Video Conferencing Standards and Terminology

There is an ever increasing number of standards, terminologies and buzz-words used within the video conferencing industry that can make understanding what is both available and compatible a minefield. We have the H.200 & 300's, the G.700's and the H.460's, not to mention ISDN, LAN, WAN, VPN, ADSL, 802.11 b/g/n wireless and SIP all mixed with High-Definition, 1080p, 720p, NTSC, PAL and CIF. To complicate matters more, we also now have to deal with media-enabled 3G/4G smartphones, tablets and their mobile data networks and how these link in with existing systems. This document explains what these standards, terminologies and buzz-words mean, how they relate to the various communications infrastructures of video conferencing and how they relate to each other.

It is assumed that the reader has a general knowledge of Video Conferencing systems. However, the following technical papers are available to provide more information:

- How do I choose a Video Conferencing system?
- IP Ports and Protocols used by H.323 Devices.
- H.323 Terminals, Gatekeepers, Gateways & MCUs.
- H.221 Framing used in ISDN Conferences.
- Cost Efficient ISDN Conferencing, including Multipoint Access.
- H.323 Dial Plan and Service Codes used by Gatekeepers etc.

International Telecommunications Union & The Internet Engineering Task Force.

Global telecommunications standards are set by the United Nations agency, International Telecommunications Union (ITU) and the Internet Engineering Task Force (IETF). Products that adhere to these standards allow users to participate in a conference, regardless of their platform. These standards for video conferencing ensure compatibility on a worldwide basis. The ITU has developed the H, G and T Series of standards whilst the IETF has developed Real-Time Protocol (RTP) & Resource Reservation Protocol (RSVP). These standards apply to different transport media.

Transport Protocols.

There are several standards based transport protocols used with conferencing, TCP, UDP & RTP. Generally, each configures the data into packets, with each packet having a ‘header’ that identifies its contents. The protocol used is usually determined by the need to have reliable or unreliable communications.

TCP is a reliable protocol designed for transmitting alphanumeric data; it can stop and correct itself when data is lost. This protocol is used to guarantee sequenced, error-free transmission, but its very nature can cause delays and reduced throughput. This can be annoying, especially with audio.

User Datagram Protocol (UDP) within the IP stack, is by contrast, an unreliable protocol in which data is lost in preference to maintaining the flow.

Real-Time Protocol (RTP) was developed to handle streaming audio and video and uses IP Multicast. RTP is a derivative of UDP in which a time-stamp and sequence number is added to the packet header. This extra information allows the receiving client to reorder out of sequence packets, discard duplicates and synchronise audio and video after an initial buffering period. Real-Time Control Protocol (RTCP) is used to control RTP.

Update on where we are today.

The speed and worldwide availability of ADSL and the Internet along with the national telephone companies has virtually stopped the availability and use of POTS as a direct means of connecting video conferencing
systems. In its place we now have Fast ADSL (Fibre), Cable (Fibre) and the forthcoming media-enabled 3G/4G smartphones and tablets as well as next generation of Codecs and Gateways to transcode the new protocols.

You also need to be aware of new and emerging standards that might have an impact on what you purchase. The latest video compression used by video conferencing systems is H.264 and its derivatives H.264 High Profile and H.264 SVC. As a guideline, basic H.264 offers twice the quality over its predecessor H.263 at the same bandwidth, or the same quality at half the bandwidth. H.264 High-Profile has even higher performance and the latest H.264 SVC is scalable and more flexible across networks. So if you are restricted in the available bandwidth, take a look at systems that support the latest video compressions.

There's also been changes in the way data collaboration is achieved; with the development of the H.239 (Dual Video) standard and 'data-showing' being favoured and replacing the old and now obsolete T.120 'data sharing' standard. H.239 defines how additional media channels are used and managed by video conferencing systems. It introduces the concept of 'data-showing', whereby the PC desktop is digitised and converted into a separate video stream and transmitted in parallel with the main 'talking heads' video stream - hence the term Dual Video. Endpoints that support H.239 will receive the dual streams and display the desktop graphics and far-end video in separate windows. Endpoints that don't support H.239 will display the shared desktop graphics instead of the far-end video.

Network, Infrastructure & Devices.

Before you start, it is useful to have an understanding of what types of networks are available.

They all have strengths and weaknesses that should be considered carefully before deciding upon which to use. Please take a look at the diagram below that tries to shows the networks, infrastructure and devices used in H.323, SIP and H.320 standards based videoconferencing.
Integrated Digital Services Network (ISDN).

There are two available ISDN connections, Basic Rate Interface (BRI) and Primary Rate Interface (PRI). Essentially, a BRI provides two 64kbps B-channels and one 16kbps D-channel. In Europe, a PRI provides 30 x
64kbps B-channels and two 64kbps D-channel - total 2048kbps; whilst in North America an PRI provides 23 x 64kbps B-channels and one 64kbps D-channel, giving at total 1544kbps.

ISDN connections usually aggregate the BRI and share the same number for both B-channels. Known as ISDN-2, this provides a line speed of 128kbps is typically used by desktop video conferencing Systems over ISDN. For increased bandwidth, ISDN-6 provides a line speed of 384kbps and is typically used by group or room-based video conferencing Systems over ISDN. With ISDN-6, the sequence in which the lines are aggregated must be known and adhered too! Furthermore, if the connection is going to use some form of 'switch', this must be configured to pass both voice and data!

The ISDN connections are usually directly into the video conferencing system and it uses the H.221 framing protocol and adheres to the H.320 standard. Less common is to use an ISDN Dial-Up modem that effectively transmits IP over ISDN, in which case the video conferencing system would have to following the H.323 standard.

In the past, most H.320 conferences would have been between just two participants as ISDN is essentially a point-to-point connection. However, multipoint technology now makes it possible for groups of people to participate in a conference and share information. To hold a multipoint conference over ISDN, participants must use either a dedicated ISDN Multipoint Control Unit - MCU that connects and manages all the ISDN lines, or an endpoint with an embedded H.320 multipoint capability.

H.320 is the ITU standard for ISDN conferencing and includes:

- **Audio**: G.711, G.722, G.722.1, G.728, AAC-LC, AAC-LD
- **Video**: H.264, H.263, H.261
- **Data**: H.239
- **Control**: H.221, H.231, H.242, H.243

LAN, WAN, VPN and Intranet.

100 Mbps LANs with switches and routers are used in most companies today and these have enough bandwidth to support desktop conferences. With a LAN offering significantly more bandwidth than ISDN, the video quality within a conference is much higher and can approach that of HD television. Technology as also helped, we now have communications advancements such as Gigabit Ethernet (1000 Mbps), Faster Switches as well as Fibre Asynchronous Digital Subscriber Lines (ADSL), Synchronous Digital Subscriber Lines (SDSL), 802.11 b/g/n wireless and 4G Mobile networks that have increased the available bandwidth, whilst IP Multicasting (routers permitting) has reduced network loading in conferences involving more than two endpoints.

Unlike ISDN networks, LANs, WANs and VPNs across the Intranet, Internet, ADSL, SDSL, Wireless and 3G/4G Mobile networks all use TCP/IP protocol and the H.323 standard defines how to assemble the audio, video, data and control (AVDC) information into an IP packet. Most companies use DHCP and allocate dynamic IP addresses to PC's. Therefore, in order to correctly identify a user, the H.323 endpoints are usually registered with a Gatekeeper and 'called' into a conference by their H.323 alias. The Gatekeeper translates the alias into the corresponding IP address. Another method of identifying H.323 users is for them to register their presence using Light Directory Access Protocol (LDAP) with a Directory Service such as Microsoft's Windows Active Directory or the freely available OpenLDAP.

To hold a multipoint conference over a TCP/IP network, H.323 systems require a Multipoint Conference Server (MCS). This is also referred to as an H.323 Multipoint Control Unit (H.323 MCU). This is not the same as an H.320 MCU; hence it is important to be clear about what you mean when using the term MCU.

To hold a large scale multipoint conference over IP, participants must use a separate dedicated MCU connected to the IP network. For small scale multipoint conferences, there are now endpoints with an embedded H.323 multipoint capability that support up to 6 endpoints in a single conference.
H.323 is the ITU standard for LAN conferencing and includes:

- **Video:** H.264 High Profile, H.264, (H.264 SVC), H.263, H.261
- **Data:** H.239
- **Control:** H.225, H.245, H.460

**Wireless 802.11 a/b/g/n networks.**

Standards based 802.11 a/b/g/n wireless networks are readily available forms of transport media for home, company and travelling users. With 802.11 b/g and now 802.11 n routers giving transmission speeds of up to 108 Mbps, there is sufficient bandwidth available to support audio, video and data sharing across wireless networks, especially when used in conjunction with the latest compression techniques and technologies.

Like with LAN above, H.323 is the ITU standard for conferencing across wireless networks.

**3G/4G mobile networks (and Public WiFi Hotspots).**

The 3G/4G cellular mobile data networks (and Public WiFi Hotspots) are a readily available form of wireless delivery and with the media-enabled Smartphones and Tablets, there is sufficient bandwidth to enable IP-based multipoint audio and video conferencing to existing H.323 video conferencing systems when used in conjunction with next generation Gateways and MCU’s that also support these new protocols.

With greater coverage, the faster 3G and the even faster 4G data networks have now made 3G-324M defunct.

Like with LAN and 802.11 wireless, H.323 is the ITU standard for conferencing across 3G/4G mobile data networks.

**Internet, ADSL, SDSL & VPN.**

With its ever increasing popularity, people have sought to use the Internet in more ways than just a means of sending email or browsing interesting sites.

Like LANs, ASDL and SDSL are other forms of TCP/IP networks for accessing the Internet and hence can be used as a transport media in video conferencing systems. Both ASDL (including Fibre ASDL) and SDSL use a modem and router (or a router with built-in modem) in order to gain access to the Internet. What each user should do is get their Internet Service Provider (ISP) to provide them with a fixed Public IP address.

Alternatively, users could register their presence with a Dynamic DNS Service Provider such as DynDNS.org. But after rebooting, a modem that is allocated a dynamic IP address could then be allocated a different IP address and any changes will take time to propagate through the DNS Service before these changes are recognised.

This is how you know or determine the address of the endpoint that you want to conference with.

For a more secure and faster connection, ISP and telecoms companies are now offering VPN over ADSL and SDSL links. A VPN provides a secure tunnel over the provider’s network by applying encryption between sites. With most Firewalls supporting VPN pass-thru, there is no need to open lots of ports. However, be wary of applying too much encryption as this can cause an unacceptable delay in the transmission between sites.

ADSL and SDSL, whilst being faster than ISDN, are only as fast as the slowest uplink when used for video conferencing. Again, users should get their DSL Service Provider to provide them with a fixed Public IP address for their xDSL Modem/Router/Firewall. Most xDSL Modems now incorporate a Router and Firewall. Depending upon whether the video conferencing system is PC or non-PC based, it can either be located behind a Firewall.
Proxy (PC-based), within their Firewalls DMZ (De-Militarised Zone) or outside on the Internet (non-PC based). Otherwise, too many Firewall ports may have to be opened in order to provide access, which defeats the objectives of having a Firewall.

**H.323** is the ITU standard used for Internet conferencing.

**Video standards:**

**H.261** - video codec for audiovisual services at p x 64Kbps.

**H.263** - video codec for narrow telecommunications channels at < 64 Kbps.

Notable elements of the standard are image size. QCIF is Quarter Common Intermediate Format and represents a 176x144 pixel image. This is the minimum size that must be supported to be H.320 compliant. CIF is the optional full-screen H.320 video image of 352x288 pixels and requires considerably more computing capability.

Note: whilst this is termed full-screen, it is nowhere near the size of a typical PC or laptop screen (1650x1050) pixels.

**H.264/AVC** - latest video codec widely used by current video conferencing systems.

In 2001, the ISO Motion Picture Experts Group (MPEG) recognised the potential of this ITU-T development and formed the Joint Video Team (JVT) that included people from MPEG and VCEG. The result is two identical standards: ISO MPEG4 Part 10 and ITU-T H.264, with the official name Advanced Video Coding (AVC).

There is little functional difference between the elements of H.264 and those of the earlier H.261 and H.263 standards. The changes that do make the difference lie mainly in the detail within each element, how well the algorithm is implemented and whether it is performed in hardware or software.

The basic technique of motion prediction works by sending a full frame followed by a sequence of frames that only contain the parts of the image that have changed. Full frames are also known as 'key frames' or 'I-frames' and the predicted frames are known as 'P-frames'. Since a lost or dropped frame can cause a sequence of frames sent after it to be illegible, new 'I-frames' are sent after a predetermined number of 'P-frames'. It is the combination of both lossy compression and motion prediction that allows H.261, H.263 and H.264 systems to achieve the required reduction in data whilst still providing an acceptable image quality.

With hundreds of experts involved in creating H.264, there were many options. Some being simpler and immediately implemented, whilst others were much more complex, but still included. Hence H.264 was organised into four profiles; Baseline, Extended, Main and High. Baseline is the simplest and uses 4:2:0 chrominance sampling and splits the picture into 4x4 pixel blocks, processing each block separately. Baseline uses Universal Variable Length Coding (UVLC) and Context Adaptive Variable Length Coding (CAVLC) techniques which have a big impact on the network bandwidth. Virtually all vendors support H.264 Baseline and some are now also supporting H.264 High Profile.

**H.264 High Profile** is the most powerful and efficient. This is achieved by using Context Adaptive Binary Arithmetic Coding (CABAC) encoding. High Profile also uses adaptive transformations to decide 'on-the-fly' how to split the picture into blocks - 4x4 or 8x8 pixels. Areas of the picture with little detail use 8x8 blocks whilst more complex and detailed areas use 4x4 blocks.

**H.264 SVC** - emerging video codec that is not yet fully interoperable between vendors.
Vendors are now introducing H.264 SVC (Scalable Video Coding) into their products. H.264 SVC is the latest adaptive technology that delivers high quality video across networks with varying amounts of available bandwidth. Formerly known as H.264 Annex G, H.264 SVC promises to increase the scalability of video networks. In stark contrast to other H.264 AVC family members (including H.264 High Profile) with which video endpoints send one resolution, one frame rate and one quality, H.264 SVC enabled video endpoints send multiple layers of resolutions (spatial), frame rates (temporal) and quality depending upon what the endpoints and network can support. This approach allows for ‘scalability’ as each endpoint can select which layers of video it needs without any additional encoding or decoding. This selecting of video layers is independent and does not affect other endpoints. It also allows each endpoint to gracefully degrade the video quality when it or network gets busy.

However, the H.264 SVC codec is only part of the interoperability equation as it also involves networking components such as signalling and error correction, which are not currently included in the standard. Hence, H.264 SVC is still essentially proprietary with vendors such as Polycom, Radvision and Vidyo each having their own flavour of SVC. Eventually, a complete standardised version of H.264 SVC will emerge that will offer true interoperability. But until then, you need to stick with the same vendor across the endpoints.

Audio standards:

G.711 - Pulse Code Modulation of voice frequencies (PCM), were 3.1 kHz analogue audio is encoded into a 48, 56 or 64 kbps stream. Used when no other standard is equally supported.

G.722 - 7 kHz audio encoded into a 48, 56 or 64 kbps stream. Provides high quality, but takes bandwidth.

G.722.1 - 7 kHz audio encoded at 24 and 32 kbps for hands-free operation in systems with low frame loss.

G.722.1 Annex C - The ITU's adoption of Polycom's Siren 14 - a 14 kHz audio codec.

G.722.2 - Coding of speech at around 16 kbps using Adaptive Multi-Rate Wideband, AMR-WB. Five mandatory modes, 6.60, 8.85, 12.65, 15.85 and 23.85 kbps.

G.723.1 - 3.4 kHz dual rate speech codec for telecommunications at 5.3 kbps & 6.4 kbps.

G.728 - 3.4 kHz Low Delay Code Excited Linear Prediction (LD-CELP) were 3.4 kHz analogue audio is encoded into a 16 kbps stream. This standard provides good quality results at low bitrates.

G.729 A/B - 3.4 kHz speech codec that provides near toll quality audio encoded into an 8 kbps stream using the AS-CELP method. Annex A is a reduced complexity codec and Annex B supports silence suppression and comfort-noise generation.
**MPEG-4 AAC-LC** - Low Complexity Advanced Audio Coding (AAC-LC) 8-96 kHz at 8-256 Kbps

**MPEG-4 AAC-LD** - Low Delay Advanced Audio Coding (AAC-LD) is the high-quality low-delay audio coding standard within MPEG-4. 22-48 kHz at 8-576 Kbps mono; 16-1152 Kbps stereo

**Data and Control standards:**

**H.221** - defines the transmission frame structure for audovisual teleservices in channels of 64 to 1920 Kbps; used in H.320

**H.223** - specifies a packet-orientated multiplexing protocol for low bit rate multimedia communications; Annex A & B handles light and medium error prone channels of the mobile extension as used in 3G-324M.

**H.224** - defines real-time control protocol for simplex applications using the H.221 LSD, HSD and HLP channels.

**H.225** - defines the multiplexing transmission formats for media stream packetisation & synchronisation on a non-guaranteed QoS LAN.

**H.231** - specifies multipoint control units used to bridge three or more H.320 systems together in a conference.

**H.233** - Confidentiality systems for audiovisual services, used by H.320 devices.

**H.234** - Encryption key management and authentication system for audiovisual services, used by H.320 devices.


**H.239** - defines role management and additional media channels for H.300-Series multimedia terminals. How data and web-enabled collaboration work in parallel with video in a conference, allowing endpoints that support H.239 to receive and transit multiple, separate media streams - typically voice, video and data collaboration.

**H.241** - defines extended video procedures and control signals for H.300-Series multimedia terminal.

**H.242** - defines the control procedures and protocol for establishing communications between audiovisual terminals on digital channels up to 2 Mbps; used by H.320.

**H.243** - defines the control procedures and protocol for establishing communications between three or more audiovisual terminals - H.320 multipoint conferences.

**H.245** - defines the control procedures and protocol for H.323 & H.324 multimedia communications.

**H.246** - Interworking of H-Series multimedia terminal.

**H.248** - Gateway Control Protocol.

**H.281** - defines the procedures and protocol for far end camera control (FECC) in H.320 calls.

**H.282** - Remote device control protocol for multimedia applications.

**H.283** - Remote device control logical channel transport.

**H.350** - Storing and retrieving video and voice over IP information from enterprise directories.
ANNEX Q - defines the procedures and protocol for far end camera control (FECC) in H.323 calls.

NAT/Firewall Traversal:

H.460.17 - defines method of discovering the ability of an H.323 entity to support this feature as well as the mechanism of encapsulating of RAS messages inside H.225.0 messages. Thus using the same transport protocol for both RAS and H.225.0 call signalling.

H.460.18 & H.460.19 - Together these define how H.323 endpoints traverse NAT/Firewall installations with no additional on-premise equipment, or alternatively, these extensions may be implemented by a proxy server to support unmodified H.323 endpoints.

H.469.18 - enables H.323 signalling to traverse NAT/Firewall installations. The H.460.18 architecture consists of a network which is divided into an internal and an external network by a NAT/Firewall. The H.323 internal endpoint and the external H.460.18 Traversal Server work together to enable bidirectional communication across the NAT/Firewall, and discover the transport addresses that have been modified by the NAT/Firewall.

H.460.19 - defines a mechanism for media communication between two H.323 entities, separated by one or more NAT/Firewall devices. It also defines a mechanism to use the same transport address for several media channels, which permits reduction of the number of “pinholes” open in the NAT/Firewall device and reduces the number of Media Channel and Media Control Channel transport addresses used by H.323 entities.

Connectivity:

ANNEX O - (URI Dialling) - defines how to utilize DNS for resolving addresses in the form of H.323 URLs.

BONDING - Bandwidth ON Demand Interoperability Group, synchronises the B-channels to transmit as one stream and attain higher data rates.

DID - Direct Inward Dialling is a method of routing H.320 incoming calls directly to H.323 endpoints without operator intervention.

DTMF - Dual Tone Multi-Frequency signals are the type of audio signals used in telephony for tone dialling.

E.164 Number - (User Number). A numeric string given to an H.323 endpoint. If this endpoint registers with a Gatekeeper, then the Gatekeeper can translate the E.164 Number into the endpoints IP address.

H.323 Alias - A logical name given to an H.323 endpoint. If this endpoint registers with a Gatekeeper, then the Gatekeeper can translate the H.323 Alias into the endpoints IP address.

IVR - Interactive Voice Response is a two-stage DID method of routing H.320 calls that is supported by the Gateway. It enables an H.320 endpoint to directly contact an H.323 endpoint using DTMF tones to control the connection.


MSN - Multiple Subscriber Numbering. When the PSTN Company assigns a group of telephone numbers to one line.

Q.931 - Signalling protocol for establishing and terminating calls.

RAS - Registration/Admission/Status. A communications protocol used between H.323 endpoints and the Gatekeeper for registration, admission and status messages.
Protocol. An IETF specification for audio and video signal management. Allows applications to synchronize audio and video packets.

**SIP** - Session Initiation Protocol.

**TCS-4** - Terminal Control Strings are another DID method of routing H.320 calls that is supported by the Gateway. The TCS-4 string contains information that is used to identify the H.323 endpoint, such as its E.164 number.

**Video and PC Window Sizes:**

- **1080i60** - Full High-Definition 1920 x 1080 pixels with interlaced (two passes) update at 60 frames per second.
- **1080p30** - Full HD 1920 x 1080 pixels using progressive update (full scan each pass) at 30 frames per second.
- **720p60** - Entry-level HD 1280 x 720 pixels using progressive update at 60 frames per second.

**NTSC** - National Television Standards Committee, used in USA, Canada & Japan. 640 x 480 pixels.

**PAL** - Phase Alternation by Line, used in Europe (except France), Africa & Middle East. 768 x 576 pixels.

**SECAM** - Sequentielle Couleur Avec Memoire, used in France & Russia.

**4CIF** - 4x Common Intermediate Format; optional for both H.263 & H.264, 704 x 576 pixels.

**CIF** - Common Intermediate Format; optional for both H.261 & H.263, 352 x 288 pixels.

**QCIF** - Quarter Common Intermediate Format; required by H.261; H.263 & H.264, 176 x 144 pixels.

**SQCIF** - Sub Quarter Common Intermediate Format; used by 3G mobiles MPEG4 video and H.263, 88 x 72 pixels.

**WSXGA+** - 1680 x 1050 pixels - used by latest high end laptops and PC workstations.

**SXGA** - 1280 x 1024 pixels - used by basic laptops and PC workstations.

**XGA** - 1024 x 768 pixels - entry-level PC or laptop resolution.

**SVGA** - 800 x 600 pixels.

**VGA** - 640 x 480 pixels.

**Misc:**

**RTFM** - A diagnostic instruction generally given to eager installers of hardware and software when things don't quite work .... *Read the FFFF.... Manual!*