Technical guide to network audio

Technologies and considerations for a world of audio possibilities.
May 2024

Introduction

Axis audio systems were not created with the audio expert in mind – they were created for those who want to install an audio system and be confident that it just works, and sounds good. Without being audio engineers. IP audio from Axis is all about smart, user-friendly solutions with everything built in. This ensures that you can navigate and resolve your audio challenges in an easier and more efficient way than with traditional audio engineering.

This *Technical guide to network audio* was created to support you in your work with Axis network audio systems and help you stay up-to-date with the changing technological landscape. The guide is especially intended for you as an Axis partner or system integrator, but really is for anyone interested in the possibilities of IP audio.

The guide provides a complete overview of network audio, including Axis network audio products and complete systems, the technologies which enable audio over IP, and the basics of audio and acoustics. If you are rather looking for step-by-step guides or information about how to set up a device or a system, check Axis support, *www.axis.com/support* or Axis Communications Academy, *www.axis.com/learning/academy*

In this guide we focus on audio use cases related to speakers and audio broadcast, such as public address, audio for security, and background music.

Table of Contents

1. Network audio: overview, purposes, and benefits

With network audio, digital audio streams are transported over an IP network. This provides many benefits compared with traditional systems that send analog audio through dedicated audio cabling. Axis network audio allows you to control all your speakers from a single user interface. From remote, you manage your audio content, schedule and reschedule the content, configure and reconfigure your audio zones, and monitor the health of all your devices.

1.1 Overview of a network audio system

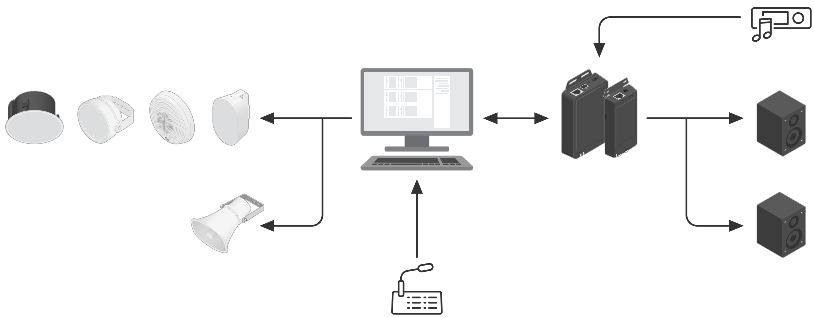

Figure 1.1a *A network audio system can be controlled from a single point via an intuitive interface. With system devices you can also control analog audio sources and analog speakers as part of the network audio system.*

Network audio uses an IP network and standard IT equipment as the backbone for transporting and managing audio. The core components are network speakers, network microphone consoles, the network itself, a server, and audio management software. Network audio from Axis can also include system devices that provide a smooth migration from a legacy (analog) audio system to network audio.

The network, the server, and the storage components are all common off-the-shelf equipment. As the speakers and microphone consoles are computer-based, they have capabilities that cannot be matched by analog audio systems. Axis network audio is digital all the way.

1.2 Purposes

Network audio is often integrated in security systems, where video surveillance and voice messages make a powerful combination. But network audio is also used for safety systems, public address announcements, and background music.

> **Improve security.**
> Network audio is a valuable addition to a video-based security installation. Perimeter protection is just one example: imagine a potential intruder climbing a fence. A network camera registers the event and alerts security personnel. They can then easily communicate a warning to the intruder, "We can see you, you're trespassing." This type of warning is often sufficient, allowing you to reserve any stronger security responses for when they are really needed to protect property. But Axis network speakers also add a new dimension to a security system with their integrated microphones. This means that you can use the speakers as active security sensors and run audio analytics for detection of, for example, breaking glass or gunshots. The microphones can also be used for two-way audio. If the placement of your speaker is not optimal for audio detection or communication, you can integrate standalone microphones into the system.

> **Increase safety.**
> With network audio, urgent safety information can be quickly and efficiently broadcast to a very large number of people. In cities, school lock-downs, or in areas that are under threat of earthquakes or other natural disasters, a public address system based on network audio can have a role in a larger safety system. Speakers can be used for emergency callouts regarding evacuation, reminders about wearing hard hats on construction sites, or other warnings to keep people safe.

> **Improve operational efficiency.**
> In public address systems, network audio provides a flexible way to manage and play non-critical information, such as informative messages and updates in schools, retail stores, hotels, and public buildings. You can make live announcements calling someone (or *paging* someone) to come to a specific area (such as a colleague to the clothing department), make scheduled announcements (for instance, about the start of the school day), or issue live or triggered announcements during an emergency. You can broadcast to single or multiple zones to make callouts to specific rooms.

Network audio can also create ambiance with background music. It is easily managed with both central and remote control, and preset volume and music choices. The system is multipurpose — background music can be combined with scheduled and live announcements for the best customer experience, and you are in control of which audio source has priority.

1.3 Video surveillance with audio

A security solution with audio and video can be monitored or unmonitored with different levels of camera integration.

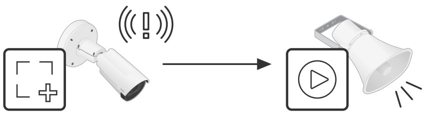

Figure 1.3a *Unmonitored solution: a camera with analytics detects motion or other events in a restricted area and sends a command to a speaker to play a prerecorded audio message.*

Figure 1.3b *Monitored solution: a camera with analytics detects motion or other events and sends a notification to an operator. The operator can choose a prerecorded audio message or speak live through the speaker.*

1.4 Two-way audio

Axis devices support two-way audio, also known as duplex audio. Axis speakers have a built-in microphone so that they can both pick up and broadcast sound, enabling conversations (talkback). The microphone also enables speaker-test functionality for remote health testing, so you know that your system is working.

Thanks to external microphone input, you can connect an additional microphone for situations where the speaker is installed at a larger distance from the intended listeners. If you have an IP camera with two-way audio, you can both see and hear what is happening on your premises and, if the need arises, speak with visitors or intruders in real-time.

With *half duplex* you can send and receive audio (talk and listen) in one direction at a time, similar to a walkie-talkie conversation. The direction is controlled automatically through voice detection software or manually through a push-to-talk button. There is no risk of echo problems because speaker and microphone are never active at the same time.

With *full duplex* you can send and receive audio (talk and listen) at the same time, similar to a phone conversation. The speaker uses advanced echo cancellation that prevents the speaker's sound from being fed back to the microphone.

1.5 Benefits of network audio

A fully digital network audio system provides many benefits and advanced functionalities.

> **Connects to standard networks.** Network audio consists of true IT devices that connect to the standard network, with no need for dedicated audio cabling. Axis network audio devices are powered through PoE (Power over Ethernet), meaning that they receive power through the same Ethernet cable that transports the audio data. This keeps installation costs low and increases system reliability. You don't need any power outlet by the speakers and no extra audio cables or technicians to calculate cable dimensions and capacity. If you prefer to keep using legacy audio cabling or analog speakers, there are specific system devices that enable you to do that while still enjoying some of the digital benefits.

> **Digital all the way.** In Axis network audio, the audio is digital all the way from the source to the speaker. The sound is digitally stored, processed, and transmitted, without any analog-to-digital or digital-to-analog conversions. Compared with analog audio, this means that the sound quality and signal strength is maintained. The integrated digital signal processor makes sure that the sound is optimized right in the speakers. No technician has to go on site to control the audio quality. The sound signal remains strong no matter how long the cables are. Audio announcements are clear and easy to understand.

> **Central accessibility from remote**. Network audio devices can be configured and accessed remotely. This enables authorized users to control them from virtually any networked location in the world.

> **Health monitoring from remote**. Axis speakers come with a self-test functionality that works from remote. This is a way to check that the speaker is in working order, by having it provide audio feedback to the system. No technician has to go on site to check the speakers, and there is no risk that a speaker is malfunctioning without you knowing.

> **Easy, future-proof integration**. The use of open-standard IP technology simplifies integration with other systems. Speakers can be integrated directly into a VMS (video management

system) or a standard Voice over IP (VoIP) phone system. With one, integrated system, it is easy to comply with changing needs.

> **Scalability and flexibility**. IP offers flexibility and broad ability for future reconfiguration, in case of any unforeseen changes in system requirements. Speaker groups and zones can be reconfigured without the need for new cabling. The audio system can be expanded or changed when needed. This is easily done in the software.

> **Everything built in**. A traditional audio system generally needs a dedicated technical room to place the rack cabinets with all audio components, such as power amplifiers, digital signal processors, and more. With Axis network audio, these tasks are done through edge processing in the speakers. Axis speakers are complete, high-quality audio systems in themselves and there is no need for a technical room. Even audio management software for handling zoning, content, and scheduling is built in.

> **Cost-effectiveness**. The many benefits of network audio contribute to a low total cost of ownership. The system needs less cabling and less components compared with analog audio. Health monitoring and maintenance from remote means less staff hours and less failure. Easy integration and scalability also help keeping costs low. Furthermore, network audio is cost-effective because you can use the same system for many purposes, such as issuing live announcements, playing prerecorded warnings, and creating atmosphere with background music.

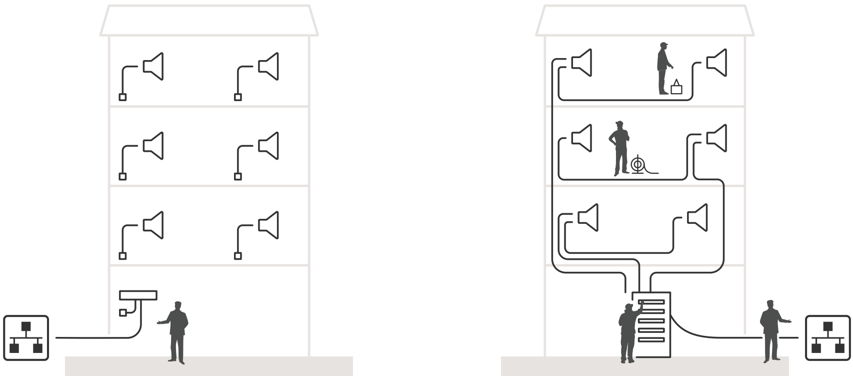

Figure 1.5a *Network audio (left) requires less cabling and less components compared with analog audio (right). Network audio devices connect to the standard network, and receive power through the same Ethernet cable that transports the audio data.*

1.6 Applications in key industry segments

Network audio can be used in an almost unlimited number of applications, both for security or safety purposes and for operational efficiency. Typical use in key industry segments include:

> **Smart cities.** Network audio can be a useful, crime-deterring complement to network video surveillance in cities. It can also increase safety by guiding people in cases of emergency.

> **Retail.** Network audio can provide retail stores with a flexible and easily managed system for making announcements and callouts to staff or customers. It can be used to deter shoplifters and loiterers through the right voice messages, but also to play commercial announcements and background music from a mobile phone or a phone system.

> **Critical infrastructure.** In power plants and heavy industries, network audio systems can provide announcements and warnings regarding safety hazards and incidents such as gas leaks, oil spill, or chemical leaks. They can issue safety reminders, such as "wear your hard hat".

> **Industrial.** In warehouses, datacenters, and logistics centers, network audio can provide triggered or live announcements about security concerns and incidents. Messages can also be related to business efficiency, such as reminders about shift changes and taking breaks, and you can play background music for staff well being.

> **Education.** In a school or on a campus, network audio can provide a flexible public address system for information callouts. It can also increase security and safety through live or triggered announcements with instructions and warnings in case of incidents. The possibilities

for two-way audio enables communication between school administration and teachers in classrooms. The network audio system can also be used for sound reinforcement in the classroom through connection of presentation equipment.

> **Transportation.** In train stations, subways, and tunnels, network audio can be used for security callouts to prevent vandalism and theft, and warnings and safety instructions in case of incidents. It can also be used to communicate valuable information about delays or schedule changes.

Figure 1.6a *Network audio can be installed for security or safety purposes, or for operational efficiency.*

2. Network audio hardware

Network speakers form the physical foundation for network audio. They can be of various types depending on the purpose of the system. Audio system devices provide possibilities to connect analog audio and network audio, and microphone consoles may be used to complete the system with live public address functionality. Axis provides everything you need for a complete audio system.

2.1 Network speakers

Axis network speakers are complete, high-quality audio systems in themselves. They are powered by Power over Ethernet (PoE) technology and connect to standard networks.

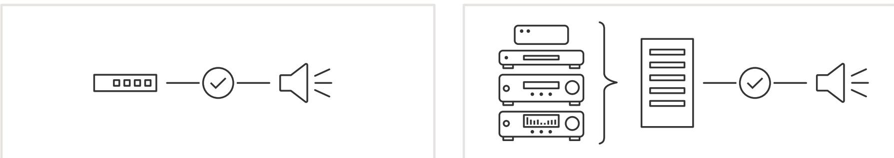

Figure 2.1a *Left: Network audio speakers are complete audio systems. Right: Traditional speakers requires additional hardware.*

Every Axis speaker has:

> Integrated amplifier and digital signal processor for preconfigured sound.

> Built-in audio management software, which provides support for live or prerecorded announcements, background music, audio content scheduling, zoning, and priority of audio sources.

> Integrated microphone.

> Two-way audio, which is enabled by the integrated microphone.

> Built-in test function for remote health verification, which is enabled by the integrated microphone and the use of test tones.

> Onboard memory for storing audio clips.

> I/O ports for integration with additional systems and devices.

> Integrated LED (in some speaker types) for visual status confirmation.

2.1.1 Speaker types

Form factors, sound pressures, and mounting possibilities vary — some speaker types are optimal for conveying clear and audible announcements in noisy outdoor areas, while others work better in small spaces.

Figure 2.1b *Axis speakers.*

Horn speaker. An Axis network horn speaker has a very high sound pressure level and maximizes the loudness of those frequencies to which the human ear is the most sensitive. This means that a message can be conveyed as clearly as possible. Due to its shape, the speaker directs all sound in one direction, which further enhances the sound pressure. A horn speaker can be used in noisy indoor areas like warehouses and plants, or in outdoor installations. It can be mounted on a pole or a wall. Axis also offers an explosion-protected network horn speaker that is certified for use in hazardous areas.

Sound projector. An Axis network sound projector has a high sound pressure level and natural, rich sound. This means that a message can be conveyed as clearly as possible, but that background music will sound good too. A sound projector can be used in outdoor installations or noisy indoor areas and can be mounted on a pole, wall, or ceiling. It can be installed in easy-to-reach locations where the risk for vandalism is higher — the sound projector is vandal-resistant and also has a sleek, minimalistic design that easily blends into the environment.

Ceiling speaker. An Axis network ceiling speaker provides a medium sound pressure level and should be used in less noisy indoor or outdoor areas, such as hospitals, retail stores, or office buildings. It can be mounted in a drop ceiling where it will be very discreet and physically well-integrated.

Pendant speaker. An Axis network pendant speaker has a medium sound pressure level and is suitable for less noisy indoor areas with high ceilings. It comes in two sizes, and the cable length can be adjusted to fit any high ceiling.

Mini speaker. An Axis network mini speaker provides a low sound pressure level and should be used in quieter indoor areas. It is small and discreet and fits into small spaces or corridors, where it can be surface mounted on a wall or ceiling. It has a wide audio coverage which means that you need fewer speakers. The mini speaker has a built in PIR sensor for motion detection, which can be set up so that the speaker automatically plays an audio message when someone is approaching.

Cabinet speaker. An Axis network cabinet speaker provides a medium sound pressure level. It can be used in most indoor areas, but is less optimal in very noisy environments. It can also be used semi-outdoors, which means it can be mounted below a roof that protects it from heavy rain. The cabinet speaker can be mounted horizontally or vertically, on a wall, in a ceiling, or with a pendant kit.

2.2 Audio system devices

There are audio system devices that enable a smooth migration from analog to network audio. These devices make it possible to combine legacy equipment, such as analog speaker systems with or without amplifiers, with network audio equipment and gain the benefits of network audio without having to replace all the equipment at once. Axis offers both a network audio bridge and a network audio amplifier for this end. There is also a volume controller that facilitates local volume adjustment and source selection.

2.2.1 Network audio bridge

The network audio bridge from Axis is a versatile, small device that connects and combines analog and network audio systems. It can be powered by PoE or use an ordinary power supply.

Analog input. The network audio bridge allows you to connect various audio equipment that uses analog interfaces and use them as input to digital systems. You can, for example, connect a laptop for a presentation, or a mobile phone for music playback, but also music players or projectors.

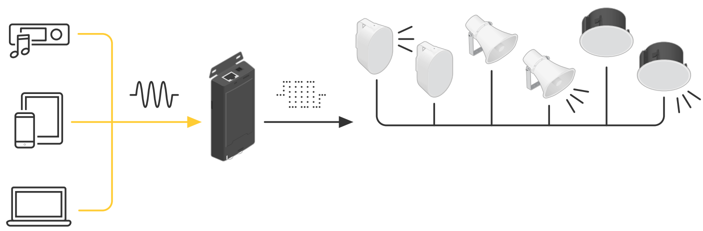

Figure 2.2a *A network audio bridge combines analog audio sources with network speakers.*

Network audio bridge also enables you to continue to use an existing, analog audio system while expanding it with network audio for added functionality and future proofing. Network audio bridge can combine both systems and provide improved network functionality even for analog systems, while protecting the possibly very large investment of the existing analog system while you gradually migrate towards network audio.

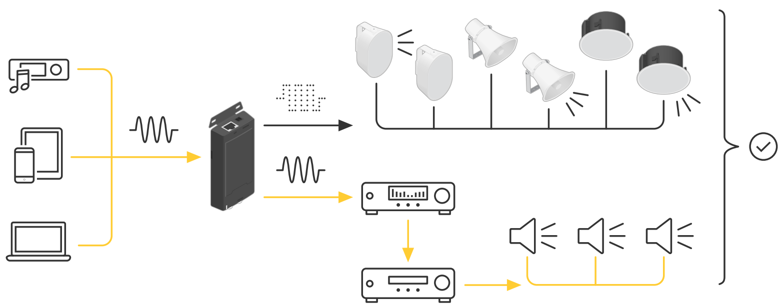

Figure 2.2b *A network audio bridge combines an analog system with a network audio system.*

Analog output. A network audio bridge can be used to connect any digital audio source to an analog speaker system. You can use the bridge to, for example, trigger announcements or make live callouts via your IP telephone system. A single audio bridge can enable IP functionality in an analog system with up to hundreds of speakers.

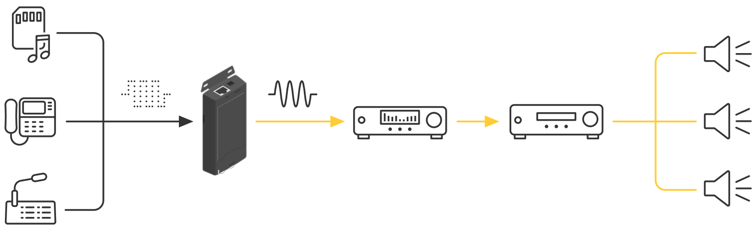

Figure 2.2c *A network audio bridge connects digital audio sources to an analog audio system.*

2.2.2 Network audio amplifier

A network audio amplifier from Axis is a small device for connecting one or multiple analog speakers. It has a built-in amplifier and digital signal processor (DSP) that delivers a total power output of 15 W. The amplifier is powered by PoE.

With the audio amplifier, passive speakers and network audio speakers can be mixed and matched in one installation and work together seamlessly. The amplifier and a passive speaker will, in all relevant aspects, act together as a network speaker. The passive speaker can then be managed through an audio management system — both network speakers and passive speakers can be controlled and managed from one location.

Ordinary speaker cables are used to connect the speaker to the amplifier, and the amplifier is then connected to the IP network in the building. If connected to only one speaker, the amplifier can be mounted right at the back of that speaker. If connected to speakers in a drop ceiling, the amplifier can also be discreetly mounted above that ceiling, away from sight. Depending on the use case and the surroundings, the amplifier is recommended to be used with up to eight speakers.

2.2.3 Volume controller

A volume controller from Axis lets you change the volume and select the audio source locally in an audio zone, by simply pushing its buttons. This device is ideal for environments with changeable occupancy or noise levels such as retail stores, warehouses, and cafeterias. It is equally well-suited for spaces that host a variety of activities and events, like community centers or nursing home common rooms. The control panel clearly indicates volume up and down, mute, and up to three preconfigured audio sources.

Figure 2.2d *A volume controller enables easy local volume adjustment for preconfigured network audio sources.*

Changes made with the volume controller only affect the selected audio source in the associated zone, and never live or scheduled announcements, paging, or public address. Voice messages always go through. Both the possible audio sources and the maximum volume are configured in the Axis audio management software, so either the headquarters or the local audio manager sets the degree of flexibility. The volume controller is easily installed by connecting it to the I/O port of a compatible Axis audio device in the associated zone. It will draw power and connectivity from the device. In large audio zones, you can place as many controllers as you need to ensure convenient access for staff.

2.3 Network microphone consoles

Axis offers a microphone console that includes a microphone, several configurable buttons, and a built-in audio management server. It can be connected to a standard network using PoE and integrated with network speakers to create a complete public address system for both live and prerecorded announcements. Using the web user interface of the microphone's built-in software, you can configure the buttons to the actions of your choice. Each button can be associated with a single zone or a combination of zones, and changed as often as you like.

2.4 Other audio related network products

Network strobe siren. This is a PoE-powered device for numerous applications. It connects to an Axis device, Axis VMS, or third-party VMS to signal and alarm using strobe lighting and siren alarms. Including various sound clips and white/RGBA light patterns, it can be set up with profiles to trigger different responses. It is ideal for deterring intruders or improving operational efficiency.

Figure 2.4a *Network strobe sirens can display various light colors and patterns, apart from siren alarms.*

Microphone kit. This is an enclosed microphone with connectivity options. It lets you hear and record audio in locations where audio quality is important and cameras might not be allowed. The kit is robust and designed to withstand outdoor challenges. It is ideal for housing the network audio bridge to get the benefits of network audio. This combination lets you easily add intelligent analytics to detect, for example, aggressive behavior, breaking glass, and gunshots.

Figure 2.4b *The microphone kit consists of an enclosed microphone with connectivity options.*

3. Network audio management software

An important aspect of an audio system is managing the audio content, as well as managing the devices. With the right audio management software it is easy to update scheduling, zoning, and audio content, and a network audio system can be efficiently managed and controlled regardless of its size and complexity, on a single site or across multiple sites.

3.1 Management features

Sophisticated network audio management software is central to the flexibility of network audio. With features such as central control, easy zone and content management from remote, and centralized health monitoring, a network audio system can be fully controlled through the software. Changes can be applied without any new cabling or anyone needing to physically go on site. The audio management software also enables integration with other systems.

> **Central control.** You control all your devices and their profiles from one location. If you have a multisite system you can manage all the sites and all their devices from one location. Guidelines, rules, and protocols can be implemented at once throughout the system without anyone having to physically visit the sites. You can update the system on the fly, in the software.

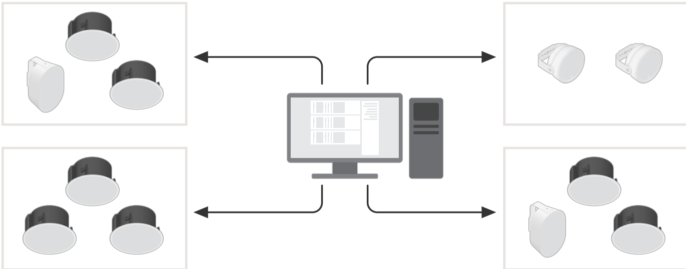

Figure 3.1a *Central control of multiple audio devices on a single site.*

> **User management.** Users can be assigned access with varying levels of privileges that fit their role in the organization.

> **Zone management.** Large spaces can be subdivided into zones, with multiple speakers in each zone. This way you can direct announcements (prerecorded or live) and music to specific areas. Zoning the system also allows volume to be tuned for the different content types separately for each zone.

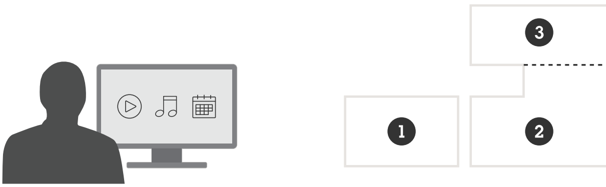

Figure 3.1b *Zones can be used to control and adapt the audio content for the right audience. It is easy to change the zones in the software when needed. In this installation, the second zone has been further divided to create a zone 3.*

> **Content management.** Local files for announcements, advertisements, music, and configuration for streaming content can be managed for all sites and zones through one single user interface. It is easy to set up and combine audio sources, configurations, and destinations.

> **Scheduling.** Recurring announcements can be played by use of scheduling, which makes sure that the announcements are played on specific times and days, and in specific parts of the system. An example could be that a specific voice announcement is played at noon on all weekdays in zones 1 and 2. Scheduling enables long-term planning for announcements and music, but it also provides flexibility and a possibility to tailor audio well in advance. In addition to calendar-based scheduling, content can be set up to play based on external events, such as when motion is detected by a video camera or by PIR sensors.

> **Prioritization.** Announcements or audio clips can be set to automatically overrule any background music played so that no announcements are missed. When the site receives a scheduled announcement, this will overrule the background music, which is muted during the announcement. If then a security announcement is initiated in the audio management system or VMS, that announcement will overrule both the scheduled announcement and the background music. More advanced management systems offer intelligent queuing in case of conflict, meaning that the overruled content will be played after the interruption.

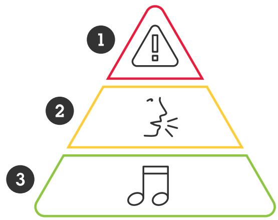

Figure 3.1c *Priority can be set to make sure that more important audio content, such as security announcements (1), always overrules less important content, such as scheduled announcements (2) and music (3).*

> **Health monitoring.** A network audio management system allows proactive troubleshooting from remote. It can detect malfunctioning speakers and send feedback about it in real time. Through the software interface you can see connection status of all speakers at once.

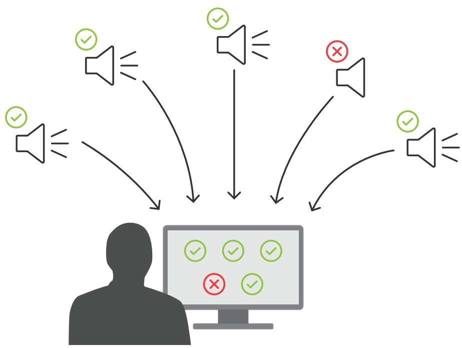

Figure 3.1d *Central system health monitoring.*

> **Secure remote access.** Centrally located IT staff can use secure remote access to rapidly diagnose an issue anywhere in the system and help get devices back online quickly.

> **System integration.** External systems can be integrated with, and provide input to, the audio system. By use of protocols, such as SIP and VAPIX, external systems can play live announcements or trigger audio based on events. See chapter 4 for more details.

3.2 Axis audio management solutions

Each network audio device from Axis comes with a built-in management software, **AXIS Audio Manager Edge.** It makes every device a complete, all-in-one sound system with no need for a separate software management server. This software is intended for low-complexity use cases on small to medium-sized sites, where it can be used to manage up to 200 devices in up to 20 zones.

For larger and more advanced use cases, there is **AXIS Audio Manager Pro** which can handle a large number of zones and thousands of devices in a single interface. It facilitates long-term scheduling and advanced priority settings with queuing in case of conflict.

AXIS Audio Manager Center is a service for remote management and monitoring of multisite systems, scaling from a few sites to several thousands. It is used together with AXIS Audio Manager Edge at each local site. Employing both cloud-based and on-premises components, this is a convenient and stable hybrid cloud solution. User workload is significantly reduced with AXIS Audio Manager Center, with a single sign-on to schedule announcements, background music, ads, and more for selected sites or zones.

4. Integration with other systems

When audio is integrated with other systems, such as video surveillance, VoIP, and access control, information from those systems can be used to trigger functions in the network audio system and vice versa. Users also benefit from having one common interface for managing the different systems.

4.1 Open standards enable integration

Axis network audio is a truly open system. It uses open standards which enable easy integration with other non-proprietary systems. For Axis, open standards mean that we provide full access to the application programming interfaces (APIs, such as VAPIX, SIP, RTP) and we work with commonly used protocols. Integration possibilities are virtually endless, and users can develop their own software for use in Axis products. Open standards make the system future-proof and allow users to make the most of their network audio installation by enabling extended use cases. Integration, both on a device level and a system level, provides flexibility with how you can use the products.

The following subsections show some examples of how audio devices and systems can be integrated with other devices and systems.

4.1.1 Video surveillance integration

Axis network audio can be easily integrated with video management systems such as AXIS Camera Station and AXIS Companion, but also with Milestone XProtect® (through the plugin AXIS Optimizer for Milestone XProtect), Genetec, and with a range of customized software solutions developed by Axis technology integration partners.

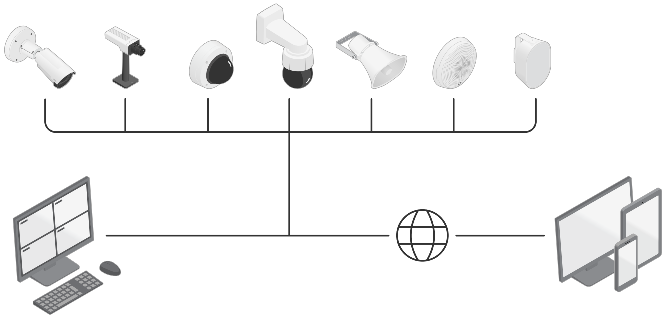

Figure 4.1a *Seamless integration of audio and video products in an IP-based security system.*

There are numerous examples of situations where a security system benefits from having both video surveillance and audio capabilities. Both video analytics and audio analytics can enhance this system integration. When trespassers are caught on video, the audio system can be triggered to send out a deterring voice message, live or prerecorded, to let the intruders know that they are being watched. It could also be the other way around, so that audio analytics applications such as glass break detection or aggression detection trigger event-based recording in the network video system.

4.1.2 VoIP integration

Voice over IP (VoIP) is a way to communicate over IP networks. VoIP applications can include voice and video elements and are used for video conferencing, call control, and instant messaging. It is enabled through the standard SIP protocol, which constitutes a way to connect, integrate, and control audio products over an IP network. You can connect as a SIP client or via SIP trunk.

Using VoIP you can integrate your network audio system with a company PBX phone system to make callouts possible from a traditional desk phone. You can also use VoIP to connect to your cloud-based collaboration solutions (such as Zoom, Cisco Webex®, RingCentral) to make callouts from your collaboration platform using your PC or mobile phone without needing to install additional applications.

4.1.3 Mass notification system integration

Emergency mass notification systems are used to manage incidents using multiple communication methods, where audio is often the most critical part. An integration of mass notification and the public address system enables the system to reach people on-site quickly and efficiently.

Mass notification systems can integrate with speakers either directly, or on system level so you can leverage the functionality of each system together or independently. Mass notification systems can easily communicate directly with Axis audio devices. This is enabled through adaptation of partner APIs, and applies to, for example, Syn-Apps Revolution and Singlewire InformaCast. If no API integration is available, basic SIP capabilities in the mass notification systems can often be used instead.

4.1.4 Fire alarm and emergency communication systems

Installation of fire alarm systems and emergency communication systems is covered by the NFPA 72 standard in the US and the EN 54 standard in the EU. These standards do not cover IP-based speakers. This means that Axis network audio devices cannot comply with the standards.

There is no direct integration between EN 54- and NFPA 72-certified systems and Axis network audio devices.

However, an Axis network audio system can co-exist with a certified system. You can even set up the systems so that the certified fire alarm triggers the I/Os of your Axis speakers to perform various actions. For example, the fire alarm could automatically mute the network audio system. The audio system could also complement the certified system by playing clear, informative messages in case of an alarm.

4.1.5 Edge-to-edge integration

One type of integration is enabled by so called edge-to-edge technology. This is a way to make IP devices communicate directly with each other. Using edge-to-edge, you can extend the functionality of an Axis network camera with audio, even if it has no built-in speaker or line-out capability.

5. Designing a network audio system

One of the main benefits of a network audio system is flexibility and scalability: the freedom to mix and match the most appropriate components and the power to optimize or expand the system to any size. Still, there are many considerations to make when you plan a network audio system.

5.1 Know the purpose of your system

Before you start designing your system, you need to be clear about its purpose. Only then can you dimension the network properly and enable the right integration possibilities. Do you need audio for security, integrated with a surveillance system? Do you need audio for safety, or audio for improving operational efficiency? Perhaps you need all at the same time? Axis network audio systems are designed to handle multipurpose use cases.

When evaluating an audio system, also look at other systems, such as video, access control, intercom, and intrusion detection, to determine if there is a way to construct an integrated system that covers all physical security needs. Axis network speakers can integrate with several video management software applications.

5.2 Plan the network

For a solid foundation, a cabled network infrastructure is recommended in network audio. If you already have an audio system installed, evaluate whether it meets your expectations and needs for today and the future. Investigate whether you will need to integrate your system with a VMS or a mass notification system, or connect it with a PBX. You may need to update your existing system or construct a new system.

Make an inventory of your IT infrastructure to see if it meets the needs or should be updated. Most networks today carry a range of network traffic types, and it is common that more and more IP-based devices are added to the network over time. If the existing network is heavily used, it can impact the quality of the audio if not configured correctly. Without appropriate action, there are risks of delay, jitter, and packet loss, which can result in audible artifacts.

The bandwidth requirements for audio applications depend on the transmitted audio quality — a higher quality generally requires more bandwidth. For example, an uncompressed stereo audio signal of 48 kHz sample rate and 16 bits per sample needs approximately 1.5 Mbit/s in bandwidth.

By default, an Axis audio system is configured for high quality audio with efficient use of bandwidth through combining efficient codecs for audio compression with multicast transmission over the network. A network for audio should use multicast transmission, but sometimes it is not possible due to equipment constraints. If you use unicast instead, each speaker will receive a separate audio stream, so all traffic is multiplied by the number of speakers.

If you have a segmented network, or you want to span your audio network across a WAN, it is recommended to use a separate VLAN for the audio traffic. This makes it easier to set up reliable communication between the audio devices and also prioritize the traffic for the audio system versus other traffic.

To ensure the right traffic prioritization, Quality of Service (QoS) should be implemented, both for resource reservation (integrated services) in terms of network resources allotted according to an application QoS request, and for prioritization (differentiated services) in terms of classified and allotted network resources according to bandwidth management policy criteria.

5.3 Plan and design the audio site

Use AXIS Site Designer to help you plan and design your audio site. On axis.com you can also find a quickguide for speaker coverage calculations, which can assist you in determining how many speakers you need. Furthermore, Axis provides GLL files for EASE Evac design tool and AXIS Plugin for Autodesk® Revit®, which can help you plan your audio solution.

5.3.1 Make a site survey

To successfully plan an audio installation, you must first be clear about the purpose of the system. If you need to make announcements in a classroom, one speaker may be enough. If instead you want to combine making announcements with playing background music in a retail store, you need several speakers for a good listening experience, even if both rooms have the same size.

When you have defined your purpose, you need to make a walkthrough of the site to make some measurements. If available, you should also use CAD drawings of the premises in a system design software.

Measure the room dimensions. Use a laser distance meter. The ceiling height and the possible mounting heights are the most important measures. Higher mounting means better coverage.

Note the reflectance. Rooms with reflective surfaces may have reverb, while larger spaces may encounter delay. This may have impact on how many speakers you need.

Measure the background noise. Use a professional dB meter. Important voice messages that everyone need to hear should be up to 12 dB above the ambient noise. If the background noise level is high, you must make sure to choose a speaker with a high enough SPL (sound pressure level).

Check the mounting possibilities. What mounting spots are available, regarding physical room characteristics as well as connectivity? Are there any limitations to how and where the speakers can be installed? Remember that it is the mounting height of the speaker, not the ceiling height, that will determine the spread of sound.

5.3.2 Find and compare speakers

The following subsections provide some guidelines to help you determine what type of speaker you need and which functionalities it should have. You can then use Product selector on axis.com to find and compare Axis products that are suitable for your installation. AXIS Site Designer is also a helpful online tool in this stage. It lets you plan and design an audio installation, by providing guidance on which speakers to use, how many speakers are needed, their optimal placement, and so on, with regard to the conditions at the site.

5.3.2.1 Speaker types

To determine which types of speakers are suitable and how many speakers are needed, the site environment and the purpose of the speaker installation must first be considered. Considerations include the following:

> **Indoor or outdoor.** If you are going to use speakers outdoors, it is important that you choose an outdoor-ready speaker. The speaker must also be approved for use in the temperature range that the particular site exhibits.

> **Area of coverage.** How far and wide the sound spreads depends on the type and model of the speaker. Horn speakers typically have a higher sound pressure and may have a long but rather narrow area of coverage, while ceiling and cabinet speakers may be designed to have lower sound pressure with wider coverage. Check the speaker specification for its SPL value and coverage information.

- **Overt or highly discreet installation.** Speakers come with very different form factors, and the shape is mainly related to the functionality. But more often than not, it is possible to find a speaker that matches your operational needs while also complying with your requirement of either a non-discreet or a discreet installation.

5.3.2.2 Speaker functionalities

Network speakers from Axis have many additional functionalities apart from providing an audio stream. Some examples are:

- **Built-in audio management software.** This provides support for live or prerecorded announcements, background music, audio content scheduling, zoning, and priority of audio sources.
- **Integrated microphone.** Makes it possible to run self-tests of the speaker functionality and use audio analytics that listen in on events.
- **Two-way audio.** Thanks to the integrated microphone and echo cancellation you can speak with visitors or intruders in real-time.
- **Mic-in connector.** Possibility to connect an external microphone.
- **Audio analytics for detection.** When integrated microphones detect sounds above a certain level, they can trigger the playing of a voice message or trigger a camera to start recording.
- **Memory card slot.** You can use SD cards for storing audio clips.
- **Input/output (I/O) connectors.** Connecting external input devices to a speaker (such as a door contact, infrared motion detector, radar device, glass-break sensor, or shock sensor) enables the speaker to react to an external event by, for example, sounding an audio message. Outputs enable the speaker or a remote operator to control external devices, for example, alarm devices, door locks, or lights.
- **Integrated LED.** Some speaker types have an integrated LED light for visual status confirmation.

5.3.3 Decide how even sound you need

Axis recommends different solutions depending on the customer requirements. Depending on how even the audio level needs to be you can choose from a basic or a premium setup, or anything in between.

The basic setup will give you the right number of speakers needed to cover a certain area. Going below this number of speakers is never recommended, since the evenness of audio levels would vary too much within the area.

The premium setup will give you twice the number of speakers compared to a basic solution. A more even audio level will be maintained throughout the area.

Figure 5.3a *Left: in a premium solution, sound is even throughout the area. Right: in a basic solution, sound is less even, and spots with lower sound volume are allowed.*

For announcements, a basic solution will be enough in most situations. If the ambient audio level is very high (like a noisy manufacturing site), a premium solution is recommended.

For background music, a basic solution will be enough in locations such as typical retail stores. If the customer experience is critical, you can choose a premium solution. Some installations might also require a combination of basic and premium. This could be a larger project with many different types of areas, such as a school campus, shopping mall, or manufacturing site.

5.3.4 Determine the speaker placement pattern

In an outdoor solution the spread of audio is better than in an indoor solution, due to less reflections, and you can often use a lower number of speakers than in an indoor solution.

When placing speakers indoors, the general rule is to, if possible, point the sound along the room. That is, if you have a rectangular room, try to place the speakers on the short walls pointing out along the longer walls. This will let the sound spread as far as possible before being reflected on the walls. However, it is not recommended to place a speaker in a corner, since that would unevenly amplify the bass sound.

Cluster placement. If you prioritize simple and low-cost installation, you can install the speakers in clusters. This will minimize cabling but might not be the best way to get a good spread of the sound.

Wall placement. If the room dimensions allow, and you do not mind the extra cabling, a wall placement solution will probably spread the sound better. With the same number of speakers as in the cluster placement example above, the installation might look like the below figure. If the room is large, however, the reach of the speakers might be too short.

Ceiling placement. If the room has a drop ceiling, or if it is possible to install built-in ceiling speakers, a ceiling placement can be a discreet solution. However, this placement is very sensitive to the ceiling height. The lower the ceiling, the more speakers you need in order to cover a certain area. In a high ceiling, on the other hand, pendant speakers with adjustable cable length can be installed at just the height you need.

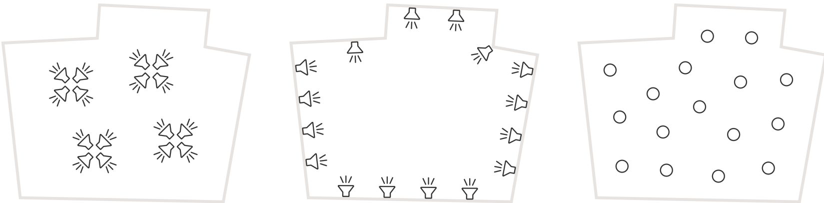

Figure 5.3b *Cluster placement, wall placement, and ceiling placement of speakers in an indoor location, seen from above.*

5.3.5 Choose speaker mounts

Many network speakers come with integrated mounts that facilitate surface installation on a wall or in a ceiling. Other types of mounts may be available as separate accessories.

Ceiling mounts. All Axis network speakers can be mounted in the ceiling, which means that the sound will go from ceiling to floor. The preferred way to get the most even coverage of audio levels is a ceiling-mounted solution. This means that if you play a message it can be clearly heard at any position. However, sometimes a ceiling mounted solution is not possible due to a very solid material of the ceiling, the height of the ceiling, or obstacles between the ceiling and the floor.

Speakers can be mounted on ceilings by use of:

> Surface mount: mounted directly on the surface of a ceiling and therefore completely visible.

> Recessed mount: mounted inside the ceiling with only parts of the speaker visible. This mount is also known as a flush mount or drop-ceiling mount.

> Pendant mount: enabling the speaker to be hung from a ceiling.

Wall mounts. Wall mounts are a common and straightforward choice for many speaker installations, both indoor and outdoor. With wall-mounted speakers the audio is directed from the speaker straight out into the room. A wall-mounted solution may require fewer speakers to spread the sound than a ceiling-mounted solution, but it may require more cabling. When considering a wall-mounted indoor solution we need to account for the distance the speaker can reach into the room. For a large room, the audio levels in the center of the room might be too low if messages are to be clearly heard at any position.

Pole mounts. Some types of speakers can be mounted on poles by use of a steel strap mount. This is often used outdoors.

5.3.6 Determine mounting height and number of speakers

The coverage of an area depends on the number of speakers and their mounting height. A quick guide for speaker coverage calculations is available on axis.com. To calculate coverage of an entire room or building, use AXIS Site Designer or other available tools.

5.3.7 Choose audio management software

Audio management software from Axis is used to manage the audio content as well as the audio devices. It features functionalities such as zone and content management, scheduling, prioritization of audio sources, and system health control. AXIS Audio Manager Edge comes built-in with all Axis audio devices.

For small and midsized systems with basic use cases, AXIS Audio Manager Edge provides all the features and functionality needed for efficient and intuitive management. It allows, for example, zoning and scheduling of audio content to different areas from a single user interface.

Larger and more complex systems are better managed through AXIS Audio Manager Pro, which can handle long-term scheduling and advanced priority settings.

Multisite systems can be managed and monitored remotely through the service AXIS Audio Manager Center. This is a hybrid cloud solution, used together with AXIS Audio Manager Edge at each local site. With a single sign-on, you can schedule announcements, background music, ads, and more for thousands of sites.

Device management is an important part of an audio management system. It helps you manage installation, security, and operational tasks of your devices, for example, device configuration and device software upgrade.

6. Network requirements

This chapter helps you specify what you need from your network to be able to connect an Axis audio system.

6.1 Glossary of network terms

Broadcast	A data transmitting method where one sender sends the same information to all receivers in a network at once.
DHCP	Dynamic Host Configuration Protocol, a protocol for automatic assignment and management of IP addresses.
IGMP	Internet Group Management Protocol, a communications protocol used on IPv4 networks to establish multicast group memberships. IGMP is an integral part of IP multicast.
mDNS	Multicast DNS, a zero-configuration protocol to resolve host names and find devices on a network. Bonjour is an example.
MQTT	Message Queuing Telemetry Transport, a standard messaging protocol for the internet of things (IoT). It is used in a wide variety of industries to connect remote devices while leaving a small code footprint and requiring minimal network bandwidth.
Multicast	A data transmitting method that allows communication with multiple receivers in a network. Multicasting reduces network traffic by sending a data stream once to many recipients.
Network port	A number assigned to uniquely identify a virtual connection endpoint and direct data to a specific service, for example HTTP that normally uses network port 80. Network ports are non-physical ports in software.

PoE	Power over Ethernet, a technology that allows switch ports to supply electric power to connected devices. This means that the devices do not need a separate power supply. Follows IEEE 802.3 af/at/bt standards that have different power limits.
RTCP	Real-Time Control Protocol, a protocol that provides out-of-band statistics and control information for an RTP session. It partners with RTP in the delivery and packaging of multimedia data but does not transport any media data itself.
RTP	Real-time Transport Protocol, a packet-based protocol that permits the transfer of real-time data, for example audio and video, between system endpoints.
SIP	Session Initiation Protocol, a signalling protocol used for initiating, maintaining, and terminating communication sessions that include voice, video, and messaging applications. SIP is one of the protocols used in Voice over IP (VoIP).
Switch port	Physical connector in a network switch where a device is connected.
Unicast	A data transmitting method for one-to-one communication in a network. With unicast you need to send multiple streams if you want to send the same information to multiple receivers.
VoIP	Voice over IP, a group of technologies that enables voice communication and multimedia sessions over IP networks.

Table 6.1a *Network terms used in this chapter.*

6.2 Network requirements are affected by the choice of audio management software

How you need to setup your local network, and which communication ports are used, depends partly on which audio management service (AXIS Audio Manager Edge or AXIS Audio Manager Pro) you choose.

If you also connect your AXIS Audio Manager Edge sites to the AXIS Audio Manager Center service, you additionally need to verify your firewall settings for internet access by the system.

6.3 General network requirements

IP address assignment. Axis audio systems support both static and DHCP-assigned IP addresses of the devices. If you use DHCP, we recommend that you always assign the same IP address, which you configure in your DHCP server.

Network switches. Axis devices use Power over Ethernet (PoE) to receive power. PoE can be supplied either by network switches or by midspans. If you use network switches, you might have to enable PoE on the switch ports where the Axis devices are installed. We recommend managed network switches for larger installations.

6.4 Network requirements for AXIS Audio Manager Edge

AXIS Audio Manager Edge is a serverless solution that you use to configure Axis audio devices into a system and set up sources and zones for music, paging, and other audio. You can also configure schedules and prioritize between sources. You select one device to be the configuration leader. From that device's web interface you launch AXIS Audio Manager Edge, with which you configure and control the system.

Protocols or services.

> The devices use Bonjour discovery (mDNS) to locate other devices on the network. You can also add devices manually.

> MQTT is used to exchange information between the devices. MQTT does not require any extra broker, but the traffic must be allowed on the network.

> For each source (paging source, music streaming source, content type) configured in the system, two network ports are used for audio streaming and control (RTP and RTCP).

> A common clock, on network port 5015, is used to make sure that audio is played synchronized between the devices.

> SIP (VoIP) can be used for paging and interfacing with other systems. If you use SIP, you must allow associated ports and traffic on the network.

Unicast/multicast.

> Axis audio systems use multicast, meaning that an audio stream can be sent from one source to many devices while keeping network traffic down. The network switch uses IGMP snooping to determine which devices should receive the stream. To use multicast, you must enable multicast and IGMP snooping in the network switches.

- AXIS Audio Manager Edge uses a single multicast address and different network ports on that address to differentiate between the streams.
- You can configure the system to use unicast instead, but this limits the system size to maximum 20 devices.

6.5 Network requirements for AXIS Audio Manager Pro

AXIS Audio Manager Pro is intended for larger systems than AXIS Audio Manager Edge, or for more complex use cases. AXIS Audio Manager Pro is a software running on a Windows server or virtual machine.

Protocols or services.

- SIP (VoIP) can be used for paging and interfacing with other systems. If you use SIP, you must allow associated ports and traffic on the network.

Unicast/multicast.

- Axis audio systems use multicast, meaning that an audio stream can be sent from one source to many devices while keeping network traffic down. The network switch uses IGMP snooping to determine which devices should receive the stream. To use multicast, you must enable multicast and IGMP snooping in the network switches.
- All devices must be connected to the same multicast domain as the server that AXIS Audio Manager Pro is installed on. Audio streaming between the server and all the devices uses multicast. AXIS Audio Manager Pro uses one multicast address per stream. The default address range is 239.0.0.0 - 239.0.0.254.

6.6 Network requirements for AXIS Audio Manager Center

AXIS Audio Manager Center is a subscription-based service for remote management and monitoring of multisite systems using AXIS Audio Manager Edge locally. The hybrid cloud solution uses both cloud-based and on-premises components for a convenient and stable solution. It significantly reduces user workloads, with a single sign-on to schedule announcements, background music, ads, and more for selected audio sites or zones.

In addition to the network requirements for each local audio site (see AXIS Audio Manager Edge), there is a network connection to the cloud service. The communication is initiated by each local audio site and normally this does not require any addition of firewall rules in the network. The communication uses encryption for security and privacy.

7. Audio capture and audio analytics

While this technical guide is focused on systems for audio broadcast, this chapter briefly presents audio capture and some of its benefits in safety and security use cases. Audio capturing provides invaluable possibilities for detecting and interpreting events and emergencies. Audio analytics can monitor the incoming audio and react when something stands out, even while maintaining privacy by neither streaming nor recording the original audio.

7.1 Audio capture

Audio capture can be deployed as a standalone technology enabling several use cases in crime prevention, protection, and forensics. One example is audio surveillance with direct operator interaction to increase scene awareness. In a hospital or care facility, audio capture can help you perceive if a patient is in distress and needs a nurse.

If used together with video surveillance, audio capture adds another dimension of information for decision making and has the potential to reinforce existing video surveillance use cases. For example, security operators can get a significantly better overview of scene events if their video stream is complemented with an audio stream.

There are several ways to implement audio capturing with Axis products. All Axis network speakers and many Axis cameras have built-in microphones. Speakers or intercoms are often placed closer to where people are, compared with a camera that might be placed in a corner of the room. Therefore, speakers or intercoms with integrated microphone can usually pick up voice better than a camera with integrated microphone. You can also connect separate, standalone microphones and place them in the strategically best locations for capturing clear and relevant audio. Microphones can be connected through a physical mic-in/line-in connector on the camera, a connectivity hub, or an audio and I/O interface with portcast technology. Pairing through edge-to-edge technology is also possible between microphones and cameras on the same network.

Figure 7.1a *A camera with built-in microphones and speaker for two-way audio, a standalone digital microphone with 3.5 mm (1/8 inch) connector, and an audio and I/O interface that uses portcast technology to seamlessly add audio capturing capability to a camera.*

7.2 Audio analytics

Audio analytics can be set to trigger automatic alarms and other actions when a microphone picks up sounds associated with people shouting or glass breaking, or other unexpected sounds. This provides early warning that enables quick responses and intervention. Audio analytics can also be used to automatically redirect a pan-tilt-zoom (PTZ) camera towards an audible incident.

Audio analytics in general do not record sound continuously. They typically just buffer it temporarily and process the audio to search for specific patterns, levels, or frequencies. But systems can be set up to record what was buffered just before and after a detection to allow security to verify the detection and, possibly, preserve the audio for forensic evidence.

7.3 Privacy and legal restrictions

Many have concerns regarding the use of microphones in surveillance. These concerns are typically linked to the recording of plain speech. For privacy reasons, audio streaming is deactivated by default in Axis products.

This means that audio that is picked up by the microphone is neither recorded nor streamed, even when it is used by integrated audio analytics. You can turn on audio streaming if you need it, but it is possible to use audio analytics to, for example, detect screams from a hospital waiting room even if no one can 'listen in' to conversations through the microphone.

Laws that regulate surveillance vary by region and by country, so make sure to know what is permitted before using audio capture or audio analytics in your surveillance system.

8. A background on audio

You don't need to be an audio expert to deploy or use Axis network audio — our speakers are preconfigured to sound good. For the interested reader, however, or those dealing with especially complex installations and solutions, this chapter provides some background information about audio physics, acoustics, and digital audio technology.

8.1 Audio fundamentals

As digital as an audio system may be, the audible sound itself consists of physical vibrations. The physics of audio includes concepts, such as frequencies, sound pressure, human sound perception, and sound measurement units.

8.1.1 What is sound?

Sound is an audible pressure wave. Vibrations from your vocal chords or the diaphragm of a speaker disturb parts of the air. These pressure differences are what we hear. The qualities of a sound — its loudness and pitch — are determined by the properties of the sound wave.

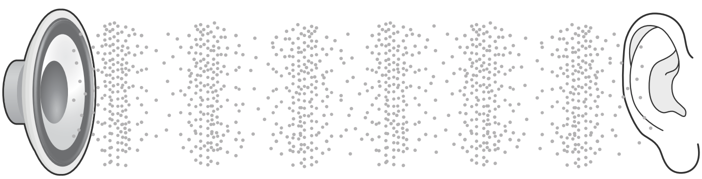

Figure 8.1a *Pressure differences in the air are interpreted as sound.*

8.1.2 Longitudinal waves

A sound wave is a longitudinal wave, which means that the air vibrates parallel to the direction of propagation of the wave. Air molecules vibrate back and forth in the same direction as the sound travels. You can visualize it as the movement that occurs when you stretch and compress a coil (typically a Slinky toy), where the distance between coils increases and decreases.

However, longitudinal waves are not as easily visualized as the other type of wave, a transverse wave. In a transverse wave, the vibration back and forth takes place in directions perpendicularly to the direction in which the wave travels. You can visualize this wave, too, with a coil, by moving the ends of the coil in a direction perpendicularly to the length of the coil. One example of a transverse wave is an electromagnetic wave (such as light), where the electric and magnetic fields vary perpendicularly to the direction of propagation.

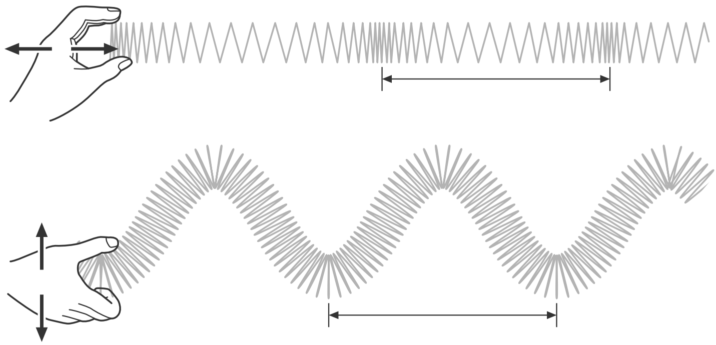

Figure 8.1b *Using a coil to visualize a longitudinal wave (top), such as a sound wave, and a transverse wave (bottom). The wavelength is marked as the distance between two minima.*

Since it is easier to visualize a transverse wave than a longitudinal wave, a sound wave is usually graphically represented by a transverse wave. Despite not being completely true to the nature of the wave, this graphical waveform provides a good understanding of a sound wave as it moves through the air over time, and it makes it easier to see the wave's fundamental characteristics: wavelength, amplitude, and frequency. Both longitudinal and transverse waves have these characteristics.

Do not be confused by a graphical representation of a sound wave, such as that on an oscilloscope which displays a transverse wave when displaying sounds. The real sound is always longitudinal.

8.1.3 Pitch

When we talk about pitch — how high or low a sound is — we are talking about the frequency of the pressure wave, that is, how many times per second the sound wave vibrates. Frequency is measured in cycles per second, unit Hz (Hertz).

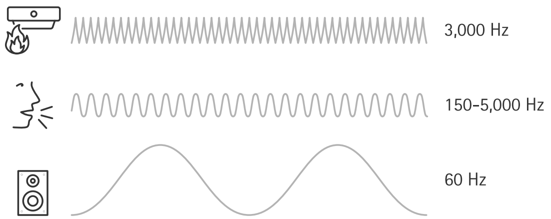

Figure 8.1c *Sounds, including human voices, can vary in pitch – vibrations per second (also known as frequency).*

8.1.4 Sound pressure

A noise can be barely audible, or extremely loud – but with the same pitch. These noises have the same frequency, but their waveforms differ instead in amplitude. The sound pressure is higher for a sound with a higher amplitude, and the changes in pressure is measured in Pascal, Pa. We can perceive sounds with as low sound pressure as 20 µPa (could be a mosquito ten feet away), and as high as 20 billion µPa (typically the launch of a space shuttle). When dealing with such a large range of numbers, it is easier to use a logarithmic scale. This is one of the reasons why sound pressure levels are more often measured in dB SPL.

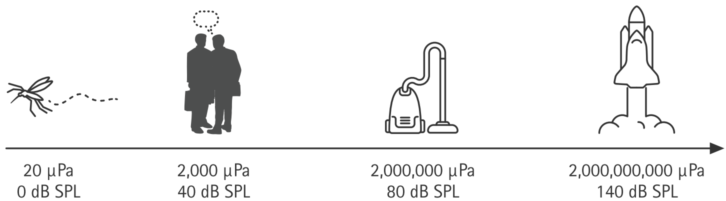

Figure 8.1d *Typical sound pressures from familiar sources, measured in both Pascal and decibel.*

8.1.5 SPL values

Sound pressure level (SPL) is the RMS (root mean square) value of the instantaneous sound pressures measured, in dB, over a specified period of time. SPL is not a constant average value of loudness but rather an average of the short peak values. An SPL value given for a speaker is assumed to be measured for a 1 kHz tone at a distance of 1 m, if nothing else is stated.

The sound pressure level of an audio source decreases with the distance from the source. Defined to start at 0 dB at 1 m from the source, the SPL decreases by 6 dB with each doubling of the distance from the source.

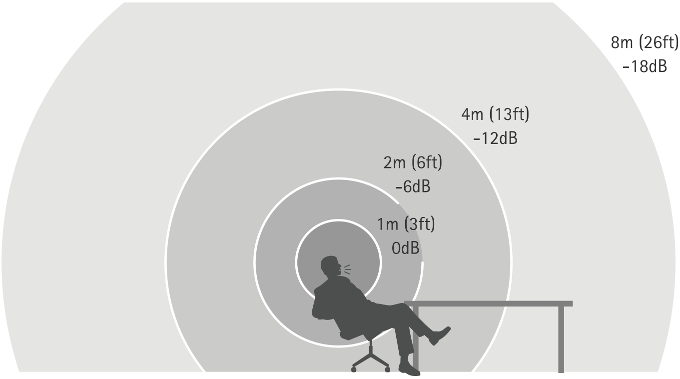

Figure 8.1e *The sound pressure level from an audio source decreases by 6 dB with each doubling of the distance from the source.*

8.1.6 Sound power

The unit of power, watt (W), is familiar from various electrical components, such as light bulbs, laptop chargers, and speakers. The unit can, however, be used in different ways, and in audio terminology we come across varieties like instantaneous power, average power, RMS (root mean square) power, and peak power.

An amplifier might be constructed to be able to deliver 300 W over a very short period of time, such as when a drum, explosion, or any other audio with a short and loud transient, will be heard. This means that the instantaneous power will increase really fast from very low to very high. The same amplifier might, however, only be rated for 50 W continuous use, since continuous use will produce a lot more heat, which impacts both the electrical components and the amplifier's performance.

8.1.7 Decibels

Because sound is perceived non-linearly, it is best measured and described using the non-linear unit decibel (dB). A doubling (measured in W) of the sound power equals to a 3 dB increase, and a doubling of the loudness equals a 10 dB increase.

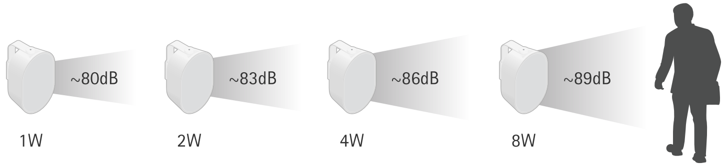

Figure 8.1f *A doubling of sound power, as measured in Watts, corresponds to a 3 dB increase.*

A sound pressure level given in the weighted dBA scale has been compensated for the human ear's frequency-dependent perception of sound. Using the unweighted dB scale, a 100 dB level at 100 Hz will, for example, be perceived to have a loudness equal to only 80 dB at 1 kHz, while 100 dBA will be perceived as equally loud at all frequencies.

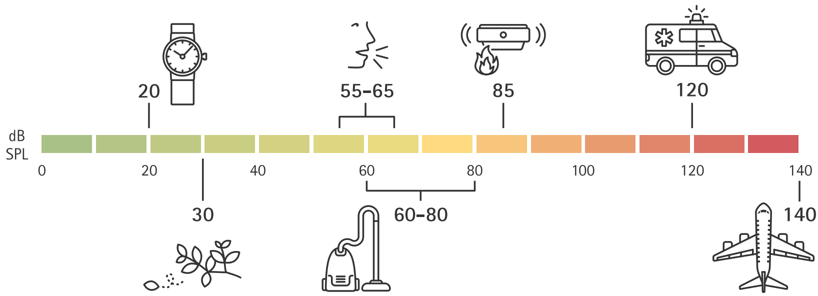

Figure 8.1g *Approximate sound levels, in decibel, from familiar audio sources.*

The decibel unit is often referring to a relative change in loudness. For expressing an absolute value, dB SPL should be used. A value of 0 dB SPL is the softest sound that the human ear can perceive.

8.1.8 Perceived loudness

The human ear is, in theory, able to perceive frequencies from 20 Hz to 20 kHz. The upper limit of 20 kHz is lowered with age but the high frequencies can still add character through overtones to audio with lower frequencies. Human speech, being complex with lots of harmonies, is scattered over frequencies from around 85 Hz (lowest for human male) to around 8 kHz (overtones for human female). In telephony, only the range of 300 Hz to 3.4 kHz is commonly used, and while it makes the voice audible, the audio will not be as clear as a full frequency range recorded voice.

Even though the ear is sensitive to all frequencies between 20 Hz and 20 kHz, the sensitivity varies with the frequency. Sounds of a specific power will thus be perceived as having different loudness at different frequencies. The loudness unit *phon* takes this sensitivity into account and, for example, a sinusoidal tone of 50 phons is perceived as equally loud at all frequencies.

The difference in sensitivity can be visualized in equal-loudness curves. One line represents the sound level that must be used, in order for the sound to be perceived at the same volume for all frequencies. The different lines represent different phon values. It is evident from the curves that the sound level must be substantially higher at the lower frequencies in order to be perceived as equally loud as higher frequencies. This is because the human ear is less sensitive to lower frequencies. The minimum of the curves is placed around 2 kHz – 5 kHz, meaning that this is the frequency range to which a human ear is most sensitive, and in which the ear can best decipher a conversation. It is also the frequency range of human speech.

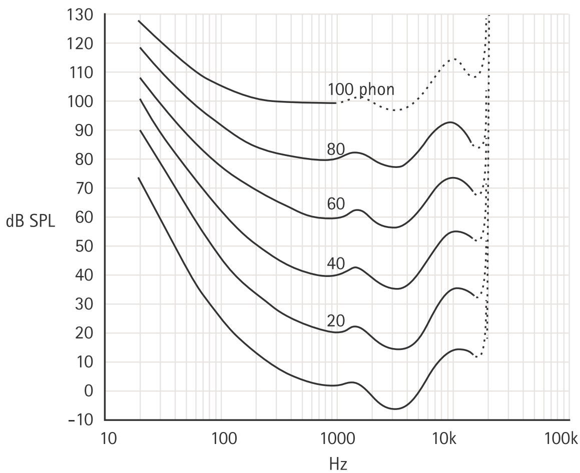

Figure 8.1h *Equal-loudness curves showing the sound pressure levels needed at different frequencies in order to make a sound perceived as equally loud over all frequencies. The curves are originally from the ISO standard ISO 226:2003.*

8.1.9 Sampling frequency

The sampling frequency is the number of audio "snapshots" taken per second of the analog input audio stream, in order to digitally reconstruct it. Lower sampling frequencies sample data less frequently. With a low sampling frequency, parts of the audio are not captured and the overall sound quality will be lower. With a higher sampling rate the audio stream can be more correctly recreated, delivering higher quality.

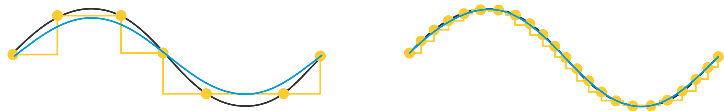

Figure 8.1i *Visualization of an analog sound wave and how it can be digitally represented. Left: with a too low sampling frequency (sampling at every dot), the original wave (black line) will be wrongly represented (as the blue line) which results in lower sound quality. Right: with a high enough sampling frequency the analog wave can be accurately reconstructed.*

The sampling frequency must be at least twice as high as the highest input audio frequency that should be reconstructed, but even higher for good quality. In audio files and CDs, 44.1 kHz is a commonly used sampling frequency, thus using 44,100 samples per second.

8.1.10 Dynamic range compression

Audio that has large differences between the quietest and the loudest parts is said to have a "large dynamic range".

Dynamic range compression is an audio signal processing operation that makes the quietest parts louder, while the loud parts either stay the same or become less loud. The operation decreases the dynamic range, and we perceive the recording as louder.

Compression of dynamic range is often applied in audio systems for restaurants, retail, and similar public environments that play background music at a relatively low volume. Apart from making the volume more constant, the compression also makes the quieter parts of the audio more audible over ambient noise.

You can also use dynamic range compression in order to increase the perceived volume of a voice announcement. This is useful for security announcements where the main focus is to make the message clearly heard over background noise, and a natural sound is less important.

8.1.11 Speaker sensitivity

A speaker's sensitivity is its ability to reproduce sound when fed a certain power. Determining the sensitivity is usually done by feeding an audio signal of 1 W (typically at 1 kHz) and then measuring the sound pressure level in dB at a 1 m distance. Common values for speakers are around 85–92 dB. The higher the sensitivity, the louder the sound will be from the speaker when fed a certain power.

The sensitivity of the speaker is sometimes an indicator of the quality of the speaker. Lower sensitivity indicates a less powerful magnet and/or a smaller and less expensive coil. Therefore, in regards to audio quality, a larger speaker is not necessarily better than a smaller speaker. The size of a speaker is, in a way, what megapixels are to a camera: unless we also have a good camera lens (or speaker sensitivity), a higher resolution (or increased speaker size) is not worth much.

You can look at the speaker sensitivity to compare analog speakers with other analog speakers, but not to compare analog speakers with active speakers such as IP speakers.

8.1.12 Polar response

Some speakers have a very narrow direction of sound in order to achieve a high sound pressure in one direction. Others are made to have as wide spread of the sound as possible. A speaker's ability to reconstruct audio is dependent on the audio frequency. Generally, lower frequencies have a wider spread while higher frequencies are more directional.

A polar diagram can visualize how frequencies spread out differently from a speaker.

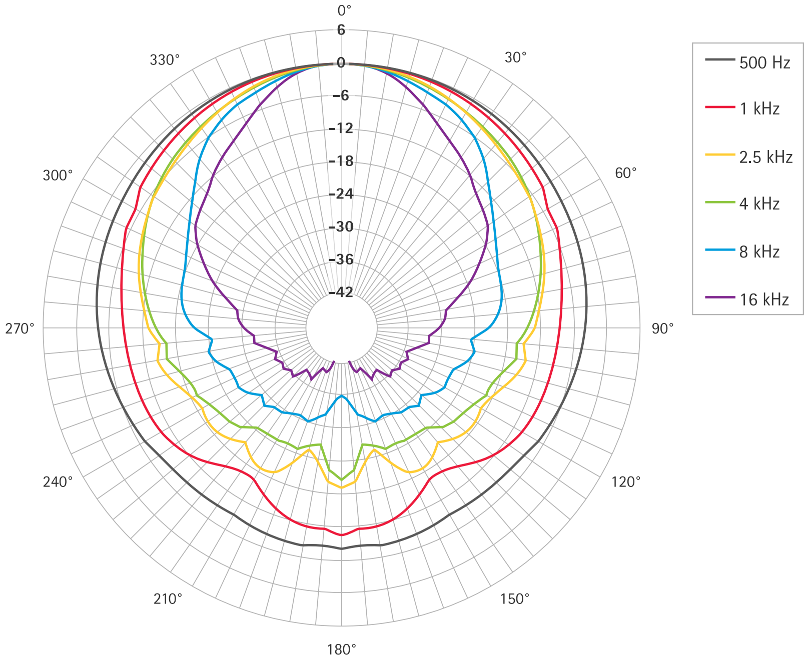

Figure 8.1j *Polar diagram showing the spread from a generic example speaker (which was located in the center of the diagram). The lower frequencies have a wider spread (even behind the speaker, at 180 degrees) while higher frequencies are more directional.*

8.2 Acoustics

The surroundings have a great impact on the sound from an audio system. While calculating the effects can be a complex task, some understanding of acoustics goes a long way.

8.2.1 Echoes

In a room that is completely empty, there will be reverb and/or delay in the sound. This is because all the flat surfaces are perfect for the audio waves to reflect against. If fabrics and uneven surfaces are added, such as sofas, curtains, and carpets, there will be less reverb, but the sound will also be perceived slightly less loud because of the absorption.

Sound waves are often reflected multiple times before reaching our ears. Knowing that the speed of sound in air is around 340 m/s (1020 feet/s), we can calculate the distance that an echo has travelled. If we hear the echo 0.25 s after the initial sound, for example, the sound has travelled around 85 m (0.25 s x 340 m/s), or 255 feet. For each reflection, the audio fades a little bit until we cannot hear it anymore.

8.2.2 The impact of room dimensions

The size of the room has a large effect on the audio experience. This is because the sound waves are reflected against walls, ceilings, and furniture. To better understand the impact of these reflections, it may be helpful to talk about the wavelength of an audio wave.

The wavelength, denoted by the Greek letter lambda (λ) is related to the speed of sound ($v=340$ m/s in air) and the frequency (or pitch, denoted by f, measured in Hz), according to $\lambda=v/f$.

A frequency of 20 kHz (20,000 vibrations per second) corresponds to a wavelength of about 1.7 cm (0.7 inches) while a lower frequency of 20 Hz (20 vibrations per second) corresponds to a longer wavelength of about 17 m (56 feet).

With wavelengths up to 17 m (56 feet) for the lowest bass, audible sound waves in a small room will be reflected against the walls before the waves have properly developed. This results in resonances and associated standing waves, causing some frequencies to be amplified (higher volume), and others to be attenuated (lower volume). We need a rather large room to hear the bass without distortion.

The impact of resonances on the experienced audio quality increases with the sound volume. With higher volume, the reflections will interfere more with the sound from the source.

8.2.3 Neutralizing room acoustics

In order to reduce annoying echoes in large or empty rooms, acoustic panels can be installed in the ceiling, on the walls, or both. The panels are made from sound-absorbing materials and create more neutral acoustics in spaces such as shopping malls, auditoriums, offices, and conference rooms. A similar effect can, however, be achieved by using curtains or other interior fabrics. Acoustic panels are usually quite effective for frequencies above 300 Hz, while the absorption capabilities gradually decrease for lower frequencies.

Figure 8.2a *Curtains and other pieces of fabric can significantly improve room acoustics.*

8.3 Audio technologies

Various audio technologies are involved in enabling network audio devices to provide high-quality sound. In security and safety installations, high quality typically means clear and intelligible speech, and this is ensured by use of several sound optimization techniques which are built into the speakers. The quality is maintained throughout the system by means of techniques for compressing, transmitting, and synchronizing audio streams.

8.3.1 Digital sound quality

In Axis network audio, the audio is digital all the way from the source to the speaker. The sound is digitally stored, processed, and transmitted, without any analog-to-digital or digital-to-analog conversions. The integrated digital signal processor makes sure that the sound is optimized right in the speakers. With predefined sound profiles for background music and voice, no technician has to go onsite to control the audio quality. Audio announcements are configured and controlled to be clear and easy to understand. The sound signal also remains strong no matter how long the cables are.

In analog audio, however, the sound waves are converted into electrical voltage. This electric signal can then be transported, stored, and eventually converted back into sound waves. When the electric signal gets passed through many steps, or over long hauls of wire, its voltage level will drop and will need to be amplified. This process can increase the noise and lower the quality, and the output audio will sound less good.

Digital audio uses numerical values instead of varying voltage to describe the waveform. These values can be stored and transmitted without changing the slightest, even in areas of electromagnetic disturbance. They will not be corrupted by noise, and when converted back into sound waves they will sound exactly the same as the original audio.

8.3.2 High quality for network audio

In the entertainment industry, the term "high-quality audio" may refer to hi-fi speakers with stereophonic sound. Such speakers are designed to reproduce audio very accurately at high loudness, and a speaker system may consist of several types of speaker elements to manage as many audible frequencies as possible. There may be a bass element that reproduces sound up to 500 Hz, a mid-range element for frequencies between 500 Hz and 9 kHz, and a treble element for frequencies above 9 kHz, for example. Entertainment systems can also create very powerful listening experiences through the use of stereophonic, or stereo, sound, where two or more independent audio channels are used in separate speakers in order to create the impression of sound coming from various directions.

Network audio from Axis, however, is not about high loudness, an extensive frequency range, or stereo sound. For security or safety purposes, our solutions are all about maximizing the loudness of the rather narrow frequency range where human speech is most discernible. Even when network audio is used for background music, the loudness is pretty low. And there is no need for stereo sound — when the intended audience consists of people moving around in a school, a hospital, or a store, there is no left or right. For these reasons, network audio uses mono speakers with relatively low loudness.

8.3.3 Sound optimization techniques

Axis network speakers come equipped with digital signal processing (DSP) capabilities. DSP for public announcement is about analyzing and manipulating sound to improve speech intelligibility. In Axis network speakers, several sound optimization techniques – such as frequency optimization, loudness compensation, and dynamic range control – are built into the speakers to deliver excellent audio quality in any environment.

> **Frequency optimization**. The edge processing in Axis network speakers means they are frequency optimized, which gives the same characteristics to every speaker. As a result, they can combine without the need for manual tuning or configuration, and the system can be easily expanded just by connecting more Axis speakers.

> **Dynamic range control**, or dynamic range compression. An audio signal will often have peaks and troughs in volume, and dynamic range control can balance these to make sure that sound is broadcast at the ideal volume for listeners.

> **Loudness compensation**. If you have a 'loudness' button on your stereo, you might be familiar with the basic concept. At low volumes, some frequencies are less perceptible to the human ear (see section 8.1.8 about perceived loudness). Loudness compensation boosts those frequencies so that the listener does not miss anything. This happens automatically in Axis speakers, which makes them suitable both for important audio messages and for background music.

8.3.4 Audio communication modes

Depending on the application, there may be a need to send audio in only one direction or both directions. This relates to three basic modes of audio communication:

> **Simplex** means that audio can be sent in one direction only. In network audio, audio is usually sent from an operator to a speaker, for example for communicating warnings or announcements through the speaker. But audio in simplex mode could instead be sent from the speaker to the operator, for example in remote monitoring applications where live audio from a monitored site is sent over a network.

> **Half duplex** means that audio can be sent and received in both directions, but only in one direction at a time. The direction is controlled either through use of a physical push-to-talk button or through voice detection software. This mode of communication is similar to a walkie-talkie conversation. Because speaker and microphone are never active at the same time, there is no risk of echo problems with half duplex.

> **Full duplex** means that users can send and receive audio (talk and listen) at the same time. This mode of communication is similar to a telephone conversation. Full duplex requires both the client (PC, SIP microphone, or VoIP phone) and the speaker to be able to handle full-duplex audio. The implementation must also support acoustic echo cancellation (AEC) in order to avoid echo effects.

Audio devices from Axis work with either half duplex or full duplex for two-way audio.

8.3.5 Echo cancellation

With half duplex technology, speaker and microphone are never active at the exact same time. This means that there is no risk of echo problems.

With full duplex technology, you can send and receive audio at the exact same time. The implementation must support acoustic echo cancellation (AEC) in order to avoid echo effects.

8.3.6 Audio codecs

An audio codec (encoder-decoder) is a software system that can digitize and compress data for transmission and decompress the received data. Axis products support various audio codecs:

AAC-LC (Advanced Audio Coding - Low Complexity), also known as MPEG-4 AAC, which requires a license. AAC-LC, particularly at a sampling rate of 16 kHz or higher and at a bit rate of 64 kbit/s or more, is the recommended codec to use when the best possible audio quality is required.

G.711 and **G.726**, which are non-licensed ITU-T standards. They have lower delay and require less computing power than AAC-LC. G.711 and G.726 are speech codecs that are primarily used in telephony and have low audio quality. Both have a sampling rate of 8 kHz. G.711 has a bit rate of 64 kbit/s. Axis G.726 implementation supports 24 and 32 kbit/s. With G.711, Axis products support only µ-law, which is one of two sound compression algorithms in the G.711 standard. When using G.711, it is important that the client also uses the µ-law compression.

Axis products that support SIP can also use Opus, L16/16000, L16/8000, speex/8000, speex/16000, G.726-32.

8.3.7 Synchronization of audio and video

Synchronization of audio and video data is handled by a media player, or by a multimedia framework such as WebRTC.

In combined audio/video products, audio and video are sent over a network as two separate packet streams. For the client or player to perfectly synchronize the audio and video streams, the audio and video packets must be time-stamped. The timestamping of video packets using Motion JPEG compression may not always be supported in a network camera. If this is the case and if it is important to have synchronized video and audio, the video format to choose is MPEG-4, H.264, or H.265 since such video streams, along with the audio stream, are sent using RTP (Real-time Transport Protocol), which timestamps the video and audio packets.

8.3.8 Synchronization between multiple network audio devices

Network audio from Axis supports inter-destination media synchronization. This means that synchronized playback is possible between multiple devices over any network, since synchronization between their clocks is maintained.

9. Networks and cybersecurity

This part of the technical guide provides information about the many network technologies behind network audio. It also presents how to protect the network and devices with cybersecurity best practices and features.

9.1 Network technologies

A network audio system employs many network technologies. Local area networks, in particular Ethernet, provide the foundation and also enables network audio products to be powered by Power over Ethernet (PoE). IP addressing and data transport protocols enable the products to be accessed over the internet, and technologies that provide Quality of Service allows different network applications to co-exist without consuming each other's bandwidth.

9.1.1 Local area networks and Ethernet

A local area network (LAN) is a group of computers that are connected in a localized area to communicate with one another and share resources such as printers. Data is sent in the form of packets, and different technologies regulate the transmission of the packets. The most widely used LAN technology is Ethernet, which is specified in the standard IEEE 802.3. Other types of LAN networking technologies include token ring and FDDI (Fiber Distributed Data Interface).

Today, Ethernet uses a star topology in which the individual nodes (devices) are connected to each other via active networking equipment such as switches. A LAN can contain several thousand networked devices.

A good rule of thumb is to always build a network with greater capacity than currently required. To future-proof it, you can design a network so that it only uses 30% of the total capacity when first put into use. As more and more applications run over networks, more and more network performance will be required. While network switches are easy to upgrade after a few years, cabling is normally much more difficult to replace.

9.1.1.1 Types of Ethernet networks

The following are the most common types of Ethernet networks used today. They can be based on twisted pair or fiber optic cables.

Fast Ethernet. Can transfer data at a rate of 100 Mbit/s, which is enough for most network audio applications. The older 10 Mbit/s Ethernet is still installed and used, but such networks may not provide the necessary bandwidth for some modern applications. Most devices are equipped with a 10BASE-T/100BASE-TX Ethernet interface, most commonly called a 10/100 interface, which supports both 10 Mbit/s and Fast Ethernet. The type of twisted pair cable that supports Fast Ethernet is called a Cat-5 cable.

Gigabit Ethernet. Supports a data rate of 1,000 Mbit/s (1 Gbit/s) and is now more commonly used than Fast Ethernet. 1 or 10 Gbit/s Ethernet may be necessary for the backbone network that connects many network devices. The type of twisted pair cable that supports Gigabit Ethernet is a Cat-5e cable, where all four pairs of twisted wires in the cable are used to achieve the high data rates. Most interfaces are backwards compatible with 10 and 100 Mbit/s Ethernet and are commonly called 10/100/1000 interfaces. For transmission over greater distances, fiber optic cables such as 1000BASE-SX (up to 550 m/1804 ft.) and 1000BASE-LX (up to 550 m with multimode optical fibers and 5 km/3 miles with single-mode fibers) can be used.

10 Gigabit Ethernet. Supports a data rate of 10 Gbit/s (10,000 Mbit/s). 10GBASE-LX4, 10GBASE-ER and 10GBASE-SR based on an optical fiber cable can be used to cover distances up to 10 km/6 miles). With a 10GBASE-T twisted pair solution, a very high quality cable (Cat-6a or Cat-7) is required. 10 Gbit/s Ethernet is mainly used for backbones in applications that require high data rates.

9.1.1.2 Connecting network devices and network switch

To network multiple devices in a LAN, network equipment, such as a network switch, is required. Its main function is to forward data from one device to another on the same network. The switch does this efficiently by directing data from one device directly to the target device, without affecting other devices on the same network.

A network switch works by registering the MAC (Media Access Control) address of each device that connects to it. Each and every networking device has a unique MAC address, made up of a series of figures and letters in hexadecimal notation, as set by the manufacturer. The address is often found on the product label. When a network switch receives data, it forwards it only to the port that is connected to the device with the appropriate destination MAC address.

Network switches typically indicate their performance in per port rates, and in backplane or internal rates (both in bitrates and in packets per second). The port rates indicate the maximum rates on specific ports. This means that the speed of a switch, for example 100 Mbit/s, is often the performance of each port.

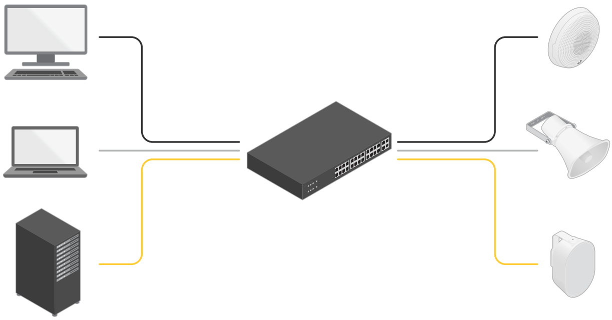

Figure 9.1a *In a network switch, data transfer is managed very efficiently as data traffic can be directed from one device to another without affecting any other ports on the switch.*

Network switches often have 10/100/1000 interfaces, thus supporting 10 Mbit/s, Fast Ethernet, and Gigabit Ethernet simultaneously. The transfer rate and mode between a port on a switch and a connected device are normally determined through auto-negotiation, whereby the highest common data rate and best transfer mode are used. A network switch also allows a connected device to function in full-duplex mode, that is, send and receive data at the same time, resulting in increased performance.

Network switches may come with different features or functions, for example, some may include router functionality. A switch may also support Power over Ethernet or Quality of Service, which controls how much bandwidth is used by different applications.

9.1.1.3 Power over Ethernet (PoE)

Power over Ethernet (PoE) is used to supply power to devices connected to an Ethernet network over the data-communication cable. Apart from its use in network audio, PoE is widely used to power IP phones, wireless access points, and network cameras.

The main benefit of PoE is the inherent cost savings. Hiring a certified electrician to install a separate power line is not required when running PoE. This is advantageous, particularly in difficult-to-reach areas. The fact that power cabling is not required can reduce costs and it also makes it easier to move a device to a new location.

Additionally, PoE makes it easier to make an audio system more secure. A system with PoE can be powered from the server room, which is often backed up by a UPS (uninterruptible power supply). This means that the system can stay operational even during a power outage.

Due to the benefits of PoE, it is recommended for use with as many devices as possible. The power available from the PoE-enabled switch or midspan should be sufficient for the connected devices and the devices should support power classification.

PoE standards

Most PoE devices today conform to the IEEE 802.3af standard. It uses Cat-5 or higher cables, and ensures that data transfer is not affected. In the standard, the device that supplies the power is referred to as the power sourcing equipment (PSE). This can be a PoE-enabled switch or midspan. The device that receives the power is referred to as a powered device (PD). The functionality is normally built into the network device, or it can be provided from a standalone splitter.

Backward compatibility to non-PoE compatible network devices is guaranteed. The standard includes a method for automatically identifying if a device supports PoE, and only when this is confirmed will power be supplied to the device. This also means that the Ethernet cable connected to a PoE switch will not supply any power if not connected to a PoE-enabled device. This eliminates the risk of electrical shock when installing or rewiring a network.

In a twisted pair cable, there are four pairs of twisted wires. PoE can use either the two 'spare' wire pairs, or it can overlay the current on the pairs used for data transmission. Switches with built-in PoE often supply power through the two pairs of wires used for transferring data, while midspans normally use the two spare pairs. A PD supports both options.

According to IEEE 802.3af, a PSE provides a voltage of 48 V DC with a maximum power of 15.4 W per port. Considering that there will be some power loss over a twisted pair cable, only 12.95 W is guaranteed as available for the PD. The standard specifies various performance categories for PDs.

PSE such as PoE-enabled switches and midspans normally supply a certain amount of power, typically 300-500 W. On a 48-port switch, this would mean 6-10 W per port, if all ports are connected to devices that use PoE. Unless the PDs support power classification, a full 15.4 W must be reserved for each port that uses PoE, which means a switch with 300 W can only supply power on 20 of the 48 ports. However, if all devices let the switch know that they are Class 1 devices, then 300 W will be enough to supply power to all 48 ports.

Other PoE standards are IEEE 802.3at (also known as PoE+) and IEEE 802.3bt. Using PoE+, the power limit is raised to at least 30 W via two pairs of wires from a PSE. For power requirements that are higher than the PoE+ standard, Axis uses the term High PoE, which raises the power limit to at least 60 W via four pairs of wires, and 51 W is guaranteed for the device (PD).

Class	Type	Minimum power level at PSE	Maximum power level used by PD
0	Type 1, 802.3af	15.4 W	0.44 W - 12.95 W
1	Type 1, 802.3af	4.0 W	0.44 W - 3.84 W
2	Type 1, 802.3af	7.0 W	3.84 W - 6.49 W
3	Type 1, 802.3af	15.4 W	6.49 W - 12.95 W
4	Type 2, 802.3at	30 W	12.95 W - 25.5 W
6	Type 3, 802.3bt	60 W	51 W
8	Type 4, 802.3bt	90 W	71.3 W

Table 9.1a *Power classifications according to IEEE 802.3af, IEEE 802.3at, and IEEE 802.3bt.*

Most network audio devices can receive power via PoE using the IEEE 802.3af standard and are normally identified as Class 3 devices.

Adding PoE support using midspans

Midspans are devices that enable an existing network to support PoE. The midspan, which injects power to an Ethernet cable, is placed between the network switch and the powered devices. To ensure that data transfer is not affected, the maximum distance between the source of the data (for example, the switch) and the network audio product must not exceed 100 m (330 ft.). The midspan is not a repeater and does not amplify the Ethernet data signal.

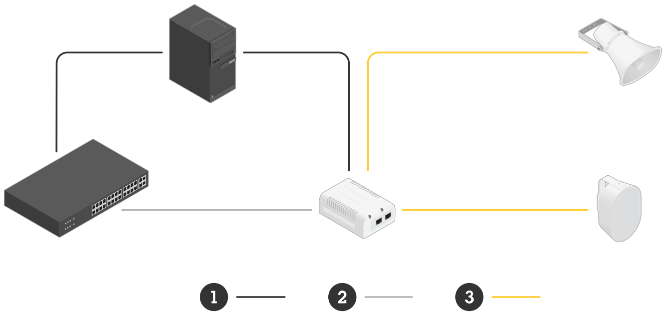

Figure 9.1b *An existing system with separate power (1) and Ethernet (2) cables can be upgraded with PoE functionality (3) using a midspan.*

9.1.2 Sending data over the internet

The basic elements of internet communication include:

Routers. To forward data packages from one LAN to another via the internet, network routers must be used. These devices route information from one network to another (hence the name) based on IP addresses. A router only forwards data packages destined for another network, and is most commonly used for connecting a local network to the internet. Routers are sometimes referred to as gateways.

Firewalls. A firewall is designed to prevent unauthorized access to or from a private network. Firewalls can be implemented in both hardware and software, or in a combination of both. Firewalls are frequently used to prevent unauthorized internet users from accessing private networks connected to the internet. Messages entering or leaving the internet pass through the firewall, which examines each message, and blocks those that do not meet the specified security criteria.

Internet connections. When connecting to the internet, the term upstream describes the transfer rate (bandwidth) at which data can be uploaded from the device to the internet. Downstream is the transfer speed for downloading files.

Sending data from a device on one LAN to a device on another LAN requires a standard method of communication, since local area networks may use different technologies. This requirement led to the development of IP addressing and the many IP-based protocols for communicating over the internet.

9.1.2.1 IP addressing

Devices wishing to communicate via the internet must have unique and appropriate IP addresses, which identify the sending and receiving devices. There are currently two IP versions: IP version 4 (IPv4) and IP version 6 (IPv6). The main difference between the two is that an IPv6 address is longer (128 bits compared with 32 bits for an IPv4 address). IPv4 addresses are the most commonly used today.

9.1.2.2 IPv4 addresses

IPv4 addresses are grouped into four blocks, with each block separated by a dot. Each block represents a number between 0 and 255; for example, 192.168.12.23.

Some blocks of IPv4 addresses have been reserved exclusively for private use. These private IP addresses are 10.0.0.0 to 10.255.255.255, 172.16.0.0 to 172.31.255.255 and 192.168.0.0 to 192.168.255.255. These addresses can only be used on private networks and are not allowed to be forwarded through a router to the internet. A device wanting to communicate over the internet must have its own individual, public IP address, which will be allocated by an internet service provider. The address can be either a dynamic IP address, which can change during a session, or a static address, which normally comes at an additional monthly fee.

Ports

A port number defines a particular service or application so that the receiving device (for example, a network speaker) will know how to process the incoming data. When a computer sends data tied to a specific application, it usually automatically adds the port number to an IP address.

Port numbers can range from 0 to 65535. Certain applications use port numbers that are pre-assigned to them by the Internet Assigned Numbers Authority (IANA). For example, a web service via HTTP is typically mapped to port 80 on a network audio device.

Setting IPv4 addresses

For a network audio device to work in an IP network, an IP address must be assigned to it. Setting an IPv4 address for an Axis network device can be done automatically using DHCP (Dynamic Host Configuration Protocol), which requires a DHCP server on the network. Another way to set the IP address is to use a management software tool such as AXIS Device Manager. Alternatively, the address can be set manually. One way to do this is to use the device’s web page to enter the static IP address, subnet mask, and the IP addresses of the default router, the DNS (Domain Name System) server and the NTP (Network Time Protocol) server.

A DHCP server manages a pool of IP addresses, which it can assign dynamically to network devices, with this function often being performed by a broadband router. The router in turn is typically connected to the internet and gets its public IP address from an internet service provider. Using a dynamic IP address means that the IP address for a network device may change from day to day. With dynamic IP addresses, it is recommended that users register a domain name for the network audio product at a dynamic DNS server, which can always tie the domain name for the product to any IP address that is currently assigned to it. A domain name can be registered using some of the popular dynamic DNS sites such as *www.dyndns.org*

Using DHCP to set an IPv4 address works as follows. When a network device comes online, it sends a query requesting configuration from a DHCP server. The DHCP server replies with the configuration requested by the network device. This normally includes the IP address, the subnet mask, and IP addresses for the router, DNS server and NTP server. The product first verifies that the offered IP address is not already in use on the local network, assigns the address to itself and can then update a dynamic DNS server with its current IP address so that users can access the device using a domain name.

With AXIS Device Manager, the software can automatically find and set IP addresses and show the connection status. The software can also be used to assign static and private IP addresses for Axis network devices. This is recommended when using audio management software to access network audio products. In a network audio system with potentially hundreds of devices, a software program, such as AXIS Device Manager, is necessary to effectively manage the system.

NAT (Network address translation)

When a network device with a private IP address wants to send information via the internet, it must do so using a router that supports NAT. Using this technique, the router translates the private IP address into a public IP address, for public exposure on the internet.

IP multicast

IP multicast allows a host to send audio streams to many other hosts, preferably in the same broadcast domain, without any unnecessary data-packet duplication. The sender can transmit to members in a group, defined by a Class D address within the range of 224.0.0.0 – 239.255.255.255. By allowing multiple users to access the same data stream, multicast conserves bandwidth compared to unicast, which requires one data stream per connected client.

9.1.2.3 IPv6 addresses

An IPv6 address is written in hexadecimal notation with colons subdividing the address into eight blocks of 16 bits each; for example, 2001:0da8:65b4:05d3:1315:7c1f:0461:7847.

The major advantages of IPv6, apart from the huge number of IP addresses it provides, include enabling a device to automatically configure its IP address using its MAC address. For communication over the internet, the host requests and receives, from the router, the necessary prefix of the public address block, as well as any additional information. The prefix and host's suffix are then used, so DHCP for IP address allocation and manual setting of IP addresses is no longer required with IPv6. Other benefits of IPv6 include renumbering to simplify switching entire corporate networks between providers, faster routing, point-to-point encryption according to IPsec, and connectivity using the same address in changing networks (Mobile IPv6).

An IPv6 address is enclosed in square brackets in a URL and a specific port can be addressed in the following way: http://[2001:0da8:65b4:05d3:1315:7c1f:0461:7847]:8081/

Setting an IPv6 address for an Axis network device is as simple as checking a box to enable IPv6 in the device. The device then receives an IPv6 address according to the configuration in the network router.

9.1.2.4 Data transport protocols

The Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) are the IP-based protocols used for sending data. These transport protocols act as carriers for many other protocols, for example, HTTP (Hyper Text Transfer Protocol), as used to browse web pages, is carried by TCP.

TCP provides a reliable, connection-based transmission channel. It ensures that data sent from one point is received at the other. TCP's reliability through retransmission may introduce significant delays, but in general TCP is used when reliable communication is preferred over reduced latency.

UDP is a connection-less protocol and does not guarantee delivery of the transmitted data, thus leaving the whole control mechanism and error-checking to the application itself. UDP does not re-transmit lost data and, therefore, does not introduce any further delay.

Protocol	Transport protocol	Port	Common usage	Network audio usage
FTP (File Transfer Protocol)	TCP	21	Transfer of files	Transfer of audio to an FTP server or an application
SMTP (Send Mail Transfer Protocol)	TCP	25	Protocol for sending email	The audio device sends images or alarm notifications via built-in email client.
HTTP (Hyper Text Transfer Protocol)	TCP	80	Browsing the web: getting web pages from web servers	The audio device functions as a web server, making audio available for the user or application server.
HTTPS (HTTP over Transport Layer Security)	TCP	443	Secure access to web pages using encryption	Secure transmission of audio to/from network audio devices.
RTP (Real-time Transport Protocol)	UDP/TCP	Not defined	RTP standardized packet format for delivering audio and video over the internet, often used in streaming media systems or video conferencing	Transmission of audio, and synchronization of audio and video. RTP provides sequential numbering and timestamping of data packets, enabling correct reassembly. Unicast or multicast.
RTSP (Real-Time Streaming Protocol)	TCP	554	Set up and control of multimedia sessions over RTP	

Table 9.1b *Common TCP/IP protocols and ports used for network audio.*

See also Axis documentation about commonly used ports at *help.axis.com/axis-os-knowledge-base#commonly-used-network-ports*

9.1.2.5 SIP

Session Initiation Protocol (SIP) is a text-based protocol, similar to HTTP and SMTP, for communication over IP networks. It is used to start, change, and terminate media stream sessions, which can include voice and video elements. SIP is the standard protocol used in Voice over IP

(VoIP) applications and unified communication platforms, for video conferencing, call control, and instant messaging. SIP constitutes a way to connect, integrate, and control Axis network products.

SIP calls can be set up in many ways, but there are three main types:

Peer-to-peer calls, also called local calls. These are calls between two devices (such as computers, network speakers, softphones, door stations, cameras, or IP desk phones) that belong to the same network. The call is made to the SIP address of the device.

SIP server calls, also called private branch exchange (PBX) calls. To make SIP server calls, the devices must be connected to a SIP server that handles the call exchanges. A SIP server, or a PBX, is a hub that works like a traditional switchboard. It can be hosted on an intranet or by a third-party service provider. The SIP-enabled devices register with the SIP server and can contact each other through their SIP addresses. A PBX can show call status, allow call transfers, handle voicemail, and redirect calls among other things.

SIP addresses, also known as SIP uniform resource identifiers (URIs) or SIP numbers, are used to identify users within a network, just like phone numbers or email addresses. Like email addresses, SIP addresses are a type of URI that includes two user-specific parts, a user ID or extension, and a domain name or IP address. Together with a prefix and the @ symbol, they make up a unique address. In the case of a peer-to-peer call, the SIP address would include the IP address rather than the domain name.

SIP trunk calls. With a service provider that offers SIP trunking, the traditional telephone network can be used to make calls, and traditional phone numbers can be assigned to the SIP devices. In this way, calls can be made from a network speaker or a network door station to a cell phone or the other way around.

9.1.3 VLANs

When a network audio system is designed, there is often a desire to keep the network separate from other networks, both for security as well as performance reasons. At first glance, the obvious choice would be to build a separate network. While the design would be simplified, the cost of purchasing, installing and maintaining the network would often be higher than using a technology called virtual local area network (VLAN).

VLAN is a technology for virtually segmenting networks, a functionality that is supported by most network switches. This can be achieved by dividing network users into logical groups. Only users in a specific group can exchange data or access certain resources on the network. If a network audio system is segmented into a VLAN, only the servers located on that VLAN can access the network audio devices. VLANs provide a flexible and more cost-efficient solution than a separate network, and can be used to, for example, separate an audio network from an office network.

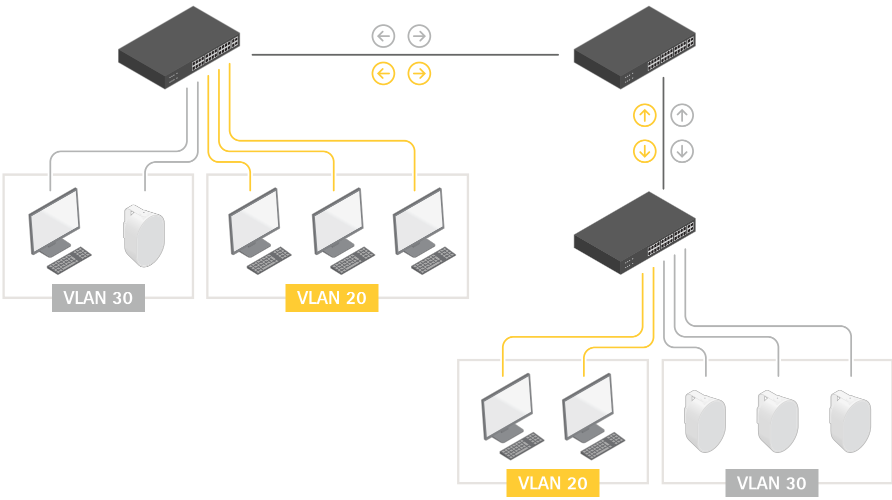

Figure 9.1c *In this illustration, VLANs are set up over several network switches. The two different LANs are segmented into VLAN 20 and VLAN 30, and only members of the same VLAN can exchange data, either within the same network or over different networks.*

The primary protocol used when configuring VLANs is IEEE 802.1Q, which tags each frame or packet with extra bytes to indicate which virtual network the packet belongs to.

9.1.4 Quality of Service

Since different applications—for example, telephone, email and audio—may be using the same IP network, there is a need to control how network resources are shared to fulfill the requirements of each service. One solution is to let network routers and switches operate differently for different kinds of services (voice, data, and audio) as traffic passes through the network. By using Quality of Service (QoS), different network applications can co-exist on the same network without consuming each other's bandwidth.

The term Quality of Service refers to a number of technologies, such as Differentiated Service Codepoint (DSCP), which can identify the type of data in a data packet and so divide the packets into traffic classes that can be prioritized for forwarding. One main benefit of a QoS-aware network include the ability to prioritize traffic to allow critical flows to be served before flows with lesser priority. Another is greater reliability in the network, by controlling the amount of bandwidth an application may use and thus controlling bandwidth competition between applications. A prerequisite for the use of QoS in an audio network is that all switches, routers and network audio products must support QoS.

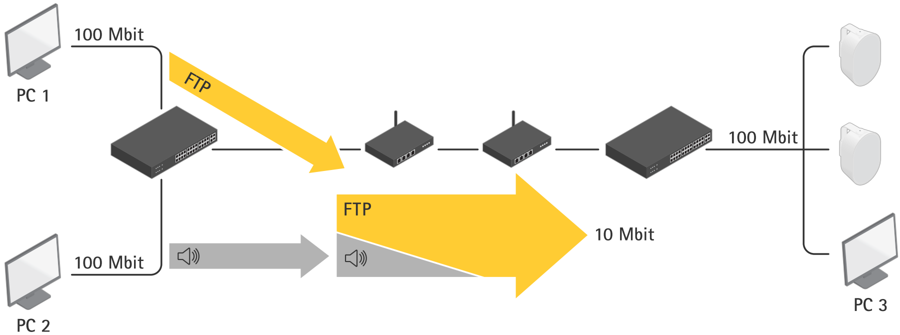

Figure 9.1d *Standard (non-QoS aware) network. In this example, FTP and audio streaming have 10 Mbit/s to share and the audio bandwidth cannot be guaranteed.*

In the illustration, PC1 is sending two audio streams to the two speakers, with each audio clip streaming at 2.5 Mbit/s. Suddenly, PC3 starts a file transfer from PC2. In this scenario, the File Transfer Protocol (FTP) will try to use the full 10 Mbit/s capacity between the routers, while the audio streams will try to maintain their total of 5 Mbit/s. The amount of bandwidth given to the audio system can no longer be guaranteed and the audio quality will probably be reduced. At worst, the FTP traffic will consume all the available bandwidth.

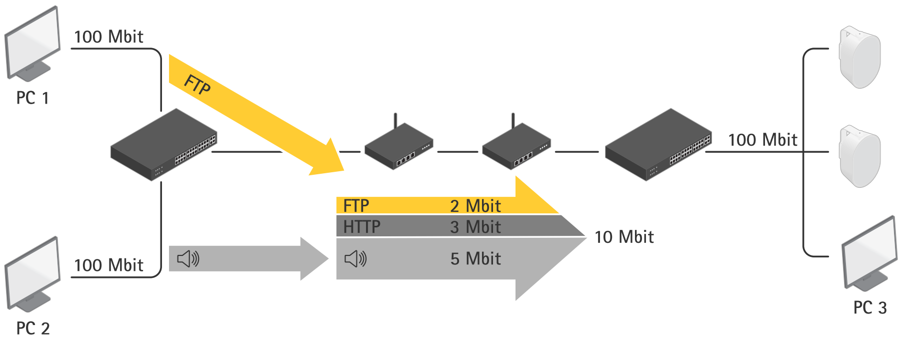

Figure 9.1e *QoS aware network. In this example, audio streaming has a guaranteed bandwidth of up to 5 Mbit/s.*

In this illustration, the first router has been configured to use up to 5 Mbit/s of the available 10 Mbit/s for streaming audio. FTP traffic can use 2 Mbit/s, and HTTP and all other traffic can use a maximum of 3 Mbit/s. Using this division, audio streams will always have the necessary bandwidth available. File transfers are considered less important and get less bandwidth, but there will still be bandwidth available for web browsing and other traffic. Note that these maximums only apply when there is congestion on the network. If there is unused bandwidth available, this can be used by any type of traffic.

9.2 System protection

Cyberthreats are commonly associated with hackers and malware, but negative impact can also be the result of unintentional misuse. When you deploy a system, it is recommended that you follow industry best practices. To be protected, a system needs to be both well configured and well maintained.

9.2.1 Network protection

A network needs protection from cyberthreats. All packages sent on the network may be collected by other computers on the same network. If the payload in the packages is sent in clear text, the data can be easily compromised, through what is called network sniffing. Another threat is network spoofing, which is when an attacking computer tries to impersonate a legitimate server, computer, or network device in order to get access to the network. Encrypted connections and CA-signed certificates provide protection.

For guidance on how to reduce the network's exposure to risks, see *Axis Network Switches Hardening Guide* at *https://help.axis.com/axis-network-switches-hardening-guide*.

For best practices about how to onboard and operate Axis devices in HPE Aruba Networking powered networks, see the integration guide *HPE Aruba Networking - Integration Guide* at *help.axis.com/axis-aruba-secure-network-integration*. The best-practice configuration uses modern security standards and protocols such as IEEE 802.1X, IEEE 802.1AR, and HTTPS.

9.2.1.1 IEEE 802.1X

Many Axis network products support IEEE 802.1X, which is a method used to protect a network against connections from unauthorized devices. IEEE 802.1X establishes a point-to-point connection or prevents access from the LAN port if authentication fails. IEEE 802.1X prevents what is called port hijacking; that is, when an unauthorized device gets access to a network through physical access to a network port/socket. IEEE 802.1X is useful in network audio applications, since network speakers are often located in public spaces where an openly accessible network socket can pose a security risk. In today's enterprise networks, IEEE 802.1X is becoming a basic requirement for anything that is connected to a network.

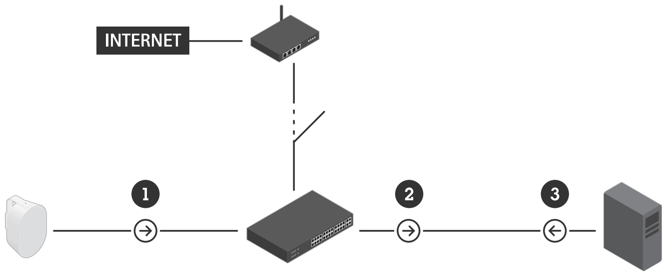

Figure 9.2a *IEEE 802.1X enables port-based security.*

1. A network speaker that is configured for IEEE 802.1X sends a request for network access to a switch.
2. The switch forwards the query to an authentication server. This can be a RADIUS (remote authentication dial-in user service) server, such as a Microsoft Internet Authentication Service server.
3. If authentication is successful, the server instructs the switch to open the port, to allow data from the network speaker to pass through the switch and onto the network.

9.2.1.2 HTTPS (HTTP over TLS)

HTTPS (Hyper Text Transfer Protocol Secure) is a secure communication method that sends HTTP inside a Transport Layer Security (TLS) connection. This means that the HTTP connection and the data itself are encrypted.

To enable an Axis network speaker to communicate over HTTPS, a digital certificate and an asymmetric key pair must be installed in the product. The key pair is generated by the Axis product. The certificate can either be generated and self-signed by the Axis product, or it can issued by a certificate authority. In HTTPS, the certificate is used for authentication and encryption, which means that the certificate allows a browser to verify the identity of the product, and it encrypts the communication using keys that are generated by public-key cryptography.

9.2.1.3 SRTP (Secure RTP)

SRTP (Secure Real-time Transport Protocol or Secure RTP) is an extension to RTP (Real-time Transport Protocol) that incorporates enhanced security features. Like RTP, it is intended particularly for VoIP (Voice over IP) communications. SRTP uses encryption and authentication to minimize the risk of denial-of-service (DoS) attacks. SRTP can achieve high throughput in diverse communications environments that include both hard-wired and wireless devices. Provisions are included that allow for future improvements and extensions.

9.2.1.4 Trusted public key infrastructure (PKI)

A trusted public key infrastructure (PKI) consists of a private or public certification authority (CA), which is a service that issues (signs) certificates to be installed in networked devices. Certificates are used for end-to-end encryption of TLS-based connections between hosts in a network. A CA-signed certificate with a validated trust chain enables applications, such as a VMS or a web browser, to validate the identity of the Axis networked device. Commonly, a CA certificate (root-certificate) signs an intermediate CA certificate, which then in return signs client certificates for end devices. For this validation to succeed, the publicly known CA certificate (root and/or intermediate certificate) must be installed and used by the application to verify client certificates that are carried by Axis networked devices. AXIS Device Manager has a built-in CA service that can cost-efficiently issue and deploy server certificates to the devices.

9.2.1.5 Network isolation

Network isolation, or network segmentation, is a way to separate critical network resources from each other in order to reduce the risk of each of them having a negative impact on each other. This is an especially relevant tactic if different resources do not need to interact with each other — or should not. Network segmentation can be virtual (VLAN) and require an infrastructure of managed switches, or the networks can be separated with different cabling and network gear.

9.2.2 Device protection

Each speaker and its resources must be duly protected. Protection may in this case refer to both physical protection, such as placing speakers out of reach, and protection of the speaker's software.

AXIS OS Hardening Guide is a continuously updated document that provides more information about best practices of device protection.

9.2.2.1 Built-in cybersecurity platform

Axis devices are safeguarded by the hardware-based cybersecurity platform Axis Edge Vault. It relies on a strong foundation of cryptographic computing modules (secure element and TPM) and SoC security (TEE and *secure boot*), combined with expertise in edge device security.

Axis Edge Vault minimizes exposure to cybersecurity risks and enables the device to be a trusted and reliable unit within the network.

> **Trusted device identity**. Being able to verify the origin of the device is key to establishing trust in the device identity. During production, devices with Axis Edge Vault are assigned a unique, factory-provisioned, and IEEE 802.1AR-compliant Axis device ID certificate. It is securely and permanently stored in the secure keystore and can be leveraged for automated secure device onboarding and secure device identification.

> **Secure key storage**. The secure keystore provides hardware-based, tamper-protected storage of cryptographic information. It protects the Axis device ID as well as customer-loaded cryptographic information, and prevents unauthorized access and malicious extraction in the event of a security breach.

> **Supply chain protection**. The cybersecurity features *secure boot* and *signed OS*, help establish a secure foundation for Axis Edge Vault by providing an unbroken chain of cryptographically validated software. *Secure boot* prevents physical supply chain tampering by ensuring that a device can boot only with Axis signed OS. *Signed OS* guarantees that the installed operating system (AXIS OS) is genuinely from Axis and ensures that any new operating system to be downloaded and installed on the device is also signed by Axis. If the device detects that the integrity is compromised or that the operating system is not signed by Axis, the upgrade will be rejected.

9.2.2.2 User account management

Device passwords tend to spread within an organization. For example, during device maintenance someone might request the password in order to adjust something. A couple of days or weeks later, someone else might have the same request. Within a short space of time, many new (or temporary) users know the password to all your devices, and you have lost control over who can access them. The strength of the password makes no difference in this scenario. Device management should involve using multiple accounts (role-based) and temporary accounts should be created for occasional maintenance and troubleshooting.

For account access, you should use the principle of least privileged accounts. This means that user access privileges are limited solely to the resources needed for a user's specific tasks. AXIS Device Manager helps you easily and efficiently manage multiple accounts and passwords, at the viewer, operator, or administrator level.

9.2.2.3 IP address filtering

Axis products provide IP address filtering, which allows or denies access to defined IP addresses. A typical configuration is to configure the network speakers to allow only the IP address of the server hosting the audio management software to access the product.

IP filtering acts like a local firewall in the speaker. The only computer or server that should be accessing speakers during normal operations is the audio management server. The speakers can be configured with an IP filter to only respond to allowlisted IP addresses, typically the audio management software server and administration clients. IP filtering helps mitigate risks if the device password is compromised from unpatched devices or from brute-force attacks.

9.2.2.4 Keeping device software up to date

Running devices with an up-to-date AXIS OS version mitigates common risks. This is because the latest version generally includes security patches for all known vulnerabilities, including those newly discovered. As a consequence, attackers cannot exploit the vulnerabilities and possibly compromise the system, application, or devices. AXIS Device Manager and AXIS Device Manager Extend help you manage upgrades of multiple devices and sites, and automatically notify you if there are newer versions.

As a CVE Numbering Authority (CNA), Axis follows industry best practices in managing and responding to security vulnerabilities discovered in our products. Vulnerabilities are scored using the widely used CVSS rating system, and patched within a specified period of time depending on the score.

To protect from the specific threat of supply chain tampering, the *signed OS* and *secure boot* features are widely available in Axis devices.

9.2.3 Physical protection

While a networked device can never be 100% physically protected, there are various physical aspects to consider. You can use a speaker design and color that blends into the environment, place the speaker out of reach, and always run the cable directly through the wall or ceiling behind the speaker. These principles provide substantial protection from both spontaneous vandalism and planned attacks. Important network equipment, such as routers and switches, and the host running the audio management software should be placed in an environment with physically and logically restricted access.

10.Online tools

Axis provides tools to facilitate audio installations at *axis.com/tools*

Find and compare products:

> **Product Selector** helps you find and compare Axis products.

> **AXIS Site Designer** helps you plan and design an audio installation (as well as a video installation), including which speakers to use and how many speakers are needed.

Plan and design sites:

> As a first step, we recommend the document **Quickguide for speaker coverage calculation**. It provides rules of thumb to help you estimate the number of speakers needed on a site.

> As the second step, use **AXIS Site Designer** which helps you plan and design an installation, including which speakers to use and how many speakers are needed.

> If you need even more advanced design help, Axis provides **GLL files for EASE® Evac design tool**. These are speaker input files which can be imported into the commercial design tool EASE® Evac to optimize speaker placement for a carefully designed sound. Similarly, you can use **Axis plugin for Autodesk® Revit®** to place Axis products in Autodesk® Revit® building plans.

Install and manage systems:

> **AXIS Device Manager.** Helps you manage all major installation, security, and operational tasks of your devices, for example, device configuration, device software upgrade, restore settings, and cybersecurity controls.

T10144682/EN/M2.18/2405

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