Estudio de H.323 y SIP

Trabajo realizado por: Abel Sáez Incertis

Introducción a VoIP

La Voz sobre IP (VoIP) abre las puertas a la convergencia de las redes de voz y datos en una única red. VoIP supone una reducción de costes en la instalación de cableado, ofreciendo además la flexibilidad de soportar nuevos servicios como la videoconferencia a través de Internet o la conexión con PCs.
No obstante, VoIP no carece de inconvenientes: Las actuales redes IP, en las que Internet está basada, no garantizan calidad de servicio. Por ello, los distintos protocolos empleados en las comunicaciones VoIP deben ser tolerantes a posibles retrasos o pérdidas de información que se puedan dar, en la medida de lo posible.
En la actualidad destacan dos tecnologías que se emplean para comunicaciones de voz sobre IP: H.323 y SIP, que se describen a continuación.

ITU-T H.323:

H.323 es un estándar creado por el grupo de estudio 16 de la ITU-T para la transmisión de voz, vídeo y datos multimedia a través de redes basadas en conmutación de paquetes sin calidad de servicio (QoS) garantizada, como las redes IP. Inicialmente, H.323 fue diseñado para transportar voz y vídeo en redes de área local, aunque posteriores revisiones del estándar habilitaron su expansión a redes de área amplia como Internet y mejoraron ciertas deficiencias del diseño inicial. H.323 es el estándar que cuenta actualmente con más difusión en el ámbito de la telefonía IP.

La arquitectura de H.323 define todo lo necesario (componentes, protocolos, señalización, códecs...etc) para llevar a cabo la comunicación y garantizar asi la compatibilidad entre dispositivos.

H.323 consta de una serie de protocolos:

- H.225 para el control de llamadas (incluyendo señalización, registro y admisión) y la sincronización y empaquetamiento de flujos de medios.
- H.235 para la seguridad y cifrado
- H.245 para la señalización de control y la apertura/cierre de canales multimedia
- H.450 para los servicios suplementarios
- RTP/RTCP para el transporte de contenido multimedia
- T.120 como protocolo de datos para conferencia multimedia

y una serie de códecs:

- G.711, G.722, G.723, G.728 y G.729 para audio (voz)
- H.261, H.263, H.264 para vídeo

Componentes de las redes H.323:

Las redes basadas en H.323 constan de cuatro tipos de dispositivos lógicos:

Los **terminales** H.323 permiten establecer conferencias bidireccionales de audio y, opcionalmente, vídeo y datos.
Cada terminal H.323 debe, como mínimo, soportar la decodificación de los formatos de audio empleados en las líneas telefónicas tradicionales (leyes a y μ) y la codificación / decodificación de audio según G.711 (PCM, 8KHz@64kbps). El soporte de G.722, G.723.1, G.728 y G.729 es opcional.
El soporte de vídeo no es imprescindible, aunque de existir, debe soportar al menos el códec H.261.
Otras funcionalidades que opcionalmente puede incluir un terminal son las indicadas en el

protocolo T.120 (transferencia de ficheros y pizarra electrónica compartida, entre otras).

Los **gatekeepers** se pueden considerar el punto central en la topología de una red H.323 y definen el concepto de zona H.323:
Una zona es un conjunto de MCUs, gateways y terminales gestionados principalmente por un único gatekeeper; no obstante, en una zona pueden existir gatekeepers secundarios por si el gatekeeper primario fallara.
Los gatekeepers no son necesarios para llamadas entre terminales H.323 dentro de una misma red, aunque sí lo son cuando se desea compatibilidad con las redes de telefonía. Es conveniente el uso de gatekeepers, puesto que proporcionan:

- Control del ancho de banda disponible en la red H.323, con el que el administrador puede limitar el número de conexiones simultáneas y así evitar problemas de congestión en la red que puedan reducir la calidad del servicio.
- Conversión de números de teléfono estándar E164 a direcciones nativas H.323. Esta funcionalidad es imprescindible cuando se pretende establecer comunicación (a través de un gateway) con la red telefónica tradicional.
- Control de admisión de gateways y terminales en una zona H.323, gestionado a través de mensajes H.225.0/RAS.
- Gestión de los elementos (terminales, gateways y MCUs) pertenecientes a la zona.

Opcionalmente, un gatekeeper puede ofrecer la señalización indirecta de llamadas entre terminales H.323 (enrutada a través del gatekeeper), restringir las llamadas que se pueden realizar, e incluso mantener una lista de las llamadas en espera.

La **MCU** o unidad multipunto es un punto final encargado de dar soporte a las conferencias entre tres o más puntos finales H.323. Una MCU consta de un controlador multipunto (MC) y uno o más procesadores multipunto (MP).
Los **MC**s, que también pueden encontrarse integrados en terminales, gateways o gatekeepers, se encargan de transmitir información de los códecs soportados por los distintos terminales para poder así negociar los códecs de audio y vídeo utilizados durante la conferencia.
Los **MP**s, por su parte, distribuyen los flujos de audio / datos / vídeo entre los distintos terminales que participan en una multiconferencia.

Un **gateway** permite conectar una red H.323 con otra red no H.323, como las redes telefónicas SCN basadas en conmutación de circuito. Sus dos funciones básicas son las de traducir los distintos protocolos de establecimiento y fin de llamada empleados por las distintas redes, y realizar la conversión de formatos de audio / vídeo oportuna.

Protocolos especificados por H.323:

H.323 especifica los protocolos que gestionan la preparación, establecimiento, control de estado, mensajería, códecs de audio/video, transferencia de datos, y fin de llamada.
Estos protocolos funcionan sobre un nivel de transporte basado en TCP y UDP y/o (tras la 5ª revisión de H.323) SCTP.
El siguiente gráfico muestra la pila de protocolos H.323:

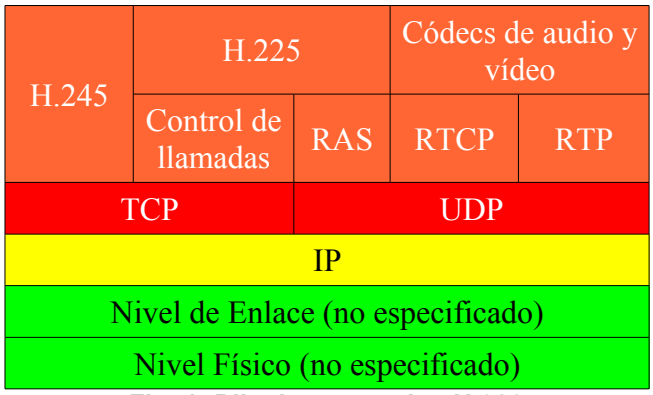

Fig. 1: Pila de protocolos H.323

H.245 es el protocolo de señalización utilizado en el canal de control, que es el único canal que siempre está abierto (en contraposición a los canales de audio y vídeo, que se abren bajo demanda después de la negociación de códecs). Se emplea básicamente para la apertura / cierre de canales lógicos y el intercambio de información sobre la capacidad de transmisión y recepción de medios de los terminales. Otras funciones para las que se utiliza H.245 son:

- Determinar el retraso de ida y vuelta entre ambos extremos de la comunicación.
- Escoger qué punto final actua como maestro y cuál como esclavo (Los papeles de maestro y esclavo sólo se aplican cuando ambos extremos pretenden realizar una acción similar).

Los mensajes de señalización de llamadas están definidos en la recomendación **H.225**, con el formato de mensaje definido en el estándar **Q.931**.
Esta señalización define cómo gestionar datos, vídeo, audio e información de control en una redes basadas en conmutación de paquete. H.225, que se emplea durante el establecimiento de las conexiones entre puntos finales H.323 (gateways y terminales), consta de dos partes: señalización de llamadas y **RAS** (**R**egistro, **A**dmisión y e**S**tado).
El canal de control de llamadas, en redes IP, se establece en el puerto TCP 1720. En este puerto se crean los mensajes necesarios para realizar, mantener y finalizar una llamada. Estos mensajes pueden enviarse directamente entre terminales H.323 (Señalización de llamada Directa) o bien utilizar un gatekeeper como gestor del establecimiento de las conexiones (en este caso se denomina Señalización Enrutada por Gatekeeper).
Por otro lado, el canal H.225.0/RAS es el primer canal que se establece en la comunicación entre un terminal (o gateway) y el gatekeeper.
RAS es el protocolo empleado para:

- Descubrir el / los gatekeepers existentes en la red H.323, tarea que se puede realizar de forma estática (con la dirección del gatekeeper conocida a priori) o dinámicamente, mediante un mensaje de petición dirigido a la dirección multicast 224.0.1.41, que los gatekeepers responden con un mensaje de confirmación.
- Registro de puntos finales (gateways o terminales) en un la zona cubierta por un gatekeeper. Para registrarse, cada punto final debe proporcionarle al gatekeeper su alias y dirección de transporte (IP:puerto en redes IP).
- Localización de puntos finales, que consiste en la traducción de un alias H.323 o número de teléfono E164 en una dirección de transporte (IP:puerto en redes IP).
- Control de admisión de puntos finales en el gatekeeper.
- Notificación de cambios de estado de la conexión o en el ancho de banda disponible.

Una opción interesante que H.225 ofrece es la de encapsular múltiples mensajes H.245 en un mensaje H.225 con el fin de reducir el tiempo de conexión de llamada y sincronizar el control de llamadas con la señalización.

Los protocolos RTP y RTCP, definidos en el RFC 3550, se utilizan para el transporte de medios y el control de transporte de medios, respectivamente. Por requisitos de tiempo real, ambos funcionan sobre el protocolo de transporte UDP (no fiable). La norma especifica que las conexiones RTP se deben establecer en un puerto par y las RTCP en el siguiente puerto impar (p.ej: 19400 para RTP y 19401 para RTCP).
La función principal de **RTCP** (Real Time Control Protocol) es la de monitorizar una conexión RTP para proporcionar información acerca de la calidad del servicio (QoS).
Para ello, obtiene estadísticas acerca de los paquetes enviados / perdidos, el jitter y el retraso de ida y vuelta (RTT) en la conexión, datos que la aplicación puede emplear para realizar ciertos ajustes (p.ej: en la tasa de bits empleada en el códec de audio).
RTP (Real Time Protocol) es el protocolo empleado para transportar flujos de audio y vídeo.

Las características principales de RTP son:

- Soporte unicast y multicast
- Calidad de servicio (QoS) no garantizada - susceptible a pérdida de paquetes
- Identificación de contenido
- Secuenciación (numeración) de paquetes, utilizada para que la aplicación pueda reordenar paquetes que no ha recibido en orden
- Monitorización de la entrega de paquetes

Procedimiento de conexión:

En este apartado se muestran una serie de diagramas en los que se puede apreciar el procedimiento clásico para establecer una conexión de voz entre terminales H.323, dependiendo del tipo de señalización escogido (directa o enrutada a través del gatekeeper) y de si los terminales se encuentran el la misma zona o distintas zonas. No se hace uso de los procedimientos Fast Connect (introducido en la 2ª revisión del estándar H.323) y Extended Fast Connect (introducido en la 5ª revisión), que reducen considerablemente el tiempo de establecimiento.

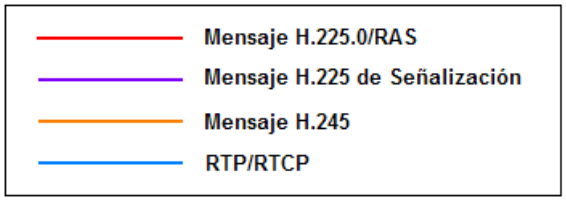

Fig. 3: Tipos de mensaje utilizados en una conversación H.323

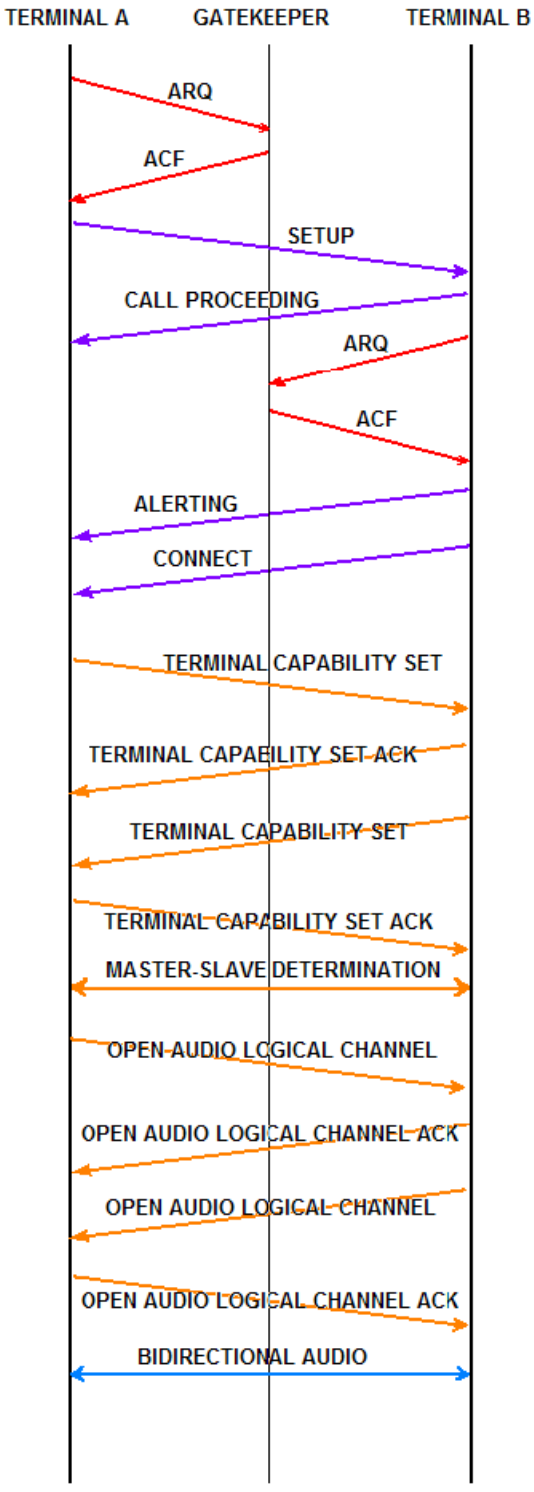

Fig. 2: Señalización directa, terminales en la misma zona

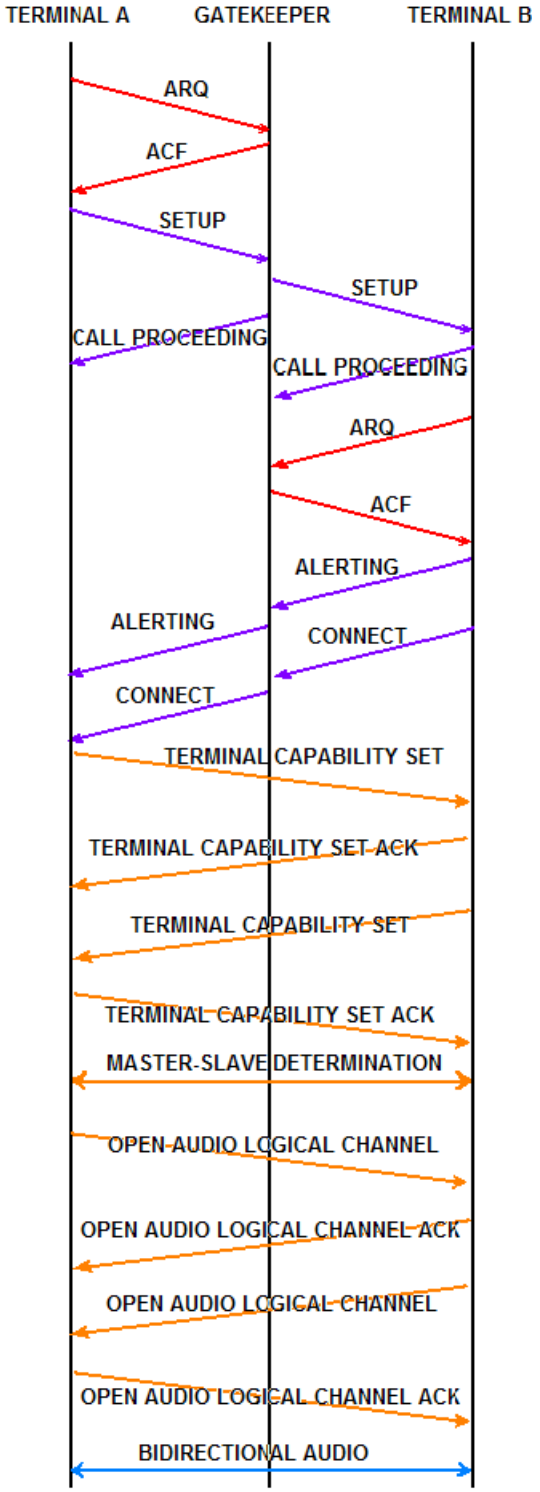

Fig. 3: Señalización enrutada por gatekeeper, terminales en la misma zona

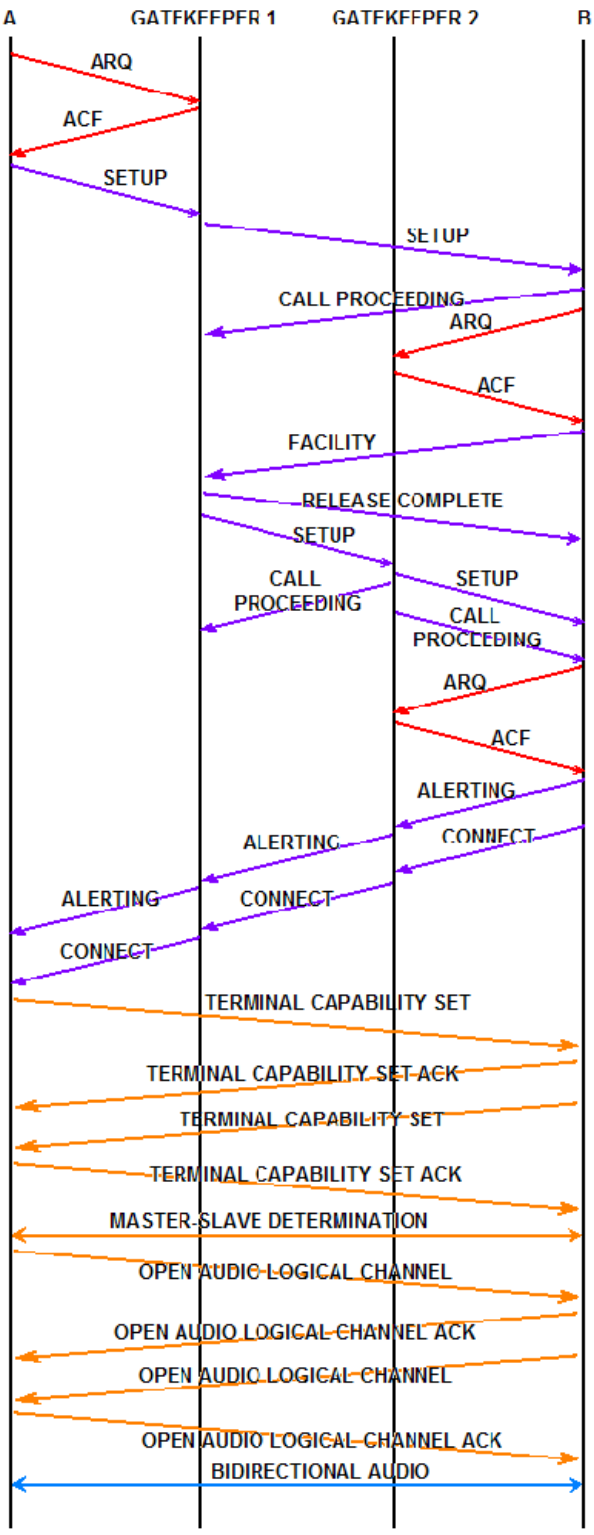

Fig. 4: Señalización directa, terminales en distintas zonas

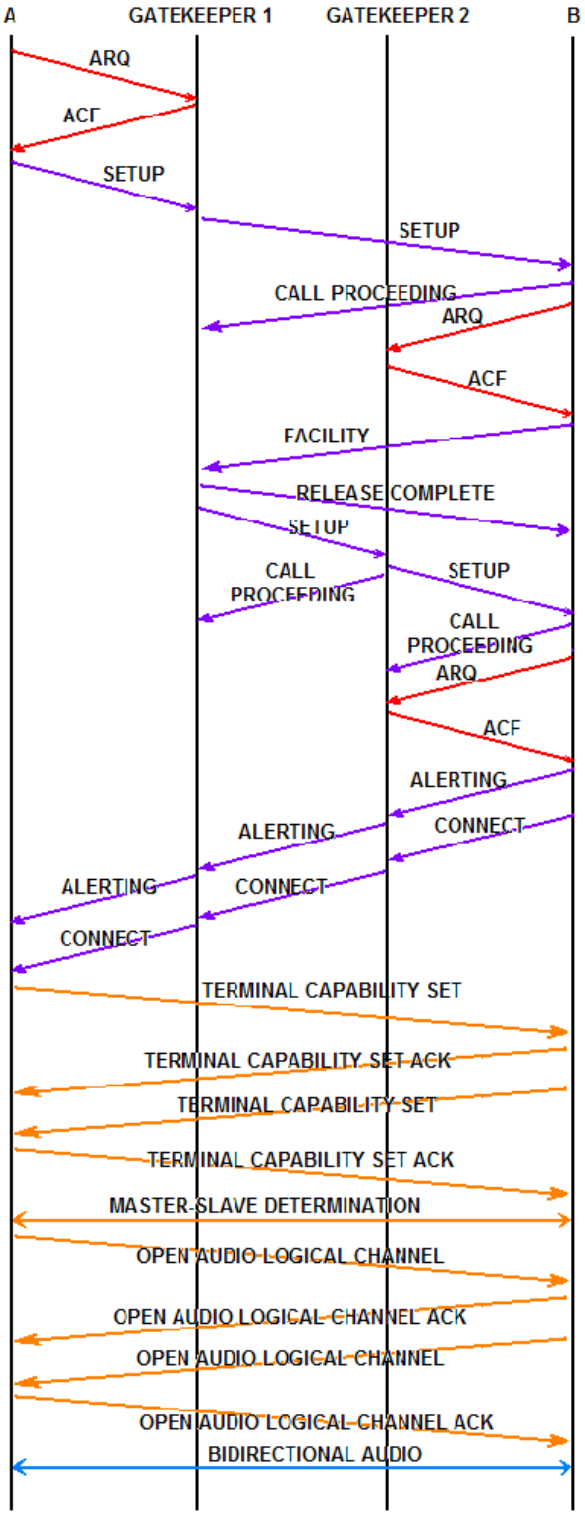

Fig. 5: Señalización enrutada por gatekeeper, terminales en distintas zonas.

Procedimiento de desconexión:

El siguiente gráfico muestra el procedimiento de desconexión entre dos terminales.

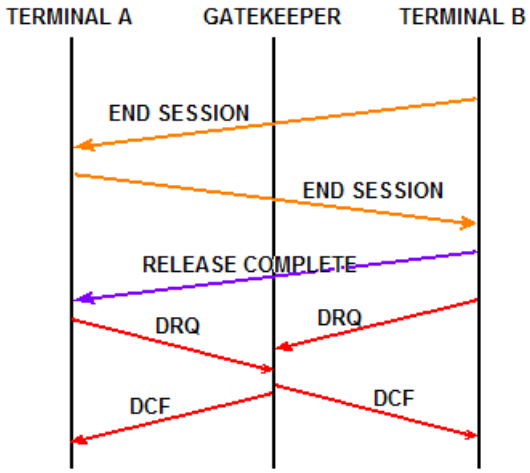

Fig. 6: Desconexión de terminales en una misma zona. Señalización directa

SIP (Session Initiation Protocol) es un protocolo de señalización (situado en el nivel ISO / OSI de aplicación) para el establecimiento, mantenimiento y terminación de sesiones interactivas entre usuarios; estas sesiones pueden tratarse de conferencias multimedia, chat, sesiones de voz o distribución de contenidos multimedia. SIP, creado en 1996 por Mark Handley y Henning Schulzrinne, ha sido estandarizado por la IETF (Internet Engineering Task Force). La especificación más reciente de SIP se puede encontrar en el RFC 3261.
SIP no define por si mismo un sistema de comunicaciones ni ofrece servicio alguno; es un protocolo flexible que se limita a ofrecer una serie de primitivas que las aplicaciones pueden utilizar para implementar servicios.

SIP ofrece funciones tales como:

- Redirección de llamadas
- Resolución de direcciones
- Determinar la disponibilidad de un punto final
- Establecer llamadas punto a punto o multipunto

Componentes SIP:

SIP define cinco componentes lógicos.
Estos componentes se pueden implementar en dispositivos físicos, tal como teléfonos IP, o bien como aplicaciones software; en cualquier caso un mismo dispositivo físico puede incluir uno o más componentes lógicos.
Todos los elementos SIP deben implementar obligatoriamente TCP y UDP. En ambos protocolos escucha en el puerto 5060.

El **Agente de Usuario** es una aplicación con arquitectura cliente / servidor que se utiliza para iniciar y terminar las sesiones.
El cliente usuario-agente (UAC) se encarga de realizar peticiones SIP, mientras que el servidor usuario-agente (UAS) notifica al usuario cuando se recibe una petición y responde a dicha petición dependiendo de la acción tomada por el usuario.

El **Servidor de Redirecciones** acepta una petición SIP y envía una respuesta al cliente que contiene las direcciones de los servidores con los que debe contactar el cliente.

El **Servidor Proxy**, que contiene funciones de servidor y cliente, actúa como un intermediario que realiza peticiones en nombre de otros clientes: para ello interpreta la cabecera del mensaje y la reescribe identificando al proxy como el que inicia la solicitud, recibe la respuesta del destinatario y se la reenvía al cliente.

Un **Servidor de Registro** almacena (o actualiza) en una base de datos la información de contacto del usuario que realiza la petición.

Un **B2BUA (Back to Back User Agent)** es una entidad que recibe una petición INVITE y la procesa como un servidor usuario-agente (UAS). Para determinar la respuesta a la petición, actúa como un cliente usuario-agente que determina cómo responder a la petición y cómo realizar llamadas salientes.
A diferencia de un proxy, un B2BUA debe mantener el estado de la llamada y participar activamente en ella, enviando peticiones y respuestas. Un B2BUA tiene un mayor control de la llamada que un proxy.

Protocolos especificados por SIP:

SIP emplea SDP para descubrir las capacidades multimedia del punto final destino y suele utilizar RTP/RTCP para el transporte de voz.

SDP (Session Description Protocol)
SDP es el protocolo empleado para describir una sesión multimedia, que consiste en un conjunto de flujos de medios (audio, vídeo o datos) que existen durante un determinado tiempo.
Los paquetes SDP contienen (entre otros campos) información acerca del ancho de banda, los protocolos de transporte empleados, los códecs utilizados en la sesión, y la dirección de contacto del iniciador de la sesión.

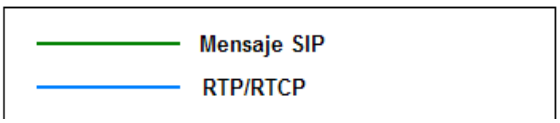

Fig. 7: Tipos de mensaje utilizados en una conversación SIP

Procedimiento de conexión/desconexión

En los siguientes diagramas se pueden apreciar las secuencias de mensajes intercambiados para el establecimientro, transcurso y finalización de una serie de sesiones SIP típicas.

Un detalle interesante es el formato de las direcciones SIP, que son idénticas a las de correo electrónico.

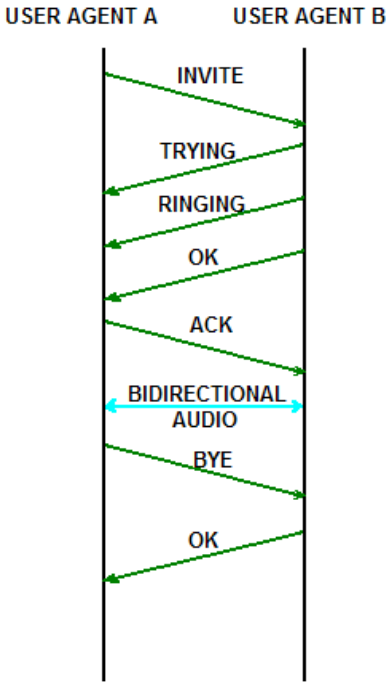

Fig. 8: Sesión SIP de audio directa entre dos agentes de usuario.

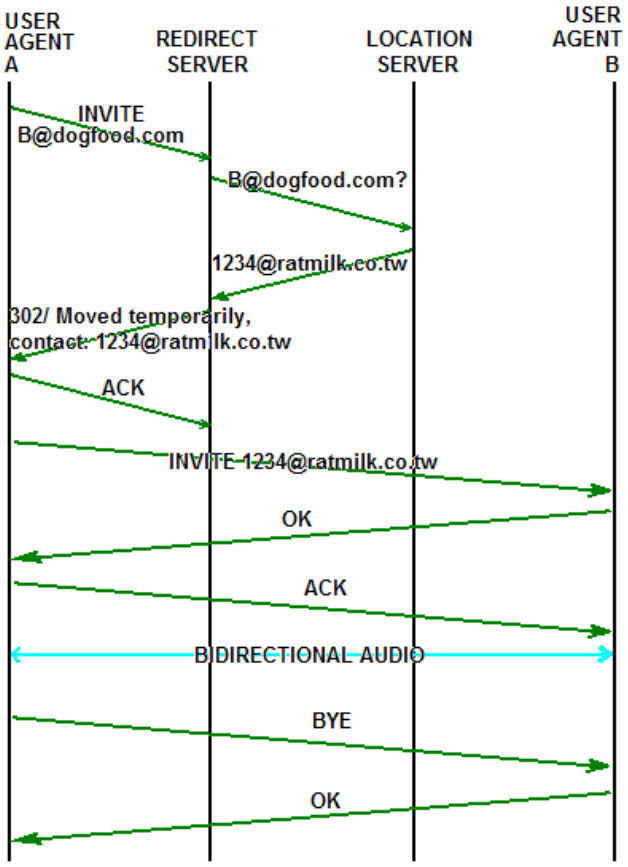

Fig. 9: Sesión SIP de audio, utilizando servidores de redirección y localización.

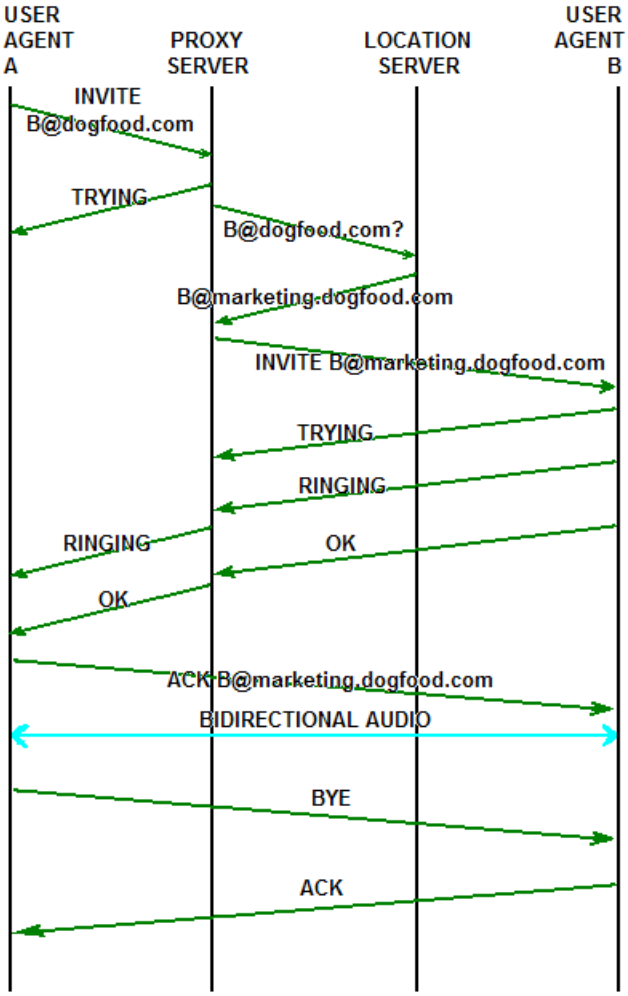

Fig. 10: Sesión SIP de audio, utilizando servidor proxy y de localización.

Comparativa de H.323 con SIP

Complejidad / Servicios ofrecidos:

- H.323 es un estándar muy complejo que describe una arquitectura de comunicaciones y servicios completa.
  Cada revisión del estándar (actualmente la 6ª) añade nuevas funcionalidades que deben implementarse obligatoriamente. Aun así, H.323 mantiene compatibilidad hacia atrás con todas las revisiones anteriores.
  Los terminales H.323 pueden ofrecer diversos servicios, pero todos son compatibles y pueden ofrecer como mínimo un servicio básico de llamadas de voz.
- SIP, en cambio, es un protocolo que al ser más abierto y flexible permite una mayor interoperabilidad con otros códecs y protocolos. Por desgracia, la flexibilidad de SIP puede derivar en la incompatibilidad de dispositivos.

CRÍTICA:

SIP fue creado inicialmente como una alternativa más simple a H.323, pero con el tiempo la especificación de SIP ha incrementado su complejidad hasta el punto de ser parecida a la de H.323: El RFC 3261 (SIP) son 269 páginas, frente a las 317 de H.323rev.6.
¡Tanto H.323 como SIP pueden ser opciones demasiado complejas para un simple servicio de telefonía o videotelefonía!.

Direccionamiento:

- H.323 soporta múltiples tipos de direcciones tales como:
  - dirección IP:puerto
  - alias H.323
  - número E164 (teléfono tradicional)
  - URL
  - ... (etc)
- SIP, por su parte, únicamente acepta direcciones del tipo URI

CRÍTICA:

Salvo por el acertado soporte de números E164 por parte de H.323, ambas aproximaciones son inapropiadas.
Por un lado, H.323 pretende soportar demasiados tipos de dirección, acumulando una complejidad que podría no verse compensada con utilidad.
SIP, por su parte, se basa en un formato de URI (Unique Resource Identifier) similar a las direcciones de correo electrónico, que son cómodas de recordar pero inadecuadas para servicios de telefonía:
Supongamos que un personaje, llamémosle 'M', tiene una dirección SIP asignada por el proveedor de servicio de telefonía VoIP, p.ej:
manolaco@clientes.sin_cobertura.es
Ahora bien, si 'M' decide cambiar su proveedor, su nueva dirección sería, por ejemplo:
manolaco@clientes.con_cobertura.es

oops... 'M' deberá avisar a todos sus contactos del cambio de dirección, o éstos no podrán llamarle. Esto resulta sumamente engorroso.
¿Qué sucede, en cambio, con los números de teléfono tradicionales E164? Como son precisamente eso - simples números, son más incómodos de recordar, pero el hecho de que ciertos números estén gestionados por distintos operadores (como sucede con los números de los teléfonos móviles) no impide que se cambie de operador sin cambiar de número - es un mero trámite administrativo transparente a los usuarios finales. Es por ello conveniente utilizar un formato basado en números que no cambian.

Formato de los mensajes:

- Los mensajes de los protocolos que recoge H.323, como H.225 y H.245, utilizan una codificación binaria, similar a la de los datagramas IP o las tramas Ethernet.
- SIP codifica los mensajes en texto plano legible por humanos, como hace HTTP o XML.

CRÍTICA:

En la práctica es mucho más adecuada la codificación binaria, pues el procesado de texto tiene un coste computacional considerable que reduce el rendimiento total. Además, por lo general los mensajes de texto ocupan más espacio y por tanto consumen un mayor ancho de banda.

Interconexión con otras redes:

- H.323 es compatible con las redes H.32x. La especificación además incluye compatiblidad con la red telefónica de circuito conmutado tradicional (PSTN); la traducción de direcciones de red H.323 a números de teléfono E164 es una función obligatoria en los gatekeepers.
- SIP no define cómo interoperar con otras redes como la red telefónica tradicional; la funcionalidad queda delegada a los dispositivos implementadores.

CRÍTICA:

La interconexión con la PSTN es de vital importancia para el éxito de la telefonía IP, pues en la vida real no se producen saltos tecnológicos drásticos que dejen atrás los anteriores sistemas. H.323 cuenta con un punto a su favor en este aspecto al estar definida la compatibilidad en el propio estándar.

Seguridad:

- Los puntos finales H.323 negocian durante la conexión los puertos que se van a emplear para:
  - H.245 “Parámetros de llamada”
  - RTP Audio
  - RTP Video
  - RTCP

  (entre cualquiera de los puertos libres entre el 1024 y el 65535).
- SIP especifica los puertos de RTP/RTCP durante el establecimiento de llamada.

CRÍTICA:

Como los puertos RTP/RTCP se escogen dinámicamente, ni H.323 ni SIP funcionan detrás de firewalls, pues éstos no saben qué puertos deben abrirse si no están configurados a priori. Así pues, la única forma de hacer que una conexión funcione es abrir todos los puertos, con el grave riesgo de seguridad que esto conlleva. No obstante, hoy en día existen firewalls “inteligentes” que reconocen los protocolos H.323 y/o SIP y pueden averiguar qué puertos se deben abrir dinámicamente inspeccionando los paquetes en los respectivos canales de control ....aunque esto supone un intento de “parchear” un diseño mal planteado y no es una verdadera solución.

Por contra, protocolos como Inter-Asterisk eXChange 2 (IAX2) solucionan el problema de los firewalls transmitiendo conjuntamente la señalización y los datos mediante UDP. IAX2 utiliza un único puerto, el 4569.

Conclusión:

Tanto SIP como H.323 son protocolos maduros que cuentan con sus partidarios y detractores en la industria. H.323 cuenta con una mayor base establecida, pero SIP está ganando aceptación por parte de algunos proveedores de servicio para el transporte de tráfico VoIP.

Es cuestión de tiempo que la industria se decante por uno o por otro, o quizá ambas tecnologías acaben coexistiendo utilizándose en escenarios diferentes.

BIBLIOGRAFÍA

H.323 vs SIP: a comparison:
http://www.packetizer.com/voip/h323_vs_sip/

Wikipedia - H.323:
http://en.wikipedia.org/wiki/H.323

Estándar H.323:
http://www.packetizer.com/voip/h323/standards.html

Wikipedia - SIP:
http://en.wikipedia.org/wiki/Session_Initiation_Protocol

RFC 3261 (Session Initiation Protocol):
http://www.ietf.org/rfc/rfc3261.txt

Foro VoIP (Voz sobre IP):
http://www.voipforo.com/

James Peters; Jonathan Davidson; Maribel Martínez Moyano (Pearson Educación) -
Fundamentos de Voz sobre IP