Capítulo 3

Aspectos de la red y de los servicios

3. Aspectos de Red y Servicios

Hasta ahora solo hemos visto los aspectos técnicos de la capa física (PHY) para transportar bits por el aire a grandes velocidades, y los aspectos técnicos de la capa de acceso al medio (MAC) para compartir los recursos de radio disponibles entre los múltiples usuarios y los distintos servicios. Estos aspectos son definitivamente los más críticos y que exigen mayor esfuerzo durante el diseño de una red inalámbrica, y de hecho la mayor parte de las especificaciones de la norma IEEE 802.16e y WiMAX tratan de estos aspectos. Pero desde el punto de vista del envío de servicios de banda ancha inalámbricas hasta el usuario final, hay muchos aspectos y desafíos que hay que tener en cuenta. Para ello necesitamos estudiar detalladamente la interfaz aérea e inalámbrica, y ver el sistema inalámbrico de banda ancha desde una perspectiva de un extremo a otro. Necesitamos ver la arquitectura de la red al completo, los protocolos de las capas altas, la interacción con los distintos elementos de red más allá de la estación móvil y la estación base.

El propósito de este capítulo es proporcionar una perspectiva de las redes de banda ancha inalámbricas desde el punto de vista de los servicios y de una red punto a punto. Vamos a ver cuatro secciones: calidad de servicio (QoS), gestión de sesión multimedia, seguridad y gestión de la movilidad.

Ya que WiMAX fue diseñada para soportar primordialmente servicios basados en IP- ya sean datos, vídeo, voz, mensajes o multimedia – vamos a tratar de estudiar parte de los protocolos y arquitectura basadas en IP y cómo son usados en los servicios punto a punto. En un principio IP no fue diseñado para ser eficiente, se diseño para ser un protocolo de máximo esfuerzo (best-effort) y no para soportar servicios que requerían calidad de servicio. La necesidad de soportar servicios multimedia y otros servicios con una determinada calidad de servicio hizo que se desarrollaran nuevas implementaciones en la arquitectura y protocolos IP. En estos progresos también incluimos la optimización de IP sobre un medio inalámbrico poco fiable y de capacidad limitada. Aunque se han logrado bastantes progresos durante los últimos años, la adaptación de IP a los servicios multimedia e inalámbricos continúa siendo una de las áreas de mayor investigación y desarrollo. En este capítulo vemos algunas de ellas.

3.1. Calidad de Servicio

En este punto vamos a ver la calidad de servicio desde una perspectiva de un extremo de la red al otro extremo. ¿Cómo se consigue proporcionar calidad de servicio a las comunicaciones entre dos puntos de una red de paquetes de banda ancha inalámbrica, que además de tener un enlace inalámbrico incluye otros enlaces interconectados mediantes routers, switches, y otros nodos de red? Los enlaces entre los nodos intermedios usarán distintas tecnologías a nivel de enlace como ATM, frame relay, y Ethernet, y cada una tendrá sus propios medios de proporcionar calidad de servicio que no vamos a ver en este apartado. Estudiaremos de forma general como proporcionan calidad de servicio las redes basadas en paquetes y nos centraremos en como proporcionan calidad de servicio las tecnologías que usan IP extremo a extremo, ya que WiMAX está diseñada para proveer servicios IP de extremo a extremo y será implementada usando un núcleo de red basado en IP. Por tanto la calidad de servicio que proporcione IP y su interacción con la capa de enlace inalámbrica son los aspectos más relevantes dentro de una red WiMAX.

La calidad de servicio es una forma de garantizar que un servicio se llevará a cabo con un cierto nivel de calidad de servicio. Este nivel esta especificado normalmente en términos de paquetes perdidos, jitter y caudal, y los requisitos variarán dependiendo de la aplicación y el servicio.

La limitación de recursos en la red es lo que hace que sea un desafío ofrecer garantías a los diferentes usuarios. Cada enlace tiene su propia limitación de ancho de banda, y una memoria limitada para almacenar los paquetes antes de ser reenviados. Una manera ineficiente y cara de proveer calidad es la construcción de la red para satisfacer toda la demanda mediante un aumento del ancho de banda y unos buffers de mayor capacidad, particularmente cuando los requisitos de calidad son muy altos. Por tanto, hay métodos más eficientes de proveer calidad de servicio que tienen en cuenta las necesidades particulares de la aplicación o servicio, y optimiza los recursos usados. Las aplicaciones necesitan diferentes mezclas de recursos, hay servicios como la VoIP que requieren menos latencia y por ello menos memoria y mayor ancho de banda. Por tanto una red que proporcione calidad de servicio debe proveer garantías particulares para varios tipos de aplicaciones y servicios haciendo un uso eficiente de los recursos de la red.

3.1.1. Mecanismos QoS en redes basadas en paquetes

Proveer calidad de servicio de extremo a extremo requiere mecanismos en los planos de control y de datos. Los mecanismos del plano de control son necesarios para permitir al usuario y la red negociar y pactar las especificaciones de calidad de servicio requeridas, para identificar qué usuarios y aplicaciones tienen derecho a qué tipo de calidad de servicio, y para permitir que la red asigne debidamente los recursos a cada servicio. Los mecanismos del plano de datos son necesarios para que se cumplan los requisitos de calidad de servicio mediante el control de la cantidad de recursos de red que cada aplicación o usuario puede consumir.

3.1.1.1. Mecanismos de QoS en el plano de control

En este plano se incluyen mecanismos de administración de QoS, de control de admisión y de señalización. La política de administración de QoS define y provee los diferentes tipos y niveles de servicios de QoS, así como gestiona qué usuario y aplicación usa qué QoS. La figura 13 muestra el sistema general de política de administración de QoS descrito en el IETF.

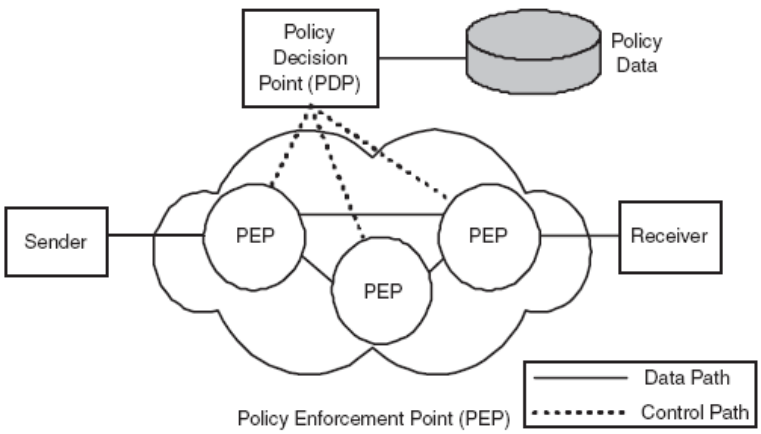

Figura 13. Esquema sistema administración QoS.

Los componentes de este sistema incluyen un almacén de datos de política, como los nombres de usuarios, aplicaciones, y los recursos de red que tiene asignados; Puntos de decisión de las políticas a aplicar (PDP-Policy Decision Points), que traducen los datos de políticas de niveles superiores a información específica de configuración para nodos individuales de la red, es decir es el elemento que oma las decisiones de autorización; Puntos de cumplimiento de política (PEP-Policy Enforcements Points), que son los elementos responsables de asegurar que solo los usuarios con permiso puedan hacer uso de los recursos protegidos; y los protocolos de comunicación entre los distintos nodos PDP y PEP.

La señalización trata los métodos de un usuario para comunicar a la red los requisitos de QoS que necesita. Los mecanismos de señalización pueden ser estáticos o dinámicos. En el caso estático, el PDP coge la información de política del nivel superior del los datos de política y crea una información de configuración que se implanta en cada PEP. Los datos de política normalmente son creados basándose en un acuerdo de los niveles requeridos por cada servicio (SLA – Service Level Agreements) entre el usuario y el proveedor de la red. En el caso dinámico, el usuario o la aplicación ira señalizando los requisitos de QoS cuando los va necesitando justo antes del flujo de datos. Para esta señalización se usa el protocolo RSVP (Resource Reservation Protocol). Cuando llega al PEP una petición de QoS, éste comprueba con el PDP la aprobación, y si es aceptada le asigna los recursos necesarios para cumplir dichos requisitos.

El control de admisión trata la habilidad de la red de controlar la admisión de nuevo tráfico en la red, basándose en los recursos disponibles. El control de admisión es necesario para asegurar que la admisión del nuevo tráfico dentro de la red no repercutirá en los tráficos existentes dentro de esta.

3.1.1.2. Mecanismos de QoS en el plano de datos

Estos mecanismos imponen la calidad de servicio acordada mediante la clasificación de los paquetes recibidos en distintas colas y asignando los recursos apropiados a cada cola. La clasificación se hace inspeccionando las cabeceras de los paquetes recibidos; la asignación de recursos se hace usando un algoritmo programado apropiado y técnicas de gestión de buffers para guardar y enviar cada paquete en cada cola.

Hay dos formas de definir estas colas, la primera llamada gestión por flujo, se basa en tener distintas colas para cada sesión individual o flujo. Para IP esto se hace mediante los cinco campos de la cabecera IP: dirección IP origen, dirección IP destino, puerto de origen y de destino, y protocolo de la capa de transporte. Esta gestión de las colas permite tener buena calidad, ya que la sesión tiene garantizadas los recursos independientemente de las demás sesiones. Sin embargo también requiere que cada nodo de la red tenga su sesión individual y aplique diferentes procesos, lo que resulta muy difícil y poco práctico cuando el número de flujos es muy grande, sobre todo en el núcleo de las redes.

La segunda opción es clasificar los paquetes en diferentes clases generales y poner cada clase en diferentes colas. Este método es llamado gestión por agregación, ya que aquí las colas consistirán en paquetes de diferentes sesiones y flujos. Para determinar a qué clase pertenece el paquete se usa la cabecera, DiffServ es un ejemplo de este tipo de gestión y este

tipo es mucho más escalable que el anterior. Con esto conseguimos calidad aunque de alguna manera está comprometida por el tráfico de los demás.

3.1.2. Tecnologías de QoS IP

Las redes tradicionales IP fueron diseñadas como redes de datos de mayor esfuerzo y no incluían ningún tipo de provisionamiento de QoS. Se podían conseguir algunas formas de QoS mediante la dependencia con algunos de los protocolos de la capa de transporte de extremo a extremo que van sobre IP. Este es el caso de TCP (Transport Control Protocol) por ejemplo, que asegura que los datos son transferidos de un extremo al otro y enviados en secuencia de manera que los flujos llegan de forma continua. Sin embargo estos protocolos de la capa de transporte no tienen ningún mecanismo para controlar el retraso de extremo a extremo o la tasa de transferencia de la red. Para asegurar el retardo y la tasa de transferencia de extremo a extremo se necesitan mecanismos de QoS en la capa de red e IP no tenía.

Es por ello que el IETF trabajó en nuevas arquitecturas y protocolos para conseguir esta calidad de servicio extremo a extremo en las redes IP. Tres de los mayores logros fueron IntServ (integrated services), DiffServ (differentiated services), y MPLS (Multiprotocol Label Switching). Estos tres elementos en conjunto consiguieron que la tradicional red IP se transformara en una red con capacidad de calidad de servicio y mucho más manejable.

3.1.2.1. Arquitectura IntServ (Integrated Services)

La arquitectura IntServ se ha diseñado para garantizar QoS sobre una base de gestión por flujo con significante granularidad mediante el uso de señalización dinámica extremo a extremo y la reserva de recursos por toda la red IP.

Intserv usa el protocolo RSVP para la señalización de los requisitos de QoS y para hacer la reserva de recursos de extremo a extremo. Los mensajes RSVP llevan información de cómo puede identificar un flujo en particular la red puede identificar, de los parámetros que describen el flujo, del servicio requerido para el flujo, y la información de política (la identidad de usuario y la aplicación).

A pesar de que la arquitectura IntServ proporciona el mayor nivel de QoS sobre IP garantizado, tiene algunas limitaciones. La primera es que al usar la gestión por flujo como comentamos anteriormente tiene problemas de escalabilidad. No podemos controlar los flujos asociados a un millón de sesiones individuales en el núcleo de la red. La segunda limitación es la necesidad de actualización periódica de la información de estado lo que puede ser muy difícil en

grandes redes. La tercera es que como RSVP no funciona con un protocolo de transporte fiable como TCP, los mensajes de señalización pueden perderse. La cuarta es la dificultad de implementación de IntServ y RSVP. La última es que RSVP requiere una infraestructura de autentificación para asegurar la validez de la solicitud de reserva. Todo esto hace que IntServ sea poco usado en grandes redes IP. Si son efectivas en pequeñas redes. Como alternativa aparece DiffServ que vemos a continuación.

3.1.2.2. Arquitectura DiffServ - Differentiated Services

Viendo los problemas de escalabilidad de la implementación de IntServ, el IETF comenzó a desarrollar otro modelo en 1997 capaz de proveer QoS sin necesidad de señalizar ni el mantenimiento del estado de la red. Se llamo servicios diferenciados, o Diffserv, este nuevo modelo se basa en la gestión por agregación. DiffServ divide el tráfico en un número pequeño de clases y trata cada clase de forma diferente. DiffServ usa el campo TOS (Type Of Service) de la cabecera IP para marcar los paquetes según las distintas clases. Dentro de este campo solo usa los 6 primeros bits, llamados DSCP (DiffServ Code Point), vemos una ilustración de esto último:

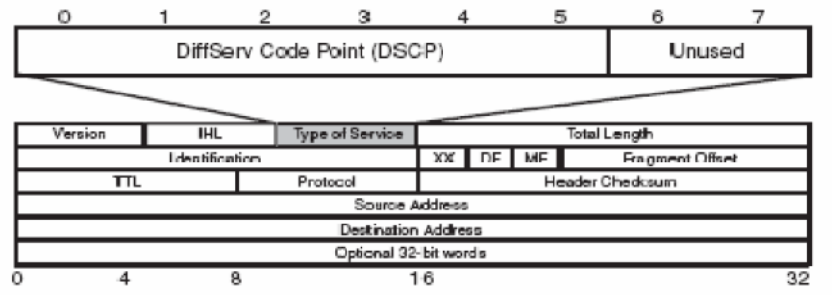

Figura 14. Campo DSCP de la cabecera IP.

Normalmente, un usuario o una aplicación enviando tráfico a una red DiffServ marca cada paquete transmitido con el DSCP apropiado para que así los routers puedan clasificar los paquetes.

Este mecanismo se usa para redes IP de gran tamaño ya que ofrecen buena calidad de servicio, es escalable y fácil de implementar. Es posible diseñar redes con partes con arquitectura IntServ y que en el núcleo tenga DiffServ para coger lo mejor de las dos arquitecturas.

3.1.2.3. MPLS MultiProtocol Label Switching

MPLS es otro de los recientes desarrollos para implementar redes IP. Originalmente fue desarrollado como método para mejorar la velocidad de los routers, ahora MPLS se usa como una herramienta de tráfico y como mecanismo para ofrecer diferentes servicios. Además MPLS permite una rigurosa integración entre IP y ATM, mejorando la actuación de IP sobre redes ATM.

Como la tecnología IP ofrece un servicio no orientado a conexión, mediante el transporte de datagramas, no mantiene el estado de la comunicación entre dos nodos, y no puede ofrecer circuitos virtuales. MPLS es una nueva tecnología de conmutación creada para proporcionar circuitos virtuales en las redes IP.

La idea básica de MPLS es insertar entre las cabeceras de la capa de enlace y la de transporte (IP) una nueva etiqueta de longitud fijada para indicar como se debe tratar al paquete dentro de una red MPLS. Dentro de la red MPLS, los paquetes no son enrutados usando la cabecera IP sino con la información de esta etiqueta, así con esta tecnología el camino que se sigue esta prefijado desde el origen.

A pesar de que no ofrece mecanismos de QoS de extremo a extremo, provee una gran infraestructura sobre la cual se pueden implementar los distintos mecanismos de QoS vistos anteriormente. Ambos IntServ y DiffServ se implementan sobre una estructura MPLS, siendo la opción MPLS-DIffServ más común. Por tanto, MPLS, rompe con el principio punto a punto del protocolo IP y pone control en las manos del operador de la red.

3.2. Gestión de la sesión multimedia

Una sesión está definida como un conjunto de comunicaciones significativas entre dos o más usuarios de una duración limitada. En el contexto de las comunicaciones multimedia, el termino sesión incluye telefonía de voz, audio y video streaming, chat y mensajes instantáneos, juegos interactivos, sesiones de realidad virtual, y demás. Una sesión suele tener múltiples conexiones asociadas con ella; por ejemplo una videoconferencia tiene dos conexiones separadas el video y el audio.

La gestión de sesión va más allá de la transferencia de bits de un transmisor a un receptor. Incluye servicios para localizar y conseguir el consentimiento de las partes implicadas en la comunicación, para la negociación de parámetros y las características de la comunicación, para modificar la etapa intermedia de las comunicaciones cuando sea necesario y para terminarlas. Para las aplicaciones IP tradicionales, como la navegación por páginas web y el

email, la gestión de la sesión es más simple. Sin embargo las comunicaciones IP multimedia necesitan un esquema de gestión de la sesión más robusta, primordialmente porque necesitan soportar una gran cantidad de aplicaciones y terminales. Las tareas de gestión de sesión como la negociación de capacidades se convierten en una parte muy importante de las comunicaciones cuando los diferentes terminales tienen distintos esquemas de codificación por ejemplo.

Claramente es necesario el uso de un protocolo de control de sesión para soportar servicios multimedia, incluyendo la telefonía IP. El estándar definido por la ITU se llama H.323 y fue el protocolo usado tradicionalmente para estos propósitos en la mayoría de los sistemas de telefonía IP y en los sistemas multimedia. Recientemente surgió un protocolo mucho más simple y ligero llamado SIP (Session Initiation Protocol), este protocolo ha surgido como el principal contendiente para estas tareas, y será el protocolo de control de sesión estándar usado en las redes WiMAX. SIP ha sido recientemente elegido como el protocolo de control de sesión para las redes de telefonía móvil de tercera generación. También se necesita un protocolo en la capa de transporte que cumpla con los requisitos multimedia tal y como RTP (Real Time Protocol) que fue diseñado para estos propósitos. SIP y RTP trabajan muy bien en conjunto para proporcionar funciones de control de sesión y de transporte de los datos multimedia que requiere una sesión IP multimedia.

3.2.1. SIP (Protocolo de Inicio de Sesión) (IETF)

SIP son las siglas en inglés del protocolo para el Inicio de Sesión, siendo un estándar desarrollado por IETF, identificado como RFC 3261. Es un protocolo de la capa de aplicación de señalización diseñado específicamente para trabajar sobre IP e Internet. El hecho de que este protocolo fuese desarrollado por la IETF implica que está más integrado en las aplicaciones y servicios de Internet, aporta más flexibilidad, es un protocolo de poco peso y extensible, y su implementación es más simple que el protocolo de la recomendación de la ITU-T, el H.323.

Su filosofía de diseño fue desacoplar el protocolo de señalización del servicio y así hacerlo útil para un amplio rango de servicios. SIP se integra perfectamente con otros protocolos basados en IP para proveer capacidades completas de sesión multimedia. Por ejemplo, usará RTP para el intercambio de servicios multimedia, TSL(Transpor-layer Security) para seguridad, y SDP (Session Description Protocol) para la descripción de la sesión, SIP puede trabajar sobre una variedad de protocolos de transporte: TCP (Transport Control Protocol), UDP (User Datagram Protocol), SCTP(Stream Control Transport Protocol) y TSL sobre TCP. Obviamente,

los flujos multimedia, como los de voz y vídeo para comunicaciones en tiempo real usan UDP para limitar los retrasos.

Una característica importante de SIP es su programabilidad. SIP sigue el modelo de programación de http, que permite a los usuarios y proveedores desarrollar servicios basados en SIP fácilmente.

3.2.1.1. Componentes y Arquitectura SIP

El protocolo se integra dentro de una arquitectura multimedia distribuida, abierta y escalable, caracterizada por su flexibilidad y su compatibilidad con el estándar H.323. Dentro de esta arquitectura genérica, podemos especificar una arquitectura concreta SIP, con dos partes fundamentales: agente de usuario y servidores (Proxy, de registro y de redirección). Los componentes básicos de una arquitectura SIP está ilustrado en la siguiente figura:

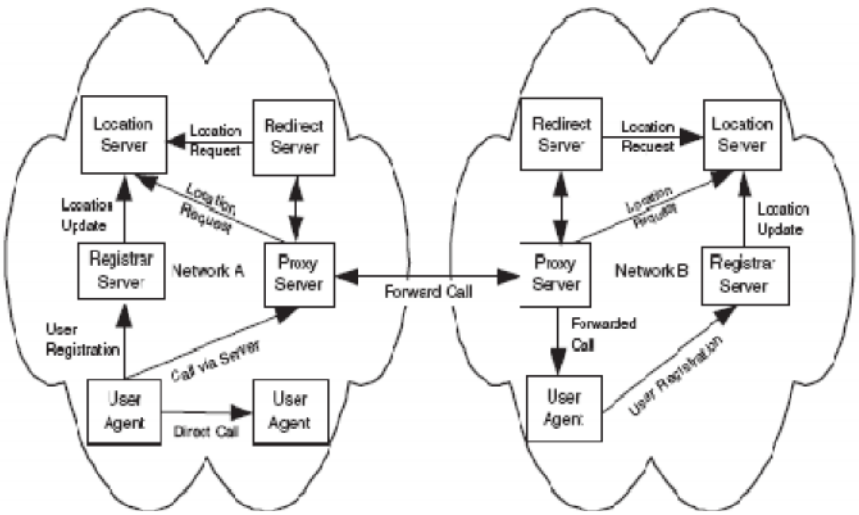

Figura 15. Arquitectura SIP.

Agente de Usuario o UA (User Agents) son los puntos finales de la estructura y son los responsables de hacer o responder las llamadas de parte de un usuario o una aplicación SIP. Todo usuario o aplicación SIP tiene asignado una dirección basada en un localizador URL (Uniform Resource Locater) con un formato del tipo “sip:roberto@192.190.132.31”, los servidores Proxy pueden resolver y transformar URIs del tipo TEL que contienen direcciones E.184: “tel:+541148277237”. De esta forma SIP integra su servicio a la Internet En este modelo se requiere el auxilio de un server de resolución de dominio DNS (Domain Name Server). El agente de usuario actúa tanto como un cliente como un servidor, dependiendo si está generando peticiones o está respondiendo a peticiones de parte del usuario. Típicamente el agente de usuario se implementa en el terminal del abonado pero también puede estar en un servidor de aplicaciones en otro lado, por ejemplo en servidor de video en la red. Puede tomar distintas

formas de acuerdo a su función: teléfono, softphone, Gateway PSTN, servidor de conferencias.

Servidor proxy SIP, que transmite la señalización de la sesión y actúa como cliente y servidor. El servidor proxy suele operar de una manera transaccional y no conserva la información del estado de la sesión. Esto hace que sea bastante escalable y robusto. Sin embargo para ciertas aplicaciones puede exigirse la conservación de la información del estado. El servidor proxy determina la localización actual del usuario preguntándole al servidor de localización. También provee servicios de autentificación y seguridad según sean necesarios e interactúa con otros proxies pertenecientes a diferentes dominios SIP.

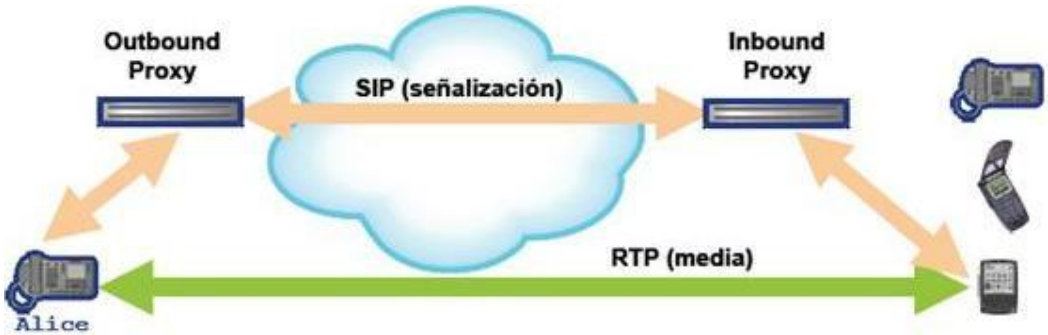

Figura 16. Servidor proxy en SIP.

Servidor de redirección SIP, que responde al agente de usuario con una respuesta de redireccionamiento notificando la localización del usuario llamado. El agente de usuario tiene que establecer una nueva sesión con la dirección indicada. Esta función es análoga a la de un servidor DNS, que provee la dirección IP de una URL dada.

Servidor de registro, que es donde el agente de usuario registra su información de localización actual y sus preferencias. La información de localización incluye usualmente la dirección IP actual del agente de usuario SIP pero también incluye otros detalles específicos de la capa de enlace, como por ejemplo la identidad de la estación base o la identidad del router de acceso. El mensaje de registro incluye también el protocolo de transporte que va a usar, TCP o UDP, el número del puerto, y detalles opcionales.

Servidor de localización, es el servidor con el que contacta el servidor proxy para determinar la localización del usuario. Debe estar en otros servidores, como por ejemplo en el servidor de registro.

El protocolo SIP tiene dos tipos de mensajes: Request y Response. El mensaje de request o petición es emitido desde el terminal cliente al terminal servidor y el response o respuesta en dirección contraria. El encabezado de ambos mensajes contiene campos similares:

- Start Line. Usada para indicar el tipo de paquete, la dirección y la versión de SIP.
- General Header. Contiene el Call-ID (se genera en cada llamada para identificar la misma); Cseq (se inicia en un número aleatorio e identifica en forma secuencial a cada request); From (es la dirección del origen de la llamada); To (es la dirección del destino de la llamada); Via (sirve para recordar la ruta del request; por ello cada proxy en la ruta añade una línea de vía) y Encryption (identifica un mensaje que ha sido encriptado para seguridad).
- Additionals. Además del encabezado general se pueden transportar campos adicionales. Por ejemplo: Expire indica el tiempo de validez de registración; Priority indica la prioridad del mensaje; etc.

SIP define 6 tipos de mensajes básicos, llamado métodos y definidos en la RFC 3261 pero permite extensiones de estos métodos básicos. Los mensajes básicos son ACK, BYE, CANCEL, INVITE, OPTIONS, y REGISTER. En la tabla 8 vemos una lista de los principales métodos usados en SIP y su descripción. Así mismo SIP también defino un número de código respuesta que el receptor de mensajes SIP puede usar para responder, dichos mensajes están listados en la tabla 9.

Método	Descripción funcional	Referencia
ACK	Confirmación de llegada de un mensaje INVITE y aceptación	RFC 3261
BYE	Terminación de una sesión	RFC 3261
CANCEL	Cancelar una sesión y las transacciones pendientes	RFC 3261
INFO	Señalización durante la sesión	RFC 2976
MESSAGE	Mensaje instantáneo sin necesidad de establecer sesión.	RFC 3428
NOTIFY	Notificación de usuario o de evento	RFC 3265
OPTIONS	Información capacidades configurada entre agentes /servidores	RFC 3261
PRACK	Acuse de recibo provisional	RFC 3262
REFER	Ordena al usuario establecer sesión con una tercera persona.	RFC 3515
REGISTER	Registra una URL en el servidor SIP.	RFC 3261
SUBSCRIBE	Permite al usuario pedir notificaciones de eventos	RFC 3265

Tabla 8. Métodos SIP principales.

Clase	Descripción	Ejemplos
1XX	Información provisional, petición en progreso pero no terminada.	100 Trying 180 Ringing, 181 Forwarding, 182 Queing, 183 In Progress
2XX	Petición completado satisfactoriamente.	200 OK
3XX	Redirección a otra localización.	301 Moved Permanently 302 Moved Temporaly
4XX	Error de cliente (error en la petición). Error de sintaxis	401 Unauthorized, 404 Not Found 420 Bad Extension, 486 Busy
5XX	Error de servidor.	503 Service unavaible
6XX	Error de la parte llamada...	600 Busy, 603 Decline

Tabla 9. Códigos de respuesta SIP.

3.2.1.2. Funcionamiento del protocolo SIP

Inscripción en la red SIP: El agente de usuario usa el método REGISTER con el fin de indicar al servidor proxy la correspondencia entre su dirección SIP y su dirección IP. La dirección IP puede ser estática u obtenida de modo dinámico por DHCP (Dynamic Host Configuration Protocol). Desde este momento el agente de usuario ya se encuentra ubicado y puede recibir llamadas.

El agente de usuario que quiere comenzar la sesión envía un mensaje INVITE hacia el servidor proxy, que interroga al servidor de localización para identificar la localización del usuario que está iniciando la sesión y encamina la llamada a su destino. El mensaje INVITE tiene varios encabezamientos obligatorios, entre los cuales están la dirección SIP de la persona que quiere iniciar la sesión, la dirección SIP del destinatario, una identificación de llamada, un número de secuencia, el número máximo de saltos. Este encabezamiento se actualiza en todas las entidades que participan en el enrutamiento de la sesión hasta que llega a su destino. Eso asegura que la respuesta siga el mismo camino que la petición. Por otra parte la petición INVITE contiene varias líneas que describen las características de la sesión que el origen necesita, esta estructura se llama SDP. El receptor responde con un mensaje de respuesta 180 RINGING hacia el agente de usuario que ha generado la llamada, y cuando acepta la sesión, manda la respuesta 200 OK hacia el agente de usuario emisor. A lo que el llamante responde con un ACK.

Los servidores participan en el encaminamiento de la señalización y cuando se establece la sesión los agentes de usuario establecen entre ellos canales RTP para el transporte de los datos multimedia en forma de paquete sin implicación de los servidores en este transporte.

Cuando uno de los agentes termina la comunicación envía BYE al otro agente, a lo cual responde con un mensaje de confirmación 200 OK.

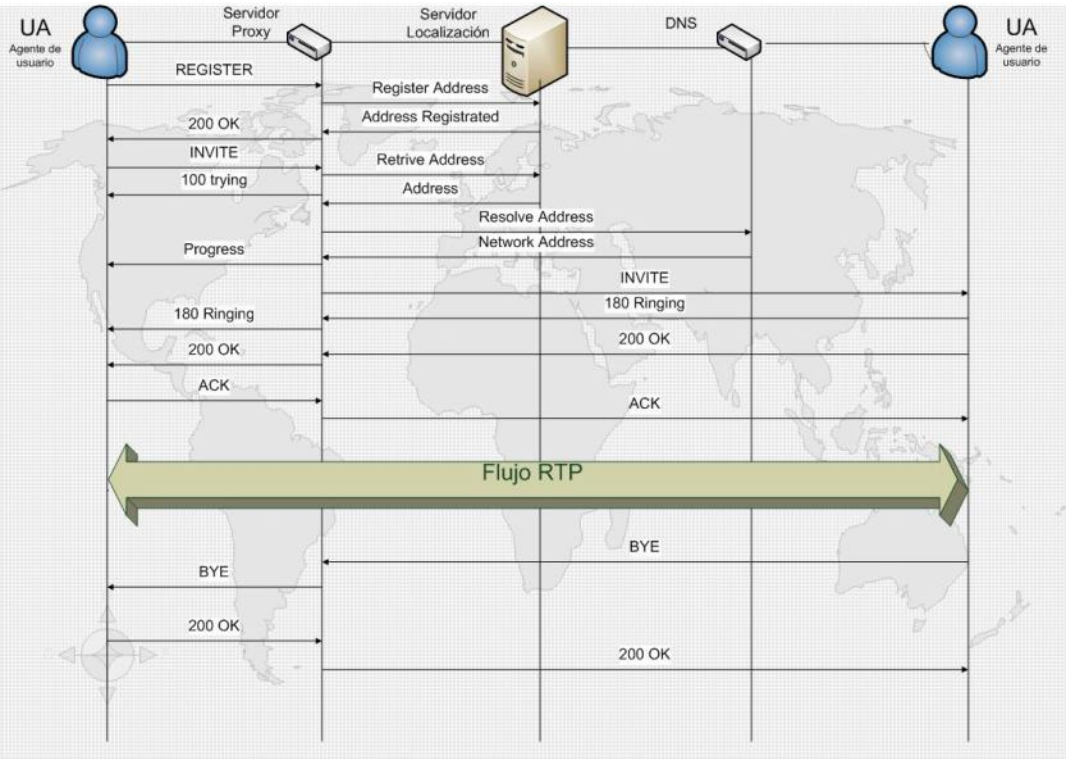

Figura 17. Funcionamiento protocolo SIP.

3.2.2. Real-Time Transport Protocol.

SIP provee las funciones necesarias de control de sesión pero no se usa para transportar los flujos de información. RTP, definido en la RFC 1889, es el protocolo de transporte más popular para la transferencia de datos en sesiones multimedia. RTP fue desarrollado porque los protocolos de transporte tradicionales, como TCP y UDP, no eran apropiados para las sesiones multimedia: TCP no ofrece límites de retardo, y UDP no garantiza retardos ni la pérdida de paquetes. Típicamente RTP trabaja sobre UDP y provee información de tiempo y de secuenciación apropiados para aplicaciones en tiempo real, como son la voz y el vídeo. La cabecera RTP contiene contenido identificativo, método de codificación de audio/vídeo, numero de secuencia, y marca de tiempo para asegurar que los paquetes son reproducidos en el orden

correcto y a una tasa constante. La información de tiempo facilita el cálculo del jitter y permite que el receptor adopte la estrategia adecuada de almacenamiento. Implementamos RTP junto con RTCP (Real-Time Control Protocol), que gestiona los flujos de tráfico, RTCP provee retroalimentación de la calidad del enlace, lo que puede usarse para modificar los esquemas de codificación si es necesario. Usando las marcas de tiempo, RCTP también facilita la sincronización de múltiples flujos, como los flujos de audio y vídeo asociados en una sesión. RTCP también proporciona el soporte para conferencias de grupos en tiempo real. RTP y RTCP no reducen el retraso total de la información en tiempo real o hacen alguna garantía de la calidad de servicio.

3.3. Seguridad

La seguridad es uno de los temas más complejos y amplios, en este punto solo vamos a ver una introducción a esta. Vamos a ver los puntos básicos de seguridad, introduciendo alguna terminología especifica, y dando una visión general de algunos de los mecanismos de seguridad, usando ejemplos relevantes en los servicios de banda ancha inalámbricos, especialmente de WiMAX. Una arquitectura de seguridad bien diseñada para una comunicación inalámbrica debe de soportar los siguientes elementos básicos:

Privacidad: Aportar protección ante escuchas no autorizadas a medida que los datos de usuario atraviesan la red desde el origen al destino.

Integridad de los datos: Asegurar que los datos de usuario, y los mensajes de control y gestión están protegidos ante interferencias durante todo el recorrido.

Autenticación: Tiene un mecanismo para asegurar que un determinado usuario es quién dice que es. En cambio, el usuario debe ser capaz de verificar la autenticidad de la red a la que se conecta.

Autorización: Mecanismo para verificar que un usuario determinado está autorizado a recibir un servicio en particular.

Control de acceso: Asegura que sólo se permite el acceso a los servicios ofrecidos a los usuarios autorizados.

Dentro de un sistema son varias las capas que se encargan de la seguridad, cada una de ellas trata un aspecto de la seguridad, aunque en algunos casos suele haber mecanismos redundantes. Como principio general de seguridad, se debe tener mecanismos de seguridad redundantes para proveer seguridad aunque uno de los mecanismos falle. En la siguiente tabla

vemos algunos de los mecanismos de seguridad que se implementan en las diferentes capas. En el nivel de enlace, debe usarse una fuerte encriptación para sistemas inalámbricos para prevenir escuchas no autorizadas en el aire que es el medio de transmisión. Además también se necesita un control de acceso para prevenir que usuarios no autorizados usen los recursos de la red. La encriptación de la capa de enlace no suele usarse en los enlaces por cable, donde la privacidad se asegura mediante mecanismos de seguridad de las capas superiores. En la capa de red, hay varios métodos que proveen seguridad, por ejemplo IPsec puede usarse para proveer servicios de autentificación y encriptación. La red por si misma debe protegerse de ataques maliciosos con el uso de firewalls. Los servicios de autentificación y autorización se suelen hacer mediante protocolos AAA (Authentication, Authorization, and Accounting) como RADIUS (Remote Access Dial-In User Service) y DIAMETER. En la capa de transporte se usa TSL (Transport Security Layer) para añadir seguridad a los protocolos y paquetes del nivel de transporte. En la capa de aplicación se implementan firmas digitales, certificados, gestión de derechos digitales, etcétera, dependiendo de la sensibilidad de la aplicación.

Capa	Mecanismo de seguridad	Notas
Enlace	Encriptación AES, autentificación de los periféricos, y autentificación de los puertos (802.1X)	Usualmente solo se realiza en los enlace inalámbricos
Red	Firewall, IPsec, infraestructura AAA (RADIUS,DIAMETER)	Protege la red y la información que va sobre ella
Transporte	TSL (Transport-layer Security)	Provee seguridad en los servicios de la capa de transporte usando certificados
Aplicación	Firmas digitales, certificados, transacciones electrónicas seguras (SET), gestión de derechos digitales (DRM)	Provee privacidad y autentificación: normalmente se basa en una estructura de clave pública.

Tabla 10.Ejemplos de mecanismos de seguridad en las diferentes capas.

En las siguientes secciones vamos a dar una visión global de algunos de estos mecanismos de seguridad que son más relevantes para WiMAX.

3.3.1. Encriptación y AES

La encriptación es el método usado para proteger la confidencialidad de los flujos de datos entre el transmisor y el receptor. La encriptación involucra coger un flujo o un bloque de datos para protegerlos, conocido como texto plano, y usando otro flujo o bloque de datos, llamado clave de encriptación, conseguir una operación matemática reversible para generar un texto cifrado. El texto cifrado es incomprensible y de ahí que pueda ser enviado por la red sin miedo de ser interceptado e interpretado por otros usuarios. El receptor debe de realizar la operación inversa, que se llama desencriptación, para extraer el texto plano del texto cifrado usando la misma clave o una diferente. Cuando se usa la misma clave para la encriptación y la desencriptación, el proceso se llama encriptación de clave simétrica. Normalmente esta clave proviene de un secreto compartido por el transmisor y el receptor y normalmente para conseguir una encriptación fuerte debe tener al menos 64 bytes de longitud. Cuando se usan diferentes claves para la encriptación y la desencriptación, el proceso se llama encriptación de clave asimétrica. Estos dos métodos de encriptación son normalmente usados en los sistemas de comunicación inalámbricos de banda ancha, cada uno es utilizado según las necesidades. En esta sección veremos un sistema de encriptación de claves simétricas llamado AES (Advanced Encryption Standard); en la siguiente sección veremos un sistema de encriptación de claves asimétricas.

AES es un nuevo estándar de encriptación de datos adaptado por el Instituto Nacional de Estándares y es un método específico de encriptación de la capa de enlace que va a ser usado en WiMAX. AES está basado en un algoritmo llamado Rijndael, que es un método de cifrado en bloques que proporciona fuertes propiedades criptográficas. Además AES es rápido, fácil de implementar en hardware o software, y requiere menos memoria que otros esquemas de encriptación comparables. La eficiencia computacional de AES ha sido una de las razones claves para su adopción masiva.

El algoritmo AES opera con bloques de datos de 128 bits, organizados en una matriz de 4x4 bytes llamada state. Las claves de encriptación pueden ser de distintos tamaños: 128, 192 o 256 bits. Específicamente WiMAX usa una clave de 128 bits.

Para el cifrado, cada ronda de la aplicación del algoritmo AES (excepto la última ronda que reemplaza la fase MixColumns por otra instancia de AddRoundKey) consiste en cuatro pasos:

1. SubBytes: En este paso se realiza una sustitución no lineal donde cada byte es reemplazado con otro de acuerdo a una tabla de búsqueda.

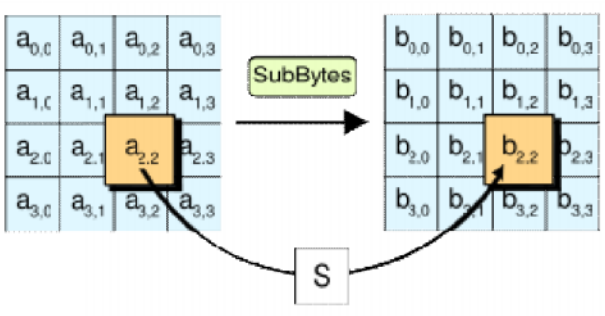

Figura 18. SubBytes.

2. ShiftRows: En este paso se realiza un transposición donde cada fila del state es rotado de manera cíclica un número determinado de veces.

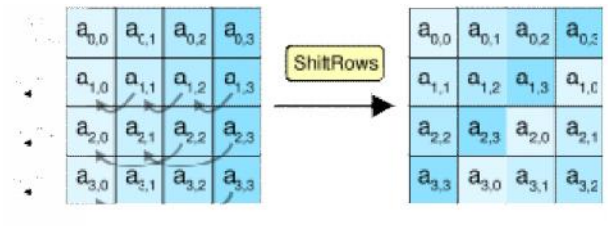

Figura 19. ShiftRows.

3. MixColumns: Operación de mezclado que opera en las columnas del state, combinando los cuatro bytes en cada columna usando una transformación lineal.

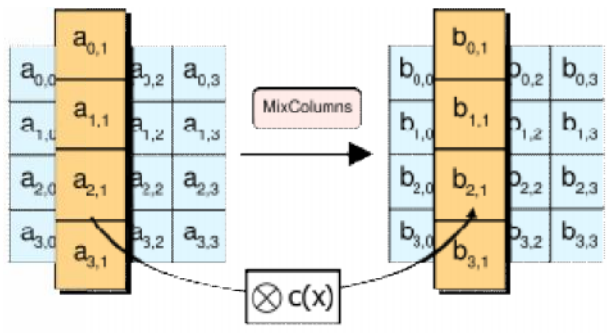

Figura 20. MixColumms.

4. AddRoundKey: cada byte del state es combinado con la clave round; cada clave round se deriva de la clave de cifrado usando una iteración de la clave.

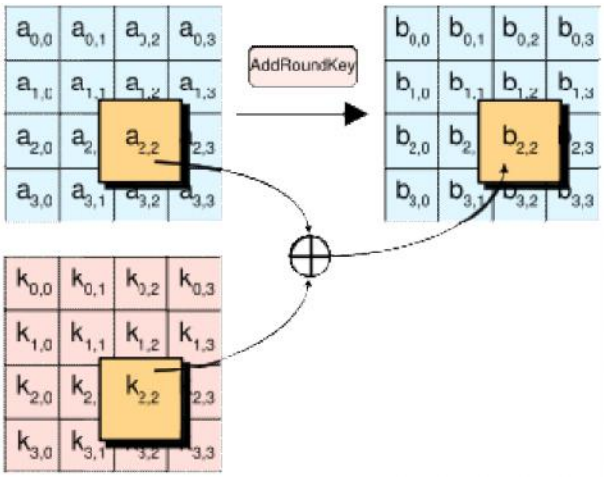

Figura 21. AddRounKey.

Para poder usar un cifrado en bloques, como es AES, se necesita un mecanismo reversible para convertir un mensaje de longitud arbitraria en una secuencia de bloques de tamaño fijo antes de ser encriptados. Este método se conoce como el modo de operación del cifrado, muchos de los cuales surgieron en base a AES. El modo de operación debe ser elegido con cuidado para no crear fallos de seguridad y teniendo en cuenta las consideraciones de implementación. El modo usado en WiMAX se llama modo contador y se muestra en la siguiente figura:

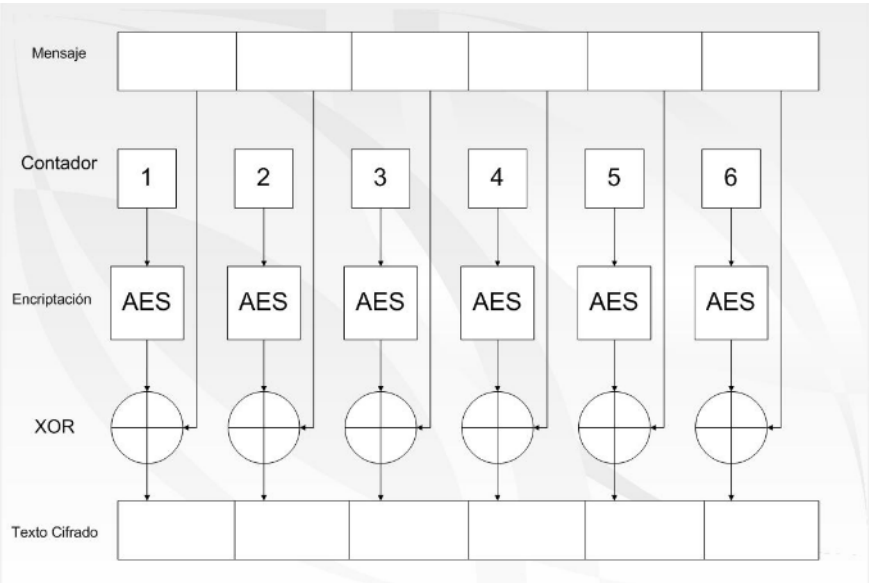

Figura 22. Modo de operación contador para AES.

En el modo contador, en vez de encriptar directamente el texto plano, un bloque arbitrario llamado "contador" es encriptado usando el algoritmo AES, y como resultado de una operación XOR con el texto plano se consigue el texto cifrado. El nombre de contador se debe a que el bloque va incrementándose en una unidad para cada bloque sucesivo, el contador puede empezar en 1 o en cualquier número arbitrario. Cambiando el número del contador para cada bloque conseguimos que el texto cifrado nunca sea igual para dos entradas iguales, por tanto provee protección ante alguien que observe los patrones de repetición en el texto cifrado.

El modo de desencriptación es exactamente igual que el de encriptación, ya que al hacer una operación XOR dos veces con el mismo número obtienes de nuevo el valor original. Más allá de esto, si el mensaje no se puede dividir en un número exacto de bloques, el último de ellos puede ser más pequeño que el resto, operando de la misma forma y enviando el número de bits requeridos por la salida. Todas estas características hacen que este modo sea una opción bastante popular para la implementación del mecanismo de cifrado AES.

3.3.2. Infraestructura de clave pública

Con la encriptación de claves simétricas, tanto el transmisor como el receptor necesitan tener la misma clave, lo que lleva a cuestionarnos como podemos transmitir la clave de forma segura. Una forma de hacerlo es estableciendo la clave con antelación por otro medio. Pero para este sistema no es muy escalable.

La encriptación de claves asimétricas es una solución elegante al problema de distribución de claves. Este método usa dos claves, una pública y otra privada. Cuando el texto cifrado es encriptado usando una de estas dos claves, este puede ser desencriptado sólo con la otra clave. Ambas claves se generan simultáneamente usando el mismo algoritmo (RSA); la clave pública se distribuye extensamente y la clave privada se mantiene en secreto. El PKI (Public Key Infrastructure), que es un método usado para la seguridad de transacciones en Internet, está basado en la idea del uso de claves asimétricas. Este sistema es útil para varias aplicaciones de seguridad:

Autentificación: Un mecanismo de autentificación es por ejemplo si queremos asegurar que los datos recibidos son del usuario B, el usuario A puede usar la clave pública y la clave privada con un número aleatorio. Si B devuelve el mismo número aleatorio que A mandó, A puede asegurar la autenticidad de B.

Distribución de la clave secreta: Para asegurar el envío de datos al usuario B, el usuario A puede usar la clave pública de B para encriptar los datos. Estos solo podrán ser

desencriptados con la clave privada del usuario B. Esta transacción segura puede ser usada para distribuir la clave secreta y compartida entre A y B, que se puede usar para encriptar el resto de las comunicaciones, usando el algoritmo de clave simétrica, como AES.

Integridad de los mensajes y no repudiación: Las claves asimétricas y el método PKI pueden usarse para probar que alguien dijo algo.

Certificados digitales: Los certificados digitales son un medio de certificar la autenticidad y la validez de las claves públicas.

3.3.3. Control de acceso y autentificación

El control de acceso es un mecanismo de seguridad que asegura que solo los usuarios con permiso accederán a la red. En términos generales, un sistema de control de acceso tiene tres elementos: una entidad que desea obtener acceso: el suplicante; una entidad que controla el acceso a la red: el autentificador; y por último una entidad que decide si el suplicante debe ser admitido: el servidor de autentificaciones.

En la siguiente figura se muestra una arquitectura de control de acceso típica usada por los proveedores. Los sistemas de control de acceso fueron primero desarrollados para su uso en módems con conexión por la red telefónica conmutada por marcación directa, y después se adapto para los servicios de banda ancha. El protocolo básico para los servicios de marcación directa fue PPP (Point-to-Point Protocol), y para los servicios de marcación remota es RADIUS. PPP se usa entre el suplicante y el autentificador, que en muchos casos será un router o un servidor de accedo a la red (NAS - Network Access Server), y RADIUS se usa entre el autentificador y el servidor de autentificación.

PPP soporta varios tipos de esquemas de autentificación si usamos EAP (Extensible Authentication Protocol) ya que originalmente solo aceptaba dos.

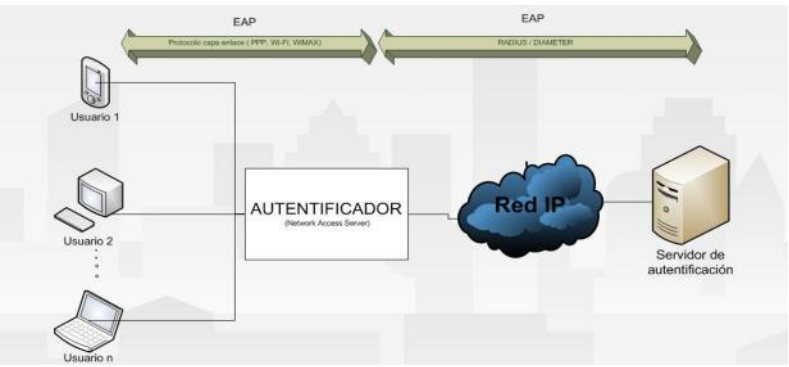

Figura 23. Arquitectura de control de acceso.

3.3.3.1. EAP - Extensible Authentification Protocol

EAP es una estructura creada por el IETF (RFC 3748) que permite intercambiar protocolos de autentificación entre el suplicante y el servidor de autentificación. EAP es una simple encapsulación que puede trabajar con PPP y con cualquier otro enlace, incluyendo WiMAX. La siguiente figura muestra la estructura EAP.

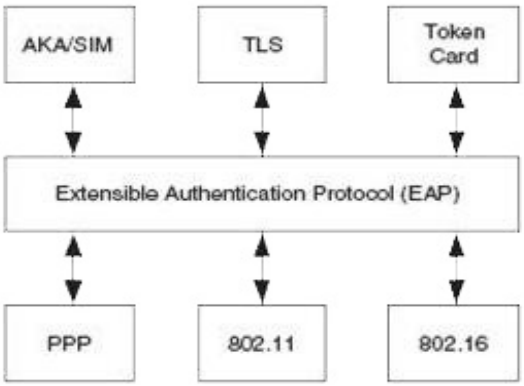

Figura 24. Arquitectura de EAP.

EAP incluye una serie de mensajes de negociación que pueden intercambiarse entre el cliente y el servidor de autentificación. El protocolo define también un conjunto de mensajes de petición y respuesta, donde el autentificador manda peticiones al servidor de autentificación; basándose en respuestas, se aprueba o se deniega el acceso al cliente. El protocolo asigna códigos tipo a varios métodos de autentificación y delega la tarea de proveer la identidad del usuario a un protocolo auxiliar, un método EAP, que define las reglas de autentificación para un usuario. Se han definido varios métodos EAP que usando una variedad de credenciales como contraseñas, certificados, testigos, y pequeñas tarjetas soportan la tarea de autentificación. Por ejemplo, PEAP (Protected EAP) o EAP protegido define un método EAP basado en contraseñas, EAP-TSL (EAP-Transport-Layer Security) define un método basado en certificados, y EAP-SIM (Subscriber Identity Module) define un método EAP basado en tarjetas SIM. EAP-TSL provee una autentificación mutua fuerte ya que cuenta con certificados en la red y en el terminal de usuario.

En los sistemas WiMAX, EAP se ejecuta entre la estación móvil y la estación base sobre el protocolo de seguridad PKMv2 (Privacy Key Management) definido en la IEE 802.16e-2005. Si el autentificador no está en la estación base, la estación base confía el protocolo de autentificación en el autentificador de la red de acceso al servicio (ASN). Cuando se hace en el sentido del autentificador al servidor de autentificación, EAP va sobre RADIUS.

3.3.3.2. RADIUS

Es el estándar más usado en las comunicaciones entre el autentificador y el servidor de autentificación, es un estándar del IETF donde se definen las funciones del servidor de autentificación y el protocolo de acceso a esas funciones. RADIUS es un aplicación cliente/servidor UDP que se ejecuta sobre IP. El servidor de autentificación es un servidor RADIUS, y el autentificador es el cliente RADIUS. Aparte de la autentificación, RADIUS también tiene funciones de autorización y contabilidad, como puede ser medir la duración y el volumen de una sesión, que puede usarse para facturar.

Sin embargo RADIUS tienen un número de deficiencias que no se pueden resolver fácilmente. Viendo esto el IETF ha desarrollado otro estándar llamado DIAMETER que ofrece mejores condiciones de seguridad, fiabilidad, seguridad y roaming.

3.4. Gestión de la movilidad

Son dos los mecanismos básicos que permiten al usuario comunicarse desde diferentes ubicaciones y mientras se mueve. La primera se llama gestión de localización, es un mecanismo que localiza a las estaciones móviles o estaciones de usuario dentro de la red. La segunda se llama gestión de traspaso de estación base y permite que el usuario mantenga una sesión al moverse del área de cobertura de una estación base a otra.

3.4.1. Gestión de la localización

Este mecanismo a su vez incluye dos procesos. El primero llamado registro de la localización, o actualización de la localización, consiste en que la estación móvil informa periódicamente a la red de su posición actual, lo que lleva a la red a autentificar al usuario y actualizar su localización en su perfil dentro de la base de datos. La localización normalmente está definida por un área que abarca el área de cobertura de una o más estaciones base. Hacer el área de localización más grande reduce el número de actualizaciones de localización ya que si no podríamos sobrecargar el sistema. Otro factor importante es la frecuencia con la que se realiza la actualización, ya que si no se hace frecuentemente puede arriesgarse a moverse de área de localización sin notificarlo a la red y que la red tenga información errónea sobre el usuario. Para soportar roaming global, la gestión de localización debe de realizarse dentro de la red y además deberá acordar con otros operadores que le presten servicio.

El segundo proceso se llama paging. Cuando por ejemplo llega una petición de inicio de sesión se busca la localización del receptor en la base de datos de localización y se informa al receptor mediante un canal de mensajería llamado “paging” que se transmite en todas las estaciones base del área de localización. Obviamente mientras más grande sea el número de estaciones base haya dentro del área de localización, mayor será la cantidad de recursos a usar para este proceso. Por tanto la red deberá realizar un compromiso entre los recursos a usar para la actualización de la localización de todas las estaciones base y los recursos de paging a usar en una gran área.

3.4.2. Gestión de traspaso

Este mecanismo requiere una actuación en tiempo real y para ciertas aplicaciones como VoIP debe realizarse de forma que no se perciba retraso en la comunicación ni perdida de paquetes. Para soportar estas aplicaciones WiMAX requiere que para la movilidad completa -hasta 120km/h- el retraso de traspaso debe ser menor de 50ms con una tasa de pérdidas de paquetes de menos de un uno por ciento.

El proceso de traspaso se puede ver como dos fases. En la primera fase el sistema detecta la necesidad del traspaso de una estación base a otra, y decide a qué estación base realiza la transición. En la segunda fase el traspaso se realiza, asegurando que la estación móvil y la estación base implicados están sincronizadas y todos los paquetes se envían correctamente, usando los protocolos apropiados.

La decisión de traspaso puede hacerla tanto la estación móvil como la red, basándose en la medida de la calidad del enlace. En WiMAX, normalmente la estación móvil toma la decisión final, mientras que la estación base hace recomendaciones sobre las estaciones base candidatas. La decisión se toma en función a la calidad de la señal recibida en la estación móvil y de la cual informa a la red, ya que normalmente la estación móvil escucha una señal de control de todas las estaciones base que rodean a su estación base y mide la calidad de la señal recibida. En WiMAX, la estación base debe actuar en este proceso proporcionando a la estación móvil una lista de vecinos y asociando los parámetros requeridos para escanear las estaciones base vecinas.

Una vez se toma la decisión de traspaso de una sesión en curso a una estación en base en concreto, se realiza una serie de pasos:

1. Antes de realizar el traspaso se hace una sincronización con las estaciones base vecinas.
2. Se establece un conexión mediante la capa física con más de una estación base al mismo tiempo para que los datos a transferir puedan ser cambiados de una a otra sin necesidad de ejecutar todo el conjunto de procedimientos de señalización de traspaso. IEEE 802.16e-2005 soporta esta funcionalidad, que es llamada intercambio rápido de estación base (FDBB-Fast Base Station Switching). En este caso, si todas las estaciones base que tienen una conexión con la estación móvil reciben paquetes de la red para dicha estación base, la perdida de paquetes cuando se realiza el intercambio puede reducirse significativamente.
3. Transferir todos los paquetes de la capa MAC no enviados que están en la cola de la estación base antigua a la nueva estación base mediante la red central para reducir la pérdida de paquetes y/o por la necesidad de retransmisiones de las capas superiores.

3.4.3. Mobile IP.

Hasta ahora hemos visto que cuando una estación móvil se mueve del área de cobertura de una estación base a otra, todo lo que necesita es mantener la conexión física para que los paquetes sigan fluyendo. Sin embargo esto no es suficiente. Para que una sesión de una aplicación quede intacta, la dirección IP de la estación móvil no debe cambiar durante toda la sesión. Si en la red se producen cambios de subredes IP, como es el caso de WiMAX cuando nos movemos a otro ASN (Access Service Network) Cuando esto ocurre, la dirección IP de la estación móvil cambia forzosamente, por lo tanto la conexión IP se corta, y las sesiones en curso terminan, aunque la conexión física se mantenga. Estos movimientos entre subredes también van a ocurrir cuando nos movamos por diferentes redes inalámbricas – por ejemplo cuando nos cambiamos de una red WiMAX a una red Wi-Fi o a una red 3G. Por tanto es necesaria una solución para mantener una sesión en curso intacta cuando nos movemos de una subred a otra, esta solución la desarrolló el IETF y se llama Mobile IP (MIP).

Mobile IP está diseñado específicamente para que sea transparente a la aplicación en el sentido de que no tiene que saber que el usuario se ha movido a una nueva subred IP, y además tiene que ser transparente a la red en el sentido de que los routers o los protocolos de enrutamiento no tienen que cambiar.

Los componentes básicos de Mobile IP se muestran en la siguiente figura. El cliente MIP se implementa en el terminal que está en movimiento y lo llamamos nodo móvil o MN (Mobile Node). El servidor IP con el que el MN se está comunicando lo llamamos nodo corresponsal o CN (Correspondent Node). Mobile IP define dos direcciones para cada MN, la primera es la dirección asignada al MN por su red, llamada HoA (Home Address) y la segunda dirección es una dirección temporal IP que se le asigna al MN en las redes visitadas y se llama CoA (Care-of-Address).

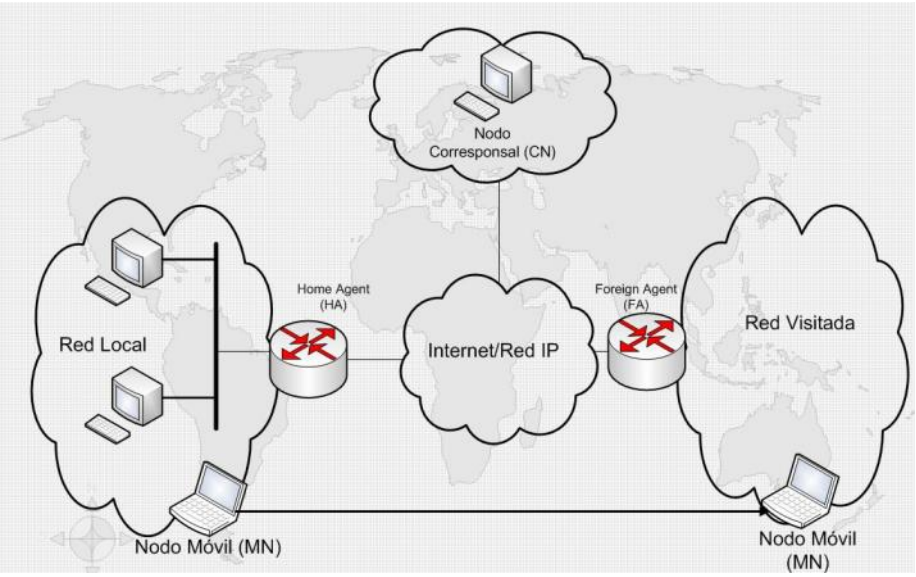

Figura 25. Componentes Mobile IP.

Para poder gestionar la movilidad, necesita una gente de movilidad llamado Home Agent (HA) que se encuentra en la red origen que trabaja conjuntamente con otro agente de movilidad llamado Foreign Agent (FA) que se encuentra en la red destino. Estos dos agentes proporcionan el mapeado dinámico de la dirección IP HoA a la dirección CoA.

Una vez el HA tiene la dirección CoA, todos los paquetes destinados al MN que llegan a la red origen se envían a la dirección CoA del FA en el que se encuentra encapsulando los paquetes con un protocolo llamado "tunneling" (entunelamiento). El FA desencapsula los paquetes y los envía al MN, por lo tanto conseguimos siempre enviar los paquetes al MN no importa donde se encuentre.

Aunque es un buen mecanismo para que los paquetes IP se envíen a un periférico que se mueve de una red a otra, Mobile IP no es suficiente para garantizar la continuidad de la sesión. Mobile IP se creó como solución a movilidades pequeñas y poco frecuentes.

En conclusión Mobile IP si es implementado correctamente y aplicando todas las mejoras posibles puede ser una solución efectiva ante la movilidad. Los retrasos creados pueden ser un problema para aplicaciones sensibles a retardos como VoIP, pero para muchos de las aplicaciones de datos esta solución es satisfactoria.

3.4.4. Mobile IP para IPv6

A diferencia de IPv4, los diseñadores de IPv6 han considerado la movilidad desde el principio, no una vez desarrollado. Por eso IPv6 tiene bastantes mejoras sobre Mobile IP en IPv4. La principal ventaja es que se ha añadido una optimización de enrutamiento; luego los paquetes no tienen que viajar a través del HA para llegar al MN. Una de las optimizaciones de enrutamiento son que las actualizaciones son enviadas a los CNs por el MN en vez de por el HA. Otras ventajas son:

- No necesita Foreign Agents.
- No necesita hacer “tunnelling”, la dirección CoA la lleva en las opciones de la cabecera de enrutamiento añadido al paquete original. Así se evita la encapsulación y reducimos el tamaño de las cabeceras.
- Se reduce la señalización. No es necesario enviar paquetes de control porque IPv6 permite añadir mensajes de control en los paquetes.

Hay bastantes diferencias más, todas se basan en el uso de características de IPv6 (como las opciones de cabecera para el destino) para realizar ciertas operaciones que en Mobile IPv4 se realizan de una forma un poco "artificial", como por ejemplo lo de encapsular los paquetes.

3.5. IP para redes inalámbricas: problemas y soluciones

IP es un protocolo a nivel de red que sigue un diseño modular que permite que pueda funcionar sobre cualquier protocolo de nivel de enlace, y a su vez transporta una gran cantidad de aplicaciones. El rotundo éxito que ha tenido Internet ha hecho que el protocolo a nivel de red IP sea la elección para muchos de los sistemas de comunicaciones modernos; no solo para la comunicación de datos sino también para comunicaciones multimedia, de voz y de video.

WiMAX ha elegido IP como protocolo para enviar todos sus servicios.

La modularidad de simplicidad de IP ha hecho que se hagan una serie de suposiciones sobre la red en la que subyace. IP supone que el nivel de enlace en la red siempre es fiable introduce muy pocos errores. IP no se esfuerza para hacer un uso eficiente de los recursos de la red; mejor supone que los recursos de la red son suficientes. Muchas de estas suposiciones no son buenas para una red inalámbrica; como resultado de usar IP sobre redes inalámbricas introduce problemas que deben solucionarse. En este apartado vamos a ver dos de estos problemas, el primero es el resultado de los errores de un enlace inalámbrico, y el segundo es sobre la escasez de ancho de banda en estos enlaces.

3.5.1. TCP en redes inalámbricas

El protocolo de control de transmisión o TCP se una en gran parte de las aplicaciones IP, como es el email, los servicios Web, y TELNET. TCP asegura que lo datos van a ser transferidos de forma fiable del origen al destino ya que es un protocolo orientado a conexión. TCP divide los datos de la capa de aplicación en segmentos y asegura que todos los segmentos van a ser enviados de forma fiable, incluyendo para ello un numero de secuencia y un checksum en la cabecera. Todo segmento TCP recibido correctamente se confirma enviando un paquete ACK al origen con el número de secuencia del siguiente paquete esperado. El receptor también informa al transmisor del tamaño de su ventana de recepción para que el numero de segmentos en transición siempre sea menor que dicha ventana.

TCP provee un mecanismo para el envío fiable de datos de punto a punto sin necesidad de recurrir a los nodos intermedios ya que para ellos TCP es transparente. Sin embargo esto lo hace realizando varias suposiciones sobre la red. Específicamente, TCP asume que toda la pérdida de paquetes, los paquetes no confirmados y los retrasos son ocasionados por congestión y que la tasa de pérdida es pequeña. Estas suposiciones en las redes inalámbricas no son validas, ya que los paquetes erróneos son muy frecuentes y la mayoría son causados por las condiciones tan pobres del canal. Si respondemos a esto disminuyendo la velocidad de envío no vamos a solucionar nada ya que no es problema de congestión, consiguiendo únicamente que TCP no consiga tener una tasa de transferencia constante.

Cuando las redes inalámbricas tienen grandes retardos, también lleva a la reducción de la tasa de transferencia. Grandes retardos y con una elevada tasa de transferencia también puede llevar a gran cantidad de datos en tránsito. Esto hace que TCP suponga que el buffer del receptor está lleno y disminuye su tasa de transferencia.

También si hay errores consecutivos en el envío de un paquete, TCP actúa ampliando su temporizador de retransmisiones, lo que conlleva a que haya largos periodos de inactividad.

Claramente, el uso de TCP sobre redes inalámbricas conlleva a una degradación innecesaria de la tasa de transferencia, y un uso ineficiente de los recursos, así como una excesiva interrupción de las transmisiones de datos. En los sistemas móviles, estos problemas se ven incrementados durante el traspaso de una estación base a otra. Una vez se detectaron todos estos problemas se ha investigado en muchos métodos para mejorar la actuación de TCP sobre las redes inalámbricas.

Hay dos propuestas para acabar con estos problemas de TCP sobre las redes inalámbricas. La primera propuesta es hacer que TCP sea consciente de que trabajando con enlaces inalámbricos debe cambiar su comportamiento. Estos esquemas intentan diferenciar entre las pérdidas basadas en congestión y las pérdidas inducidas por el canal, y hacen que TCP no aplique control de congestión cuando se trata de errores del canal. Ejemplos de estos esquemas son TCP-Santa Cruz, Freeze-TCP y otros. Ya que muchos de estos esquemas requieren hacer cambios en todos los servidores, se consideran soluciones poco prácticas.

La otra alternativa es hacer que el enlace inalámbrico se adapte a las necesidades de TCP. Una manera obvia de hacer que TCP trabaje bien sobre estos enlaces es hacer en enlace más fiable. Esto puede hacerse mediante el uso de una corrección de errores fuerte y mediante los esquemas de retransmisión de la capa de enlace (ARQ). Muchas de las redes de banda ancha inalámbricas actuales, incluyendo WiMAX, tienen este esquema de retransmisión ARQ. WiMAX soporta ARQ hibrido en la capa física además del estándar en la capa MAC.

ARQ en la capa de enlace hace que el enlace inalámbrico tenga relativamente pocos errores, pero introduce retrasos variables en el envío de paquetes, lo que puede causar problemas en los tiempos de vencimiento de TCP. Como resultado, es probable, por ejemplo que TCP asuma que se pierde un paquete mientras está siendo correctamente retransmitido usando un proceso ARQ en la capa de enlace. Esto es de nuevo un gasto del ancho de banda del enlace inalámbrico. Teniendo una coordinación más cercana entre la capa de enlace y TCP, esta solución puede llegar a ser muy efectiva.

3.5.2. Compresión de cabeceras

En aplicaciones IP tales como VoIP, mensajes, y juegos interactivos, el tamaño de la carga útil de los paquetes tiende a ser bastante pequeña. Para estos paquetes, el tamaño de la cabecera es una gran fracción del tamaño total del paquete. Por ejemplo, un paquete de voz es

normalmente de un tamaño de 20 a 60 bytes, mientras que la cabecera asociada es de 40 bytes. Ya que las cabeceras, que contienen la dirección IP y puerto origen y destino, el numero de secuencia, los identificadores de protocolos, etcétera, varían muy poco de un paquete a otro dentro de un mismo flujo, es posible comprimirlas para ahorras más del 80 por ciento del ancho de banda (ver tabla 11). Aparte de ahorrar ancho de banda, la compresión de las cabeceras reduce la pérdida de paquetes ya que los paquetes mientras más pequeños, menos pérdidas de errores de bits van a sufrir dado una BER (Bit Error Ratio), y a su vez mejoraran los tiempos de respuesta.

Tipo de protocolo	Tamaño cabecera (bytes)	Tamaño mínimo de cabecera comprimida (bytes)	Ahorro de ancho de banda (%)
IPv4/TCP	40	4	90
IPv4/UDP	28	1	96.4
IPv4/UDP/RTP	40	1	97.5
IPv6/TCP	60	4	93.3
IPv6/UDP	48	3	93.75
IPv6/UDP/RTP	60	3	95

Tabla 11. Ahorro de ancho de banda con la compresión de cabeceras.

La compresión de cabeceras usa el concepto de contexto de flujo, que es una recopilación de información sobre los campos estáticos y dinámicos, y los cambios en la cabecera del paquete. Este contexto lo usan tanto el compresor como el descompresor para conseguir una compresión máxima. Los primeros paquetes de un flujo se envían sin comprimir para construir este contexto en ambos lados, el número de estos paquetes depende de la BER del enlace y del tiempo de ida y vuelta de un paquete. Una vez esta establecido el contexto, lo paquetes comprimidos son enviados con un identificador de contexto prefijado.

Durante muchos años se han desarrollado muchas técnicas de compresión, aquí solo vamos a ver una de ellas que es soportada en WiMAX y se llama, ROHC (Robust Header Compression). Se trata de una técnica complicada, pero que funciona bien bajo condiciones de grandes BER y largos tiempos de ida y vuelta, y puede reducir el tamaño de la cabecera hasta un mínimo de 1 byte. ROCH puede usarse para gran variedad de cabeceras, incluyendo

IP/UDP/RTP para VoIP.

Al principio de un flujo, se envía un mensaje de actualización estático que contiene todos los campos que no se esperan que cambien, como la dirección IP destino y origen. Los campos dinámicos se envían sin comprimir al principio y cuando cambian. De otra manera, los campos dinámicos son enviados comprimidos. ROHC incluye un proceso de recuperación de errores en el descompresor, mostrado en la figura 26. En cada cabecera comprimida se envía un código de redundancia cíclica o CRC que es válido para las cabeceras descomprimidas. Si el CRC falla después de la descompresión, el descompresor trata de interpolar los datos que faltan de las cabeceras previas y lo comprueba de nuevo. Esto se intenta varias veces; si no termina con éxito, se pide una actualización del contexto. Este esquema de recuperación de errores es lo que hace que ROHC sea robusto. ROCH es reconocido como una pieza importante en cada red IP inalámbrica, y el IETF sigue investigando y añadiendo mejoras.

Una consecuencia negativa de usar la compresión de cabeceras en un enlace aéreo es que el ancho de banda requerido de una aplicación es diferente si va sobre el aire o sobre el resto de la red. Esto hace que sea difícil para una aplicación hacer la petición de ancho de banda de extremo a extremo en los parámetros de calidad de servicio.

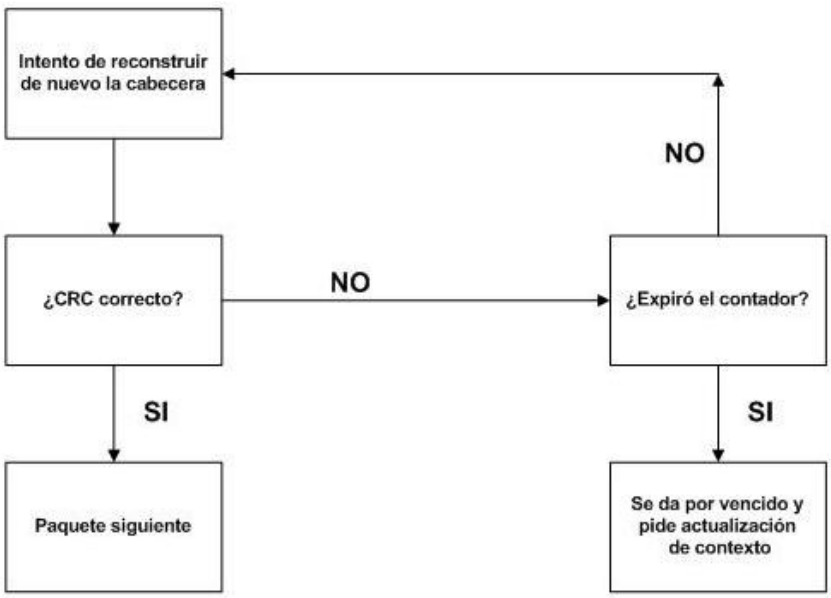

Figura 26. Proceso de recuperación del descompresor de ROHC.

3.6. Conclusiones

En este capítulo hemos visto una serie de aspectos de las redes inalámbricas de banda ancha, vamos a resumirlas a continuación:

Hay dos tipos de QoS: una basada en la gestión por flujo y otra basada en la gestión por agregación. La gestión por flujo ofrece mejor QoS pero tiene problemas de escalabilidad. La mayoría de las redes IP hoy en día confían en la gestión por agregación.

El IETF ha desarrollado numerosas arquitecturas y protocolos para proveer QoS en una red IP. Tres de las mayores tecnologías de QoS de IP son IntServ, DiffServ, y MPLS.

SIP es un protocolo simple, flexible y un potente protocolo que rápidamente se ha establecido como el protocolo elegido para el control de sesiones multimedia en redes IP.

El diseño de redes inalámbricas debe incluir soporte para los mecanismos básicos de seguridad, como encriptación, autentificación, y control de acceso. El IEEE 802.16e-2005 conjuntamente con la arquitectura WiMAX, soporta mecanismos de seguridad robustos y flexibles.

Para soportar usuarios en movimiento, las redes inalámbricas deben incorporar mecanismos de gestión de la localización y gestión de traspaso de una estación base a otra. Desarrollar buenos mecanismos de traspaso es fundamental para las redes móviles.

Aparte de los mecanismos de la capa física y la capa MAC para soportar los traspasos, hay que desarrollar Mobile IP para transferir sesiones en curso de una subred a otra en las redes WiMAX.

TCP no fue diseñado para ejecutarse sobre enlaces ruidosos y con limitación de ancho de banda, por lo que actúa pobremente sobre enlaces inalámbricos Hay bastantes soluciones a estos problemas.

La compresión de cabeceras puede mejorar la eficiencia de la tasa de transferencia de los enlaces inalámbricos que tienen un ancho de banda limitado. El estándar WiMAX soporta una compresión de cabeceras bastante robusta.