UNIVERSIDAD TECNOLÓGICA DE PANAMÁ

FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES

CALIDAD DE SERVICIOS EN REDES

Prof. Vladimir Villarreal

2014

CONTENIDO

Módulo I Aspectos Fundamentales de la Calidad de Servicio en Redes

Objetivos:

- Conceptualizar el tema calidad de servicios en redes
- Presentar una perspectiva histórica de la calidad de servicios en redes
- Comparar la calidad de servicios versus las clases de servicios

¿De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje se presentan los aspectos fundamentales de la Calidad de Servicios en redes (QoS). De identifican los conceptos básicos que componen este aspecto en las redes de comunicaciones. Se plantea una perspectiva histórica que resalta la importancia del QoS en la actualidad.

1. Aspectos fundamentales de la Calidad de Servicio en Redes

1.1 Definición de la Calidad en redes.

La calidad del servicio (QoS) se define como la capacidad de una red para proporcionar diversos niveles de servicio a los diferentes tipos de tráfico. Al contar con QoS es posible asegurar una correcta entrega de la información, dando preferencia a aplicaciones de desempeño crítico, donde se comparten simultáneamente los recursos de red con otras aplicaciones no críticas. QoS hace la diferencia al proveer un uso eficiente de los recursos en caso de presentarse congestión en la red, seleccionando un tráfico específico de ésta, priorizándolo según su importancia relativa, y utilizando métodos de control y evasión de la congestión para darles un tratamiento preferencial. Implementando QoS en una red, se logra un rendimiento de ésta más predecible y una utilización de ancho de banda más eficiente.

En una red se debe garantizar un cierto nivel de calidad de servicio para un nivel de tráfico, el cual sigue un grupo detallado de reglas o parámetros, las cuales establecen un contrato de intercambio de información entre el usuario y la red.

También se puede definir dependiendo del contorno de la red en que se aplique la Calidad de Servicio. En el ámbito de las telecomunicaciones, este término se define como: “el efecto colectivo del rendimiento de un servicio que determina el grado de satisfacción del usuario de dicho servicio”, esta definición está totalmente ligada con la apreciación del usuario al obtener este servicio, debido a que es quien establece unos requerimientos mínimos para cualificar. En el campo de la telemática, QoS se define como la capacidad que posee un componente de red (Aplicación, Servidor, Usuario, switch, etc.) para asegurar que el tráfico y requisitos establecidos utilizados en su red se manejan de la manera más óptima y por ende queden satisfechos; también es definida como el grupo de tecnologías que dejan que los administradores de la red decidan ante las consecuencias o resultados de la congestión del tráfico en la red (antes de ampliar consecutivamente capacidad en la red) utilizando las distintas técnicas que ésta posee.

1.2 Definición del Servicio.

Definimos Clase de Servicio (CS) como el conjunto de parámetros de calidad de transmisión que delimitan las características de un cierto flujo de información. En principio, cada una de las aplicaciones posibles (cuyo número es en principio indefinidamente grande), tendrá asociada una o más CS. Cada uno de los flujos de información generados y que deben ser transmitidos por la aplicación tendrá asignada una CS. Las conexiones asociadas a una CS generarán información siguiendo un cierto patrón de tráfico.

Definimos patrón de tráfico como la estadística con la que una cierta conexión genera paquetes de información. Los modelos de generación de tráfico se caracterizan por variables aleatorias y procesos estocásticos. Por tanto, el conocimiento completo de un cierto patrón de tráfico implica la definición de todos los momentos estadísticos de las variables aleatorias que definen el patrón.

Es evidente que el planteamiento de un sistema de comunicaciones que pueda dar cabida a la infinidad de patrones de tráfico posibles, cada uno de ellos con sus requerimientos de calidad, representa una tarea poco menos que inabordable. Es por ello que debe arbitrarse una solución viable y aplicable a un entorno real. Una solución posible consiste en la definición de un conjunto acotado de CS a las que deban acogerse todas las conexiones activas en el sistema y sus correspondientes aplicaciones. Este conjunto debe ser lo suficientemente amplio como para abarcar, de un modo suficientemente preciso, a la práctica totalidad de las conexiones que puedan requerir servicio del sistema, y a su vez lo suficientemente restringido como para simplificar en lo posible la implementación real de los mecanismos de gestión de recursos.

Por tanto, cuando una aplicación quiera ser servida por el sistema de transmisión, deberá analizar cada una de sus conexiones o flujos de información activos. Este análisis debe permitir decidir cuál de las CS definidas en el sistema resulta más adecuada a las necesidades de calidad de transmisión y se ajusta mejor al patrón de tráfico de cada conexión. Esta decisión deberá hacerse siempre basándose en un análisis conservador de los requerimientos, para asegurar así el cumplimiento de los requisitos necesarios para todas y cada una de las conexiones activas. En

este sentido, el uso de conformadores de tráfico para 'suavizar' el tráfico ofrecido al sistema es una técnica frecuentemente utilizada en los sistemas actuales.

De este modo, el sistema de comunicaciones verá las conexiones de todas las aplicaciones como un conjunto acotado y determinado de Clases de Servicio, cuyas características son conocidas a priori, lo que permite una planificación eficiente de la gestión del tráfico. Tanto el dimensionado de los accesos, enlaces y redes de comunicaciones, como la gestión de todo el funcionamiento del sistema para poder garantizar la calidad de servicio de las conexiones activas resultan realizables con un grado de complejidad abordable.

1.3 Perspectiva histórica

A continuación se presenta una reseña histórica de las redes de computadoras, es importante conocer la historia de las redes de computadoras, ya que nos indica cómo con el transcurrir de los años van surgiendo aplicaciones que exigen además de redes más y más robustas técnicas que permitan a esas aplicaciones tener un funcionamiento aceptable, una de esas técnicas es la Calidad de Servicio.

La gran industria de las redes de computadoras se empezó a formar en los años 80, con el surgir de diversos inventos relacionados con esta área, como es el caso de Alto Alhoa Network de Bob Metcalfe y Boggs, luego convertida en Ethernet y aplicada por la empresa 3Com en la primera LAN (1983); otro punto importante que se dio en esta década fue que se establecieron las normas OSI (Open Systems Interchange) por la Organización Estándar Internacional (ISO), también en esta misma década surge el Token Ring, una red local de datos inventada por la IBM.

La IEEE (Institute of Electrical and Electronic Engineers) designa un comité encargado de establecer normas para la transmisión de datos (el comité 802).

La interconexión entre redes, y la dilución de límites en ambientes locales convertidos en globales se da hasta 1985 con el surgimiento de los routers, pero el fortalecimiento y formalización del mercadeo en el campo de las redes se presentó en 1988, con la aparición de OpenView, la plataforma de administración y gestión de redes de Hewlett-Packard; y del Lan Manager, el sistema operativo de red de Microsoft que sustituía al MS-Net.

En la década del 90 surge la conmutación rápida de paquetes, y se inventa una tecnología nueva, la "Frame Relay", además se inicia la utilización del correo electrónico con la tecnología Token Ring. El 92 se inició con la tecnología ATM en un switch para redes privadas desarrollado por Network Equipment y fue aquí donde el término Calidad de Servicio se definió por primera vez en los protocolos de comunicaciones de esta tecnología (ATM), un año después, National Semiconductor implanta la tecnología Isonet, la cual admite la transmisión totalizada de servicios multimedia y toleraba protocolos Ethernet y RDSI. Después de ésta tecnología surgió la Fast Ethernet, basada en la norma 100 Base T, la cual contribuía con beneficios parecidos a las de Any LAN. Pero es sólo hasta finales de 20 los 90 que se desata la utilización de las redes, lo que consolidó el término Calidad de Servicio, debido a la incorporación de funciones de voz en redes de datos.

En estos últimos años se ha intensificado el manejo de funciones de seguridad, como la encriptación, la autenticación de usuarios, el firewalls, entre otros.

Estos sucesos acontecidos hasta el día de hoy ratifican que dentro de algunos años la voz gastará sólo una mínima porción del ancho de banda, y cualquier dificultad para los responsables del área de sistemas estará en gestionar apropiadamente un flujo de datos cada vez más denso y relevante.

1.4 Calidad de Servicio versus Clases de Servicios

De cara a la definición de las CS soportadas por el sistema, se hace necesario establecer los parámetros de transmisión que delimitan la calidad del servicio. Estos parámetros deberán tener una relación directa con la percepción que el usuario final (no necesariamente un ser humano) deba tener de la calidad de la conexión. Como usuario final se entiende cualquier nivel superior del sistema de comunicaciones que tenga unas necesidades de calidad determinadas. Así por ejemplo, para el caso de una aplicación de transmisión de voz en tiempo real, deberán establecerse relaciones entre los parámetros mesurables de la transmisión (retardo máximo de los paquetes, diferencia máxima entre retardos de paquetes, tasa máxima de

paquetes perdidos, tasa máxima de errores en los bits de los paquetes, etc.) y la percepción subjetiva de inteligibilidad del habla.

Será por tanto el tipo de aplicación o usuario al que se deba dar servicio lo que condicionará el tipo y los valores de los parámetros que marcarán la definición de cada CS. A continuación se presenta un conjunto acotado de parámetros de servicio, que se detallan a continuación:

- **Retardo medio de los paquetes de información:** se entiende por retardo de cada paquete el tiempo transcurrido desde que el bloque de información llega a la capa MAC hasta que es transmitido correctamente por la capa física.
- **Varianza del retardo de los paquetes de información:** el retardo de cada paquete, tal y como se ha definido en el punto anterior, es una variable aleatoria de la que podemos obtener su desviación típica y su varianza.
- **Retardo máximo de los paquetes de información:** se puede establecer un cierto tiempo de vida máximo de los paquetes, de tal modo que cuando el retardo de uno de ellos es superior a este tiempo de vida, el paquete es descartado.
- **Tasa máxima de paquetes perdidos:** porcentaje de paquetes descartados a causa de que su retardo ha superado el tiempo de vida prefijado para ellos.
- **Tasa de error media en los bits de información:** puede definirse antes o después de codificación, e indica el número relativo de bits erróneos (medido en porcentaje o como una probabilidad) que pueden admitirse por la aplicación.
- **Velocidad media de transmisión garantizada:** normalmente medida en Kbps, indica la velocidad media de transmisión para intervalos 'largos' de tiempo. Un intervalo largo se define como un número suficientemente grande de unidades de tiempo del sistema. Este número deberá ser grande en comparación con el tiempo en el que pueden variar las condiciones del tráfico ofrecido.

- **Velocidad mínima instantánea de transmisión:** también medida en Kbps, indica la velocidad mínima de transmisión de datos, si la hay, que se le garantiza a una conexión determinada. Si este valor es mayor que cero, indica que se está reservando una cierta cantidad de recursos de transmisión mínimos en exclusiva para la conexión, independientemente de la carga restante del sistema.

- **Velocidad máxima instantánea de transmisión:** indica la máxima velocidad de transferencia que le es permitida a una cierta conexión. Este valor puede usarse para impedir que una única conexión pueda copar una cantidad excesivamente grande de recursos del sistema, y evitar los problemas que de este hecho pudieran derivarse.

Normalmente, una CS estará definida por un subconjunto de estos parámetros, así como los valores correspondientes para cada uno de ellos. Y el QoS se referirá a la capacidad de una red de proporcionar un mejor servicio al tráfico de la red seleccionada sobre las diversas tecnologías, manteniendo los parámetros para cada clase de servicio.

Bibliografía

- Tanenbaum. Andrew S. (2003). Redes de Computadoras. Pearson Educación. México.
- J.A. Jimenez Toro. UF1875: Gestión de recursos, servicios y de la red de Comunicaciones. Editorial Elearning SL. Edición 5.0. ISBN: 978-84-16199-01-3. España.
- Álvarez, S., González, A. Estudio y Configuración de Calidad de Servicio para Protocolo IPV4 e IPV6 en una Red de Fibra Óptica WDM. Universidad Técnica Federico Santa María.

Caso de Estudio

1. TEST 1 (Red Básica)

Este test consiste en evaluar el comportamiento de la red, sin ningún método de control de acceso al medio, ni haber instalado políticas de calidad de servicio.
Se debe ejecutar el script “tráfico” en las estaciones clientes y evaluar la calidad de la navegación web con el programa *echoping*. Simultáneamente los clientes deben comenzar a generar tráfico.
Ejecute el siguiente comando en las estaciones del mesón 1

#trafico

Ejecute el siguiente comando en todas las estaciones

#/usr/local/bin/echoping -h / -n 20 http://192.168..1.1

Tome nota de los resultados:
Minimum time: _____
Maximum time: _____
Average time: _____
Standard deviation: _____
Median time: _____

Módulo 2 Estándares y Organismos de Estandarización

Objetivos:

- Conocer el concepto de estándar y los organismos que regulan los estándares en las comunicaciones.
- Resaltar la importancia de la aplicación de estándares en las comunicaciones.

¿De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje abarcamos lo concerniente a estándares y los organismos internacionales que los representan. Resaltamos la importancia de los estándares la cual permite la compatibilidad de los productos y reconocemos los estándares más sobresalientes de internet.

2. Estándares y Organismos de estandarización

En los primeros tiempos en que apareció el concepto de redes, cada compañía fabricante de equipos tenía sus protocolos de comunicación, por lo que dos equipos de diferentes compañías no podían comunicarse entre sí. Esta incompatibilidad provocó la exigencia, por parte de los usuarios, para que se estableciera una estandarización al respecto, evitando así caer en los mercados cautivos de las distintas compañías.

La estandarización no solamente facilitará la comunicación entre ordenadores o equipos construidos por diferentes compañías, sino que también beneficiará a los productos que se acojan a la norma ya que el mercado de dichos productos será mucho más amplio. Se facilitará una producción masiva de los equipos y por tanto, al poder utilizar técnicas de producción mejores la tendencia será a disminuir el precio de los productos y facilitar la aceptación de los mismos.

Las administraciones de cada país deberán dictar normas de obligado cumplimiento por parte de los equipos de telecomunicación que deberán ser homologados cuando las cumplen. El acelerado desarrollo de redes, servicios y aplicaciones a nivel mundial no ha hecho sino aumentar la necesidad de la coordinación y reglamentación internacional.

Un estándar se puede definir, por ejemplo, el tipo de conector a emplear, las tensiones e intensidades empleadas, el formato de los datos a enviar, etc. En resumen, un estándar es un conjunto de normas, acuerdos y recomendaciones técnicas que regulan la transmisión de los sistemas de comunicación.

2.1 La importancia de los estándares

El empleo de estos estándares presenta las siguientes ventajas:

- Los productos de diferentes fabricantes que cumplen los estándares son totalmente compatibles y, por tanto, pueden comunicarse entre ellos sin necesidad de utilizar adaptadores.
- El mercado se amplía, ya que al existir compatibilidad entre los productos de diferentes fabricantes, la oferta de productos será mayor, pudiendo derivar

en precios más competitivos. Esto se traduce en una mayor flexibilidad a la hora de elegir y utilizar dispositivos.

- Se asegura la compatibilidad con productos futuros empleando la misma tecnología.
- Se reducen los costes de los productos.
- De esta forma, la estandarización evita que las empresas posean arquitecturas cerradas que derivan en monopolios, favoreciendo la interoperabilidad entre dispositivos de varios fabricantes y la flexibilidad del mercado.

2.2 Estándares y regulación.

Existen dos tipos de estándares:

- De facto: son estándares con gran aceptación en el mercado, establecidos normalmente por grupos de empresas y organizaciones, pero que aún no son oficiales.
- De iure: son estándares definidos por organizaciones o grupos oficiales.

Puede ocurrir que una empresa o corporación posea una normativa establecida para el desarrollo de sus productos y servicios, siendo ésta propiedad absoluta de la empresa o corporación. Esta manera de actuar es seguida por muchas empresas con la intención de atar a los clientes a sus productos. A esta normativa con frecuencia se le denomina “estándar propietario”, y si alcanza una penetración en el mercado considerable, puede llegar a convertirse en estándar de facto e incluso de iure.

En este sentido, los estándares pueden clasificarse, atendiendo a la propiedad, en dos tipos, abiertos y cerrados. Al primer tipo pertenecen los estándares de facto y iure, ya que pueden ser consultados por cualquiera. No obstante, existen organismos que cobran una cuota por acceder a sus estándares prohibiendo su distribución. A este tipo de estándares se les denomina estándares de distribución restringida. En el otro extremo se sitúan los estándares cerrados, también

denominados propietarios, que representan normas únicamente accesibles para los miembros de la empresa propietaria.

Centrándose en los estándares abiertos, existen dos tipos de organizaciones que pueden definirlos, los consorcios de fabricantes y los organismos oficiales.

Los consorcios de fabricantes están formados por grupos de empresas que cooperan para establecer acuerdos y reglas que permitan obtener la interoperabilidad de sus productos empleando una tecnología determinada. Asegurando dicha interoperabilidad, se consigue un aumento del mercado que se traduce en un mayor número de clientes potenciales para sus productos. En este caso, las empresas o personas interesadas pueden unirse al consorcio y participar en los grupos de trabajo que definen los documentos técnicos de la norma. ADSL Forum, ATM Forum, Zigbee Alliance, y PLC forum son ejemplos de consorcios de este tipo.

Por otra parte, los organismos oficiales están formados por consultores independientes, miembros de los departamentos o secretarías de estado de diferentes países y otros miembros, son ejemplos de organismos oficiales:

- **ISO (International Organization for Standardization):** La organización internacional para la normalización es una agencia internacional sin ánimo de lucro con sede en Ginebra (Suiza), cuyo objetivo es el desarrollo de normalizaciones que abarcan un amplio abanico de materias. Esta organización ha definido multitud de estándares de diferentes temáticas, que van desde el paso de los tornillos hasta arquitecturas de comunicaciones para la interconexión de sistemas abiertos (OSI - Open Systems Interconnection).

  ISO está formada por organismos de estandarización de diversos países (ANSI en EEUU, DIN en Alemania, AENOR en España) y por un grupo de organizaciones observadoras, que no poseen capacidad de voto. A pesar de ser una organización no gubernamental, la mayoría de sus miembros son instituciones gubernamentales. Se fundó en 1946 y actualmente reúne a más de 100 países.

- **IEEE (Institute of Electrical and Electronic Engineers):** IEEE es la mayor asociación profesional para el avance de la innovación y la excelencia tecnológica en busca del beneficio de la humanidad. IEEE y sus miembros inspiran una comunidad global que innove hacia un mejor mañana a través de sus publicaciones enormemente citadas, conferencias, estándares tecnológicos, y actividades profesionales y educativas. Fue fundada en 1884 y desde entonces desarrolla estándares para las industrias eléctricas y electrónicas. Desde el punto de vista de las redes de datos son muy interesantes los trabajos del comité 802, que desarrolla estándares de protocolos de comunicaciones para la interfaz física de las conexiones de las redes locales de datos.
- **IETF (Internet Engineering Task Force):** Este Grupo de Trabajo de Ingeniería de Internet, es una organización internacional abierta de normalización, que tiene como objetivos el contribuir a la ingeniería de Internet, actuando en diversas áreas, como transporte, encaminamiento, seguridad. Fue creada en EE.UU. en 1986. El IETF es mundialmente conocido por ser la entidad que regula las propuestas y los estándares de Internet, conocidos como RFC (Request For Comments).

  Es una institución sin fines de lucro y abierta a la participación de cualquier persona, cuyo objetivo es velar para que la arquitectura de Internet y los protocolos que la conforman funcionen correctamente. Se la considera como la organización con más autoridad para establecer modificaciones de los parámetros técnicos bajo los que funciona la red. El IETF se compone de técnicos y profesionales en el área de redes, tales como investigadores, integradores, diseñadores de red, administradores, vendedores, entre otros.

  Dado que la organización abarca varias áreas, se utiliza una metodología de división en grupos de trabajo, cada uno de los cuales trabaja sobre un tema concreto con el objetivo de concentrar los esfuerzos.
- **ANSI (American National Standards Instituto):** El Instituto Americano de Normas Nacionales. Organización sin ánimo de lucro encargada de supervisar el desarrollo de estándares que se aplica en los Estados Unidos de América.

- **TIA (Telecommunications Industry Association):** La Asociación de la Industria de las Telecomunicaciones Organización formada por representantes de las industrias más importantes del sector de las telecomunicaciones y que ha desarrollado también numerosos estándares a nivel internacional relacionados con el mundo de las redes en colaboración con ANSI y la antigua EIA
- **ETSI (European Telecommunications Standardas Institute**): Las siglas ETSI hacen referencia al instituto europeo de estándares de las telecomunicaciones.
  ETSI es una organización independiente sin ánimo de lucro que produce estándares aplicables globalmente para las tecnologías de la información y comunicación. Este instituto es reconocido por la Unión Europea como una organización de estándares europeos. Posee 766 organizaciones miembro procedente de 63 países de los cinco continentes.
  El ETSI ha tenido gran éxito al estandarizar el sistema de telefonía móvil GSM.
  Cuerpos de estandarización significativos dependientes del ETSI son 3GPP (para redes UMTS) o TISPAN (para redes fijas y convergencia con Internet).
  El ETSI fue creado en 1988.
- **CEN (Comité Europeo de Normalización):** En francés Comité Européen de Normalisation, es una organización no lucrativa privada cuya misión es fomentar la economía europea en el negocio global, el bienestar de ciudadanos europeos y el medio ambiente proporcionando una infraestructura eficiente a las partes interesadas para el desarrollo, el mantenimiento y la distribución de sistemas estándares coherentes y de especificaciones.
  El CEN fue fundado en 1961. Sus veintinueve miembros nacionales trabajan juntos para desarrollar los estándares europeos (EN) en varios sectores.

2.3 Estándares de Internet

El *World Wide Web Consortium* (*W3C*) desarrolla Estándares Web o Recomendaciones que tienen por finalidad conseguir que las tecnologías que conforman la Web sean interoperables, eficientes, confiables, accesibles y fáciles

de usar, lo que a su vez repercutirá en el desarrollo de aplicaciones cada vez más robustas.

Estas recomendaciones son el fruto de un proceso neutro, transparente y consensuado en el que toman parte los miembros del W3C (más de 400 organizaciones en la actualidad), su equipo de trabajo, expertos y aquellos usuarios de la Web que deseen colaborar.

Los Estándares Web han surgido de la necesidad de evitar la fragmentación de la Web así como de mejorar la organización de la información ofrecida en ella, y muchos de ellos han ido sentando las bases de su desarrollo y fomentando su éxito. Algunos de los estándares Web más conocidos y ampliamente utilizados son el lenguaje de etiquetado para hacer páginas Web HTML (*HyperText Markup Language*), el lenguaje para crear estructuras de documentos XML (*eXtensible Markup Language*), y el lenguaje de hojas de estilos CSS (*Cascading Style Sheets*), que permiten controlar la presentación de los documentos (X)HTML.

El primer paso a la hora de crear un Estándar Web es llevar cabo un proceso inicial controlado en el que intervienen todos los usuarios de las tecnologías, con el fin de aportar conocimientos y opiniones que contribuyan a la mejora de los documentos.

A continuación, se obtienen unos estándares de calidad, los cuales pueden emplearse de forma libre en la comunidad Web al estar sujetos a la Política de Patentes del W3C, mientras que las especificaciones sufren un proceso de refinamiento exhaustivo antes de que se consideren *Recomendaciones*. Al utilizar las mismas tecnologías, las máquinas se entienden entre sí y cualquier usuario puede interactuar con el resto.

Ventajas de la Utilización de Estándares Web

A continuación se indican las principales ventajas que conlleva la aplicación de Estándares en el desarrollo de un sitio Web:

- **Código más sencillo**: Un código limpio, válido, modular y semánticamente correcto facilita su comprensión y reutilización por parte de cualquier

desarrollador, ayudando asimismo a que las aplicaciones puedan convertirlo de forma sencilla a otro formato.

- **Compatibilidad**: Los Estándares Web garantizan la compatibilidad del código independientemente del navegador o plataforma empleado. Además, se consigue una mayor estabilidad del sitio Web de cara al futuro y a la aparición de nuevas herramientas.
- **Mejora de la accesibilidad**: Los Estándares Web ayudan a hacer el contenido de un sitio Web accesible a un mayor número de usuarios, independientemente del idioma, localización geográfica, cultura, limitación técnica, física, psíquica o sensorial de éstos, cumpliéndose las directrices y sin que se sacrifique el aspecto visual o el rendimiento del mismo.
- **Mejora del posicionamiento**: Los sitios Web desarrollados en base a Estándares tendrán una mejor posición en los motores de búsqueda. En el caso de que se emplee un código complejo, los robots de búsqueda localizarán e indexarán los contenidos con más dificultad.
- **Mejor adaptación al dispositivo final**: El empleo de Estándares permite que la información sea interpretada por diferentes tipos de dispositivo (navegadores visuales y sólo textos, lectores de pantalla, lectores Braille, dispositivos móviles, etc).
- **Mejor adaptación al usuario**: El usuario puede ajustar la presentación del sitio según sus preferencias o necesidades.
- **Mejora en la impresión**: A través de los Estándares se proporciona de una forma sencilla versiones para imprimir de todas las páginas Web.
- **Mejora del mantenimiento**: La separación de contenido y presentación mediante el empleo de hojas de etilo CSS facilita futuros cambios. Así, resulta más sencillo efectuar modificaciones en un único documento (CSS) que en todas las páginas (documentos (X)HTML) en las que se hayan incluidos estilos.
- **Ahorro de ancho de banda y carga de páginas más rápida:** Los sitios basados en Estándares hacen uso de un menor ancho de banda, lo cual

implica a su vez un ahorro en los gastos de alojamiento Web. Por otra parte, la adecuación gramatical de las páginas de un sitio, contribuye a que se muestren más rápido a los usuarios, lo que mejora la experiencia de éstos.

- **Mayor confianza en la Web:** La Web es un medio colaborativo, donde los usuarios interactúan y se relacionan, siendo necesaria la confianza entre sí. Para ello, se han desarrollado tecnologías como las firmas digitales de documentos, la encriptación de datos confidenciales o las políticas de privacidad de datos de los sitios Web.
- **Mayor carga semántica:** Se proporcionan mecanismos para añadir significado a los recursos, haciendo posible que una máquina pueda interpretar los datos de la Web de forma análoga como lo hacen los seres humanos. De este modo, también se consigue una mejora del rendimiento y eficiencia de la Web, beneficiando a los usuarios a través de una mayor precisión en sus búsquedas y operaciones.

- **Competitividad:** La aplicación de Estándares aporta una mayor ventaja competitiva en el mercado.

Estándar XML

XML es una especificación de carácter genérico derivada del Estándar SGML (Standard Generalized Markup Language) que permite definir lenguajes de marcado. Es lo que se denomina un metalenguaje: no se usa directamente, sino que sirve para definir otros lenguajes. Su importancia reside en su capacidad para expresar el significado de un contenido con independencia del formato de documento final que se presente al usuario gracias a una serie de etiquetas.

Un documento XML puede ser procesado por un sistema automático o transformado en un formato adaptado al usuario. No se ve limitado por las características o capacidades del usuario ni la forma de presentación. Por su flexibilidad, XML es aplicable a una gran diversidad de campos como pueden ser el intercambio de mensajes de datos entre diferentes sistemas, dibujos vectoriales, correo por voz,

subtítulos para multimedia, fórmulas matemáticas, partituras de música, y páginas Web.

Por ser un formato estandarizado existe un gran conocimiento y mucha experiencia en su uso. Existen numerosas herramientas para el procesamiento y transformación de XML desde editores de etiquetas hasta aplicaciones especializadas para el dominio de una aplicación concreta como podría ser un programa de dibujo. Todos trabajan con el mismo formato subyacente.

XML se establece como una tecnología que se rodea de un conjunto de tecnologías paralelas que la complementan, entre las cuales caben destacar las siguientes:

- XSL: Familia de lenguajes basados en el estándar XML (XSLT, XSL-FO y XPath) que permite definir una presentación o formato para un documento XML.
- XSLT: lenguaje empleado para transformar la información en el formato final más apropiado para el usuario.
- XSL-FO: lenguaje que permite describir la forma en que se presentan los componentes de un documento XML.
- XPath: Lenguaje que permite identificar de forma inequívoca cualquier elemento o atributo de un documento XML.
- XLink: lenguaje creado para poder definir de forma estándar hipervínculos en archivos XML.
- XPointer y XFragments: lenguajes para apuntar a partes de un archivo XML.
- XQuery: lenguaje de consulta similar a SQL para colecciones de datos XML.
- XSchema: lenguaje de esquema empleado para describir la estructura y contenido adecuados de los elementos incluidos en los documentos XML.
- CSS: lenguaje de hojas de estilos que permite controlar la presentación de documentos (X)HTML y XML.

A continuación se detallan las principales características del lenguaje XML:

- XML permite guardar la información en un formato independiente del documento final que recibe el usuario.

- Posee una estructura sencilla que facilita su comprensión, aprendizaje y empleo.
- Posee una arquitectura abierta y extensible a través de la definición de nuevas etiquetas, lo que garantiza su correcto funcionamiento bajo cualquier tipo de navegador (antiguo, presente o futuro).
- XML se establece como Estándar para el intercambio de información estructurada entre diferentes aplicaciones y plataformas de un modo sencillo, seguro y fiable.
- Se trata de un lenguaje flexible que agrupa un amplio abanico de aplicaciones (páginas Web, bases de datos, etc).
- XML marca la semántica o significado de cada elemento: sea una persona, un código de barras, o un círculo, describiendo las relaciones entre los elementos.
- Se encuentra estructurado, lo que permite el modelado de datos de diferentes niveles de complejidad y facilita su procesamiento.
- Los documentos XML pueden ser validados contra una DTD.
- Mediante el empleo de XML se obtiene un comportamiento más estable y actualizable de las aplicaciones Web.
- Los documentos XML proporcionan metainformación sobre sí mismos, lo que repercutirá en búsquedas más precisas.
- El análisis de un documento XML es un proceso estandarizado, lo que permite utilizar cualquier analizador, evitando de este modo errores y optimizando el desarrollo de aplicaciones.

Estándar HTML

HTML (*HyperText Markup Language*) es el lenguaje de marcado empleado universalmente para crear páginas Web. Se trata de un lenguaje de hipertexto constituido por un conjunto de etiquetas que marcan la apertura y el cierre de cada elemento, mediante el cual es posible incluir de forma estructurada textos, imágenes, objetos programados y scripts.

El hecho de que HTML sea un Estándar del W3C, permite que cualquier página Web creada a través de dicho lenguaje pueda ser visualizada de forma homogénea, con independencia del navegador o plataforma empleados (siempre que estos sean fieles a los estándares).

Así, los orígenes del HTML se remontan a 1980, año en el que Tim Berners-Lee, trabajador del CERN (*European Laboratory for Particle Physics*), comienza a elaborar un sistema de hipertexto para Internet, no siendo hasta el 1990 cuando definiera el lenguaje HTML como un subconjunto del poderoso lenguaje de etiquetado SGML. En 1991, Tim Berners-Lee publica la primera descripción formal de HTML, conocida como HTML Tags, en la que se recogen los 22 primeros elementos del lenguaje.

En 1993, el organismo IETF (*Intenet Engineering Task Force*) elabora una propuesta para estandarizar HTML. Si bien, no se llega a establecer como Estándar ninguna de las dos propuestas existentes en el momento (HTML y HTML+). Será el 22 de Septiembre del año 1995 cuando el IETF logre publicar el Estándar HTML 2.0 como primer Estándar oficial de HTML, creado con fines divulgativos y académicos, y donde prevalecía el contenido por encima del diseño.

Con todo, HTML 2.0 no permitía controlar el diseño de las páginas ni añadir elementos multimedia, a lo que la empresa Netscape responde definiendo nuevas etiquetas en el estándar. Por otro lado, el consorcio internacional W3C, creado en Marzo de 1995, comenzó a desarrollar un borrador para la versión HTML 3.0, no siendo bien acogido debido al elevado número de elementos y atributos que se definieron en él, lo que le hacía muy complejo para poder desarrollarse mediante la tecnología del momento y finalmente fue abandonado.

El 14 de Enero de 1997 es la fecha elegida por el W3C para publicar HTML 3.2, que es oficialmente su primera recomendación. En ella se abandonan muchas de las características de HTML 3.0 y se incluyen los últimos avances desarrollados por los navegadores *Internet Explorer* y *Netscape Navigator*, como por ejemplo los applets de Java o el texto flotado.

Sin embargo, el avance más notorio se observa en HTML 4.0, recomendación publicada por el W3C el 18 de Diciembre de 1997 y revisada el 24 de Abril de 1998. Mediante esta versión de HTML se pretende dar soporte a marcos, hojas de estilo CSS, scripts y tablas complejas. También se introducen mejoras en los formularios y en la accesibilidad general de las páginas, así como a nivel de código, especificándose un conjunto de elementos desaprobados y obsoletos. Posteriormente, esta versión sufre una revisión que da lugar a HTML 4.01, la última especificación oficial de HTML, publicada el 24 de Diciembre de 1999 y que es muy similar a su antecesora.

A partir de este momento, el W3C deja aparcado el desarrollo del Estándar HTML para centrarse en una nueva vía, el XHTML (*eXtensible HyperText Markup Language*), una versión más estricta y limpia de HTML preparada para su uso con herramientas basadas en XML. Este cambio de rumbo motiva, de la mano de integrantes de Mozilla Foundation, Opera Software y Apple, la creación en el año 2004 de la asociación WHATWG (*Web Hypertext Application Technology Working Group*), cuyo objetivo es implementar el nuevo Estándar HTML 5, del que ya existe un borrador desde el 22 de Enero de 2008.

Este nuevo contexto hace que el W3C retome el desarrollo de HTML en Marzo del 2007, si bien se trabaja de forma simultánea en la implementación de XHTML, publicándose su primera recomendación, el **XHTML 1.0** el 26 de Enero de 2000, y su segunda recomendación, el XHTML 1.1 el 31 de Mayo de 2001, que es una versión modularizada de XHTML 1.0. Por último, cabe destacar que también existe un borrador de la novedosa especificación **XHTML 2.0**, la cual data del 26 de Julio del 2006.

2.4 La Unión Internacional de Telecomunicaciones

En 1865 veinte países forman la International Telegraph Union (ITU). En 1885 la ITU asume las competencias en la regulación de la conexión telegráfica internacional. En 1906 se celebra la primera conferencia radiotelegráfica

internacional, y a partir de entonces la ITU fue normalizando el funcionamiento de las grandes redes públicas de conmutación (telefonía, radiodifusión, televisión, etc.)
La UIT tiene su sede en ginebra y es dependiente de las Naciones Unidas (191 estados miembros) desde 1947. De los 5 órganos importantes en la ONU uno de ellos es el económico y social que se encarga de fomentar grandes estándares que afecten al progreso económico y social, para ello, se apoya en agencias especializadas, como la UIT.
Actualmente la UIT tiene tres órganos principales, que se ocupan sobre todo de la difusión internacional de radio y de los sistemas telefónicos y de comunicación de datos. La UIT tiene varias clases de miembros: administraciones de correos y teléfonos nacionales, organizaciones científicas e industriales, otras organizaciones internacionales.
La organización ITU (UIT en castellano, Unión Internacional de Telecomunicaciones) es la organización más importante de las Naciones Unidas en lo que concierne a las tecnologías de la información. Esta organización representa un foco global para los gobiernos y el sector privado en el desarrollo de redes y servicios. ITU coordina el uso del espectro radioeléctrico, promoviendo la cooperación internacional para la asignación de órbitas de satélites, trabajando para mejorar las infraestructuras de comunicación mundiales, estableciendo estándares mundiales para la interconexión de un enorme rango de sistemas de comunicación, y haciendo frente a problemas actuales, como el cambio climático y la seguridad en el ciberespacio.
Esta organización está compuesta por tres sectores o comités:

- **ITU-R**: que se encarga de promulgar estándares de comunicaciones que emplean el espectro electromagnético.
- **ITU-D:** que se encarga de la organización, coordinación técnica y actividades de asistencia.
- **ITU-T:** que se encarga de desarrollar estándares para la telefonía, la telegrafía, interfaces, redes y otros aspectos de las telecomunicaciones.

Las recomendaciones de la ITU-T se agrupan en series que tratan sobre distintos temas. Algunas de ellas son:

- Serie B: Significado de símbolos, definiciones
- Serie C: Estadísticas generales de telecomunicación.
- Serie D: Principios de tarificación.
- Serie F: Otros servicios no telefónicos.
- Serie G: Sistemas y medios de transmisión, sistemas digitales y redes.
- Serie H: Líneas de transmisión de señales no telefónicas.
- Serie I: Red digital de servicios integrados.
- Serie J: Transmisión de señales de sonido y T.V.
- Serie P: Calidad de transmisión telefónica, instalación de teléfonos.
- Serie Q: Señalización y conmutación.
- Serie R: Transmisión telegráfica.
- Serie T: Características de los terminales y protocolos de alto nivel telemático.
- Serie U: Conmutación telegráfica.
- Serie V: Comunicación de datos sobre línea telefónica (V.24).
- Serie X: Redes de datos y comunicación de sistemas abiertos.
- Serie Z: Lenguajes de programación.

Las normas de la ITU-T se denominan recomendaciones, y se identifican con un número tras la serie correspondiente. Algunas normas de la ITU-T, son:

G.711: Modulación por impulsos codificados (MIC) para frecuencias vocales

G.729: Codificación de la voz a 8kb/s.

G.732: Características del equipo multiplexor MIC primario a 2048 kb/s

G.774: Jerarquia Digital Síncrona (SDH) –

G.991.1: Transceptores de línea digital de abonado de alta veloc. binaria (HDSL)

G.992.1: Transceptores de línea de abonado digital asimétrica (ADSL)

V. 34: Modem de 33.600 bit/s para uso en la red telefónica conmutada

V.90: Modem de datos a 56.000 bit/s en sentido descendente y 33.600 en ascendente

V.92: Mejoras a la V.90 que permiten conseguir 48.000 bit/s en ascendente.

H.323: Sistemas de comunicación multimedia basados en paquetes (VoIP)

Q.931: Señalización de RDSI

La ITU-R también elabora recomendaciones sobre técnicas de gestión de espectro, servicio fijo por satélite, servicio de radiodifusión…

2.5 Estándares IEEE 802

La IEEE fue fundada en 1963. Es la organización profesional más grande del mundo. Esta institución de origen estadounidense cuenta en la actualidad con más de 380.000 miembros en 150 países distintos. El IEEE engloba, entre otras, áreas tan diversas como la ingeniería de computadoras, tecnología biomédica y telecomunicaciones.

Además de publicar numerosas revistas, que representan el 30% de las publicaciones a nivel mundial sobre ingeniería eléctrica, computadoras y control automático, anualmente programa un número importante de conferencias (más de 300 conferencias técnicas) y cuenta con 900 estándar activos.

Las normas 802 del IEEE, para una red de área local ha tenido especial relevancia, con normas como 802.3 Ethernet, 802.5 Token Ring, 802.11 Wireless LAN.

En total el comité 802 está formado por 13 grupos de trabajo que podemos agrupar de la siguiente manera:

- 802.1: Panorámica y Arquitectura, Puentes, redes locales virtuales (VLANs).
- 802.2: LLC, Logical Link Control (actualmente en hibernación e inactivo).
- 802.3,.4,.5,.6,.9,.11,.12,.14: métodos de acceso y señalización física para tipos concretos de tecnologías LAN y MAN.
- 802.7 y 802.8: Grupos técnicos asesores en redes de banda ancha y en fibras ópticas, respectivamente.(actualmente en hibernación e inactivo)
- 802.10: Niveles de seguridad en estándares 802

Los grupos de trabajo especializados en métodos de acceso corresponden a las siguientes tecnologías:

- 802.3: CSMA/CD (Etherent)

- 802.4: Token Bus (actualmente en hibernación e inactivo)
- 802.5 Token Ring
- 802.6: DQDB, Distributed Queue Dual Bus (actualmente en hibernación e inactivo)
- 802.9: Servicios Integrados (Iso-Ethernet)
- 802.11: Redes inalámbricas
- 802.12: Demand Priority (100VG-AnyLAN)
- 802.14: Redes de televisión por Cable (actualmente en desarrollo del primer estándar)

A título de ejemplo se detalla a continuación algunos de los proyectos más relevantes del comité 802:

- 802.1D: puentes transparentes
- 802.1p: Filtrado por clase de tráfico (Calidad de Servicio)
- 802.1Q: Puentes en redes locales virtuales
- 802.3u: Fast Ethernet
- 802.3x. Ethernet Full dúplex y control de flujo
- 802.3z: Gigabit Ethernet
- 802.3ab: Gigabit Ethernet en cable UTP-5 (en desarrollo)

Todos los estándares IEEE 802 son más tarde aprobados por ANSI y por la ISO. El estándar IEEE 802.x tiene un estándar equivalente ISO 8802-x Normalmente un estándar IEEE es aprobado más tarde por ISO bajo la denominación 8802.x, convirtiéndose así en estándares internacionales; así por ejemplo el estándar ISO 8802.3 es equivalente al IEEE 802.3

Bibliografía

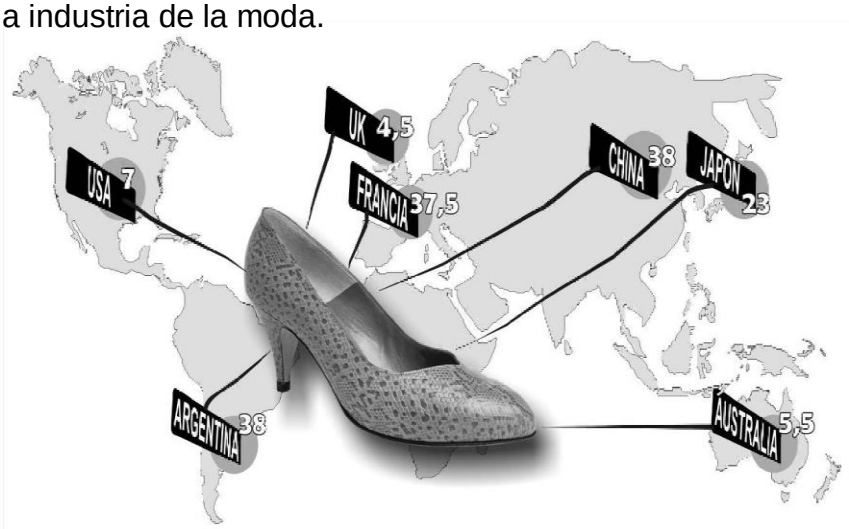
- José Antonio Merlo Vega. Organizaciones de Normalización en Internet. Revista Española de Documentación Científica, jul.-sep. 2000, vol 23, n.2, p. 327-340.
- Guía Práctica de Comprobación de Accesibilidad: ESTÁNDARES WEB. 2010. Instituto Nacional de Tecnología de la Comunicación (INTECO). España.

Caso de Estudio

1. Importancia de los estándares

Los estándares son acuerdos que estructuran cualquier actividad o industria. Son reglas o guías que todos aplican. Asimismo, constituyen una forma de medir, describir o clasificar productos o servicios.

Una de las formas más sencillas de entender la utilidad de los estándares es pensar en lo que ocurre cuando ellos no existen o no se aplican normas. Tomemos el ejemplo del tamaño del calzado. Un zapato de mujer que sea número 7 en Nueva York , será un tamaño 38 en Shangai, un número 4,5 en Londres, un 37,5 en París, un número 23 en Tokio, un 5,5 en Sydney y un 38 en Buenos Aires. Esto resulta inconveniente y dificultoso para un turista que desee ir de compras, pero resulta increíblemente inconveniente y dificultoso para las compañías que fabrican calzado o están en la industria de la moda.



Debido a que no existen estándares globales para los tamaños de calzado, las empresas tienen que marcar los mismos zapatos de manera diferente y deben especificar la referencia del tamaño de manera correcta en todas las órdenes de compra, facturas y remitos de entrega para cada país. Además, dado que lleva más tiempo prestar atención a todas estas especificaciones, resulta más engorroso desde el punto de vista de los procesos de producción. Acarrea mayores costos en las fábricas que luego deben ser trasladados a los consumidores y todo ello se traduce en un calzado con precios más altos para el comprador final.

1. Aplique el ejemplo de los calzados a los estándares que regulan las telecomunicaciones, y explique la importancia de su implementación en estas.

Módulo 3. Calidad de Servicio en Internet

Objetivos:

- Conocer algunos de los requerimientos, los métodos y técnicas para la calidad de servicio en internet.
- Reconocer las características principales de los mecanismos y arquitecturas que podemos implementar para ofrecer QoS en la red de internet.

¿De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje abarcamos lo concerniente a la calidad de servicios en internet y las principales arquitecturas, métodos o técnicas que se pueden implementar en la red para que de esta forma los servicios ofrecidos alcanzan un nivel de QoS aceptable.

3. Calidad de Servicio en Internet

Se define la calidad de servicio (CdS o QoS) como la capacidad que tiene un sistema de asegurar, con un grado de fiabilidad preestablecido, que se cumplan los requisitos de tráfico para un flujo de información dado.

La Calidad de Servicio puede ser implementada de diferentes formas, una de ellas es sobre el protocolo de Internet IP, aunque es más conocido comúnmente como TCP/IP. También existe una especificación para implementar el concepto de calidad de servicio en un nivel más bajo de capa de red, en el nivel de enlace con la tecnología ATM, del cual se deriva una variante que también es capaz de hacer diferenciación de flujos en MPLS. Se menciona también como el protocolo de administración de redes SNMP el cual puede usarse para proveer información de calidad de servicio.

La forma más simple de implementar Calidad de Servicio es diferenciando el tráfico en clases de servicio, el cual es definido solamente en el encabezado del paquete de información. Para implementar Calidad de Servicio de una forma más completa se requieren algunas modificaciones extras que pueden incluir el agregar hardware o software en los componentes de la red.

La Calidad de Servicio (QoS) se ha convertido en un asunto que deben tener en cuenta los proveedores de servicios y las redes de área extensa (WAN) de las empresas, que están añadiendo más tráfico de voz y de imágenes al tráfico de datos en constante crecimiento. Por ejemplo, el tráfico esencial y sensible a retrasos, como el tráfico de voz, puede necesitar mayores garantías de calidad de servicio que el tráfico menos sensible a retrasos, como puede ser la transferencia de archivos o el correo electrónico. Es probable que la mayoría de usuarios se hayan encontrado con costes prohibitivos para el acceso WAN de alta velocidad, simplemente ampliar el ancho de banda no es una opción para muchos puntos de una red. De aquí que la utilización óptima y el uso diferenciado del ancho de banda existente se convierte en un asunto de importancia fundamental.

La omnipresencia del Protocolo Internet (IP) en los puestos de trabajo lo ha convertido en el protocolo más utilizado para aplicaciones extremo a extremo de

voz, vídeo y datos que están apareciendo en el mercado. Por lo tanto, el reto de los administradores y los arquitectos de redes ha sido construir redes que puedan soportar estas nuevas aplicaciones de voz, imágenes y datos basadas en IP junto con las aplicaciones tradicionales orientadas a circuitos sobre redes WAN, que utilizan variedad de medios.

Algunos fabricantes ofrecen exhaustivos mecanismos extremo a extremo de QoS que satisfacen estas demandas independientemente del/los medio(s) utilizado(s) para construir su WAN, como por ejemplo ATM, Frame Relay y SONET

3.1 Requerimientos de QoS en Internet

Entre los requerimientos de QoS en Internet se pueden mencionar los siguientes:

1. Cada elemento de red que acepta un requerimiento de servicio de carga controlada debe reservar los recursos suficientes que le permitan cumplir los compromisos de QoS. Los recursos más importantes incluyen el ancho de banda del enlace, el espacio en buffers y la capacidad de cómputo en la sección de direccionamiento de paquetes.
2. Un elemento de red puede usar herramientas estadísticas para decidir la aceptación de un nuevo flujo. Dependiendo de la conducta pasada que hayan mostrado los flujos puede disponer de recursos que le permitan asegurar a un nuevo flujo sin perjudicar a los existentes.
3. Un elemento de red puede hacer uso de los algoritmos de ordenamiento apropiados para cumplir los compromisos de servicio. El algoritmo implementado debe disponer un ancho de banda mayor que el especificado en el TSpec a fin de superar los momentos de ráfaga del flujo, de no hacerlo el retardo de encolamiento se incrementaría permanentemente. El algoritmo de ordenamiento puede implementar esta condición en una forma explícita, "pidiendo prestado" ancho de banda, o en forma implícita mediante técnicas de multiplexaje. Similarmente, la implementación debe reservar mayor espacio en buffers que el especificado en TSpec con el objetivo de reducir las pérdidas de paquetes. Los servicios de carga controlada no rediseñan el

TSpec en cada nodo, distorsionándose éste a su paso por los puntos de encolamiento, distorsión que se produce especialmente en momentos de ráfagas. El algoritmo implementado puede usar técnicas de multiplexaje estadístico para cumplir esta condición.

4. Un dispositivo de red no puede fragmentar los paquetes de un flujo que recibe el servicio de carga controlada. Si los paquetes son más grandes que la máxima unidad de transporte (MTU) del enlace, éstos no son

3.2 Aseguramiento de los recursos

Los modelos de QoS para Internet son estándares abiertos definidos por le IETF, existen dos modelos de calidad de servicio normalizados: IntServ y DiffServ. Estos dos modelos mejoran el servicio sobre las redes IP que siguen un sistema de mejor servicio o Best-Effort el cual se describe en el RFC 1812, Best-Effort presenta complicaciones para la prestación de servicios de red que requieran la transmisión de datos en tiempo real, puesto que la llegada de datos desordenados o la pérdida de información pueden ser críticas. El modelo IntServ, donde las aplicaciones cuyo tráfico requieren tratamiento diferencial señalizan la red para requerir y garantizar los recursos necesarios para el adecuado funcionamiento de la aplicación, y garantiza las condiciones de operación de cada una de las sesiones que se establecen. Y por último el modelo DiffServ, en el cual la infraestructura de la red es la que reconoce los diferentes tipos de tráfico y aplica políticas diferenciadas para cada clase de tráfico, este es más escalable y flexible en su implementación

3.3 Diferenciación de Servicios

Se trata de diferenciar cada paquete y darle un trato dependiendo del servicio que necesite. Los paquetes se marcan y clasifican para recibir un tratamiento específico por salto en la ruta. Esta política de clasificación sólo se implementa en las fronteras de la red y no en los nodos intermedios.

Se trata de dividir los paquetes en distintas clases que requerirán distintos servicios. Los 4 básicos son:

- **PHB por defecto:** Es el menos riguroso y equivale a enviar si se puede y si no, descartar.
- **PHB selector de clase:** Si no hay congestión, se asegura el envío y si la hay, no.
- **PHB de reenvío explícito**: Se garantiza un ancho de banda, se asegura que no hay pérdidas, poca latencia y variación de retardo (para videoconferencia, etc.).
- **PHB de reenvío asegurado:** Se garantiza que no hay pérdida de paquetes.

3.4 Arquitecturas y Mecanismos para proveer QoS en Internet

3.4.1 Asignación de Recursos

Una vez que se tiene el tráfico clasificado, y por tanto se saben qué parámetros de QoS se deben cumplir, hay que asignar los recursos en la interfaz. Hay que permitir que los paquetes se transmitan al medio (el aire o un cable).

La fase de clasificación es común a todos los tipos de interfaz que necesitan garantizar la QoS, pero la principal diferencia viene en la fase de asignación de recursos. Existen dos mecanismo “QoS a nivel 3 (L3QoS o IPQoS)” y “QoS a nivel 2 (L2QoS o MACQoS)”.

- **L3QoS: QoS a nivel IP:** Las técnicas que se usan en este tipo de mecanismos de QoS son los típicos de los conformadores de tráfico o *traffic shapers* (TS). El TS clasifica el tráfico que entra en función de los criterios que se establezcan para cada una de los contratos de QoS. Es también conocida como *QoS a nivel IP*.

  Una vez que el tráfico está clasificado, el TS asigna de una forma estadística los recursos de transmisión al medio. Por ejemplo si la cola de un servicio de baja latencia está muy llena, intentará vaciarla lo más rápido posible o por ejemplo si la cola de un servicio con tasa mínima garantizada tiene paquetes, intentará mantener en promedio a la salida esa tasa.

Estas técnicas de QoS a nivel 3, a veces llamados a nivel IP, son las clásicas basadas en colas de prioridades asociadas al DSCP o al TOS de las cabeceras IP, por ejemplo.

El problema que presenta la técnica L3QoS es que no se conoce con exactitud la capacidad y la disponibilidad del medio sobre el que se transmiten. Imaginemos que tenemos un medio sin cables. El tráfico bruto puede depender del usuario al que se transmita, ya que podrían estar más lejos u obstruidos. Usar técnicas de L3QoS en estos casos, al desconocer la capacidad real por usuario destino, por ejemplo, lleva a una ineficiencia insalvable: “No se puede garantizar una QoS en términos absolutos, solo relativos”. Esto quiere decir que si tenemos un servicio de 1Mbps y otro de 2Mbps, la única garantía que puede hacer un sistema de L3QoS es que el tráfico del primero va a ser la mitad que el del segundo, pero no puede garantizar cuál va a ser en realidad ese mínimo, ya que desconoce el estado y disponibilidad del medio. Este problema aún se agrava mucho más en el caso en el que el medio está gestionado en contienda (WiFi, ethernet…) En estos casos el propio uso del medio es estadístico, ni siquiera el nivel 2 puede saber si podrá transmitir en un momento dado. Es más, en el caso de que la red empiece a cursar mucho tráfico, es posible que un paquete jamás sea transmitido debido a las continuas colisiones.

- **L2QoS: QoS a nivel MAC:** Cuando la asignación de recursos se hace a nivel 2, el sistema que va asignando los slots de transmisión conoce en todo momento tanto la disponibilidad del medio como la calidad o tráfico neto que es capaz de transmitir para cada usuario. Es también conocida como *QoS a nivel MAC*. Esto hace posible implementar algoritmos que permitan garantizar de forma absoluta la asignación de tráfico. Por ejemplo, WiMAX es un sistema de L2QoS. La estación base es el nodo maestro de la red, que asigna la transmisión de datos tanto en la bajada hacia los usuarios (Downlink) como en la subida desde los usuarios (Uplink). El tener un nodo central permite eliminar la contienda, lo que garantiza que la BS puede, si así

se desea, conocer en todo momento la disponibilidad del medio radio. Además la BS WiMAX conoce la calidad del enlace de cada uno de los clientes que tiene conectados, con lo que puede asignar de una forma totalmente determinista el tráfico, tanto en bajada como en subida. Por supuesto la calidad de servicio a nivel MAC no es exclusiva de WiMAX, por ejemplo DVB-RCS, un protocolo estándar para el acceso múltiple vía satélite, es un esquema parecido: un nodo central que asigna tráfico, un conocimiento exhaustivo de la capacidad y disponibilidad del medio… una QoS que se puede garantizar.

3.4.2 Optimización del Rendimiento

Ingeniería de tráfico

La Ingeniería de tráfico de Internet está definida como el aspecto de la Ingeniería de la Red de Internet que trata con el problema de la evaluación y optimización del rendimiento de las Redes IP operativas. La Ingeniería de Tráfico abarca la aplicación de tecnología y principios científicos para la medición, caracterización, modelado y control del Tráfico de Internet. [RFC-2702, Awduche2] en [RFC-3272]. Otra definición, según Jesús García “La Ingeniería de Tráfico se define como el proceso de controlar los flujos de datos a través de una red”, es decir, es el proceso de optimizar la utilización de los recursos disponibles por parte de los distintos flujos y por tanto, optimizar el uso global de los recursos y las prestaciones de la red [XIAO, 1999] y [XIAO,2000].

La Ingeniería de Tráfico, trata de resolver, uno de los mayores problemas de las redes IP actuales: ajustar el tráfico IP para hacer un mejor uso del ancho de banda, así como enviar flujos específicos por caminos específicos. IETF, ha propuesto varias técnicas para proveer calidad de servicio QoS en Internet. Las redes como Internet, tienen tres características significativas:

1) proporcionan servicios en tiempo real,

2) son de misión crítica, y

3) sus entornos operativos son muy dinámicos, desde este punto de vista, resulta complejo modelar, analizar y resolver los problemas asociados al mantenimiento, gestión y afinamiento de las redes.

Objetivo de la Ingeniería de Tráfico

El objetivo global de la IT, es mejorar el rendimiento de una red operacional, minimizando la congestión en una red al mismo tiempo que se intenta incrementar la eficiencia de la utilización de sus recursos.

Congestión en una red

Cuando muchos paquetes están presentes en la red, su rendimiento se reducirá. Esta situación es llamada congestión. Cuando el número de paquetes acumulados en la red por los hosts está dentro de su capacidad de carga, ellos son todos entregados y el número entregado es proporcional al número enviado. Sin embargo como el tráfico incrementa más, el router pierde los paquetes. Esto tiende a hacer los problemas peores. En un tráfico alto, el rendimiento colapsa completamente y casi todos los paquetes no son entregados.

La congestión en una red, puede deberse a muchos factores:

- Insuficiencia de recursos en la red (por ejemplo, capacidad de enlaces).
- Utilización ineficiente de los recursos debido al mapeado del tráfico.

El primer caso, se podría resolver aumentando la capacidad de los recursos; para el segundo caso, la Ingeniería de Tráfico adapta los flujos de tráfico a los recursos físicos de la red, tratando de equilibrar de forma óptima la utilización de esos recursos, de manera que no haya algunos que estén subutilizados, y otros sobre utilizados que crean cuellos de botella. Solucionar los problemas de congestión en costos razonables es uno de los mayores objetivos de la ITE.

Tareas de la Ingeniería de Tráfico

Caracterización de la demanda de tráfico

Se realiza mediante modelos que aproximan el comportamiento estadístico de la red. Los *modelos de tráfico,* adoptan suposiciones simplificadas acerca de los procesos de tráfico que usualmente son complicados.

Usando estos modelos, el tráfico se caracteriza por un conjunto limitado de parámetros (solamente los parámetros que sean relevantes para determinar el impacto de la demanda de tráfico sobre el rendimiento de la red, por ejemplo: media, varianza, índice de dispersión de contadores, etc.)

Las *métricas de tráfico* son definidas para validar los modelos; éstas métricas estiman el valor de los parámetros por cada segmento de red durante cada período de tiempo. Como complemento al modelado de tráfico y métricas de tráfico, se requiere un *sistema de predicción de tráfico* para propósitos de planeación y dimensionamiento, esto permitirá pronosticar las demandas de tráfico según períodos de tiempo anteriores.

Objetivos del Grado de Servicio (GoS, Grade of Service)

Los objetivos del Grado de Servicio GoS se derivan de los requerimientos de Quality of Service (QoS). La calidad de servicio (QoS) es el rendimiento de extremo a extremo de los servicios electrónicos tal como lo percibe el usuario final. Los parámetros de QoS son: el retardo, la variación del retardo y la pérdida de paquetes. El GoS, se refiere al número de parámetros de ingeniería de tráfico que proporcionan una medida adecuada o suficiente de la infraestructura bajo condiciones específicas; estos parámetros GoS puede ser expresados como una probabilidad de bloqueos o de retardos, etc. El bloqueo y el retardo pueden ser causados por: la capacidad limitada de manejo de tráfico de la red o de un componente de ella y la naturaleza estocástica de la demanda de tráfico.

Controles de tráfico y dimensionamiento

Una vez que las demandas de tráfico han sido caracterizadas y los objetivos de GoS han sido establecidos, la IT debe proveer un diseño de operación de la red que garantice el soporte de la demanda de tráfico mientras los objetivos de GoS son satisfechos.

Las entradas para el diseño y operación de la red son: el dimensionamiento de la red y los controles de tráfico. El *dimensionamiento* (de los elementos de la red física y lógica) asegura que la red tenga suficientes recursos para atender la demanda de tráfico. Los *controles de tráfico,* incluyen: enrutamiento de tráfico, controles de

gestión de tráfico de red, métodos de protección de servicio, supervisión de tráfico a nivel de paquetes, controles de señalización y redes inteligentes.

Monitoreo del rendimiento

Una vez que la red es operacional, se requiere un monitoreo continuo de los GoS. Aunque la red sea correctamente dimensionada, hay situaciones de sobrecarga y fallos no considerados (sobre todo cuando se toman acciones de gestión de tráfico en períodos de tiempo cortos, minutos, horas). El monitoreo del GoS es necesario para detectar errores o aproximaciones incorrectas durante el dimensionamiento y para producir una retroalimentación para la caracterización de tráfico y diseño de la red. Dependiendo de los problemas detectados, las reconfiguraciones de la red, los cambios en los patrones de enrutamiento o el ajuste de los patrones de control de tráfico, se pueden realizar en plazos de tiempo medios (semanas, meses).

Bibliografía:

- García T. (2007).”Análisis de los Modelos de Servicios Diferenciales y Servicios Integrales para Brindar QoS en Internet”. Tesis para optar al título de Ingeniero en Computación. México.
- España. M. (2003). Servicios Avanzados de Telecomunicación. Ediciones Díaz de Santos, S.A. España
- Valdiviezo, J. (2012). Propuesta de una mejor alternativa de cumplimiento de servicio (QoS) en internet, para la utilización de los servicios diferenciados (DiffServ) y los servicios integrados (IntServ). Pontificia Universidad Católica del Ecuador. Tesis para optar al título de Ingeniero en Sistemas y Computación. Ecuador.
- Cabrera, A., Carrillo, J., Abad, M., Jaramillo, D., y Poma, A. (2015). Modelo de calidad de servicio QoS en entornos Cloud. International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC), Vol. 2, Num. 2, pp. 70-80. Consultado el [02/10/2018] en www.ijisebc.com.
- J.A. Jimenez Toro. UF1875: Gestión de recursos, servicios y de la red de Comunicaciones. Editorial Elearning SL. Edición 5.0. ISBN: 978-84-16199-01-3. España.

Módulo 4. Servicios Integrados (IntServ)

Objetivos:

- Conocer el modelo de servicio de internet IntServ, sus componentes y principales características.
- Distinguir entre aplicaciones elásticas e inelásticas, a través de sus principales características.
- Analizar el funcionamiento del protocolo RSVP y sus herramientas para garantizar QoS.

¿De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje nos enfocamos en el modelo de servicio de internet IntServ y sus componentes para garantizar calidad de servicios a los diferentes tipos de aplicaciones ya sean elásticas o inelásticas. Además haremos un análisis del protocolo RSVP el cual es el encargado de proporcionar los servicios de este modelo de internet.

4. Servicios Integrados (IntServ)

El surgimiento de nuevas aplicaciones multimedia así como su extensa difusión ha sido fomentado principalmente por tres tecnologías:

a) el desarrollo de poderosas estaciones de trabajo equipadas con hardware dedicado a voz y video,

b) el desarrollo de aplicaciones sofisticadas (por ejemplo tele-educación, tele inmersión1, etc.), y

c) la propiedad de multicastde IP.

Sin embargo, la diferencia de retardo de encolamiento y las pérdidas por congestión presentes en el actual Internet son los mayores problemas que tienen que afrontar las aplicaciones en tiempo real, haciéndose necesaria la implementación de calidad de servicio extremo a extremo en tiempo real. Ha surgido además la necesidad de administrar el ancho de banda del enlace controlando la porción de recurso entregado a cada usuario en condiciones de sobrecarga.

Los Servicios Integrados (IntServ) son un modelo de servicio de Internet que incluye el actual servicio best effort, el servicio en tiempo real y la compartición controlada del enlace. La Arquitectura de Servicios Integrados no fue diseñada para reemplazar al básico servicio IP, en cambio añade nuevos componentes y mecanismos que permiten implementar Calidad de Servicio.

En la arquitectura IntServ el concepto de flujo juega un papel fundamental. El flujo se define como un tráfico continuo de datagramas relacionados entre sí que se produce como consecuencia de una acción del usuario y que requiere una misma QoS. Un flujo es unidireccional y es la entidad más pequeña a la que se le puede aplicar una determinada QoS. Los flujos pueden agruparse en clases; todos los flujos pertenecientes a una misma clase reciben la misma QoS.

IntServ define tres tipos de servicios.

- **Servicio Garantizado:** garantiza un caudal mínimo y un retardo máximo. Cada router que se encuentre en el trayecto del datagrama debe ofrecer las garantías solicitadas, aunque a veces no es posible por las características del medio físico.

- **Servicio de Carga Controlada:** este servicio debe ofrecer una calidad comparable a la de una red de datagramas poco cargada, en general deben proporcionar un buen tiempo de respuesta, pero sin garantías estrictas. Eventualmente de pueden producir retardos graves.
- **Servicio Mejor Esfuerzo:** este servicio no tiene ninguna garantía, ni ofrece QoS.

El modelo IntServ dispone del protocolo RSVP (Resource Reservation Protocol, Protocolo de Reservación de Recursos), se basa en este protocolo para señalizar la QoS deseada para cada flujo de datos en la red. Debido a que la información de estados para cada reservación necesita ser mantenida por cada router a lo largo de la ruta del datagrama.

IntServ provee a las aplicaciones de un nivel garantizado de servicio, negociando parámetros de red de punto a punto. La aplicación solicita el nivel de servicio necesario para ella con el fin de operar apropiadamente, y se basa en la QoS para que se reserven los recursos de red necesarios antes de que la aplicación comience a operar. Estas reservaciones se mantienen habilitadas hasta que la aplicación termina o hasta que el ancho de banda requerido por ésta sobrepase el límite reservado para la aplicación.

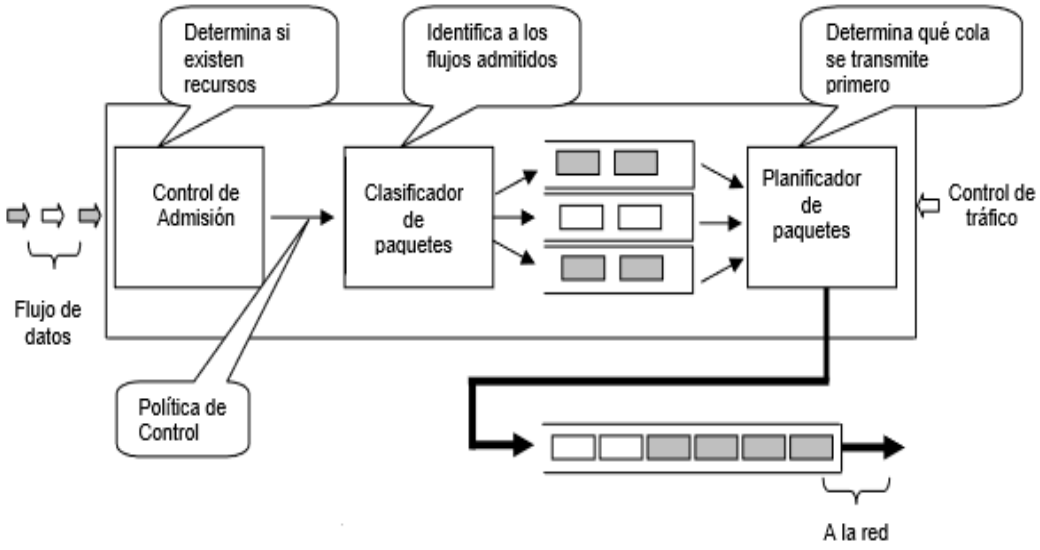


Figura No. 1: Funcionamiento de IntServ

4.1 Aplicaciones Elásticas e Inelásticas

El requerimiento principal de un flujo es la prontitud de envío de sus paquetes, por consiguiente la única forma cuantitativa de entregar calidad de servicio es la fijación de un retardo mínimo y máximo que recibirá el flujo. Mediante la sensibilidad al retardo se puede distinguir dos tipos de aplicaciones: aplicaciones inelásticas y aplicaciones elásticas.

Las aplicaciones inelásticas incluyen la mayoría de aplicaciones playback, que son aquellas cuya forma de funcionamiento implica empaquetamiento, envío, desempaquetamiento y reproducción de una señal original. Para su operación se define el punto de playback correspondiente a un tiempo de compensación desde que se envía el paquete hasta que se reproduce; todos los paquetes que tienen un retardo menor al punto de playback son datos útiles, mientras los paquetes con un retardo mayor son desechados. El retardo de compensación se define en función el retardo de los primeros paquetes enviados o del compromiso de la red cuando ofrece algún nivel de calidad de servicio.

Las medidas del rendimiento de estas aplicaciones son la latencia y la fidelidad, relacionadas con el retardo y la pérdida de paquetes respectivamente. El retardo es el responsable directo de la latencia pero también de la fidelidad ya que una aplicación pierde fidelidad directamente a través de la pérdida de paquetes o en forma de distorsión si acepta los paquetes tardíos desplazando el punto de playback. La latencia es crítica en aplicaciones interactivas como la telefonía, mientras la fidelidad es importante en aplicaciones de difusión de audio y video.

Existen aplicaciones inelásticas que no resisten distorsión ni pérdida de paquetes, y que exigen de la red el menor retardo de compensación, el cual será superior al retardo de cualquier paquete de la aplicación, para estas aplicaciones intolerantes existe el Servicio Garantizado. Otras aplicaciones inelásticas no son tan exigentes y aceptan un retardo de compensación variable entre ciertos límites e incluso soportan alguna pérdida a cambio de mayor eficiencia de transmisión y menores costos, para estas aplicaciones tolerantes se definió el Servicio de Carga Controlada.

Las aplicaciones elásticas, a diferencia de las aplicaciones inelásticas, no son tan sensibles al retardo. Se caracterizan porque siempre esperan la llegada de todos sus datos y éstos se procesan inmediatamente después de su arribo.

Entre las aplicaciones elásticas se encuentran aquellas que presentan un comportamiento de ráfaga interactiva como Telnet y NFS (Network File System), aplicaciones de transferencia interactiva en grandes proporciones como FTP y aplicaciones de transferencia asincrónica en grandes proporciones como e-mail y fax. Las aplicaciones elásticas no necesitan control de admisión, siendo el tradicional servicio best effort el más adecuado para cubrir sus necesidades.

A pesar que se ha intentado clasificar en cierta forma el tráfico, cualquier aplicación puede solicitar el nivel de calidad de servicio que desee basándose en sus requerimientos de fidelidad, latencia y costo.

4.2. Principios para garantizar QoS

- **Requerimientos De Calidad De Servicio:** La base del modelo del servicio se refiere casi exclusivamente al tiempo de entrega de paquetes. Así, el retardo por paquete es la cantidad central sobre la cual la red hace la responsabilidad de calidad de servicio. Se asume una restricción limitada: la única cantidad sobre la cual se hace la responsabilidad cuantitativa del servicio es en los límites de máximo y mínimo retardo.
- **Descarga De Paquetes:** Hasta ahora, se ha asumido implícitamente que todos los paquetes dentro de un flujo son igualmente importantes. Sin embargo, en muchos flujos de audio y video, algunos paquetes son más valiosos que otros. Por lo tanto se propone aumentar el modelo del servicio con un servicio de paquete prioritario, por el que algunos de los paquetes dentro de un flujo se pudieran marcar como prioritarios. Cuando la red estaba en peligro de no resolver algunas de sus comisiones cuantitativas del servicio, se podría ejecutar en ciertos paquetes opciones de prioritario y desechar el paquete (no simplemente lo retarda, desde entonces introduciría problemas fuera de orden). Desechando estos paquetes prioritarios, un

direccionador puede reducir el retardo de los paquetes no prioritarios. Además, uno puede definir una clase de paquetes que no esté sujeto a control de admisión. En el panorama descrito donde los paquetes prioritarios son descargados solamente cuando la comisión cuantitativa de servicio está en peligro de ser violados, la expectativa es que los paquetes prioritarios serán entregados casi siempre y deben estar incluidos en la descripción de tráfico usada en control de admisión. Sin embargo, podemos ampliar la prioridad al extremo de un caso de paquetes prescindibles (este término se utiliza como término extremo de prioridad), donde está la expectativa que muchos de estos paquetes prescindibles no pueden ser entregados. Uno puede entonces excluir los paquetes prescindibles de la descripción del tráfico usada en control de admisión; es decir, los paquetes no se consideran parte del flujo desde la perspectiva de control de admisión, puesto que la comisión no los entregará.

- **Uso Del Feedback:** Otro tema importante en el servicio es el modelo para el uso del Feedback, también conocido como contabilidad, que tiene la función de prevenir el abuso de los recursos de la red. El servicio anteriormente descrito se puede utilizar para proporcionar límites impuestos administrativamente en uso. Sin embargo, un modelo de acceso de red de libre mercado requerirá la contrapresión en los usuarios para los recursos reservados de la red.
- **Modelos De Reserva:** Los modelos de reserva describen cómo una aplicación negocia un nivel de Calidad de Servicio (QoS). El modelo más simple de aplicación pide una QoS particular y la red admite o rechaza la aplicación; la situación será a menudo más compleja. Muchas aplicaciones podrán conseguir servicio aceptable de una gama de niveles de QoS, generalmente, desde cualquier parte de la región del espacio. Por ejemplo, simplemente rechazando la petición, la red puede conceder un nivel más bajo del recurso e informar la aplicación de la cual la QoS habrá garantizado realmente. Un ejemplo más complejo es el modelo de reserva

- **Mecanismos del Control de Tráfico:** Primero se examina muy brevemente los mecanismos posibles del control de tráfico.En la trayectoria de paquete avanzado, hay actualmente un conjunto muy limitado de acciones que un router puede tomar. Dado un paquete particular, un router debe seleccionar una ruta para este paquete; adicionalmente el router puede dejar avanzar o descargar el paquete, y puede ser que el router reordene el paquete con respecto a otros paquetes que esperan para salir.

4.3 Componentes de los Servicios Integrados

Los cuatro componentes básicos de la arquitectura IntServ se describen así:

- **El control de admisión:** comprueba que existen recursos suficientes para soportar el servicio solicitado.
- **Clasificador de paquetes:** analiza los campos de direcciones y puertos para determinar la clase a la que pertenece el paquete.
- **Planificador de paquetes:** aplica algoritmos de encolado que gestiona la transmisión de los paquetes por un enlace de salida.
- **El protocolo RSVP:** para que una aplicación ´pida un determinado servicio a la red. El protocolo entrega la petición al control de tráfico de cada router, que comprobará si es viable la petición.

4.4 Protocolo RSVP

El Protocolo de Reservación de Recursos RSVP es usado para establecer y mantener reservaciones de recursos que permitan suministrar calidad de servicio a un flujo de datos. Los requerimientos son evaluados en cada router en el camino del flujo, siendo RSVP el medio de transporte de éstos.

Las principales características de este protocolo son:

- RSVP permite reservaciones para aplicaciones unicast y multicast.
- RSVP es simplex, es decir hace reservaciones a flujos de datos unidireccionales.

- RSVP es orientado al receptor, es decir el receptor de un flujo de datos inicia y mantiene la reservación de recursos para el flujo.
- RSVP mantiene un estado "suave" en los routers y hosts, soportando los cambios dinámicos de membresías y adaptándose dinámicamente a los cambios de rutas.
- RSVP no es un protocolo de enrutamiento pero depende de los actuales y futuros protocolos de enrutamiento.
- RSVP transporta y mantiene los parámetros de control de tráfico y de políticas.
- RSVP provee varios estilos de reservación.
- RSVP provee operación transparente a través de todos los routers que no lo soportan.
- RSVP soporta las dos versiones del Protocolo de Internet, IPv4 e IPv6.

RSVP define como sesión al flujo de datos que tiene un destino particular y un protocolo de capa transporte. Una sesión está definida por tres campos: (DestAddress, Protocol Id, DstPort) en donde, DestAddress es la dirección de destino IP (unicast o multicast); Protocol Id es la identificación del protocolo IP; y, DstPort representa un puerto de destino generalizado que podría ser el puerto de destino UDP/TCP. Este tercer parámetro es opcional en ciertas ocasiones, pero es necesario cuando se debe distinguir entre varios flujos unicast direccionados al mismo destino, además debe ser consistente en todos los nodos a lo largo del camino.

El flujo de datos dentro de una sesión RSVP puede ser multicast o unicast. Si es muiticast se envía una copia de cada paquete de la fuente a cada destino, si la sesión es unicast habrá un solo host destino R, pero pueden existir varias fuentes.

Modelo de reservación

El proceso RSVP empieza cuando una fuente genera un mensaje Path, el cual es enviado a través de la red para documentar la ruta de la reservación de extremo a extremo. El paquete contiene información que permite identificar el flujo que demanda el servicio y los parámetros que definen el servicio esperado. En un punto dado el paquete contiene la dirección IP del salto previo e información de la capacidad y retardo introducido en el nodo.

Luego que este mensaje llega al destino, el receptor tiene una idea clara de la ruta, de los servicios y capacidades relativas que la red puede ofrecer, las cuales comparándolas con la descripción del tráfico de la fuente, genera un mensaje de reservación Resv. Este mensaje contiene el mismo clasificador que permite identificar al flujo y además información que describe el tipo de reservación, la cual puede ser garantizada o de carga controlada.

En cada nodo se realizan las pruebas de control de admisión y control de políticas, la primera para verificar si hay recursos disponibles y la segunda verifica si el transmisor tiene permiso para hacer la reservación; si ambas pruebas son exitosas entonces se realiza la reservación.

Se debe observar que quien realmente empieza la reservación es el receptor, pues aunque aparentemente lo más obvio sería que la inicie el transmisor, ya que éste conoce las características del tráfico que enviará, en cambio el receptor sabe lo que desea (o lo que puede) recibir. Si se dejara la tarea de iniciación de la reservación al transmisor, ello provocaría problemas de escalamiento en árboles multicast grandes, dinámicos, y con receptores heterogéneos.

Estos problemas de escalamiento se resuelven dejando que el receptor sea el que inicie la reservación, de este modo se manipula fácilmente una reservación para receptores heterogéneos ya que cada receptor simplemente solicita la reservación para sí mismo, y si se presentan reservaciones diferentes, simplemente éstas se combinan dentro de la red. La iniciación del receptor también es consistente con el multicast del protocolo IP, en el cual un grupo multicast se crea implícitamente por los receptores que se unen a él.

En el modelo de reservación, el receptor envía el mensaje Resv y cada nodo en el camino acepta o rechaza la reservación, esta forma de reservación puede resultar difícil cuando se desea hallar un determinado servicio extremo a extremo. Para evitar ello se puede utilizar el objeto ADSPEC que recoge información en cada nodo y que permite hacer una predicción de la calidad de servicio extremo a extremo. Esta información puede ser usada por el receptor para construir o ajustar dinámicamente la reservación.

Bibliografía

- D. Fonseca, P. Morales. 2001. Estudio de Factores Técnicos y Operativos que Interviene en la Infraestructura de Calidad de Servicio en Internet. Tesis para optar al grado de Ingeniero en Electrónica y Telecomunicaciones. Universidad Politécnica Nacional. Quito, Ecuador.
- T. García. 2007.“Análisis de los Modelos de Servicios Diferenciales y Servicios Integrales para Brindar QoS en Internet. Tesis para optar al grado de Ingeniero en Computación. Universidad Tecnológica de la Mixteca. México

Módulo 5. Servicios Diferenciados (DiffServ)

Objetivos:

- Conocer el modelo de servicio de internet DiffServ, sus componentes y principales características.
- Realizar una comparación entre los servicios IntServ y DiffServ.
- Analizar el funcionamiento del protocolo MPLS y sus herramientas para garantizar QoS.

¿De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje nos enfocamos en los servicios Diferenciados DiffServ y sus componentes para garantizar calidad de servicios a los diferentes tipos de servicios que ofrecen las redes. Además haremos un análisis del protocolo MPLS el cual es el encargado de proporcionar el QoS en los Servicios Diferenciados.

5. Servicios Diferenciados (DiffServ)

5.1 Definición

DiffServ se basa en un marco de trabajo arquitectónico que reconoce la entidad relevante para servicios garantizados efectivos en Internet, en el dominio administrativo de un único operador de red. Entonces el modelo está orientado hacia un servicio borde a borde a través de un dominio único, con un apropiado Acuerdo de Nivel (Level Agreement, LA) que se asume en los bordes del dominio. El énfasis ha sido en desarrollar bloques de construcción de QoS antes que los servicios, reconociendo la necesidad de mecanismos altamente escalables con un mínimo impacto en los elementos de los caminos donde van los datos de los routers del núcleo, los cuales manejan enlaces de multi gigabits.

Los Servicios Diferenciados satisfacen requisitos como proporcionar altas prestaciones, escalabilidad, permitir el crecimiento sostenido del tamaño de las redes y su ancho de banda, etc. La filosofía empleada en su diseño se basa en situar el proceso complejo y la gestión de los recursos en los límites de la red, al mismo tiempo que mantiene el reenvío de paquetes en el núcleo de la red de la manera más sencilla posible. En los nodos del núcleo de la red, no se mantiene el estado de las conexiones, sino que el tratamiento se basa únicamente en los códigos DS de los paquetes, que designan la clase de calidad que deben recibir.

Para DiffServ es indispensable diferenciar los tráficos normales de los tráficos diferenciados, en otras palabras, la esencia de DiffServ está dada por un esquema de prioridades relativas ya que ofrece QoS relativa a agregados. La principal suposición de DiffServ es saber que la mayoría del tráfico en las redes son besteffort o el mejor esfuerzo. DiffServ otorga servicios a cada uno de los usuarios que desean que sus flujos tengan un tratamiento especial o con QoS. A diferencia de su antecesor, IntServ, el cual tenía que marcar un camino e ir reservando recurso de extremo a extremo, sin posibilidad de escalabilidad práctica; DiffServ ofrece una amplia capacidad de satisfacer las necesidades de calidad de servicio desde los flujos, tramas, paquetes y/o datagramas.

5.2 Arquitectura Básica

En la arquitectura definida por Diffserv aparece nodos extremos DS de entrada y salida, así como nodos DS internos. Este conjunto de nodos definen el dominio Diffserv y presenta un tipo de políticas y grupos de comportamiento por salto (PHB - Per Hop Behavior) que determinarán el tratamiento de los paquetes en la red. Debemos tener en cuenta que un dominio Diffserv puede estar formado por más de una red, de manera que el administrador será responsable de repartir adecuadamente los recursos de acuerdo con el contrato de servicio (SLA– Service Level Agreement) entre el cliente y el proveedor del servicio.

Se analiza las diferentes funciones que deben realizar los nodos DS:

Nodos extremos DS: Será necesario realizar diferentes funciones como el acondicionamiento de tráfico entre los dominios Diffserv interconectados. De esta manera debe clasificar y establecer las condiciones de ingreso de los flujos de tráfico en función de: dirección IP y puerto (origen y destino), protocolo de transporte y DSCP, este clasificador se conoce como MF (Multi Field Classifier). Una vez que los paquetes han sido marcados adecuadamente, los nodos internos deberán seleccionar el PHB definido para cada flujo de datos. Los nodos DS de entrada serán responsables de asegurar que el tráfico de entrada cumple los requisitos de algún TCA (Traffic Conditioning Agreement), que es un derivado del SLA, entre los dominios interconectados.

Por otro lado los nodos DS de salida deberán realizar funciones de acondicionamiento de tráfico o TC (Traffic Conformation) sobre el tráfico transferido al otro dominio DS conectado.

Nodos internos DS: Podrá realizar limitadas funciones de TC, tales como remarcado de DSCP. Los nodos DS internos solo se conectan a nodos internos o a nodos externos de su propio dominio. A diferencia de los nodos externos para la selección del PHB solo se tendrá en cuenta el campo DSCP, conocido como clasificador BA (Behavior Aggregate Classifier).

5.3 Diseño de Servicios Diferenciados basado en el protocolo IP

En IPv4, los 6 primeros bytes del campo Type of Service (ToS) y Traffic Class de IPv6 son algunos ejemplos de clasificación por tipo de tráfico que permiten la diferenciación de servicios gracias a la codificación de servicios DSCP (Definition of the Differentiated Services Code Point, RFC 2474). La codificación DSCP asocia a cada paquete un PHB: El campo DSCP consta de los primeros 6 bits, permitiendo 64 posibilidades distintas de códigos DSCP. Sin embargo se divide en grupos. El primer grupo es de uso estándar el cual permite de 32 códigos determinados por la estandarización definida por la IETF. El grupo 2 abarca 16 códigos que han sido reservados para uso experimental o local. Por último el grupo 3, también tiene uso local y experimental, sin embargo de ser necesarios más códigos que los que soporta el grupo 1, el grupo 3 puede ser utilizado como extensión.

El rango de códigos posibles en el campo DSCP no definen un PHB en específico, a un código DSCP puede ser asignado cualquier PHB, lo que hace que el número de PHB sea ilimitado. Los PHB son mapeados en los códigos del campo DSCP. Los PHB más usados y mayormente aceptados son descritos a continuación.

- **Default PHB:** Este campo es recomendado codificarlo con todos los bits del campo DSCP en 0, y es asignado a paquetes que no necesitan ningún tratamiento especial. Los paquetes marcados con este PHB son enviados tan pronto estén disponibles los recursos para su transmisión o procesamiento.
- **Expedited Forwarding (EF):** Este PHB implica que los paquetes deben ser tratados con baja latencia, baja perdida y bajo jitter. Para tal fin el tráfico perteneciente a este PHB tiene prioridad en el encolamiento sobre otro tipo o clase de tráfico. Para que no exista congestión debido a altos volúmenes de paquetes EF, son necesarias políticas de admisión estrictas que limiten la cantidad de tráfico que puede ser codificado con este PHB. El IETF define estos PHB en el RFC 3246.
- **Assured Forwarding (AF):** Los PHB de este grupo permiten controlar la entrega de paquetes en donde se determinan unas cuotas de uso justo para cada clase. De presentarse congestión los paquetes tienen una probabilidad

de ser descartados, según la prioridad del código DSCP. Los PHB también son divididos en cuatro clases con diferentes prioridades, cada clase tiene códigos que determinan la probabilidad de ser descartados en una congestión. El IETF define estos PHB en los RFC 3260 y 2597.

- **Class Selector (CS):** Anteriormente se utilizaba el campo Precedence en el campo Type of Service de IPv4 para determinar prioridad. Estos PHB son utilizados para mantener compatibilidad hacia dispositivos antiguos.
- **Voice Admit (VA):** En el RFC 5865 se define este PHB y básicamente tiene el mismo principio de funcionamiento que EF aplicado a tráfico de llamadas de voz (VoIP).

El IETF también define unas reglas para la asignación de codigos a los PHB en el RFC 3140, en resumen un PHB especifica el tratamiento por cada enrutador dentro de un dominio DiffServ (DS). Se denomina un dominio DS a un grupo de enrutadores contiguos los cuales trabajan con una política de servicio común implementada en cada enrutador. Todos los enrutadores tienen reglas de redireccionamiento basadas en los valores de DS de los paquetes, los cuales, son comparados con su correspondiente valor de PHB. En caso de no poseer ningún tratamiento de QoS el campo DSCP será igual a 0. Los límites de la región la determinan los enrutadores o host de frontera, estos son los encargados de clasificar los paquetes entrantes en una determinada clase y se aseguran que estén correctamente etiquetados usando PHB por todo el dominio. Un dominio DS generalmente consiste en una red o un conjunto de redes que utilizan los mismos códigos PHB. Un domino también se consideran como una unidad administrativa.

Ahora bien se denomina una región DS como un conjunto de dominios de DS que aseguran las rutas de servicios por todas las redes y dominios que abarca. Cada dominio dentro de una región puede contar con una definición y mapeo igual o diferente de PHB, de ser diferente es necesario el uso de condicionadores de tráfico hacer las respectivas traducciones de los mapeos y definiciones de PHB.

El campo ECN o Notificación Explicita de Congestión permite identificar en que segmentos de red existen congestiones, permitiendo nuevas funcionalidades como un enrutamiento inteligente basado en estas características.
El campo ECN está definido en el RFC 3168. ECN utiliza una codificación simple a través de 4 diferentes códigos que se explican a continuación.

Código	Funcionalidad
00	Indica que un nodo no está utilizando o no tiene habilitado el uso de ECN.
01/10	Estos dos códigos tienen el mismo tratamiento indicando que el ECN está habilitado entre las dos nodos (receptor-transmisor). El código 01 es llamado ECT(0) y el 10 ECT(1).
11	Indicador de congestión, un enrutador envía este código para notificar que está presentando congestión.

Figura 2: Campos ECN y su funcionalidad

El uso de este campo permite no solo el desarrollo de protocolos de enrutamiento dinámicos más inteligentes, sino que, un enrutador antes de que llegue a congestión y empiece a descartar o perder paquetes, notifica la congestión para que disminuya la pérdida de paquetes.

5.4 Comparación de Intserv y DIffServ

En esta sección se pretende comparar el desempeño del modelo de servicios diferenciados con el modelo de servicios integrados, analizando parámetros como el aislamiento del tráfico, ámbito de QoS, complejidad en la configuración, y la escalabilidad.

Parámetros	DiffServ	IntServ
Aislamiento del tráfico	Por clase de tráfico, agregados de varios flujos	Por flujo
Ámbito de QoS	Dentro del dominio	Entre origen y destino
Complejidad en la configuración	Configuración realizada a largo plazo para cada categoría de forma estática	Configuración realizada por flujo, en el momento en el que se necesita, de forma dinámica. Existen mensajes de señalización entre los routers
Escalabilidad	En los routers frontera se mantiene información para cada flujos o agregados de flujos, en los routers del núcleo se mantiene información para cada clase	Cada routers mantiene información de estado por cada flujo.

Tabla 1: Comparación entre Parámetros de DiffServ e IntServ

SERVICIOS INTEGRADOS	SERVICIOS DIFERENCIADOS
Son aplicables en redes pequeñas	Presenta un buen desempeño tanto en redes pequeñas como grandes
Funciona en el nivel 4 del modelo OSI	Trabaja en el nivel 3 del modelo OSI, el cual lo hace transparente para el usuario
Deja que los usuarios puedan realizar explícitamente peticiones de QoS	Tiene solo 12 posibilidades de servicio
Permite solicitudes de calidad de servicio con gran granularidad	Los tipos de servicios son permanentes
Necesita periódicamente refrendar el tipo de servicio	Los recursos son asignados en el router de frontera
Utiliza un protocolo de reserva para designar recursos	Los nodos procesan los paquetes de acuerdo al campo DS
Posee un mecanismo más complejo y exigente	Tiene una forma sencilla de clasificar y priorizar el tráfico.

Tala 2: Comparación de características de DiffServ e IntServ

5.5 Protocolo MPLS

El multiprotocolo de conmutación de etiquetas (*MPLS*) reduce significativamente el procesamiento de paquetes que se requiere cada vez que un paquete ingresa a un enrutador en la red, esto mejora el desempeño de dichos dispositivos y del desempeño de la red en general. Dicho protocolo se puede considerar en desarrollo constante ya que en los últimos años la demanda de esta tecnología ha ido creciendo.

Una red MPLS consiste de un conjunto de Enrutadores de Conmutación de etiquetas LSR) que tienen la capacidad de conmutar y rutear paquetes en base a la etiqueta que se ha añadido a cada paquete. Cada etiqueta define un flujo de paquetes entre dos puntos finales. Cada flujo es diferente y es llamado Clase de Equivalencia de Reenvío (FEC), así como también cada flujo tiene un camino específico a través de los LSR de la red, es por eso que se dice que la tecnología MPLS es “orientada a conexión”. Cada FEC, además de la ruta de los paquetes contiene una serie de caracteres que define los requerimientos de QoS del flujo. Los routers de la red MPLS no necesitan examinar ni procesar el encabezado IP, solo es necesario reenviar cada paquete dependiendo el valor de su etiqueta. Esta es una de las ventajas que tienen los routers MPLS sobre los routers IP, en donde el proceso de reenvío es más complejo.

En un router IP cada vez que se recibe un paquete se analiza su encabezado IP para compararlo con la tabla de enrutamiento (routing table) y ver cuál es el siguiente salto (next hop). El hecho de examinar estos paquetes en cada uno de los puntos de tránsito que deberán recorrer para llegar a su destino final significa un mayor tiempo de procesamiento en cada nodo y por lo tanto, una mayor duración en el recorrido.

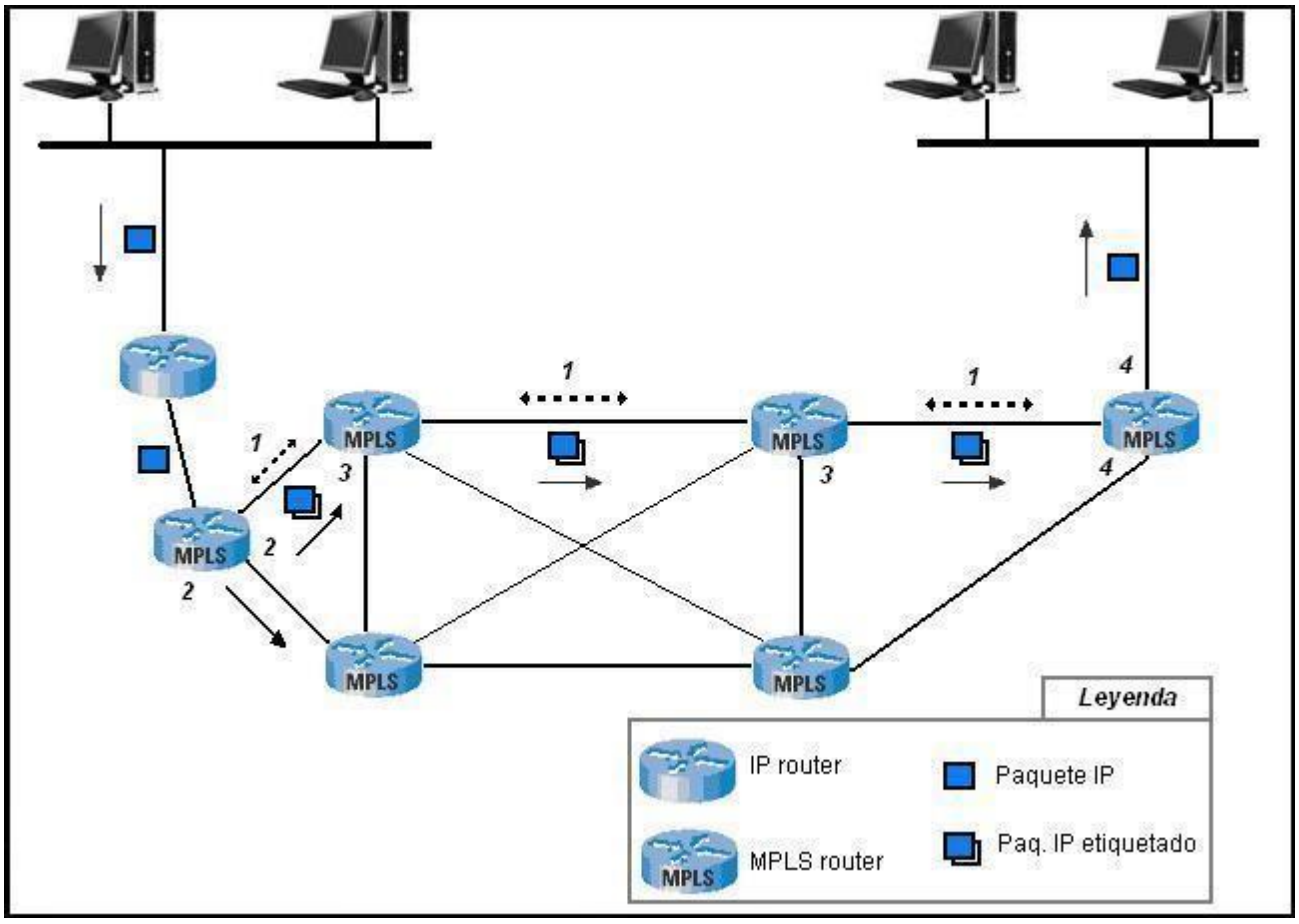


Figura 3. Funcionamiento MPLS

A continuación los pasos que sigue el flujo de paquetes MPLS en la Figura 3:

1. Antes de mandar la información por el flujo es necesario establecer un Camino de Conmutación de Etiquetas (LSP) entre los routers que van a transmitir la FEC. Dichos LSP sirven como túneles de transporte a lo largo de la red MPLS e incluyen los parámetros QoS específicos del flujo. Estos parámetros sirven para determinar dos cosas:

- La cantidad de recursos a reservar al LSP.
- Las políticas de desechado y la cola de procesos en cada LSR.

Para lograr los puntos anteriores se utilizan dos protocolos para intercambiar información entre los routers de la red.

2. Se le asignan etiquetas a cada flujo FEC particular para evitar el uso de etiquetas globales que dificultan el manejo y la cantidad de las mismas. Por esta razón las etiquetas solo hacen referencia al flujo específico. La asignación de nombres y rutas se puede realizar manualmente o bien se puede utilizar el Protocolo de Distribución de Etiquetas (LDP).

3. En esta sección el paquete entra al dominio MPLS mediante un LSR frontera que determina que servicios de red requiere, definiendo así su QoS. Al terminar dicha asignación el LSR asigna el paquete a una FEC y a un LSP particular, lo etiqueta y lo envía. Si no existe ningún LSP, el router frontera trabaja en conjunto con los demás LSRs para definirlo.

4. En este momento el paquete ya está dentro del dominio MPLS, cuando los routers contiguos del LSR reciben el paquete se llevan a cabo los siguientes procesos:

- Se deshecha la etiqueta de entrada y se le añade la nueva etiqueta de salida al paquete.
- Se envía el paquete al siguiente LSR dentro del LSP.

El LSR de salida “abre” la etiqueta y lee el encabezado IP para enviarlo al destino final.

Ventajas específicas de MPLS

En este momento ya es posible identificar algunas de las ventajas internas más importantes que MPLS presenta:

1. Un domino MPLS consiste de una serie de routers habilitados con MPLS continuos y contiguos. El tráfico puede entrar por un punto final físicamente conectado a la red, o por otro router que no sea MPLS y que esté conectado a una red de computadoras sin conexión directa a la nube MPLS.
2. Se puede definir un Comportamiento por Salto (PHB) diferente en cada router de la FEC. El PHB define la prioridad en la cola y las políticas de desechado de los paquetes.
3. Para determinar el FEC se pueden utilizar varios parámetros que define el administrador de la red:

- Dirección IP fuente o destino y/o las direcciones IP de la red.
- Utilizar el ID del protocolo IP.
- Etiqueta de flujo IPv6.
- Numero de puerto de la fuente o del destino.
- El punto de código (codepoint) de los servicios diferenciados (DSCP).

4. El reenvío de la información se lleva a cabo mediante una búsqueda simple (lookup) en una tabla predefinida que enlaza los valores de las etiquetas con las direcciones del siguiente salto (next hop).

5. Los paquetes enviados de mismos endpoints pueden tener diferente FEC, por lo que las etiquetas serán diferentes y tendrán un PHB distinto en cada LSR. Esto puede genera diferentes flujos en la misma red

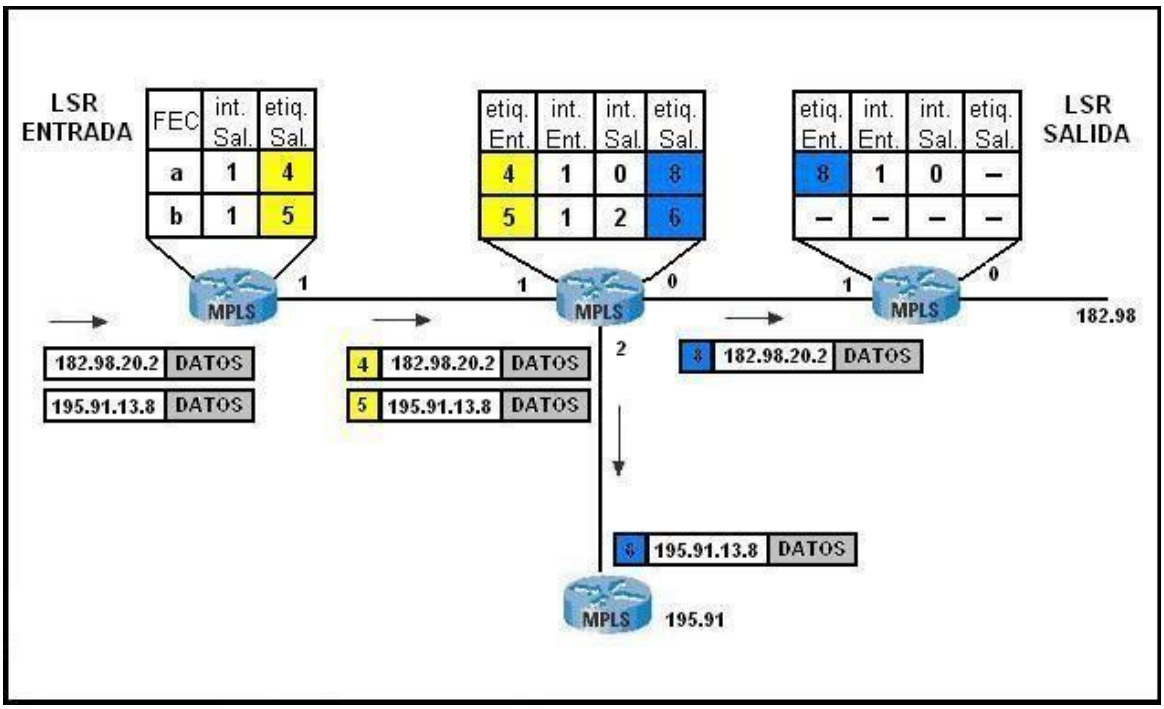


Figura 4. Reenvío de Paquetes

Como se puede observar en la Figura 4 cada LSR tiene una tabla de reenvío para cada LSP que pasa por sus interfaces. Dichas tablas manejan diferentes tipos de datos, la tabla del LSR de entrada maneja la FEC, la interfaz de salida y etiqueta de

salida, los LSR siguientes manejan etiqueta e interfaz, ambas de entrada y de salida.

Posteriormente se muestra como llegan los datos (a y b) sin etiqueta al LSR de entrada, el cual les asigna una etiqueta de salida y lo manda al siguiente LSR (next hop LSR). El LSR siguiente deshecha las etiquetas de entrada y les añade nuevas y las manda a los LSR correspondientes, es aquí donde se ve la escalabilidad de la tecnología, ya que las etiquetas solo tienen significado local.

Otra de las funciones del LSR de entrada es asignarle una FEC a cada paquete sin etiquetar que entra, y en base a esto asigna cada paquete a un LSP particular. En la Figura 4 tenemos dos FECs (a y b) cada uno con su LSP particular.

Apilamiento de Etiquetas (Label Stacking)

Una de las características más importantes que tiene MPLS es el apilamiento de etiquetas que maneja, un paquete etiquetado puede contener varias etiquetas organizadas en modo Ultimo en Entrar Primero en Salir (LIFO). El procesado de etiquetas en MPLS siempre se basa en la etiqueta superior, por lo que en cualquier LSR se puede añadir (push) o remover (pop) una etiqueta. La ventaja principal del apilamiento de etiquetas es que permite añadir rutas parciales dentro de la red a un LSP existente, creando así túneles.

Al principio de cada túnel los LSR asignan la misma etiqueta a los paquetes que van entrando mediante la operación push que mencionamos anteriormente. Al final de cada túnel pasa lo inverso, el LSR de salida remueve la etiqueta superior (añadida a la entrada del túnel) para mostrar la etiqueta original con el fin de que siga su trayectoria original. Esta operación se puede realizar indefinidamente formando así una red de túneles dentro de cada LSP original. Esta es una característica que ATM maneja, sin embargo solo maneja apilamiento de un nivel.

Formato de etiquetas MPLS

Figura 5. Formato de Etiqueta

Una etiqueta MPLS está conformada por 32 bits, divididos como se muestra en la Figura 5, y contiene los siguientes elementos:

- Valor de la etiqueta: Etiqueta de 20 bits con valor local.
- Experimental: Son los 3 bits siguientes reservados para uso experimental.
- S: Es el bit de posición de pila:
  - Cuando es 1 denota que es la entrada más antigua en la pila.
  - Cuando es 0 denota que es cualquier otra entrada.
- Tiempo de Vida (TTL): Son los últimos 8 bits del paquete y se utilizan para codificar el valor del conteo de saltos (IPv6) o de tiempo de vida (IPv4).

Procesando el TTL

Un elemento clave en el encabezado de un paquete IP es el campo TTL y el Límite de Saltos (Hop limit). En un ambiente común de Internet (basado en IP), dicho campo va disminuyendo uno a uno hasta que llega a cero y se elimina el paquete. Esto es una medida de prevención de los paquetes se ciclen (looping) o estén demasiado tiempo en el Internet debido a un ruteo mediocre.

En el ruteo MPLS no se lee el encabezado de los paquetes, es por eso que se añaden estos 8 bits que manejan el TTL para evitar que ocurra lo mencionado anteriormente.

Reglas para procesar el campo TTL:

1. Cuando un paquete IP llega al router de entrada de un dominio MPLS, solo se añade una etiqueta de entrada a la pila, el valor de TTL de este campo se obtiene del valor original del TTL en IP. En este paso se da por supuesto que el campo ya fue disminuido, como parte del proceso IP. Cuando un paquete MPLS llega a uno de los LSR internos, el valor del campo TTL de la etiqueta del primer elemento en la pila es disminuido. Entonces:

- **a)** Si el valor es 0, no se reenvía el paquete, dependiendo del valor que tenga la etiqueta del paquete puede ser desechado o es enviado al nivel de red para procesamiento de errores.
- **b)** Si el valor es positivo, se le añade a la nueva etiqueta de la pila en el campo TTL y es reenviado al siguiente salto. El valor del campo TTL del paquete reenviado esta dado en función del valor del campo de Tiempo de Vida del paquete original.

2. Cuando un paquete MPLS llega a un LSR de salida, el valor del campo TTL en la etiqueta es disminuido (uno por uno) y posteriormente se quita la etiqueta de la pila, lo que deja una pila vacía.

Entonces:

- **a)** Si el valor es 0, no se reenvía el paquete, dependiendo del valor que tenga la etiqueta del paquete puede ser desechado o es enviado al nivel de red para procesamiento de errores.
- **b)** Si el valor es positivo, se coloca en el campo TTL del encabezado IP y es enviado utilizando ruteo IP tradicional.

3. Es importante mencionar que cuando el valor del campo TTL llega a 0 y el paquete no ha llegado a su destino predefinido en el valor de la etiqueta, dicho paquete es desechado y se envía un mensaje del Protocolo de Control de Mensajes de Internet (ICMP) al remitente. Esto con el fin de evitar que un paquete no entregado se quede circulando en el Internet.

4. En teoría TTL esta medido en segundos, aunque cada equipo (host) que pase el paquete debe reducir el Tiempo de Vida en al menos una unidad. El

campo TTL

es disminuido en una unidad en cada salto, es por eso que en IPv6 a este decremento de unidades en cada salto se le llama Conteo de Saltos (hop count).

Pila de etiquetas

La última sección del formato de las etiquetas es la sección S, en donde está contenida la información del orden en la pila. Cuando S = 1 indica que es la última etiqueta y que al salir quedará vacía la pila, esto generalmente ocurre en el router de salida, cuando es S = 0 indica que por lo menos hay otra etiqueta antes, en la pila.

Figura 6. Formato de paquete MPLS

Lo anterior se puede ver claramente en la Figura 6. Es muy importante considerar que cuando un Enrutador de Etiquetas Frontera (LER) saca el último encabezado MPLS del paquete este debe de mandar la información (payload) fuera de la nube MPLS al destino contenido en el encabezado IP, previamente obtenido por el router de entrada. La importancia radica en que los routers MPLS no cuentan con tablas de búsqueda de etiquetas (label lookup tables).

Para entender esto podemos ver que cuando a un router MPLS le llega el valor S=1 en el encabezado MPLS se sabe que el siguiente encabezado es el encabezado de red y que debe usarlo para reenviar el paquete conforme al mecanismo de ese tipo de red. Como se menciona anteriormente MPLS soporta múltiples protocolos de red, en realidad todos, pero un encabezado IP no tiene la misma estructura que un encabezado Ethernet, por lo que aunque

el router de salida sepa que lo siguiente en la pila es un encabezado de red, no sabe de qué tipo es y no puede interpretarla.

Esto se soluciona leyendo los valores reservados del campo de valor de la etiqueta de 20 bits, este indicará el tipo de encabezado de red para que así pueda “entender” lo que este dice.

Los valores reservados para el campo “valor de etiqueta” de la primera etiqueta que se añade al paquete (la etiqueta con el valor S=1) son los siguientes:

- Label “0”: El paquete proviene de una red IPv4.
- Label ”2”: El paquete proviene de una red IPv6.
- Label ”4” – “15” reservados para uso futuro por la Agencia de Asignación de Números de Internet (IANA).

Las etiquetas de la pila de etiquetas MPLS van después de los encabezados de la capa de enlace de datos (modelo OSI), pero antes de los encabezados de la capa de red.

Relación entre FECs, LSPs y Etiquetas.

Para entender el funcionamiento de la tecnología MPLS, es necesario comprender y analizar la importancia de la relación entre etiquetas, el LSP y la FEC. La característica principal de la funcionalidad de la tecnología MPLS es que el tráfico se divide en “canales privados” (Clase de Equivalencia de Reenvío). El tráfico de cada FEC viaja por medio de un LSP dentro del dominio MPLS. Los paquetes dentro de un FEC solo tienen validez en dicho canal, ya que las etiquetas son de significado local.

A lo largo de la nube MPLS en cada LSR se reenvían los paquetes etiquetados a la ruta especifica de la etiqueta, cada vez que entra el paquete al LSR este reemplaza el valor de la etiqueta de entrada con el nuevo valor de la etiqueta de salida. Esto se lleva acabo sucesivamente hasta que dicho paquete llega al router se salida.

Requisitos para el tráfico MPLS

Para que esto se lleve a cabo satisfactoriamente se deben de cumplir los siguientes requisitos:

1. Todo tráfico debe de asignarse a un FEC específico.
2. Se necesita un protocolo de ruteo para determinar la topología y las condiciones del dominio para que las LSPs puedan ser asignadas a un FEC. Adicionalmente, el protocolo de ruteo debe de recolectar información y utilizarla para proveer los requerimientos de QoS del FEC.
3. Cada LSR debe de conocer las LSPs de cada FEC para poder asignarles una etiqueta de entrada y deben de comunicarla a todos los demás LSR en la ruta de dicho LSP.

Ingeniería de Tráfico

Es la habilidad de definir rutas dinámicamente y planear la asignación de recursos con base en la demanda, así como optimizar el uso de la red. MPLS facilita la asignación de recursos en las redes para balancear la carga dependiendo de la demanda y proporciona diferentes niveles de soporte dependiendo de las demandas de tráfico de los usuarios. El protocolo IP provee una forma primitiva de Ingeniería de Tráfico al igual que el protocolo del Camino Más Corto Primero (OSPF) que permite a los enrutadores cambiar la ruta de los paquetes cuando sea necesario para balancear la carga. Sin embargo esto no es suficiente ya que este tipo de ruteo dinámico puede llevar a congestionar la red y no soporta QoS.

Todo tráfico entre dos puntos finales (endpoints) sigue la misma ruta y puede ser cambiada si ocurriera congestión, sin embargo este cambio solo ocurre solo cuando hay congestión que es algo que siempre se trata de evitar. En MPLS a diferencia de OSPF no se ve paquete por paquete sino flujos de paquetes con su respectivo QoS y demanda tráfico predecible. Con este protocolo es posible predecir rutas en base a flujos individuales, pudiendo haber diferentes flujos

entre canales similares pero dirigiéndose a diferentes enrutadores.
Si llegase a amenazar congestión en la red, las rutas MPLS pueden ser re-ruteadas inteligentemente, de esta manera se pueden cambiar las rutas de flujo de paquetes dinámicamente conforme a las demandas de tráfico de cada flujo.

Beneficios principales de ingeniería de trafico MPLS

- Permite al eje troncal (backbone) expandirse sobre las capacidades de la ingeniería de tráfico de las redes de Modo de Transferencia Asíncrona (ATM) y Frame Relay (FE) de Capa 2.
- La ingeniería de tráfico es esencial para los ejes troncales de proveedores de servicios. Dichos ejes deben soportar un uso elevado de su capacidad de transmisión.
- Utilizando MPLS las capacidades de ingeniera de tráfico son integradas a la Capa 3 (OSI) lo que optimiza el ruteo de trafico IP gracias a las pautas establecidas por la topología y las capacidades de la troncal.
- La ingeniería de trafico MPLS rutea el flujo de tráfico a lo largo de la red basándose en los recursos que dicho flujo requiere y en los recursos disponibles en toda la red.
- MPLS emplea la ruta más corta que cumpla con los requisitos del flujo de tráfico, que incluye: requisitos de ancho de banda, de medios y de prioridades sobre otros flujos.

Con un buen manejo del tráfico en las redes, se pueden evitar congestionamientos, mejorar el desempeño general y reducir la latencia y el desechado de paquetes. En pocas palabras se maximiza la capacidad de la red y se minimizan los costos.

Bibliografía:

- R. Montes, K. Tous. 2006. “Servicios Integrados y Diferenciados de Internet”. Tesis para optar por el título de Ingeniero Electrónico. Universidad Tecnológica de Bolívar. Colombia.
- J.A. Jimenez Toro. UF1875: Gestión de recursos, servicios y de la red de Comunicaciones. Editorial Elearning SL. Edición 5.0. ISBN: 978-84-16199-01-3. España.
- Benítez, M., Castellar A. (2011). Differentiad Services (DIFFSERV): Ventajas, Desventajas y Casos de Estudio. Universidad Tecnológica de Bolívar. Tesis para optar al Título de Ingeniero en Sistemas. Colombia.

Caso de Estudio

1. Indique cómo tienen que configurarse los parámetros de una red DiffServ para que un flujo TCP y uno UDP que compiten obtengan el mismo caudal. Indique cómo sería la compartición de caudal en una red sin DiffServ.

Solución: Hay que limitar la tasa máxima permitida al tráfico UDP a la mitad del caudal disponible. De no existir una red Diffserv, UDP mantendría su tasa constante mientras que TCP tendería reducir su tasa debido al mecanismo de control de congestión de TCP.

2. Implementando Servicios usando DiffServ

Los routers de Entrada al dominio DiffServ deben tener un clasificador, que clasifica el tráfico entrante, un gestor que controle ancho de banda y prioridades, y un sistema de colas dependiendo de la clasificación del tráfico. Todo el tráfico dentro del dominio se gestiona de acuerdo a la clasificación que de él se hace en los routers de entrada.

Para el desarrollo de la práctica, se van a hacer dos topologías dentro del laboratorio, cada una compuesta a su vez por tres routers (dos routers remotos, que serán de entrada y salida al dominio y uno central). Los extremos serán las interfaces F0/0 de los routers remotos.

En este ejemplo, queremos dar QoS de extremo a extremo a varios tipos diferentes de clases de tráfico utilizando Cisco IOS Differentiated Services. Las clases de tráfico que se van a utilizar son las siguientes:

- Clase Premium: tráfico de voz
- Clase Oro: son sesiones TACACS con tráfico marcado con DSCP 12 y 14.
- Clase plata: consta de Telnet, SMTP y sesiones FTP.
- Clase bronce: es tráfico web y tráfico marcado con DSCP 28 y 30.
- Cualquier otro tráfico se considera perteneciente a la clase de tráfico best-effort.

Las características de funcionamiento deben ser:

- La clase Premium debe ser enviada al siguiente nodo con el menor retardo posible, hasta un máximo de 500kBps durante periodos de congestión.
- La clase oro debe ser tratada preferencialmente con respecto a la clase plata, y esta con más preferencia que la bronce. Estas tres clases deben tener un 35%, 25% y 15% respectivamente, del ancho de banda de la interfaz con unas garantías mínimas de ancho de banda.
- En provisión de clases de tráfico distintas, el tráfico necesita ser clasificado basándose en valores DSCP en un dominio DiffServ. Ya que ese tráfico se basa en valores DSCP, debe ser premarcado con los DSCP adecuados en el momento de entrar en la red. Esto se debe hacer en las interfaces F0/0 de los routers extremos. Este marcado puede conseguirse con una política de entrada.

La siguiente tabla muestra los valores usados para marcar las distintas clases de tráfico que entran en la red de ejemplo:

CLASE DE TRÁFICO	TIPO DE TRÁFICO	VALOR DSCP
Premium	Voz	46
Oro	Tacacs	10
Plata	Telnet	18
	SMTP	20
	FTP	22
Bronce	http	26
Best-effort (defecto)		

Módulo 6. Políticas de aseguramiento de la calidad de servicio en redes.

Objetivos:

- Identificar las políticas para el aseguramiento de la calidad de servicio en redes
- Evaluar las distintas disciplinas de encolamiento y sus principales características.
- Comprender el funcionamiento de las técnicas de control de congestión en la red.

De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje nos enfocamos en las políticas para garantizar calidad de servicio en las redes. Veremos las técnicas de colas, así como también algunas técnicas para el control de congestión en la red.

6. Políticas de aseguramiento de la calidad de servicio en redes.

Al planificar la política de calidad de servicio (QoS) se debe revisar, clasificar y después priorizar los servicios que proporciona la red. También se debe evaluar la cantidad de ancho de banda disponible para determinar la tasa a la que cada clase de tráfico se transfiere en la red.

6.1 Disciplina de Colas

6.2 Técnicas de disciplinas de colas

6.2.1 Primero en llegar, primero en servir (FCFS)

Es el tipo más simple de encolamiento, consiste en un búfer sencillo que retiene los paquetes salientes hasta que la interfaz de transmisión pueda enviarlos. Los paquetes se envían fuera de la interfaz en el mismo orden en el que llegaron al búfer.

Como el tráfico de red llega a un punto de entrada o salida, como a una interfaz de enrutador, debe ser capaz de procesar adecuadamente el tráfico como este está siendo recibido. El encolamiento primero entra / primero sale FIFOQ (First In / First Out Queuing) es el enfoque más básico para ordenar tráfico en una comunicación adecuada. Una cola FIFO ubica todos los paquetes en una línea simple como van ingresando a la interfaz.

Los paquetes son procesados por el enrutador en el mismo orden que ingresan a la interfaz. No se asigna una prioridad determinada a los paquetes. La razón importante del uso del encolamiento FIFO es que durante el proceso de enrutamiento, cuando un paquete se dirige de una interfaz de enrutador a otra, este a menudo cambia el tipo de interfaz y la velocidad. Por ejemplo, se considera un flujo de comunicación simple yendo de una interfaz de 100BaseT FastEthernet hacia una conexión serial a 512 Kbps. El flujo encuentra un desajuste de velocidad. El segmento FastEthernet alimenta al flujo del enrutador en 100 Mbps, mientras que la conexión serial de salida envía el flujo en 512 Kbps. La cola FIFO es usada para ordenar los paquetes y mantenerlos hasta que el enlace serial pueda procesarlos correctamente.

La cola FIFO permite al enrutador procesar comunicaciones de muy alta velocidad de salida a través de una velocidad media más baja. En los casos donde la comunicación FastEthernet está compuesta de pequeñas ráfagas, la cola FIFO manipula todos los paquetes sin dificultad. Sin embargo, una mayor cantidad de tráfico de alta velocidad proveniente de la interfaz FastEthernet puede a menudo causar que la cola FIFO se desborde. Esta situación es conocida como caída de la cola, porque los paquetes son descartados desde la parte de atrás de la cola. La cola continuará con el descarte de paquetes en la parte de atrás hasta que procese los paquetes de adelante, liberando así el espacio dentro de la cola para acomodar los nuevos paquetes de entrada desde el fin de la parte de atrás.

La desventaja del encolamiento FIFO viene dado por la simplicidad. Puesto que no tiene un mecanismo para distinguir los paquetes que manipula, no tiene manera de asegurar que procese los paquetes justa y equitativamente. Este encolamiento simplemente procesa los paquetes en el mismo orden que ingresan a la cola. Esto significa que los protocolos de tráfico alto, como el protocolo de transferencia de archivos FTP, pueden usar porciones significantes de la cola FIFO, dejando a los protocolos sensibles de tiempo, como Telnet, con un pequeño ancho de banda para operar. En tal caso, la sesión Telnet debería parecer interrumpida y sin respuesta, ya que la mayor parte de la cola es usada para transferir el FTP.

Es evidente que FIFO es un mecanismo de encolamiento muy básico que permite al enrutador ordenar y procesar paquetes de acuerdo a como concurren para salir de una interfaz. Los paquetes pueden venir de una o múltiples interfaces conectadas al enrutador. Es bueno señalar que este principio de cola simple es la base de los otros mecanismos de encolamiento, los cuales se construyen sobre este principio para ofrecer una mejor calidad de servicio dependiendo de los requerimientos de tráfico.

Se tiene claro que FIFO no parece ser un sofisticado o incluso deseable método de encolamiento, considerando las ricas características de otros mecanismos de encolamiento. Sin embargo, FIFO puede ser un método de encolamiento muy eficiente en ciertas circunstancias. Por ejemplo, un segmento Ethernet 10BaseT

conectado a un enrutador que a su vez se conecta a una WAN a través de un segmento E1; en este caso, no existe opción que la comunicación de 10 Mbps de entrada pueda postrar al tubo de 2 Mbps de salida. El enrutador todavía requiere la cola FIFO para ordenar los paquetes en una línea simple con el fin de alimentarlos a la interfaz E1 para el procesamiento. El uso de un mecanismo simple de encolamiento reduce el retardo experimentado por los paquetes como el enrutador los procesa. En las aplicaciones sensibles al retardo, como voz o video, esto puede ser un factor inapropiado.

Una consecuencia negativa cuando los paquetes entran en el proceso de caída de la cola es que las retransmisiones son requeridas en capas superiores del modelo OSI.

6.2.2 Colas Basadas en Prioridad (PQ)

Es un sencillo enfoque para ofrecer un tratamiento preferencial a los paquetes identificados. Los paquetes que llegan a la interfaz se separan en cuatro colas: baja, normal, media y alta prioridad. La salida de estas cuatro colas alimenta un búfer de transmisión de la interfaz. Los paquetes siempre se sirven desde las primeras colas de alta prioridad; este mecanismo se ajusta a condiciones donde existe un tráfico importante, pero puede causar la total falta de atención de colas de menor prioridad.

El encolamiento de prioridad PQ (Priority Queuing) permite a los administradores de red priorizar tráfico basado en criterios específicos. Estos criterios incluyen tipos de protocolo o subprotocolo, interfaces origen, tamaño de paquetes o cualquier parámetro identificado a través de una lista de acceso. PQ ofrece cuatro diferentes colas:

- Prioridad baja
- Prioridad normal
- Prioridad media
- Prioridad alta

A través de la configuración adecuada de PQ, cada paquete es asignado a una de estas colas. Si no es asignada una clasificación a un paquete, este es ubicado en la cola de prioridad normal. La prioridad de cada cola es absoluta. Cuando los

paquetes son procesados, PQ examina el estado de cada cola, siempre sirviendo a las colas de más alta prioridad antes que las colas de prioridad más baja. Esto significa que mientras exista tráfico en la cola de prioridad más alta, las colas de prioridad más baja no serían procesadas. Por tanto, PQ no utiliza un reparto equitativo de recursos entre sus colas. Estrictamente PQ las atiende en base de las clasificaciones de prioridad configuradas por el administrador de red. La figura 7 muestra el PQ en acción.

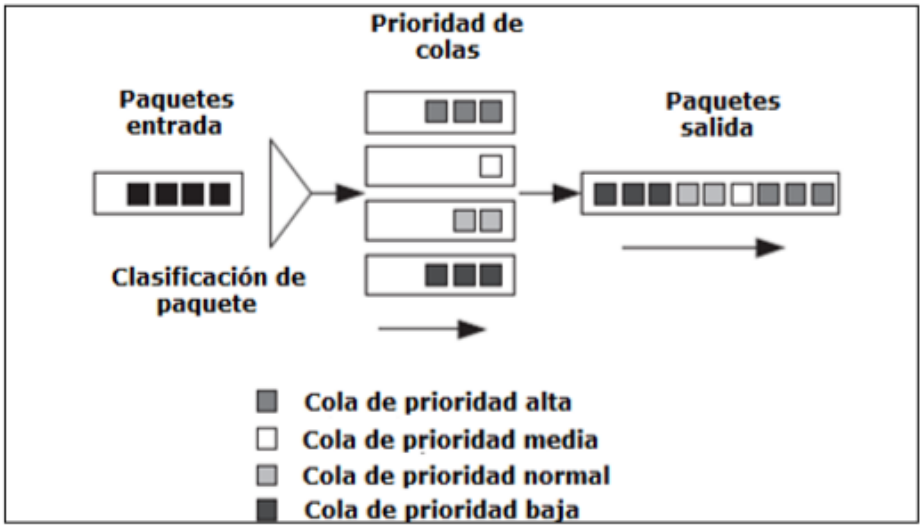


Figura 7. Funcionamiento de PQ

Cada una de estas colas actúa como leaky bucket individual la cual es propensa al descarte de la cola. Los tamaños de cola pueden ser ajustados manualmente de 0 a 32767 paquetes.

El encolamiento de prioridad puede parecer un enfoque tosco para la priorización de tráfico, pero permite dar a ciertas clases de tráfico prioridad sobre otras. Por ejemplo, muchos sistema heredados tal como mainframes usan arquitectura de red de sistemas SNA (Systems Network Architecture) como método de transporte. SNA es muy susceptible a los retardos por lo que sería un excelente candidato para una cola de prioridad alta. Si Telnet es el negocio central de una empresa, podría también ser colocado en la cola de alta prioridad sobre otros tipos de tráfico cualquiera. Esto asegura que los protocolos de volumen alto como FTP no impacten negativamente a las aplicaciones críticas del negocio.

Hay que recordar que la configuración de PQ dispone cómo el proceso de encolamiento operaría en ese enlace. Si nuevas aplicaciones usando protocolos nuevos son desplegadas dentro del ambiente de red, simplemente PQ ubicaría estos protocolos en la cola de prioridad normal. Por consiguiente, la configuración de PQ debería ser periódicamente revisada para asegurar la validez de la configuración de encolamiento.

Cuando se usa PQ, se debe dar una consideración seria a la priorización de tráfico. Si el tráfico asignado a la cola de alta prioridad es pesado, las colas de más baja prioridad nunca serían útiles. Esto conduce a que el tráfico en estas colas nunca está siendo transmitido y el tráfico adicional asignado a estas colas está siendo descartado de la cola.

6.2.3 Colas Basadas en Clases (CBQ)

Permite al administrador priorizar el tráfico sin los efectos laterales de inanición de las colas de baja prioridad, especificando el número de paquetes o bytes que deben ser atendidos para cada cola. Se pueden crear hasta 16 colas para categorizar el tráfico, donde cada cola es atendida al estilo Round- Robin. CQ ofrece un mecanismo más refinado de encolamiento, pero no asegura una prioridad absoluta como PQ. CQ se utiliza para proveer a tráficos particulares de un ancho de banda garantizado en un punto de posible congestión, asegurando una porción fija del ancho de banda y permitiendo al resto del tráfico utilizar los recursos disponibles.

Se ha visto como el encolamiento de prioridad permite asignar tráfico a diferentes colas, cada cola es tratada estrictamente dependiendo de su prioridad. El encolamiento personalizado CQ (Custom Queuing), desplaza el servicio de colas desde un mecanismo absoluto basado en la prioridad hacia un efecto round-robin, atendiendo cada cola secuencialmente.

El encolamiento personalizado permite la creación de más de 16 colas de usuario, cada cola es atendida en secuencia por el proceso CQ. También existe una cola adicional, conocida como cola 0 (es una cola especial usada por el sistema para pasar paquetes de control de red, como por ejemplo paquetes de señalización, entre

otros; tiene prioridad sobre todas las otras colas y así es vaciada antes de cualquier cola definida por el usuario), la misma que es creada automáticamente por el proceso de CQ. Esta cola es configurada por el usuario, pero esto no es recomendable. Cada una de las colas configurables por el usuario, e incluso la cola 0, representan un leaky bucket individual, el cual también es susceptible a los descartes de la cola.

El encolamiento personalizado asegura que cada cola sea atendida, evitando así la situación potencial en la cual cierta cola nunca sea procesada. Este encolamiento lleva su nombre del hecho que los administradores de red pueden controlar el número de colas en los procesos de encolamiento. Adicionalmente, la cantidad de bytes o la cuenta de bytes para cada cola pueden ser ajustadas con el fin de gastar más tiempo en ciertas colas en los procesos de CQ. Por lo tanto, el encolamiento personalizado puede ofrecer un mecanismo de encolamiento más refinado, pero no puede asegurar prioridad absoluta como el encolamiento de prioridad.

El encolamiento personalizado opera mediante el servicio de colas configuradas por el usuario, individuales y secuenciales, para una cantidad específica de bytes. La cuenta de byte por defecto para cada cola es 1500 bytes, sin ninguna personalización, CQ debería procesar 1500 bytes de la cola 1, después 1500 bytes de la cola 2, luego 1500 bytes de la cola 3, y demás.

El tráfico puede ser clasificado y asignado a cualquier cola a través de los mismos métodos como en el encolamiento de prioridad, esto es, tipos de protocolo o subprotocolo, interfaces origen, tamaño de paquete o cualquier otro parámetro identificable a través de una lista de acceso. La figura 8, muestra la operación del encolamiento personalizado.

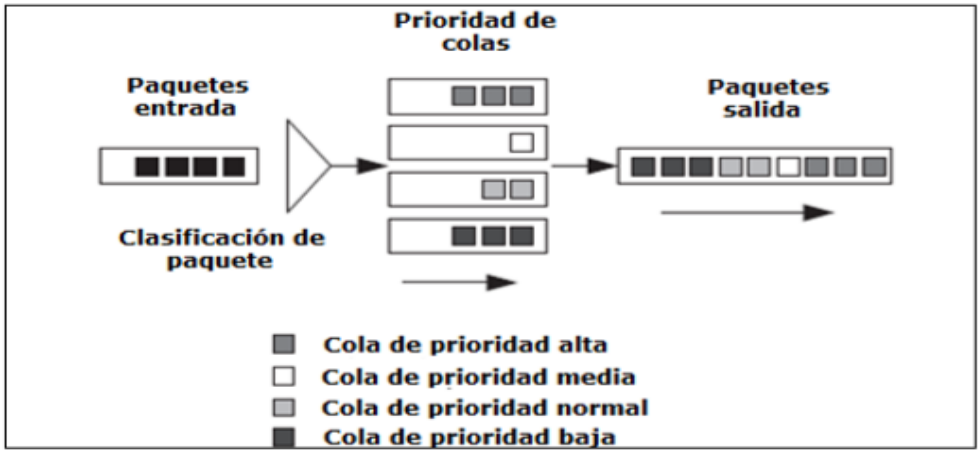


Figura 8: Encolamiento personalizado de CQ

A través de un prudente uso de la cuenta de byte de cada cola, es posible realizar asignaciones de ancho de banda usando encolamiento personalizado. Por ejemplo, en una comunicación de datos determinada, se quiere restringir el tráfico de navegación al 30% del total del ancho de banda, el tráfico de SSH a un 20% del total del ancho de banda y el 50% restante del ancho de banda para cualquier otro tráfico; para lo cual se podría configurar el encolamiento personalizado con tres colas, la cola 1 debería manipular todo el tráfico de navegación con una cuenta de byte por defecto de 1500 bytes, la cola 2 debería manipular todo el tráfico SSH también con una cuenta de byte por defecto de 1500 bytes, por último la cola 3 debería manipular todo el tráfico restante, pero sería manualmente asignada un valor de byte de 3000 bytes

En cuanto a las interacciones de protocolos con el encolamiento personalizado es importante entender que este no proporciona garantías absolutas con respecto a la asignación de ancho de banda. El encolamiento personalizado soporta los protocolos de red, pero también este depende de las operaciones de estos protocolos. El encolamiento personalizado es un excelente mecanismo para realizar asignación de ancho de banda en enlaces de tráfico alto. Este permite a los administradores de red controlar el flujo de paquetes y proporcionar un troughput asegurado a servicios preferidos. Este mecanismo de encolamiento asegura que cada cola es atendida secuencialmente, no se adapta automáticamente a los

cambios de ambiente de red. Todos los protocolos nuevos que no están definidos en la configuración CQ serían asignados a la cola por defecto para el procesamiento de la información.

6.2.4 Round Robin (RR)

El Round Robin es uno de los algoritmos más antiguos, sencillos y equitativos en el reparto de la CPU entre los procesos lo que significa que evita la monopolización de uso de la CPU, y es muy válido para entornos de tiempo compartido.

El algoritmo consiste en definir una unidad de tiempo pequeña, llamada quantum o ¿cuánto? de tiempo, la cual es asignada a cada proceso que está en estado listo. Si el proceso agota su quantum (Q) de tiempo, se elige a otro proceso para ocupar la CPU. Si el proceso se bloquea o termina antes de agotar su quantum también se alterna el uso de la CPU.

Es por ello que surge la necesidad de un reloj en el sistema. El reloj es un dispositivo que genera periódicamente interrupciones. Esto es muy importante, pues garantiza que el sistema operativo (en concreto la rutina de servicio de interrupción del reloj) coja el mando de la CPU periódicamente. El quantum de un proceso equivale a un número fijo de pulsos o ciclos de reloj. Al ocurrir una interrupción de reloj que coincide con la agotación del quantum se llama al despachador, el cual le cede el control de la CPU al proceso seleccionado por el planificador.

Un proceso puede abandonar la CPU por 2 criterios:

- Libremente, si su tiempo de ejecución en la CPU es < Q (quantum).
- Después de una interrupción, si su tiempo de ejecución en la CPU es > Q (quantum) o si el proceso se bloquea.

6.2.5 Round Robin Ponderado (WRR)

El Round Robin ponderado (WRR) garantiza que los paquetes en todas las colas se programen a su vez. WRR programa los paquetes en colas según el peso de cada cola.

Si tres contadores bancarios utilizan la programación de WRR y los pesos de los tres contadores bancarios son 50%, 25% y 25% respectivamente, los otros dos contadores bancarios pueden proporcionar transacciones comerciales de un usuario cuando el contador bancario para VIPs procesa transacciones comerciales de dos usuarios Se sirve a los usuarios de todos los contadores bancarios, pero el tiempo de espera de los VIP es más corto.

Supongamos que tres contadores bancarios se consideran como tres colas. La programación es la siguiente.

Un switch programa paquetes a su vez según el peso de cada cola. Después de una ronda de programación, los contadores disminuyen en 1. La cola de la cual el contador se reduce a 0 no se puede programar. Cuando los contadores de todas las colas se reducen a 0, comienza la siguiente ronda de programación.

Los contadores se inicializan primero: Count [1] = 2, Count [2] = 1, y Count [3] = 1.

Primera ronda de programación de WRR:

El paquete 1 se toma de la cola 1, con el conteo [1] como 1. El paquete 5 se toma de la cola 2, con el conteo [2] como 0. El paquete 8 se toma de la cola 3, con el conteo [3] como 0.

Segunda ronda de la programación de WRR:

El paquete 2 se toma de la cola 1, con el recuento [1] como 0. Las colas 2 y 3 no participan en esta ronda de planificación de WRR porque el recuento [2] y el recuento [3] son 0.

Count [1], Count [2] y Count [3] son todos 0. Los contadores se inicializan de nuevo. Entonces el conteo [1] es 2, el conteo [2] es 1 y el conteo [3] es 1.

Tercera ronda de la programación de WRR:

El paquete 3 se toma de la cola 3, con el conteo [1] como 1. El paquete 6 se toma de la cola 2, con el conteo [2] como 0. El paquete 9 se toma de la cola 3, con el conteo [3] como 0.

Cuarta ronda de programación de WRR:

El paquete 4 se toma de la cola 1, con el recuento [1] como 0. Las colas 2 y 3 no participan en esta ronda de programación de WRR porque el recuento [2] y el recuento [3] son 0.
Count [1], Count [2] y Count [3] son todos 0. Los contadores se inicializan de nuevo. Entonces el conteo [1] es 2, el conteo [2] es 1 y el conteo [3] es 1.

Las estadísticas muestran que el recuento de la programación de paquetes en cada cola está en proporción directa al peso de esta cola. Un peso mayor indica un mayor conteo de programación de paquetes. Si el ancho de banda de la interfaz es de 100 Mbit / s, la cola con el peso más bajo puede obtener un ancho de banda de al menos 25 Mbit / s. La programación de WRR evita que los paquetes en colas de baja prioridad no se sirvan en el modo de programación de PQ, pero la programación de WRR no puede garantizar que los paquetes de servicios sensibles a la demora se transmitan preferentemente.

6.2.6 Colas basadas en ponderación (WFQ)

El encolamiento equitativo ponderado WFQ (Weighted Fair Queuing) clasifica dinámicamente el tráfico de red dentro de flujos individuales y asigna a cada flujo una participación equitativa del total de ancho de banda. Cada flujo es clasificado como un flujo de ancho de banda alto o un flujo de ancho de banda bajo. Los flujos de ancho de banda bajo como por ejemplo el Telnet, obtienen prioridad sobre los flujos de ancho de banda alto como el tráfico FTP.
Si múltiples flujos de ancho de banda alto ocurren simultáneamente, estos compartirán el ancho de banda restante uniformemente una vez que los flujos de ancho de banda bajo han sido atendidos. Cada uno de estos tráficos es ubicado dentro de una cola individual que sigue la analogía del leaky bucket. Si los paquetes de un flujo específico excedieron la capacidad de la cola a la cual es asignado, esta cola es sujeta al descarte de la cola como todas las otras colas.
Los enrutadores equipados con tarjetas de procesador de interface versátil VIP33 (Versatile Interface Processor) pueden descargar el proceso WFQ a estas tarjetas.

En este caso, el proceso es referido como encolamiento equitativo ponderado distribuido DWFQ (definido posteriormente). Delegar el proceso WFQ a la tarjeta VIP crea memoria y ciclos de CPU adicionales desde el procesador principal disponible del enrutador. Esta arquitectura distribuida permite enrutadores de alta potencia para realizar un número largo de tareas concurrentes sin exceso en el procesador del enrutador.

En cuanto al funcionamiento, WFQ primero identifica cada flujo individual y lo clasifica como flujo de ancho de banda alto o bajo. Cada flujo es caracterizado usando la información del Anexo A2.

Una vez clasificados, los flujos son ubicados en una cola equitativa. El número por defecto de las colas dinámicas es 256. Cada cola es atendida en una manera round-robin, como en el encolamiento personalizado, dando prioridad a las colas de ancho de banda bajo. Cada cola es configurada con un umbral de descarte congestivo por defecto que limita el número de mensajes en cada cola, este valor por defecto para cada cola es de 64 paquetes.

Para flujos de ancho de banda alto, los mensajes que intentan entrar en la cola una vez alcanzado el umbral de descarte, son descartados. Sin embargo, los mensajes de ancho de banda bajo todavía pueden entrar a la cola aunque el umbral de descarte congestivo es excedido por esta cola. Los límites para las colas dinámicas y el umbral de descarte congestivo pueden ser ajustados sobre un valor de 4096 paquetes.

Hasta ahora el proceso descrito muestra un tratamiento igual de todas las conversaciones ocurridas en una interface de salida. Aparte de la diferenciación entre flujos de velocidad alta y baja, estas conversaciones no tiene ninguna prioridad o peso la una sobre la otra. Por lo tanto, este proceso sería referido como encolamiento equitativo.

Ahora bien, el factor ponderado empieza a afectar el proceso de encolamiento cuando el campo ToS o el campo de precedencia IP son diferentes. WFQ toma en cuenta la precedencia IP y da tratamiento preferencial a los flujos de precedencia

más alta ajustando sus pesos. Si todos los paquetes tienen el mismo valor de precedencia por defecto, entonces el factor ponderado no afecta el proceso WFQ.

El peso de los flujos, donde los valores presentes de ToS son diferentes, es calculado mediante la adición de 1 a la precedencia del paquete. El peso total de todos los flujos representa el ancho de banda total a ser dividido entre los flujos individuales. Por ejemplo, si tres flujos utilizan una precedencia IP por defecto de 0, cada flujo tiene un peso de 1 (0+1). El peso del ancho de banda total es 3 (1+1+1), y cada flujo representa un tercio del total del ancho de banda. En cambio, si dos flujos tiene una precedencia IP de 0, y un tercer flujo tiene una precedencia de 5, el peso total es 8 (1+1+6). Los primeros dos flujos representan cada uno un octavo del ancho de banda, mientras que el tercer flujo recibe seis octavos del ancho de banda.

Cabe indicar que cuando WFQ es configurado en un enlace, el protocolo de reservación de recurso RSVP hace uso de diferentes colas dentro del proceso de WFQ con el fin de asegurar que los requerimientos de QoS de las conversaciones RSVP sean respetados. El número por defecto reservado de las colas RSVP es 0. Esto significa que con el fin de que WFQ soporte adecuadamente RSVP, este debe ser configurado manualmente a otro valor que el valor por defecto.

WFQ es simple de implementar, es un mecanismo de encolamiento dinámico el cual asegura que toda conversación en la red alcance una compartición equitativa del ancho de banda. A diferencia de PQ y CQ, los cuales necesitan ser configurados manualmente, WFQ se adapta dinámicamente a los cambios de la red, incluyendo nuevos protocolos y aplicaciones. Si no existe tráfico crítico que debe ser dado prioridad sobre otro tráfico, WFQ es un método fácil y eficiente para proporcionar el mejor nivel de servicio a todo usuario de red.

6.3 Manejo de Congestión

6.3.1 Descarte (Tail Drop)

Tail Drop significa la ausencia completa de un gestor de la memoria de la cola. Cuando un paquete llega al final de una cola completamente llena. El paquete se descarta al igual que todos los que lleguen tras él hasta que hay espacio disponible en la cola.

Los beneficios del gestor de cola Tail Drop incluyen:

- Tail Drop es fácil de implementar por los proveedores y de entender por parte de los clientes.
- Tail Drop puede reducir el número de los paquetes descartados con un incremento del tamaño de las colas, sin embargo esto hará aumentar el retardo de extremo a extremo de todos los flujos que atraviesen la cola.

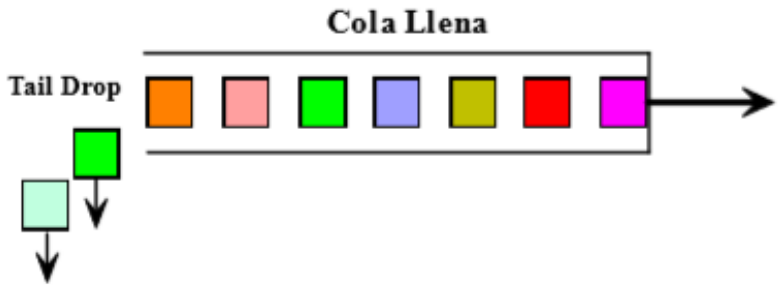


Figura 9: Funcionamiento de Tail Drop

Las limitaciones de Tail Drop son:

- Tail Drop no descarta paquetes hasta que la cola está completamente llena y los recursos consumidos completamente. Esto significa que la cola no puede absorber ráfagas de tráfico hasta que no haya espacio disponible en la cola. Esto puede dar como resultado un comportamiento de cierre, debido a la falta de espacio de buffer para almacenar los paquetes entrantes. En consecuencia un pequeño número de flujos puede monopolizar toda la capacidad del buffer e impedir a las sesiones existentes o nuevas acceder a la cola.

- Tail Drop permite a las colas permanecer llenas o casi llenas durante largos periodos de tiempo, ya que los host no reconocen la congestión (mediante el descarte de paquetes) hasta que las colas no alcanzan el 100% de su capacidad y se consumen completamente los recursos.
- Tail Drop es un algoritmo extremadamente pobre para tráfico basado en TCP. Aproximadamente del 85 al 95% del tráfico de las redes IP es TCP. TCP supone que si se tira un paquete en un router es debido a la congestión. Esto permite a las sesiones TCP controlar su propia tasa de transferencia. Sin embargo, Tail Drop produce que todas las sesiones TCP que atraviesan la cola congestionada reduzcan sus tasas de transmisión al mismo tiempo resultando un proceso conocido como sincronización global TCP. Esto produce oscilaciones drásticas en el tráfico que dan como resultado un uso ineficiente del ancho de banda de salida debido a que muchas sesiones dividen por dos sus ventanas de transmisión al mismo tiempo.
- Las sesiones individuales de TCP se recuperan más lentamente de descartes de paquetes múltiples que se un descarte individual. Esto puede reducir significativamente el caudal global de los flujos de los clientes.
- Una media de la profundidad de la cola grande incrementa el retardo de extremo a extremo experimentado por los flujos de los clientes que atraviesan la red.

Gestión Activa de la Memoria de las Colas

Tail Drop es la forma más simple de gestionar la memoria de la cola ya que representa la ausencia total de un gestor de la memoria de la cola. Los gestores de la memoria de cola activos permiten a un router responder a la congestión de forma activa si la media de los tamaños de sus colas comienza a incrementarse. En vez de esperar a que se congestione la cola y se desborde y realizar Tail Drop con todos los paquetes que lleguen, los gestores de memoria de cola activos responden a la congestión marcando o descartando los paquetes antes de que los recursos de memoria de la cola se consuman completamente.

Los beneficios de la gestión activa de las colas comparadas con Tail Drop incluyen:

- La eliminación de la sincronización global de fuentes TCP que da como resultado un uso más eficiente del ancho de banda de la red.
- El soporte de fluctuaciones momentáneas en el tamaño de la cola, que permiten absorber ráfagas sin descartar paquetes y causar que los host reduzcan sus caudales cuando reducen sus tasas de transmisión.
- La habilidad para controlar el tamaño de la cola influyendo en la media del retardo de encolamiento a través del router.

6.3.2 Random Early Detection (RED)

Monitorea el tamaño de la cola y cuando ésta alcanza un umbral determinado, selecciona aleatoriamente flujos TCP individuales de los cuales descarta paquetes con el objetivo de indicar al emisor que debe disminuir la tasa de envío.

La detección temprana aleatoria RED (Random Early Detection) es un mecanismo propuesto por Sally Floyd y Van Jacobson a principios de los 90s para direccionar la congestión de red en una respuesta más bien de manera reactiva. Lo fundamental en este mecanismo es la premisa que la mayor parte de tráfico se ejecuta sobre las implementaciones de transporte de datos que son sensibles a la pérdida y temporalmente se retardaría cuando algo de su tráfico sea descartado. TCP, el cual responde apropiadamente, incluso con firmeza, al descarte de tráfico mediante el retardo de su transmisión de tráfico, efectivamente permite el comportamiento de descarte de tráfico de RED para trabajar como un mecanismo de señalización de prevención de congestión.

Al considerar la utilidad de RED cuando transportes robustos como TCP son generalizados, es importante considerar seriamente las implicaciones negativas de emplear RED cuando un porcentaje significante de tráfico no es robusto en respuesta a la pérdida de paquetes.

En definitiva, RED es un mecanismo que previene situaciones de congestión mediante el tratamiento de comunicaciones de red cuando el enlace comienza a presentar signos tempranos de saturación. En consecuencia, con RED habilitado,

un enlace nunca debería alcanzar el punto de congestión porque este mecanismo limitará el flujo de paquetes antes que esto suceda. Esto también tiene como efecto la normalización del ancho de banda usado en un enlace y mantenerlo en la capacidad pico.

Funcionamiento

RED trabaja por descarte aleatorio de paquetes de diferentes conversaciones. Utiliza ventana deslizante de TCP/IP y mecanismos de recuperación rápida para forzar la comunicación a reducir la velocidad en la cual se están transmitiendo paquetes, en consecuencia reduce el uso de ancho de banda de esa conversación particular. Mediante la aplicación de este principio aleatorio a varias comunicaciones en marcha, RED puede retrasar las cosas ya que detecta que un enlace se aproxima a un estado de congestión. RED no es apropiado en situaciones donde el tráfico UDP es predominante, esto porque RED no tiene efectos apreciables sobre este.

Con el fin de comprender como opera RED, es importante entender el mecanismo fundamental que RED utiliza para reducir las comunicaciones.

A medida que el remitente envía trenes de paquetes, el receptor reconoce el último paquete del tren e informa que la transmisión fue satisfactoria. Además, instruye al remitente que puede aumentar el número de paquetes por tren o tamaño de ventana, en su siguiente transmisión. Si no se controla, las sesiones TCP incrementarán su tamaño de ventana hasta que un paquete es descartado y un NAK es enviado por el receptor, o hasta que una salida de secuencia ACK es recibida por el remitente. En este punto, TCP recupera en la última secuencia ACK satisfactoria y reduce el tamaño de ventana en un intento de lograr una comunicación exitosa.

Cuando múltiples sesiones TCP operan sobre un enlace común, todas aumentarán el tamaño de sus ventanas deslizantes tanto como las ACKs satisfactorias son recibidas. Gradualmente esta progresión sincronizada consume el ancho de banda del enlace hasta que el enlace este congestionado. En este punto, todas las

conversaciones TCP experimentan un error de transmisión, resultando un descarte considerable en el uso de
ancho de banda tal como todas las conexiones TCP se mueven a tamaños de ventana deslizante más pequeños simultáneamente. Este proceso es llamado sincronización global, y crea problemas sobre el enlace debido al hecho que todas las corrientes de entonces comenzarán a retroceder simultáneamente, guiando a otra situación de congestión. Este ciclo continúa una y otra vez, creando picos y valles de utilización de ancho de banda en el enlace.

RED trata de prevenir esta fluctuación en ancho de banda mediante el descarte aleatorio de paquetes de varias conexiones mientras el enlace se aproxima a un estado de congestión. Por lo tanto, las ventanas de las conexiones TCP se reducen una por una tal como el algoritmo aleatorio de RED desecha paquetes desde sus conexiones. Esto resulta en una normalización de tráfico de red cerca al punto de congestión del enlace, en lugar de tener retornos masivos tal como todas las conexiones TCP descartan paquetes cuando alcanzan el punto de congestión de la conexión.

6.3.3 RED ponderado (WRED)

La detección temprana aleatoria ponderada WRED (Weighted Random Early Detection) combina las capacidades del algoritmo RED con la característica de la precedencia IP de proporcionar un tratamiento preferencial para el tráfico que incluye paquetes de mayor prioridad. Selectivamente, WRED puede descartar tráfico de prioridad más baja cuando la interfaz comienza a estar congestionada y proveer características de funcionamiento diferenciado para diferentes clases de servicio.
Para interfaces configuradas para utilizar RSVP, WRED escoge paquetes de otros flujos para descartar en lugar de los flujos con RSVP. Además, la precedencia IP domina cuáles paquetes son descartados, el tráfico que está con una precedencia

baja tiene una tasa de descarte mayor y por lo tanto es más probable que se reduzca.

WRED difiere de otras técnicas de prevención de congestión como las estrategias de encolamiento porque procura anticipar y evitar la congestión además de controlar la congestión una vez que esta se produzca.

Beneficios

WRED efectúa detección temprana de congestión y suministra para múltiples clases de tráfico. También protege contra la sincronización global. Por estas razones, WRED es útil en cualquier interfaz de salida donde se espera que ocurra saturación. Sin embargo, por lo general WRED es utilizado en los enrutadores de núcleo de una red en lugar de los enrutadores de borde sobre esta. Los enrutadores de borde asignan precedencias IP a los paquetes mientras van ingresando a la red de comunicación. WRED utiliza estas precedencias para determinar los diferentes tipos de tráfico.

WRED establece los umbrales por separado y pondera las diferentes precedencias IP, permitiendo proporcionar diferentes calidades de servicio en lo que respecta al descarte de paquetes para diferentes tipos de tráfico. El tráfico estándar puede ser descartado más frecuentemente que otra clase de tráfico durante los períodos de congestión.

Funcionamiento

Al descartar paquetes aleatoriamente antes de los períodos de congestión, WRED anuncia al origen de paquetes para disminuir su velocidad de transmisión. Si el origen de paquete está utilizando TCP, disminuye su velocidad de transmisión hasta que todos los paquetes alcanzan su destino, lo cual indica que la congestión está clareada.

Generalmente WRED descarta paquetes basado en la precedencia IP. Los paquetes con una precedencia IP mayor son menos probables para ser descartados

que los paquetes con una precedencia menor. En consecuencia, si mayor es la prioridad de un paquete, mayor es la probabilidad de que el paquete será entregado.

WRED reduce las posibilidades de descarte de cola de forma selectiva descartando paquetes cuando la interfaz de salida empieza a mostrar signos de congestión. Al descartar paquetes tempranamente en lugar de esperar hasta que la cola se llene, WRED permite descartar números largos de paquetes a la vez y minimizar las posibilidades de la sincronización global. Por lo tanto, WRED permite a la línea de transmisión ser usada completamente en todo tiempo.

Adicionalmente, estadísticamente WRED descarta más paquetes de los grandes usuarios que de los pequeños. Por lo tanto, los orígenes de tráfico que generan el mayor tráfico son más probables a ser retardados que los orígenes de tráfico que generan tráfico pequeño.

WRED evita los problemas de globalización que ocurren cuando el descarte de la cola es utilizado como mecanismo de prevención de congestión. La sincronización global se manifiesta cuando múltiples hosts TCP reducen sus velocidades de transmisión en respuesta al descarte de paquetes, entonces aumenta sus velocidades de transmisión una vez más cuando la congestión es reducida.

WRED solo es útil cuando la mayoría de tráfico es tráfico TCP/IP. Con TCP, paquetes descartados indican congestión, por lo que el origen de paquete reducirá sus velocidades de transmisión. Con otros protocolos, los orígenes de paquete pueden no responder o pueden reenviar paquetes descartados en la misma tasa. En consecuencia, el descarte de paquetes no disminuye la congestión.
WRED trata al tráfico no IP como precedencia 0, la más baja precedencia. Por tanto, el tráfico no IP, en general, es más probable a ser descartado que el tráfico IP.

Bibliografía:

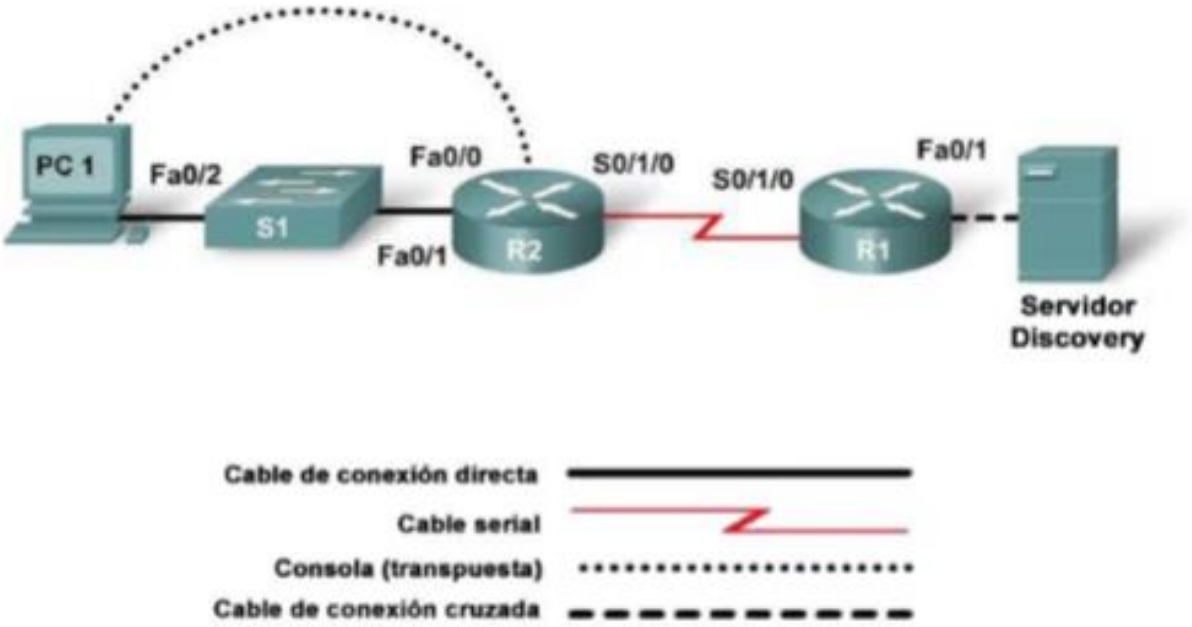
- Cadena C. (2010).Escuela Politécnica del Ejército. Tesis para obtener el grado de Ingeniero. Control de Tráfico en Redes TCP/IP Fundamentado en Procedimientos y Técnicas de Calidad de Servicio a lo Largo de una Infraestructura de Telecomunicaciones. Ecuador.
- Rubio, G. (2016). QoS en routers y switches Cisco. ISBN: 978-1-329-84032-4.
- Alarcón R. (2003). Estudio e Implementación de Mecanismos de Calidad de Servicio sobre una Arquitectura de Servicios Diferenciados. Tesi para optar al título de Ingeniería Técnica de Telecomunicación, especialidad Telemática. Colombia.
- J.A. Jimenez Toro. UF1875: Gestión de recursos, servicios y de la red de Comunicaciones. Editorial Elearning SL. Edición 5.0. ISBN: 978-84-16199-01-3. España.

Casos de Estudio

1. Colas de prioridad

Objetivo

• Explicar dónde se puede implementar la calidad de servicio para afectar el flujo de tráfico



Designación del dispositivo	Nombre del dispositivo	Dirección	Máscara de subred
Servidor Discovery	Servicios de red	172.17.1.1	255.255.0.0
R1	ISP	Fa0/1 172.17.0.1 S0/1/0 10.10.0.1	255.255.0.0 255.255.255.252
R2	FC-CPE-1	Fa0/0 10.0.0.1 S0/1/0 10.10.0.2	255.255.255.0 255.255.255.252
S1	FC-ASW-1	—	—
PC1	Host1	10.0.0.200	255.255.255.0

1. Información básica / Preparación

FilmCompany es una compañía de publicidad en crecimiento que se está desplazando a los medios de publicidad interactivos, incluidas las presentaciones de video. Esta compañía acaba de obtener un importante contrato de soporte de video con StadiumCompany. Con este nuevo contrato, FilmCompany espera que su

negocio crezca aproximadamente un 70%. La actualización de red requerida para soportar este crecimiento en los negocios tendrá que cargar una gran variedad de tipos de tráfico de datos. Algunos de estos tipos de datos requerirán acceso de prioridad a recursos de red para garantizar su entrega eficaz y útil. En esta práctica de laboratorio, el usuario examina y aplica algunos de los comandos IOS de Cisco para configurar la cola de prioridad en el router.

Paso 1: Realizar el cableado y configurar la red

NOTA: Si la PC utilizada en esta práctica de laboratorio también está conectada a su red LAN de la Academia o a Internet, asegúrese de documentar las conexiones del cable y configuraciones TCP/IP para que éstas puedan ser reestablecidas al finalizar la práctica.

a. Conecte y configure los dispositivos de acuerdo con la topología y configuración dadas.

- Debe configurarse el enrutamiento a través del enlace serial WAN para establecer la comunicación de datos.
- Configure el acceso a Telnet en cada router. NOTA: El instructor debe sustituir el servidor Discovery por un servidor equivalente para esta práctica de laboratorio.

b. Haga un ping entre el Host1 y el servidor Discovery para confirmar la conectividad de la red.

- Confirme la conectividad de la capa de aplicación con la utilización de telnet desde R2 a R1.
- Realice el diagnóstico de fallas y establezca la conectividad si los pings o Telnet fallan. c. Luego de confirmar las configuraciones iniciales, mantenga una conexión de sesión de terminal de la consola con R2.

Paso 2: Examinar los comandos de la cola de prioridad Configuración de la cola de prioridad La configuración de la cola de prioridad (PQ, priority queueing) consta de dos pasos obligatorios y uno opcional:

1. Definir la lista de prioridad (Obligatorio)

2. Asignar la lista de prioridad a una interfaz (Obligatorio)

3. Monitorear las listas de la cola de prioridad (Opcional) La lista de prioridad contiene las definiciones para un conjunto de colas de prioridad. La lista de prioridad específica en qué cola se va a colocar un paquete y, opcionalmente, la longitud máxima de las distintas colas. Para realizar una cola utilizando una lista de prioridad, debe asignar la lista a una interfaz. La misma lista de prioridad puede aplicarse a múltiples interfaces. O bien, puede crear distintas políticas de prioridad para aplicar a distintas interfaces.

Definición de la lista de prioridad

La lista de prioridad se define:

1. Asignando paquetes a colas de prioridad

2. Especificando el tamaño máximo de las colas de prioridad (Opcional) Los paquetes se asignan a las colas de prioridad de acuerdo con el tipo de protocolo y de la interfaz en donde los paquetes entran al router.

Los comandos priority-list se leen en orden de aparición, hasta que se encuentre un protocolo o tipo de interfaz que coincida. Cuando se encuentra la coincidencia, se asigna el paquete a la cola correspondiente y la búsqueda culmina. Los paquetes que no coinciden con otras reglas de asignación, se asignan a la cola predeterminada. Los siguientes comandos del modo de configuración global se utilizan para especificar en qué cola se coloca un paquete.

El formato del comando es **priority-list list-number**

Utilice un número de lista de 1 y vea las opciones disponibles.

- Ingrese el siguiente comando y vea las opciones disponibles.
  FC-CPE-1(config)#priority-list 1 ?
- Vea algunas de las opciones de protocolo disponibles.
  FC-CPE-1(config)#priority-list 1 protocol ?

Paso 3: Configurar una cola de prioridad de ejemplo

Desde el modo de configuración global, introduzca los siguientes comandos.

FC-CPE-1(config)#priority-list 1 protocol http high

FC-CPE-1(config)#priority-list 1 protocol ip normal tcp ftp
FC-CPE-1(config)#priority-list 1 protocol ip medium tcp telnet
¿Qué establecen estos comandos?

Paso 4: Examinar el funcionamiento de las colas de prioridad
a. En el Host 1, inicie el explorador Web y escriba la URL http://172.17.1.1 para acceder a los servicios Web configurados en el servidor.
b. Utilice el FTP para descargar un archivo. En Host 1, inicie una nueva ventana de explorador Web y escriba la URL ftp://172.17.1.1, o desde la línea de comandos introduzca ftp 172.17.1.1
c. Descargue un archivo grande del servidor; por ejemplo, el archivo de programa de instalación de Thunderbird. d. Introduzca el siguiente comando desde el modo EXEC privilegiado: FC-CPE-1#show queueing interface s0/1/0

Paso 5: Determinar los requisitos de la cola de prioridad para el estudio del caso
a. Al utilizar el estudio del caso FilmCompany, cuáles cree que serán los requisitos de la cola de prioridad?
b. Analice y compare sus prioridades con los demás estudiantes.
c. Corrija las sentencias de su lista de prioridades de modo que incluyan el tráfico asociado con la actualización de la red propuesta.

Módulo 7. Calidad de Servicio basado en el protocolo IPv6

Objetivos:

- Conocer las principales características del protocolo IPv6 y sus herramientas para proporcionar calidad de servicio en las redes.

De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje abordaremos el tema del protocolo IPv6, sus características en comparación a su versión anterior el IPv4, su arquitectura y funcionamiento que le ´permiten a este proveer de calidad de servicio a la red donde opera.

7. Calidad de Servicio basado en el protocolo IPv6

7.1 Definición

El protocolo IPv6 soluciona todas las limitaciones de IPv4 en materia de escalabilidad y flexibilidad. El objetivo del diseño en IPv6 era crear un protocolo que pudiera manejar el gran crecimiento de Internet y suplir los requerimientos de servicios, movilidad y seguridad.

Las principales mejoras de IPv6 se pueden resumir en:

- **Extensión del espacio de direccionamiento:** IPv4 tenía un espacio de direccionamiento de 32 bits el cual en IPv6 es incrementado a 128 bits, espacio suficiente para asignar millones de direcciones IPv6 a cada habitante del planeta hoy en día. Adicionalmente permite una distribución de direcciones en el mundo más organizada, logrando un uso eficiente del espacio de direccionamiento y un enrutamiento más eficiente.
- **Auto configurable:** En IPv6 un dispositivo que desee conectarse a una red puede utilizar su propia dirección de capa 2 y tomar un prefijo para auto asignarse una única dirección IPv6. En IPv4 era necesaria la utilización del protocolo DHCP para asignar la dirección IPv4 a cada dispositivo.
- **Simplificación de la cabecera:** IPv6 fija la longitud de la cabecera a 40 bytes, elimina algunos campos de IPv4 e introduce otros, 32 bytes están destinados a la información de direcciones de origen y destino, los restantes 8 bytes están destinados a información general de la cabecera. IPv6 al tener una longitud de cabecera fija hace más eficiente el procesamiento.
- **Soporte mejorado a opciones y extensiones:** Las opciones en IPv4 eran incluidas en la cabecera, IPv6 redirecciona estas opciones a extensiones de cabecera adjuntas a la cabecera fija, y son utilizadas solo cuando es necesario para no comprometer la velocidad de procesamiento. La especificación base describe un conjunto de cabeceras de extensión, cabeceras para enrutamiento, movilidad, calidad de servicio y seguridad. El uso de extensión de cabecera permite agregarle nuevas características al protocolo IPv6 para futuras necesidades.

7.2 Encabezados de QoS en IPv6

El tamaño del encabezado que el protocolo IPv6 añade a los datos es de 320 bit, el doble que en la versión 4. Sin embargo, este nuevo encabezado se ha simplificado con respecto al anterior. Algunos campos se han retirado de la misma, mientras que otros se han convertido en opcionales por medio de las extensiones. De esta manera los routers no tienen que procesar parte de la información del encabezado, lo que permite aumentar de rendimiento en la transmisión.

- **Versión:** Número de versión del protocolo IP, que en este caso contendrá el valor 6. Tamaño: 4 bit.
- **Prioridad:** Contiene el valor de la prioridad o importancia del paquete que se está enviando con respecto a otros paquetes provenientes de la misma fuente. Tamaño: 4 bit.
- **Etiqueta de flujo:** Campo que se utiliza para indicar que el paquete requiere un tratamiento especial por parte de los routers que lo soporten. Tamaño: 24 bit.
- **Longitud:** Es la longitud en bytes de los datos que se encuentran a continuación del encabezado. Tamaño: 16 bit.
- **Siguiente encabezado:** Se utiliza para indicar el protocolo al que corresponde el encabezado que se sitúa a continuación de la actual. El valor de este campo es el mismo que el de protocolo en la versión 4 de IP. Tamaño: 8 bit.
- **Límite de existencia:** Tiene el mismo propósito que el campo de la versión 4, y es un valor que disminuye en una unidad cada vez que el paquete pasa por un nodo. Tamaño: 8 bit.
- **Dirección de origen:** El número de dirección del host que envía el paquete. Su longitud es cuatro veces mayor que en la versión 4. Tamaño: 128 bit.
- **Dirección de destino:** Número de dirección de destino, aunque puede no coincidir con la dirección del host final en algunos casos. Su longitud es cuatro veces mayor que en la versión 4 del protocolo IP. Tamaño: 128 bit.

7.3 Funcionamiento de QoS en IPv6

El protocolo IPv6 tiene dos campos que pueden ser utilizados como herramientas para implementar QoS: Etiqueta de Flujo y Clase de Tráfico.

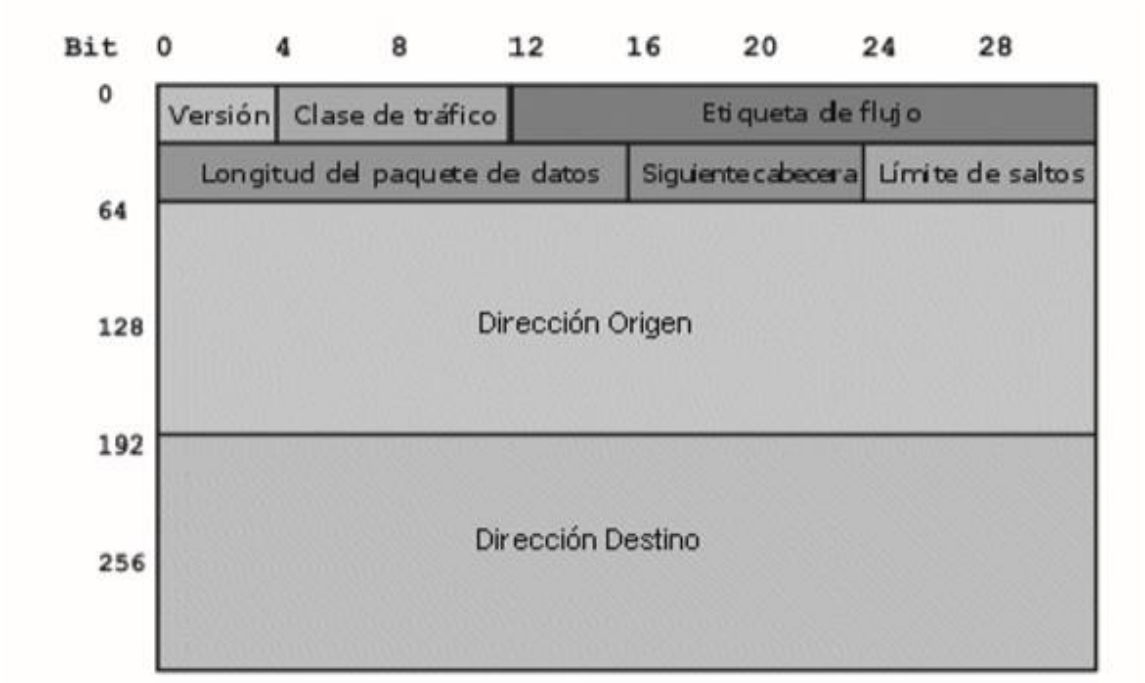


Figura 10: Arquitectura de IPv6

El campo Etiqueta de Flujo de 20 bits en la cabecera IPv6 se agrega para permitir el etiquetado de paquetes que pertenecen a flujos de tráfico particulares y puede ser usado por el origen para etiquetar secuencias de paquetes para las cuales solicita un manejo especial por parte de los enrutadores IPv6, tal como la calidad de servicio no estándar o el servicio en tiempo real. Se exige a los hosts o a los enrutadores, que no dan soporte a las funciones del campo Etiqueta de Flujo, poner el campo en cero al enviar un paquete, pasar el campo inalterado al reenviar un paquete e ignorar el campo al recibir un paquete.

El campo de ocho bits Clase de Tráfico en la cabecera IPv6 es utilizado por los nodos origen y/o enrutadores intermedios para identificar y distinguir entre las diferentes clases o prioridades de paquetes IPv6; su función es similar al campo ToS de IPv4.

Los siguientes requisitos generales se aplican al campo Clase de Tráfico:

- La interface de servicio para el servicio IPv6 dentro de un nodo debe proporcionar un medio para que un protocolo de capa superior proporcione el valor de los bits Clase de Tráfico en los paquetes originados por éste. El valor por defecto debe ser cero para todos los ocho bits.
- Los nodos que soportan un uso específico de algunos o todos los bits Clase de Tráfico se les permite cambiarlos en los paquetes que los nodos originan, reenvían o reciben, como sea requerido para ese uso específico. Los nodos deben ignorar y dejar sin alterar los bits del campo Clase de Tráfico para los cuales no dan soporte a un uso específico.
- Un protocolo de capa superior no debe asumir que el valor de los bits Clase de Tráfico en un paquete recibido es el mismo que el valor enviado por el origen del paquete.

Bibliografía

- Salcedo J., López D., Ríos A. Desempeño de la Calidad del Servicio (QoS) sobre IPv6. Revista Tecnura vol. 15. Enero-Junio 2011. Colombia.
- Ramírez G. (2013). Agente Administrador de QoS sobre Redes IPv6. Pontificia Universidad Javeriana. Tesis para optar al título de Magister en Ingeniería Electrónica. Colombia.

Caso de Estudio

1. IPv6 – Análisis del encabezado de los paquetes.

En este ejercicio vamos a analizar el encabezado del protocolo IPv6 e intentaremos descubrir algunas diferencias con respecto a IPv4.

Vamos a capturar los paquetes de pings, v4 y v6, enviados de Sx1 a Rx2, en el router Rx2.

- En el router Rx2:

```
[root@RX2 ~]# ip addr show eth2
4: eth2: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP qlen 1000
    link/ether 00:18:51:6F:B7:31 brd ff:ff:ff:ff:ff:ff
    inet 172.2x.4.1/30 brd 172.2x.4.3 scope global eth2
    inet6 fe80::218:51ff:fe6f:b731/64 scope link
       valid_lft forever preferred_lft forever

[root@RX2 ~]# tcpdump -i eth2 -s 0 -w /captura/exerc04.pcap
tcpdump: listening on eth2, link-type EN10MB (Ethernet), capture size 65535 bytes
```

- En el servidor Sx1:

```
[root@SX1 /]# ping -c 5 172.2x.4.1
PING 172.2x.4.1 (172.2x.4.1) 56(84) bytes of data.
64 bytes from 172.2x.4.1: icmp_seq=0 ttl=64 time=2.06 ms
64 bytes from 172.2x.4.1: icmp_seq=1 ttl=64 time=0.031 ms
64 bytes from 172.2x.4.1: icmp_seq=2 ttl=64 time=0.081 ms
64 bytes from 172.2x.4.1: icmp_seq=3 ttl=64 time=0.038 ms
64 bytes from 172.2x.4.1: icmp_seq=4 ttl=64 time=0.044 ms

--- 172.2x.4.1 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4002ms
rtt min/avg/max/mdev = 0.031/0.451/2.061/0.805 ms, pipe 2
[root@SX1 /]# ping6 -c 5 fe80::218:51ff:fe6f:b731 -I eth0
PING fe80::218:51ff:fe6f:b731(fe80::218:51ff:fe6f:b731) from fe80::218:51ff:fe32:cc5f eth0:
56 data bytes
64 bytes from fe80::218:51ff:fe6f:b731: icmp_seq=0 ttl=64 time=0.061 ms
64 bytes from fe80::218:51ff:fe6f:b731: icmp_seq=1 ttl=64 time=0.085 ms
64 bytes from fe80::218:51ff:fe6f:b731: icmp_seq=2 ttl=64 time=0.035 ms
64 bytes from fe80::218:51ff:fe6f:b731: icmp_seq=3 ttl=64 time=0.089 ms
64 bytes from fe80::218:51ff:fe6f:b731: icmp_seq=4 ttl=64 time=0.036 ms

--- fe80::218:51ff:fe6f:b731 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4010ms
rtt min/avg/max/mdev = 0.035/0.061/0.089/0.023 ms, pipe 2
```

Compare los paquetes IPv4 e IPv6... Identifique cada uno de los campos del encabezado IP en los dos casos y observe sus valores. Compare también el ICMP. Responda las siguientes preguntas:

- ¿Cuál es la diferencia de tamaño entre el encabezado IPv4 y el encabezado IPv6?
- ¿Hay también diferencias en el encabezado ICMP? ¿Cuáles?

Módulo 8. Calidad de Servicios en Redes ATM

Objetivos:

- Comprender el modo de funcionamiento de las redes ATM, su arquitectura y modos de conexión que proporcionan calidad de servicio.

¿De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje estudiaremos las redes ATM y como esta ofrece calidad de servicio a sus clientes, ya que esta es una red basada en conmutadores, lo cual tiene sus ventajas sobre el bus de datos como son: Reservar ancho de banda, Mayor ancho de banda, velocidades flexibles y procedimientos de conexión bien definidos. Está y otras características la hacen ideal para prestar muchos servicios garantizados.

8. Calidad de Servicio en Redes ATM

8.1 Definición

El Modo de Transferencia Asíncrona es un protocolo de transporte de alta velocidad, actualmente se encuentra implementado principalmente en redes locales en compañías que requieren altas velocidades para transferencia de datos. ATM en Redes de Área Amplia (*WAN)* proporciona un backbone de conmutación de las redes que así lo requieran y tiene facilidad de conexión a redes de alta velocidad (Como carriers y proveedores de servicios). Los anchos de banda soportados por ATM permiten el transporte de vídeo, voz y datos.

Esta tecnología define dos velocidades de transmisión, STM-1 (155Mbps) y STM-4 (620Mbps). Actualmente esta tecnología es utilizada ampliamente, sin embargo está siendo sustituida por medios de transmisión síncronos y ópticos.

ATM es una tecnología de conmutación de paquetes relativamente nueva, basada en la Red Digital de Servicios Integrados de Banda Ancha (B- ISDN). La característica más relevante en cuanto al envío de paquetes es que estos son de longitud fija, 53 bytes, de los cuales 48 son la información (payload) y los 5 restantes son el encabezado (header) que es donde se lleva acabo el direccionamiento. Esta característica permite diferentes tipos de tráfico en la misma red ya que la información es transportada de una manera segura y predecible gracias a la longitud física de sus paquetes.

Otra característica es que ATM está basado en conmutadores, lo cual tiene sus ventajas sobre el bus de datos como son: Reservar ancho de banda, Mayor ancho de banda, velocidades flexibles y procedimientos de conexión bien definidos.

8.2 Modos de Conexión

Existen principalmente cuatro tipos de conexiones en ATM:

- **Conexiones virtuales permanentes:** La conexión se efectúa por mecanismos extremos, principalmente a través del gestor de red, por medio del cual se programan los elementos de conmutación entre fuente y destino.

- **Conexiones virtuales conmutadas:** La conexión se efectúa por medio de un protocolo de señalización de manera automática. Este tipo de conexión es la utilizada habitualmente por los protocolos de nivel superior cuando operan con ATM. Dentro de estas conexiones se pueden establecer dos configuraciones distintas:
    1. **Conexión punto a punto:** Se conectan dos sistemas finales ATM entre sí, con una comunicación uni o bidireccional.
    2. **Conexión punto multipunto**: Conecta un dispositivo final como fuente con múltiples destinos finales, en una comunicación unidireccional.

8.3 Arquitectura Básica

La arquitectura ATM está dividida en tres capas:

- **Adaptación:** Divide los diferentes tipos de datos en el *payload*.
- **Capa intermedia:** Añade los datos de la Capa de Adaptación (*OSI*) con los 5 bytes del encabezado y garantiza que el paquete será enviado por la conexión adecuada.
- **Capa física:** Define las características físicas del enlace entre las interfaces de red. ATM no está ligado a un tipo de transporte físico en particular, este puede ser par trenzado, coaxial u óptico.

En el campo de las VPN's el Modo de Transferencia Asíncrona reserva circuitos Virtuales Permanentes (PVC) con un ancho de banda determinado para cada uno de los puntos a conectar. Los PVC son líneas virtuales punto a punto que se interconectan a través de un circuito establecido.

Como ya se mencionó, ATM define una celda de tamaño fijo con una longitud de 53 bytes. Consta de dos partes: la carga útil o payload de 48 bytes que transporta la información generada por un emisor o transmisor, y el encabezamiento o header de 5 bytes que contiene la información necesaria para la transferencia de la celda. Las

celdas son enviadas sobre una estructura de transmisión física, como por ejemplo el DS1, DS3 o SONET de Norte América; el E1, E3, E4 o STM de Europa.

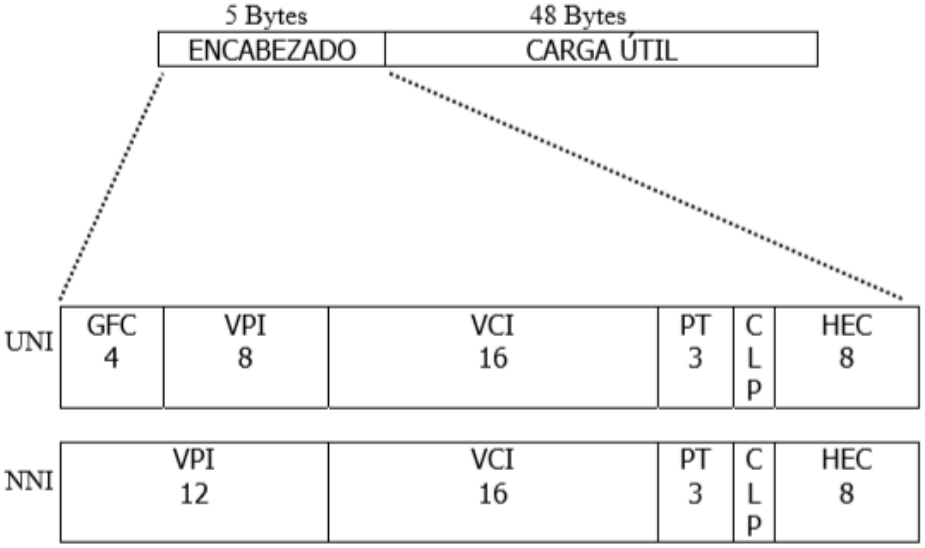


Figura 11: Arquitectura básica de ATM

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

Descripción de los campos de encabezamiento de una celda ATM

- **Control de Flujo Genérico (CFG, Generic Flow Control):** Este campo consta de 4 bits que corresponden a los más significativos del primer byte. Este campo se puede utilizar desde la interfaz del usuario, para asegurar el acceso apropiado de varios terminales sobre un medio compartido (bus o anillo).
- **Identificador de Camino Virtual (VPI, Virtual Path Identifier) y de Canal Virtual (VCI, Virtual Circuit Identifier):** Representa una dirección lógica que identifica el circuito virtual al cual la celda está conectada. Los campos de

identificación VCI y VPI son esenciales para el enrutamiento y la multiplexación.

- **Tipo de Carga Útil (PT, Type Payload):** Este campo de información permite que el nodo ATM pueda distinguir en una conexión, si el contenido de las celdas corresponde a un usuario o es información de control y gestión de la red. El campo PT contiene 3 bits, en la tabla 1 se describen cada uno de los ocho valores posibles de este campo.
- **Prioridad de Perdida de la Celda (CLP, Cell Loss Priority):** Este campo se emplea para indicarle a la red cuáles celdas, dentro de una conexión determinada, son más sensibles a una pérdida que otra. Aquellas celdas que contienen el bit CLP en 1 tienen prioridad más baja y por tal motivo serán descartadas de la red en primera opción frente a una posible congestión de tráfico.
- **Control de Errores de Encabezamiento (HEC, Header Error Check):** Proporciona un CRC de los primeros 40 bits que detecta todos los errores simples y la mayoría de los errores múltiples. En el encabezamiento de las celdas, el transmisor calcula el valor HEC para la totalidad del encabezamiento de la celda, excluido el campo HEC.

8.4 Clases de Servicio

Las clases de servicios en Redes ATM se detallan de la siguiente forma:

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

  La categoría de servicio CBR soporta aplicaciones en tiempo real, requiriendo una cantidad fija de ancho de banda. CBR soporta

ajustadamente los parámetros MCTD (Máximo retardo de transferencia de una celda) y CDVT (Tolerancia en la variación en el retardo de una celda). CBR es perfecto para aplicaciones que no puedan tolerar variaciones en la demora, como aquellas que manejan voz y vídeo en forma constante.

- **Servicio VBR (Variable Bit Rate, Tasa de Bit Variable):** Está pensado para cuando se prevé una elevada cantidad de tráfico de forma continuada.
  Tiene dos modalidades:
  RT-VBR (Real Time VBR), con requerimientos de bajo retardo y jitter para cuando se trata de aplicaciones en tiempo real (videoconferencia, vídeo bajo demanda, etc.), y
  NRT-VBR (Non Real Time VBR) para cuando se trata de aplicaciones de tráfico elevado pero donde el retardo no es tan importante, por ejemplo correo multimedia o transmisión de ficheros MPEG por la red que son vistos luego por el usuario localmente de forma asíncrona en su computador. En VBR el usuario especifica un ancho de banda medio pero, en función de sus necesidades y del estado de la red, podrá en muchas ocasiones utilizar anchos de banda superiores, lo cual da mayor flexibilidad y permite al usuario ajustar más este recurso a sus necesidades medias reales. En algunos servicios VBR el tráfico excedente sale marcado con el bit CLP. Desde el punto de vista de la red VBR tiene una complejidad superior a CBR.
- **Servicio ABR (Available Bit Rate, Tasa de Bit Disponible).** ABR está pensado para tráfico a ráfagas, se supone que habrá instantes de gran demanda de capacidad seguidos de otros de total inactividad. La meta de este servicio es el de proveer dinámicamente el ancho de banda que actualmente no está en uso por otras categorías de servicios a usuarios que pueden ajustar sus transmisiones a esa tasa. ABR permite establecer un mínimo garantizado en el ancho de banda, y fijar un máximo orientativo. ABR es la única categoría de servicio ATM en la que se pronostica que la red suministre control de flujo al emisor para que reduzca el ritmo en caso de congestión; esta circunstancia hace de ABR la categoría de servicio más

apropiada para tráfico de datos, por ejemplo para enviar datagramas IP cuando no se utilicen aplicaciones isócronas, ABR también es recomendado en las interconexiones del tipo LAN, trasferencias de archivos de alta prestaciones, archivos de bases de datos y navegadores web. Sin embargo debido a su funcionalidad ABR es la categoría de servicio más compleja de implementar.

- **Servicio UBR (Unspecified Bit Rate, Tasa de Bit No Especificada).** Se puede considerar el de más baja calidad. No existe ningún tipo de garantías en cuanto al retardo o ancho de banda, y tampoco se informa al emisor en caso de congestión. UBR utiliza la capacidad sobrante en la red de las demás categorías de servicio. Puede utilizarse para emulación de LAN, IP sobre ATM y tráfico de misión no crítica.

Bibliografía:

- Redes e Internet de Alta velocidad; Rendimiento y Calidad de Servicio, segunda edición, William Stalling, Ed. Pearson/Prentice Hall, 2002.
- Calidad de Servicio en Redes IP, ATM Y FRL. Mecanismos, protocolos y Aplicaciones. Edición 2002- ISBN 84-89416-31-1. 199 pág. https://lmdata.es/reports/qos.htm

Caso de Estudio

1. Red Multicampus de la Universidad George Mason

El objetivo es proporcionar buen acceso a internet y a intranet a gran cantidad de usuarios distribuidos en un entorno inseguro. Asimismo, era necesario sustituir y unificar las diferentes tecnologías y protocolos de las LANs existentes para poder proporcionar interoperabilidad entre sus usuarios.

La solución adoptada fue recablear con fibra óptica las conexiones entre todos los edificios formando un backbone ATM.

Mediante esta red se soluciona la conexión inter-campus, y la conexión a Internet se soluciona mediante un enlace de cada campus a 45 Mb/s al GigaPop más cercano, teniendo como backup una conexión SMDS a 10 Mb/s con Bell Atlantic.

Dado que la red tiene que transportar tanto aplicaciones críticas como acceso de estudiantes a internet, es preciso proveer de algún tipo de seguridad al sistema. La forma de hacerlo es instalando firewalls a la entrada/salida de los sistemas críticos en vez de proteger mediante firewall el acceso a toda la red, dado que el control y configuración necesario se complicaría demasiado.

Como protección frente a los usuarios internos, es decir, los propios estudiantes, se decidió asignar a cada boca una dirección IP fija, que permita hacer un traceo de los paquetes IP hasta su origen.

Por último, la gestión de la red se realiza mediante un sistema de supervisión SNMP, algo lógico dada la gran diversidad de equipamiento.

En este ejemplo se ve cómo es posible la coexistencia en una misma red de tráfico interno con tráfico de internet, garantizando la seguridad de los sistemas críticos mediante firewalls. La utilización de tecnología ATM garantiza los niveles deseados de QoS y el ancho de banda para las aplicaciones críticas.

1. Explique de que forma se proveerá QoS a esta red.
2. Defina las clases de servicios predominantes en esta red.

Módulo 9. Mediciones y Monitorización de la Calidad de Servicio

Objetivos:

- Conocer métodos y procedimientos que nos ayuden a medir o monitorear la calidad de servicio que proporciona la red.
- Identificar las medidas intrusivas y las no intrusivas que se pueden poner en prácticas para el monitoreo o medición del nivel de QoS en una red.

¿De qué trata esta sesión de aprendizaje?

En esta sesión de aprendizaje nos enfocamos en los métodos y procedimientos para medir y monitorear la calidad de servicio que proporcionan las redes de computadoras. Distinguiremos las características de las medidas intrusivas y no intrusivas como métodos para monitorear y medir la calidad de servicio.

9. Mediciones y Monitorización de la Calidad de Servicio

Una vez que se habilita QoS en una Red mediante algunas de las Arquitecturas estudiadas anteriormente es necesario e imprescindible desarrollar y conocer técnicas que permitan comprobar cómo está trabajando esa red.

Para lograr esto, existen dos técnicas de medidas: activas y pasivas. Las técnicas activas o intrusivas son aquellas en las que se inyecta tráfico en la red con el objetivo de realizar las medidas, mientras que las técnicas pasivas o no intrusivas, se limitan a observar el tráfico existente en la red.

9.1 Mediciones Intrusivas

Las medidas intrusivas sirven para medir el retardo de transferencia de paquetes en un sentido (IPTD), la variación del retardo de paquetes IP en un sentido (IPDV), la tasa de pérdida de paquetes (IPLR) y la tasa de errores de los paquetes (IPER). En la recomendación ETSI TS 185 001 no se especifica ningún protocolo para realizar las medidas, sólo se habla del uso de paquetes de prueba.

Para llevar a cabo medidas de QoS en redes de telecomunicaciones, existen muchas herramientas de aplicación libres que generan tráfico y lo inyectan en la red, como lo son entre otras:

Herramienta generadora de tráfico	Plataforma	Licencia
Nemesis. Nemesis es una utilidad de inyección y generación de paquetes por línea de comandos para probar intrusión en redes, firewalls, etc. Nemesis puede generar paquetes e inyectar tráfico nativo ARP, DNS, Ethernet, ICMP, IGMP, IP, OSPF, RIP, TCP y DUP. http://nemesis.sourceforge.net/	Linux, Windows	GPL
SCAPY. Scapy es un programa de manipulación de paquetes interactivo. Este es capaz de descifrar paquetes de un amplio número de protocolos, enviando estos por la red, capturándolos, y mucho más. Este fácilmente realiza la mayoría de tareas clásicas como scanning, tracerouting, probing, unit tets, attacks o network discovery. Este también realiza tareas	Linux	GPL

específicas que la mayoría de otras herramientas no pueden realizar como envió de tramas invalidas, inyección de tramas 802,11, combinación de técnicas (VLAN hopping + ARP cache poisoning, VOIP decoding on WEP encrypted channel, etc.) http://www.secdev.org/projects/scapy		
Distributed Internet Traffic Generator. D-ITG (generador de tráfico internet distribuido) es una plataforma capaz de producir tráfico a nivel de paquetes con gran exactitud replicando apropiadamente procesos estocásticos para ambos IDT (inter Departure Time) y las variables PS (packet Size) aleatorias (exponencial, uniforme, cauchy, normal, pareto). D-ITG soporta generación de tráfico IPv4 e IPv6 y es capaz de generar tráfico a nivel de red, transporte y Aplicación. http://www.grid.unina.it/software/ITG/index.php	Linux/ Windows	Other
pktgen. Pktgen es una herramienta de prueba de alto desempeño incluida en el kernel de linux. Siendo parte del kernel es actualmente la mejor forma de probar el proceso TX del driver del dispositivo y NIC. Pktgen permite también ser usado para generar paquetes ordinarios para probar otros dispositivos de red. Especialmente de interes es el uso de pktgen para probar routers o bridges que usen el stack de red en Linux. Ya que pktgen está en el kernel, este puede generar una rata alta de paquetes y con poca saturación en el sistema de los dispositivos de red tales como routers o bridges.	Linux (kernel)	GPL
Packet Generator. Packet Generator es una herramienta simple para medir carga en la red y reproducir tráfico de red observado. Este es un software para transmitir tráfico vía Ethernet 10/100M desde un computador Windows. El software soporta un modo de paquetes simple para enviar repetidamente el mismo paquete y un modo buffer para regenerar tráfico capturado de la red actual. http://www.clearsightnet.com/products-packetgenerator.jsp	Windows	Commercial
Packgen. Packgen es un simple generador de paquetes de red escrito en Ruby. Esta marca manualmente servicios diferenciados (diffserv), útil para medir ancho de banda de red y QoS, esta puede generar varios flujos de	Ruby	GPL

datos, cada uno con sus propiedades tales como: nombre, destino, ancho de banda, tamaño del paquete, DSCP (Differentiated Services Code Point), y rangos de tiempo. http://packgen.rubyforge.org/files/README.html		
GASP. Gasp es un sistema analizador y generador de protocolos. Este permite construir paquetes a mano para probar el comportamiento de sus programas cuando enfrentan algún paquete desconocido. GASP está dividido en dos partes +: un compilador que toma las especificaciones del protocolo y genera el código manualmente de tal protocolo, este código es un nuevo comando Tcl. como GASP está construido sobre Tcl/Tk y extendido a script, facilidades proveídas por Tcl. http://laurent.riesterer.free.fr/gasp/	Linux/ Windows	GPL
Gspoof 3.0. Gspoof es una herramienta que con exactitud y facilidad construye y envía paquetes TCP-IP. Esta trabaja desde la consola (línea de comando) y tiene una interfase gráfica fácil de usar escrita en GTK+ too. Soporta manipulación de cabecera ethernet, manipulación de cabecera IP, manipulación de cabecera tcp, carga útil Tcp, torrentes, soporta notificación de congestión. http://gspoof.sourceforge.net/	Linux	GPL
Harpoon. Harpoon es un generador flujo a nivel de tráfico. Este usa un set de parámetros distribucionales que pueden ser automáticamente extraídos de trazas netflow para generar flujos que exhiben las mismas cualidades estadísticas presentes en las trazas medidas de Internet, incluyendo características temporales y espaciales. Harpoon puede ser usado para generar tráfico representativo Background para probar aplicaciones ó protocolos, o para probar swicht y routers. http://www.cs.wisc.edu/~jsommers/harpoon/http://wail.cs.wisc.edu/waildownload.py	Linux, Solaris 8, FreeBSD, MAcOSX	Other (non-commercial research purposes only)

Tabla 2: Herramientas para medir QoS

9.2 Mediciones No Intrusivas

El uso de medidas pasivas se propone para medir IPER e IPLR en cualquier enlace establecido entre dos encaminadores. Dichos enlaces seleccionados para una medición particular, se denominan “población de interés”

Los métodos no intrusivos no requieren de señales extra y son adecuados para monitorear la calidad en servicio. Dependiendo del tipo de entrada al método se pueden clasificar como basados en señales, la entrada es la señal transmitida por la red, o basados en parámetros donde las entradas son parámetros de la red de comunicación y parámetros de la señal en cuestión.

A diferencia de los métodos intrusivos antes presentados donde el servicio debe ser interrumpido para inyectar las señales, los métodos no intrusivos pueden ser utilizados durante el servicio. Aquí cabe aclarar que no siempre es posible utilizar estos métodos en servicio, debido a que si bien no utilizan señales extra, sí pueden inyectar algún tipo de tráfico para estimar el estado de la red.

Estos métodos se pueden clasificar en basados en parámetros o basados en señales. Los últimos predicen la calidad utilizando la señal distorsionada sin necesidad de referencia. A este tipo de método se los denomina Null Reference. Los otros predicen la calidad a partir del valor de parámetros de la red IP (por ejemplo probabilidad de pérdida, jitter, retardo) y de parámetros no específicos de la red (códec utilizado, eco, tasa de bits del video, etc.). Ejemplo de estos métodos son el E-Model y el uso de redes neuronales.

El E-Model es un modelo empírico matemático estandarizado por la ITU en la recomendación G.107 [ITUT05]. Es un conjunto de fórmulas que tienen como entrada parámetros de la red tradicional de circuitos conmutados y de la red de paquetes conmutados, y tiene como salida el factor de calidad el cual se puede mapear en MOSc. Si bien es una herramienta para la planificación de redes, actualmente es muy utilizada para predecir calidad percibida en VoIP.

Las redes neuronales se utilizan para aproximar la relación no lineal que existe entre calidad percibida (mejor dicho MOS) y el conjunto de parámetros considerado. Un conjunto de parámetros de entrada posible sería el formado por: la probabilidad de

pérdida, retardo, jitter, códec utilizado, tasa de bits del video, lenguaje en audio, etc. Para lograr el mapeo deseado se debe generar una base de entrenamiento que consiste en un conjunto de valores de los parámetros y el correspondiente valor de calidad obtenido mediante tests subjetivos.

El obtener una buena base de entrenamiento, es decir un rango considerable de variación de los parámetros, es el principal limitante debido al costo de los tests subjetivos.

La caracterización fundamental de un servicio es mediante el conocimiento de la latencia, el jitter, la contabilidad, la capacidad de absorber ráfagas y el volumen del tráfico. Al conocer para un servicio cada uno de estos parámetros, se podrá anticipar el comportamiento de la red.

Monitoreo del retardo o latencia

La latencia es uno de los parámetros de mayor importancia en el tránsito de los paquetes a través de una red. La idea más simple de una latencia está relacionada con la habilidad que posee el servidor para responder. A fin de cuantificar una verdadera latencia, se suele implementar en la red tres clases de pruebas:

PRUEBA 1:

PING: una prueba de este tipo mide la latencia de la red (considerando la ida y vuelta de un paquete, RTT Round-Tríp Time), además de dar muestras claras de la habilidad del servidor de responder al mencionado PING.

PRUEBA 2:

Prueba TCP: esta prueba incluye la configuración de una sesión TCP y la medición de la respuesta de tiempo, y otorga resultados específicos sobre el proceso realizado en el servidor.

PRUEBA 3:

Prueba de Aplicación: esta prueba indica el verdadero comportamiento (comportamiento final) del servicio.

Monitoreo del jitter

El Jitter o la variación del retardo es un parámetro muy importante para aplicaciones en tiempo real. Existen varias propuestas para medir el Jitter, una de ellas proviene del ITU (International Telecommunícation Union) y requiere la inyección de paquetes a intervalos regulares de tiempo para medir, luego, la variabilidad en los tiempos de arribo. De esta forma el jitter corresponde al Rango Intercuartil IQR1 (ínter Quartile Range) de la distribución de frecuencias de las mediciones del tiempo de respuesta.

Altos niveles de jitter tienden a indicar fluctuación en la profundidad de las colas, lo cual indica un tratamiento de tráfico muy pobre en puntos de congestión de la red, que para ser controlado pueden implementarse los algoritmos de control de congestión.

En tanto que un nivel bajo de jitter para una clase particular de tráfico es una muestra clara de la implementación de una diferenciación o tratamiento especial al tráfico en mención, debiendo tomarse muy en cuenta cuando estos valores sean sumamente pequeños, ya que pueden provocar que otras clases de tráfico sean servidas con valores de jitter muy exagerados y por lo tanto su desempeño sea pésimo.

Monitoreo en diferentes protocolos

Los distintos protocolos utilizados en las redes de transmisión de datos suelen utilizar cada uno de ellos diferentes tipos de métricas para tener conocimiento de los valores de parámetros como el retardo, el porcentaje de paquetes perdidos, la velocidad de datos, etc. En la siguiente sección se analiza las métricas o formas que utilizan cada uno de los protocolos a fin de controlar los parámetros antes mencionados.

- **Protocolo TCP:** Utiliza el descarte de paquetes y el deterioro de la respuesta de tiempo, para detectar congestión. Una vez que ésta es detectada el protocolo reacciona reduciendo la ventana de transmisión de datos, es decir reduciendo el número de paquetes circulantes entre origen y destino para esa sesión TCP. Las métricas que caracterizan una sesión TCP son: el

tamaño actual de la ventana de transmisión, un promedio de paquetes perdidos y el tiempo RTT.

- **Protocolos basados en reservaciones:** Los sistemas que basan su accionamiento en reservación de recursos, tal como el protocolo RSVP, mantienen un acuerdo centralizado sobre la capacidad existente como también una lista de reservaciones de flujos actuales mediante el control de admisión y de las negociaciones PEP/PDP. Así por ejemplo para una red que habilite Servicios Integrados mediante el Protocolo de Reservación de Recursos (RSVP) y que su administración esté centralizada, tendrá en el servidor COPS el dispositivo en el cual se encuentre la mayor cantidad de información sobre el estado de Calidad de Servicio de la mencionada red.
- **Protocolo ATM:** ATM incluye un gran número de métricas que son utilizadas para caracterizar el servicio que presta a cierta clase de tráfico y que suelen ser acordadas entre el proveedor y el usuario de la red. Este tipo de métricas depende de la clase de servicio que se haya contratado. A fin de hacer posible la realización de contratos de tráfico concretos, el estándar ATM define una serie de parámetros de Calidad de Servicio cuyos valores los pueden negociar el cliente y la portadora. Los tres primeros parámetros especifican la rapidez a la que quiere transmitir el usuario.

  Tasa pico de celdas (PCR, Peak Cell Rate): es la rapidez máxima con que el transmisor planea enviar sus celdas. Este parámetro puede ser menor que lo permitido por el ancho de banda de la línea. Si el transmisor planea sacar celdas cada 4 useg, su PCR es de 250 000 celdas/seg, aún si el tiempo real de transmisión de las celdas puede ser 2.7 useg.

  Tasa sustentable de celdas (SCR, Sustained Cell Rate): es la tasa esperada o requerida de celdas promediada en un intervalo de tiempo grande. La razón PCR/SCR es una medida de las ráfagas de tráfico.

  Tasa mínima de celdas (MCR, Mínimum Cell Rate): es la tasa mínima de cetdas/seg que el cliente considera aceptable.

Tolerancia de variación de retardo de celdas (CVDT, Cell Varíation Delay Tolerance): indica la cantidad de variación que habrá en los tiempos de transmisión de las celdas; se especifica en forma independiente tanto para la velocidad PCR como para SCR.

Las características de la red, cuantificables en el receptor son descritas por los siguientes tres parámetros, mismos que son negociables:

Razón de pérdida de celdas (CLR, Cell Loss Ratio): mide la fracción de las celdas transmitidas que no se entregan en absoluto o que se entregan demasiado tarde, siendo por lo tanto inservibles para servicios en tiempo real, por ejemplo.

Retardo de transferencia de celdas (CTD, Cell Transfer Delay): es el tiempo promedio de tránsito de las celdas desde el origen al destino.

Variación en el retardo de celdas (CDV, Ce*ll* Delay Variation): es un parámetro que mide la uniformidad con que se entregan las celdas.

Acuerdos de nivel de servicio (SLAs, Service Level Agreements)

Los acuerdos de nivel de servicio (SLAs) son contratos celebrados entre el cliente y su proveedor de servicio, mediante el cual se especifica el tratamiento que recibirá de la red el tráfico del cliente. Contienen un conjunto de parámetros que definen por un lado, las condiciones que debe cumplir el tráfico del cliente, y por otro los recursos que asignará la red del proveedor de servicios si se cumplen las condiciones. En la actualidad los SLAs son firmados especificando valores acordados de parámetros tales como disponibilidad, latencia, niveles de ráfagas, jitter.

Una organización (cliente) que contrata un servicio mediante un SLA desea tener siempre la certeza de que el ISP está sirviendo el tráfico de sus aplicaciones bajo los parámetros acordados, para ello muchas organizaciones utilizan promedios de múltiples PINGs. Siendo cada día más el número de empresas que utilizan estadísticas de respuesta de tiempo generadas por los PINGs para medir el rendimiento de sus aplicaciones.

Además es obligación de los proveedores de servicio, mantener disponibles documentos en los cuales se detallen aspectos como los siguientes:

- Interpretación de los reportes y de sus estadísticas.
- Técnica utilizada para la recolección de datos.
- Recomendaciones para optimizar la red a fin de limitar lo siguiente:
  - ✓ Inversión de capital (routers, FRADs, etc.).
  - ✓ Costos de transmisión recurrentes.
  - ✓ Optimización del ancho de banda, realizando la identificación de usuarios no autorizados o de aplicaciones que estén monopolizando los recursos.

Para ilustrar un Acuerdo de Nivel de Servicio y la forma de determinarlo considérese el siguiente ejemplo, en el cual se oferta un servicio de video MPEG2 libre de jitter. "libre de jitter implica la necesidad de eliminar completamente las variaciones de los tiempos de envío entre los paquetes de video variaciones del retardo de propagación. Sin embargo, el proveedor de servicios puede estimar un tamaño razonable para los buffers de recepción, los cuales disminuyen el jitter y permiten alguna flexibilidad. Ya que la aplicación es un flujo unidireccional de video continuo, puede tolerar el retardo de reproducción (ptayback) resultante, a diferencia de una conversación bidireccional.

La calificación de video MPEG2 caracteriza una imagen con resolución y velocidad de imagen de alta calidad, lo cual indica necesidad de gran ancho de banda. Sin embargo, no especifica un tamaño de imagen, lo que significa que no se define exactamente el requerimiento de ancho de banda. Se asume calidad de televisión digital DTV, resolución de 576x720 y una velocidad de imagen de 60 imágenes por segundo, lo cual genera un flujo de video comprimido de alrededor de 4 Mbps. Así, se requiere un buffer de recepción de alrededor de 16 Mb para permitir 4 segundos de jitter.

Estos puntos ilustran la necesidad de políticas en un Acuerdo de Nivel de Servicio bilateral que se traduzca, sin ambigüedades, en acciones específicas en este caso,

los parámetros del tráfico de la red son cuantificables y tal vez incluyan requerimientos del usuario final como el tamaño del buffer.

Bibliografía:

- Recomendaciones de la Unión Internacional de Telecomunicaciones (UIT-T, P.561) (2002). Dispositivos de medida en servicio no intrusivas.
- Grupo de Expertos NGN. (2001). “Medición de la Calidad del Servicio”. Interactic. Documentos NGN No. 01. CINTEL (Centro de Investigación de las Telecomunicaciones).

Casos de estudio

1. Captura y análisis de paquetes.

Para la captura de paquetes en los servidores y routers Linux utilizaremos el comando “tcpdump”. Para realizar los análisis utilizaremos el programa “Wireshark”, previamente instalado en las computadoras portátiles de los participantes.

Para familiarizarnos con el uso de las herramientas vamos a monitorear el tráfico en la interfaz eth2 del router Rx2, y a ejecutar un traceroute de Sx1 a Sx2.

- En el router Rx2:

```
labnicX:~$ router x2
Password:
entered into CT 112
[root@RX2 /]#
[root@RX2 /]# tcpdump -i eth2 -s 0 -w /captura/exerc02.pcap
tcpdump: listening on eth2, link-type EN10MB (Ethernet), capture size 65535 bytes
```

- En el servidor Sx1:

```
labnicX:~$ server x1
Password:
entered into CT 110
[root@SX1 /]# traceroute 172.2X.10.2
traceroute to 172.2X.10.2 (172.2X.10.2), 30 hops max, 46 byte packets
 1  172.2X.4.1 (172.2X.4.1)  3.027 ms  0.026 ms  0.025 ms
 2  172.2X.3.2 (172.2X.3.2)  0.642 ms  0.663 ms  0.651 ms
 3  172.2X.10.2 (172.2X.10.2)  1.851 ms  0.277 ms  0.275 ms
[root@SX1 /]#
```

- Nuevamente, en el router Rx2:

```
[root@RX2 /]# tcpdump -i eth2 -s 0 -w /captura/exerc02.pcap
tcpdump: listening on eth2, link-type EN10MB (Ethernet), capture size 65535 bytes
[CTRL + C]
127 packets captured
127 packets received by filter
0 packets dropped by kernel
[root@RX2 /]#
```

Aproximadamente en 1 minuto, como máximo, un script copiará este archivo a un directorio compartido vía web en nuestro servidor de administración. Aguarde unos instantes y, usando un navegador en su computadora portátil, acceda a la dirección http://[xxxx:xxxx:x:xxxx::xxx] (la misma que utilizó con ssh).

Abra el archivo en Wireshark. Aplique el filtro

ip.addr=="dirección de origen del traceroute", si lo desea, para facilitar la visualización, y responda las 2 preguntas siguientes:

1 – ¿Qué protocolo se utiliza para enviar los mensajes por el origen?

2 – ¿Cuántos paquetes se envían para cada valor de TTL?