CAPÍTULO 1: ADQUISICIÓN DE CONOCIMIENTOS BÁSICOS

Voz sobre Protocolo de Internet, también llamado Voz sobre IP, VozIP, **VoIP** (por sus siglas en inglés), es un grupo de recursos que hacen posible que la señal de voz viaje a través de Internet empleando un **protocolo IP** (Internet Protocol). Esto significa que se envía la señal de voz en forma digital en paquetes en lugar de enviarla (en forma digital o analógica) a través de circuitos utilizables sólo para telefonía como una compañía telefónica convencional o PSTN (*Public Switched Telephone Network*).

Los protocolos que son usados para llevar las señales de voz sobre la red IP son comúnmente referidos como protocolos de Voz sobre IP o protocolos IP. Pueden ser vistos como implementaciones comerciales de la "Red experimental de Protocolo de Voz" (1973), inventada por ARPANET [1].

El tráfico de Voz sobre IP puede circular por cualquier red IP, incluyendo aquellas conectadas a Internet, como por ejemplo redes de área local (LAN).

Es muy importante diferenciar entre Voz sobre IP (VoIP) y Telefonía sobre IP.

- **VoIP** es el conjunto de normas, dispositivos, protocolos, en definitiva *la tecnología* que permite la transmisión de la voz sobre el protocolo IP.
- **Telefonía sobre IP** es el conjunto de *nuevas funcionalidades* de la telefonía, es decir, en lo que se convierte la telefonía tradicional debido a los servicios que finalmente se pueden llegar a ofrecer gracias a poder portar la voz sobre el protocolo IP en redes de datos.

La principal ventaja de este tipo de servicios es que **evita los altos costes** de telefonía (principalmente de larga distancia) que son usuales de las compañías de la Red Pública Telefónica Conmutada. Algunos ahorros en el costo son debidos a utilizar una **misma red para llevar voz y datos**, especialmente cuando los usuarios tienen sin utilizar toda la capacidad de una red ya existente en la cual pueden usar para VoIP sin un costo adicional. Las llamadas de VoIP a VoIP entre cualquier proveedor son generalmente **gratis**, en contraste con las llamadas de VoIP a PSTN que generalmente cuestan al usuario de VoIP.

El desarrollo de **codecs** para VoIP (aLaw, g.729, g.723, etc.) ha permitido que la voz se codifique en paquetes de datos de cada vez menor tamaño. Esto deriva en que las comunicaciones de voz sobre IP requieran anchos de banda muy reducidos. Junto con el avance permanente de las conexiones ADSL en el mercado residencial, éste tipo de comunicaciones, están siendo muy populares para llamadas internacionales.

La calidad de los sistemas de VoIP es a veces un problema con el que nos solemos encontrar. No escuchamos bien, la comunicación se entrecorta, etc. **Los problemas son muchas veces inherentes a la utilización de la red** (Internet y su velocidad y ancho

de banda). Pero si conocemos las causas que pueden producir estos problemas quizás podamos mejorar la calidad.

Por otro lado, debido al acierto en el mercado de VoIP, muchas fueron las empresas que se interesaron por la reducción de pérdidas de telecomunicaciones a causa de mover el tráfico de voz a redes de paquetes. Mientras estas redes de telefonía de paquetes y las dependencias de interconexió aparecían, llegó a ser claro que la industria necesitaba protocolos VoIP estándares. Cuatro son los diferentes protocolos de control de llamadas y señalización para VoIP:

- H.323
- Protocolo de Control Gateway Media (MGCP)
- Protocolo de Iniciación de Sesión (SIP)
- Control Gateway Media / H.248 (MEGACO)

1. Elementos de un Red VoIP

Una infraestructura de VoIP puede tener varias alternativas, dependiendo del tipo de red en que se desarrollen y las funcionalidades que se ofrezcan. A continuación se detalla un esquema VoIP completo, capaz de establecer llamadas en los distintos tipos de redes posibles, de forma que podamos tener un conocimiento más amplio de los distintos elementos que pueden intervenir. Éste esquema se basa en el uso del protocolo H.323, aunque funcionalmente es equivalente al resto de protocolos como SIP,…

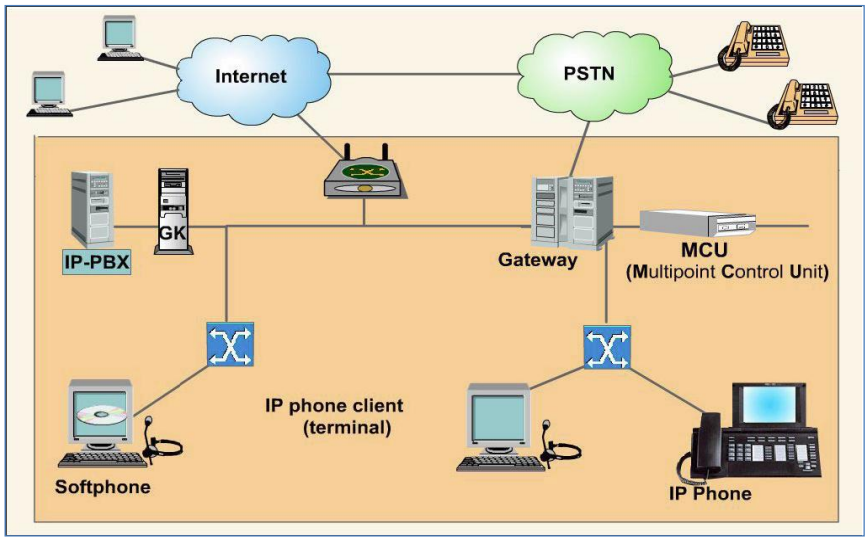

Figura 1: Elementos de una red VoIP
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

1.1.Terminal

El usuario final necesita un teléfono IP, también conocido como **Terminal.** Aunque VoIP implica la transmisión digital de voz en paquetes, el propio teléfono puede ser analógico, como los que usamos habitualmente, o digital, es decir, un teléfono IP. La voz puede ser digitalizada tanto antes como durante la **paquetización** (proceso en el que porciones del flujo de bits son agrupadas en paquetes IP), produciéndose siempre el proceso de **codificación** en el terminal llamante, y la **decodificación** en el terminal que recibe la llamada. A continuación haremos mención especial al papel que desempeñan los codecs para la realización de los procesos anteriores.

1.1.1. CODECS: Funcionamiento y características básicas

1.1.1.1. Introducción

Con respecto a la voz sobre IP, un códec es un algoritmo utilizado para **codificar y decodificar la conversación de voz** [2]. Puesto que la voz y el sonido que escuchamos se transmiten como una señal analógica, tiene que ser convertida (o codificada) a un formato digital apto para su transmisión a través de Internet. Una vez en el otro extremo, necesita ser decodificada de nuevo para la otra persona pueda oír lo que se está diciendo. Hay una variedad de maneras diferentes en los que se puede hacer esta codificación y decodificación - muchos utilizan la compresión para reducir el ancho de banda requerido de la conversación. La cantidad de ancho de banda suele ser directamente proporcional a la calidad de los datos transmitidos. Es importante recordar que según la codificación que hagamos, podemos introducir más o menos retardo en la conversación. Así, debemos tener en cuenta que no sólo tenemos que tener una buena calidad compresión, sino que hay que ser capaz de hacer la codificación y decodificación en una cantidad mínima de tiempo.

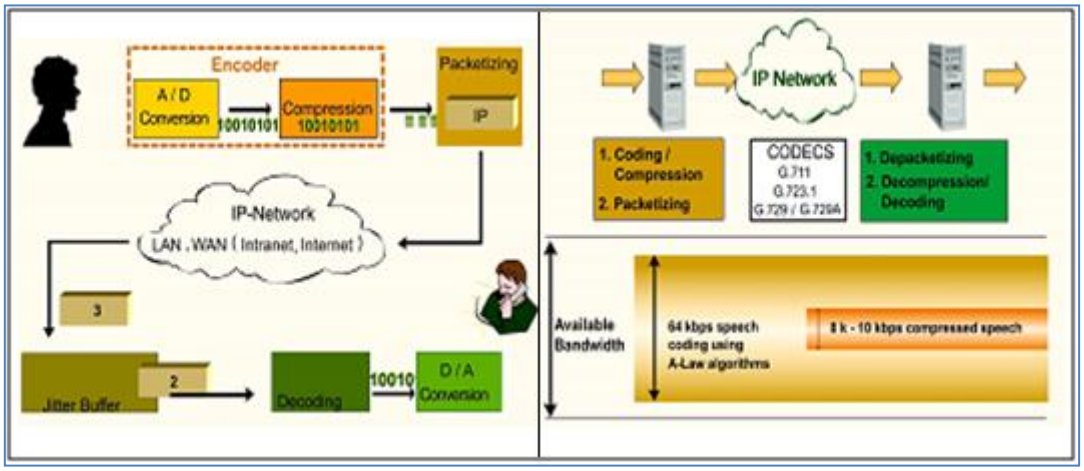

Figura 2: Codificación y decodificación
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

Es importante saber que los diferentes clientes de VoIP soportan diferentes codecs, y cada proveedor de VoIP sólo admitirá un subconjunto de codecs también. En general, cuando se establece una llamada VoIP, se tendrá que utilizar un códec que ambas partes y el proveedor soporten. Este tipo de negociación se maneja automáticamente, pero conociendo los detalles nos permitirá forzar o fomentar la utilización de ciertos codecs.

Características básicas

A continuación comentamos las características más importantes que diferencian a unos codecs de otros.

Ancho de Banda

- Los valores del ancho de banda representan en la mayoría de los datos la **carga de los paquetes IP**.
- Estos valores nos indican el ancho de banda en cada dirección, no la suma del de subida y de bajada.
- Asumen transmisión continua de voz en ambas direcciones sin supresión de silencio.
- El “ancho de banda nominal” indica el ancho de banda Ethernet típico que podemos esperar de un códec a usar.

Frecuencia de muestreo

La frecuencia de muestreo es la tasa con la que se muestrea la señal analógica. El **Teorema de Nyquist** establece que para registrar una cierta frecuencia, el muestreo debe ser al menos el doble de esa frecuencia. Por tanto, cuanto mayor es la frecuencia de muestreo, mayor es el rango de frecuencias en el flujo de audio codificado. El oído humano es capaz de escuchar desde 20 Hz hasta 20.000 Hz. Normalmente, la voz está entre unos 100-4,000 Hz. Por lo tanto, necesitaremos al menos una frecuencia de muestreo de unos 8kHz para codificar correctamente la voz humana. Tasas de muestreo mayores capturarán frecuencias más altas, pero también aumentará el ancho de banda, ya que hay más muestras para codificar y transmitir.

Tamaño de carga útil

El tamaño de la carga útil de cada paquete de voz codificada influye en dos cosas: retardo y ancho de banda. Cada paquete codificado que se envía incurre en gastos generales de ancho de banda (debido a IP y otras cabeceras añadidas a los datos en la red). Por lo tanto, mayores cargas útiles incurren en sobrecargas proporcionalmente menores, reduciendo así la utilización de ancho de banda nominal. Sin embargo, mediante el uso de cargas útiles más grandes, más audio se requiere para construir un solo paquete, que a su vez aumenta la cantidad de tiempo que tarda el paquete en ir desde el principio hasta llegar al otro extremo y ser decodificado, lo que aumenta el retraso en la conversación. Se trata del típico "*trade-off*" de VoIP. La mayoría de los codecs usan tamaños de carga útil de 10-40ms. (2)

1.1.1.2. Tipos de codecs

A continuación se muestra una **tabla resumen** [3] con los codecs más utilizados actualmente:

- **Bit Rate**: indica la cantidad de información que se manda por segundo.
- **Sampling Rate:** es la frecuencia de muestreo de la señal vocal (cada cuánto se toma una muestra de la señal analógica).
- **Frame size:** indica cada cuántos milisegundos se envía un paquete con la información sonora.

- **MOS:** indica la calidad general del códec (valor de 1 a 5).

Nombre	Estandarizado	Descripción	Bit rate (kb/s)	Sampling rate (kHz)	Frame size (ms)	Observaciones	MOS (Mean Opinion Score)
G.711	ITU-T	Pulse code modulation (PCM)	64	8	Muestreada	Tiene dos versiones u-law (US, Japan) y a-law (Europa) para muestrear la señal	4.1
G.721	ITU-T	Adaptive differential pulse code modulation (ADPCM)	32	8	Muestreada	Obsoleta. S e ha transformado en la G.726.	
G.722	ITU-T	7 kHz audio-coding within 64 kbit/s	64	16	Muestreada	Divide los 16 Khz en dos bandas cada una usando ADPCM	
G.722.1	ITU-T	Codificación a 24 y 32 kbit/s para sistemas sin manos con baja perdida de paquetes	24/32	16	20		
G.723	ITU-T	Extensión de la norma G.721 a 24 y 40 kbit/s para aplicaciones en circuitos digitales.	24/40	8	Muestreada	Obsoleta por G.726. Es totalmente diferente de G.723.1.	
G.723.1	ITU-T	Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s	5.6/6.3	8	30	Parte de H.324 video conferencing. Codifica la señal usando linear predictive analysis-by-synthesis coding. Para el codificador de high rate utiliza Multipulse Maximum Likelihood Quantization (MP-MLQ) y para el de low-rate usa Algebraic-Code-Excited Linear-Prediction (ACELP).	3.8-3.9
G.729	ITU-T	Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)	8	8	10	Bajo retardo (15 ms)	3.92
GSM 06.10	ETSI	RegularPulse Excitation LongTerm Predictor (RPE-LTP)	13	8	22.5	Usado por la tecnología celular GSM	
LPC10	Gobierno de USA	Linear-predictive codec	2.4	8	22.5	10 coeficientes.La voz suena un poco "robotica"	
Speex			8, 16, 32	2.15-24.6 (NB) 4-44.2 (WB)	30 (NB) 34 (WB)		
iLBC			8	13.3	30		
DoD CELP	American Department of Defense (DoD) Gobierno de USA		4.8		30		
DVI	Interactive Multimedia Association (IMA)	DVI4 uses an adaptive delta pulse code modulation (ADPCM)	32	Variable	Muestreada		

Tabla 1: Resumen características codecs

1.2. Servidor IP-PBX

El elemento central es **servidor IP-PBX** (Public Branch Exchange Server) junto con un Gatekeeper si el sistema es H.323, un servidor SIP si el sistema es SIP o un Call Agent si el sistema es MEGACO. Ambos son responsables de la gestión de solicitud de conexión y conmutación.

1.3. Gateway

Es un dispositivo que traduce un protocolo a otro. Se necesita un **Gateway** para establecer una conexión orientada a paquetes entre los elementos de la red IP y la Red Pública Telefónica Conmutada, PSTN.

Podemos considerarlo como una caja que por un lado tiene un interface LAN y por el otro dispone de uno o varios de los siguientes interfaces:

- FXO. Para conexión a extensiones de centralitas ó a la red telefónica básica.
- FXS. Para conexión a enlaces de centralitas o a teléfonos analógicos.
- E&M. Para conexión específica a centralitas.
- BRI. Acceso básico RDSI (2B+D)
- PRI. Acceso primario RDSI (30B+D)
- G703/G.704. (E&M digital) Conexión especifica a centralitas a 2 Mbps.

Existen dos protocolos de control de gateways:

- MGCP: Media Gateway Control Protocol (IETF). Utilizado por los controladores de gateways para establecimiento, control y término de las llamadas.
- MEGACO: Media Gateway Control Protocol (IETF/UIT-T). Tiene la misma finalidad del MCGP, sin embargo fue desarrollado para ser una alternativa a ese protocolo, adecuándose también a controladores distribuidos de gateways, a controladores multipunto (conferencia) y a unidades interactivas de respuesta audible.

1.4. Red IP

La **Red IP** proporciona conectividad entre todos los terminales. Ésta puede ser una red IP privada, una intranet o Internet.

En el tránsito de la red IP se producen los principales problemas de calidad de servicio en VoIP[1] y vienen derivados de dos factores principalmente:

- **Internet es un sistema basado en conmutación de paquetes** y por tanto la información no viaja siempre por el mismo camino a lo largo de la red IP. Esto produce efectos como la pérdida de paquetes o el jitter.
- **Las comunicaciones VoIP son en tiempo real,** lo que produce que efectos como el eco, la pérdida de paquetes y el retardo o la latencia sean muy molestos y perjudiciales y deban ser evitados.

2. Protocolos de señalización

Un protocolo es un conjunto de reglas y acuerdos que los computadores y dispositivos deben seguir para que puedan comunicarse entre ellos. Más concretamente, un protocolo de señalización es el que se encarga de gestionar los mensajes y procedimientos utilizados para establecer una comunicación. Los más usados, y por tanto, en los que nos centraremos son el protocolo **H.323** y el **SIP (Seccion Initiation Protocol).**

2.1. SIP (Session Initiation Protocol)

2.1.1. Arquitectura y elementos funcionales

El protocolo SIP (Session Initiation Protocol) fue desarrollado por el grupo MMUSIC (Multimedia Session Control) del IETF, definiendo una arquitectura de señalización y control para VoIP. Inicialmente fue publicado en febrero del 1996 en la RFC 2543 [4], ahora obsoleta con la publicación de la nueva versión RFC 3261 que se publicó en junio del 2002.

SIP fue diseñado de acuerdo al modelo de Internet. Es un protocolo de señalización extremo a extremo que implica que toda la lógica es almacenada

[1] En el apartado "*3. QoS: Parámetros que afectan y mecanismos de medición*", se detallarán con más detalle los factores que afectan a la calidad del servicio mencionados en los párrafos anteriores.

en los dispositivos finales (salvo el enrutado de los mensajes SIP). El estado de la conexión es también almacenado en los dispositivos finales**.** El precio a pagar por esta capacidad de distribución y su gran escalabilidad es **una sobrecarga en la cabecera de los mensajes** producto de tener que mandar toda la información entre los dispositivos finales.

SIP es un protocolo de señalización a nivel de aplicación para establecimiento y gestión de sesiones con múltiples participantes. Se basa en mensajes de petición y respuesta y reutiliza muchos conceptos de estándares anteriores como **HTTP** y **SMTP**.

2.1.2. Protocolos

El propósito de SIP es la comunicación entre dispositivos multimedia. SIP hace posible esta comunicación gracias a dos protocolos:

- **RTP/RTCP(Real-Time Transport Protocol / Real-Time Transport Control Protocol):** El protocolo RTP se usa para transportar los datos de voz en tiempo real (igual que para el protocolo H.323)
- **SDP:** Se usa para la negociación de las capacidades de los participantes, tipo de codificación, etc.)

2.1.3. Señalización

Se trata de un protocolo basado en **petición-respuesta** (request-response), al igual que HTTP o SMTP. SIP maneja mensajes de petición que se estructuran en tres bloques:

- Request Line
- Cabecera
- Cuerpo

Y mensajes de respuesta, también en tres:

- Status Line
- Cabecera
- Cuerpo

En ambos casos el cuerpo es independiente de SIP y puede contener cualquier cosa. A efectos de estandarización se definen métodos para describir las áreas de especificación. SIP define los siguientes métodos:

- **INVITE**: este método es usado para establecer sesiones y anunciar las capacidades de los nodos SIP.
- **ACK**: es usado para confirmar que el cliente solicitante ha recibido una respuesta final desde un servidor a una solicitud INVITE, reconociendo la respuesta como afirmativa.
- **OPTIONS**: es usado para preguntar a un nodo SIP por sus capacidades, sin que ningún canal multimedia haya sido establecido aún.
- **BYE**: cuando la llamada es completada, es decir, cuando alguno de los extremos involucrados en la comunicación desea finalizar la llamada.
- **CANCEL:** cancela una solicitud pendiente, pero no afecta a una solicitud ya completada. Este método finaliza una solicitud de llamada incompleta.
- **REGISTER:** notifica al servidor SIP en qué terminal SIP puede ser alcanzado un usuario.
- **INFO**: es usado para transmitir señales de aplicación de telefonía a través del canal usado por la señalización SIP (por ejemplo dígitos marcados).
- **PRACK:** este método es usado en lugar de ACK para notificar al otro extremo que se está estableciendo una llamada.
- **SUBSCRIBE:** este método provee una forma de establecer manejadores de eventos dentro de aplicaciones de telefonía SIP.
- **NOTIFY:** este método entrega mensajes entre extremos SIP, tales como eventos ocurridos durante la llamada.

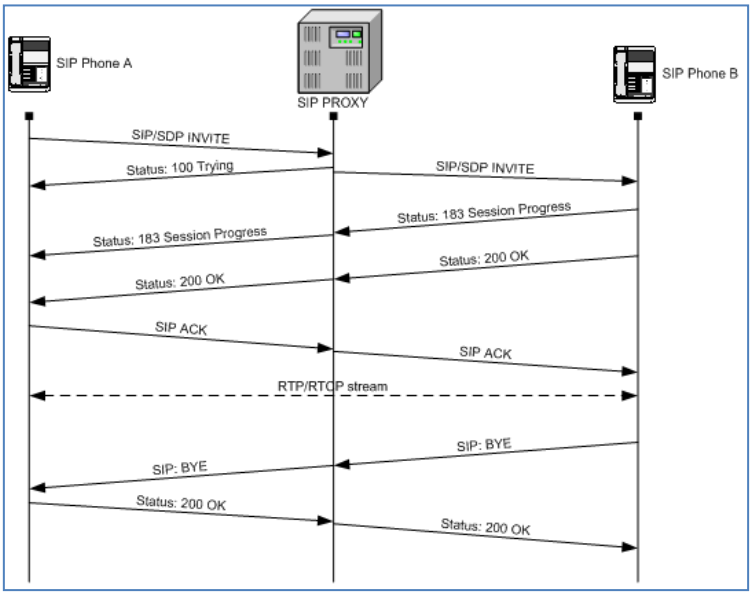

Figura 3: Establecimiento de llamada SIP.
(Fuente: http://messenger.es/sip_445)

Las respuestas son del tipo HTTP:

- 1xx Informational (100 Trying, 180 Ringing, 181 Call is being forwarded).
- 2xx Successful (200 OK, 202 Accepted).
- 3xx Redirection (300 Multiple choices, 301 Moved Permanently, 302 Moved Temporarily).
- 4xx Client Error (400 Bad Request, 404 Not Found, 482 Loop Detected, 486 Busy here).
- 5xx Server Failure (500 Server Internal Error, 501 Not Implemented).
- 6xx Global Failure (600 Busy Everywhere, 603 Decline).

Está pensado para ser independiente de los niveles inferiores; sólo necesita un servicio de datagramas no fiable, con lo cual se puede montar sobre UDP o TCP. Sobre ese servicio no fiable se monta un transporte con RTP/RTCP.

2.2. H.323

2.2.1. Arquitectura y elementos funcionales

H.323 fue el primer estándar internacional de comunicaciones multimedia, que facilitaba la convergencia de voz, video y datos. Fue inicialmente construido para las redes basadas en conmutación de paquetes, en las cuales encontró su fortaleza al integrarse con las redes IP, siendo un protocolo muy utilizado en VoIP.

Los diseñadores de H.323 lo definieron de tal manera que las empresas que manufacturan los equipos pueden agregar sus propias especificaciones al protocolo y pueden definir otras estructuras de estándares que permiten a los dispositivos adquirir nuevas clases de características o capacidades.

H.323 establece los estándares para la compresión y descompresión de audio y vídeo, asegurando que los equipos de distintos fabricantes se intercomuniquen. La norma H.323 hace uso de los procedimientos de señalización de los canales lógicos contenidos en la norma H.245. Un punto importante es que se deben determinar las capacidades de los sistemas, de forma que no se permita la transmisión de datos si no pueden ser gestionados por el receptor.

Los componentes más relevantes que define H.323 serían:

- Terminal
- Gateway
- Gatekeeper
- Unidad de Control Multipunto
- Controlador Multipunto
- Procesador Multipunto
- Proxy H.323

2.2.2. Protocolos

A continuación se explican los protocolos más significativos para H.323:

- **RTP/RTCP (Real-Time Transport Protocol / Real-Time Transport Control Protocol)**: Protocolos de transporte en tiempo real que proporcionan servicios de entrega punto a punto de datos.

- **RAS (Registration, Admission and Status):** sirve para registrar, control de admisión, control del ancho de banda, estado y desconexión de los participantes.
- **H225.0:** protocolo de control de llamada que permite establecer una conexión y una desconexión.
- **H.245:** protocolo de control usado en el establecimiento y control de una llamada.

  En concreto presenta las siguientes funcionalidades:

  - Intercambio de capacidades: Los terminales definen los codecs de los que disponen y se lo comunican al otro extremo de la comunicación.
  - Apertura y cierre de canales lógicos: Los canales de audio y video H.323 son punto a punto y unidireccionales. Por lo tanto, en función de las capacidades negociadas, se tendrán que crear como mínimo dos de estos canales. Esto es responsabilidad de H.245.
  - Control de flujo cuando ocurre algún tipo de problema.
  - Multitud de otras pequeñas funciones.
- **Q.931** (Digital Subscriber Signalling): este protocolo se define para la señalización de accesos RDSI básico.
- **RSVP** (Resource ReSerVation Protocol): protocolo de reserva de recursos en la red para cada flujo de información de usuario.
- **T.120:** la recomendación T.120 define un conjunto de protocolos para conferencia de datos.

Entre los codecs que recomienda usar la norma H.323 se encuentran principalmente:

- **G.711:** de los múltiples codecs de audio que pueden implementar los terminales H.323, este es el único obligatorio.
- **H.261y H.263:** los dos codecs de video que propone la recomendación H.323. Sin embargo, se pueden usar otros.

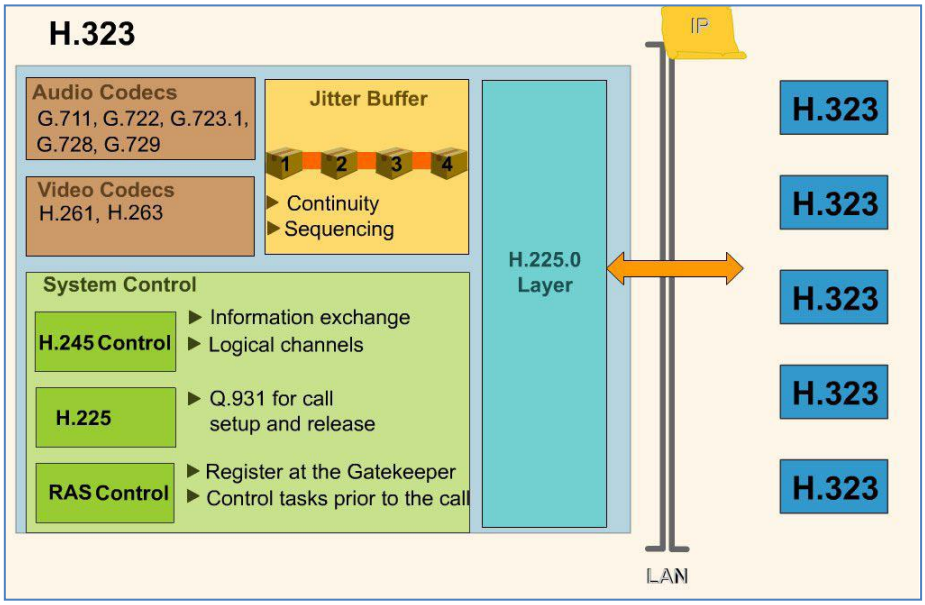

Figura 4: Torre de protocolos H.323
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

2.2.3. Señalización

La función de señalización está basada en la recomendación **H.225**, que especifica el uso y soporte de mensajes de señalización **Q.931/Q932**. Las llamadas son enviadas sobre TCP por el puerto 1720. Sobre este puerto se inician los mensajes de control de llamada Q.931 entre dos terminales para la conexión, mantenimiento y desconexión de llamadas.

Los mensajes más comunes de Q.931/Q.932 usados como mensajes de señalización H.323 son:

- **Setup**: es enviado para iniciar una llamada H.323 para establecer una conexión con una entidad H.323. Entre la información que contiene el mensaje se encuentra la dirección IP, puerto y alias del llamante o la dirección IP y puerto del llamado.
- **Call Proceeding**: enviado por el Gatekeeper a un terminal advirtiendo del intento de establecer una llamada una vez analizado el número llamado.
- **Alerting**: indica el inicio de la fase de generación de tono.
- **Connect**: indica el comienzo de la conexión.
- **Release Complete**: enviado por el terminal para iniciar la desconexión.

- **Facility**: es un mensaje de la norma Q.932 usado como petición o reconocimiento de un servicio suplementario.

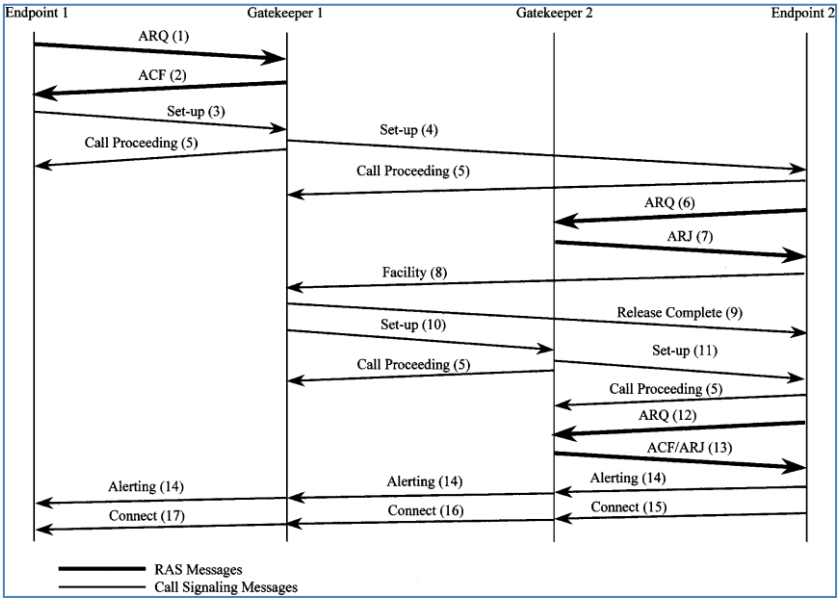

Figura 5: Señalización de llamada entre dos gatekeepers mediante H.323
(Fuente: Bur Goode. Voice Over Internet Protocol)

Función de control H.245

El canal de control H.245 es un conjunto de mensajes ASN.1 usados para el establecimiento y control de una llamada. Unas de las características que se intercambian más relevantes son:

• **MasterSlaveDetermination (MSD):** este mensaje es usado para prevenir conflictos entre dos terminales que quieren iniciar la comunicación. Decide quién actuará de Master y quién de Slave.

• **TerminalCapabilitySet (TCS):** mensaje de intercambio de capacidades soportadas por los terminales que intervienen en una llamada.

• **OpenLogicalChannel (OLC):** mensaje para abrir el canal lógico de información contiene información para permitir la recepción y codificación de los datos. Contiene la información del tipo de datos que serán transportados.

• **CloseLogicalChannel (CLC):** mensaje para cerrar el canal lógico de información.

3. QoS: Parámetros que afectan y mecanismos de medición

3.1.Introducción

El auge de la telefonía IP es algo evidente y la principal razón es el reaprovechamiento de los recursos y la disminución en el coste de llamadas a través de Internet.

Sin embargo, si de algo adolece todavía la VoIP es de la calidad de los sistemas telefónicos tradicionales**.** Los problemas de esta calidad son muchas veces inherentes a la utilización de la red (Internet y su velocidad y ancho de banda) y podrán irse solventando en el futuro. Mientras tanto, cuanto mejor conozcamos los problemas que se producen y sus posibles soluciones mayor calidad disfrutaremos.

Los principales problemas en cuanto a la calidad del servicio (QoS) de una red de VoIP, son la **latencia**, el **jitter**, la **pérdida de paquetes** y el **eco**. En VoIP estos problemas pueden ser resueltos mediante diversas técnicas que se explican en los siguientes apartados.

3.2.Jitter

3.2.1. Causas

El jitter es un efecto de las redes de datos no orientadas a conexión y basadas en conmutación de paquetes. Como la información se discretiza en paquetes, cada uno de los paquetes puede seguir una ruta distinta para llegar al destino.

El jitter se define técnicamente como la **variación en el tiempo en la llegada de los paquetes, causada por congestión de red, perdida de sincronización o por las diferentes rutas seguidas por los paquetes para llegar al destino**.

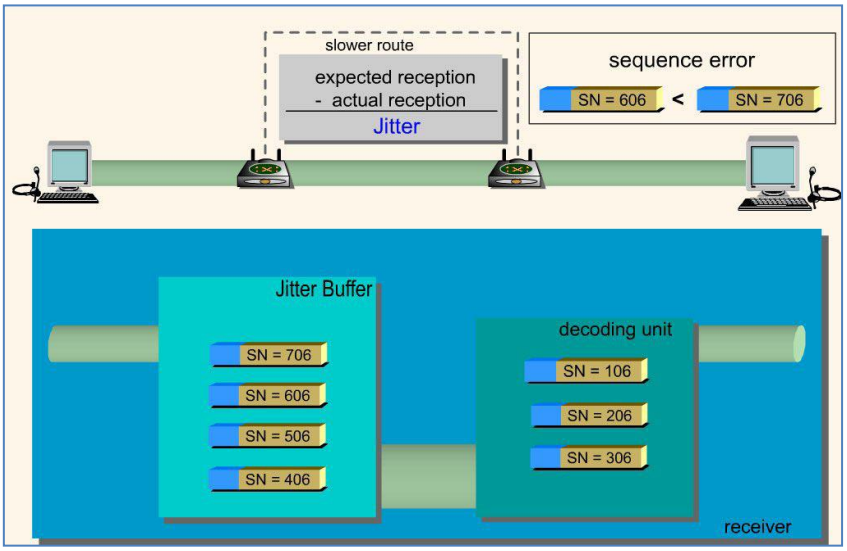

Figura 6: Jitter
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

Las comunicaciones en tiempo real (como VoIP) son especialmente sensibles a este efecto. En general, es un problema frecuente en enlaces lentos o congestionados. Se espera que el aumento de mecanismos de QoS como prioridad en las colas, reserva de ancho de banda o enlaces de mayor velocidad (100Mb Ethernet, E3/T3, SDH) puedan reducir los problemas del jitter en el futuro aunque seguirá siendo un problema por bastante tiempo.

3.2.2. Valores recomendados

El jitter entre el punto inicial y final de la comunicación debiera ser **inferior a 100 ms**. Si el valor es menor a 100 ms el jitter puede ser compensado de manera apropiada. En caso contrario debiera ser minimizado.

3.2.3. Posibles soluciones

La solución más ampliamente adoptada es la utilización del **jitter buffer**. El jitter buffer consiste básicamente en asignar una pequeña cola o almacén para ir recibiendo los paquetes y sirviéndolos con un pequeño retraso. Si algún paquete no está en el buffer (se perdió o no ha llegado todavía) cuando sea necesario se descarta. Normalmente en los teléfonos IP (hardware y software) se pueden modificar los buffers. Un aumento del buffer implica menos perdida de paquetes pero más retraso. Una disminución implica menos retardo pero más pérdida de paquetes.

3.3.Latencia

3.3.1. Causas

A la latencia también se la llama **retardo**. No es un problema específico de las redes no orientadas a conexión y por tanto de la VoIP. Es un problema general de las redes de telecomunicación. Por ejemplo, la latencia en los enlaces vía satélite es muy elevada por las distancias que debe recorrer la información.

La latencia se define técnicamente en VoIP como el **tiempo que tarda un paquete en llegar desde la fuente al destino.**

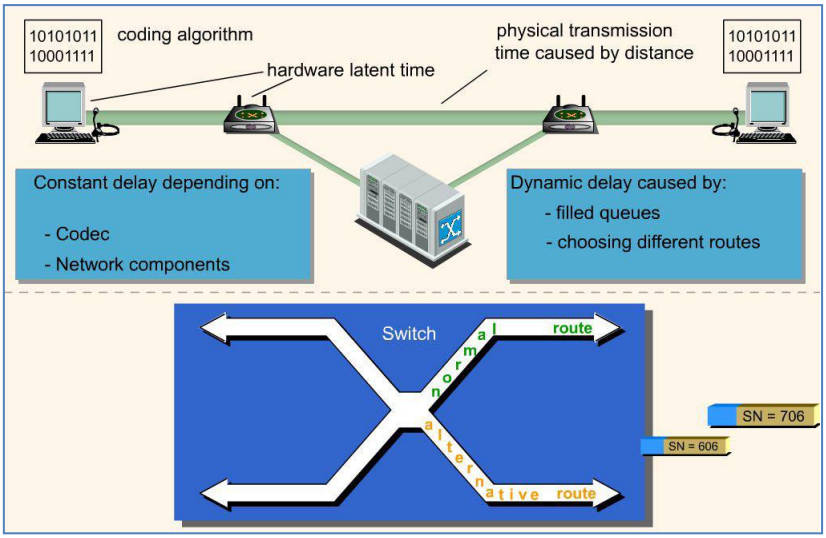

Figura 7: Latencia
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

Las comunicaciones en tiempo real (como VoIP) y full-dúplex son sensibles a este efecto. Es el problema en el que "se nos pisa la comunicación". Al igual que el jitter, es un problema frecuente en enlaces lentos o congestionados.

3.3.2. Valores recomendados

La latencia o retardo entre el punto inicial y final de la comunicación debiera ser **inferior a 150 ms**. El oído humano es capaz de detectar latencias de unos 250 ms, 200 ms en el caso de personas bastante sensibles. Si se supera ese umbral la comunicación se vuelve molesta.

3.3.3. Posibles soluciones

No hay una solución que se pueda implementar de manera sencilla. Muchas veces depende de los equipos por los que pasan los paquetes, es decir, de la red misma. Se puede intentar **reservar un ancho de banda de origen a destino** o **señalizar los paquetes con valores de ToS (Type of Service)** para intentar que los equipos sepan que se trata de tráfico en tiempo real y lo traten con mayor prioridad pero actualmente no suelen ser medidas muy eficaces ya que no disponemos del control de la red.

Si el problema de la latencia está en nuestra propia red interna podemos aumentar el ancho de banda o velocidad del enlace o priorizar esos paquetes dentro de nuestra red.

3.4.Eco

3.4.1. Causas

El eco se produce por un fenómeno técnico que es la conversión de 2 a 4 hilos de los sistemas telefónicos o por un retorno de la señal que se escucha por los altavoces y se cuela de nuevo por el micrófono. El eco también se suele conocer como **reverberación**.

El eco se define como una reflexión retardada de la señal acústica original.

El eco es especialmente molesto cuanto mayor es el retardo y cuanto mayor es su intensidad con lo cual se convierte en un problema en VoIP puesto que los retardos suelen ser mayores que en la red de telefonía tradicional.

3.4.2. Valores recomendados

El oído humano es capaz de detectar el eco cuando su retardo con la señal original es igual o superior a 10 ms. Pero otro factor importante es la intensidad del eco ya que normalmente la señal de vuelta tiene menor potencia que la original. **Es tolerable que llegue a 65 ms y una atenuación de 25 a 30 dB.**

3.4.3. Posibles soluciones

En este caso hay dos posibles soluciones para evitar este efecto tan molesto:

- **Supresores de eco:** Consiste en evitar que la señal emitida sea devuelta convirtiendo por momentos la línea full-dúplex en una línea half-dúplex de tal manera que si se detecta comunicación en un sentido se impide la comunicación en sentido contrario. El tiempo de conmutación de los supresores de eco es muy pequeño. Impide una comunicación full-dúplex plena.

- **Canceladores de eco:** Es el sistema por el cual el dispositivo emisor guarda la información que envía en memoria y es capaz de detectar en la señal de vuelta la misma información (tal vez atenuada y con ruido). El dispositivo filtra esa información y cancela esas componentes de la voz. Requiere mayor tiempo de procesamiento.

3.5.Pérdida de Paquetes

3.5.1. Causas

Las comunicaciones en tiempo real están basadas en el protocolo UDP. Este protocolo no está orientado a conexión y si se produce una pérdida de paquetes no se reenvían. Además la perdida de paquetes también se produce por descartes de paquetes que no llegan a tiempo al receptor.

Sin embargo la voz es bastante predictiva y si se pierden paquetes aislados se puede recomponer la voz de una manera bastante óptima. El problema es mayor cuando se producen pérdidas de paquetes en ráfagas.

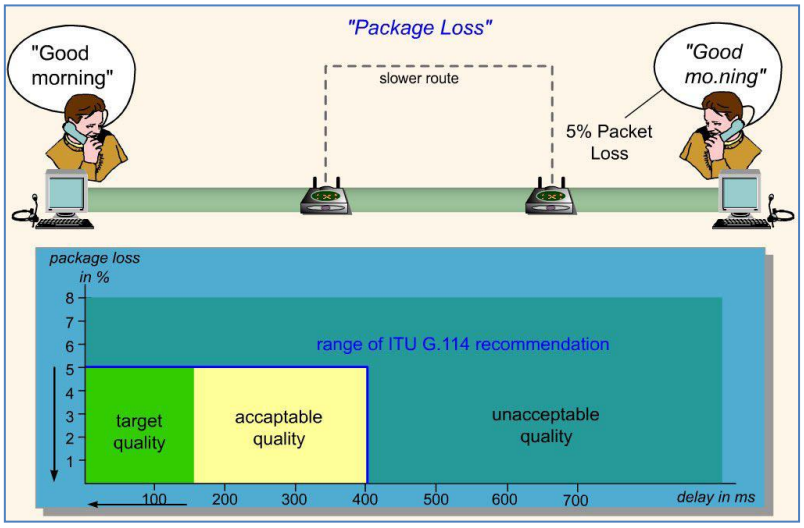

Figura 8: Pérdida de paquetes
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

3.5.2. Valores recomendados

La perdida de paquetes máxima admitida para que no se degrade la comunicación deber ser **inferior al 1%.** Pero es bastante dependiente del códec que se utiliza. Cuanto mayor sea la compresión del códec más pernicioso es el efecto de la pérdida de paquetes. Una pérdida del 1% degrada más la comunicación si se usa el códec G.729 en vez del G.711.

3.5.3. Posibles soluciones

Para evitar la pérdida de paquetes una técnica muy eficaz en redes con congestión o de baja velocidad **es no transmitir los silencios**. Gran parte de las conversaciones están llenas de momentos de silencio. Si sólo transmitimos cuando haya información audible liberamos bastante los enlaces y evitamos fenómenos de congestión.

De todos modos este fenómeno puede estar también bastante relacionado con el jitter y el jitter buffer.

3.6.Pruebas

Para poder comprobar la calidad de vuestro enlace para VoIP existen algunas herramientas y webs muy interesantes que se pueden consultar[2].

4. Elementos de la arquitectura SIP (Session Initiation Protocol)

4.1.Componentes

SIP soporta funcionalidades para el establecimiento y finalización de las sesiones multimedia: localización, disponibilidad, utilización de recursos, y características de negociación.

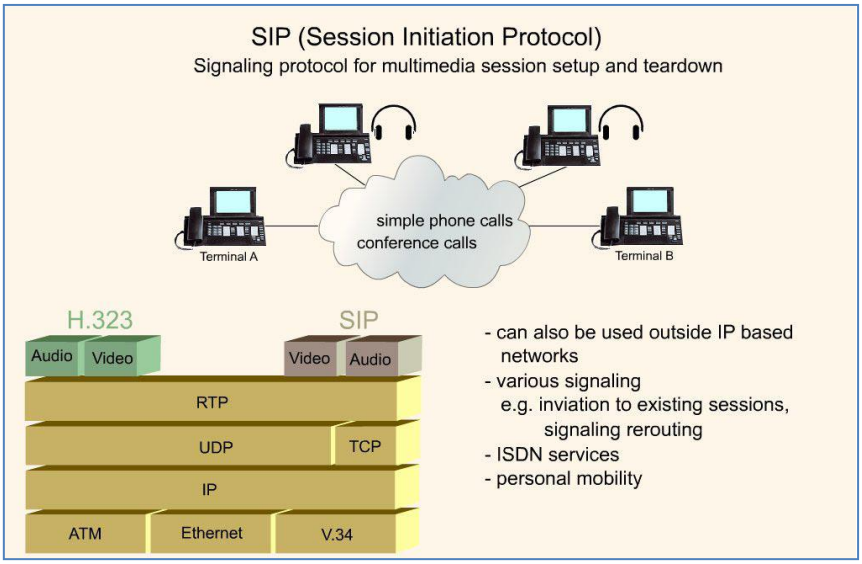

Figura 9: Arquitectura SIP
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

Para implementar estas funcionalidades, existen varios componentes distintos en SIP. Existen dos elementos fundamentales, los **agentes de usuario (UA)** y los **servidores SIP (de registro, proxy o de redirección)**.

4.1.1. Agentes de Usuario

Los usuarios, que pueden ser seres humanos o aplicaciones de software, utilizan para establecer sesiones lo que el protocolo SIP denomina "Agentes de usuario". Estos no son más que los **puntos extremos del protocolo**, es decir,

[2] Para comprobar si existe algún problema en los routers o firewalls que impida progresar las llamadas VoIP se recomiendoa consultar la página web http://www.bandwidth.com/tools/voipTest

son los que emiten y consumen los mensajes del protocolo SIP. Un videoteléfono, un teléfono, un cliente de software (softphone) y cualquier otro dispositivo similar es para el protocolo SIP un agente de usuario. **El protocolo SIP no se ocupa de la interfaz de estos dispositivos con el usuario final, sólo se interesa por los mensajes que estos generan y cómo se comportan al recibir determinados mensajes.**

Los agentes de usuario se comportan como **clientes** (UAC: *User Agent Clients*) y como **servidores** (*UAS: User Agent Servers*). Son UAC cuando realizan una petición y son UAS cuando la reciben. Por esto los agentes de usuario deben implementar un UAC y un UAS.

4.1.2. Servidores de Registro o *Registrar*

El protocolo SIP permite establecer la **ubicación física de un usuario determinado**, esto es, en qué punto de la red está conectado. Para ello se vale del **mecanismo de registro**. Este mecanismo funciona como sigue:

Cada usuario tiene una **dirección lógica** que es invariable respecto de la ubicación física del usuario. Una dirección lógica del protocolo SIP es de la forma *usuario@dominio*, es decir, tiene la misma forma que una dirección de correo electrónico. La dirección física (denominada "dirección de contacto") es dependiente del lugar en donde el usuario está conectado (de su dirección IP). Cuando un usuario inicializa su terminal (por ejemplo conectando su teléfono o abriendo su software de telefonía SIP) el agente de usuario SIP que reside en dicho terminal envía una petición con el método REGISTER a un **Servidor de Registro**, informando a qué dirección física debe asociarse la dirección lógica del usuario. El servidor de registro realiza entonces dicha asociación (denominada *binding*). Esta asociación tiene un período de vigencia y si no es renovada, caduca. También puede terminarse mediante un desregistro. La forma en que dicha asociación es almacenada en la red no es determinada por el protocolo SIP, pero es vital que los elementos de la red SIP accedan a dicha información.

4.1.3. Servidores Proxy y de Redirección

Para encaminar un mensaje entre un agente de usuario cliente y un agente de usuario servidor normalmente se recurre a los servidores. Estos servidores pueden actuar de dos maneras:

- Como **Proxy**, encaminando el mensaje hacia destino.
- Como **Redirector** (*Redirect*), generando una respuesta que indica al origen la dirección del destino o de otro servidor que lo acerque al destino.

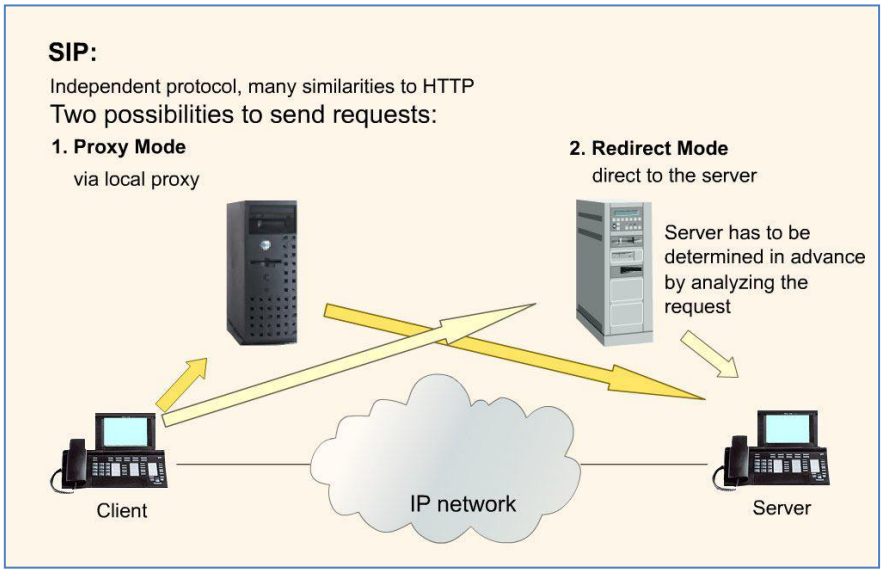

Figura 10: Servidor SIP como Proxy o Redirector
(Fuente: Business Interactive. Web-Based Training VoIP Basics)

La principal diferencia es que el servidor proxy queda formando parte del camino entre el UAC y el (o los) UAS, mientras que el servidor de redirección una vez que indica al UAC cómo encaminar el mensaje ya no interviene más.

Un mismo servidor puede actuar como Redirector o como Proxy dependiendo de la situación.

4.1.3.1. Casos típicos de servidores

Un conjunto de usuarios que pertenecen a una compañía o proveedor de servicios de comunicaciones, conforman un **dominio**. Este dominio, que se indica en una dirección SIP después del carácter "@" es normalmente atendido por un servidor (o más de uno). Este servidor recibe las peticiones hacia sus usuarios. Este servidor será el encargado de determinar la dirección física del usuario llamado. Un servidor que recibe las peticiones destinadas a un dominio específico es denominado **servidor entrante (*Inbound Server*)**.

Es habitual también, que exista un servidor que reciba las peticiones originadas por los usuarios de un dominio hacia otros dominios. Este recibe el nombre de **Servidor Saliente (*Outbound Server*).**

Un agente de usuario normalmente encamina todos sus pedidos hacia un servidor de su propio dominio. Es este quien determina (por sus propios medios o valiéndose de otros servidores) las ubicaciones de los usuarios que son llamados por el agente de usuario en cuestión.

4.2.Cabecera

Las cabeceras se utilizan para transportar información necesaria a las entidades SIP. A continuación, se detallan los campos:

- **Via**: Indica el transporte usado para el envío e identifica la ruta del request, por ello cada proxy añade una línea a este campo.
- **From**: Indica la dirección del origen de la petición.
- **To**: Indica la dirección del destinatario de la petición.
- **Call-Id**: Identificador único para cada llamada y contiene la dirección del host. Debe ser igual para todos los mensajes dentro de una transacción.
- **Cseq:** Se inicia con un número aleatorio e identifica de forma secuencial cada petición.
- **Contact**: Contiene una (o más) dirección que pueden ser usada para contactar con el usuario.
- **User Agent**: Contiene el cliente agente que realiza la comunicación.

Message Header

Via: SIP/2.0/UDP 192.168.0.100:5060;rport;branch=z9hG4bK646464100000007343c52679000020a600000e45
Content-Length: 0
Call-ID: 911D32E5-EEDF-4572-B0B2-61B294636E88@192.168.0.100
CSeq: 1 ACK
From: "Prueba"<sip:20000@miasterisk.com>;tag=8922404614682
Max-Forwards: 70
Route: <sip:20001@192.168.0.1>
To: <sip:20001@miasterisk.com>;tag=as0a27b928
User-Agent: SJphone/1.60.289a (SJ Labs)
Contact: <sip:20100@192.168.0.100:5060>;expires=3600

4.3.Direccionamiento

Una de las funciones de los servidores SIP es la localización de los usuarios y resolución de nombres. Normalmente, el agente de usuario no conoce la dirección IP del destinatario de la llamada, sino su e-mail.

Las entidades SIP identifican a un usuario con las SIP URI (Uniform Resource Identifiers) definido en el RFC 2396[3]. Una SIP URI tiene un formato similar al del e-mail, consta de un usuario y un dominio delimitado por una @, como muestra los siguientes casos:

usuario@dominio, donde dominio es un nombre de dominio completo.
usuario@equipo, donde equipo es el nombre de la máquina.
usuario@dirección_ip, donde dirección_ip es la dirección IP del dispositivo.
número_teléfono@gateway, donde el gateway permite acceder al número de teléfono a través de la red telefónica pública.

La solución de identificación de SIP, también puede ser basada en el DNS descrito en el RFC 3263, donde se describen los procedimientos DNS utilizados por los clientes para traducir una SIP URI en una dirección IP, puerta y protocolo de transporte utilizado, o por los servidores para retornar una respuesta al cliente en caso de que la petición falle.

4.4.RTP-RTCP

RTP es la abreviación de Real-time Transport Protocol, por su denominación en inglés. Es un estándar creado por la IETF para la transmisión confiable de voz y video a través de Internet. La primera versión fue publicada en 1996 en el documento RFC 1889 y fue reemplazado por el estándar RFC 3550 en 2003.

Muy importante en Voz sobre IP, RTP es el **protocolo responsable de la transmisión de los datos**. La digitalización y compresión de la voz y el video es realizada por el códec. Para el manejo de señalización o establecimiento de llamada se usa el protocolo SIP.

Dentro del estándar RFC 3550 se define un protocolo adicional para el envío de datos de control y datos de mediciones realizadas durante la transmisión. Se conoce como **RTCP** *(RTP Control Protocol)*. Los paquetes RTCP se envían periódicamente dentro de la secuencia de paquetes RTP.

[3] Para más información sobre las RFCs consultar http://www.ietf.org/rfc.html

Aunque RTP tiene algunas características de protocolo de nivel de transporte (Según el modelo OSI), es transportado usando UDP. UDP no maneja sesiones ni mecanismos que garanticen la recepción de los paquetes, pero es usado por RTP en lugar de TCP debido a que **reduce el tiempo de envío de los paquetes a través de la red**. En aplicaciones de voz y video es más importante una transmisión rápida que la pérdida de algunos paquetes durante el recorrido.

RTP implementa dos mecanismos principales para garantizar una transmisión de voz: El uso de **número de secuencia** y un **registro de tiempo**. En redes IP es común que los paquetes tomen caminos diferentes para llegar al destino. En aplicaciones de datos esto no es demasiado importante pero para voz y video puede representar una falla detectable por el oído del usuario final. Por esto RTP usa el número de secuencia para reorganizar los paquetes en caso de que lleguen en desorden y el Registro de tiempo es usado para ajustar los intervalos de muestreo de acuerdo a la secuencia original.

RTCP es utilizado para enviar datos de control entre el emisor y receptor de una secuencia RTP. Los paquetes RTCP son enviados aproximadamente cada cinco segundos, y contienen datos que ayudan a verificar las condiciones de transmisión en el extremo remoto.

4.5.SDP

El protocolo SDP (*Session Description Protocol*) RFC 2327 se **utiliza para describir sesiones multicast en tiempo real,** siendo útil para invitaciones, anuncios, y cualquier otra forma de inicio de sesiones.

La propuesta original de SDP fue diseñada para anunciar información necesaria para los participantes y para aplicaciones de multicast MBONE (Multicast Backbone). Actualmente, su uso está extendido para el anuncio y la negociación de las capacidades de una sesión multimedia en Internet.

Puesto que SDP es un protocolo de descripción, los mensajes SDP se pueden transportar mediante distintos protocolos con SIP, SAP, RTSP, correo electrónico con aplicaciones MIME o protocolos como HTTP. Como el SIP, el SDP utiliza la codificación del texto. Un mensaje del SDP se compone de una serie de líneas, denominados campos, dónde los nombres son abreviados por una sola letra, y está en una orden requerida para simplificar el análisis. El SDP no fue diseñado para ser fácilmente extensible.

La única manera de ampliar o de agregar nuevas capacidades al SDP es definir

un nuevo atributo. Sin embargo, los atributos desconocidos pueden ser ignorados.

Un ejemplo de código de sesión SDP:

Session Description Protocol Version (v): 0
Owner/Creator, Session Id (o): Cisco-SIPUA 26425 12433 IN IP4 192.168.0.100
Owner Username: Cisco-SIPUA
Session ID: 26425
Session Version: 12433
Owner Network Type: IN
Owner Address Type: IP4
Owner Address: 192.168.0.100
Session Name (s): SIP Call
Connection Information (c): IN IP4 192.168.0.100
Connection Network Type: IN
Connection Address Type: IP4
Connection Address: 192.168.0.100
Time Description, active time (t): 0 0
Session Start Time: 0
Session Stop Time: 0
Media Description, name and address (m): audio 17338 RTP/AVP 0 8 18 101
Media Type: audio
Media Port: 17338
Media Proto: RTP/AVP
Media Format: ITU-T G.711 PCMU
Media Format: ITU-T G.711 PCMA
Format: ITU-T G.729
Media Format: 101
Media Attribute (a): rtpmap:0 PCMU/8000
Media Attribute (a): rtpmap:8 PCMA/8000
Media Attribute (a): rtpmap:18 G729/8000
Media Attribute (a): rtpmap:101 telephone-event/8000
Media Attribute (a): fmtp:101 0-15

4.6.Ejemplo Comunicación SIP

A continuación se analizará detalladamente una llamada. **En una llamada SIP hay varias transacciones SIP**. Una transacción SIP se realiza mediante un intercambio

de mensajes entre un cliente y un servidor. Consta de varias peticiones y respuestas y para agruparlas en la misma transacción está el parámetro "CSeq".

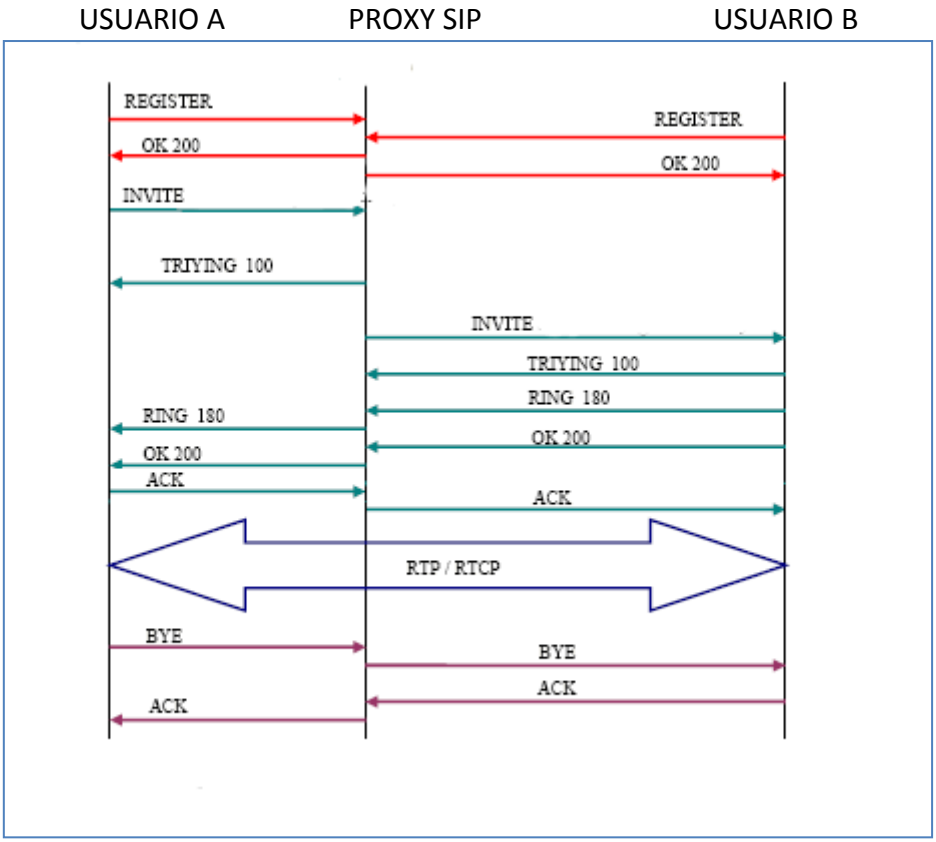

Figura 11: Ejemplo comunicación SIP
(Fuente: http://www.voipforo.com/SIP/SIPejemplo.php)

Las dos primeras transacciones corresponden al **registro de los usuarios**. Los usuarios deben registrarse para poder ser encontrados por otros usuarios. En este caso, los terminales envían una petición REGISTER, donde los campos "from" y "to" corresponden al usuario registrado. El servidor Proxy, que actúa como Register, consulta si el usuario puede ser autenticado y envía un mensaje de OK en caso positivo. La siguiente transacción corresponde a un **establecimiento de sesión**. Esta sesión consiste en una petición INVITE del usuario al proxy. Inmediatamente, el proxy envía un TRYING 100 para parar las retransmisiones y reenvía la petición al usuario B. El usuario B envía un Ringing 180 cuando el teléfono empieza a sonar y también es reenviado por el proxy hacia el usuario A. Por último, el OK 200 corresponde a aceptar la llamada (el usuario B descuelga). **En este momento la llamada está establecida y pasa a funcionar el protocolo de transporte RTP** con los parámetros (puertos, direcciones, codecs, etc.) establecidos en la negociación mediante el protocolo SDP. La última transacción corresponde a una **finalización de sesión**. Esta finalización se lleva a cabo con una única petición BYE enviada al Proxy, y posteriormente reenviada al usuario B.

Este usuario contesta con un OK 200 para confirmar que se ha recibido el mensaje final correctamente.