PQ Priority Queuing

Priority Queuing (PQ, Colas con Prioridad) fue uno de los primeros mecanismos que permitieron la clasificación de paquetes en clases múltiples. El itinerario se hace con prioridad estricta. PQ puede clasificar los paquetes en una de cuatro colas:

1. La cola de prioridad alta
2. La cola de prioridad media
3. La cola de prioridad normal (la cola predefinida)
4. La cola de prioridad baja

El itinerario prefiere los paquetes en el mismo orden. Cada clase usa una cola FIFO, donde los paquetes son desechados si la cola se llena. (Ver Fig. 11)

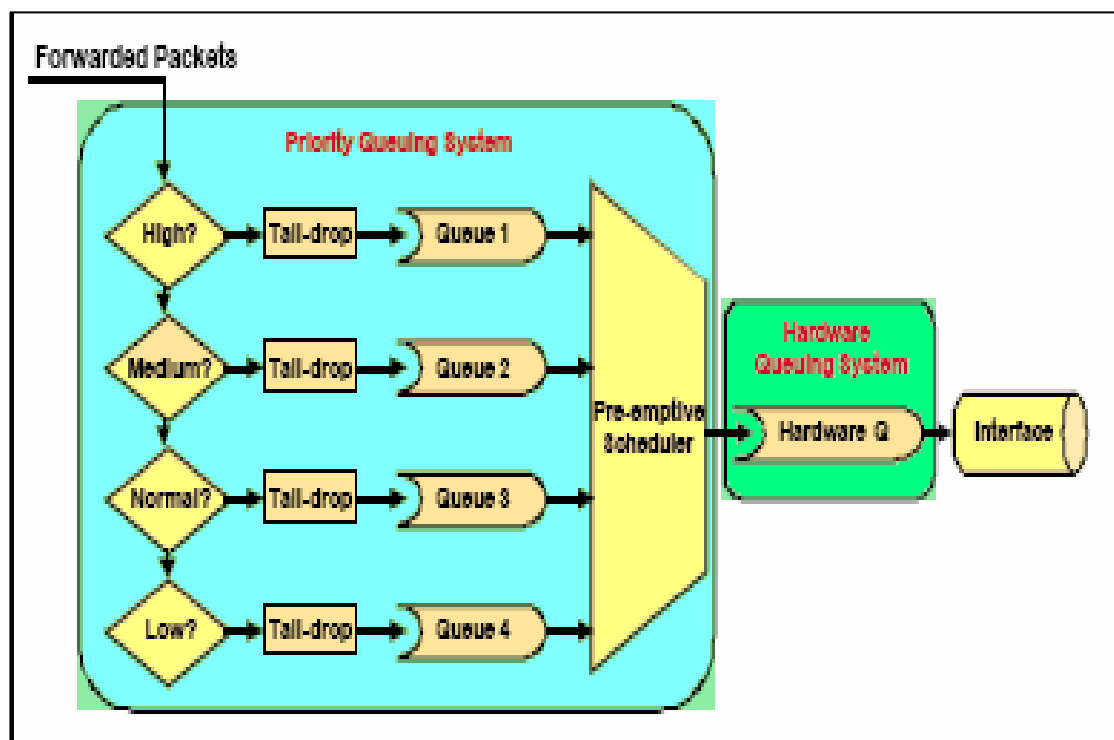


FIG. 11 Encolamiento Con Prioridad De Cola - Tomado de Diseño e Implementación de Qos.

Los paquetes siempre se sirven desde las primeras colas de alta prioridad. Si los paquetes están esperando en dichas colas, se enviarán al buffer de transmisión, si la cola de alta prioridad está vacía, se envía al buffer cualquier paquete de prioridad media, si las colas de alta y media prioridad están vacías, se envían al buffer de transmisión los paquetes de la cola normal, y así sucesivamente.

La cola de prioridad consigue los requisitos QoS de VoIP, pero deja mucho que desear en el campo de la distribución de un ancho de banda equitativo para el resto del tráfico. Con esta técnica de encolamiento, el tráfico de alta prioridad incurre lo menos posible en la latencia y las fluctuaciones de fase, pero no hay provisiones para distribuir un ancho de banda entre el tráfico con las mismas prioridades. Dentro de una prioridad, el tráfico es de cola FIFO. La limitación más significativa es que si alguna de las colas tiene un flujo constante de tráfico, las colas de alta prioridad están completamente desnutridas de ancho de banda.

Normalmente no se recomienda la cola de prioridad para VoIP, pero es útil cuando debe tratarse con diferentes tipos de tráfico de alta prioridad. Si tiene por ejemplo VoIP y tráfico SNA (IBM System Network Architecture) en un mismo enlace necesita asegurarse que los paquetes VoIP siempre estarán al frente de la cola, pero los de SNA deberían ir justo detrás de los de VoIP par asegurar un tráfico fluido de datos. LLQ (Low Latency Queuing). Y WFQ (Weighted Fair Queuing ) son incapaces de lograr este objetivo, la cola de prioridad también puede conseguir las necesidades de encolamiento de VoIP mezclado con tráfico de Video IP.

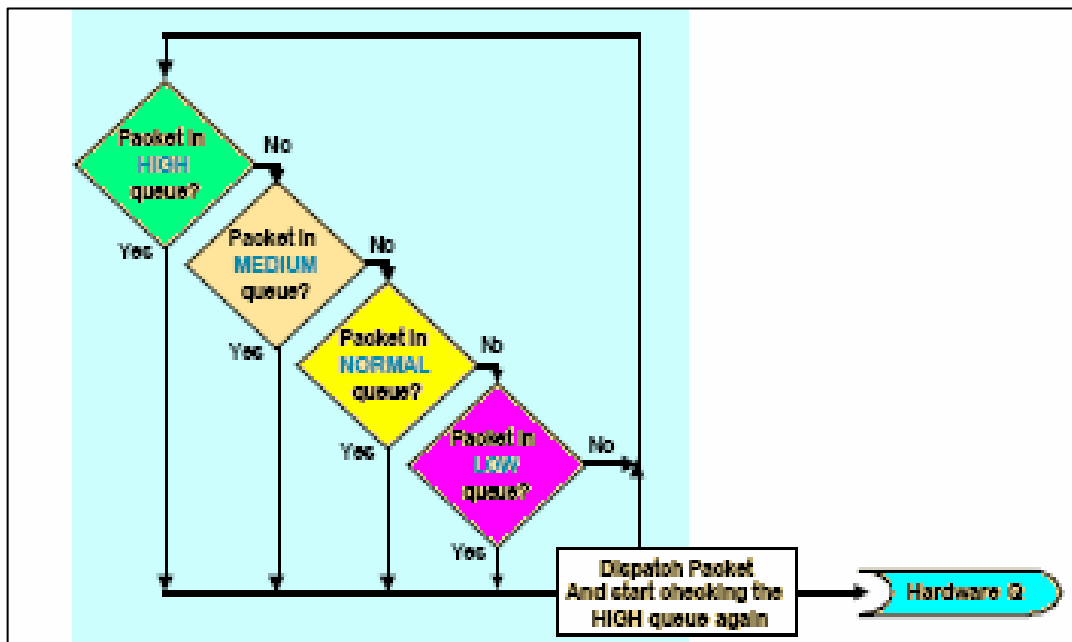


FIG.12. Programación en el tiempo de las colas con prioridad - Tomado de Diseño e Implementación de Qos.

Priority Queuing puede clasificar los paquetes de IP con las herramientas siguientes:

1. Comparación directa y apareamiento con la dirección de origen del interfaz.
2. Listas de control de acceso estándares o extendidas comparando los siguientes parámetros:
  - Dirección de origen IP
  - Dirección de destino IP
  - Puertos TC y UDP de origen y destino
  - Bits de precedencia IP ( ToS) y los 4 bits DTRC del ToS
  - Bits DSCP ( los 6 bits de mas alto orden del nuevo campo ToS)
  - Fragmentos
  - Las banderas de TCP (ACK, SYN, RST, URG, PSH)
  - Apareamiento directo con puertos TCP y UDP
  - Apareamiento directo de fragmentos
  - Apareamiento directo de paquetes basándose en el tamaño del mismo.

**PQ** es un mecanismo multi-protocolo de QoS porque soporta las herramientas de clasificación de otros protocolos no IP. Algunos de los protocolos de Capa-3 soportados así como otros de capa superior son: IPX, CLNS , DECNET, DLSw.

**1.5.2.1 PX:** El protocolo Intercambio de Paquetes Entre Redes (IPX) es la implementación del protocolo IDP (Internet Datagram Protocol) de Xerox. Es un protocolo de data gramas rápido orientado a comunicaciones sin conexión que se encarga de transmitir datos a través de la red, incluyendo en cada paquete la dirección de destino.

**1.5.2.2 CLNS (Servicio No Orientado a Conexión).** En telecomunicaciones es un servicio que establece la comunicación entre entidades sin necesidad de establecer una conexión entre ellas. Cuando una entidad tiene información para transmitir, sencillamente la envía, (tramas, paquetes, bloques, etc.).

**1.5.2.3 DECNET:** Es un grupo de productos de Comunicaciones, desarrollado por la firma Digital Equipment Corporation. La primera versión de DECnet se realiza en 1975 y permitía la comunicación entre dos mini computadoras PDP-11 directamente. Se desarrolló en una de las primeras arquitecturas de red Peer-to-peer.

**1.5.2.4 DLSW:** fue desarrollado para soportar SNA y NetBios en Routers multiprotocolo. En función de que tanto SNA como NetBios son protocolos orientados a la conexión, el procedimiento de control de enlace de datos usado sobre la LAN es IEEE 802.2 LLC (Logical Link Control) tipo 2. Además, DLSW también transporta protocolo SNA sobre un enlace de WAN usando el Protocolo SD

Aunque se apoyan otros protocolos, las opciones de la clasificación no son tan poderosas como con IP. La mayoría de los protocolos puede usar sus correspondientes listas de acceso para el clasificado de paquetes. Emparejado con el tamaño del paquete también está disponible en todos los protocolos apoyados.

Ahora estudiaremos la política de inserción, programación del itinerario y configuración del encolamiento con prioridad de las colas.

#### **1.5.2.5 Política de inserción en el encolamiento con Prioridad de las colas.**

Priority Queuing es básicamente una colección de cuatro colas tipo FIFO paralelas. Cada cola padece todos los problemas de FIFO, aislados a la clase (desigual, inanición, jitter). Cada cola usa el esquema del desecho del paquete IP cuando la cola está llena. Cada uno de las cuatro colas puede configurarse con el número máximo de paquetes que puede retener.

#### **1.5.2.6 Programación del itinerario en el encolamiento con Prioridad de las colas.**

Priority Queuing usa itinerario de prioridad estricta, mientras haya paquetes en la cola de alta prioridad ninguna otra cola se servirá, si la

cola alta está vacía el enrutador comienza a servir la cola media. La congestión en cualquiera de las colas, excepto en la cola baja, causa un tipo diferente de debilidad. Una cola de muy alta prioridad congestionada causa que todas las colas de más baja prioridad sufran de hambre (el exterminio por inanición de la clase).

#### **1.5.27 configuración del encolamiento con Prioridad de las colas**

La configuración de Priority Queuing puede dividirse en 4 pasos:

1. Clasificar los datos en cuatro clases
2. Asignar una cola a cada clase
3. Establecer el tamaño de cola máximo (si el valor predeterminado no es apropiado)
4. Aplicar la prioridad en la cola a uno o más interfases

#### **1.5.2.8 Ventajas y desventajas en el encolamiento con Prioridad de las colas.**

##### **Ventajas**

1. Provee bajo retraso de propagación al paquetes de alta prioridad
2. Soportada en todas las plataformas y marcas
3. **PQ** es un mecanismo multi-protocolo de QoS porque soporta las herramientas de clasificación de otros protocolos no IP. Algunos de los protocolos de Capa-3 soportados así como otros de capa superior.
4. Cada uno de las cuatro colas puede configurarse con el número máximo de paquetes que puede retener.

##### **Desventajas**

1. Todas las desventajas del encolamiento FIFO con una clase simple, pero extendido a 4 clases.

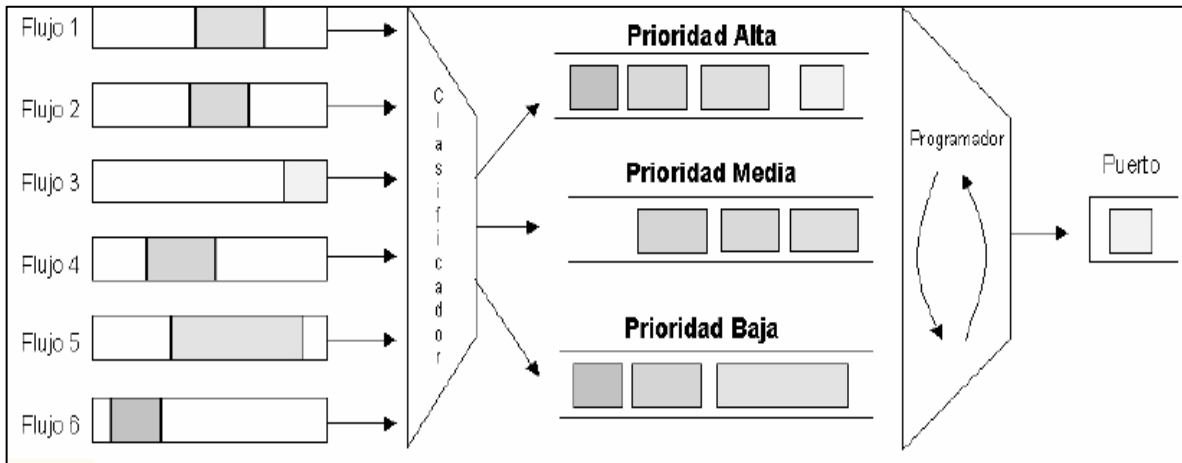


2. Cada clase puede experimentar debilidad, retardo y jitter si uno o mas flujos en la clase causan congestión
3. Inanición de las clases de baja prioridad cuando las clases altas se congestionan
4. En todas las marcas se configura manualmente por salto.

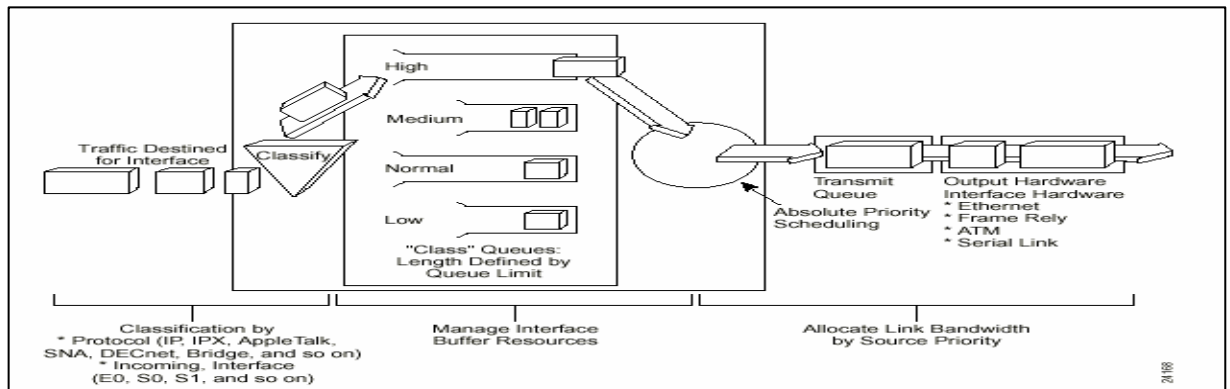
Resumiendo las características de Priority Queuing podemos decir que La cola de prioridad es un sencillo enfoque para ofrecer un tratamiento preferencial a los paquetes identificados. Esta técnica ordena los paquetes en cuatro colas de diferente prioridad: alta, media, baja y normal.

El algoritmo de vaciado sigue un orden de prioridad absoluta dando siempre servicio a la cola de mayor prioridad en detrimento del resto de colas. Hasta tanto no se evacue la totalidad de la cola de mayor prioridad la siguiente cola no puede servirse. En caso de vaciado de la cola prioritaria, el orden de servicio pasa a la media y así sucesivamente hasta la cuarta cola, todo en función de la evacuación total de la cola precedente.

Por la forma de procesar las colas, esta técnica proporciona total preferencia a las aplicaciones sensibles al retardo y jitter, caso VoIP, sin embargo adolece de deficiencias al no permitir la evacuación de tráfico de menor prioridad. Para hacer viable el uso de esta técnica para aplicaciones convergentes, hay que controlar los tráficos prioritarios con métodos de acondicionamiento de tráfico ver fig.13 y fig. 14.



**FIG.13 Encolamiento Con Prioridad De Cola - Tomado de Diseño e Implementación de Qos.**



**FIG14. PQ Priority Queuing - Tomado de Diseño e Implementación de Qos.**

La Figura 13 y 14 son otra representación grafica de Queing Priority o encolamiento por prioridad, en donde observamos como se clasifican los paquetes dependiendo de la prioridad para luego ser sometidos a la salida por una estrategia de encolamiento FIFO en donde salen en el mismo orden de llegada.

### 1.5.3 CQ Custom Queuing

La cola personalizada es un algoritmo de encolamiento configurado equitativamente. El trafico se clasifica en dos colas separadas y sirve al

estilo round-robin para asegurarse de que ninguna esta desocupada, se muestra la estructura de la cola personalizada. (Vea FIG. 15)

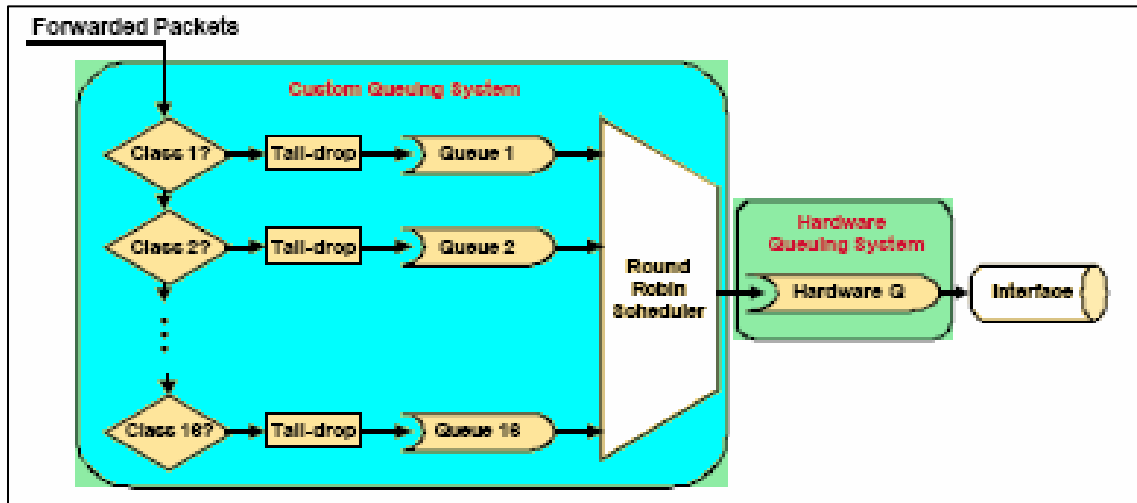


FIG. 15 Encolamiento Custom Queing (colas definidas por el usuario) - Tomado de Diseño e Implementación de Qos.

Custom Queuing (**CQ**), como se denomina en ingles es similar a Priority Queuing en el sentido de que es configurado y además en el soporte de opciones de clasificación, la programación de itinerarios sin embargo es completamente diferente.

**CQ** usa hasta 16 colas como muestra la *Figuran 14*, las cuales pueden usarse para las clases definidas por el usuario, las opciones de clasificación son idénticas a las de la estrategia de colas con prioridad (PQ).

El mecanismo de planificación usa el servicio round-robin dónde a cada cola le es permitido remitir un cierto número de bytes (no de paquetes). El desecho del último paquete todavía se usa dentro de cada cola individual. Custom Queuing (similar a Priority Queuing) puede clasificar paquetes IP con las mismas herramientas ya mencionadas en Priority Queuing.

### 1.5.3.1 Políticas de inserción en el encolamiento definido por el usuario

Una vez el paquete es clasificado un enrutador tiene que determinar si el paquete puede ser encolado, el paquete se desecha si la cola está llena.

Cada cola, a menos que se haya configurado de otra manera, puede almacenar en el buffer hasta 20 paquetes antes que el primer paquete se deseché.

### 1.5.3.2 Programación del itinerario en el encolamiento definido por el usuario.

Custom Queuing usa programación de itinerario round-robin, (ver FIG. 16) donde cada paquete consigue servicio (ancho de banda). Cada cola se configure con el número de bytes que puede enviar en una ronda (count-byte). El ultimo paquete es siempre enviado aun si la cantidad total de bytes en una ronda esta por encima del límite (byte-count). El enrutador entonces inicia el proceso de la siguiente cola.

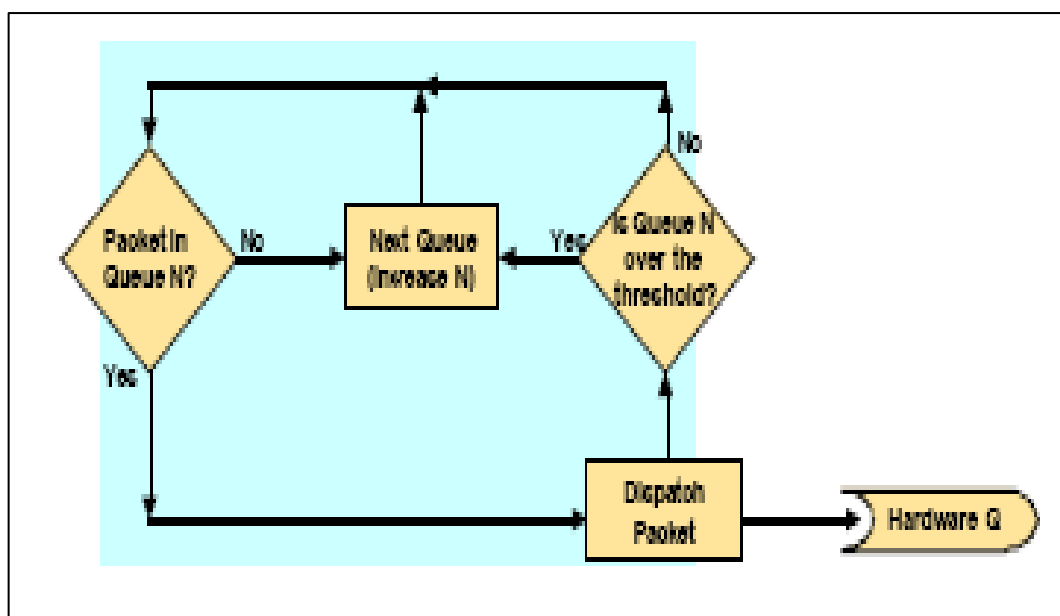


FIG. 16 Itinerarios en la estrategia CQ - Tomado de Diseño e Implementación de Qos.

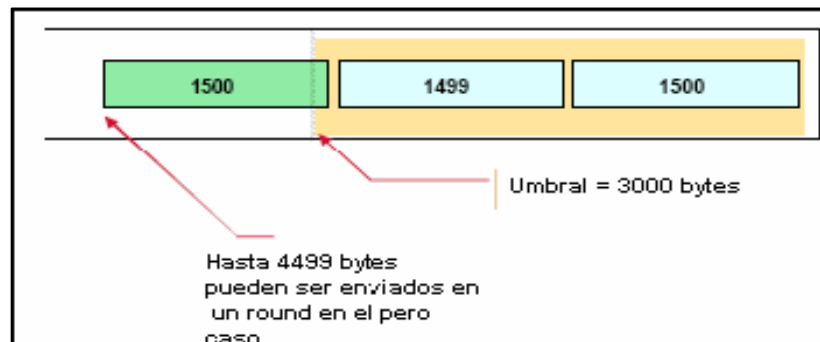


FIG. 17 Escenario de caso en el itinerario de CQ - Tomado de Diseño e Implementación.

La FIG. 17 ilustra el peor escenario en un caso dónde fueron usados los siguientes parámetros para llevar a cabo a Formación en estrategia CQ en una interfaz:

1. MTU de la interfaz es 1500 bytes
2. El count-byte es 3000 bytes (dos veces el MTU)

El ejemplo muestra cómo el enrutador envió dos paquetes primero con un tamaño total de 2999 los bytes. Desde que esta todavía dentro del límite (3000)

El enrutador puede enviar el próximo paquete (clasificado según tamaño del MTU).

El resultado fue que la cola recibió casi 50% más de lo debido de ancho de banda en la ronda.

Éste es uno de las desventajas de formación de colas de espera por la estrategia de usuario, no asigna el ancho de banda con precisión, el límite o peso de la cola se configuran en los bytes del count-byte, la exactitud de CQ depende del peso (el count-byte) y el MTU.

Si la proporción entre el count-byte y el MTU en CQ es demasiado pequeña no asignará el ancho de banda con precisión.

Si la razón entre el count-byte y el MTU es demasiado grande CQ causará grandes retardos

### **1.5.3.3 Configuración de la estrategia CQ**

Debe configurar manualmente:

1. Que trafico se asocia con cada una de las colas.
2. Cuantos paquetes pueden esperar en cada cola.
3. Que cantidad de ancho de banda se destina a cada cola.

Puede ser difícil configurar la distribución del ancho de banda para cada una de las colas. En lugar de configurar directamente la distribución del ancho de banda, puede definir el numero de bytes que van a ser transmitidos desde cada cola cuando estén en servicio. El número relativo de bytes que se asignan a cada cola identifica el porcentaje de ancho de banda distribuido para esas colas.

Aunque la capacidad de transmisión de cada cola se especifica en bytes, la cola transmite paquetes. La cola seguirá transmitiendo paquetes mientras se alcance el límite del contador de bytes, y finalizara con el paquete que este a mitad de la transmisión. Esto significa que si el contador de bytes se ajusta a 3.050, y hay tres paquetes IP de 1.500 bytes cada uno, la cola transmitirá los tres paquetes por un total de 4.500 bytes. Debe tener cuidado cuando determine los valores del contador de bytes para cada cola, o la distribución del ancho de banda actual puede ser muy diferente a cuando lo planifico. Esto puede resultar muy embarazoso, especialmente desde que algunos protocolos como IPX ajustan dinámicamente el tamaño de la trama dependiendo de la velocidad del enlace.

La cola personalizada afronta el tema del ancho de banda equitativo, pero no se ajusta bien para proporcionar QoS a flujos de tráfico específico. Como cada una de las colas se sirve al estilo round-robin, no hay una cola que tenga una prioridad mayor que la otra (excepto la cola del sistema

para los routers activos). Algunas colas pueden tener un mayor rendimiento que otras, pero ninguna puede ofrecer un tratamiento especial con respecto a la latencia o las fluctuaciones de fase. Esto hace que la cola personalizada sea una mala elección para VoIP.

#### **1.5.3.4 Asignación del retardo versus asignación del ancho de banda.**

La figura 11 ilustra cálculos de la muestra de garantías del ancho de banda y el

El retardo máximo. El tiempo que toma para completar una ronda depende del ancho de banda de la interfaz, el tamaño del MTU y la suma de todos los byte-counts de la cola. Los parámetros de estudio de caso son:

1. La primera cola usa un count-byte de 4500 (tres veces los bytes del MTU), 5999 bytes puede enviarse en el peor caso.
2. La segunda cola usa un count-byte de 3000 (dos veces el MTU)-4499 bytes pueden enviarse en el peor caso.
3. La tercera cola usa count-byte de 1500 (los bytes de MTU), 2999 bytes pueden enviarse en el peor caso.

Las primeras muestras del cálculo dan que la primera cola debe recibir 50% del ancho de banda aproximadamente.

El segundo cálculo muestra el retardo del round-robin de 562ms para la Cola 1 cuando todas las clases están congestionadas.

El tercer cálculo muestra el retardo del round-robin de 937ms para la Cola 1 cuando todas las clases están congestionadas y se las arreglan para enviar el número máximo de bytes (el count-byte + MTU - 1) en una ronda. Aunque este peor caso es muy improbable que suceda, es también improbable que esa clase use el máximo configurado exacto.

#### 1.5.4 Encolamiento WFQ ( Weighted Fair Queuing)

Weighted Fair Queuing (WFQ) es un sofisticado proceso y estrategia de construcción de colas en un enrutador el cual requiere muy poca configuración porque el proceso detecta dinámicamente los flujos entre aplicaciones y automáticamente asigna y maneja colas separadas para cada flujo. En términos de WFQ esos flujo son conversaciones a nivel de sesión, Una conversación pudiera ser un sesión Telnet , una transferencia de archivos FTP, un video sobre IP, una descarga de una página Web, o cualquier otro flujo TCO o UDP entre un cliente y servidor. Particularmente WFQ considera que un conjunto de paquetes son parte de la misma conversación si todos tienen la misma dirección IP de origen y destino y el mismo puerto TCP/UDP de origen y destino. Es decir, si el cuarteto de identificación del socket es idéntico para todos los paquetes. Además puede considerar en algunas modalidades usar el campo de precedencia ToS o los bits DTRC del mismo capo. También puede trabajar con grupos QoS como elemento de identificación para la conversación.

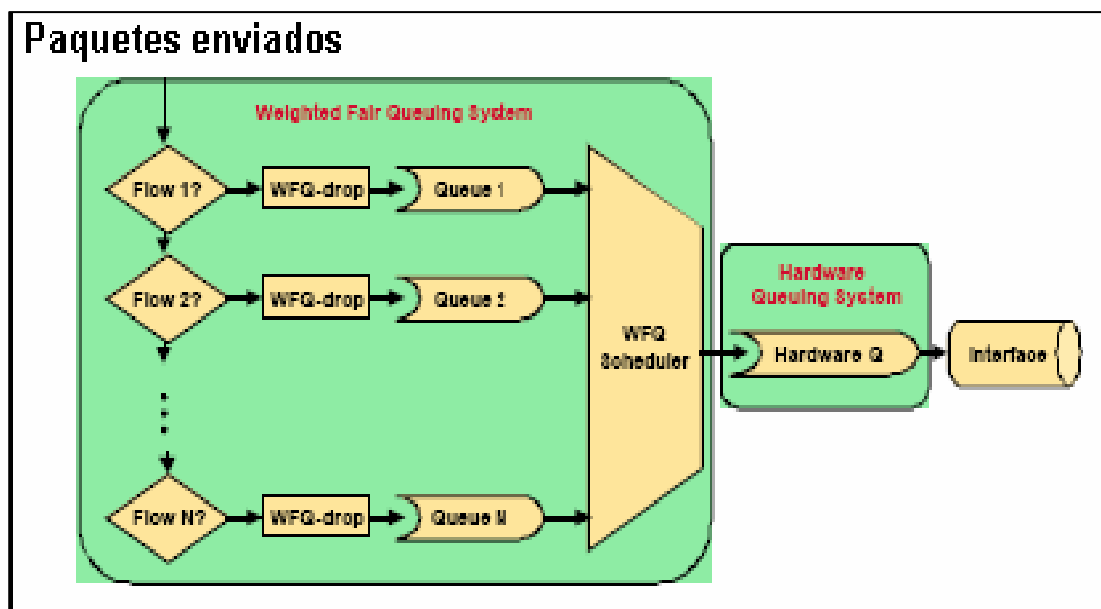
Cuando WFQ detecta una conversación y determina que esa conversación necesita QoS o que los paquetes de esa conversación deben ser encolados automáticamente crea una cola para la conversación en el interfaz. En presencia de múltiples conversaciones crea automáticamente múltiples colas una por cada conversación y clasifica automáticamente otra vez cada paquetes en la cola que le corresponde acuerdo a la conversación a que pertenezca. Debido al automatismo de WFQ en la creación de colas y clasificación de paquetes no necesita que se le creen manualmente listas para identificar los varios tipos de tráfico, como sucede en PQ y CQ, siendo mucho más fácil de implementar. Como otros tipos de encolamientos WFQ solo es invocado por el sistema operativo del enrutador cuando hay congestión en una interfaz de salida. Weighted Fair Queuing (WFQ) se presentó como una solución a los problemas de los mecanismos de formación de colas de espera siguientes:



1. FIFO que causa inanición en la cola, además retardo y jitter,
2. PQ que causa que la inanición de colas de baja prioridad y padece de todos los problema de FIFO dentro de cada uno de las cuatro colas.
3. CQ que causa grandes retardos y también sufre de todos los problemas de FIFO dentro de cada uno de las 16 colas.

La idea de WFQ es como muestra la *figura 18*:

1. Tener una cola especializada para cada flujo (ninguna inanición, retardo o jitter dentro de la cola),
2. Asignar justamente y con precisión el ancho de la banda entre todo los flujos (el retardo mínimo por la planificación del itinerario, el servicio garantizado),
3. Usar la precedencia IP como el peso para asignar el ancho de la banda.



**FIG. 18 Estrategia De Encolamiento WFQ ( Weighted Fair Queuing) – Tomado de Estudio y Configuración de Calidad de Servicio.**

Otras características de **WFQ**:

1. WFQ usa clasificación automática. No se pueden definir clases manualmente.
2. WFQ usa colas FIFO por flujo, siendo esta su característica más importante
3. El descarte en WFQ (dropping de paquetes) no es tan simple como el último de la cola sale. WFQ descarta los paquetes de los flujos más agresivos.
4. EL planificador de itinerarios de WFQ es un simulador de un sistema TDM (time -división multiplexer).
5. El ancho de banda es equitativamente distribuido a todos los flujos activos.

#### **1.5.4.1 Implementación de WFQ**

Algunas recomendaciones de implementación:

1. WFQ es soportada por todas las plataformas de enrutamiento de alto vuelo.
2. Clasificación identifica un flujo y asigna una cola a ese flujo.
3. Un peso se usa para planificar el itinerario con el fin de dar un MAS proporcionado ancho de banda a los flujos con una mayor precedencia IP.
4. El esquema de descarte de paquetes de la cola se mejora para desechar paquetes de los flujos más agresivos en proporción a su agresividad.

En resumen hay que tener en cuenta para la configuración e implementación de WFQ los siguientes puntos:

1. Un número fijo de colas por flujo debe configurarse
2. Una función hash se usa para traducir parámetros en números de cola.

3. Hay colas para paquetes del propio sistema en el encolamiento WFQ (8 colas) y flujo de la aplicación RSVP (si se configura) y ellas son mapeadas en colas separadas.
4. Dos o mas flujos pudieran Mapear la misma cola, dando como resultado un ancho de banda mas bajo por flujo.
5. Es muy Importante que el número de colas configuradas sea mas grande que el número de flujos esperado.
6. WFQ usa un número fijo de colas. La función hash es usada para asignar una cola a un flujo. Hay 8 colas adicionales para los paquetes del sistema y opcionalmente hasta 1000 colas para flujos RSVP.
7. WFQ usa 256 colas por defecto. El número de colas puede ser configurado en un rango de entre 16 y 4096 (el número debe ser un potencia de 2).

Si hay un gran número de flujos concurrentes es muy probable que 2 flujos puedan terminar en la misma cola. Es recomendable tener tantas colas como flujos haya (en promedio). Aunque esto pudiera no ser posible en ambientes de muy grandes enrutadores centrales o de distribución donde el número de flujos concurrentes es de miles y decenas de miles.

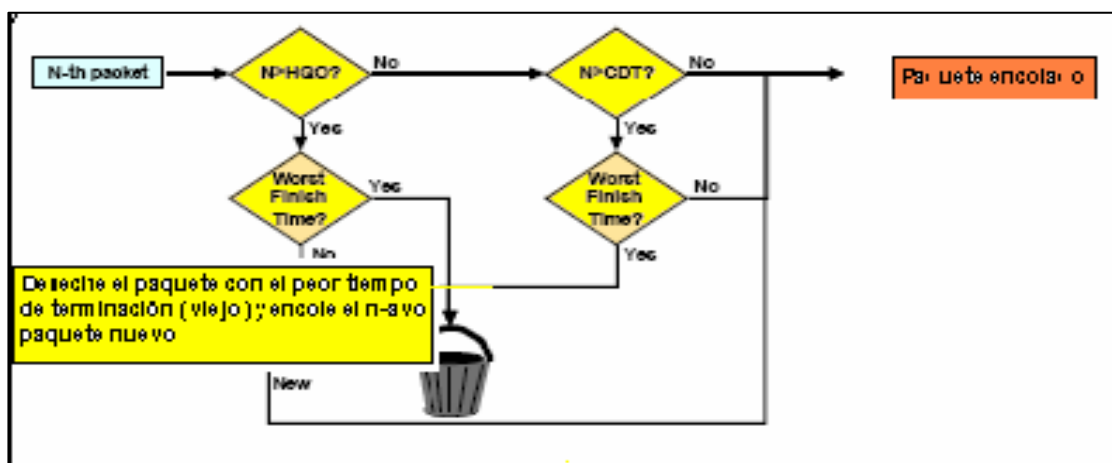
#### **1.5.4.2 Políticas de inserción y descarte de paquetes en WFQ**

WFQ tiene 2 modalidades de descarte, el primero se llama descarte temprano y se ejecuta cuando el umbral de descarte por congestión ( CDT= congestión discard threshold por sus siglas en ingles) y la otra descarte agresivo cuando el HQO ( hold queue out) es alcanzado. WFQ por defecto siempre descarta los paquetes de los flujos mas agresivos.

Flujos	64	128	256
5	15%	8%	4%
10	52%	30%	16%
15	83%	57%	34%
20	96%	79%	53%
25	100%	92%	70%
30	100%	98%	83%
35	100%	99%	91%
40	100%	100%	96%

**Tabla 2. Cifras probabilísticas de ocurrencia de 2 flujos terminando en la misma cola para 3 tamaños de colas en la estrategias WFQ: colas de 64, 128 y 256 paquetes - Tomado de Estudio y Configuración de Calidad de Servicio.**

La fig. 19 nos ilustra el esquema de descarte de WFQ. El proceso puede ser dividido en la siguiente etapa.



**FIG. 19 Esquema de descarte de WFQ - Tomado de Estudio y Configuración de Calidad de Servicio.**

1. Etapa 1 Descarta un nuevo paquete si el sistema WFQ esta lleno HQO (hold-queue out limit alcanzado) y el nuevo paquete tiene el peor “finish time” (es decir es el ultimo en todo el sistema).
2. Etapa 2 Descarta el paquete con el peor “finish time” en el sistema WFQ si el sistema esta lleno y encola un Nuevo paquete.

3. Etapa 3 Descarta el nuevo paquete si la cola donde el paquete debe ser encolado según su clase es la mas larga (no en paquetes si no en el “finish time” del nuevo paquete) y no hay en sistema WFQ mas paquetes que el número del CDT.
4. Etapa 4 Si llega a este punto encola el nuevo paquete.

#### **1.5.4.3 Políticas de programación de itinerario en WFQ**

Cada paquete es marcado con su *Finish time en un sistema* TDM virtual.

El planeador de itinerarios (el scheduler) selecciona el paquete con el mas temprano “finish time tag”, ese es el paquete que deja el TDM virtual primero

La longitud de la cola (para propósitos de planeación del itinerario) no es en paquetes sino en el tiempo que al sistema le tomaría transmitir todos los paquetes en la cola.

#### **1.5.4.4 WFQ basado en flujo**

Con WFQ basado en flujo, los paquetes son clasificados por el flujo. Los paquetes con la misma dirección fuente IP y dirección de destino IP, puerto fuente TCP o UDP, puerto de destino TCP o UDP, protocolo, y el mismo tipo de campo de servicio (ToS) pertenecen al mismo flujo.

Cada flujo corresponde a una cola separada de salida. Cuando un paquete se asigna a un flujo, se coloca en la cola para ese flujo. Durante los períodos de congestión, WFQ asigna una porción igual del ancho de banda a cada cola activa.

WFQ basado en flujo también se llama “formación de colas de espera justa (FQ por Fair Queue) porque todo los flujos tienen igual peso.

#### **1.5.4.5 Ventajas y desventajas de WFQ**

##### **Ventajas**

1. La configuración Simple (la clasificación no tiene que ser configurada).

2. Garantiza el throughput a todos los flujos.
3. Descarta los paquetes de la mayoría de los flujos agresivos.
4. Soporta la mayoría de las plataformas.
5. Diseñado para permitir que varias aplicaciones comportarían la red.
6. Garantiza un ancho de banda mínimo.
7. Garantiza unos valores aceptables de retardo.
8. En este método se comporta el ancho de banda de manera proporcional, sin dejar tráfico fuera de servicio.

#### **Desventajas**

1. Todas las desventajas de FIFO que hace cola dentro de una sola cola.
2. Los Múltiples flujos pueden terminar en una sola cola.
3. No soporta la configuración manual de clasificación.
4. No puede proporcionar las garantías fijas del ancho de banda.
5. Limitaciones en el rendimiento debido a la clasificación compleja y los mecanismos fijados.

De este método podemos decir que corrige la deficiencia del método anterior asociando en forma dinámica flujos de tráfico con colas. Bajo este método, todo el tráfico es supervisado y separado en dos categorías: los que requieren una gran cantidad de ancho de banda (caso datos) y los que requieren una cantidad relativamente pequeña de ancho de banda (caso VoIP).

Para efectuar esta separación, el clasificador del nodo analiza el header del paquete y define la existencia de un flujo que se asocia a un tipo de tráfico, pudiendo ser este IP (direcciones IP de fuente y destino y puerto TCP/UDP), Frame Relay (DLCI) o ATM (VPI/VCI).

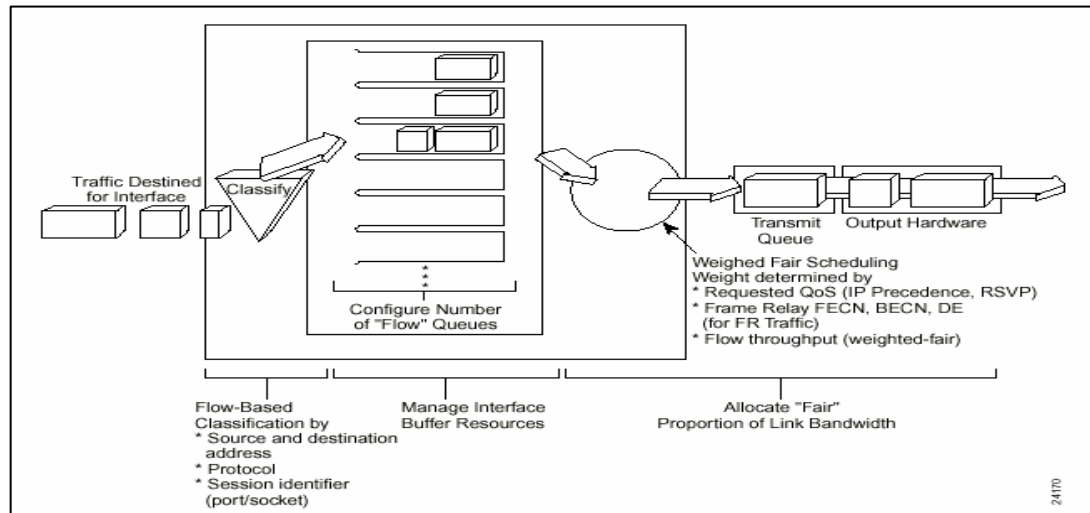
En definitiva, el algoritmo WFQ permite asignar dinámicamente colas para las dos categorías de tráfico, establecer umbrales de descarte por congestión (profundidad de cola) y priorizar colas basados en los bits de

IP Precedente y protocolo de reserva RSVP.

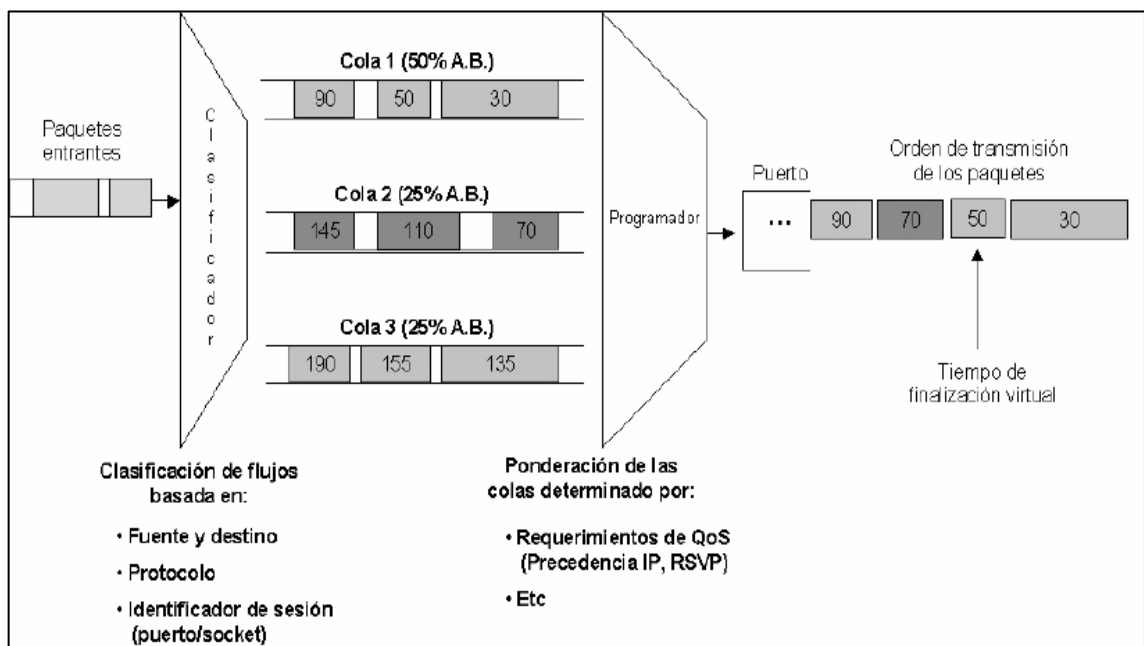
El objetivo de WFQ es asegurar que las aplicaciones que requieren de un ancho de banda reducido reciban un tratamiento preferente en la obtención del acceso a una interfaz de salida, a la vez que se permita que las aplicaciones que consumen mayor ancho de banda se beneficien del remanente que queda libre en proporción a sus pesos. WFQ asocia un peso a cada flujo basado en la información que extrae del header y en base a este peso secuencia el orden de salida de los paquetes a través de una única cola de salida. Las colas son servidas tomando primero las colas de reserva (RSVP) y luego en función decreciente de IP Precedente, dejando IP Precedente 0 para best effort.

Pese a que WFQ proporciona una distribución dinámica y equitativa de ancho de banda en función de la prioridad y que todos los flujos son tramitados en cada ciclo, incluso los de baja prioridad, en interfaces que tienen una gran cantidad de colas, aún los flujos con alta prioridad sufren retardos que podrían llegar a no ser aceptables. Para proporcionar QoS este método debe complementarse con la Prioridad IP RTP, en especial para aplicaciones que son procesadas a través de interfaces seriales de baja velocidad, menores a 2 Mbps.

## APOYO GRAFICO A Weighted Fair Queuing.



**FIG. 20. Estrategia De Encolamiento WFQ - Tomado de Estudio y Configuración de Calidad de Servicio.**



**FIG.21. Weighted Fair Queuing - Tomado de Estudio y Configuración de Calidad de Servicio.**

En la *Figura 20* y la *Figura 21* observamos como en el proceso de



encolamiento por WFQ se clasifican los flujos basándose en el TX y el RX y el protocolo, para luego ponderar las colas con los requerimientos de QoS en función de la reserva de los recursos.

### 1.5.5 Encolamiento en dWFQ (Distributed Weighted Fair Queuing)

La estructura de WFQ basado en Flujo distribuido (el dWFQ) es similar al WFQ mostrado en el anterior párrafo, sin embargo hay unas diferencias claves.

Hay 4 versiones diferentes de “distributed WFQ” tres de las cuales son discutidas en este párrafo:

- ToS-based dWFQ
- QoS-group-based dWFQ o
- QoS-based Dwfq

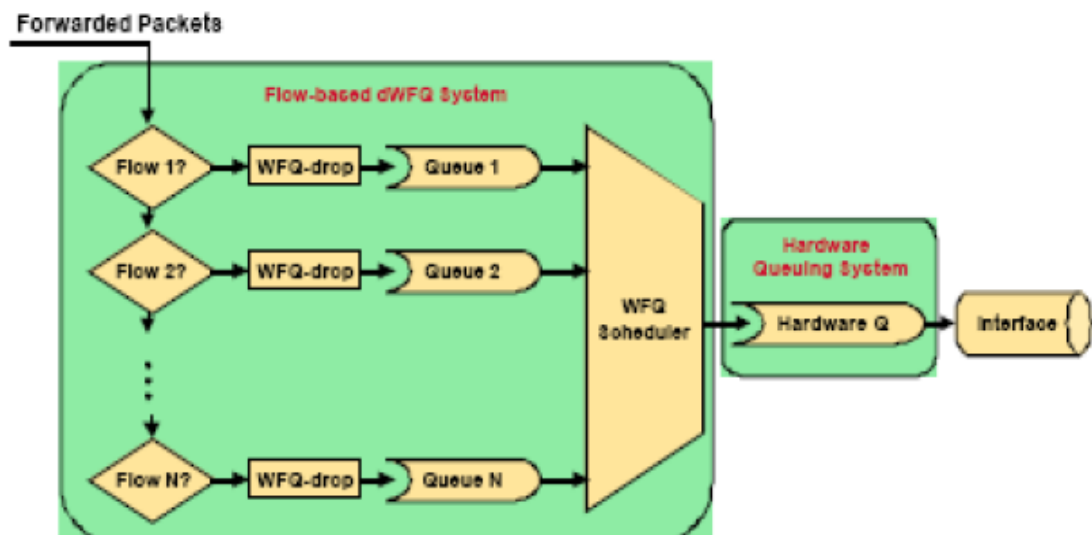


Figura 22: La estrategia dWFQ para QoS - Tomado de Estudio y Configuración de Calidad de Servicio.

#### 1.5.5.1 Clasificación en dWFQ

La clasificación identifica los flujos pero no usa el campo de ToS. Usa todos los otros parámetros que identifican un flujo (o conversación):

- Dirección de origen IP
- Dirección de destino IP
- Puertos y números de protocolo TC y UDP de origen y destino
- El número de colas es 512 y no puede cambiarse.

Continuación veremos una grafica que nos muestra el proceso de clasificación en d.C. ver fig. 23

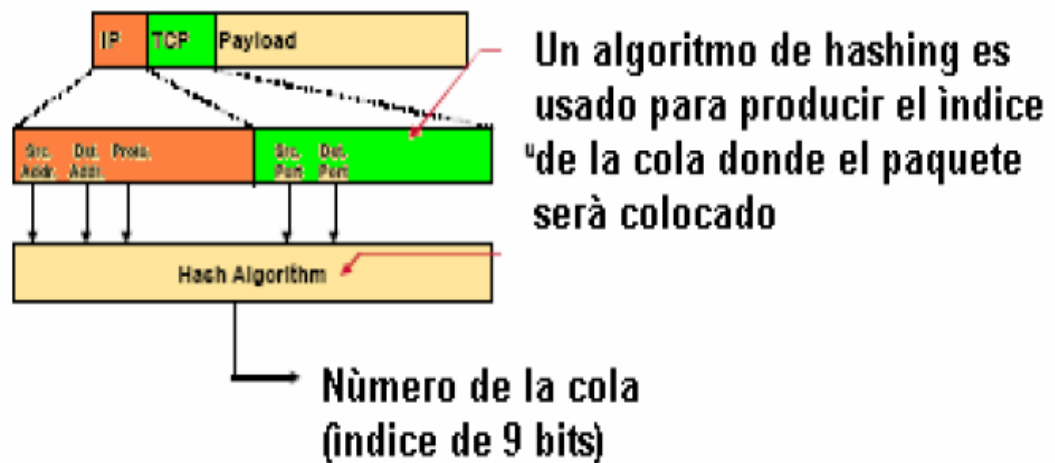


Figura 16: Proceso de clasificación en dWFQ - Tomado de Estudio y Configuración de Calidad de Servicio.

#### 1.5.5.2 Políticas de inserción y descarte de paquetes en dWFQ

Las políticas de inserción reposan en 3 puntos básicos:

- dWFQ descarta los paquetes cuando ambos el límite de la cola individual (IQL) y el límite agregado de la cola (AQL) se alcanza
- El dWFQ no es tan estricto con los flujos agresivos como WFQ no-distribuido
- La inserción y la política de descarte son las mismas para todas las tres versiones de dWFQ (basada en flujo, basada en ToS, y basada en grupo QoS)

#### 1.5.6 dWFQ basado en clases (Class-based Weighted Fair Queuing)

Es el WFQ basado en clase, en el cual se asignan los paquetes a colas diferentes basadas en su grupo de QoS o la precedencia IP en el campo

de ToS.

El encolamiento class-based weighted fair queuing (CB-WFQ) incorpora ideas de las colas personalizadas al formato del encolamiento weighted fair (WFQ). Se han introducido considerables mejoras al respecto a las técnicas de encolamiento más antiguas. El número Máximo de colas personalizadas (denominadas clases) ha aumentado de 16 a 64. Sigue existiendo mucha flexibilidad a la hora de asignar el tráfico a cada clase. Cada una de ellas puede usar el método tail-drop de administración de la profundidad de la cola (es decir, posibilita que las colas se llenen y sobrecarguen durante la congestión), o bien WRED puede configurarse para cada clase independientemente. Con CB-WFQ, puede especificarse directamente la cantidad de ancho de banda que quiere destinar a cada clase, y CB-WFQ ajustará los parámetros internos para que esto ocurra. Con WFQ, es difícil especificar con precisión la cantidad de ancho de banda destinada a un flujo. El problema es que el ancho de banda actual por flujo depende del número de flujos, los cuales cambian constantemente. El parámetro anchura influye en el destino del ancho de banda en WFQ, pero solo existen seis anchos que se pueden especificar (precedencia IP de 0 a 5), más uno por flujo RSVP. Dado que WFQ es un algoritmo de encolamiento, cualquier ajuste de los flujos o niveles de precedencia IP afecta a todos los otros niveles. Esto se suma a la complejidad de la administración del ancho de banda con WFQ. CB-WFQ soluciona el problema de la asignación del ancho de banda a las clases individuales, usando el concepto de cola personalizada. En una cola personalizada, las colas individuales se configuran con un valor de cuenta de bytes que controla la cantidad de ancho de banda usada en cada cola. La implementación CB-WFQ es más ajustada, ya que puede especificarse directamente la cantidad de ancho de banda que quiere asignar a cada clase. El algoritmo ajusta el ancho asignado a la clase en función de los ajustes de ancho de banda de cada clase y el total de ancho de banda disponible en el enlace.

Como recomendación global para IP, debe utilizar una de las siguientes técnicas de encolamiento para interfaces de ancho de banda bajo, dependiendo de sus necesidades: WFQ, CB-WFQ, Ambas técnicas ofrecen una distribución de ancho de banda equitativa y deben usarse con un mecanismo para priorizar el tráfico VoIP. WFQ requiere prioridad IP RTP, y CB-WFQ una clase de tráfico priorizada, denominada LLQ. Puede decidir que técnica utilizar basándose en los requisitos de asignación del ancho de banda y el nivel de complejidad de la configuración. WFQ apenas necesita configuración, pero CB-WFQ requiere configuración avanzada.

### 1.5.7 Encolamiento MDR (Modified Deficit Round-robin)

Los proveedores de Internet y redes corporativas generalmente no tienen soporte para estrategias de encolamiento como WFQ, PQ o CQ. La razón que tienes que ver con los miles de rutas, y gran cantidad de tarea que uno de estos enrutadores debe realizar, en razón a dedicarse a las tareas complejas de estas estrategias les quitaría ciclos de CPU muy críticos para su tarea principal como es el enrutamiento a redes. Por ello estos enrutadores implementan MDDR (Modified Deficit Round-robin) para el manejo de congestión todavía bajo la carpa de enrutamiento basado en políticas. MDDR paradójicamente puede ser implementado en interfases salientes y entrantes para esto ver Fig. . 22

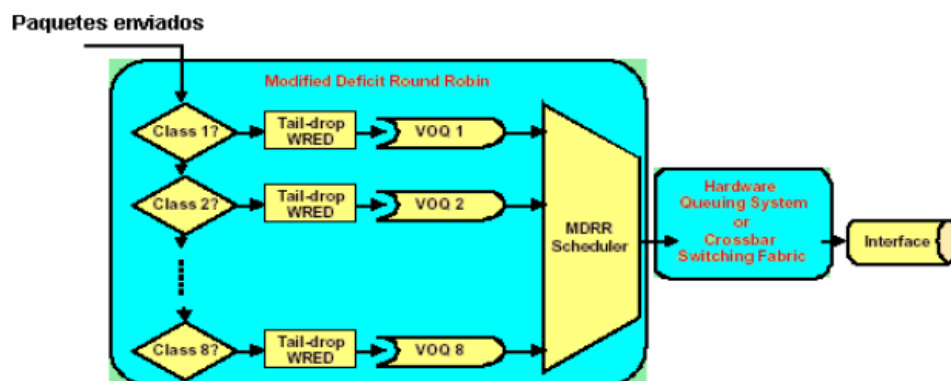


Figura 22: Arquitectura de MDDR - Tomado de Estudio y Configuración de Calidad de Servicio.

DRR ( Déficit Round Robin) es el mecanismo de encolamiento basado en clases y el modificado adiciona el encolamiento de baja latencia y una capacidad de soporte para 8 clases.

El MDRR clasifica paquetes basados en los valores de los bits de precedencia IP, por esta razón solo puede soportar 8 clases de colas. Esa colas son servidas en modo round robin y a cada cola se le permite transmitir un número fijo de bytes cada vez que es servida ( el quantum de la cola). Un contador de déficit sigue la pista de cuantos bytes ha transmitido una cola en cada vuelta. Cuando las colas son configuradas inicialmente se establece que el contador de déficit de cada cola sea igual al quantum de cada cola. Cuando los paquetes son transmitidos desde una cola sus longitudes son restadas de el contador de déficit , cuando este contador llega a 0, el programador de itinerario se mueve a la siguiente cola. En cada nueva vuelta el contador de déficit de una cola no vacía es incrementado en el valor del quantum.

Fuera de las colas basadas en IP-precedencia ToS MDRR implementa un cola especial (la cola 0) que puede ser atendida en una de dos maneras:

**1.5.7.1 Modo de estricta prioridad:** En la cual la cola puede ser servida si no esta vacía. O dicho de otro modo se atiende hasta vaciarla. Modo de estricta prioridad significa que la cola especial suministra bajo retardo, pero si muchos paquetes son asignados a ella entonces una gran cantidad de paquetes de las otras colas pueden sufrir de inanición.

**1.5.7.2 Modo alternado:** Significa que se atiende alternativamente la cola especial y una de las otras colas.

MDRR puede implementarse en las interfases de salida (como en todas las otras estrategias de conformación de colas de espera) o por delante de la matriz de conmutación en la interfaz de entrada.

Otra características es que Deficit Round Robin (DRR) utiliza 8 Virtual

Output Queues (VOQ) o colas para prevenir bloqueo a la cabeza de la cola

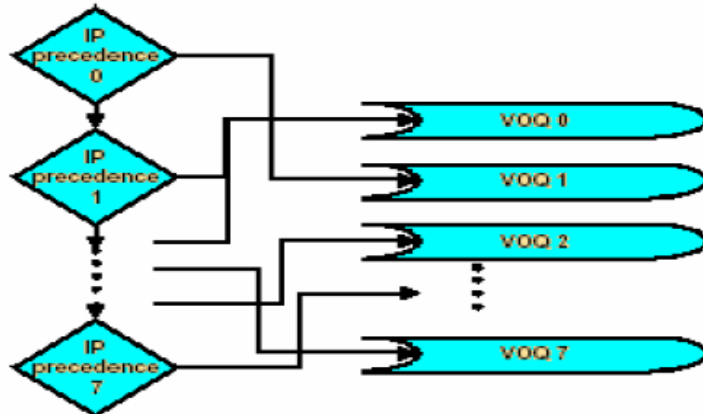


Figura 23 : El esquema de clasificación de MDRR - Tomado de Estudio y Configuración de Calidad de Servicio.

La clasificación se hace mediante la lectura de los 3 bits de precedencia IP. Una de las colas la VOQ0 IP prec.=000 es la cola expedita de baja latencia.

#### 1.5.7.3 Clasificación en MDRR

Cada cola puede ser configurada para soportar WRED o tail-drop en el descarte, para prevenir congestión.

#### 1.5.7.4 Planeación de itinerario MDRR

En DRR la planeación de itinerarios es similar a la de Custom Queuing, excepto en que es mas exacta. DRR recuerda el número de bytes enviados por encima del umbral en la ronda previa (déficit).ver fig 24.

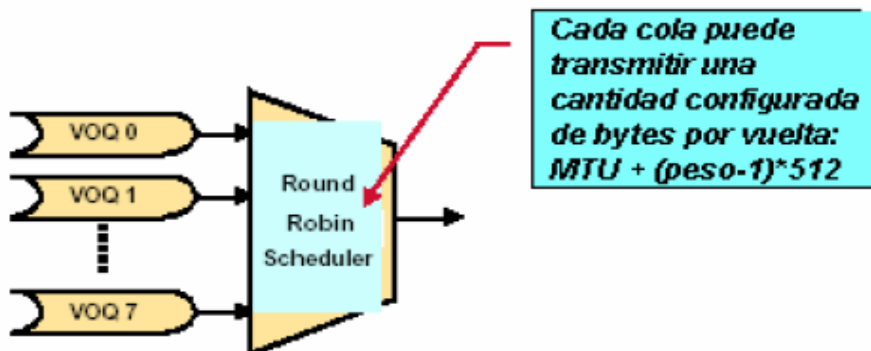


Figura 24: La operación del scheduler en MDRR - Tomado de Estudio y Configuración de Calidad de Servicio.

Política de servicio para una cola en una ronda:

1. Adicionar  $MTU+(Weight-1)*512$  tokens al deposito de tokens.
2. Transmitir paquetes hasta que todos los tokens se usen o hasta que la cola se vacíe.
3. Colocar el deposito de token a 0 si la cola esta vacía. De lo contrario recuerde el déficit (cuantos tokens por encima los disponibles fueron usados).
4. Empezar a servir la siguiente cola.

### 1.5.8 Priorización de paquetes IP RTP

Hasta aquí nos hemos referido a las técnicas de encolamiento desde la perspectiva de un escenario de las colas simples o múltiples.

El modelo de encolamiento de múltiples etapas es útil para proveer una solución a los conflictivos requisitos de la QoS, como son flujos específicos y el ancho de banda equitativo para todos los flujos (véase la figura 25). La cola del primer escenario identifica el trafico de prioridad alta que no puede tolerar retrasos y fluctuaciones de fase. El trafico restante pasa a través de la segunda cola, que proporciona un tratamiento equitativo a los distintos flujos de datos.

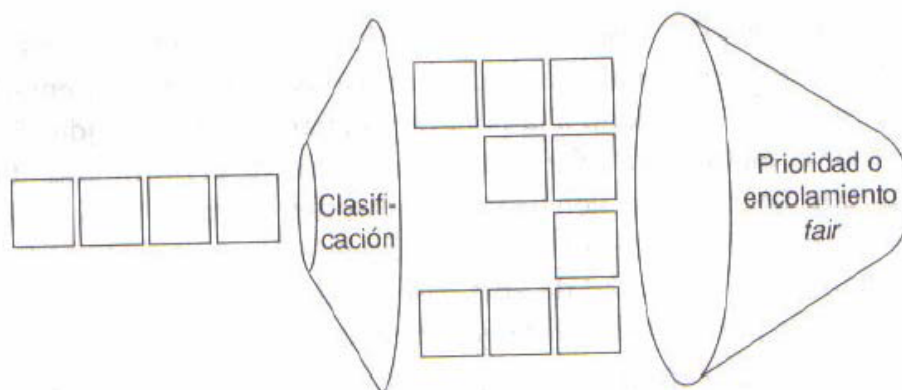
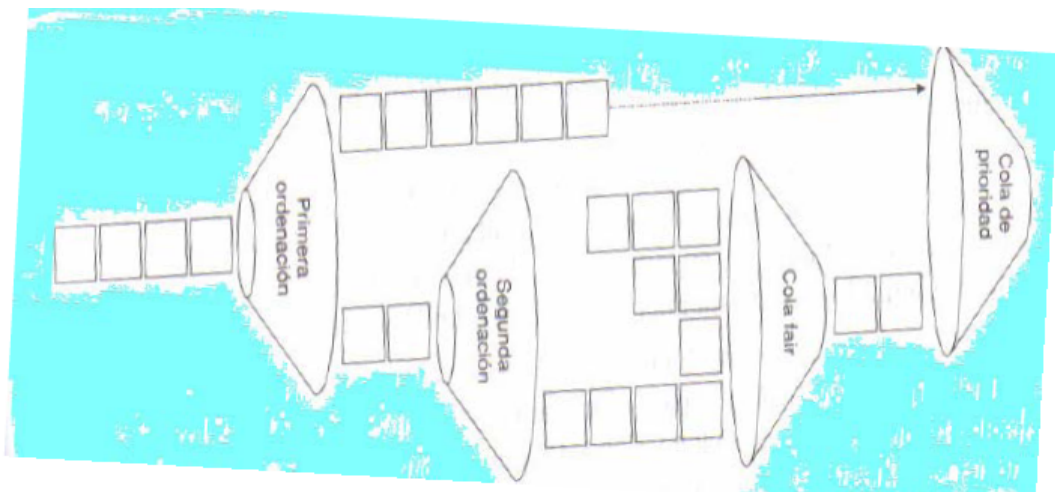


Figura 25 : El modelo de encolamiento sencillo puede proporcionar distribución QoS o de ancho de banda equitativo – Tomado de Conformador de Trafico de un Nodo, Eduardo de La Cruz Gamez

Para VoIP, este modelo de encolamiento de múltiples etapas se implementa con la función prioridad IP RTP. Cuando esta función esta

habilitada, todo el tráfico RTP que usa puertos UDP específicos se coloca aparte en la cola de escenario simple y se envía directamente al buffer transmitido. El tráfico restante utiliza WFQ o CB-WFQ para asignar de forma equitativa el ancho de banda a los flujos de tráfico o clases. Observe que la prioridad RTP proporciona un retraso y fluctuaciones de fase bajas en la ruta de audio de la conversación VoIP, pero la configuración de la llamada, destrucción y cualquier otro mensaje de señalización debe seguir cruzando la cola del segundo escenario. En el caso de interfaces ocupadas, CB-WFQ mejora el rendimiento de los mensajes de señalización de llamada .ver. Fig. 26



**Figura 26. Múltiples etapas del modelo de encolamiento IP RTP suministrando QoS y distribución equitativa de ancho de banda - - Tomado de AURA, Ganz, Multimedia Wires Network; Standard and Qos**



### 1.5.8.1 Clasificación de IP RTP

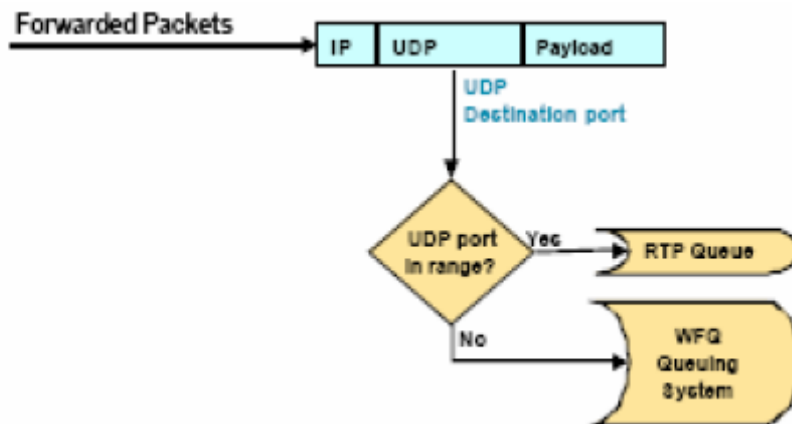


Figura 27 : El procedimiento de clasificación de IP TRP – Tomado de Conformador de Trafico de un Nodo, Eduardo de La Cruz Gamez.

IP RTP clasifica los paquetes basado en puertos UDP. Es decir si el puerto UDO de destino del segmento UDP esta en el rango permitido el paquete es clasificado y se le asigna una cola de alta prioridad.

### 1.5.8.2 Ventajas y desventajas de IP RTP

#### Ventajas

- Agrega baja latencia a las colas WFQ y CBWFQ.
- Previene la inanición de las otras colas de tráfico no prioritario aunque no con mucha eficiencia, lo intenta hacer.

#### Desventajas

- Presenta pobres opciones de clasificación.
- La estrategia Class-based Low-latency hizo obsoleta la estrategia IP RTP.

## 1.6 Enrutamiento basado en clases

- Class-based Weighted Fair Queuing
- Class-based Low-latency Queuing,

### **1.6.1 Class-based Weighted Fair Queuing** (Clase ponderado basado en la FERIA cola )

CBWFQ extiende la funcionalidad estándar WFQ para prestar apoyo a las definidas por el usuario clases de tráfico. Para CBWFQ, usted define las clases de tráfico sobre la base de criterios, entre ellos coinciden con los protocolos, listas de control de acceso (ACLs), y las interfaces de entrada. Los paquetes que cumplen los criterios para coincidir con una clase constituyen el tráfico de dicha categoría. A la cola se reserva para cada clase, y el tráfico que pertenece a una clase está dirigido a la cola de dicha categoría.

Una vez que la clase ha sido definida en función de sus criterios de partido, puede asignarle características. Para caracterizar una clase, se le puede asignar el ancho de banda, el peso, y límite máximo de paquetes. El ancho de banda asignado a una clase es el ancho de banda garantizado entregado a la clase durante la congestión.

Para caracterizar una clase, usted también especificar el límite de la cola de esa clase, que es el número máximo de paquetes permitido acumular en la cola para la clase. Los paquetes pertenecientes a una clase están sujetos a la cola de ancho de banda y límites que caracterizan a la clase. Si una clase por defecto está configurada con la política de ancho de banda de ruta clase de configuración de mando, todos sin clasificar el tráfico se pone en una sola cola y el tratamiento dado en función de la configuración de ancho de banda. Si una clase por defecto está configurado con el aire f-cola de mando, todos sin clasificar flujo de tráfico es clasificado y dado mejor esfuerzo tratamiento. Si no hay clase por defecto está configurado, por defecto, entonces el tráfico que no coincide con ninguna de las clases está configurado flujo clasificado y dado mejor esfuerzo tratamiento. Una vez que un paquete se clasifica, todos los mecanismos estándar que puede utilizarse para diferenciar entre el servicio de las clases de aplicación.

Flujo de clasificación es el tratamiento estándar WFQ. Es decir, los paquetes con la misma dirección IP de origen, destino, dirección IP,

fuentes de Transmisión Control Protocol (TCP), o User Datagram Protocol (UDP) del puerto, o el destino TCP o UDP puerto están clasificados como pertenecientes al mismo flujo. WFQ asigna una parte igual de ancho de banda para cada flujo. Flow-WFQ basado también se le llama justo cola porque todas las corrientes son igualmente ponderadas.

Para CBWFQ, que se extiende la norma WFQ justo cola, el peso especificado para la clase se convierte en el peso de cada paquete que satisfaga los criterios del partido de la clase. Los paquetes que llegan a la interfaz de salida se clasifican de acuerdo a criterios coinciden con los filtros que definir, entonces cada uno se le asigna la debida importancia. El peso de un paquete que pertenecen a una clase específica se deriva del ancho de banda asignado a la clase cuando se configura, en este sentido, el peso de una clase de usuario es configurable.

Después de que el peso de un paquete es asignado, el paquete es enqueued en la clase correspondiente cola. CBWFQ utiliza las ponderaciones asignadas a la cola los paquetes para asegurarse de que la cola de clase cuenta con los servicios de manera equitativa.

Configuración de una clase política-por lo tanto, configurar CBWFQ-implica estos tres procesos:

- Definición de clases de tráfico para especificar la clasificación política (clase mapas).

Este proceso determina cuántos tipos de paquetes se ha de diferenciarse el uno del otro.

- Asociar las políticas-es decir, la clase de características con cada clase de tráfico (política mapas).

Este proceso implica la configuración de las políticas que deben aplicarse a los paquetes que pertenecen a una de las categorías previamente definidas a través de una clase de ruta. Para este proceso, puede configurar una política de ruta que especifica la política de cada clase de tráfico.

- Adjuntar a las políticas de interfaces (las políticas de servicio).

Este proceso requiere que usted asocie una política existente de ruta, o la política del servicio, con una interfaz de aplicación del particular conjunto de políticas para el mapa para que la interfaz.

CBWFQ permite especificar la cantidad exacta de ancho de banda para ser asignado a una clase específica de tráfico. Teniendo en cuenta el ancho de banda disponible en la interfaz, puede configurar hasta 64 clases y controlar la distribución entre ellos, que no es el caso de flujo basado en WFQ.

Flow-WFQ basado en pesos se aplica al tráfico de clasificación en las conversaciones y determinar cuánto ancho de banda de cada conversación se permite en relación con las demás conversaciones. Para el flujo de base WFQ, estos pesos, y el tráfico de clasificación, dependen de y se limita a los siete niveles de precedencia IP.

CBWFQ permite definir lo que constituye una clase sobre la base de criterios que superan los confines de flujo. CBWFQ permite usar listas de control de acceso y protocolos de interfaz de entrada o nombres para definir cómo el tráfico se clasifica, usted no necesita mantener el tráfico de clasificación en un flujo. Además, puede configurar hasta 64 clases discretas en una política de servicio.

- Si configura una clase política en un mapa para el uso de paquetes WRED para soltar en lugar de la cola caída, debe asegurarse de que WRED no está configurado sobre la interfaz a la que le atribuyen la intención de que la política del servicio.

#### **1.6.1.1 La definición de clase mapas**

Para crear una clase de mapa que contiene los criterios de partido en contra de un paquete que se verifica para determinar si pertenece a una clase-y de manera eficaz a crear la clase política cuya pueden especificarse en uno o varios mapas de la política de utilizar el primer

comando en el modo de configuración global para especificar el nombre de ruta de clase, entonces uno de los siguientes comandos en la clase mapa modo de configuración:

Configuración de la clase política en la política de ruta

Para configurar una política de ruta y crear clases políticas que conforman la política del servicio, utilice la política de ruta para especificar el nombre de ruta política, a continuación, utilice uno o varios de los siguientes comandos para configurar la política de una clase estándar o la falta de clase:

- Clase
- Ancho de banda
- Justo-cola (para la clase de clase sólo por defecto)
- - cola de limitar o aleatorios detectar

Para cada clase que usted defina, puede utilizar uno o más de los que figuran los comandos para configurar la clase política. Por ejemplo, puedes especificar el ancho de banda para una clase y tanto el ancho de banda y la cola de límite de otra clase.

El valor por defecto de la clase política mapa (comúnmente conocida como la clase de la clase por defecto) es la clase a la que el tráfico se dirige en caso de que el tráfico no cumple los criterios de coincidir con otras clases cuya política se define en la política de ruta.

Puede configurar las políticas de clase para el mayor número de clases como se definen en el router, hasta un máximo de 64. Sin embargo, la cantidad total de ancho de banda asignada para todas las clases incluidas en una política de ruta no debe exceder del 75 por ciento del ancho de banda disponible en la interfaz. El otro 25 por ciento se utiliza para el control y encaminamiento del tráfico. (Para anular la limitación del 75 por ciento, utiliza el máximo de ancho de banda reservado comando.) Si no de todos es el ancho de banda asignado, el resto de ancho de banda es asignado proporcionalmente entre las clases, sobre la base de su ancho de banda configurado.

### 1.6.2 LLQ Low Latency Queuing

Esta técnica usa un manejo mixto de colas. Una cola específica para tráfico de voz o tráfico DSCP Expedited Forwarding bajo modalidad PQ y el resto de colas a través de clases asociadas a CBWFQ.

La diferencia con IP RTP PRIORITY es que no está limitada a filtrado por port UDP/RTP.

Mientras (WFQ) provee una porción justa de ancho de banda a cada flujo, y proporciona planificación justa de sus colas, no puede garantizar ancho de banda y un retardo bajo a las aplicaciones seleccionadas. Por ejemplo, el tráfico de Voz no puede competir con otros flujos agresivos en el sistema WFQ de colas el cual no tiene planeador de itinerario de prioridad para las aplicaciones sensibles al retardo.

IP RTP es un protocolo diseñado para garantizar la prioridad del tráfico de la voz. Sin embargo, porque sólo puede priorizar puro tráfico RTP trafican (IP RTP usa un rango heurístico de puertos UDP para distinguir RTP del resto del tráfico), y carece de flexibilidad en las políticas, no presenta una solución viable cuando tráfico non-RTP de múltiples aplicaciones críticas y sensibles al retardo de tiempo se despliegan en la red.

El encolamiento de Baja latencia basado en clases (CB-LLQ) es un método, usado dentro de la estructura CBWFQ, que puede priorizar los flujos de tráfico con la flexibilidad que le pueda soportar la interfaz del sistema operativo del enrutador

VoIP y otras aplicaciones multimedia no son igualitarias con las colas de datos. El sistema de prioridades de IP RTP tiene problemas como el hecho de que solo trabaja en UDP, no soporta TCP y otros protocolos, solo maneja una prioridad la alta prioridad. Aparece CB-LLQ como la estrategia que remueve las anteriores deficiencias en IP RTP.

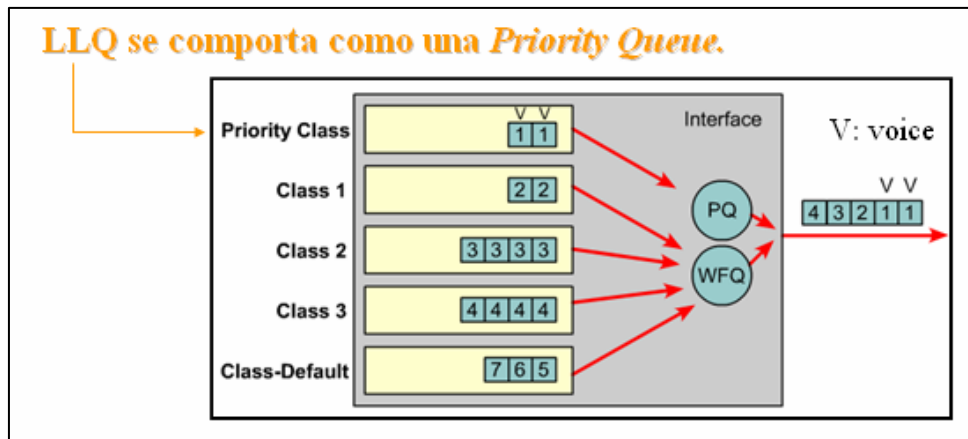


FIG.24. LLQ Low Latency Queuing - Tomado de Estudio y Configuración de Calidad de Servicio.

### 1.7 TABLAS DE RELACION DE DIFERENTES TIPOS DE MECANISMOS DE ENCOLAMIENTO.

En las siguientes tablas podemos observar como al cambiar los mecanismos de encolamiento en los equipos de comunicación el tiempo de respuesta llega a variar.

		Mecanismo de encolamiento						
		FIFO		PQ		WFQ		
		Tiempo de respuesta promedio (milisegundos)	Intervalo de Confianza (95%)	Tiempo de respuesta promedio (milisegundos)	Intervalo de Confianza (95%)	Tiempo de respuesta promedio (milisegundos)	Intervalo de Confianza (95%)	
Tamaño paquete.	Bytes	64	0.529800	0.000879	0.532567	0.001869	0.528033	0.001213
		128	0.781133	0.002358	0.749400	0.001202	0.751300	0.002302
		256	1.202667	0.001864	1.201000	0.002174	1.203667	0.002737
		512	2.154000	0.002224	2.135667	0.002605	2.122667	0.002809
		1024	3.951000	0.001720	3.989667	0.001754	4.269259	0.098073
		1280	4.867667	0.002034	4.856667	0.002713	4.865000	0.002441
		1518	5.717333	0.002289	5.720333	0.002393	5.720333	0.002894

TABLA.3 Tiempo de respuesta en Mecanismo de encolamiento – Tomado Paket Shaper, Aplicaciones de Red.

		Método de encolamiento									
		FIFO			PQ			WFQ			
		Caudal Eficaz (Mbps)	Jitter (ms)	% Paquetes Perdidos	Caudal Eficaz (Mbps)	Jitter (ms)	% Paquetes Perdidos	Caudal Eficaz (Mbps)	Jitter (ms)	% Paquetes Perdidos	
Datos enviados	Mbps	1	0.998	0.097	0	1.000	0.445	0	1.000	0.447	0
		2	2.000	0.515	0	2.000	0.528	0	2.000	0.537	0
		3	3.000	0.545	0	2.997	0.525	0	3.000	0.553	0
		4	4.000	0.524	0	3.993	0.550	0.159	3.993	0.549	0.159
		5	4.894	0.559	2.114	4.895	0.564	2.099	4.882	0.560	2.237
		6	5.791	0.699	3.464	5.794	0.700	3.412	5.794	0.700	3.412
		7	6.610	0.532	5.547	6.602	0.539	5.656	6.613	0.541	5.502
		8	7.408	0.398	7.387	7.404	0.406	7.408	7.397	0.408	7.512
		9	8.372	0.958	6.857	8.448	0.298	6.163	8.438	0.306	6.268
		10	8.911	0.917	10.635	8.977	0.938	10.221	8.990	0.916	10.099

**TABLA.4** Parámetros de respuesta en Mecanismo de encolamiento.



## CONCLUSIONES.

1. Los recursos finitos de los nodos, tales como la capacidad de los buffers de entrada y salida, capacidad de procesamiento de paquetes por segundo de los nodos deben ser precautelados dando al tráfico crítico un tratamiento diferencial que vaya acorde con la calidad de servicio pactado para cada uno de ellos. Esta opción gana mayor validez si consideramos que en una red multiservicios conviven varios tipos de tráficos y varios de ellos con estrictos requisitos de calidad de servicio, como son las aplicaciones de voz y multimedia.
2. Las colas de espera tienen un papel muy importante ya que entre mas tiempo pasen los paquetes en estas colas mayor es el tiempo total de la comunicación, existen diferentes mecanismos de encolamiento basados en algoritmos que varían desde lo más simple hasta lo más complejo.
3. Podemos concluir que las técnicas de manejo de colas son fundamentales para proporcionar Factor de Calidad (**QoS**) en redes convergentes; las mismas entran en acción cuando se ha producido una congestión y es necesario precautelar el transito acelerado del tráfico prioritario a través de los nodos congestionados.
4. En definitiva podemos decir que las diferentes técnicas de encolado ordenan los paquetes en colas bajo diferentes enfoques de prioridad, esto genera en cada nodo una selección del orden y cantidad de bytes que se tomaran de cada cola a transmitir en el proceso de transmisión hacia el siguiente nodo.

5. El encolamiento en los enrutadores es necesario para acomodar las ráfagas de paquetes que se presentan cuando la rata de llegada es mayor que la rata de salida, podemos concluir que esto se debe a que el interfaz de entrada es más rápido que el interfaz de salida o que la interfaz de salida está recibiendo paquetes de múltiples interfaces de entrada simultánea.
6. Concluimos que la cola por software se adiciona en un paquete cuando la cola por hardware se llena a su máxima capacidad, mientras la cola por hardware tenga espacio los paquetes no van a la cola por software, evidentemente mientras la cola por Hardware no se llene no hay oportunidad de hacer ningún procedimiento de QoS, lo cual es natural, pues si la cola por hardware está con cupo no hay ningún problema de congestión y no hay necesidad de QoS.
7. Nos damos cuenta que la cola FIFO es un buffer sencillo que retiene los paquetes salientes hasta que la interfaz los retire en el mismo orden en el que llegaron al buffer, Esta técnica de encolamiento no proporciona ninguna QoS para el flujo, ni distribuye un ancho de banda equitativo entre flujos que compartan un enlace. Concluimos que los periodos de congestión, se llama el buffer y los paquetes se descartan sin importar el tipo de paquete o las solicitudes de la aplicación asociada, esto es muy malo para aplicaciones como VoIP.
8. Podemos analizar que el método de encolamiento Priority Queuing le da prioridad en un sencillo enfoque donde ofrece un tratamiento preferencial a los paquetes identificados, ordenando los paquetes en cuatro colas de diferente prioridad: alta, media, baja y normal.

9. Encontramos que Priority Queuing es básicamente una colección de cuatro colas tipo FIFO paralelas, en donde cada cola padece todos los problemas de FIFO, aislados a la clase (desigual, inanición, jitter), por lo tanto Cada cola usa el esquema del desecho del paquete IP cuando la cola está llena.
10. Podemos concluir que el método Priority Queuing trata de mejorar la deficiencia que presenta el método de encolamiento Priority Queuing en la asignación equitativa de ancho de banda.
11. También analizamos que el método Priority Queuing permite que ningún protocolo monopolice los servicios del nodo más allá de una determinada porción de ancho de banda total, para esto el método establece hasta 16 buffer de encolado y la asignación de capacidad se asigna en términos de bytes por cola que se debe vaciar antes de pasar a atender la siguiente.
12. Analizamos que bajo el encolamiento WFQ el tráfico es supervisado y separado en dos categorías: los que requieren una gran cantidad de ancho de banda (caso datos) y los que requieren una cantidad relativamente pequeña de ancho de banda (caso VoIP).
13. En definitiva, podemos decir que el algoritmo WFQ permite asignar dinámicamente colas para las dos categorías de tráfico, establecer umbrales de descarte por congestión (profundidad de cola) y priorizar colas basados en los bits de IP Precedente y protocolo de reserva RSVP.
14. También concluimos que el objetivo de WFQ es asegurar que las aplicaciones que requieren de un ancho de banda reducido reciban un tratamiento preferente en la obtención del acceso a una interfaz

de salida, a la vez que se permita que las aplicaciones que consumen mayor ancho de banda sean beneficiadas.

15. Concluimos que las diferentes técnicas de encolado ordenan los paquetes en colas bajo diferentes enfoques de prioridad. Esto genera en cada nodo una selección del orden y cantidad de bytes que se tomarán de cada cola a transmitir en el proceso de transmisión hacia el siguiente nodo.

## BIBLIOGRAFIA.

- 1 THEODORES, Raparot, Wirreles Communications; Segunda Edición, Prentice Hall. 2002.
- 2 AURA, Ganz, Multimedia Wires Network; Standard and Qos, Edition Prentice Hall. 2004.
- 3 NEILE, Presand, Wland System and Wireles IP for next Generation of communication, Ed artch Houes. 2004.
- 4 MUÑOS, Luis, Wlan y Wpan. Edición. Ed artch Houes. 2004.
- 5 Tanenbaum, A. S.: Redes de Computadoras. Prentice Hall. (1997).
- 6 Acampora, A.: An Introduction to Broadband Networks. Plenum Press New York (1994).
- 7 Dutton, H., Lenhard, P.: High-Speed Networking Technology. IBM Prentice Hall New Jer-sey
- 8 <http://www.axis.com/es/documentacion/Las%20redes%20IP.pdf>
- 9 [www.idg.es/comunicaciones/pdf/Soluzio06.pdf](http://www.idg.es/comunicaciones/pdf/Soluzio06.pdf)
- 10 [www.gcr.tsc.upc.edu/publications/proceedings/1999/Gestion del mecanisdetraspaso.pdf](http://www.gcr.tsc.upc.edu/publications/proceedings/1999/Gestion_del_mecanisdetraspaso.pdf)