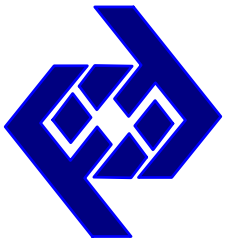
UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE INGENIERIA
ESCUELA DE INGENIRÍA ELECTRICA
COMISION DE ESTUDIOS DE POST-GRADO



ESTRATEGIAS PARA GARANTIZAR CALIDAD DE SERVICIO EN LA INTEGRACION DE VOZ, VIDEO Y DATOS EN REDES CORPORATIVAS

Trabajo de Grado presentado ante la Universidad Central de Venezuela
por
Luis Roberto Madera Blanco
C.I. V-11488049
Como requisito para optar al título de
Especialista en Ingeniería Eléctrica Mención
Comunicaciones y Redes de Comunicación de Datos.

Septiembre de 2004

TRABAJO DE GRADO

ESTRATEGIAS PARA GARANTIZAR CALIDAD DE SERVICIO EN LA INTEGRACION DE VOZ, VIDEO Y DATOS EN REDES CORPORATIVAS

Autor : Ing. Luis Roberto Madera Blanco

Tutor : Ph.D. Luis Fernández

Septiembre, 2004

Hecho el Deposito de Ley
Deposito Legal lft487200462158

DEDICATORIA

Este trabajo de grado se lo dedico a Dios
y a mi Madre por ser la fuente de toda
Mi inspiración...

AGRADECIMIENTOS

A dios y a mi Madre pues sin ellos no podría lograr ninguna meta en mi vida.

A mis hermanas por apoyarme en todas las cosas en las que deseo triunfar.

A mi Tutor Prof. Luis Fernández por orientarme no solo en mi tesis de grado sino durante el transcurso de toda la Especialización.

Muy en especialmente a la señora Gipsy (Sec. Postgrado Ingeniería UCV). Por todo el apoyo prestado durante la duración de la especialización y por ser la persona que motoriza muchas de las cosas importantes para los alumnos del postgrado, **Señora Gipsy Muchas Gracias.**

A Mis amigos más cercanos, Héctor Plaza, Jhonder Rosales, Néstor Hernández, por estar siempre en el lugar y en el momento indicado.

A Roxana Ojeda. Por haber llegado a mi vida en el momento que más lo necesitaba y por sus palabras de apoyo, amor y ternura las cuales necesita todo hombre para seguir adelante.

A todos mis compañeros del postgrado por permitirme ser parte de su grupo y por el apoyo brindado.

A la Universidad Central de Venezuela, por permitirme tener la dicha de estar en sus aulas y ser un miembro mas de la comunidad UCVISTA.

A todos los que de una u otra forma contribuyeron a que esta meta se lograra satisfactoriamente.

UNIVERSIDAD CENTRAL DE VENEZUELA
COMISION DE ESTUDIOS DE POST-GRADO
FACULTAD DE INGENIERIA
ESCUELA DE ELECTRICA

TRABAJO ESPECIAL DE GRADO

Especialización: Comunicaciones y Redes de Comunicación de Datos
Tema: Estrategias para Garantizar Calidad de Servicio en la Integración de Voz, Video y Datos en Redes Corporativas.
Ponente: Luis R. Madera B.

Resumen:

En la actualidad los diferentes servicios de información están tendiendo a ser transmitidos a través de redes convergentes; estas redes transportan Voz, Video y Datos de manera integrada entre localidades de una corporación e incluso entre corporaciones. Esta tendencia trae consigo muchísimas ventajas que permiten ofrecer servicios cada vez más eficientes y efectivos a los usuarios de la red, incidiendo de manera directa en la productividad empresarial. Para que una red pueda integrar servicios diferentes esta debe cumplir con los requerimientos necesarios para un funcionamiento optimo de cada uno de ellos. Uno de los aspectos más importantes a considerar para lograr una integración de servicios mediante una red convergente, es la Calidad de Servicio (QoS) ofrecida, para eso existen muchas tecnologías que son desarrolladas por los diferentes fabricantes de dispositivos de Interconectividad, sin embargo a la hora de decidir cual de ellas debe ser aplicada surgen grandes confusiones pues no existen metodologías claras que faciliten la toma de decisiones para lograr una buena implantación.

En este proyecto se presentan las estrategias necesarias para lograr la integración de Voz, Video y Datos en una red Convergente que garantice la calidad de servicio que se requiere para un optimo funcionamiento. En tal sentido durante el desarrollo del trabajo especial de grado se hizo un estudio de las características de la Voz, el Video y la transmisión de Datos para conocer los requerimientos para su correcta operación. Se analizaron los métodos de eficiencia de enlaces para optimizar el funcionamiento de los enlaces digitales de baja velocidad, los métodos de compresión y codificación de Voz y Video, las técnicas para evitar congestión, los métodos de prioritización de trafico, los procedimientos usados en las teorías de manejo de colas y el uso de los servicios diferenciados y servicios integrados para garantizar calidad de servicio en una red Convergente. Al final se estudiaron las tecnologías para la transmisión de Voz, Video y Datos sobre redes de paquetes y como los equipos de Cisco Systems implementan las técnicas disponibles según las estrategias planteadas.

LISTA DE ACRONIMOS

ABR	Available Bit Rate
ACL	Access Control List (Lista de Control de Acceso)
ANSI	American National Standards Institute (Instituto Americano de Normalización)
AQoS	Automatic Quality of Service (Calidad de Servicio Automática)
ARP	Address Resolution Protocol (Protocolo de Resolución de Direcciones)
ATM	Asynchronous Transfer Mode (Modo de Transferencia Asíncrono)
Be	Excess Burst Size (Medida de Rafaga en Exceso)
BECN	Backward Explicit Congestion Notification (Notificación explícita de congestión de vuelta)
Bc	Committed Burst Size
CAC	Control Admission Connection
CBT	Computer Based training (Entrenamiento Basado en Computador)
CBWFQ	Class Based Weighted Fair Queuing
CCS	Channel Common Signaling (Señalización por Canal Común)
CCITT	Consultative Committee on International Telephone and Telegraph (Comité Consultivo Internacional para Telefonía y Telegrafía)
CFI	Canonical Format Identifier
CIP	Classical IP (IP Clásico)
CIR	Committed Information Rate
CoS	Class of Service (Clase de Servicio)
CRTP	Compression RTP
DiffServ	Difference Services
DLCI	Data Link Connection Identifier (Identificador de Conexión del enlace de datos)
FECN	Forward Explicit Congestion Notification (Notificación explícita de congestión hacia adelante)
FIFO	First-In First-Out (Primero en Entrar Primero en Salir)
FR	Frame Relay
FRAD	Frame Relay Access Device (Dispositvo de acceso a redes Frame Relay)
FRF	Forum Frame Relay
FXO	Foreign Exchange Office
FRTS	Frame Relay Traffic Shape
FTP	File Transfer Protocol
FXS	Foreign Exchange Subscriber (IBM) Foreign Exchange Station
GTS	Generic Traffic Shape
GUI	Graphical User Interface (Interfaz gráfica de usuario)
HDLC	High-Level Data Link Control (Protocolo de control de enlace de datos de alto nivel)

IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IntServ	Integrated Services (Servicios Integrados)
IP	Internet Protocol (Protocolo de Internet)
IP/VC	IP Video Conferencing (Video Conferencia sobre IP)
IP/TV	IP Television (Television sobre IP)
IPX	Internet Protocol Exchange
ISDN	Integrated Services Digital Network (Red Digital de Servicios Integrados)
ISO	International Organization for Standardization (Organization Internacional de Estándares)
ITU	International Telecommunications Union (Unión International de Telecomunicaciones)
LAN	Local Area Network (Red de área local)
LAPB	Link Access Procedure Balanced
LFI	Link Fragmentation and Interleaving
LLQ	Low Latency Queuing
LMI	Layer Management Interface
MAC	Media Access Control (Control de Acceso al Medio)
MCR	Minimum Cell Rate
MGCP	Media Gateway Control Protocol (Protocolo de Control de Gateway)
MIB	Management Information Base (Base de información de administración)
MLPPP	Multi Link PPP
MOS	Mean Opinion Score (Puntaje de opinión media)
NBAR	Network Based Application Recognition (Reconocimiento de Aplicaciones Basadas en Red)
NETBIOS	Network Basic Input Output System (Sistema Básico de entrada salida de red)
OSI	Open System Interconnection (Interconexión de sistemas abiertos)
OSPF	Open Short Path First (Primer camino abierto más corto)
PABX	Private Automated Branch Exchange
PBR	Policy Based Routing (Enrutamiento Basado en Políticas)
PBX	Public Branch Exchange
PCM	Pulse Code Modulation (Modulación por Pulsos Codificados)
PCR	Per Cell Rate
PDU	Protocol Data Unit (Unidad de datos del Protocolo)
PHB	Per Hop Behavior (Comportamiento por Salto)
PSTN	Public Switched Telephone Network
PVC	Permanent Virtual Circuit (Circuito Virtual Permanente)
QDM	Quality Device Manager (Gestor de Dispositivos de Calidad)
QoS	Quality of Service (Calidad de Servicio)
QPM	Quality Policy Manager (Gestor de Políticas de Calidad)
RED	Random Early Detection

RFC	Request For Comments (Petición de comentarios)
RMON	Remote Network Monitor (Monitoreo remoto de red)
RSVP	Resource Reservation Protocol (Protocolo de Reservación de Recursos)
RTP	Real Time Protocol (Protocolo de Tiempo Real)
SAA	Service Assurance Agent
SAN	Store Area Network
SIP	Session Initiation Protocol (Protocolo de Iniciación de Sesión)
SLA	Service Level Agreement (Acuerdo de Nivel de Servicio)
SNA	System Network Architecture (Arquitectura de sistemas en red)
SNMP	Simple Network Management Protocol (Protocolo Simple de Administración de Red)
SRB	Source Routing Bridging
STP	Shielded Twisted Pair Cable
SVC	Switched Virtual Circuit (Circuito Virtual Conmutado)
TCP	Transmission Control Protocol (Protocolo de control de transmisión)
TDM	Time Division Multiplexing (Multiplexación por division en el tiempo)
ToS	Types of Service (Tipos de Servicio)
UDP	User Datagram Protocol (Protocolo de datagramas de usuario)
UIT	Union Internacional de Telecomunicaciones
UNI	User Network Interface (Interfaz usuario-red)
UTP	Unshielded Twisted Pair Cable
VC	Virtual Channel (Canal virtual) Virtual Circuit (Circuito virtual)
VCC	Virtual Channel Connections (Conexión de canal virtual)
VFRAD	Voice Frame Relay Access Device
VLAN	Virtual Local Area Network (Red de área local virtual)
VoATM	Voice Over ATM (Voz sobre ATM)
VoFR	Voice over Frame Relay (Voz sobre Frame Relay)
VoIP	Voice over IP (Voz sobre IP)
VPC	Virtual Path Connection (Conexión de camino virtual)
WAN	Wide Area Network (Red de área amplia)
WFQ	Weighted Fair Queuing
WRED	Weighted Random Early Detection

INDICE GENERAL

INTRODUCCIÓN

Los sistemas de telecomunicaciones tradicionalmente han transmitido Voz, Video y Datos mediante redes paralelas, en la mayoría de los casos existe una red para servicios de Video, una red para servicios de Voz y otra red para la transmisión de Datos. Esta forma de transportar información trae como consecuencia la necesidad de gestionar tres redes, requiriéndose personal especializado para la administración y monitoreo de cada red y cubrir los costos asociados al mantenimiento y operatividad de los servicios, a estos factores se debe adicionar los costos de infraestructura y pagos a proveedores, por ejemplo, pagos por servicios de telefonía transportados por redes publicas para permitir la comunicación telefónica entre sucursales de una misma empresa.

El uso del Video como medio para la difusión de eventos corporativos, información de interés para los miembros de una empresa, establecimiento de videoconferencias, telemedicina y actividades del tipo e-Learning o CBTs, están indicando cada día con mayor fuerza que es necesario la implementación de sistemas de telecomunicaciones para la transmisión de Video en la red corporativa, lo cual requeriría de una red aparte para su transporte si se usa el esquema tradicional.

Otro de los servicios que ha tenido una gran importancia y un gran crecimiento es la transmisión de Datos mediante el protocolo IP, caracterizado por el envío y la recepción de información entre localidades remotas mediante redes basadas en conmutación de paquetes con servicios de mejor esfuerzo.

Es importante señalar que cada uno de estos servicios es de naturaleza distinta y por lo tanto es lógico pensar que deben ser tratados de forma distinta. En tal sentido se hacen grandes esfuerzos para que estas tres redes (publicas y/o

privadas) cumplan con los requerimientos de calidad, disponibilidad y alcanzabilidad que se requiere para un optimo funcionamiento.

El planteamiento anterior conduce a pensar que si se lograra la adecuación de una red que cumpla con los requerimientos de calidad de servicio, ancho de banda, eficiencia de enlaces, retardo controlado y redundancia, se podría tener una red totalmente integrada para el transporte de los servicios de información que tradicionalmente se transportan en tres redes, lo cual permitiría un control centralizado de los servicios, un ahorro sustancial en los costos por concepto de telefonía publica, una reducción de costos por entrenamiento, costos de propiedad, numero de dispositivos de telecomunicaciones, la estandarización de las tecnologías usadas en la red, en fin, una optimización de los recursos y un aprovechamiento de la infraestructura usada para transportar servicios por conmutación de paquetes, la cual mediante estrategias de QoS podría funcionar como una red integrada para la transmisión de Voz, Video y Datos.

Los avances en las tecnologías de información y las telecomunicaciones, el gran crecimiento y la popularidad de los protocolos TCP/IP y las técnicas de calidad de servicio, permiten integrar Voz, Video y Datos en una red convergente que contribuya con la optimización de recursos y con un mejor uso de las tecnologías actuales y futuras para proveer servicios de información de calidad y con alta disponibilidad.

Las técnicas de calidad de servicios, los mecanismos para evitar congestión, los métodos para dar eficiencia a enlaces de baja velocidad, los esfuerzos de los proveedores de servicio para cumplir con los SLAs y las aplicaciones para el monitoreo del optimo funcionamiento de la QoS, permiten implementar redes convergentes siempre que se sigan estrategias que utilicen las tecnologías de una manera acertada.

En este trabajo se presentan las estrategias necesarias para lograr la integración de Voz, Video y Datos en una red convergente que garantice la calidad de servicio necesaria. En tal sentido durante el desarrollo del trabajo especial de grado se muestra un estudio de las características de la Voz, el Video y la transmisión de Datos, para conocer los requerimientos que garantizaran su optimo funcionamiento. Se analizan los métodos de eficiencia de enlaces para mejorar el desempeño de los enlaces digitales de baja velocidad, los métodos de compresión y codificación de Voz y Video, las técnicas para evitar congestión, los métodos de prioritización de trafico, los procedimientos usados en las teorías de manejo de colas y el uso de los servicios diferenciados y servicios integrados para garantizar calidad de servicio en una red convergente. Se estudiaron las tecnologías para la transmisión de Voz, Video y Datos sobre redes de paquetes, los equipos de Cisco Systems para mostrar como implementan las técnicas disponibles y como deben ser usados según las estrategias planteadas.

OBJETIVOS

Objetivo General

Presentar estrategias que permitan garantizar calidad de servicio en la integración de voz, video y datos en redes corporativas con tecnologías de Cisco Systems.

Objetivos Específicos

- Estudiar las características de la voz, el video y los datos.
- Analizar los métodos de compresión y codificación de voz y video.
- Analizar los métodos de eficiencia de enlaces para optimizar el funcionamiento de los enlaces digitales de baja velocidad.
- Presentar las técnicas para evitar congestión, los métodos de prioritización de trafico y los procedimientos usados en las teorías de manejo de colas.
- Uso de los servicios diferenciados y servicios integrados para garantizar calidad de servicio en una red convergente.
- Estudiar las tecnologías más usadas para la transmisión de voz y video sobre redes de paquetes.
- Estudiar los equipos de Cisco Systems para conocer como implementan las técnicas disponibles de QoS.
- Presentar estrategias para optimizar una red en convergencia o para implementarla desde su inicio.

INTRODUCCIÓN A QoS BASADA EN IP

1.1. – Que es Calidad de Servicio basada en IP

El conjunto de protocolos del TCP/IP fue diseñado originalmente para manejar aplicaciones de correo de texto, transferencia de archivos y telnet, bajo la filosofía del mejor esfuerzo. Pero las redes de la actualidad están generando demandas cada vez mayores de ancho de banda, lo que produce fuertes retrasos que impactan en la productividad de las empresas y producen descontento en los usuarios. Servicios como Voz y Video necesitan tiempos de espera predecibles para evitar el fenómeno de saltos en imagen o interrupciones de sonido.

Una red IP es un elemento importante en el diseño de una red empresarial con características convergentes. En muchos casos, la red IP se encuentra en funcionamiento, sin embargo es necesario un apropiado acondicionamiento para soportar de manera efectiva el trafico de Voz, Video y Datos de forma integrada.

En otros casos, la red IP es diseñada e implementada desde el comienzo para soportar trafico convergente.

Las redes IP poseen muchas características de funcionamiento, entre las cuales se encuentran algunas que deben ser controladas pues determinan la calidad de servicio. Los siguientes son elementos de las redes IP que deben ser controlados:

- Delay.
- Jitter.
- Pérdida de paquetes.
- Ancho de Banda.

1.1.1.- Delay

El delay de extremo a extremo también llamado delay fijo, es causado por los retrasos en la conmutación, propagación y serialización en la red. El delay fijo es una característica de la red que debe ser mantenida en los niveles correctos. El valor de 150 mseg. es comúnmente usado como el delay deseado en una vía. El usuario percibe una degradación de la calidad de la llamada si el umbral de 150 mseg. es excedido. Este valor esta definido en la recomendación de la UIT G.114.

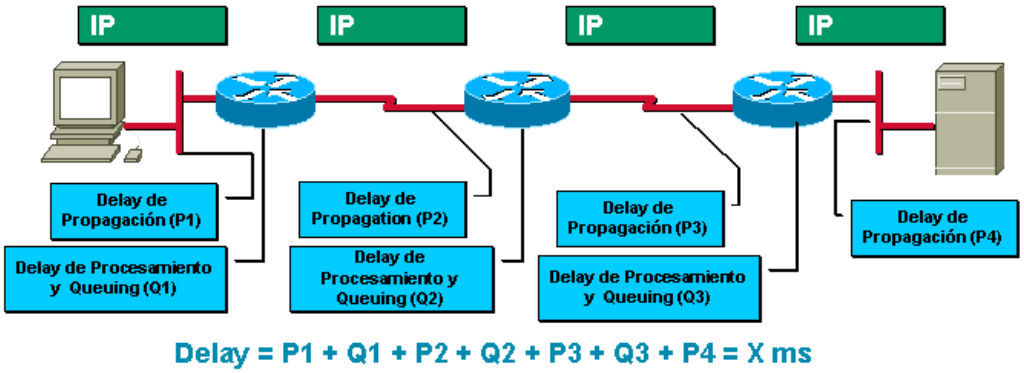


Figura 1. Calculo del Delay.

El delay de extremo a extremo es la suma del delay de propagación, delay de procesamiento y delay de encolamiento. El delay de propagación es fijo, mientras que el delay de procesamiento y de queuing es impredecible en redes con servicios de mejor esfuerzo.

El delay de procesamiento, es el tiempo que tarda un dispositivo de comunicaciones en tomar un paquete de la interfaz de entrada y colocarlo en la cola de la interfaz de salida.

El delay de encolamiento (Queing Delay), es el tiempo que dura un paquete en la cola de la interfaz de salida de un dispositivo de comunicaciones.

El delay de propagación o de serialización, es el tiempo que se toma un dispositivo de comunicaciones en transmitir un paquete.

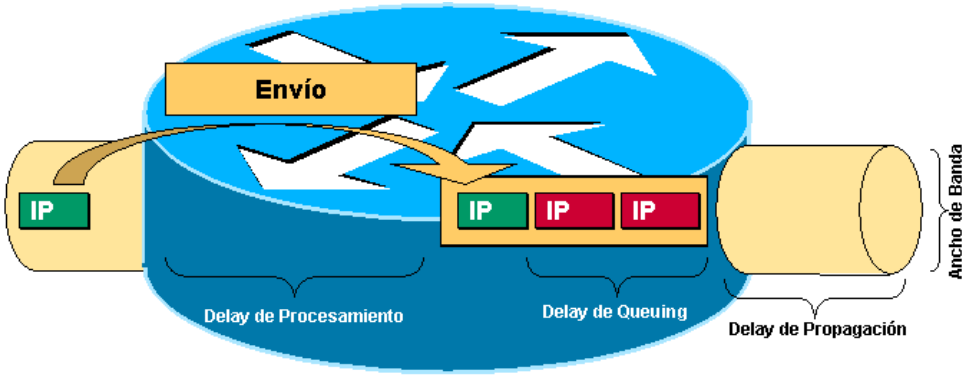


Figura 2. Representación del Delay Total.

1.1.2.- Jitter

El Jitter es la variación del retardo y la latencia, que tiene un impacto significativo en la Calidad de Servicio para Voz y Video. El Jitter tiene muchas causas tales como:

- Variación en el tamaño de la cola.
- Variación en el tiempo de procesamiento necesario para reordenar los paquetes que lleguen fuera de orden.
- Variación en el tiempo de procesamiento necesario para reensamblar los paquetes que fueron segmentados en la fuente antes de su transmisión.

1.1.3.- Pérdida de Paquetes

Los dispositivos de interconectividad, tales como los routers y switches, ponen los paquetes de datos en un buffer, donde esperan en cola para ser transmitidos por un enlace. Si la ruta está muy congestionada, el buffer se desborda y se perderán los datos. Los paquetes perdidos son a menudo retransmitidos, si por ejemplo se usa TCP en vez de UDP, aumentando así la demora total.

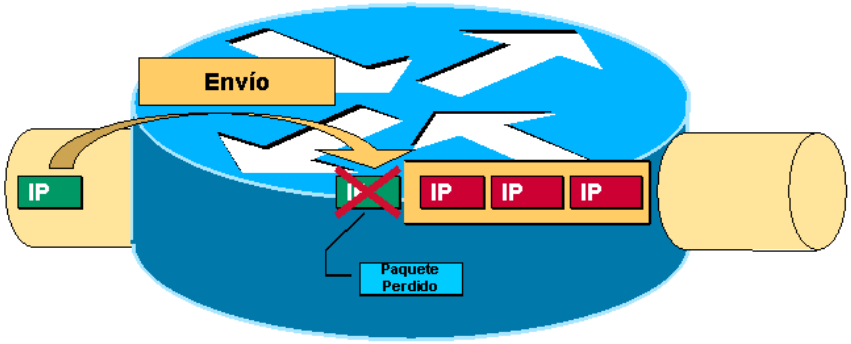


Figura 3. Pérdida de Paquetes.

La pérdida de paquetes ocurre cuando la cola de salida está llena. Estas son las pérdidas más comunes cuando un enlace esta congestionado.

También hay otros tipos de pérdidas de paquetes (pérdidas en la cola de entrada, paquetes ignorados, paquetes fuera del limite, buffer no disponible, etc.). Estas pérdidas de paquetes son usualmente un resultado de la congestión de un equipo de comunicaciones.

1.1.4.- Ancho de Banda

El término ancho de banda, aplicado a un medio de transmisión, se refiere a la gama de frecuencias que el canal transmite con atenuación constante. Aplicado a

una señal, este termino representa la gama de frecuencias ocupadas por el espectro de frecuencia de la señal.

En el mundo digital, el término ancho de banda se usa a menudo como tasa de transmisión, lo cual es formalmente incorrecto.

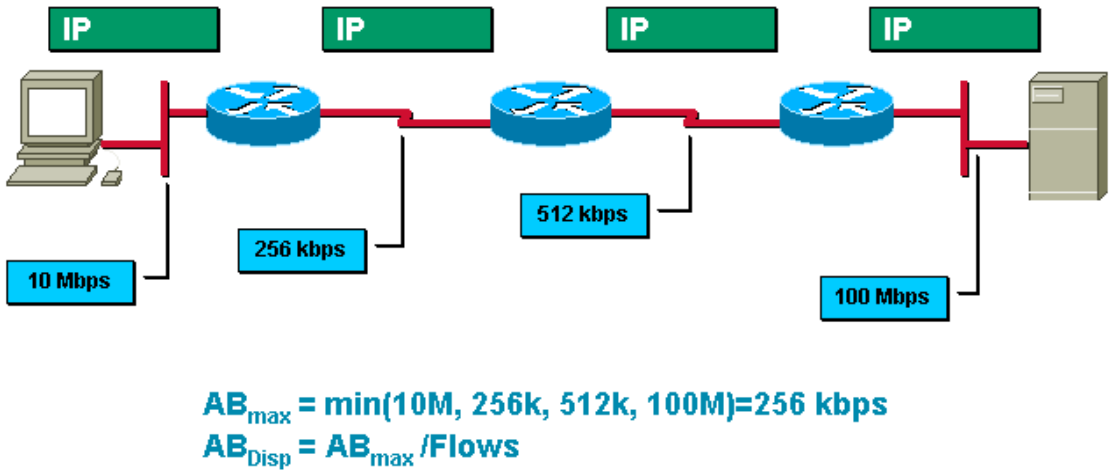


Figura 4. Ancho de Banda.

El ancho de banda máximo en una ruta de transmisión es el menor ancho de banda en la ruta. Múltiples flujos compiten por el mismo ancho de banda, resultando en un menor ancho de banda disponible para una aplicación.

1.1.5. - Definición de Calidad de Servicio

La Calidad de Servicio trata diferentes tipos de trafico de forma diferente, dependiendo del tipo y de las necesidades del servicios. La QoS es la habilidad de la red para proveer diferentes servicios a diferentes aplicaciones. La QoS se enfoca en las herramientas para permitir a la Voz, el Video y los Datos ser tratados inteligentemente a medida que se transmiten sobre la red.

En términos generales la calidad de servicio (QoS, Quality of Service) de una red se refiere a su capacidad para transportar la información desde la fuente al destino de forma rápida y confiable.

La meta principal de la calidad de servicio es proveer ancho de banda dedicado, Jitter y delay controlado, y reducción de las pérdidas de paquetes.

Una red tradicional trata el trafico como de mejor esfuerzo, el envío del trafico por los dispositivos de red es FIFO. La calidad de servicio prioritiza el trafico en diferentes niveles de servicio y provee tratamiento preferencial al envío de cierto tipo de trafico a expensas del trafico de menor prioridad.

1.2. – Por que es necesario usar Calidad de Servicio

Las redes empresariales de hoy están constituidas por múltiples localidades, granjas de servidores, acceso remoto y usuarios que trabajan desde sus casas. El uso incrementado de la red para el transporte de Datos críticos, Video y Voz sobre una red IP, requieren una más alta calidad de servicio que la requerida para redes de transmisión de Datos simples.

Las estadísticas de diferentes firmas especializadas, indican que el trafico Wan corporativo entre el año 1998 y 2002 creció 600 % aproximadamente. Con ese crecimiento, los costos de operación están proyectados para incrementarse por 250 % a escala mundial.

Hasta ahora se ha direccionado el manejo de diversos servicios de información con redes paralelas: una para Voz, una para Datos y una para Video, generando mayores gastos y requiriendo de personal adicional.

La integración de múltiples servicios permite crear una poderosa red que es mucho más fácil y menos costosa para gestionar y operar. La consolidación de Datos, Voz y Video en una red convergente permite no solo un control de costos, sino también, ofrecer a la compañía una nueva forma para hacer negocios y una mejor posición para competir. Con una red multiservicios, una empresa puede desarrollar e implementar un gran numero de nuevas aplicaciones nunca antes posible con redes separadas.

La QoS provee la habilidad de gestionar una red convergente de Voz, Video y Datos. Cada aplicación requiere diferentes servicios de la red y se comporta de manera distinta.

Las aplicaciones críticas para el negocio tal como Oracle, SAP y PeopleSoft, se comportan erráticamente si el tiempo de respuesta ida y vuelta excede cierto umbral. En ese punto la aplicación va en un modo de retrasmisión excesiva que afecta el flujo de Datos sobre la red. Por otro lado, esas aplicaciones de planeación de recursos (ERP), son sensibles no solo al tiempo de round-trip promedio, sino también a picos de velocidad. Las aplicaciones ERP, deben recibir su información con un mínimo de retardo para poder realizar su trabajo de forma eficiente.

La Voz requiere servicios de la red con bajo retardo, mínimo Jitter y poca pérdida de paquetes. El trafico de Voz no requiere en si mismo mucho ancho de banda, ya que por ejemplo una conversación puede ser comprimida a 8 kbps; pero el retardo y la pérdida de paquetes afectan la calidad de la Voz.

El trafico multimedia también debe arribar a tiempo y en orden. Al igual que la Voz, el Video requiere baja latencia y mínimo Jitter, además debe ser protegida de ráfagas de trafico de datos.

La tecnología de QoS es un camino viable para trabajar con un ambiente de recursos limitados y para acondicionar la red para ofrecer a cada aplicación los recursos que necesita para desempeñarse apropiadamente.

1.3. – Beneficios de la Calidad de Servicio

Las aplicaciones de tiempo real tales como aplicaciones de Voz y Video tienen diferentes características y requerimientos que las aplicaciones tradicionales de Datos. Debido a que son basadas en tiempo real, toleran variaciones mínimas en la cantidad de delay en la entrega de los paquetes. Ese trafico es también intolerante a la pérdida de paquetes y al Jitter, ambos degradan la calidad de la transmisión de la Voz y el Video entregada a un usuario final. Para transportar efectivamente el trafico de Voz y Video sobre IP, requieren de mecanismos que aseguren la entrega de los paquetes con baja latencia y sin pérdidas. Las características de la calidad de servicio incorporan colectivamente las técnicas necesarias para lograr esos requerimientos, ofreciendo lo principal para proveer prioritización de servicios que reúnen los estrictos requerimientos para la entrega de paquetes de Voz, Video y Datos de mission-critical.

1.4. – Evolución de la Calidad de Servicio

La calidad de servicio ha evolucionado desde los servicios de mejor esfuerzos donde la red no garantiza ancho de banda ni controla el retardo, hasta la implementación de QoS de forma inteligente y automatizada (AQoS). Esta evolución ha permitido el desarrollo de aplicaciones en red mucho más amigables y de mejor calidad.

Con la implementación del modelo de servicios integrados, la red realiza la reservación del ancho de banda y la verificación del retardo antes de que el flujo de información sea transmitido; con los servicios diferenciados el trafico es

clasificado y marcado para obtener un tratamiento diferente según las políticas empresariales y las características del trafico.

Cualquiera que sea el modelo de servicio utilizado, siempre es necesario la configuración manual de este, lo cual en muchos casos resulta sumamente complicado y se pone en riesgo el optimo funcionamiento de los servicios ofrecidos por aplicaciones mission – critical.

En la actualidad la QoS está en una etapa donde la configuración ya no se hace de forma manual sino que los equipos de comunicaciones tienen ciertas capacidades para automatizar algunos procesos involucrados en la implementación de calidad de servicio. En la figura # 5 se muestra la evolución de la QoS desde los años 90 y su tendencia actual.

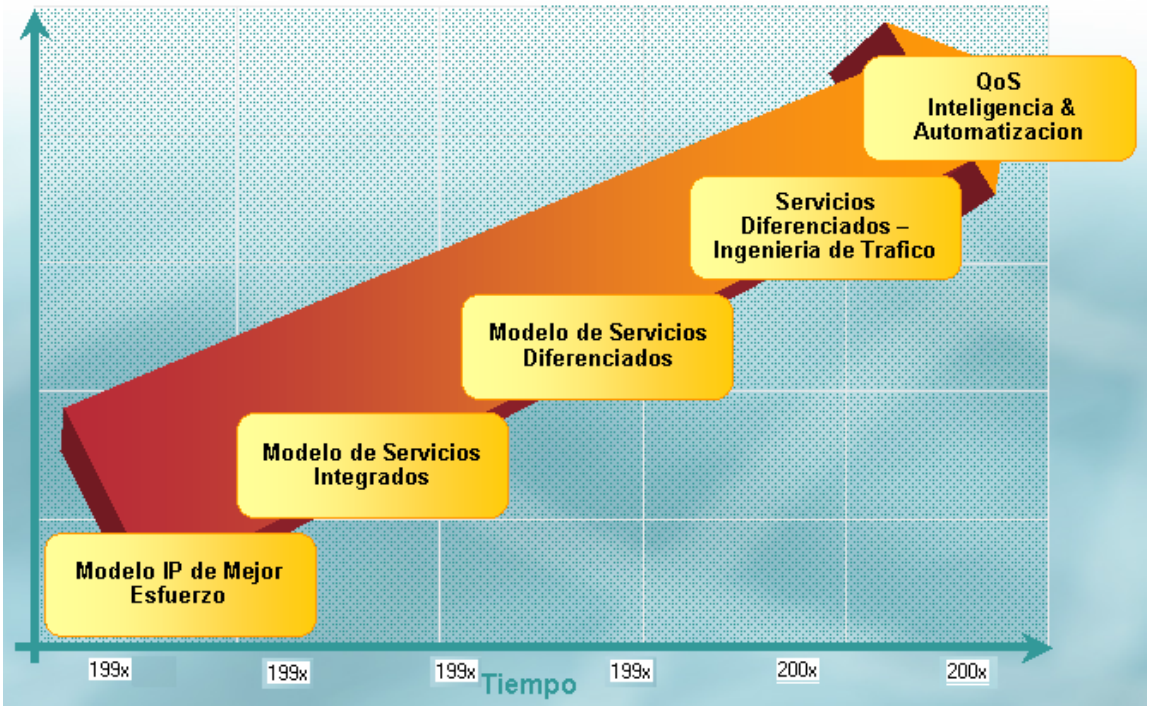


Figura 5. Evolución de la Calidad de Servicio.

II. TIPOS DE CALIDAD DE SERVICIO

La calidad de servicio de extremo a extremo es la habilidad de la red para entregar los servicios requeridos por tráficos específicos de red desde un extremo a otro.

Las redes empresariales están típicamente enfocadas en proveer QoS a las aplicaciones críticas para cumplir con sus requerimientos de funcionamiento. En la siguiente tabla se muestra un ejemplo de los requerimientos de algunas aplicaciones en términos de ancho de banda, delay, perdidas de paquetes y Jitter.

	Ancho de Banda	Delay	Perdidas	Jitter
Interactivas (Telnet)	Bajo	Bajo	Bajo	No es Importante
Tx en lote (FTP)	Alto	No es Importante	Bajo	No es Importante
Fragil (SNA)	Bajo	Bajo	Ninguna	No es Importante
Voz	Bajo	Bajo y Predecible	Bajo	Bajo
Video	Alto	Bajo y Predecible	Bajo	Bajo

Tabla 1. Requerimientos de QoS para Aplicaciones comunes.

Estos requerimientos son satisfechos mediante diferentes tipos de servicios, los cuales se encargan de proveer la calidad de servicio necesaria para la operación de las aplicaciones.

Generalmente estos servicios son clasificados en tres modelos: Servicios de Mejor Esfuerzo, Servicios Integrados y Servicios Diferenciados.

2.1. – Servicios de Mejor Esfuerzo

El Mejor Esfuerzo es un modelo de servicio sencillo en el cual una aplicación envía datos cuando quiera, en cualquier cantidad y sin ningún requerimiento para un tratamiento especial en la red. Para servicios de Mejor Esfuerzo, la red entrega datos si puede, sin ningún aseguramiento de la fiabilidad, delay o ancho de banda.

La implementación de QoS de Mejor Esfuerzo usa la técnica de Queing First-in, First out (FIFO).

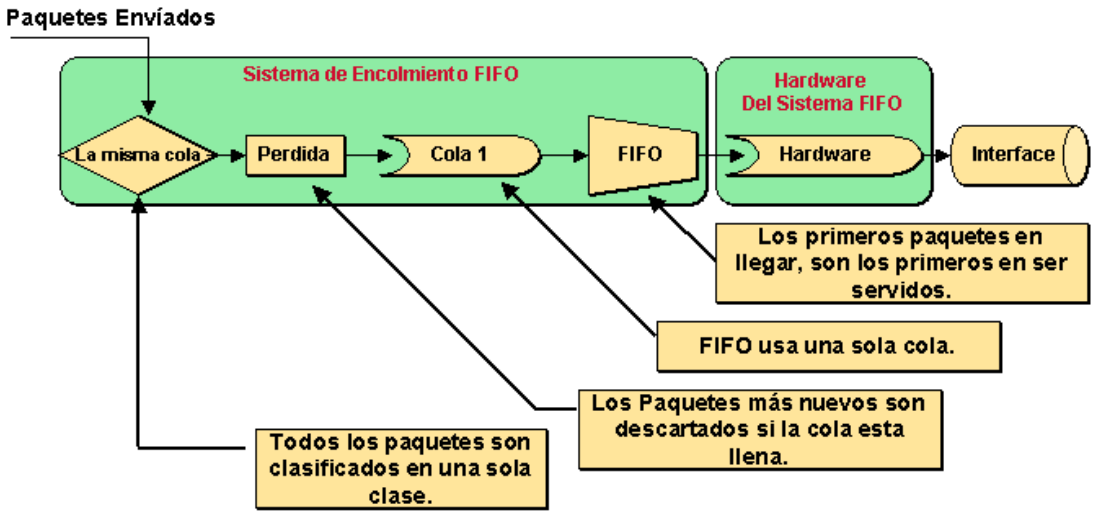


Figura 6. Sistema de Encolamiento FIFO.

Los servicios de Mejor Esfuerzo son utilizados para un amplio rango de aplicaciones en red tales como la transferencia de archivos o el correo electrónico.

2.2. – Servicios Integrados

Los Servicios Integrados (IntServ), es un modelo de servicios múltiples que puede organizar múltiples requerimientos de Calidad de Servicio. En este modelo las aplicaciones requieren una clase de servicio especifica de la red antes de enviar la información. El requerimiento es realizado a través de una señalización explicita; la aplicación le informa a la red su perfil de trafico y solicita una clase particular de servicio que puede satisfacer sus requerimientos de ancho de banda y delay. La aplicación solo enviará la información luego de obtener una confirmación de la red.

Los Servicios Integrados (RFC 1633), son un modelo inherentemente orientado a conexión, cada comunicación individual debe especificar explícitamente a la red su descripción de trafico así como los recursos requeridos. El equipo de comunicaciones final desempeña control de admisión para asegurar que los recursos son suficientes en la red. El modelo de IntServ asume que los equipos de comunicaciones a lo largo de la ruta configuran y mantienen el estado para cada comunicación individual.

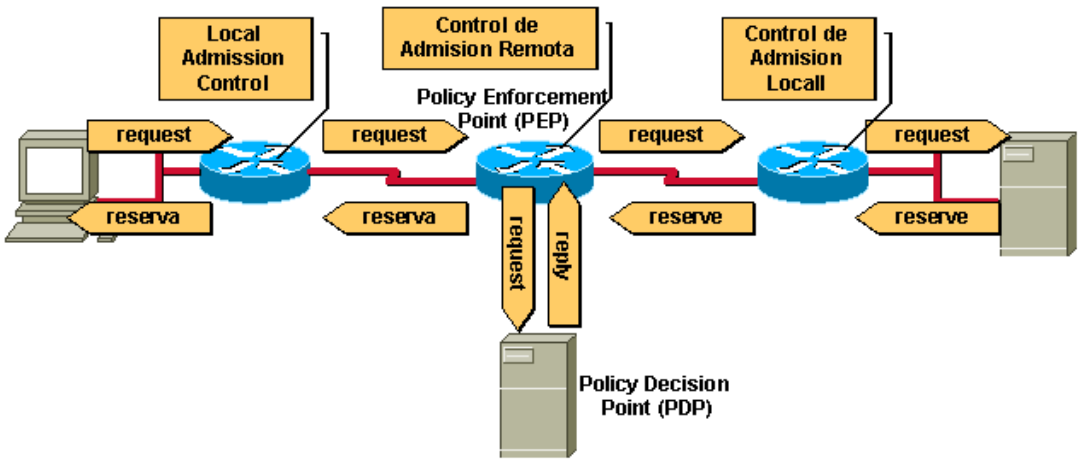


Figura 7. Modelo de IntServ en Bloques.

El rol del protocolo de reservación de recurso (RSVP, Resource Reservation Protocol) es proveer el control de admisión de recursos en redes convergentes. Si los recursos están disponibles, RSVP acepta una reservación e instala un clasificador de trafico en la tabla de envío de QoS. El clasificador de trafico le informa a la ruta de envío de QoS como clasificar los paquetes de un flujo particular y que tratamiento de envío proveer. La instalación de un clasificador de trafico y tratamiento de flujo es la interfaz entre RSVP y el modelo de servicios diferenciados DiffServ. El RSVP es una característica de planeación de control que limita la carga de trafico convergente aceptada que la red puede soportar.

La red desempeña control de admisión basado en la información de la aplicación y los recursos de red disponibles. La red se compromete a reunir los requerimientos de QoS de la aplicación siempre que estos se mantengan dentro de las especificaciones del perfil. La red garantiza el mantenimiento del estado por flujos, desempeñado clasificación de paquetes, policing y encolamiento inteligente basado en el estado de cada flujo.

La QoS utilizada por ejemplo, de equipos de comunicaciones de Cisco Systems, incluyen las siguientes características que proveen servicios de carga controlada, la cual es una clase de Servicio Integrado:

- RSVP. Puede ser utilizado por aplicaciones para indicar sus requerimientos de calidad de servicio a la red.
- Mecanismos de encolamiento inteligente. Son utilizados con RSVP para proveer servicios de velocidad garantizada y servicios de carga controlada.
- Los servicios de Velocidad Garantizada, permiten a las aplicaciones reservar ancho de banda para satisfacer sus requerimientos. Por ejemplo, una aplicación de Voz sobre IP (VoIP) puede reservar 32 Kbps de extremo a extremo usando esta clase de servicio.

- Servicios de Carga Controlada, permiten a las aplicaciones tener bajo delay y alto throughput durante los momentos de congestión. Por ejemplo, aplicaciones de tiempo real adaptativas tal como la reversa de una conferencia grabada pueden utilizar estos servicios.

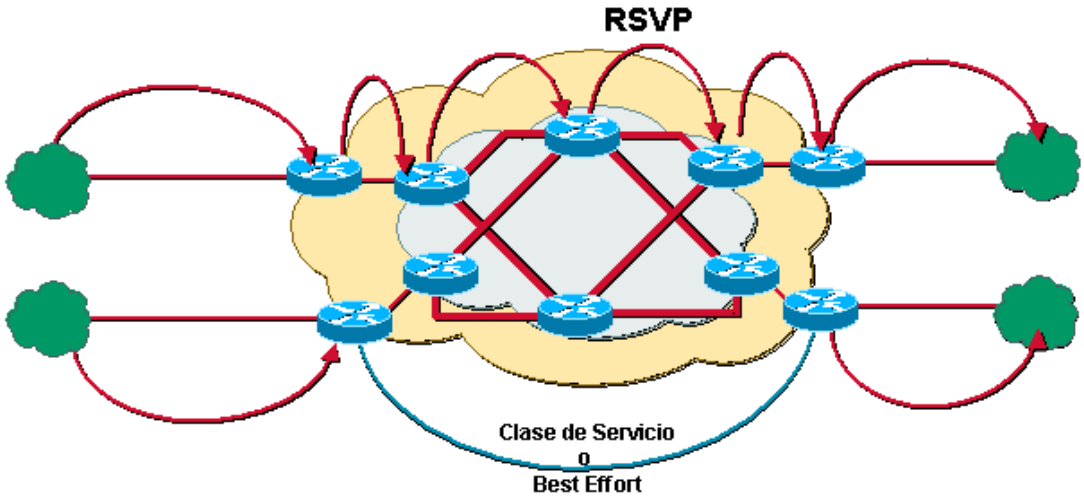


Figura 8. Opciones de Implementación de RSVP.

El modelo de Servicios Integrados tiene varias formas de implementación; a continuación se presentan tres opciones:

1. RSVP Explicito en cada nodo de la red.
2. RSVP 'pass-through' y transporte de CoS.
   - Mapear RSVP a CoS al final de la red.
   - Requerimientos de RSVP pass-through para salir.
3. RSVP al final de la red y 'pass-through'con envío de best-effort en el núcleo de la red, si hay ancho de banda suficiente en el núcleo.

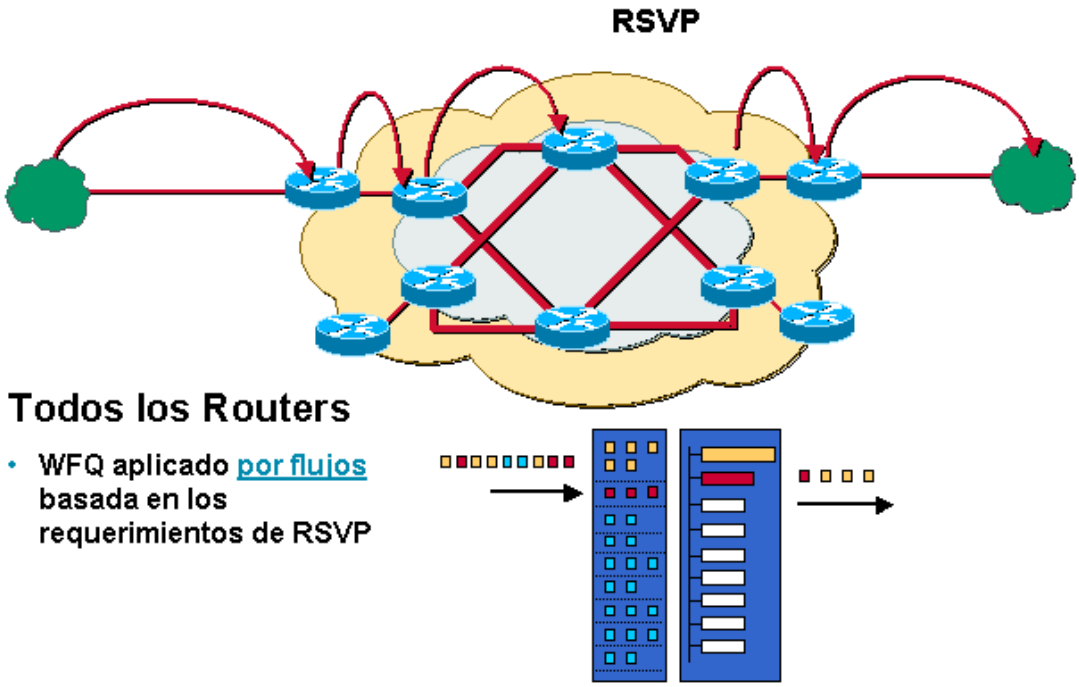


Figura 9. RSVP Explicito.

Los requerimientos explícitos de RSVP permiten el transporte de los servicios integrados de extremo a extremo. En este modelo los routers implementan la técnica WFQ por flujos basada en los requerimientos de RSVP.

Esto permite que cada aplicación obtenga la QoS necesaria mediante IntServ según las solicitudes que el protocolo de reservación de recursos solicita a todos los routers de la red, permitiendo una QoS desde el origen hasta el final de cada flujo de información asociado a las aplicaciones presentes en la red. Los equipos de comunicaciones como los routers usan WFQ como técnica de encolamiento.

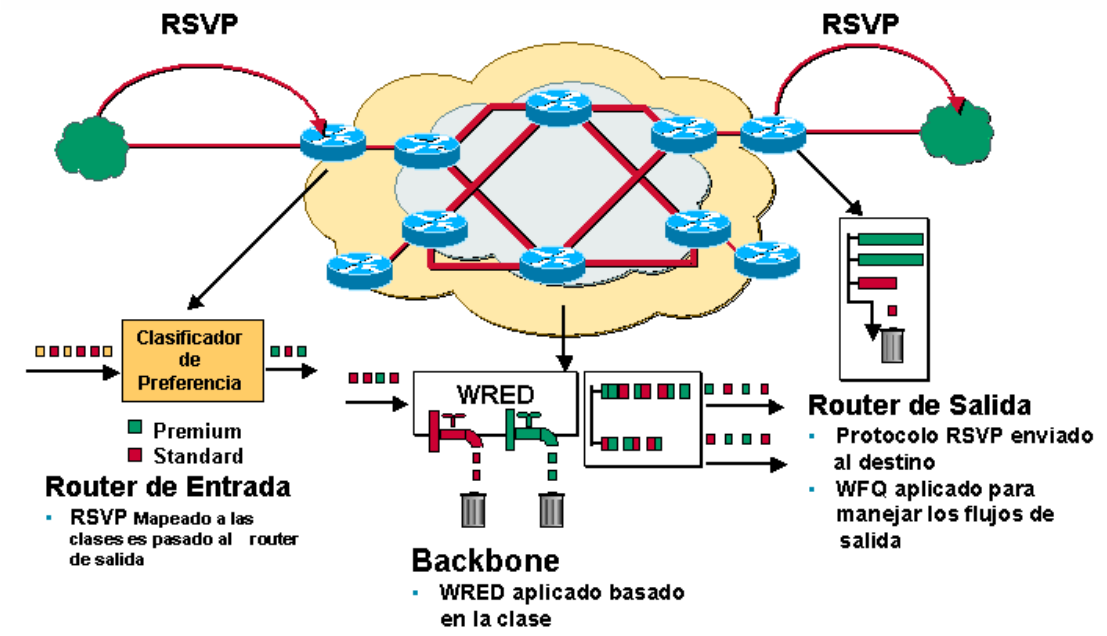


Figura 10. RSVP Pass-Through.

En el RSVP pass-through, los routers pasan los requerimientos de las aplicaciones al router de salida, en el núcleo de la red los routers aplican WRED basado en la clase de servicio configurada. En el router de salida el RSVP es enviado al destino y se aplica WFQ para manejar los diferentes flujos de información.

El protocolo de Reservación de Recursos (RSVP), es una característica que permite a los sistemas finales o hosts en cualquier lado de un router de red, establecer una ruta con ancho de banda reservado entre ellos para predeterminar y asegurar la QoS para sus transmisiones de información.

Actualmente RSVP tiene el único protocolo de señalización estándar diseñado para garantizar ancho de banda de red de extremo a extremo para redes IP.

RSVP es un protocolo estándar de Internet de la IETF (RFC 2205) para permitir a una aplicación reservar dinámicamente ancho de banda en la red. RSVP habilita a las aplicaciones para requerir una QoS especifica para un flujo de información.

Hosts y routers usan RSVP para entregar requerimientos de QoS a los routers alrededor de las rutas del flujo y para mantener el estado de routers y hosts para que provean el servicio requerido, usualmente ancho de banda y latencia. RSVP usa una velocidad de datos principal, la más larga cantidad de datos que el router mantendrá en cola, y la mínima QoS para determinar la reservación de ancho de banda.

Weighted Fair Queuing (WFQ) o WRED actúan como el caballo de batalla para RSVP, configurando la clasificación de paquetes y la programación requerida para flujos reservados. Usando WFQ, RSVP puede entregar un servicio garantizado de IntServ. Usando WRED, RSVP puede entregar un servicio de carga controlada.

2.3. – Servicios Diferenciados

Los servicios diferenciados son un modelo de servicio múltiple que puede satisfacer diferentes requerimientos de QoS. El modelo de servicio diferenciado describe los servicios con clases de trafico. Para Servicios Diferenciados, la red trata de entregar una clase particular de servicio basado en la QoS especificada para cada paquete. Esta especificación puede ocurrir en diferentes formas, por ejemplo, usando el bit de precedencia IP configurado en paquetes IP de la dirección IP fuente o destino. La red usa la especificación de QoS para clasificar, shape, aplicar políticas de trafico y desempeñar un encolamiento inteligente.

Figura 11. Uso del Campo ToS de IP en DiffServ.

El modelo de servicios diferenciados es usado para diversas aplicaciones de misión-critical y para proveer calidad de servicio de extremo a extremo.

Este modelo de servicio es implementado a través del campo ToS de la cabecera IP. Este campo llamado DSCP utiliza un código de 6 bits denominado DS como se muestra en la figura 11.

Un flujo es definido como una corriente de datos individual y unidireccional entre dos aplicaciones, y es únicamente indefinido por la dirección IP fuente, numero de puerto, dirección IP destino, numero de puerto destino y el protocolo de transporte.

La arquitectura de servicios diferenciados está basada en un modelo sencillo donde el trafico entrante a una red es clasificado y posiblemente condicionado en el borde de la red. La clase de trafico es entonces identificada con un código DS en la cabecera del paquete IP. En el núcleo de la red, los paquetes son enviados de acuerdo al comportamiento por salto (per-hop behavior) asociado con el punto de código DS.

Uno de los principios primarios del modelo de servicios diferenciados es que se puede marcar los paquetes lo mas cerca posible del borde de la red. Es bastante difícil y es una tarea que consume mucho tiempo entender a cual clase de trafico pertenece un paquete, por lo cual se puede clasificar la información en el menor tiempo posible. Al marcar el trafico en el borde de la red, los dispositivos del núcleo y otros dispositivos de la ruta del trafico podrán determinar rápidamente la clase de servicio (CoS) apropiada a ser aplicada a un determinado flujo de trafico.

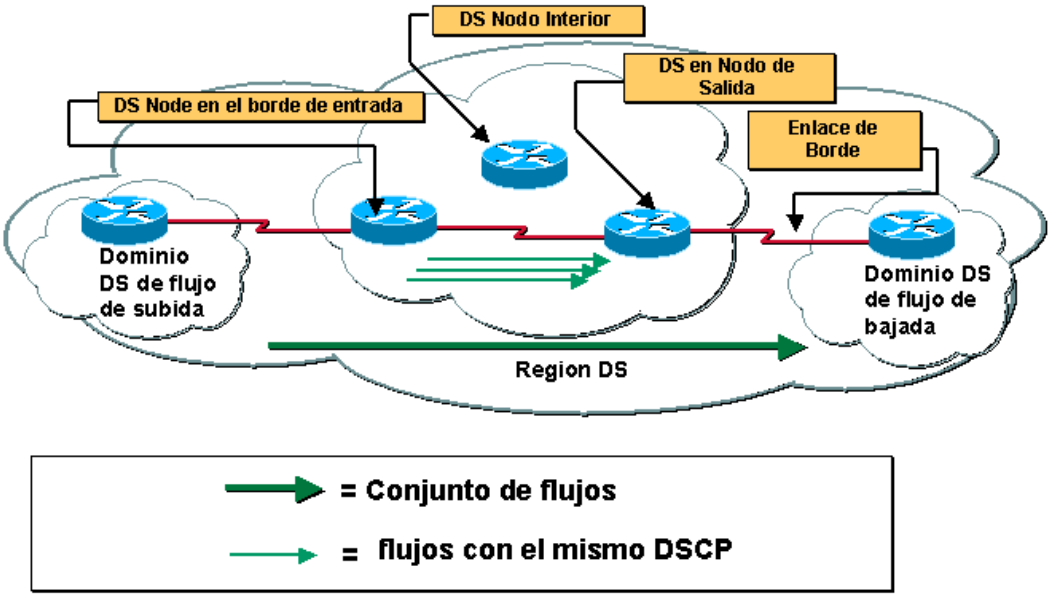


Figura 12. Terminología en Topología de DiffServ.

Un domino de DS consiste en nodos en el borde y en el interior de la red implementando el modelo de servicios diferenciados. Los nodos que implementan DS (DiffServ) en el borde de la red interconectan dominios DS u otros dominios no DS. Los nodos que implementan códigos DS solo interconectan a otros nodos interiores que también usan DS o a otros nodos DS dentro del mismo dominio. Ambos nodos en el borde y en el interior deben ser capaces de aplicar el PHB apropiado a los paquetes basados en el punto de código DS. De otra forma ocurrirá un comportamiento impredecible.

Los nodos que implementando códigos DS en los bordes de la red actúan como nodos de entrada y de salida para diferentes direcciones de trafico. El tráfico entra a un dominio de DS en el nodo de entrada y abandona el dominio DS en el nodo de salida. Una región de servicios diferenciados es un conjunto de uno o más dominios DS continuos. Las regiones DS son capaces de soportar servicios diferenciados en la ruta la cual se expande en el dominio DS. En la figura 13 se muestra una arquitectura de Servicios Diferenciados.

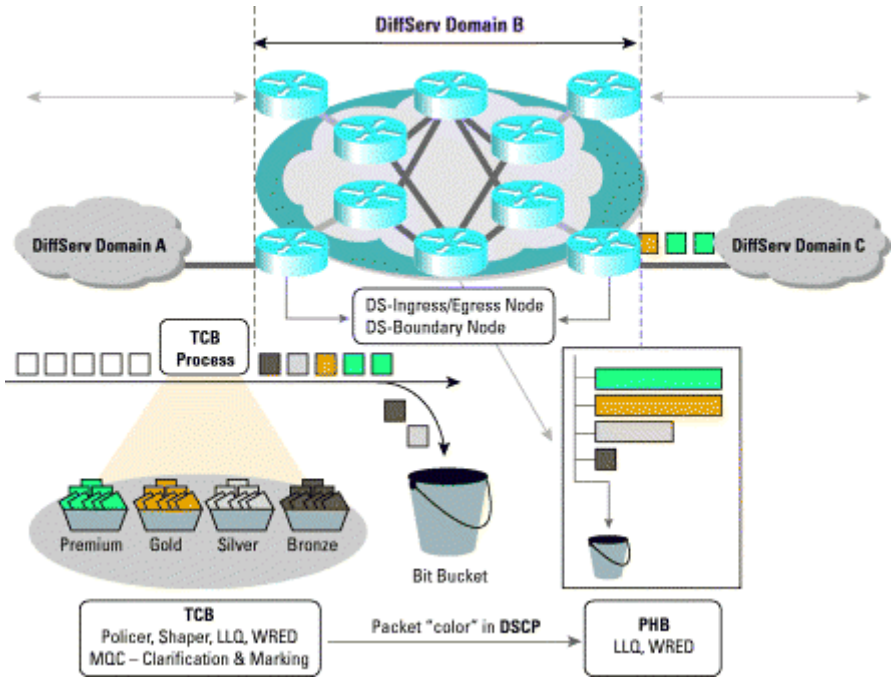


Figura 13. Arquitectura de Servicios Diferenciados

La red utiliza las especificaciones de la QoS para clasificar, marcar, limitar, aplicar políticas y desempeñar encolamiento inteligente a través de características que proveen la funcionalidad necesaria para implementar Servicios Diferenciados. A continuación se indican esas características las cuales serán presentadas con mayor detalle durante el desarrollo de este trabajo.

La clasificación y marcado es una técnica que permite la especificación de un trafico independientemente de la política de Calidad de Servicio.

Se pueden definir clases de servicio basadas en la marcación de paquetes, esto provee la posibilidad de diferenciar paquetes al diseñar diferentes valores de identificación para cada paquete.

La gestión de la congestión es un mecanismo de programación usado para proveer un mínimo ancho de banda garantizado para clases de trafico durantes momento de congestión.

Esta característica permite limitar la velocidad de entrada o salida de una clase de trafico basada en los criterios definidos por el administrador de red. También le permite al sistema marcar paquetes al configurar el valor de la Precedencia IP.

Los servicios diferenciados utilizan Weighted Random Early Detection (WRED) a través del valor de DSCP o IP Precedente cuando calcula la probabilidad de descarte.

III. FUNCIONES DE LA CALIDAD DE SERVICIO

En la figura 14 se muestra las cinco categorías de herramientas y su descripción en términos sencillos de como contribuyen con la QoS.

Figura 14. Funciones de la QoS.

La clasificación y el marcado es la identificación, división del trafico en diferentes clases y el marcado de acuerdo al comportamiento y a las políticas empresariales. Algunas herramientas para clasificación y marcado son: NBAR, PBR, ACL/Route Map, CAR y MQC (Modular Quality of Service command-line interfaz).

La gestión de la congestión es la prioritización, protección y aislamiento del trafico basado en el marcado. Herramientas que realizan estas funciones son IP RTP Priority, CBWFQ y LLQ.

La técnica de evitar congestión descarta paquetes específicos basados en el marcado para evitar congestión en la red. WRED es un ejemplo de herramientas utilizadas para evitar congestión.

El condicionamiento del trafico verifica el trafico para detectar comportamientos fuera de lo común y mantener la integridad de la red. Las técnicas de GTS y FRTS son ejemplos de esta función.

La eficiencia de enlace fragmenta el trafico a velocidades de transmisión y comprimen las cabeceras para mejorar la eficiencia en la WAN. Herramientas de fragmentación incluyen LFI y FRF.12. Herramientas de compresión incluyen CRTP.

3.1 Clasificación y Marcado

La clasificación es la identificación y división del trafico en diferentes clases. El marcado, es realizado con cualquiera de estos tres métodos: Modular QoS, command-line interfaz (MQC), policer o CAR.

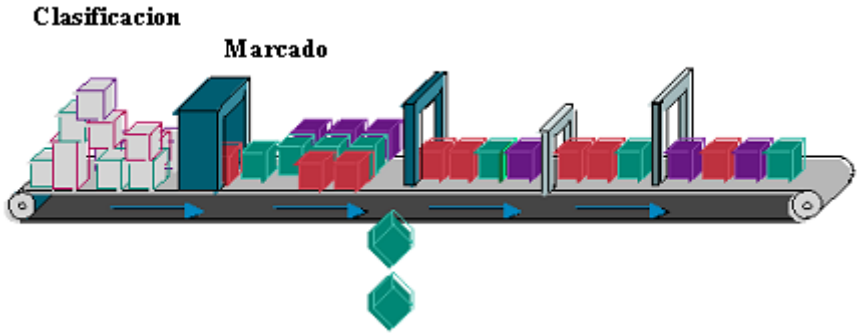


Figura 15. Clasificación y Marcado.

Como se muestra en la figura 15 la clasificación consiste en la identificación y particionamiento del trafico en diferentes clases y el marcado es el proceso de marcar para diferenciar el trafico de acuerdo a su comportamiento y a las políticas de la empresa.

El reconocimiento de aplicaciones basadas en red (NBAR) es una herramienta de clasificación que puede reconocer una amplia variedad de aplicaciones, incluyendo aplicaciones basadas en Web y aplicaciones cliente/servidor, que asignan dinámicamente números de puertos TCP o UDP. Una vez que la aplicación es reconocida, la red puede invocar servicios específicos para esta aplicación en particular. NBAR actualmente trabaja con características de QoS para asegurar que

el ancho de banda de la red es usado de la mejor manera para cumplir los objetivos del cliente. Estas características incluyen la habilidad para garantizar ancho de banda a aplicaciones criticas, limitar el ancho de banda a otras aplicaciones, descartar paquetes selectivamente, para evitar congestión y marcar paquetes apropiadamente para que la red pueda proveer la QoS requerida.

IP Precedence es uno de los métodos de Servicios Diferenciados para la ubicación de recursos en la red. Los Tres bits en la cabecera IP designados como el campo de tipo de servicio (ToS) puede ser manipulado para informar a los dispositivos de red que una prioridad debería ser dada al paquete IP a medida que este viaja a través de la red.

Para que los métodos de QoS puedan ser utilizados en la red, el trafico debe ser clasificado con diferentes propiedades**.** Cada clasificación debe estar marcada para identificación y así poder aplicar los métodos de calidad de servicio.

La clasificación de paquetes usa un descriptor de trafico (por ejemplo, IP Precedence o DSCP) para categorizar un paquete con un grupo especifico con el fin de definir el paquete. Después de que el paquete ha sido definido (clasificado), es entonces accesible para un manejo con QoS en la red.

Usando clasificación de paquetes, se puede particionar el trafico de red en múltiples niveles de prioridad o clases de servicio. Cuando descriptores de trafico son usados para clasificar el trafico, la fuente acepta adherirse a los términos contratados y la red compromete una calidad de servicio para ese trafico.

El marcado está relacionado a la clasificación. El marcado permite clasificar un paquete o trama basado en un descriptor de trafico especifico. La marcación de un paquete o trama con su clasificación permite configurar información en las cabeceras de las capas 2, 3 o 4, o configurar información en la carga útil

(payload) de un paquete, con el fin de que el paquete o trama pueda ser identificada y distinguida de otros paquetes o tramas.

El estándar IEEE 802.1p es un suplemento del IEEE 802.1D, el cual es un estándar para la conexión de LANs a través de Bridges. El estándar 802.1p soporta ocho clases de servicios (CoS). Un valor de CoS de 0 (cero) significa prioridad ninguna.

El estándar 802.1Q es una especificación de la IEEE para la implementación de virtual LANs (VLANs) en Switches capa 2.

Una trama Ethernet puede no ser compatible con 802.1p/Q. Si este es el caso, entonces un campo Tag de cuatro bytes es insertado en la trama Ethernet. El campo Tag sirve el propósito de ambos estándares (802.1p y 802.1Q). Los primeros tres bits, son conocidos como los bits de prioridad (bits 802.1p), el resto del campo Tag es para los campos CFI (Canonical Format Identifier) y VLAN ID, los cuales son campos 802.1Q. El campo Tag es algunas veces llamado el campo 802.1p/802.1Q, o el campo 802.1p/Q.

Entonces cuando una trama Ethernet es marcada para prioridad, los tres bits 802.1p son configurados a un valor de 0-7.

Antes que la IETF definiera métodos de QoS en IP (capa 3), la ITU-T, el ATM Forum y el Frame Relay Forum ya habían realizado estándares para manejar QoS en la capa 2 en redes ATM y Frame Relay. El estándar ATM define una infraestructura de QoS muy rica por soportar contratos de trafico, QoS ajustable (tales como Peak Cell Rate PCR, Minimum Cell Rate MCR y otras), señalización y Control de Admisión de Conexión (CAC).

Frame Relay, por otro lado provee un mecanismo más simple de QoS tales como CIR, notificación de congestión y la fragmentación de Frame Relay (FRF.12).

A través de estos ricos mecanismos de QoS existentes en tecnologías de transporte capa 2, no es posible alcanzar una verdadera calidad de servicio de extremo a extremo, a menos que una solución capa 3 esté presente. Los proveedores de servicio que ofrecen ambos servicios ATM/Frame Relay e IP pueden proveer robustas soluciones de QoS a sus clientes.

El mapping de QoS capa 3 a QoS capa 2 es el primer paso para lograr una completa solución que no dependa de alguna tecnología capa 2. Ambos IntServ y DiffServ pueden ser implementados sobre infraestructuras de QoS con ATM y Frame Relay. Por ejemplo, el servicio de carga controlada puede ser implementado usando RSVP (IntServ) sobre un circuito virtual ATM VBR-rt (variable bit rate, real time). Con DiffServ, paquetes marcados con diferente valor de ToS pueden ser enviados sobre diferentes PVC o SVC ATM. Como un ejemplo, el trafico de alta prioridad puede ir sobre un circuito virtual VBR-nrt y todo el otro trafico puede ir sobre un circuito virtual available bit rate (ABR). Similarmente Frame Relay Traffic Shaping (FRTS), FRF.12 puede ser usado para complementar la calidad de servicio IP.

La precedencia IP usa los tres bits de precedencia en la cabecera del protocolo IPv4 para especificar la clase de servicio para cada paquete, como se muestra en la figura # 16. Se puede particionar el trafico en 6 clases de servicios usando IP Precedence. Las tecnologías de cola a través de la red pueden entonces usar esta señal para proveer el manejo apropiado del trafico.

Características tales como enrutamiento basado en política (Policy-Based Routing), marcado basado en clases (Clase-Based Marking) y velocidad de acceso acordada (CAR) pueden ser usadas para configurar precedencia basada en clasificación de listas de acceso extendidas. Esto permite una flexibilidad para la asignación de precedencia, incluyendo asignación por aplicación o usuario, o por subred fuente o destino. Típicamente estas funcionalidades son desarrolladas tan cerca como sea

posible del borde de la red para que cada elemento subsiguiente de la red pueda proveer servicios basados en una política determinada.

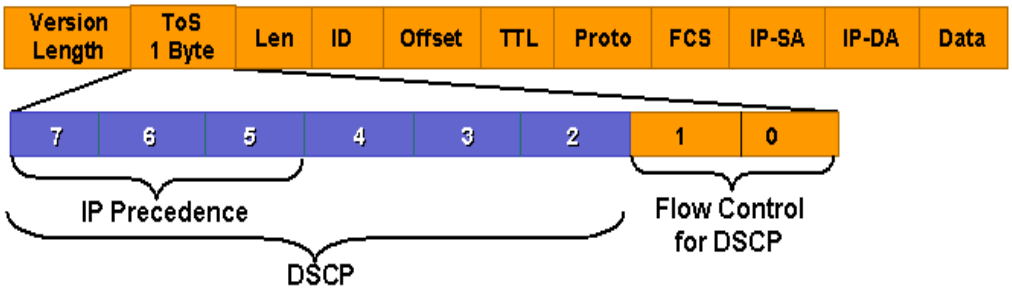


Figura 16. Uso del Campo ToS.

La precedencia IP puede también ser configurada en el host o cliente de red, y esta señalización puede ser usada opcionalmente; sin embargo, esto puede ser sobre escrito por políticas ya existentes en la red.

La precedencia IP habilita clases de servicios a ser establecidas usando los mecanismos de cola existentes en la red (por ejemplo WFQ o WRED) sin ningún cambio en las aplicaciones o complicados requerimientos de red. Estas posibilidades son también aplicables en IPv6 usando su campo de prioridad.

DiffServ es un nuevo modelo que reemplaza, y es compatible con IP Precedence. DiffServ usa seis bits de prioridad, lo cual permite clasificaciones de hasta 64 valores (0 a 63). Un valor de DiffServ es llamado un Differentiated Services Code Point (DSCP).

Con DiffServ, el campo DS suplanta el octeto ToS de IPv4 y el octeto clase de trafico de IPv6. Seis bits del campo DS son usados como los puntos de códigos DS para seleccionar el Per-Hop Behavior (PHB) en cada interfaz.

3.1.1 Donde Marcar

La clasificación podría toma lugar en el borde de la red, típicamente en el gabinete de cableado, en el dispositivo de video o en el teléfono IP mismo. En la figura 17 se demuestra esto con un teléfono IP. Los paquetes pueden ser marcados como importantes usando la configuración del CoS en los bits de prioridad de usuarios dentro de la porción 802.1p del campo 802.1PQ o en la precedencia IP/bits DSCP en el campo ToS/DS en la cabecera IPv4.

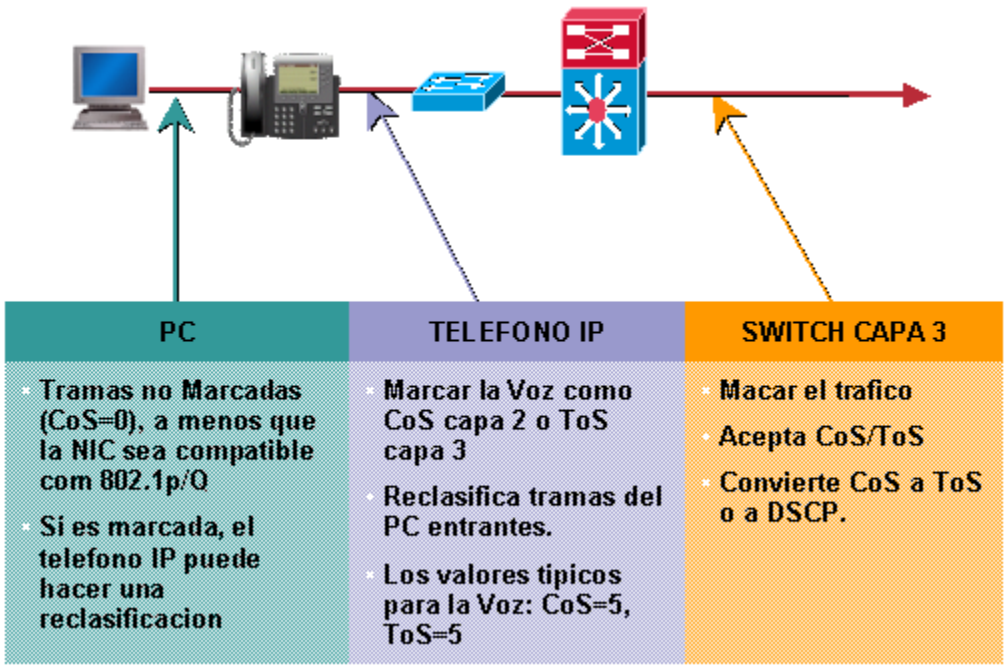


Figura 17. Donde Marcar

Los teléfonos IP (por ejemplo, los de Cisco Systems) pueden marcar paquetes de Voz como de alta prioridad usando el campo CoS o ToS. Por defecto, el teléfono envía paquetes con la bandera 802.1p colocada con el valor CoS y ToS configurado a 5.

Debido a que muchos PCs no tienen una tarjeta de red compatible con el 802.1Q, ellas envían los paquetes sin esa bandera. Esto significa que las tramas no tienen

un campo 802.1p. Por lo tanto, a menos que la aplicación ejecutándose en el PC envíe paquetes con un valor especifico de CoS, este campo será cero (0). Un caso especial es donde el stack del TCP/IP en el PC ha sido modificado para enviar todos los paquetes con un valor de ToS diferente de cero. Típicamente esto no sucede y el valor del ToS es cero.

El switch usa su cola para almacenar tramas entrantes antes de enviarlas a la maquina de conmutación. El switch usa los valores del CoS para poner las tramas en una cola apropiada. El switch también puede emplear mecanismos tales como WRED para hacer descartes inteligentes en una cola y WRR para proveer más ancho de banda a algunas colas que a otras.

3.1.2 Herramientas para Clasificar y Marcar

Existen varias técnicas para realizar el procedimiento de Clasificación y Marcado del trafico en una red, estas pueden ser del tipo de interfaz de línea de comando (CLI) o Modulares. A continuación se explican algunas de ellas:

NBAR: Network Based Application Recognition (NBAR) es una interfaz de línea de comando para QoS habilitada para realizar clasificación y descubrimiento de protocolos. NBAR puede determinar mezclas de trafico en la red, lo cual es importante para aislar problemas de congestión.

PBR: Policy Based Routing (PBR) puede ser usado con route map sobre las interfaces para clasificar y marcar paquetes. El route map realiza la clasificación (chequea el patrón en la dirección o en otras características del paquete), entonces se marca (se altera la precedencia IP) de acuerdo a esas características.

ACL: Access Control List (ACL), una lista de control de acceso puede ser usada para permitir o negar acceso a interfaces o líneas, y un route map (comúnmente

usada para permitir o negar acceso a una interfaz receptora, pueden ser usadas juntas en la implementación de políticas de QoS.

Dial Peers: Para darle una alta prioridad al trafico de voz sobre otros tráficos en la red, se puede asociar el trafico de voz a un Dial Peer particular de voz usando la precedencia IP.

CAR: Además de sus capacidades para establecer políticas, CAR puede ser usada para clasificar y marcar trafico.

3.2 Gestión de la Congestión

Las características de gestión de la congestión operan para controlar la congestión una vez que ocurre. Una vía para que los elementos de red manejen el sobre flujo del trafico entrante es mediante el uso de algoritmos de encolamiento, para ordenar el trafico y determinar algunos métodos de prioritizarlo en el enlace de salida. Cada algoritmo de encolamiento fue diseñado para solucionar un problema de trafico de red especifico y tiene un efecto sobre el desempeño de la red.

Algunos algoritmos usados para la gestión de la congestión son:

- FIFO- First-In, First-Out
- PQ- Priority Queuing
- CQ- Custom Queuing
- WFQ- Flow Base Weighted Fair Queuing
- CBWFQ- Class Based WFQ
- IP RTP Priority- PQ/WFQ
- LLQ- Low Latency Queuing

Cada interfaz en los equipos de comunicaciones (Por ejemplo, Cisco Routers) tiene su sistema de encolamiento basado en hardware y en software, el sistema de encolamiento basado en hardware siempre utiliza FIFO. El sistema de encolamiento basado en software puede ser seleccionado y configurado dependiendo de la plataforma y la versión de sistema operativo.

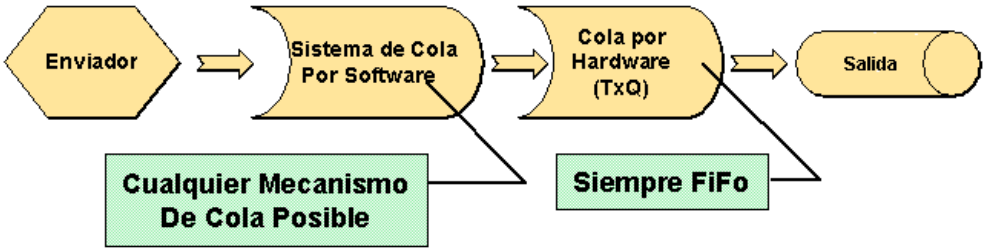


Figura 18. Encolamiento en Hardware/Software

Cuando un paquete esta siendo transmitido el router no usará el sistema de encolamiento basado en software si la cola está vacía o si la cola basada en hardware no esta llena. En la figura 19 se muestra un diagrama de flujo que representa el uso del encolamiento basado en Software.

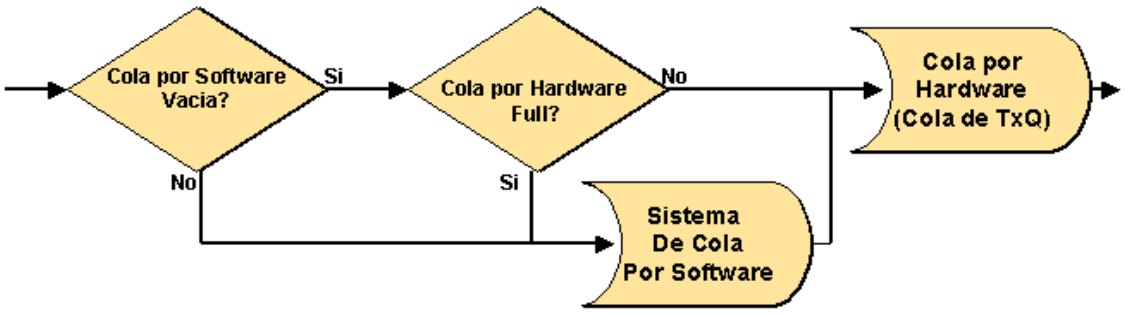


Figura 19. Uso del Encolamiento Basado en Software.

Los routers determinan la longitud de la cola basada en hardware mediante el ancho de banda configurado en la interfaz. Un gran tamaño de la cola de Transmisión puede resultar en un pobre desempeño de la cola basada en software, mientras que un tamaño corto de la cola de transmisión podría resultar en un gran numero de interrupciones lo cual generaría un alto uso del CPU y un bajo uso del enlace.

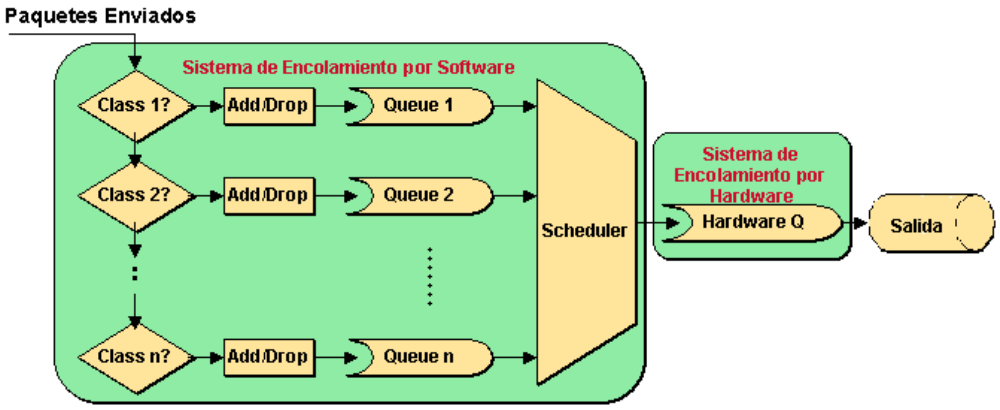


Figura 20. Componentes de un Sistema de Encolamiento

Cada sistema de encolamiento tiene tres componentes principales:

- Clasificación: Selecciona la clase de trafico
- Política de Inserción: Determina si un paquete puede o no ser puesto en cola.
- Política de Servicio: Se planifica la colocación del paquete en la cola de salida (Encolamiento basado en Hardware).

3.2.1 FIFO

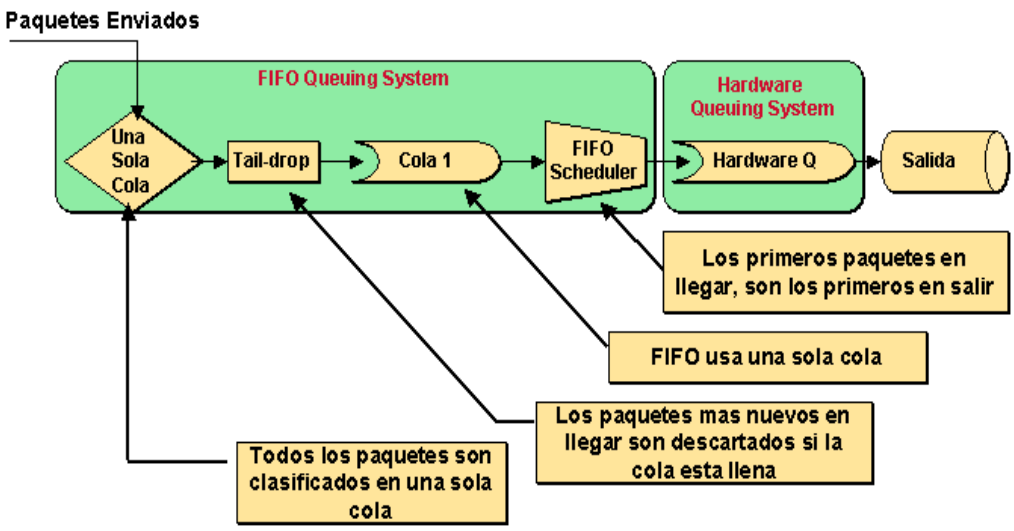


Figura 21. Sistema de Encolamiento FIFO

FIFO es el más común y el más simple algoritmo de encolamiento, trabaja bien si el enlace no está congestionado. Este es el método de encolamiento por defecto para cualquier interfaz con más de 2 MB de ancho de banda. El primer paquete en ser colocado en la interfaz de salida es el primer paquete en salir. El principal problema con FIFO es que cuando una estación comienza la transferencia de archivos, esta puede consumir todo el ancho de banda de cualquier interfaz, en detrimento de sesiones interactivas. FIFO es efectivo en ambientes con grandes anchos de bandas con poco delay y mínima congestión.

Características:

- Más rápido y más simple.
- Soportado en todas la plataformas
- Soportados en todas las rutas de conmutación
- Causa Jitter

- Permite monopolización de enlaces
- No maneja prioritización de trafico

3.2.2 Priority Queuing

PQ es aplicable a interfaces seriales que operan a velocidades de línea de hasta un E1 (2.048Mbps). PQ satisface los requerimientos de encolamiento para el trafico de Voz, pero cada cola puede perjudicar colas de baja prioridad. Esto significa, que las colas de baja prioridad puede que nunca se les permita transmitir paquetes si las colas de más alta prioridad siempre tiene trafico para transmitir (como se muestra en la figura 23).

En general, se podrían usar una de las técnicas de encolamiento más nuevas tales como IP RTP Priority o LLQ para asegurar que los paquetes sensitivos al retardo obtengan la más alta prioridad. Sin embargo, el PQ tradicional es más flexible si se tiene múltiples tipos de tráficos de alta prioridad. Por ejemplo, si se tiene trafico de VoIP y de SNA fluyendo a través del mismo enlace, entonces se podría configurar los paquetes de VoIP para que tengan una prioridad más alta, y asignar el trafico SNA a una prioridad media. De esta forma, el trafico SNA está todavía prioritizado sobre todo el trafico de datos, pero no impacta sobre la calidad de la Voz.

Priority Queuing también es apropiado cuando el trafico de Voz está mezclado con trafico de Video y datos. Se puede asegurar que el trafico de Voz reciba el mejor servicio, y el trafico de IPVC (IP Videoconferencing) es prioritizado sobre el trafico regular de datos sin impactar el trafico de Voz. Si se desea transmitir Voz, Video y trafico SNA sobre el mismo enlace, se puede configurar PQ por defecto como bajo para tener tres diferentes colas con un tratamiento de alta prioridad. VoIP podría estar siempre en la cola de más alta prioridad, y colocar el trafico de Video y SNA en colas de prioridades medias y normales respectivamente.

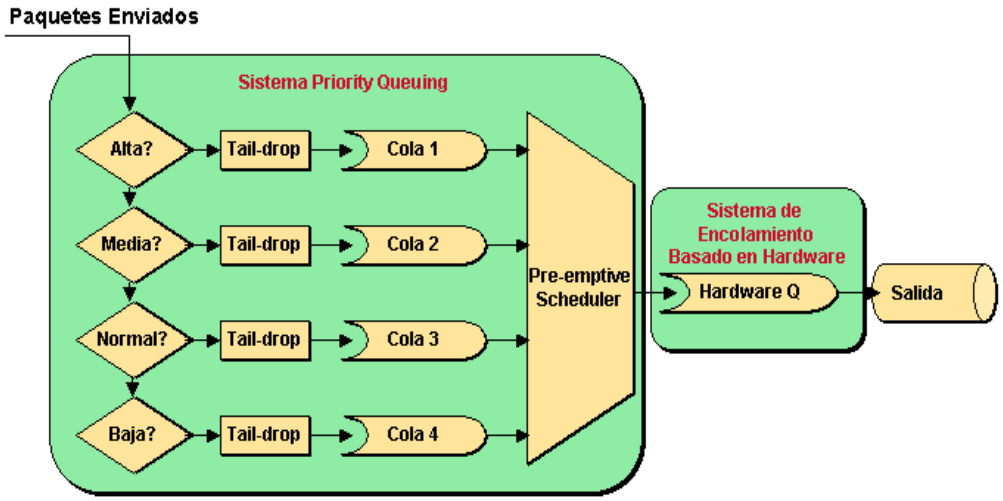


Figura 22. Sistema de Encolamiento Priority Queuing

Como se muestra en la figura 22, el sistema PQ utiliza cuatro colas FIFO con diferentes prioridades.

Priority Queuing soporta clasificación de paquetes IP para:

- Interfaz fuente
- Lista de Acceso
- Tamaño del paquete
- Fragmentos
- Numero de puertos TCP origen y destino
- Numero de puertos UDP origen y destino

También soporta clasificación para los siguientes protocolos:

- IPX
- CLNS
- DECNet

- Apple Talk
- VINES
- DLSw

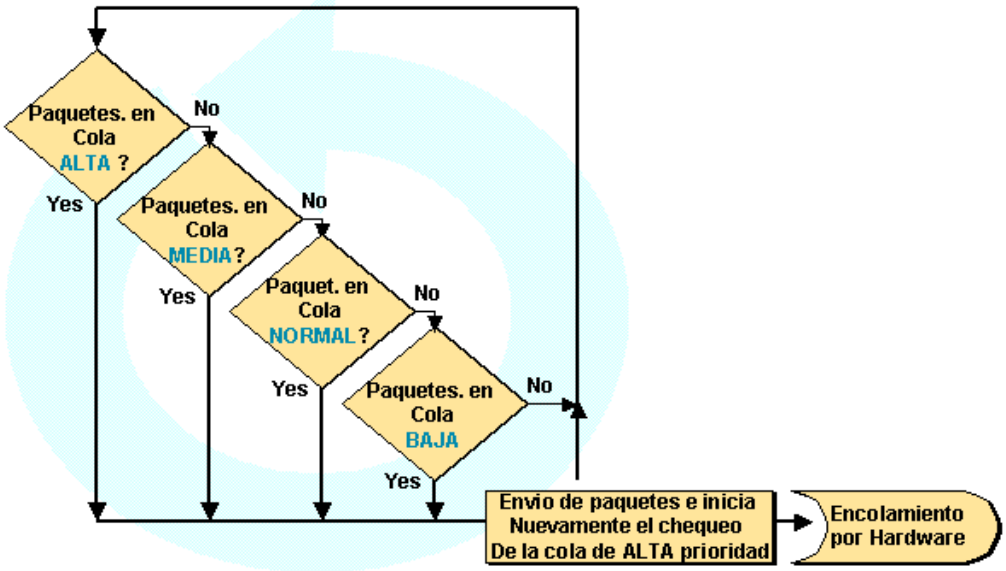


Figura 23. Planificación de la Cola de Prioridad

3.2.3 Custom Queuing

Al igual que Priority Queuing, Custom Queuing es aplicable a interfaces seriales que operan a velocidades de línea de hasta un E1 (2.048 Mbps). Custom Queuing es una forma de habilitar de manera flexible la distribución del ancho de banda entre clases de trafico definidas, sin perjudicar a ninguna de las clases de trafico.

Las estrategias de Fair Queuing no son bien ubicadas en redes de VoIP u otros tráficos sensitivos al Jitter, por que los paquetes en la cola incurren en un delay variable mientras esperan por que cada cola sea servida. El impacto adverso en la transmisión de Voz y Video es mucho mayor cuando el numero de colas o el numero de bytes aceptados para cada cola se incrementa.

Las colas son servidas una a una, y cada una tiene permitido transmitir el numero de bytes configurado, antes de pasar el control a la próxima cola. Si la cantidad máxima es alcanzada antes de terminar de transmitir un paquete, la cola tiene permitido finalizar la transmisión del paquete. Por ejemplo, si una cola que contiene dos paquetes de 1500 bytes es configurada para transmitir 1501 bytes, entonces ambos paquetes son transmitidos (para un total de 3000 bytes). Se debe considerar el efecto del tamaño del paquete cuando se está configurando la cuenta de bytes para asegurar que el ancho de banda sea distribuido como se espera.

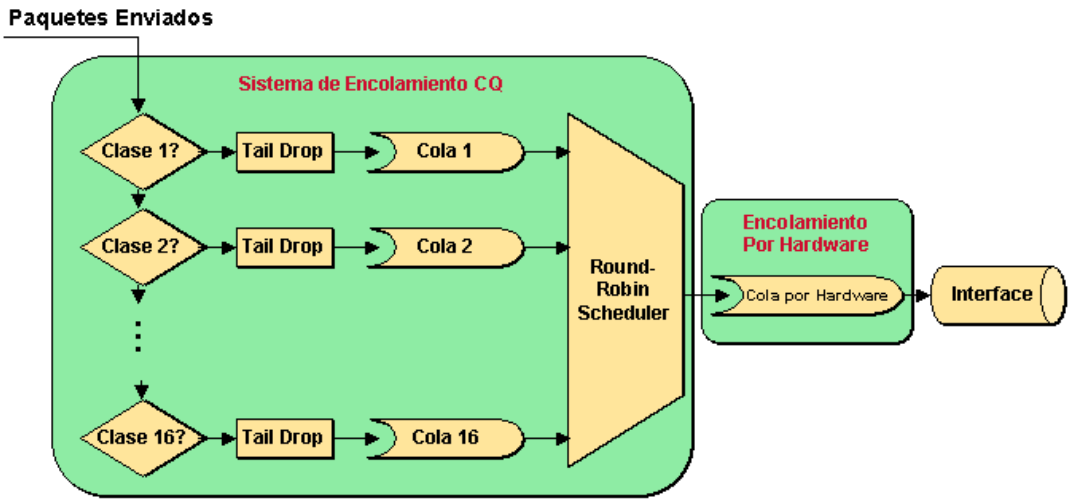


Figura 24. Sistema de Encolamiento Custom Queuing

Como se muestra en la figura 24 CQ maneja hasta 16 diferentes colas, la planificación de entrega de los paquetes a la interfaz de salida se realiza una cola tras otra y se repite el proceso.

3.2.4 Flow Based Weighted Fair Queuing

Weighted Fair Queuing (WFQ) está diseñado para proveer un igual tratamiento a todos los flujos de trafico. Un flujo es definido por la dirección origen/destino y el numero de puerto. FIFO permite que flujos con un alto volumen de paquetes (por ejemplo, FTP) consuman una cantidad desproporcionada del ancho de banda de la interfaz. Flow Based WFQ soluciona ese problema al establecer una cola separada para cada flujo de trafico, y transmitir algo de trafico de cada cola una a una. De esta forma, los flujos con alto volúmenes de paquetes no bloquean la transmisión de flujos con pocos paquetes, tales como sesiones Telnet u otras aplicaciones interactivas.

Si se desea usar WFQ en interfaces que pasan trafico interactivo, se debe configurar también IP RTP Priority para asegurar un adecuado desempeño para esos paquetes.

A diferencia de Priority o Custom Queuing, WFQ es fácil de configurar; por que WFQ es el método por defecto para interfaces a velocidades de un E1 o menores.

Para configurar WFQ el primer parámetro que se debe especificar es la profundidad para cada flujo de trafico. Se puede ubicar hasta 4096 paquetes por colas de flujo. El segundo parámetro a especificar es el máximo numero de colas disponibles (hasta 4096) para clasificar conversaciones. El ultimo parámetro a considerar es el numero de colas que pueden ser reservadas por RSVP. Hay que tomar en cuenta que capacidades extras consumen memoria del sistema, por lo cual no es recomendable maximizar estos parámetros.

3.2.5 Class Based WFQ

CB-WFQ, es un modelo de encolamiento muy robusto que combina los mejores elementos de PQ, CQ y WFQ con WRED. CB-WFQ es más complejo de configurar que otros mecanismos de encolamiento. Esta complejidad permite la libertad para organizar clases de trafico de acuerdo a una variedad de características y aplicar políticas de encolamiento diferente a cada clase.

Se puede asignar directamente ancho de banda a cada clase, de manera opuesta al método arcaico de cuenta de bytes requerido por CQ. Se puede especificar una o múltiples clases para que tengan un tratamiento de alta prioridad, sin limitar a las otras clases en la estructura de prioridad. Se puede usar WFQ para tener múltiples colas de conversación en una clase, mientras todavía se impone una limitación de ancho de banda para la clase en su totalidad. Por un lado se puede definir desde el tipo de encolamiento y al ancho de banda para cada clase, como también se puede especificar WRED para cada una de ellas. De esta forma se puede usar WRED donde se requiera, y ajustar parámetros (umbral de congestión y probabilidades de descarte) basado en las características de cada clase de trafico.

Existen tres pasos principales en el proceso de implementación de CB-WFQ:

- Clasificar el trafico en clases
- Aplicar políticas a las clases
- Asignar una política de servicio a una interfaz

Clasificando el trafico en clases:

Una clase de trafico está definida por un map class. Todo el trafico asignado a un map class debería tener los mismos requerimientos de QoS. Después de crear un

map class, se puede asignar ese trafico basado en la interfaz de entrada, el tipo de protocolo, o en una lista de acceso.

Aplicando políticas a las clases:

Después de que el trafico es clasificado en clases, se debe definir las políticas de trafico que serán aplicadas a cada clase. Se puede definir políticas para la clase de trafico por defecto (que no cae en ninguna clase existente). La colección de clases de trafico y las políticas que aplican a ellas es llamado política de servicio. La política de servicio debe ser aplicada a la interfaz de un Router para que pueda tener efecto. Las siguientes opciones son ejemplo de políticas aplicables:

- Mínimo ancho de banda durante periodos de congestión
- Profundidad de la cola FIFO para clases definidas
- Comportamiento de WRED durante congestión
- Umbral de congestión y probabilidades de descarte para clases WRED
- Tratamiento prioritario para clases en su totalidad
- Comportamiento FIFO o WFQ para las clases por defecto
- Numero de conversaciones WFQ para la clase defecto

Asignar una política de servicio a una interfaz:

Para asignar una política de servicio a las interfaces de debe entrar en el modo de configuración de interfaz y especificar cual política de servicio estará usando la interfaz en particular.

3.2.6 IP RTP Priority

El flujo basado en WFQ es un buen mecanismo de encolamiento para el trafico de datos, sin embargo no es una buena opción para el trafico sensible al Jetter. Cuando el numero de conversaciones de Voz y el trafico de datos incrementan, los paquetes sensibles a las distintas formas de retardo deben esperar a que las colas sean servidas una a una. Un buen arreglo podría ser habilitar colas de prioridad para darle un mejor tratamiento al flujo de Voz y Video, mientras que otros flujos (por ejemplo, datos) reciben el tratamiento normal con WFQ. Esto es exactamente la funcionalidad proveída por IP RTP Priority.

Para habilitar este sistema de encolamiento (por ejemplo, en Routers Cisco), el primer parámetro a usar es el numero de puerto UDP para el comienzo del rango que tendrá prioridad. Los Routers de Cisco usan el puerto 16384 como el inicio del rango de RTP para paquetes de audio. El segundo parámetro especifica el numero de puertos UDP a ser prioritizados después del puerto inicial. El parámetro final es la cantidad máxima de ancho de banda permitida para la cola de prioridad. El ancho de banda asegura que la cola prioritaria no perjudique otras colas de conversación que usen WFQ.

Cuando la cola prioritaria excede el ancho de banda permitido, ese trafico es descartado. No es posible controlar cuales paquetes se pierden cuando el ancho de banda se excede, por lo cual todas los flujos de esa cola pueden experimentar una reducción en la calidad.

La solución a este problema es usar RSPV u otro mecanismo para Control de Admisión de Conexión (CAC). Esto permite que si se quiere establecer una sesión de Videoconferencia pero no existe ancho de banda suficiente entonces no se establece la sesión.

3.2.7 Low Latency Queuing

Al igual que CBWFQ, Low Latency Queuing (LLQ) es también un algoritmo de encolamiento híbrido. LLQ es una combinación de Priority Queuing, Custom Queuing y WFQ. En efecto LLQ fue originalmente llamado PQ-CBWFQ, pero ese nombre fue cambiado a LLQ por razones de mercadeo.

LLQ adiciona una estricta cola de prioridad a la planificación que realiza CBWFQ. LLQ está diseñado exclusivamente para Voz y aplicaciones de Video interactivo. La cola estricta de prioridad es completamente vaciada antes de que las colas CBWFQ sean servidas.

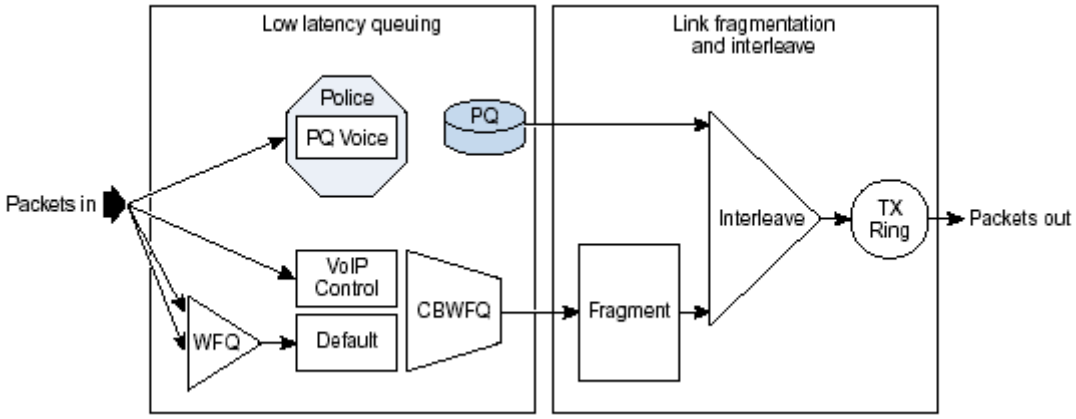


Figura # 25. Subsistema lógico de encolamiento LLQ/CBWFQ

El subsistema de encolamiento para LLQ/CBWFQ es mostrado en la figura 25. En esa figura, el caudal de trafico de VoIP es colocado en el LLQ, comúnmente conocido como la cola de prioridad (PQ). Cuando hay trafico presente en LLQ, este es servido con tratamiento preferencial exhaustivamente (primero hasta que no exista trafico). La cola de prioridad tiene la política de que el trafico prioritario no

puede afectar a las otras clases de trafico. Por lo cual LLQ debe ser implementado correctamente. La acción de esta política podría afectar negativamente la calidad de la voz y el video si el valor de la política es configurado más bajo que la cantidad total de trafico de voz o video a ser servido por LLQ, en cuyo caso el trafico excedido seria descartado.

Esta figura también muestra que el trafico de control de VoIP es colocado en una clase basada en WFQ de su propiedad para que un ancho de banda pequeño pueda ser garantizado. Esto asegura que el trafico de control de llamadas sea transmitido aun durante periodos de congestión.

3.3 Evitando la Congestión

La técnica de evitar congestión monitorea la carga de trafico en la red como un esfuerzo para anticipar y evitar congestión en cuellos de botella comunes. Esto es logrado a través del descarte de paquetes.

Existen dos caminos para que la red pueda responder a la congestión:

- Tail Drop: Es el método de encolamiento por defecto ante congestión. Cuando la cola de salida está llena y Tail Drop está funcionando, todos los paquetes tratando de entrar la cola los descarta hasta que la congestión es eliminada y la cola ya no esté llena.
- WRED: Weighted Random Early Detection (WRED), incrementa la probabilidad de que una congestión será evitada por descartar paquetes de baja prioridad de una manera aleatoria con el fin de evitar reacción de sincronización global. Cuando estaciones TCP hacen drop de paquetes, esto causa a las estaciones que transmiten bajen su velocidad con el fin de evitar congestión.

Es importante señalar que WRED no es recomendado para colas que manejan trafico de Voz. La red no debería estar diseñada para descartar paquetes de Voz debido a que la perdida de paquetes de Voz reduce la calidad de la voz, porque, los paquetes de Voz son transportados mediante UDP. WRED controla la congestión por impactar otros trafico basados en TCP evitando congestión, lo cual ayuda a asegurar la calidad de servicio para Voz y Video.

Existen varias técnicas para evitar la congestión, entre ellas tenemos:

- Tail Drop
- RED
- WRED
- FRED

3.3.1 Tail Drop

Tial Drop es el método de encolamiento por defecto cuando existe congestión. Tail Drop trata todo el trafico igual y no hace diferencia entre clases de servicio. Las colas colapsan durante periodos de congestión. Cuando la cola de salida está llena y Tail Drop está en efecto, los paquetes son descartados hasta que la congestión sea eliminada y la cola deje de estar llena.

Tail Drop tiene varios defectos, tales como:

- Sincronización del TCP
- Afecta el normal funcionamiento del TCP
- Alto delay y Jitter
- Descartes no diferenciados
- Poco feedback para TCP
- Un uso alto sobre el buffer de manera constante genera delay

- Flujos agresivos perjudican a otros flujos
- Buffer variable causa Jitter

Estos defectos son mostrados en la figura 26.

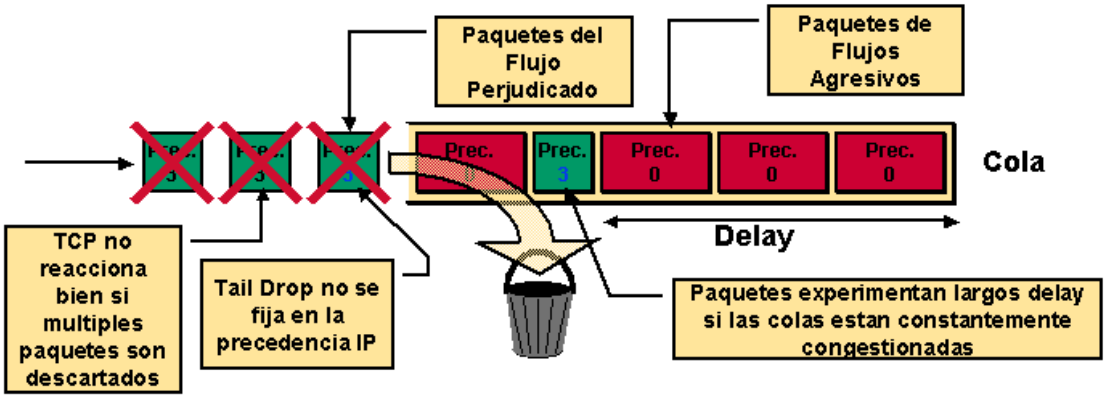


Figura # 26. Defectos de Trail Drop

En conclusión, Trail Drop no debería ser implementado.

3.3.2 RED

Random Early Detection es un mecanismo que descarta paquetes de forma aleatoria antes de que la cola esté llena. Esto permite que los paquetes seleccionados aleatoriamente bajen su velocidad de transmisión. Si el paquete origen está usando TCP, éste decrementa su velocidad de transmisión hasta que todos los paquetes alcanzan su destino, indicando que la congestión ha sida eliminada.

La probabilidad de que un paquete sea descartado está basada en tres parámetros de configuración:

1. Umbral Mínimo: Cuando el promedio de profundidad de la cola está por arriba del umbral mínimo, RED comienza a descartar paquetes. La velocidad del descarte de paquetes es incrementada linealmente de acuerdo al incremento en el tamaño de la cola, hasta que el tamaño promedio de la profundidad de la cola alcanza el umbral máximo.
2. Umbral Máximo: cuando el tamaño promedio de la cola está sobre el umbral máximo, todos los paquetes son descartados. Si la diferencia entre el umbral máximo y el umbral mínimo es muy pequeña, muchos paquetes pueden ser descartados de una vez, resultando en sincronización global.
3. Denominador de Probabilidad Marcada: Esta es la fracción de paquetes descartados cuando la profundidad promedio de la cola está en el umbral máximo.

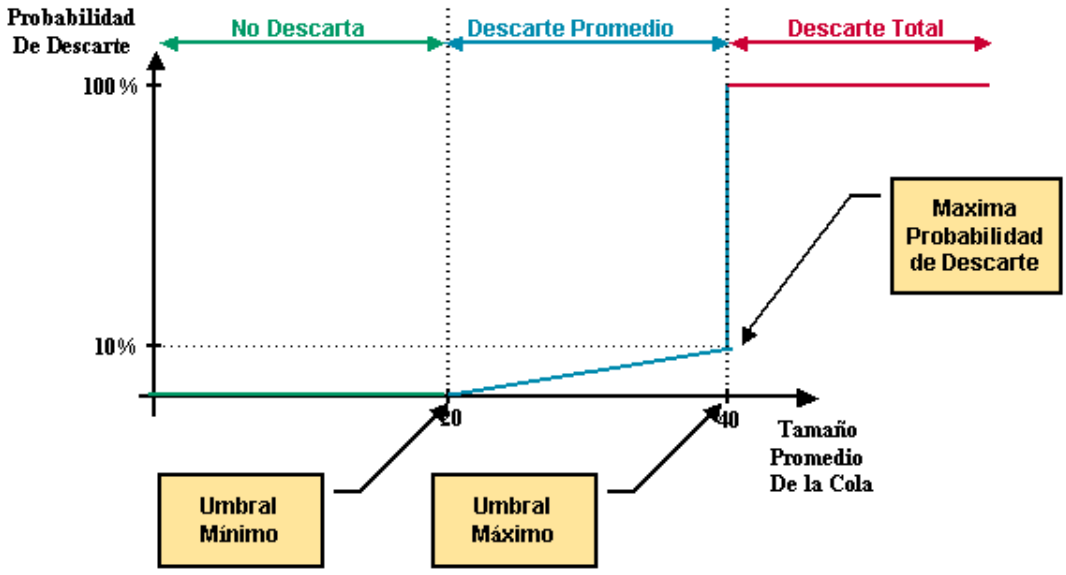


Figura # 27. Perfil de RED

3.3.3 WRED

Weighted Random Early Detection es una implementación RED de Cisco. WRED combina RED con IP Precedence o DSCP y descarta paquetes basado en la precedencia IP o en el marcado DSCP.

WRED monitorea la profundidad promedio de la cola en el router y determina cuando comenzar a descartar paquetes basado en la profundidad de la cola. Cuando la profundidad promedio de la cola cruza el umbral mínimo definido, WRED comienza a descartar paquetes (TCP y UDP) con una cierta probabilidad. Si la profundidad promedio de la cola cruza el umbral máximo definido, entonces WRED se convierte en Tail Drop, donde todos los paquetes entrantes pueden ser descartados. La idea de usar WRED es mantener la profundidad de la cola entre el umbral mínimo y máximo.

WRED puede hacer descartes selectivos de tráficos de prioridad baja cuando la interfaz comienza a congestionarse y provee características de desempeño diferenciadas para diferentes clases de servicio.

Para interfaces configuradas para usar RSVP, WRED escoge paquetes de otros flujos para descartar distintos al flujo de RSVP. También, la precedencia IP o el DSCP gobiernan cuales paquetes serán descartados, es decir, trafico con menor prioridad tiene una mayor velocidad de descarte.

WRED evita los problemas de sincronización global que ocurren cuando Trail Drop es usado como el mecanismo para evitar congestión. La sincronización global se manifiesta cuando múltiples hosts TCP simultáneamente reducen su velocidad de transmisión en respuesta a los descartes de paquetes, luego incrementa su velocidad de transmisión una vez que la congestión es reducida.

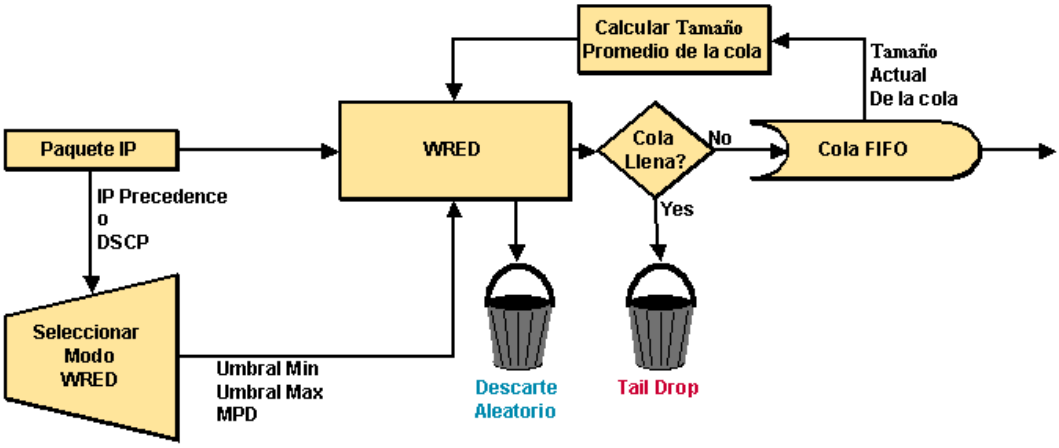


Figura # 28. Bloques Funcionales de WRED.

3.3.4 Flow Based WRED

FRED es una característica de WRED que obliga a WRED a producir una mayor imparcialidad a todos los flujos en una interfaz con respecto a como los paquetes son descartados.

FRED usa la clasificación de paquetes y el estado de la información para asegurar que cada flujo no consuma más buffer del que tiene permitido. FRED determina cuales flujos monopolizan los recursos y le aplica una señalización más dura a esos flujos.

FRED mantiene una cuenta del numero activo de flujos que existen a través de una interfaz de salida. Dado el número de flujos activos y el tamaño de la cola, FRED determina el número de buffers disponible por flujo. Para permitir más aplicación, FRED aumenta el número de buffers por flujos multiplicando por un factor configurado permitiendo a cada flujo activo tener un cierto numero de paquetes en la cola de salida. Este factor de escala es común para todos los flujos. Esto tiene un limite y cuando un flujo excede ese limite tendrá una más alta

probabilidad de que sus paquetes sean descartados. En la figura 29 se muestra el funcionamiento de FRED.

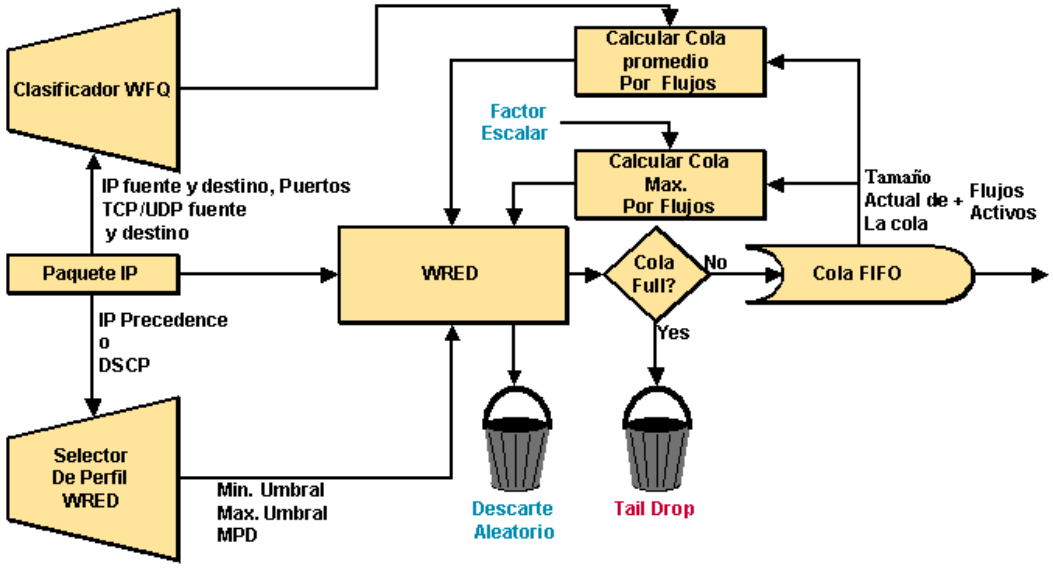


Figura # 29. Bloques Funcionales de FRED.

3.4 Condicionamiento del Trafico

El condicionamiento del trafico ocurre en la red. Esta es una técnica de calidad de servicio que permite controlar los flujos de trafico para que se comporten bajo ciertos rangos con el fin de ayudar a evitar congestión.

Existen dos mecanismos principales para hacer condicionamiento de trafico, estos son:

- Policing
- Shaping

Con Policing, la velocidad a la que puede fluir el trafico es limitada sin almacenaje en buffer (buffering). Esto es realizado con el trafico entrante con el fin de controlar que tan rápido es enviado el trafico.

Con Shaping, las ráfagas son reducidas a un flujo más constante de datos. La reducción de ráfagas ayuda a reducir la congestión en el núcleo de la red.

3.4.1 Policing

Policing es el componente de la calidad de servicio que limita el flujo de trafico a una velocidad de bits configurada.

Puede ser aplicado a través de dos técnicas principales:

- CAR
- Policing Basada en Clases

CAR permite especificar una política que determina cuales paquetes podrían ser asignados a cual clase de trafico. La cabecera IP ya provee un mecanismo para hacer eso, llamado ToS. CAR permite la configuración de políticas basadas en la información de la cabecera IP o TCP, tales como: Dirección IP, puerto de aplicación, puerto físico, subinterfaces, etc., para decidir como los bits de la precedencia IP podrían ser marcados. Una vez marcados, se le puede dar un tratamiento apropiado en el backbone para asegurar que paquetes premium obtengan un servicio premium en términos de ancho de banda, control del delay y otros beneficios de QoS.

Otro de los métodos para aplicar políticas que permitan condicionar el trafico en la red es mediante la aplicación de políticas basadas en clases (Class Based Policing).

Las políticas basadas en clases permiten una completa implementación de técnicas compatibles con los DiffServ.

La aplicación de políticas a una clase con una acción configurada (Conforme, Excedida o Violada) permite simular un comportamiento WRED en la clase y reducir la posibilidad de que una multitud de sesiones TCP/UDP sean descartadas durante la presencia de congestión.

Class Based Policing tiene tres funciones básicas:

- Clasificación de paquetes
- Marcado de paquetes
- Gestión de ancho de banda

3.4.2 Shaping

Traffic Shaping es una técnica para controlar el acceso al ancho de banda disponible, con el fin de asegurar que el trafico se comporte de acuerdo a las políticas establecidas, y regular el flujo de trafico para evitar la congestión que puede ocurrir cuando el trafico transmitido excede la velocidad de acceso del otro extremo.

Traffic Shaping permite controlar el trafico saliente de una interfaz con el objetivo de que el flujo se corresponda con la velocidad del extremo remoto. El trafico adherido a una política particular puede ser limitado para cumplir con los requerimientos de velocidades de bajada, eliminando cuellos de botella en topologías con velocidades de datos disparejas.

Existen dos métodos de Shaping, a continuación se explica cada uno de ellos:

Generic Traffic Shaping (GTS): GTS limita el trafico mediante la reducción del flujo del trafico salida para evitar congestión por contener el trafico a una velocidad de bits usando un mecanismo llamado Token Bucket. GTS se aplica a cada interfaz deseada y se puede usar listas de acceso para seleccionar el trafico a limitar. GTS trabaja con una variedad de tecnologías de la capa 2 del modelo OSI, incluyendo Frame Relay, ATM, SMDS y Ethernet.

Sobre subinterfaces Frame Relay, GTS puede ser configurado para adaptarse dinámicamente al ancho de banda disponible mediante la integración de señales BECN. GTS también puede ser configurado sobre una interfaz ATM para responder al RSVP señalizado sobre circuitos virtuales permanentes (PVCs) configurados estáticamente. En la figura 30 se muestra un diagrama de flujo que resume la operación de GTS.

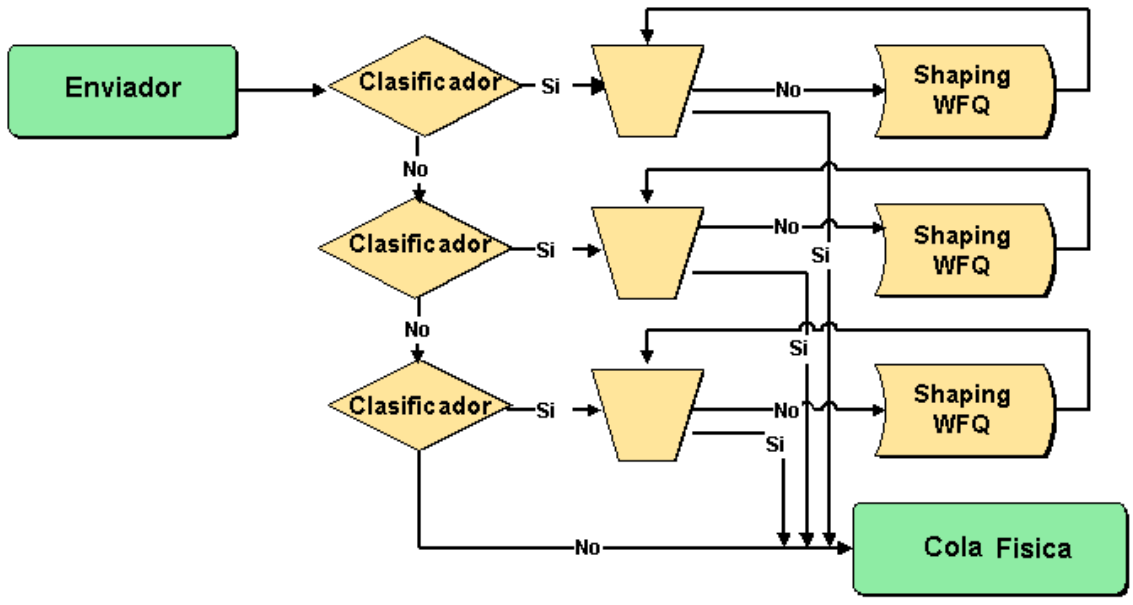


Figura # 30. Operación de GTS.

Frame Relay Traffic Shaping: FRTS provee parámetros que son muy útiles para manejar congestión de trafico en la red. Esto incluye Committed Information Rate (CIR), Forward y Backward Explicit Congestion Notification (FECN/BECN), y el bit Discard Eligibility (DE).

Sin el uso de FRTS, el switch del lado del proveedor de servicio podría tomar decisiones para descartar paquetes que excedan el CIR durante momentos de congestión. Los paquetes en el PVC son considerados iguales y no se le da precedencia a un flujo sobre otros. FRTS habilita a los dispositivos de borde configurar el DE basándose en el flujo de trafico y para asegurar que paquetes con alta precedencia no sean descartados durante congestión. Esta forma de condicionar el trafico es mostrada en la figura 31.

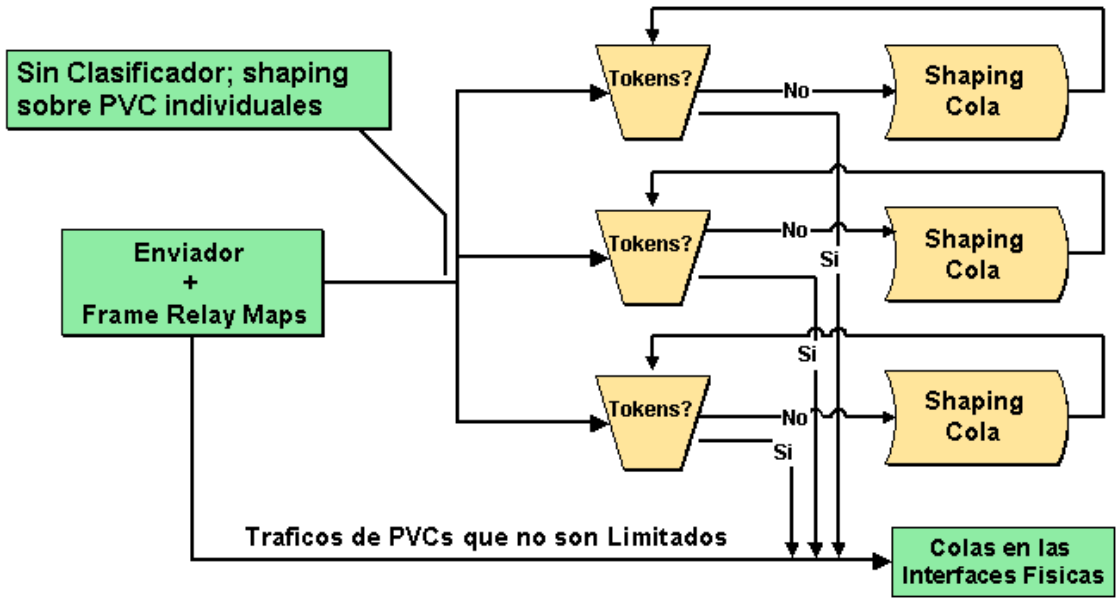


Figura # 31. Frame Relay Traffic Shaping

3.5 Mecanismos de Eficiencia de Enlace

El retardo de serialización es el tiempo que se tarda un dispositivo en colocar bits en un circuito. La velocidad de circuito más alta tiene menos retardo de serialización. Por ejemplo toma aproximadamente 5 microsegundos enviar un paquete de 1024 bytes sobre una línea T1 (1.544 Mbps). El tipo de compresión usada tiene un efecto importante sobre la serialización. Por ejemplo, si paquetes de Voz son comprimidos a tamaños de 8 Kbps (usando G.729), el retardo resultante podría estar alrededor de 20 microsegundos.

Existen retardos de encolamiento inaceptables para pequeños paquetes de tiempo real, sin importar la característica de QoS utilizada o el algoritmo de compresión. Aun con las mejores técnicas de compresión y QoS existirá un retardo en tiempo real por que el overhead de los paquetes es muy largo y una larga unidad máxima de transmisión (MTU) es necesitada para producir una transmisión eficiente.

Para permitir una transferencia de información optima en la red es necesario la utilización de técnicas de eficiencia de enlace tales como: Payload Compression, TCP Header Compression, RTP Compression, LIF, MLP, FRF.11, FRF.12 o CRPT. Ellas permiten proveer eficiencia a enlaces digitales de baja velocidad sin importar el modelo de QoS.

3.5.1 Payload Compression

La Calidad de Servicio requiere ancho de banda. La compresión de la carga útil permite aumentar el ancho de banda disponible, debido a que utiliza algoritmos para comprimir las tramas de capa 2. Mediante esta técnica se logra incrementar la transferencia de información y decrementa el retardo percibido.

Características:

- La compresión reduce el tamaño de la carga útil de la trama
- El paquete IP es comprimido completamente
- La compresión adiciona retardo de acuerdo a su complejidad
- El retardo de serialización es reducido y la latencia total es disminuida

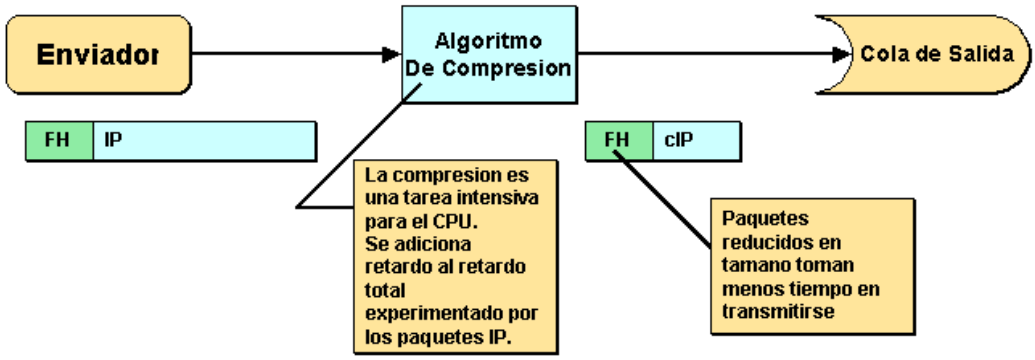


Figura # 32. Payload Compression.

La compresión produce dos beneficios principales, como se muestra a continuación:

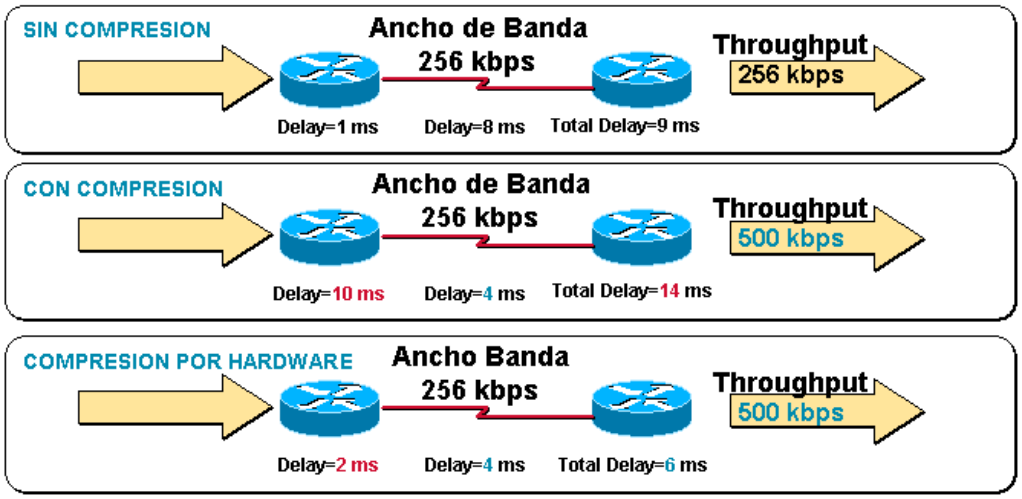


Figura # 33. Efectos de la Compresión.

3.5.2 Header Compression

La compresión del Header reduce el overhead mediante la compresión de paquetes IP y segmentos TCP/UDP. TCP Header Compression comprime la cabecera IP y el segmento TCP, mientras que RTP Header Compression comprime las cabeceras IP, UDP y RTP.

La compresión del Header es especialmente eficiente en enlaces de baja velocidad con trafico interactivo o sensible al retardo.

Características de los Algoritmos para Comprimir Cabecera:

TCP Header Compression (RFC 1144):

- Usado para comprimir la cabecera de segmentos TCP
- Más efectivos sobre enlaces de baja velocidad con muchas sesiones TCP

RTP Header Compression (RFC 1889):

- Usado para reducir el retardo e incrementar el throughput
- Mejora la calidad de la Voz
- Más efectivo sobre enlaces de baja velocidad

3.5.3 Link Fragmentation e Interleaving

LIF (Link Fragmentation and Interleaving) es usado en enlaces lentos para asegurar que los paquetes de Voz puedan ser enviados sin un gran retardo. Mientras que un paquete de datos es serializado (enviado bit a bit sobre un enlace serial), el paquete de Voz debe esperar. Paquetes de datos largos puede significar un tiempo relativamente alto esperando sobre un enlace WAN. Mediante la

fragmentación de largos paquetes de datos, LIF permite a los paquetes de Voz ser enviados más pronto, reduciendo la latencia total.

En la figura 34 se muestra de manera grafica lo expresado anteriormente.

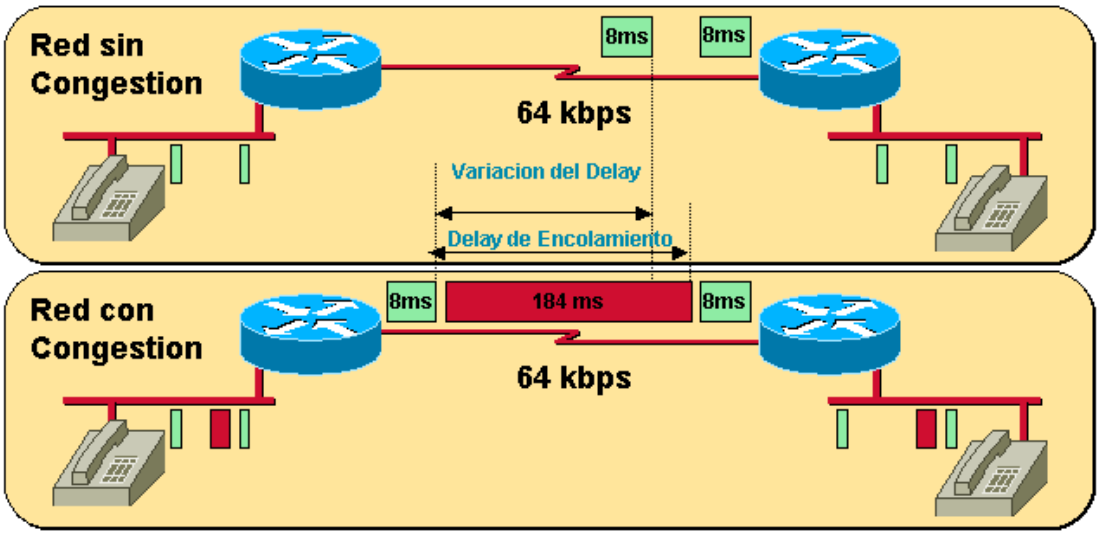


Figura # 34. Retardo de Encolamiento y Serialización

Como se menciono anteriormente el retardo es el tiempo que le toma a un paquete alcanzar el punto final después de ser transmitido desde el otro extremo. Este tiempo es llamado retardo de extremo a extremo, y esta constituido por dos componentes: Retardo fijo y retardo variable.

Basado en medios WAN, existen varias técnicas de LFI tales como: MPL Interleaving, FRF.12, FRF.11.

MPL Interleaving, permite que los paquetes grandes sean encapsulados en múltiples enlaces y fragmentados en tamaños de pequeñas cantidades para satisfacer el retardo requerido para el trafico de tiempo real.

La característica de interleaving también provee una cola de transmisión especial para los paquetes más pequeños sensibles al retardo, habilitándolos para que sean transmitidos más pronto que otros flujos.

FRF.12 (Frame Relay Fragmentation), es una implementación definida para soportar Voz, Video y otros trafico sensibles al retardo sobre enlaces de baja velocidad. FRF.12 acomoda variaciones en tamaños de trama de manera que se permita manejar trafico de tiempo real o no.

FRF.12 estipula que cuando la fragmentación está operando sobre un DLCI, solo las tramas que excedan el tamaño especificado son fragmentadas. Este arreglo permite a los paquetes pequeños (los cuales no son fragmentados por su tamaño) ser intercalados como tramas entre paquetes grandes que han sido fragmentados en tramas más pequeñas. Esto mejora el retardo de serialización ya que permite intercalar los paquetes y prevenir que paquetes sensibles al retardo tengan que esperar mientras los paquetes grandes sean procesados.

FRF.11 es una especificación que describe la forma en que sen transmiten los datos y la Voz sobre un mismo DLCI. Solo las tramas de datos son fragmentadas. Frame Relay distingue tramas de Voz de tramas de datos debido a que FRF.11 especifica el tipo de trama.

IV. TRANSMISIÓN DE VOZ Y VIDEO SOBRE REDES DE PAQUETES

El Protocolo de Internet (IP) es un protocolo general para el envío de información digital empaquetada entre interfaces identificadas por direcciones IP. En sí, el protocolo es transparente para la información transportada en los paquetes. Voz sobre IP se refiere al uso del Protocolo de Internet entre aplicaciones que manejan señales para la transmisión en tiempo real de información de voz sobre una red IP.

En el punto transmisor la información de voz es codificada en una representación digital adecuada, que es dividida en paquetes y enviada a la dirección IP del destinatario. En el lado receptor, la información es desempaquetada y decodificada hasta recuperar la señal de voz. Para reducir la necesidad de ancho de banda en la red, algoritmos de compresión son generalmente usados como parte de la codificación y decodificación.

La transmisión de Video sobre redes de paquetes se realizo por mucho tiempo mediante la utilización de tecnologías como: ISDN, Frame Relay y ATM ya que estas presentan características favorables para el transporte de servicios de Video.

Con la popularidad del protocolo IP y los avances en las técnicas de calidad de servicio, actualmente se utilizan redes IP para el transporte de Video de manera muy eficiente. Existen en el mercado tecnologías denominadas TV sobre IP y Video conferencia sobre IP que funcionan de manera más sencilla que otras tecnologías tradicionales; esto también permite la unificación de tecnologías lo cual se traduce en mejor gestión de los servicios de información.

4.1 Estándares y Protocolos: H.323.

El estándar H.323 especifica los componentes, protocolos y procedimientos que proveen los servicios de comunicaciones multimedia (audio en tiempo real, video y datos) sobre redes de paquetes, incluyendo el protocolo IP.

H.323 es parte de las recomendaciones de la ITU llamadas H.32X que suministran servicios de telecomunicaciones multimedia sobre una gran variedad de redes.

Las redes de paquetes incluyen redes IP, IPX, redes de área local (LAN), redes empresariales, redes de área metropolitana (MAN) y redes de área amplia (WAN). Además, H.323 suministra los mecanismos que pueden ser utilizados en telefonía IP, video, telefonía, audio y datos.

También soporta comunicaciones del tipo multimedia multipunto, por lo que puede ser aplicado a una gran variedad de escenarios.

H.323 es un estándar ITU-U para videoconferencias multimedia sobre una red de paquetes conmutados. La recomendación H.323 Versión 1 está formalmente titulada "*Visual Telephone Systems and Equipment for Local Networks Which Provide a Non-Guaranteed Quality of Service*". Desde entonces la voz es un elemento de multimedia; el estándar cubre la mayoría de las cosas que se deben hacer para manejar voz sobre redes IP corporativas y la Internet.

En particular, el estándar H.323 define las siguientes funciones:

- Cómo los puntos terminales hacen y reciben llamadas.
- Cómo los puntos terminales negocian un conjunto de capacidades de audio, video y datos.

- Cómo la información de audio y video es estructurada y enviada sobre la red.
- Cómo el audio y el video son sincronizados.
- Cómo los puntos terminales se comunican con sus respectivos *gatekeepeers,* los cuales proveen control de admisión de llamadas (CAC) en la red.

El estándar H.323 es un estándar que hace referencia a otras recomendaciones.

A continuación se mencionan los estándares más importantes:

- H.323 – Sistema de documentación.
- H.225 – Señalización de llamada, paquetización, y el protocolo de registro de *gatekeeper*, admisión y estados (RAS).
- H.245 – Protocolo de control.
- T.120 – Control de datos y conferencia.
- RTP – Protocolo de transporte de tiempo real (IETF).
- RTCP – Protocolo de transporte de tiempo real (IETF).

H.323 usa un conjunto compatible de recomendaciones H.320; asi los puntos terminales de H.323 todavía pueden comunicarse con antiguas estaciones basadas en ISDN y sistemas de video conferencia.

Las recomendaciones H.323 han resultado de una colaboración sin precedentes entre las empresas de telecomunicaciones y las empresas de computación. Esto ha permitido que H.323 se haya desarrollado rápidamente.

Ya han sido publicadas dos recomendaciones:

- **H.323v1:** La primera recomendación ITU-T H.323**,** publicada en Mayo de 1996, se titula: "*Visual Telephone Systems and Equipment for Local Networks Which Provide a Non-Guaranteed Quality of Service*".

- **H.323v2:** La segunda recomendación ITU-T H.323**,** publicada en Enero de 1998, se titula: "*Packet Based Multimedia Communications Systems*".

El estándar fue especificado por el grupo de estudios de la ITU-T; la primera versión de la recomendación solo tomaba en cuenta sistema de telefonía visual (video conferencia) y sus correspondientes equipos de red local que no manejaban calidad de servicio (QoS).

El surgimiento de las aplicaciones de voz sobre IP y la telefonía IP ha obligado a la revisión de las especificaciones. La ausencia de un estándar para la voz sobre IP resultó en productos que fueron incompatibles entre sí. Con el desarrollo de VoIP nuevos requerimientos surgieron, tales como la comunicación entre PC y sistemas telefónicos convencionales a través de las redes de circuitos conmutados. Dichos requerimientos forzaron la necesidad para un estándar de la telefonía IP; la versión 2.0 de H.323 fue definida para acomodar esos requerimientos adicionales y fue aceptada en Enero de 1998.

Nuevas características son incluidas al estándar H.323 las cuales lo conllevan a la versión 3.0. Estas incluyen Fax sobre redes de paquete, comunicaciones entre Gatekeepers y mecanismos de conexión rápida.

H.323 se relaciona con otros estándares a través de las siguientes recomendaciones:

- H.342 sobre SCN

- H.320 sobre ISDN
- H.321 y H310 sobre B-ISDN
- H.322 sobre redes de área local (Lan) con calidad de servicio.

4.1.1 Componentes H.323

El estándar H.323 especifica cuatro clases de componentes, los cuales son integrados en una red, suministrando servicios de telecomunicaciones multimedia, punto a punto y punto-multipunto.

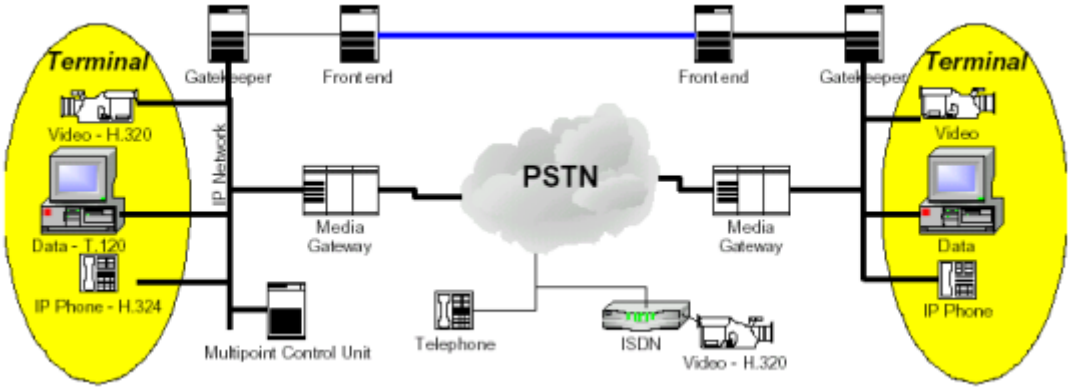


Figura # 35. Componentes de una Red H.323

Los componentes son:

1. Terminales.
2. Gateways.
3. Gatekeepers.
4. MCU's (Unidades de Control Multipunto).

Terminales

Son utilizados para comunicaciones bidireccionales en tiempo real. Un terminal H323 puede ser un dispositivo "standalone" o un PC corriendo una aplicación H323. Estos soportan comunicaciones de audio y pueden soportar opcionalmente video y datos. Debido al servicio básico que suministran, juegan un papel muy importante en los servicios de telefonía IP.

La función principal de H323, es la de interoperabilidad, entre terminales multimedia.

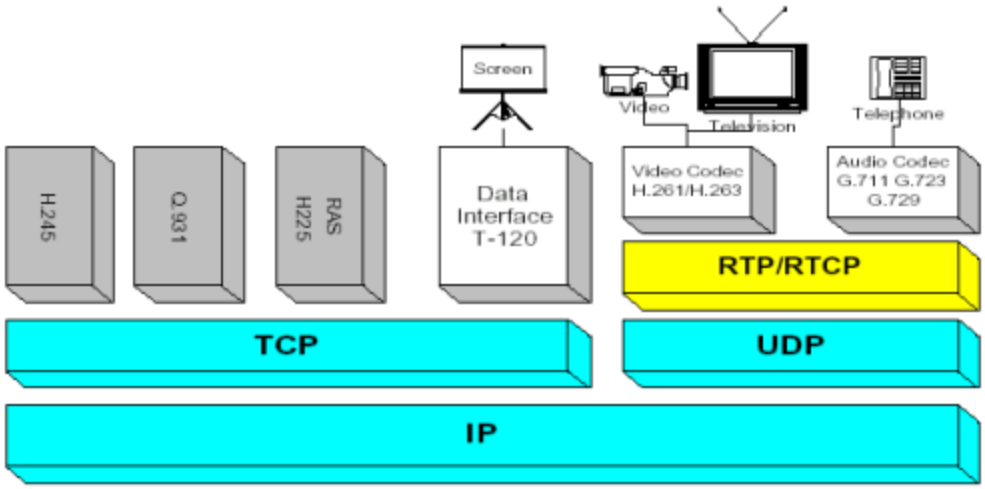


Figura # 36. Modelo de Software H.323.

Gateways

Un gateway o puente de enlace, conecta dos redes distintas. Un gateway H.323 suministra la conectividad entre una red H.323 y una red que no es H.323. Por ejemplo un gateway puede conectar y proveer comunicación entre un terminal

H.323 y una red de teléfonos conmutados (una red de tipo PSTN). Esta conectividad de diferentes redes es lograda a través de protocolos de traducción para la configuración y liberación de las llamadas (Call Setup and Release).

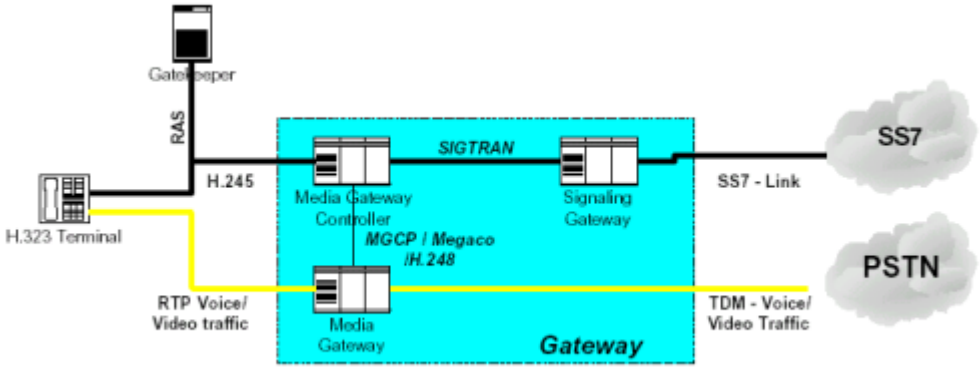


Figura # 37. Gateways H.323

Gatekeepers

Es considerado el cerebro de una red H.323. Este es el punto focal para todas las llamadas H323. Los gatekeepers suministran servicios importantes tales como direccionamiento, autorización y autentificación de terminales y gateways, administración de ancho de banda, tarificación y servicio de enrutamiento de llamadas.

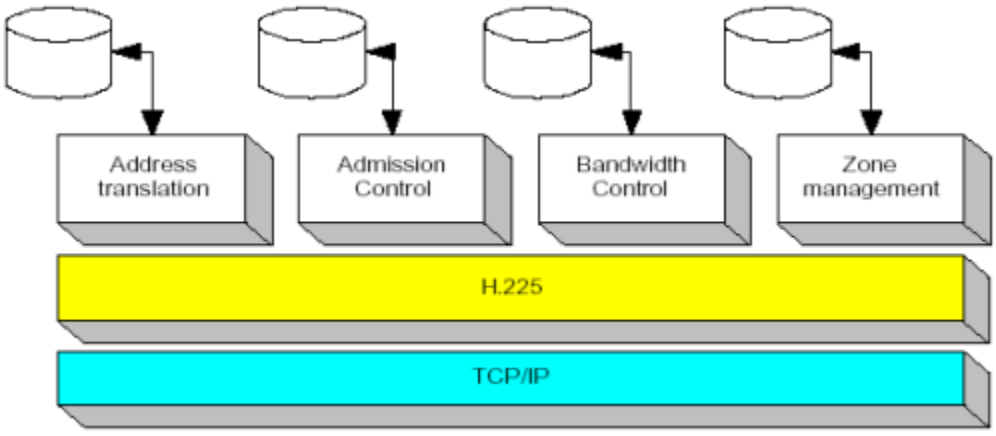


Figura # 38. Módulos de un Gatekeeper.

Unidades de Control Multipunto (MCU's)

Las unidades de control multipunto suministran soporte para conferencias de tres o más terminales H.323. Todos los terminales participan en conferencias establecidas a partir de una conexión con el MCU.

El MCU maneja los recursos de conferencia y negociación entre terminales con el propósito de determinar los CODEC's que van a ser utilizados y el manejo de la cadena (string) de datos.

4.2 Session Initiation Protocol: SIP

Las redes de VoIP que no son H.323 son actualmente desarrolladas para soportar servicios de voz. Una tecnología con una popularidad que se incrementa cada día mas es SIP, el cual fue estandarizado por la IETF en marzo de 1999, como se definió en el RFC 2543 y su actualización en junio de 2002, en el RFC 3261. SIP esta madurando rápidamente en términos de características, estabilidad e interoperatividad entre fabricantes. Ciertas características de H.323 no existen hoy en día en productos SIP; sin embargo SIP esta avanzando mucho para soportar todas las características de H.323.

Similar a H.323, SIP usa un protocolo peer-to-peer que resulta en la inteligencia existente en el dispositivo final. Esto significa que cada dispositivo final llamado SIP User Agent (UAs), puede tener varios niveles de inteligencia. Sin embargo en muchos casos es necesario un Proxy Server o un Softswitch para facilitar la sesión de VoIP.

4.2.1 Componentes de la Arquitectura de Red SIP

El modelo SIP usa dos componentes, clientes y servidores, para describir el protocolo de señalización. Dos categorías de componentes SIP son User Agente (UAs) e intermediarios. UAs son puntos finales que inician o terminan diálogos y transacciones SIP, sin importar si los intermediarios son servidores, tales como Proxy y servidores de redireccionamiento, que ayudan a enrutar las llamadas y proveer algunos servicios de valor agregado, tales como seguridad, alrededor de la ruta.

Un cliente SIP se refiere simplemente a la entidad que origina un requerimiento, y un servidor SIP es la entidad que recibe la solicitud. Un UAs de SIP y un Proxy de SIP consisten ambos de un cliente SIP y de un servidor SIP. En la figura numero 39 se muestra la arquitectura de una red SIP.

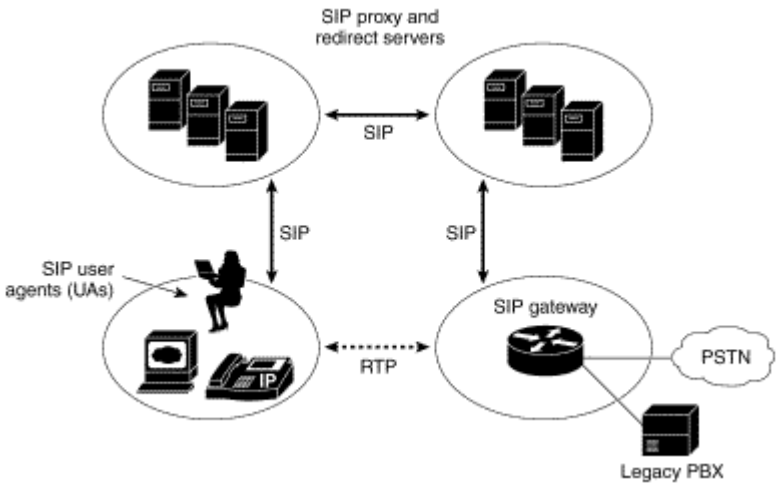


Figura # 39. Arquitectura de Red SIP

SIP User Agents: SIP es un protocolo peer-to-peer que puede establecer y desconectar llamadas entre puntos finales SIP. Estos puntos finales son llamados User Agents. Hay dos tipos de UAs: User Agent Client (UAC) y User Agent Server (UAS). User Agent de SIP origina y termina requerimientos SIP. Estos clientes tienen ambos un UAC y un UAS para originar y terminar solicitudes SIP. Un User Agent de SIP puede tener varias formas, incluyendo las siguientes o cualquier dispositivo que tenga conectividad a Internet y soporte UA de SIP:

- Teléfonos SIP
- Clientes PC basados en SIP (Softphones)
- Gateways VoIP SIP
- PDAs basadas en SIP
- Dispositivos inalámbricos basados en SIP

SIP Gateways: Los gateways en la red SIP desempeñan la misma función que los gateways en una red H.323. Para que un gateway se comunique con un servidor SIP, los dial peers en el gateway deben ser configurados para usar SIP.

SIP Servers: Cada teléfono SIP en la red debe registrarse con un servidor de registro para notificarle al servidor la dirección donde este puede ser alcanzado, esto se realiza mediante el envío de un mensaje SIP Register al Servidor. Después del registro, un cliente SIP puede enviar un requerimiento SIP INVITE a un Servidor SIP. El requerimiento INVITE establece la llamada SIP. El servidor puede ser o un SIP Proxy Server o un SIP Redirect Server. Existen tres tipos de servidores SIP:

- Proxy Server: Desempeña el enrutamiento de los mensajes SIP para controlar el enrutamiento de la llamada, un Proxy Server desempeña otras funciones vitales tales como autenticación, autorización y determinación del próximo salto en la ruta.

- Redirect Server: Provee a los clientes SIP información sobre la localización del próximo salto para la terminación del punto final SIP. Un Redirect Server provee resolución de direcciones similar a un gatekeeper.
- Registar Server: Los clientes SIP registran su localización con el Registar Server usando su dirección (única) SIP. Esta dirección SIP puede ser un numero E.164 u otra, por ejemplo: sip:177055512@gateway.com; user=phone.

Mensajes de Señalización SIP: SIP es un protocolo de señalización que ha sido definido por la IETF para soportar sesiones basadas en Internet que requieran voz o multimedia. SIP es un protocolo punto a punto que usa mensajes de texto codificados, en vez de los mensajes binarios usados por el protocolo H.323. SIP sirve para aplicaciones en la Internet, tales como: Instant Messaging y VoIP.

Una de las ventajas de SIP es que puede usar protocolos basados en Internet para complementar su protocolo de señalización. Debido a que SIP especifica solo como las sesiones son creadas, modificadas y desconectadas, las funcionalidades adicionales son soportadas por otros protocolos definidos por la IETF.

4.3 MGCP

MGCP (Media Gateway Control Protocol), describe un modelo que usa una visión abstracta de recursos lógicos y físicos encontrados en una red de VoIP. Dos entidades abstractas son definidas: puntos finales y conexiones. Un punto final es un recurso lógico que existe en un gateway que es requerido para soportar servicios de voz y datos. Ejemplos de puntos finales son un canal DS0, una línea analógica, un servidor de anuncio, o un circuito virtual de ATM. Un punto final puede ser asociado con uno o más entidades externas, tales como un circuito físico PRI. Una relación debe existir entre puntos finales para completar una llamada. La entidad de conexión provee esta función. Una conexión crea una relación entre

dos o más puntos finales para habilitar una llamada de voz o datos a ser creada en el modelo MGCP.

MGCP usa un modelo maestro/esclavo que define el gateway para transportar un conjunto limitado de transacciones iniciadas por un MGC Call Agent. Este modelo maestro/esclavo es similar a la arquitectura tradicional de conmutación de circuitos TDM donde toda la inteligencia de control de llamada ha sido desarrollada en una manera centralizada, similar a un Switch clase 4 o clase 5. Similar a los Switches clase 4/5, MGCP permite el control centralizado de llamadas. Sin embargo, a diferencia de los Switches, MGCP provee un estándar abierto para una clara separación entre la entidad centralizada de control de llamadas y las entidades que terminan las troncales y líneas.

MGCP es un protocolo ampliamente desarrollado por la IETF (RFC 2705bis) que habilita a los Softswitches para controlar gateways en una red VoIP. Por esto el control inteligente de llamadas reside en el Softswitch.

4.4 IP/VC

La Video Conferencia sobre IP es una solución que está basada en el estándar H.323. Esta solución provee la capacidad de transmitir Video sobre una red IP.

Como se indicó en el punto 4.1, H.323 es un estándar ITU-U para videoconferencias multimedia sobre una red de paquetes conmutados. La recomendación H.323 Versión 1 está formalmente titulada "*Visual Telephone Systems and Equipment for Local Networks Which Provide a Non-Guaranteed Quality of Service*". Desde entonces la voz es un elemento de multimedia, el estándar cubre la mayoría de las cosas que se deben hacer para manejar voz sobre redes IP corporativas y la Internet.

En particular, el estándar H.323 define las siguientes funciones:

- Cómo los puntos terminales hacen y reciben llamadas o sesion de video.
- Cómo los puntos terminales negocian un conjunto de capacidades de audio, video y datos.

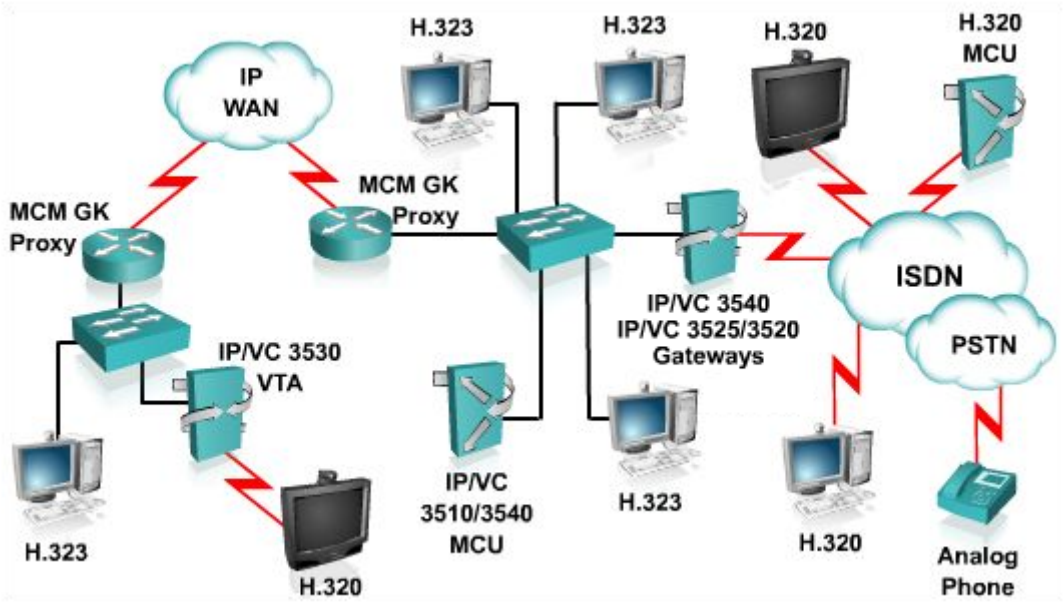


Figura # 40. Arquitectura de Red IP/VC

Como se muestra en la figura numero 40, los componentes de esta arquitectura son los siguientes:

- MCU
- Gateways
- Gatekeepers
- Terminales H.323
- Terminales H.320
- Terminales Analógicos

Los cuales cumplen la misma función indicada al inicio de este capitulo.

4.5 IP/TV

La solución de IP/TV es un sistema completo para la transmisión de video en vivo con calidad de TV, incluyendo broadcasts, programas de entrenamiento, clases universitarias, negocios, tecnologías satelitales y otras aplicaciones, las cuales pueden ser implementadas desde PCs, salas de clases o cuartos de reuniones.

Esta solución utiliza protocolos estándares existentes en redes IP, además de tecnologías adoptadas por la mayoría de fabricantes tales como H.323, MPEG-X y Microsoft Media Technologies.

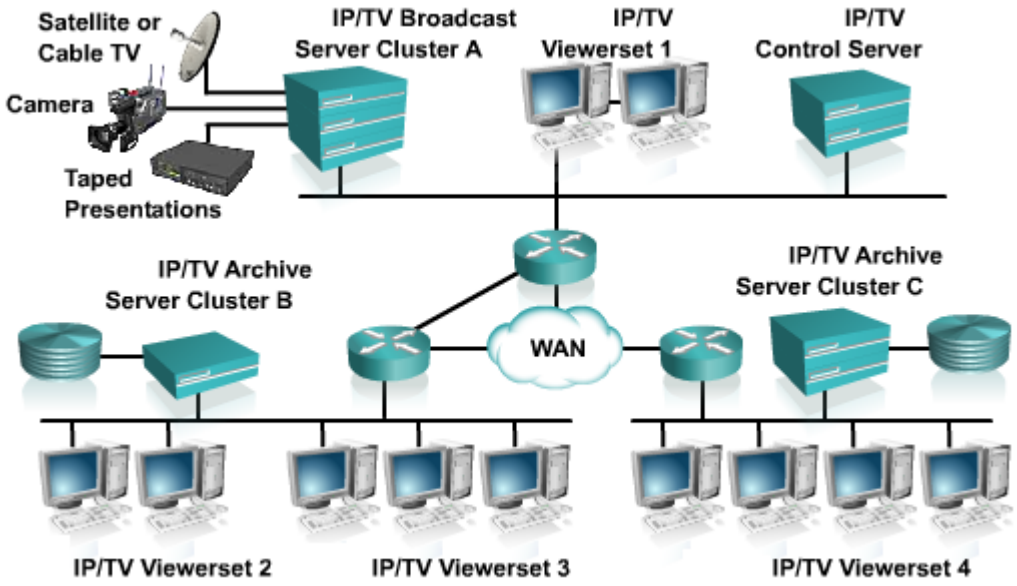


Figura # 41. Componentes Red IP/TV

Como se muestra en la figura numero 41, los componentes de una solución IP/TV son:

- IP/TV Broadcast Server
- IP/TV Archive Server
- IP/TV Control Server
- IP/TV Viewer

El IP/TV Broadcast Server se encarga de la transmisión de programas de acuerdo a las direcciones que recibe del Control Server. En el caso de Cisco, soporta dos formatos, uno para bajos anchos de banda tales como Microsoft Windows Media y para formato MPEG. Los Broadcasts Server son usados primariamente para multicastiing en vivo o para programas pregrabados desde dispositivos como cámaras de video, VCRs, DVDs, satélites, MP3 y archivos MPEG.

El IP/TV Control Server es el administrador de políticas. Este comunica la información programada, los formatos de video disponibles, las consideraciones de ancho de banda y otras mas al Broadcast Server, también la información de la programación a los IP/TV Client Viewer. El IP/TV Control Server realiza balanceo de carga de video en la red de manera automática, sin requerir de intervención adicional del administrador de la red.

Para requerimientos de broadcast programadas, se puede usar un Archive Server, el cual se ubica en el borde de la red para entregar video pregrabado o video sobre demanda.

V. CALIDAD DE SERVICIO EN ARQUITECTURAS CISCO

La calidad de servicio implantada sobre arquitecturas de Cisco Systems responde a un conjunto de tecnologías, herramientas y equipos de comunicaciones que unidos permiten la aplicación de calidad de servicio de manera integrada. En la figura 42 se muestra un diagrama que agrupa los elementos que constituyen una solución de Cisco para el manejo de la calidad de servicio.

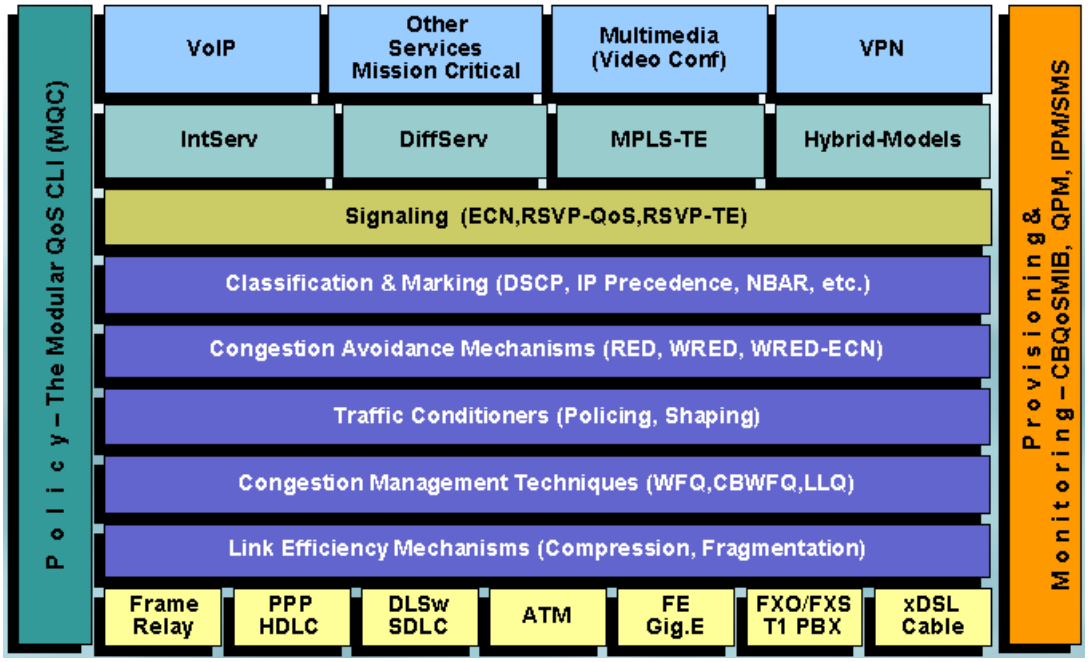


Figura # 42. Solución de Calidad de Servicio de Cisco Systems.

Esta solución es llamada Cisco AVVID (Arquitectura de Voz, Video y Datos Integrados de Cisco). El Cisco AVVID es una implementación en capas funcionales que asegura la calidad de servicio sobre cualquier tecnología de capa 2 (Frame Relay, ATM, xDSL, Ethernet, etc.).

Cisco AVVID utiliza mecanismos de eficiencia de enlace como compresión y fragmentación para dar un tamaño máximo a las tramas que serán transmitidas, esto con el fin de mejorar la eficiencia de enlaces de baja velocidad.

En la capa siguiente se muestran las técnicas para el manejo de la congestión, entre las cuales se encuentra WFQ, CBWFQ y LLQ. Estas técnicas permiten un manejo de las colas de transmisión para que cada servicio (Voz, Video y Datos) utilice la cola que mejor desempeño ofrezca según sus requerimientos. Como se explico en capítulos anteriores, el manejo de colas permite un control del trafico cuando exista congestión en la red y asegura que los servicios sensibles al retardo tengan un tratamiento preferencial en esos momentos.

Cisco AVVID toma en consideración los métodos para evitar congestión ya que estos reducen la probabilidad de que los paquetes de Voz y Video sean descartados durante periodos de congestión. La estrategia para evitar congestión utilizada por Cisco incluye RED y WRED.

Con el fin de facilitar la diferenciación de los flujos de información, Cisco plantea la utilización de técnicas de clasificación y marcado tales como NBAR, DSCP y Precedencia IP. La clasificación y el marcado es un punto muy importante como estrategia de calidad de servicio ya que permite diferenciar flujos sensibles de flujos no sensibles al retardo, Jitter y/o perdida de paquetes. Esto le indica a los dispositivos de red que deben darle un tratamiento diferenciado a cada flujo de información.

Para que la estrategia de QoS funcione de manera optima es necesario utilizar una señalización que permita indicarle a los equipos de comunicaciones que deben reservar anchos de banda y retardos mínimos a ciertos flujos de información. La señalización también permite verificar la existencia o no de los recursos necesarios

para el establecimiento de una conexión en la red. Cisco AVVID plantea la utilización de RSVP como método de señalización.

La implementación de calidad de servicio de Cisco responde a uno o varios modelos de servicio (IntServ, DiffServ). El uso de uno u otro dependerá de la estrategia de calidad diseñada para satisfacer los requerimientos de los diferentes tipos de tráficos que serán transmitidos en la red. En muchos casos se plantean combinaciones de modelos de servicios Integrados y Diferenciados en una misma infraestructura.

Todas estas capas permiten la transmisión de VoIP, VoFR, VoATM, Video y datos sobre la misma infraestructura de comunicaciones. Para tener un control de las políticas de calidad de servicio y del comportamiento de los dispositivos es necesario la utilización de herramientas de monitoreo y aprovisionamiento de QoS. Cisco AVVID plantea el uso herramientas propietarias para este fin, tales como Cisco QPM y MQC.

5.1. Implementación de QoS en Equipos de Comunicaciones de Cisco

La arquitectura para Voz, Video y Datos Integrados esta constituida por cuatro capas (como se muestra en la figura 42). Cada una de las capas tiene una función dentro de la arquitectura. La capa "A" es la que representa a todos los equipos de acceso (Routers, Switches) y servicios Wan (Frame Relay, ATM xDSL). La capa "B" esta constituida por los equipos que permiten la comunicación entre redes diferentes (por ejemplo, PSTN, Intranets). En la capa "C" se encuentran los dispositivos usados por el usuario, tales como SoftPhone, Teléfonos IP, Gestor de Llamadas y dispositivos de Video. La capa "D" representa todos los servidores para las aplicaciones que manejan los servicios en convergencia tales como: Servidores de Video, de Aplicaciones y mensajeria unificada.

Para que estos componentes funcionen de manera adecuada es necesario la implementación de estrategias de calidad de servicio que garanticen los recursos y un tratamiento adecuado a cada servicio soportado por la red.

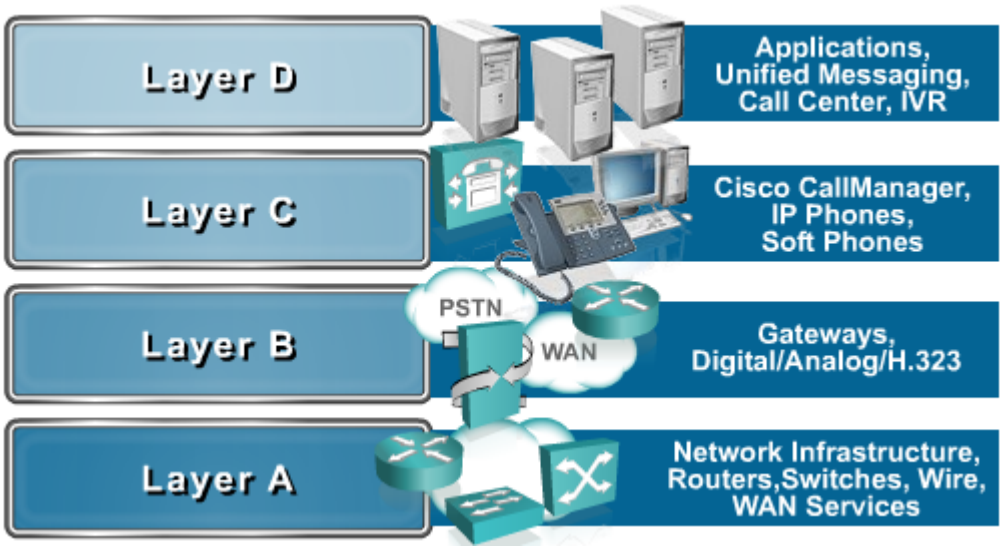


Figura # 42. Componentes de Cisco AVVID.

5.1.1. Capa "A". Infraestructura.

En la capa de Infraestructura se maneja todo lo que tiene que ver con conectividad, sistemas de cableado y servicios Wan. Los equipos de Cisco utilizados en esta capa son los siguientes:

- Routers
- Switches Capa 2 y 3
- Sistemas de Cableado Estructurado. Normas 568 A, 569
- Enlaces Digitales tales como: Frame Relay, ATM, HDLC, xDSL

Estos elementos cumplen funciones distintas y son necesarias para el funcionamiento de la red.

Los Routers se encargan de realizar la interconexión de localidades remotas y de realizar el proceso de selección de la mejor ruta hacia un destino. Algunos de estos routers son llamados Voice Enabled Router (Router Habilitado para Voz) ya que también tienen la capacidad de manejar VoFR, VoIP o VoATM.

Los Switches son elementos de red que operan en la capa dos del modelo OSI, su función principal es la de permitir el acceso a los usuarios de la red. Los switches son considerados la puerta de entrada a la red y entre sus funciones está la de establecer comunicación entre los usuarios de manera local. A diferencia de los Routers, tienen una mayor densidad de puertos y menor inteligencia. Los Switches llamados capa 3 son equipos que realizan funciones de capa 2 y capa 3 (enrutamiento), estos dispositivos permite la comunicación de redes Lan virtuales y en algunos casos pueden cumplir la función del Router. Una de las ventajas de los Switches capa 3 sobre los Routers es que el enrutamiento es realizado en hardware, mientras que en los Routers es por Software. A lo Switches no solo se conectan las estaciones de trabajo, sino también se pueden conectar teléfonos IP, dispositivos de Video, y otros necesarios para la arquitectura integrada.

Los Sistemas de Cableado Estructurado son una parte muy importante dentro de la infraestructura. A través de ellos fluye la información de manera digital de un punto a otro y de manera transparente para los usuarios. Para que una solución de integración de servicios funciones es imperativo contar con una infraestructura de cableado certificada que garantice la transmisión de información de manera eficiente. Si el cableado no está en condiciones adecuadas, entonces los servicios no operaran de manera optima.

Los enlaces digitales son sistemas de telecomunicaciones (generalmente públicos) que se utilizan para la transferencia de información entre localidades remotas. En la mayoría de los casos los enlaces son de baja velocidad lo cual limita la cantidad y el tipo de información que se puede transmitir por un enlace determinado, el uso

de mayores anchos de banda depende en la mayoría de los casos de elementos económicos y no tanto técnicos; esto es debido a que el costo de los enlaces son más elevados de forma proporcional a la velocidad de acceso contratada. En tal sentido la utilización de un ancho de banda debe responder a un requerimiento técnico pero se debe contar con los recursos económicos necesarios para tal fin. Los enlaces digitales entregados por los proveedores de Telecom permiten la transmisión de Voz, Video y Datos, entre localidades remotas. La cantidad de sesiones de Video o de llamadas telefónicas va ha depender del ancho de banda disponible, es por eso que las técnicas de eficiencia de enlace son tan importantes.

5.1.2. Capa "B". Gateways

En el Cisco AVVID, los Gateways son utilizados para el manejo de la comunicación con redes diferentes, (por ejemplo, comunicación con la PSTN). Estos equipos pueden manejar un conjunto diverso de interfaces para comunicarse con PBX, Switches telefónicos, Equipos de Video, Teléfonos analógicos y digitales, etc. Por tanto su utilización dentro de una red convergente es prácticamente obligada. En muchos casos los Routers habilitados para voz se comportan como Gateway.

La infraestructura para la comunicación entre redes remotas es generalmente soportada sobre tecnologías como Frame Relay y ATM, la transmisión de información es realizada por conmutación de paquetes en la cual no se desperdicia el ancho de banda. Estos enlaces actualmente son utilizados como canal de comunicaciones principal para la transferencia de información entre localidades. La PSTN es la red utilizada para comunicaciones de Voz, además de ser esa su función principal también es posible la transferencia de datos a baja velocidad y en muchos casos para enlaces de respaldo.

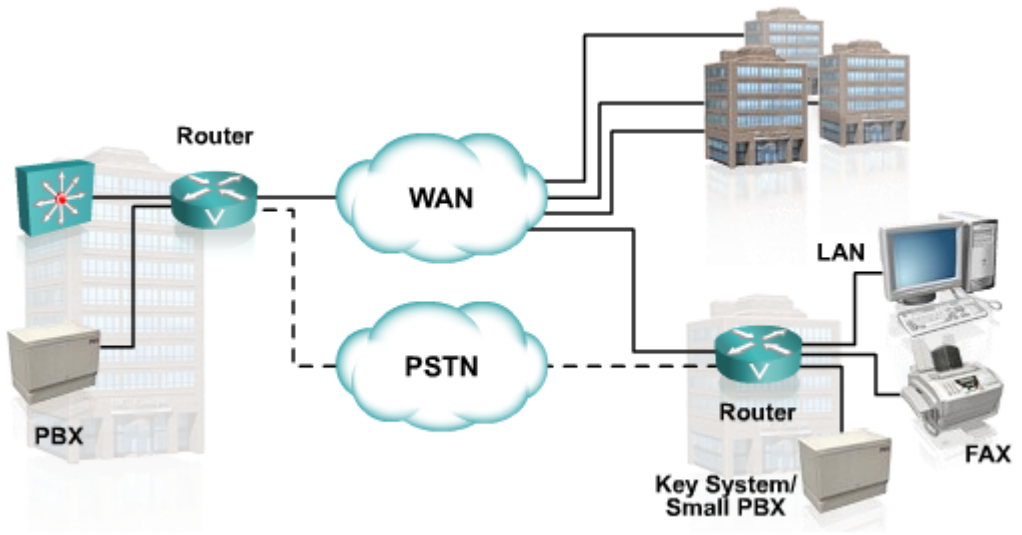


Figura # 43. Infraestructura de Red con Conexión a la PSTN.

5.1.3. Capa "C". Dispositivos de Acceso

Los dispositivos de acceso utilizados en el Cisco AVVID están constituidos por los siguientes elementos:

- Computadores Personales
- Teléfonos IP
- Teléfonos por Software (SoftPhone)
- Cámara de Video

El PC es el dispositivo que el usuario utiliza para la realización de sus actividades empresariales y como elemento final dentro de una red local. Estos equipos utilizan sistemas operativos diversos y protocolos de comunicaciones también diversos que deben ser manejados por la infraestructura de comunicaciones.

Los teléfonos IP son equipos para comunicaciones telefónicas mediante la utilización de conmutación de paquetes y son gestionados por un servidor de

telefonía (Cisco Call Manager). La comunicación entre estos dispositivos y la red de acceso es mediante técnicas de señalización como H.323, MGCP o SIP.

Los SoftPhone son teléfonos IP implementados en Software que se ejecutan en computadores Personales, los SoftPhone tienen menos funciones que los teléfonos IP tangibles.

Las cámaras de Video son equipos que se encargan de capturar imágenes para luego ser tratadas por una aplicación y ser transmitidas a través de la infraestructura de comunicaciones. Las cámaras implementan técnicas de codificación y compresión con el fin de lograr la utilización de la menor cantidad de ancho de banda posible. Estos elementos pueden estar instalados en una computadora o funcionando de manara independiente y conectada a la Red.

También existen equipos de Videoconferencia que no solo capturan y transmiten Video sino que también envía los sonidos que se generan en cierto lugar o sala. Todas las informaciones son introducidas a la red y mediante el Cisco AVVID se manipulan de acuerdo a sus características y requerimientos de ancho de banda y delay.

5.1.4. Capa "D". Servidores y Mensajeria Unificada.

La capa D es implementada mediante todos los equipos que se encargan de gestionar las diferentes solicitudes de servicios de información que realizan los clientes. En esta capa se encuentran servidores de aplicaciones, servidores de archivo, servidores de bases de datos, controladores de video, servidores de telefonía y otros equipos que permiten manejar la mensajeria unificada.

Esta capa es una de las más criticas dentro de la arquitectura, debido a que en ella se encuentran todos los servidores que se encargan de manejar los servicios de los

clientes, si alguno de estos servidores dejara de funcionar, entonces ese servicio se vería comprometido; para solventar esta posibilidad de falla se crean arreglos de servidores como clusters, servidores de respaldo, discos espejos, o implementaciones de redes SAN.

VI. GESTION DE LA CALIDAD DE SERVICIO

La calidad de servicio es una parte importante dentro de las estrategias para integrar Voz, Video y Datos; sin una correcta implementación de QoS dentro de una red convergente no será posible la operación eficiente de los diferentes servicios transportados sobre esta. En tal sentido es necesario realizar una buena implementación de técnicas de calidad de servicio y mantener su operación durante el tiempo.

La gestión de la calidad de servicio busca garantizar el correcto comportamiento de las políticas de QoS implementadas en la red y facilitar su aplicación en los dispositivos de comunicaciones. La gestión de la calidad de servicio cumple esta función mediante herramientas sencillas de monitoreo y configuración de las políticas de calidad de servicio definidas en una organización.

Para realizar esta tarea existen diferentes software que se encargan de facilitar a los administradores de red la configuración de políticas de QoS y su posterior monitoreo; esto mediante interfaces graficas y en muchos caso con aplicaciones basadas en Web.

En esta área no existe mucha normalización por lo cual dependiendo del fabricante de los equipos de comunicaciones se tendrán las herramientas de gestión de la calidad de servicio. Sin embargo hay corporaciones que se dedican a realizar software para la gestión de redes mediante diferentes protocolos de gestión como SNMP y otros. En el caso de Cisco Systems existen cuatro herramientas principales que permiten una gestión optima de la calidad de servicio aplicada en equipos de comunicaciones de Cisco, las cuales se nombran a continuación:

- Cisco QMD. Cisco Quality Device Manager
- Cisco QPM. Cisco Quality Policy Manager

- Cisco SAA. Cisco Service Assurance Agent
- Cisco IPM. Cisco Internetwork Performance Monitor

Estas cuatro herramientas permiten realizar una gestión de la calidad de servicios sobre infraestructuras Cisco.

6.1. Cisco Quality Device Manager

Configurar QoS no es suficiente para un exitoso desarrollo de Calidad de Servicio: Los administradores de red deben monitorear el trafico de la red antes y después de que las políticas sean aplicadas para asegurar que los efectos deseados están ocurriendo. Las tendencias históricas son esenciales para asegurar que a medida que la empresa crece y cambia, los niveles de servicio se mantengan.

QDM es una herramienta que permite configurar y monitorear políticas de QoS en un dispositivo sencillo. QDM es la herramienta obvia para experimentar con QoS y observar su impacto sobre una red pequeña o experimental. Una vez que el administrador de la red ha confirmado que las políticas correctas han sido configuradas en el dispositivo y que se corresponde con los resultados esperados, se puede aplicar en toda la red empresarial.

QDM tiene la ventaja de un interfaz Web con comandos accionados por medio de un click, eliminando los posibles errores por el uso de interface de línea de comando (CLI).

6.2 QPM

QPM es una aplicación de gestión de redes que soporta configuración de rango empresarial de funcionalidades de QoS en dispositivos de Cisco basados en políticas definidas por el usuario. QPM provee a los administradores de redes la

habilidad para configurar políticas consistentes y centralizadas en la red. QPM ofrece varia ventajas, tales como:

- Carga de Dispositivo: Permite agregar la configuración a un dispositivo nuevo mediante las políticas existentes en la base de datos de QPM.
- Verificación de dispositivo: Permite verificar que la un política de QoS que ha sido previamente desarrollada está sin alteración en la configuración del dispositivo.
- Reversión de Configuración: Permite revertir rápida y fácilmente una configuración previamente realizada en caso de que una política incorrecta haya sido desarrollada o en el caso de que un conjunto de políticas ya no será implementada.
- Desarrollo Externo: Permite aplicar QoS mediante el uso de scripts u otro mecanismo externo a QPM.

QPM y QDM son productos complementarios. Ambos productos agregan valor a los administradores de red con un conocimiento de calidad de servicio para una definición de políticas de QoS fáciles y rápidas mediante el uso de interfaces tipo GUI (Graphical User Interface). QDM desempeña configuración y monitoreo de la QoS en routers Cisco al mismo tiempo que permite un buen punto de inicio para la implementación de Calidad de Servicio en una red.

6.3 Cisco Service Assurance Agent

Otra herramienta para la gestión de redes es Service Assurance Agent (SA Agent). SA Agent genera un trafico que es utilizado para monitorear el desempeño de la red mediante la medición de los niveles de servicio, métricas tales como tiempo de respuesta, recurso de red, disponibilidad, Jitter, tiempo de conexión, perdida de paquetes y métricas de desempeño de aplicaciones, dando una representación indicativa de las condiciones de la red. Los datos que son capturados por el SAA

son usados por las aplicaciones Internetwork Performance Monitor (IPM) y Service Management Solution (SMS), para proveer gráficos que facilitan la creación de baseline, monitoreo, tendencia y solución de problemas. La información que provee el SAA permite cerrar el lazo de la gestión de redes.

6.4 Cisco Internetwork Performance Monitor

IPM permite a los ingenieros de redes manejar proactivamente problemas de tiempo de respuesta. IPM notifica cuando el tiempo de respuesta en la red se degrada o cuando un enlace siendo monitoreado pasa a inactivo.

IPM permite la medición del desempeño automático para todas las rutas entre dos dispositivos en la red o para cada salto en la ruta entre dos dispositivos de red. Los administradores de red pueden fácil y rápidamente identificar la fuente de un problema de desempeño en un punto de la red. Como resultado, el diagnostico rápido de problemas aumenta la disponibilidad de la red.

IPM provee reportes gráficos históricos y de tiempo real del tiempo de respuesta entre dos dispositivos de red. IPM también notifica proactivamente a los administradores de red con un trap SNMP cuando un tiempo de respuesta excede el umbral predefinido o cuando un enlace deja de estar disponible.

VII. ESTRATEGIAS PARA GARANTIZAR CALIDAD DE SERVICIO

Las estrategias de Calidad de Servicio están orientadas a cumplir con los requerimientos de cada servicio presente en la red para que opere de manera eficiente y de forma integrada. En tal sentido la estrategia principal que se presenta en este trabajo es la metodología que se debe seguir para integrar Voz, Video y Datos en una misma red.

Esta metodología consta de cuatro procesos fundamentales que permiten la integración de los servicios de información corporativos. A continuación se indican estos procesos:

1. Determinar las prioridades y políticas de QoS de la empresa: Consiste en determinar que trafico es importante y que nivel de servicio requiere.
2. Caracterizar la Red: Conducir análisis para capturar información sobre el uso actual de Voz, Video y Datos.
3. Implementar políticas: Consiste en la definición de listas de control de acceso, clases y políticas para las clases.
4. Monitorear la Red: Capturar trafico en la red para hacer estudios de desempeño y manejar las fallas, lo cual permitirá corregir comportamientos fuera de la line base.

Las características que debe tener una solución robusta que sirva de estrategia principal para garantizar calidad de servicio en la integración de Voz, Video y Datos son las siguientes:

- Políticas de Extremo a Extremo: La QoS debe ser aplicada de extremo a extremo. Consecuentemente, esta debe ser independiente de plataforma, dispositivo y medios de transporte, abierta al nivel de la capa 3 y más arriba para asegurar funcionalidad de extremo a extremo a través de múltiples

dispositivos de red (tales como, Routers, Switches, Firewalls, Servidores de Acceso y Gateways) enlaces de datos (por ejemplo, ATM, Frame Relay, o Ethernet).

- Múltiples Parámetros: Las políticas deben estar basadas en como las personas usan la red. Los dispositivos deben tener la flexibilidad para aplicar QoS basada en diversos parámetros que pueden reflejar las políticas de calidad definida por el administrador de la red. Estos parámetros distinguen flujo de trafico basado en IP, direcciones MAC, aplicaciones, usuario, hora del día, o localización en la red.

  Los administradores de red deben definir políticas usando una combinación de esos parámetros. Por ejemplo una política puede ser "Ningún trafico de Webcast esta permitido a través de cierto enlace a 56 Kbps hacia la oficina principal". Una política más compleja puede ser, "Permitir trafico de video conferencia en un segmento Lan definido los lunes entre las 3 y las 5 de la tarde, solo para los usuarios A, B, C y D. En cualquier otro momento o para otros usuarios esta prohibido ese trafico.

- Clasificación: Por definición, se necesita configurar diferentes niveles de Calidad de Servicio para el trafico de acuerdo en lo especificado en la política.

- Control Centralizado: Una red basada en políticas debe ser un desarrollo consistente y eficaz manejado de forma central.

- Sofisticadas Herramientas de QoS: Debido a que existen muchos elementos de red diferentes y muchos parámetros requeridos para exitosamente desarrollar e implementar políticas de QoS de extremo a extremos, el

conjunto asociado de herramientas es necesariamente complejo. Ellas deben permitir proveerle a la red la inteligencia que necesita.

7.1 Prioridades y políticas de QoS de la Empresa

Las políticas deben de estar basadas en como las personas usan la red. Los dispositivos deben tener la flexibilidad para aplicar QoS basada en diferentes parámetros que reflejan las políticas definidas por los administradores de red. Estos parámetros distinguen el flujo del trafico basado en IP, dirección MAC, aplicación, usuario, hora del día, o localización en la red.

Para esto es necesario responder las siguientes preguntas: Cuál tipo de servicio es apropiado para desarrollar en la red?. Es un servicio Integrado?. Es un servicio Diferenciado?. Esto depende de diversos factores:

1. La aplicación o el problema que se esta tratando de solucionar: Cada uno de los tres tipos de servicios es apropiado para ciertas aplicaciones. Esto no implica que se deba migrar a servicios diferenciados y entonces a servicios garantizados. Un servicio diferenciado o de mejor esfuerzo puede ser apropiado dependiendo de los requerimientos de las aplicaciones.
2. La velocidad en la cual se puede realmente actualizar la infraestructura. Hay una ruta de actualización natural de tecnologías necesitadas para proveer servicios diferenciados que la que se necesita para servicios integrados.
3. El costo de implementación y desarrollo de servicios integrados es comúnmente mayor que para un servicio diferenciado.

Se deben definir dominios de QoS en la red para indicar cuales dispositivos realizaran cuales tareas. La instancia más común de un dominio de QoS es una demarcación entre una Lan y una Wan. El dispositivo (usualmente un router) que enlaza la Lan y la Wan se convierte en el dispositivo de borde del dominio de QoS,

mientras que los dispositivos en el backbone de la Lan y la Wan son dispositivos de core (núcleo). Los dispositivos de borde deben manejar las actividades requeridas para monitoreo, reconocimiento y clasificación del trafico; los dispositivos de core son optimizados para transmitir trafico de alta prioridad a gran velocidad y sin retardos.

En cada dominio se puede definir políticas de QoS especificas. Por ejemplo, los servicios garantizados pueden ser importantes para soportar trafico de aplicaciones de criticas en la Wan entre la oficina corporativa y sus oficinas regionales. En la Lan, donde el ancho de banda es mayor, los servicios diferenciados pueden ser aplicados, excepto en la división de ingeniería, la cual puede necesitar servicios garantizados para sus aplicaciones de modelaje y simulación.

Cuando se categoriza el trafico en clases, es importante notar que, mientras la categorización basada en el tipo de aplicación puede ser la manera más intuitiva (como se muestra en la figura numero 44), hay otros caminos para dividir el trafico en clases.

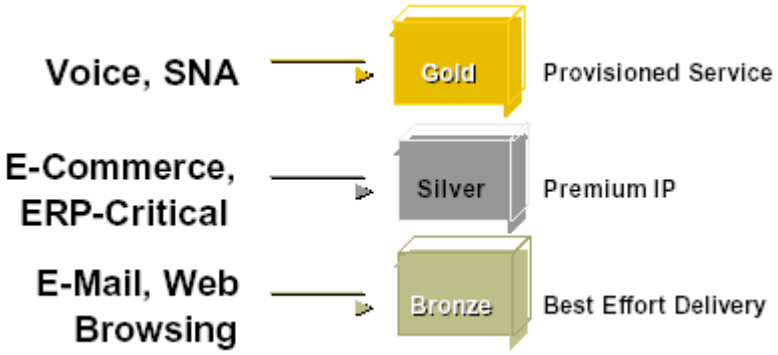


Figura # 44. Categorización según el tipo de Aplicación.

Otra forma de categorizar es mediante clases de servicio. Se puede separar la asignación de ancho de banda y los espacios en buffer, en clases especificas según las necesidades de los servicios. Por ejemplo, la clase gold podría ser usada para trafico de voz: Una asignación de un gran ancho de banda asegura que existe

suficiente ancho de banda para todos los paquetes de voz en una cola de voz, mientras un moderado tamaño del buffer limita el potencial para retardo de paquetes.

También se puede categorizar el trafico basado en políticas de la siguiente manera:

- Nivel de Servicio: Tratar el trafico gold con el más alto nivel de servicio que el silver y bronze.
- Evitando Congestión: El trafico Bronze o Silver será descartado cuando exista congestiona. El trafico gold será enviado sin ser afectado siempre que no se exceda las velocidades contratadas.
- Máximo Ancho de Banda Garantizado/Prioridad para un Clase: Asegurar que la clase gold obtenga un tratamiento prioritario y las clases silver y bronze obtengan el menor ancho de banda garantizado.
- Máxima Limitación de Velocidad: Asegurar que el trafico bronze no obtendrá más que X Kbps del ancho de banda en cualquier momento.

7.2 Caracterización de la Red

La caracterización de la red permite determinar que aplicaciones están ejecutándose en la red, cuanto ancho de banda consumen, que tiempo de respuesta obtienen los usuarios, que nuevas aplicaciones serán implementadas, cual es la utilización actual del ancho de banda por enlaces, cual es la velocidad actual de perdida de paquetes, donde están los puntos de congestión, cual es el patrón de trafico de voz, cuales son los requerimientos del trafico de video.

La caracterización del trafico juega un papel crucial en el análisis de desempeño y diseño de redes de comunicación. En el capitulo tres, se mostró la herramienta llamada NBAR la cual permite realizar la clasificación del trafico en una

infraestructura Cisco, esta aplicación incluye un protocolo de descubrimiento que permite detectar cuales aplicaciones está ejecutándose en la red. Esto ayuda a definir las clases de QoS y las políticas, tales como cuanto ancho de banda proveer a aplicaciones críticas y ayuda a determinar a cuales protocolos aplicar políticas.

Luego de analizar el trafico hay que hacer una lista de las aplicaciones importantes, tales como:

- Voz
- Video
- Aplicaciones de Finanzas, tales como SAP y People Soft
- Aplicaciones de Negocio
- Puntos de Venta
- Respaldo o Sincronización de servidores
- Transacciones de base de datos, tales como Oracle y SQL.

Este procedimiento es conocido como clasificación.

7.3 Aplicación de Políticas

La meta básica de las políticas es asegurar que el ancho de banda de la red sea usado eficientemente, mediante el uso de características de calidad de servicio para proveer:

- Ancho de Banda garantizado/bajo retardo
- Limitar el ancho de banda vía policing
- Traffic Shaping
- Marcado de paquetes

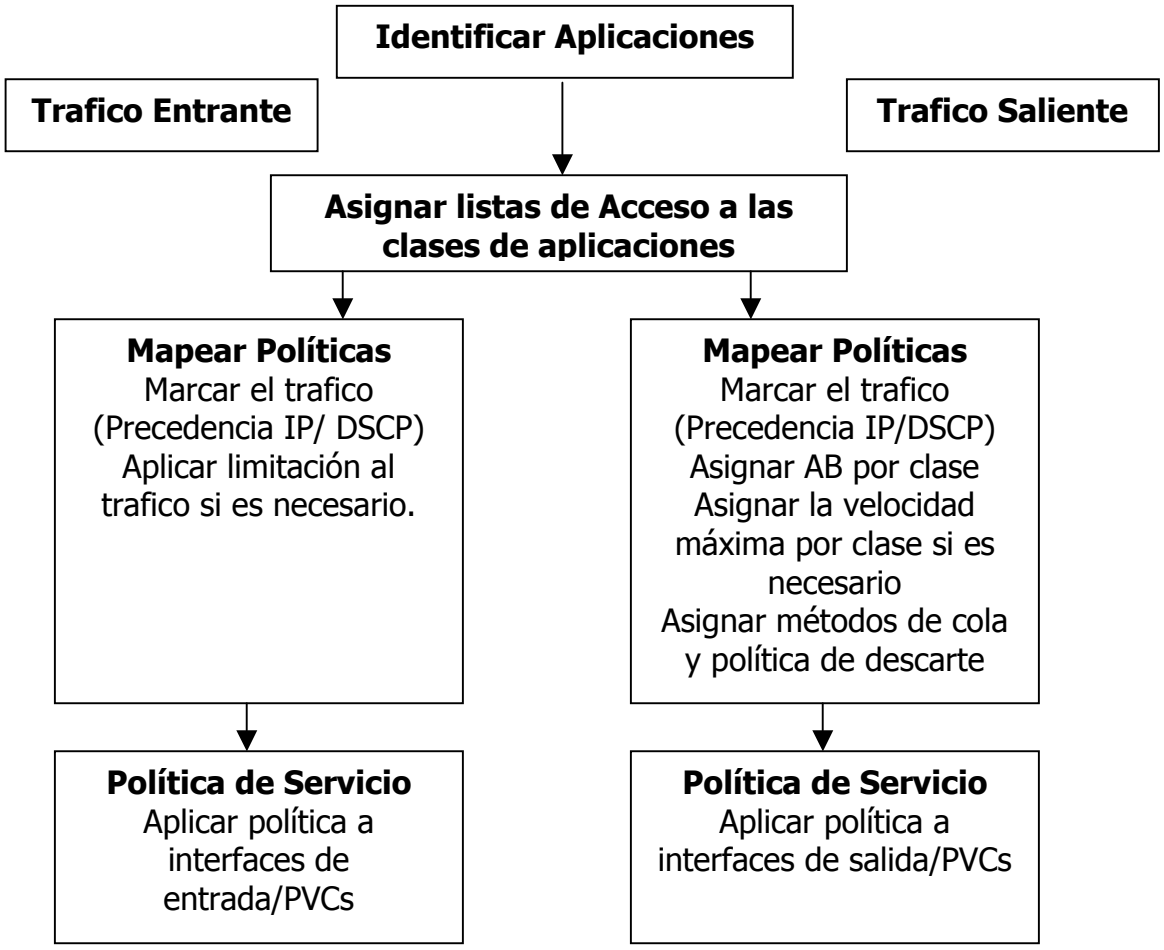


Figura # 45. Pasos para implementar QoS.

La calidad de Servicio puede aplicarse en dos sentidos, al trafico entrante o al trafico saliente (como se muestra en la figura 45). Dependiendo del sentido en que el trafico está fluyendo se aplican ciertas técnicas y políticas de calidad. Sin importar si es entrante o saliente, se debe buscar una manera de poderlo identificar, categorizar y luego asignarle una clase; una forma de realizar esta tarea es mediante listas de control de acceso. Una vez clasificado es necesario realizar una marcación mediante precedencia IP o mediante los bits del DSCP, luego se debe limitar el ancho de banda a usar (si es necesario), al final se aplican

las políticas de QoS definidas. Para el caso del trafico saliente es necesario incluir un método de encolamiento y una política de descarte.

Estos pasos no son suficientes para garantizar QoS en la integración de redes, es necesario realizar un monitoreo con algunas de las herramientas antes estudiadas para poder controlar la correcta aplicación de las políticas así como el diagnostico que permita evitar problemas de disponibilidad de los servicios.

7.4 Gestión de la Calidad de Servicio

Como se explicó en el capitulo 6, la gestión de la calidad de servicio busca garantizar el correcto comportamiento de las políticas de QoS implementadas en la red y facilitar su aplicación en los dispositivos de comunicaciones. La gestión de la calidad de servicio cumple esta función mediante herramientas sencillas de monitoreo y configuración de las políticas de calidad de servicio definidas en una organización.

Para realizar esta tarea existen diferentes software que se encargan de facilitar a los administradores de red la configuración de políticas de QoS y su posterior monitoreo; esto mediante interfaces graficas y en muchos caso con aplicaciones basadas en Web.

Si se están integrando redes o servicios en una infraestructura compuesta por equipos de comunicaciones de Cisco se puede usar la solución de gestión integrada por las aplicaciones: QDM, QPM, IPM y SAA. Para otos productos, existen aplicaciones que están basadas en el protocolo SNMP y que se encargan de leer las variables MIBs de cada dispositivo y de acuerdo a consideraciones matemáticas permiten monitorear en tiempo real o mantener históricos del comportamiento de los servicios de información, luego de aplicada alguna política de calidad en una parte o en toda la red.

7.5 Consideraciones de QoS para Voz

Existen varios factores que inciden en una buena comunicación de Voz, esos factores son el delay, el Jitter y la perdida de paquetes.
Para lograr integrar en una red de Voz, Video y Datos basada en paquetes es necesario tomar en cuenta las siguientes consideraciones:

7.5.1 Retardo

Los elementos que contribuyen con el retardo están constituidos por elementos fijos y elementos de tipo variable. Los elementos fijos son el retardo de propagación, el retardo de codificación, el retardo de paquetizacion y el retardo de serialización. Estos valores deben tener como máximo 0-150 mseg en una sola vía. Este valor es producto de la recomendación ITU G.114. la cual indica que cuando se diseña una solución de Voz el retardo máximo en una sola vía debe ser de 150 mseg. Esto es un valor practico. Es posible encontrar valores por encima o por debajo, dependiendo del ambiente de red existente.

7.5.2 Jitter. Retardo Variable

Las colas de salida congestionadas, y el retardo de serialización en las interfaces de red pueden causar retardo variable en el envío de paquetes. Sin prioridad o colas de baja latencia como LLQ, es imposible la transmisión de Voz sobre una red de paquetes. Esto es causado por el envío de paquetes de datos muy largos que monopolizan el ancho de banda disponible, haciendo que paquetes más pequeños (por ejemplo, paquetes de Voz) sean afectados durante su transmisión. Para solucionar este problema se debe utilizar fragmentación de los enlaces e intercalar los paquetes de Voz, Video y Datos que son colocados en el buffer de salida.

7.5.3 Transmisión de Voz sobre enlaces Wan

En términos generales se pueden seguir las siguiente recomendaciones para la transmisión de Voz sobre enlaces de alta y/o baja velocidad:

Para alta velocidad:

- Controlar el trafico con Frame Relay Traffic Shaping
- Utilizar como técnica de encolamiento LLQ
- Comprimir largas cabeceras con CRTP (si es necesario)

En enlaces con velocidades hasta un E1, es necesario usar prioridad para la Voz con el fin de garantizar a la Voz la estricta prioridad que requiere. Si el ancho de banda para Voz es un asunto critico, se puede comprimir las cabeceras RTP con CRTP. Debido a que son enlaces de alta velocidad, no es necesario fragmentar largos paquetes de datos para intercalarlos con los de Voz.

Para Baja Velocidad:

- Realizar clasificación y Marcado con IP Precedente/DSCP.
- Usar Frame Relay Traffic Shaping
- Usar LLQ
- Implementar FRF.12
- VAD
- DTMF Relay (dependiendo del CODEC).

7.6 Consideraciones de QoS para Video

Las tecnologías punto a multipunto permiten la transmisión de programas para grandes audiencias sin colocar un innecesario cuello de botella en el servidor o en

la red. Sin embargo, la infraestructura de red debe soportar enrutamiento y protocolos de grupo multicast. El trafico de video unidireccional no es interactivo mientras que el bidireccional envuelve transacciones interactivas entre fuente y destino.

Mientras que el Video sobre IP tiene consistente y demostrables valores máximos de retardo y mínima perdida de paquetes, estos valores son mucho más bajos que para Voz sobre IP. Por ejemplo, un valor típico de retardo en una sola vía para una cadena de Video en formato MPEG-2 es de 400 a 500 mseg antes de que la calidad del Video se vea afectada.

Debido a su más alta tolerancia al retardo, los paquetes de Video podrían no ser colocados en la cola de prioridad orientada a la Voz en los elementos de red.

Otra razón más compleja para no unir paquetes de Voz y Video en la misma cola es que cada punto final de Video transmitirá paquetes de Video usando el tamaño máximo del MTU de la interfaz. Cada paquete de Voz es absolutamente más pequeño para minimizar el retardo a medida que viaja por un nodo de la red.

La solución a este dilema es la creación de una clase para el trafico de Video usando un valor determinado del DSCP para clasificar el Video sobre IP. Esta clasificación debe ofrecer la posibilidad de una más alta clasificación para asegurar el envío del trafico de Video, mientras también se especifique el menor porcentaje de perdida de paquetes.

Para el caso de Video bidireccional a baja velocidad se debe utilizar Priority Queuing y especificar un ancho de banda de 384 Kbps. Para el caso de Video en una sola vía usar CBWFQ.

Para transmitir Voz y Video sobre líneas dedicadas se debe seguir la siguiente estrategia:

- Mecanismo de Encolamiento: LLQ
- Fragmentación e Intercalado: MLPPP
- Compresión RTP: CRTP
- VAD
- IP RTP Priority
- Fax Relay
- DSCP

7.7 Estrategias para garantizar QoS en el Caso Frame Relay

En el escenario Frame Relay hay que definir la estrategia en dos partes, una del lado de la red del operador de servicio Frame Relay y la otra desde el lado de los equipos de acceso de la red corporativa.

a.- Lado del operador: Para poder garantizar una calidad de servicio aceptable para los servicios de voz, video y datos, se debe solicitar al operador del servicio Frame Relay, que ofrezca PVCs con calidad de servicio, es decir, PVCs que están definidos dentro de la red Frame Relay como los de mayor prioridad, por lo que al presentarse congestión, no se descartarán las tramas de éstos, además que las tramas de estos PVCs tendrán acceso siempre a los buffers de salida de los conmutadores Frame Relay. Fuera de estos requerimientos, se debería establecer un contrato de garantía de niveles de servicio entre el operador de la red Frame Relay y la red corporativa, es decir, un SLA.

Los parámetros que se definirían son los siguientes:

- % de disponibilidad de los PVCs (availability)
- Retardo máximo (delay)
- Desempeño (throughput)
- MTTR (Mean time to restore o repair)
- MTTr (Mean time to respond)
- Costo.

Para el caso particular de este escenario se debería estipular los siguientes valores:

*.- % de disponibilidad: 99,5% mensual
*.- Retardo ida y vuelta máximo (round trip): 140 ms.
*.- Desempeño: 99% del valor del CIR promedio durante el mes
*.- MTTR: 4 horas
*.- MTTr: 4 horas

b.- Lado de los equipos de acceso: Una vez que se cuenta con la mejor condición de calidad de servicio dentro de la red Frame Relay, hay que definir las herramientas y facilidades para otorgar calidad de servicio a las aplicaciones de voz y vídeo. Estas facilidades residen en los equipos de acceso Frame Relay, entre las cuáles se definirán las siguientes:

Clases de tráfico VBR-rt, VBR-rt, UBR. Estos tipos de tráfico, que son característicos de ATM, son implementados en los equipos multimedia FRADs, para así garantizar la calidad de servicio de las aplicaciones. En el caso de la voz y videoconferencia, se usará VBR-rt, y para el caso de datos se usará UBR. Además de las clases de tráfico, los equipos de acceso deben permitir hacer la

fragmentación de las tramas. Por último se deben realizar funciones de manejo de prioridad tanto en el acceso al enlace de salida, así como en el acceso a las colas de las interfaces de entrada y salida.

7.8 Estrategias para Garantizar Qos en el Caso de ATM

El escenario ATM es el que posee mayores facilidades para garantizar la calidad de servicio de las aplicaciones, basado en sus funciones y herramientas con las que cuenta. Si se dimensiona correctamente una red con el ancho de banda necesario para las aplicaciones que serán transportados sobre ella, entonces implementando las herramientas de calidad de servicio de ATM, se podrá garantizar la calidad de servicio de las aplicaciones.

El escenario ATM también fue analizado desde dos puntos de vista, una desde el punto de la red ATM y en segundo lugar la de los equipos de acceso o concentradores.

a.- Punto de vista de la RED: LA red ATM cuenta con un tratamiento especial para las aplicaciones sensibles a retardos, sólo hay que definir los PVC con las características particulares de servicio, y si la red acepta esa definición entonces estará garantizada la calidad de servicio. Sin embargo para las aplicaciones de voz y vídeo se usa el establecimiento de las llamadas mediante los SVCs al momento que los clientes (gateway de videoconferencia y gateway de voz), intenten hacer una llamada, ésta se hará a través de un SVC, el cuál en el momento del *setup*, negociará los parámetros de calidad de servicio como son CDV (cell delay variation), PCS (peak cell rate), SCR (sustained cell rate) y MBS (maximun cell rate). En caso de que la red a través de las técnicas de control de admisión acepte los parámetros, entonces será garantizada la calidad de servicio dentro de la red.

La red también cuenta con herramientas de "*policy*", en las cuáles se supervisa que cada unas de las conexiones establecidas dentro de la red esté ajustada a los valores previamente negociados, evitando de esa manera que ninguna conexión monopolice los recurso de la red y afecte la calidad de servicio del resto de las conexiones.

b.- Punto de vista del acceso: A nivel del equipo de acceso ATM, se requiere que éste clasifique las aplicaciones de acuerdo a la calidad y clase de servicio definida, para luego establecer la conexión (SVC) con los parámetros de calidad de servicio correspondiente. Una vez establecido el SVC, el equipo de acceso tendrá la obligación de manejar una regla de prioridad donde la prioridad para la salida a la red ATM la tendrán las aplicaciones de voz y vídeo, quedando la ultima prioridad para las aplicaciones de datos. Es responsabilidad del equipo de acceso iniciar el establecimiento de las conexiones virtuales a través de la red ATM (usando SVCs), con los parámetros de calidad de servicio asociados a cada una de las aplicaciones.

7.9 Estrategias para garantizar QoS en el Caso IP

El escenario IP es el más complejo para garantizar calidad de servicio a las aplicaciones de voz y vídeo. Aunque este protocolo cuenta hoy en día con una gran variedad de herramientas de calidad de servicio, siempre será una de los puntos a vencer para que de alguna manera se pueda ofrecer una calidad de servicio aceptable. Debido a la naturaleza de connectionless, el protocolo IP tiene una debilidad en que no se puede garantizar al 100% la calidad de servicio de las aplicaciones sensibles a retardos. Es por ello que para transportar este tipo de aplicaciones sobre una red IP, se debe hacer un trabajo de ingeniería bien detallado para que de esa manera el diseño resultante ofrezca el mejor ambiente para garantizar en lo mejor posible la calidad de servicio. En muchos casos hay

que combinar varias herramientas de calidad de servicio para que el resultado de esa combinación ofrezca ventajas para las aplicaciones.

Del mismo modo como se hizo en el escenario ATM y *Frame Relay*, aquí se definirán las estrategias desde dos puntos de vista. A nivel de la red y al nivel del equipo de acceso.

a.- Desde el punto de vista de la red: A nivel del backbone IP de la red corporativa, se deben definir las políticas apropiadas para garantizar la calidad de servicio a las aplicaciones, entre esas políticas se tienen:

- Políticas para evitar la congestión
- Reconocimiento del campo TOS
- Asignación de colas a paquetes con prioridad basados en clases de tráfico.
- Reserva de ancho de banda.

b.- Desde el punto de vista del equipo de acceso: El equipo de acceso es el que tiene mayor responsabilidad en el proceso de garantizar la calidad de servicio. A continuación se detallan las políticas a definir en los equipos:

- Clasificación: Esta primera fase se encarga de marcar los paquetes de acuerdo a la clase de servicio a la que pertenecen, y esto se logra modificando el campo ToS.
- Asignación de colas a paquetes con prioridad basados en clases de tráfico.
- Reserva de ancho de banda.

CONCLUSIONES

La integración de servicio de Voz, Video y Datos requiere de la aplicación correcta de políticas de Calidad orientadas a satisfacer los requerimientos necesarios para un funcionamiento optimo de los mismos. Para poder cumplir con ese objetivo es necesario realizar un conjunto de pasos estratégicos que garanticen que la aplicación de QoS sea realizada de manera adecuada para permitir una integración eficiente de Voz, Video y Datos.

Durante este proyecto de grado se pudo estudiar y poner en practica un gran conjunto de procedimientos, no solo para lograr aplicar políticas de Calidad de Servicio, sino también la metodología a seguir para definir esas políticas y su posterior monitoreo.

Actualmente existe un abanico muy amplio de opciones para aplicar QoS en una red, sin embargo es necesario saber en que momento y por que usar servicios diferenciados o servicios integrados, que técnica de encolamiento es la mas adecuada para un servicio, a que servicio darle prioridad, cuanto ancho de banda se requiere para transmitir Voz y Video, etc. Todas estas interrogantes deben tener una respuesta que contribuya con la mejora de la calidad de algún servicio sin perjudicar el correcto funcionamiento de otro.

Es importante conocer el funcionamiento de las aplicaciones y servicios presentes en la red, pues este conocimiento permitirá la selección correcta de la estrategia a seguir para garantizar la calidad necesaria para la Voz y el Video sin afectar el funcionamiento adecuado de aplicaciones críticas.

En la actualidad las redes de paquetes son soportadas por muchos protocolos, estos protocolos pueden operar en una o en varias capas del modelo OSI. Durante esta investigación se determino que la opción más flexible y con mayor aceptación

es el uso del protocolo IP para la transmisión de Voz, Video y Datos, esto permite tener todos los servicios críticos operando bajo la misma tecnología de paquetes.

En tal sentido es necesario realizar un conjunto de pasos que permitan garantizar el tratamiento correcto a cada servicio. Es importante la clasificación de los paquetes para poder determinar que tipo de aplicación es y en función a sus características garantizar los requerimientos mínimos de operación. Luego de la clasificación es importante marcar cada paquete para poderlo identificar en toda la red y asignarle una respectiva clase de servicio que se adecue a sus requerimientos de ancho de banda y retardo.

La clasificación y el marcado no es suficiente para garantizar QoS, es necesario tomar en cuenta la posibilidad de que se produzca congestión en la red y por supuesto controlarla, por lo cual es imperativo usar mecanismos que manejen la congestión tales como Priority Queuing o LLQ; de ser posible se debería tratar de evitar congestión mediante el uso de algoritmos como WRED o FRED, pero tomando en consideración que no se puede descartar paquetes de Voz o de Video.

La mayoría de las empresas cuentan con varias localidades remotas las cuales deben disfrutar de los mismos servicios corporativos presentes en la red principal, para lograr eso se realiza una interconexión mediante enlaces de tipo Wan, esto enlaces en la mayoría de los casos es de tipo Frame Relay o ATM. Debido a que los enlaces representan un costo fijo alto las empresas usan el menor ancho de banda posible. Para lograr un uso eficiente de estos anchos de banda se deben aplicar mecanismos de eficiencias de enlaces que optimizan el funcionamiento de enlaces digitales de baja velocidad. Se puede usar compresión de cabeceras en conjunto con técnicas como interleaving.

Una vez que el paquete de Voz, o Video sale de la red corporativa entra en una red publica, en este punto es importante que la red del proveedor de servicio

posea mecanismos de QoS tan o más eficientes que los aplicados internamente, por lo tanto es necesario llegar a acuerdos de servicio (SLA) con los proveedores que permitan garantizar un tratamiento adecuado a cada servicio de información que es transmitido hacia la red publica y que los paquetes que la red publica entrega en el punto final conserve las características de marcado ya realizadas en la red privada para así asegurar una QoS de extremo a extremo.

El punto final en la estrategia es realizar un seguimiento del comportamiento de los servicios de información para garantizar su optimo funcionamiento y el cumplimiento de las políticas de QoS diseñadas para una operación satisfactoria de la red convergente.

RECOMENDACIONES

1.- La primera recomendación es implementar la red multiservicios basada en IP, ya que representa la mejor relación beneficios versus inversión.

2.- Si se realizara una integración en una red ya existente, se recomienda verificar que la red tenga la capacidad para aplicarle QoS.

3.- Para el soporte de QoS en los equipos Cisco se recomienda escoger la versión de IOS (Internetworking Operating System) adecuada, ya que no todas las versiones soportan QoS.

4.- Proveer QoS en la red Lan mediante el uso de los campos CoS de la trama Ethernet.

5.- Debido a que las estrategias presentadas pueden ser aplicadas sobre la red corporativa por ser de administración privada, es necesario llegar a acuerdos de niveles de servicios (SLA) mínimos entre la empresa y el proveedor de servicio Wan. Esto debido a que si se aplica QoS en la red Lan y luego se quiere establecer comunicaciones de Voz y/o Video entre localidades remotas y no existe QoS en la Wan, entonces se perderá el trabajo de QoS realizado internamente.

6.- La ultima recomendación es seguir las estrategias indicadas en el capitulo 7.

BIBLIOGRAFÍA

1. Voice-Enabling the Data Network, Cisco Press, 2003.

2. Cisco Voice over Frame Relay, ATM, and IP, Cisco Press, 2001.

3. Integrating Voice and Data Networks, Cisco Press 2000.

4. Carlos Marcos Sánchez, "Gestión del ancho de banda para garantizar la calidad del servicio: la anarquía deja paso al orden", Internet 99, Madrid, Febrero de 1999.

5. Vicenzo Mendillo, Software Educativo: "Compresión de Datos y Codificación de Audio y Video", Abril de 2002.

6. Arquitecturas para Voz, Video y Datos Integrados, Cisco AVVID, 2004.

7. Deploying Cisco Quality of Services, Cisco Training, 2004.

8. *Frame Relay* Forum, "Voice over *Frame Relay* Implementation Agreement FRF.11", Mayo 1997.

9. IMTC, " Voice over IP Forum Service Interoperability Implementation Agreement 1.0", Diciembre 1997.

10. Soluciones De Integración De Voz, Video y Datos En Redes Corporativas. María A. Cruells. Marzo de 2004.