THE FUTURE OF VISUAL COMMUNICATIONS: APPLICATIONS FOR EDUCATION AND THEIR IMPACT ON THE IT INFRASTRUCTURE

Stefan Karapetkov, Polycom, Inc., 3553 North First Street, San Jose, California 95134, United States of America, stefan.karapetkov@polycom.com.

KEYWORDS

Visual Communications, VC2, Video, Content Sharing, Telepresence, High Definition, People On Content, Live Music Mode, Video Recording, Video Streaming, Lost Packet Recovery, Bandwidth Management, QOS Policies.

EXECUTIVE SUMMARY

The emergence of high-speed IP networks such as Internet2 in the USA and GEANT2 in Europe are creating opportunities for new kinds of real-time applications in universities and schools around the world. Bandwidth is no longer limiting, but rather driving new applications that were impossible even few years ago. The most profound changes are in the area of bandwidth-hungry visual communication applications.

Distributed visual communication will enable new levels of interaction among scientists, researchers and students. It will be easy to use and accessible anywhere and anytime. Telepresence - a new kind of visual communication that provides a 'just as being there' experience – will bring distributed teams closer together, and will be deployed by leading organizations to extend the reach of their higher education programs.

Content sharing – an essential component of team collaboration – will go beyond sharing slides and spreadsheets, and allow for sharing video clips, animation, and three-dimensional images.

Recording, content management, and streaming applications will emerge as a natural complement to real-time communication. These applications will make visual communication much more scalable and accessible from any computer and mobile device.

This paper explores the future of collaboration, and focuses on the applications for the education market. It introduces new technologies that led to the emergence of telepresence, and discusses how these technologies change and improve the user experience.

The paper also discusses on the impact of these new applications on the IP network, and the mechanisms IT has to implement in order to support them. It focuses on end-to-end quality of service, and introduces new technologies in the area packet loss recovery. The paper concludes with a discussion of the required interaction between application and IT infrastructure.

1. VISUAL COMMUNICATIONS

Visual communication is the most natural form of human interaction. Long before writing was invented people were painting pictures on cave walls, and telling stories through these pictures. Throughout human history, paintings - and later photographs - captured the key events of humanity in much more powerful way than any writing can do. Television and moving images captured the imagination of people in the 20th century. Figure 1 summarizes these social developments.

Figure 1: Visual Communication Progression

With every new generation of technology, visual communication became more powerful and true to reality. Today's high definition video is the pinnacle of communications technology and the result of progress over centuries.

2. BUSINESS COMMUNICATIONS

Business communication up until the 19th century was limited to face to face meeting and an occasional letter carrier. This special limitation led to mostly local business interactions. The invention of the telegraph gave business people means to communicate across countries and even continents. Missing was the human interaction – telegraphs did not transmit any emotional information, so communication partners were not able to develop trust.

The invention of the telephone improved the situation. Voice communication provided more context - business partners could hear each other's voices and guess emotions. Although voice communication became ubiquitous in the 20th century and became wireless towards the end of the century, it never fully solved the problem of transmitting emotions across distances. Figure 2 summarizes the development of business communication over the centuries.

Figure 2: Business Communication Progression

Visual communication promises to deliver the ultimate communication and collaboration experience. New technologies such as telepresence make the visual communication life-like and equivalent to a face-to-face meeting while personal video simplifies the use and makes video easily accessible in the everyday business life.

3. THE FUTURE OF VISUAL COMMUNICATIONS

There will be a dramatically different marketplace for video in the years ahead. Social, economic, and technological trends are aligning to create a unique opportunity for new and innovative forms of visual communication. This combination of factors will bring video into the mainstream and make visual communication essential in both our personal and professional lives. Users in this emerging marketplace demand new and richer experiences, and they will apply visual communications in unforeseen ways. This will require that traditional applications like video conferencing, video training, and more are transformed into broader and more impactful forms of visual communication. We call this transformation VC2.

VC2 is both about convergence of the user experiences and about network convergenvce. Figure 3 is a preview of the convergence that is already beginning to happen in enterprises and in educational organizations.

Figure 3: Communications Convergence

The EUNIS presentation will use several video clips to illustrate the new wave of visual applications: pervasive video communications in the office and on the road, the use of next-generation VC2 technologies to reach the right balance between personal life and work, new ways for visual design and collaboration using video, and a dinner that extends the boundaries of time and space through telepresence.

4. INTEGRATED VISUAL COMMUNICATION

As video goes mainstream it becomes a pervasive tool across organizations. It gets seamlessly integrated in the everyday communication tools and merges with voice communication, instant messaging, and presence applications, i.e. it becomes part of the integrated visual communication. Future integrated visual communication will be able to address any audience, and will be ubiquitous, i.e. it will happen at any time and place, and by the use of any device. The key elements of integrated visual communication are described in Figure 4.

Figure 4: Integrated Visual Communication

The first element is about scalability. Visual communication systems of the future will support individuals (e.g. through personal telepresence), teams (e.g. through room telepresence systems and multipoint conferencing), and scale to thousands of users. New applications such as personal telepresence and desktop video will necessarily impact IT due to the increased bandwidth consumption in the IP network. High scalability also means thousands of video-enabled devices in the IP network, and the need for robust and scalable device provisioning and management.

The second element will give users choices how to access video. In addition to real-time calls, the visual communication system of the future will support new applications such as streaming, recording, and video content management, as well as scalable distribution through Content Delivery Networks and IP Multicast.

The third element gives users the choice where to communicate. While today's enterprise video communication is mostly in conference rooms, future visual communication will allow users to access video on the road and from home.

The fourth element is about the increased choices users will have which device to use for video. Room-based and personal telepresence system will be complemented by PCs and mobile devices, all seamlessly communicating.

5. VC2 CHARACTERISTICS

How are VC2 visual communications different from traditional video conferencing? Figure 5 provides a comparison.

Traditional Video	VC2
Difficult for user	Easy & intuitive
Reasonable quality	High Definition
Primarily room	Room, Desktop & Mobile
Telecom-oriented	IT integrated
Limited scale	Very large scale
Difficult to manage	Common provisioning & management
Proprietary	Open, standards-based

Figure 5: VC2 Characteristics

Numerous user surveys are telling us that video is perceived as difficult to use. The user interface is fairly complex and there are a lot of moving parts (codecs, cameras, displays, microphones) that can - if not set up correctly - impair the user experience. VC2 in comparison is easy and intuitive for users. One way of achieving that is through integration of video into collaboration tools that are used daily.

While traditional video only provides reasonable quality (from CIF to Standard Definition) VC2 makes High Definition video ubiquitous. HD enables new set of applications such as remote teaching, telemedicine, telehealth, telejustice, and remote quality assurance that require crisp images. HD audio is essential to achieving immersive telepresence experience and to transmitting music performances over IP networks. The main impact on IT is that HD video requires a lot of bandwidth (minimum 1.2Mbps per call at 720p quality).

Today's video communication is mostly in conference rooms. VC2 includes group conferencing but make video also available on desktops and mobile devices. This transition will increase the importance of standards and open platforms, as users will be connected through terrestrial, mobile, and other networks in multi-party video calls. IT will need to provide the network connectivity and create policies that allow users to access the network and use VC2 services.

Video has been overlay application with own management tools for long time. In the VC2 future, video will be integrated in the IT infrastructure and will use standard IT management and provisioning tools. This will enable high scalability and reliability that are essential for rolling out large video deployments. New applications will emerge for monitoring and managing Video QOS in IP networks. There are already efforts in the industry to define methodology for objectively measuring video quality and to create management tools that can process Video QOS reports.

6. VISUAL COMMUNICATIONS IN EDUCATION

Visual Communication and collaboration enables new levels of interaction among scientists, researchers, and students.

Research is usually distributed across many universities in different locations. Telepresence combined with high definition content sharing is a key technology to connect researchers in high-performance teams. The quality of the shared content is paramount for successful research.

Recent survey of 431 corporate HR directors showed that companies are looking for people with strong collaboration skills. Universities responded by encouraging close collaboration in student teams. Content sharing is a key enabler for student collaboration while immersive telepresence can be used very efficiently for class-to-class meetings. Since classes can be anywhere between 10 and 25 people, it is important that telepresence solutions offer immersive experience for large groups.

University administration is another opportunity for visual communications: dean-to-dean and dean-to-staff virtual meetings can be much more effecting through the use of HD video and audio. Visual communications have a solid place in project management – for IT, academic, and departmental projects.

University staff needs continuous education and most of it can be accomplished through distance learning. Recording, playback, and streaming are key visual communications technologies that enable effective distance learning. For example, the Washington State University in the USA uses these technologies extensively in their Engineering and Management Technology Program.

Many universities are tied with medical centers and hospitals for medical student education. Visual communications open new learning opportunities to medical students – they can for example observe a surgery in crisp HD quality from a remote location and see a close-up of the surgeon's hands through a separate HD camera. In addition, microscopes, CAT scans, or X-rays machines can be connected to the video equipment and their input can be transmitted to the remote site via that content channel.

A new trend in the USA is that university medical centers are gradually transforming into regional health education centers. They invite experts over video and organize lectures about e.g. managing diabetes, managing kidney disease, or nutritional education. In this way, universities broaden their scope outside standard education to get additional funding from public health programs. This trends leads to the convergence of visual applications developed for education and healthcare.

7. TELEPRESENCE

Telepresence – from personal to immersive - is a new kind of visual communication that provides a 'just as being there' experience and brings distributed teams closer together. This technology is being deployed by leading organizations to extend the reach of their higher education programs.

Polycom has just introduced its telepresence solutions to the educational space, and this has created a lot of excitement in the USA. Best suited for telepresence are high-calibre MBA programs with remote campuses. An example is the MBA program at Duke University's School of Business in North Carolina who deployed Polycom RPX telepresence systems for class-to-class immersive collaboration.

Another interesting application for telepresence is job interviews for graduating student placement. Corporations often have long-standing relationships with particular universities, and can deploy telepresence technology to interview prospective employees remotely thus saving the time and cost for travel.

Key characteristics of telepresence are the high definition image quality, the use of life-sized images, and the superb high definition audio. Examples for telepresence solutions from Polycom are in Figure 6.

Figure 6: Telepresence Solutions

In addition, immersive telepresence offers the ability to keep eye contact throughout the meeting and guarantees HD audio quality through microphones and speakers built into the walls of the telepresence room. Best example for such immersive experience is the Polycom Real Presence Experience (PRX) which is a complete room, including walls, monitors, tables, and even chairs, all optimized for supreme audio-visual experience. For example, eye contact is achieved through installing invisible small cameras between the large displays. This configuration allows users to look directly into the eyes of their communication partners, instead of looking to the left, right or bottom, as it is the case with cameras mounted at the periphery of the screen wall.

HD audio quality includes the use of super wide-band audio codecs such as Siren™22 audio codec that transmits up to 22 kHz audio (CD quality) and has very short delay of 40ms for excellent interactivity. HD audio also requires stereo capabilities, so that movement of the immersive telepresence users across the room can be detected acoustically at the remote site. Special acoustic design requires microphones and speakers to be placed in particular parts of the wall and ceiling (through ceiling microphones).

8. CONTENT SHARING

Content sharing is essential to team collaboration and is critical in the educational environment. Dual Video Streams allow a 'presentation' (sometimes also called 'content') audio-video stream to be created in parallel to the primary 'live' audio-video stream. This second stream is used to share any type of content: slides, spreadsheets, X-rays, video clips, etc. Polycom's pre-standard version of this technology is called People+Content. H.239 is heavily based on intellectual property from Polycom People+Content and became the ITU-T standard that allows interoperability between different vendors. Figure 7 explains the concept.

Figure 7: Content Sharing

While the function works well on single-monitor systems, it is especially powerful in multi-screen setups (video endpoints can support up to 4 monitors). In the example in Figure 7, a Polycom HDX 4000 personal video system is on a live call with a Polycom HDX 9000 Executive Collection with two flat screen monitors. The live stream is shown on the right monitor.

The user of the HDX 4000 uses a laptop directly connected to HDX 4000 or running Polycom content sharing software to activate content sharing to the HDX 9000 Executive Collection. A 'presentation' stream is created in parallel to the 'live' stream, and the content is displayed on the left screen of the receiver system.

The benefit of this functionality is that users can share not just slides or spreadsheet but also moving images: Flash video, movie clips, commercials, etc. The 'presentation' channel has flexible resolution, frame rates, and bit rates. For dynamic images, it can support full High-Definition video, and for static content, such as slides, can work for example at 3 frames per second, and save bandwidth in the IP network. Another major benefit of using a video channel for content sharing is that the media is encrypted (by AES in H.323 and by SRTP in SIP). In addition, once the firewall and NAT traversal works for the 'live' stream, it works for the 'presentation' channel as well and there is no need for separate traversal solution.

If still images are transmitted, the content channel can use few kbps; for high quality video, it may require more than 1 Mbps. Future content sharing solutions will include the capability to share HD 1080p and 3D content that will require even higher bit rates.

9. PEOPLE ON CONTENT

People On Content (POC) allows displaying a speaker with different background (content) which is fed into the video system from a DVD player or a computer. Standard application for POC is a professor presenting slides as background. However, the POC technology allows for dynamic backgrounds, and this allows for another application: the professor is presenting and the near-site students are shown in the background – this configuration helps students at remote site feel connected and involved in the lecture.

POC can also be used for school branding – the picture of the school building or campus or a "Welcome to …" background can be applied throughout the presentation.

The ability for the presenter to become part of the presentation is unique feature in Polycom HDX video endpoints, and finds wide acceptance in schools and universities across the USA.

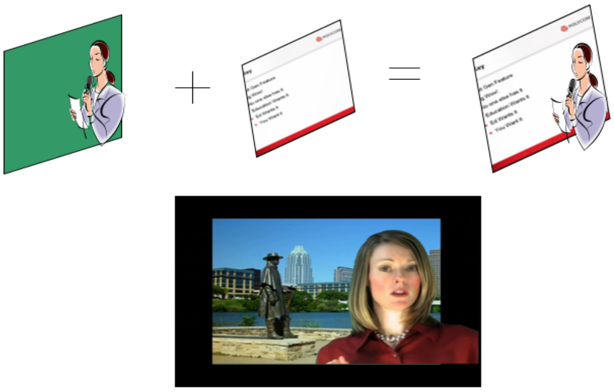

Figure 8: People On Content

The POC technology does not require any special capabilities in the video equipment at the remote (receiving) site because the image of the speaker and the background are digitally merged in the video equipment at the near (sending) side and are sent to the remote site as a single video stream.

10. LIVE MUSIC PERFORMANCES OVER IP NETWORKS

When the Manhattan School of Music (MSM) in New York approached Polycom with the request for equipment that can be used to deliver their live music performances to remote schools across the Unites States, we suggested using our high-end audio and video codecs (endpoints). These systems deliver the highest level of voice and video quality for applications such as video conferencing, telepresence, and vertical applications for the education, healthcare, and government markets.

Very soon we realized that the technologies that guarantee true music reproduction are often in conflict with the technologies developed to improve the user experience in a standard bi-directional voice call. Key technologies are Automatic Gain Control (AGC), Automatic Noise Suppression (ANS), Noise Fill, and Acoustic Echo Cancellation (AEC). In order to allow transmission of live music performances, Polycom made series of modifications and created the so-called Live Music Mode (LMM). Figure 9 depicts the configuration for transmitting live music.

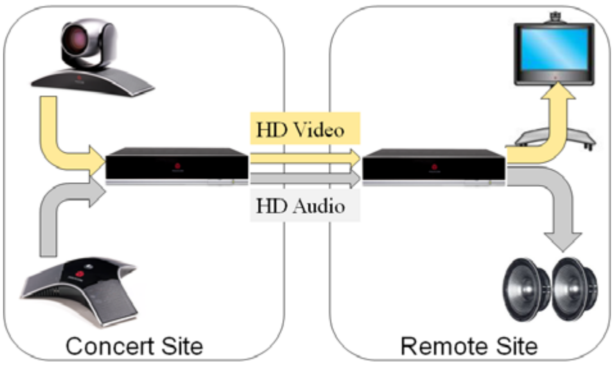

Figure 9: Transmitting Live Music

While secondary to audio, the video component is important for the overall audience experience. Usage of audio-video technology allows music students at remote locations to ask questions, and for the performers at the concert site to answer them, all over a HD video connection, if necessary accompanied by content sharing.

Polycom presented its technology for transmitting live music performances over high-speed IP networks at the TERENA Networking Conference in Bruges, Belgium (May 2008) http://tnc2008.terena.org/programme/presentations/show.php?pres_id=38 and at the BroadcastAsia conference http://www.broadcast-asia.com/BCA_Int_Conference.htm#session10 in Singapore (June 2008).

11. NON-REAL-TIME COMMUNICATIONS

Recording, content management, and streaming applications emerged as a natural complement to real-time communication. While less problematic from a QOS perspective, these applications increase the bandwidth consumption and lead to a requirement in the area of IP Multicast. Figure 10 includes the most common use scenarios.

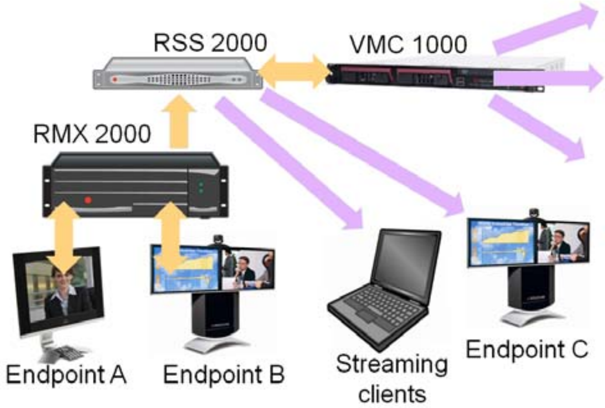

Figure 10: Video Recording, Playback and Streaming

'Non-real-time' refers to the fact that the audio and video information is not immediately sent to the remote site but rather recorded and made available for later consumption. The delay may be just a coupe of seconds (we talk about near-real-time communications) or may be in days or months.

The first step is recording the content. In the example in Figure 10, this is accomplished by the use of Polycom RSS 2000 recorder, which can record and simultaneously stream video over the IP network in Microsoft's Media Player or Real Networks' Real Player formats. Note that the RSS 2000 recorder can record video from a single endpoint or – as shown in Figure 10 – a multipoint call through a conference server such as RMX 2000. As a rule, the stream reaches the streaming clients on PCs within 10-15 seconds. The quality is usually set to standard definition or lower – this can be controlled by the selected streaming bit rate - but new streaming formats allows streaming in HD quality.

Video endpoint such as Endpoint C in Figure 10 can connect to RSS 2000 and playback the recording over a real-time connection. This delivers the highest quality since both endpoints and RSS 2000 support HD quality video and audio.

Once the video is recorded it can be automatically moved to a video content management server (here VMC 1000 from Polycom) that allows organizing the video recordings in programs and channels, searching for content, and scalable streaming through unicast (point-to-point) or multicast (point-to-multipoint) connections.

How can this technology be used in universities? Recording guest lecturer for later viewing comes to mind – tools also allow uploading presentation slides and supporting documents to the server. The recordings can then be streamed to dormitories and homes – this is excellent method for universities that have a lot of online and remote students.

The technology saves time because professors can direct students to particular chapter of the video recording and ask them to only download documents related to this chapter. Chaptering is very important feature in video content management solutions such as VMC 1000.

Another application is 'electronic portfolio'. When students are close to graduation and are looking for jobs, they can record themselves speaking about their goals and achievements, and upload their resumes, sample projects, etc. to the server. Hiring companies can then access the server securely over the Internet, and use this information when evaluating candidates.

Finally, university faculty can use the technology for continuous training – existing tools allow tracking who watched which video content.

12. THE ROLE OF HIGH-SPEED NETWORKS

With regards to network requirements, latency, packet loss, and jitter are important as for any form of real-time communication over IP, e.g. Voice over IP. Specific to visual communications are the high bandwidth requirements.

The emergence of high-speed IP networks such as Internet2 in the USA and GEANT2 in Europe are creating opportunities for new kinds of real-time applications in universities and schools around the world. Bandwidth is no longer limiting, but rather driving new applications that were impossible even a few years ago. The most profound changes are in the area of bandwidth-hungry visual communication applications. Figure 11 summarizes the bandwidth requirements for video. The vertical axle of the diagram in figure shows the video quality while the horizontal axle shows the necessary bandwidth to transmit video at this quality across the network.

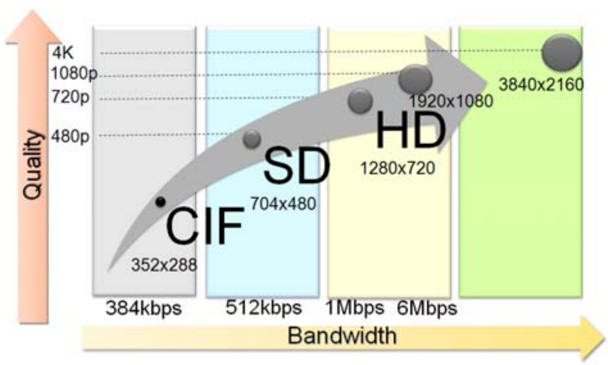

Figure 11: Bandwidth Requirements for IP Video

The Common Intermediate Format (CIF) has resolution of 352x288 pixels. Current video technology allows CIF quality video to be transmitted at 30 fps over 384 kbps. Standard Definition (SD) is the quality of a regular TV, and has resolution of 704x480 pixels. Current video technology allows transmitting SD quality video at 30 fps over 512 kbps-1 Mbps.

The latest generation of telepresence products support High Definition (HD), starting with 720p HD, i.e. 1280 x 720 pixels with progressive scan. Current video codec technology allows transmitting HD quality video at 30 fps over 1-2 Mbps. Immersive telepresence requires multiple codecs and uses more bandwidth in the IP network. Polycom demonstrated HD 1080p at the InfoComm trade show in Las Vegas (June 2008). A 1080p connection at 30 fps requires about 3Mbps bandwidth in the IP network.

Audio – even at the highest quality of 22 kHz stereo – does not require more than 128 kbps and is therefore comparatively light load on the IP network.

13. LOST PACKET RECOVERY (LPR)

Packet loss is a common problem in IP networks because IP routers and switches drop packets when links get congested and when their buffers overflow. Real-time streams, such as voice and video, are sent over the User Datagram Protocol (UDP) which does not retransmit packets. Even if an UDP/IP packet gets lost, retransmission does not make much sense, since the retransmitted packet would arrive too late; playing it will destroy the real-time experience.

Lost Packet Recovery (LPR) is a new method of video error concealment for packet based networks, and is based upon Forward Error Correction (FEC). Additional packets that contain recovery information are sent along with the original packets in order to reconstruct packets that were lost during transmission.

For example, suppose you have 2% packet loss at 4 Mbps (losing approx. 8 packets for every 406 packets in a second). After engaging LPR, the packet loss can be reduced to e.g. less than 1 packet /5 minutes = 1 packet/(406 packets/sec*300 sec) = 1 packet/121,800 packets = .00082%. Figure 12 depicts the functions in LPR – both sender side (top) and receiver side (bottom).

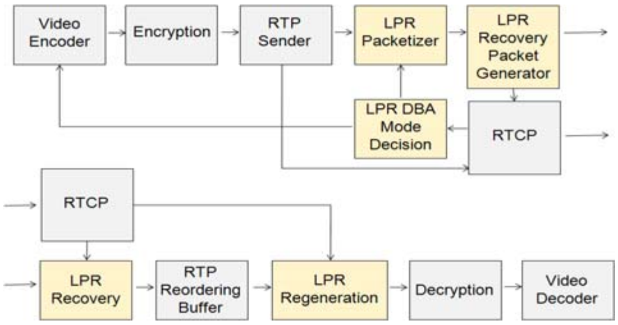

Figure 12: Lost Packet Recovery

LPR includes a network model which estimates the amount of bandwidth the network can currently carry. This network model is driven from the received packet and lost packet statistics. From the model, the system determines the optimal bandwidth and the strength of the FEC protection that is required to protect the media flow from irrecoverable packet loss. This information is fed back to the transmitter via the signalling channel, which then adjusts its bit rate and FEC protection level to match the measured capacity and quality of the communication channel. The algorithm is designed to adapt extremely quickly to changing network conditions – such as cross congestion from competing network traffic over a Wide Area Network (WAN).

The LPR encoder takes RTP packets from the RTP Tx channel, encapsulates them into LPR data packets and inserts the appropriate number of LPR recovery packets. The encoder is configured by the signalling connection, usually when the remote LPR decoder signals that it needs a different level of protection.

The LPR decoder takes incoming LPR packets from the RTP Rx channel. It reconstructs any missing LPR data packets from the received data packets and recovery packets. It then removes the LPR encapsulation (thereby converting them into the original RTP packets) and gives them back so they can be processed and forwarded onto the video decoder. The decoder has been optimized for compute and latency.

LPR has advanced dynamic bandwidth allocation capabilities. Figure 13 illustrates the behaviour

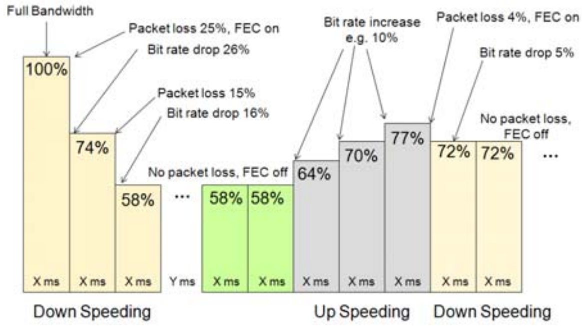

Figure 13: LPR DBA Example

When packet loss is detected, the bit rate is initially dropped by approximately the same percentage as the packet loss rate. At the same time, we turn on FEC and begin inserting recovery packets. This two-pronged approach provides the fastest restoration of the media streams that loss creates, ensuring that there is little or no loss perceived by the user. The behaviour can be modified by the system administrator through configuration.

When the system determines that the network is no longer congested the system reduces the amount of protection, and ultimately goes back to no protection. If the system determines that the network congestion has lessened, it will also increase the bandwidth back towards the original call rate (Up-speed). This allows the system to deliver media at the fastest rate that the network can safely carry. Of course, the system will not increase the bandwidth above the limits allowed by the system administrator. The up-speed is gradual, and the media is protected by recovery packets when the up-speed occurs. This ensures that it no impact on the user experience. In the example in Figure 13, up-speeds are in steps of 10% of the previous bit rate.

If the packet loss does not disappear, the system continues to monitor the network, and finds the protection amount and bandwidth that delivers the best user experience possible given the network condition.

A recent evaluation of Polycom LPR by the independent analysts from Wainhouse Research http://www.wainhouse.com/ concluded: 'While most of the video systems on the market today include some form of packet loss / error concealment capability, Polycom LPR is one of only two error protection schemes available today that uses forward error correction (FEC) to recover lost data. One of LPR' differentiators and strengths is that it protects all parts of the video call, including the audio, video, and content / H.239 channels, from packet loss.'

14. APPLICATION AWARE IP NETWORKS

Application-aware IP networks detect the type of application running on the network and provide the quality of service required for this application to function properly. This is very important for real-time applications and especially for bandwidth hungry video applications. The IP network can detect the type of application by analyzing the traffic patters and the used ports or the application itself can notify the policy engine of the IP network about its requirements. Figure 14 summarizes the visual application functions, the IP network functions and the interaction between the two.

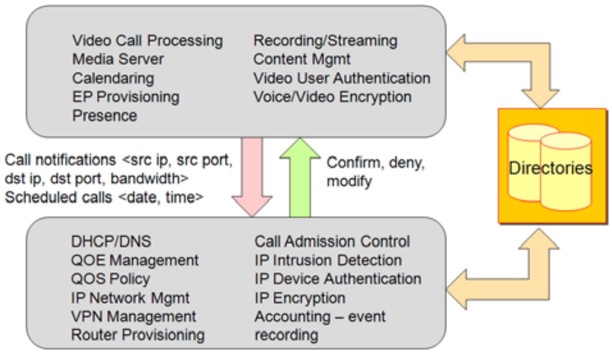

Figure 14: Interaction between Visual Apps and IP Network

Application-aware IP networks are necessary for proper support of visual communication applications. The work in this area is ongoing and the presentation at the EUNIS conference will provide status of the current developments.

CONCLUSION

High-speed IP networks such as Internet2 in the USA and GEANT2 in Europe allow new kinds of visual communication applications at universities and schools.

The abundance of bandwidth allows HD video and audio deployments for advanced collaboration, telepresence, content sharing, transmission of live music performances, etc.

Polycom is driving these market and technology developments and is promoting the use of video in educational environments.