Enhancing the Quality of Low Bit-Rate Real-Time Internet Communication Services

Schubert Foo, Siu Cheung Hui, See Wai Yip
School of Applied Science, Nanyang Technological University, Singapore

Abstract

The Internet environment, with its packet-switched network and lack of resource reservation mechanisms, has made the delivery of low bit-rate real-time communication services, such as audio and video, particularly difficult and challenging. The high potential transmission delay and data packet loss under varying and uncontrollable network conditions will almost certainly lead to unpleasant and unintelligible audio and jerky video play-out.

The Internet TCP/IP protocol suite can be extended with new mechanisms in an attempt to tackle such problems. In his research, an integrated transmission mechanism that incorporates a number of existing techniques to enhance the quality and deliver ‘acceptable’ real-time services is proposed. These techniques include the use of data compression, data buffering, dynamic rate control, packet lost replacement, silence deletion and virtual video play-out mechanism. The proposed transmission mechanism is designed as a generic audio and video communication system providing proficiency for each different type of service so that it can be used in different systems and conditions. This approach has been successfully implemented and demonstrated using three separate systems that include the Internet Phone, Web Video and Video-Conferencing Tool.

Introduction

Internet technology has evolved to an extent in enabling low-cost real-time communication services to become a reality. Internet has been successfully used as a transmission medium to exchange real-time multimedia information such as audio and video data between users across the world. This is evidenced by the numerous interpersonal communication systems that have sprung up in the last two years, spanning from the humble but yet extremely useful audio-only Internet telephony systems to the more complex video-conferencing systems. Some of the existing Internet telephony and video-conferencing systems include Internet Phone (VocalTec Communications Ltd., 1999), NetMeeting (Microsoft Corporation, 1999), RealPlayer (RealNetworks Inc., 1999), WebMedia (University of Ulm, 1999), CU-SeeMe (Cornell University, 1999).
IRIS Phone (IRIS Systems, 1999), WebPhone (Netspeak Corporation, 1999) and VDOPhone (VDOnet Corporation, 1999).

However, the Internet presents a relatively harsh environment for real-time services especially for those that require large bandwidths such as video data. The poor Quality of Service (QoS) that is encountered in most existing systems has largely prevented them from gaining widespread acceptance. The potential high transmission delay and data packet loss of the Internet environment that is characteristic of a packet switched network without resource reservation mechanisms, have made real-time communications services difficult. Under varying and uncontrollable network load conditions, the data packets will inevitably suffer varying degrees of delay between senders and recipients. The variance in delay produces unpleasant and undesirable play-out for both audio and video stream for real-time services. Furthermore, the unreliability of the network can additionally give rise to data packet loss that will deteriorate the quality of real-time communication services even further.

To tackle this problem, the Transmission Control Protocol/Internet Protocol (TCP/IP) (Comer, 1995) suite of Internet protocol can be extended with new mechanisms to deliver acceptable real-time Internet services. However, such an approach will at best simulate real-time but does not guarantee real-time delivery due to the underlying nature of existing protocol.

Many of the proposed techniques are not new. For example, buffering mechanisms have been used to adjust the play-out time of arriving data packets at the recipient’s machine to minimise the impact of delay jitters, while adaptive rate control mechanisms have been used to eliminate or minimise the impact of network data packet loss (Bolot and Turletti, 1994; Bolot and Turletti 1996; Kanakia et al., 1993).

However, the use of individual techniques in isolation is inadequate to tackle the harsh reality of the network conditions. With a congested network at different times of the day at different parts of the world, occasional packet losses can result in intolerable delays when using only the buffering mechanism. Similarly, the use of the dynamic rate control mechanism by itself can result in the deterioration of the transmission rate to the point that results in unintelligible audio and very jerky video play-out due to the low transmission rates. The insufficiency of individual mechanisms in isolation has prompted the necessity of investigating a new approach to integrate the various mechanisms and enhance the quality of real-time communication services.

Therefore, this research proposes an integrated transmission mechanism that integrates a number of techniques to enhance the quality of real-time communication services. It is also desirable to realise these techniques even in low bit-rate situations. Even if it is possible to send live video data across the Internet, it is not cost-effective and not the best means of utilising precious bandwidth. For example, in the situation of a typical video-conferencing application, the video frames are essentially a series of ‘head and shoulder’ images with little changes in-between frames so that sending 25 such frames per second (of say 250 x 175 pixel resolution per frame) implies 8.75 Mbps of information. Thus,
the main objective of this integrated mechanism should be to deliver ‘acceptable’ low bit-rate real-time services so that it could be utilised by the masses on Internet.

The techniques used in this research include the use of data compression, data buffering, dynamic rate control, packet loss replacement, silence deletion, and virtual video play-out mechanism. The buffering mechanism is used to minimise delay jitters. Dynamic rate control is used to eliminate the impact of audio/video data packet loss. Packet lost replacement is used to simulate the lost packets through the use of redundant data in subsequent audio packets (for audio stream) or through information interpolation (for video stream). Silence deletion is used for audio to eliminate the transmission of silent audio packets, thereby decreasing unnecessary bandwidth usage. Finally, in cases when no video frames are transmitted during a congested network condition, a virtual play-out mechanism can be used to play out past video frames instead of freezing the play-out. The use of virtual play-out attempts to give the user an impression of continuous video transmission, thereby achieving a more ‘natural’ session while minimising the effect of jerky playback. This is based on the assumption that there are generally little changes in a series of captured video frames as the objects involved in the scene do not move about very often and that there are no sudden change in scenes. Obviously, if constant and dramatic changes in scenes are anticipated, virtual play out can be turned off to avoid providing misleading information to the users.

The proposed transmission mechanism is designed as a generic audio and video communication system providing proficiency for each different type of service so that it can be used in different systems such as the Internet Phone (audio only), Web-Video (video only), Internet Video-Conferencing (audio and video), Picture Phone (audio and image), and so on.

Internet Protocol for Data Delivery

A number of protocols are available for delivery of real-time data over the Internet. The TCP/IP protocol suite comprises two protocols, namely, Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). Additionally, the Real Time Protocol (RTP) (Schuizrinne et al., 1996) that is based on the TCP/IP protocol suite, has been introduced to facilitate multiparty conferencing applications. These protocols are introduced and contrasted for their suitability in real-time services:

Transmission Control Protocol (TCP)

The TCP verifies that the data is delivered in order and without corruption. Associated with this feature is the extra overhead needed in the generation and maintenance of a connection. TCP provides for the transmission of reliable, connection-oriented stream of bytes. TCP’s reliability comes from its inclusion of checksum into each packet of data transmitted. On reception, a checksum is generated and compared to the checksum included in the header of data packet. If the checksum does not match, the receiver communicates that fact to the sender, and data is automatically resent. TCP is considered connection-oriented because the two end-points of communications exchange a
handshaking dialogue before data transmission can begin. This handshaking guarantees the sender that the receiver is alive and ready to accept data.

**User Datagram Protocol (UDP)**

The UDP allows data to be transferred over the network with a minimum of overhead. UDP’s overhead is low because it can only provide unreliable data delivery. There is no method in the protocol to verify that the data has reached the destination exactly as it was sent. The data may be lost, duplicated, or arrived out of order. However, this limitation does not make UDP useless. In fact, the low overhead in UDP transmission and the lack of reliability can make UDP efficient in transmitting real-time data across Internet!

**Real-time Protocol**

The Real-Time Transport Protocol (RTP) uses the multicasting backbone network to facilitate multiparty conferencing. It provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. These services include payload type identification, sequence numbering, time-stamping and delivery monitoring. Applications typically run RTP on top of UDP to make use of its multiplexing and checksum services.

RTP itself does not provide any mechanism to ensure timely delivery or provide other QoS guarantees, but relies on lower-layer services to do so. It does not guarantee delivery or prevent out-of-order delivery, nor does it assume that the underlying network is reliable and delivers packets in sequence. The sequence numbers included in RTP allow the receiver to reconstruct the sender’s packet sequence. RTP consists of two closely-linked parts:

- Real-Time Transport Protocol (RTP) specifies how the audio-video data is packetised. This protocol can be used to transport different type of real-time media in some standard formats; and
- RTP Control Protocol (RTCP) that monitors the quality of service and to convey information about the participants in an on-going session. The latter aspect of RTCP may be sufficient for “loosely controlled” sessions (i.e. where there is no explicit membership control and set-up).

RTP is intended to be malleable to provide the information required by a particular application and will often be integrated into the application processing rather than being implemented as a separate layer. RTP is a protocol framework that is deliberately not complete. Unlike conventional protocols in which additional functions might be accommodated by making the protocol more general or by adding an optional mechanism that would require parsing, RTP is intended to be tailored through modifications and/or additions to the headers as necessary.
Choice of Protocol

The various real-time communication services require real-time, continuous media service to transfer audio and video data between senders and recipients. This implies that TCP and other reliable transport protocols are inappropriate due to the following reasons:

- It cannot support multicasting (e.g. audio conferencing);
- Even for unicast, reliable transmission is inappropriate for delay-sensitive data such as real-time audio. By the time the sender has discovered that the receiver is missing a packet and retransmitted it, at least one round-trip time (time between sending data from sender to recipient and back to sender) or more has elapsed. The receiver has to wait for the retransmission and this increases delay and incurs an audible gap in playback. Retransmitted packets cannot be discarded since standard TCP implementation forces the receiver application to wait, so that the packet losses would always produce increased delay.
- The TCP congestion control mechanisms decrease the congestion window when packet losses are detected. On the other hand, audio has ‘natural’ rates that cannot be suddenly decreased without starving the receiver.

Therefore, UDP is the preferred protocol for transmitting real-time data and accordingly adopted in this research. However in non-real-time audio transmission (such as a voice mail system), TCP may be more appropriate since it requires less programming effort and at the same time, guarantees the delivery of every data packet. Nonetheless, with sufficiently long buffering and adequate average network throughput, near-real-time delivery using TCP can be successful as demonstrated in Netscape’s Web browser.

In addition to UDP, RTP is used in conjunction to provide the required packetisation of the audio and video data while RTCP is used to exchange network information that is necessary for dynamic QoS control.

System Architecture

Figure 1 shows the basic system architecture that consists of the Transmitter Module and Receiver Module. The Transmitter Module is responsible for data acquisition and preparation of data for transmission over the network to the recipient. Dynamic rate control is used to adapt to changing network condition to better utilise the network. The Receiver Module is responsible for managing the received data packets. This involves the use of buffering and packet lost replacement mechanisms to improve the play-out performance.
System Components

The Transmitter Module is made up of three main components, namely, Data Acquisition and Compression, RTP Packets Generation, and Dynamic Rate Control.

Data Acquisition and Compression

Video and audio data are acquired through their respective capturing devices and compressed before transmission to their destination. Compression is necessary for efficient transmission of data packets and to reduce the required bandwidth.

The size of video frames can be greatly reduced with the use of compression methods such as discrete cosine transform (DCT) used in JPEG (Pennebaker and Mitchell, 1993) and motion compensation algorithm used in MPEG (ISO Standard 11172-1, 1993). A compression ratio of 1:15 is usually achievable with JPEG and MPEG compression. As MPEG compression is computationally intensive and real-time MPEG video compression requires the use of specialised MPEG encoder card to achieve acceptable compression rate of at least 15 frames per second (fps), the JPEG video codec that provides high compression ratio is chosen for implementation. Furthermore, JPEG compression is less computational intensive than MPEG. Transmitting MPEG streams using UDP can be complicated as the lost of a principal frame in the MPEG stream can render the uselessness of other intermediate frames due to their inter-dependency. The format used for the JPEG image is JFIF (JPEG File Interchange Format) (Pennebaker and Mitchell, 1993).
that has been widely accepted and is one of the recognised codecs specified in the H.323 standard (ITU, 1996). Different image compression qualities and compression ratio can be achieved by varying the JPEG compression parameters. This is a useful feature when dynamic adjustment of compression quality and ratio is required to adapt to different network conditions.

Audio data are captured through the microphone connected to the sound card and compressed using various audio codecs. The compression algorithms reduce the audio sample size in raw audio PCM format by encoding it in another format. Common compression algorithms include A-law, µ-law (CCITT, 1984), adaptive differential pulse code modulation (ADPCM) (I-ETS, 1992), Groupe Speciale Mobile (GSM) (I-ETS, 1992) and linear predictive coding (LPC) (Campbell and Tremain, 1986). The first four formats are able to maintain almost the same quality as the raw format. However, as LPC produces a lower audio quality at a lower information rate, it can be conveniently used to compress previous audio packets into ‘redundant’ audio information and bundled together with the current sample to form an audio data packet. When the need arises, the redundant information is used for packet lost replacement to replace previously lost data (see the subsequent section). Decompression performed at the recipient’s end reverts the compressed audio samples (both redundant and non-redundant audio) to its uncompressed format that is ready to be played.

Additionally, the audio data acquisition supports silence deletion to eliminate, and thus prevent useless silent audio packets from being transmitted to the Receiver Module. This is an essential mechanism to improve the QoS by reducing significant bandwidth since audio communication is often carried out in half duplex mode where only one party is speaking while the other is listening and remaining silent. Silent deletion is achieved through the two operations of threshold calculation (to obtain a threshold value of background noise) and application of an appropriate silent deletion algorithm. In this work, the HAM algorithm (Claypool and Riedl, 1994) has been chosen for its superior performance and simplicity among existing silent deletion algorithms.

**RTP Packets Generation**

Each audio and video data packet is linearly time-stamped to indicate the instant of sampling. Although the transmission is carried out in sequential order, the packets may arrive out of order due to the different network paths traversed by the data packets and varying network delays. The time-stamp allows the packets arriving at the recipient’s host computer to be ordered in the correct time sequence before being played out.

The Real-time Protocol (RTP) format of the H.323 standard is used for real-time data packetisation. The time-stamp is stored in the time-stamp field of the RTP header of each video packet. Since each video frame is discrete, a single frame should not be packetised into multiple packets, as the original video frame will not be recoverable once a packet is lost. This is especially true when the Internet is used as the transmission medium. Hence, a video packet should encapsulate one complete video frame before transmission.
Audio data are packetised in intervals of 20ms so that the time-stamp mechanism uses 20 as the base value with increments of 20 for each new audio packet generated. The time-stamp is stored in the field preceding the actual and redundant audio data to form an audio packet that is transmitted across the network.

**Dynamic Rate Control**

Dynamic rate control obtains the necessary information regarding the network conditions in an attempt to adjust control parameters (such as sampling rates, choice of compression algorithms, compression ratios, and so on) and provide an optimum level of performance based on existing network conditions. In order to do this, the network condition must be analysed periodically. This is accomplished by a continuous feedback loop between the *Transmitter Module* and the *Receiver Module*.

Feedback information is passed to the RTCP (Real-Time Control Protocol) Packets Generation for the generation of sender reports from the *Transmitter Module* and receiver reports from the *Receiver Module*. As defined in the RTP, these reports are generated periodically and processed by the *Transmitter Module*. Busse’s algorithm (Busse et al., 1996) is adopted to use the feedback information given in these reports. Information from these reports enables the computation of packet loss rate, packet delay jitters and round-trip delay time. With this, the network congestion state can be determined. This can be classified as unloaded, loaded or congested. From this, the *Transmitter Module* makes the decision of increasing, holding or decreasing the bandwidth.

For example, if the network condition is congested, a higher compression ratio with lower quality JPEG compression can be selected to compress video images in order to reduce the bandwidth used. Lower quality audio compression can also be used under such circumstances. If audio transmission has been assigned top priority and the network condition is not improving, the system may decide to drop video transmission altogether and use virtual video play-out instead. Thus, dynamic rate control provides the necessary information and mechanism to adopt and optimise bandwidth usage.

The *Receiver Module* is made up of three main components, namely, *Buffer Management*, *Packet Lost Replacement* and *Play-Out Controller*.

**Buffer Management**

The purpose of buffer management is to cushion the out-of-order, late delivery and jitters experienced by the data packets. Incoming audio and video data packets are ordered and arranged in the correct order according to the time-stamp and stored using a ring buffer.

Information pertaining to each individual video frame such as time-stamp is also stored to facilitate the play-out process. The buffers used for the implementation of the buffering mechanism can either reside in main memory or on disk. Memory buffers allow fast data accessing for efficient video play-out. However, each video frame takes up considerable amount of space for storage so that there is a limit to the number of video frames that can
be feasibly stored in memory buffers. In contrast, the storage capacity of file buffers is only limited to the available disk space. Although accessing to file buffers will be slower as compared to memory buffers, the access time for modern day hard disk is sufficient for efficient video play-out. Therefore, file buffers are chosen for implementing the video buffering mechanism. In addition, a double buffering mechanism is used to control the use of buffers efficiently as well as allow uninterrupted supply of video frames during virtual play-out.

![Audio Buffer Management and Packet Lost Replacement](image)

**Figure 2. Audio Buffer Management and Packet Lost Replacement**

As the size of audio data is significantly less than video data, the audio buffer can be implemented using memory buffers directly. In this instance, the audio data is organised into slots of fixed sizes. Each slot size is equivalent to an uncompressed audio sample size as shown in Figure 2. Incoming audio packets are decompressed, identified by its associated time-stamp and placed accordingly into the correct location in the buffer. The total buffer space allocated is dependent on the amount of jitters experienced by the packets. Although the amount of jitters will vary with time, it is a complex process to dynamically keep changing buffer size during an on-going session to cater for these variations. As such, a margin of safety of 2 has been incorporated so that the total buffer space is set to be twice the number of packets that can fit within an average jitter period. With the amount of space allocated for buffering, the play-out point occurs when the ring buffer is half full. This margin of safety cushions the variations of delay exceeding the jitter period used in the calculation of buffer space.

**Packet Lost Replacement**

Some form of packet lost replacement mechanism is necessary to cater for any underlying unreliable network transmission that results in data packets lost or data packets delay beyond a tolerable limit. This mechanism is activated by the *Play-Out Controller* when missing data packets are detected. Redundant information transmitted
in audio packets is used for audio packet lost replacement. Dynamic video frames reconstruction is used to cater for video packet lost replacement as well as low video transmission rates in order to provide smooth transmission between frames.

**Audio Redundancy.** Hardman’s (Hardman et al., 1995) approach of audio redundancy has been used for audio packet lost replacement. In this approach, each voice segment is encoded into two packets so that in the event of a packet lost, a duplicated encoding in the following packet can be played out. In order to reduce overheads due to duplicate voice encoding, the first packet of a voice segment uses toll-grade compression, whereas the duplicated encoding of the same voice segment uses a simpler form of encoding to reduce the cost in both processing power and bandwidth. Hence, this implementation produces toll-grade audio quality inter-mix with sub-standard audio quality. This redundant information is combined to form an audio packet. In the event that a packet that is scheduled to be played is not present in the buffer, it will replace this lost packet with the redundant audio information encoded in the next packet. However, if consecutive packets are lost, no redundant audio data can be used to replace the missing links. When this happens, this period of time will be replaced with silence at the receiver’s output. This situation can also arise when silent deletion from the Transmitter Module stops silent audio packets from being sent to Receiver Module.

**Dynamic Video Frames Reconstruction.** Dynamic video frames reconstruction is a lightweight mechanism for reconstructing intermediate frames to smoothen out jerky video play-out due to low video transmission rate and data packet lost during the video transmission process. Different algorithms can be used to generate intermediate frames based purely on two successive frames. In this work, the interpolated transparency algorithm (Foley et al., 1996) is adopted due to its simplicity in implementation and its capability in providing satisfactory performance to produce the necessary effects. Conventionally, transparency algorithms have been used for producing fading effects in computer graphics. This fading effect produces a smooth transition between two different scenes by employing a linear interpolation of pixel values between two frames.

**Play-Out Controller**

The play-out controller determines the play-out scheme to be used during audio and video play-out.

In audio play-out, a check is performed to verify if any missing audio packets are detected in the audio buffer, in which case, packet lost replacement is activated to replace missing data packets with corresponding lower-grade audio information or silence.

In video play-out, two types of play-out schemes, namely, real-time play-out and virtual play-out, are supported based on existing network conditions as shown in Figure 3.

In real-time play-out (Figure 3a), video frames will be played out immediately once they are received. In addition, the video frames will also be stored in the buffer to be ready for use in virtual play-out. Real-time play-out is only feasible when the transmission rate is
high and the transmission latency is within acceptable range. In this work, real-time play-out is used whenever the play-out controller detects a reception rate of more than 1 fps. In conjunction with the video frames reconstruction mechanism, acceptable video play-out is possible from such low transmission rate.

![Diagram of video play-out sequences](image)

Figure 3. Video Play-Out Sequences

When no video frames are received (Figure 3b), the play-out controller will switch from real-time play-out to virtual play-out. In this case, video frames previously stored in the buffer will be played out one at a time. Intermediate frames will also be constructed and played out with the original frames. After playing out each video frame, the system will check for the availability of any newly received video frames. If new video frames are received, real-time play-out will be used. If not, virtual play-out will continue (Figure 3c).

By using virtual play-out, the user at the Receiver Module will still perceive a real-time play-out even when no video frames are actually received from the Transmitter Module. The number of video frames available for play-out will be based on the number of video frames previously stored in the buffer and its buffer size.
Video frames stored in the buffer can be played out in either the forward or reverse direction. Figure 4 shows two play-out modes are possible during virtual play-out, namely, unidirectional and bi-directional play-out modes. In the unidirectional play-out mode, video frames in the buffer are played out from the first frame till the last frame with the process repeating itself (Figure 4a). In the bi-directional play-out mode, video frames are played out in the reverse direction after the last frame has been played back (Figure 4b). The bi-directional play-out mode is preferred and used since it gives a better and more natural visual effect and smoother transition of video play-out from the last video frame to the next frame in the play-out sequence.

**Performance Evaluation**

In this section, the audio and video capability of the system is evaluated. The performance of the audio capability is compared with four other Internet telephony systems which include Internet Phone (VocalTec Communications Ltd., 1999), NetMeeting (Microsoft Corporation, 1999), CoolTalk (Netscape Communications, 1999) and Speak Freely (Walker, 1999). These Internet telephony systems have been selected on the basis of common system requirements and their availability on the Internet. All are Microsoft Windows applications that support full-duplex communication.

In this experiment, a framework (Foo and Hui, 1998) which combines a feature and functionality appraisal together with a set of quantitative and qualitative tests is used for the performance evaluation. Two Pentium 200 MHz PCs are placed at two different remote locations to act as caller and recipient stations. A set of microphone and speakers
are connected to the sound card. The Sound Blaster sound card is to be plugged into the PC driven by its full-duplex software driver. In addition, 6 evaluators have been utilised in conducting the tests. Although this can potentially give rise to less reliable results, an analysis of the test statistics confirms that they are acceptable and able to distinguish among systems tested.

The results of the tests showed that Internet Phone came out tops, followed by our system, CoolTalk, NetMeeting and Speak Freely. The performance of the Internet Phone is slightly better than our system. As Internet Phone is a commercial product, we are unable to gather any technical information about the techniques used in the development of the product. One possible reason is the proprietary compression technique adopted in Internet Phone that outperforms the ADPCM compression used in our system. The performance of our system is better than the other Internet telephony systems tested.

In addition, experiments are also conducted to measure the video performance of the system. The measurements were designed to gauge whether dynamic video reconstruction can sustain a high video play-out rate so as to minimise the jerky effects. The environment used for performance analysis comprises two Pentium 200 MHz PCs connected over a LAN (local area network). A video sequence consisting of 225 frames captured at a resolution of 160x120 using Indeo compression technique (Bunzel and Morris, 1994) is used as the video stream for testing. The use of a pre-captured video sequence enables various transmission rates and packet loss rates to be simulated. These experiments showed that our system was able to achieve a relatively constant play-out rate at around 23 fps (frames per second) even under packet loss conditions.

**Applications of the Underlying System**

The underlying system has been successfully developed at the School of Applied Science, Nanyang Technological University, and applied to a number of applications in a range of computing platforms. As the system is generic, it can be easily adopted and used as a stand-alone audio system, stand-alone video system, or a combination of both systems.

**Internet Phone**

The stand-alone audio system is basically an Internet telephony system such as the Internet Telephone Software System (Foo et al., 1996; Chin et al., 1997). The system allows real-time full duplex voice communication to take place between two parties through the Internet using either a pair of PC or UNIX computers with audio capturing devices.
A stand-alone video-only application, WebVideo (Hui et al., 1998) has been developed to deliver live video over the World Wide Web. The system consists of two main components, namely, WebVideo server and WebVideo client. The WebVideo server is basically the Transmitter Module for video capturing, packetisation and video data transmission to the WebVideo client over the Internet. The WebVideo client is basically the Receiver Module that runs on a Web browser. The system is currently being used to remotely monitor the activities in the Object Technology Laboratory of the School of Applied Science from the laboratory manager’s room as shown in Figure 5. The monitoring can obviously take place in any part of the world since it only requires a Web browser and the Uniform Resource Locator (URL) of the Web Video server.

**Video-Conferencing Tool**

The combined audio and video system results in a video-conferencing system that can be tailored to operate at different modes according to the nature of the application. Such a concept has been utilised in the delivery of a Web-based intelligent help desk support for a multinational company in Singapore (Hui et al., 1997; Liu et al., 1998).
In addition to using the facilities provided by the Web interface, customers can communicate with service engineers in the company through Internet video-conferencing as shown in Figure 6. Not only does such a system exhibit time and cost benefits, it enhances the overall customer support process by allowing users to select different operating modes by defining the audio and video priority under different conditions:

**High video and audio priority.** With the presence of adequate bandwidth, some of the mechanism designed for an unreliable network such as dynamic rate control, dynamic video frames reconstruction and virtual video playback becomes unnecessary and can therefore be de-activated. However, audio and video compression is still used to enable effective management of data packets by the underlying network layer.

**High video and low audio priority.** When good quality video is desired, lower quality audio quality compression can be used or even terminated to allow for better video transmission. This is useful for sending a sequence of video frames to the service engineers for their evaluation. This mode of operation has been utilised to remotely monitor the action of faulty machines at the customer’s site.

In the case where a static but larger and higher-resolution image is desired, the video-conferencing mode can be switched to the Picture Phone mode so that only an enhanced single frame (of different size) is captured and transmitted as shown in Figure 7. Such a facility will allow images of critical components to be zoomed in and out during the
diagnostic process. In this instance, TCP can be used for the transmission since the additional delay will not incur any inaccuracy to the communication process.

![Figure 7. Picture Phone Interface](image)

**Low video and high audio priority.** In this mode, emphasis is placed on the audio communication aspect so that video play-out is used to provide a better communication session. Since audio takes a higher priority, dynamic video frames reconstruction is employed to allow for low video transmission rate with minimal jerky play-out. Virtual video play-out is used to simulate video presence when no video frames are transmitted. Such a facility is useful for service engineers to listen to a sequence of sound that is generated by a faulty machine at the customer’s site. If necessary, this sequence of sound can be recorded beforehand and transmitted via TCP. A sound wave analyzer can be used (not shown) to observe the wave pattern of the transmitted signals. This allows important information, such as threshold sound values and cycle times to be determined easily at the service engineer’s end.

**Conclusion**

In this paper, a number of techniques for enhancing low-bit rate real-time communication services over the Internet have been described and integrated together to form the basis of delivering such services. These include data compression, dynamic rate control, data buffer management, data packet lost replacement, audio silence deletion, and dynamic video frames reconstruction to support virtual video play-out.
By using appropriate Internet protocols and the proposed system architecture of the Transmitter and Receiver Modules, these mechanisms have been successfully integrated into three systems and utilised to deliver acceptable audio and video QoS. These systems, that include the Internet Phone, Web Video and Video Conferencing Tool, adequately demonstrate the feasibility of using such an integrated approach to deliver a cost-effective real-time Internet communication solution that utilises limited bandwidth.

Currently, the work is being extended to support an Internet multimedia gateway (IMG) for real-time fax, voice and video communications over the Internet. IMG aims to integrate the Internet with traditional Public Switched Telephone Network (PSTN) to support voice, fax and video communications. IMG comprises three separate gateways, namely, Fax Gateway, Telephony Gateway and Video Gateway. Fax Gateway supports Internet faxing between conventional fax machines and a mail client or Web browser. Telephony Gateway supports real-time voice communication between conventional telephones and Internet telephones. Video Gateway supports multi-party video conferencing over the Internet.

References


I-ETS: European Telecommunications Standards Institute (1992), European digital cellular telecommunications system (Phase 1); Full-rate speech transcoding (GSM 06.10), I-ETS 300 036, February.


VDOnet Corporation (1999), VDOPhone, online document available at <URL: http://www.vdo.net>.

University of Ulm (1999), *Web Media*, online document available at <URL: http://www-vs.informatik.uni-ulm.de/soft/wv/>.