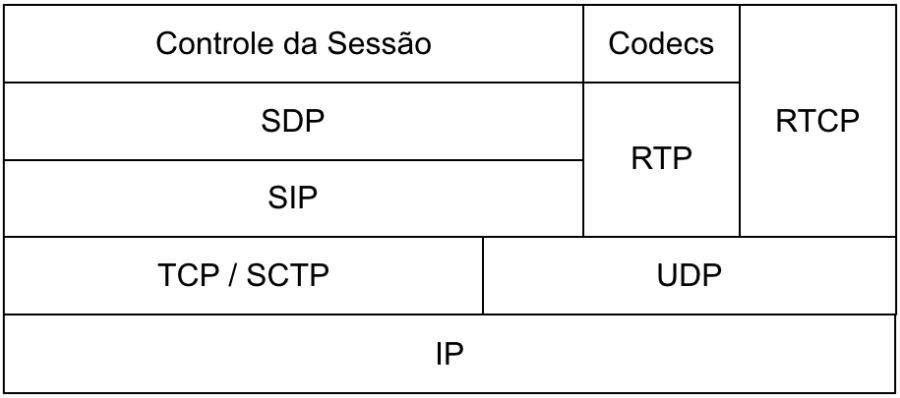
Arquitetura SIP

A arquitetura SIP é destinada às mesmas aplicações englobadas pela arquitetura H.323, ou seja, àquelas utilizadas para comunicações multimídia em tempo real na Internet. Da mesma forma, para a operação da arquitetura SIP, são utilizados inúmeros protocolos com objetivos específicos. Combinando as funcionalidades desses protocolos, essa arquitetura de comunicação é capaz de oferecer um serviço flexível e ao mesmo tempo robusto para uma série de aplicações multimídia de escopos distintos. Os principais protocolos constituintes dessa arquitetura de comunicação são os protocolos SIP (Session Initiate Protocol), o SDP (Session Description Protocol), o RTP e, opcionalmente, o RTCP, além de outros protocolos menos utilizados, tais como o SAP (Session Announcement Protocol) e o RTSP (Real Time Streaming Protocol). A Figura a seguir apresenta a arquitetura SIP de forma simplificada.



Da mesma forma que ocorre com o H.323, os protocolos de transporte padrões dessa arquitetura são o TCP e o UDP. Contudo, isso não impede que o protocolo SCTP seja utilizado, embora sua utilização seja ainda “experimental” na arquitetura SIP.

Quando comparamos a arquitetura SIP com a arquitetura H.323, podemos observar algumas diferenças interessantes. Por exemplo, não existe um protocolo H.323 propriamente dito, uma vez que o H.323 é uma recomendação de como alguns protocolos devem operar entre si para obter um serviço de comunicação em tempo real satisfatório. Já a arquitetura SIP é formada também por um conjunto de protocolos, contudo há um protocolo SIP. Ou seja, o SIP é tanto um protocolo quanto uma arquitetura de comunicação multimídia em tempo real.

Enquanto o H.323 se desenvolveu bastante na área de aplicações de videoconferência, o SIP ganhou espaço na área de telefonia IP. Muitas aplicações nessa área estão utilizando a arquitetura SIP para gerência das comunicações multimídia, com gateways SIP/H.323 e SIP/PSTN[1] sempre que necessário. Na Internet atual, pode-se dizer que o SIP é o padrão de sinalização adotado para comunicações multimídia em tempo real.

Protocolo SIP

O protocolo SIP foi especificado pelo grupo de trabalho MMUSIC do IETF, inicialmente com a RFC 2543, publicada em 1999. A versão posterior do protocolo SIP, contida na RFC 3261, foi publicada no ano de 2002.

O protocolo SIP possui uma série de funcionalidades, sendo algumas delas contidas na lista a seguir.

- Pedido de abertura de uma comunicação multimídia. No escopo do protocolo SIP, esse pedido é chamado de convite;
- Controle e encerramento de comunicações em tempo real (*sessões* SIP);
- Autorização de chamadas, utilizando serviços específicos da arquitetura;
- Localização de usuário. Como o SIP considera que os host podem ser móveis, ele implementa um mecanismo para descobrir o endereço IP atual desses hosts.

Para o desenvolvimento do protocolo SIP, dois protocolos de aplicação da Internet foram considerados como referência: o HTTP (Hypertext Transfer Protocol) e o SMTP (Simple Mail Transfer Protocol). O SIP utiliza endereços URL, como o protocolo HTTP, e uma estrutura de cabeçalho semelhante a presente em mensagens SMTP. Além disso, o protocolo SIP é estruturado textualmente, como esses dois protocolos. Isso significa que uma mensagem SIP enviada na rede pode ser lida por pessoas diretamente. Em comparação, as mensagens dos protocolos da arquitetura H.323 são codificadas diretamente em formato binário (ASN.1): uma pessoa não consegue ler diretamente uma mensagem do H.245, por exemplo.

[1] PSTN (Public Switched Telephone Network) representa a rede de telefonia convencional.

Assim com o H.323, o SIP utiliza portas de comunicação pré-definidas para a abertura de comunicações. As portas utilizadas são 5060 e 5061, que podem ser, por padrão, UDP ou TCP, embora formas não padrões possam utilizar o SCTP. Essa característica do protocolo SIP garante mais flexibilidade à comunicação. No padrão H.323, por outro lado, as sinalizações dos protocolos H.225 e H.245 ocorrem obrigatoriamente utilizando o protocolo TCP. Utilizar protocolos confiáveis em comunicações em tempo real pode ser ligeiramente prejudicial, na medida em que certo tempo é gasto com a abertura de conexão ao nível de transporte.

Utilizando apenas o protocolo SIP para sinalização, em muitos casos sobre o UDP, as comunicações realizadas sobre essa arquitetura são mais rápidas que aquelas realizadas utilizando as versões iniciais do H.323. Como o protocolo UDP não oferece controle de erro, o protocolo SIP utiliza mecanismos para retransmitir mensagens perdidas ou recebidas com erros, quando esse protocolo de transporte for empregado.

A Tabela a seguir apresenta as principais mensagens do protocolo SIP, juntamente com suas finalidades.

Mensagem	Funcionalidade
INVITE	Indicar um pedido de abertura de comunicação SIP
ACK	Confirmar recebimento de mensagem
CANCEL	Indicar cancelamento de pedidos
BYE	Indicar o encerramento de uma comunicação SIP
REGISTER	Realizar registros e consultas a usuários em servidores Registrar
OPTIONS	Requisitar informações de servidores SIP
SUBSCRIBE	Registrar o usuário junto a elementos SIP especiais
NOTIFY	Informar sobre a ocorrência de algum evento previamente solicitado

Protocolo SDP

O protocolo SDP (RFC 2327), obrigatório para a arquitetura SIP, é utilizado para descrever sessões multimídia. Com o SDP, aplicações participantes de comunicações com

suporte multimídia podem trocar informações sobre suas capacidades de processamento de mídia, permitindo assim compatibilizar os codecs a serem empregados nas comunicações.

Algumas das informações presentes nas mensagens SDP são endereços IP, portas UDP, TCP ou SCTP utilizadas, codecs suportados, título e assunto da sessão multimídia, informações de contatos, entre outras. Todas essas informações são disponibilizadas textualmente seguindo a especificação desse protocolo.

As informações transmitidas pelo protocolo SDP trafegam encapsuladas em mensagens SIP.

Assim como o SIP, o protocolo SDP utiliza uma codificação textual. Uma mensagem SDP é composta por campos de controle como mostra a Tabela a seguir. Cada um desses campos, que podem ser representados de forma completa ou reduzida, como na Tabela, deve estar contido em uma linha separada, segundo a codificação do SDP.

Parâmetro	Significado
v	Número da versão do protocolo SDP
o	Identificador do iniciador da comunicação
s	Assunto/nome da sessão
c	Informações de controle da comunicação
t	Tempo de início e fim da sessão
m	Descrições da mídia
a	Atributos da mídia

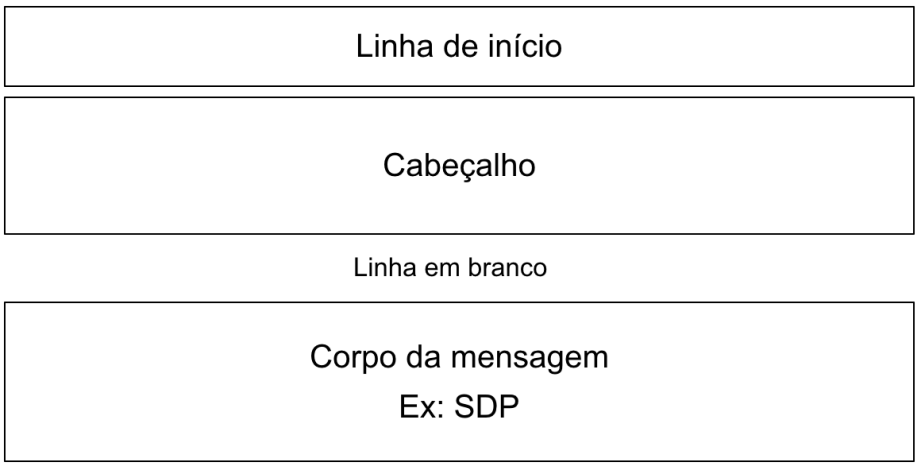
Mensagens SIP

As mensagens SIP possuem um formato textual de codificação. As informações contidas nessas mensagens podem estar presentes em três partes:

- Parte I - Linha de início. Essa parte define o tipo da mensagem e sua natureza, que pode ser um pedido ou uma resposta. Além disso, informações do destino da mensagem também podem estar presentes.

- Parte II – Cabeçalho. Essa parte contém campos de controle diversos, relacionados à operação das mensagens.
- Parte III – Corpo. Essa parte, opcional, vem separada do cabeçalho por uma linha em branco. Um exemplo típico de informações contidas nessa área são mensagens SDP.

A Figura a seguir apresenta a organização típica de uma mensagem SIP.



Vamos analisar uma mensagem SIP típica para descrever alguns dos campos mais comuns utilizados em mensagens de protocolo. A mensagem de exemplo que será considerada será a INVITE.

Uma mensagem INVITE, enviada para iniciar uma conexão SIP, contém informações sobre o tipo de comunicação multimídia sendo requisitada. A mensagem a seguir apresenta campos típicos de uma mensagem INVITE.

```
INVITE sip: daniel@uefs.br SIP/2.0
Via: SIP/2.0/UDP uefs.br:5060
Max-Forwards: 70
To: Daniel <sip:daniel@uefs.br>
From: Carlos <sip: carlos@rnp.br>
Call-ID: 5412365@rnp.br
CSeq: 1 INVITE
Subject: Comunicações multimídia em tempo real
Contact: <sip:carlos@rnp.br
```

```
Content-Type: application/sdp
Content-Length: 160

v=0
o=Tesla 2890844526 2890844526 IN IP4 rnp.br
s=Phone Call
c=IN IP4 200.137.5.98
t=0 0
m=audio 49170 RTP/AVP 0
a=rtpmap:0 PCMU/8000
```

Essa mensagem contém as informações necessárias para a abertura de uma conexão SIP. Essa funcionalidade da mensagem INVITE é semelhante àquela presente no protocolo H. 245.

Nesse pedido SIP, verificamos que o INVITE é endereçado ao terminal SIP indicado pelo endereço “sip:daniel@uefs.br". Nessa primeira linha, está indicada que a versão 2 do protocolo SIP é utilizada: a versão 2 é a versão atual desse protocolo.

A partir da primeira linha inicia-se a parte do cabeçalho da mensagem SIP. O primeiro campo dessa parte é o campo Via. Esse campo é utilizado para indicar o número de elementos SIP que essa mensagem “passou” até chegar ao seu destino. Assim, cada dispositivo SIP que retransmite uma mensagem SIP deve acrescentar seu endereço em um campo Via. É claro que uma mensagem SIP também pode ser transmitida diretamente ao ponto final da comunicação, sem encaminhamento por elementos intermediários. O emissor de uma mensagem SIP sempre coloca seu próprio endereço no campo Via.

Outras informações interessantes presentes nesse campo são a porta de comunicação utilizada e o protocolo de transporte que encapsula a mensagem SIP. No caso do exemplo considerado, a porta é 5060 UDP.

O campo seguinte ao campo Via é o Max-Forwards. Esse campo, que deve ser iniciado preferencialmente com um valor alto (70 é o recomendado), é decrementado por cada servidor SIP que recebe e encaminha uma mensagem SIP. Quando esse valor atingir zero, a mensagem deve ser descartada. A idéia é evitar que mensagens SIP fiquem indefinidamente em *loop*. Campos semelhantes a esse são o TTL dos datagramas IPv4 e HOP LIMIT dos datagramas IPv6.

Os campos To e From apresentam, respectivamente, o destino e a origem da mensagem, seguindo o formato dos endereços SIP. O campo seguinte, Call-ID, é usado para identificador uma sessão de comunicação, de forma a permitir que mensagens sejam classificadas como pertencente à determinada comunicação. Para construir esse identificador, um número aleatório é criado pelo host que transmitiu o INVITE. A esse identificador é então adicionado o símbolo @ e o domínio (ou endereço IP) desse mesmo host. Identificadores opcionais (*tags*) podem ser adicionados aos campos To e From para auxiliar na identificação de uma comunicação particular.

O próximo campo, na seqüência, é o CSeq. Esse campo contém um identificador numérico seguido pelo nome da mensagem, nesse caso o INVITE. Este identificador é incrementado a cada nova requisição enviada e é utilizado para a associação de pedidos e respostas.

O campo Contact contém a url do emissor da mensagem, indicando o endereço que pode ser utilizado para encaminhar mensagens, tais como respostas, diretamente para a origem. Na maior parte do tempo, esse campo conterá o mesmo conteúdo que o campo From.

Na seqüência, o campo Subject contém um texto explicativo informando o "tema" da comunicação que será iniciada. O uso desse campo é opcional, e seu conteúdo não possui qualquer ligação com a operação do protocolo.

Os campos Content-Type e Content- Length servem para indicar o conteúdo da mensagem. No exemplo considerado, o conteúdo da mensagem INVITE é uma mensagem SDP, contendo 160 octetos de dados.

Uma linha em branco deve estar presente para separar o cabeçalho SIP do corpo da mensagem SDP. Como vimos anteriormente, o SDP é utilizado para descrever os atributos de mídia que a origem da mensagem deseja utilizar para a comunicação. No exemplo considerado, uma conexão de áudio (parâmetro m) é solicitada, utilizando o RTP para a transmissão da mídia codificada no formato PCM μ-law, indicada pelo payload RTP 0. O valor "0", que será indicado no campo PT dos datagramas RTP, indica o codec G.711 com codificação PCM u-law.

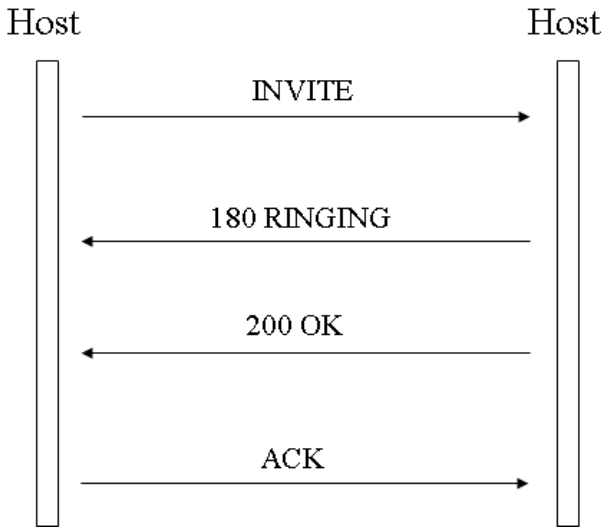
As mensagens SIP de resposta utilizam um código numérico que identifica a situação do pedido feito anteriormente. Cada código está contido em uma faixa geral de descrição,

como mostra a Tabela a seguir. Esse esquema de informação por códigos é semelhante àquele utilizado pelo protocolo HTTP.

Faixa	Significado
100 – 199	Ação ou situação temporária
200 – 299	Indicação de sucesso
300 – 399	Informação de redirecionamento
400 – 499	Erro do lado cliente da comunicação
500 – 599	Erro do lado servidor da comunicação
600 – 699	Indicação de falha global

Abertura de conexão SIP

Para iniciar uma comunicação seguindo a arquitetura SIP, uma conexão deve ser estabelecida com o protocolo SIP, sobre TCP, UDP ou SCTP (experimentalmente). As mensagens utilizadas nesse procedimento são as INVITE e ACK, e as mensagens de resposta. A Figura a seguir apresenta uma troca típica de mensagens SIP para a abertura de uma conexão, que não utiliza qualquer elemento auxiliar da arquitetura SIP.



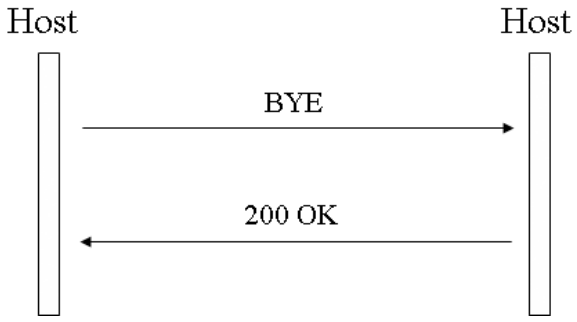
O pedido de abertura de conexão pode ser originado por qualquer terminal SIP que implemente essa funcionalidade, sendo esses elementos representados na Figura simplesmente como “host”.

A mensagem de requisição INVITE e a mensagem de resposta 200 OK carregam informações SDP, cada uma referente ao seu emissor. Dessa forma, é possível identificar um denominador comum entre os codecs possíveis, da mesma forma que ocorre com o H.245.

Diferente do padrão H.323, que determina que canais lógicos sejam abertos explicitamente pelo protocolo H.245 para a transmissão de datagramas RTP, a arquitetura SIP não especifica essa necessidade. Sendo assim, logo após a abertura de uma conexão SIP, os participantes já podem iniciar a troca de pacotes RTP contendo as mídias codificadas.

Encerramento de conexão SIP

Para encerramento de uma conexão SIP, mensagens específicas são utilizadas. A Figura a seguir apresenta uma troca de mensagens típica para encerramento de uma comunicação SIP.



Uma associação pode ser encerrada a qualquer momento da comunicação. Caso elementos especiais da arquitetura SIP estejam sendo utilizados, trocas de mensagens adicionais podem ser necessárias para o encerramento de uma comunicação SIP.

Elementos da Arquitetura SIP

A arquitetura SIP, assim como a arquitetura H.323, possui uma série de elementos que desempenham papéis específicos em comunicações multimídia em tempo real. Uma

arquitetura SIP típica é formada por terminais e opcionalmente por servidores proxies, de registro e de redirecionamento, além de agentes de notificação e gateways. O uso de nenhum desses elementos é obrigatório, com exceção dos terminais SIP.

As seções seguintes tratam das características e funcionalidades dos elementos da arquitetura SIP.

Terminais SIP

Os terminais SIP, assim como os terminas H.323, são os host que possibilitam que usuários acessem o serviço de comunicação multimídia em tempo real. Os destinos finais dos pacotes contendo dados multimídia são preferencialmente os terminais.

Esses elementos SIP são formados por agentes clientes e servidores, sendo eles, respectivamente, os produtores e consumidores efetivos do tráfego multimídia em tempo real, como voz e vídeo. Para desempenhar essas tarefas, a arquitetura SIP definiu os sub-elementos *client user agent* e *server user agent*. Esses sub-elementos são partes integrantes de um terminal SIP: para ser um terminal SIP, um host deve possuir um desses dois agentes, podendo ter inclusive os dois. Devido à presença desses agentes, os terminais SIP são, em alguns casos, chamados simplesmente de “agentes de usuário”.

Para poder iniciar uma conexão SIP, com o envio de uma mensagem INVITE, um client user agent deve ser implementado pelo terminal. Por outro lado, para ser capaz de receber um pedido de abertura de conexão, um server user agent deve ser implementado.

A mesma idéia que define os terminais SIP está presente na arquitetura H.323, porém sem essa terminologia. Nesse padrão, um terminal H.323 deve ser capaz de enviar e receber pedidos de abertura de conexão, não sendo possível apenas uma das possibilidades. Para o SIP, é possível que um terminal ou só receba ou só envie pedidos de abertura de conexão, embora as duas opções sejam possíveis.

Servidores de presença

Os servidores de presença, descritos na RFC 3265, são elementos da arquitetura SIP capazes de gerar notificações do estado de outros dispositivos, através de mensagens

NOTIFY. Os elementos que desejem esse serviço devem enviar um pedido de “registro” junto ao servidor de presença, através de mensagens SUBSCRIBE.

Algumas das informações fornecidas por esses servidores SIP são as situações de usuários, como ausente ou online, por exemplo.

Gateways

Os gateways, assim como ocorre com o padrão H.323, são elementos SIP utilizados para compatibilizar comunicações entre arquiteturas diferentes, como H.323 e o SIP. Outra utilização comum desses elementos é na interoperabilidade de redes SIP e PSTN, notadamente em ambientes de telefonia IP.

SIP Proxy

Os SIP Proxy são utilizados como intermediários em comunicações, de forma semelhante à proxies web. Eles recebem requisições, por exemplo, mensagens INVITE, de um terminal ou outro proxy e encaminham ou respondem as mensagens. Numa comunicação SIP, uma mensagem pode atravessar zero, um ou vários proxies.

Quando um proxy recebe uma requisição, ele pode utilizar uma base de dados ou um serviço de localização auxiliar para determinar qual é o próximo elemento SIP que ele encaminhará a mensagem.

Um SIP proxy não gera mensagens próprias, sendo apenas responsável pelo encaminhamento de mensagens. Uma exceção a esse procedimento são as mensagens CANCEL, que podem ser criadas diretamente pelo proxy.

Como vimos anteriormente, quando um mensagem é encaminhada por um proxy, seu endereço é adicionado ao campo Via da mensagem. Com isso, um terminal consegue saber, por exemplo, o caminho que determinada mensagem seguiu.

Há dois tipos de operação de um servidor proxy: *stateless* ou *statefull*. Os servidores proxy stateless não possuem qualquer informação a respeito das comunicações que estão utilizando esse proxy como ponto intermediário. Assim, para saber qual o próximo destino a encaminhar uma mensagem, os proxy SIP utilizam preferencialmente o campo Via contido

nos cabeçalhos das mensagens. Já os proxy statefull armazenam informações sobre as comunicações, o que permite o encaminhamento de mensagens de forma mais eficiente.

Servidores de redirecionamento

Os servidores SIP de redirecionamento são utilizados para informar onde localizar determinado recurso, quando este não está mais disponível na localização requisitada. Quando um servidor SIP mudar de endereço, por exemplo, o servidor de redirecionamento pode informar a localização atual desse servidor.

Algo interessante a se notar em relação aos servidores de redirecionamento é que eles não encaminham mensagens SIP, como fazem os servidores proxy. Esses elementos estão apenas responsáveis por responder requisições, informando sobre mudança de endereços. As mensagens de redirecionamento, como vistas anteriormente, possuem identificação na classe de redirecionamento, com código de identificação contido na Faixa 300 a 399.

SIP Registrar

Os servidores de registro, conhecidos como SIP Registrar, desempenham um papel importante na arquitetura SIP. Eles possuem uma função semelhante aos gatekeepers H.323, no tocante a tradução de endereços simbólicos (alias) e endereços IP.

Assim como ocorre com os gatekeepers, os usuários podem se registrar junto a um Registrar. Para criar e consultar registros nesses servidores, é utilizado a mensagem SIP REGISTER. Essa mensagem contém informações importantes, como o endereço SIP do usuário, seu endereço IP atual e endereços IP alternativos, caso existam.

Os serviços de mobilidade da arquitetura SIP estão fortemente baseados nos servidores Registrar. Os host que são móveis, e que por isso podem trocar de endereço IP com certa freqüência, devem manter atualizado seu registro nos Registrar, de forma a serem localizados para novas comunicações.

Para encontrar um servidor de registro, um terminal SIP pode utilizar o endereço do grupo multicast 224.0.1.75.

Leitura complementar

RFC 3261 – SIP: Session Initiate Protocol

Endereço: http://www.ietf.org/rfc/rfc3261

RFC 2327 – SDP: Session Description Protocol

Endereço: http://www.ietf.org/rfc/rfc2327.txt

Página do SIP Forum

Endereço: http://www.sipforum.org/

Dr. Daniel G. Costa:

http://www.uefs.br/danielgcosta

danielgcosta@uefs.br