SIP – An introduction
1. **Introducing SIP**

Session Initiation Protocol (SIP) is used to initiate, maintain and terminate multimedia sessions between different parties. Usually these sessions consist of audio, but sometimes they consist of video. SIP is the standard protocol used in Voice over IP (VoIP) applications and unified communication platforms.

SIP is for example supported by AXIS C3003-E Network Horn Speaker and AXIS A8004-VE Network Video Door Station and constitute a new way to connect, integrate and control your Axis network products.

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 looks like an email address with a "sip:" prefix, for example, sip:bob@biloxi.ex.com [sip:<user>@<provider>]. This identifier can be used across a number of devices and is analogous to the use of a telephone number linked to a SIM card that 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 look like sip:<local-ip>, for example, sip:192.168.0.90

In order to make a peer-to-peer call from one UA to another on a local network all that is needed is the SIP address containing the unit’s IP address. Note that not all SIP clients support peer-to-peer calls.

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 registrar, Private Branch Exchange (PBX) or SIP server, as a central hub. The SIP UAs register with the registrar and can then reach each other's SIP UAs simply by dialling an extension on the PBX. A typical SIP address in this case would look like sip:<user>@<domain> or sip:<user>@<registrar-ip> such as 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 an intranet or it can be hosted by a third-party provider. 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 can connect to a SIP server running either locally or hosted by a third-party provider. The server handles the setup and termination of calls between SIP devices on the local network or over the Internet.

In this setup the SIP address of the device is independent of its IP address and the SIP server makes the device accessible as long as it is registered to the server.

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

2.3 Using SIP trunks – assigning a telephone number

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

Example:
Using a SIP trunk with a service provider you can assign an external phone number to your device. This way you can make calls between a network speaker or door station and regular telephones.

When used with a SIP trunk the device connects to the server in the way described above. The service provider will usually charge extra for external numbers.

3. 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 sends a request or an INVITE to the recipient’s 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, 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 decides to pick 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 to where the media streams should be sent.

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.

3.1 SDP – negotiating what format is used

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 3G mobile phone, PCMU is be a better choice.

3.2 Calls in complex SIP infrastructures

In a more complex SIP infrastructure setup, the initiation looks a bit 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, instead travelling directly between the different parties as in the previous example.
4. 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 numbers 0-9, letters A-D, *, and #.

For example, while in a call to a SIP-enabled door station, 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. However, there is 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 it has wide support.

5. Complex environments and higher security

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

5.1 NAT traversal – navigating complex networks

In a more complex network environment, it may be necessary to utilize Network Address Translation (NAT) traversal techniques. NAT is a way of translating IP addresses on a private local network to a public representation. This means that all units in a private subnetwork 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 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, it is not possible to call, hold or hang up.
> No media stream(s).
> One directional media stream(s).
To solve these issues, SIP supports three different NAT traversal 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 it 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 is not working.

- **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.

### 5.2 Using encryption together with SIP

The SIP signalling traffic is normally sent over the connectionless UDP protocol. It can also be sent over TCP and in that 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”; that is, 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 safety 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.

### 6. Summary

The session initiation protocol (SIP) provides an additional interface for system integration for security products. It is a widely adopted standard in the telecommunications industry. Having this interface provides increased flexibility for interconnectivity 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 made to be used both for security and communication.

Setting up a SIP system can be very 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 of the installer or technician.
## 7. The SIP dictionary

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3G</td>
<td>Third generation mobile telecommunications technology</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>codec</td>
<td>coder-decoder</td>
</tr>
<tr>
<td>DTMF</td>
<td>Dual-Tone Multi-Frequency signaling</td>
</tr>
<tr>
<td>Hardphone</td>
<td>Hardware that makes telephone calls, that is, a phone</td>
</tr>
<tr>
<td>ICE</td>
<td>Interactive Connectivity Establishment</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>Mobile client</td>
<td>Software program on a mobile device that makes telephone calls</td>
</tr>
<tr>
<td>NAT</td>
<td>Network Address Translation</td>
</tr>
<tr>
<td>PBX</td>
<td>Private Branch Exchange</td>
</tr>
<tr>
<td>PSTN</td>
<td>Public Switched Telephone Network, that is, the normal telephone network</td>
</tr>
<tr>
<td>RTP</td>
<td>Real-time Transport Protocol</td>
</tr>
<tr>
<td>SDP</td>
<td>Session Description Protocol</td>
</tr>
<tr>
<td>SIM</td>
<td>Subscriber Identity Module</td>
</tr>
<tr>
<td>SIP</td>
<td>Session Initiation Protocol</td>
</tr>
<tr>
<td>SIP server</td>
<td>Main component of an IP PBX Handles call setup and call tear down Also called SIP Proxy or Registrar</td>
</tr>
<tr>
<td>SIPS</td>
<td>Secure SIP</td>
</tr>
<tr>
<td>SIP URI (SIP address)</td>
<td>Uniform Resource Identifier The unique address of the SIP UA</td>
</tr>
<tr>
<td>Softphone</td>
<td>Software program that makes telephone calls</td>
</tr>
<tr>
<td>STUN</td>
<td>Session Traversal Utilities for NAT</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TLS</td>
<td>Transport Layer Security</td>
</tr>
<tr>
<td>TURN</td>
<td>Traversal Using Relays around NAT</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UA</td>
<td>User Agent Both end points of a communication session</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
</tbody>
</table>
About Axis Communications

Axis offers intelligent security solutions that enable a smarter, safer world. As the global market leader in network video, Axis is driving the industry by continually launching innovative network products based on an open platform - delivering high value to customers through a global partner network. Axis has long-term relationships with partners and provides them with knowledge and ground-breaking network products in existing and new markets.

Axis has more than 2,000 dedicated employees in more than 40 countries around the world, supported by a network of over 75,000 partners across 179 countries. Founded in 1984, Axis is a Sweden-based company listed on NASDAQ Stockholm under the ticker AXIS.

For more information about Axis, please visit our website www.axis.com.