WHITE PAPER

SIP — an introduction

September 2024

AXIS COMMUNICATIONS

Summary

The session initiation protocol (SIP) provides an additional interface for system integration for security products. SIP is a widely adopted standard in the telecommunications industry, and it provides increased flexibility for inter-connectivity and everyday use. Open, standardized interfaces are requested by system integrators, developers, and end-users, increasing the value offered to them, as products can be used in a variety of systems. Axis products with SIP support are intended for use in both security solutions and communication solutions.

Setting up a SIP system can be easy. However, in the case of complicated network topologies or when security requirements and extra call handling functionality is required, SIP server and NAT traversal techniques need to be used, requiring more technical understanding on the part of the installer or technician.

Table of Contents

1 Introduction

Session Initiation Protocol (SIP) is used to initiate, maintain, and terminate multimedia sessions between different parties. These sessions usually consist of audio, but sometimes they also include video. SIP is the standard protocol used in Voice over IP (VoIP) applications and Unified Communications (UC) platforms (see Section 3).

SIP is a way to connect, integrate, and control your Axis network products. It's supported by all Axis network speakers, all Axis network intercoms, and selected Axis system devices and Axis cameras.

2 How does it work?

In order to communicate using SIP, at least two SIP clients are needed. A SIP client can be a SIP hardphone, softphone, mobile client, or SIP-enabled Axis product.

Each SIP client is assigned its own SIP address. A SIP address is similar to an email address, but with the prefix "sip:"

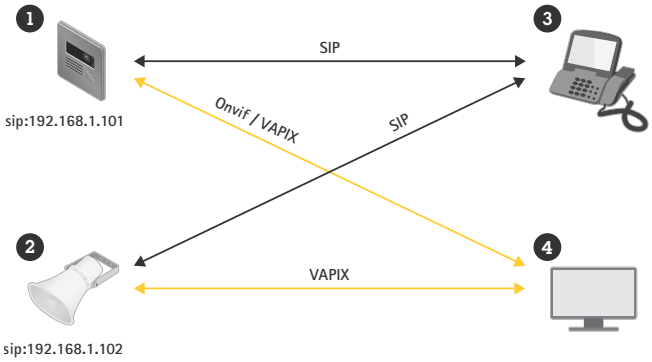
For example, sip:bob@axis.com [sip:<user@]<provider>. This identifier can be used across a number of devices and is analogous to a telephone number linked to a SIM card, which can be used in a number of devices.

2.1 Peer-to-peer setup – the simple way

A SIP system can take many forms. In its simplest form, the system consists of two or more SIP User Agents (UA) communicating directly with each other. This can be called a peer-to-peer setup, a direct call setup, or local setup. A typical SIP address in this case would take the form sip:<local-ip>, for example, sip:192.168.0.90

Example: In a simple setup these Axis products (1, 2) can use SIP for setting up audio and/or video communication with other SIP devices (3) on the same network, without needing a server or PBX.

At the same time they can be connected as any other Axis device to the video management system (4) using the open APIs VAPIX, or ONVIF Profile S.



To make a peer-to-peer call from one UA to another on a local network, all that's needed is the SIP address containing the unit's IP address.

2.2 Using a SIP server (PBX) – adding more possibilities

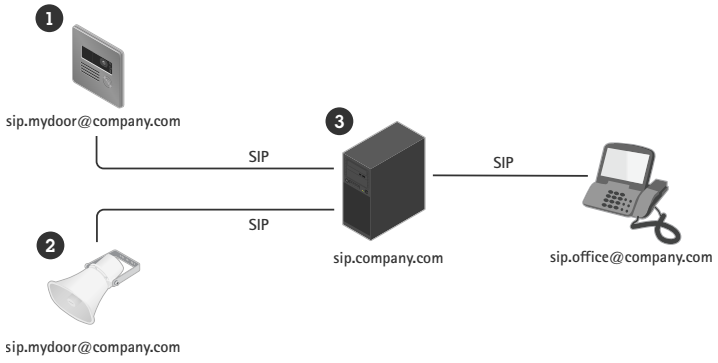
A SIP-based VoIP infrastructure scales very well. The next step up in size is to use a SIP server, or Private Branch Exchange (PBX), as a central hub. The SIP UAs register with the server's registrar and can then reach other UAs simply by dialling an extension on the PBX.

A typical SIP address in this case would use the form sip:<user>@<domain>. Alternatively, it could be sip:<user>@<registrar-ip> for example sip:6007@ mysipserver.net. A PBX works like a traditional switchboard, showing the clients' current status, allowing call transfers, voicemail, redirections, and much more.

A SIP server usually includes proxy, registrar, and redirect functionality. Proxies route calls and provide additional logic to incoming calls. Registrars accept register requests and act as a location service for the domain that it handles. Redirect servers redirect the client to contact an alternative SIP address.

The SIP server can be set up as a local entity or it can be located offsite. It can be hosted on-prem or in the cloud. When making SIP calls across sites, calls are normally initially routed through a set of SIP proxies. These proxies query the location of the SIP address to be reached.

Example: Axis products (1, 2) can connect to a SIP server (3) locally or offsite. The server handles the setup and termination of calls between SIP devices on the local network or over the internet. In this setup the device's SIP address is independent of its IP address and the SIP server makes the device accessible as long as it is registered to the server.



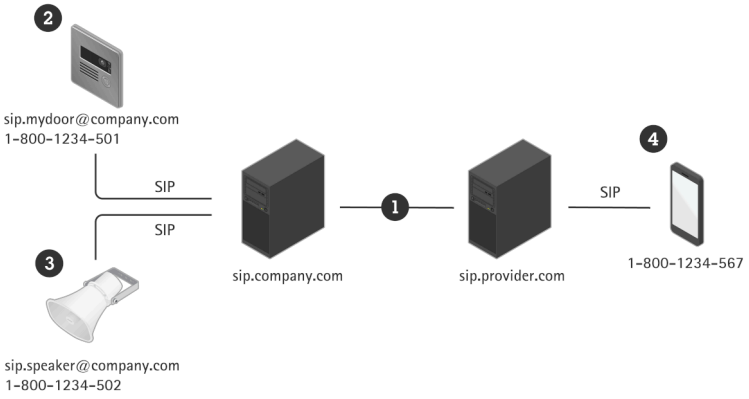
To use your device with a SIP server you must create an account on the server with a specified user ID and password. To register your device with the server, you need to set up an account on the device, entering the server address, user ID, and password.

2.3 Using a SIP trunk – assigning a telephone number

Using a SIP trunk, SIP UAs can be switched to the traditional telephone network (PSTN). In this way you can even assign a regular telephone number to the SIP UA.

Cloud-based SIP trunking is a modern approach that leverages the internet to deliver calls and other communication services. This method eliminates the need for physical phone lines, making it easier to integrate with cloud solutions and VoIP systems.

Example: Using a SIP trunk (1) with a service provider you can assign external phone numbers to your devices (2, 3). In this way you can make calls between a network speaker or network intercom and regular telephones (4).



When used with a SIP trunk, the device connects to the server in the way described above.

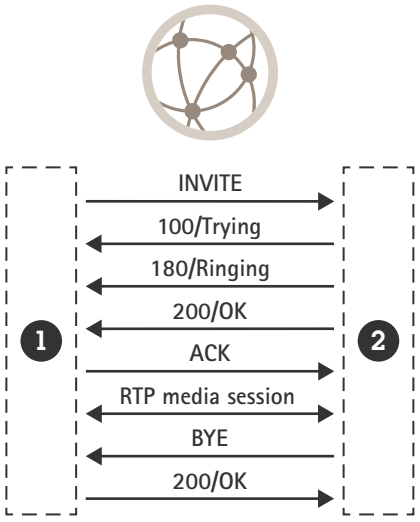
3 Unified Communications (UC)

Unified Communications (UC) refers to the integration of various communication tools and technologies into a single, cohesive system. SIP plays a vital role in UC by enabling seamless interaction between different communication channels, such as voice, video, instant messaging, and presence information. By leveraging SIP, organizations can create a unified communication environment that enhances collaboration, improves productivity, and provides a consistent user experience across multiple devices and platforms.

UC can be provided locally in on-prem solutions or via cloud solutions (UC as a service – UCaaS). Examples of service providers for the cloud solutions are Cisco Webex, Microsoft Teams, and Zoom.

4 Inside a normal SIP call

In order to make a SIP call a sequence of steps are performed to exchange information between the UA initiating and receiving the call.



When initiating a call, the Initiator UA (1) sends a request or an INVITE to the Recipient UA (2) SIP address. The INVITE contains a Session Description Protocol (SDP) body describing the media formats available and contact information for the initiator of the call.

Upon receiving the INVITE, the recipient immediately acknowledges this by answering with a 100 TRYING response.

The receiving UA then compares the offered media formats described in the SDP with its own. If a common format can be decided on, the UA alerts the recipient that there is an incoming call and sends a provisional response back to the initiating UA - 180 RINGING.

When the recipient picks up the call, a 200 OK response is sent to the initiator to confirm that a connection has been established. This response contains a negotiated SDP indicating to the initiator which media formats should be used and where to send the media streams.

The negotiated media streams are now set up using the Real-time Transport Protocol (RTP) with parameters based on the negotiated SDP, and the media travels directly between the two parties. The initiator sends an acknowledgement (ACK) via SIP to acknowledge that it has set up the media streams as agreed. The SIP session is still active but it is no longer involved in the media transfer.

When one of the parties decides to end the call, it sends a new request – BYE. Upon receiving a BYE, the receiving party acknowledges this with a 200 OK and the RTP media streams are then stopped.

4.1 SDP – negotiating the format to use

Session Description Protocol (SDP) is a format for describing streaming media initialization parameters. The SDP body contains information about which media formats (that is, codecs) are supported by the clients and the clients' preferred codec selection order.

Typical audio codecs used for SIP calls are PCMU, PCMA, G.722, G.726, and L16. If multiple overlapping codecs are supported by both the initiator and the recipient, the codec with the highest priority on the recipient side will normally be selected. The choice of codecs ultimately affects the bandwidth, so careful consideration should be taken to meet compatibility requirements to other SIP UAs and to maintain a bandwidth requirement that suits the use case. For example, in a local network where all clients support L16, the choice of uncompressed audio works well. However, if the SIP UA is to be accessed via the internet through a mobile phone, PCMU is a better choice.

4.2 Calls in complex SIP infrastructure

In a more complex SIP infrastructure setup, the initiation looks a little different, as the SIP session is set up step-by-step for each hop. However, once the SIP session is set up, traffic is normally not routed, but instead travels directly between the different parties, as in the previous example.

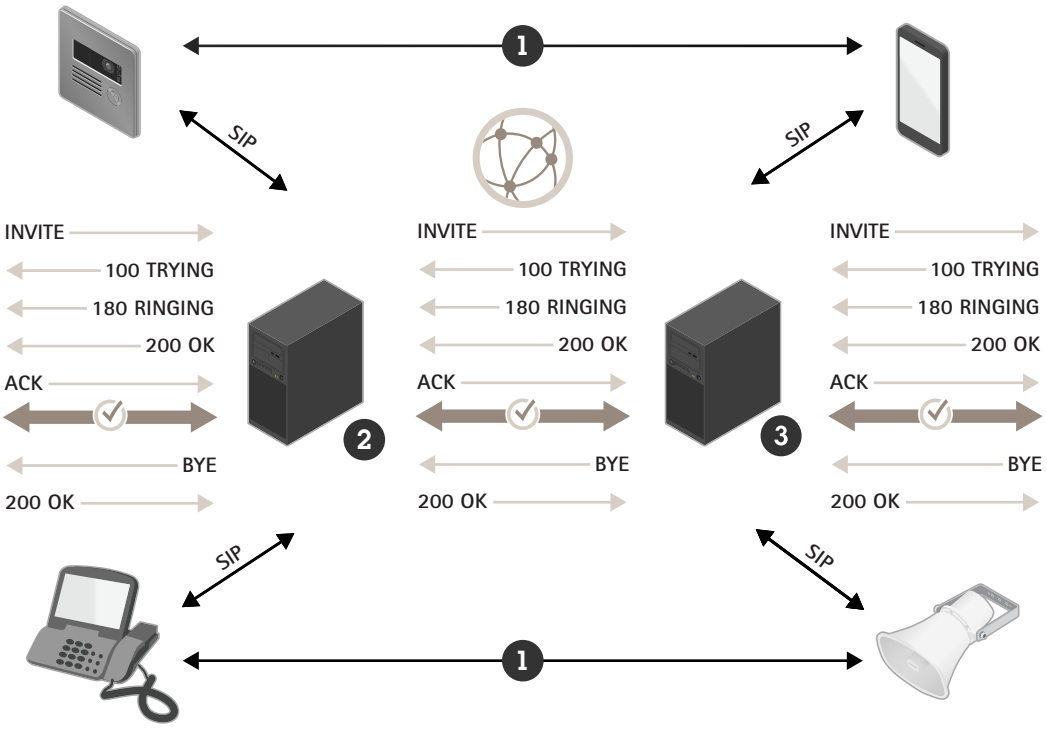


Figure 1. Call setup in a complex SIP infrastructure. Media RTP streams (1) travel directly between the parties once the SIP session has been set up between initiator (2) and recipient (3).

5 DTMF – sending commands in SIP calls

Dual-Tone Multiple-Frequency (DTMF) is a format used to send information over a telephone connection. DTMF signals can be sent in SIP calls and can be used to give instructions to a SIP device. The DTMF character range consists of digits 0-9, letters A-D, * and #.

For example, in a call to a SIP-enabled intercom, the DTMF character '5' could be sent from the phone's keypad, which can be configured to be interpreted by the receiver as the command for unlocking the door.

There are three different ways of sending DTMF in a SIP call:

- The traditional in-band method, where the signal is actually an audio pulse interleaved with the audio stream. However, this is unreliable and only works with non-compressed codecs.
- The SIP INFO method, where the DTMF character is sent in a SIP message in the signaling stream. This method is very reliable and out-of-band, but only has limited support.
- The RTP method (RFC2833), where the DTMF character is encoded as an RTP package and sent out-of-band. This is the de-facto standard and enjoys broad support.

6 Complex environments and greater security

Complex network environments, such as corporate networks, can cause difficulties when using SIP. The same is true if you wish to use encryption.

6.1 NAT traversal – navigating complex networks

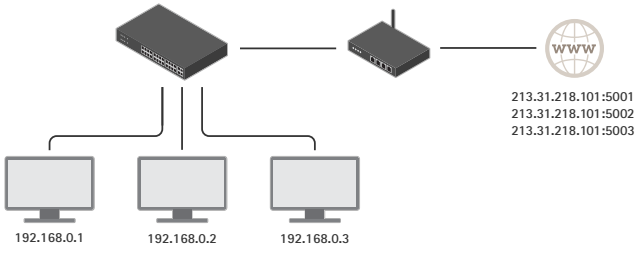
In a more complex network environment, it may be necessary to utilize Network Address Translation (NAT). NAT is a way to publicly represent IP addresses located on a private local network. This means that all units in a private sub-network share a common IP address prefix, for example, 192.168.1.XXX. This is the address they use when communicating with each other. When they communicate with another network, this address is translated to the router's public address, and appended with a port mapping

- 192.168.1.24 => 184.13.12.33:44221
- 192.168.1.121 => 184.13.12.33:24325, and so on.

As the translation table is stored in the router, in most cases it is not possible for an external user to learn the address of a NAT:ed device. When communicating over SIP, this can result in one of the following problems:

- Unable to initiate, update or terminate a session, that is, not possible to call, hold or hang up.
- No media stream(s).
- One-directional media stream(s).

NAT traversal: NAT changes the source address of each packet to a public IP address with different source ports.



To solve these issues, SIP supports three different NAT techniques:

- STUN – This is a way of asking a server at a known location what the unit's public address is. The STUN server returns the public IP and port mapping used to make the request. The result is then used in the signaling and media transfer, and this works in most situations.

- TURN – When using TURN, all traffic is relayed through a known server. This adds an extra overhead as the machine hosting the TURN server must be powerful enough to route all media for each client using the service. This is a more expensive solution, but can work in some situations where STUN does not.
- ICE – The ICE protocol gathers all IP addresses it can find that are related to a SIP UA and then tries to calculate which one should be used. When used in combination with STUN and TURN on both the initiating and receiving SIP UA, it increases the chances of successfully establishing SIP calls.

6.2 Using encryption with SIP

SIP signalling traffic is normally sent over the connection-less UDP protocol. It can also be sent over TCP, in which case it can also be encrypted with Transport Layer Security (TLS).

To ensure that a secure connection is used for a call, the SIP protocol utilizes an addressing scheme called Secure SIP (SIPS), which requires that the transport mode is set to TLS. When making a call, the dialled SIP address is prefixed with "sips:" rather than "sip"; for example: sips:bob@biloxi.ex.com instead of sip:bob@biloxi.ex.com. This mandates that each hop must be secured with TLS and requires the receiving end to employ the same security level. Calling a sip-prefixed address when using TLS only ensures that the first hop is encrypted.

To obtain the highest level of security, the following measures should be taken:

- Transport mode should be set to TLS.
- The sips prefix should be used at all times.
- SIP INFO should be used for sending DTMF tones, as this is sent in the encrypted channel.

Note that not all clients support Secure SIP.

7 SIP terminology

API	Application Programming Interface
Codec	Coder-decoder
Hardphone	Hardware that makes telephone calls, that is, a phone
ICE	Interactive Connectivity Establishment
IP	Internet Protocol
Mobile client	Software program on a mobile device that makes telephone calls
NAT	Network Address Translation
PBX	Private Branch Exchange
PSTN	Public Switched Telephone Network, that is, the regular telephone network
RTP	Real-time Transport Protocol
SDP	Session Description Protocol
SIM	Subscriber Identity Module
SIP	Session Initiation Protocol

SIP server	Main component of an IP PBX. Handles call setup and call tear down. Also called SIP proxy or registrar.
SIPS	Secure SIP
SIP URI (SIP address)	Uniform Resource Identifier. The unique address of the SIP UA.
Softphone	Software program that makes telephone calls
STUN	Session Traversal Utilities for NAT
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TURN	Traversal Using Relays around NAT
UA	User agent. Both end-points of a communication session.
UC	Unified Communications
UDP	User Datagram Protocol
VoIP	Voice over IP

T10175991/EN/M4.6/2409

About Axis Communications

Axis enables a smarter and safer world by creating solutions for improving security and business performance. As a network technology company and industry leader, Axis offers solutions in video surveillance, access control, intercom, and audio systems. They are enhanced by intelligent analytics applications and supported by high-quality training.

Axis has around 4,000 dedicated employees in over 50 countries and collaborates with technology and system integration partners worldwide to deliver customer solutions. Axis was founded in 1984, and the headquarters are in Lund, Sweden