

ANEXO B

ARQUITECTURA DE QoS EN 3GPP Y 3GPP2

Introducción de QoS E2E

EL concepto de QoS ha sido creado y definido con el fin de proveer diferente tratamiento a tráfico de diversos servicios y/o usuarios. QoS es un amplio concepto que puede ser desagregado en QoS provista en los diferentes dominios y niveles del sistema, pero que tiene significancia únicamente cuando es considerada en E2E, dado que al final, la QoS alcanzada está siempre determinada por el nivel más débil.

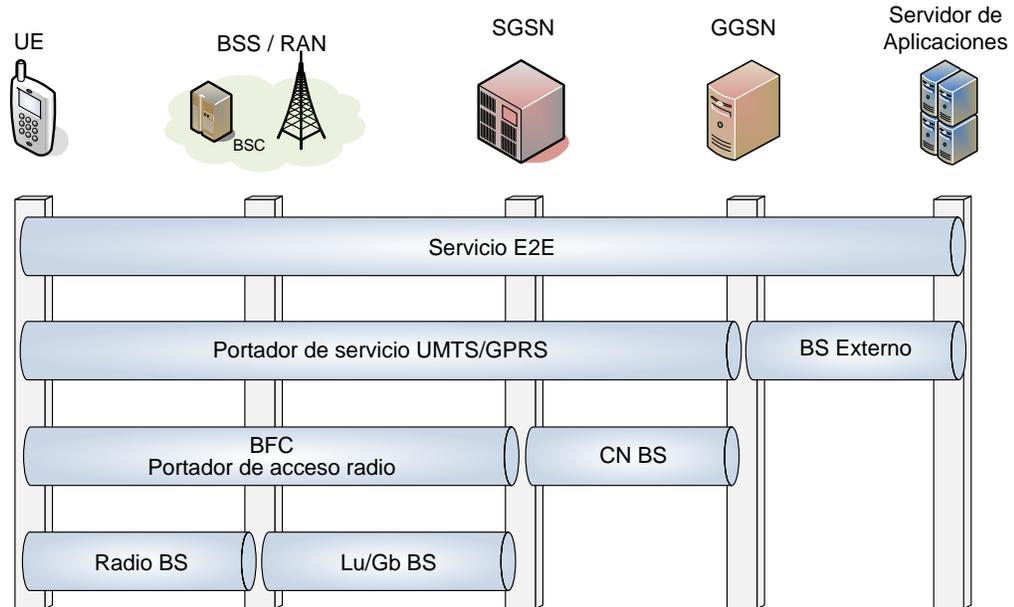
Específicamente, desde el punto de vista de la red, los mecanismos de QoS son responsables del manejo de tráfico y la diferenciación de servicios, por ejemplo, de acuerdo a ciertos requerimientos de calidad especificados por la aplicación, las funcionalidades de la red tratan cada servicio acorde a sus propios requerimientos. Sin embargo, la percepción de QoS desde el punto de vista del usuario final es un tema amplio y subjetivo, definido también como QoE (Quality of Experience, Experiencia de calidad). En cualquier caso, desde la perspectiva de la red celular, la mejor forma de tratar de asegurar un buen QoE es proveer un buen QoS E2E.

La arquitectura de administración de QoS definida en los diferentes grupos de Estandarización tales como el 3GPP y 3GPP2 han establecido la QoS E2E como una combinación de QoS provista por los diferentes niveles del subsistema. La aproximación común a esta definición se muestra en la Figura B1.

Como se ilustra en la Figura B2, la QoS E2E es lograda por medio del mapeo de los requerimientos E2E sobre una apropiada configuración a través de los diferentes mecanismos disponibles en la red. Por ejemplo, la RAN incluye mecanismos conocidos de QoS como: control de admisión, asignación de canal, programador de paquetes, control de potencia o control de handover, que puede diferenciar muchos servicios y/o usuarios. En el núcleo de red, existen muchos otros mecanismos como: el control de congestión, autorización de recursos, reservación de recursos para servicios de tiempo real, etc. Adicionalmente, el operador puede incluir métodos de QoS basados en IP dentro de su propia Intranet, como también un apropiado mapeo entre el móvil y las redes externas.

El servicio E2E es definido en la cima de los portadores UMTS/EDGE/GPRS y portadores externos. Los portadores UMTS/EDGE/GPRS siguen una filosofía Bottom-Up, estando definidos los portadores superiores sobre la capacidad suministrada por los portadores inferiores. Como por ejemplo, un portador UMTS está compuesto por portadores de acceso radio y portadores de núcleo de red; mientras el portador de acceso radio está compuesto de un portador radio y un portador Iu (Gb en versiones anteriores).

Figura B1. Arquitectura de QoS E2E



BS: Base Station, Estación Base.

BFC: Block Flow Context, Contexto de Flujo de Paquetes.

Lu: Interfaz encargada de conectar la UTRAN con el SGSN en una red UMTS.

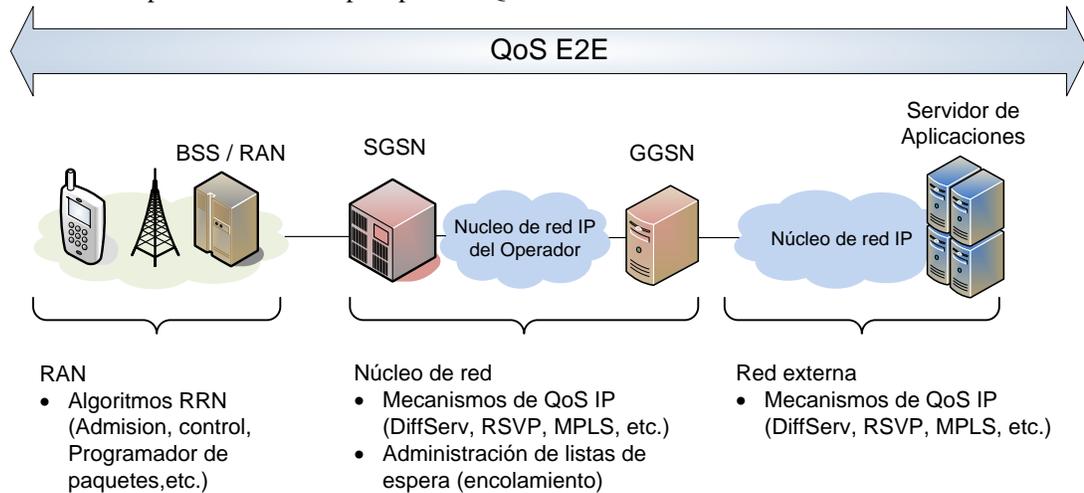
Cada uno de los portadores mostrados en la Figura B1 tienen sus propios mecanismos de QoS, por consiguiente la QoS provista puede ser también nivelada o clasificada por mecanismos de QoS para cada portador, siendo la QoS E2E el resultado de la combinación de los diferentes mecanismos de QoS a través de los diferentes sistemas portadores.

La arquitectura de QoS descrita en este anexo es parte de la revisión 6 de los estándares del 3GPP y 3GPP2. Sin embargo, esta estructura es el resultado de la evolución de las especificaciones, las cuales provienen de soluciones más primitivas, donde mecanismos menos elaborados fueron implementados y únicamente unos pocos portadores estaban disponibles. Actualmente, la mayor herencia de las existentes redes de 2G y 2.5G trabajan de forma fija en una aproximación de BE (Best Effort, Mejor Esfuerzo) o usan una muy básica diferenciación de QoS basada en prioridades.

EVOLUCIÓN DE QoS EN LOS RELEASES 3GPP

La especificación 3GPP apunta a proveer un marco de trabajo común para los diferentes estándares de telecomunicaciones. Estos estándares están en constante evolución de acuerdo a los requerimientos del mercado y acondicionamiento tecnológico. Esta evolución puede verse también en la arquitectura de las especificaciones 3GPP, las cuales ha evolucionado a través de los diferentes release de acuerdo a las necesidades de los nuevos servicios.

Figura B2. Principales mecanismos para proveer QoS E2E

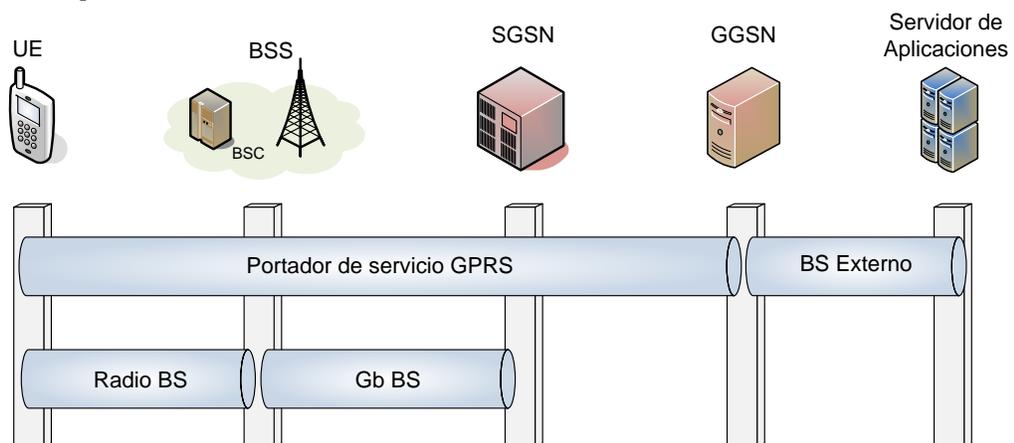


RRM: Radio Resource Management, Administrador de Recursos Radio.
 DiffServ: Differentiated services, Servicios Diferenciados.
 RSVP: Resource Reservation Setup Protocol, Protocolo de Reserva de Recursos.
 MPLS: Multi-Protocol Label Switching, Conmutación de Etiquetas Multiprotocolo.

Release 97/98

La introducción del servicio portador GPRS en el Rel'97 se entiende como el primer paso en la evolución del manejo de la arquitectura de QoS para las redes móviles. Los servicios portadores provistos por GPRS fueron basados en el concepto de contexto PDP, el cual es, en esencia, una conexión lógica establecida entre un MS y un GGSN para transportar todo el tráfico IP generado hacia y desde esa MS. La arquitectura de QoS en el Rel'97/98 se describe en la Figura B3.

Figura B3. Arquitectura de QoS en el Release 97/98



La arquitectura de QoS se basa en el contexto PDP para proveer diferenciación entre los paquetes de datos. Los parámetros de QoS son establecidos en base al contexto PDP, por lo

cual todo el tráfico que se envía por el mismo contexto PDP tiene la misma QoS, aún si el tráfico proviene de diferentes aplicaciones ejecutadas por el mismo usuario. En orden de lograr un tratamiento diferenciado de QoS para la misma MS, muchos contextos PDP pueden ser activados, entonces, cada uno de ellos puede tener su propio perfil de QoS. Sin embargo, en este release, el terminal recibe una nueva dirección PDP por contexto PDP, el cual es un claro factor limitante para servicios como los que requieren de un flujo continuo de datos (streaming), donde muchos contextos PDP con diferentes perfiles de QoS deben ser activados simultáneamente a la misma aplicación para transportar tráfico interactivo y de tiempo real.

Los atributos de QoS son negociados entre el terminal y la red durante el procedimiento de activación del contexto PDP. En el Rel'97/98 el perfil de QoS es definido de acuerdo a las siguientes clases de atributos: Precedence (Precedencia), Delay (Retardo), Reliability (Fiabilidad) y Throughput¹ (Velocidad Efectiva). Ver Tabla B1. El terminal es el único capaz de seleccionar los valores de los atributos de QoS de acuerdo a las necesidades de la aplicación. En caso de que cualquiera de los valores no sea especificado por el terminal, la red se encarga de seleccionar los valores por defecto en el perfil QoS.

En el Rel'97/98, únicamente son soportados servicios que no sean de tiempo real (non-real-time), como son e-mail, buscadores web, transferencia de archivos, MMS o servicios WAP.

Tabla B1: Atributos de QoS en el Rel. 97/98

Atributo QoS	Descripción
Precedence class (Clase Precedencia)	Hay 3 diferentes niveles de precedencia del servicio, los cuales indican la prioridad de mantener el servicio. Altos niveles de precedencia garantizan que el servicio esté delante de los otros niveles de precedencia. Niveles de precedencia normales garantizan que el servicio esté delante de los usuarios con bajas prioridades. Bajos niveles de precedencia reciben el servicio después de que los compromisos con las prioridades normal y alta han sido cumplidos.
Delay class (Clase Retardo)	El atributo delay se refiere al retraso de transferencia E2E a través del sistema GPRS. Hay 3 clases de retardo predictivo y una clase BE.
Reliability class (Clase Fiabilidad)	Hay 5 clases diferentes de fiabilidad. La fiabilidad de los datos es definida en términos de la RER ² (Residual Error Rate, Velocidad Residual de Error) para los siguientes casos: probabilidad de pérdida de los datos, probabilidad de datos entregados fuera de secuencia, probabilidad de duplicar datos entregados y probabilidad de datos corruptos.
Throughput	El Throughput de datos del usuario es especificado en términos de un

¹ Throughput: Se llama así al volumen de trabajo o de información que fluye a través de un sistema.

² RER: Es un parámetro que describe la exactitud del servicio, basado en la frecuencia de dañados, duplicados o perdidos SDUs entregados a la interface usuario-red.

class (Clase Velocidad Efectiva)	conjunto de clases de throughput, y distinguido por el ancho de banda esperado que requiere un contexto PDP. El throughput es definido por las clases de throughput: pico y promedio. El throughput pico especifica la velocidad máxima a la cual se espera que los datos sean transmitidos a través de la red para un contexto PDP individual. No hay garantía de que esa velocidad pico pueda ser alcanzada o sostenida por un periodo de tiempo; esto depende de las capacidades de la MS y de los recursos de radio disponibles. El throughput pico es independiente de las clases de retardo que determinan el retardo de tránsito por paquete en la red GPRS. El throughput medio especifica la velocidad promedio a la cual se espera transmitir los datos a través de la red GPRS durante el tiempo de vida restante del contexto PDP ya activado.
---	--

Release 99

En el Rel'99 se habilita el uso de muchos contextos PDP por dirección PDP, cada uno tiene un perfil de QoS de su propiedad. El primer contexto PDP activado para una dirección PDP es llamado primario o contexto por defecto. Los subsiguientes contextos PDP activados por la misma dirección son llamados contextos PDP secundarios mientras están conectados al mismo APN del contexto primario. De otra manera, un nuevo contexto PDP primario es activado para ese terminal asignándole un APN diferente.

La capacidad de establecimiento de muchos contextos PDP simultáneos con diferentes características de QoS utilizando solo una dirección IP es factible gracias a la introducción del TFT (Traffic Flow Template, Plantilla de Flujo de Tráfico). La TFT está basada en diferentes combinaciones de filtros como: fuentes de direcciones y máscaras de subred, rango de puertos destino y/o fuente, etc., los cuales son definidos por el terminal en el momento de la activación de un contexto PDP. La TFT es asignada a todos los contextos PDP secundarios con el fin de direccionar el tráfico entrante del DL al correcto contexto PDP en el GGSN, siendo el PDP primario el usado por defecto (en caso de que el tráfico entrante no coincida con cualquiera de los filtros existentes)

En adición, el Rel'99 trata de armonizar las estructuras de administración de QoS en la GERAN y UTRAN, por ejemplo, para alinear los portadores GPRS con el definido para UMTS y proveer los mismos parámetros y procedimientos de QoS. El Rel'99 introduce nuevos atributos, los cuales son descritos en la Tabla B2 y clases de QoS en UMTS, así: conversational, Streaming, Interactive y background. El principal factor de distinción entre estas clases es qué tanto es afectado el tráfico por los retardos y paquetes perdidos. La clase Conversational fue asignada para el tráfico altamente sensible a los retardos mientras que la clase Background fue encaminada al tráfico menos sensible a los retardos.

Las clases Conversational y Streaming fueron desarrolladas principalmente para transportar flujos de datos en tiempo real. Los servicios conversacionales en tiempo real, como VoIP, videoconferencia sobre IP o juegos en red de tiempo real, son aplicaciones altamente sensibles a los retardos, y son transportados en la clase conversacional. La clase Streaming

es también empleada en aplicaciones sensibles a los retardos, cadenas de video y audio, pero distintas a las aplicaciones conversacionales, pueden compensar algunas variaciones de retardos con el mecanismo de buffering en el cliente. La principal diferencia con las aplicaciones conversacionales es la naturaleza unidireccional de las cadenas, la cual disminuye la interactividad.

Tabla B2: Atributos de QoS en el Rel. 99

Atributo QoS	Descripción
Traffic Class	Tipo de aplicación para el cual el servicio RAB (Radio Access Bearer, Portador de Acceso Radio) es optimizado: ('conversational', 'Streaming', 'Interactive', o, 'Background').
Delivery Order	Indica si el portador proveerá en secuencia de entrega o no los SDU (Service Data Unit, Unidad de Datos de Servicio).
Máximo SDU size (octetos)	Define el máximo tamaño permitido del SDU.
SDU format Information (bits)	Lista de posibles tamaños exactos de SDU's. Si se utiliza protección de errores impar por un servicio RAB, la información del formato SDU define los formatos de subflujos exactos de la carga SDU.
Delivery of erroneous SDUs	Indica si SDUs con errores detectados serán entregados o no. En caso de protección de errores impar, el atributo es determinado a través de los subflujos.
Residual Bit Error Rate (BER)	Especifica la BER no detectada por cada subflujo en la entrega de SDUs. En caso de protección de errores par, únicamente es necesario un valor. Si no hay solicitud de detección de errores para un subflujo, la BER residual indica la BER en ese subflujo de la SDU entregada.
SDU Error Ratio	Define la relación de SDUs perdidas o detectadas como erróneas. La relación de error SDU es definida únicamente para la conformación de tráfico. En caso de protección de errores impar, la relación de errores SDU se establece a través de los subflujos y representa la relación de error en cada subflujo. La relación de error SDU se establece únicamente para los subflujos que se solicitó la detección de errores.
Transfer Delay	Indica el máximo retardo para el percentil 95 de la distribución de los retardos, para todos los SDUs entregados durante el tiempo de vida o servicio portador, donde el retardo para un SDU es definido como el tiempo desde la solicitud para transferir un SDU a un SAP (Service Access Point, Punto de acceso al Servicio) hasta su entregas a otro SAP.
Maximum Bit Rate (Kbps)	Máximo número de bits entregados a un SAP dentro de un periodo de tiempo, dividido por la duración del periodo.
Guaranteed bit Rate	Número garantizado de bits entregados a un SAP dentro de un periodo de tiempo (suponiendo que hay datos para entregar), dividido por la duración del periodo.

Traffic handling priority	Especifica la importancia relativa para el manejo de todas las SDUs pertenecientes al RAB, comparado con los SDUs de otros portadores.
Allocation/retention priority	Especifica la importancia relativa comparada con otro RAB para asignación y retención del RAB. El atributo de prioridad de asignación y retención es un parámetro de suscripción, el cual no es negociado desde el terminal móvil.

Las clases Interactive y Background son principalmente utilizadas por aplicaciones que no tienen las restricciones de los retardos para tiempo real, y que no son tolerantes a errores, como buscadores web o WAP, MMS o e-mail. Debido a los requerimientos permisivos de retardos, en comparación con las clases conversacionales, ambas proveen mejores velocidades de error por la forma de los esquemas de retransmisiones. La principal diferencia entre las clases Interactive y Background es que la clase Interactive es utilizada para obtener mejores tiempos de ida y vuelta para los servicios que requieren solicitudes/respuestas entre las dos partes, como los buscadores web y WAP. La clase Background es utilizada para servicios que tienen muy pocas exigencias en los requerimientos de retardos y se pueden satisfacer con esquemas básicos BE.

Las principales características y posibles aplicaciones de las diferentes clases de tráfico se muestran en la Figura B4.

Figura B4: Clases de QoS 3GPP en UMTS

	Clases de tráfico	Principales características	Aplicaciones Posibles
Tiempo Real	Conversational	<ul style="list-style-type: none"> • Riguroso y retardos bajos. • Reduce la variación de retardos. 	<ul style="list-style-type: none"> • Llamadas de voz. • Llamadas de videoconferencia.
	Streaming	<ul style="list-style-type: none"> • Reduce la variación de retardos. 	<ul style="list-style-type: none"> • Cadenas de audio. • Cadenas de video.
No Tiempo Real	Interactive	<ul style="list-style-type: none"> • Formato de solicitudes y respuestas. • Minimiza la velocidad de error de bit. 	<ul style="list-style-type: none"> • Buscadores Web. • Buscadores WAP.
	Background	<ul style="list-style-type: none"> • El destino no espera los datos dentro de un cierto tiempo. • Minimiza la velocidad de error de bit. 	<ul style="list-style-type: none"> • e-Mail. • Descargas de archivos.

Aparte de las clases de tráfico, existen otros atributos de QoS principalmente relacionados a los retardos, velocidad de bit y requerimientos de confiabilidad, que se describen en la Tabla 2.

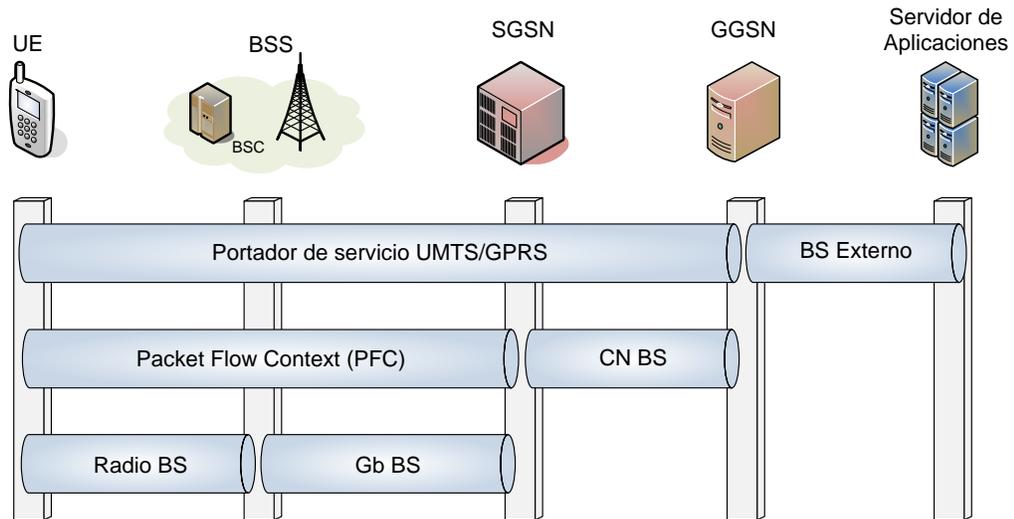
El Rel'99 también incluye algunas características para permitir un mejor control en la provisión de QoS en el dominio radio. Los requerimientos de QoS pueden ser enviados desde SGSN-2G al BSS por medio de un PFC (Packet Flow Context, Contexto de Flujo de Paquetes) al BSS. Un PFC BSS puede ser compartido por muchos contextos PDP activados con requerimientos similares de QoS. La transferencia de datos relacionada al contexto PDP que comparten el mismo PFC constituye un agregado flujo de paquetes, cuyos requerimientos de QoS son llamados perfiles agregados de QoS BSS. El perfil agregado de QoS BSS define los QoS que deben ser provistos por la BSS para un flujo de paquetes dado entre la MS y el SGSN, ej. Para la combinación de las interfaces Um y Gb.

La arquitectura QoS en Rel'99 es descrita en la Figura B5.

El perfil de QoS de una conexión en UMTS es el resultado de una negociación de QoS entre la QoS de UMTS solicitada por el terminal y la concedida por la red.

La negociación de QoS en UMTS ocurre típicamente en el establecimiento del contexto PDP, aunque puede forzarse a ser renegociado por medio de procedimientos de modificación del contexto PDP.

Figura B5: Arquitectura de QoS en el Rel 99



Durante el proceso de activación del contexto PDP, el UE solicita al CN el acceso a una red IP externa, identificada por un APN, e incluye un perfil de QoS UMTS para el contexto PDP. Antes de que el CN responda positivamente, necesita:

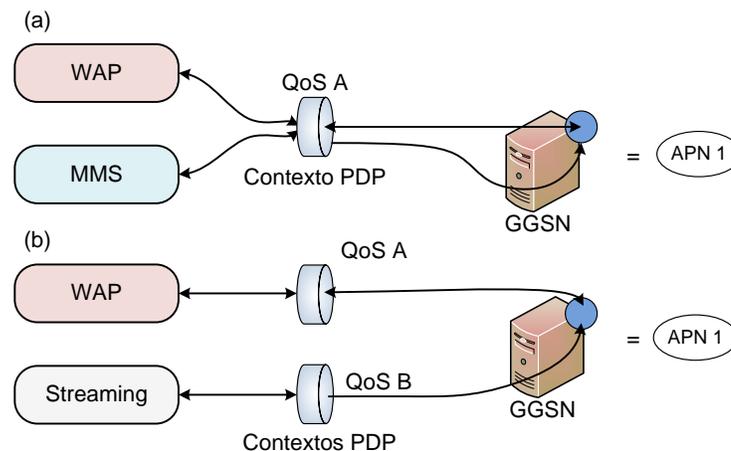
- Validar la solicitud del perfil de QoS por el APN con la suscripción en el HLR. En el HLR, la información de suscripción de usuario es almacenada en asociaciones donde por cada APN autorizado, es indicado el máximo perfil de QoS que puede ser concedido. Si la solicitud del perfil QoS es mayor que la almacenada en el HLR, entonces la solicitud de QoS no puede ser concedida como tal.

- Ejecutar control de admisión y reservación de recursos en el CN
- Iniciar el establecimiento de portadores y comprobación de QoS

Después de obtener resultados positivos de portadores inferiores (RAB y CN), el contexto PDP es establecido satisfactoriamente. Sin embargo es posible que la solicitud de QoS UMTS y la QoS UMTS concedida no sean las mismas, debido a las limitaciones introducidas por el control de admisión en los diferentes elementos de red y portadores inferiores.

Es importante notar que un contexto PDP ha sido asociado a un perfil de QoS específico, pero puede transportar datos de dos o más aplicaciones diferentes. Por consiguiente esas aplicaciones experimentarán la misma QoS, como se ilustra en la Figura B6(a). Idealmente, para dos aplicaciones que tienen muchas diferencias en los requerimientos de QoS, se recomienda usar diferentes contextos PDP, aún si están conectadas al mismo APN. Con esta aproximación, diferentes perfiles de QoS por contexto PDP pueden ser aplicados. Esto es representado en la Figura B6(b).

Figura B6: Relación entre contextos PDP, APN y perfil de QoS. (a) Un contexto PDP para un APN 1 con QoS A para dos diferentes aplicaciones. (b) Dos contextos PDP para un APN 1 con diferentes perfiles de QoS para dos diferentes aplicaciones.



Sin embargo, no es posible el escenario donde muchas aplicaciones están compartiendo el mismo contexto PDP pero encaminadas a diferentes APNs.

No todos los terminales tienen la posibilidad de establecer múltiples contextos PDP, porque la complejidad impuesta es muy alta en el lado del terminal. Esta es la razón por la que hay esfuerzos en investigaciones para permitir la utilización de un único APN para diferentes aplicaciones, pero manteniendo la provisión de diferenciación de QoS. Todos esos estudios apuntan a la instalación de funcionalidades en el GGSN para analizar el tráfico del plano de usuario y tratarlo de forma diferente dependiendo de la aplicación de origen.

Release 5

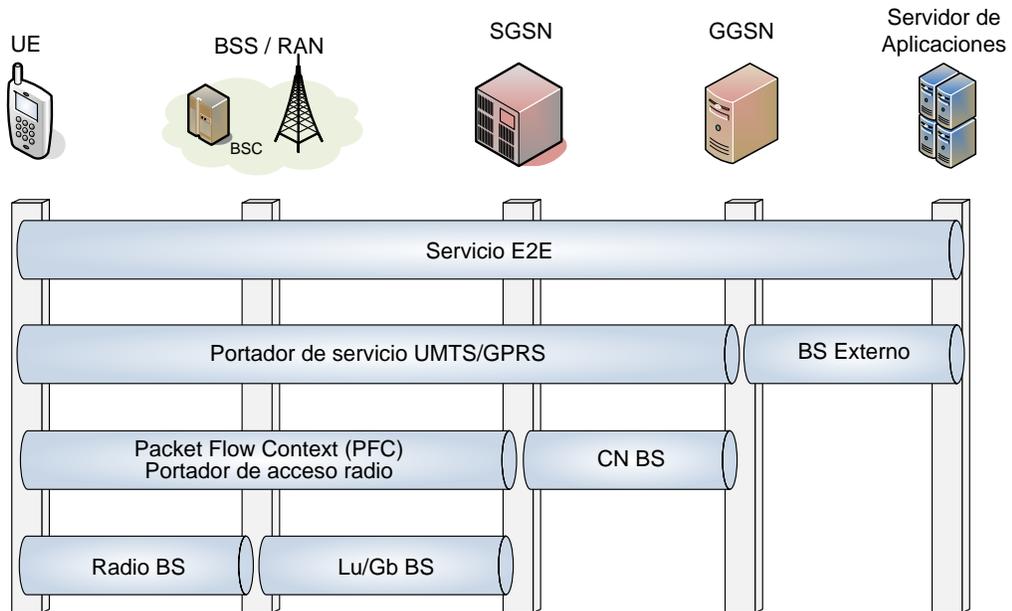
El Rel'5 permite la negociación de los requerimientos de QoS en el nivel de aplicación, desde una perspectiva E2E. Con Rel'5 se introducen los servicios IMS y es estandarizada la autorización de QoS por portadores. Esta autorización provee una nueva forma de activar un contexto PDP secundario con atributos de QoS específicos por medio de una interacción entre la MS y el GGSN.

La arquitectura de QoS en el Rel'5 es descrita en la Figura B7. Un asunto importante en la negociación de QoS (que también es aplicable a los anteriores release) es la ejecución de un correcto mapeo de QoS desde los servicios portadores de niveles altos a los bajos. De forma que los requerimientos de QoS de niveles altos son negociados en el nivel de aplicación, requiriendo un apropiado mapeo en los perfiles de QoS del contexto PDP así como también atributos externos de servicios portadores. De la misma manera, perfiles de QoS deberán ser mapeados sobre atributos QoS RAB (o PFC para GPRS) y atributos portadores en el CN.

El RAB en UMTS o su equivalente PFC en GPRS, permite un mejor control de QoS en la RAN. Los atributos de QoS para un RAB son comunes, pero los valores que son solicitados serán adaptados a las capacidades disponibles de la tecnología radio.

Los atributos de QoS UMTS, vistos como los atributos globales, usualmente conservan el mismo valor a cuando se mapea sobre atributos QoS RAB, sin embargo, hay tres de ellos que usualmente se reducen: BER residual, porcentaje de error SDU y retardo de transferencia. La razón de esto es que servicios portadores CN pueden introducir errores y retardos, los cuales no serán contados en el nivel RAB.

Figura B7: Arquitectura de QoS en el Rel 5



Por ejemplo, si se considera el caso donde el atributo QoS UMTS 'delay transfer' es 350ms. Este retardo puede ser dividido en 250ms requeridos para el lado de la RAN y 100ms para el CN. De manera que transferirá al atributo QoS RAB 'transfer Delay' el valor 250ms.

El portador CN establecido entre SGSN y GGSN, se basa típicamente en el transporte basado en IP o ATM. Para el núcleo de red basado en IP, es obligatorio usar DiffServ. El mapeo entre clases QoS UMTS y DSCP son controladas por las decisiones de los operadores.

Release 6

En el Rel'6, la P-CSCF, que es un elemento de la arquitectura IMS, se convierte en AF (Application Function, Función de Aplicaciones). Similarmente a la P-CSCF, el AF ofrece servicios que requieren el control de recursos portadores IP, mapas QoS-relacionados con parámetros en el nivel de aplicación, (ej. SDP) sobre políticas de configuración de información, y envía esta información al PDF (Policy Decision Function, Función de Decisión de Política). En el caso particular de IMS, el AF es la P-CSCF la cual está en el mismo dominio que el GGSN. Se utiliza una nueva interfaz estandarizada Gq, especificada en 3GPP TS 23.002, para los servicios basados en políticas de configuración de información intercambiada entre el PDF y el AF. Esta información es utilizada por el PDF para servicios basados en decisiones de políticas locales.

La principal diferencia entre el AF y el P-CSCF es que el AF puede estar en el mismo dominio del PDF o en uno diferente, mientras que P-CSCF está siempre en el mismo dominio del PDF, que está a su vez en el mismo dominio del GGSN

3GPP VS 3GPP2 en QoS

Una vez descritos los mecanismos que son especificados en el 3GPP, se dará una vista a los soportes existentes para la provisión de QoS dados por 3GPP2, el cual es el cuerpo de estandarización encargado de CDMA2000.

Actualmente, la arquitectura de red CDMA2000 muestra muchos inconvenientes que se pueden evitar soportando múltiples clases de tráfico en la RAN, como se hace en el 3GPP. Sin embargo, algunas modificaciones de arquitectura están bajo estudio para poder proveer QoS E2E en un futuro y alcanzar el objetivo de armonizar las arquitecturas de red IMS 3GPP/3GPP2

Para ilustrar como esos inconvenientes surgen en la red CDMA2000, primero se describe su arquitectura y los servicios básicos de paquetes de datos prestados por su red. La Figura B8 muestra la arquitectura de red actual en CDMA2000 sin soporte de QoS, sus componentes y el stack de protocolos para cada uno de ellas.

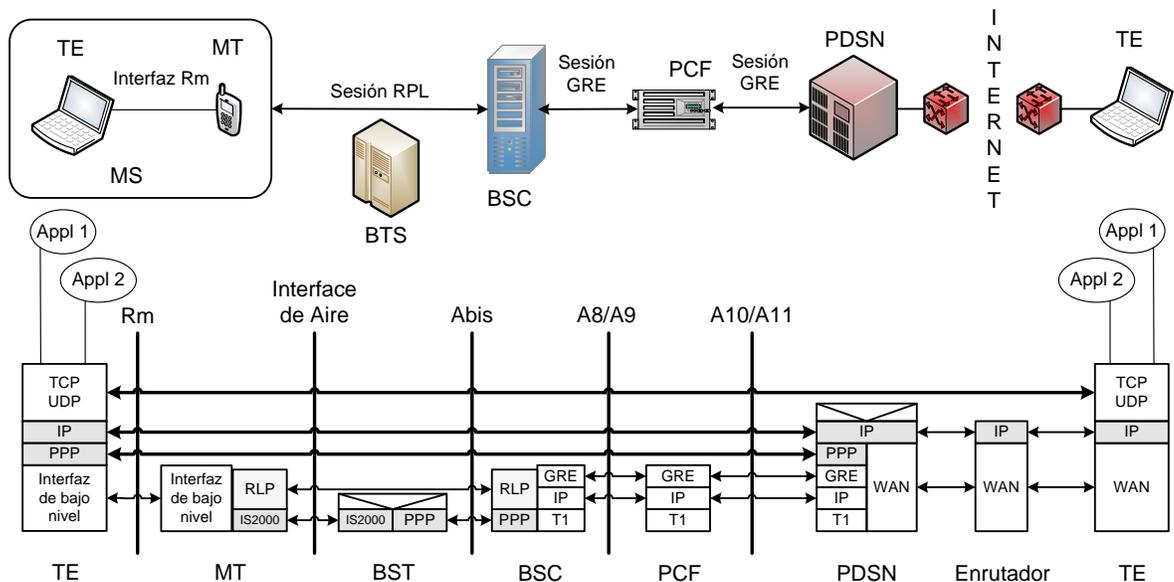
La MS consiste de dos componentes: TE como por ejemplo un computador portátil, y el MT (Mobile Terminal) como en tarjeta PCMCIA CDMA2000. La BS CDMA2000, consiste del BTS y el BSC y es conectada a la MS a través de la interfaz aire o radio. La

BS junto con la PCF (Packet Control Function, Función de Control de Paquetes) representan la RAN que conecta la MS a los PDSN (Packet Data Serving Node, Nodo Servidor de Paquetes de Datos) que trabaja como pasarela (gateway) para las redes IP.

Para establecer una conexión de datos hacia Internet, es necesario configurar primero una conexión PPP (Point-to-Point Protocol, Protocolo Punto a Punto) entre el TE y el PDSN. Los paquetes de datos son transportados desde el TE al PDSN en forma de tramas PPP; en la dirección opuesta, paquetes desde Internet recibidos en el PDSN hacia el TE son transportados sobre la misma conexión PPP. Un asunto clave de esta forma de operación es que el tráfico para todas las aplicaciones ejecutadas en el TE van a través de la misma conexión PPP entre el TE y el PDSN.

Este modo de operación significa una barrera para la provisión de QoS. Una vez los paquetes de datos han sido encapsulados en tramas PPP es imposible hacer distinción de esos paquetes y manejarlos dependiendo de las necesidades de QoS de la aplicación. Las tramas PPP recibidas en el MT (en el UL, o dirección contraria) y en el BSC (en el DL) son transportados sobre una sesión RLP (Radio Link Protocol, Protocolo Enlace de Radio) al BSC y MT respectivamente. Hay una sesión RLP por sesión PPP. Los datos enviados a través del RLP en cualquier dirección son segmentados en tramas RLP antes de ser transmitidos a través de la interfaz aire o radio.

Figura B8: Arquitectura CDMA2000



La principal funcionalidad del PFC es direccionar o enrutar las solicitudes de conexión PPP desde el TE a un apropiado PDSN que maneje la conexión. Para transportar el tráfico entre el BSC y la PCF se utiliza un túnel GRE (Generic Routing Encapsulation, Encapsulación de Enrutamiento Genérico) por sesión PPP. Las tramas GRE entre el BSC y el PCF son normalmente transportadas sobre una red privada, por ejemplo una red IP. Los datos entre

el PCF y el PDSN son transportados sobre otro conjunto de túneles GRE, nuevamente uno por sesión PPP. La red que conecta la PCF con el PDSN puede ser la red pública IP.

Para proveer QoS E2E, algunos parámetros de QoS deben ser agregados al nivel de aplicación y mapeados hacia abajo sobre los niveles inferiores, ej. Niveles de enlace y de red. Aunque la negociación de parámetros de QoS en el nivel de aplicación puede ejecutarse sobre la arquitectura de red existente mediante la utilización de SIP y PDP, los niveles actuales de enlace y red no soportan múltiples sesiones RLP y túneles GRE para transportar paquetes de diferentes servicios en una única sesión PPP, ej. No hay forma de convertir los parámetros del nivel de aplicación en parámetros de nivel de red y enlace solo porque no existen. Este es el porqué de la necesidad de algunas modificaciones en la arquitectura con el fin de implementar mecanismos de QoS en los niveles inferiores.

Para aclarar los inconvenientes de CDMA2000, se considera el caso de un TE enviando paquetes de datos al PDSN. Cuando un TE envía tramas PPP al otro extremo, el podría ejecutar algunos tipos de programación y priorización en los caminos de paquetes que vienen de las aplicaciones con necesidades de QoS acordadas. Este proceso se puede llevar a cabo por la clasificación de paquetes, de acuerdo a información TCP o IP, por ejemplo puerto fuente TCP. Pero una vez las tramas PPP son convertidas a cadenas de byte y enviadas sobre la interfaz Rm y luego como tramas RLP al BSC, esas tramas son entregadas en secuencia al BSC. Esto es asegurado por el número de secuencia de las tramas RLP. Por otro lado, cuando las cadenas de byte son convertidas dentro de tramas GRE y enviadas sobre los túneles GRE entre el BSC y la PCF y luego entre la PCF y el PDSN, dichas tramas GRE son también entregadas en secuencia a los puntos de entrega. Entonces, no hay forma de manejar apropiadamente el tráfico en tiempo real y de no tiempo real que es parte de la misma trama RLP. Como cada sesión RLP tiene parámetros como número de retransmisiones que aplican a todas las tramas RLP en la sesión, las retransmisiones son ejecutadas para tráfico de tiempo real aunque ello no tenga sentido.

Finalmente, por las entregas ordenadas de RLP, si una trama RLP que lleva tráfico de no tiempo real se pierde, las tramas RLP de tiempo real que siguen a las tramas perdidas no se entregarán a los niveles superiores, aún si estos se reciben en el BSC, ej. La entrega de esas tramas RLP de tiempo real pueden pasar sólo después de que la trama RLP es retransmitida y posteriormente se recibe en el BSC.

Como se dijo anteriormente, en el caso de un TE enviando datos al PDSN, las tramas RLP que son recibidas en la BSC son encapsuladas en tramas GRE y enviadas a la PCF. Debido a que las tramas GRE son también entregadas en secuencia, significa que no se puede proveer diferenciación de servicio dentro de una red IP que transporta tráfico entre el BSC y la PCF. Lo mismo aplica con la red IP que interconecta la PCF y el PDSN, porque otro túnel GRE ha sido utilizado.

Para resumir, no se puede ejecutar diferenciación de QoS si todos los paquetes de datos para y desde un TE son mapeados sobre un sesión RLP y luego sobre una sesión GRE.

El uso de múltiples sesiones RLP y múltiples túneles GRE cambia esencialmente el transporte de tráfico desde y hacia un TE, permitiendo la diferenciación de servicio. Actualmente, muchos estándares dan soporte a los requerimientos de QoS E2E. Estos nuevos estándares introducen un nuevo concepto, la instancia de servicio. Este nuevo estándar propone que diferentes tramas PPP son enviadas a una sesión PPP, cada una transportando diferentes clases de servicio, sobre diferentes instancias de servicios entre el MT y el BSC. Cada instancia de servicio es implementada para una sesión RLP separada. En este nuevo estándar, también se añade que entre el BSC y la PCF y entre la PCF y el PDSN, hay múltiples sesiones GRE por TE, cada uno de los cuales transporta una clase diferente de tráfico.

Una vez que cada una de las clases de tráfico se transporta en separadas sesiones RLP, las sesiones RLP pueden ser utilizadas para priorizar las clases de servicio entre el MT y el BSC. Similarmente, diferentes sesiones GRE son utilizadas para manejar diferentes clases de tráfico y servicios.

El soporte de QoS en las redes inalámbricas IP CDMA2000 puede ser provisto por medio de una o más instancias de un servicio de paquetes de datos. Los recursos radio pueden ser asignados por instancias de servicio y una instancia de servicio puede solicitar unos atributos específicos de QoS en el enlace radio. Los tipos de esas instancias se definen como una instancia principal de servicio o una instancia auxiliar de servicio, recordando el concepto del 3GPP de contexto primario PDP y contexto secundario PDP. Como ejemplo de uso, se puede decir que el propósito de una instancia principal de servicio es la provisión de recursos en la red inalámbrica IP CDMA2000 para lograr los requerimientos QoS de las aplicaciones que pueden requerir únicamente soporte de QoS BE. Sin embargo, para lograr las peticiones de QoS de aplicaciones que requieren más que QoS BE, se puede utilizar una instancia auxiliar de servicio.

La asignación de recursos para una instancia auxiliar de servicios es selectiva y basada en la caracterización de requerimientos de QoS asociados con una aplicación. Una o más instancias de servicio auxiliares pueden ser establecidas por la MS basándose en el número de aplicaciones utilizadas por una MS cada una requiriendo diferente QoS.

Por un lado, la instancia principal de servicio es una instancia de servicio de paquetes de datos que se configura durante el establecimiento inicial de un servicio de paquetes de datos. Esta instancia de servicio de paquetes de datos normalmente tiene características por defecto ya establecidas, que son basadas en el perfil del suscriptor y en políticas internas. Por otro lado, la instancia auxiliar de servicio es una instancia de servicio de paquetes de datos que se configura bajo petición para soportar una mejor QoS que las características de QoS por defecto configuradas por la instancia principal de servicio. Esta instancia de servicio de paquetes de datos tiene características de QoS que son basadas en las solicitudes de los usuarios y limitadas por el perfil de QoS de usuario y por las políticas internas.

REFERENCIAS

- G. Gómez, R. Sánchez, End-to-End Quality of Service over Cellular Networks - Data Services Performance and Optimization in 2G3G. John - Wiley & Sons, 2005.