

The Standards

A videoconference link requires:

- transmitting and receiving equipment at each site for more details see the VTAS guide Videoconferencing Audio and Video Equipment?
- an intervening network to carry the signals.

In the case of IP based conferences other network related equipment is normally required to establish a connection, namely gatekeepers. The role of these devices is explained fully in the factsheet H323 Videoconferencing Components.

The network to be traversed can involve one or more of the following:

- Local Area Network (LAN), e.g. a university campus?
- Regional Network, supporting a city or region?
- Wide Area Network (WAN), extending to national and international sites.

These networks may comprise a number of physical transmission methods including: fibreoptic cables, coaxial transmission lines, copper twisted pair cable, satellite and high frequency radio. The latter includes longrange terrestrial microwave links up to 320km and the newer shorrange (10cm10m) cableless connection systems.

To enable transmission of information over these networks different data transmission methods are available that may broadly be broken down into two categories:

1. Switched Circuit Networks (SCN) that include:
 - NISDN Narrowband Integrated Services Digital Network (NISDN) (used to transfer data over digital telephone lines)?
 - General Switched Telephone Network (GSTN), a very narrow bandwidth method using existing analogue telephone lines.
2. Packet Based Networks (PBN) that include:
 - Internet Protocol (IP) (sometimes referred to as packet based format).

It is frequently a requirement to 'bridge' more than one network type to achieve a link, e.g. one organisation with IP capable equipment may need to communicate to another that only has an ISDN connection. Gateways (sometimes termed bridges) are pieces of equipment that can transparently translate the communication between different network types.

The ITU-T has produced several umbrella videoconferencing standards, collectively known as the H.3xx videoconferencing standards.

Table 1: The H.3xx Umbrella Videoconferencing Standards

Network Type	ITU-T Standard	Description
--------------	----------------	-------------

ATM	H.310	Broadband conferencing over A
N-ISDN (Narrowband ISDN)	H.320	Narrowband conferencing over
B-ISDN (Broadband ISDN)	H.321	An adaptation of H.320 enabling
GQoS (Guaranteed Quality of Service)	H.322	Guaranteed Quality of Service o networks
IP (Internet Protocol)	H.322	Narrowband conferencing over
GSTN (General Switched Telephone Network)	H.324	Low bit rate (very narrow band telephone lines

Within these umbrella standards are several substandards specific to a particular area of the signal, e.g. G.72x defines the audio coding and H.26x the video coding.

Videoconferencing Standards

This is the umbrella standard for IP conferencing. It includes several sub standards: H.261 defines the mandatory video Coder/Decoder (CODEC)* standard, whereas H.263 and H.264 define optional video CODECs.

Similarly G.711 is the mandatory audio CODEC and G.729 one of several audio CODEC options. The complex operation of managing the data streams from the CODEC including calling, establishing a call, and controlling the various component parts i.e. video, audio and data is defined by two standards, H.225 and H.245.

The capacity available to an application on a basic IP network varies with the amount of data traffic carried, so with the basic standard there is no Guaranteed Quality of Service (GQoS) i.e. the received quality can vary a great deal, from acceptable to very poor. Within the H.323 standard there are however suggested methods for maintaining quality, to overcome this limitation.

** A CODEC provides the compression and signal processing to enable high bandwidth sound and vision signals to be transmitted and received over low bandwidth transmission paths.*

Figure 1: H.323 Conferencing Standard

H.322

H.322 is the umbrella standard for IP conferencing that provides a guaranteed quality of service within LANs. Other methods are now in use to provide GQoS beyond local networks and these are covered in section 3.9.

H.310

This is the umbrella standard for broadband conferencing over ATM networks.

H.324

This is the umbrella standard for very low bandwidth conferencing, e.g. over GSTN (i.e. telephone networks).

H.324/M

This is the standard for visual telephone terminals over mobile radio. It is not really applicable to videoconferencing but included for completeness.

H.320

The umbrella standard for N-ISDN (usually abbreviated to ISDN) videoconferencing includes separate sub standards for video coding, audio coding and data format. Options include improved video CODECs, still image transfer, far end camera control, multipoint control, data sharing/exchange etc.

Figure 2: H.320 Conferencing Standards

H.321

The umbrella standard for BISDN? this adapts H.320 (N-ISDN) narrowband ISDN to work within ATM environments.

Mandatory Standards

Within each ITU-T umbrella standard, minimum mandatory standards are defined that will guarantee compatibility, albeit at a basic level, e.g. within H.320 provision must be made for H.261 video coding, G.711 audio coding and H.221, H.230, H.242 communications protocols. Similarly for H.323, the corresponding mandatory standards are H.261, G.711 and H.225/H.245 communication protocols. These mandatory requirements will allow all compliant products to communicate easily and effectively.

Some substandards are common throughout the range of umbrella standards, e.g. H.261 video coding and G.711 audio coding are mandatory in H.320, H.321, H.323, H.324 and H.310.

Optional Standards

Other, optional, substandards are defined to allow enhanced performance, e.g. H.243 provides for multipoint control function, i.e. when two or more sites conference there is provision for sending signals through a Multipoint Control Unit (MCU). H.281 provides for far-end camera control from the local site, H.282/H.283 provide the requirements for remote control of devices other than the camera, and T.120 provides for data exchange.

Proprietary Standards

Manufacturers may also choose to include proprietary enhancements, e.g. Polycom's Siren Audio extends the audio bandwidth up to 14kHz to improve the sound quality. These proprietary enhancements are not international standards so only provide a benefit when used between products from the same manufacturer. Proprietary standards should not be confused with 'options' within the ITU-T standards. The options are not mandatory but when incorporated will allow improved compatible communication between dissimilar equipment without problems.

Videoconferencing Substandards

The substandards most likely to be met with in practice are detailed below:

Video Coding Standards

H.261 Video CODEC

For audio visual services? this defines the way in which the picture information is compressed and coded to enable transmission over low bandwidth networks. It is the baseline coding which is mandatory for most videoconferencing systems to ensure interoperability at a basic level.

H.261 Annex D Graphics

The coding format for transmission of still images over H.320 conferencing at a screen resolution up to a maximum of 704 x 576 pixels, i.e. 4 x CIF. (See also 3.6.4.1 below)

H.262 (MPEG2)

Video coding used in broadband, i.e. H.310 ATM, conferencing systems.

H.263 Video CODEC

For audio visual services, a variation of the H.261 CODEC but specifically designed for low bit rate transmission, i.e. H.324 (GSTN) and H.323 (IP) networks at 64128kbit/s.

H.263+ Video CODEC

H.263+ is an enhanced version of H.263 coding giving improved coding efficiency at the expense of increased CODEC complexity.

H.263++ Video CODEC

H.263++ is an even more efficient CODEC, particularly for pictures containing movement.

H.264 Video CODEC

H.264 is also known as MPEG4 Advanced Video Coding (AVC). The latest video CODEC developed jointly between the ITU-T and ISO/IEC. It uses more sophisticated compression techniques than H.263 coding and is designed to require less bandwidth for an equivalent quality signal using other compression algorithms.

Audio Coding Standards

G.711

To ensure interoperability between systems G.711 is the baseline audio coding algorithm. It is mandatory in most videoconferencing systems. This coding produces an upper frequency limit of 3.4kHz/s (i.e. telephone quality) and occupies up to 64kbit/s of data.

G.722.1

An improved coding for audio signals giving higher quality signals with an upper frequency limit of 7kHz/s but only occupying 4856/64kbit/s of data.

G.723.1

Coding for ultra low bandwidth applications and occupying only 5.3/6.3kbit/s.

G.728

Low bit rate coding producing 3.4kHz upper frequency limit but occupying only 16kbit/s of bandwidth.

G.729

Coding for very low bandwidth applications and occupying 8kbit/s.

Structure for Communication (i.e. data stream formats)

H.221

Defines the frame structure for 641920kbit/s audio visual channels, i.e. videoconferencing up to 1920kbit/s (in H.320 systems).

H.224

A protocol for real time simplex control, i.e. one-way communication.

H.225.0

Call signalling and packet multiplex protocols for packet based (i.e. H.323) conference systems.

H.230

Frame control and indicating signals for conferencing equipment.

H.231

Multipoint control signals for conferencing channels up to 1920kbit/s (i.e. for communication between three or more sites conferencing up to 1920kbit/s).

H.233, H.234, H.235

Encryption option for H.3xx conferences.

H.241

Extended video procedures for H.3xx series terminals.

H.242

System for establishing communication between terminals in H.320 conference systems up to 1920kbit/s.

H.243

Protocol for communication between three or more conferencing units up to 1920kbit/s, i.e. multipoint conferencing.

H.245

Control protocol used in H.310 and H.323 conferencing systems.

H.281

Far end camera control, i.e. control of the remote site's camera from the local site.

H.282, H.283

Remote control of devices other than a camera.

H.323 Annex Q

Far end camera control within H.323 systems. This has now been superseded and is included within the latest (07/2003) H.323 recommendations.

H.331

Broadcasting type audio visual multipoint systems and terminal equipment.

Still Image Transfer Formats

H.261 Annex D and T.81

H.320 systems can offer the option for transferring still images at a resolution greater than the basic H.261 video resolution. It is H.261 Annex D coding. This provides a resolution up to a maximum of 4xCIF i.e. 704 x 576 pixels. While these still images are being transmitted then the normal motion videoconference images are suppressed. Alternatively some products offer Joint Photographic Expert Group (JPEG) still image coding (see 3.8.1 below) which is defined by ITU-T standard T.81.

ITU-T Substandards Applied to an H.323 CODEC

Figure 3 shows the components of a typical H.323 videoconferencing CODEC. A simplified diagram, it is intended to illustrate how the various standards apply within a videoconferencing system. The flow lines are bidirectional.

The vision transmit path starts at the local camera (video input), the output video signal then being coded and compressed by the video Coder (part of the video CODEC) before being multiplexed with the audio and other data streams. It then feeds to the network (IP in this case).

The inverse path takes an IP data signal arriving (via the network) from the remote site? it is demultiplexed into separate video, audio, data and control signals and then directed to the relevant DECoder e.g. to the video DECoder (part of the video CODEC). The decoded video finally feeds the local picture monitor (video output) to give an image from the remote site. The ITU-T standards that are relevant at each stage are shown on the diagram.

Figure 3: Block Diagram of an H.323 CODEC (*In this diagram, 'RAS control' refers to Registration, Admission and Status control)

H.235 Security and Encryption for H.323 Conferences

This ITU standard defines the security and encryption for H.323 and other H series connections that utilise the H.245 control protocol. H.323 networks by their nature do not guarantee either Quality of Service (QoS) or security of the data. The two main concerns are authentication and privacy. Authentication enables an endpoint to verify that a caller is who they say they are. The privacy of data can also be a worry during conferences as without precautions an H.323 network is relatively easy to interrogate.

The standard has been developed over several years and has three versions: 1, 2 and 3. Each iteration supersedes its predecessor. In common with other ITU standards there are mandatory, recommended and optional requirements. Within the standard are Annexes defining interoperability at specific levels of security:

- Annex D defines the baseline measures that are utilised in managed environments with symmetric keys/passwords assigned among the entities (terminal-gatekeeper, gateway-gatekeeper). This method uses a simple but secure password profile protection. It may also incorporate voice encryption for secure speech transmission.
- Annex E is an optional suggested signature security profile deploying digital signatures, certificates and a public-key infrastructure. As administration of passwords is not required between entities it enables much more efficient connection to a final endpoint via gatekeepers, gateways and MCUs on the network. It may incorporate annex D voice encryption and/or random number data encryption for messages.

To achieve maximum interoperability the ITU recommends that CODECs should have the ability to negotiate and to be selective concerning the cryptographic techniques utilised, and the manner in which they are used.

T.120 Document and Data Sharing Standards

The main standard in use for data sharing within videoconferencing is T.120.

Equipment that is T.120 compliant interleaves the data sharing information within the pass band of the H.320, H.323 etc. conferencing channel. This is an asset as sound, vision and data are shared across a single channel, but it can also be a hindrance as with low bandwidth channels, e.g. ISDN2, the T.120 data exchange part can degrade the audio and video signals to an unacceptable degree.

For further information, see the VTAS guide, Data Sharing within Videoconferencing.

Figure 4: T.120 Umbrella Standard for Document and Data Sharing

The T.120 standard for data exchange includes its own group of substandards e.g. T.127 defines the standard for file transfer under T.120. T.120 is designed to fit within the data stream of the conferencing system, i.e. H.320, H.321, H.323 and H.324 – an umbrella within an umbrella.

Figure 5: T.120 Interleaving Other Signals within the H.3xx Data Stream

T.120 Substandard

T.121

Generic application template to which application software must conform to operate under T.120.

T.122

Defines the transport of control and data sharing in multipoint conferencing.

T.123

Defines the protocol standard for each particular network supported, i.e. ISDN, GSTN, IP etc.

T.124

Generic conference control for the start, finish and control of conferencing.

T.125

Multipoint communications protocol.

T.126

Multipoint still image and annotation protocol, i.e. to enable the use of a whiteboard and shared applications.

T.127

Multipoint file transfer, i.e. to enable file transfer during a multisite conference.

T.128

Audio and video control.

T.140 Text Conversation

Not included within T.120 but sometimes seen in videoconferencing products. Equipment designed to T.164 is compliant with the protocol for multimedia text conversation.

Other Standards

JPEG

ISO/IEC standard 109181 (also defined by ITU.T standard T.81). This is an international

standard for the compression and coding of continuous tone still images. This standard includes several methods of compression depending on the intended application. JPEG is a 'lossy' method of compression as it loses some detail during the coding/decoding process. It can be adjusted however to be very economical in terms of data rate. 'Lossless' algorithms on the other hand can be decoded to reproduce the original detail but require higher data rates for transmission.

MPEG-1

This is a popular standard for the compression and coding of moving images and sound. It is the format used to store material on CDROM and CDI? the maximum data rate obtained is 1.5Mbit/s.

MPEG1 has three elements:

- MPEG1 ISO/IEC 11172-1 defines the MPEG1 multiplex structure, i.e. the way in which the digital audio/video/control data is combined?
- MPEG1 ISO/IEC 11172-2 defines the MPEG1 video compression and coding?
- MPEG1 ISO/IEC 11172-3 defines the MPEG1 audio coding.

MPEG-1 is a widely used compression format and has been used for CDROM production. It has an upper video resolution of 352 x 288 pixels (i.e. CIF) which while adequate for many applications represents only a quarter of the SDTV (Standard Definition Television) resolution of 704 x 576. Because of this limitation, to meet the needs of the broadcasters the MPEG-2 standard was developed.

MPEG-2

- MPEG-2 ISO/IEC 138181 defines MPEG2 data stream formats.
- MPEG2 ISO/IEC 138182 defines MPEG2 video coding.
- MPEG2 ISO/IEC 138183 defines MPEG2 audio coding.

Basically MPEG-2 is a 'compression toolbox' which uses all the MPEG-1 tools but adds new ones. MPEG-2 is upwardly compliant, i.e. it can decode all MPEG-1 compliant data streams. MPEG-2 has various levels of spatial resolution dependent on the application.

- Low level, i.e. 352 x 288 pixels (CIF resolution)
- Main level, i.e. 720 x 576 pixels (Programmable Array Logic (PAL) TV resolution)
- High level, i.e. 1440 x 1152 pixels (high definition TV)
- High level wide screen, i.e. 1920 x 1152 pixels.

MPEG-2 has further options regarding the algorithms used for coding and compressing the information these are known as 'profiles'.

- Simple Profile uses a simple Encoder and Decoder but requires a high data rate.
- Main Profile requires a more complex Encoder and Decoder at a greater cost but requires a lower data rate.
- Scalable Profiles which allow a range of algorithms to be transmitted together e.g. basic encoding for decoding by an inexpensive decoder and enhanced encoding, which can be accessed by a more sophisticated and more expensive decoders.

- High Profile to cater for High Definition Digital Television Video (HDTV) broadcasts.

The most common MPEG-2 set is Main profile, Main level, used for television broadcasting. Depending on the quality required the data rate can vary from 49Mbit/s. Data rates for the whole MPEG-2 family can vary between 1.5 and 100Mbit/s.

MPEG-4

MPEG-4 is a comprehensive format that builds on the MPEG-1 and MPEG-2 standards. It is designed to provide a mechanism whereby multimedia content can be exchanged freely between the producers (e.g. the broadcasters and record companies), the distributors (telephone companies, cable networks, Internet Service Providers (ISPs) etc.) and the consumers. This content can be audio, video and/or graphic material. The delivery can be oneway or interactive and may be streamed in real time. This all-encompassing standard spans digital broadcasting, interactive graphics and multimedia over the Internet and includes 3G multimedia phones. The standard has numerous profiles for audio, video, graphics etc.

MPEG4

AVC, sometimes referred to as 'MPEG4 part 10', is the one most likely to be met with in videoconferencing.

MPEG-4 AVC

The ISO/IEC has collaborated with the ITU to develop this new standard also known as H.264. It is expected to eventually replace MPEG-2 and MPEG-4 standards in many areas due to its more efficient coding algorithms. It is claimed that bandwidth can be reduced by 50% when compared to H.263 compression. Another big advantage of H.264 is its inbuilt IP adaption layer, allowing it to integrate into fixed IP, wireless IP and broadcast networks with ease. It is also expected to find new applications in areas such as Asymmetric Digital Subscriber Line (ADSL).

Motion Joint Photographics Expert Group (MJPEG)

While MPEG encoding is now used extensively it does have some serious limitations for some applications and particularly for videoconferencing. The MPEG encoding/compression process in common with H.261/H.263 coding of video signals functions by eliminating a high proportion of both redundant spatial and temporal picture elements. In doing this it requires a considerable amount of time to actually complete the process (termed latency). In practice this 'latency' demands that the audio signal be delayed by a similar amount so that lip synchronisation can be preserved within a conference. To ensure realism 'echo cancellers' then have to be incorporated to reduce echo between sites to an acceptable level.

If the temporal structure of the vision signal is left intact and JPEG frames are joined together the resultant coding is called MJPEG or Motion JPEG. This signal format can overcome most of the latency problems. Unfortunately no single standard has yet evolved for joining the JPEG frames together so MJPEG itself is not an international standard as are the JPEG and MPEG formats.

MJPEG coding/compression reduces the redundant spatial picture elements but does not affect the temporal elements (i.e. redundancy between successive frames). The process

therefore generates far less latency than MPEG or H.261 systems and it is found that echo cancellers are generally not necessary. This reduces cost and has the potential to improve sound quality.

This reduction in latency is quite marked. For MJPEG CODECs end-to-end audio delay is typically 60 microseconds whereas for an ISDN-2 system (i.e. 128kbit/s) the delay could be as much as 400 microseconds (i.e. almost half a second). MJPEG encodes only vision signals, so another coding algorithm (usually G.711) or high quality Pulse Code Modulation (PCM) is used for the audio information.

Guaranteed Quality of Service (GQoS)

The increased popularity of IP (H.323) based services due mainly to the lower cost of connection has spawned a great deal of development to produce effective methods of delivering high quality videoconference (and telephone traffic) over the IP infrastructure. The Internet Engineering Task Force (IETF) has been particularly active in defining standards in this area while the major network equipment manufacturers have produced workable network solutions.

IETF RFC 2205 Resource Reservation Set up Protocol

This IETF recommendation modifies the normal routing control protocols and allows a host to request specific Qualities of Service (QoS) for the audio/video content of a conference. It is also used by routers to deliver QoS requests to all nodes on the network and to manage the QoS state.

IP Precedence (IPP)

IP Precedence enables an endpoint to prioritise the video/audio data into five Types of Service (ToS) with respect to that of other traffic on the network.

These choices are available:

- Maximum throughput?
- Maximum reliability?
- Minimum delay?
- Minimum cost?

in addition to the default 'Normal' that has no priority.

The ToS tag commands routers to prioritise data and so low priority traffic may have to be dropped at busy times to enable a reliable conference.

Differentiated Services (DiffServ)

DiffServ is a more sophisticated method than IP precedence of specifying Type of Service (ToS). With DiffServ there are 63 separate ToS available.

Intelligent Packet Loss Recovery (IPLR)

In cases where despite the best efforts of QoS procedures data packets are still lost then some more advanced CODECs attempt to minimise the visual effect of the loss sometimes downspeeding the conference to regain stability. Polycom®'s PVEC and Tandberg's IPLR are two examples. The Polycom® solution is proprietary and so needs a similar CODEC at each end whereas the Tandberg is H.323 and H.320 compliant and so works with any compliant endpoint or MCU.

Source URL: <https://community-stg.jisc.ac.uk/library/videoconferencing-booking-service/standards>