

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS QUÍMICAS E INGENIERÍA
Programa de posgrado de Maestría y Doctorado en Ciencias e
Ingeniería (MYDCI)

DOCTORADO EN INGENIERIA



“MODELO DE RED DE SIGUIENTE GENERACIÓN (NGN) APLICADO AL ENTORNO RURAL DE MÉXICO”

T E S I S

“Que presenta para obtener el grado de DOCTOR EN INGENIERIA”

ALBERTO COLÍN GONZÁLEZ

DIRECTOR DE TESIS: Dr. Arturo Serrano Santoyo

CO-DIRECTOR DE TESIS: Dr. Eduardo Álvarez Guzmán)

Tijuana, B. C.

Febrero del 2015

Dedicado a mí Creador

“Oh mi Dios infinito, le agradezco con toda el alma el haber estado conmigo todo este tiempo, por haberme dado fuerza y entendimiento, por librarme de las enfermedades mortales, por ponerme a las personas indicadas en mi camino, por cuidar a mi familia y a un servidor en los momentos más difíciles, y por hacer que hoy pueda cumplir uno de mis más grandes sueños, es por esta razón que este trabajo especialmente se lo dedico humildemente a Usted Padre bondadoso”

*A Andrew, Alberto, Alessa y
Alexander las razones de esta tesis y
de toda mi vida.*

RESUMEN de la Tesis de Alberto Colín González, presentada como requisito parcial para la obtención del grado de DOCTOR EN INGENIERIA. Tijuana, Baja California, México. Febrero de 2015.

“A MODEL FOR NEXT GENERATION NETWORKS (NGN) APPLIED TO RURAL COMMUNITIES OF MEXICO”

Resumen aprobado por:

Dr. Arturo Serrano Santoyo

Director de tesis

Dr. Eduardo Álvarez Guzmán

Co-Director de tesis

RESUMEN

El objetivo principal de esta tesis es el diseño de un modelo de red rural de siguiente generación con capacidad de integración de las principales funciones de las redes de acceso móviles de tercera y cuarta generación. Esta nueva arquitectura de red permite interconectar a las diferentes redes inalámbricas fijas y móviles, simplificando su implementación y reduciendo los costos asociados a inversión y operación. El modelo propuesto incluye una arquitectura abierta constituida de las siguientes entidades: La red troncal, el subsistema IMS (el corazón del modelo), las interfaces para interconectar redes de acceso inalámbricas de tercera y cuarta generación, y la red que constituye a los servidores de aplicación. Los aspectos más relevantes que se consideraron en el diseño son: la calidad del servicio, el control de las pérdidas de la red, la movilidad (*handover*) y las tasas de tráfico a operar.

Las aportaciones principales de esta tesis, dentro del objetivo general, se pueden resumir de la siguiente manera. En primer lugar, se aporta un modelo nuevo de red de siguiente generación para las zonas rurales y urbanas marginadas de México, con capacidad para proveer servicios multimedia (voz, datos y video). La segunda aportación, es una arquitectura para una central telefónica basada en IPv6 móvil, denominada IMS-AGC, que incluye la definición de todas las entidades de red que la componen, así como sus interfaces y protocolos, que permiten ofrecer los servicios de acceso a la red, movilidad de usuario (*handover*), transferencia de datos, acceso a servidores de aplicación, sincronización y localización de usuarios. La arquitectura propuesta se describe a través

de diagramas de bloques, secuencia de mensajes SIP, y se valida con los cálculos de los parámetros más importantes, y el análisis del comportamiento dinámico de las entidades principales y sus funciones proporcionadas por las redes de acceso.

La red propuesta, se evaluó utilizando un modelo matemático robusto y funcional basado en el protocolo SIP del Grupo de Trabajo de Ingeniería de Internet (IETF por sus siglas en Inglés), para las funciones del subsistema multimedia, y se utilizó la herramienta de software MatLab, con la cual se calcularon múltiples parámetros tales como: el ancho de banda, y la calidad de servicio, que permitan comunicaciones claras y efectivas entre usuarios. También, con la finalidad de eliminar retardos se utilizó otro modelo matemático que localiza servidores saturados, sustituyéndolos por otros de mayor capacidad, evitando con esto los cuellos de botella que generan retardos y que afecten la calidad de los servicios ofrecidos.

Los resultados obtenidos cumplen con el objetivo de la investigación realizada, y se contribuye además a la planeación del despliegue de infraestructura en telecomunicaciones en zonas rurales en el marco del proyecto México Conectado.

Palabras clave: sip; ims; ngn; e-utran; red rural; lte_advanced; 3gpp; Mexico Conectado

ABSTRACT of the thesis, presented by Alberto Colin Gonzalez, in order to obtain the DOCTOR IN ENGINEERING DEGREE in ELECTRONICS ENGINEERING. Tijuana, Baja California, Mexico. December, 2014.

“MODEL FOR NEXT GENERATION NETWORK (NGN) APPLIED TO RURAL COMMUNITIES OF MEXICO”

Approved by:

Dr. Arturo Serrano Santoyo
Thesis advisor

Dr. Eduardo Alvarez Guzman
Co-thesis advisor

ABSTRACT

The objective of this thesis is to develop a model for next generation rural networks capable of integrating the major functions of mobile access networks of third and fourth generation. This new network architecture allows interfacing to different fixed and mobile wireless networks, simplifying its implementation and reducing costs associated with investment and operation. The proposed model includes an open architecture consisting of the following entities: The backbone network, the IMS subsystem (the core of the model), interfaces to interconnect wireless access networks of third and fourth generation, and the network that constitutes the servers application. The most relevant aspects considered in the design are: The quality of service, the control of network losses, mobility (handover) and traffic data rates.

The main contributions of this thesis can be summarized as follows: First, a new model for next generation networks for underserved rural and urban areas of Mexico is presented. The second contribution is the proposal of architecture for a mobile telephone system based on IPv6, called IMS-AGC, which includes the definition of all the network entities, their interfaces and protocols to provide network access services, user mobility, data transfer, access to application servers, timing and users' location. The proposed architecture is described using block diagrams and the sequence of SIP messages; it is validated with calculations of the most important parameters and analysis of the dynamic behavior of the main entities and functions provided by the access network.

The proposed network architecture was evaluated using a robust and functional mathematical model based on the SIP protocol for the multimedia subsystem operation. MatLab software was employed to calculate the bandwidth and quality of service that allow clear and effective communications between users. Additionally, another

mathematical model was developed to eliminate delays in server saturation conditions affecting the offered quality of service.

The results fully comply with the objective of the proposed research, and also contribute to the process of planning and deployment of telecommunications infrastructure in rural areas under the Mexico Conectado Program.

Keywords: sip; ims; ngn; e-utran; red rural; lte_advanced; 3gpp; Mexico Conectado

Agradecimientos

Deseo agradecer a todas aquellas personas que con su ayuda invaluable me ha permitido materializar esta tesis, en especial a:

Mis padres †Irene y †Pedro por su excelente apoyo espiritual durante su vida, el cual me dio fortaleza para concluir mi proyecto de vida, descansen en paz.

A Graciela mi amada esposa que con su apoyo incondicional fue posible concluir este trabajo.

A mis queridos hijos Jonatan y Daniela por su apoyo y sobre todo por su comprensión en momentos difíciles.

A mis directores de tesis los Dres. Arturo Serrano Santoyo y Eduardo Álvarez Guzmán los cuales sin sus conocimientos, ayuda y apoyo en la realización de este trabajo hubiera sino imposible. Conocí de su calidad humana y profesional, la cual me impulso y me dio fuerzas para poder concluir el proyecto de tesis a pesar de las adversidades.

Al Dr. Wagdy Anis investigador del Cairo, Egipto, que sin conocerme me ofreció su ayuda incondicional en los trabajos de investigación de esta tesis. Dios bendiga al Dr. Aziz.

Al personal de la FCQI de la UABC Campus Tijuana, particularmente al Dr. Luis E. Palafox Maestre y al Dr. José M. Cornejo, por su apoyo incondicional, a la Dra. Edith García por su apoyo moral y consejos que fueron de gran ayuda, a la Sra. Patricia Carpio por su gran labor que realiza apoyando a todos los estudiantes de posgrado y a todas y cada una de las personas que de alguna manera formaron parte de este proyecto.

ÍNDICE

página
1

1. Introducción

- 1.1 Objetivos de la tesis doctoral.
- 1.2 Entorno de esta tesis.
- 1.3 Estructura del documento.

2. Marco Teórico

- 2.1 Antecedentes.
- 2.2 Evolución de la tecnología
- 2.3 Avances tecnológicos que limitan la solución del problema.
 - 2.3.1 Introducción a las arquitecturas GSM y GPRS.
 - 2.3.2 Arquitectura UMTS y limitaciones
- 2.4 Propuesta de solución.
- 2.5 Red LTE Advanced del grupo 3GPP
 - 2.5.1 Arquitectura de red del sistema LTE
 - 2.5.1.1 Introducción
 - 2.5.1.2 Arquitectura de los sistema celulares
 - 2.5.1.3 Arquitectura de red de los sistemas 3GPP
 - 2.5.1.4 Arquitectura de la red LTE
 - 2.5.1.5 Red de acceso evolucionada: E-UTRAN
 - 2.5.1.5.1 Arquitectura de E-UTRAN
 - 2.5.1.5.2 Unidades de red e interfaces
 - 2.5.1.5.2.1 Nodo B evolucionado (eNB)
 - 2.5.1.5.2.2 Interfaz de radio
 - 2.5.1.5.2.3 Interfaz eNB↔EPC (S1)
 - 2.5.1.5.2.4 Interfaz eNB↔eNB (X2)
 - 2.5.1.6 Comparación entre las arquitecturas E-UTRAN y UTRAN
 - 2.5.1.7 Red Troncal de paquetes evolucionada: (EPC)
 - 2.5.1.7.1 Arquitectura de EPC
 - 2.5.1.8 Unidades de red y sus interfaces
 - 2.5.1.8.1 Unidades de administración de movilidad (MME)
 - 2.5.1.8.2 Pasarela de servicio (S-GW)
 - 2.5.1.8.3 Pasarela PDN (P-GW)
 - 2.5.1.8.4 Servidor suscriptor de usuarios (HSS)
 - 2.5.1.8.5 Interfaz P-GW↔ redes externas (SGi)
 - 2.5.1.8.6 Interfaz P-GW↔S-GW (S5 y S8)
 - 2.5.1.8.7 Interfaz MME↔S-GW (S11)
 - 2.5.1.8.8 Interfaz MME↔MME (S10)
 - 2.5.1.8.9 Interfaz HSS↔MME (S6a)
 - 2.5.1.9 Configuraciones de la red EPC
 - 2.5.1.10 Soporte de itinerancia
 - 2.5.1.9 Configuraciones de la red EPC
 - 2.5.1.10 Soporte de itinerancia
- 2.6 Subsistema Multimedia (IMS)
- 2.7 Flujo de tráfico de datos y dimensionamiento de redes IMS
- 2.8 Interconectividad de LTE con redes heterogéneas

- 2.8.1 Generalidades.
- 2.8.2 Fundamento de soluciones de interconexión
 - 2.8.2.1 Escenario de Referencia
- 2.8.3 Arquitectura genérica de las redes de acceso
- 2.8.4 Terminales o teléfonos multimodal utilizados en la red
- 2.8.5 Niveles de interconexión
- 2.8.6 Mecanismos de interconexión
 - 2.8.6.1 Mecanismos AAA entre redes
 - 2.8.6.2 Mecanismos para la transferencia de datos entre redes
 - 2.8.6.3 Mecanismos de movilidad en capa de red
- 2.9 Interconexión entre LTE y otras redes 3GPP
 - 2.9.1 Marco general
 - 2.9.2 Interconexión mediante interfaces propias del sistema GPRS.
 - 2.9.3 Interconexión de GPRS y EPC utilizando nuevas interfaces
 - 2.9.4 Movilidad entre redes de acceso
 - 2.9.4.1 Movilidad de teléfonos en estado desconectado
 - 2.9.4.2 Movilidad entre redes de acceso 3GPP en estado activo
 - 2.9.5 Solución de interconexión para los servicios de voz
- 2.10 Interconexión entre LTE y otras redes no-3GPP.
- 2.10.1 Descripción general
- 2.11 Arquitectura de red rural NGN propuesta
- 2.12 Modelo matemático de la red NGN para el entorno rural de México
 - 2.12.1 Introducción
 - 2.12.2 Arquitectura IMS
 - 2.12.3 Protocolo SIP en subsistema IMS
 - 2.12.3.1 Retransmisiones para invitación a conectarse (*Invite*).
 - 2.12.3.2 Retransmisiones *no-Invite*.
 - 2.12.3.3 Fases de establecimiento de una sesión IMS
 - 2.12.3.4 Modelado del establecimiento de la sesión IMS
 - 2.12.3.5 Servidores de red matriz y de red visitada
 - 2.12.3.6 Modelado de la retransmisión de mensajes SIP
 - 2.12.3.6.1 Modelado de la retransmisión de la solicitud *Invite*
 - 2.12.3.6.2 Pérdidas durante la fase *Invite*
 - 2.12.3.6.3 Modelado de retransmisión de solicitudes *no-Invite*
 - 2.12.3.6.4 Pérdidas durante la fase *no-Invite*
 - 2.12.3.7 Consumo de ancho de banda del establecimiento de sesión IMS en una red con pérdidas.
 - 2.12.3.7.1 Intentando fase *Invite*/100 (*Invite*)
 - 2.12.3.7.2 Progreso de sesión fase 183/PRACK/200 OK (*Invite*)
 - 2.12.3.7.3 Fase de negociación UPDATE/200
 - 2.12.3.7.4 Fase de timbrado 180/PRACK/200 OK
 - 2.12.3.7.5 Fase de respuesta final

200 OK/ACK (*no-Invite*).

2.12.3.7.6 Ancho de banda total.

2.12.4 Estimación del retardo en la puesta de sesión IMS.

2.12.4.1 Retardo de la fase *no-Invite*.

2.12.4.2 Retardo de la respuesta provisional
(183, 180).

2.12.4.3 Retardo total.

2.13 Limitaciones del modelo de red rural.

2.13.1 Limitaciones de la tasa de tráfico.

3. Marco de Simulación

3.1 Experimentación numérica del modelo de red NGN.

3.1.1 Cálculo del ancho de banda.

3.1.2 Cálculo de flujo del tráfico y dimensionamiento
de la red.

4. Resultados

4.1 Análisis de resultados

5. Conclusiones

5.1 Introducción

5.2 Contribuciones de la Tesis

5.2.1 Publicación de Resultados

5.3 Trabajo Futuro

Referencias

Apéndices

Acrónimos

Índice de figuras

Figura

Página

- 1.1.- Modelo de arquitectura de red NGN propuesta.
- 2.1.- Arquitectura GSM con módulo GPRS agregado.
- 2.2.- Dominios de redes UTRAN.
- 2.3.- Arquitectura de red UTRAN.
- 2.4.- Arquitectura principal de 3GPP *release99*.
- 2.5.- Arquitectura de la red 3GPP *release 4*.
- 2.6.- Arquitectura de la red 3GPP *release 5*.
- 2.7.- Arquitectura lógica de la red del grupo 3GPP *release R4/5*.
- 2.8.- Arquitectura genérica de un sistema celular 2G/3G.
- 2.9.- Escenarios de provisión de servicios itinerantes en redes celulares múltiples.
- 2.10.- Escenarios de servicios de comunicaciones entre redes celulares y otras redes.
- 2.11.- Unidades que constituyen la arquitectura de los sistemas 3GPP.
- 2.12.- Arquitectura de red del sistema LTE del organismo 3GPP.
- 2.13.- Red de acceso E-UTRAN.
- 2.14.- Mecanismos de transferencia de información en la interfaz radio.
- 2.15.- Control de los servicios portadores radio y S1 a través de la interfaz S1-MME.
- 2.16.- Arquitecturas de las redes de acceso E-UTRAN y UTRAN.
- 2.17.- Arquitectura básica de la red troncal EPC.
- 2.18.- Tipos de interconexión a través de la interfaz SGi.
- 2.19.- Posibles configuraciones de la red EPC.
- 2.20.- Encaminamiento de tráfico a través de la red matriz (itinerancia).
- 2.21.- Encaminamiento de tráfico a través de la red visitada, con acceso a los servicios de red visitada (Itinerancia).
- 2.22.- Encaminamiento de tráfico a través de la red visitada, con acceso a los servicios de la red matriz.
- 2.23.- Arquitectura simplificada del modelo de red NGN.
- 2.24.- Arquitectura simplificada del subsistema IMS.
- 2.25.- Posibles escenarios de conexión entre dos redes telefónicas diferentes.
- 2.26.- Escenario de Soluciones de Interconexión de Redes Heterogéneas.
- 2.27.- Arquitectura genérica de una red de acceso de radio.
- 2.28.- Arquitectura de protocolos de un teléfono multinodal.
- 2.29.- Niveles y mecanismos de interconexión asociados.
- 2.30.- Marco de interconexión entre redes de acceso 3GPP.
- 2.31.- Interconexión mediante interfaces propias del sistema 3GPP.
- 2.32.- Interconexión a través de nuevas interfaces del sistema 3GPP.
- 2.33.- Selección de celda de teléfono en estado desocupado en cubrimiento muti-RAT.
- 2.34.- Procedimiento de gestión para localizar un usuario en despliegue multi-RAT.
- 2.35.- Gestión para la localización en despliegue multi-RAT en el modo ISR.
- 2.36.- Procedimiento de movilidad entre las redes E-UTRAN a UTRAN.
- 2.37.- Modelo de interconexión en la entidad VCC.
- 2.38.- Arquitectura de interconexión de la organización 3GPP con diferentes redes.
- 2.39.- Modelo de red rural NGN (arquitectura general IMS-AGC).
- 2.40.- Modelo de red rural NGN.

- 2.41.- Escenario del establecimiento de sesiones SIP.
- 3.1.- Relación entre R_i y λ para diferentes valores de l (Dos dimensiones).
- 3.2.- Relación entre R_i y λ para diferentes valores de l (tres dimensiones).
- 3.3.- Rutas de re-transmisión de mensajes de la fase *Invite* (salto por salto)
- 3.4.- Pérdidas l_e vs l .
- 3.5.- Cálculo de P_i vs porcentaje de pérdidas (l) y tasa de tráfico λ
- 3.6.- P_{100} vs porcentaje de pérdidas de la red (l).
- 3.7.- Solicitudes (*SIE*) vs pérdidas (l) y η saltos (tres dimensiones).
- 3.8.- Solicitudes (*SIE*) vs pérdidas (l) y número de saltos η (dos dimensiones).
- 3.9.- Cálculo de R_0 vs λ y l (dos dimensiones).
- 3.10.- Cálculo de R_0 vs λ y l (tres dimensiones).
- 3.11.- Pérdidas L_e vs η y l (fase *no-Invite*).
- 3.12.- Ancho de Banda vs λ y l , con parámetros $\eta = \mu = 5$ (tres dimensiones).
- 3.13.- Ancho de Banda vs λ y l , con parámetros $\eta = \mu = 5$ (dos dimensiones).
- 4.1.- Recorrido de las re-transmisiones *no-Invite*.
- 4.2.- Ancho de banda (B) generado por el IMS con llamadas locales.
- 4.3.- Ancho de banda (B) generado por el IMS con llamadas itinerantes.

Índice de tablas

Tabla

Página

- 2.1. Centros educativos y unidades médicas por estado.
- 2.2. Centros educativos nivel básico en zona rurales.
- 2.3. Unidades de red e interfaces 3GPP para E-UTRAN.
- 2.4. Unidades de red e interfaces de EPC para el acceso desde E-UTRAN.
- 2.5. Escenarios posibles de llamadas telefónicas entre dos redes diferentes.
- 2.6. Cadena de ruteo del flujo de señalización.
- 2.7. Distribución de tráfico y embotellamiento.
- 2.8. Propuesta de distribución de tráfico en la red.
- 2.9. Retransmisiones de solicitudes *Invite* debido a pérdidas de la red IMS.
- 2.10. Retransmisiones solicitudes *no-Invite* debido a pérdidas en redes IMS.
- 2.11. Tamaño de los mensajes SIP para UDP.
- 3.1. Solicitudes *Invite* $R_i(l, \lambda, n)$ generadas en horario normal de trabajo.
- 3.2. Número de solicitudes *Invite* $R_i(l, \lambda, n)$ (horas pico).
- 3.3. Número de solicitudes *Invite* en casos de emergencia $R_i(l, \lambda, n)$.
- 3.4. Resumen de datos calculados de l_e vs l (rango 0 al 20%).
- 3.5. Solicitudes *no-Invite* $R_o(l, \lambda, n)$ vs l , λ (horario normal de trabajo).
- 3.6. Solicitudes *no-Invite* $R_o(l, \lambda, n)$ vs l , y λ (horas Pico de trabajo).
- 3.7. Resultados de $R_o(l, \lambda, n)$ vs l , y λ (Emergencias).
- 3.8. Resultados numéricos de L_e vs l y η .
- 3.9. Rutas posibles de los usuarios durante el establecimiento de una llamada.
- 3.10. Cadena de ruteo del flujo de tráfico de señalización.
- 4.1. Cálculo R_i experimento #1 (fase *Invite*).
- 4.2. Pérdidas l_e de extremo a extremo (dos direcciones).
- 4.3. Resultados numéricos de las P_i vs l y λ .
- 4.4. Solicitudes (SIE) exitosas.
- 4.5. Solicitudes R_o vs λ y l (fase *no-Invite*).
- 4.6. Cálculo de L_e vs l y η
- 4.7. Anchos de banda (B) para las tres comunidades rurales.
- 4.8. Visualización del embotellamiento de servidores CSCF.
- 4.9. Servidores participantes en la red con diferentes escenarios posibles.
- 4.10. Propuesta de segmentación del porcentaje de tráfico.
- 4.11. Distribución de tráfico propuesta para diferentes escenarios.

Capítulo 1

1. Introducción

Actualmente un buen porcentaje de las zonas rurales remotas y aisladas de México se encuentran vulnerables, debido a que no disponen de salud y educación, y comunicaciones efectivas con que cuentan las zonas urbanas. La disponibilidad de redes de comunicación en contextos rurales crea condiciones para la provisión de servicios de Tele-educación y Telesalud entre otros. Estas redes tienen el potencial de detonar desarrollo económico, así como también dar acceso a herramientas que puedan contribuir al mejoramiento de la calidad de vida de los habitantes de estas zonas. Los diferentes órdenes de gobierno intentan proveer a los ciudadanos de los elementos necesarios para resolver sus necesidades básicas de comunicación a través de estas redes. Sin embargo, limitaciones geográficas y de infraestructura del contexto rural, aparte de otros factores financieros, tecnológicos y regulatorios, dificultan su despliegue. Es en este escenario en donde tecnologías y arquitecturas emergentes pueden apoyar la dotación de servicios de comunicación a las zonas rurales. Las Redes de Siguiete Generación (NGN por sus siglas en Inglés), constituyen una alternativa que permite la conectividad a diferentes redes de acceso, estas redes facilitan una plataforma tecnológica flexible, interoperable y adaptable que resulta adecuada para su aplicación en contextos rurales. Es por esto, que a través de este trabajo de investigación se propone un modelo de NGN, capaz de proporcionar acceso telefónico a zonas rurales remotas y aisladas de México. Por otro lado, las comunicaciones se han convertido en un recurso estratégico para los países que buscan su desarrollo, y la integración de sus sitios públicos (escuelas, centros de salud, centros comunitarios y oficinas de gobierno) hasta alcanzar la cobertura universal en todo el país.

En el caso de México se ha reportado que seis de cada diez mexicanos carecen de estos servicios básicos de comunicaciones. Adicionalmente, de las mil localidades urbanas con mayor población, solo treinta y ocho por ciento de las viviendas cuentan con una computadora y de estas solo el setenta y cinco por ciento están conectadas a Internet. Específicamente, existen doscientos cincuenta mil sitios públicos que carecen de este servicio. De las ciento noventa mil localidades rurales, solo el siete por ciento de las viviendas cuenta con una computadora y menos del cincuenta por ciento están conectadas a Internet [1]. El modelo de red NGN propuesto en este documento proporciona una plataforma tecnológica que contribuye a conectar las diferentes zonas rurales y los sitios públicos mencionados a la red de banda ancha nacional de telecomunicaciones, utilizando diferentes redes de accesos inalámbricos (fijos y móviles) y alámbricas (*legacy*). El trabajo desarrollado en esta tesis basa su diseño de arquitectura y su modelo matemático logístico, en las recomendaciones de organizaciones internacionales como el Grupo de Socios de la 3ra Generación (3GPP por sus siglas en Inglés) [2]. Así también, el compromiso de la Unión Internacional de Telecomunicaciones (UIT), como oficina de la Organización de las

Naciones Unidas, es una pieza clave en la regulación y la estandarización de las Tecnologías de la Información y la Comunicación (TIC), elaborando directrices, normas y recomendaciones para Redes NGN. México por su parte, ha hecho un compromiso de apoyar a las zonas rurales y urbano marginadas, a través de su oficina gubernamental conocida como Coordinación de la Sociedad de la Información y del Conocimiento (CSIC), para desarrollar un plan estratégico para proveer servicio de Internet de banda ancha a sitios y espacios públicos incluyendo a las zonas rurales de México, dando inicio al programa denominado México Conectado [1]. El objetivo de dicho sistema es detonar el desarrollo socioeconómico y cultural de las personas más desprotegidas, ofreciendo conectividad de Internet y acceso de servicios de telecomunicaciones aplicados a la salud, educación y eventos culturales de interés para las diferentes zonas rurales y urbanas más desatendidas. Entre los objetivos se incluyeron a 3,200 municipios, escuelas e instituciones públicas de salud, por la necesidad de que estos servicios sean provistos a través de la Banda Ancha de Internet. Así también, las grandes empresas desarrolladoras de tecnologías en telecomunicaciones han llevado a cabo importantes proyectos y han aportado inversiones significativas para posicionarse como jugadores clave en el entorno de las nuevas tecnologías de redes. Para esto se han aliado a las organizaciones internacionales de estandarización, buscando tener las mejores arquitecturas de red y protocolos cada vez más interoperables. En esta carrera por el liderazgo tecnológico de más de dos décadas, se puede observar que la convergencia tecnológica ha impactado el ecosistema de comunicaciones de una manera importante.

A este respecto, las principales organizaciones internacionales encargadas de las recomendaciones en el ámbito de las telecomunicaciones son las siguientes: UIT, 3GPP, 3GPP2, IEEE, ETSI, IETF. El compromiso de la UIT como oficina de la Organización de las Naciones Unidas es y ha sido una pieza clave en la regulación y la estandarización de las TIC, elaborando directrices, normas y recomendaciones para redes NGN tales como las recomendaciones de la serie Y.2000, las cuales proporcionan los fundamentos de las NGN [3]. Particularmente, la recomendación Rec. UIT-T Y.2001 (12/2004) de la UIT define a las Redes de Sigüiente Generación como: Red basada en paquetes que permite prestar servicios de telecomunicación utilizando múltiples tecnologías de transporte de banda ancha propiciadas por la Calidad de Servicio (QoS por sus siglas en Inglés), y en la que las funciones relacionadas con los servicios son independientes de las tecnologías subyacentes relacionadas con el transporte. Permite a los usuarios el acceso sin dificultad a redes y a proveedores de servicios y/o servicios de su elección. Puede soportar movilidad generalizada que permitirá la prestación coherente y ubicua de servicios a los usuarios [4]. Así mismo el Proyecto 3GPP y 3GPP2 colocaron las bases de un sistema global de comunicaciones de tercera generación (3G) para sistemas móviles basándose en las especificaciones del sistema de comunicaciones móviles globales (GSM) dentro del marco del proyecto internacional de telecomunicaciones móviles 2000 de la UIT. Con el fin ubicar el área donde se aplicará el proyecto se describe a continuación el entorno genérico del mismo.

1.1 Objetivos de la tesis doctoral

El objetivo específico y principal del presente trabajo de tesis es el diseño de una arquitectura de red de siguiente generación aplicada a zonas rurales y urbanas marginadas de México, con capacidad de integrarse al Sistema México Conectado de la República Mexicana. La red propuesta permite integrar diferentes tecnologías inalámbricas, así como también alámbricas de circuitos conmutados y de paquetes, redes inalámbricas con tecnologías 3G y 4G operando el Protocolo Internet (IP) de acuerdo a las recomendaciones de ETSI, IETF, 3GPP, 3GPP2, IEEE y UIT.

Los aspectos más relevantes y que son aportaciones derivadas de éste trabajo son:

- Un modelo de red de siguiente generación aplicable a zonas rurales y urbanas marginadas de México que ofrezca servicios de telefonía y multimedia a través del Protocolo de Internet.
- El modelo de la red está basado en el Subsistema Multimedia de 3GPP (IMS por sus siglas en Inglés).
- El modelo está constituido de una arquitectura de red abierta para permitir conectar a las diferentes tecnologías de acceso alámbricas e inalámbricas (RTCP, ISDN, WLAN, WiMAX, GERAN, UTRAN y E-UTRAN).
- Capacidad de integrarse al sistema México Conectado como red complementaria.
- La Arquitectura de red esta validada con un modelo matemático, el cual se implementó utilizando la herramienta de simulación numérica (MatLab).

Con el fin ubicar el área donde se aplicará el proyecto se describe a continuación el entorno genérico del mismo.

1.2 Entorno de esta tesis

El presente trabajo propone una arquitectura abierta de red digital inteligente con capacidad de gestión y provisión de servicios con aplicación al entorno rural. Además, puede soportar redes de acceso que utilicen diferentes tecnologías de circuitos conmutados y de paquetes usando el protocolo IP. Por tratarse de una red NGN, el modelo contempla la capacidad de interconectar redes de acceso 3GPP, 3GPP2, IEEE, y RTCP entre otras. La arquitectura está diseñada como una red de cuatro niveles (acceso, transporte, control y aplicación) como lo muestra la figura 1.1.

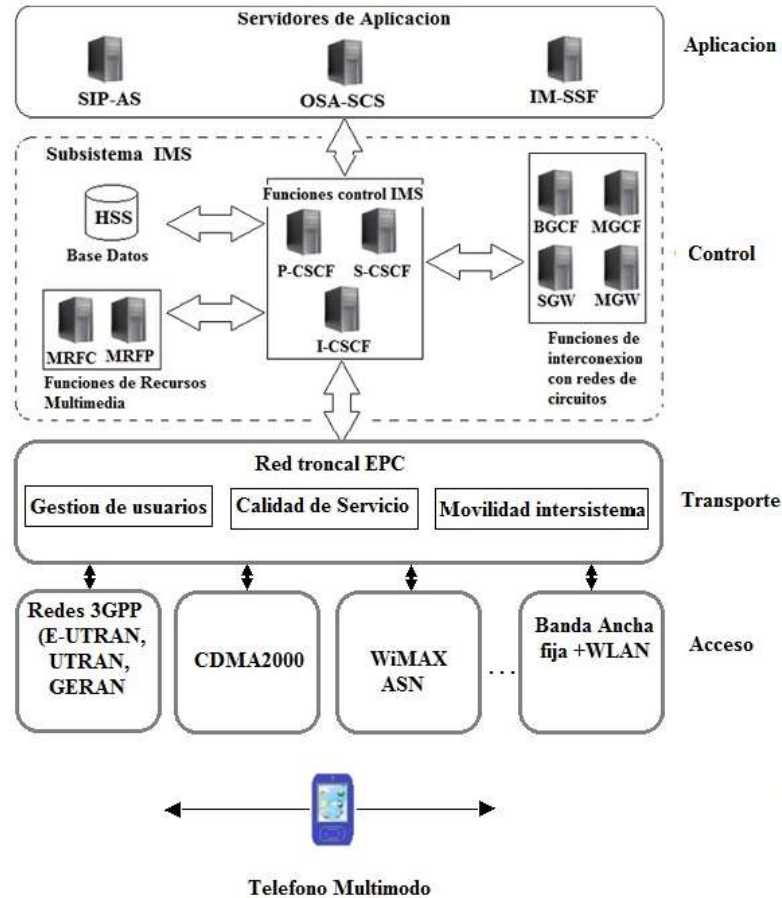


Figura 1.1. Modelo de arquitectura de red NGN propuesta

La arquitectura de la figura 1.1, ofrece una serie de servicios multimedia en el ámbito de las comunicaciones móviles, desde los servicios telefónicos hasta servicios de videoconferencia, acceso a Internet y soporte a diferentes medios. Debido a que el modelo de red propuesto es abierto y aunque las funciones de sincronización, movilidad (*handover*), transmisión de datos de usuario, localización, etc., son similares en las diferentes redes de acceso, el uso de protocolos de enlace, acceso, red y aplicación son específicos a cada tecnología de radio. El proyecto 3GPP ha diseñado interfaces particulares para cada una de las diferentes redes de acceso a su red troncal (EPC por sus siglas en Inglés). El entorno físico del modelo de red propuesto se localizará en cada ciudad capital de las treinta y dos entidades federativas mencionadas, de acuerdo a lo propuesto en el programa México Conectado.

1.3 Estructura del documento

Esta tesis está organizada de la siguiente manera: en el capítulo 1 se plantea el problema a resolver, sus objetivos generales y específicos, y la estructura del documento. En el capítulo 2 se mencionan los antecedentes, la evolución de la tecnología, y los avances tecnológicos que han limitado la solución del problema. Se plantea también, la propuesta de solución y la arquitectura para una red rural. Así mismo, describe la

arquitectura del sistema LTE y del subsistema IMS (corazón de la red NGN). También se presenta un método para medir la eficiencia del flujo de tráfico que permita identificar los cuellos de botella de la red y dimensionar su capacidad. Describe la interconectividad de la red LTE con otras redes heterogéneas. Finalmente, propone el modelo de arquitectura de red NGN rural y su modelo matemático para validar la arquitectura propuesta, mencionando las limitaciones del modelo de red propuesta. En el capítulo 3 se presenta el marco de simulación, en el cual se realizan los experimentos numéricos, con la herramienta MatLab. Con estos experimentos se valida el modelo de red, y se obtienen conclusiones de sus ventajas y limitaciones. En el capítulo 4 se realiza el análisis de los experimentos realizados, observando el efecto que tienen todos y cada uno de los parámetros que participan en la red, específicamente el ancho de banda generado durante el proceso del establecimiento de las llamadas, para garantizar una Calidad de Servicio (QoS por sus siglas en Inglés) en base a las normas internacionales del organismo UIT. Durante el análisis se hace mención de los parámetros que participan a la entrada a la red y los parámetros resultantes de la red (ancho de banda, retardos y variaciones de retardo). Se menciona como los parámetros intrínsecos de la red, afectan la calidad de servicio que los usuarios detectan al realizar llamadas (llamadas locales o itinerantes). El capítulo 5 trata sobre los trabajos futuros a realizar, y finalmente se escriben las conclusiones de la tesis.

Capítulo 2

2. Marco teórico

Como se ha expuesto en el párrafo anterior, la imperativa necesidad de realizar un proyecto de telecomunicaciones que permita comunicar a las diferentes zonas rurales y urbanas marginadas de México ha sido una inquietud de la población en general. La penetración del servicio de telecomunicaciones a las zonas marginadas ha quedado al margen de los planes de expansión de las empresas concesionarias, debido a que en la mayoría de los casos la población con mejores capacidades de pago se encuentra concentrada en las grandes ciudades. Sin embargo es importante considerar que el desarrollo integral de cualquier país se logra unificando a las zonas rurales al entorno global de desarrollo. Así que los servicios de telecomunicaciones son vitales para este proceso de integración.

2.1 Antecedentes

El informe anual de la organización UIT [5], documento que trata de la Medición de la sociedad de la información del año 2013 indica el sitio donde se localiza México en la actualidad a nivel mundial en relación a los servicios de telecomunicaciones. Se observa claramente en dicho reporte que la banda ancha móvil a través de teléfonos inteligentes y tabletas se ha incrementado. Las cifras mostradas a finales de 2013 indican una gran

demanda en todo el mundo en lo que se refiere a productos y servicios de tecnologías TIC, así mismo se observa que los precios de los servicios celulares y de banda ancha que utilizan tecnologías 3G y 3G+ han aumentado también. Según estos estudios a finales de 2013 hubo 6,800 millones de usuarios móviles celulares aproximadamente, casi tantos como habitantes en el planeta. Sin embargo se estima que 2,700 millones de personas también se conectaron a Internet, aunque las velocidades y los precios varían mucho de un país a otro. Según el organismo de la UIT las conexiones de banda ancha móvil a través de redes 3G y 3G+ crecen un 40 por ciento de la media anual, lo que equivale a 2,100 millones de abonados de banda ancha móvil y una tasa de penetración mundial de casi 30 por ciento. Casi 50 por ciento de todos los habitantes del mundo disponen ahora de una cobertura de red de 3G. El índice de desarrollo de las TIC (IDI) del organismo UIT clasifica 157 países de los 195 que reconoce la ONU en función de su nivel de acceso, utilización y conocimientos de las tecnologías TIC. Los 30 países más avanzados son países de altos ingresos, lo que demuestra que existe una fuerte relación entre ingresos y avance de las tecnologías TIC. El reporte muestra que México se encuentra fuera de este contexto, y que existen grandes diferencias entre países desarrollados y en desarrollo, y los valores del índice IDI son en promedio dos veces más elevados en los países desarrollados que en los países en desarrollo. Es por esta razón que México a través de su oficina de la Sociedad de la Información y del Conocimiento busca un mayor y mejor acceso a la utilización de las tecnologías TIC en sectores de Salud, Educación, Empleo y Seguridad en las zonas rurales y urbanas marginadas. El reto más importante es instalar servicios y redes de siguiente generación que permita integrar las comunidades rurales de indígenas y zonas marginadas para superar la pobreza. El reporte menciona que a principios de 2013 casi 80 por ciento de todos los hogares del mundo tenían un televisor, en comparación con 41 por ciento de hogares con computador y 37 por ciento con acceso a Internet. Así mismo, muestra que el número de hogares con acceso a Internet aumenta en todos los países, pero sigue habiendo grandes diferencias, con tasas de penetración que al final de 2013 alcanzarán prácticamente 80 por ciento en los países desarrollados, en comparación con 28 por ciento en los países en desarrollo. Se estima que 1,100 millones de hogares en todo el mundo todavía no están conectados a Internet, y 90 por ciento de ellos están en países en desarrollo. Lo cual significa que países como México un gran porcentaje de hogares y sobre todo de las zonas rurales no cuenta con este servicio que es vital para su desarrollo. Se observa también que la proporción de hogares con acceso a Internet en los países en desarrollo ha pasado de 12 por ciento en 2008 a 28 por ciento en 2013, una tasa de crecimiento anual del 18 por ciento. El porcentaje de usuarios de Internet en la población ha crecido a una media de dos cifras durante los últimos 10 años. El porcentaje de la población en línea en los países desarrollados alcanzo el 77 por ciento a finales de 2013, en comparación con el 31 por ciento en los países en desarrollo. El estudio muestra que los países en desarrollo están todavía muy lejos de alcanzar un crecimiento razonable. Investigaciones del organismo UIT muestran que los gastos de capital (CAPEX) de los operadores de telecomunicaciones (Concesionarios) terminaron en 2008, con una inversión global total de 290,000 millones de dólares.

El Grupo 7 del organismo mencionado define las zonas rurales o remotas como aquellas áreas desatendidas por los servicios de telecomunicaciones. *“Estas zonas consisten en asentamientos dispersos, pueblos o ciudades pequeñas, y pueden estar ubicados varios cientos de kilómetros de una ciudad o centro urbano. En algunos casos*

de acuerdo con su situación geográfica y condiciones culturales de algunas zonas suburbanas se consideran rurales. En particular, podemos decir que un área rural tiene una o más de las siguientes características: La escasez o ausencia de habitantes y servicios, condiciones topográficas adversas, como lagos, ríos, colinas, montañas o desiertos que hacen muy costoso construir redes de línea fija. Además de las condiciones climáticas, duras que pueden afectar gravemente el funcionamiento de los equipos, bajo nivel de actividad económica, basada principalmente en la agricultura, la pesca, la artesanía, el ingreso per cápita bajo, bajo desarrollado en infraestructura, y escasez de servicios de telefonía.” [5]

Por otro lado también se contempla proveer los servicios TIC a las zonas urbanas marginadas, particularmente a las instituciones de seguridad social, tales como: Hospitales y/o clínicas del ISSSTE, Seguro Social, escuelas federales, Secretarías de Seguridad, Instancias del gobierno en sus tres niveles: Federal, Estatal y Municipal. Según datos de la ENDUTIH de INEGI, y de la Dirección de Información Estadística de Mercados de la extinta COFETEL, al 2009 en México había 34 millones de usuarios de computadora, 27 millones de usuarios de Internet y 79 millones de usuarios de telefonía móvil. De los hogares mexicanos: 97% contaban con servicio de TV. 83% con radio, 79% con servicios de telefonía, 27% con computadora y 18% con conexión a internet. No obstante, las condiciones económicas, sociales y culturales de nuestro país en los últimos años, y su heterogeneidad en todo el territorio nacional, han dado como resultado un lento avance en el abatimiento de la brecha digital, acercándose a un punto que permitirá que sólo las personas y comunidades con niveles de ingresos altos y medios puedan acceder al uso de las tecnologías de la información y de las telecomunicaciones, dejando excluidas a las personas y comunidades en condiciones de vulnerabilidad y en zonas marginadas. Esto ha impedido que millones de mexicanos puedan acceder a las telecomunicaciones, internet y a sus contenidos y servicios digitales que les permitan mejorar sus condiciones de vida.

Por esta razón, la Estrategia Nacional de Conectividad, suscrita bajo la Agenda Digital México Conectado, [1] se centra en la construcción de las Redes de Banda Ancha de Cobertura Social, cuyo objetivo es sentar las bases para una transición acelerada del país hacia la sociedad de la información y el conocimiento.

Los objetivos específicos del sistema de conectividad de banda ancha son:

- Garantizar el acceso universal de banda ancha gubernamental y comunitario en todo el territorio nacional, especialmente en las escuelas y espacios educativos, bibliotecas, centros de salud, oficinas gubernamentales de los tres niveles de gobierno, y en la red de puntos comunitarios de acceso ubicados en zonas urbanas de alta marginación y localidades geográficamente remotas.
- Reducir los costos de acceso a la conectividad y las telecomunicaciones de las distintas instancias gubernamentales y comunidades rurales, con el fin de sentar las bases para la implantación generalizada de servicios digitalizados y en línea.
- Facilitar el proceso de adopción tecnológica de las instituciones educativas, de los servicios de salud, de las instancias gubernamentales, especialmente a nivel municipal, y de los servicios de desarrollo social que se impartan a través de los espacios de acceso digital comunitario.
- Crear las condiciones para acelerar el proceso de adopción y apropiación tecnológica de los mexicanos, especialmente aquellos en condiciones de vulnerabilidad y marginalidad.

Tabla 2.1 Centros educativos y unidades médicas por estado

No	Estados	Centros Educativos	Unidades Medicas
1	Aguascalientes	1137	163
2	Baja California	2192	463
3	Baja California Sur	538	143
4	Campeche	992	271
5	Chiapas	11012	1702
6	Chihuahua	3889	578
7	Coahuila	2059	481
8	Colima	623	237
9	Distrito Federal	3133	938
10	Durango	1565	533
11	Guanajuato	6023	983
12	Guerrero	5030	1106
13	Hidalgo	4320	1151
14	Jalisco	5714	1043
15	México	11148	1657
16	Michoacán	3512	1300
17	Morelos	1487	409
18	Nayarit	1983	493
19	Nuevo León	3514	761
20	Oaxaca	10729	1864
21	Puebla	7828	1393
22	Querétaro	1234	315
23	Quintana Roo	951	237
24	San Luis potosí	2515	600
25	Sinaloa	1584	648
26	Sonora	2445	730
27	Tabasco	2569	721
28	Tamaulipas	3596	721
29	Tlaxcala	1427	242
30	Veracruz	12851	1738
31	Yucatán	1756	395
32	Zacatecas	3496	555
	Total General	122,852	24,632

Fuente: Sitio web de SER, Sitio Web de Secretaría de Salud.

Universo de usuarios potenciales, la tabla 2.1 presenta el total de centros educativos y unidades médicas del país por entidad federativa, así como las escuelas de educación básica ubicadas en zonas de muy alta y alta marginación conforme a su población. Presenta a 122,852 centros educativos ubicados fuera del entorno urbano y 24,632 unidades médicas (clínicas y dispensarios) que se encuentran privados del servicio de las comunicaciones básica. Así mismo muestra la necesidad urgente de integrarlos a los

medios de comunicaciones de banda ancha, para mejorar su nivel y promedio de vida al contar con mejores servicio de salud eficientes.

La tabla 2.2 muestra a las escuelas de educación básica ubicadas en zonas de muy alta y alta marginación conforme a su población.

Tabla 2.2 Centros educativos de nivel básico en zona rurales

Grado de Marginación	Entidad Federativa	Número de escuelas por rango de habitante y grado de marginación										
		1-49			50-99			100-499			Total 100-499	Total General
		Muy Alto	Alto	Total 1-49	Muy Alto	Alto	Total 50-99	Muy Alto	Alto			
Muy Alto	Chiapas	850	1159	2009	860	521	1381	2106	3041	5147	8537	
	Guerrero	515	179	694	520	212	732	1338	1574	2912	4338	
	Oaxaca	215	152	367	317	344	661	1347	2810	4157	5185	
Alto	Campeche	29	89	118	21	28	49	76	342	418	585	
	Michoacán	374	427	801	245	417	662	145	921	2066	3529	
	Hidalgo	95	112	207	188	286	474	361	1937	2298	2979	
	Puebla	72	140	212	134	209	343	581	1939	2520	3075	
	S.L.P.	236	204	440	199	360	559	350	2345	2695	3694	
	Tabasco	19	27	46	10	57	67	22	669	691	804	
	Veracruz	308	517	825	439	636	1075	1409	5080	6489	8389	
	Yucatán	78	41	119	36	83	119	29	439	468	706	
	Medio	Durango	374	193	567	175	207	382	147	479	626	1575
Guanajuato		124	188	312	125	332	457	168	2047	2215	2984	
Nayarit		170	50	220	75	56	131	129	220	349	700	
Querétaro		35	46	81	68	113	181	123	611	734	996	
Sinaloa		154	113	267	134	196	330	137	784	921	1518	
Tlaxcala		6	36	42		30	30		106	106	178	
Zacatecas		33	121	154	15	155	170	31	638	669	993	
Bajo		Aguascalientes	1	15	16		13	13		21	21	50
		B.C.S	12	39	51	2	16	18	4	33	37	106
	Chihuahua	394	152	546	280	134	414	237	354	591	1551	
	Colima	5	6	11		12	12		41	41	64	
	Jalisco	203	272	475	105	389	494	81	767	848	1817	
	México	6	26	32	22	64	86	87	866	953	1071	
	Morelos	16	47	63	6	25	31	5	96	101	1951	
	Quintana Roo	28	6	34	22	23	45	41	263	304	383	
	Sonora	18	47	65	25	66	91	37	336	373	529	
	Tamaulipas	23	91	114	22	140	162	22	421	443	719	
	Muy Bajo	Baja California	4	25	29		17	17		38	38	84
Coahuila		12	20	32	8	50	53	3	159	162	242	
D.F.												
	Nuevo León	43	121	164	39	148	187	9	324	333	684	
TOTAL		4452	4661	9113	5339	5339	9426	9025	30701	39726	58265	

Fuente: Dirección General de Política de Telecomunicaciones y Radiodifusión, con datos de INEGI

Como resultado de la evidencia presentada, es clara la necesidad de implementar una estrategia de inclusión de las comunicaciones en México, sin embargo, esta necesidad solo puede cubrirse con la tecnología actual, la cual debe permitir adecuarse a los requerimientos y retos tecnológicos de las redes NGN.

2.2 Evolución de la tecnología

En cuanto a la evolución de la tecnología, en las anteriores dos décadas, las redes inalámbricas han pasado por tres generaciones en su evolución [6]. La primera generación de los sistemas celulares fue diseñada y optimizada para la transmisión analógica de señales de voz hacia y desde abonados móviles. Operando en modo de conmutación de circuitos, estos sistemas ofrecen la transmisión de datos en banda de voz. Redes que operan en las bandas de frecuencias de 450 y 800 MHz utilizan esquemas derivados de la modulación FDMA. Por otra parte, el inter-funcionamiento entre redes diferentes difícilmente se aplica. Es decir, un abonado no puede utilizar los servicios en una red distinta a la que está suscrita. Posteriormente, la segunda generación de los sistemas celulares conocida como Sistema Global para las Comunicaciones Móviles (GSM por sus siglas en Inglés) fue un gran paso adelante. En su forma original, la tecnología GSM opera en las bandas de 900, 1,800 y 1,900 MHz en las cuales utilizan un esquema modulación TDMA para la transmisión "en modo circuito" de la voz digitalizada y los datos digitales en hasta 9.6 kbits/s. La introducción de tarjetas de identificación de subscritor (SIM) y el protocolo GSM como parte de la Aplicación Móvil (MAP), permite habilitar sin defectos el inter-funcionamiento entre redes diferentes, facilitando a los suscriptores comunicarse por todo el mundo. Las redes GSM son ampliamente utilizadas en el mercado de telefonía móvil (60%) [7]. Dado que las redes GSM están diseñadas para los circuitos conmutados y servicios de voz ofrecen velocidades de datos bajas, no son adecuadas para soportar servicios conmutados de paquetes de Internet. Finalmente, con la tecnología 2.5G, se inicia la tercera generación, la cual ofrece mejoras destinadas a las comunicaciones de paquetes, tales como:

- GPRS, que se añade a las redes GSM para apoyar de manera eficiente los servicios de conmutación de paquetes. Soporta hasta 160 kbit/s de datos móviles de conmutación de paquetes junto con la telefonía por conmutación de circuitos. Sin embargo, es compatible con los servicios de paquetes en tiempo no real conmutada solamente.
- EDGE (Velocidades de Datos Mejoradas para redes de GSM Evolucionadas), que admite velocidades de datos de hasta 384 kbits/s.
- GERAN, (Red de Acceso Radio GSM/EDGE), que es la fase 2 de GSM/EDGE evolucionada. Ofrece velocidades de datos de hasta 1920 kbits/s para voz en paquetes y servicios en tiempo real.

La introducción del Sistema Universal de Telecomunicaciones Móviles de Tercera Generación (UMTS por sus siglas en Inglés), basado en tecnología de acceso múltiple por división de código de banda ancha (WCDMA por sus siglas en Inglés), es un paso más hacia la satisfacción de la creciente demanda de servicios de datos/Internet. El sistema UMTS se caracteriza por sus tasas de bits de usuario: puede alcanzar hasta 384 kbps en conexiones de conmutación de circuitos, y alcanza hasta 2 Mbps en conexiones de conmutación de paquetes [2]. Este sistema universal es la propuesta del Instituto de Estándares Europeos de Telecomunicaciones (ETSI por sus siglas en Inglés) para las Telecomunicaciones Móviles Internacionales (IMT) 2000, la Unión Internacional de Telecomunicaciones (UIT) inicia los objetivos del programa para establecer un sistema de comunicación de orden universal. Se han iniciado los estudios desde finales de 1980 por el esfuerzo y el financiamiento de universidades e industrias. En enero de 1998 el Grupo

Especial Móvil (SMG) del instituto ETSI seleccionó dos tecnologías de radio para la interfaz de acceso aérea de radio terrestre (UMTS UTRA): CDMA de banda ancha (WCDMA) en las bandas de frecuencias apareadas para la operación División de Frecuencia Dúplex (FDD por sus siglas en Inglés), y CDMA de División de Tiempo (TDCDMA) para una operación con la banda impar utilizando el mecanismo de División de Tiempo Dúplex (TDD) [8]. Esta decisión fue la base para el acceso terrestre (UTRA) propuesta del sistema UMTS presentada por el instituto ETSI a la organización UIT como la tecnología de transmisión de radio IMT2000. En países como Japón, Estados Unidos y Corea, eligieron de forma independiente sus propias tecnologías de acceso de radio 3G al mismo tiempo que el IMT-2000. El Proyecto del Grupo de Socios de la Tercera Generación (3GPP por sus siglas en Inglés) se creó en 1998 para tener un único foro para la estandarización de una especificación común UTRA. El grupo 3GPP es una organización sin precedentes en todo el mundo que reúne a expertos de los diversos organismos regionales de normalización, y están constituidos por los siguientes grupos:

- Instituto Europeo de Normas de Telecomunicaciones (ETSI),
- Asociación de Negocios de la Industria de Radio (ARIB) / Japón,
- Grupo de Estándares de Telecomunicaciones Inalámbricas de China (CWTS)
- Comité de Estandarización de Estados Unidos (T1-Telecommunications)
- Asociación de Tecnología de Telecomunicaciones (TTC/Japón)

Las tres generaciones de tecnologías analizadas en este punto, han permitido grandes avances en cuanto a comunicaciones inalámbricas se refiere, sin embargo, también han limitado la solución del problema. Debido a que los diferentes fabricantes de tecnologías y los creadores de las recomendaciones internacionales han tenido que pasar por el proceso de convergencia y el liderazgo tecnológico.

2.3 Avances tecnológicos que limitan la solución del problema

2.3.1 Introducción a las arquitecturas GSM y GPRS

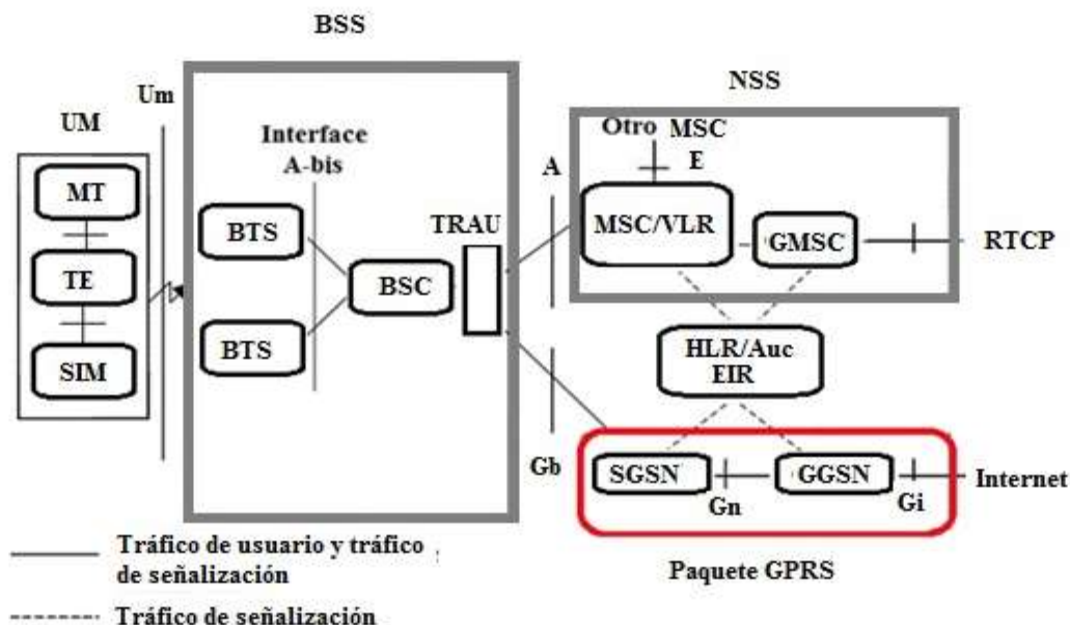


Figura 2.1. Arquitectura GSM con módulo GPRS agregado.

La arquitectura GSM conocida como arquitectura de segunda generación (G2) ha emigrado al siguiente nivel (3G) ya que sus servicios se encuentran limitados en cuanto a la velocidad, tráfico y esquemas de acceso al medio para comunicaciones de voz y datos. La figura 2.1, muestra la arquitectura simplificada de GSM con módulo de GPRS agregado.

- Unidad Móvil (UM):
 Equipo físico del suscriptor; lo constituyen el equipo móvil y una tarjeta conocida como Modulo de Identidad del Suscriptor (Tarjeta *SIM* por sus siglas en Inglés).

 - Subsistema de Estación Base (BSS):
 Realiza funciones requeridas para mantener las comunicaciones de radio con la Estación Móvil: Codificación/Decodificación de voz, ajuste de la velocidad desde y hacia la red inalámbrica. Este subsistema se conecta con la entidad MSC del módulo NSS, y con el nodo SGSN del Módulo GPRS a través de las interfaces A y Gb respectivamente. Las redes GSM contienen muchos módulo BSS, los cuales a su vez consisten de varios controladores BTS's y un Controlador de Estación Base (BSC).
1. Estación Transceptora Base (BTS): Componente de la red que atiende una celda. Se ocupa de la interfaz de aire de señalización (*Um*), del cifrado y el procesamiento de la voz.

2. Controlador de Estación Base (BSC): Es el elemento central del subsistema BSS y su función es controlar una o más estaciones BTS.
 3. Unidad de transcodificación y de Adaptación (TRAU): Es el elemento BSS que se encarga de la transcodificación de voz, es decir, es capaz de convertir la voz de un formato de codificación digital a otro y viceversa.
- Subsistema de Red (NSS)
Su función es conectar a las redes inalámbricas con la red pública, permite el handover entre diferentes subsistemas BSS, y comprende funciones para la localización en todo el mundo de usuarios, y controla los cargos, contabilidad, y servicios de itinerancia de usuarios entre diferentes proveedores en diferentes países. El subsistema NSS está constituido de los siguientes módulos:
 1. Centro de Conmutación Móvil (MSC): Este elemento es la parte principal del módulo NSS desde el punto de vista de control de llamadas. MSC es la interfaz entre el sistema de radio y las redes fijas legacy. Controla todas las funciones requeridas para manejar el servicio de los circuitos conmutados hacia y desde las estaciones móviles.
 2. Pasarela MSC (GMSC): Participa en la administración de la movilidad y las comunicaciones, y conexiones a otras redes.
 3. Registro de Localización de Usuario (HLR): Controla la localización central y la base de datos de gestión, que almacena la información del suscriptor (identidad de la Unidad Móvil internacional, el perfil de usuario, etc.), así también controla la información sobre la ubicación dinámica en relación con la zona de ubicación actual del equipo EM.
 4. Registro de Localización de Visitantes (VLR): Localiza y gestiona la base de datos asociada a cada pasarela MSC. Proporciona un almacenamiento local para todas las variables y funciones necesarias para gestionar las llamadas desde y hacia los suscriptores móviles en el área relacionada con la pasarela MSC.
 5. Centro de Autenticación (AUC): Gestiona la autenticación y el cifrado para cada suscriptor.
 6. Registro de Identidad del Equipo (EIR): Realiza un seguimiento de las estaciones móviles y sus identidades a fin de evitar el uso de equipos robados.
 - Núcleo de Paquete GPRS: Su función principal es proporcionar la transferencia de datos en modo de paquete. Requiere dos nodos adicionales:
 1. Nodo de Soporte GPRS de Servicios (SGSN): Controla la gestión de la movilidad, la seguridad y autorización.
 2. Nodo de Soporte GPRS de Cabecera (GGSN): Se encarga de la unidad de inter-funcionamiento entre las redes GPRS y externa de paquetes de datos (PDN). El nodo GGSN es responsable de la gestión de direcciones IP, la gestión de calidad de servicio y funciones de pasarelas externas.

Agregando las funciones GPRS a la tecnología GSM se pueden proporcionar servicios de comunicación básica para tráfico de conmutación de circuitos y paquetes, estos servicios son la base de la red central UMTS.

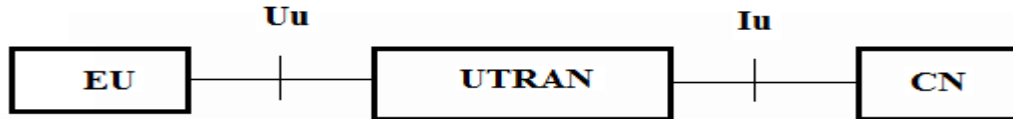


Figura 2.2. Dominios de redes UTRAN

2.3.2 Arquitectura UMTS y limitaciones

Las redes UMTS consisten de tres dominios que interactúan:

1. Equipo de Usuario (EU)
2. Red de Acceso de Radio Terrestre UMTS (UTRAN).
3. Red Trocal (CN).

La arquitectura básica UMTS se puede observar en la figura 2.2.

- Equipo de Usuario (EU)

Se compone de dos partes: Equipo Móvil (EM), y el módulo de Identificación del Subscriber UMTS (USIM).

1. EU: Teléfono o terminal utilizado para la comunicación de radio.
2. USIM: Tarjeta inteligente que contiene la identidad del abonado y realiza una serie de funciones de seguridad.

- UTRAN (UMTS Red de Acceso Radio Terrestre)

La tarea principal de la UTRAN es crear y mantener Portadores de Acceso de Radio (RAB) para la comunicación entre el usuario UE y el núcleo de red CN.

- CN (Red Central)

La red Central (CN) controla la función de handover entre sistemas, así también las pasarelas de acceso a otras redes (fijas o inalámbricas), y las funciones de gestión de red. El objetivo de la red central 3G CN es actuar como la red principal universal para controlar el acceso de diferentes tipos de redes (radio y fijas).

- La arquitectura 3GPP versión 99 (3GPP Release99)

La primera versión de especificaciones UMTS, 3GPP R99 presenta la nueva red UTRAN de acceso de radio basada en el acceso de banda ancha WCDMA, y describe la red principal del sistema GSM modificada con tecnología GPRS.

- Arquitectura de Red UTRAN (figura 2.3)

Los bloques funcionales básicos de la arquitectura UTRAN son el nodo B y el controlador de red de radio (RNC). UTRAN está conectado al usuario a través de la interfaz de radio Uu, y se comunica con la red central a través de la interfaz Iu.

Nodo B: Convierte los flujos de datos entre las interfaces Iub y Uu, y participa en la gestión de recursos de radio [6]. Se puede decir que es equivalente a la estación BTS GSM.

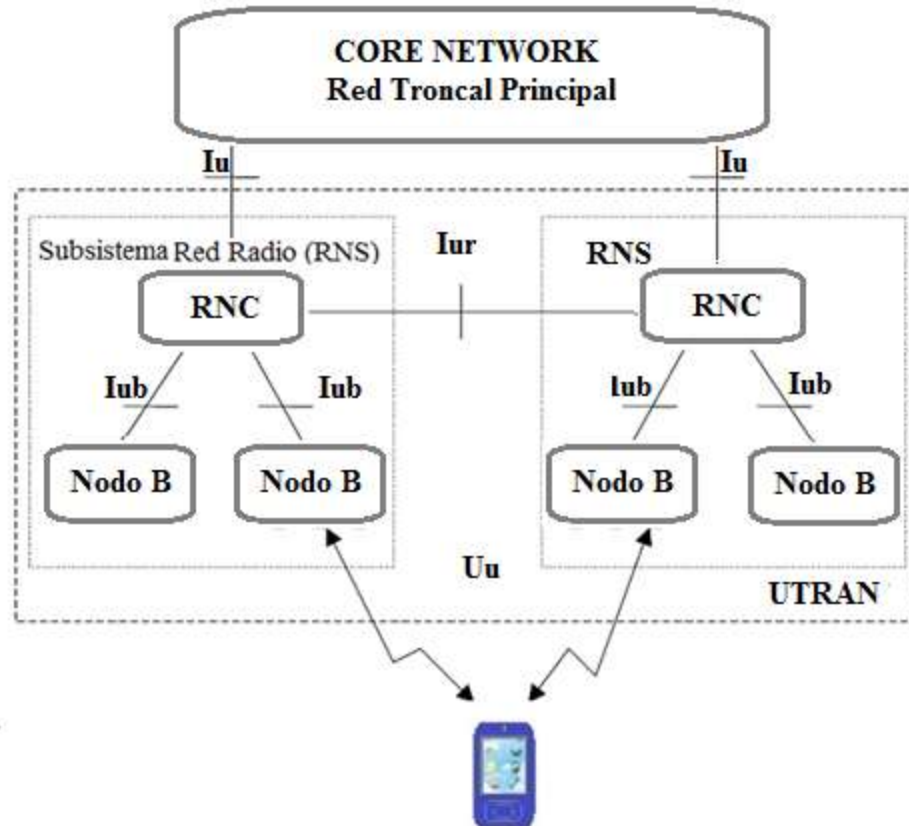


Figura 2.3. Arquitectura de red UTRAN

- **Controlador de Red de Radio (RNC):** El controlador RNC es el punto de acceso para todos los servicios que la arquitectura UTRAN proporciona a la red central. El controlador RNC es equivalente a un nivel de pares con el controlador BSC GSM. Es responsable de controlar los recursos asociados con un número de nodos B, y negociar con la red principal aspectos tales como portadoras y calidad de servicio. La *release 99* de la arquitectura UMTS del grupo 3GPP describe la nueva interfaz de radio Iur, de la arquitectura UTRAN. Esta interfaz conecta dos controladores de red de radio (RNC) vecinos juntos y se utiliza para la nueva función basada en el acceso WCDMA implementado en el controlador RNC. Estaciones base (BS) están conectados a la entidad RNC a través de la interfaz Iub.

Red Principal (CN)

La *release99* del grupo 3GPP minimiza los cambios necesarios para que el subsistema de red GSM /GPRS pueda soportar las dos generaciones. La red central UMTS tiene un dominio de conmutación de circuito (CS-CN) y un dominio de conmutación de paquetes (PS-CN). La arquitectura UTRAN se conecta con el dominio de

conmutación de circuitos a través de la interfaz IuCS y al dominio de conmutación de paquetes a través de la interfaz IuPS. Ver figura 2.4.

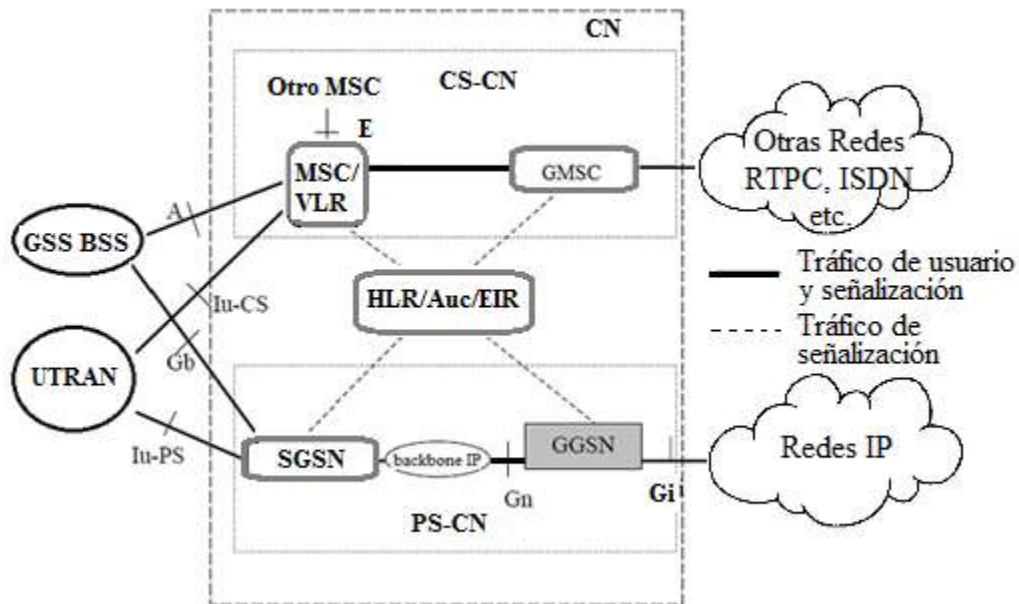


Figura 2.4. Arquitectura principal de 3GPP release99

- Red Principal (CS-CN)
 Dominio de conmutación de circuitos se desarrolló a partir de la técnica GSM NSS. La red CS-CN soporta la conectividad a la red telefónica pública conmutada y la Red Digital de Servicios Integrados para servicios de conmutación de circuitos. Es compatible además con los servicios tradicionales de telefonía, tales como voz y fax, y también es compatible con servicios mejorados, tales como servicio de mensajes cortos [7]. Los componentes importantes del módulo CS-CN son las entidades MSC, VLR y GMSC.
- Dominio de Conmutación de Paquetes (PS-CN)
 Dominio de conmutación de paquetes evolucionó a partir de la red principal GPRS basada en el protocolo IP. El módulo PS-CN soporta la conectividad a las redes de datos por paquetes, como Internet, para los servicios de conmutación de paquetes.
 Los componentes importantes de este módulo PS-CN son los nodos SGSN y el GGSN. La red de transporte que conecta los nodos SGSN y el GGSN se llama red troncal IP, que puede considerarse como un "Intranet" privada y por lo tanto se separa en realidad de otras redes por la funcionalidad de servidor de seguridad [9].
- HLR y AuC
 El servidor HLR mantiene el perfil de abonado tanto para el circuito y los servicios de paquetes. El servidor AuC es compatible con funciones de autenticación para ambos dominios. Por lo tanto estos componentes pertenecen a ambos dominios.

Arquitectura de la red de 3GPP Release 4

La *release 4* de del grupo 3GPP no cambia mucho, pero la red de acceso actualiza notablemente la red central, especialmente en el dominio de conmutación de circuitos. Ver figura 2.5.

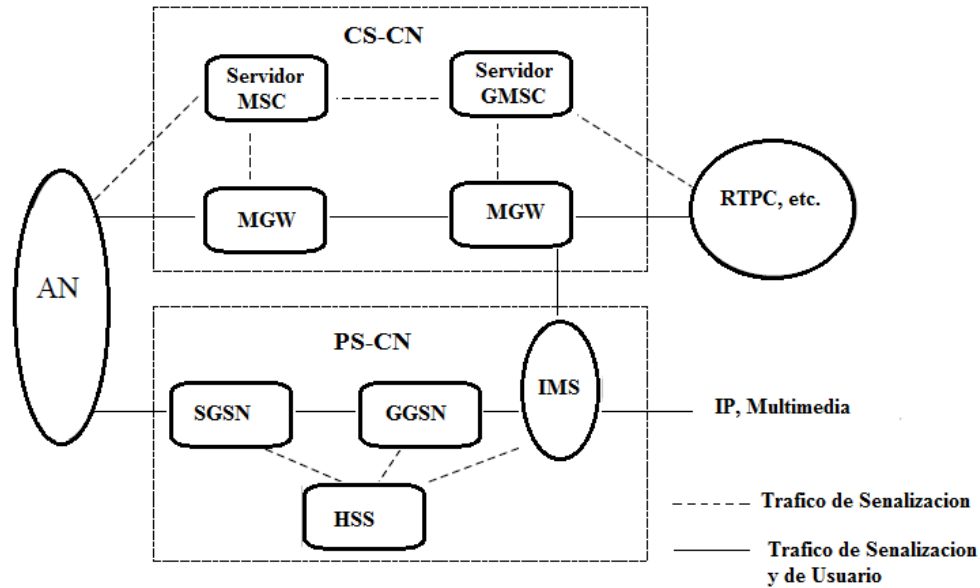


Figura 2.5. Arquitectura de la red 3GPP *release 4*

Una de las características más sobresalientes de R4 de *CS-CN* es la separación del plano de usuario y el plano de control: el módulo *MSC/VLR* se divide en un servidor *MSC* y una Pasarela de *Multimedia* (*MGW*), y la entidad *GMSC* se divide de la misma manera como la entidad *MSC*, con un servidor de *GMSC* y una pasarela de medios.

- **Servidor MSC**

La función principal del servidor *MSC* es gestionar la conexión, es decir, que maneja toda la señalización y controla la pasarela de medios. El servidor *MSC* también contiene un servidor *VLR* para guardar los datos de servicio del abonado móvil.

- **Media Gateway**

La pasarela *MGW* es la entidad dominio de circuitos *CS* y por lo tanto puede ser denominado como pasarela *CS-MGW*. Se mueve la comunicación multimedia (por ejemplo, voz) de un portador de circuito de conmutación a un portador de *IP*, y también opera las funciones de cancelación de eco y transcodificación. [10].

Esta arquitectura separa los planos de usuario (el portador de datos) y el plano de control (el portador de la señalización y control) obteniendo los siguientes beneficios [10]:

1. La traducción de los medios de comunicación hace posible el uso de una red *IP* común tanto para la conmutación de paquetes como para la conmutación de

circuitos de datos, esta red IP también podría ser utilizado para otros tipos de tráfico simultáneamente.

2. Las entidades MSC o GMSC puede controlar varias pasarelas de medios, se mejora la flexibilidad y la escalabilidad de la arquitectura. Además, la señalización y la capacidad de datos de usuario se pueden escalar de forma independiente. Otro aspecto importante en la arquitectura de la *release R4* del grupo de 3GPP de la red CN es la adición del Subsistema Multimedia Basado en IP (IMS), que permite a los operadores de Redes Móviles Terrestres Publicas (PLMN) ofrecer a sus suscriptores los servicios multimedia basados y construidas sobre las aplicaciones de Internet, servicios y protocolos. Es importante hacer notar que el subsistema multimedia IP utiliza el dominio de paquetes PS para transportar señalización y tráfico multimedia de datos. Es independiente del dominio de circuitos CS, aunque algunos elementos de red pueden ser comunes con este dominio [11]. Esto significa que no es necesario desplegar un dominio de CS a fin de apoyar a una red de subsistema multimedia IP. El subsistema multimedia IP consta de todos los elementos de la red principal para la prestación de servicios multimedia.

Arquitectura de la Red 3GPP, *release 5*

En la arquitectura de red del grupo 3GPP *release 5*, correspondiente a la red de acceso los cambios se incrementan (Ver figura 2.6). Aquí el problema principal es el transporte de IP en la red de acceso. En la *release 99*, el protocolo ATM implementa la transmisión dentro de la red de acceso debido a su fuerza de soporte de la QoS. Conforme pasa el tiempo, IP como una tecnología de transporte controlará los mecanismos de la calidad de servicio que proporcionan el apoyo necesario para servicios tales como voz y vídeo en tiempo real. Por lo tanto, en la *release 4* de las recomendaciones del grupo 3GPP está especificando el uso de IP como una alternativa a la ATM. En la *release 5* del grupo 3GPP todo el tráfico que viene de la red UTRAN es basada en el protocolo IP y por lo tanto maneja conmutación de paquetes. Esto permite a los operadores a converger sus redes desde el paquete tradicional de conmutación de circuitos y la infraestructura a una operación de "todo-IP". En esta fase, el criterio principal de selección para la tecnología de acceso de radio es ofrecer suficiente ancho de banda para el servicio utilizado y la tecnología de radio en sí se vuelve menos importante [9]. La visión futura es que la red central 3G cuente con interfaces para varias tecnologías de acceso de radio, por ejemplo, las redes de acceso GSM / EDGE (GERAN), CDMA2000, WCDMA y de la red inalámbrica de área local (WLAN), etc.

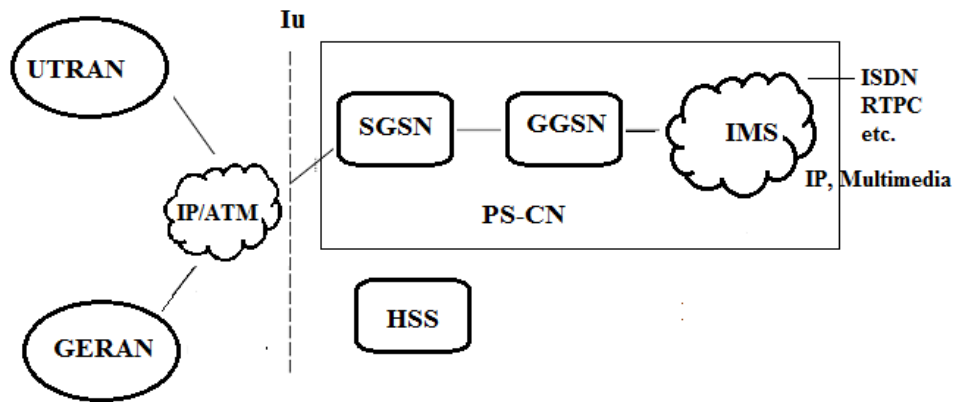


Figura 2.6. Arquitectura de la red 3GPP release 5

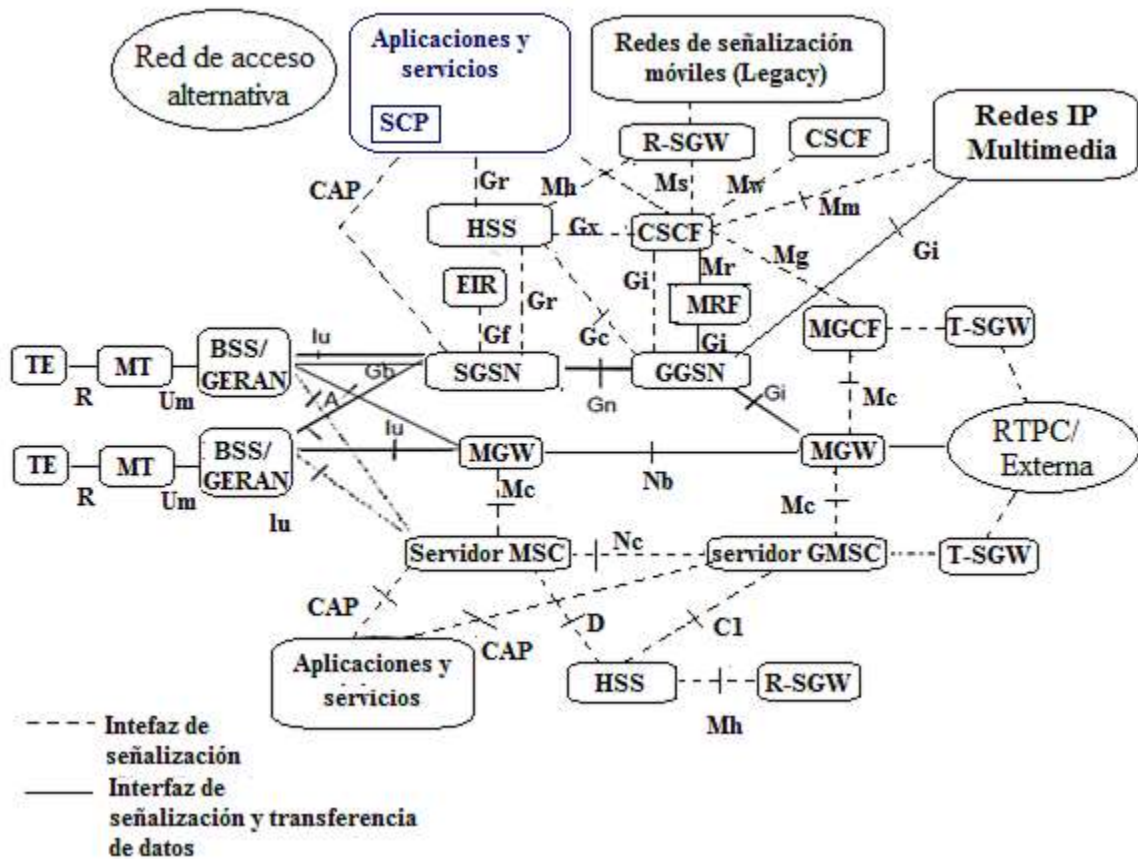


Figura 2.7. Arquitectura lógica de la red del grupo 3GPP release R4/5

Puede observarse que todas estas tecnologías de redes inalámbricas pueden operar en forma independiente, con sus tecnologías y protocolos propios. Pero debido a sus limitaciones tecnológicas, jamás podrá ofrecer los servicio que ofrece una red de siguiente generación. Las cuales ofrecen servicios de voz, datos y video en forma horizontal ajustándose a las recomendaciones Y.2000 de la organización UIT. Es decir, operar una red abierta y de alta velocidad para todas las redes de acceso fija y móvil, operando sobre

dos capas principales: Transporte y aplicación, dejando con esto abierta la posibilidad de nuevos proveedores de servicios. De acuerdo al grupo 3GPP la arquitectura lógica de la red (release R4 y R5) quedaría de acuerdo a la figura 2.7.

2.4 Propuesta de solución

En base a los antecedentes mencionados los cuales se componen de elementos teóricos y científicos basados en estudios de la Unión Internacional de Telecomunicaciones (UIT) oficina de las Naciones Unidas, el Instituto de Estándares Europeos de Telecomunicaciones (ETSI), información estadística de las diferentes entidades federales de México, recopilada y publicada por la Sociedad de la Información y del Conocimiento de México. Así como también las Especificaciones del Proyecto de Socios de la 3ª Generación (3GPP) formado por organizaciones tales como: ARIB, ATIS, CCSA, ETSI, TTA, TTC que desarrollan estándares. La organización 3GPP incluye recomendaciones para las tecnologías de redes de telecomunicaciones móviles, incluidas las de acceso de radio, la red básica de transporte, capacidades de servicio, códec, seguridad, calidad de servicio (QoS), entre otras. También ofrece especificaciones para el acceso de redes alámbricas de servicios conmutados de circuitos (*legacy*), y redes inalámbricas: WiMAX y WLAN. Se propone un modelo de red de cuatro capas, las cuales se clasifican como: Acceso, transporte, control y aplicación, que se ajuste a la definición que hace la organización UIT en cuanto a los siguientes servicios: Red basada en paquetes que permita prestar servicios de telecomunicación utilizando múltiples tecnologías de transporte de banda ancha propiciadas por la Calidad de Servicio (QoS), y en la que las funciones relacionadas con los servicios son independientes de las tecnologías subyacentes relacionadas con el transporte. Permite a los usuarios el acceso sin dificultad a redes y a proveedores de servicios y/o servicios de su elección. Puede soportar movilidad generalizada que permitirá la prestación coherente y ubicua de servicios a los usuarios [4]. Diseñada para manejar tráfico rural y urbano marginado, es decir de baja y mediana velocidad, y que pueda conectarse al proyecto México Conectado.

2.5 Red LTE Advanced del grupo 3GPP

La arquitectura del modelo de red rural propuesta está basada en las siguientes tecnologías: Arquitectura de Red del Sistema LTE, arquitectura del Subsistema Multimedia (IMS), arquitectura y dimensionamiento de redes IMS, arquitectura de interconectividad de LTE con redes heterogéneas, modelo matemático para validar arquitectura de red rural NGN, arquitecturas de Redes Inalámbricas de Siguierte Generación.

2.5.1 Arquitectura de red del sistema LTE

2.5.1.1 Introducción

Para describir la arquitectura de la red de comunicaciones móviles LTE de la organización 3GPP, se analizará la arquitectura genérica de los sistemas celulares 2G/3G

actuales. Se identificarán las partes fundamentales de la arquitectura de red de los sistemas (GSM, UMTS, LTE) del grupo 3GPP. Así también se describen los nuevos componentes creados por LTE relacionados a las redes GSM y UMTS. El concepto de la tecnología LTE fue creado por la organización 3GPP para estudiar la evolución de la red de acceso universal móvil UMTS, llamada tecnología UTRAN, aunque también se le llama red LTE a la tecnología E-UTRAN. Cuando se refiere a la red troncal, el grupo 3GPP utiliza el término Evolución de la Arquitectura del Sistema (SAE por sus siglas en Inglés), cuando se refiere a la red troncal evolucionada de conmutación de paquetes. Oficialmente a esta red troncal se le llama Red de Paquetes Evolucionada (EPC por sus siglas en Inglés). También se le llama Dominio de Conmutación de Paquetes 3GPP Evolucionado (EPS por sus siglas en Inglés). El término LTE se utiliza como sinónimo de sistema EPS, la primera especificación del sistema EPS fue presentada en la *release 8* por la organización 3GPP. Antes de presentar la arquitectura de red propuesta, se analizarán conceptos relacionados a las diferentes tecnologías de acceso, y de las redes troncales utilizadas para el transporte de la información, así como también se analizará la arquitectura de los subsistemas IMS que forman parte del corazón de la red, todas estas tecnologías pertenecientes al grupo 3GPP.

2.5.1.2 Arquitectura de los sistemas celulares

La arquitectura de los sistemas celulares representa un modelo de red de alto nivel, constituido de tres componentes básicos:

- Equipo del usuario.
Dispositivo por medio del cual el usuario accede a los servicios de la red. Incluye una tarjeta inteligente de Circuito Integrado Universal (UICC por sus siglas en Inglés), la cual contiene la información necesaria para que el usuario se pueda conectar a la red y utilizar sus servicios. Posteriormente, el usuario se conecta a través de la red de acceso, utilizando una interfaz de radio.
- Red de Acceso.
Esta red es la encargada de proporcionar la conectividad entre los equipos de usuario a la red troncal, para esto utiliza sistemas de comunicaciones de radio. La red de acceso tiene como función gestionar el uso de los recursos de radio disponibles para proveer servicios portadores de una manera eficiente. Los recursos de transmisión en la red de acceso son controlados desde la red troncal. La red de acceso está constituida por las radio bases, y en los sistemas móviles actuales G2 y G3 por equipos controladores de las estaciones base.
- Red Troncal.
Esta red tiene como función controlar el acceso a la red celular (autenticar los usuarios que están adscritos al sistema), gestión de la movilidad de los usuarios, gestión de las sesiones de datos o circuitos que transportan la información de los usuarios, mecanismos de interconexión con otras redes, etc. También forman parte de la red troncal las funciones asociadas con el control de los servicios finales ofrecidos a los usuarios (control y señalización asociada al servicio de telefonía). La red troncal también puede tener equipos que realicen las funciones de conmutación de circuitos, encaminamiento de

paquetes (ruteo), bases de datos, etc. La figura 2.8, muestra la arquitectura genérica de un sistema celular.

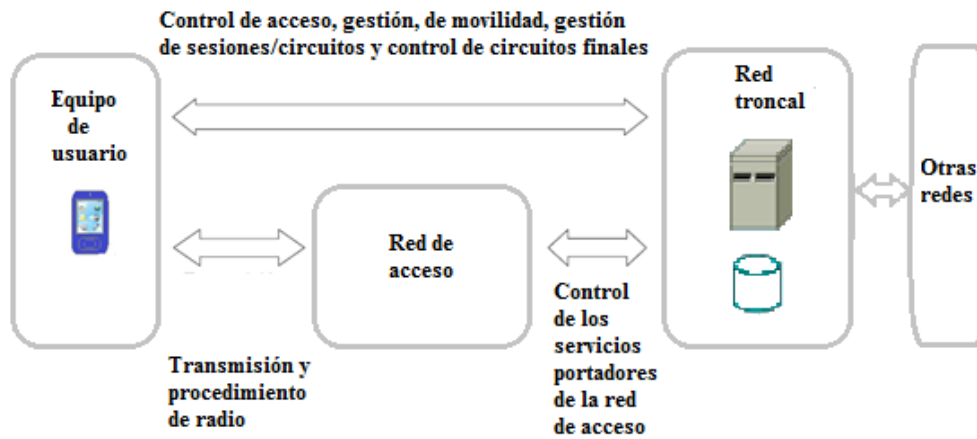


Figura 2.8. Arquitectura genérica de un sistema celular 2G/3G

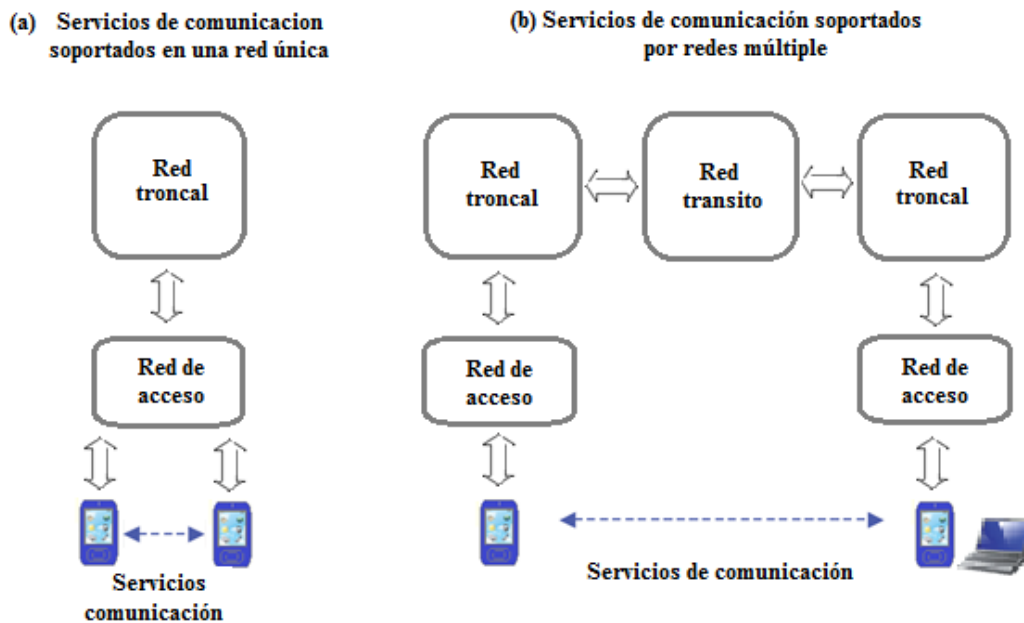


Figura 2.9. Escenarios de provisión de servicios itinerantes en redes celulares múltiples.

La figura 2.9, muestra dos escenarios para proveer el servicio de comunicaciones a través de redes celulares móviles.

El escenario (a) contiene dos usuarios conectados a la misma red. En (b) se observa la provisión de los servicios entre equipos de usuario operando en redes celulares diferentes interconectadas entre sí, a través de redes de tránsito. La idea de estas figuras es mostrar como la organización 3GPP introduce sus redes troncales EPC con capacidad de poder intercomunicar a usuarios de diferentes tecnologías a sus sistemas de transporte de datos, para permitirles servicios de itinerancia con tecnologías diferentes. Estos servicios son de gran utilidad para las comunidades rurales.

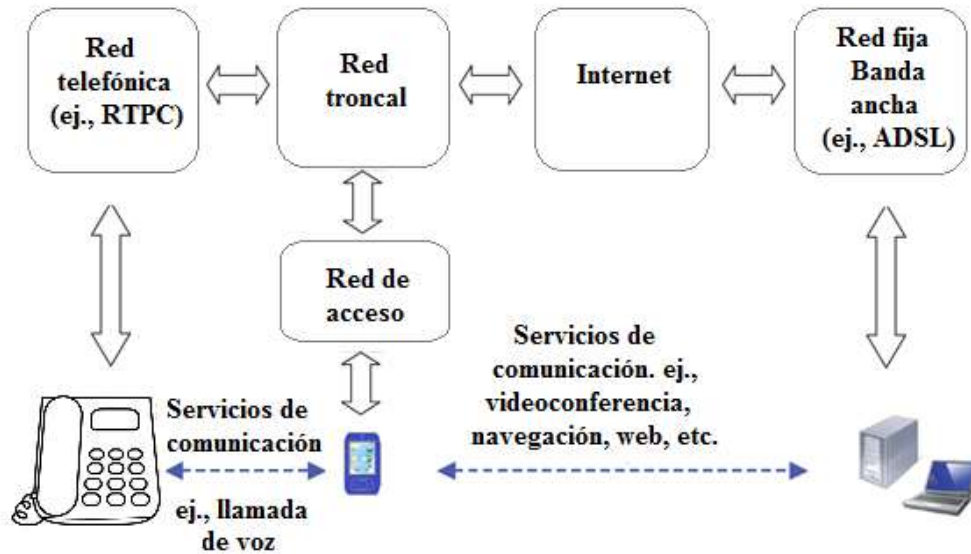


Figura 2.10. Escenarios de servicios de comunicaciones entre redes celulares y otras redes.

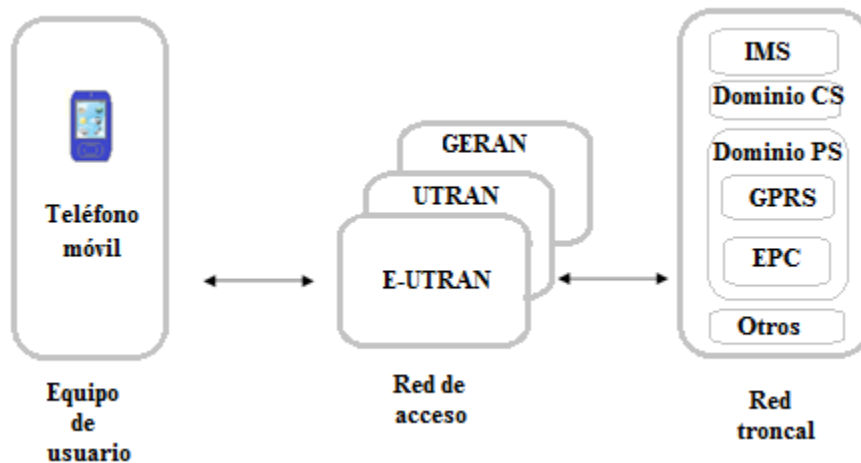


Figura 2.11. Unidades que constituyen la arquitectura de los sistemas 3GPP

La figura 2.10, muestra la provisión de servicios entre equipos conectados a redes celulares y equipos localizados en otras redes (red ADSL, Internet, etc.). La figura 2.11, muestra también que las redes de transporte de la organización 3GPP pueden intercomunicar a las diferentes redes alámbricas e inalámbricas con su tecnología, ya que ha diseñado interfaces particulares para las diferentes tecnologías de acceso.

2.5.1.3 Arquitectura de red de los sistemas 3GPP

La figura 2.11, presenta a las unidades que constituyen la arquitectura de alto nivel de los sistemas 3GPP (GSM, UMTS, y LTE). El equipo de usuario (EU), la red troncal (CN), y las redes de acceso propias de 3GPP (AN) [13].

El equipo de usuario del grupo 3GPP.

Está formado de los siguientes elementos: Equipo móvil (EM), tarjeta UICC o Modulo de Identificación del Subscriptor (SIM por sus siglas en Inglés) utilizada en los sistemas GSM y SIM universal (USIM) en los sistemas UMTS y LTE. A través de esta separación entre terminal y tarjeta se permite que un usuario pueda utilizar diferentes terminales o teléfonos para comunicarse a la red.

Red de acceso

La organización 3GPP ha identificado tres tipos de red de acceso diferentes: Redes de Acceso de Radio GSM/EDGE (GERAN por sus siglas en Inglés), Redes de Acceso Terrestre de Radio UMTS (UTRAN por sus siglas en Inglés), y UTRAN evolucionadas (E-UTRAN por sus siglas en Inglés). Las redes de acceso GERAN y UTRAN forman parte del sistema 3G UMTS mientras que la arquitectura E-UTRAN es la nueva red de acceso del sistema LTE. Cada red de acceso define su propia interfaz de radio para comunicarse con los equipos de usuarios: GERAN (GSM) utiliza TDMA, UTRAN utiliza WCDMA y E-TRAN utiliza la tecnología OFDMA. La interconexión de las redes de acceso a la red troncal se realiza mediante las interfaces AN-CN.

Red Troncal

La red troncal se divide de forma lógica en dos dominios: dominio de circuitos (CS), dominio de paquetes PS (redes IP y X.25) y el Subsistema IP Multimedia (IMS). El dominio CS contiene todas las unidades de la red troncal basadas en conmutación de circuitos. Es decir, recursos asignados en forma dedicada hasta que el servicio sea terminado (servicios de voz y videoconferencia en redes UMTS). El dominio de circuitos de la red troncal es accesible a través de las redes de acceso UTRA y GERAN. A diferencia de la red E-UTRAN, la cual proporciona todos los servicios a través del dominio PS. Existen dos dominios PS: GPRS y EPC. La tecnología GPRS pertenece al dominio de paquetes PS desarrollado en el contexto de las redes GSM y también forma parte del sistema UMTS. Los servicios GPRS son accesibles a través de las redes UTRAN y GERAN. La entidad EPC por el contrario es una evolución de la tecnología GPRS, la cual fue optimizada para ofrecer el servicio de conectividad IP a los equipos de usuario a través de la red E-UTRAN. El dominio EPC también ofrece el acceso al servicio de conectividad IP desde otras redes de acceso 3GPP (UTRAN y GERAN) así como de redes no 3GPP (CDMA200, *Mobile WiMAX*, etc.). En la red troncal se encuentra el servidor que contiene la información de los usuarios del sistema (HSS por sus siglas en Inglés), el cual apoya la operación de los dominios CS y PS, así como también de la entidad IMS. El subsistema IMS como parte de la red troncal contiene los elementos relacionados con la provisión de servicios IP multimedia, basados en el protocolo SIP del grupo IETF. El módulo IMS atiende la señalización para proveer los servicios multimedia y utiliza como transporte el servicio de transferencia de datos proporcionado por el dominio PS. Así el IMS, constituye el plano de control de dichos servicios, quedando separado las funciones asociadas al transporte de la información (dominio PS) y las funciones de la capa de control de los servicios (señalización a nivel de aplicación), que corresponde a las funciones de las redes de siguiente generación.

2.5.1.4 Arquitectura de la red LTE

Se presenta la arquitectura de la red LTE debido que el modelo de red rural propuesto en esta tesis está basado en la arquitectura del organismo 3GPP, en cuanto a que su tecnología ofrece una arquitectura abierta a todas las demás tecnologías de acceso (alámbricas e inalámbricas). A la Arquitectura de Red del Sistema LTE, también se le conoce como el sistema EPS, y está constituida por las siguientes unidades: La red de acceso E-UTRAN y la red troncal de paquetes EPC, y el subsistema IMS. Estos componentes fueron diseñados para proveer todo tipo de servicios de telecomunicaciones mediante mecanismo de conmutación de paquetes, en este sentido la entidad EPC es una versión evolucionada del sistema GPRS. La red de acceso E-UTRAN y la red troncal EPC ofrecen servicios de transferencia de paquetes IP entre los equipos de usuario y redes de paquetes externas (IMS), y otras redes de telecomunicaciones como Internet. La figura 2.12, muestra la arquitectura de red del sistema LTE.

El servicio de transferencia de paquetes IP que ofrece la red LTE entre el equipo de usuario y una red externa se conoce como Servicio Portador (EPS por sus siglas en Inglés). Así mismo, el servicio de transferencia de paquetes que ofrece la red de acceso E-UTRAN se denomina como Portadora de Acceso de Radio E-UTRAN (ERAB). El sistema LTE contempla también el acceso a sus servicios a otras redes de acceso que no pertenecen a las recomendaciones de 3GPP (CDMA2000, *Mobile WiMAX*, Redes 802.11, etc.). Estas características fueron fundamentales para seleccionar la arquitectura generalizada de la red

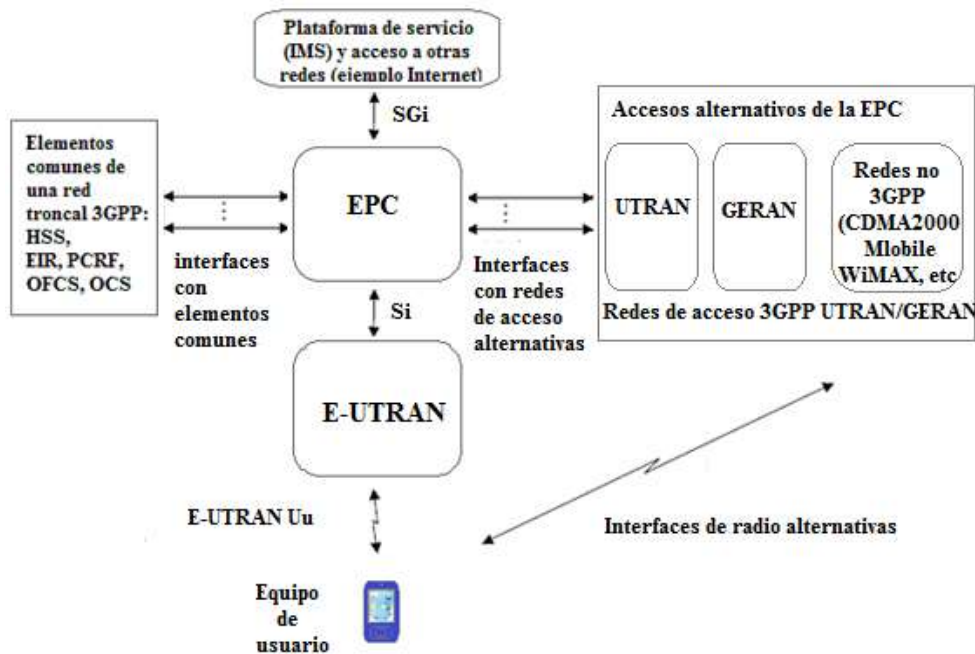


Figura 2.12. Arquitectura de red del sistema LTE del organismo 3GPP

3GPP, para ser la base de nuestra arquitectura de red rural propuesta.

La infraestructura de una red LTE, además de los equipos propios que implementan las funciones del estándar 3GPP, también cuenta con otros elementos de red propios de las redes IP tales como; ruteadores, servidores Protocolo de configuración Dinámica del servidor *Host* (DHCP por sus siglas en Inglés) para la configuración automática de las direcciones IP de los equipos de la red LTE y Servidores de Nombre de Dominio (DNS por sus siglas en Inglés) para asociar los nombres de los equipos con sus direcciones IP. La figura 2.12 es clave, debido a que esta presenta en forma genérica la arquitectura de las redes NGN, características que la hacen útil para este propósito.

2.5.1.5 Red de acceso evolucionada: *E-UTRAN*

2.5.1.5.1 Arquitectura de *E-UTRAN*

La arquitectura de la red de acceso E-UTRAN está constituida de una sola entidad de red denominada *NodeB* Evolucionado (eNB por sus siglas en Inglés). Esta estación base de E-UTRAN integra toda la funcionalidad de la red de acceso. La figura 2.13, muestra la red de acceso que ofrece la conectividad entre los equipos de usuario (UE) y la red troncal EPC. Un nodo eNB se comunica con los otros elementos del sistema a través de tres interfaces: E-UTRAN, Uu, S1 y X2. La arquitectura E-UTRAN, se describen a detalle en [14] y [15].

Cada interfaz de la red de acceso tiene una función, por ejemplo la interfaz Uu o interfaz de radio LTE, tiene como función específica transferir información por el canal de radio entre el nodo eNB y los equipos de usuario. Así la interfaz S1 conecta al nodo eNB a la red troncal EPC. Esta interfaz está compuesta de dos partes diferentes: S1-MME que sustenta

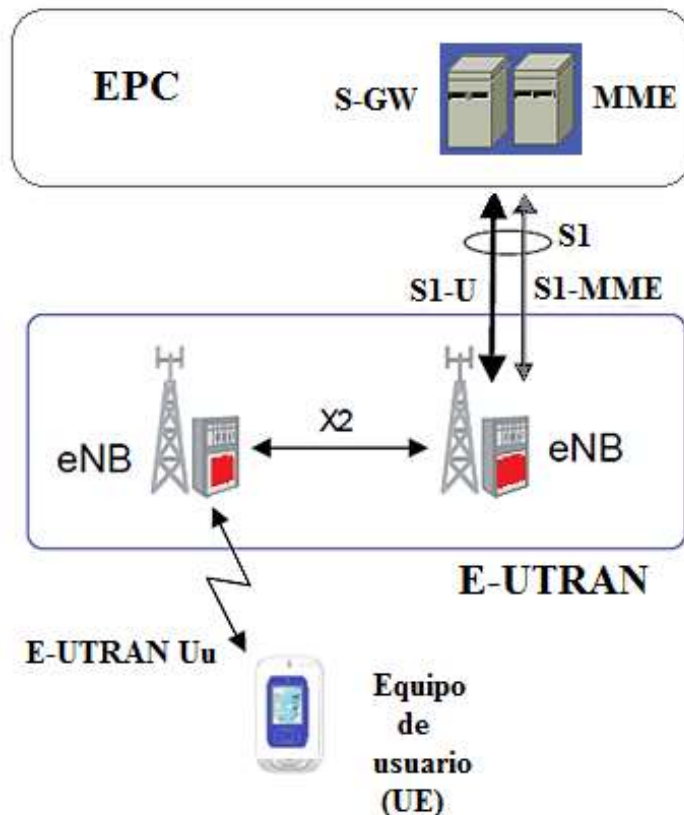


Figura 2.13. Red de acceso E-UTRAN

el plano de control y la interfaz S1-U soporta el plano de usuario. Es decir, que el plano de usuario de una interfaz es la pila de protocolos utilizada para el envío de tráfico del usuario a través de dicha interfaz S1-U. El plano de control es la pila de protocolos requerida para sustentar las funciones y procedimientos necesarios para gestionar la operación de dicha interfaz o la entidad correspondiente. Esta separación de los planos de control y usuario en la interfaz S1 permite realizar la conexión del nodo eNB con dos nodos diferentes de la red troncal. Así que mediante la interfaz S1-MME, el nodo eNB se puede comunicar con una entidad de red del módulo EPC encargada de las funciones del plano de control (*Mobility Management eNB UE element Entity MME*). Al mismo tiempo, a través de la interfaz S1-U, el nodo eNB se comunica con otra entidad de red encargada de procesar el plano de usuario (*Serving Gateway, S-GW*). También los nodos eNB se conectan entre sí mediante la interfaz X2. Con esta interfaz los nodos eNB se intercambian mensajes de señalización que permiten la gestión eficiente del uso de los recursos de radio (reducir interferencias entre nodos eNB) y tráfico de los usuarios del sistema cuando estos se desplazan de un nodo eNB a otro durante un proceso de movilidad (*handover*). La tabla 2.3, presenta las unidades de red e interfaces de la red E-UTRAN e indica las especificaciones del organismo 3GPP relacionadas con cada una de ellas

Tabla 2.3. Unidades de red e interfaces 3GPP para E-UTRAN

Unidades de Red	Denominación	Descripción	Referencias
	<i>Evolved NodeB</i> (eNB)	Estación base de la red de acceso E-UTRAN	TS36.300[4] TS36.401[5]
Interfaces	Denominación	Unidades de red asociadas	
	E-UTRAN Uu (también denominada LTE Uu o interfaz radio)	eNB UE	TS 36.300[4] Documentos TS36.2xx TS 36.3xx
	X2	eNB eNB	TS 29.281[26]
	S1-MME	eNB Red Troncal EPC (MME)	Documentos TS 36.41x
	S1-U	eNB Red Troncal EPC (S-GW)	

2.5.1.5.2 Unidades de red e interfaces

En este apartado se describirán las tres Unidades de red que forman el núcleo de la red troncal EPC (MME, S-GW y P-GW), así mismo la entidad HSS la cual es común al sistema 3GPP y a las interfaces SGi, S5/S8, S11, S10 Y S6a. Las estructuras de las interfaces S1 y X2 se describen en las referencias [16] y [17].

2.5.1.5.2.1 Nodo B evolucionado (eNB)

El nodo eNB integra todas las funciones de la red de acceso, en este nodo terminan todos los protocolos específicos de la interfaz de radio. Mediante estos protocolos el eNB transmite los paquetes IP hacia/desde los equipos de usuario junto con los mensajes de señalización que se requieren para el control de la interfaz de radio.

El nodo eNB contiene funciones de control de admisión de los servicios portadores de radio, control de movilidad (*handover*), asignación dinámica de los recursos de radio, en ambos enlaces; ascendente y descendente (funciones de calendarización [*scheduling*]), control de interferencias entre estaciones base, control de la realización y del envío de medidas desde los equipos de usuario que pueden ser útiles en la gestión de recursos, etc. El nodo eNB puede también realizar la selección dinámica de la entidad MME de la red troncal EPC, cuando un teléfono se registra en la red LTE. Esta función le asigna un grado de flexibilidad importante en la operación de la red. En la redes E-UTRAN el nodo eNB puede estar conectado simultáneamente a múltiples entidades MME de la red troncal. Un nodo eNB puede enviar/recibir paquetes IP de los usuarios a los que sirve a través de diferentes pasarelas S-GW de la red troncal EPC. El nodo eNB contiene funciones de encaminamiento de tráfico de los usuarios hacia la pasarela de red S-GW correspondiente.

2.5.1.5.2.2 Interfaz de radio

La interfaz de radio puede operar tres tipos de mecanismos de transferencia de la información en el canal radio: difusión de señalización de control, envío de paquetes IP y transferencia de señalización de control dedicada entre un equipo de usuario y el nodo eNB. La figura 2.14, muestra los tres mecanismos mencionados.

En relación al envío de paquetes de usuario, cada servicio portador tiene asociado un perfil de calidad de servicio (QoS) que se satisface configurando los protocolos de radio correctamente y con la operación adecuada de los mecanismos de gestión de recursos de radio (*scheduling*).

2.5.1.5.2.3 Interfaz eNB ↔ EPC (S1)

Al plano de usuario de esta interfaz se le conoce como S1-U, y su función es proporcionar un servicio de transferencia de datos de usuario entre el nodo eNB y la pasarela S-GW sin garantizar la entrega. El plano de control conocido como S1-MME (S1-C) realiza un conjunto de funciones y procedimientos de control entre el nodo eNB y la entidad MME de la red troncal. Los procedimientos que realiza la interfaz S1 son:

- Procedimientos para establecer, modificar y liberar recursos de los servicios portadores en las interfaces de radio y S1. En la figura 2.15, se muestra el control del plano de usuario por parte de la entidad MME.
- Procedimiento de movilidad (*handover*) entre el nodo eNB. Si la red E-UTRAN decide que un teléfono debe cambiar de nodo eNB en el transcurso de una conexión, y no existe una interfaz X2 entre los nodos eNB que participan de la conexión, entonces se utilizará la interfaz S1-MME que se encargara de realizar la movilidad (*handover*).

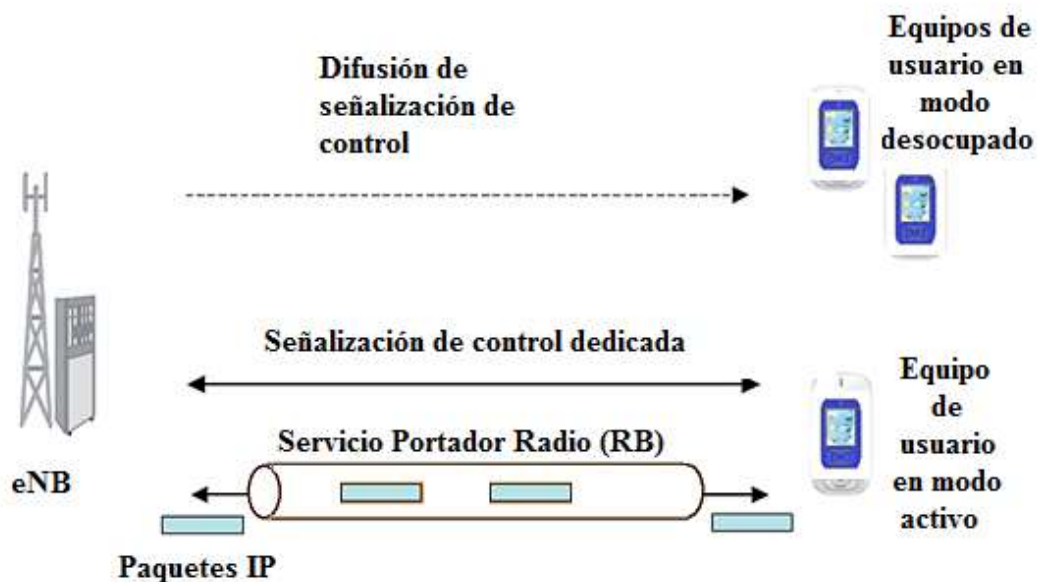


Figura 2.14. Mecanismos de transferencia de información en la interfaz radio.

- Procedimiento de aviso (*Paging*). Una de las funciones básicas de la entidad MME es la gestión de la localización de los equipos de usuario en la red. Por medio de la localización se puede conocer en que el nodo eNB se encuentra un usuario disponible (que no tenga establecida una conexión de control RRC con ningún nodo eNB).
- Procedimiento de envío de forma transparente entre la entidad MME y nodo eNB de los mensajes de señalización de control que existen entre dicha entidad y el equipo de usuario. Estos mensajes corresponden a los protocolos denominados protocolos conocidos como *Stratum Sin Acceso* (NAS por sus siglas en Inglés).

La interfaz S1 permite que un nodo eNB pueda estar conectado simultáneamente a múltiples equipos de la red troncal (MMEs y S-GWs). Esta característica se le conoce como S1-flex y ofrece las siguientes ventajas:

- Mayor seguridad de la red troncal cuando los nodos de la red fallan, pues cuando un nodo eNB falla estos pueden seguir ofreciendo servicios a través de otros nodos de la red troncal.
- Un nodo eNB puede proporcionar acceso a nodos troncales de diferentes operadores de red que comparten la red de acceso.
- Todos los teléfonos registrados en la red LTE tienen asociado un nodo MME que efectúa el seguimiento del estado de dicho teléfono.
- Está permitido el balanceo de carga entre la red E-UTRAN y los nodos de la red troncal.

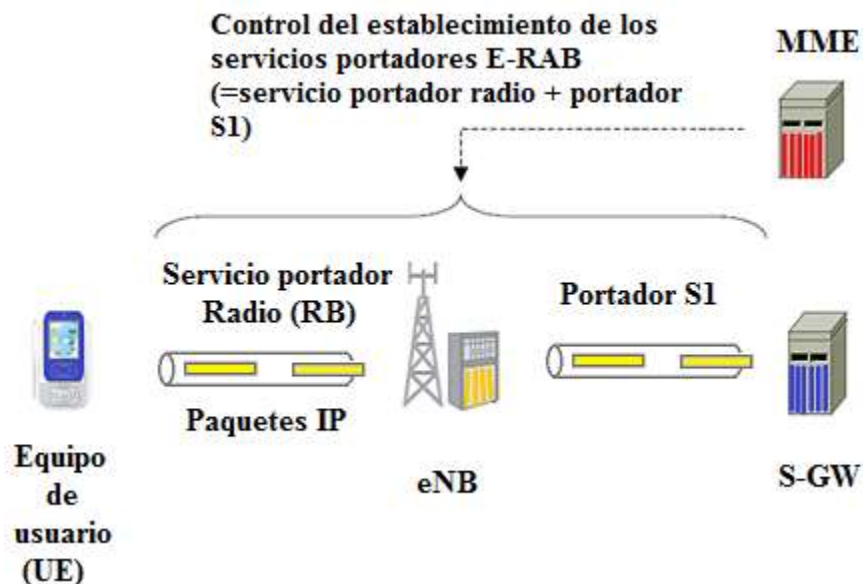


Figura 2.15. Control de los servicios portadores radio y S1 a través de la interfaz S1-MME

2.5.1.5.2.4 Interfaz eNB ↔ eNB (X2)

El plano de usuario de la interfaz X2 permite la transferencia de datos de usuario entre los nodos eNBs sin garantizar la entrega y sin soporte de mecanismos de control de errores y de control de flujo. A través de la movilidad se realiza la transferencia de datos de usuario entre los nodos eNBs. El plano de control, de las funciones y procedimientos realizados por la interfaz X2 se menciona a continuación:

- A través del plano de control se realiza la transferencia de datos de un usuario del nodo eNB anterior al actual y controla la transferencia de paquetes IP en el plano de usuario de la interfaz X2.
- Con esta interfaz se puede mostrar el *estatus* del nodo eNB (de la carga), y los nodos eNB que tengan celdas vecinas puede realizar funciones de gestión de recursos de radio.

2.5.1.6 Comparación entre las arquitecturas E-UTRAN y UTRAN

Con el fin de comparar las arquitecturas de E-UTRAN y UTRAN se presenta un gráfico de ambas arquitecturas que muestran las unidades que las constituyen. La figura 2.16, muestra ambas arquitecturas.

Las redes UTRAN tienen una arquitectura jerárquica, donde la red de acceso se constituye en dos tipos de nodos: estaciones base (denominados nodos B) y sus equipos controladores (RNC) en redes UTRAN. La arquitectura UTRAN se describe en detalle en el documento 3GPP TS 25.401 [18]. En esta arquitectura, los equipos controladores incluyen el plano de control de la interfaz de radio (señalización de control del enlace de radio) así como múltiples funciones del plano de usuario (funciones de la capa de acceso al medio, control de enlace, compresión de cabeceras, etc.). Las estaciones base realizan las

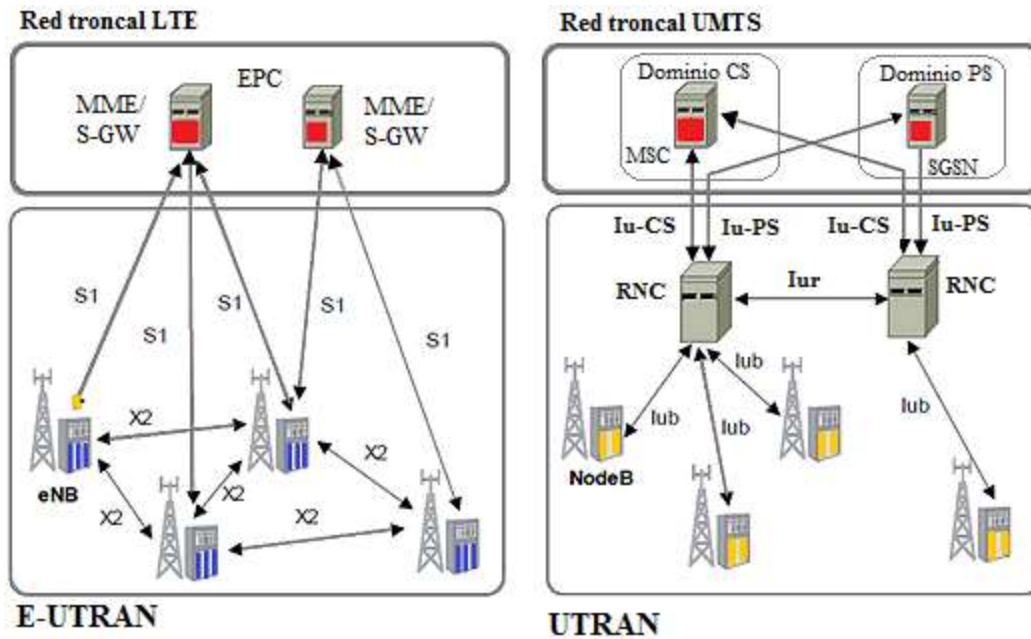


Figura 2.16. Arquitecturas de las redes de acceso E-UTRAN y UTRAN

funciones de transmisión de radio (procesado de capa física) y su operación se realiza en forma remota desde los equipos controladores. La interconexión de estaciones y controladores se realiza a través de la interfaz lub, permitiendo que la topología de red resultante a nivel lógico sea una topología de estrella. Es importante mencionar que los equipos controladores también se pueden conectar entre sí a través de la interfaz lur. La interconexión de la red de acceso a la red troncal se realiza a través de los equipos controladores utilizando interfaces lu-PS, entre controladores RNCs y los nodos SGSN del dominio de paquetes, y lu-CS, entre controladores RNCs y las centrales de conmutación MSC del dominio de circuitos. La figura 2.16, muestra que la red E-UTRAN sigue una arquitectura plana, sin niveles de jerarquía, donde los protocolos de radio se realizan en los nodos eNB (no requieren equipo adicional tal como el controlador RNC de la red UTRAN). La interconexión de la red E-UTRAN con la red troncal se lleva cabo en las estaciones base (eNB) utilizando la interfaz S1. En las redes E-UTRAN las estaciones base se pueden conectar entre ellas, permitiendo la transferencia de información de control y de tráfico de usuarios eficiente. La [19] trata sobre las redes celulares UMTS, y [20] trata de las arquitecturas de radio 4G.

2.5.1.7 Red troncal de paquetes evolucionada: (EPC)

2.5.1.7.1 Arquitectura de EPC

Cuando la organización 3GPP diseñó la red troncal EPC, incluyó un servicio de conectividad IP, que pueda explotar las capacidades que ofrece la red de acceso E-UTRAN. Así también, que tenga capacidad de poder dar acceso a las redes 3GPP (UTRAN, GERAN) y a otras redes que no pertenecen a la organización 3GPP (cdma2000, WiMAX, 802.11). La descripción completa de la red EPC se presenta en los documentos 3GPP TS 23.401 [21] y 3GPP TS 23.402 [22]. La especificación TS 23.401 presenta la

arquitectura de la red troncal EPS cuando la red de acceso es E-UTRAN, así como la utilización de redes de acceso 3GPP alternativas o complementarias como UTRAN y GERAN. Finalmente, la recomendación TS 23.402 extiende la arquitectura de la red troncal EPC para soportar el acceso a otras redes no 3GPP. La figura 2.17, muestra las unidades de red que forman el núcleo de la red troncal EPC, para proveer servicios de conectividad IP a través de una red de acceso E-UTRAN, junto con las unidades de red e interfaces que soportan las funciones relacionadas con el control del servicio de conectividad (control de calidad de servicio) y de los mecanismos de tarificación.

El núcleo del sistema EPC está constituido de las siguientes unidades: Entidad de Administración de Movilidad (MME), Pasarela de Servicio (S-GW), y Pasarela de Red de Paquete de Datos (P-GW). Estas tres unidades junto con la base de datos principal del sistema 3GPP denominada Servidor Subscriptor de Hogares (HSS por sus siglas en Inglés), constituyen los elementos básicos para la provisión del servicio de conectividad IP entre los equipos de usuario conectados a través de la red E-UTRAN y redes externas a la que se conecta la red troncal EPC. Las funciones asociadas con el plano de usuario se concentran en las pasarelas (S-GW y P-GW) mientras que la entidad MME se encarga de las funciones y señalización del plano de control. La interconexión de la red de acceso E-UTRAN a la entidad EPC se realiza a través de la interfaz S1. Particularmente, la interfaz S1-MME que maneja el plano de control termina en la entidad MME, mientras que la interfaz S1-U del plano de usuario termina en la pasarela S-GW.

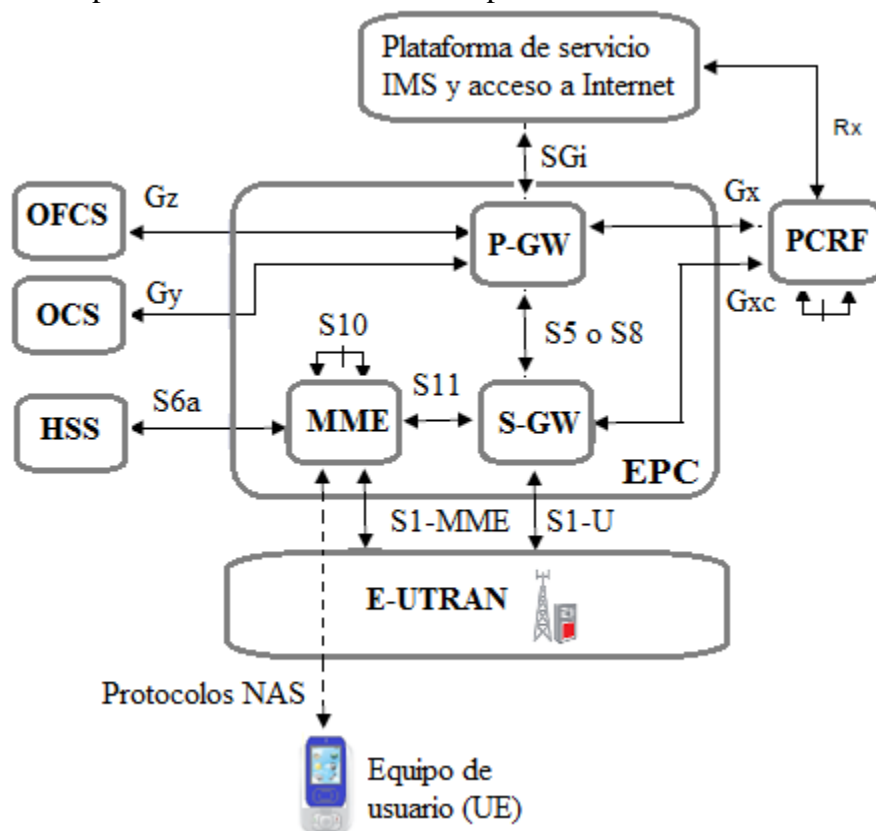


Figura 2.17. Arquitectura básica de la red troncal EPC

La entidad MME termina el plano de control de los equipos de usuario conectados a la red LTE a través de los protocolos NAS y controla las funciones de transferencia del

plano de usuario de la red LTE a través de la interfaz S11 con la pasarela S-GW. Del mismo modo, la entidad MME se conecta a la entidad HSS a través de la interfaz S6a para tener acceso a la información asociada a los usuarios de la red que estén autorizados a establecer conexiones a través de la red E-UTRAN. Las unidades MME también pueden comunicarse entre ellas a través de la interfaz S10.

La entidad P-GW permite la interconexión de la entidad EPC con redes externas o plataformas de servicio (plataformas IMS) utilizando la interfaz SGi. Así también realiza funciones de asignación de direcciones IP a los equipos de usuario y mecanismos de control de los parámetros de calidad de servicio, de las sesiones de datos que se establecieron a través de la red LTE. Internamente, la pasarela P-GW se conecta a la pasarela S-GW a través de la interfaz S5, cuando ambas pasarelas pertenecen al mismo operador, y mediante la interfaz S8, cuando estas pertenecen a redes de operadores diferentes y se ofrece un servicio de itinerancia.

La entidad conocida como Función de política y Reglas de Cargos (PCRF por sus siglas en Inglés) es parte del marco funcional denominado Control de Cargos y Políticas (PCC por sus siglas en Inglés) [23], utilizado para el control de los servicios portadores de la red LTE (activación y determinación de los parámetros de calidad de servicio asociados a cada servicio portador) y de realizar el control de la tarificación.

Tabla 2.4. Unidades de red e interfaces de EPC para el acceso desde E-UTRAN.

	Denominación	Descripción	Referencias 3GPP
Unidades de red EPC	MME	Nodo que canaliza el plano de control de la red <i>LTE</i>	TS 23.401 [5]
	S-GW	Punto de anclaje del plano de usuario en la red Troncal	TS 23.401 [14]
	P-GW	Pasarela para la interconexión con redes externas	TS 23.401 [14]
Unidades comunes a las redes 3GPP	Denominación	Descripción	
	HSS	Base de datos global del sistema (contiene, entre otros, los datos de suscripción de los usuarios).	TS 23.002 [1] TS 23.008 [16]
	PCRF	Elemento central del sistema de control de uso de la red (política de control) y control de tarificación	TS 23.203 [27]
	OCS	Nodo de control para la provisión de servicios que requieren tarificación	TS 23.203 [27] TS 32.240 [28]

		online	
	OFCS	Nodo de recogida de la información de tarificación para su posterior transferencia al sistema de facturación.	TS 23.203 [27] TS 32.240 [28]
Interfaces	Denominación	Unidades de red asociadas	
	S1-MME	MME E-UTRAN (eNB)	Documentos TS 36.41x
	S1-U	S-GW E-UTRAN (eNB)	TS 29.281 [26]
	SGi	P-GW Redes externas	TS 29.061 [17]
	S6a	<u>MME HSS</u>	TS 29.272 [25]
	S5/S8	<u>P-GW S-GW</u>	TS 29.274 [21] (opción GTP) TS 23.275 [24] (opción PMIPv6)
	S11	<u>MME S-GW</u>	TS 29.274 [21]
	S10	<u>MME MME</u>	TS 29.274 [21]
	Señalización NAS	<u>UE MME</u>	TS 24.301 [37] [37]
	Rx	PCRF Plataformas servicios	TS 29.214 [29]
	S9	PCRF PCRF	TS 29.215 [30]
	Gx/Gxc	P-GW/S-GW PCRF	TS 29.212 [31]
	Gz/Gy	P-GW OFCS/OCS	Documentos TS 32.2xx

De esta manera con la interfaz Gx, el servidor PCRF gestiona los servicios portadores de la red LTE. La entidad PCRF es accesible desde las plataformas externas IMS a través de la interfaz Rx. Por ejemplo, si un usuario establece un servicio de videoconferencia a través del subsistema IMS, el elemento que controla la provisión del servicio en el IMS puede indicar a través de la interfaz Rx cuáles son los parámetros de calidad de servicio que debe proporcionar el servicio portador de la red LTE para transferir correctamente la información de la videoconferencia.

Las Unidades Sistema de Carga Sin-conexión (OFCS por sus siglas en Inglés) y el Sistema de Carga con-Conexión (interactúan con la pasarela P-GW a través del servicio CS) representan el núcleo del sistema de tarificación de la red [24]. Las dos unidades interactúan directamente con la pasarela P-GW a través de la interfaz Gz (caso OFCS) y Gy (caso OCS).

La tabla 2.4, muestra las unidades de red e interfaces propias de la entidad EPC, así como otras unidades comunes de los sistemas 3GPP y sus especificaciones correspondientes.

2.5.1.8 Unidades de red y sus interfaces

En esta sesión se describirá con detalle las funciones soportadas por las tres Unidades de red que constituyen el núcleo de la red troncal EPC (MME, S-GW y P-GW) y la entidad HSS común al resto de sistemas 3GPP, así mismo se describirán las interfaces más importantes SGi, S5/S8, S11, S10 y S6a.

2.5.1.8.1 Unidad de administración de movilidad (MME)

La entidad MME es el elemento principal del plano de control de la red LTE, y su función es gestionar la interconexión de los teléfonos y/o terminales a través de la red E-UTRAN. Es importante mencionar que todo teléfono que se encuentre registrado en la red LTE y sea accesible a través de la red E-UTRAN tiene una entidad MME asignada, de otro modo no puede comunicarse. Cuando el teléfono se registra en la red se elige la unidad que lo va a gobernar, esto dependerá de su ubicación geográfica del teléfono en la red (cada MME atiende a un conjunto de nodos eNB). Esta unidad contiene los datos de los usuarios (tales como identificadores de usuario, conexiones y servicios portadores EPS activos, claves de seguridad, datos de localización del usuario en la red, etc.) y realiza todas las gestiones relacionadas a dicho usuario (establecimiento de servicios portadores EPS, etc.). La unidad MME asignada a un usuario permite la movilidad de dicho usuario dentro de la zona de servicio de la red. Las funciones principales de la entidad MME son las siguientes:

- Autenticar y autorizar el acceso de los usuarios a través de la red E-UTRAN. Utilizando los datos de usuario adquiridos desde el servidor HSS, el módulo MME controla el acceso a la red a través de la identificación, autenticación y autorización de los usuarios que se conectan a través de la red E-UTRAN.
- Gestión de los servicios portadores EPS. La entidad MME administra la señalización requerida para establecer, mantener, modificar y liberar los servicios portadores EPS sobre los cuales se realiza el envío de paquetes IP entre los equipos de usuario y la red externa.
- El módulo MME es el encargado de dar el seguimiento de la ubicación de los usuarios dentro del área de servicio de la red. Para lograr esto, se definen las áreas de seguimiento (*tracking areas*) y unos procedimientos asociados (*Tracking Update*) que permiten disponer de información de localización de todos los usuarios que se encuentren registrados en la red LTE. Gestión de movilidad de los usuarios en el modo desocupado (*idle*) (teléfonos que no tienen ninguna conexión de control establecida con la red E-UTRAN).
- Señalización para el soporte de movilidad entre el sistema EPS y redes 3GPP. La unidad MME de la red principal EPC y el nodo SGSN intercambian información relativa a los equipos de usuario conectados a través de la red E-UTRAN o de las redes UTRAN/GERAN para operar, mecanismos de movilidad conjunta (la red troncal GPRS así como la red troncal EPC pueden intercambiar información relativa a las áreas de seguimiento).
- Terminación de los protocolos NAS. Los protocolos NAS operan a través del equipo de usuario y la unidad MME que tenga asignada. A través de estos protocolos se controlan las funciones de control de acceso a la red LTE, la gestión de las conexiones a redes externas y el establecimiento de los servicios

portadores EPS, y la gestión de movilidad de los teléfonos que se encuentran en el modo desocupado (*idle*).

2.5.1.8.2 Pasarela de servicio (S-GW)

La entidad *Pasarela de Servicio* (S-GW por sus siglas en Inglés) es la pasarela del plano de usuario entre la red E-UTRAN y la red troncal EPC. Como sucede con la entidad MME, un usuario suscrito en la red LTE dispone de una entidad S-GW asignada en la EPC a través de la cual trabaja su plano de usuario. La idea de asignar una pasarela S-GW es con el fin de operar criterios geográficos y el balanceo de cargas. Las funciones principales de la pasarela S-GW son:

- Ofrecer un punto de anclaje en la red troncal EPC con respecto a la movilidad del teléfono entre los nodos eNBs. Para que durante un proceso de movilidad (*handover*) entre dos nodos eNBs, el cambio del plano de usuario únicamente afecte el cambio del servicio portador S1 entre los nodos eNBs implicados y la pasarela S-GW, manteniéndose sin cambios el resto del plano de usuario (camino entre las pasarelas S-GW y P-GW).
- La funcionalidad de punto de anclaje también se aplica a la gestión de movilidad con las otras redes de acceso 3GPP (UTRAN y GERAN). Permitiendo a los equipos de usuario que se conecten a la red LTE a través de las redes UTRAN o GERAN, dispongan también de una pasarela S-GW asociado en la red troncal EPC por el que fluye su plano de usuario.
- Almacenamiento temporalmente de paquetes IP de usuarios cuando los teléfonos se encuentren en modo desocupado (*idle*). En una red LTE, el plano de usuario entre la pasarela S-GW y el equipo de usuario puede “desactivarse” cuando no haya tráfico para transmitir.
- Cuando se recibe tráfico de la red externa a un usuario en modo desocupado (*idle*), este tráfico llega hasta la pasarela S-GW a cargo de ese usuario, que almacena temporalmente los paquetes IP e el establecimiento del plano de usuario.
- Una pasarela S-GW, almacena la información y las funciones de encaminamiento necesarias para dirigir el tráfico de subida (tráfico IP proveniente de los equipos de usuario) hacia la pasarela P-GW que corresponda y el tráfico de bajada (proveniente de las pasarelas P-GW) hacia el nodo eNB a través del cual se encuentra conectado el equipo de usuario.

2.5.1.8.3 Pasarela PDN (P-GW)

La unidad PDN es responsable de proporcionar conectividad entre la red LTE y las redes Externas; denominadas como Redes de Paquetes de Datos (PDN por sus siglas en Inglés). Es decir, a través de la unidad P-GW, un usuario conectado al sistema LTE resulta “visible” en la red externa. Por tanto, los paquetes IP generados por el usuario se envían la red externa a través de esta pasarela y, viceversa, todo el tráfico IP dirigido a un teléfono

LTE proveniente de la red externa va a ser encaminado hasta la pasarela P-GW. Funciones principales de la pasarela P-GW:

- Maneja el control de políticas y el control de tarificación a los servicios portadores que tengan los teléfonos.
- Asignación de la dirección IP a los teléfonos en una determinada red externa. La dirección puede ser una dirección IPv4, IPv6 o bien un par de direcciones (IPv4, IPv6).
- La pasarela P-GW actúa como punto de anclaje para la gestión de movilidad entre la red LTE y redes no 3GPP. La pasarela guarda funciones de Agente de Casa (HA por sus siglas en Inglés) para proporcionar continuidad de servicio en caso de utilizar el protocolo *Mobile IPv4* (MIPv4) [25] para gestionar la movilidad entre la red LTE y, por ejemplo, una red WiMAX. Además del protocolo MIPv4, la pasarela incluye soporte de movilidad para los protocolos *Dual Stack MIPv6* (DSMIPv6) y *Proxy MIPv6* (PMIPv6).
- El tráfico IP que viaja por la pasarela P-GW es procesado a través de un grupo de filtros que asocian cada paquete IP con el usuario y servicio portador EPS correspondiente.

2.5.1.8.4 Servidor subscriptor de usuarios (HSS)

El servidor HSS se utiliza como la base de datos del sistema principal 3GPP que almacena la información de los usuarios de la red. Contiene información de los usuarios, así como información necesaria para la operación propia de la red. Esta base de datos es consultada y modificada desde las diferentes Unidades de red encargadas de ofrecer los servicios de conectividad o servicios finales (desde la entidad MME de la red troncal, el nodo SGSN de la red GPRS, la entidad MSC del dominio de circuitos y también desde servidores de control del sub-sistema IMS). El servidor HSS contiene información fija que únicamente puede ser modificada a través de procesos administrativos, así como información temporal tal como localización del teléfono dentro de la zona de servicio del sistema. Las informaciones más importantes que almacena el servidor HSS son: identificadores universales del usuario (Identificador de Subscriber Móvil Internacional, IMSI), identificadores de servicio (Unidad Móvil ISDN, MSISDN); información de seguridad y cifrado (vectores de autenticación); información de localización del usuario en la red (identificación de la entidad de control, por ejemplo, la entidad MME, que proporciona el plano de control hacia un determinado usuario); e información necesaria para la provisión de los servicios de acuerdo con las condiciones establecidas en el contrato de suscripción (identificador de la red externa y parámetros de calidad de servicio del servicio portador). El servidor HSS se estandarizó por la organización 3GPP R5 tomando como base la integración de dos Unidades definidas en las redes GSM denominadas Registro de Localización de Hogar (HLR por sus siglas en Inglés) y el Centro de Autenticación (AuC por sus siglas en Inglés), a las cuales se le añadieron funciones adicionales para soportar el acceso y la operativa del sistema LTE.

- En la versión (*Release 8*), que corresponde al sistema LTE, el servidor HSS incluye:

- Las funciones de los servidores HLR/AuC necesarias para el funcionamiento del dominio de paquetes EPC, así como las redes GPRS.
- El acceso a los servidores HSS desde la red EPC se realiza desde la entidad de red MME mediante la interfaz S6a.
- El subconjunto de funciones de las unidades HLR/AuC necesarias para el funcionamiento del dominio de circuitos CS.
- Funciones de soporte asociadas a las funciones de control del subsistema IMS y el almacenamiento de perfil de usuario asociado a servicios multimedia.

La información almacenada en el servidor HSS se describe en la especificación de la organización 3GPP TS 23.008 [26]. Las unidades de red que se conectan a la base de datos del servidor HSS para gestionar el acceso al servicio de conectividad de la red troncal EPC son las siguientes: La unidad MME interactúa con la base de datos a través de la interfaz S6a (cuando el acceso se realiza a través de redes E-UTRAN). Y cuando el acceso es a través de las redes UTRAN o GERAN, la comunicación al servidor HSS se realiza desde el Servidor *GPRS Support Node* (SGSN) a través la interfaz S6d. Cuando la comunicación es a través de redes no 3GPP, el acceso se canaliza a través del servidor AAA utilizando la interfaz SWz.

2.5.1.8.5 Interfaz P-GW ↔ redes externas (SGi)

A través de la interfaz SGi se realiza la interconexión de la pasarela P-GW de la red LTE con redes externas IP. La red externa puede ser tanto una red pública (Internet) como cualquier otra red privada (intranet corporativa, red de un ISP, red interna del propio operador para la provisión, por ejemplo, de servicios IMS). La interfaz SGi es equivalente a la interfaz Gi especificada para la interconexión de la pasarela GGSN del dominio GPRS con redes externas [27]. La interfaz SGi soporta la interconexión tanto a redes IPv4 como

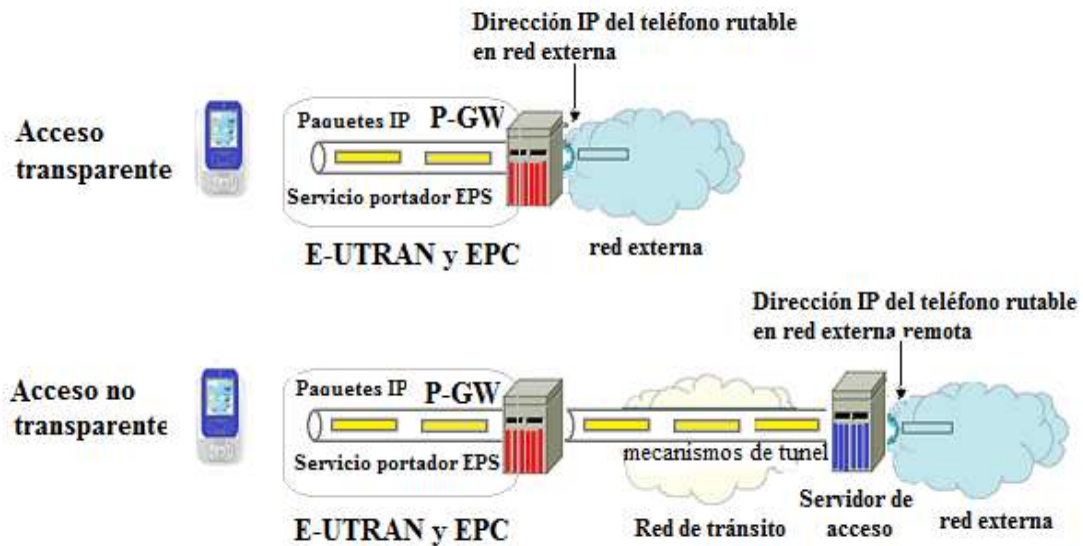


Figura 2.18. Tipos de interconexión a través de la interfaz SGi.

IPv6. Desde la perspectiva de la red externa, la pasarela P-GW es vista como un ruteador IP convencional. Sobre esta base, existen dos modelos básicos de interconexión de la red LTE con la red externa: acceso transparente y acceso no transparente. Ambos modelos se muestran en la Figura 2.18.

Bajo el modelo de interconexión transparente, la dirección IP asignada al teléfono es válida en la propia interfaz SGi con la red externa, de forma que el teléfono es “visible” en la red externa a la que proporciona acceso la pasarela P-GW a través de dicha dirección. En cambio, en el modelo no transparente, la red LTE ofrece un acceso a una red externa remota de forma que el espacio de direcciones utilizado por los teléfonos pertenece al espacio de direcciones de la red externa remota. La conexión entre la red LTE y la red remota admite diferentes soluciones tales como el establecimiento de un túnel IP (IPsec, GRE, etc.) entre la pasarela P-GW y el servidor de acceso remoto. En el modelo no transparente, puede ser necesario que la pasarela P-GW participe en funciones de autenticación del usuario y asignación de direcciones dinámicas pertenecientes a la red remota

2.5.1.8.6 Interfaz P-GW ↔S-GW (S5 y S8)

Las interfaces S5 y S8 se utilizan para la transferencia de paquetes de usuario entre las pasarelas S-GW y P-GW. La interfaz S5 actúa cuando ambas pasarelas pertenecen a la misma red, mientras que la interfaz S8 es utilizada en caso de escenarios de itinerancia donde la pasarela S-GW pertenece a la red visitada (conocida como red PLMN visitada). Ambas interfaces aceptan dos implementaciones diferentes: Una basada en el protocolo GTP y otra basada en el protocolo PMIPv6. La implementación basada en el protocolo de tunelización GPRS (GTP por sus siglas en Inglés) ofrece funciones de creación, eliminación, modificación, y cambio del servicio portador del plano de usuario entre las pasarelas S-GW y P-GW de los usuarios conectados a la red LTE. En el caso de la implementación basada en el protocolo PMIPv6, no se soporta ninguna gestión de servicios portadores entre las pasarelas P-GW y S-GW ya que, estrictamente el servicio portador EPS de la red LTE se extiende desde el equipo de usuario hasta la pasarela S-GW (y no hasta la pasarela P-GW). Esto se debe a que el protocolo PMIPv6 está diseñado para ofrecer exclusivamente un servicio de movilidad entre las pasarelas S-GW y el P-GW y no dispone de los mecanismos necesarios para señalar parámetros de calidad de servicio. Por lo tanto las interfaces S5 y S8 básicamente ofrecen servicio de conectividad entre las pasarelas implicadas de forma que todos los flujos de datos son llevados a través del mismo túnel (independientemente de los servicios portadores). La descripción funcional de la interfaz S5/S8 basada en el protocolo GTP se describe en la recomendación TS 23.401 [21].

2.5.1.8.7 Interfaz MME↔ S-GW (S11)

Esta interfaz controla la operación del plano de usuario en la red troncal EPC desde la entidad de red MME. Es decir, controla los procedimientos soportados en esta interfaz permitiendo la creación, eliminación, modificación, y cambio de los servicios portadores que los teléfonos tienen establecidos a través de la red troncal LTE. En este

sentido, dado que la entidad MME es la entidad de control que termina los protocolos NAS con los usuarios, la interfaz S11 permite establecer el enlace entre el plano de control con las funciones del plano de usuario de la red troncal LTE. Las acciones a realizar sobre la interfaz S11 tienen su origen en los diferentes eventos relacionados con la señalización entre el terminal y la red troncal (registro de un teléfono en la red LTE, incorporación de un nuevo servicio portador EPS a una conexión, establecimiento de una nueva conexión con otra red externa, indicación de movilidad [*handover*], etc.) así como en eventos originados desde la propia red troncal EPC (notificación de que existe tráfico almacenado en la pasarela S-GW pendiente de ser enviado a un teléfono en modo desocupado (*idle*), modificación de servicios portadores iniciada por la pasarela P-GW, etc.). La interfaz S11 también da soporte al proceso de reubicación de la pasarela S-GW asociada a un teléfono mediante la transferencia de contextos entre la pasarela antigua y la nueva. Asimismo, durante la ejecución de un procedimiento de movilidad que implique una reubicación de la pasarela S-GW, el nodo MME controla los mecanismos de transferencia de paquetes entre las dos posibles pasarelas involucradas. La funcionalidad de esta interfaz se encuentra en las recomendaciones de la organización 3GPP TS 23.401 [18] y el protocolo GTPv2-C que da soporte a esta interfaz se especifica en TS 29.274 [28].

2.5.1.8.8 Interfaz MME↔ MME (S10)

La Interfaz S10 se define entre dos unidades MME. Su principal función es el soporte del mecanismo de reubicación de la entidad MME. Es decir, cuando la entidad MME que controla a un determinado equipo de usuario debe cambiarse (debido, por ejemplo, a su movilidad), la interfaz S10 realiza la transferencia del contexto de dicho usuario entre entidades MMEs. En este caso, el contexto de un teléfono en una entidad MME abarca toda la información relacionada con la gestión de los servicios portadores (e.g., configuración de las conexiones activas) y la gestión de movilidad (área de seguimiento donde se encuentra localizado el teléfono), junto con otra información de seguridad (claves de cifrado), características del teléfono, parámetros de suscripción del usuario obtenidos desde el servidor HSS, etc. La funcionalidad de esta interfaz se describe en la recomendación TS 23.401 [18] y el protocolo GTPv2-C que da soporte a esta interfaz se especifica en TS 29.274 [28].

2.5.1.8.9 Interfaz HSS↔ MME (S6a)

Esta interfaz permite la transferencia de información entre la base de datos HSS y la entidad del plano de control MME de la red troncal EPC. A través de la interfaz S6a se da soporte a las siguientes funciones:

- Mantenimiento de información de gestión de la localización. La base de datos del servidor HSS mantiene unos campos que contienen la identificación del nodo MME que controla a cada usuario registrado en la red. Esta información la actualiza el nodo MME correspondiente a través de la interfaz S6a. Esta información permite que cuando un teléfono se conecta a una entidad MME nueva, dicha entidad MME puede recuperar información relativa al nodo MME que previamente dio servicio al teléfono para realizar la reubicación pertinente.

- Autorización de acceso a la red LTE. La base de datos del servidor HSS almacena los datos de suscripción de los usuarios que condicionan el acceso a los servicios que ofrece la red. El perfil de suscripción de un usuario se transfiere desde el HSS al nodo MME, que es la entidad encargada de ejecutar las comprobaciones pertinentes. Así, a modo de ejemplo, la entidad MME puede autorizar o no la conexión de un usuario a una red externa en función de que dicha red externa forme parte de la lista de redes externas permitidas de acuerdo a la suscripción del usuario.
- Autenticación de los usuarios. A través de la interfaz S6a, de la entidad MME se descarga la información que permite llevar a cabo el procedimiento de autenticación desde la entidad MME. Esta información se denomina vector de autenticación EPS.
- Notificación y descarga de la identidad de la pasarela P-GW que utiliza un usuario en una conexión. El almacenamiento en la base de datos HSS de información relativa a las pasarelas P-GW que dan servicio al usuario se utiliza para proporcionar soporte a los mecanismos de movilidad entre la red LTE y otras redes no 3GPP.

La interfaz también soporta escenarios de itinerancia donde una entidad MME, donde desde la red de un operador puede acceder a la base de datos HSS de otro operador. La interfaz S6a se basa en el protocolo *Diameter*. La funcionalidad de la interfaz se detalla en TS 23.401 [18] y la especificación de la extensión (i.e., *Diameter application*) del protocolo se detalla en la recomendación TS 29.272 [29].

2.5.1.9 Configuraciones de la red EPC

La red troncal EPC cuenta con cuatro posibles implementaciones, basadas a la ubicación física de las tres principales unidades de red que la constituyen: MME, S-GW y P-GW. La figura 2.19, muestra las posibles implementaciones de la red troncal. Una implementación posible de la red troncal EPC consiste en integrar las tres unidades funcionales en un único equipo de red. En esta opción el número de saltos o puntos de procesamiento del plano de usuario en la red LTT sea únicamente de dos (NB y equipo de la red troncal S-GW+P-GW), con una mejora en los términos de latencia. Sin embargo, esta configuración no permite dimensionar por separado los recursos necesarios para soportar el plano de control y el de usuario de forma que el número de equipos de red troncal necesarios debe contemplar el peor de los casos. Debido a que el dimensionamiento de los recursos del plano de control depende principalmente del número de usuario, mientras que el dimensionamiento del plano de usuario está asociado al volumen de tráfico, la proporción entre el número de recursos necesarios para soportar ambos planos puede abarcar un amplio rango de valores atendiendo a la relación y evolución del número de usuarios y del volumen de tráfico que genera cada usuario en la red. Similarmente, el montante de recursos necesarios para soportar la funcionalidad de punto de anclaje del plano de usuario (S-GW) y la funcionalidad de pasarela con redes externas (P-GW) tampoco guarda una relación de proporcionalidad clara. El número de equipos de red que alberguen la funcionalidad de la pasarela P-GW puede depender en gran medida del número y tipología de las redes externas a las que se debe proporcionar servicio mientras

que el número de equipos S-GW está vinculado más directamente al número de usuarios y distribución geográfica de la red de acceso. Por tanto, implementaciones de equipos que alberguen ambas funciones, tal como sería el caso de la opción (A), limitan claramente la versatilidad y escalabilidad del sistema.

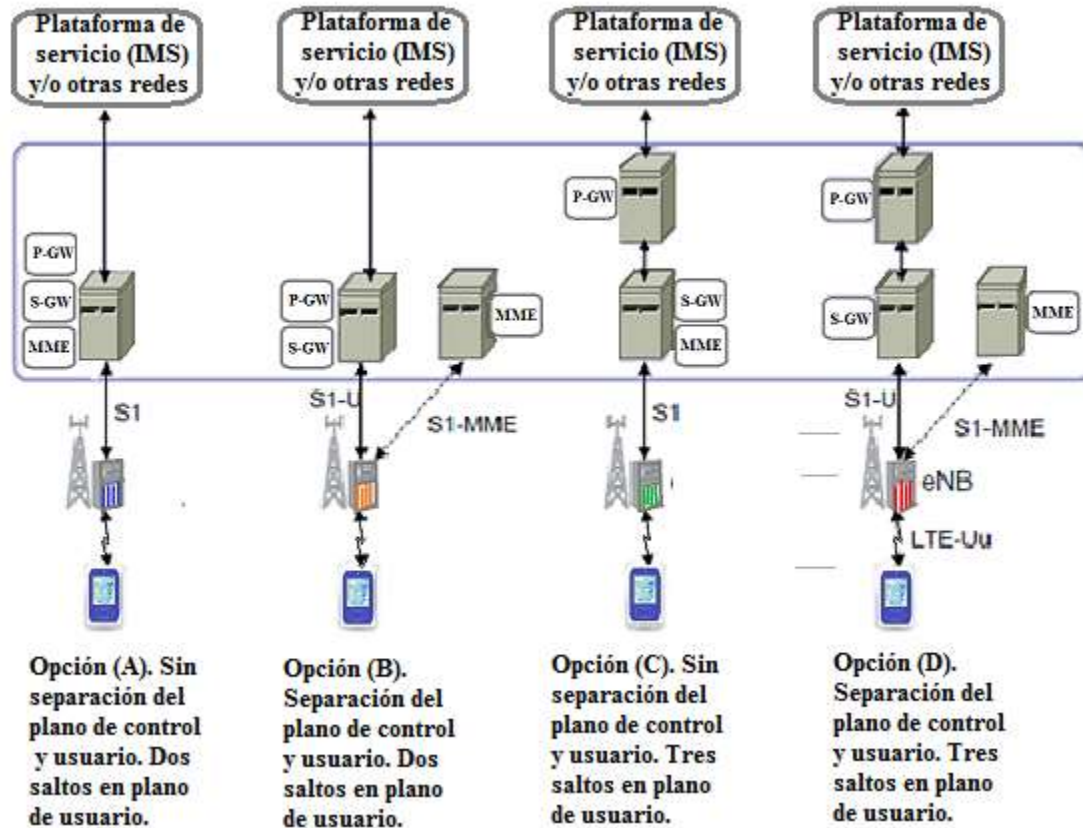


Figura 2.19. Posibles configuraciones de la red EPC.

2.5.1.10 Soporte de itinerancia (*roaming*)

Una función importante de las redes de comunicaciones móviles es el servicio de itinerancia (*roaming*). Este servicio permite que los usuarios puedan acceder a sus servicios de telecomunicación a través de las redes de otros operadores con los que no tienen establecida ninguna relación de contrato (suscripción). El operador con el que el usuario tiene establecida la relación contractual para la prestación de servicios se conoce como operador matriz, y a la red de dicho operador se le llama la red matriz (*Home Network*). La red de otro operador a la que el usuario puede tener acceso se denomina red visitada. El sistema LTE especifica tres posibles configuraciones para la implementación de un servicio de itinerancia. Las diferentes configuraciones dependen de qué pasarela P-GW se utiliza para encaminar el tráfico con la red externa y de la capacidad de proporcionar acceso a los servicios propios del operador matriz. A continuación se mencionan las tres configuraciones son:

- Encaminamiento de tráfico a través de la red matriz, con acceso a los servicios de la red matriz.

- b. Encaminamiento de tráfico a través de la red visitada, con acceso a los servicios de la red visitada.
- c. Encaminamiento de tráfico a través de la red visitada, con acceso a los servicios de la red matriz.

La arquitectura de red de la opción (a) se muestra en la Figura 2.20. Se observa que el nodo MME en la red visitada responsable de la terminación del plano de control con la terminal (protocolos NAS) accede a la base de datos HSS de la red matriz (a través de la interfaz S6a) para poder obtener la información necesaria del usuario en itinerancia.

De esta forma, el usuario en itinerancia puede autenticarse en la red visitada a partir de las credenciales que le han sido otorgadas por su operador matriz. Por otro lado, el establecimiento de los servicios de conectividad se realiza mediante la utilización de

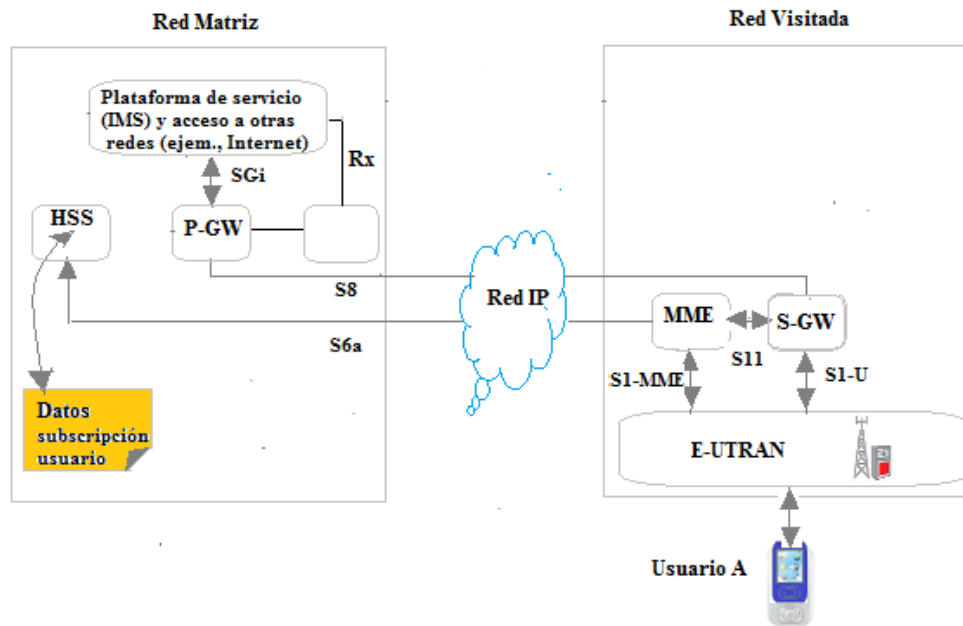


Figura 2.20. Encaminamiento de tráfico a través de la red matriz (itinerancia)

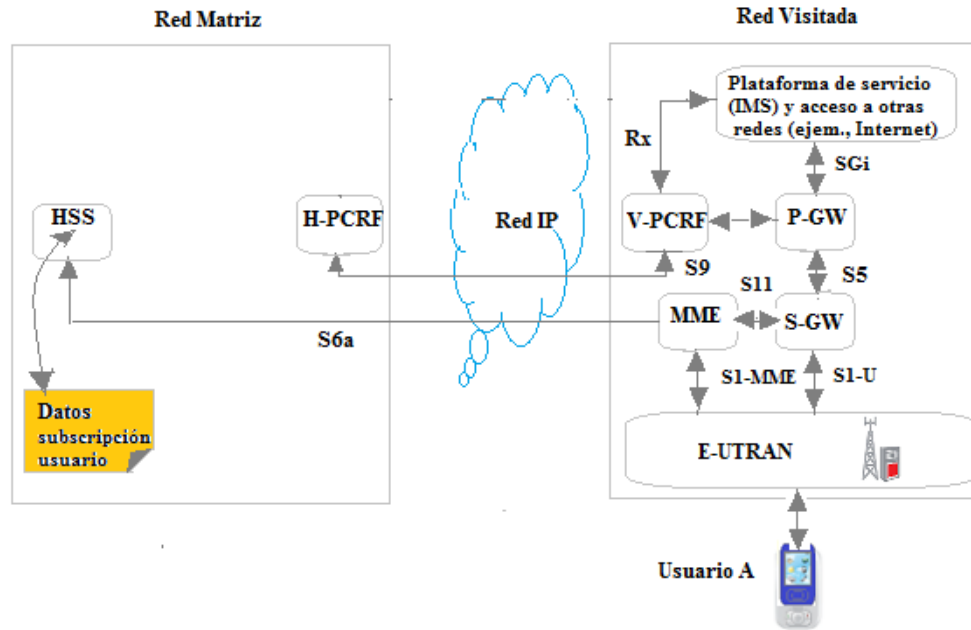


Figura 2.21. Encaminamiento de tráfico a través de la red visitada, con acceso a los servicios de red visitada (itinerancia)

pasarelas P-GW del operador matriz. Para ello, se establece un túnel (a través de la interfaz S8) entre la pasarela S-GW que actúa de punto de anclaje en la red visitada y la pasarela P-GW que proporciona la interconexión con la red externa en la red matriz. Esta configuración permite que el usuario acceda a todos sus servicios como si estuviera conectado a través de la red de acceso de su operador matriz.

La Figura 2.21, muestra la arquitectura de la red para la opción (b). Esta configuración permite que el tráfico generado por los usuarios en itinerancia se atienda de forma local en las redes visitadas. Así, el acceso a las redes externas y/o plataformas de servicio se realiza mediante pasarelas P-GW pertenecientes a la red visitada. En cualquier caso, nótese como, al igual que en la opción (a), el nodo MME de la red visitada accede directamente a la base de datos HSS de la red matriz para obtener la información relativa al usuario en itinerancia. Asimismo, el control de las reglas de uso de la red y de tarificación, atendiendo a que cada operador puede establecer sus propias estrategias comerciales, puede realizarse en base al acceso al sistema PCC de la red matriz. De esta forma, las reglas de uso (parámetros de calidad de servicio) que se aplicarían en la red visitada vendrán determinadas por el operador matriz. Las reglas de uso pueden transferirse desde la red matriz de un usuario a la red visitada donde está recibiendo servicio a través de la interfaz S9 diseñada para tal efecto.

Finalmente, la arquitectura correspondiente a la opción (c) se muestra en la figura 2.22. Esta resulta útil en el caso de que se pretenda que el usuario en itinerancia tenga acceso a las plataformas de servicio de su operador (el control de los servicios se realiza a través del operador matriz) pero evitándose el encaminamiento de todo el tráfico a través de la red matriz. De esta forma, el tráfico de usuario que no necesariamente tenga que atravesar la plataforma de servicios puede acceder a las redes externas (Internet) sin necesidad de que tenga que ser transferido desde la red visitada a la red matriz.

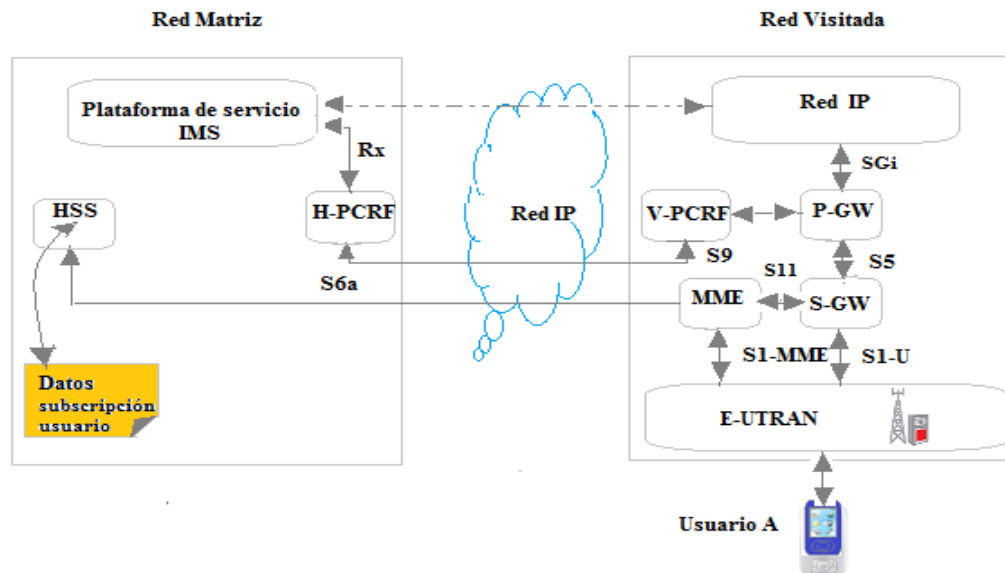


Figura 2.22. Encaminamiento de tráfico a través de la red visitada, con acceso a los servicios de la red matriz.

2.6. Subsistema Multimedia (IMS)

El subsistema Multimedia basado en el Protocolo de Internet (IMS) se encarga de los mecanismos de control para proveer servicios de comunicación multimedia basados en la utilización del protocolo IP a los usuarios de la red LTE. Así, este subsistema IMS es el corazón de las redes de siguiente generación por su diseño, el cual permite manejar tráfico multimedia de alta velocidad con confiabilidad. Su arquitectura está implementada a base de servidores SIP, bases de datos, pasarelas de acceso y de red que se comunican entre sí mediante los protocolos SIP, SDP y *Diameter*, para gestionar la provisión de servicios de voz y video sobre IP, presencia y mensajería instantánea, servicios de llamadas en grupo, etc. La forma de acceso de parte de las terminales inteligentes y/o teléfonos al subsistema IMS se realiza a través de los servicios de conectividad IP que proporciona la red LTE. Este sistema de comunicaciones tiene la capacidad de interconectar redes de circuitos conmutados y de paquetes. Fue diseñado originalmente para conectar las comunicaciones móviles, sin embargo sus diferentes soluciones tecnológicas, han permitido que puedan conectarse las redes PSTN, ISDN, GSM, GPRS, *Mobile WiMAX*, y finalmente *LTE Advanced* (E-UTRAN). Ha adoptado protocolos desarrollados por la asociación de ingenieros del IETF para la provisión de servicios multimedia en sistemas de comunicaciones móviles, los cuales permiten reducir considerablemente el ciclo de desarrollo de las terminales y equipos inteligentes que utilizarán los diferentes usuarios para comunicarse entre sí. La organización 3GPP escogió el Protocolo de Iniciación de Sesión (SIP) definido en la recomendación RFC3261 del grupo IETF como protocolo de base para soportar la señalización asociada al subsistema IMS. El subsistema IMS [31] tiene sus orígenes en el conjunto de especificaciones correspondientes a la revisión (*Release 5*) del sistema UMTS. El modelo de red NGN basado en el subsistema IMS se divide en cuatro capas: acceso, transporte, control y aplicación, tal como se muestra en la figura 24. La capa de transporte representa la infraestructura de red IP, dependiente de la

tecnología de acceso, que proporciona el encaminamiento de los paquetes IP entre las terminales y demás elementos de la red. Por tanto, el servicio de

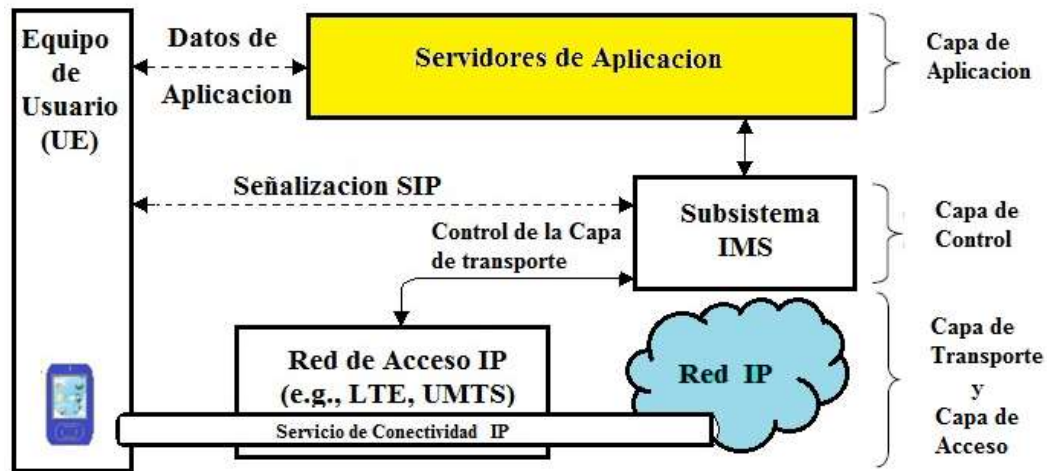


Figura 2.23. Arquitectura simplificada del modelo de red NGN

conectividad IP que ofrece la red LTE constituye claramente una realización válida de la capa de transporte en este modelo de provisión de servicios. En la capa de control se ubican los elementos especializados en la gestión de sesiones tales como los servidores de señalización SIP, así como otros elementos específicos para la interacción con redes telefónicas convencionales (pasarelas VoIP, controladores, etc.). Esta capa de control contiene un subsistema IMS como el mostrado en la figura 2.23, donde los servicios de conectividad IP proporcionados por la capa de transporte se controlan desde esta capa. Por último, en la capa de aplicación residen los servidores que operan la lógica y datos asociados a los diferentes servicios proporcionados a través de subsistema IMS (servicio de mensajería instantánea y presencia, etc.). En la capa de aplicación también se pueden encontrar elementos ligados a otras plataformas de servicios como redes inteligentes y pasarelas Parlay/OSA [40]. A través de estas plataformas se posibilita la provisión de servicios desde proveedores de aplicaciones externos, denominados como Proveedores de Servicio de Aplicación (ASP).

La puesta y liberación de llamadas a través del subsistema IMS se realiza utilizando el Protocolo de Iniciación de Sesión (SIP por sus siglas en Inglés). Este protocolo se concibió en un principio para controlar el establecimiento y liberación de sesiones multimedia (telefonía, videoconferencia, etc.) sobre redes IP entre dos o más participantes. Por su flexibilidad, se utiliza en una gran cantidad de aplicaciones mucho más extensa tales como: mensajería instantánea y presencia, juegos distribuidos, control remoto de dispositivos, etc. Además del protocolo SIP, en el subsistema IMS se emplean otros protocolos de señalización, tales como *Diameter*, y *MEGACO/H.248*, para escenarios de comunicación en los que intervienen pasarelas de medios (por ejemplo, pasarelas VoIP para interconectar redes IP con redes telefónicas conmutadas convencionales). Las especificaciones de la arquitectura funcional del subsistema IMS se encuentran en el documento 3GPP 23.228 [31]. La figura 2.24, muestra una arquitectura de las unidades del subsistema IMS. El núcleo del subsistema IMS lo forman los

servidores llamados Servidores de Llamado (CSCF por sus siglas en Inglés), y proporcionan las siguientes funciones:

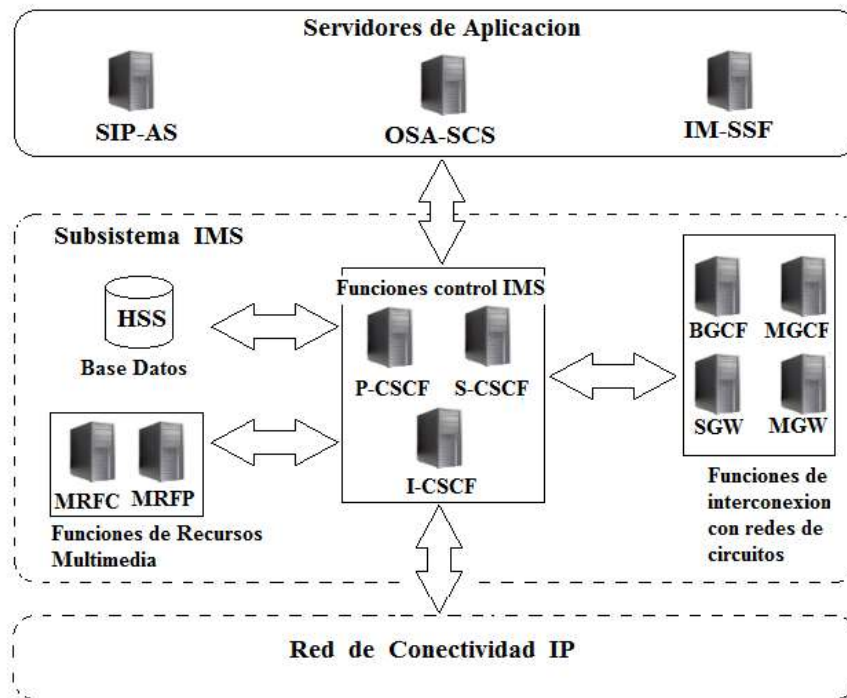


Figura 2.24. Arquitectura simplificada del subsistema IMS

- Servidor *Serving* (*S-CSCF*) actúa como el nodo central de la señalización en la puesta de sesiones *IMS*. El servidor *S-CSCF* actúa como servidor de registro *SIP* de forma que, cualquier mensaje de señalización *SIP* dirigido al usuario mediante la dirección *SIP* correspondiente, siempre termina pasando por un servidor *S-CSCF* donde el usuario se encuentra registrado. El servidor *S-CSCF* puede proporcionar por si solo algunos servicios al usuario tales como: servicio de re-direccionamiento de llamadas, listas de marcado, etc., o bien encaminar la señalización *SIP* hacia los servidores de aplicación tales como: servidores que implementan un servicio de mensajería instantánea.
- Servidor *Proxy* (*P-CSCF*) es un servidor *SIP* que actúa como la puerta de entrada al subsistema *IMS* desde la red de conectividad *IP*: toda su señalización *SIP* de las redes *LTE/UMTS* transcurre a través de este servidor. Entre otras funciones, el servidor *P-CSCF* es el elemento que interacciona con las funciones del servidor *PCC* (*Policy and Charging Control*) de la red de conectividad (a través de la entidad *PCRF*). Por tanto, a través del servidor *P-CSCF*, el subsistema *IMS* puede controlar la operativa de la capa de transporte (i.e., servicios portadores *EPS* en el caso del sistema *LTE*).
- Servidor *Interrogating CSCF* (*I-CSCF*) es un servidor *SIP* que actúa como puerta de entrada de la señalización *SIP* proveniente de redes externas. En este sentido, la dirección *IP* del servidor *I-CSCF* es la que está incluida en los servidores de nombres de las redes externas (servidores *DNS* de Internet) de forma que la resolución de nombres de dominio (usuario@operadorLTE.com) conduce al envío de los mensajes

SIP a la dirección IP del servidor I-CSCF quien, posteriormente, re-dirige los mensajes a los servidores S-CSCF correspondientes.

Los servidores CSCF tienen acceso a la base de datos HSS. En dicha base de datos, junto con la información necesaria para acceder a las redes 3GPP también se almacena información necesaria para soportar sesiones multimedia sobre el subsistema IMS (servidor S-CSCF asignado a cada usuario). El subsistema IMS opera mecanismos que permiten la interoperabilidad de los servicios IMS con servicios equivalentes ofrecidos a través de redes de conmutación de circuitos como la red telefónica convencional. De esta forma, un usuario de la red LTE puede establecer una llamada de voz a través del subsistema IMS con un usuario de la red telefónica conmutada. En una llamada de estas características, la transmisión de voz del usuario LTE se soporta mediante una solución de voz sobre IP a través del servicio de conectividad proporcionado por la red LTE, y el subsistema IMS se encargaría de realizar la conversión de códec necesaria (teléfono) con un códec de Tasa Multimedia Adaptada (AMR por sus siglas en Inglés), y circuitos que operan bajo la recomendación G.711) y el establecimiento del circuito telefónico. Las funciones de interconexión con redes de circuitos se realizan a través de Recursos Multimedia (MRFP por sus siglas en Inglés). El servidor MRFC actúa como un agente de usuario SIP y controla los recursos del servidor MRFP a través de una interfaz H.248. Finalmente, la lógica y datos de los servicios proporcionados mediante el subsistema IMS se ubica en los servidores de aplicación, tal como se ilustra en la figura 2.24. Básicamente, un servidor de aplicación (denominado como, AS, en el contexto del subsistema IMS) es el servidor que contiene la lógica y ejecuta los servicios SIP. El sistema IMS identifica tres tipos diferentes de servidores de aplicación: Servidor de Aplicación SIP (SIP AS), Servidor Servicio de Acceso-Servicio de Capacidad Abierto (OSA-SCS por sus siglas en Inglés) y Función de Conmutación de Servicios Multimedia (IM-SSF por sus siglas en Inglés). El servidor de aplicaciones SIP (SIP AS) representa un servidor SIP genérico encargado de proveer un servicio final. Es importante destacar que la especificación de servicios finales basados en SIP queda fuera del ámbito de las redes 3GPP. Mediante el servidor de aplicaciones OSA-SCS, el subsistema IMS también ofrece la posibilidad de interactuar con el entorno de provisión de servicios OSA/Parlay [30]. Asimismo, a través del servidor IM-SSF el subsistema IMS también permite que los servicios IMS puedan interactuar con los mecanismos de red inteligente que soportan las redes 3GPP (servicios CAMEL, *Customised Applications for Mobile network Enhanced Logic*, [33]). Las pasarelas de medios (MGW por sus siglas en Inglés) proporcionan la conectividad entre el plano de transporte de la red de paquetes (conexiones voz sobre IP) y los circuitos conmutados de 64 kbps. La pasarela de medios se maneja desde un nodo de control llamado Función de Control Pasarela de Medios (MGCF por sus siglas en Inglés) a través del protocolo H.248 [32]. El controlador de la pasarela de medios es donde se realiza la conversión de la señalización SIP a la señalización utilizada en la red telefónica y conocida como Parte de Usuario de Señalización (ISUP por sus siglas en Inglés). El subsistema IMS integra también un conjunto de funciones de recursos multimedia. A modo de ejemplo, estas funciones permiten, bajo control de los servidores SIP, poner locuciones o tonos a usuarios de los servicios IMS. También permiten realizar la agregación de tráfico en sesiones SIP con más de dos interlocutores así como conversiones

de códec. Estas funciones se soportan en las unidades de red denominadas como Controladores de Función.

2.7 Flujo del tráfico de datos y dimensionado de redes IMS

El dimensionamiento de la red dependerá de varios parámetros: tasa de llamadas por segundo (carga de tráfico), porcentaje de pérdida de paquetes debido a sobre carga de servidores, retardos debido a las pérdidas de paquetes y variaciones de retardo, entre otros. Debido a que el modelo será de siguiente generación, deberá operar los servicios de voz, datos y video en protocolo IP (arquitectura horizontal de paquetes). Considerando estos elementos y parámetros involucrados los organismos 3GPP y 3GPP2 han diseñado un subsistema multimedia basado en el protocolo IP (IMS) como parte principal del elemento de gestión y control de la red. Así mismo se ha escogido al Protocolo de Iniciación de Sesión (SIP) para que realice el establecimiento y control de llamadas en el subsistema IMS. Los diferentes escenarios de enrutamiento del tráfico de datos, y la tasa de llamadas en horas pico, producirán retardos que afectaran la calidad del servicio (QoS) que perciben los usuarios en sus llamadas. Con el fin de poder prevenir estos problemas se propone un modelo que permita medir el ancho de banda requerido para la carga de tráfico rural esperada. Con el modelo de red propuesto se provee de un algoritmo matemático que calcula el ancho de banda estimado generado por el IMS durante el proceso de la puesta de cada sesión, permite además balancear la red en caso de saturación de los servidores SIP, para esto se contemplan los retardos y el número de mensajes SIP requeridos para establecer y liberar una llamada en la red, tomando en cuenta las pérdidas y retardos. Cuando se solicita el establecimiento de una llamada en SIP, el tráfico de señalización pasa varias veces a través de las diferentes unidades funcionales involucradas. Bajo este principio, cada entidad funcional (servidores SIP-CSCF's) en la cadena de encaminamiento es atravesada por el flujo de tráfico de señalización varias veces. Una vez atravesados los servidores SIP-CSCF (o HSS) estos proveerán algún servicio y luego reenviarán el tráfico. Con el propósito de administrar adecuadamente la red, se considera a los servidores SIP como nodos involucrados, bajo esta condición, cada servidor tiene una cola formada por diferentes flujos de tráfico en espera de ser enviados. Mediante este análisis se podrán conocer a los servidores que están comportándose como los cuellos de botella de la red. Se sabe en promedio que durante el establecimiento y liberación de una llamada SIP los

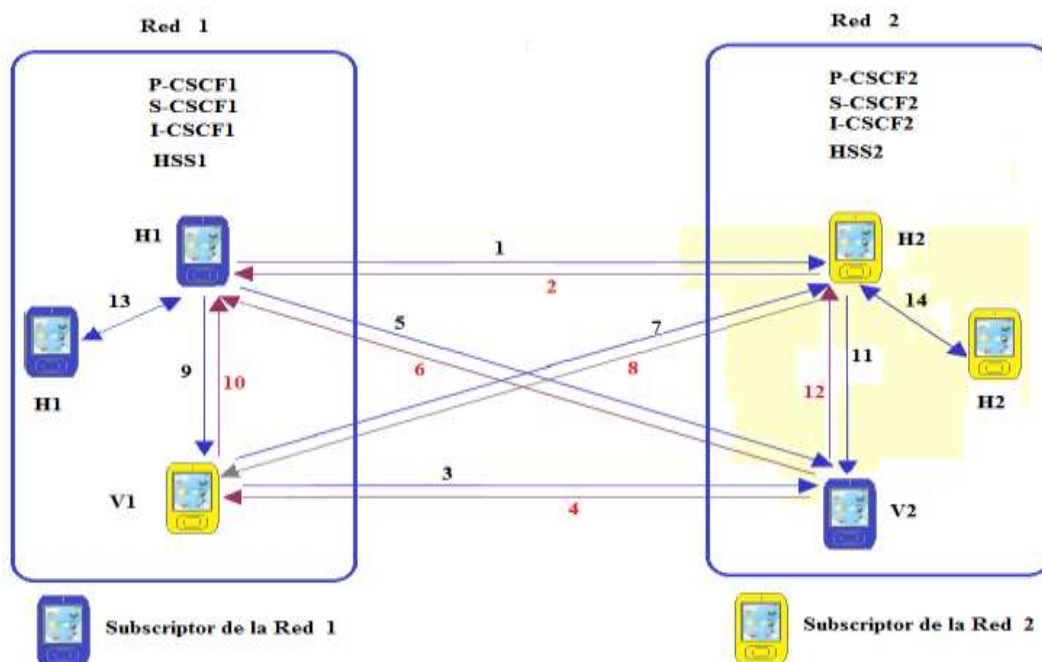


Figura 2.25. Posibles escenarios de conexión entre dos redes telefónicas diferentes.

Tabla 2.5. Escenarios posibles de llamadas telefónicas entre dos redes diferentes

	Escenario	Descripción
1	H1 → H2	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 2 (sin itinerancia)
2	H2 → H1	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 1 (sin itinerancia)
3	V1 → V2	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 1 con itinerancia en Red 2
4	V2 → V1	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 2 con itinerancia en Red 1
5	H1 → V2	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 1 con itinerancia en Red 2
6	V2 → H1	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 1 (sin itinerancia)
7	V1 → H2	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 2 (sin itinerancia)
8	H2 → V1	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 2 con itinerancia en Red 1
9	H1 → V1	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 2 con itinerancia en Red 1
10	V1 → H1	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 1 (sin itinerancia)
11	H2 → V2	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 1 con itinerancia en Red 2

12	$V2 \rightarrow H2$	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 2 (sin itinerancia) llama a
13	$H1 \rightarrow H1$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 1 (sin itinerancia)
14	$H2 \rightarrow H2$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 2 (sin itinerancia)

servidores P-CSCF y S-CSCF son tocados 11 veces, el servidor I-CSCF 12 veces y el servidor HSS es tocado una vez. La tabla 2.5, muestra los posibles escenarios para el establecimiento de llamadas entre dos redes.

Para el análisis de flujo del tráfico se analizaron dos redes diferentes, ambas tienen un servidor P-CSCF, S-CSCF y el I-CSCF cada una, y el modo de servicio que se brinda al usuario será: el primero que llega es el primero que se atiende. El tiempo de servicio promedio del servidor P-CSCF es \bar{X}_p , el tiempo de servicio del servidor S-CSCF es \bar{X}_s , el tiempo medio de servicio del servidor I-CSCF es \bar{X}_i , y el tiempo medio de servicio del servidor HSS es \bar{X}_h ; en base a estas consideraciones se identificará que servidor será el cuello de botella durante los diferentes escenarios de rutas de flujo. La figura 2.25 muestra los posibles escenarios de conexión.

H1: Subscriber de la Red 1 (sin itinerancia)

H2: Subscriber de la Red 2 (sin itinerancia)

V1: Subscriber de la Red 2 con itinerancia en la Red 1

V2: Subscriber de la Red 1 con itinerancia en la Red 2

La tabla 2.6, muestra la cadena de ruteo del flujo de señalización.

Tabla 2.6. Cadena de ruteo del flujo de señalización.

	Escenario	Cadena de Ruteo
1	$H1 \rightarrow H2$	P-CSCF1 ↔ S-CSCF1 ↔ I-CSCF2 ↔ HSS2 ↔ S-CSCF2 ↔ P-CSCF2
2	$H2 \rightarrow H1$	P-CSCF2 ↔ S-CSCF2 ↔ I-CSCF1 ↔ HSS1 ↔ S-CSCF1 ↔ P-CSCF1
3	$V1 \rightarrow V2$	P-CSCF1 ↔ S-CSCF2 ↔ I-CSCF1 ↔ HSS1 ↔ S-CSCF1 ↔ P-CSCF2
4	$V2 \rightarrow V1$	P-CSCF2 ↔ S-CSCF1 ↔ I-CSCF2 ↔ HSS2 ↔ S-CSCF2 ↔ P-CSCF1
5	$H1 \rightarrow V2$	P-CSCF1 ↔ S-CSCF1 ↔ I-CSCF1 ↔ HSS1 ↔ S-CSCF1 ↔ P-CSCF2
6	$V2 \rightarrow H1$	P-CSCF2 ↔ S-CSCF1 ↔ I-CSCF1 ↔ HSS1 ↔ S-CSCF1 ↔ P-CSCF1
7	$V1 \rightarrow H2$	P-CSCF1 ↔ S-CSCF2 ↔ I-CSCF2 ↔ HSS2 ↔ S-CSCF2 ↔ P-CSCF2

8	$H2 \rightarrow V1$	$P-CSCF2 \leftrightarrow S-CSCF2 \leftrightarrow I-CSCF2 \leftrightarrow HSS2 \leftrightarrow S-CSCF2 \leftrightarrow P-CSCF1$
9	$H1 \rightarrow V1$	$P-CSCF1 \leftrightarrow S-CSCF1 \leftrightarrow I-CSCF2 \leftrightarrow HSS2 \leftrightarrow S-CSCF2 \leftrightarrow P-CSCF1$
10	$V1 \rightarrow H1$	$P-CSCF1 \leftrightarrow S-CSCF2 \leftrightarrow I-CSCF1 \leftrightarrow HSS1 \leftrightarrow S-CSCF1 \leftrightarrow P-CSCF1$
11	$H2 \rightarrow V2$	$P-CSCF2 \leftrightarrow S-CSCF2 \leftrightarrow I-CSCF1 \leftrightarrow HSS1 \leftrightarrow S-CSCF1 \leftrightarrow P-CSCF2$
12	$V2 \rightarrow H2$	$P-CSCF2 \leftrightarrow S-CSCF1 \leftrightarrow I-CSCF2 \leftrightarrow HSS2 \leftrightarrow S-CSCF2 \leftrightarrow P-CSCF2$
13	$H1 \rightarrow H1$	$P-CSCF1 \leftrightarrow S-CSCF1 \leftrightarrow I-CSCF1 \leftrightarrow HSS1 \leftrightarrow S-CSCF1 \leftrightarrow P-CSCF1$
14	$H2 \rightarrow H2$	$P-CSCF2 \leftrightarrow S-CSCF2 \leftrightarrow I-CSCF2 \leftrightarrow HSS2 \leftrightarrow S-CSCF2 \leftrightarrow P-CSCF2$

Durante cada procedimiento para colocar una llamada y liberarla, los servidores P-CSCF y S-CSCF involucrados son tocados 11 veces, y el servidor I-CSCF es tocado 12 veces. Si λ_i es la carga de tráfico para cada escenario, entonces $\lambda = \sum \lambda_i$ es la carga total de tráfico (llamadas/unidad de tiempo).

El Apéndice B, muestra un método completo para identificar el cuello de botella de la red cuando se conoce la carga de tráfico de cada escenario. Así mismo el Apéndice C presenta un programa de cálculo que genera la tabla 2.7, la cual muestra la distribución de tráfico y embotellamiento de la red [34].

La tabla 2.7, muestra los datos de los servidores que se comportan como el cuello de botella cuando las dos redes interactúan entre sí. Para que un nodo se convierta en el cuello de botella de la red, este dependerá de la distribución del tráfico en dicha red. Es decir, cuando el tráfico entre suscriptores y visitantes de la Red-1 (sin itinerancia) se hace más pesado, el servidor S-CSCF1 será el cuello de botella (situación 6), sin embargo, si el tráfico entre suscriptores de la Red-2 sin itinerancia y visitantes se hace más pesado, el servidor P-CSCF2 es el cuello de botella (situación 2). Si aumentamos la carga de tráfico entre suscriptores de la red 1 sin itinerancia y con itinerancia, el servidor S-CSCF1 es el cuello de botella (situación 5), etc. Y debido a que existen posibilidades infinitas de distribuciones de tráfico, teóricamente, cada nodo puede ser el cuello de botella en determinada situación. El Apéndice B hace un análisis completo del comportamiento de una sola red y analiza lo que sucede en esta. De este análisis se puede concluir que si el tráfico hacia y desde los suscriptores itinerantes es más pesado, entonces el servidor S-CSCF será el cuello de botella, si el tráfico desde y hacia los visitantes es más pesado, entonces el servidor P-CSCF es cuello de botella.

Tabla 2.7. **Distribución de tráfico y embotellamiento.**

	1	2	3	4	5	6
λ_1/λ	0.2	0.1	0.05	0.05	0.1	0.1
λ_2/λ	0.2	0.1	0.05	0.05	0.1	0.1
λ_3/λ	0.05	0.05	0.05	0.05	0.05	0.05
λ_4/λ	0.05	0.05	0.05	0.05	0.05	0.05
λ_5/λ	0.05	0.05	0.15	0.05	0.15	0.05
λ_6/λ	0.05	0.05	0.15	0.05	0.15	0.05
λ_7/λ	0.05	0.05	0.05	0.15	0.05	0.05
λ_8/λ	0.05	0.05	0.05	0.15	0.05	0.05
λ_9/λ	0.03	0.03	0.05	0.05	0.05	0.15
λ_{10}/λ	0.03	0.03	0.05	0.05	0.05	0.15
λ_{11}/λ	0.02	0.12	0.05	0.05	0.05	0.05
λ_{12}/λ	0.02	0.12	0.05	0.05	0.05	0.05
λ_{13}/λ	0.1	0.1	0.1	0.1	0.05	0.05
λ_{14}/λ	0.1	0.1	0.1	0.1	0.05	0.05
Cuello de botella	P-CSCF1	P-CSCF2	S-CSCF1	S-CSCF2	S-CSCF1	P-CSCF1

Tabla 2.8. Propuesta de distribución de tráfico en la red.

1a	1b	2a	2b	3a	3b
10%	10%	25%	25%	15%	15%

En base a este análisis, se puede ver cómo mejorar la capacidad de la red. Analizando el caso particular de una red rural con una carga de tráfico normal de $\lambda = 10$ llamadas/segundo, con un tiempo de servicio promedio de los servidores P-CSCF, S-CSCF e I-CSCF son iguales a $\overline{X}_p = \overline{X}_s = \overline{X}_i = \overline{X}_h = 0.01s$. La tabla 2.8, muestra distribución de tráfico en la red sugerida:

De acuerdo a la distribución de tráfico propuesto para los diferentes escenarios, la utilización de cada entidad SIP podría calcularse de la siguiente manera:

$$\text{P-CSCF: } \rho_p = (10+10+15+15) \% \times \lambda \times 11 \times \overline{X}_p = 0.55;$$

$$\text{S-CSCF: } \rho_s = (10+10+25+25) \% \times \lambda \times 11 \times \overline{X}_s = \mathbf{0.77};$$

$$\text{I-CSCF: } \rho_i = (10+25) \% \times \lambda \times 12 \times \overline{X}_i = 0.42;$$

$$\text{HSS: } \rho_h = (10+25) \% \times \lambda \times \overline{X}_h = 0.035;$$

Los cálculos anteriores indican que el servidor S-CSCF es el que produce el cuello de botella en el tráfico. Para corregir este problema multiplicaremos la carga de

tráfico mediante un factor α_1 , así se obtendrá la máxima carga de tráfico de la red cuando la utilización del servidor S-CSCF alcance el valor de 1.

Probando con diferentes valores de α_1 encontramos que cuando $\alpha_1 = 1.3$, $\rho_s = (10+10+25+25) \% \times \lambda \times \alpha_1 \times 11 \times \overline{X}_s = 1$, con esta operación vemos que la capacidad máxima de la red sería de 13 llamadas/s Si agregamos otro servidor S-CSCF a la red tendremos la carga balanceada entre los dos servidores S-CSCF, bajo estas circunstancias el tiempo promedio de servicio del servidor S-CSCF será la mitad del tiempo de servicio original, es decir, $\overline{X}_s = \overline{X}_s/2$, y la utilización de cada nodo quedara de la siguiente manera:

$$\text{P-CSCF: } \rho_p = (10+10+15+15) \% \times \lambda \times 11 \times \overline{X}_p = \mathbf{0.55};$$

$$\text{S-CSCF: } \rho_s = (10+10+25+25) \% \times \lambda \times 11 \times \overline{X}_s = 0.385;$$

$$\text{I-CSCF: } \rho_i = (10+25) \% \times \lambda \times 12 \times \overline{X}_i = 0.42;$$

$$\text{HSS: } \rho_h = (10+25) \% \times \lambda \times \overline{X}_h = 0.035;$$

Se puede observar que el rendimiento de la red lo determina el servidor P-CSCF en lugar del servidor S-CSCF. Ahora vamos a incrementar el rendimiento multiplicando la carga del tráfico por un factor α_2 . Así con esto obtendremos la capacidad máxima de la red, cuando la utilización del servidor P-CSCF alcance el valor de 1.

Con $\alpha_2 = 1.82$, $\rho_p = (10 + 10 + 25 + 25) \% \lambda \times \alpha_1 \times 11 \times \overline{X}_p = 1$, por lo tanto la capacidad de la red ahora es 18.2 llamadas/segundo.

Así que el rendimiento adicional es igual a $(18.2-13)$ llamadas/segundo = 5.2 llamadas/segundo. Y la capacidad de la red se incrementa en $(5.2/13) \times 100\% = 40\%$. Esto significa que las llamadas entrantes se incrementarán en un 40% debido al rendimiento extra. Conociendo el costo de un servidor S-CSCF, podremos calcular cuántos ingresos adicionales se obtendrán añadiendo un servidor en el cuello de botella. Es decir, si el precio de un servidor S-CSCF es equivalente al 5% de los ingresos obtenidos, entonces el aumento de los ingresos será $(40\% - 5\%) = 35\%$. Este pequeño análisis le permitirá al operador dimensionar su red analizando la carga de tráfico total en horas pico y así poder dimensionar el tamaño de la red sin comprometer la calidad del servicio.

2.8 Interconectividad de LTE con redes heterogéneas

2.8.1 Generalidades

La red principal (*Core Network*) del modelo NGN rural, objeto de esta tesis, está basada en recomendaciones de la organización 3GPP, la cual cuenta con gran capacidad de interconexión con las diferentes redes de acceso alámbricas e inalámbricas. El grupo de socios 3GPP ha diseñado las interfaces adecuadas para que las diferentes redes de acceso UMTS, PSTN, ISDN, WLAN, etc., puedan ser interconectadas entre sí a través de la red troncal EPC con tecnología *LTE Advanced*. Los concesionarios de servicios móviles actualmente ofrecen sus servicios utilizando las tecnologías: GSM, UMTS R99, UMTS HSPA y LTE, y en un futuro, tecnologías *LTE Advanced*. Así mismo existen tecnologías complementarias que se utilizan en determinados entornos de operación, (tecnologías para proveer acceso inalámbrico en exteriores o bien en interiores de

edificios). Por esta razón el modelo propuesto por tratarse de un modelo de última generación debe permitir la convivencia de múltiples y diversas tecnologías adecuación a diferentes aplicaciones y escenarios de operación. Como ejemplo de estas tecnologías existen soluciones Wi-Fi ampliamente utilizadas para proporcionar acceso de radio a redes fijas de comunicaciones de banda ancha, estas redes representan una alternativa interesante para la provisión de los servicios propios de las redes celulares en entornos domésticos, oficinas, hoteles, aeropuertos, etc. De acuerdo a la manera tan rápida que evolucionan los sistemas de telecomunicaciones, se busca utilizar las tecnologías más adecuadas para cada caso, permitiendo la coexistencia de diferentes tecnologías de radio en los sistemas de comunicaciones móviles de última generación. El concepto de redes heterogéneas crea la necesidad de contar con teléfonos inteligentes multimodales, es decir, teléfonos que sean capaces de conectarse a la red troncal utilizando varias tecnologías de acceso de radio. Se conoce que la mayoría de teléfonos que hay actualmente en el mercado son de tecnología 2G/GSM, pero en el futuro los usuarios dispondrán de teléfonos LTE multimodales que soporten algunas de las tecnologías actuales (ejemplo terminales de banda ancha móvil que operen HSPA y LTE). En particular, el servicio de conectividad IP de la red troncal EPC puede atender a las redes de acceso 3GPP anteriores a las redes LTE, es decir, las redes UTRAN y GERAN. La interconexión de estas redes de acceso a una red LTE se realiza a través de interfaces específicas diseñadas por la organización 3GPP con la nueva red troncal EPC de tal manera que garantiza la continuidad de los servicios cuando los teléfonos cambian de red de acceso (es decir, se permite los mecanismos de movilidad (*handover*) entre tecnologías 3GPP). Además, el sistema LTE contempla también la posibilidad de utilizar redes de acceso no estandarizadas por la organización 3GPP, redes

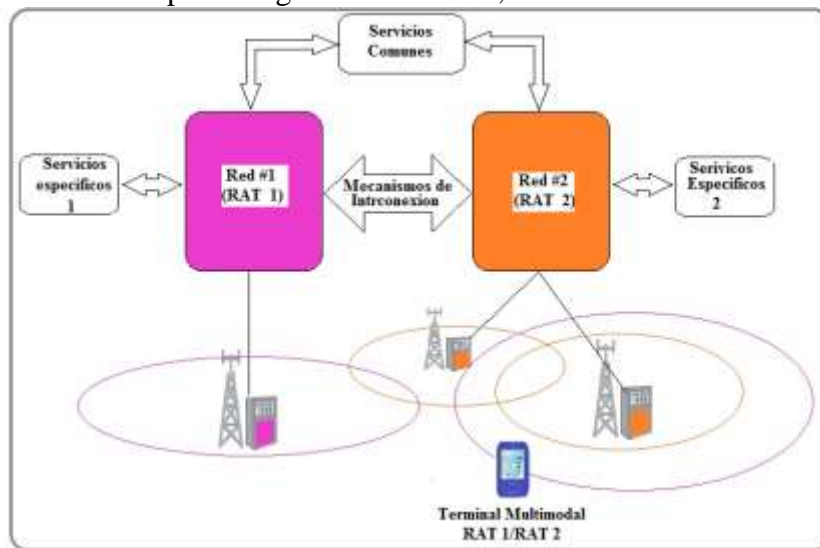


Figura 2.26. Escenario de soluciones de interconexión de redes heterogéneas.

denominadas como redes no 3GPP, para la provisión de los servicios de la red troncal EPC. Dentro de las redes no 3GPP, cabe mencionar especialmente las redes CDMA2000 desarrolladas por el organismo 3GPP2 (1xRTT y EVDO) [35].

2.8.2 Fundamentos de soluciones de interconexión

2.8.2.1 Escenario de Referencia

La figura 2.26, muestra la interconexión entre dos redes (redes #1 y #2), con tecnologías de radio diferente, RAT 1 y RAT 2.

Ambas redes pueden proporcionar acceso a un conjunto común de servicios (cualquiera de las dos redes puede ofrecer el servicio de Internet). Se asume que cualquiera de las redes puede proporcionar en forma exclusiva acceso a una serie de servicios específicos). Es decir, cuando existe dos redes con tecnologías diferentes (LTE y UMTS) la red tipo LTE podría utilizarse para proporcionar un servicio de video bajo demanda de alta calidad que no sería accesible mediante redes UMTS, y el servicio telefónico podría ofrecerse únicamente a través de redes UMTS. Si ambas redes (1 y 2) pertenecen a concesionarios diferentes el acceso de los usuarios debe estar sujeto a los acuerdos de itinerancia entre operadores. A la red a la que se encuentra suscrito un usuario se le denomina como red matriz y la red a la que puede conectarse un usuario sin tener una suscripción en ella se le denomina como red visitada. La figura 27 muestra dos tipos de redes de acceso diferentes (GSM y UMTS) que pertenecen al mismo operador de red.

2.8.3 Arquitectura genérica de las redes de acceso

Las redes de acceso que se conectarán al modelo de red NGN rural propuesto están diseñadas para proporcionar servicios de conectividad de paquetes con calidad de servicio (QoS). Para lograr esto, la provisión de servicios finales se realiza a través de plataformas especializadas externas a la red de acceso tales como los subsistemas IMS que constituirán el corazón de la red y que la convierten en red NGN. La figura 2.27, muestra los nodos

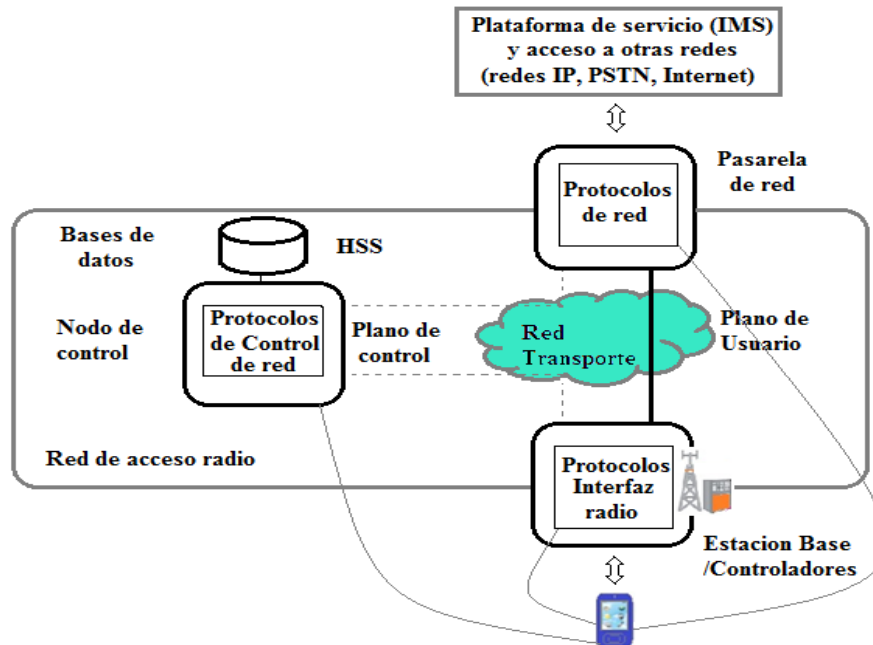


Figura 2.27. Arquitectura genérica de una red de acceso de radio.

básicos de una arquitectura de red representativa de la mayoría de los sistemas actuales de comunicaciones móviles basados en conmutación de paquetes (GPRS/UMTS, LTE, CDMA2000 1xEV-DO, *Mobile WiMAX* y redes de acceso WLAN). Así mismo muestra la ubicación en la red de las diferentes torres de protocolos con las que interactúan las terminales inteligentes para obtener el servicio de conectividad

En relación a la red de acceso, su plano de control se encarga de lo siguiente: control de autenticación y autorización, control de la gestión de movilidad (registro, actualización de la localización), así también la gestión de seguridad (gestión de claves y control de los servicios de confidencialidad e integridad) y finalmente la gestión de las sesiones (activación de los servicios de conectividad de la red). Estas funciones relacionadas al plano de control de la red se puede observar en la figura 28. Se observa al nodo de control y la pila de protocolos a través de la cual la red interactúa con las terminales. A la entidad MME se le conoce como el nodo de control, en redes GPRS/UMTS estas funciones se integran en el nodo SGSN. La red de acceso también cuenta con las bases de datos necesarias para sustentar la operación de las funciones de control tales como: almacenamiento de datos de localización, parámetros de seguridad, perfiles de calidad de servicio, etc. Las funciones y conjuntos de protocolos utilizadas en el plano de control dependen de cada tecnología de acceso de radio que se utilice, por ejemplo en redes LTE, la señalización de control entre la red y la terminal se lleva a cabo mediante los protocolos NAS diseñados por el organismo 3GPP. Es importante mencionar que la interconexión entre elementos de la red tales como: nodos de control, estaciones base y pasarelas de red se realizar mediante la utilización de redes de transporte IP/MPLS o Ethernet, tecnologías completamente independientes de la tecnología de acceso radio. La idea que se propone en este proyecto es que todas las interfaces de la red LTE se soporten sobre pilas de protocolos IP.

2.8.4 Terminales o teléfonos multimodal utilizados en la red

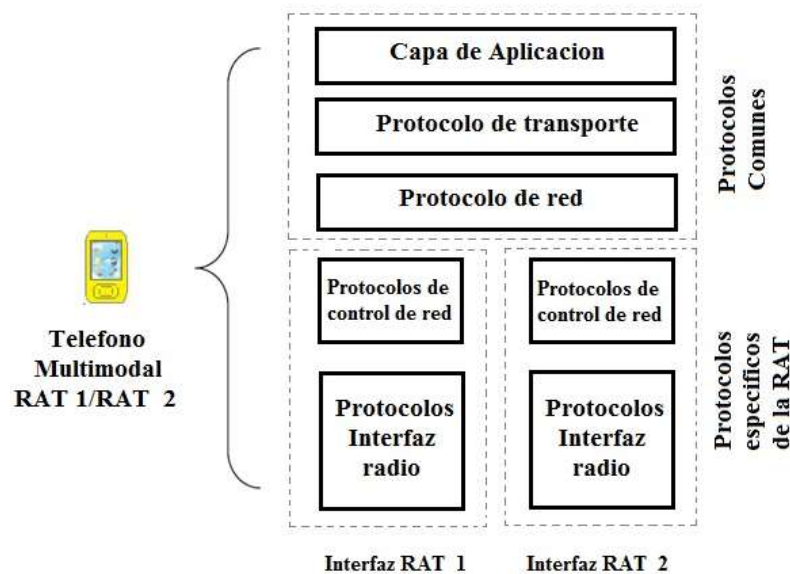


Figura 2.28. Arquitectura de protocolos de un teléfono multinodal.

La figura 2.28, muestra los elementos que constituyen una terminal o teléfono multimodal, así también muestra la pila de protocolos que lo hace una terminal multimodal para redes de acceso como las descritas anteriormente. La pila de protocolos debe estar estructurada de un conjunto de pilas de protocolos propios de cada una de las tecnologías de radio soportadas y un conjunto de protocolos comunes para las capas de red y superiores (transporte y aplicación). En cuanto a los protocolos específicos de cada tecnología radio, la terminal incluye los protocolos propios de la interfaz radio de las capas física y de enlace y los protocolos de control de red, tales como: protocolos NAS en redes 3GPP, protocolos EAP para autenticación en redes IEEE 802, etc. La estructura de la terminal mencionada puede operar en la capa de red IP permitiendo que diferentes protocolos de la capa de transporte tales como TCP y UDP y de aplicación como SIP trabajen en forma independiente a la tecnología de acceso de radio. Es importante mencionar que la capa de red en el modelo de interconexión tiene un papel muy importante ya que constituye la red dorsal entre las diferentes interfaces de comunicaciones que pueda tener este dispositivo multimodal y las capas de protocolos superiores. Otra función de la terminal multimodal es el soporte de funciones asociadas a la gestión de movilidad entre redes. Particularmente, puede seleccionar la mejor red de acceso en un momento determinado, atendiendo a la disponibilidad de la misma y al tipo de servicios realizado y las preferencias del usuario. Esta función para seleccionar la red se basa en la información que proporcionan las diferentes capas de radio (niveles de señal en cada interfaz) y la inicialización de los procedimientos para cambio de red cuando esta sea necesaria, conocido como mecanismos de movilidad. Con el fin de incorporar determinados mecanismos de interconexión en su capacidad para operar dos interfaces de radio en forma simultánea. Esto incrementa enormemente la complejidad del diseño de estos teléfonos cuando se busca el acoplamiento en radio frecuencia, cuando las frecuencias utilizadas en las dos tecnologías de radio se encuentran relativamente cercanas. Las terminales multimodal se clasifican en dos categorías: terminales multimodal *dual-radio*, con la capacidad de mantener operativas las dos interfaces radio a la vez, y terminales multimodal de un radio (*single-radio*), que solamente soportan la transmisión/recepción por una interfaz de forma simultánea. Una terminal de dos radios (*dual-radio*) puede conectar en forma simultánea a las dos redes de acceso de forma que el tiempo de interrupción del servicio podría ser mínimo (debido únicamente a procesos de cambios de encaminamientos dentro de las redes).

2.8.5 Niveles de interconexión

Los niveles de intercomunicación establecen los requerimientos básicos que condicionan la elección y/o el diseño del tipo de mecanismos de intercomunicación necesarios. Con este planteamiento se separa la definición de los niveles de interconexión de aspectos particulares de arquitectura, tecnología y funciones de la red. Bajo este enfoque, se definen los siguientes cuatro niveles de interconexión entre redes:

Primer Nivel: Acceso a servicios a través de una red visitada. Este nivel de interconexión permite a un usuario poder utilizar los servicios disponibles en una red visitada mediante el uso de la identidad y credenciales otorgadas por otra red, la red matriz donde el usuario tiene su suscripción. Usuario de una red UMTS/LTE, equipado con una terminal portátil con interfaces de red celular UMTS/LTE y Wi-Fi, puede conectarse a Internet a través de

una red Wi-Fi de acceso público (aeropuerto, restaurante, etc.) identificándose con su tarjeta de identificación del usuario UMTS/LTE (USIM).

Segundo Nivel: Acceso a servicios de la red matriz desde una red visitada. Bajo este nivel, un usuario que se conecta a través de una red visitada podría acceder a servicios disponibles en la red de su operador matriz. Como ejemplo, retomando el caso del usuario con un teléfono portátil con interfaces UMTS/LTE y Wi-Fi, el usuario podría acceder a los servicios UMTS/LTE (servicios de presencia o mensajería instantánea entre un grupo cerrado de usuarios) mientras se encuentra conectado a través de la red Wi-Fi.

Tercer Nivel: Continuidad de servicio entre redes. Este nivel extiende los dos niveles anteriores de forma que el usuario no tendría que re-establecer sus servicios al producirse un cambio en la red de acceso. Sin embargo, en este nivel se admite la posibilidad de que la calidad de servicio pueda sufrir un cierto deterioro durante la transición entre redes debido al propio mecanismo de cambio de red. Así pues, siguiendo con el ejemplo del usuario del teléfono portátil UMTS/LTE y Wi-Fi, el usuario podría encontrarse descargando un archivo de datos a través del acceso Wi-Fi mientras se está desplazando y pierde la cobertura de esa red. En ese momento, automáticamente el teléfono cambia a la red UMTS/LTE que proporciona cobertura en exteriores y la descarga del archivo continua a través de la nueva red, considerando que en la transición de una red a otra se detenga brevemente la descarga.

Cuarto Nivel: Continuidad transparente de servicio entre redes. Este nivel de interconexión entre redes permitiría la continuidad de los servicios sin que se produzca una degradación perceptible al usuario durante la transición entre redes. En este nivel de interconexión entre redes heterogéneas, el tiempo de interrupción requerido en un proceso de movilidad entre las redes LTE y UMTS debe ser inferior a 300 ms para servicios en tiempo real [35]). Siguiendo el ejemplo anterior, si el usuario tiene un teléfono con tecnologías inalámbricas UMTS/LTE y Wi-Fi (es la marca de la *Wi Fi Alliance*) e inicia una conexión de voz sobre IP (VoIP) estando conectado a través de tecnología Wi-Fi, dicha conexión debería transferirse de forma transparente a la red LTE/UMTS si el usuario sale de la cobertura de la red de área local.

2.8.6 Mecanismos de interconexión

La figura 2.29, muestra los mecanismos de interconexión más importantes sobre los que se sustentan los niveles de interconexión mencionados en el párrafo anterior.

2.8.6.1 Mecanismos AAA entre redes

Estos mecanismos ofrecen la opción para que suscriptores tengan el acceso a redes visitadas haciendo uso de las credenciales otorgadas por su red matriz a la que están suscritos, sin que la red visitada tenga ningún conocimiento de dicho usuario previamente a la solicitud del servicio. También, permiten transferir información para tarificar el acceso a los servicios a través de redes visitadas (la utilización de los recursos que el usuario hace en la red visitada se envía a la red matriz que se encarga de la facturación de sus suscriptores). Estos mecanismos se denominan de forma genérica como mecanismos (Autenticación, Autorización y Contable). Entre los mecanismos AAA principales para la realización del Nivel A se encuentran los siguientes:

- Protocolo de Autenticación Extensible (EAP por sus siglas en Inglés) definido en la recomendación RFC 3748 [36] proporciona el soporte básico para el transporte confiable de diferentes protocolos de autenticación. A través del protocolo EAP se permite la autenticación del usuario mediante credenciales simples del tipo usuario/clave de acceso así como mediante la utilización de certificados digitales. Todos estos marcos de autenticación y autorización basados en protocolo EAP han sido adoptados en el sistema LTE para el acceso básico a través de redes no 3GPP. El protocolo EAP también forma parte del sistema de control de acceso soportado en equipos Wi-Fi, y en redes móviles WiMAX.

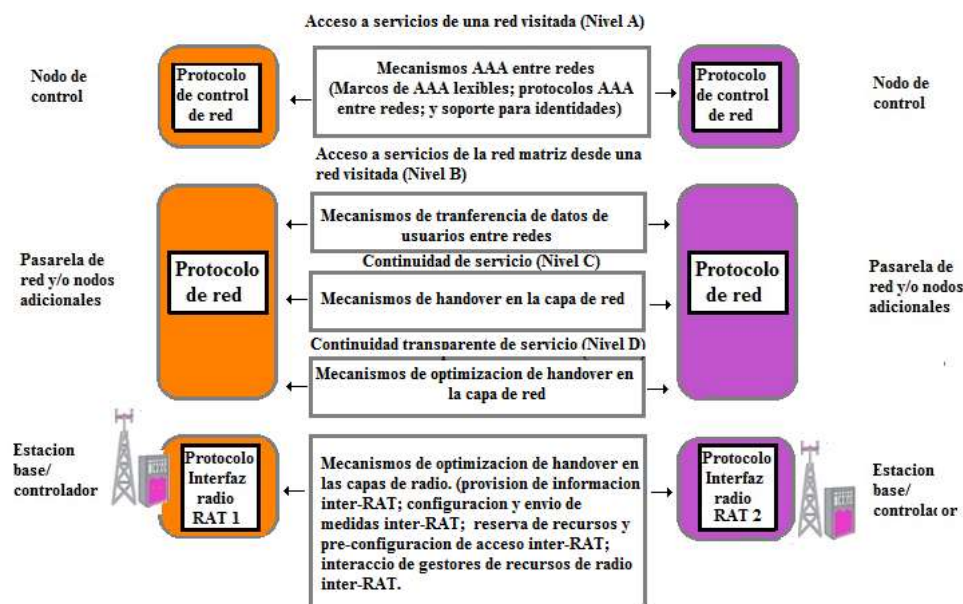


Figura 2.29. Niveles y mecanismos de interconexión asociados.

- Funciones del protocolo *Proxy/Relay* para la transferencia de información AAA entre redes, junto con los correspondientes protocolos de señalización. En este contexto, el protocolo *Diameter* definido en la recomendación RFC 3588 [37] está creado para poder cumplir los requisitos mínimos de un protocolo de AAA entre redes. Tal como se verá más adelante, el sistema LTE incorpora el uso de este protocolo en las soluciones de interconexión con redes no 3GPP.

2.8.6.2 Mecanismos para la transferencia de datos entre redes.

Permiten la transferencia entre redes de la información generada o destinada a un usuario de forma que el usuario pueda tener acceso a un determinado servicio ofrecido en una red diferente de la que se encuentre conectada, y constituyen la base del Nivel B. En redes de paquetes, se basan en la utilización de protocolos de túnel (*Tunnelling Protocol*) [L2TP] de capa 2, definido en la recomendación RFC 2661 [38], y el modo túnel del

protocolo *IPsec* definido en la recomendación RFC 4301 [39]. A modo de ejemplo, las especificaciones 3GPP para acceso mediante WLAN a servicios UMTS, establece la necesidad de que los terminales utilicen el protocolo *IPsec Encapsulating Security Payload* (ESP) descrito en la recomendación RFC 4303 [40] para establecer un túnel entre la propia terminal y un determinado nodo de la red UMTS. Mediante este túnel, un teléfono multimodal (terminal) conectado en una red Wi-Fi podría acceder a los servicios ofrecidos en la red UMTS de su operador celular (servicios en plataformas IMS).

2.8.6.3 Mecanismos de movilidad en capa de red

Los mecanismos de movilidad en la capa de red constituyen la base para un Nivel C de interconexión. En redes IPv6, referencia RFC 3775 [11] se especifica un protocolo denominado MIPv6 que diseñado para aprovechar las capacidades ofrecidas por la capa de red IPv6. Así mismo, con el fin de facilitar el uso de protocolos de movilidad IP en escenarios donde co-existan redes IPv4 e IPv6, la extensión de doble pila (*Dual Stack MIPv6 DSMIPv6*) del protocolo MIPv6 permite soportar tanto direcciones IPv4 como IPv6.

2.9 Interconexión entre LTE y otras redes 3GPP.

2.9.1 Marco general

La tecnología LTE ha sido diseñada para soportar enlaces de telecomunicaciones de red donde co-existan tecnologías E-UTRAN, UTRAN y/o GERAN. La solución de interconexión especificada permite un Nivel D de Interconexión para proveer el servicio de conectividad IP a través de los diferentes accesos 3GPP con teléfonos de radio simple. Debe tomarse en cuenta las diferentes prestaciones de cada una de las redes de acceso en cuanto a latencias y tasas de transferencia en bps. La solución de Interconexión entre redes de acceso 3GPP se concentra en un punto de anclaje en la red troncal EPC que permite la movilidad entre las redes de acceso sin que el servicio de conectividad IP se vea afectado, es decir, la dirección IP asignada al teléfono no se pierde aun cuando esta cambie de red de acceso. La interconexión de las redes UTRAN y GERAN a la red troncal EPC se realiza mediante el soporte de funciones del dominio de paquetes GPRS de la red UMTS. Ver figura 2.30.

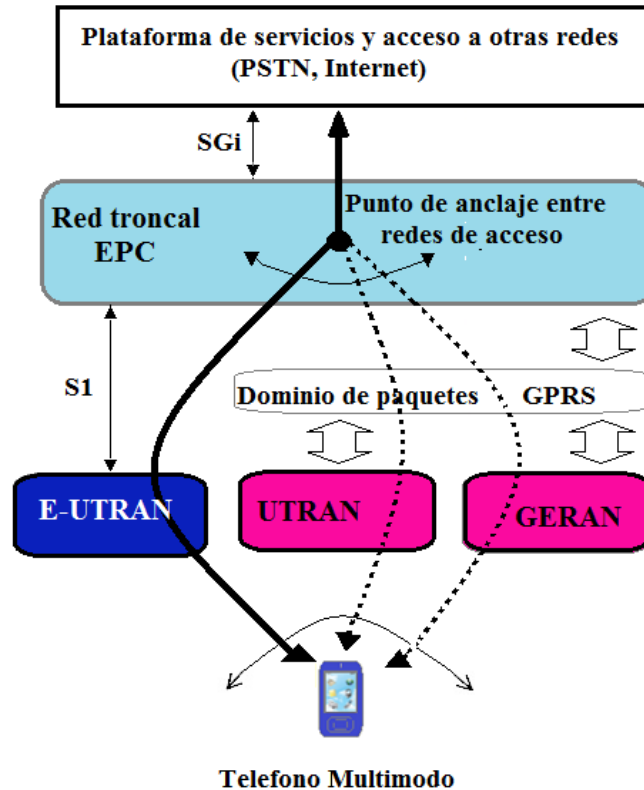


Figura 2.30. Marco de interconexión entre redes de acceso 3GPP

2.9.2 Interconexión mediante interfaces propias del sistema GPRS

La figura 2.31, muestra la arquitectura de una red troncal EPC a la que se interconectan redes de acceso UTRAN [41] y GERAN [42] mediante el soporte de las interfaces propias del sistema GPRS. Se describen las unidades de red e interfaces del dominio de paquetes GPRS y de las redes de acceso UTRAN y GERAN que forman parte de la solución de interconexión. Esta figura muestra la interconexión entre las redes GERAN y UTRAN y donde la red troncal EPC se efectúa a través del nodo SGSN, el cual forma parte del dominio de paquetes GPRS [43] y constituye el punto de entrada de la red troncal a los servicios GPRS ofrecidos a través de las redes de acceso UTRAN y GERAN. El nodo SGSN integra tanto las funciones de plano de control como de plano de usuario para la provisión del servicio GPRS. Comparando la arquitectura de la red troncal GPRS con la red troncal EPC, la entidad SGSN del protocolo GPRS sería equivalente a las unidades MME y S-GW de la entidad EPC. La red UTRAN se conecta al nodo SGSN mediante una interfaz denominada Iu-PS mientras que la red GERAN puede hacerlo mediante la interfaz Gb o bien también a través de una interfaz Iu-PS. El nodo SGSN canaliza toda la señalización entre la red GPRS y el equipo de usuario relativo a las funciones de gestión de movilidad y gestión de las

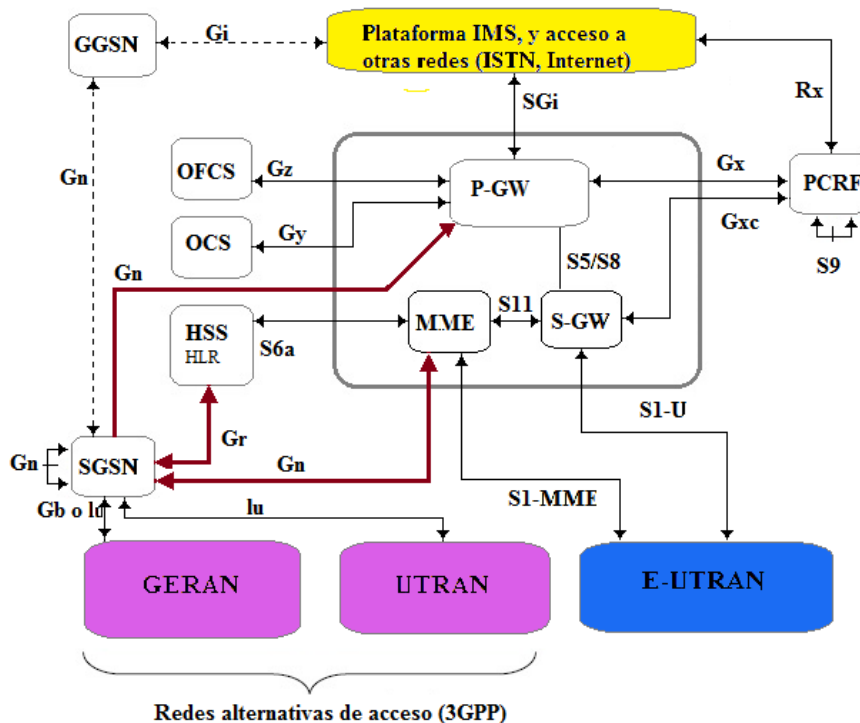


Figura 2.31. Interconexión mediante interfaces propias del sistema 3GPP

sesiones. Para ello, el nodo SGSN se conecta a la base de datos que contiene la información de los usuarios del sistema, denominada registro de localización de casa (HLR2 por sus siglas en Inglés) en redes GPRS, mediante la interfaz Gr. La interfaz Gr es una interfaz basada en el protocolo MAP y que se transporta sobre redes de señalización SS7 de la Red RTPC o PSTN (Por sus siglas en Inglés). Además del nodo SGSN, la implementación del servicio GPRS se basa en una entidad de red adicional denominada Nodo de Soporte GPRS de Cabecera (GGSN por sus siglas en Inglés).

El nodo GGSN es el nodo que interconecta la red troncal GPRS con la red externa de paquetes a la que se proporciona acceso a la Red de Paquete de Datos (PDN por sus siglas en Inglés). Siguiendo la analogía con la red troncal EPC, la entidad GGSN de la red GPRS es equivalente a la entidad P-GW de la red EPC. La interconexión entre las unidades SGSN y GGSN se realiza mediante una interfaz denominada Gn a través de la cual se sustentan los procedimientos de gestión de sesiones (establecimiento de servicios portadores en GPRS, denominados contextos PDP) y movilidad en GPRS (cambio del nodo SGSN sin perder conectividad con la red externa). La interfaz Gn también se utiliza entre el nodo SGSN para sustentar el cambio de nodo SGSN en una conexión en curso. Así, mediante esta interfaz, un usuario de la red GPRS puede pasar a ser servido por otro nodo SGSN sin que esto afecte a la continuidad del servicio que en todo momento sigue proporcionándose a través del mismo nodo GGSN. Para esto, los nodos SGSNs involucrados se intercambian la información de control necesaria (contextos asociados al servicio de un usuario) y, de forma opcional, pueden activar un plano de usuario provisional entre ellos para transportar los paquetes de datos del usuario con el fin de minimizar la pérdida de información durante el traspaso. En la parte superior izquierda de la figura 2.31, se muestra el papel de la entidad GGSN dentro de la arquitectura GPRS,

aunque es importante destacar que dicha entidad no forma parte de la arquitectura de interconexión de las redes UTRAN y GERAN con la red troncal EPC. La solución de interconexión se logra a través de las interfaces mostradas en rojo en la figura 2.31. De esta forma, mediante la re-utilización de una interfaz Gn entre el nodo SGSN y la pasarela P-GW de la red troncal EPC, la pasarela P-GW es la encargada de realizar la función de punto de anclaje que permite la movilidad entre las redes de acceso 3GPP. Desde la perspectiva del acceso GPRS, la pasarela P-GW se comporta como un nodo GGSN. Por tanto, a las terminales LTE/UMTS que accedan al servicio de conectividad a través de las redes UTRAN o GERAN, la red GPRS (el nodo SGSN) debe asignarles una pasarela P-GW en lugar de un nodo GGSN convencional. De esta forma, cuando los terminales duales tengan cobertura de la red de acceso E-UTRAN, el servicio de conectividad puede reestablecerse a través de la red E-UTRAN manteniendo la misma pasarela P-GW para la interconexión con la red externa. El cambio de una red de acceso UTRAN o GERAN a otra red E-UTRAN comporta también el cambio del nodo SGSN por las unidades de la red troncal EPC correspondientes. En este caso, dicho cambio se sustenta en la implementación de la interfaz Gn entre las entidades SGSN y MME. La entidad MME contiene el plano de control de la red troncal EPC y a través de ella, puede articularse el cambio hacia o desde la red de acceso E-UTRAN. A diferencia de lo que ocurre en la red GPRS, en la interfaz Gn entre las entidades SGSN y MME no se transportan paquetes de usuario. La interfaz Gn es una interfaz basada en una versión del Protocolo Túnel GPRS (GTP por sus siglas en Inglés) definida en la recomendación TS 29.060 [44], a diferencia de las interfaces internas del sistema LTE que utilizan una nueva versión del protocolo definida en la recomendación 29.274 [45].

2.9.3 Interconexión de GPRS y EPC utilizando nuevas interfaces

La organización 3GPP ha creado tres nuevas interfaces para conectar la tecnología GPRS con la entidad EPC a través del módulo SGSN. Estas interfaces son: S3, S4 y S6d. Las interfaces S3 y S4 son interfaces basadas en la misma versión del protocolo GTP que se utiliza en las interfaces internas de la red troncal EPC. En este caso, la interfaz S3 se utiliza para interconectar el nodo SGSN con la entidad MME del plano de control de la red troncal EPC y la interfaz S4 se utiliza entre el nodo SGSN y la pasarela S-GW que constituye el punto de anclaje del plano de usuario dentro de la red EPC. En la misma línea, la interfaz S6d es la adaptación de la interfaz S6a para el acceso a la base de datos HSS desde el dominio GPRS. El uso de esta interfaz basada en *Diameter* evita la necesidad de soportar la interfaz Gr basada en la señalización MAP/SS7 para el acceso al servidor HLR. Además de estas tres interfaces, la solución planteada permite también la explotación de una interfaz directa entre la red UTRAN y la red troncal EPC denominada S12.

En la figura 2.32 se muestra la solución de interconexión especificada, resaltando en color rojo las interfaces a través de las cuáles se provee del servicio LTE a través de las diferentes redes de acceso. Para un usuario conectado a través de las redes GERAN o UTRAN, el plano de usuario transporta datos a través de un nodo SGSN y una pasarela S-GW hasta llegar a la pasarela P-GW. Esta opción es importante en escenarios de itinerancia donde la pasarela S-GW de la red visitada siempre constituye un punto de anclaje único de todo el tráfico de usuario.

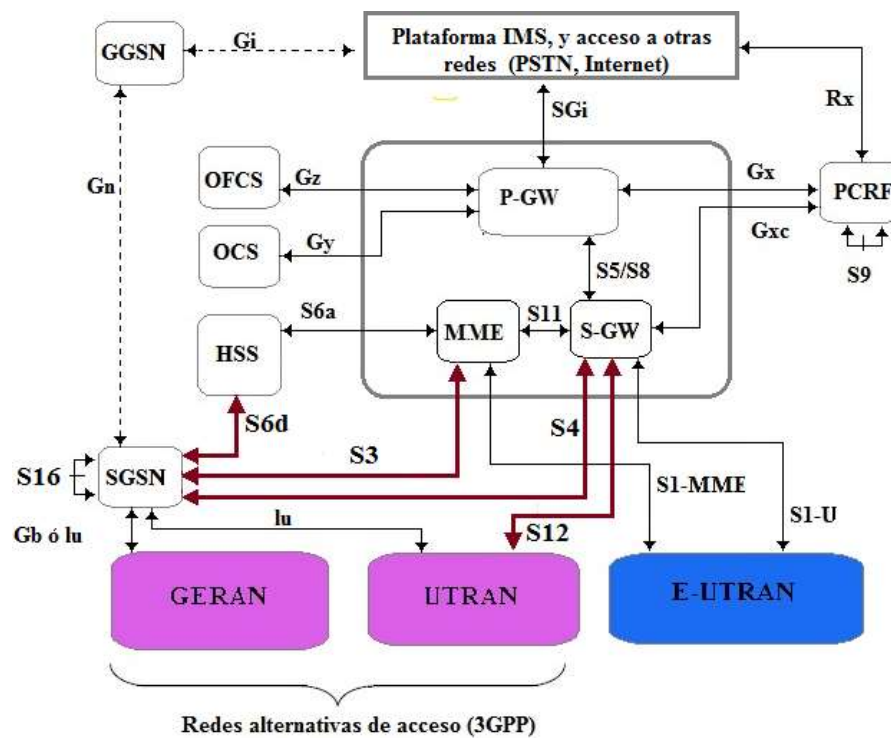


Figura 2.32. Interconexión a través de nuevas interfaces del sistema 3GPP

La utilización de un punto de anclaje único e independiente de la red de acceso en la red visitada facilita al operador de esa red la monitorización y control del tráfico que cursa un usuario en itinerancia. No obstante, la canalización del tráfico a través de una pasarela S-GW de la red troncal conlleva que el tráfico cursado a través de redes UTRAN y GERAN sea procesado en un elemento de red adicional en comparación con la solución anterior de forma que la latencia de la red puede verse penalizada (en esta solución el plano de usuario atraviesa tres nodos de la red troncal GPRS/EPC mientras que en la anterior solución únicamente eran dos). En el caso de redes UTRAN, la utilización de la interfaz S12 permite reducir el número de saltos del plano de usuario de forma que se sorte el procesamiento del plano de usuario en el nodo SGSN. La interfaz S12 permite el establecimiento del túnel GTP entre el controlador RNC y la pasarela S-GW, evitándose que el tráfico transcurra a través del nodo SGSN. Es importante destacar que, aunque se establezca el túnel del plano de usuario entre la red UTRAN y la pasarela S-GW, la señalización de control sigue realizándose a través del nodo SGSN. En el caso de la red GERAN no se contempla esta opción. Además de facilitar la operación en escenarios de itinerancia, la utilización de las interfaces S3 y S4 en lugar de la interfaz Gn permite incorporar una nueva funcionalidad denominada Reducción del Modo de Señalización Desocupado (ISR por sus siglas en Inglés), para gestionar de forma más eficiente la movilidad de las terminales en modo *desocupado* que pueden estar conmutando entre celdas UTRAN/GERAN y E-UTRAN.

2.9.4 Movilidad entre redes de acceso

Los mecanismos a través de los cuales se realiza la movilidad entre redes de acceso 3GPP dependen del estado en que se encuentre el teléfono (conectado o desconectado) [46].

2.9.4.1 Movilidad de teléfonos en estado desconectado

Considerando que la red a utilizar debe contemplar un contexto de múltiples redes de acceso, decimos que un teléfono se encuentra en el estado desocupado cuando está registrado en la red EPC, es decir tiene establecido como mínimo un servicio portador EPS por defecto, aunque no se encuentra conectado a través de una red de acceso. Esto es, no ha establecido una conexión a través de ninguna red de acceso. Considerando un escenario donde existen varias tecnologías de acceso 3GPP, y para el caso del sistema E-UTRAN, el sistema de información soportado a través de los canales de radiodifusión (*broadcast*) envía periódicamente mensajes con información sobre las frecuencias y celdas vecinas de redes UTRAN, GERAN y CDMA2000 (i.e., mensajes SIB6, SIB7 y SIB8). La Figura 2.33, muestra el mecanismo de re-selección de celda de un teléfono en estado *desocupado* en un despliegue *multi-RAT*.

En las redes de acceso GERAN y UTRAN, la gestión de localización se realiza en base a unas áreas de seguimiento denominadas Áreas de Ruteo (RA por sus siglas en Inglés) y su actualización la lleva a cabo el nodo de soporte SGSN. La base de datos del sistema HSS contiene el identificador de las entidades de red SGSN o MME donde se encuentra localizado el usuario. La figura 2.34, muestra el procedimiento de gestión de localización descrito. Tal como se observa en el gráfico (a) de dicha figura, cuando un usuario sintoniza una celda UTRAN o GERAN, el nodo SGSN dispone de un contexto de movilidad asociado a dicho usuario donde se almacena el identificador de RA. Asimismo, el contexto de movilidad del usuario en la base de datos HSS contiene el identificador del nodo SGSN correspondiente. Además, en el caso de la solución de intercomunicación basada en la utilización de las interfaces S3/S4/S6d, la pasarela S-GW también dispone del identificador del nodo SGSN, tal como se ilustra en la figura 2.34(a). En el estado desocupado, el plano de usuario únicamente permanece establecido entre la pasarela P-GW y S-GW, de forma que cuando se detecta la presencia de tráfico descendente en la pasarela S-GW, ésta debe saber a qué nodo debe notificarlo, si los nodo SGSN o MME

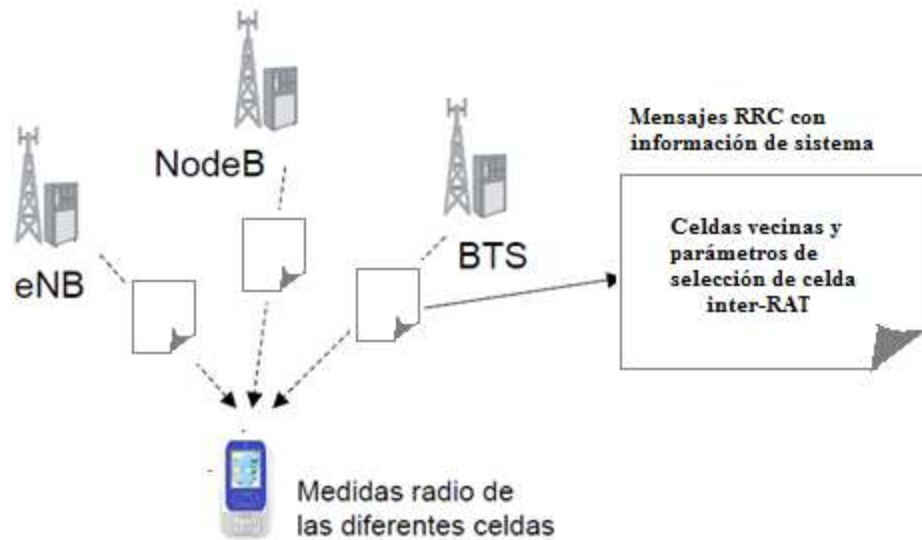


Figura 2.33. Selección de celda de teléfono en estado desocupado en cubrimiento multi-RAT

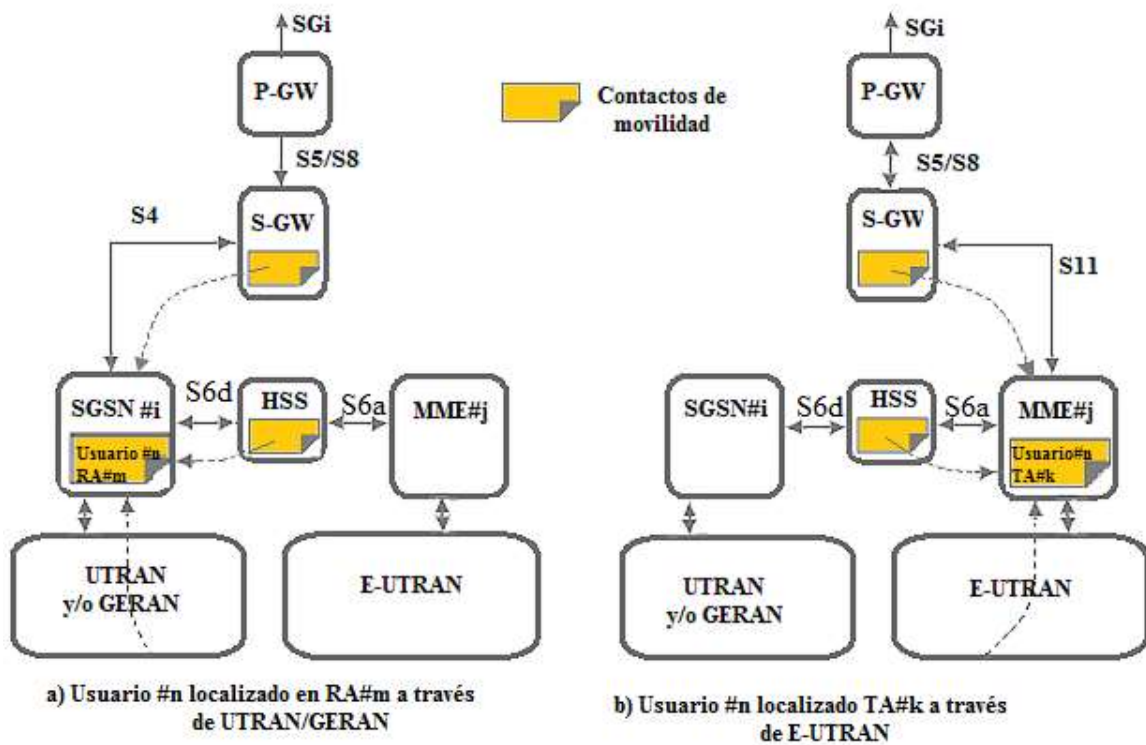


Figura 2.34. Procedimiento de gestión para localizar un usuario en despliegue multi-RAT para iniciar el procedimiento de aviso. Análogamente, en la gráfica de la figura 2.34 (b) se muestra la situación en que el usuario se encuentra localizado a través de la red E-UTRAN.

La red LTE cuenta con la funcionalidad conocida como Reducción de la Señalización en Modo Desocupado (ISR por sus siglas en Inglés) la cual permite reducir el número de actualizaciones del RA y TA que se producen en el caso de re-selecciones

frecuentes entre celdas de las redes E-UTRAN y UTRAN/GERAN. La idea básica de la funcionalidad ISR es permitir que el teléfono pueda mantener dos registros en paralelo, es decir, que pueda ser localizada simultáneamente mediante un área de rutado y un área o lista de áreas de seguimiento. Para ello, es necesario que los nodos SGSN, MME, HSS y S-GW soporten dicha funcionalidad y que las interfaces entre el nodo SGSN y las unidades de la red troncal sean S3 y S4 en lugar de Gn. La activación del modo ISR se produce cuando un teléfono, ya registrado en una RA inicia el registro a una TA, o viceversa. En el procedimiento de actualización de localización el teléfono avisa que puede soportar la funcionalidad ISR. Si los elementos de la infraestructura de red citados anteriormente también lo soportan, el procedimiento de actualización termina de forma que tanto el nodo SGSN como la entidad MME mantienen activo de forma simultánea en un contexto de movilidad asociado al usuario, en el cual en el contexto del nodo MME se guarda el TA y el identificador GUTI y en el contexto del SGSN se guarda el RA y P-TMSI. Asimismo, en la base de datos del servidor HSS se guardan los identificadores tanto del nodo SGSN como de la entidad MME. Por otro lado la entidad S-GW que termina el plano de usuario de la terminal en estado desocupado, se establecen ahora dos conexiones de control: una conexión con la entidad MME donde la terminal se encuentra registrada a través de la interfaz S11 y otra con el nodo SGSN donde también existe un registro activo a través de la interfaz S4. De esta forma, cuando se recibe tráfico en la pasarela S-GW, éste notifica tanto a la entidad MME como al nodo SGSN correspondiente, quienes, en paralelo, inician el mecanismo de aviso. Por tanto, el teléfono

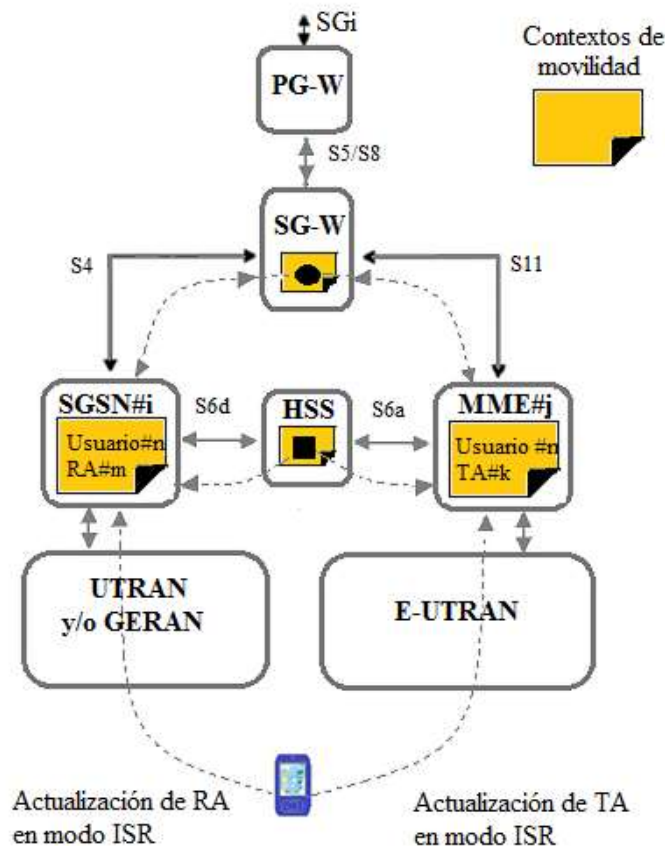


Figura 2.35. Gestión para la localización en despliegue multi-RAT en el modo ISR

recibirá el mensaje de aviso a través de la red de acceso en la que se encuentre conectada en ese momento e iniciará un procedimiento de solicitud de servicio que terminará con el restablecimiento del plano de usuario a través de las redes E-UTRAN, UTRAN o GERAN. Debido a que el restablecimiento del plano de usuario supone la creación de un túnel GTP entre la pasarela S-GW y la red de acceso correspondiente (en un nodo eNB para el caso de la red E-UTRAN, y el nodo SGSN para el caso de las redes UTRAN y GERAN o el controlador RNC para el caso de la red UTRAN con soporte de la interfaz S12), la misma pasarela S-GW se encarga de ordenar la finalización del mecanismo de búsqueda en la entidad MME o el nodo SGSN donde no se encontraba accesible al teléfono. La figura 2.35, muestra la configuración de los contextos de movilidad cuando un usuario opera en el modo ISR.

2.9.4.2 Movilidad entre redes de acceso 3GPP en estado activo.

La gestión de movilidad en un escenario multi-RAT, cuando un teléfono se encuentra conectado, es decir, tiene establecida una conexión de radio con una de las tres posibles redes de acceso 3GPP, la movilidad del teléfono entre celdas de diferentes tecnologías se gestiona mediante procedimientos de movilidad inter-RAT. Este mecanismo de movilidad inter-RAT se controla siempre desde la red de acceso origen, la cual decide el cambio hacia una celda de otra red de acceso teniendo en cuenta medidas enviadas desde los equipos de usuario, tal como la movilidad controlada por la red y asistida por el teléfono. Previamente al iniciar la movilidad, la red de acceso origen inicia la preparación del mismo hacia la red de acceso destino a través de la red troncal EPC4. La fase de preparación permite reservar recursos en la red de acceso destino, tales como el planteamiento antes de realizar el rompimiento mientras el teléfono sigue conectado en la celda de la red de acceso origen. Así mismo, la señalización intercambiada en la fase de preparación permite que en el momento del cambio de celda, el teléfono disponga de información específica de la operación de las capas de radio en la celda destino. Esta información se refiere a los parámetros de configuración de los canales de acceso, información de sistema, etc., que le permita re-establecer la conexión en la nueva red con el mínimo tiempo de interrupción posible. Es importante mencionar que el inicio de la ejecución del cambio de celda en el teléfono siempre lo ordena la red de acceso origen. Los procedimientos de movilidad entre redes E-UTRAN y UTRAN, y entre redes E-UTRAN y GERAN, en cualquiera de los sentidos, son bastante similares. En la figura 2.36 se muestra un procedimiento de movilidad desde una red E-UTRAN a otra red UTRAN como un caso representativo de los procedimientos de movilidad inter-RAT. Los diferentes pasos del procedimiento se describen a continuación.

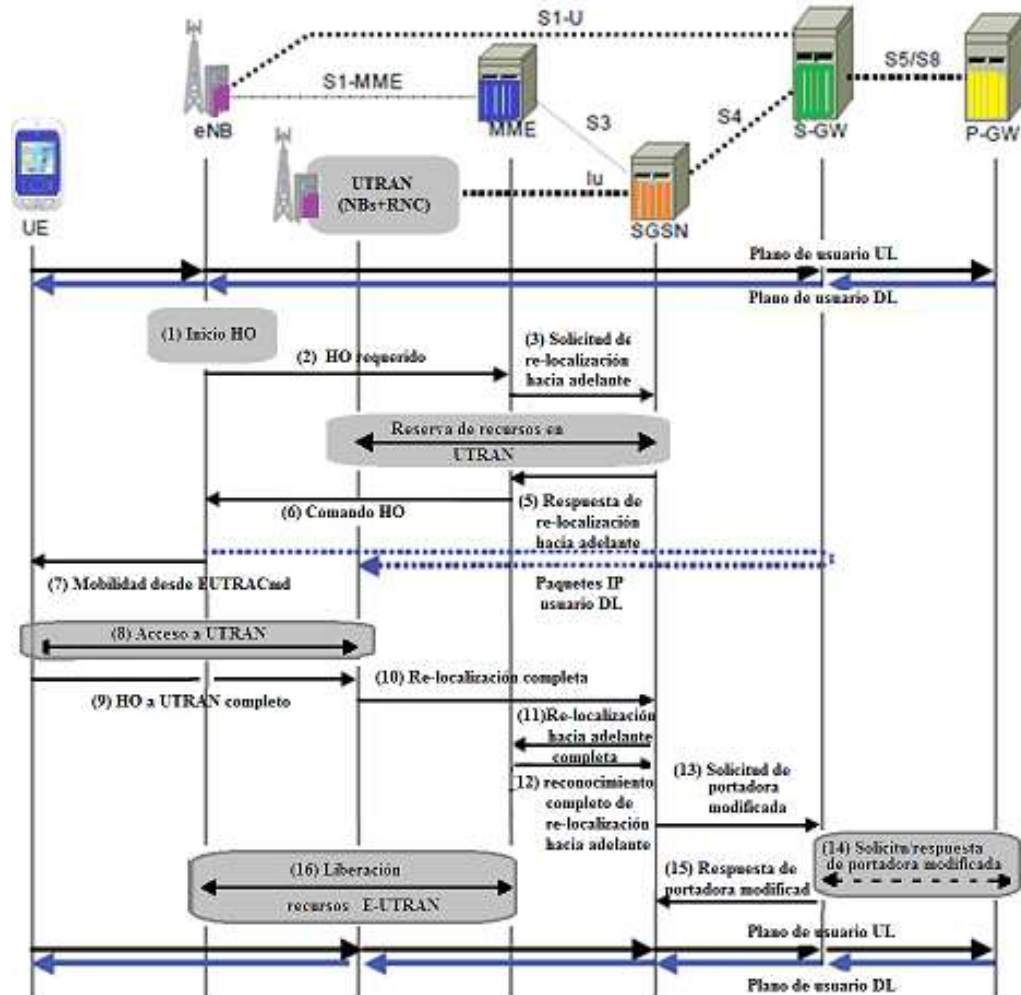


Figura 2.36. Procedimiento de movilidad entre las redes E-UTRAN a UTRAN

(1) Iniciación de la movilidad (HO por sus siglas en Inglés)

El procedimiento de movilidad lo inicia el nodo eNB a través del cual se conecta el teléfono. La decisión de realizar la movilidad puede ser debida a que, por ejemplo, las medidas recibidas desde el teléfono indican la recepción de señal en el teléfono de una celda de la red UTRAN con mejores condiciones que la celda actual en la red E-UTRAN.

(2)-(6) Preparación de la movilidad.

El nodo eNB origen envía la solicitud de movilidad (mensaje “HO requerido”) hacia la entidad de movilidad MME de la red troncal. El mensaje contiene información que permite identificar la red de acceso UTRAN, es decir, el identificador del equipo controlador RNC de la red UTRAN donde se encuentra la celda hacia la que se pretende dirigirla movilidad. A partir de esta información, la entidad MME determina que se trata de movilidad inter-RAT entre las redes E-UTRAN y UTRAN e inicia el procedimiento llamado relocalización directa (*Forward Relocation*) con el nodo SGSN correspondiente. Este procedimiento permite transferir el contexto del usuario, información tal como: conexiones PDN, servicios portadores establecidos, contexto de movilidad, etc., de la entidad MME origen hacia el nodo SGSN destino. De esta forma, el nodo SGSN destino conoce las características de los servicios portadores activos. Tales como los parámetros

de calidad de servicio y puede proceder a reservar los recursos necesarios en la red de acceso destino UTRAN, recursos tales como la activación de Portador de Acceso de Radio (RABs por sus siglas en Inglés), en la red UTRAN. El nodo SGSN es el encargado de mapear los parámetros de calidad de servicio utilizados en los servicios portadores del sistema EPS con los parámetros de calidad de servicio utilizados en la red UMTS. El procedimiento de reserva de recursos en la red UTRAN paso (4) de la figura 2.36 se soporta mediante los procedimientos existentes en la interfaz Iu entre los nodos SGSN y los controladores RNC. De hecho, el procedimiento completo de relocalización directa (*Forward Relocation*) es una extensión del procedimiento ya soportado en redes UMTS para realizar movilidad entre nodos SGSNs. Una vez finalizada la reserva de recursos en la red UTRAN, el nodo RNC destino queda ya preparado para la recepción de paquetes IP del usuario en la dirección descendente. Finalmente, el nodo SGSN destino responde a la petición de relocalización directa (*Forward Relocation*) a la entidad MME origen y ésta a su vez señala al nodo eNB cuando termina la etapa de preparación de la movilidad. Los mensajes (5) y (6) pueden transportar de forma transparente información relativa a parámetros de configuración de la nueva conexión radio de la terminal con la red UTRAN. En este momento, se pueden establecer túneles en la red de forma que el tráfico IP en sentido descendente que llegue al nodo eNB origen sea redirigido hasta el controlador RNC destino. (7)-(12) la ejecución de la movilidad en la interfaz de radio del teléfono que recibe la orden de conmutar su conexión de radio hacia una celda de la red UTRAN mediante la señalización de control RRC (mensaje *MobilityFromEUTRACmd* mostrado en el paso (7). A través de esta señalización, el teléfono puede recibir los parámetros de configuración de radio necesarios para proceder al acceso a través de la red UTRAN. A partir de este instante, el teléfono sincroniza su interfaz de radio con la celda correspondiente de la red UTRAN e inicia el acceso a la red destino. Una vez que el teléfono ha establecido una nueva conexión en la red UTRAN, el controlador RNC indica la finalización del proceso de relocalización mediante el mensaje “Relocalización Completa”. En este momento, la entidad SGSN destino puede ya recibir tráfico en dirección ascendente proveniente de la red UTRAN que se encamina hacia la pasarela S-GW. Finalmente, el nodo SGSN informa a la entidad MME origen del resultado de la movilidad para que pueda proceder a la liberación de recursos en la red de acceso E-UTRAN origen.

(13)-(15) Actualización de las conexiones en la red troncal EPC La entidad SGSN interactúa con la pasarela S-GW para actualizar el contexto de datos asociado al usuario y conmutar el túnel GTP todavía establecido en la interfaz S1 con el nodo eNB origen hacia la entidad SGSN a través de la interfaz S4. En caso de que la red UTRAN soporta la funcionalidad denominada modo directo (*Direct Mode*), el túnel GTP podría también establecerse entre la pasarela S-GW y el controlador RNC de la red UTRAN, evitándose que el tráfico de usuario atraviese la entidad SGSN.

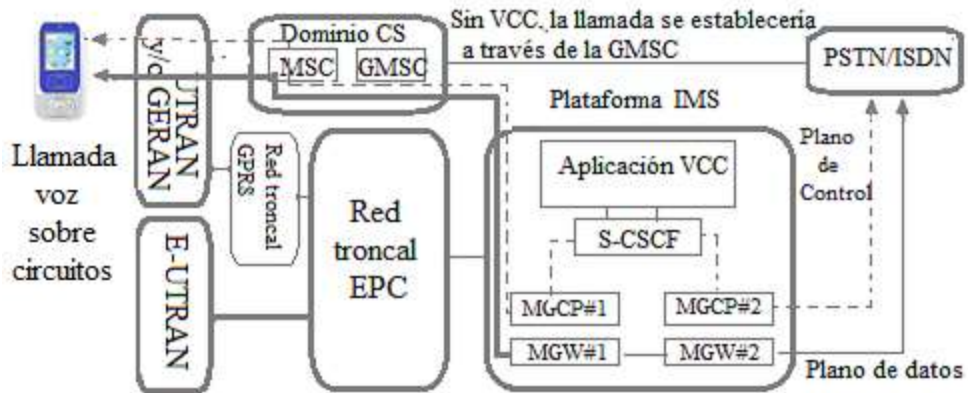
(16) Liberación de recursos

Finalmente, la entidad MME indica al nodo eNB origen que la movilidad se ha realizado con éxito y que puede proceder a la eliminación del contexto y recursos asociados al usuario. El procedimiento mostrado en la figura 2.36, corresponde a una solución de interconexión basada en la utilización de las interfaces S3/S4/S6d en tecnologías GPRS. Para mayor información de la movilidad ver referencia [46].

2.9.5 Soluciones de interconexión para los servicios de voz

En redes de telecomunicaciones rurales se considera que el servicio básico para este propósito es el servicio de voz, es por esto que analizaremos su funcionamiento desde el punto de vista de la tecnología de redes 3GPP. Dicha tecnología propone dos soluciones diferentes para este fin: una se basa en la utilización del dominio de conmutación de circuitos de la red troncal GSM/UMTS y la otra se realiza en el subsistema IMS, accesible a través del dominio de conmutación de paquetes GPRS y EPC. En el caso de las redes de acceso UTRAN y GERAN, ambas redes proporcionan acceso tanto al dominio de circuitos como al dominio de paquetes de forma que la provisión del servicio de voz admite ambas soluciones. Por otro lado la red de acceso E-UTRAN no proporciona acceso al dominio de circuitos por lo que la solución IMS constituye la elección natural para soportar el servicio de voz. El servicio de voz que se ha estandarizado para operar en el subsistema IMS se conoce como Telefonía Multimedia (MMTel por sus siglas en Inglés) [47]. Tratando de ofrecer un servicio de voz con calidad de servicio en redes E-UTRAN dicho servicio debe complementarse con el servicio de voz del dominio de circuitos en las zonas donde la red E-UTRAN no tenga cobertura. Teniendo en cuenta que el mecanismo conocido como *Continuidad de la Llamada de Voz* (VCC por sus siglas en Inglés) permitirá proporcionar continuidad del servicio de voz entre el dominio de circuitos de GSM/UMTS y el servicio MMTel soportado en el subsistema IMS. El servicio VCC se ha desarrollado como una aplicación del subsistema IMS que se ejecuta en la plataforma IMS del operador matriz. A través de dicha aplicación, las llamadas de una terminal que soporte el servicio VCC, denominado simplemente como terminal VCC, pueden ser ancladas en la plataforma multimedia IMS del operador matriz, tanto si las llamadas se inician desde el dominio de circuitos como a través del subsistema IMS. El anclaje se consigue mediante la utilización de mecanismos de re-encaminamiento de llamadas que ya existen en el dominio de circuitos. De esta forma, una llamada originada por un terminal VCC en el dominio de circuitos se encamina hacia la plataforma IMS de su operador, y a partir de aquí, se continúa el encaminamiento de la llamada hasta el usuario destinatario. Debido a este anclaje resulta posible que una llamada de voz iniciada en la red UTRAN por una terminal VCC a través del dominio de circuitos sea transferida hacia el dominio de paquetes al subsistema IMS sin interrumpir el servicio y viceversa. Las transferencias entre dominios de circuitos CS y multimedia IMS pueden realizarse en las dos direcciones y en el transcurso de una sesión, la aplicación del servicio VCC puede soportar múltiples cambios.

(a) Servicios de voz a través del dominio de circuitos y anclaje en IMS



(b) Servicio de voz a través del IMS

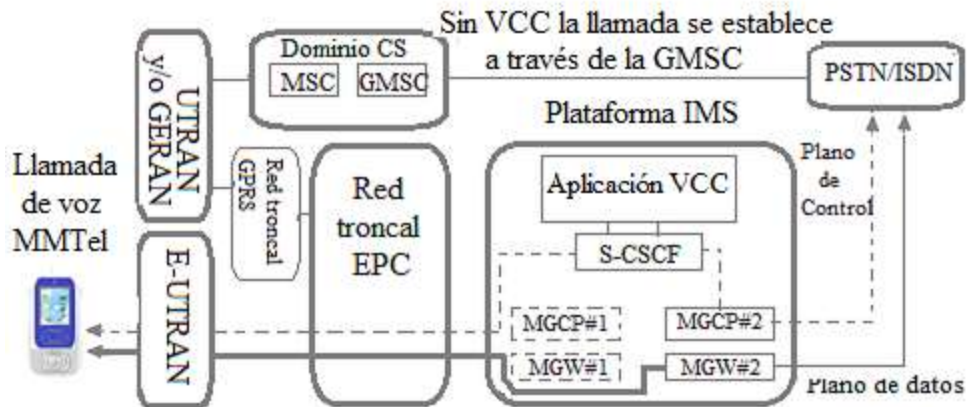


Figura 2.37. Modelo de interconexión en la entidad VCC

La figura 2.37 muestra un modelo de interconexión de la entidad VCC donde una llamada establecida a través del dominio de circuitos de GSM/UMTS, en lugar de ser encaminada a través esta red de circuitos lo hace a través de las pasarelas *Gateway MSC* y GMSC, se encamina hacia la plataforma IMS. Allí, la aplicación VCC realiza funciones de anclaje de la señalización asociada a la llamada. De esta forma, si durante el transcurso de la llamada la terminal cambia su conexión hacia una red de acceso E-UTRAN, el servicio de voz con la red telefónica puede mantenerse a través del subsistema IMS, tal como se ilustra en la gráfica (b) de la figura 2.37. Una descripción detallada de la arquitectura funcional del servicio VCC y los procedimientos asociados se proporcionan en el documento 3GPP TS 23.206 [48]. Es importante destacar que el servicio VCC actualmente sólo se utiliza para llamadas de voz, y por tanto no aplica a otros servicios proporcionados sobre el dominio de circuitos como, por ejemplo video llamadas. La solución VCC no es propia del sistema LTE sino que se introdujo para resolver la misma problemática de continuidad de llamadas entre el dominio de circuitos y el subsistema IMS en el contexto de redes UMTS. Una limitación importante de la solución VCC es que esta solución del subsistema IMS constituye la elección

natural para soportar el servicio de voz. Una limitación importante de la solución VCC es que está planteada para utilizarse con terminales multimodal de radio-doble. El hecho de utilizar terminales radio-dual hace que no se requiera una interacción entre las redes troncales de conmutación de circuitos y de conmutación de paquetes GPRS o EPC para soportar el cambio de red de acceso. Sin embargo, tal como sucede con las actuales terminales multimodal 2G/3G, sabemos que la mayoría de terminales multimodal que ofrezcan soporte para redes E-UTRAN son terminales de radio-simple. Debido a estas situaciones el organismo 3GPP ha especificado la solución conocida como VCC de radio simple (SRVCC Por sus siglas en Inglés) [49]. La solución SRVCC se fundamenta en el mismo principio de funcionamiento de la solución VCC descrito en este párrafo con la salvedad de que la solución puede ser aplicada en el caso de terminales radio-simple. Para ello es necesaria la introducción de nuevas interfaces de red que interconecten la entidad de red MME de la red troncal EPC con las unidades de conmutación MSC del dominio de circuitos. Una explicación más detallada de las soluciones VCC y SRVCC, así como otras propuestas que se han planteado para el soporte del servicio de voz sobre LTE quedan fuera del alcance de esta tesis.

2.10 Interconexión entre LTE y otras redes no 3GPP

2.10.1 Descripción general

La red troncal EPC se considera como la Red Principal Evolucionada y ha sido diseñada para proporcionar servicios de conectividad IP a terminales que utilicen redes de acceso de radio o redes de cable que son diferentes a las especificadas por la organización 3GPP [50]. En la figura 2.38 se ilustra un diagrama a bloques de la interconexión del sistema LTE con otras redes de acceso que no son del organismo 3GPP. La red troncal EPC, junto con otros elementos comunes estandarizados por el 3GPP, constituida de las unidades: base de datos HSS, subsistema PCC, subsistemas de facturación, etc., constituyen el núcleo de la arquitectura de red heterogénea donde se muestran las funciones requeridas para la provisión del servicios y que son independientes de la red o redes de acceso utilizadas. La arquitectura del grupo 3GPP, realiza las siguientes funciones relacionadas con la gestión de usuarios: almacenamiento y acceso a los datos de suscripción, control de acceso, el control de calidad de servicio y tarificación del servicio de conectividad, y el soporte a la movilidad entre las diferentes redes de acceso. La arquitectura propuesta por la organización 3GPP no impone ninguna restricción a las redes de acceso que no pertenecen a su organización, salvo el hecho de que sean redes que proporcionen un mecanismo de transferencia de paquetes IP. Ejemplos de estas redes de acceso son: redes de acceso radio CDMA2000, redes de banda ancha cableadas con terminaciones inalámbricas, conocidas como las redes tradicionales (*Legacy* por ejemplo el servicio ADSL con puntos de acceso inalámbricos IEEE 802.11) y redes de acceso Móviles WiMAX.

El marco planteado no presupone ninguna interacción entre las diferentes redes de acceso, es decir, las redes de acceso pueden estar completamente desacopladas unas de otras. Por tanto, la red troncal EPC actúa como punto de anclaje del servicio de conectividad IP, ofrecido a través de cualquiera de las redes de acceso 3GPP y no 3GPP,

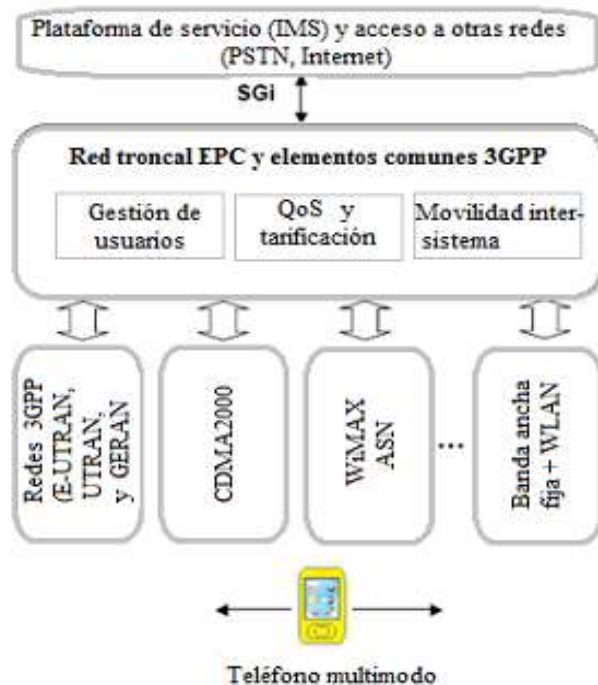


Figura 2.38. Arquitectura de interconexión de la organización 3GPP con diferentes redes.

garantizándose la continuidad del servicio al cambiar de red de acceso asegurando que la dirección IP asignada a la terminal se mantiene aunque ésta cambia de red de acceso). Considerando que las redes de acceso que se conectarán a la red EPC son de una gran diversidad, así como las diferentes opciones tecnológicas existentes para permitir la movilidad en la capa de red entre sistemas, el organismo 3GPP ha diseñado las especificaciones para aceptar varias soluciones de interconexión entre el sistema LTE y las redes de acceso que no son del organismo 3GPP. Las posibles soluciones pueden clasificarse en torno a dos aspectos fundamentales:

- Redes de acceso que no pertenecen al grupo 3GPP se clasifican en confiables y no confiables.
- El protocolo de movilidad de capa de red utilizado network-based o host-based.

A continuación se describen ambos aspectos. El término confiable se utiliza para indicar que la red de acceso que no es del grupo 3GPP dispone de los mecanismos de seguridad apropiados que hacen que el operador de la red troncal EPC a la que se conecta la considere como tal, estos mecanismos son: de autenticación del usuario, servicios de confidencialidad e integridad, etc. Las especificaciones del organismo 3GPP no definen los requerimientos técnicos que debe cumplir una red no 3GPP para ser considerada como confiable. Así, la consideración como tal está sujeta al criterio del operador que ofrece su servicio de conectividad IP a través de dicha red. Ejemplos de redes confiables podrían ser redes de acceso CDMA2000 o móviles (*Mobile WiMAX*) conectadas a la infraestructura de la red troncal EPC mediante una red segura. En la figura 2.38 se muestra un escenario de interconexión con una red de acceso móvil (*Mobile WiMAX* denominada Red de Servicio de Acceso según las especificaciones *WiMAX NWG Release 1* [51]). También muestra como la información cursada por un teléfono conectado a través de la

red de acceso Mobile WiMAX (utilizando las credenciales de su USIM) es encaminada desde la pasarela ASN GW (*ASN Gateway*) de la red de acceso hasta la red troncal EPC desde donde se accede a las redes externas (e.g., Internet, plataformas de servicios del subsistema IMS).

2.11 Arquitectura de red rural NGN propuesta

La arquitectura propuesta está basada en las recomendaciones de la organización 3GPP para Subsistemas Multimedia, en lo propuesto por la Fuerza de Trabajo de Ingeniería de Internet (IETF) para protocolo de señalización SIP, en lo definido por el instituto UIT para modelo de Calidad de Servicio; así como en los organismos 3GPP e IEEE para enlaces inalámbricos (GSM, UMTS y LTE Advanced). Esta arquitectura inicia como una plataforma de servicios con tecnología basada en el Protocolo de Internet (IP por sus siglas en Inglés) y servicios inalámbricos y de movilidad de redes, así mismo permite la convergencia con las redes fijas (tradicionales) [52] [53] [54]. En este trabajo se utiliza el protocolo SIP del grupo IETF para las funciones de señalización y control que requieren las NGN. Las figuras 2.39, y 2.40, muestran la arquitectura de red rural propuesta en esta tesis, la cual está estructurada en cuatro capas de red (acceso, transporte, control y aplicación). Esta arquitectura conectará a usuarios rurales a través de las diferentes redes de acceso utilizando redes inalámbricas de siguiente generación (NWGN por sus siglas en inglés) tales como GERAN, UTRAN, y E-UTRAN; así como también redes conmutadas de circuito (RTPC) y redes WLAN. Las Centrales Telefónicas denominadas (IMS-AGC) mostradas en la figura 2.39 están constituidas por la Red Troncal de Transporte (EPC por sus siglas en Inglés), la red de gestión y control (IMS), así como los servidores de aplicación. Esta propuesta considera la utilización del modelo E del organismo UIT [Rec. G.107] [55] para garantizar la calidad de servicio de voz.

2.12.1 Introducción

El modelo matemático propuesto para esta tesis está basado en el análisis sistemático del establecimiento y liberación de sesiones SIP sobre redes IMS, y provee un método para calcular el Ancho de Banda requerido para el establecimiento de las sesiones SIP sobre redes UDP. Sin embargo, el protocolo SIP puede ser utilizado sobre otros protocolos de transporte tales como: TCP o SCTP. El modelo contempla las pérdidas generadas por los servidores sobrecargados o enlaces con pérdidas los cuales producen retardos en el establecimiento de sesiones afectando la calidad de servicio percibida por los usuarios. El modelo considera como entradas características de tráfico; el número de sesiones por segundo y el tiempo de retención promedio, y características de la red, tales como pérdidas y retardos de propagación. La salida del modelo proporciona el ancho de banda y el retardo necesario para establecer y liberar con éxito una sesión cuando se utiliza el protocolo SIP sobre redes UDP.

2.12.2 Arquitectura IMS

La figura 2.24, muestra la arquitectura simplificada del subsistema IMS, la cual será validada utilizando el modelo matemático utilizado para este propósito.

2.12.3 Protocolo SIP en el subsistema IMS

Todas las rutas de señalizaciones de sesiones de voz IP y servicios multimedia en el subsistema IMS se realizan a través del protocolo SIP para permitir la introducción de nuevos servicios y la integración con los servicios de red IP fijos. Los mensajes SIP están constituidos de una solicitud con su respuesta, lo cual implica un tiempo específico de transmisión de datos. Una solicitud indica que un usuario está iniciando o terminando una sesión (solicitud de inicio de conexión *Invite*, solicitud de terminación *BYE*). La solicitud *Invite* la crea el remitente para establecer una sesión entre dos usuarios, y la solicitud *BYE* pone fin a este período de sesiones en un diálogo. La respuesta puede ser definitiva o provisional. Respuestas finales indican que una solicitud se recibió exitosamente y se atendió correctamente por su destinatario. Así mismo, una respuesta final indica que la solicitud no pudo ser procesada por el destinatario o por algún servidor *proxy* o que la sesión no se puede establecer a tiempo por alguna razón. Respuestas provisionales indican que el establecimiento de sesión está en curso, por ejemplo, el teléfono del destino está

sonando pero el usuario no lo ha contestado aun. Los servidores *proxy* del protocolo SIP pueden operar en dos modos (con estado o sin estado). En el primer caso, el servidor proxy reenvía una solicitud a su destino y mantiene información del estado acerca de la solicitud enviada hasta que se recibe una respuesta de esta solicitud o el tiempo programado para la respuesta haya expirado. Cuando se utiliza un protocolo de transporte como UDP, si el servidor proxy no recibió una respuesta después de algún tiempo, volverá a enviar la solicitud. Por otro lado, el modo sin estado, el servidor proxy envía la solicitud pero no mantiene la información del Estado. En este caso, el remitente retransmite la solicitud si no recibió una respuesta. El protocolo SIP utiliza un comportamiento de retransmisión exponencial. Así que si un remitente de un mensaje SIP no recibe una respuesta después de algún tiempo, volverá a enviar la solicitud después de un tiempo de espera. En caso de que no haya recibido una respuesta para la retransmisión, el remitente incrementa el tiempo de espera e intenta otra vez hasta un cierto número de retransmisiones, [56] [57]. En general se puede distinguir entre dos modos de retransmisión en el protocolo SIP:

2.12.3.1 Retransmisiones para invitación a conectarse (*Invite*)

Este nombre se le da a las solicitudes *Invite*, así como algunos otros mensajes intercambiados durante el establecimiento de la sesión. En este modo, el remitente retransmite un mensaje después de un periodo de T1 segundos si no ha recibido confirmación de recepción. El temporizador de retransmisión se incrementa de forma exponencial hasta un tiempo máximo llamado *TimerB* [32 seg.]. Una vez que este tiempo es alcanzado el remitente elimina el mensaje y detiene la retransmisión.

2.12.3.2 Retransmisiones *no-Invite*

Esta acción se aplica a todas las solicitudes que no sean *Invite*. En este modo el remitente retransmite un mensaje de confirmación, si no se recibió después de T1 segundos. El temporizador aumenta de forma exponencial hasta un máximo de tiempo denominado en el modelo como T2 segundos. Una vez que el contador es alcanzado el remitente continúa la retransmisión de la solicitud cada T2 segundos hasta un límite máximo denominado *TimerF*.

2.12.3.3 Fases de establecimiento de una sesión IMS

La figura 2.41 muestra el protocolo SIP requerido para poder establecer una sesión IMS entre la persona que llama (Agente de Usuario Cliente UAC) o también remitente y el (Servidor Agente de Usuario UAS) o destinatario de la llamada. El ejemplo se basa en la hipótesis prevista en la propuesta de [58]. El establecimiento de la sesión IMS se

produce por el envío de una solicitud *Invite* y la respuesta correspondiente. En general, el establecimiento de sesión se compone de ocho fases, a saber:

[Primera Fase].- Intentando Invite / 100, (1-14)

Esta fase se inicia con el envío de una solicitud *Invite* y termina cuando el cliente recibe una respuesta provisional o definitiva. Esta fase consiste en realidad de diferentes partes, la comunicación entre el cliente UAC y el P-CSCF, P-CSCF y S-CSCF. Una vez que el P-CSCF recibe la solicitud *Invite* del UAC, este responde con una respuesta provisional, es decir, 100 o 180 para el remitente. Lo mismo se aplica entonces en cada salto hasta que el *Invite* llega al UAS. En caso de que el *Invite* o la respuesta provisional se pierdan, el UAC retransmitirá la solicitud después que ha vencido el tiempo T1 de espera de la llamada. Si el *Invite* retransmitido por el P-CSCF al S-CSCF o la respuesta provisional retransmitida por el S-CSCF al P-CSCF se pierden, el P-CSCF retransmitirá el *Invite* después de T1 segundos. Lo mismo se aplica en todos los saltos entre servidores SIP, hasta que el receptor de la llamada reciba el *Invite*. Se debe tener en cuenta que la solicitud *Invite* es la única solicitud en que se envían los acuses de recibo salto por salto. En todas las otras fases descritas aquí, el UAC es responsable de la generación de los acuses de recibo correspondientes de extremo a extremo.

[Segunda Fase].- Progreso de Sesión 183 / PRACK (15-26):

En esta fase el remitente y el destinatario de la llamada negocian los códigos de audio y vídeo que se utilizaran en la llamada, así como los criterios de calidad de servicio. La respuesta provisional 183 se envía confiablemente. Es decir, el UAS de la llamada la retransmitirá utilizando el modo de *Invite* (hasta que se haya recibido un PRACK).

[Tercera Fase].- PRACK/ 200 OK (22-31):

En esta fase la persona que llama reconoce la recepción de la respuesta 183. Luego el que llama retransmitirá la solicitud PRACK (hasta que se reciba un 200 OK) utilizando el modo de retransmisión *no-Invite*. Debe tomarse en cuenta que el PRACK iniciará su *Timer* de retransmisión tan pronto como reciba la respuesta 183. Además, la retransmisión PRACK no se retransmitirá cada vez que se reciba una respuesta 183 sino hasta que el *Timer* de retransmisión se active (es decir, T1, 3T1, etc.).

[Cuarta Fase].- UPDATE / 200 OK (32-41):

En esta fase el remitente y el destinatario de la llamada negocian el código y la QoS. En el caso de que se pierda la solicitud UPDATE o la confirmación 200 OK la persona que llama retransmitirá nuevamente la solicitud UPDATE hasta que reciba el mensaje de confirmación 200 OK utilizando el modo retransmisión *no-Invite*.

[Quinta Fase].- Timbrado 180 / PRACK (42-52):

En esta fase, el destinatario de la llamada informa a la persona que llama que está siendo alertado sobre la llamada. La respuesta provisional 180 es enviada confiablemente. Es decir, el destinatario de la llamada, la retransmitirá utilizando el modo de retransmisión *Invite*, pero hasta que haya recibido un PRACK.

[Sexta Fase].- PRACK/ 200 OK (48-57):

En esta fase la persona que llama reconoce la recepción de la respuesta 180. La solicitud PRACK será retransmitida por la persona que llama hasta que reciba el mensaje de confirmación 200 OK utilizando el modo de retransmisión no-*Invite*.

[Séptima Fase].- Respuesta Final (200 OK) / ACK (58-68):

En esta fase, el UAS informa al UAC que la llamada fue aceptada. En el caso de que se pierda la respuesta o el ACK, el UAS retransmitirá la respuesta 200 OK utilizando el modo de retransmisión *no-Invite*, esto hasta que un ACK se haya recibido.

[Octava Fase].- BYE / 200 OK:

Durante esta fase, uno de los usuarios termina la llamada. En caso de que las respuestas BYE o 200 OK se pierdan, el remitente de la solicitud BYE retransmitirá esta hasta que reciba el mensaje de confirmación 200 OK en el modo de retransmisión no-*Invite*.

2.12.3.4 Modelado del establecimiento de la sesión IMS

Con base en los parámetros presentados, es posible desarrollar un modelo de red para el establecimiento de sesión que considere las fases correspondientes. A continuación, se presenta un modelo gráfico teórico de las técnicas de retransmisión SIP en las redes de datos con pérdida, también se calculará el ancho de banda para la puesta de sesión IMS creada y se estimará del retardo de dicha sesión. La figura 2.41, muestra el escenario del establecimiento de una sesión SIP básica, donde una sesión o llamada telefónica tiene que atravesar cinco servidores proxies, durante el proceso de sesión. Cada enlace de la red representada tiene una tasa de pérdida de (l) y un retardo de propagación de (D) segundos.

La gráfica horizontal en la figura 2.41, describe el recorrido a través de los diferentes servidores SIP que son tocados por los mensajes de solicitudes y respuestas retransmitidos (provisionales y finales), creados durante el proceso temporal de señalización cuando un suscriptor (en estado de itinerancia) desea contactar con otro. La gráfica vertical muestra las ocho fases (Invite y no-Invite) que deben realizarse y que forman parte del proceso SIP en el establecimiento de una sesión o llamada. Además muestra a los diferentes servidores SIP mencionados cuando se encuentran en proceso de solicitud de sesión, hasta localizar al usuario buscado pasando por diferentes redes que participan en el proceso.

2.12.3.5 Servidores de red matriz y de red visitada

Para propósito de descripción de las rutas de señalización que realiza el protocolo SIP a través de los servidores CSCF durante el establecimiento de una sesión o llamada de un usuario para conectarse a la red, denotaremos los términos servidores matriz, a los servidores que pertenecen a la red de casa, y servidores de red visitada aquellos que se encuentra fuera de casa visitando otras redes.

2.12.3.6 Modelado de la retransmisión de mensajes SIP

2.12.3.6.1 Modelado de la retransmisión de la solicitud Invite

Para las solicitudes Invite, el comportamiento exponencial de retransmisión dura hasta alcanzar un temporizador llamado TimerB. Eso es una solicitud que se retransmite en los siguientes instantes de tiempo $T1$, $3T1$, $7T1$, $15T1$, hasta alcanzar TimerB. Esto puede representarse como una serie como lo muestra la ecuación (2.1):

$$(2^1 - 1)T, (2^2 - 1)T1, (2^3 - 1)T1 \dots \dots (2^{N_i} - 1)T \quad (2.1)$$

donde el instante $(2^{N_i} - 1)T1 = TimerB$, de ésta manera, el número máximo de solicitudes *Invite* retransmitidas N_i se expresa por la ecuación (2.2). Donde *TimerB* representa el número máximo de segundos para alcanzar el estado estable y $T1$ representa el

$$N_i = \left\lceil \log_2 \left(\frac{TimerB}{T_1} + 1 \right) \right\rceil \quad (2.2)$$

instante temporal de envío de solicitudes. Si consideramos que la carga total de tráfico en la red es λ (sesiones/unidad de tiempo) y que l es el porcentaje de pérdida sobre una tasa λ de solicitudes *Invite* emitidas, se pierden $\lambda \times l$ paquetes en promedio por T_1 segundos. Estos paquetes deben retransmitirse T_1 segundos después. Los paquetes retransmitidos también sufren de una pérdida y tendrán que ser retransmitidos posteriormente. Por lo tanto, la tasa de generación de llamada (R_i) se puede representar como se muestra en la Tabla 2.9. En el instante de tiempo 0, se envían λ solicitudes por cada T_1 segundo. Después de T_1 segundos los remitentes continuarán generando un número λ de nuevas solicitudes *Invite* por cada T_1 segundos y retransmitirán las solicitudes pérdidas ($l \times \lambda$), por ejemplo, $\lambda + (l \times \lambda)$ serán enviadas. Fuera de esos ($l \times (\lambda + (l \times \lambda))$) se perderán. Estas se retransmitirán en un tiempo $3T_1$. En el tiempo $2T_1$ serán enviadas λ solicitudes nuevas, más las solicitudes que se perdieron T_1 segundos antes; es decir, ($l \times \lambda$) solicitudes. Fuera de la solicitud enviada en el tiempo T_1 , donde ($l \times (\lambda + (l \times \lambda))$) se perderá. Entonces las solicitudes nuevas se retransmitirán en el tiempo $4T_1$ y así sucesivamente. El número de solicitudes *Invite* (R_i) enviados por el remitente en cualquier instante de tiempo (n), se determina a través de la ecuación (2.3) [59]: Donde observamos que λ es la tasa de sesiones por segundo, l es el porcentaje de pérdidas, y $k = \log_2(n + 1)$ representa el límite de la secuencia de retransmisiones, n puede ser maximizada como la relación ($\frac{TimerB}{T_1}$) que es el número de sesiones cuando se alcanza el tiempo límite.

Tabla 2.9. Retransmisiones de solicitudes *Invite* debido a pérdidas de la red IMS [59]

Tiempo	R_i	Pérdida
0	λ	$l \times \lambda$
1 T_1	$\lambda + l \times \lambda$	$l \times \lambda + l^2 \times \lambda$
2 T_1	$\lambda + l \times \lambda$	$l \times \lambda + l^2 \times \lambda$
3 T_1	$\lambda + l \times \lambda + l^2 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda$
4 T_1	$\lambda + l \times \lambda + l^2 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda$
5 T_1	$\lambda + l \times \lambda + l^2 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda$
...
7 T_1	$\lambda + l\lambda + l^2 \times \lambda + l^3 \times \lambda$	$l\lambda + l^2\lambda + l^3 \times \lambda + l^4 \times \lambda$
8 T_1	$\lambda + l\lambda + l^2 \times \lambda + l^3 \times \lambda$	$l\lambda + l^2\lambda + l^3 \times \lambda + l^4 \times \lambda$
...

En este estado el número de nuevas pérdidas, las pérdidas de solicitudes recién generadas, deben igualarse al número de retransmisiones que serán terminadas como el número máximo de intentos que fueron ya realizados. En esta etapa el sistema alcanza el estado estable.

$$R_i(l, n) = \lambda \times \left(1 + \sum_{m=1}^{m=k} l^m\right) = \lambda \times \sum_{m=0}^{m=k} l^m \quad (2.3)$$

2.12.3.6.2 Pérdidas durante la fase *Invite*

Es importante recordar que en un procedimiento de solicitud de sesión (o llamada) por parte de un subscriptor a la red local donde se encuentra, participaran dos tipos de solicitudes *Invite* y *no-Invite*. Las solicitudes *Invite* se analizaran primero en esta sesión. Las solicitudes *Invite* se definen como aquellas solicitudes que se consideran como “solicitudes seguras” en el proceso de interconexión, ya que operan en base a un procedimiento de salto por salto. Por tanto, si la solicitud *Invite* enviada por el remitente (UAC) o la respuesta 100 enviada por el primer proxy de la red visitada originalmente (P-CSCF) se pierde, el UAC retransmitirá la solicitud *Invite*. Lo mismo se aplica entre cada uno de los dos proxys y entre el último proxy de la Red Matriz (P-CSCF) y el Servidor Agente de Usuario (UAS). Los saltos realizados por los mensajes son independientes unos de otros. Se considera una solicitud exitosa para cada salto si la solicitud y su respuesta llegan a su destino con éxito. Se deben considerar las pérdidas en ambas direcciones, y se determina con la ecuación (2.4).

$$l_e = 1 - (1 - l)^2 \quad (2.4)$$

Por ejemplo, si en el estado estable se envían 10 mensajes por segundo, y la tasa de transmisión es 5 mensajes por segundo entonces significa que en promedio 2 mensajes pueden pasar confiablemente

$$P_i = \frac{R_i(l_e, N)}{\lambda} \quad (2.5)$$

Así que P_i representa el número de solicitudes *Invite* necesarios para establecer una sesión SIP, y se determina (ecuación (2.5)) como el número de mensajes enviados después de alcanzar el estado estable, dividido por el número de mensajes originales, es decir, los mensajes enviados no retransmitidos. N es cualquier punto del tiempo y se calcula con la ecuación (2.6)

$$N = \frac{TimerB1}{T1} \quad (2.6)$$

La transmisión de las respuestas únicamente se activará después de la recepción de una solicitud *Invite*. Con pérdidas de l en una dirección y una tasa de transmisión continua de $R_i = (R_i(l_e, N))$, únicamente $R_i(l_e, N) \times l$ solicitudes *Invite* se pueden recibir. Así pues podemos definir a P_{100} como el número de respuestas enviadas durante el establecimiento de una sesión y se obtiene de la ecuación (2.7), R_i representa el número

$$P_{100} = \frac{R_i(l,N)}{\lambda} \quad (2.7)$$

de solicitudes *Invite* enviados por el remitente en cualquier instante de tiempo (N), y λ es la tasa de sesiones/s. Para enviar exitosamente una solicitud *Invite* a través de η saltos, se tienen que transmitir $(\eta \times P_i + \eta \times P_{100})$ mensajes SIP en promedio.

2.12.3.6.3 Modelado de retransmisiones de solicitudes *no-Invite*.

En esta sesión describiremos a las solicitudes *no-Invite*, las cuales operan con un comportamiento de retransmisión exponencial hasta un *Timer* denominado $T2$ y luego cada $T2$ segundos hasta el instante llamado *TimerF* que corresponde al tiempo máximo permitido para una solicitud *no-Invite*. Esta es una solicitud que se retransmite en el momento de los tiempos $T1$, $3T1$, $7T1$, $15T1$ y hasta $T2$ segundos. Luego a $2T2$, $3T2$, y hasta *TimerF*. Estas retransmisiones pueden ser representadas como una serie representada por la ecuación (2.8).

$$(2^1 - 1)T1, (2^2 - 1)T1, (2^3 - 1)T1 \dots (2^{N_n^e} - 1)T1, 2(2^{N_n^e} - 1)T1 \dots \text{TimerF} \quad (2.8)$$

Cuando $(2^{N_n^e} - 1)T1 = T2$, donde (N_n^e) está definida por la ecuación (2.9) y representa el número de retransmisiones realizadas, descrito en forma exponencial hasta el instante de $T2$ segundos.

$$N_n^e = \left\lceil \log_2 \left(\frac{T2}{T1} + 1 \right) \right\rceil \quad (2.9)$$

Después de $T2$ segundos, el tiempo de espera de retransmisión se mantiene constante a $T2$. Dado que el *Timer* responde por una parte como una función exponencial y por otra como una función lineal, es importante definir el número máximo de retransmisiones *no-Invite* de una solicitud N_n y éste se puede determinar con la ecuación (2.10)

$$N_n = N_n^e + N_n^l = \left\lceil \log_2 \left(\frac{T2}{T1} + 1 \right) \right\rceil + \left\lceil \frac{\text{TimerF} - T2}{T2} \right\rceil \quad (2.10)$$

donde N_n también se puede representar de función (2.11)

$$= \left\lceil \log_2 \left(\frac{T2}{T1} + 1 \right) \right\rceil + \left\lceil \frac{\text{TimerF} - (2^{N_n^e} - 1)T1}{T2} \right\rceil \quad (2.11)$$

N_n^e representa la parte exponencial y N_n^l a la parte lineal, con $(2^{N_n^e} - 1) \times T1$ que indica el instante en el que el remitente va desde una respuesta exponencial a un valor de tiempo constante. Para las solicitudes *no-Invite*, el comportamiento de la transmisión se puede observar en la Tabla 2.10. Así mismo, la recomendación RFC 3261 [60] indica que el

Tabla 2.10. Retransmisiones solicitudes *no-Invite* debido a pérdidas en redes IMS. [59]

Tiempo	Ri	Pérdidas
0	λ	$l \times \lambda$
1 T1	$\lambda + l \times \lambda$	$l \times \lambda + l^2 \times \lambda$
2 T1	$\lambda + l \times \lambda$	$l \times \lambda + l^2 \times \lambda$
3 T1	$\lambda + l \times \lambda + l^2 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda$
4 T1	$\lambda + l \times \lambda + l^2 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda$
5 T1	$\lambda + l \times \lambda + l^2 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda$
...
7 T1	$\lambda + l \times \lambda + l^2 \times \lambda + l^3 \times \lambda$	$l \times \lambda + l^2\lambda + l^3 \times \lambda + l^4 \times \lambda$
8 T1	$\lambda + l \times \lambda + l^2 \times \lambda + l^3 \times \lambda$	$l \times \lambda + l^2\lambda + l^3 \times \lambda + l^4 \times \lambda$
...
15T1	$\lambda + l \times \lambda + l^2 \times \lambda + l^3 \times \lambda + l^4 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda + l^4 \times \lambda + l^5 \times \lambda$
16T1	$\lambda + l \times \lambda + l^2 \times \lambda + l^3 \times \lambda + l^4 \times \lambda$	$l\lambda + l^2 \times \lambda + l^3 \times \lambda + l^4 \times \lambda + l^5 \times \lambda$
...
23T1	$\lambda + l \times \lambda + l^2 \times \lambda + \dots l^5 \times \lambda$	$l\lambda + l^2 \times \lambda + \dots l^6 \times \lambda$
24T1	$\lambda + l \times \lambda + l^2 \times \lambda + \dots l^5 \times \lambda$	$l\lambda + l^2 \times \lambda + \dots l^6 \times \lambda$
...
31T1	$\lambda + l \times \lambda + l^2 \times \lambda + \dots l^6 \times \lambda$	$l\lambda + l^2 \times \lambda + \dots l^7 \times \lambda$
...

valor de T2 es de 4 segundos y T1 es fijado a 0.5 segundos. En estas condiciones, el número de solicitudes *no-Invite* (R_0) enviadas por el remitente en cualquier punto del tiempo (n) se puede determinar con la ecuación (2.12)

$$R_0(l, n) = \begin{cases} \lambda \times (1 + \sum_{m=1}^{m=k} l^m), & n \leq \frac{T_2}{T_1} \\ \lambda \times (1 + \sum_{m=1}^{m=k} l^m) + \sum_{m=k+1}^{m=q} l^m & \text{en otro caso} \end{cases} \quad (2.12)$$

donde $k = \left\lfloor \frac{\ln(n+1)}{\ln(2)} \right\rfloor$, y cuando $(n \leq \frac{T_2}{T_1})$, k debe ser maximizada a N_0^e , y q se obtiene de la función (2.13).

$$q = \left\lfloor \frac{(n-2^{N_n^e}-1) \times T_1}{T_2} \right\rfloor \quad (2.13)$$

lo cual asegura que q se incrementa cada T_2 segundos. Note que el valor máximo de n aquí es $n = \left(\frac{TimerF}{T_1}\right)$ en el cual se alcanza su estado estable, por ejemplo, número de retransmisiones nuevas iguala al número de re-transmisiones terminadas.

2.12.3.6.4 Pérdidas durante la fase *no-Invite*

Debido a la dependencia del desempeño de la red a las pérdidas producto de las características de las redes, es importante proponer el modelo que describe a las distintas pérdidas, en el caso de la fase *no-Invite* se procede como se describe a continuación. Después de enviar una respuesta final, el *UAC* espera recibir una confirmación, antes de T_1 segundos. Por lo tanto, la relación entre la respuesta final y ACK es similar a la que existe entre una *Invite* y una respuesta provisional. A diferencia de las solicitudes *Invite*, la relación entre la respuesta final y ACK es una respuesta de extremo a extremo, por ejemplo, los servidores *proxis* no retransmiten los mensajes perdidos entre ellos. Por lo tanto para determinar el número de la respuesta final (P_f) y la solicitud ACK (P_a) que se necesita para crear una sesión, se pueden utilizar las mismas ecuaciones que antes, pero teniendo en cuenta el retardo de extremo a extremo. Suponiendo que todas las solicitudes siguen el mismo camino, es decir, los mismos *Proxis* registran sus rutas en las solicitudes SIP y con una tasa de pérdida de l sobre cada enlace, la pérdida (L) de extremo a extremo en una dirección de una solicitud que atraviesa η saltos se obtiene de la ecuación (2.14).

$$L = 1 - (1 - l)^\eta \quad (2.14)$$

La pérdida de extremo a extremo para una solicitud más su respuesta se obtiene de la ecuación (2.15).

$$L_e = 1 - (1 - L)^2 = 1 - (1 - l)^{2\eta} \quad (2.15)$$

Es importante mencionar que el exponente al cuadrado en la expresión anterior se relaciona con el proceso de ida y vuelta de los paquetes de señalización durante el proceso del establecimiento de la llamada. En base a este modelo propuesto se calculará el consumo del ancho de banda de la Señalización SIP en el IMS cuando se establece y posteriormente se libera una sesión o llamada en una red con pérdidas. Antes de iniciar el cálculo se distinguirá entre las 8 fases del establecimiento de una llamada en términos de los tipos de retransmisión *Invite* y *no-Invite*.

2.12.3.7 Consumo de ancho de banda del establecimiento de sesión IMS en una red con pérdida.

Con base en la propuesta del modelo, es viable obtener el comportamiento de la tasa de datos generada, considerando los paquetes, pérdidas y tiempos límite involucrados en el proceso de puesta de sesión, por lo tanto es importante describir matemáticamente dicho proceso, considerando las solicitudes *Invite* y *no-Invite* que participan en dicho proceso.

2.12.3.7.1 Intentando fase *Invite* /100 (*Invite*)

Con el propósito de conocer el ancho de banda estimado que se generara durante una llamada o puesta de sesión entre dos usuarios de la red, se requiere identificar el tráfico de paquetes durante la puesta de sesiones, y para esto empezaremos con las solicitudes *Invite*. Utilizando la función (2.16) se obtiene $R_i(l_e)$ que representa las retransmisiones

$$R_i(l_e) = \lambda \times \sum_{m=0}^{m=N_i} l_e^m = \lambda \times \sum_{m=0}^{m=N_i} (1 - (1 - l)^2)^m \quad (2.16)$$

Invite con pérdidas en dos direcciones al solicitar el servicio de conexión de un usuario con otro. Recordemos que todas las solicitudes 100, pertenecen a la fase *Invite*, y en el presente caso, las variables de la ecuación (2.17) se calculan con la ecuación (2.2), donde

$$R_{100}(l) = \lambda \times \sum_{m=0}^{m=N_i} (l)^m \quad (2.17)$$

N_i representa el número máximo de retransmisiones *Invite* y es directamente proporcional al *TimerB* e inversamente proporcional a T1. Recordemos que el *TimerB* de acuerdo a la recomendación RFC 3261 es igual a 64*T1 y T1 es igual a un tiempo de 500 [ms].

2.12.3.7.2 Progreso de sesión fase *183/PRACK/200 OK* (*Invite*)

Una vez determinado el proceso de retransmisión es necesario identificar su efecto en el progreso de la sesión solicitada, ya que éste progreso afecta el consumo de ancho de banda en el proceso de la sesión; de esta manera la fase 183 y PRACK tienen su *timer* de transmisión independiente propio; tan pronto como 183 es recibido por el UAS este inicia su propio *Timer* de retransmisión *PRACK*; el UAC retransmite el mensaje 183 en una forma de retransmisión *Invite*. La recepción de un 183 retransmitido no hará que el destinatario de la llamada retransmita el mensaje *PRACK*. Como cualquier solicitud *no-Invite*, el *PRACK* se retransmite periódicamente hasta un intervalo máximo de 4 segundos, y siempre que no se haya recibido una respuesta 200 OK. Por lo tanto el número de mensajes 183 enviados dependerá de las pérdidas en una dirección, como del tiempo necesario para que un mensaje *PRACK* alcance al remitente, el cual también depende de las pérdidas en una dirección entre el destinatario de la llamada y la persona que llama. La

respuesta provisional confiable se retransmite periódicamente, e inicia con un intervalo de retransmisión de 500 [ms], y se duplica este intervalo de tiempo después de cada retransmisión, hasta un máximo de 32 segundos. Esto refleja el comportamiento de las respuestas *Invite* en [61]. Si el PRACK no se ha recibido para esa respuesta después de 96 segundos se considera una falla de la red. La ecuación (2.18) calcula la respuesta provisional 183 (R_{183}) considerando las pérdidas de extremo a extremo en una dirección.

$$R_{183} = \sum_{m=0}^{m=N_{183}} (L)^m = \sum_{m=0}^{m=N_{183}} (1 - (1 - l)^{\eta})^m \quad (2.18)$$

Donde N_{183} representa el número máximo de retransmisiones *Invite* 183 que genera el UAC cuando trata de conectarse con el UAS, y se calcula con la ecuación (2.19). Donde

$$N_{183} = \left\lceil \log_2 \left(\frac{T_{183} + T_{PRACK}}{T1} + 1 \right) \right\rceil \quad (2.19)$$

T_{183} se obtiene de la función (2.20), y es directamente proporcional a T1 segundos, y $(2^{N_i} - 1)$ es número de intentos que realiza R_{183} hasta alcanzar al remitente (UAC).

$$T_{183} = T1 \times (2^{N_i} - 1) \quad (2.20)$$

Las variables de la ecuación (2.20) se obtienen con la ecuación (2.2), y los valores de los *timers* T1 y el *TimerB*, cuyos valores están definidos por omisión en el apéndice A. El tiempo T_{PRACK} dependiente de $T1$ se obtiene ecuación (2.21).

$$T_{PRACK} = T1 \times (2^{N_o^e} - 1) + \text{Max}(0, T2 \times N_n^l) \quad (2.21)$$

Así mismo T_{PRACK} se obtiene de la expresión (2.22), y será directamente proporcional a T1.

$$T_{PRACK} = T2 + (\text{TimerF} - T2) = T2 + 64T1 - 8T1 = 64T1 \quad (2.22)$$

Considerando que el temporizador responde por una parte como una función exponencial y por otra como una función lineal, y N_n se define como el número máximo de retransmisiones *no-Invite* de una solicitud y se expresa matemáticamente de acuerdo a como lo muestra la ecuación (2.10). R_{PRACK} representa a las retransmisiones *no-Invite* de extremo a extremo en dos direcciones, y se obtiene utilizando la función (2.23). Así, cuando el servidor de agente de usuario (UAS) reciba la respuesta R_{PRACK} el agente de usuario cliente (UAS) enviará la respuesta 200 OK.

$$R_{PRACK} = \sum_{m=0}^{m=N_n} (L_e)^m = \sum_{m=0}^{m=N_n} (1 - (1 - l)^{2\eta})^m \quad (2.23)$$

La respuesta R_{200} *no-Invite* se calcula con la ecuación (2.24), e indica que el estatus del proceso para el establecimiento de la sesión está operando correctamente en las diferentes fases del mismo. Se conoce que (L) representa las pérdidas de extremo a

extremo en una dirección para una solicitud que atraviesa η saltos, y L_e representa a las pérdidas de extremo a extremo de dicha solicitud más su respuesta.

$$R_{200} = \sum_{m=0}^{m=N_n} (L)^m = \sum_{m=0}^{m=N_n} (1 - (1 - l)^\eta)^m \quad (2.24)$$

2.12.3.7.3 Fase de negociación UPDATE/200

R_{UPDATE} se obtiene con la ecuación (2.25) y se define como la fase de retransmisión *no-Invite* donde se reservan los recursos a utilizar en el enlace (código y QoS).

$$R_{UPDATE} = \sum_{m=0}^{m=N_0} (L_e)^m = \sum_{m=0}^{m=N_0} (1 - (1 - l)^{2\eta})^m \quad (2.25)$$

Los mensajes R_{200} se calculan con la función (2.26), y representan la respuesta *no-Invite* con pérdidas de extremo a extremo en una dirección.

$$R_{200} = \sum_{m=0}^{m=N_0} (L)^m = \sum_{m=0}^{m=N_0} (1 - (1 - l)^\eta)^m \quad (2.26)$$

Recordando que (L) representa las pérdidas de extremo a extremo en una dirección de una solicitud que atraviesa η saltos de los servidores que toca la respuesta, y que las variables descritas en la ecuación (2.26) están representadas por las expresiones (2.10), (2.14) y (2.15).

2.12.3.7.4 Fase de timbrado 180/PRACK/200 OK

En la fase 180/PRACK/200 OK el UAS informa al UAC que el usuario está siendo alertado acerca de la llamada, sin embargo la solicitud PRACK será retransmitida por el UAC hasta que este reciba una señal 200 OK. La ecuación (2.27) calcula el timbrado R_{180} .

$$R_{180} = \sum_{m=0}^{m=N_{180}} L^m = \sum_{m=0}^{m=N_{180}} (1 - (1 - l)^\eta)^m \quad (2.27)$$

Puesto que la señal de timbrado R_{180} está asociada a los tiempos T_{180} y T_{PRACK} , a través del número máximo de retransmisiones realizadas en forma exponencial N_{180} , la cual se obtiene a través de la ecuación (2.28).

$$N_{180} = \left\lceil \log_2 \left(\frac{T_{180} + T_{PRACK}}{T_1} + 1 \right) \right\rceil \quad (2.28)$$

Donde los tiempos T_{180} y T_{PRACK} se obtienen de las funciones (2.29) y (2.30).

$$T_{180} = T_1 \times (2^{N_i} - 1) \quad (2.29)$$

y

$$T_{PRACK} = T1 \times (2^{N_n^e} - 1) + \text{Max}(0, T2 \times N_o^l) \quad (2.30)$$

Se calcula el número de retransmisiones *no-Invite* (R_{PRACK}) en función de las pérdidas de extremo a extremo, utilizando la ecuación (2.31).

$$R_{PRACK} = \sum_{m=0}^{m=N_n} (L_e)^m = \sum_{m=0}^{m=N_n} (1 - (1 - l)^{2\eta})^m \quad (2.31)$$

Finalmente en esta fase de timbrado el UAS responde con una respuesta R_{200} , la cual representa una retransmisión *no-Invite* con pérdidas de extremo a extremo en una dirección, y que las variables descritas en la función (36) están representadas por las expresiones (21), (12) y (13).

$$R_{200} = \sum_{m=0}^{m=N_n} L^m = \sum_{m=0}^{m=N_n} (1 - (1 - l)^\eta)^m \quad (2.32)$$

Donde las variables descritas en la función (2.32) están representadas por las expresiones (2.10), (2.14) y (215.)

2.12.3.7.5 Fase de respuesta final 200 OK/ACK (*no-Invite*)

Esta fase pertenece a la última etapa del proceso de señalización que SIP realiza para el establecimiento de una puesta de sesión sobre IMS; y se utiliza para indicar que dicho proceso ha sido exitoso y las partes involucradas (remitente y destinatarios) se han puesto de acuerdo en el protocolo que requiere el establecimiento de una llamada SIP (QoS solicitada, códec a utilizar, etc.). En el caso de que la respuesta *200 OK* o el *ACK* se perdieran en el camino, el destinatario retransmitirá la respuesta *200 OK* utilizando el modo de retransmisión *no-Invite*. Dicha retransmisión R_{200} (con pérdidas de extremo a extremo en dos direcciones) se obtiene a partir de la función (2.33), una vez que el

$$R_{200} = \sum_{m=0}^{m=N_n} L_e^m = \sum_{m=0}^{m=N_n} (1 - (1 - l)^{2\eta})^m \quad (2.33)$$

remitente de la llamada ha recibido un R_{200} correctamente, enviará la retransmisión R_{ACK} (*no-Invite*) con pérdidas de extremo a extremo en una dirección, de acuerdo a la función (2.34).

$$R_{ACK} = \sum_{m=0}^{m=N_n} L^m = \sum_{m=0}^{m=N_n} (1 - (1 - l)^\eta)^m \quad (2.34)$$

Esto significa que el destinatario de la llamada finalmente acepta la llamada con la solicitud *ACK*. En esta etapa el establecimiento de la sesión, la persona que llama y el destinatario pueden comenzar a intercambiar tráfico de medios. Las variables descritas en la función (2.34) están representadas por las expresiones (2.10), (2.14) y (215.).

2.12.3.7.6 Ancho de Banda Total

Dado que el parámetro de interés en este caso es el ancho de banda requerido para obtener la calidad de servicio deseada que permita ofrecer un servicio adecuado. Es importante definir el procedimiento que determina teóricamente el ancho de banda requerido en el subsistema IMS. La cantidad total de ancho de banda (B) requerido por la señalización SIP en el establecimiento de llamadas de usuarios rurales con una tasa de λ mensajes por segundo, η saltos y un tamaño de paquete SIP de S bytes definidos en la tabla 11. Estos valores son compatibles con [62] y [63] estas referencias evaluaron el rendimiento de una sesión basado en SIP bajo UDP en diferentes tipos de redes, por ejemplo UMTS, etc.). El cálculo del Ancho de Banda se obtendrá con la función (2.35).

$$\begin{aligned}
 B = & \eta \times (R_i \times S_i + R_{100} \times S_{100}) \\
 & + (R_{183} \times S_{183} + R_{prack} \times S_{prack} + R_{200} \times S_{200}) \\
 & + (R_{update} \times S_{update} + R_{200} \times S_{200}) + \mu \times (R_{180} \times S_{180} + R_{prack} \times S_{prack} + R_{200} \times \\
 & S_{200}) + (R_{200} \times S_{200} + R_{ACK} \times S_{ack}) \quad (2.35)
 \end{aligned}$$

Se puede observar de la expresión anterior que el ancho de banda depende de la tasa de (λ) llamadas/segundo, del número de saltos (η) que realizan los paquetes de datos a través de los servidores SIP, del porcentaje de pérdidas con el que opera la red (l), y del número de timbrados que escucha el remitente durante el proceso de la puesta de sesión (μ). Y para el modelo bajo análisis propuesto, se identifican las siguientes variables relevantes:

- μ Es el número de timbrados (3) que escuchara el remitente UAC durante el establecimiento de la sesión.
- η Es el número de saltos que realizan los paquetes de datos cuando UAC solicita comunicarse con el UAS. Nuestro modelo requiera realizar 5 saltos por tratarse de una sesión en itinerancia.
- l Es la probabilidad de pérdidas entre saltos (5% a 20%). Es decir, por tratarse de una red con pérdidas, se consideró que por cada vez que un paquete de datos atraviesa un servidor existe la posibilidad de pérdidas de los datos por diferentes razones (saturación en horas pico, retardos propios de las redes, etc.) Estos retardos están contemplados en las solicitudes de sesiones *Invite* y *no-Invite*.
- L Pérdidas de extremo a extremo en un sentido. Estas pérdidas son las que ocurren en un viaje completo en una dirección, después de atravesar todos los servidores que forman parte de las rutas que tomarán los paquetes de datos.
- Le Pérdidas de extremo a extremo en dos sentidos. Lo mismo que L solo que ahora se consideran las pérdidas de extremo a extremo de ida y vuelta (UAC hacia el UAS y viceversa).
- S Tamaño de los mensajes SIP (definidos en la Tabla 11)

- λ Tasa de llamadas o sesiones por segundo. Este parámetro representa al número de llamadas/segundo que realizan los suscriptores cuando utilizan la red.
- B Ancho de Banda necesario para el establecimiento de sesiones IMS. Este parámetro es de gran importancia ya que indica el ancho de banda estimado que la red requiere para poder atender la tasa de llamadas de los suscriptores.
- T1 500 [ms] por omisión especificado en la recomendación [60]. Es el *Timer* utilizado en las solicitudes *Invite*.
- T2 4 segundos por omisión especificada en la recomendación [60]. Es el *Timer* utilizado en las solicitudes *no-Invite*.

TimerB 64T1 [60]. Temporizador máximo permitido en las solicitudes *Invite*.

TimerF 64T1 [60]. Temporizador máximo permitido en las solicitudes *no-Invite*.

Todas estas variables participan en un determinado momento en el cálculo del ancho de banda estimado producido durante el flujo de tráfico telefónico realizado por usuarios de las diferentes zonas rurales. Por lo tanto el operador de la red deberá estar atento a los parámetros que miden la densidad de tráfico para evitar los cuellos de botella, buscando que la red opere en forma balanceada evitando con esto la saturación de la misma. El administrador de la red también debe cuidar que la QoS se mantenga mayor a 70. Los valores de los tamaños de los mensajes de la señalización SIP del modelo, se incluyen en la Tabla 2.11, estos valores muestran los tamaños que tendrán los mensajes SIP que se utilizaran durante el proceso de la puesta de sesiones.

Tabla 2.11. Tamaño de los mensajes SIP para UDP. [62]

Mensaje SIP	Tamaño (Bytes)
<i>SIP INVITE</i>	810
<i>SIP REGISTER</i>	225
<i>183 SESSION PROGRESS</i>	260
<i>SIP 180 RINGING</i>	260
<i>SIP PRACK</i>	260
<i>SIP 100 TRYING</i>	260
<i>SIP UPDATE</i>	260
<i>SIP 200 OK</i>	100
<i>SIP SUBSCRIBE</i>	100
<i>SIP NOTIFY</i>	100
<i>SIP 401 UNAUTHORIZED</i>	100
<i>SIP ACK</i>	60

2.12.4 Estimación del retardo en la puesta de sesión IMS

En toda red de telecomunicaciones existen retardos debido a sus características intrínsecas de la misma, como por ejemplo, debido a la saturación de los servidores por

las altas tasas de llamadas, al tiempo requerido para atravesar diferentes servidores durante las diferentes puestas de sesión, y otros parámetros como los mencionados en la sesión anterior. Es por esto que es importante conocer el retardo total estimado que las redes operaran en promedio, con el fin de mantener la calidad del servicio que se ofrece a los usuarios de la red. El trabajo de esta tesis no contempla el análisis profundo de los retardos en la redes, pero si mencionara la metodología propuesta por [64] el cual especifica lo siguiente: Las solicitudes Invite se retransmiten a través de cada servidor que participa durante la solicitud de sesión. Sin embargo, cada servidor no espera hasta que el anterior servidor haya recibido con éxito una respuesta. Por ejemplo, la persona que llama (UAC) envía una solicitud de sesión al Proxy P-CSCF. P-CSCF recibe la solicitud Invite y devuelve una respuesta 100 al remitente. Si se pierde la respuesta 100, el remitente (UAC) podría retransmitir la solicitud después de $T1$ segundos. Sin embargo, el servidor P-CSCF no necesita tener en cuenta esto y reenviara la solicitud Invite al siguiente servidor en la cadena de enlace (S-CSCF) e iniciara su propio tiempo de espera de retransmisión. Así que para determinar el retardo de las solicitudes Invite, hay que tener en cuenta la pérdida en un solo sentido (l) entre los saltos. De este modo, el retardo de la fase *Invite* (T_i) en los saltos entre el remitente y el P-CSCF, así como entre el P-CSCF y el S-CSCF se calcula con la ecuación (2.36).

$$T_i = T1 \times (2^{N_i} - 1) + \sum_{m=1}^{m=\eta} D_m \quad (2.36)$$

donde D se define como el retardo de propagación que existe entre cada salto cuando se establece una sesión entre el remitente y el destinatario, y N_i será el número máximo de solicitudes *Invite* retransmitidas y está determinado por la ecuación (2.2).

2.12.4.1 Retardo de la fase *no-Invite*

Las fases *no-Invite* se llevan a cabo de una manera de extremo a extremo. El retardo que se produce en estas fases es igual al generado por el retardo de propagación de ida y vuelta con η saltos ($2 \times \eta \times D$) más el tiempo entre las retransmisiones; de esta manera los retardos promedio para completar la fase *no-Invite* (T_n) se calcula con las expresiones de la (2.37) a la (2.39).

$$T_{PRACK} = T1 \times (2^{N_n^e} - 1) + \text{Max}(0, T2 \times N_n^l) + \sum_{m=1}^{m=\eta} D_m \quad (2.37)$$

$$T_{UPDATE} = T1 \times (2^{N_n^e} - 1) + \text{Max}(0, T2 \times N_n^l) + \sum_{m=1}^{m=\eta} D_m \quad (2.38)$$

$$T_{200} = T1 \times (2^{N_n^e} - 1) + \text{Max}(0, T2 \times N_n^l) + \sum_{m=1}^{m=\eta} D_m \quad (2.39)$$

Las variables de las ecuaciones (2.37), (2.38), y (2.39) se determinan con la ecuación (2.10). Donde P_o es el número de solicitudes *no-Invite* necesarias para configurar la sesión. El primer término de la ecuación describe el retardo de ida y vuelta, el segundo retardo causado mediante la retransmisión durante la fase exponencial de retroceso y el último describe el retardo al retransmitir la solicitud cada T2 segundos. Téngase en cuenta que si los mensajes de respuesta se pierden se tienen que retransmitir nuevamente la solicitud, y por lo tanto, cualquier retardo adicional causado por la pérdida ya se representó en el aumento del número de solicitud de retransmisiones. Es decir, que el envío de la respuesta únicamente agrega el retardo de propagación al retardo total. Podemos aplicar la ecuación (32) para *PRACK/200* y *UPDATE/200 OK*. Finalmente la respuesta *200 OK/ACK* y la fase *BYE (BYE / 200 OK)* con los retardos T_{PRACK} , T_{UPDATE} , T_f y T_{BYE} .

2.12.4.2 Retardo de la respuesta provisional (183, 180)

El retardo de la respuesta provisional confiable se retransmite periódicamente iniciando en 500 [ms], y se duplica después de cada retransmisión hasta un máximo de 32 segundos. Es similar al retardo de la respuesta *Invite*. Sin embargo si el *PRACK* no ha sido recibido por el UAS después de 96 segundos de haberse enviado la respuesta provisional *183*, se considerará una red o punto de falla. Así entonces N_{183} es el número máximo de solicitudes *Invite 183* retransmitidas. T_{183} es el retardo de la respuesta provisional *183* y se calcula con la siguiente expresión (2.40)

$$T_{183} = T1 \times (2^{N_i} - 1) + \sum_{m=1}^{m=\eta} D_m \quad (2.40)$$

donde T_{PRACK} representa el retardo de la respuesta *PRACK*, la cual se calcula con la función (2.41) y se retransmitirá únicamente cuando el *Timer* de retransmisión sea activado (por ejemplo, *T1, 3T1, ...*)

$$T_{PRACK} = T1 \times (2^{N_n^e} - 1) + \text{Max}(0, T2 \times N_n^l) + \sum_{m=1}^{m=\eta} D_m \quad (2.41)$$

N_{180} se calcula con la función (2.28) y corresponde al número máximo de solicitudes *Invite 180* retransmitidas. El retardo T_{180} corresponde al retardo de la respuesta provisional confiable *180* y se calcula con la expresión (2.42).

$$T_{180} = T1 \times (2^{N_i} - 1) + \sum_{m=1}^{m=\eta} D_m \quad (2.42)$$

y el retardo T_{PRACK} se calcula con la expresión (2.43) e incluye el retardo de ida y vuelta, es el retardo generado mediante la retransmisión durante la fase exponencial de retroceso y el último retardo que se produce al retransmitir la solicitud cada T2 segundos.

$$T_{PRACK} = T1 \times (2^{N_n^e} - 1) + \text{Max}(0, T2 \times N_n^l) + \sum_{m=1}^{m=\eta} D_m \quad (2.43)$$

2.12.4.3 Retardo Total

Es importante conocer el retardo total producido durante el establecimiento de una sesión debido a que un retardo mayor a 130 segundos, afectará la QoS esperada. El sistema de administración de la red propuesta debe tener la capacidad de medir este parámetro, y las causas que lo originan, para que el administrador de la red pueda tomar las medidas requeridas para corregir el retardo total. Finalmente el retardo para el establecimiento de la sesión total promedio para el caso de η saltos se calcula de la función (2.44):

$$T_t = \eta \times T_i + T_{183} + T_{PRACK} + T_{UPDATE} + \mu \times (T_{180} + T_{PRACK}) + T_{200} \quad (2.44)$$

2.13 Limitaciones del modelo de red rural

2.13.1 Limitaciones de la tasa de tráfico

Por tratarse de un modelo NGN aplicado a entornos rurales y urbanos marginados, la red en si está diseñada para atender un tráfico de 10 [sesiones/s] en horario normal de trabajo, 20 [sesiones/s] en horas pico, y 100 [sesiones/s] en casos de emergencia. Estas características limitan a la red en cuanto a su capacidad de tráfico, ya que fue diseñada para zonas rurales de baja y mediana densidad de tráfico. También lo limita el hecho de que la red utiliza el protocolo de datagramas (UDP) como protocolo de transporte. Se conoce que UDP carece de un protocolo de señalización propio que le permita la gestión y control de las puestas de sesiones telefónicas; para realizar esta función, el sistema se auxilia del protocolo de señalización SIP. A diferencia de TCP que si cuenta con la capacidad de señalización para establecer sesiones. Esta situación limita la operación de la red, debido que el ancho de banda se hace mayor, generando retardos innecesarios en la red. En este trabajo no están contemplados los costos de CAPEX y OPEX, característicos en la implementación de redes de telecomunicaciones, pero es importante mencionar que dichos costos de inversión y operación son los que permitirán a los concesionarios de redes públicas decidir construir o no el modelo. Considérese que se trata de zonas marginadas, donde sus habitantes tienen salarios bajos y los inversionistas analizarán cuidadosamente el Periodo de Recuperación de Inversión (PRI). Aunque de acuerdo el nuevo concepto de NGN, las redes de acceso pueden ser alámbricas o inalámbricas, y estarán sujetas a nuevas regulaciones del gobierno, permitiendo la entrada a nuevos actores (participantes). Es decir, no solo los concesionarios podrán implementarlas y operarlas; de acuerdo a la UIT las NGN operarán dos capas independientes unas de otra (transporte y servicios). Este concepto permite posibilidades inmensas de inversión, debido a que hace que los servicios sean completamente independientes del transporte de la información.

Capítulo 3

Marco de Simulación

3.1. Experimentación numérica del modelo de red NGN

Una vez expuestas la teoría relevante que soportan los modelos de arquitectura y matemático, se analizarán las condiciones en las cuales el modelo propuesto es válido y las restricciones que diferentes escenarios pueden imponer sobre la implementación de la arquitectura mencionada. Para esto, se calculan los diferentes parámetros que participan en el comportamiento de la red y la magnitud del ancho de banda requerido con este propósito. Primero, definiendo los parámetros que operará el modelo y posteriormente calculando las variables más significativas.

No	Parámetro	Descripción
1	μ	Es el número de timbrados durante el establecimiento de la sesión.
2	η	Es el numero de saltos (5 saltos por tratarse de una sesion en itinerancia).
3	l	Es la probabilidad de pérdidas de la red entre dos saltos (típicamente de 5% a 20%, pues por encima del 20% no es viable su aprovechamiento).
4	L	Pérdidas de extremo a extremo en un sentido. Estas pérdidas son las que ocurren en una dirección, después de atravesar todos los servidores que forman parte de las rutas que tomaran los paquetes de datos, desde el servidor agente de usuario cliente (UAC) que es el remitente al servidor agente de usuario (UAS) que es el que recibe la llamada.
5	Le	Pérdidas de extremo a extremo en dos sentidos.
7	λ	Tasa de llamadas o sesiones por segundo.
8	B	Ancho de Banda necesario para el Establecimiento de Sesiones IMS expresado en bytes.
9	$T1$	El tiempo de retransmisión de una solicitud <i>INVITE</i> (típicamente de 500 [ms] por omisión referencia [60]).
10	$T2$	El tiempo de retransmisión de una solicitud <i>NO-INVITE</i> (típicamente de 4 [s] valor por omisión según [60]).
11	$TimerB$	El tiempo máximo permitido para solicitudes <i>INVITE</i> , antes de que el proceso de puesta de sesión, se considere fallido (típicamente de 64T1

de acuerdo con [60].

- 12 *TimerF* El tiempo máximo permitido para solicitudes no-*Invite*, antes de que el proceso de puesta de sesión, se considere fallido (típicamente de 64T1 de acuerdo con [60]).

Para la realización de los experimentos numéricos de este trabajo de tesis se utilizó la herramienta MatLab, por ser una herramienta práctica y de uso común en el ámbito científico

Experimento numérico # 1: Solicitudes R_i (Fase *Invite*)

Con el fin de observar los cambios producidos en el ancho de banda total durante el establecimiento de las sesiones SIP, primero se analizara el comportamiento de las solicitudes *Invite*. Para esto, se estudiaran cuatro escenarios de red donde el primer escenario contempla conectar una red rural de hasta 250 usuarios suscritos a la misma (se considera un caso hipotético de zona rural de aproximadamente 1000 habitantes). En este experimento numérico se analizara el comportamiento de las solicitudes *Invite* (R_i), que se generan por el Servidor de Agente de Usuario Cliente (UAC) en cualquier punto (n) del tiempo, para conocer el número de solicitudes *Invite* que se requieren generar por parte del UAC cuando este desea establecer una llamada o sesión telefónica. El cálculo se realiza utilizando los parámetro de tráfico (λ) y el porcentaje de pérdidas de la red (l). Para esto se analizaran cuatro escenarios:

El primer escenario contempla el comportamiento de R_i considerando un tráfico con tasas de sesiones de $\lambda = (10, 20, 30, 40, 50, \text{y } 100)$ [sesiones/s] diferentes provenientes de los usuarios suscritos a la red, y en todos los escenarios se consideran porcentajes de pérdidas igual a ($l = 0, 0.05, 0.10, 0.15, 0.20$).

En el segundo escenario se considera una tasa de tráfico de 10 [sesiones/s] en horas normales de trabajo. En el tercer escenario se analiza el comportamiento de R_i con los diferentes porcentajes de pérdidas a una tasa fija de 20 [sesiones/s]. Finalmente, en el cuarto y último escenario se analiza el comportamiento de R_i con los diferentes porcentajes de pérdidas l a una tasa fija de 100 [sesiones/s] con el porcentaje de pérdidas mencionadas.

Parámetros considerados del primer escenario

$\lambda = (10, 20, 30, 40, 50, \text{y } 100)$ [sesiones/s]

$l = 0, 0.05, 0.10, 0.15, 0.20$ (porcentaje de pérdidas de la red)

Las retransmisiones *Invite* $R_i(l, n)$ se calculan con la ecuación (3.1) con datos de λ y l ,

$$R_i(l, n) = \lambda \times \left(1 + \sum_{m=1}^{m=k} l^m\right) = \lambda \times \sum_{m=0}^{m=k} l^m \quad (3.1)$$

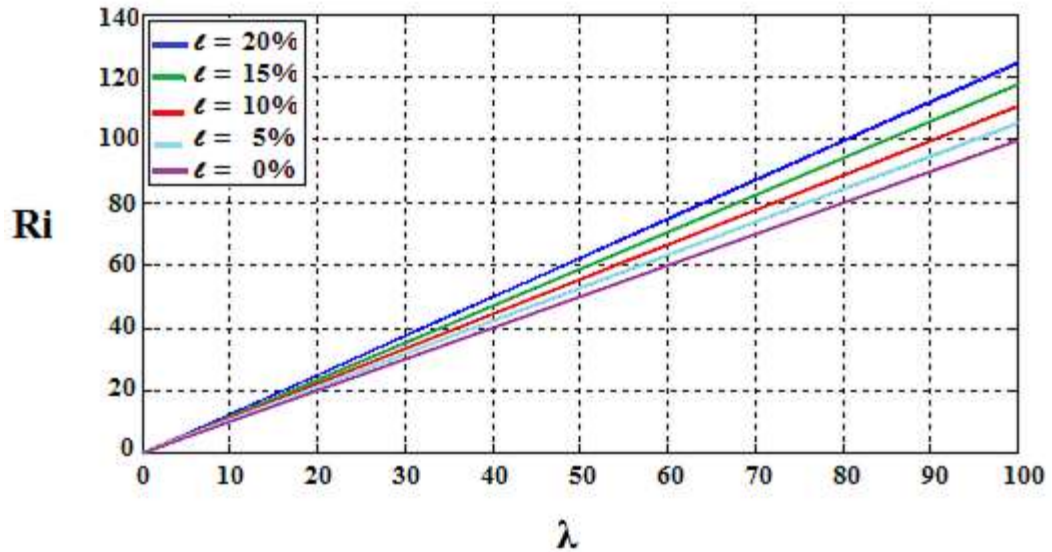


Figura 3.1. Relación entre R_i y λ para diferentes valores de l (dos dimensiones)

con $k = \log_2(n + 1)$ y $n_{MAX} = \frac{TimerB1}{T1}$, las figuras 3.1 y 3.2, muestran en dos y tres dimensiones la relación que existe entre el número de solicitudes *Invite* (R_i) vs diferentes valores de la tasa de llamadas y diferentes porcentajes de pérdidas l .

$R_i(l, n)$ se determina empleando la ecuación (3.1) para los siguientes escenarios.

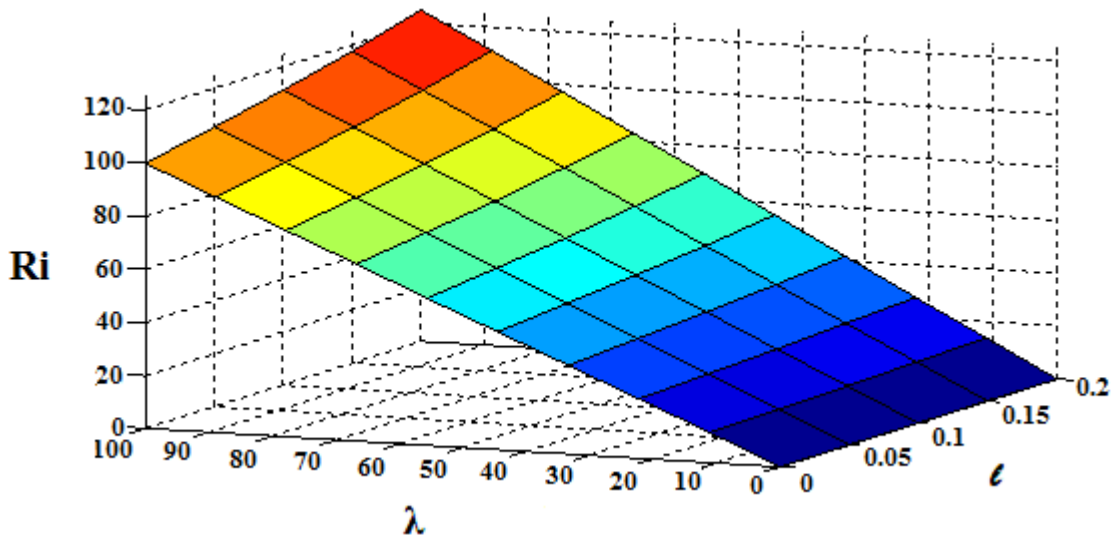


Figura 3.2. Relación entre R_i vs λ para diferentes valores de l (tres dimensiones)

Tabla 3.1. Solicitudes *Invite* $R_i(l, \lambda, n)$ generadas en horario normal de trabajo.

% Pérdidas	Tasa llamadas/s	$R_i(l, \lambda, n)$	incremento de $R_i(l, \lambda, n)$ (%)
0	10	10.0000	0
5	10	10.5263	5.26
10	10	11.1111	11.11
15	10	11.7646	17.64
20	10	12.4992	24.99

Datos del segundo escenario

$\lambda = 10$ [sesiones/s] (tasa fija de tráfico)

$l = (0, 0.05, 0.1, 0.15, 0.2)$ (% de pérdidas de la red)

Las respuestas R_i vs la tasa de llamadas y el porcentaje de pérdidas de la red se muestran en la tabla 3.1.

Datos del tercer escenario

$\lambda = 20$ [sesiones/s] (tasa fija de tráfico)

$l = (0, 0.05, 0.1, 0.15, 0.2)$ (% de pérdidas de la red)

La tabla 3.2, muestra las solicitudes *Invite* R_i calculadas en las horas pico de trabajo.

Datos del cuarto escenario

$\lambda = 100$ [sesiones/s](tasa fija de tráfico)

$l = (0, 0.05, 0.1, 0.15, 0.2)$ (% de pérdidas de la red)

Tabla 3.2. Número de solicitudes *Invite* $R_i(l, \lambda, n)$ (horas pico)

% Pérdidas	Tasa llamadas/s	$R_i(l, \lambda, n)$	incremento de $R_i(l, \lambda, n)$ (%)
0	20	20.0000	0
5	20	21.0526	5.26
10	20	22.2222	11.11
15	20	23.5291	17.64
20	20	24.9984	24.99

La tabla 3.3, muestra las solicitudes *Invite* R_i calculadas en casos de emergencia.

Tabla 3.3. Número de solicitudes *Invite* en casos de emergencia $R_i(l, \lambda, n)$

% Pérdidas	Tasa llamadas/s	$R_i(l, \lambda, n)$	incremento de $R_i(l, \lambda, n)$ (%)
0	100	100.0000	0
5	100	105.6000	5.60
10	100	111.4667	11.14
15	100	118.0233	11.80
20	100	125.3984	12.53

Experimento numérico # 2: (Porcentaje de pérdidas fase *Invite*)

En este experimento numérico se determinarán los porcentajes de las pérdidas de los mensajes que viajan de extremo a extremo y de ida y vuelta que se realizan durante el proceso de una solicitud *Invite* (saltos de los mensajes de servidor a servidor y viceversa para las respuestas provisionales). Esto permitirá conocer el resultado de las pérdidas que tienen los mensajes que viajan a través de los servidores SIP durante el establecimiento de una sesión IMS, en dos direcciones (Ver figura 3.3). Para esto se define al parámetro (l) como la probabilidad de pérdidas que sufren los mensajes cuando viajan de un servidor a otro en una dirección, y a (l_e) como las pérdidas que sufren los mensajes al transportarse de un servidor a otro de ida y vuelta (solicitudes y respuestas).

Parámetros del experimento

$l = (0.00, 0.05, 0.10, 0.15, 0.20)$ porcentajes de pérdidas de la red.

Los porcentajes de pérdidas de ida y vuelta se calcula con la ecuación (2.4), y se grafican (l_e vs l) en la figura 3.4.

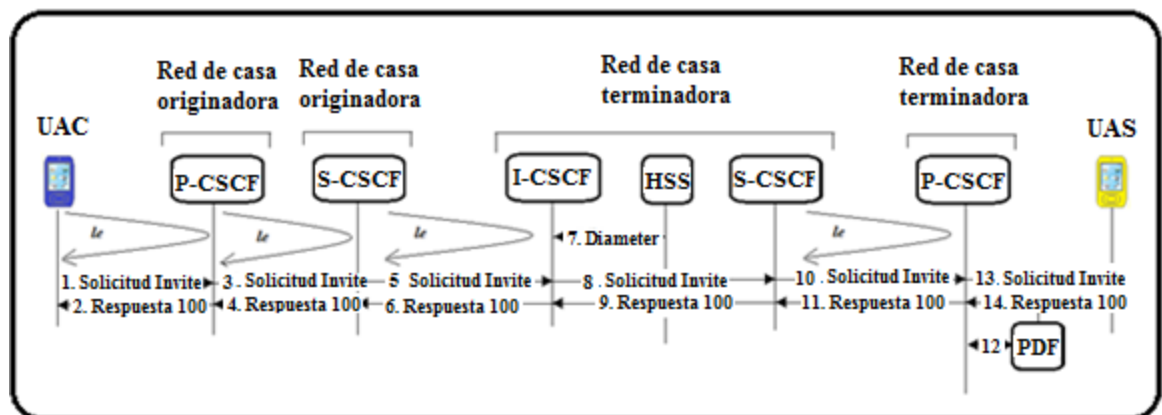


Figura 3.3. Rutas de re-transmisión de mensajes de la fase *Invite* (salto por salto)

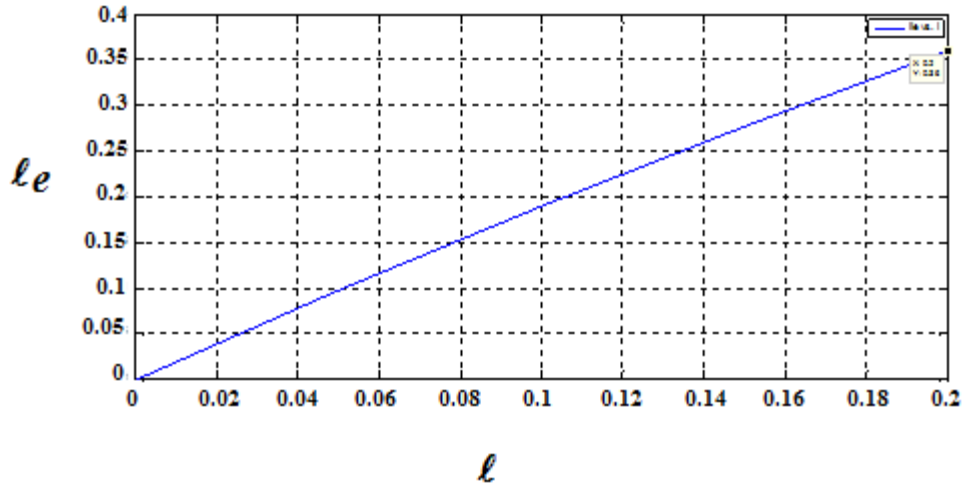


Figura 3.4. Pérdidas l_e vs l

La tabla 3.4, muestra los cálculos del porcentajes de pérdidas (0 a 20%).

Experimento numérico #3: Puestas de sesión SIP exitosas (fase Invite)

En este experimento y con el fin de conocer la eficiencia que presenta la red IMS en cuanto a los mensajes (*Invite*) re-transmitidos durante las puestas de llamadas, se calcularon las (P_i) , y (P_{100}) , las cuales corresponden a las solicitudes y respuestas exitosas, generadas una vez que el sistema alcance el estado estable. Las P_i se definen textualmente como el número de mensajes R_i enviados por el remitente (UAC) después de que el sub-sistema alcance el estado estable, dividido mediante el número de mensajes originales que fueron enviados y que no han sido retransmitidos (λ).

Así mismo, las P_{100} se definen como el número de respuestas exitosas generadas por una solicitud *Invite* generada durante el establecimiento de una sesión, después de que el sistema alcance el estado estable. Con esta información se conocerá la eficiencia del subsistema IMS durante la fase *Invite*. La figura 2.46 muestra cuando el remitente (UAC) genera una solicitud y el servidor P-CSCF, regresa una respuesta a esta solicitud, ambas (solicitud y respuesta) deben alcanzar el estado estable. Es decir, con este experimento se

Tabla 3.4. Resumen de datos calculados de l_e vs l (rango 0 al 20%)

No	(l) Pérdidas %	(l_e) Pérdidas %
1	0	0
2	5	9.75
3	10	19
4	15	27.75
5	20	36

conocerá cuantas R_i se deben generar por el IMS para que los mensajes re-transmitidos y sus respuesta correspondientes durante la fase *Invite* tengan éxito. Como ejemplo se puede mencionar el caso de una zona rural que tiene un tráfico promedio de 10 [mensajes/s], y para lograr el estado estable del sistema, se requirieron enviar por lo menos 20 [mensajes/s]; esto significa que en promedio por cada dos mensajes enviados únicamente uno tendrá éxito. Con el fin de verificar numéricamente este ejemplo, y considerando que N es cualquier instante en el tiempo. Se calculan las P_i utilizando la función (3.2) con los siguientes parámetros.

Parámetros propuestos

$\lambda = (10, 20, 30, 40, 50, \text{y } 100)$ tasa de [sesiones/s]

$R_i = (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100)$

$$P_i = \frac{R_i(l_e, N)}{\lambda} = \frac{\lambda \times \sum_{m=0}^{m=k} (1 - (1-l)^2)^m}{\lambda} = \sum_{m=0}^{m=k} (1 - (1-l)^2)^m \quad (3.2)$$

Los resultados numéricos de la figura 3.5, muestran que las P_i dependen únicamente de las pérdidas de la red (l), y no se ven afectadas por el tráfico λ .

Ahora se calcularán las respuestas correspondientes a las solicitudes mencionadas. La generación de las respuestas únicamente serán realizadas después de la recepción de una solicitud *Invite*. Con un retardo l en una dirección y una tasa de transmisión estable de $R_i(l_e, N)$, únicamente $R_i(l, N) \times l$ solicitudes *Invite* pueden ser recibidas. Así que, si definimos a P_{100} como el número de respuestas recibidas durante el establecimiento de una sesión, estas se pueden calcular con la ecuación (3.3).

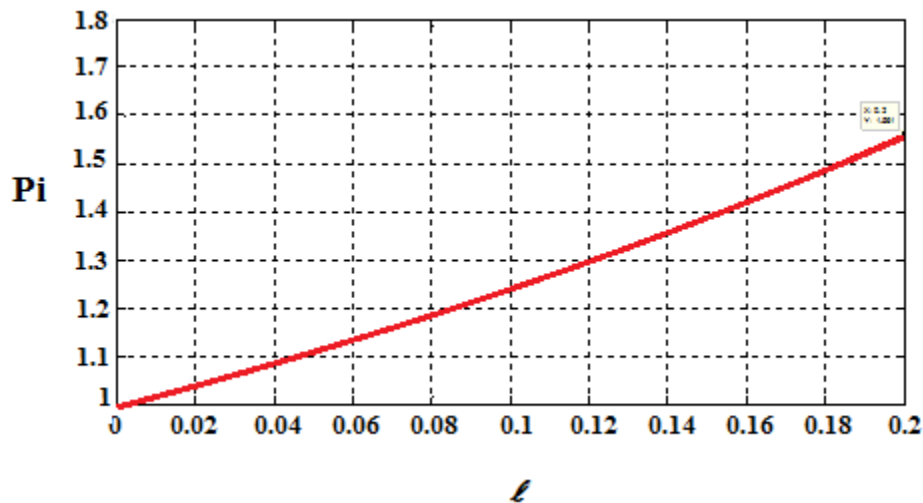


Figura 3.5. Cálculo de P_i vs porcentaje de pérdidas (l) y Tasa de Tráfico λ

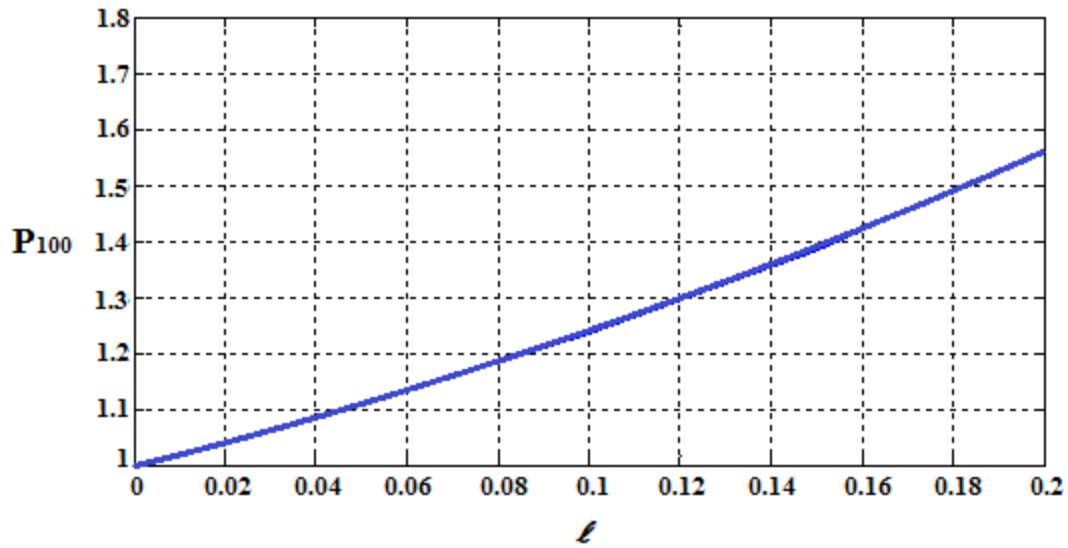


Figura 3.6. P_{100} vs porcentaje de pérdidas de la red (l)

$$P_{100} = \frac{R_i(l,N)}{\lambda} = \frac{\lambda \times \sum_{m=0}^{m=k} l^m}{\lambda} = \sum_{m=0}^{m=k} l^m \quad (3.3)$$

Los resultados numéricos de la figura 3.6, muestran que las P_i dependen únicamente de las pérdidas de la red (l), y no se ven afectadas por el tráfico λ .

Experimento numérico # 4: Solicitudes + respuestas exitosas (SIE) (fase Invite)

Con este experimento numérico se conocerá el número de respuestas exitosas, de las solicitudes (*Invite*) colocadas, después de que el sistema alcance el estado estable. Se sabe que estas respuestas se activan únicamente después de recibir una solicitud exitosa. Esta es la siguiente etapa del experimento numérico #3, que forma parte del proceso para el establecimiento de una solicitud *Invite*. El objetivo de este experimento es conocer el porcentaje de solicitudes *Invite* que se requieren generar para poder enviar una solicitud válida con su respuesta correspondiente después de alcanzar el estado estable. Se conoce que P_i representa al número de solicitudes *Invite* exitosas obtenidas cuando el sistema ha alcanzado su estado estable, y posterior a esto dichas solicitudes deben estar acompañadas de una respuesta válida (P_{100}) para que la solicitud este completa. En nuestro experimento se calcularán dichas respuestas, para diferentes números de saltos ($\eta = 6$). La transmisión de las respuestas únicamente será disparada después de la recepción de un *Invite*. Con este propósito definimos a ($SIE = \eta \times P_i + \eta \times P_{100}$), como el número de solicitudes *Invite* exitosas enviados con sus respectivas respuestas, después de haber alcanzado el estado estable, y que han tocado η servidores en la red IMS. Así que, para poder enviar una *SIE* exitosa en la red a través de η saltos sobre los mensajes SIP, se requieren haber transmitido ($\eta \times P_i + \eta \times P_{100}$) mensajes promedio (incluyendo solicitud y respuesta). Las *SIE* se calculan utilizando la ecuación (3.4) la cual es el número de mensajes *Invite* y sus respuestas después de alcanzar el estado estable.

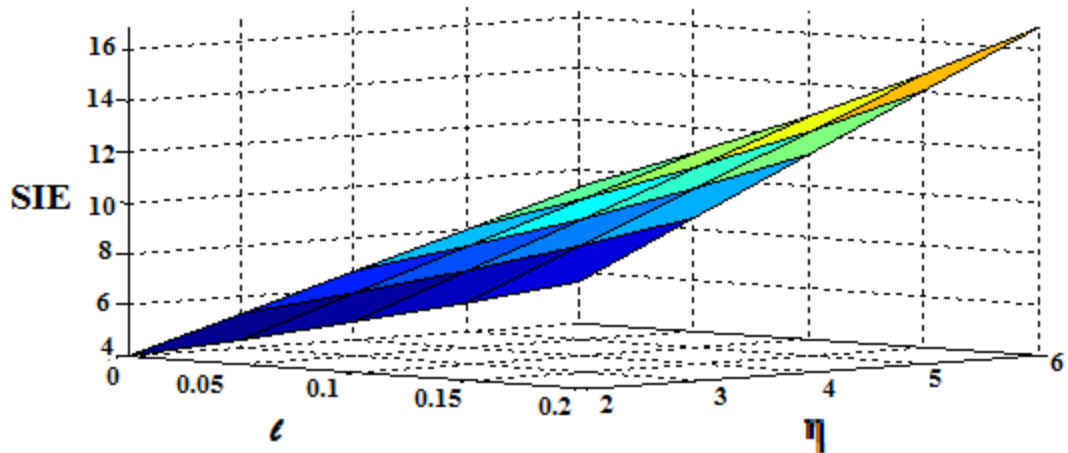


Figura 3.7. Solicitudes (SIE) vs pérdidas (l) y η saltos (tres dimensiones)

$$SIE = \eta \times P_i + \eta \times P_{100} = \eta(P_i + P_{100}) \quad (3.4)$$

Las gráficas de las figuras 3.7 y 3.8, muestran la relación que existe entre las solicitudes *Invite* exitosas y $(P_i + P_{100})$ mensajes promedio, cuando el sistema alcanza el estado estable para diferentes valores de η (saltos).

Los resultados numéricos de las gráficas de analizan con detalle en el capítulo 4 de este documento.

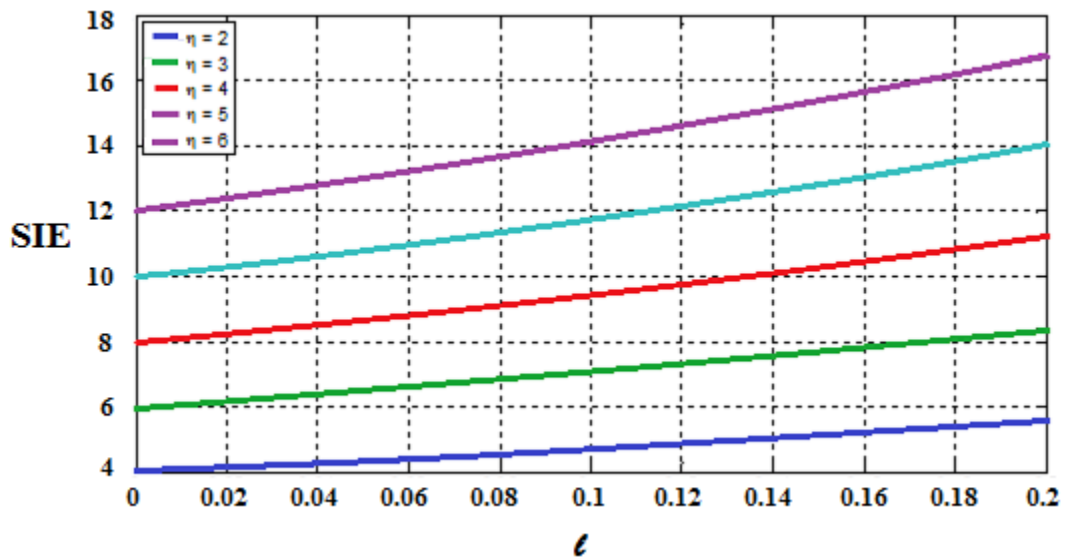


Figura 3.8. Solicitudes (SIE) vs pérdidas (l) y número de saltos η (dos dimensiones)

Experimento numérico # 5: Solicitudes R_0 (fase *no-Invite*)

En este experimento numérico se determinó el número de solicitudes *no-Invite* (R_0) enviadas por el remitente en cualquier punto del tiempo (n). Recordemos que para establecer una sesión SIP en las redes IMS, dicha sesión se compone de dos fases, la primera conocida como *fase Invite*, y por otro lado la *fase no-Invite*. En este modo el remitente UAC retransmite un mensaje y si no recibe una respuesta después de $T1$ [s], el temporizador de retransmisión se incrementa exponencialmente a un tiempo de retransmisión máximo llamado $T2$ [s]. Una vez que el temporizador es alcanzado el remitente continua retransmitiendo la solicitud cada $T2$ [s] hasta un temporizador máximo llamado (*TimerF*). Una vez que este temporizador es alcanzado el remitente tira el mensaje y detiene la retransmisión). En los experimentos del 1 al 4 se analizaron solicitudes del tipo *Invite*, que son las que inician la puesta de sesión SIP, en este experimento y el experimento 6, se analizaran solicitudes del tipo *no-Invite*, las cuales concluyen el proceso de puesta de sesión. Los mensajes en estas solicitudes viajan de extremos a extremo (ida y vuelta), y emplean temporizadores diferentes a los utilizados en las solicitudes *Invite*. El experimento utiliza diferentes tasas de sesión (λ) y porcentajes de pérdidas de paquetes (l) como se muestran a continuación:

Parámetros del experimento

$\lambda = (10, 20, 30, 40, 50, \text{y } 100)$ [sesiones/s]

$l = (0, 0.05, 0.10, 0.15, 0.20)$ porcentaje de pérdidas.

La ecuación matemática (2.12) se utilizó para el cálculo de R_0 vs λ para diferentes valores de (l). Las figuras 3.9 y 3.10, muestran el resultado de la ecuación mencionada.

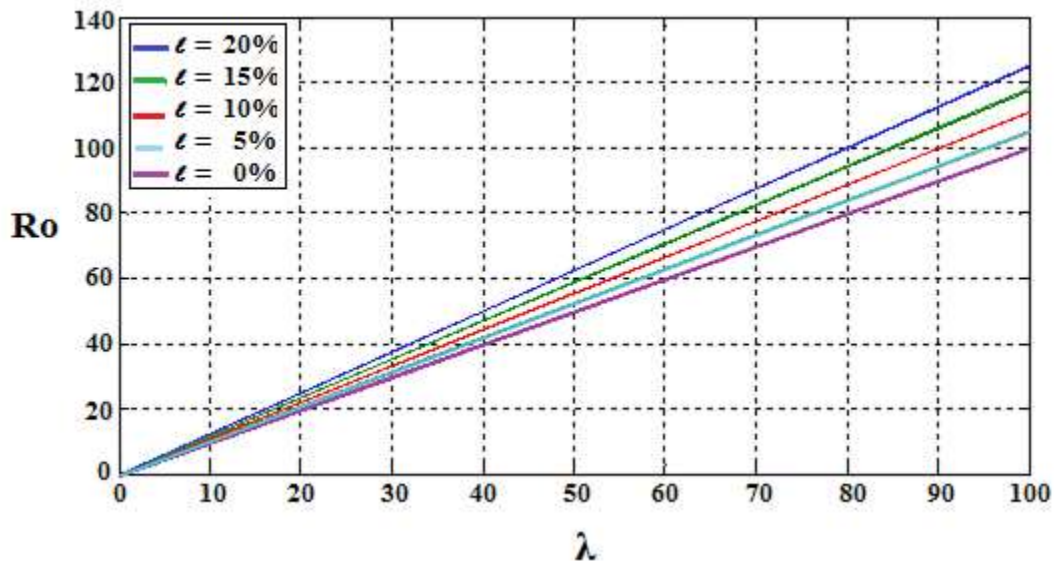


Figura 3.9. Cálculo de R_0 vs λ y l (dos dimensiones)

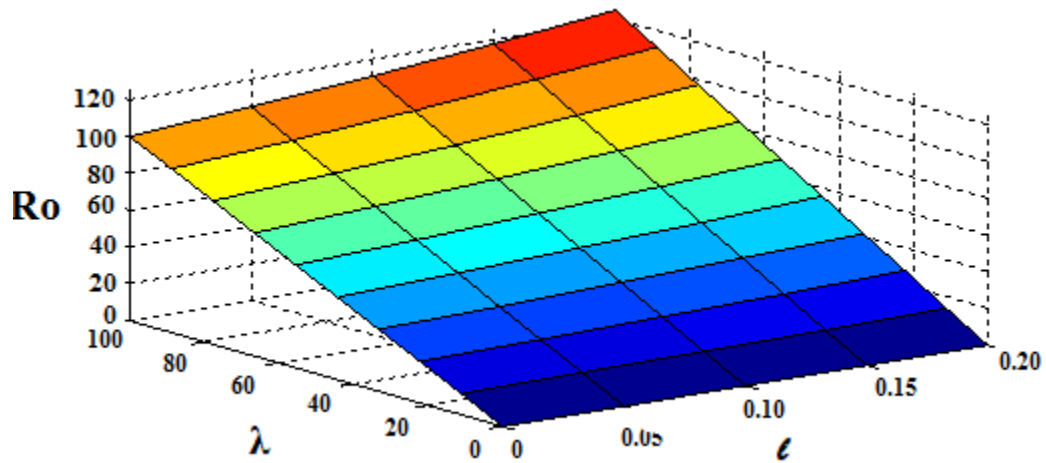


Figura 3.10. Cálculo de R_0 vs λ y l (tres dimensiones)

Datos del segundo escenario

$\lambda = 10$ [sesiones/s] tasa fija

$l = (0, 0.05, 0.1, 0.15, 0.2)$ porcentaje de pérdidas de la red

La tabla 3.5 muestra los datos de las solicitudes R_0 vs el porcentaje de pérdidas y la tasa de llamadas en un horario normal de trabajo.

Respuestas

Datos del tercer escenario

$\lambda = 20$ [sesiones/s] tasa fija

$l = (0.05, 0.1, 0.15, 0.2)$ porcentaje de pérdidas de la red.

Tabla 3.5. Solicitudes *no-Invite* $R_0(l, \lambda, n)$ vs l, λ (horario normal de trabajo)

% Pérdidas l	Tasa [sesiones/s] (λ)	$R_0(l, \lambda, n)$	incremento de $R_0(l, \lambda, n)$ (%)
0	10	10	0
5	10	10.5263	5.26
10	10	11.1111	11.11
15	10	11.7646	17.64
20	10	12.4992	24.99

Tabla 3.6. Solicitudes *no-Invite* $R_0(l, \lambda, n)$ vs l , y λ (horas Pico de trabajo)

% Pérdidas l	Tasa [llamadas/s] (λ)	$R_0(l, \lambda, n)$	incremento de $R_0(l, \lambda, n)$ (%)
0	20	20	0
5	20	21.0526	5.26
10	20	22.2222	11.11
15	20	23.5291	17.64
20	20	24.9984	24.99

La tabla 3.6, muestra las solicitudes R_0 (solicitudes *no-Invite* en horas pico de trabajo).

Datos del cuarto escenario

$\lambda = 100$ [sesiones/s]

$l = (0, .05, 0.1, 0.15, 0.2)$ porcentaje de pérdidas de la red.

La tabla 3.7, muestra las solicitudes R_0 (solicitudes *no-Invite* en casos de emergencia).

Experimento numérico # 6: Pérdidas (L) durante (fase no-Invite)

En el experimento #2 se calcularon las pérdidas durante la fase *Invite*, en este experimento por el contrario se calculan las pérdidas durante la fase *no-Invite*, donde se conocerán las pérdidas de extremo a extremo para una solicitud *no-Invite* más su respuesta. A diferencia de la fase *Invite* la relación entre la respuesta final y el ACK es una de extremo a extremo. Es decir, en esta fase los *proxis* no retransmiten entre si los mensajes perdidos. Para el cálculo de respuesta final (P_f), y solicitudes ACK (P_a) que se

Tabla 3.7. Resultados de $R_0(l, \lambda, n)$ vs l , y λ (Emergencias)

% Pérdidas l	Tasa [llamadas/s] λ	$R_0(l, \lambda, n)$	incremento de $R_0(l, \lambda, n)$ (%)
0	100	100	0
5	100	105.6000	5.60
10	100	111.1110	11.11
15	100	117.6457	11.76
20	100	124.9920	12.49

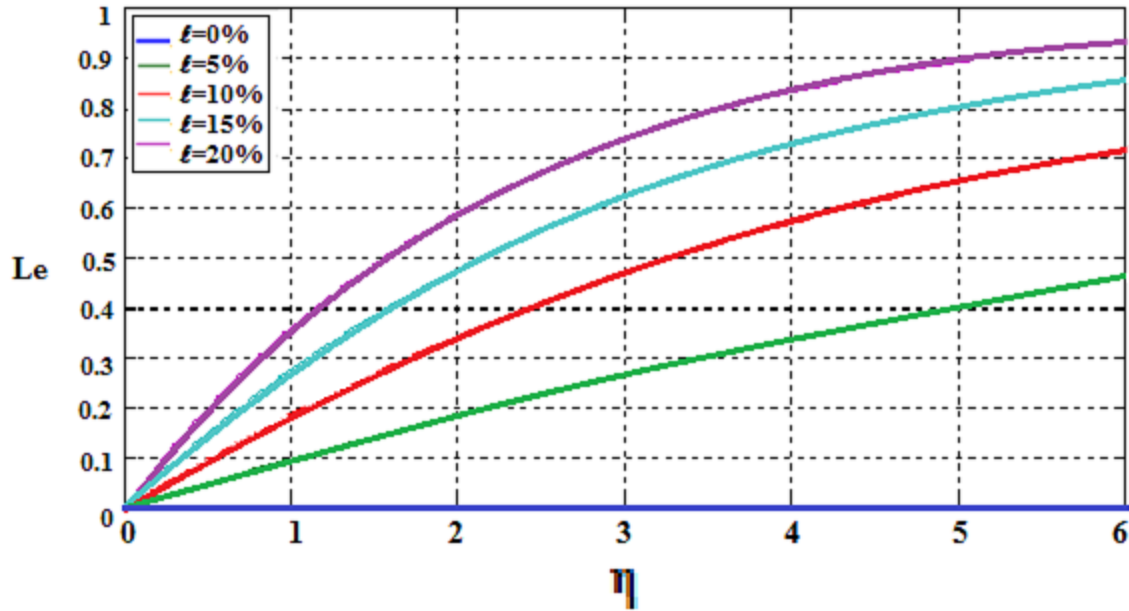


Figura 3.11. Pérdidas L_e vs η y l (Fase *no-Invite*)

requieren en promedio para colocar una sesión, se utilizan las mismas ecuaciones pero tomando en cuenta los retardos de extremo a extremo. Se asume que todas las solicitudes siguen la misma trayectoria (Los *proxis* registran la ruta ellos mismos) con una tasa de pérdida (l) en cada salto η , así que entonces se define a (L_e) como las pérdidas de extremo a extremo en una dirección que viajan a través de η saltos para una solicitud y su respuesta y se calculan con la ecuación (2.15). En este experimento numérico calcularon las pérdidas durante la fase *no-Invite*, para ver el comportamiento de dichas pérdidas cuando los paquetes viajan directamente de extremos a extremo realizando diferentes saltos y su respuesta. El propósito de este experimento es conocer el porcentaje de pérdidas que se producen cuando los paquetes de mensajes realizan un viaje completo en la fase *no-Invite*. Se conoce que los retardos generados durante una puesta de sesión están determinados por los mensajes de la fase *no-Invite*.

Parámetros del experimento

$l = (0, 0.05, 0.10, 0.15, 0.20)$ porcentaje de pérdidas.

$\eta = (2, 3, 4, 6)$ número de saltos (servidores SIP tocados por los mensajes).

Resultados gráficos de L_e se muestran en la figura 3.11, y la tabla 3.8, muestra los resultados numéricos de L_e .

Tabla 3.8. Resultados numéricos de L_e vs l y η

No	% Pérdidas (l)	η saltos	L_e
1	0	2	0
2	5	2	0.1855
3	10	2	0.3439
4	15	2	0.4780
5	20	2	0.5904
6	0	3	0
7	5	3	0.2649
8	10	3	0.4686
9	15	3	0.6229
10	20	3	0.7379
11	0	4	0
12	5	4	0.3366
13	10	4	0.5695
14	15	4	0.7275
15	20	4	0.8322
16	0	6	0
17	5	6	0.4596
18	10	6	0.7176
19	15	6	0.8578
20	20	6	0.9313

Experimento numérico #7: Ancho de Banda (AB)

3.1.1 Cálculo del ancho de banda

La finalidad de este experimento es conocer el Ancho de Banda (AB) que se genera durante la puesta de sesiones telefónicas en redes IMS que operan sobre el protocolo de transporte UDP, este parámetro influye directamente en la QoS que los usuarios perciben cuando se comunican a través de centrales de telecomunicaciones NGN como la propuesta en este proyecto. La estimación del cálculo del AB requerido para establecer una sesión SIP a través de redes IMS sobre UDP, depende de diferentes parámetros de tráfico tales como la tasa de sesiones/segundo, el tiempo de retención promedio y características de la red, particularmente las pérdidas y retardos de propagación, así como el número de saltos de los servidores SIP entre otros. La salida del modelo proveerá detalles sobre el ancho de banda y los retardos requeridos para el establecimiento exitoso de una sesión cuando se utiliza el protocolo SIP sobre redes de transporte UDP en subsistemas multimedia IMS. El Protocolo SIP utiliza ocho fases (*Invite* y *no-Invite*) para establecer una sesión telefónica. Ambas fases *Invite* y *no-Invite* están gobernadas por diferentes temporizadores. En la ecuación del AB se calculan en

forma independiente las diferentes fases y posteriormente se suman. Se asume que los usuarios que solicitan una puesta de sesión ya han realizado la fase de registro en las centrales telefónicas correspondientes.

Parámetros a considerar para el experimento:

$l = (0, 0.05, 0.1, 0.15, 0.20)$ porcentaje de pérdida

$\lambda = (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100)$ tasa de [sesiones/s]

$\eta = 5$ número de saltos (servidores tocados por los mensajes)

$\mu = 3$ timbrados que escuchara el remitente durante el proceso de sesión.

La ecuación (3.5) se utiliza para calcular la primera fase.

$$AB_{f1} = \lambda \times [\eta \times (\sum_{m=0}^{m=6} (1 - (1-l)^2)^m \times 810 + \sum_{m=0}^{m=6} l^m \times 260)], \quad (3.5)$$

Las fases 2 y 3 se obtienen de la ecuación (3.6).

$$AB_{f23} = \lambda \times [\sum_{m=0}^{m=6} (1 - (1-l)^\eta)^m \times 260 + \sum_{m=0}^{m=6} (1 - (1-l)^{2\eta})^m \times 260 + \sum_{m=0}^{m=6} (1 - (1-l)^\eta)^m \times 100], \quad (3.6)$$

La cuarta fase se obtiene utilizando la ecuación (3.7).

$$AB_{f4} = \lambda \times [\sum_{m=0}^{m=6} (1 - (1-l)^{2\eta})^m \times 260 + \sum_{m=0}^{m=6} (1 - (1-l)^\eta)^m \times 100] \quad (3.7)$$

Las fases 5 y 6 se obtienen de la ecuación (3.8).

$$AB_{f56} = \lambda \times [\mu \times (\sum_{m=0}^{m=6} (1 - (1-l)^\eta)^m \times 260 + \sum_{m=0}^{m=6} (1 - (1-l)^{2\eta})^m \times 260 + \sum_{m=0}^{m=6} (1 - (1-l)^\eta)^m \times 100)] \quad (3.8)$$

Finalmente la fase 7 se obtiene con la ecuación (3.9)

$$AB_{f7} = \lambda \times [(\sum_{m=0}^{m=6} (1 - (1-l)^{2\eta})^m \times 100 + \sum_{m=0}^{m=6} (1 - (1-l)^\eta)^m \times 60)] \quad (3.9)$$

Todas estas fases se derivan de la ecuación (2.35), tal que el ancho de banda total estará representado por la suma de todas las fases como:

$$\text{Ancho de Banda} = \sum_{x=1}^7 \text{Fase } x \quad (3.10)$$

Las gráficas de las figura 3.12 y 3.13, muestran el ancho de banda estimado que se requiere durante una puesta de sesión IMS, para diferentes valores de λ , l , μ y η (valores compatibles a los utilizados en redes rurales).

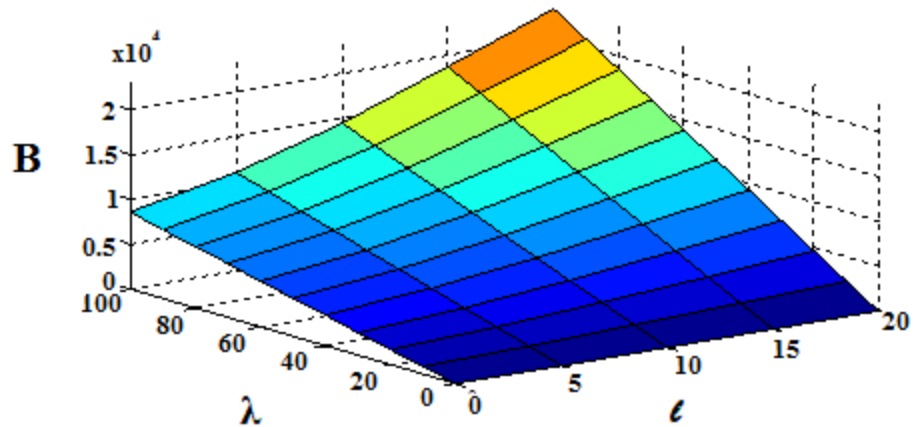


Figura 3.12. Ancho de Banda vs λ y l , con parámetros $\eta = \mu = 5$ (tres dimensiones)

El análisis numérico de las diferentes corridas de datos se analiza en el capítulo 4.

Experimento numérico #8: Cálculo de flujo de tráfico

3.1.2 Cálculo de flujo del tráfico y dimensionamiento de la red

Se sabe en promedio que durante el establecimiento y liberación de una llamada SIP, los servidores P-CSCF y S-CSCF son tocados 11 veces, el I-CSCF 12 veces y el HSS es tocado una vez por los mensajes SIP. Para el análisis de flujo se analizaron dos redes diferentes con operadores diferentes, cada una de estas redes tienen sus propios servidores

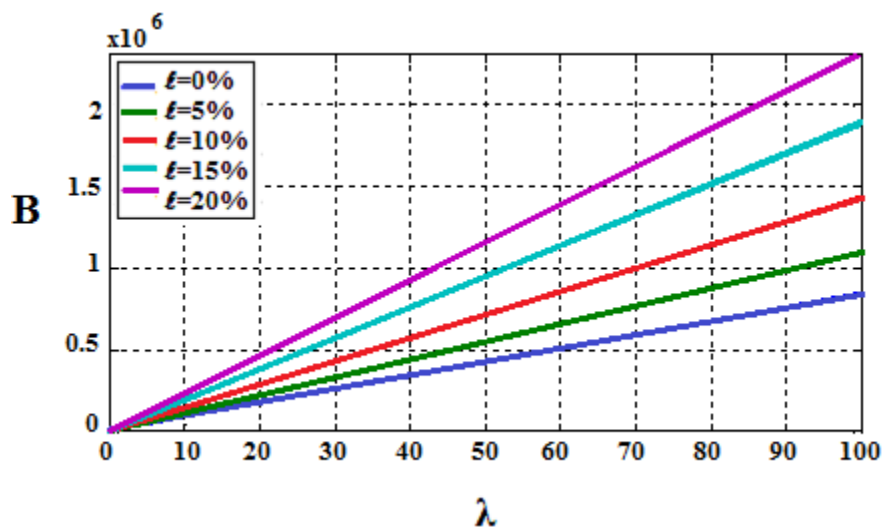


Figura 3.13. Ancho de Banda vs λ y l , con parámetros $\eta = \mu = 5$ (dos dimensiones)

Tabla 3.9. Rutas posibles de los usuarios durante el establecimiento de una llamada

	Escenario	Descripción
1	$H1 \rightarrow H2$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 2 (sin itinerancia)
2	$H2 \rightarrow H1$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 1 (sin itinerancia) llama a
3	$V1 \rightarrow V2$	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 1 con itinerancia en Red 2
4	$V2 \rightarrow V1$	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 2 con itinerancia en Red 1
5	$H1 \rightarrow V2$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 1 con itinerancia en Red 2
6	$V2 \rightarrow H1$	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 1 (sin itinerancia)
7	$V1 \rightarrow H2$	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 2 (sin itinerancia)
8	$H2 \rightarrow V1$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 2 con itinerancia en Red 1
9	$H1 \rightarrow V1$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 2 con itinerancia en Red 1
10	$V1 \rightarrow H1$	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 1 (sin itinerancia)
11	$H2 \rightarrow V2$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 1 con itinerancia en Red 2
12	$V2 \rightarrow H2$	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 2 (sin itinerancia) llama a
13	$H1 \rightarrow H1$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 1 (sin itinerancia)
14	$H2 \rightarrow H2$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 2 (sin itinerancia)

P-CSCF, S-CSCF e I-CSCF, y la manera de ofrecer el servicio que se brinda al usuario será el siguiente: el primero que llega es el primero que se atiende. El tiempo de servicio promedio de P-CSCF es $\overline{X_p}$, el tiempo de servicio de S-CSCF es $\overline{X_s}$, el tiempo medio de servicio de la I-CSCF es $\overline{X_i}$, y el tiempo medio de servicio de HSS es $\overline{X_h}$; en base a estas consideraciones identificaremos que servidor será el cuello de botella con diferentes escenarios de rutas de flujo. Sean todos los posibles escenarios de conexión entre dos redes diferentes como lo muestra la figura 2.25, y la tabla 3.9.

La tabla 3.10, muestra el flujo de tráfico de señalización a través de ambas redes.

Tabla 3.10. Cadena de ruteo del flujo de tráfico de señalización.

	Escenario	Cadena de Ruteo
1	$H1 \rightarrow H2$	P-CSCF1↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2
2	$H2 \rightarrow H1$	P-CSCF2↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1
3	$V1 \rightarrow V2$	P-CSCF1↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF2
4	$V2 \rightarrow V1$	P-CSCF2↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF1
5	$H1 \rightarrow V2$	P-CSCF1↔S-CSCF1↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF2
6	$V2 \rightarrow H1$	P-CSCF2↔S-CSCF1↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1
7	$V1 \rightarrow H2$	P-CSCF1↔S-CSCF2↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2
8	$H2 \rightarrow V1$	P-CSCF2↔S-CSCF2↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF1
9	$H1 \rightarrow V1$	P-CSCF1↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF1
10	$V1 \rightarrow H1$	P-CSCF1↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1
11	$H2 \rightarrow V2$	P-CSCF2↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF2
12	$V2 \rightarrow H2$	P-CSCF2↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2
13	$H1 \rightarrow H1$	P-CSCF1↔S-CSCF1↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1
14	$H2 \rightarrow H2$	P-CSCF2↔S-CSCF2↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2

H1: Subscriptor de la Red 1 (sin itinerancia)

H2: Subscriptor de la Red 2 (sin itinerancia)

V1: Subscriptor de la Red 2 con itinerancia en la Red 1

V2: Subscriptor de la Red 1 con itinerancia en la Red 2

Tenemos que por cada procedimiento de colocar una llamada y liberarla, los servidores P-CSCF y S-CSCF involucrados son tocados 11 veces, y el servidor I-CSCF es tocado 12 veces; Suponga que la carga total de tráfico es λ [sesiones/s], la carga de tráfico para cada escenario es λ_i , y $\lambda = \sum \lambda_i$. Entonces de acuerdo a la tabla 21 se tiene que el total de veces que es tocado cada nodo quedaría:

P-CSCF1:

$$\Lambda_1 = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8) \times 11 + \lambda_9 + \lambda_{10} + \lambda_{13})22; \quad (3.11)$$

P-CSCF2:

$$\Lambda_2 = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8) \times 11 + (\lambda_{11} + \lambda_{12} + \lambda_{14})22; \quad (3.12)$$

S-CSCF1:

$$\Lambda_3 = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_9 + \lambda_{10} + \lambda_{11} + \lambda_{12}) \times 11 + (\lambda_5 + \lambda_6 + \lambda_{13})22, \quad (3.13)$$

S-CSCF2:

$$\Lambda_4 = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_9 + \lambda_{10} + \lambda_{11} + \lambda_{12}) \times 11 + (\lambda_7 + \lambda_8 + \lambda_{14}); \quad (3.14)$$

I-CSCF1:

$$\Lambda_5 = (\lambda_2 + \lambda_3 + \lambda_5 + \lambda_6 + \lambda_{10} + \lambda_{11} + \lambda_{13}) \times 12; \quad (3.15)$$

I-CSCF2:

$$\Lambda_6 = (\lambda_1 + \lambda_4 + \lambda_5 + \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{12} + \lambda_{14}) \times 12; \quad (3.16)$$

HSS1:

$$\Lambda_7 = (\lambda_2 + \lambda_3 + \lambda_5 + \lambda_6 + \lambda_{10} + \lambda_{11} + \lambda_{13}); \quad (3.17)$$

HSS2:

$$\Lambda_8 = (\lambda_1 + \lambda_4 + \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{12} + \lambda_{14}); \quad (3.18)$$

Calculando la utilización que se hace de cada entidad o servidor seria:

$$P - CSCF1: \rho_p = \Lambda_1 \times \overline{X_p} \quad (3.19)$$

$$P - CSCF2: \rho_p = \Lambda_2 \times \overline{X_p} \quad (3.20)$$

$$S - CSCF1: \rho_s = \Lambda_3 \times \overline{X_s} \quad (3.21)$$

$$S - CSCF2: \rho_s = \Lambda_4 \times \overline{X_s} \quad (3.22)$$

$$I - CSCF1: \rho_i = \Lambda_5 \times \overline{X_i} \quad (3.23)$$

$$I - CSCF2: \rho_i = \Lambda_6 \times \overline{X_i} \quad (3.24)$$

$$HSS1: \rho_h = \Lambda_7 \times \overline{X_h} \quad (3.25)$$

$$HSS2: \rho_h = \Lambda_8 \times \overline{X_h} \quad (3.26)$$

Si asumimos que $\overline{X_p} = \overline{X_s} = \overline{X_i} = \overline{X_h}$, se obtendrá el cuello de botella si se conoce la carga de tráfico de cada escenario. El programa de cálculo del cuello de botella en este sistema de dos redes se adjunta como apéndice C al final de la tesis.

CAPITULO 4

Resultados

4.1 Análisis de Resultados

Experimento numérico #1: Cálculo de las solicitudes R_i (fase Invite)

En este primer experimento se calcularon las solicitudes (R_i) de la fase *Invite* generadas por el subsistema IMS durante el establecimiento de una sesión telefónica. Para el cálculo se utilizaron los siguientes parámetros como entradas: tráfico de llamadas ($\lambda = 10, 20$ y 100) [sesiones/s] y porcentajes de pérdidas de la red $l = (0, 0.05, 0.10, 0.15, 0.20)$. Para propósitos de análisis los datos obtenidos de R_i durante las corridas se muestran en la tabla 4.1, y se grafican en las figuras 2.42 y 2.43.

Para propósito de análisis se considera que el modelo de red rural diseñado opera un porcentaje de pérdida de la red del 5 por ciento (0.05). Además, se considera que el parámetro de mayor importancia de la red es la QoS y esta dependerá del ancho de banda que la red IP genere durante la puesta de sesiones. Por esta razón se analizarán los efectos que producirán en el “ancho de banda” los diferentes parámetros (R_i , l_e , P_i , P_{100} , SIE , R_0 , y L_e) que participan en el establecimiento de una sesión telefónica SIP (durante las fases Invite y no-Invite). Con esto se conocerá el efecto de los parámetros mencionados en QoS de la red.

Tabla 4.1. Cálculo R_i experimento #1 (fase *Invite*).

No	(l) %	λ [sesiones/s]	Solicitudes R_i	Incremento R_i %
1	0	10	10	0
2	5	10	10.5	5
3	10	10	11.1	11
4	15	10	11.7	17.6
5	20	10	12.4	24
6	0	20	20	0
7	5	20	21	5.2
8	10	20	22.2	11.1
9	15	20	23.5	17.6
10	20	20	24.9	24.5
11	0	100	100	0
12	5	100	105.2	5.2
13	10	100	111.1	11.1
14	15	100	117.6	17.6
15	20	100	124.9	24.9

Análisis de solicitudes *Invite* (R_i), generadas cuando se coloca una llamada telefónica, y la red está configurada para atender un tráfico de (10, 20 y 100) [sesiones/s], con un porcentaje de pérdida del 5 por ciento. Para este análisis se considera como base de referencia el valor de R_i generada por una red sin pérdidas. Con el fin de conocer un indicador estadístico que muestre la desviación de los datos (incremento de R_i) con respecto a la media aritmética. Esto con el fin de visualizar los mismos lo más cercano a la realidad durante su interpretación. Se concluye que bajo las condiciones de tráfico y porcentaje de pérdidas mencionados, las variables R_i se incrementan en un porcentaje menor del 0.1% (0.0942). Por lo que estos valores aseguran una calidad de servicio que permite la comunicación entre usuarios de forma clara y segura, lo cual le da validez al modelo propuesto.

Experimento numérico #2: Cálculo de pérdidas entre servidores (fase *Invite*)

En este experimento numérico se calcularon las pérdidas que sufren los mensajes cuando viajan de servidor a servidor durante la fase *Invite*. La figura 2.44, muestra un gráfico del enlace donde se medirán las pérdidas de extremo a extremo. Para esto se definió a l como la probabilidad de pérdidas que sufren las solicitudes *Invite* (cuando viajan de un servidor a otro en una dirección), y l_e como las pérdidas que sufren las solicitudes *Invite* más su respuesta correspondiente (pérdidas de extremo a extremo entre servidores SIP). El resultado de las corridas con ($l = 0, 0.05, 0.10, 0.15, 0.20$) se muestran en la Tabla 4.2. El objetivo del experimento es mostrar que las pérdidas generadas durante el proceso de la puesta de sesión (en la fase *Invite*), inciden en los retardos propios del IMS. Se observa de la tabla 4.2 que para diferentes porcentajes de pérdidas de la red en una dirección, existirá un porcentaje de pérdida de la red en dos direcciones, el cual es menor del doble contemplado para una dirección. Se sabe que a mayores pérdidas en las redes (debido a su arquitectura interna), mayor serán las solicitudes *Invite* que debe generar el IMS, produciendo un ancho de bando mayor y por lo tanto creando retardos en la operación. Las pérdidas originadas por mal funcionamiento de servidores o fallas en la arquitectura de la red, se reflejarán en la QoS detectada por los usuarios en la misma. El modelo de red propuesto en este proyecto, opera con un

Tabla 4.2. Pérdidas l_e de extremo a extremo (dos direcciones).

No	(l) Pérdidas %	(l_e) Pérdidas %
1	0	0
2	5	9.75
3	10	19
4	15	27.75
5	20	36

porcentaje de pérdidas promedio del 5%. Esto significa de acuerdo a la tabla 4.2, que las pérdidas en dos direcciones serán de 9.75%, esta información le indica al operador de la red que durante el diseño de la misma, éste deberá considerar pérdidas menores al 5% para poder ofrecer una calidad de servicio aceptable.

Experimento numérico #3: Cálculo de las P_i y las P_{100} (fase Invite)

En este experimento se analizan las funciones (P_i y P_{100}). Las P_i se definen como el número de mensajes R_i enviados por el remitente (UAC) después de que el subsistema IMS alcanza el estado estable, dividido mediante el número de mensajes originales que fueron enviados y que no han sido retransmitidos; así mismo, las P_{100} se definen como el número de respuestas exitosas generadas durante el establecimiento de una sesión, después de que el subsistema haya alcanzado el estado estable. La figura 2.44, muestra cuando el remitente (UAC) genera una solicitud y el servidor P-CSCF, regresa una respuesta P_{100} a esta solicitud, ambas (solicitud y respuesta) se obtienen cuando se alcanza el estado estable del sistema. Las ecuaciones (2.5) y (2.7) se utilizan para los cálculos correspondientes. El propósito del experimento es conocer cuantas solicitudes *Invite* R_i se deben generar por el IMS una vez que éste alcance el estado estable (cuando los mensajes transmitidos son exitosos). Los datos numéricos obtenidos del experimento (tabla 4.3) ayudarán a conocer en qué porcentaje se verá afectado el ancho de banda generado por el IMS, por el efecto de las solicitudes y respuestas *Invite* mencionadas.

Tabla 4.3. Resultados numéricos de las P_i vs l y λ

No	(l) Pérdidas %	Tasa sesiones (λ)	Solicitudes P_i	Respuestas P_{100}
1	0	10	1	1
2	5	10	1.1080	1.0526
3	10	10	1.2345	1.1111
4	15	10	1.3835	1.1765
5	20	10	1.5591	1.2499
6	0	20	1	1
7	5	20	1.1080	1.0526
8	10	20	1.2345	1.1111
9	15	20	1.3835	1.1765
10	20	20	1.5591	1.2499
11	0	100	1	1
12	5	100	1.1080	1.0526
13	10	100	1.2345	1.1111
14	15	100	1.3835	1.1765
15	20	100	1.5591	1.2499

Se observa de la tabla 4.3 que los parámetros P_i y P_{100} no se ven afectadas por cambios en la tasa de tráfico λ , pero si se afectan por los cambios en porcentajes de pérdida de paquetes (0.05, 0.10, 0.015, 0.20). Sin embargo, debido a que el modelo de red propuesto opera con un porcentaje fijo de pérdidas (5%), bajo estas circunstancias los parámetros P_i y P_{100} se mantienen sin cambios para las tres muestras de tráfico mencionadas. Se concluye del análisis que las solicitudes *Invite* (R_i), generadas por el IMS durante el proceso de las puestas de sesiones operando tasas de tráfico de (10, 20 y 100) [sesiones/s], no se verán afectadas en forma alguna ($\sigma = 0$); ya que el ancho de banda no se altera en redes sin pérdidas.

Experimento numérico #4: Cálculo de solicitudes *Invite* exitosas SIE (fase *Invite*)

Tabla 4.4. Solicitudes (SIE)

No	(<i>l</i>) pérdidas %	# saltos (η)	Solicitudes (SIE) exitosas
1	0	2	4.0000
2	5	2	4.3213
3	10	2	4.6912
4	15	2	5.1198
5	20	2	5.6180
6	0	3	6.0000
7	5	3	6.4820
8	10	3	7.0369
9	15	3	7.6797
10	20	3	8.4271
11	0	4	8.0000
12	5	4	8.6427
13	10	4	9.3825
14	15	4	10.2396
15	20	4	11.2361
16	0	5	10.0000
17	5	5	10.8033
18	10	5	11.7281
19	15	5	12.7995
20	20	5	14.0451
21	0	6	12.0000
22	5	6	12.9640
23	10	6	14.0737
24	15	6	15.3594
25	20	6	16.8541

En el experimento anterior se calcularon los parámetros P_i y P_{100} con los cuales se conocerán el número de solicitudes *Invite* R_i que el subsistema IMS genera cuando alcance el estado estable (ida y vuelta). En este experimento, se calcularon las solicitudes *Invite* R_i cuando se hace una llamada en estado de itinerancia. Es decir, que dependiendo del tipo de llamada que el usuario desee realizar será el número de saltos que los mensajes tendrán que realizar a través de las diferentes redes, hasta alcanzar al usuario que se desea contactar. Con la ecuación (3.4) se calcula el número de solicitudes *Invite* exitosas, después de que el sistema haya alcanzado el estado estable, y que los mensajes hayan atravesado η servidores SIP en las redes IMS. La figura 2.44, indica la ruta que siguen los mensajes mencionados. Las figuras 2.48 y 2.49, muestran las gráficas en tres y dos dimensiones de los resultados del *SIE* vs el porcentaje de pérdidas y el número de saltos. La tabla 4.4 muestra el número de solicitudes *Invite* requeridas para colocar una llamada exitosa en estado de itinerancia. Véase el siguiente ejemplo, se desea colocar una sesión telefónica a través de redes IMS con porcentaje de pérdidas del 5%. La ruta de tráfico seleccionada por las redes requiere tocar dos servidores SIP para poder conectarse con el otro usuario buscado. La tabla 4.4, muestra que el IMS debe generar **4.3** solicitudes *Invite*, para lograr que el remitente de la llamada se conecte exitosamente con el receptor. Analizando ahora otro caso cuando un usuario de la red desea realizar una llamada telefónica utilizando la misma red del ejemplo anterior, solo que ahora la ruta de tráfico cambia, y esta vez los mensajes tocarán a tres servidores SIP para poder tener contacto al usuario buscado. En esta situación el IMS debe generar **6.4** solicitudes *Invite*, para lograr su objetivo. Véase ahora el caso cuando un usuario se encuentra fuera de casa (itinerancia) y desea colocar una sesión telefónica de larga distancia, para esto las solicitudes *Invite* tocan hasta seis servidores SIP para poder establecer la llamada a otro usuario ubicado en otra central. En esta situación los subsistemas IMS que participan en el enlace deben generar en total **12.9** solicitudes *Invite* para poder conectar a estos dos usuarios. Resumiendo el análisis se puede decir que para que dos usuarios inscritos en la misma red puedan conectarse a través de una llamada telefónica. El IMS de la red deberá generar hasta cuatro solicitudes *Invite* para poder interconectarlos. Véase ahora el caso cuando dos usuarios localizados en ciudades diferentes, e inscritos en sus propias redes de casa, deseen conectarse entre sí (llamada itinerante). Los subsistemas IMS de las redes telefónicas que participen en el enlace deberán generar aproximadamente **13** solicitudes *Invite*, para poder lograr su objetivo. Esto significa que entre más lejos se encuentre un usuario de la red donde está suscrito, mayor será el número de solicitudes *Invite* que se generará, durante el proceso de una puesta de sesión.

De la tabla 4.4 se observa que el número de *Solicitudes Invite Exitosas* que se requieren generar cuando se coloca una llamada en la red, dependerán del porcentaje de pérdida de la misma, y del tipo de llamada a realizar (local o itinerante), si es local ($\eta = 2$) o si es itinerante ($\eta = 3, 4, 5, \text{ o } 6$), y considerando como referencia el valor de las *SIE* generadas cuando la red no tiene pérdidas. Con estos parámetros se puede obtener un

indicador estadístico que muestre la dispersión de los datos generados (*SIE*) con respecto a la media aritmética, con el fin de visualizar los resultados lo más cercano a la realidad posible.

Con todo esto se concluye que cada vez que un usuario (UAC) solicita realizar una llamada telefónica con otro, el IMS genera 2.5 ($\sigma = 2.4157$) *Solicitudes Invite Exitosas* para poderse conectarse con la contraparte (UAS), siempre y cuando la red opere con un porcentaje de pérdidas del cinco por ciento.

Experimento numérico #5: Cálculo de las solicitudes R_0 (fase *no-Invite*)

En este experimento se calcularon las solicitudes (R_0) de la fase *no-Invite* generadas por el IMS durante el establecimiento de una sesión telefónica, para esto se utilizaron parámetros de entrada tales como: tráfico de llamadas λ (10, 20 y 100) [sesiones/s] y porcentajes de pérdidas de la red l (0, 0.05, 0.1, 0.15, y 0.2). Los datos obtenidos durante las corridas se presentan en la tabla 4.5 para propósitos de análisis, y se grafican en las figuras 2.50 y 2.51. Para este propósito se utilizó la ecuación (2.12).

Es importante mencionar que las re-transmisiones $R_{PRACK}/200$, $UPDATE/200$, $200/ACK$ también pertenecen a la fase *no-Invite*, y todas estas afectaran también el comportamiento del ancho de banda que genera el IMS durante el proceso de puesta de sesión. De la tabla 4.5 se comprueba que el número de solicitudes R_0 es directamente proporcional a la tasa de tráfico (λ) y a los diferentes porcentajes de pérdida de la red (l).

Tabla 4.5. Solicitudes R_0 vs λ y l (fase *no-Invite*)

No	(l) %	λ [sesiones/s]	Solicitudes R_0	R_0 Incremento %
1	0	10	10.0000	0
2	5	10	10.5263	5.26
3	10	10	11.1111	11.11
4	15	10	11.7646	17.64
5	20	10	12.4992	24.99
6	0	20	20.0000	0
7	5	20	21.0526	5.26
8	10	20	22.2222	11.11
9	15	20	23.5291	17.64
10	20	20	24.9984	24.99
11	0	100	100.0000	0
12	5	100	105.6000	5.60
13	10	100	111.1110	11.11
14	15	100	117.6457	17.64
15	20	100	124.9920	24.99

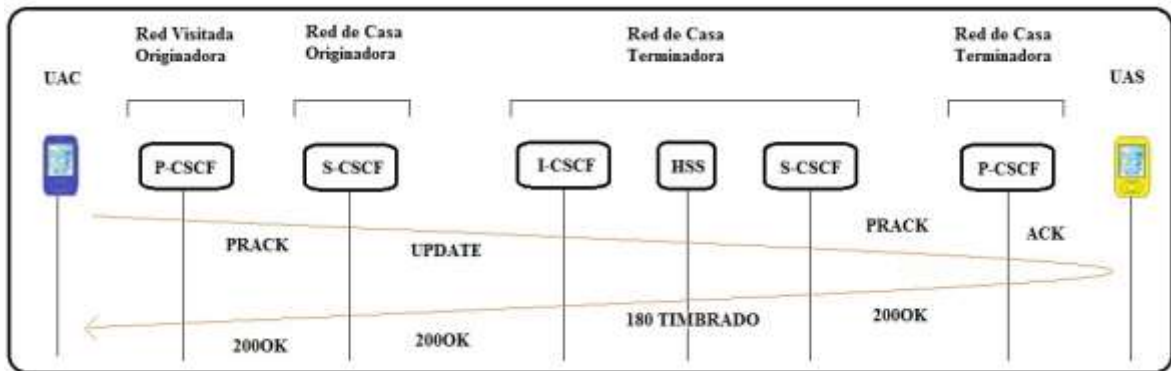


Figura 4.1. Recorrido de las re-transmisiones *no-Invite*

Considerando que la QoS depende del ancho de banda que la red genere durante los procesos de las sesiones, y que la red propuesta opera pérdida del 5 %. Bajo estas circunstancias, cuando los usuarios (UAC) colocan sesiones telefónicas (en cualquier horario), se obtiene el número de solicitudes R_0 generadas que el IMS, y el porcentaje de cambio que se producirá en estas, debido al tráfico cursado por la red y las pérdidas de la misma. Para esto se consideró como referencia el valor de R_0 para una red sin pérdidas. Con esta información se obtendrá un indicador estadístico que muestre la desviación de los datos R_0 de la tabla 4.5, con respecto a la media aritmética, con el fin de visualizarlos, lo más cercano a la realidad durante su interpretación. Se concluye que bajo las condiciones de tráfico y porcentaje de pérdidas mencionados, las variables R_0 se incrementan en un porcentaje mínimo ($\sigma = 0$), lo cual hace al subsistema IMS estable y por lo tanto predecible.

Experimento numérico #6: Cálculo de porcentaje de pérdida L_e (fase *no-Invite*)

El propósito de este experimento es conocer el porcentaje de pérdidas que se producen cuando los paquetes de mensajes realizan un viaje completo en la fase *no-Invite*. Para esto se calcularon las pérdidas de extremo a extremo de las re-transmisiones *no-Invite* y su respuesta correspondiente (mensajes que viajan a través de η saltos en dos direcciones). Para visualizar el comportamiento de dichas pérdidas se estudió el recorrido que realizan los mensajes, cuando participa la fase *no-Invite* durante el proceso de la puesta de sesión. En la gráfica de la figura 4.1, se muestra el recorrido que hacen los mensajes generados durante la fase *no-Invite* (PRACK/200OK, UPDATE/200OK, TIMBRADO 180 Y ACK) y la secuencia de retransmisiones donde se muestran los recorridos que hacen las solicitudes desde el UAC hasta el UAS y viceversa, todo esto como parte del proceso para establecer una sesión SIP. Estas re-transmisiones (*no-Invite*) se inician en la fase tres y continúan hasta la fase 7 terminando con la re-transmisión ACK (última fase del proceso), que es la fase donde se reconoce que ambos usuarios han sido conectados.

Analizando cuidadosamente la tabla 4.6, y utilizando el ejemplo cuando un usuario remitente (UAC) desea colocar una llamada telefónica *local* a través de una red con pérdidas (0.05) para comunicarse con otro usuario (UAS), la trayectoria de las retransmisiones *no-Invite* tocarán por lo menos dos servidores SIP ($\eta = 2$). Durante el proceso de la puesta de sesión se producirán pérdidas del 18% de los mensajes durante el viaje completo. Analizando ahora el caso cuando un UAC ubicado fuera de su red de casa (otra ciudad) realiza una llamada telefónica (en itinerancia) a un UAS localizado en otra ciudad, donde los mensajes viajan a través de seis servidores SIP ($\eta = 6$) durante la puesta de sesión, bajo estas condiciones los IMS participantes generan un porcentaje de error del 45% por la trayectoria del viaje que realizaron las re-transmisiones *no-Invite* durante el proceso de puesta de sesión.

La tabla 4.6, muestra los datos numéricos calculados de L_e vs l y η .

Se concluye entonces que el ancho de banda generado durante el proceso de la puesta de sesión lo determinan en su mayoría el número de re-transmisiones *no-Invite*.

Tabla 4.6. Cálculo de L_e vs l y η

No	(l) Pérdidas %	η	L_e
1	0	2	0.0000
2	5	2	0.1855
3	10	2	0.3439
4	15	2	0.4780
5	20	2	0.5904
6	0	3	0.0000
7	5	3	0.2649
8	10	3	0.4686
9	15	3	0.6229
10	20	3	0.7379
11	0	4	0.0000
12	5	4	0.3366
13	10	4	0.5695
14	15	4	0.7275
15	20	4	0.8322
16	0	6	0.0000
17	5	6	0.4596
18	10	6	0.7176
19	15	6	0.8578
20	20	6	0.9313

Experimento numérico #7: Análisis del ancho de banda (fases *Invite* y *no-Invite*)

La finalidad de este experimento es conocer la cantidad total de Ancho de Banda (B) que se genera mediante la señalización SIP, cuando se establecen llamadas IMS, operando un tráfico de llamadas rurales (λ), y ejecutando un número de (η) saltos de los mensajes a través de los servidores SIP, que forman partes de los sub-sistemas IMS que participaran en los enlaces. Particularmente, los experimentos numéricos realizados se enfocan en el comportamiento del ancho de banda (B) de tres comunidades rurales. La primera comunidad tiene hasta 1,500 usuarios adscritos al servicio telefónico, la segunda 15,000 usuarios, y la tercera contempla hasta 35,000. Estos valores son típicos en México para describir densidad de usuarios rurales. La primera comunidad atiende un tráfico de 10 [sesiones/s], la segunda 20 [sesiones/s] y la tercera hasta 100 [sesiones/s]. Los tres casos contemplan porcentajes de pérdida de paquetes del 5% y del 20%.

¿Con que fin se realiza este análisis?

La finalidad del experimento es conocer cuál será el comportamiento del ancho de banda generado durante el establecimiento simultáneo de llamadas telefónicas locales e itinerantes, cuando se instalan redes NGN en comunidades rurales de diferente tamaño.

¿Por qué es importante conocer el Ancho de Banda?

Porque el Ancho de Banda garantiza la calidad de Servicio (QoS) de la red, y con esta se garantiza una red eficiente y funcional. Por lo tanto, conociendo el tráfico de la red y el ancho de banda que se genera durante el establecimiento de sesiones telefónicas, así como también sus características intrínsecas de la misma, se podrá garantizar una red dimensionada con calidad de servicio y sin congestiones.

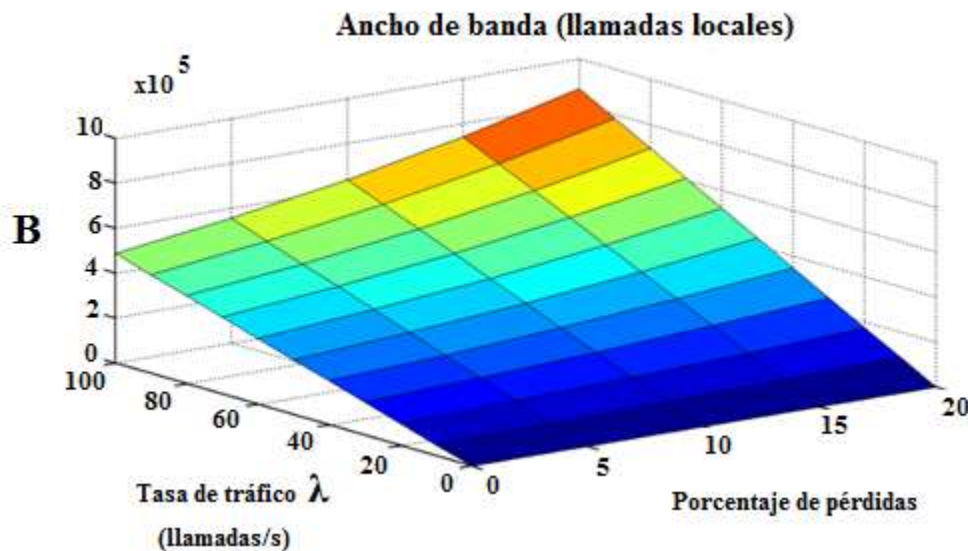


Figura 4.2. Ancho de banda (B) generado por el IMS con llamadas locales

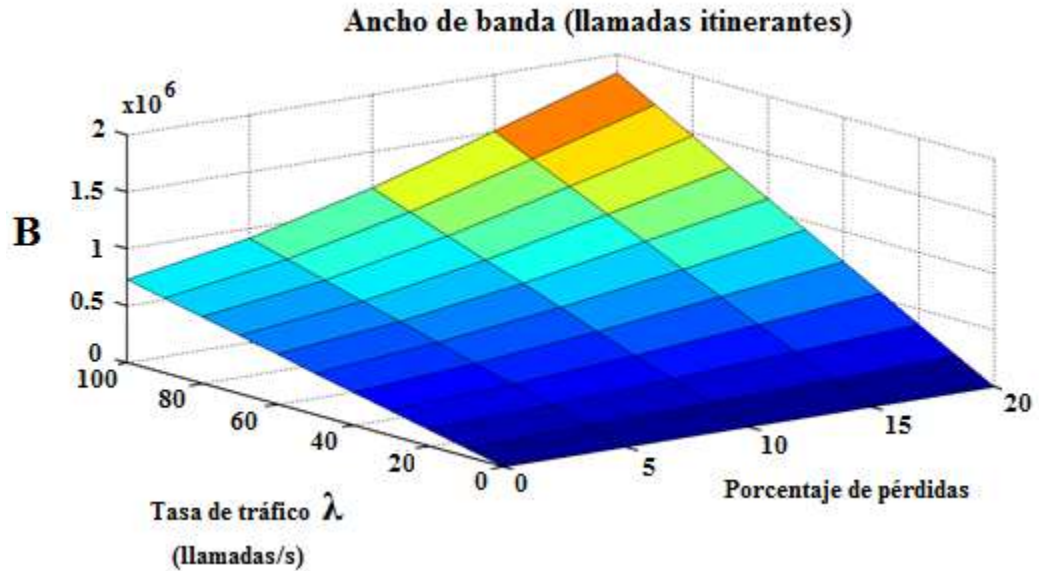


Figura 4.3. Ancho de banda (B) generado por el IMS con llamadas itinerantes

Con este propósito, la figura 4.2, muestra la gráfica del ancho de banda generado por los usuarios que realizan llamadas telefónicas locales, también muestra los cambios producidos por los efectos de los parámetros de tráfico (10, 20 y 100 [sesiones/s]), y los diferentes porcentajes de pérdidas de la red (0, 0.05, 0.1, 0.15 y 0.2).

La Figura 4.3, muestra la gráfica del ancho de banda generado cuando los usuarios realizan llamadas itinerantes, también muestra los cambios producidos por los efectos de los parámetros de tráfico, y los diferentes porcentajes de pérdidas de la red mencionados.

La Tabla 4.7, muestra los datos de B, cuando la red opera porcentajes de pérdida de 0.05 y 0.20, y tasas de tráfico de (10, 20 y 100) [sesiones/s] para las tres comunidades respectivamente. Se observa de la tabla 4.7, que para una red (con pérdidas del 5%) instalada en una comunidad rural de 1500 usuarios, cuando los usuarios realizan llamadas *locales*, el ancho de banda generado por el sistema será de 56 kB; el cual es un valor razonable y sencillo de generar; mientras que para el caso de una red instalada para una comunidad de mediana capacidad (15000 usuarios) y con las mismas condiciones de operación anteriores, el ancho de banda generado por la red es de 111 kB; resulta interesante que éste incremento exige cerca del doble de ancho de banda que el sistema anterior; por último, analizando la red instalada en la comunidad de 35000 usuarios, se puede observar que la red demanda un ancho de banda de 556 kB. El dimensionamiento requerido para ésta comunidad exigirá de dotar al sistema con un número mayor de servidores SIP en el IMS para ofrecer una calidad de servicio adecuado.

Cuando se analiza el comportamiento de la red operando con 20% de pérdidas, se puede observar para el caso de una comunidad rural con 1,500 usuarios inscritos a la red (servicios itinerantes únicamente), que la red genera un ancho de banda de 184 kB;

Tabla 4.7. Anchos de banda (B) para las tres comunidades rurales

Número de (usuarios)	Tipo de llamadas	Perdidas l (%)	Tasa de Tráfico λ [sesiones/s]	Ancho Banda B Kilobytes (kB)
1, 500	Locales	5	10	56
15, 000	Locales	5	20	111
35, 000	Locales	5	100	556
1, 500	Itinerantes	20	10	184
15, 000	Itinerantes	20	20	368
35, 000	Itinerantes	20	100	1838

es decir supera las características del caso de 15,000 usuarios con pérdidas del 5% (sin itinerancia). En el caso de la red en la comunidad de 15,000 usuarios, se observa que la red debe generar un ancho de banda de 368 kB, y por último analizando la comunidad de 35,000 usuarios, la red exige un ancho de banda de 1,838 kB. Esta velocidad equivale a la tercera parte de una línea OC-1 de SONET, y el sistema la puede generar sin necesidad de aumentar el número de servidores SIP en el IMS. Comparando ahora entre los casos sin y con itinerancia se observa en el caso en que la red opera con un porcentaje de pérdidas del 20% el ancho de banda se incrementa aproximadamente hasta 330% con respecto a la red cuando opera al 5%. Si se analiza la dispersión que sufre el B respecto a su valor medio cuando opera las diferentes tasas de tráfico mencionadas en la tabla 4.7, se obtiene una desviación de $\sigma = 224$ kB respecto a su valor medio 241 kB (para la red con 5% de pérdidas). Y una desviación de $\sigma = 740$ kB respecto a su valor medio de 797 kB (para la red con 20% de pérdidas). Estos valores del ancho de banda representan cantidades satisfactorias para servicios de VoIP sobre redes IMS con QoS, ya que no superan la capacidad mínima de una línea óptica OC-1.

En el capítulo 2 se propuso un método para calcular el flujo del tráfico que opera la red. Esto con el fin de visualizar el rendimiento de la misma identificando los cuellos de botella, dimensionando y balanceando la red. La tabla 4.8, muestra los servidores SIP que actúan como cuellos de botella cuando dos redes se conectan entre sí, bajo diferentes situaciones de tráfico. Esto significa que cuando el tráfico entre subscriptores y visitantes de la Red-1 (sin itinerancia) se hace más pesado, el servidor S-SCF1 será el cuello de botella (situación 6), sin embargo, si el tráfico entre subscriptores de la Red-2 (sin itinerancia) y visitantes se hace más pesado, entonces P-CSCF2 será el cuello de botella (situación 2). Si se aumenta la carga de tráfico entre subscriptores de la red #1 (sin itinerancia y con itinerancia), S-CSCF1 será el cuello de botella (situación 5), etc. Y debido a que existen posibilidades infinitas de distribuciones de tráfico, teóricamente, cada nodo puede ser el cuello de botella en determinada situación.

Experimento numérico #8: (Rutas para la distribución del tráfico)

Tabla 4.8. Visualización del embotellamiento de servidores CSCF

	1	2	3	4	5	6
λ_1/λ	0.2	0.1	0.05	0.05	0.1	0.1
λ_2/λ	0.2	0.1	0.05	0.05	0.1	0.1
λ_3/λ	0.05	0.05	0.05	0.05	0.05	0.05
λ_4/λ	0.05	0.05	0.05	0.05	0.05	0.05
λ_5/λ	0.05	0.05	0.15	0.05	0.15	0.05
λ_6/λ	0.05	0.05	0.15	0.05	0.15	0.05
λ_7/λ	0.05	0.05	0.05	0.15	0.05	0.05
λ_8/λ	0.05	0.05	0.05	0.15	0.05	0.05
λ_9/λ	0.03	0.03	0.05	0.05	0.05	0.15
λ_{10}/λ	0.03	0.03	0.05	0.05	0.05	0.15
λ_{11}/λ	0.02	0.12	0.05	0.05	0.05	0.05
λ_{12}/λ	0.02	0.12	0.05	0.05	0.05	0.05
λ_{13}/λ	0.1	0.1	0.1	0.1	0.05	0.05
λ_{14}/λ	0.1	0.1	0.1	0.1	0.05	0.05
Servidores CSCF	P-CSCF1	P-CSCF2	S-CSCF1	S-CSCF2	S-CSCF1	P-CSCF1

Con el fin de conocer la eficiencia de la red, se realiza el análisis de flujo de tráfico del subsistema multimedia de la red-1, la cual consiste de las siguientes entidades: P-CSCF1, S-CSCF1, I-CSCF1, HSS1. Debido a la naturaleza de las redes (modo de operación en su área de cubrimiento), los servidores SIP (CSCF) en una red dada, requieren operar con los siguientes escenarios. Con el propósito de facilitar el análisis, en la tabla 4.9, se enlistan los servidores CSCF que participan de esta red, y no se mencionan

Tabla 4.9. Servidores participantes en la red con diferentes escenarios posibles

1	Subscriber localizado en su área de servicio
	a. Subscriber como remitente $\dots \leftrightarrow$ P-CSCF \leftrightarrow S-CSCF $\leftrightarrow \dots$
	b. Subscriber c/ receptor $\dots \leftrightarrow$ I-CSCF \leftrightarrow HSS \leftrightarrow S-CSCF \leftrightarrow P - CSCF $\leftrightarrow \dots$
2	Subscriber con itinerancia en una red visitada
	a. Subscriber como remitente $\dots \leftrightarrow$ S-CSCF $\leftrightarrow \dots$
	b. Subscriber c/receptor $\dots \leftrightarrow$ I-CSCF \leftrightarrow HSS \leftrightarrow S-CSCF $\leftrightarrow \dots$
3	Visitante no suscrito con itinerancia en su área de servicio
	a. Visitante como remitente $\dots \leftrightarrow$ P-CSCF $\leftrightarrow \dots$
	b. Visitante como receptor $\dots \leftrightarrow$ P-CSCF $\leftrightarrow \dots$

Tabla 4.10. Propuesta de segmentación del porcentaje de tráfico

1a	1b	2a	2b	3a	3b
A1%	A2%	B1%	B2%	C1%	C2%

otras unidades funcionales o servidores de otras redes. Se sabe que la carga de tráfico total en la red es λ , y suponiendo que el porcentaje de carga de cada escenario se distribuye de acuerdo a la tabla 4.10, y con el propósito de calcular la utilización de cada nodo se utilizan las funciones (4.1), (4.2) y (4.3).

$$\text{P-CSCF: } \rho_p = (A1+A2+C1+C2) \% \times \lambda \times 11 \times \overline{X_p}; \quad (4.1)$$

$$\text{S-CSCF: } \rho_s = (A1+A2+B1+B2) \% \times \lambda \times 11 \times \overline{X_s}; \quad (4.2)$$

$$\text{I-CSCF: } \rho_i = (B1+B2) \% \times \lambda \times 12 \times \overline{X_i}; \quad (4.3)$$

$$\text{HSS: } \rho_h = (B1+B2) \% \times \lambda \times \overline{X_h}; \quad (4.4)$$

Se considera que los tiempos de servicio promedio son iguales, y que la utilización más grande de los servidores P-CSCF y S-CSCF en la red depende de los valores de (B1 + B2) y (C1 + C2). Cuando el tráfico hacia y desde los subscriptores itinerantes es más pesado, entonces el S-CSCF se convierte en el cuello de botella, por otro lado, si el tráfico desde y hacia los visitantes es más pesado, entonces P-CSCF será el cuello de botella. (El programa de cálculo del cuello de botella en la red y el tiempo de espera de cada servidor se puede encontrar en el apéndice C). En base a este análisis, se puede observar cómo mejorar el rendimiento de la red, para esto, se analiza el caso particular de una red rural con una carga de tráfico normal de $\lambda = 10$ [sesiones/s], con un tiempo de servicio promedio de los servidores P-CSCF, S-CSCF e I-CSCF iguales a $\overline{X_p} = \overline{X_s} = \overline{X_i} = \overline{X_h} = 0.01s$. Se supone que la distribución de tráfico en la red queda como lo muestra la tabla 4.11.

Con la distribución de tráfico propuesta en los diferentes escenarios, la utilización de cada entidad SIP se calcula de la siguiente manera:

$$\text{P-CSCF: } \rho_p = (10+10+15+15) \% \times \lambda \times 11 \times \overline{X_p} = 0.55;$$

$$\text{S-CSCF: } \rho_s = (10+10+25+25) \% \times \lambda \times 11 \times \overline{X_s} = \mathbf{0.77};$$

$$\text{I-CSCF: } \rho_i = (10+25) \% \times \lambda \times 12 \times \overline{X_i} = 0.42;$$

$$\text{HSS: } \rho_h = (10+25) \% \times \lambda \times \overline{X_h} = 0.035;$$

Los cálculos muestran que bajo estas condiciones el S-CSCF actúa como cuello de botella de la red. Una vez que el administrador de la red detecta esta falla, procede a corregirla multiplicando la carga de tráfico especificada mediante un factor α_1 con el fin de balancearla, buscando con esto obtener la carga de tráfico de la red máxima. Esto se obtiene con la ecuación (4.2) que permite conocer la utilización del servidor S-CSCF alcanzando el valor de 1. Así de esta manera, probando con diferentes valores de α_1 hasta encontrar el valor de $\alpha_1 = 1.3$, se obtiene que $\rho_s = (10+10+25+25) \% \times \lambda \times \alpha_1 \times 11 \times \overline{X_s} = 1$, con esta operación se observa que la capacidad de la red máxima se

Tabla 4.11. Distribución de tráfico propuesta para diferentes escenarios.

1a	1b	2a	2b	3a	3b
10%	10%	25%	25%	15%	15%

incrementa a 13 [llamadas/ segundo]. Para lograr esto el administrador agregará otro servidor S-CSCF a la red, logrando con esto balancear la red entre los dos S-CSCF, bajo estas circunstancias el tiempo promedio de servicio del servidor S-CSCF se reduce a la mitad del tiempo de servicio original, es decir, $\bar{X}_s = \bar{X}_s/2$, quedando la utilización de cada nodo de la siguiente manera:

$$\text{P-CSCF: } \rho_p = (10+10+15+15) \% \times \lambda \times 11 \times \bar{X}_p = 0.55;$$

$$\text{S-CSCF: } \rho_s = (10+10+25+25) \% \times \lambda \times 11 \times \bar{X}_s = \mathbf{0.385};$$

$$\text{I-CSCF: } \rho_i = (10+25) \% \times \lambda \times 12 \times \bar{X}_i = 0.42;$$

$$\text{HSS: } \rho_h = (10+25) \% \times \lambda \times \bar{X}_h = 0.035;$$

Analizando las ecuaciones anteriores se observa que el rendimiento de la red ahora lo determina el servidor P-CSCF en lugar del S-CSCF. Entonces el administrador de la red procederá al balanceo para incrementar el rendimiento de la misma. Para esto, multiplicará la carga del tráfico en la ecuación ρ_p por un factor α_2 . Así con esto obtendrá la capacidad máxima de la red, cuando la utilización del servidor P-CSCF alcance el valor de 1. Con $\alpha_2 = 1.82$, y aplicando la ecuación (4.1), $\rho_p = (10 + 10 + 15 + 15) \% \lambda \times \alpha_2 \times 11 \times \bar{X}_p = 1$, por lo tanto la capacidad de la red se incrementó a 18.2 [llamadas/s]. Así que el rendimiento adicional es igual a $(18.2-13) = 5.2$ [llamadas/segundo]. Y la capacidad total de la red se incrementa en $(5.2/13) \times 100\% = 40\%$. Esto significa que la red ahora tiene una capacidad de 40% más debido al rendimiento obtenido al colocar dos servidores adicionales a los existentes.

Con el fin de verificar los beneficios del cálculo de la capacidad de tráfico de la red realizaremos este ejercicio; suponga que se conoce el costo de un servidor S-CSCF, se calculará cuántos ingresos adicionales se obtendrían añadiendo un servidor en el nodo que está actuando como cuello de botella. Es decir, si el precio de un S-CSCF es equivalente al 5% de los ingresos obtenidos, entonces el aumento de los ingresos será $(40\% - 5\%) = 35\%$. Se concluye que con este método los concesionarios pueden detectar anomalías en su red y proceder a repararlas sin grandes inversiones de capital.

5. Conclusiones

5.1 Introducción

En este trabajo se presentó un modelo de red de siguiente generación enfocado a comunidades rurales en México para proveerles servicios de telecomunicaciones multimedia. El modelo calcula diferentes parámetros, tales como el ancho de banda, calidad de servicio, retardos por saturación de servidores, etc., los cuales son generados durante el proceso del establecimiento de las llamadas telefónicas, provenientes de diferentes comunidades rurales.

Los resultados obtenidos en los cálculos de esta investigación muestran que el parámetro ancho de banda, es el más importante a considerar, ya que de este parámetro depende la calidad de servicio que disfrutarán los usuarios de la red. Se puede observar de los datos obtenidos, que cuando se opera la red con cinco por ciento de pérdidas, y una tasa máxima de tráfico de cien sesiones por segundo, el ancho de banda máximo generado por el subsistema multimedia será de la mitad de un mega byte, operando llamadas locales. Sin embargo, cuando la red opera llamadas itinerantes, el ancho de banda generado por el IMS, se incrementa y será menor a dos megabytes, en ambos casos los anchos de banda generados son técnicamente razonables y sencillos de obtener. Este resultado es importante para el operador de la red, ya que no requerirá grandes inversiones de capital para proveer el servicio multimedia a las comunidades rurales mencionadas en este documento, debido a que tan solo requerirá de conectar en la central telefónica un servidor *Proxy*, un *Serving*, un *Interrogating* y un servidor subscriber de usuarios como entidades del IMS.

También se observa del análisis de resultados del ancho de banda, que éste siempre se verá afectado por la tasa de tráfico, en forma lineal y proporcional, resultado favorable para los operadores de redes, ya que de esta manera podrán predecir el comportamiento de la red, cuando el tráfico aumenta súbitamente, y tienda a saturar los servidores. Por otro lado, al incrementarse el porcentaje de pérdidas de la red de cinco a veinte por ciento; el ancho de banda se triplica, esta información resulta útil para el diseñador de redes.

Se observa del análisis, que la eficiencia de la red está determinada principalmente por los porcentajes de pérdidas, los cuales deben ser menores a cinco por ciento para una red funcional, información útil para el diseñador de redes.

Se identifica de los resultados, que el modelo matemático utilizado para visualizar las diferentes rutas de los paquetes, durante el proceso de establecer llamadas, ofrece información al administrador para identificar el comportamiento de las rutas que siguen los mensajes SIP a través del subsistema multimedia, y conocer el número de veces que dichos mensajes tocan a los servidores. Identificando las posibles rutas, que siguen los mensajes durante las puestas de sesiones, y cuanto tiempo promedio emplean los mensajes al pasar a través de los servidores. Esta información es importante para valorar los retardos y variaciones de retardo que afectan la calidad de servicio.

Los resultados también muestran que si el tráfico entre suscriptores y visitantes (sin itinerancia) de la primera red, se incrementa, entonces el servidor S-CSCF1 se convertirá en el cuello de botella de dicha red. Sin embargo, si el tráfico (sin itinerancia) entre suscriptores de la segunda red y visitantes, se hace más pesado, entonces el servidor

P-CSCF2 será el cuello de botella de la segunda red. Y en el caso que se aumente la carga de tráfico entre suscriptores de la primera red (con y sin itinerancia), el servidor S-CSCF1 actuará como cuello de botella de la red mencionada. Esta información le permite al operador visualizar el rendimiento de la red, para mantenerla balanceada en casos de saturación.

Se concluye también que debido a que existen diversas posibilidades de distribución de tráfico, en teoría cualquier nodo puede actuar como cuello de botella en determinada situación. Sin embargo, el administrador de la red tiene capacidad de identificar los servidores saturados y substituirlos por otros de mayor capacidad, pudiendo con esto, incrementar la capacidad de la red, hasta cuarenta por ciento más de su capacidad básica inicial instalada.

Otra conclusión importante es que a pesar de que la arquitectura de ésta red rural NGN está diseñada para dar servicio a redes inalámbricas fijas y móviles, también cuenta con interfaces abiertas para reconocer y conectar diferentes tecnologías alámbricas (legacy), sin comprometer la QoS. Esto es posible debido a que el diseño del modelo está basado en las recomendaciones internacionales de la UIT, 3GPP, 3GPP2, ITSE, IETF, e IEEE.

Finalmente se concluye, que los resultados obtenidos cumplen con el objetivo de la investigación realizada, la cual consistió en diseñar una arquitectura de red de siguiente generación capaz de proveer servicios multimedia a las comunidades rurales de México, y contribuye a la planeación de despliegue de infraestructura en telecomunicaciones en zonas rurales, en el marco del proyecto México Conectado, al considerar las densidades de cada población y los efectos en diferentes escenarios, ya que permite planificar su impacto en el desarrollo sustentable de éstas comunidades.

5.2 Contribuciones de la Tesis

Las contribuciones de esta tesis se centran en las redes de siguiente generación, aplicadas a las comunidades rurales y urbanas marginadas de México, particularmente, la creación de un modelo de red VoIP basado en el subsistema multimedia, con arquitectura abierta para interconectar a las diferentes redes de acceso inalámbricas de tercera y cuarta generación, así como también a las redes alámbricas (*legacy*), para conectarse a la red de telecomunicaciones de banda ancha del sistema México Conectado, implementado por la Secretaria de Comunicaciones y Transportes del Gobierno Federal Mexicano. Esta tesis aporta el modelo de red de última generación, a partir del cual se puedan crear redes adaptables en tamaño y capacidad a diferentes comunidades rurales y urbanas marginadas de México. El modelo está diseñado en cuatro capas o estratos (acceso, transporte, control y aplicación) de acuerdo a las especificaciones internacionales de redes NGN, e incluye interfaces con una arquitectura abierta para conectar las diferentes tecnologías de acceso (alámbrica e inalámbricas) en forma transparente. Se propuso un modelo teórico de red NGN diseñado para atender las comunidades rurales mencionadas, con capacidad de interconectar diferentes tecnologías de acceso a la central telefónica IMS-AGC. La arquitectura propuesta incluye una red troncal EPC para interconectar a las tecnologías de acceso GERAN, UTRAN, E-UTRAN de 3GPP, y las redes de acceso (WiMAX, WLAN) de la IEEE, de acuerdo a las recomendaciones de ETSI, IETF, 3GPP, 3GPP2, IEEE y UIT. También contempla capacidad de interconexión de redes alámbricas (xDSL, RTPC, ISDN,

etc.) conectadas directamente al subsistema multimedia (IMS). Permite la interconexión a través de Internet a las diferentes centrales telefónicas (IMS-AGC) instaladas en las zonas de mayor densidad de usuarios.

Se seleccionó y aplico un modelo matemático funcional para validar la arquitectura de la red NGN propuesta, el cual está basado en el protocolo SIP, y permite el cálculo del ancho de banda generado durante el establecimiento de llamadas telefónicas, realizadas por las diferentes comunidades rurales. El modelo de red está diseñado para operar un tráfico de 10 [sesiones/s] en horas normales de trabajo, 20 [sesiones/s] en horas pico, así como 100 [sesiones/s] en casos de emergencia, con porcentajes de pérdida del 0.05, 0.1, 0.15 y 0.20 por ciento y capacidad para atender simultáneamente llamadas telefónicas locales e itinerantes.

La central telefónica IMS-AGC es el corazón de la red y está basada en las recomendaciones de 3GPP; así mismo dicha central está constituida de la red de transporte, el subsistema Multimedia, los servidores de aplicación (AS) y las interfaces a las diferentes redes de acceso mencionadas. Su arquitectura incluye cuatro niveles o capas de red, los cuales son: red de acceso, red de transporte o troncal, red de gestión de control, y servidores de aplicación. La central IMS-AGC se ajusta a su modelo matemático para cálculo de rutas de tráfico, con el cual se identifican a los servidores SIP que producen retardos por saturación en los nodos de la red, debido a los cuellos de botella. Con este modelo se localizan dichos servidores y se sustituyen por otros de mayor capacidad. Así mismo se realizaron múltiples experimentos de simulación numérica, para estimar el ancho de banda que requieren las redes que atienden el tráfico de llamadas, provenientes de comunidades rurales de tamaños y capacidades diferentes. El modelo propuesto permite conectar desde teléfonos alámbricos análogos, hasta teléfonos digitales inteligentes inalámbricos móviles de siguiente generación, las cuales utilizan múltiples transmisores y receptores (MIMO por sus siglas en Inglés) para su operación.

5.2.1 Publicación de Resultados

Como resultado de esta tesis, se publicaron los siguientes artículos:

IEEE 2011 Electronics, Robotics and Automotive Mechanics Conference.

- Colin, A. ; Serrano, A. ; Alvarez-Guzman, E. “Exploring a Framework for the Application of the Next Generation Networks (NGN) in the Mexican Rural Context” Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2011 IEEE.
DOI: 10.1109/CERMA.2011.68
Año de publicación: 2011, Página(s): 371 - 376

IEEE Conference Publications

- Colin, A. ; Serrano, A. ; Alvarez-Guzman, E. “Next Generation Networks (NGN) in the Mexican Rural Context”
Publicado en:

Computing Technology and Information Management (ICCM), 2012 8th
International Conference on (Volume: 2)
Fecha de la Conferencia: 24-26 Abril 2012
Páginas(s): 692 - 695
Impresión ISBN: 978-1-4673-0893-9
Lugar de la conferencia: Seoul, Korea (South)
Publisher: IEEE

Revista IEEE Latin America Transactions

- Colín, A. ; Serrano, A. ; Alvarez-Guzman, E. , Modelo NGN para el Contexto Rural de Mexico Aplicado al Sistema e-Mexico. (NGN Model for the Mexican Rural Context Applied to the e-México System). Publicado en:
IEEE Latin American Magazine, 2015
Volume: 13, Issue: 1
Fecha de la Publicación: enero 2015
Páginas(s):
Impresión ISBN:
Publisher: IEEE

5.3 Trabajo Futuro

En cuanto a trabajos futuros, si bien los resultados de las simulaciones son muy prometedores en cuanto a las posibilidades de proveer servicios multimedia a contextos rurales, se contempla analizar el modelo considerando otros factores de impacto. Es decir, analizando más a fondo el uso de un modelo más complejo con pérdida, considerando escenarios de tráfico de mayor densidad, e investigando los efectos de los servidores sobrecargados. También se sugiere utilizar un modelo que pueda soportar un protocolo de transporte TCP; con sus mecanismos de retransmisión propios; calculando el número de mensajes enviados y los retardos que pudieran afectar la Calidad de Servicio de la red. También resulta importante analizar los retardos producidos durante el proceso del establecimiento de la llamada con tasas de tráfico mayores.

Otro aspecto a considerar para trabajos futuros es que SIP continuará aumentando su complejidad, en cuanto al tamaño de los mensajes, dando lugar a la posible fragmentación de los mismos. Considerando que SIP no solo realiza llamadas a destinos individuales, sino que SIP soporta llamadas bifurcadas. Es decir, permite a los proxy transmitir una solicitud a diferentes destinos y esperar a sus respuestas. Todo esto agregaría complejidad adicional para poder determinar las pérdidas y los retardos, y evaluar la calidad de servicio en tales condiciones.

REFERENCIAS

- [1] M. Aspe (2014, Feb 21). México Conectado (1st Ed.) [Online]. Available:<http://www.mexicoconectado.gob.mx>
- [2] 3GPP, Technical Specification Group Services and System Aspects, IP Multimedia Subsystem (IMS) - Stage 2 (Release 5), TS 23.228 v5.6.0, 2002-09.
- [3] Telecommunication Standardization Sector of ITU. "Y.2000. Y series: Global information infrastructure, Internet protocol aspects and next-generation networks. Y.2000-Y.2999: Next Generation Networks. International Telecommunication Union. 12/2004.
- [4] Telecommunication Standardization Sector of ITU. "Y.2001, Next Generation Network- Frameworks and functional architecture models. General" "overview of NGN". International Telecommunication Union. 12/2004.
- [5] Informe final del Grupo Temático 7 del UIT-D. "Nuevas tecnologías para aplicaciones rurales" Unión Internacional de Telecomunicaciones, septiembre 2000, pp. 6-7.
- [6] "Wireless Communications Beyond 3G" Alcatel Telecommunications Review 1st Quarter, 2001.
- [7] Narayan Parameshwar, Lachu Aravamudhan and Chris Reece, Advanced SIP Series: SIP and 3GPP Operations, Award Solutions, Inc.
- [8] P. Sehier, J-M.Gabriagues, A. Urie "Standardization of 3G mobile systems" Alcatel Telecommunications Review 1st Quarter 2001.
- [9] Jeikki Kaaranen, Ari Ahtiainen, etc., UMTS Networks: Architecture, Mobility and Services John Wiley & Sons, Ltd 2001.
- [10] The Third Generation of Mobile Services (AU-SYSTEM white paper).
- [11] 3GPP TS 23.228: "IP Multimedia Subsystem (IMS); Stage 2".
- [12] Antti Toskala and Harri Holma, WCDMA for UMTS John Wiley & Sons, Ltd 2002.
- [13] 3GPP 23.002, "Network architecture".
- [14] 3GPP 36.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); overall description; Stage 2".
- [15] 3GPP 36.401, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Architecture description".
- [16] 3GPP TS 36.410, "S1 General Aspects and Principles".
- [17] 3GPP TS 36.420, "X2 General Aspects and Principles".
- [18] 3GPP TS 25. 401, "UTRAN overall description".
- [19] Bosch, H.G.P. et al., "Flat Cellular (UMTS) Networks", Wireless Communications and Networking Conference (WCNC), marzo 2007, Kowloon, China.
- [20] Rajeev Agrawal and Anand Bedekar, "Network Architectures for 4G: Cost Considerations", IEEE Communications Magazine, diciembre 2007.

- [21] 3GPP TS 23.401, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”.
- [22] 3GPP TS 23.402, “Architecture enhancements for non-3GPP accesses”.
- [23] 3GPP TS 23.203, “Policy and charging control architecture”.
- [24] 3GPP TS 32.240, “Telecommunication management; Charging management; Charging architecture and principles”.
- [25] “Mobile Inter-networking with IPv6: Concepts, Principles and Practices”, Rajeev S. Koodli y Charles E. Perkins, Editorial Wiley-Interscience, 2007.
- [26] 3GPP TS 23.008, “Organization of subscriber data”.
- [27] “UMTS Networks: Architecture, Mobility and Services”, H. Kaaranen et al., Editorial John Wiley & Son, segunda edición 2005.
- [28] 3GPP TS 29.274, “3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3”.
- [29] 3GPP TS 29.272, “Evolved Packet System (EPS); Mobility Management Entity (MME) and Serving GPRS Support Node (SGSN) related interfaces based on Diameter protocol”.
- [30] “Parlay/OSA: From Standards to Reality”, Musa Unmehopa, Kumar Vemuri, Andy Bennett, Editorial John Wiley & Sons, 2006.
- [31] 3GPP TS 23.228, “IP Multimedia Subsystem (IMS); Stage 2”.
- [32] The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds, Gonzalo Camarillo y Miguel Ángel García, Editorial Wiley, Tercera Edición 2008.
- [33] 3GPP TS 29.278, “Customized Applications for Mobile network Enhanced Logic (CAMEL); CAMEL Application Part (CAP) specification for IP Multimedia Subsystems (IMS)”.
- [34] ZHU, BEI. Analysis of SIP in UMTS IP Multimedia Subsystem, Computer Engineering of North Carolina State University; Bottleneck Analysis of IP Multimedia Subsystem; pp 56-66, 2003.
- [35] CDMA Development Group, “CDMA Migration Path to LTE”, presentación disponible online en www.cdg.org.
- [36] IETF RFC 3748, “Extensible Authentication Protocol (EAP)”.
- [37] IETF RFC 3588, “Diameter Base Protocol”.
- [38] IETF RFC 2661, “Layer Two Tunneling Protocol “L2TP”.
- [39] IETF RFC 4301, “Security Architecture for the Internet Protocol”.
- [40] IETF RFC 4303, “IP Encapsulating Security Payload (ESP)”.
- [41] 3GPP TS 25. 401 “UTRAN overall description”.
- [42] 3GPP TS 43.051, “GSM/EDGE Radio Access Network; Overall description - Stage 2”.

- [43] 3GPP TS 23.060, "General Packet Radio Service (GPRS); Service description; Stage 2".
- [44] 3GPP TS 29.060, "GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface".
- [45] 3GPP TS 29.274 "3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3.
- [46] 3GPP TS 23.401, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access".
- [47] 3GPP TS 22.173, "IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1".
- [48] 3GPP TS 23.206, "Voice Call Continuity (VCC) between Circuit Switched (CS) and IP Multimedia Subsystem (IMS); Stage 2".
- [49] 3GPP TS 23.216, "Single Radio Voice Call Continuity (SRVCC); Stage 2 (Release 8)".
- [50] 3GPP TS 23.402, "Architecture enhancements for non-3GPP accesses".
- [51] WiMAX Network Forum Architecture, Stage 2, 3GPP-WiMAX interworking, Release 1, version 1.2, January 2008.
- [52] 3GPP, Technical Specification Group Services and System Aspects, IP Multimedia (IM) Subsystem IMS; Stage 2 (Release 9). TS 23.228, v9.4.0 (2010-2009).
- [53] 3GPP, Technical Specification Group Services and System Aspects, Network Architecture (Release 12), TS 23.002, v12.2.0, 2013-06.
- [54] 3GPP, Technical Specification Group Services and System Aspects, IP Multimedia Core Network System (IMS), Multimedia Telephony Service and Supplementary Services; Stage 1 (Release 12). TS 22.173, v14.4.0 (2013-06).
- [55] ITU-T, Recommendation G.107, "The E-model, a Computational Model for use in Transmission Planning", (12/2011).
- [56] J. Rosenberg , H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session Initiation Protocol, RFC3261. IETF, June 2002.
- [57] 3GPP, Technical specification group core network and terminals, IP Multimedia Subsystem (IMS) Sh interface; Signaling flows and message contents. TS 29.328, 2008.
- [58] Sisalem, D.; Floroiu, J.; Kuthan, J.; Abend, U.; H. Schulzrinne, " SIP Security ", 2009.
- [59] Sisalem, D.; Liisberg, M.; Rebahi, Y., "A Theoretical Model of the Effects of Losses and Delays on the Performance of SIP", Globecom 2008, New Orleans, December 2008.
- [60] J. Rosenberg, H. Schulzrinne , G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol." Internet Engineering Task Force, RFC 3261, June 2002.

- [61] Intel and Open Cloud performed the SIP Application Server tests in May 2006.
- [62] V. Kueh, R. Tafazolli, B. Evans, "Performance evaluation of SIP-based session establishment over satellite-UMTS." Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual, Vol: 2, 22-25 Apr 2003, pp: 1381 – 1385.
- [63] V. Kueh, R. Tafazolli, B. Evans, "Performance evaluation of SIP-based session establishment over satellite-UMTS." Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual, Vol: 2, 22-25 Apr 2003, pp: 1381 – 1385.
- [64] H. Fathi, S. S. Chakraborty, R. Prasad, "On SIP session setup delay for VoIP services over correlated fading channels," IEEE Transactions on Vehicular Technology, vol. 55, no 1, pp. 286–295, January 2006.

APENDICE A

Tabla valores de timers proceso SIP, recomendación RFC3261

Rosenberg, et. al.

Standards Track

[Page 264]

RFC 3261

SIP: Session Initiation Protocol

June 2002

A Table of Timer Values

Table summarizes the meaning and defaults of the various timers used by this specification.

Timer	Value	Section	Meaning
T1	500ms default	Section 17.1.1.1	RTT Estimate
T2	4s	Section 17.1.2.2	The maximum retransmit interval for non-INVITE requests and INVITE responses
T4	5s	Section 17.1.2.2	Maximum duration a message will remain in the network
Timer A	initially T1	Section 17.1.1.2	INVITE request retransmit interval, for UDP only
Timer B	64*T1	Section 17.1.1.2	INVITE transaction timeout timer
Timer C	> 3min	Section 16.6 bullet 11	proxy INVITE transaction timeout
Timer D	> 32s for UDP 0s for TCP/SCTP	Section 17.1.1.2	Wait time for response retransmits
Timer E	initially T1	Section 17.1.2.2	non-INVITE request retransmit interval, UDP only
Timer F	64*T1	Section 17.1.2.2	non-INVITE transaction timeout timer
Timer G	initially T1	Section 17.2.1	INVITE response retransmit interval
Timer H	64*T1	Section 17.2.1	Wait time for ACK receipt
Timer I	T4 for UDP 0s for TCP/SCTP	Section 17.2.1	Wait time for ACK retransmits
Timer J	64*T1 for UDP 0s for TCP/SCTP	Section 17.2.2	Wait time for non-INVITE request retransmits
Timer K	T4 for UDP 0s for TCP/SCTP	Section 17.1.2.2	Wait time for response retransmits

APENDICE B

Modelo matemático para calcular servidores (CSCF) saturados del IMS

El dimensionamiento de la red dependerá de varios parámetros: tasa de llamadas por segundo (carga de tráfico), porcentaje de pérdida de paquetes debido a sobre carga de servidores, retardos debido a las pérdidas de paquetes y variaciones de retardo, entre otros. Debido a que el modelo será de siguiente generación, deberá operar los servicios de voz, datos y video en protocolo IP (arquitectura horizontal de paquetes). Considerando estos elementos y parámetros involucrados 3GPP y 3GPP2 han diseñado un subsistema multimedia (IMS) basado en el protocolo IP como parte principal de la capa de control de la red. Así mismo se ha escogido al Protocolo de Iniciación de Sesión (SIP) para que se encargue del establecimiento y control de llamadas en el IMS. Los diferentes escenarios de enrutamiento del tráfico de datos, y la tasa de llamadas en horas pico, generan retardos que pueden afectar la calidad del servicio (QoS) que perciben los usuarios en sus llamadas telefónicas. Con el fin de prevenir estos inconvenientes se propone un modelo de red NGN que permita medir el ancho de banda estimado que se generará cuando los usuarios adscritos a la red realicen llamadas telefónicas y/o envío de tráfico multimedia, para que de esta manera la operación de la red no se vea afectada por saturación debido al tráfico. Con el modelo de red propuesto se proveerá de un método para calcular el ancho de banda estimado que se requiere generar por el IMS, durante el proceso de las puestas de las sesiones, balancear la red en caso de saturación de los servidores SIP, medir los retardos y el número de mensajes SIP requeridos para establecer y liberar una llamada en la red, tomando en cuenta las pérdidas de la red. Cuando se solicita el establecimiento de una llamada, el tráfico de señalización pasa varias veces a través diferentes entidades funcionales involucradas. Bajo este principio, cada entidad (servidores SIP-CSCF's) en la cadena de encaminamiento es tocada por el flujo de tráfico de señalización varias veces. Cuando las solicitudes SIP cruzan los servidores CSCF (o HSS) estos proveen algún servicio y luego reenvían el tráfico. Con el propósito de administrar adecuadamente la red, cada servidor SIP tiene una cola formada por diferentes flujos de tráfico en espera de ser enviados. Mediante este análisis se pueden identificar los diferentes servidores que están actuando como cuellos de botella en la red. Normalmente, durante el establecimiento y liberación de una llamada SIP los servidores P-CSCF y S-CSCF son tocados 11 veces, el I-CSCF 12 veces y el HSS es tocado una vez. Para el análisis de flujo del tráfico se analizan dos redes diferentes con operadores diferentes, ambas redes cuentan por lo menos con un servidor (P-CSCF, S-CSCF y un I-CSCF) por red, y el modo de servicio que se brinda al usuario será: el primero que llega es el primero que se atiende. El tiempo de servicio promedio de P-CSCF es \bar{X}_p , el tiempo de servicio de S-CSCF es \bar{X}_s , el tiempo medio de servicio de la I-CSCF es \bar{X}_i , y el tiempo medio de servicio de HSS es \bar{X}_h ; en base a estas consideraciones se conocerá que servidor será el cuello de botella utilizando diferentes escenarios de rutas de flujo. La figura 2.25, muestra todos los escenarios posibles de conexión entre dos redes diferentes. Considérese lo siguiente:

- H1: Subscriptor de la Red 1 (sin itinerancia)
- H2: Subscriptor de la Red 2 (sin itinerancia)
- V1: Subscriptor de la Red 2 con itinerancia en la Red 1
- V2: Subscriptor de la Red 1 con itinerancia en la Red 2

Tabla B1. Escenarios posibles de llamadas telefónicas entre 2 dos redes diferentes

	Escenario	Descripción
1	$H1 \rightarrow H2$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 2 (sin itinerancia)
2	$H2 \rightarrow H1$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 1 (sin itinerancia) llama a
3	$V1 \rightarrow V2$	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 1 con itinerancia en Red 2
4	$V2 \rightarrow V1$	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 2 con itinerancia en Red 1
5	$H1 \rightarrow V2$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 1 con itinerancia en Red 2
6	$V2 \rightarrow H1$	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 1 (sin itinerancia)
7	$V1 \rightarrow H2$	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 2 (sin itinerancia)
8	$H2 \rightarrow V1$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 2 con itinerancia en Red 1
9	$H1 \rightarrow V1$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 2 con itinerancia en Red 1
10	$V1 \rightarrow H1$	Subscriber de la Red 2 con itinerancia en Red 1 llama a Subscriber de la Red 1 (sin itinerancia)
11	$H2 \rightarrow V2$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 1 con itinerancia en Red 2
12	$V2 \rightarrow H2$	Subscriber de la Red 1 con itinerancia en Red 2 llama a Subscriber de la Red 2 (sin itinerancia) llama a
13	$H1 \rightarrow H1$	Subscriber de la Red 1 (sin itinerancia) llama a Subscriber de la Red 1 (sin itinerancia)
14	$H2 \rightarrow H2$	Subscriber de la Red 2 (sin itinerancia) llama a Subscriber de la Red 2 (sin itinerancia)

Tabla B2. Cadena de ruteo del flujo de tráfico de señalización

	Escenario	Cadena de Ruteo
1	$H1 \rightarrow H2$	P-CSCF1↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2
2	$H2 \rightarrow H1$	P-CSCF2↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1
3	$V1 \rightarrow V2$	P-CSCF1↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF2
4	$V2 \rightarrow V1$	P-CSCF2↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF1
5	$H1 \rightarrow V2$	P-CSCF1↔S-CSCF1↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF2
6	$V2 \rightarrow H1$	P-CSCF2↔S-CSCF1↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1

7	$V1 \rightarrow H2$	P-CSCF1↔S-CSCF2↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2
8	$H2 \rightarrow V1$	P-CSCF2↔S-CSCF2↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF1
9	$H1 \rightarrow V1$	P-CSCF1↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF1
10	$V1 \rightarrow H1$	P-CSCF1↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1
11	$H2 \rightarrow V2$	P-CSCF2↔S-CSCF2↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF2
12	$V2 \rightarrow H2$	P-CSCF2↔S-CSCF1↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2
13	$H1 \rightarrow H1$	P-CSCF1↔S-CSCF1↔I-CSCF1↔HSS1↔S-CSCF1↔P-CSCF1
14	$H2 \rightarrow H2$	P-CSCF2↔S-CSCF2↔I-CSCF2↔HSS2↔S-CSCF2↔P-CSCF2

Tenemos que por cada procedimiento para colocar una llamada y liberarla, los servidores P-CSCF y S-CSCF involucrados son tocados 11 veces, y el servidor I-CSCF es tocado 12 veces; Suponga que la carga total de tráfico es λ [llamadas/s], la carga de tráfico para cada escenario es λ_i , y $\lambda = \sum \lambda_i$. Entonces de acuerdo a la tabla B2, se tiene que el total de veces que es tocado cada nodo quedaría:

$$\begin{aligned}
\text{P-CSCF1: } \Lambda_1 &= (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8) \times 11 + (\lambda_9 + \lambda_{10} + \lambda_{13}) \times 22; \\
\text{P-CSCF2: } \Lambda_2 &= (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8) \times 11 + (\lambda_{11} + \lambda_{12} + \lambda_{14}) \times 22; \\
\text{S-CSCF1: } \Lambda_3 &= (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_9 + \lambda_{10} + \lambda_{11} + \lambda_{12}) \times 11 + (\lambda_5 + \lambda_6 + \lambda_{13}) \times 22; \\
\text{S-CSCF2: } \Lambda_4 &= (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_9 + \lambda_{10} + \lambda_{11} + \lambda_{12}) \times 11 + (\lambda_7 + \lambda_8 + \lambda_{14}) \times 22; \\
\text{I-CSCF1: } \Lambda_5 &= (\lambda_2 + \lambda_3 + \lambda_5 + \lambda_6 + \lambda_{10} + \lambda_{11} + \lambda_{13}) \times 12; \\
\text{I-CSCF2: } \Lambda_6 &= (\lambda_1 + \lambda_4 + \lambda_5 + \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{12} + \lambda_{14}) \times 12; \\
\text{HSS1: } \Lambda_7 &= (\lambda_2 + \lambda_3 + \lambda_5 + \lambda_6 + \lambda_{10} + \lambda_{11} + \lambda_{13}); \\
\text{HSS2: } \Lambda_8 &= (\lambda_1 + \lambda_4 + \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{12} + \lambda_{14});
\end{aligned}$$

Calculando la utilización que se hace de cada entidad o servidor sería:

$$\begin{aligned}
\text{P-CSCF1: } \rho_p &= \Lambda_1 \times \overline{X_p} \\
\text{P-CSCF2: } \rho_p &= \Lambda_2 \times \overline{X_p} \\
\text{S-CSCF1: } \rho_s &= \Lambda_3 \times \overline{X_s} \\
\text{S-CSCF2: } \rho_s &= \Lambda_4 \times \overline{X_s} \\
\text{I-CSCF1: } \rho_i &= \Lambda_5 \times \overline{X_i} \\
\text{I-CSCF2: } \rho_i &= \Lambda_6 \times \overline{X_i} \\
\text{HSS1: } \rho_h &= \Lambda_7 \times \overline{X_h} \\
\text{HSS2: } \rho_h &= \Lambda_8 \times \overline{X_h}
\end{aligned}$$

Si asumimos que $\bar{X}_p = \bar{X}_s = \bar{X}_l = \bar{X}_h$, localizaremos el cuello de botella si se conoce la carga de tráfico de cada escenario. El programa de cálculo del cuello de botella en este sistema de dos redes se adjunta como Apéndice C. Veamos algunos ejemplos:

Tabla B3. Distribución de tráfico y embotellamiento

	1	2	3	4	5	6
λ_1/λ	0.2	0.1	0.05	0.05	0.1	0.1
λ_2/λ	0.2	0.1	0.05	0.05	0.1	0.1
λ_3/λ	0.05	0.05	0.05	0.05	0.05	0.05
λ_4/λ	0.05	0.05	0.05	0.05	0.05	0.05
λ_5/λ	0.05	0.05	0.15	0.05	0.15	0.05
λ_6/λ	0.05	0.05	0.15	0.05	0.15	0.05
λ_7/λ	0.05	0.05	0.05	0.15	0.05	0.05
λ_8/λ	0.05	0.05	0.05	0.15	0.05	0.05
λ_9/λ	0.03	0.03	0.05	0.05	0.05	0.15
λ_{10}/λ	0.03	0.03	0.05	0.05	0.05	0.15
λ_{11}/λ	0.02	0.12	0.05	0.05	0.05	0.05
λ_{12}/λ	0.02	0.12	0.05	0.05	0.05	0.05
λ_{13}/λ	0.1	0.1	0.1	0.1	0.05	0.05
λ_{14}/λ	0.1	0.1	0.1	0.1	0.05	0.05
Cuello de botella	P-CSCF1	P-CSCF2	S-CSCF1	S-CSCF2	S-CSCF1	P-CSCF1

La tabla B3, muestra a los servidores que se comportan como cuellos de botella cuando dos redes interactúan entre sí. La saturación de tráfico en los servidores depende de la distribución de dicho tráfico en la red. Es decir, cuando el tráfico entre subscriptores y visitantes de la Red-1 sin itinerancia se hace más pesado, el servidor S-CSCF1 será el cuello de botella (situación 6), sin embargo, si el tráfico entre subscriptores de la Red-2 sin itinerancia y visitantes se hace más pesado, P-CSCF2 se convierte en el cuello de botella (situación 2). Si aumentamos la carga de tráfico entre subscriptores de la Red-1 sin itinerancia y con itinerancia, S-CSCF1 es el cuello de botella (situación 5), etc. Y debido a que existen posibilidades infinitas de distribuciones de tráfico, teóricamente, cada nodo puede ser el cuello de botella en determinada situación. Analizando a la red mostrada en la figura 2.25 se identifica que servidor CSCF del IMS se convierte en el cuello de botella de la red por saturación de tráfico. Con el fin de calcular que servidores SIP se saturan durante la operación de la red, se analizan aquellos que participan en dicha red con diferentes escenarios posibles. La tabla B4 muestra a los diferentes servidores que participan en una red particular.

Tabla B4. Servidores que participan en una red con diferentes escenarios posibles

1	Subscriber localizado en su área de servicio
	c. Subscriber como remitente $\dots \leftrightarrow$ P-CSCF \leftrightarrow S-CSCF $\leftrightarrow \dots$
	d. Subscriber c/ receptor $\dots \leftrightarrow$ I-CSCF \leftrightarrow HSS \leftrightarrow S-CSCF \leftrightarrow P \leftrightarrow CSCF $\leftrightarrow \dots$
2	Subscriber con itinerancia en una red visitada
	c. Subscriber como remitente $\dots \leftrightarrow$ S-CSCF $\leftrightarrow \dots$
	d. Subscriber c/receptor $\dots \leftrightarrow$ I-CSCF \leftrightarrow HSS \leftrightarrow S-CSCF $\leftrightarrow \dots$
3	Visitante no suscrito con itinerancia en su área de servicio
	c. Visitante como remitente $\dots \leftrightarrow$ P-CSCF $\leftrightarrow \dots$
	d. Visitante como receptor $\dots \leftrightarrow$ P-CSCF $\leftrightarrow \dots$

Suponga que la carga de tráfico total en la red es λ [sesiones/s], y que el porcentaje de carga de cada escenario será de acuerdo a la tabla B4. Para obtener la utilización de cada nodo se lograría mediante las siguientes funciones:

$$\text{P-CSCF: } \rho_p = (A1+A2+C1+C2) \% \times \lambda \times 11 \times \overline{X_p};$$

$$\text{S-CSCF: } \rho_s = (A1+A2+B1+B2) \% \times \lambda \times 11 \times \overline{X_s};$$

$$\text{I-CSCF: } \rho_i = (A2+B2) \% \times \lambda \times 12 \times \overline{X_i};$$

$$\text{HSS: } \rho_h = (A2+B2) \% \times \lambda \times \overline{X_h};$$

Suponiendo que los tiempos de servicio promedio sean iguales, entonces si la utilización más grande de los servidores P-CSCF o S-CSCF en una red dada depende de los valores de (B1 + B2) y (C1 + C2). Entonces, si el tráfico hacia y desde los subscribers itinerantes es más pesado, el S-CSCF será el cuello de botella, pero si el tráfico desde y hacia los visitantes es más pesado, entonces el P-CSCF será el cuello de botella. (El programa de cálculo para identificar los servidores saturados en la red y el tiempo de espera de cada servidor se encuentra en el apéndice C). En base a este análisis, se puede ver cómo mejorar la capacidad de la red. Analice el caso particular de una red rural con una carga de tráfico normal de $\lambda = 10$ [llamadas/s], con un tiempo de servicio promedio de los servidores P-CSCF, S-CSCF e I-CSCF iguales a $\overline{X_p} = \overline{X_s} = \overline{X_i} = \overline{X_h} = 0.01s$. Así mismo suponga que la distribución de tráfico en la red queda como lo muestra la tabla 4.11. De acuerdo a la distribución de tráfico propuesto para los diferentes escenarios, la utilización de cada entidad SIP podría calcularse de la siguiente manera:

$$\text{P-CSCF: } \rho_p = (10+10+15+15) \% \times \lambda \times 11 \times \overline{X_p} = 0.55;$$

$$\text{S-CSCF: } \rho_s = (10+10+25+25) \% \times \lambda \times 11 \times \overline{X_s} = \mathbf{0.77};$$

$$\text{I-CSCF: } \rho_i = (10+25) \% \times \lambda \times 12 \times \overline{X_i} = 0.42;$$

$$\text{HSS: } \rho_h = (10+25) \% \times \lambda \times \overline{X_h} = 0.035;$$

Los cálculos anteriores indican que el S-CSCF es el que produce el cuello de botella en el tráfico. Para corregir este problema multiplicaremos la carga de tráfico mediante un factor α_1 , así se obtendrá la máxima carga de tráfico de la red cuando la utilización del servidor S-CSCF alcance el valor de 1. Probando con diferentes valores de α_1 encontramos que cuando $\alpha_1 = 1.3$, $\rho_s = (10+10+25+25) \% \times \lambda \times \alpha_1 \times 11 \times \bar{X}_s = 1$, con esta operación se observa que la capacidad máxima de la red sería de 13 llamadas/ segundo. Si agregamos otro servidor S-CSCF a la red tendremos la carga balanceada entre los dos S-CSCF, bajo estas circunstancias el tiempo promedio de servicio del servidor S-CSCF será la mitad del tiempo de servicio original, es decir, $\bar{X}_s = \bar{X}_s/2$, y la utilización de cada nodo queda de la siguiente manera:

$$\text{P-CSCF: } \rho_p = (10+10+15+15) \% \times \lambda \times 11 \times \bar{X}_p = \mathbf{0.55};$$

$$\text{S-CSCF: } \rho_s = (10+10+25+25) \% \times \lambda \times 11 \times \bar{X}_s = 0.385;$$

$$\text{I-CSCF: } \rho_i = (10+25) \% \times \lambda \times 12 \times \bar{X}_i = 0.42;$$

$$\text{HSS: } \rho_h = (10+25) \% \times \lambda \times \bar{X}_h = 0.035;$$

Ahora se observa del resultado que el rendimiento de la red lo determina el servidor P-CSCF en lugar de S-CSCF. Multiplicando la carga del tráfico por un factor α_2 se logra incrementar el rendimiento. Ahora se logra la capacidad máxima de la red, cuando la utilización del servidor P-CSCF alcance el valor de 1.

Con $\alpha_2 = 1.82$, $\rho_p = (10 + 10 + 25 + 25) \% \lambda \times \alpha_1 \times 11 \times \bar{X}_p = 1$, por lo tanto la capacidad de la red ahora es 18.2 llamadas/s.

Se observa del cálculo que el rendimiento adicional es igual a $(18, 2-13)$ llamadas/s = 5.2 llamadas/s, y la capacidad de la red se incrementa en $(5.2/13) \times 100\% = 40\%$. Esto significa que las llamadas entrantes se incrementan en un 40% debido al rendimiento extra. Conociendo el costo de un servidor S-CSCF, se puede calcular cuántos ingresos adicionales se obtendrán añadiendo un servidor adicional en el sitio donde existe el cuello de botella. Es decir, si el precio de un S-CSCF es equivalente al 5% de los ingresos obtenidos, entonces el aumento de los ingresos será $(40\% - 5\%) = 35\%$. Este pequeño análisis le permitirá al operador dimensionar su red analizando la carga de tráfico total en horas pico y así poder dimensionar el tamaño de la red sin comprometer la calidad del servicio.

APENDICE C

Programa para localizar el cuello de botella en un sistema de dos redes (programado en lenguaje *Java*)

Calcula el nodo que será el cuello de botella basado en la distribución de tráfico.

```
Public static void main(String args[]){
//double d1 to d14 represents the ratio of traffic load of each scenario to
//the total traffic load.
double[] hitCount=new double[9];
double d1=0.1;
double d2=0.1;
double d3=0.05;
double d4=0.05;
double d5=0.15;
double d6=0.15;
double d7=0.05;
double d8=0.05;
double d9=0.05;
double d10=0.05;
double d11=0.05;
double d12=0.05;
double d13=0.05;
double d14=0.05;
//hitCount1 is the total hits on the P-CSCF1 for calls setup and release
hitCount[1]=(d1+d2+d3+d4+d5+d6+d7+d8)*11+(d9+d10+d13)*22;
//hitCount2 is the total hits on the P-CSCF2 for calls setup and release
hitCount[2]=(d1+d2+d3+d4+d5+d6+d7+d8)*11+(d11+d12+d14)*22;
//hitCount3 is the total hits on the S-CSCF1 for calls setup and release
hitCount[3]=(d1+d2+d3+d4+d9+d10+d11+d12)*11+(d5+d6+d13)*22;
//hitCount4 is the total hits on the S-CSCF2 for calls setup and release
hitCount[4]=(d1+d2+d3+d4+d9+d10+d11+d12)*11+(d7+d8+d14)*22;
//hitCount5 is the total hits on the I-CSCF1 for calls setup and release
hitCount[5]=(d2+d3+d5+d6+d10+d11+d13)*12;
//hitCount6 is the total hits on the I-CSCF2 for calls setup and release
hitCount[6]=(d1+d4+d5+d7+d8+d9+d12+d14)*12;
//hitCount7 is the total hits on the HSS1 for calls setup and release
hitCount[7]=d2+d3+d5+d6+d10+d11+d13;
//hitCount8 is the total hits on the HSS2 for calls setup and release
hitCount[8]=d1+d4+d7+d8+d9+d12+d14;
for (int l=1;l<=8;l++) {
System.out.println("hitCount["+l+"]="+hitCount[l]);
}
double temp=0;
int k=0;
```

```
for (int m=1; m<=8; m++) {  
    if (temp<hitCount[m]){  
        temp=hitCount[m];  
        k=m;  
    }  
}  
  
String bottleneck=" ";  
switch (k) {  
    case 1: bottleneck="P-CSCF1"; break;  
    case 2: bottleneck="P-CSCF2"; break;  
    case 3: bottleneck="S-CSCF1"; break;  
    case 4: bottleneck="S-CSCF2"; break;  
    case 5: bottleneck="I-CSCF1"; break;  
    case 6: bottleneck="I-CSCF2"; break;  
    case 7: bottleneck="HSS1"; break;  
    case 8: bottleneck="HSS2"; break;  
}  
System.out.println("Bottleneck is "+bottleneck);  
}  
}
```

ACRÓNIMOS

UIT	Unión Internacional de Telecomunicaciones.
3GPP	Grupo de Socios de la Tercera Generación. <i>Third Generation Partnership Project</i>
3GPP2	Grupo de Socios de la Tercera Generación 2. <i>Third Generation Partnership Project 2</i>
IEEE	Instituto de Ingenieros Eléctricos y Electrónicos. <i>Institute of Electrical and Electronics Engineers</i>
ETSI	Instituto Europeo de Estándares de Telecomunicaciones. <i>European Telecommunications Standards Institute.</i>
ARIB	Asociación Japonesa de Industrias y Empresas de Radiocomunicaciones. <i>Japan's Association of Radio Industries and Businesses.</i>
ATIS	Alianza para la Industria de Telecomunicaciones y Soluciones.
CCSA	Asociación de Normalización de las Comunicaciones de China. <i>China Communications Standardization Association.</i>
TTA	Asociación Tecnológica de las Telecomunicaciones de la República de Corea. <i>Republic of Korea's Telecommunications Technology Association.</i>
TTC	Comité Tecnológico de las Telecomunicaciones de Japón. <i>Japan's Telecommunication Technology Committee.</i>
IETF	Grupo de Trabajo de Ingeniería de Internet. <i>Internet Engineering Task Force.</i>
TIC	Tecnologías de la Información y la Comunicación. <i>Information and communications Technology.</i>
NGN	Redes de Siguiete Generación. <i>Next Generation Network.</i>
UIT-T	UIT- Sección Telecomunicaciones.
Y.2000	Recomendaciones de UIT de Redes de Siguiete Generación.
QoS	Calidad de Servicio. <i>Quality of Service.</i>
2G	Segunda Generación. <i>Second Generation.</i>

3G	Tercera Generación. <i>Third Generation.</i>
4G	Cuarta Generación. <i>Fourth Generation.</i>
4G LTE	Cuarta Generación LTE. <i>Fourth Generation LTE.</i>
CSIC	Coordinación de la Sociedad de la Información y del Conocimiento. México Conectado.
IP	Protocolo Internet. <i>Internet Protocol.</i>
IPv6	IP versión 6.
IPV4	IP versión 4.
RTCP	Red Telefónica Conmutada Pública.
EPC	Red de Paquetes Evolucionada. <i>Evolved Packet Core.</i>
ISDN	Red Digital de Servicios Integrados. <i>Integrated Services Digital Network.</i>
WLAN	Red de Área Local Inalámbrica. <i>Wireless Local Area Network.</i>
WiMAX	Interoperabilidad Global para Acceso Microondas. <i>World Interoperability for Microwave Access.</i>
GSM	Sistema Global para las Comunicaciones Móviles. <i>Global System for Mobile Communications.</i>
GPRS	Servicio General de Paquetes de Radio. <i>General Packet Radio Service.</i>
EDGE	Velocidades de Datos Mejoradas para la Evolución de GSM. <i>Enhanced Data Rates for GSM Evolution.</i>
GERAN (GSM/EDGE)	Red de Acceso de radio GSM/EDGE. Radio Access Network.
UTRAN	Redes de Acceso de Radio Terrestres UMTS. <i>UMTS RAN.</i>
E-UTRAN	Redes UTRAN Evolucionadas. <i>Evolved Universal Terrestrial Access Network.</i>
UMTS	Sistema de Telecomunicaciones Móviles Universales.

	<i>Universal Mobile Telecommunications System.</i>
WCDMA	CDMA de Banda Ancha. <i>Wideband CDMA.</i>
IMT	Telefonía Móvil Internacional. <i>International Mobile Telephony.</i>
SMG	Grupo Especial Móvil de ETSI <i>Special Mobile Group.</i>
IMT-2000	Telefonía Móvil Internacional. <i>International Mobile Telephony 2000.</i>
LTE Advanced	LTE Avanzado. <i>Long Term Evolution Advanced.</i>
IDI	Dispositivos de Interconexión. <i>Interconnection Devices.</i>
CAPEX	Gastos de Capital. <i>Capital Expenditures.</i>
OPEX	Gastos Operativos. <i>Operational Expenditure.</i>
ISSSTE	Instituto de Seguridad y Servicios Sociales para los Trabajadores del Estado.
ENDUTIH	Encuesta Nacional Sobre Disponibilidad y Uso de las Tecnologías de la Información en los Hogares.
IFT Mexicano.	Instituto Federal de Telecomunicaciones del Gobierno Federal
FDMA	Acceso Múltiple por División de Frecuencia. <i>Frequency Division Multiple Access.</i>
TDMA	Acceso Múltiple por División de Tiempo. <i>Time Division Multiple Access.</i>
SIM	Módulo de Identidad del Suscriptor. <i>Subscriber Identity Module.</i>
MAP	Punto de Anclaje de Movilidad. <i>Mobility Anchor Point.</i>
EM	Equipo Móvil. <i>Movil Equipment.</i>

BSS	Subsistema de Estación Base. <i>Base Station Subsystem.</i>
MSC	Centro de Conmutación Móvil. <i>Mobile Switching Center.</i>
NSS	Subsistema de Conmutación de Red. <i>Network switching Subsystem.</i>
SGSN	Nodo de Soporte Servidor de GPRS. <i>Serving GPRS Support Node.</i>
GGSN	Nodo de Soporte Pasarela GPRS. <i>Gateway GPRS Support Node.</i>
BSC	Controlador de Estaciones Base. <i>Base Station Controller.</i>
TRAU	Unidad de Transcodificación y de Adaptación. <i>Transcoding and Rate Adaptation Unit.</i>
GMSC	Pasarela MSC. <i>Gateway MSC.</i>
HLR	Registro de Suscriptores Propios. <i>Home Location Register.</i>
EM	Equipo Móvil.
VLR	Registro de Ubicación de Visitantes. <i>Visitor Location Register.</i>
AUC	Centro de Autenticación. <i>Authentication Center.</i>
EIR	Registro de Identidad de Equipos. <i>Equipment Identity Register.</i>
PDN	Red de Paquete de Datos. <i>Packet Data Network.</i>
CN	Red Principal. <i>Core Network.</i>
EU	Equipo de Usuario. <i>User Equipment.</i>
RNC	Controlador de Red de Radio. <i>Radio Network Controller.</i>

IuCS	Permite conectar a UTRAN con el Dominio de Conmutación de Circuitos.
IuPS	Permite conectar a UTRAN con el dominio de conmutación de Paquetes.
CS-CN	Conmutación de Circuitos-Red Principal. <i>Circuit Switching-Core Network.</i>
MGW	Servidor Pasarela de Multimedia. <i>Media Gateway.</i>
CS	Conmutación de Circuitos. <i>Circuit Switching.</i>
PS	Conmutación de Paquetes. <i>Packet Switching.</i>
PLMN	Red Móvil Pública Terrestre. <i>Public Land Mobile Network.</i>
QoS	Calidad de Servicio. <i>Quality of service.</i>
ATM	Modo de Transferencia Asíncrono. <i>Asynchronous Transfer Mode.</i>
CDMA	Acceso Múltiple por División de Código. <i>Code Division Multiple Access.</i>
CDMA2000	Red Inalámbrica Basada en CDMA. <i>Wireless Network Based in CDMA.</i>
WCDMA	CDMA de Banda Ancha. <i>Wideband CDMA.</i>
IMS	Subsistema Multimedia IP. <i>IP Multimedia Subsystem.</i>
EPC	Red de Paquetes Evolucionada. <i>Evolved Packet Core.</i>
MME	Entidad de Administración de Movilidad del IMS. <i>Mobility Management eNB.</i>
S-GW	Pasarela de Servicio. <i>Serving Gateway.</i>
P-GW	Pasarela de la Red de Datos de Paquetes del IMS.
EPS	Red Troncal de Paquetes Evolucionada.

ADSL	Línea de Abonado Digital Asimétrica. <i>Asymmetric digital subscriber line.</i>
IP	Protocolo de Internet.
X.25	Estándar de UIT para redes de área amplia de conmutación de paquetes.
HSS	Servidor de Suscriptores Propios. <i>Home Subscriber Server.</i>
ERAB	Portadora de Acceso de Radio E-UTRAN. <i>E-UTRAN Radio Access Bearer.</i>
Mobile	
WiMAX	Interoperabilidad mundial para acceso por microondas.
Redes	
802.11	Estándar 'IEEE 802.11' para WLAN.
NodeB	Estación Transceptora Base en GSM.
Interfaz X2	Interfaz que Permite Transferencia de Datos de Usuario entre los eNBs.
Interfaz Uu	Interfaz que Permite la Comunicación entre UTRAN y el Equipo de Usuario.
Interfaz S1	Interfaz que permite Conectar el eNB a la Red Troncal EPC.
Interfaz S1-U	Permite que el eNB se Comunique con Plano de Usuario (S-GW).
S1-MME	Interfaz que Maneja el Plano de Control que Termina en la Entidad MME.
S1-U	Interfaz que Soporta el Plano de Usuario.
RRC	Protocolo de Control de Recursos Radio. <i>Resource Control Protocol.</i>
UDP	Protocolo de Datagramas de Usuario <i>User Datagram Protocol.</i>
UAC	Agente de Usuario Cliente. <i>User Agent Client.</i>
UAS	Servidor de Agente de Usuario. <i>User Agent Server.</i>
CSCF	Función de Control de Sesión de llamada. <i>Call Session Control Function.</i>

P-CSCF	<i>Servidor Proxy SIP (P-CSCF).</i>
S-CSCF	<i>Servidor Serving (S-CSCF).</i>
I-CSCF	<i>Servidor Interrogating (I-CSCF).</i>
VoIP	Voz con IP. <i>Voice IP.</i>
MIMO	Entradas Múltiples Salidas Múltiples. <i>Multiple Input Multiple Output.</i>
ISO	Organización Internacional de Estándares. <i>International Estandar Organization.</i>
MEGACO	Protocolo de Control Pasarela de Medios. <i>Media Gateway Control Protocol.</i>
MGC	Controlador Pasarela de Medios. <i>Media Gateway Controller.</i>
MGCP	Protocolo de Control Pasarela de Medios. <i>Media Gateway Control Protocol.</i>
MPLS	Conmutación de etiquetas multiprotocolo. <i>Multiprotocol Label Switching.</i>
PSTN	Red Telefónica Conmutada Pública. <i>Public Switching Telephone Network.</i>
GPT	Protocolo de tunelización GPRS