Abstract — This paper presents a JPEG2000 based real-time scalable video communication system developed in Sony, named as “BEAM” video system. Two practical objectives are emphasized in this system: scalable video communication and real-time communication. We extend IETF RTP and RTSP streaming protocol to take care of scalable delivery of video data from a single layered coding data to heterogeneous devices and different resolution display such as HDTV, standard TV, PDA and mobile phone. Utilizing an advantage of low coding delay JPEG2000 codec, real-time ARQ (RT-ARQ) is proposed to resolve the QoS issues for real-time applications. In addition to this protocol, various network adaptive control techniques, including rate control and error control, are presented to achieve high quality video transmission.

Keywords— JPEG2000, Real-time communication, Scalable delivery, Rate control, Error Control, FEC, ARQ, Transport Protocol

I. Introduction

With the development of broadband network, demands for video communication are increasing as one of the most important Internet applications. However, the current IP-based network provides only a single class best effort service. Video packets, that are not delivered due to bad network condition and/or due to the exceeding maximum delay threshold, can be regarded as packet losses by a video decoder. Many researches have been done how to guarantee the end-to-end video transmission quality. From the video coding point, scalable codec is proposed to cope with the fluctuation of the available network bandwidth [1, 2]. From the error control point, unequal loss protection is recommended to alleviate the effect of the lost packets [3]. At the transport layer level, TCP-friendly rate control has been proposed as the congestion control solution to share the network resource fairly with the dominated network TCP traffic and maintain the sending UDP media rate smoothly as much as possible [4,5].

In this paper we introduce a JPEG2000 based real-time scalable video communication system. In this system we considered two practical issues of video communications: transport packet format and real-time requirements. IETF RTP and RTSP streaming protocol are extended to take care of scalable delivery of video data from a single layered coding data to heterogeneous devices and different resolution such as HDTV, standard TV, PDA and mobile phone. Utilizing an advantage of low coding delay JPEG2000 codec, Real-time ARQ (RT-ARQ) is proposed to resolve the QoS issues for real-time applications. Moreover, various network adaptive control techniques, rate control and error control, are included in the system to achieve high transmission quality.

II. BEAM Scalable video communication system

BEAM Group in the Information Technologies Laboratories of Sony has been dedicated to propose and develop a robust & high quality video communication system. Figure 1 shows the overview of our implemented real-time video communication system, which realizes not only a one-way streaming application but also two-way interactive video transmission such like TV conference and TV phone. Along the data path, we introduce the function of each module. At first, the raw video is captured and compressed by the video codec. Since each frame is encoded independently within its own frame, JPEG2000 is chosen as a video codec for BEAM system. There are two factors for the selection. One is its network friendliness: high scalability and error resilience. Because of inherent layered JPEG2000 bit-stream, it is easy to provide and extract the required data as the user wants for an appropriate device. The feature of error resilience lies not only in the error resilient code itself, embedded in the bit-stream, but also in the intra frame coding. By intra frame coding, you can avoid error propagation among frames. Another factor that we consider is the JPEG2000’s low coding delay. Less than just 1.5 frame delay, or about 50ms, can be seen on encoding process in our system. Since the receiver can be HDTV, standard TV, PC, PDA or mobile phone, the sender needs to parse required data from all the layered data to match the different capable equipments. To do that efficiently, we
propose new RTP packet format to satisfy such service on the packetizer module, providing the indication of priority for each packet and scalable requirements. Focusing on keeping real-time constraints and high quality video on the receiver, various control techniques are supplied in the Control Technology module: network status monitoring, rate control, FEC, and Real-time ARQ (RT-ARQ). In this system, RTP [6] above UDP is served as the transport protocol for the network multimedia application while RTCP is accompanied to convey the feedback information for the sender to monitor the network status and adjust the control techniques to adapt to the network conditions. Loss prediction based TCP-friendly Congestion Control is proposed to react to the network congestion in the early stage and smooth the rate of the video data. In the application layer, FEC and RT-ARQ are adaptively selected or combined to recover the lost packets at a maximum. When the Round Trip Time (RTT) is small, RT-ARQ is then activated to recover the lost packets. As RTT increases and exceeds a given threshold, FEC becomes the dominated error control means for the real-time constraints of the video communication.

Our system consists of standard codec JPEG2000 and protocols for real-time applications defined in IETF.

[1]. JPEG2000 intra coding

JPEG 2000 is a new coding system that uses state-of-the-art wavelet based compression techniques and provides enhanced spatial-frequency resolution in the transformed representation of the image. By the discrete wavelet transform, the coder decomposes the spatial image into a series of subbands, which are divided into rectangular arrays of code blocks afterwards. Based on the EBCOT coding algorithm [7], each-code block is generated a separate highly scalable (or embedded) bit-stream. Such coding strategy allows a post-compression process to extract only the essential information for the targeting application and form the final transmitting bit-stream in a rate-distortion optimized order.

[2]. IETF streaming protocols

The Internet Engineering Task Force (IETF) is a voluntary standards body that is dedicated to making recommendations for how to communicate information over the IP networks. The IETF has recommended a number of methods based on RTP [6] and RTSP [8] for the delivery of bit-streams with real-time requirements.

RTP is an application layer component based on UDP as a transport mechanism. It relies on upper layer to handle the network congestion and packet loss. It is therefore suitable for delivering data with time deadline. RTP includes a sub-component known as RTCP that is used to convey control information between the sender and the receiver. RTSP is a session-oriented protocol that uses SDP to carry all description information associated with the streaming session.

III. Technologies for scalable & real-time delivery

This section introduces the techniques that have been developed in our BEAM scalable video system, mainly including two parts: RTP format for JPEG2000 delivery and real-time control technologies.

A. RTP format for JPEG2000 delivery

Figure 2 shows our proposed RTP payload header for JPEG2000 video stream. The detail of each field can be referred by [11]. Here we focus on explaining about priority field how to realize a JPEG2000 scalable delivery. The priority field included in the RTP packet indicates the importance of the JPEG2000 packet (jp2-packet). Typically, a jp2-packet in the lower layers and the lower sub-bands has higher priority. The priority value assigned to each jp2-packet can be defined in the priority mapping table. Default priority value in the table is assigned that is related to a jp2-packet sequence number. The reason why the sequence number is chosen as the priority value comes from the concept of JPEG2000 code streams that former jp2-packets should be more important than the later ones. The user can define this unique mapping table according to different application. The Priority field of RTP packet is embedded by the sender. If the RTP packet consists of one jp2-packet, of course the Priority field is filled with the priority value of this jp2-packet. For case that there are multiple jp2-packets in one RTP packet, the lowest priority value of the jp2-packets is set as the representative value of this whole RTP packet.

![Image](image-url)

Figure 2. Proposed RTP header for JPEG2000 video streams.

To facilitate streaming the same video to different capable devices, like HDTV, PDA, etc. We extended signaling parameters in RTSP (Real Time Streaming Protocol) by defining a new header field. An “X-
Scalability” header is proposed in RTSP PLAY method. Its syntax is simple as below,

   X-Scalability: <Resolution start-end >
   <Layer start-end>

Our extended parameters can be safely ignored by the receiver if they can’t be recognized. With this header, clients are possible to request scalable data and retrieve them from a server via “X-Scalability”. Figure 3 illustrates an example of its usage in the case of VoD applications.

   C \rightarrow S: PLAY rtsp://vodserver/contents/videoA.mj2
   RTSP/1.0
   Cseq: 4
   Session: 12345678
   X-Scalability : R0-R3 L1-L4

   S \rightarrow C: RTSP/1.0 200 OK
   Cseq: 4
   Session: 12345678

   C: Client, S: Server

Figure 3.: Usage of Proposed RTSP Play header: X-Scalability

In this example, the client requests a video file “videoA.mj2”, and set scalable range of each resolution and layer as R0-R3 and L1-L4. The range for the client are corresponding to portion of data of a frame segmented by 2-dimentional box ($ (R0-R3) 	imes (L1-L4) $) described in Figure 4, consisting of totally 16 jp2-packets.

On the server side, since these jp2-packets are ordered sequentially in the file along the packet number, the scalable server process can just easily read and skip the data continuously according to the client’s request quality..

B. Adaptive control technology

Rate control is very important for video delivery or communication applications in the Internet due to two aspects. One is that rate control has to compete the bandwidth fairly with the dominant traffic in the Internet, e.g. TCP, namely in TCP-friendly manner. To guarantee the stable and high quality video rendered in the receiver, the second function of rate control is to avoid the packet loss events. In this system we have proposed and implemented a TCP-friendly rate control (TFRC) with loss prediction method. TFRC is a congestion control mechanism designed for unicast flows operating in the Internet environment. A TFRC sender estimates the available network bandwidth based on the feedback information periodically received from a receiver, and adapt the sending rate by AIMD (Additive-Increase Multiplicative-Decrease) rate control. We have made a modification to this TFRC algorithm and added packet loss protection. Our method consists of two steps. Based on periodical feedback from the receiver, for example the packet lost ratio $ p $ and timestamp in the receiver, the sender can estimate the available bandwidth using the commonly used TCP throughput equation. To avoid packet losses, it is necessary to predict the congestion status and suppress the sending rate in the early stage so as to alleviate network congestion. Therefore we introduce the ratio of long term to the short term exponential moving average of the RTT as a signal to indicate the congestion status. It is intuitive that if the indicator is less than 1, it seems that the network is getting congested; while the indicator is greater than 1, it suggests that the network is getting sparse so that we can send more packets into the network. More technical details of this rate control can be referred to [5].

Fig. 5 shows the performance of our loss prediction rate control by comparing the video quality between the case with rate control and without rate control. The vertical axis shows the DSCQS (Double Stimulus Continuous Quality Scale) values, which is the quality measurement defined in ITU-R recommendation BT.500-10. The Less value, the better video quality is [9]. In the experiments, during the period from 30 to 60 seconds, TCP traffic is jammed, which results increasing its value sharply. The DSCQS value increase about 22 % without rate control, while the increase percentage is limited about 10 % with rate control. The reason is that the rate will retreat when loss prediction algorithm predicts the network congestion. Thus the number of lost packets is less than that in the case without rate control so as to achieve better quality. This experiment shows that the video quality can be improved by combining JPEG2000 and TFRC with loss prediction method.
FEC has been highly recommended for real-time transmission since it assures constant and low delay at the cost of overhead. The idea of FEC across the packets is to transmit the redundant packets to help the receiver to recover lost packets. Reed-Solomon (RS) code is selected in our system, which is perfectly suited for recovering the erasure errors of packet loss. Across the packets, $RS(n, k)$ encodes $k$ information symbols (each symbol per packet) into $n$ symbols so as to construct the $n-k$ parity packets. The key point of FEC is how to decide protection degree. We adaptively increase or decrease the redundant packets based on feedback $RTT$ information and lost condition history so as to assure robust transmission and small overhead.

Since ARQ induces the redundancy only when the packets are lost, as error correction capability ARQ has shown more effective performance than FEC under actual network environment. However, traditionally ARQ is not considered suitable for real-time applications, which produces additional retransmission delay. In the BEAM system, we have implemented Real-time ARQ (RT-ARQ) to meet the time constraint and to achieve high quality video transmission. With the deployed buffer, say the buffer time 190ms, it is possible to request the lost packets retransmission request when $RTT$ is smaller than 126ms. It should be mentioned that RT-ARQ works in the application layer level then RT-ARQ is enabled only in the case that the number of lost packets is larger than the number of parity packets for FEC block. For the packet loss detection, three timings are proposed in this protocol: 1) At packet arrival timing to check a gap in the sequence number of each packets; 2) A periodical interval timer when receiver checks a gap within a fixed period of a specified timestamp or sequence number in periodical interval to monitor the success of the retransmission; 3) The “Last Chance” timing, defined as the play out time minus the $RTT$ value, which is the final retransmit request in time for play out to keep real-time constraints of communication applications. The essence of proposed RT-ARQ is to try to recover the lost packets before the deadline that applications require. More details can be referred to [10].

IV. Conclusions

In this paper we described a JPEG2000 based real-time scalable video communication system. In this system two great features have been realized for video communication applications: scalable delivery and keeping real-time constraints. IETF RTP and RTSP streaming protocols are extended to handle scalable delivery of video data from a single layered coding data to heterogeneous devices and different resolution display such as HDTV, standard TV, PDA and mobile phone. Utilizing an advantage of low coding delay JPEG2000 codec, real-time ARQ (RT-ARQ) is proposed to resolve the QoS issue for real-time applications. In addition to this protocol, various network adaptive control technologies, including rate control and FEC error control, are presented to achieve high quality video transmission. This system can provide best quality at a given network condition by combination of JPEG2000 based scalable delivery, rate control, FEC and RT-ARQ error controls.
Acknowledgment

The authors are grateful to Professor Yutaka Takahashi of Kyoto University, Dept. of Systems Science, Graduate School of Informatics, for collaborative research and all the supportive work on the analysis of Real-time ARQ.

Reference


