

# Protocolos de Voz sobre IP

---

## Introducción

Existen una gran cantidad de protocolos que proponen formas distintas de establecer y controlar comunicaciones voz sobre redes IP. Mucho se habla de ellos, y la gran cantidad de acrónimos y abreviaturas hacen que reiteradas veces se confundan las funciones y el alcance de cada uno de ellos. La idea de este documento es clarificar los conceptos elementales para poder clasificar de manera precisa el alcance y las funciones básicas de los principales protocolos utilizados para el transporte de audio sobre redes IP.

Queda excluido del objetivo de este documento profundizar en los detalles de las normas que definen los protocolos aquí mencionados, sino simplemente clarificar algunos conceptos básicos sobre señalización y transporte de voz en redes IP. Por ello se han pasado por alto muchos detalles específicos de los protocolos con el objetivo de simplificar las explicaciones.

## Clasificación de los protocolos

Si quisiéramos definir en forma teórica, independizándonos de los protocolos ya existentes, un modelo del procedimiento para establecer una comunicación de voz entre dos terminales sobre una red IP, lo primero que deberíamos hacer es definir los distintos tipos de negociaciones que deberían intercambiar las terminales para lograr la comunicación.

La primer idea que surge es la de informar al terminal llamado que deseo establecer la una comunicación de voz. Luego el terminal llamado responderá de alguna forma, aceptando o rechazando dicha comunicación. A este tipo de intercambio de información se la suele llamar *señalización de llamada* (call signalling).

Por tratarse de una comunicación de voz sobre una red IP, la voz se transmite codificada en paquetes. Existen una gran variedad de codificadores y hoy en día los más utilizados son G.729, G.711, GSM, entre otros. Además en la mayoría de los casos la voz se transporta sobre segmentos UDP, lo que hace necesario la negociación de los puertos UDP donde el receptor espera recibir el audio. Debido a esto, es necesario intercambiar mensajes donde se negocien estas cuestiones y otras más específicas de cada protocolo. Para el intercambio de este tipo de información se definen los protocolos de *control de señalización de llamada* (Call control signaling)

Una vez establecida la comunicación, se debe enviar el audio codificado en paquetes IP. Las redes IP suelen tener variaciones de retardo altos respecto a las redes de telefonía tradicionales ya que no fueron diseñadas para el transporte de voz. Y además, por ser una red de datagramas, los paquetes de voz podrían llegar desordenados. Debido a estas características de la red IP, se necesita empaquetar la información de voz sobre algún protocolo que minimice o controle estos efectos. A

éstos protocolos se los denomina protocolos de *transporte de “media”* (media transport protocols). Estos protocolos están asociados con los protocolos de *control de transporte de “media”* (media transport control protocols) cuya función es la de enviar entre los terminales intervinientes en la comunicación estadísticas sobre jitter, paquetes enviados, paquetes recibidos, paquetes perdidos, etc. La RFC3550 define el protocolo RTP y RTCP que son hoy en día los más utilizados para el transporte y control de la “media”.

A esta altura parecería que tenemos todos los elementos necesarios para poder establecer y controlar una comunicación de voz entre dos terminales. Esto es cierto y de hecho en algunas topologías de red chicas con esto es suficiente. Cuando la red empieza a crecer y ya no son solo terminales los que se quieren comunicar sino que también gateways para interconectarse con la red de telefonía pública tradicional, se hace necesario centralizar cierto tipo de información para que la red sea escalable. Para lograr esto se coloca un dispositivo de control que posee la inteligencia de la red, es decir, capacidades de ruteo, transcoding de señalización y localización de dispositivos entre otras funciones. A éste dispositivo se lo suele denominar *softswitch*. Como consecuencia se hace necesaria la comunicación entre gateways o terminales y el dispositivo de control, el softswitch. A este tipo de comunicación le llamaremos protocolos de *registro y control*. Cabe destacar que esta clasificación es un poco ambigua ya que a veces la definición de los protocolos de registro y control está embebida como parámetros dentro de los protocolos de señalización de llamadas.

Ahora sí, podemos empezar a hablar específicamente de los protocolos existentes hoy en día y clasificarlos según las funciones definidas. Las principales entidades que definen las normas o protocolos de VoIP que hoy se utilizan son la IETF por medio de las RFCs y la ITU-T. En la siguiente tabla se muestra la clasificación:

	SIP	H.323	MeGaCo/ H.248	MGCP
	<i>IETF</i>	<i>ITU-T</i>	<i>IETF/ITU-T</i>	<i>IETF</i>
<b>Señalización de llamada</b>	SIP	H.225/Q.931	MeGaCo	MGCP
<b>Control de Señalización de llamada</b>	SDP	H.245	SDP	SDP
<b>Registro y control</b>	SIP	H.225/RAS	MeGaCo	MGCP
<b>Tranporte de audio</b>	RTP	RTP	RTP	RTP
<b>Control de transporte de audio</b>	RTCP	RTCP	RTCP	RTCP
<b>SoftSwitch</b>	SIP server	Gatekeeper	Call Agent MGC	Call Agent o MGC

**Tabla 1**

*Nota: Tanto RTP, como RTCP, como SDP, están definidos por la IETF.*

Se han colocado en la tabla los cuatro protocolos más utilizados al día de hoy: SIP, H.323, MEGACO/H.248 y MGCP. Nótese un concepto importante, el transporte del audio se hace siempre por RTP independientemente de la señalización utilizada. Entonces si a la pregunta “¿Qué protocolo de VoIP se está utilizando?” se responde RTP, ésta respuesta estaría incompleta ya que estaría faltando definir el protocolo de señalización utilizado. Si la respuesta hubiera sido SIP, o H.323, la respuesta hubiera alcanzado por que en estas normas se define al RTP como el protocolo de transporte de audio. Además prácticamente en todos los casos se utiliza RTP para transportar el audio. Nótese también que solo H.323 define un protocolo aparte e independiente para la registraci3n y el control. En el caso de SIP, para la registraci3n se definen mensajes

específicos y para el control se utilizan los mismos mensajes de señalización de llamada. El caso de MEGACO/H.248 y MGCP es diferente y será tratado más adelante ya que no admite llamadas sin el dispositivo de control (softswitch), es decir, no se pueden enviar mensajes de señalización entre gateways ya que presenta una topología de maestro/esclavo.

Se agregó también a la tabla el nombre del dispositivo de control que venimos llamando softswitch de forma general, según se define en cada norma. En la práctica el softswitch suele tener capacidad para realizar transcoding de señalización, es decir, podríamos pensar como que el softswitch tiene embebido un gatekeeper, un SIP Server y un call agent, y que traduce la señalización para que terminales y gateways que trabajan con distintos protocolos de señalización puedan comunicarse entre sí.

A continuación se describen las funciones de los distintos tipos de protocolos.

**Nota:** Existen protocolos propietarios para el transporte de voz y señalización. Estos son desarrollados por los “vendors” y suelen ser muy eficientes pero condicionan el diseño de la red a un único “vendor” o a ciertas topologías muy específicas

### Protocolos de señalización de llamada

Para simplificar la explicación vamos a utilizar un ejemplo de una llamada directa entre dos terminales (teléfonos IP o softphones)

Supongamos que un usuario quiere establecer una llamada, y para ello, digita una dirección IP (no suele ser lo más común pero ésta sería la forma más simple si no hay algún tipo de dispositivo de control que traduzca nombres o números en direcciones IP) que indica el teléfono destino al que se quiere llamar. El teléfono IP llamante envía un paquete al teléfono IP llamado indicándole que quiere establecer una comunicación. El teléfono llamado responde indicando que recibió la llamada y la está procesando. Cuando el llamado avisa por medio del timbre local (ringing) que ha recibido una llamada entrante, envía un mensaje al llamante para avisar. Una vez que se levanta el teléfono, se avisa al llamante con otro mensaje y a partir de ese momento la comunicación queda establecida. Todo este envío de mensajes se realiza a través de los protocolos de *señalización de llamada*. Estos son SIP y H.323 (H.225/Q.931)

A manera de ejemplo se muestra el flujo de mensajes en la figura 1.

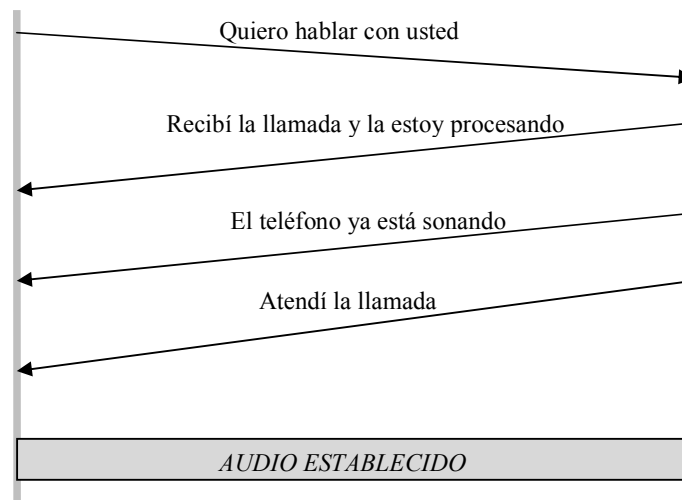


Figura 1

H.323

En el caso de H.323, se utilizan los mensajes que ya habían sido definidos en la norma Q.931 para ISDN. Como estos mensajes no fueron originalmente definidos para VoIP, carecen de parámetros relacionados con el “mundo” IP. Por ello, se definen en el protocolo H.225 los parámetros IP y son transportados sobre el protocolo Q.931 en el campo UIIE (User to User Information Element). H.225/Q.931 corre sobre TCP, por lo que requiere un establecimiento de conexión inicial. En este tipo de llamadas, el llamante trabaja como cliente, y el llamado como servidor en el puerto TCP 1720.

La figura 2 muestra un ejemplo del flujo de llamadas para H.323



Figura 2

*Notas aclaratorias:* Entre el connect y el releaseComplete estaría el audio (RTP). En estos escenarios el ringing se generó localmente pero puede ser generado en banda mediante mensajes de progreso.

SIP

En el caso de SIP, utiliza sus propios mensajes que fueron definidos desde el inicio en un entorno IP. Es importante notar que SIP suele correr sobre UDP por ello aparece el mensaje de ACK confirmando el mensaje de 200OK. Al igual que en H.323, trabaja con un modelo cliente servidor donde el llamante es el cliente y el llamado un servidor en el puerto 5060 de UDP.

La figura 3 muestra el flujo de una llamada en el caso de SIP

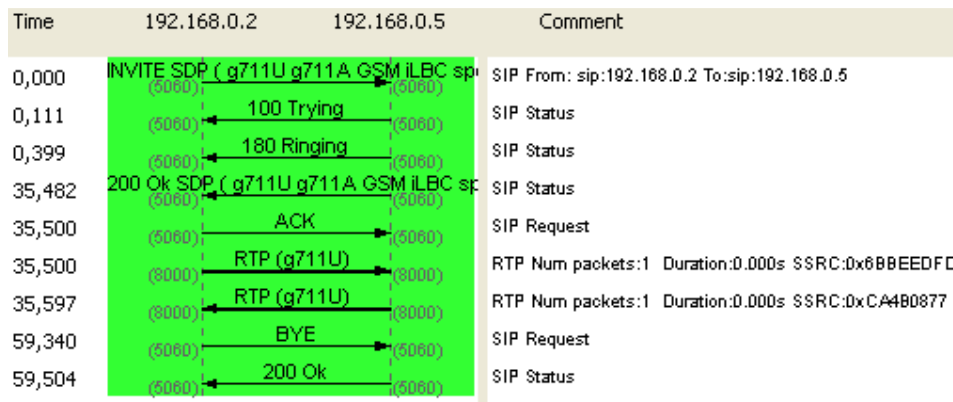


Figura 3

En la tabla 2 se resumen los significados básicos de los nombres de los mensajes de para SIP y H.323

	SIP	H.323 (H.225/Q.931)
Quiero hablar con ud.	<i>INVITE</i>	<i>SETUP</i>
Recibí la llamada y la estoy procesando	TRYING	CALL PROCEEDING
El teléfono ya está sonando	RINGING	ALERTING
Atendí la llamada	200 OK + <i>ACK</i>	CONNECT

**Tabla 2**

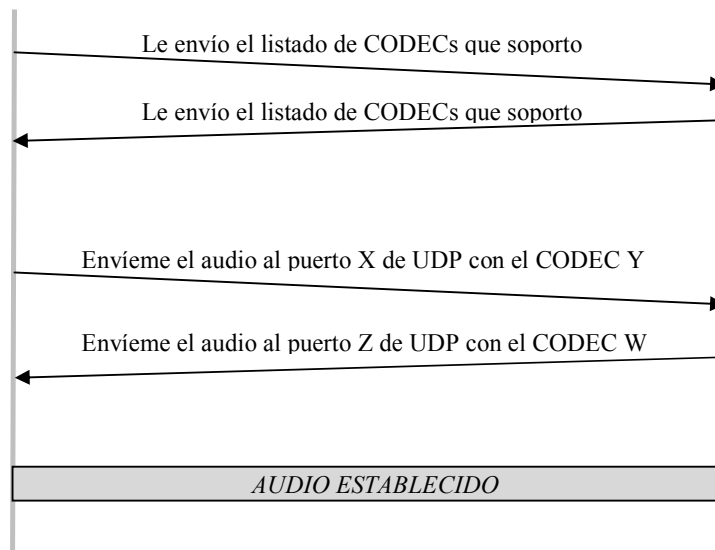
*Nota: los mensajes en cursiva van de llamante ha llamado.*

### Protocolos de control de señalización de llamada

Hasta aquí vimos de manera muy resumida y simplificada como se utilizan los protocolos de señalización de llamada. Estos mensajes llevan principalmente identificadores de la llamada, datos del llamante, y datos del llamado, pero no llevan información de cómo debe enviarse el audio. Esto es, como ya habíamos mencionado, para el transporte del audio se utiliza RTP, que corre sobre UDP, y utiliza puertos efímeros que deben ser negociados. Además, existen una gran cantidad de codificadores diferentes para enviar el audio y esto también debe ser negociado. Como conclusión entonces, podemos decir que se necesitan mensajes para negociar entre otras cosas: 1) Los puertos a donde se enviará el RTP, es decir, los puertos de UDP en los cuales el receptor del audio trabajará como servidor; 2) El CODEC a utilizar en la llamada.

*Nota: H.323 negocia normalmente además de esto la selección de maestro/esclavo con este protocolo.*

En la figura 4, se muestra un modelo simplificado para los protocolos de control de señalización de llamada



**Figura 4**

Entonces, para las funciones ya mencionadas, se utilizan los protocolos que solemos llamar protocolos de *control de señalización de llamada*. SIP utiliza el protocolo SDP (Session Description Protocol), mientras que H.323 utiliza H.245.

## *SIP*

En el caso de SIP como vemos en la *figura 3*, en los mensajes de INVITE y 200OK, se envía el SDP (Session Description Protocol). Notar que el INVITE va de llamante al llamado y el 200OK en el otro sentido. Dentro del mensaje de SDP sobre el INVITE, se envía principalmente, el listado de CODECs con los que desea trabajar el llamante en orden de prioridad, y además, la IP y Puerto a donde el llamado quiere recibir el RTP. En el mensaje de 200OK se envían los mismos parámetros pero desde el llamado al llamante. No siempre el 200OK lleva el SDP, ya que muchas veces, algún mensaje anterior lo lleva. Esto ocurre por ejemplo cuando se quiere enviar la señal de *ringing* a través de la red y no generarla en forma local.

## *H.323*

El caso de H.323 es muy similar, con la diferencia que originalmente se negociaban los CODECs y las direcciones de transporte (IP:Puerto UDP) una vez establecida la llamada. Esto generaba un retardo en el inicio del envío del audio, así que se definió lo que se conoce como FAST START o FAST CONNECT que trabaja de forma similar a SIP transportando el H.245 sobre los mensajes H.225.

Como se ve en la figura 5, el H.245 posee tres mensajes principales:

- 1) TCS (Terminal Capability Set): Negociación de tablas de CODECs principalmente.
- 2) MSD (Master/Slave Determination)
- 3) OLC (Open Logical Channels): La principal función es la de establecer los canales lógicos. Esto es, negociar al igual que el SDP, los puertos de UDP por donde se recibirá el audio.

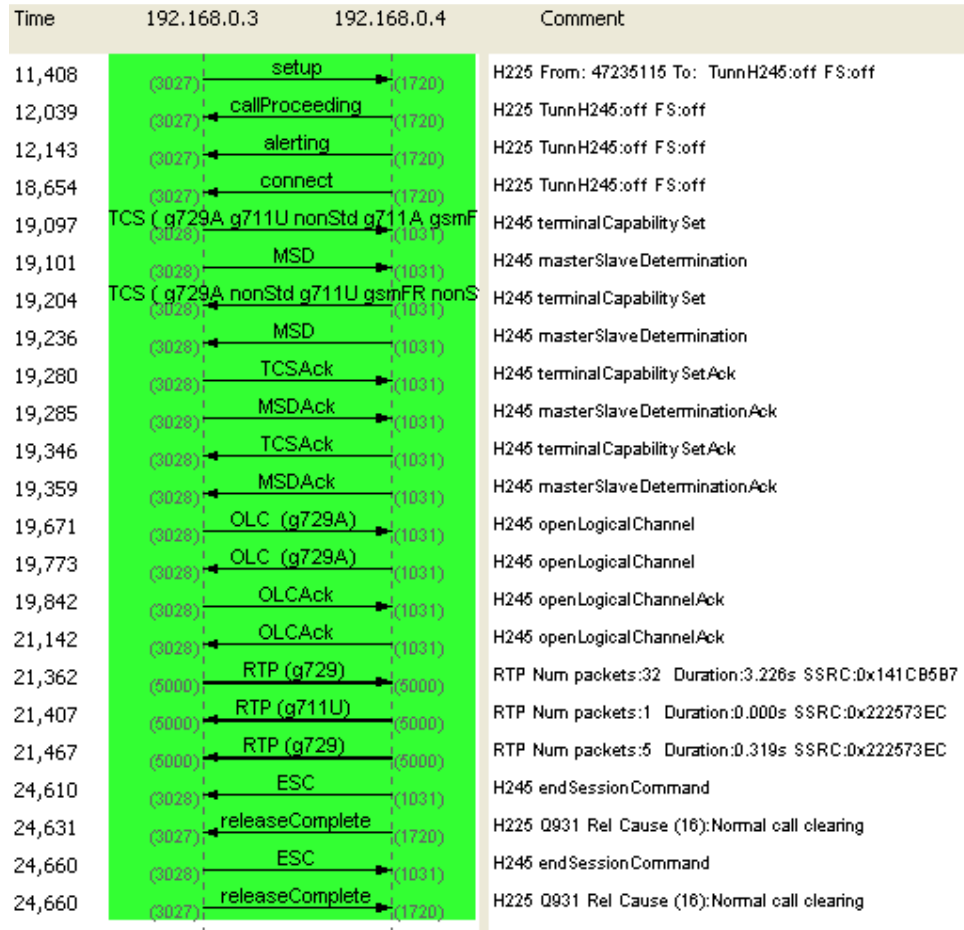


Figura 5

En la siguiente tabla se resumen las diferencias entre SIP y H.323

	SIP (SDP)	H.323 (H.245)
Negociación Maestro/Esclavo	No lo negocia	MSD
Negociación de tablas de capacidades	Parámetro de SDP	TCS
Apertura de canales lógicos	Parámetro de SDP	OLC

Tabla 3

*Nota:* La idea de este documento como ya dijimos, no es profundizar de manera rigurosa sino simplemente introducir los conceptos básicos, por ello, se han simplificado algunas cuestiones para que se haga más simple la comprensión. Por ejemplo aquí solo se habla del establecimiento de la llamada, pero como se ven en las capturas existen mensajes para la terminación de llamadas, y también para el mantenimiento de las mismas.

### Protocolos de transporte de media:

Como ya venimos mencionando independientemente del protocolo utilizado para el establecimiento de la llamada, el protocolo más utilizado para el transporte de audio es el RTP. La principal función del RTP es la de transportar el audio codificado sobre UDP. Para ello posee dos campos fundamentales: 1) Etiqueta de tiempo (TimeStamp):

Este campo mide tiempo en unidades de 125us y se utiliza para que el receptor sepa exactamente en que instante de tiempo debe reproducir el audio recibido; 2) Número de secuencia (Sequence Number): Un número que se incrementa de a uno y sirve para ordenar y determinar si se han perdido paquetes en la red.

La RFC3550 define el formato y funcionamiento de los mensajes de RTP, y de igual forma define los mensajes RTCP que no transportan audio sino que su función es la de enviar estadísticas de lo recibido y enviado a los usuarios con quienes se estableció la comunicación. Esto es: Cantidad de octetos y paquetes enviados, fracción de paquetes perdidos, cantidad de paquetes perdidos, jitter interarribo, entre otros.

## Protocolos de registración y control

Hasta el momento, la llamada se hacía directo sin ningún dispositivo de control interviniendo. Existen varias razones para pensar en colocar un dispositivo que controle y rutee las llamadas. Los motivos más evidentes son:

- 1) Sería poco práctico tener que recordar y discar direcciones IP para llamar a un destino, y además es muy probable que los usuarios no tengan siempre la misma. (El usuario usa una computadora portátil y si conecta en distintas redes o recibe su dirección por DHCP por ejemplo)
- 2) Es importante tener controlado si ciertos usuarios pueden o no llamar o ciertos otros, sobre todo si el destino final está en una red tarifada. Además es importante no solo permitir o no las llamadas, sino que también controlar la duración de las mismas para poder facturar en el caso que sea necesario.
- 3) Si la red se encuentra conectada a la PSTN, es importante tener un dispositivo que se encargue del ruteo en función de los prefijos discados. Si esto no estuviera centralizado, cada telefono IP o gateway debería tener una tabla completa para saber por que gateway salir a la PSTN en función del prefijo discado.

Este dispositivo de control suele conocerse con distintos nombres dependiendo del protocolo con el que se comunique. Esto es, si hablamos de SIP, suele llamarse SIP Server (SIP Proxy, Redirect Server, Location Server, Registrar), si hablamos de H.323 gatekeeper, en MGCP y MeGaCo Media Gateway Controller o Call Agent. Muchas veces en la práctica se le suele llamar SoftSwitch al dispositivo genérico que trabaja como Gatekeeper, SIP Proxy y Call Agent realizando el ruteo y transcoding de señalización. (NO transcoding de audio ya que en este tipo de topologías, el audio no pasa a través del dispositivo de control)

El modelo de llamada ahora cambia respecto al que vimos en el ejemplo anterior, y es diferente la forma de trabajo según el protocolo (H.323, SIP, MEGACO, MGCP) que se utilice.

Lo primero que hay que realizar en todos los casos es registrar el Terminal (telefono IP) o gateway al dispositivo de control o softswitch. Una vez registrado el terminal, el softswitch ya puede colocarlo dentro de la tabla de los usuarios en línea, y además conoce la ubicación (dirección de transporte - IP:PUERTO-). Lo que se hace con la registración en el caso de terminales, es relacionar un número de teléfono o nombre, con una dirección de transporte. En el caso de los gateways, el hecho de registrarlo, alerta al softswitch de que está en línea y que puede utilizarlo para rutear las llamadas que en función del prefijo saldrán por este Gateway según corresponda en su tabla de ruteo.

H.323

En el caso de H.323, el protocolo que se utiliza entre los terminales y gateways con el gatekeeper es H.225 canal RAS (Registration, Admisión, Status). Existen mensajes para generar la registración, mensajes de admisión que se utilizan al momento de realizar la llamada, y mensajes de status para controlar el estado de las llamadas activas.

Como se muestra en la figura 6, los usuarios con IPs 192.168.0.3 y 192.168.0.5 se registran(mensajes de registración) en el Gatekeeper cuya IP es 192.168.0.4. Luego el usuario 192.168.0.3 quiere llamar al 192.168.0.5 pero antes de enviar el SETUP, solicita permiso al Gatekeeper (mensajes de admisión) quien acepta respondiendo con la dirección de transporte del usuario al que se quiere llamar. De la misma forma, el usuario 192.168.0.5 solicita autorización al gatekeeper antes de responder con el mensaje de CONNECT.

*Nota: Existen distintos modelos de llamadas en H.323 en los que, los mensajes de H.225, los de H.245 y hasta el RTP podría forzarse a pasar por el GateKeeper En este ejemplo se utiliza el modelo de llamada directo donde la señalización de llamada va entre terminales. Se utiliza solo el establecimiento de llamada como ejemplo. De manera similar se realiza la finalización de la llamada.*

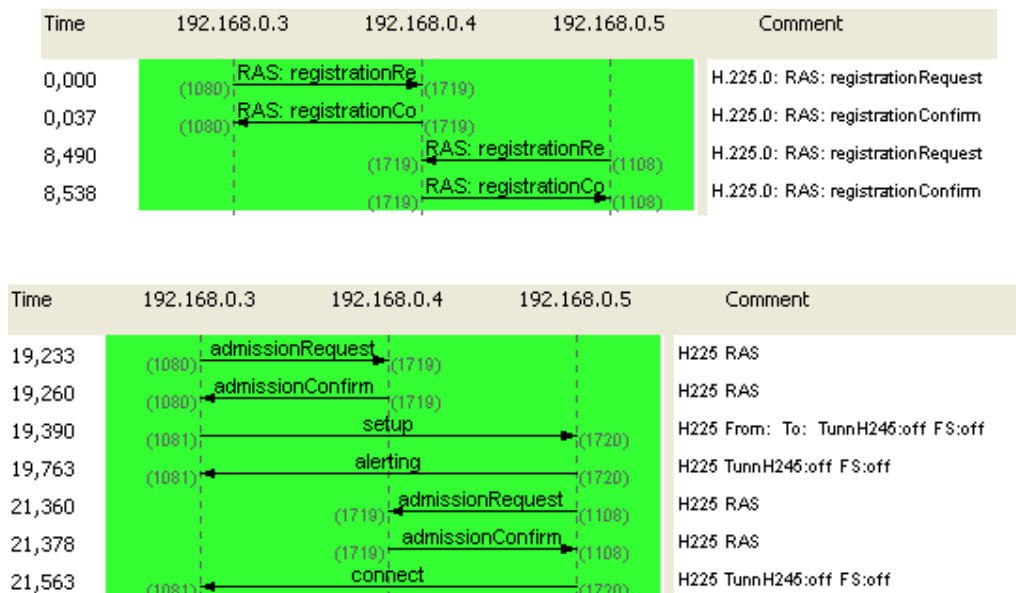


Figura 6

SIP

En el caso de SIP, se utiliza el mensaje de REGISTER para realizar la registración y luego los mismos mensajes de INVITE para realizar la llamada, que en vez de ir hacia el telefono o GW destino pasan primero por el SIP PROXY quien sabe como rutear la llamada y puede o no solicitar autenticación para realizarla. Cabe aclarar que en SIP se define un Registrar que es el dispositivo a donde se registran los usuarios, y por el otro lado un SIP PROXY que es quien rutea las llamadas, en este caso, estamos suponiendo por simplicidad que el Registrar Server y el SIP Proxy Server están localizados en el mismo dispositivo algo que es usual en muchas topologías.

Como se puede ver en la figura, a diferencia de H.323, el mensaje de INVITE no genera una respuesta en el SIP Proxy de vuelta hacia el terminal para indicarle al llamante donde encontrar al usuario destino para enviarle el audio, sino que lo único que hace en principio el SIP Proxy es reenviar el mensaje al usuario destino. Por ello, se podría pensar que en realidad el SIP Proxy funciona como si fuera un router a nivel de SIP, y entonces la llamada sería muy similar al del ejemplo anterior sin SIP Proxy pero con un salto de ruteo en este.

*Nota:* No es totalmente cierto que el SIP Proxy pase los datos sin modificarlos ya que en realidad agrega y en muchos casos modifica campos. Pero es útil pensarlo de esta forma para comprender el concepto básico.

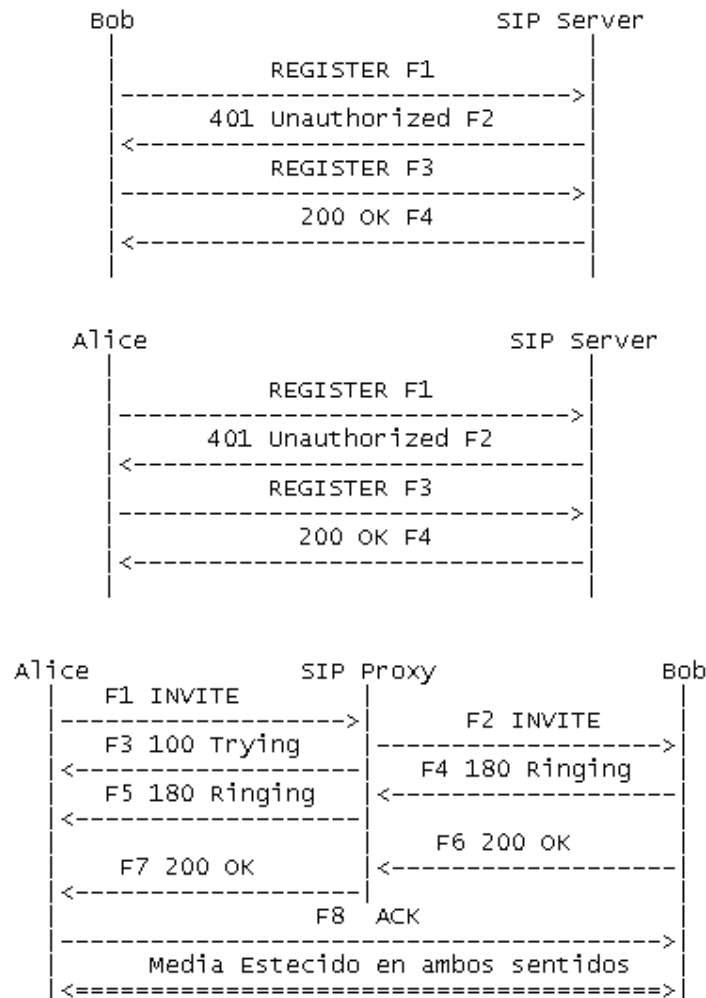


Figura 7

MGCP y MEGACO (H.248)

Hasta el momento todos los ejemplos fueron con SIP y H.323. Tanto MGCP como MEGACO son protocolos con topologías Maestro/Esclavo, es decir, no pueden funcionar sin un dispositivo de control (softswitch). Al contrario tanto SIP y H.323 son protocolos peer-to-peer, y como ya vimos en los primeros ejemplos, se pueden realizar llamadas entre terminales sin necesidad de utilizar el dispositivo de control. Tanto MGCP como MEGACO, son protocolos que fueron pensados únicamente para

gateways y no para terminales, y el concepto es simplificar lo mas posible la inteligencia de estos gateways llevándola al Call Agent (softswitch). Se dice que los gateways son dispositivos “tontos” y que toda la inteligencia se concentra en el Call Agent o softswitch.

Pongamos un ejemplo para clarificar: si se quiere realizar una llamada con un teléfono analógico conectado a una interfaz FXS de un gateway, el procedimiento sería levantar el tubo y esperar a tener tono de discado. En el caso de que el gateway trabaje con SIP o H.323, este generará localmente el tono de discado y el usuario digitara el número al que quiere llamar. Una vez discado el número destino, el gateway por medio del mensaje de INVITE en SIP o ARQ en H.323 enviara este número al softswitch quien resolverá el destino de la llamada. En el caso de que el gateway fuera MGCP o MEGACO, el Call Agent, luego de que el gateway se registre, le enviará un mensaje solicitandole que si detecta el tubo levantado se lo notifique. Una vez que se levanta el tubo, el gateway simplemente le enviará un mensaje notificando el evento. En el caso normal, el Call Agent responderá con un mensaje indicándole al gateway que ponga tono de discado y recoja los dígitos marcados, el Gateway no pondrá el tono hasta no recibir este mensaje. Esto explica que no podría existir una llamada peer-to-peer con este modelo ya que siempre el gateway estaría esperando las instrucciones del Call Agent.

Si entonces queremos clasificar de alguna manera a estos protocolos, tanto MGCP como MEGACO definen los mensajes de *señalización de llamada y registración y control*. Es importante aclarar que en ambos casos se usa SDP como protocolo de *control de señalización de llamada*, y RTP para el *transporte del audio*.

**Tabla de resumen de protocolos en VoIP**

	SIP	H.323	MeGaCo	MGCP
Señalización de llamada	<b>SIP</b> INVITE, Trying, Ringing, BYE, etc	<b>H.225/Q.931</b> SETUP, Call Proceeding, Alerting, Connect	MeGaCo	MGCP
Control de Señalización de llamada	SDP	H.245	SDP	SDP
Registración y control	<b>SIP</b> REGISTER	H.225 canal RAS	MeGaCo	MGCP
Tranporte de audio	RTP	RTP	RTP	RTP
Control de transporte de audio	RTCP	RTCP	RTCP	RTCP

Tabla 4

**Comparación y comentarios finales:**

	H.323	SIP	MGCP	MEGACO
Especificación	ITU-T	IETF	IETF	IETF / ITU-T
Arquitectura	Distribuida	Distribuida	Centralizada	Centralizada
Versión Actual	H.323v6	SIP 2.0 RFC 3261	MGCP 1.0 RFC 2705	H.248.1 RFC 3525

Transporte de señalización	UDP (canal RAS), TCP el resto	TCP o UDP UDP el más utilizado	UDP	TCP o UDP
Codificación	ASN.1	Texto	Texto	ASN.1 o Texto

*H.323 y SIP*

H.323 fue el primero de estos protocolos en establecerse como líder por lo que está maduro y es escalable. Compite directamente con SIP por trabajar ambos protocolos con el mismo tipo de arquitectura. En este último tiempo SIP le fue quitando mucho protagonismo y la mayoría de las soluciones nuevas se implementan con SIP.

*MGCP y MEGACO*

Son protocolos utilizados principalmente por los proveedores de servicios de telefonía ya que permiten controlar de manera eficiente gran cantidad gateways que a su vez poseen gran cantidad de abonados POTS. De manera similar que entre H.323 y SIP, MGCP se utiliza mucho aunque de a poco MEGACO fue quitándole terreno en soluciones con troncales o grandes cantidades de abonados POTS. La especificación de MGCP no se desarrolla más y toda modificación sobre este tipo de arquitectura se realiza sobre la especificación de MEGACO/H.248.

**Referencias:**

- “SIP: Session Initiation Protocol”, RFC 3261
- “SDP: Session Description Protocol”, RFC 2327
- “MGCP: Media Gateway Control Protocol”, RFC 2705
- “MeGaCo: Gateway Control Protocol”, RFC 3525
- “RTP: A Transport Protocol for Real-Time Applications”, RFC 3550
- “H323v6: Packet-based multimedia communication systems”

**Acrónimos:**

- RFC: Request For Comments
- IETF: Internet Engineering Task Force
- ITU-T: International Telecommunication Union – Telecommunication Standardization Sector.
- MGCP: Media Gateway Control Protocol.
- SIP: Session Initiation Protocol
- SDP: Session Description Protocol
- RTP: Real-time Transport Protocol
- RTCP: Real-time Transport Control Protocol
- POTS: Plain Old Telephone System
- RAS: Registration, Admission, Status
- RRQ: Registration Request
- ARQ: Admission Request