Dual-video-scheme Videoconferencing

Junhui Zhou, Bernard Cousin, Hong Wang.

Abstract

Videoconferencing system delivers real time multimedia data through data network. Among data traffic of videoconferencing, video stream has comparatively high data rate, sometimes much higher than all other traffic streams, so in videoconferencing, the issue of video transport is peculiarly important. Video streams generated in videoconferencing will be multicast ~~to~~ over a multicast tree. A source usually generates a single video stream. Multicasting a single data stream in a heterogeneous network such as the Internet where the links have different bandwidths may cause some links to be overloaded and some links to be under-loaded. Layered coding is a solution to this problem. In this paper, we propose another solution: delivering two type data streams in two different schemes which have different bandwidth requirements to adapt to heterogeneous network environments. We refer to videoconferencing delivering two type video streams of different schemes as dual-video-scheme videoconferencing. The two sub-trees over which the two type streams are sent and two approaches of video translation between two video schemes are first described. Then the architecture for dual-video-scheme videoconferencing is discussed. A dual-video-scheme system can be formed by two mono-video-scheme sub-systems operating in two different video schemes with a gateway. This allows us to use existing systems to implement dual-video-scheme videoconferencing. The dual-video-scheme system IVS-H261/JPEG is implemented in this way. IVS-H261/JPEG consists of two mono-video-scheme videoconferencing system IVS-H261 and IVS-JPEG operating in interframe coding scheme H261 and intraframe coding scheme JPEG, respectively, with a gateway IVS-SERV. This paper finally presents the architecture of IVS-H261/JPEG and its typical application.

1 Introduction

The past years have been seeing rapid development of videoconferencing technology. The videoconferencing system delivers real time multimedia data such as video and audio data through a data network. It allows people at different locations to attend a videoconference. This will bring great facility to us in modern society.

Among data traffic of videoconferencing, the video stream has comparatively high data rate, sometimes much higher than all other traffic streams, so in videoconferencing, the issue of video transport is peculiarly important. The video streams generated by participants of the videoconference will be sent in multicast mode to a multicast tree where all the participants spread. A source usually generates a single video stream. Sending a single data stream in multicast is suitable for the homogeneous network where the bandwidth is identical everywhere. However for heterogeneous networks such as the Internet where the links may possess different bandwidths, multicasting a single data stream to a multicast tree may cause some links to be overloaded and some links to be under-loaded. A solution to this problem is limiting the data rate of the stream to the minimun bandwidth of the network. But this solution is not fair because those who are not at the narrow-band links are forced to lose some video quality that they are able to obtain. Another solution is delivering multiple data streams which have different bandwidth requirements to adapt to different network bandwidths.

Two layered coding[15] is such an idea. The sender encodes video data in two layers: the basic layer and the enhanced layer. The basic layer ensures basic video quality, while the basic layer plus the enhanced layer gives a better quality. The data stream of the basic layer is delivered over all the multicast tree, and that of the enhanced layered is delivered over a sub-tree of the multicast tree. This sub-tree has enough bandwidth to accommodate those two streams. Clearly layered coding streams can achieve better utilization of the bandwidth resources of the heterogeneous network than a single data stream.

In this paper, we propose another approach to produce and deliver multiple video streams in videoconferencing.

Because of the diversity of video coding techniques, we can find a lot of video coding schemes, and some of them have become international standards such as JPEG, MPEG, and H261[5]. A video source, using different coding schemes, may generate video streams of different data rates and have different bandwidth requirements. For example, intraframe coding scheme generally produces video streams of high data rate, and in contrast, interframe coding scheme produces video streams of low data rate because it takes advantage of correlation of consecutive frames. Data streams in intraframe coding need much greater bandwidth than those in interframe coding. But intraframe coding is superior to interframe coding in capacity of tolerating data loss. Data loss may happen in the network at any time for various reasons. That will impose grave influence upon interframe coding, but don't impose much influence upon intraframe coding[5]. Hence intraframe coding can ensure better video quality in the presence of data loss. When a source sends video data to a multicast tree in a heterogeneous network, we construct two sub-trees: one consists of broadband links and the other is the rest of the multicast tree, and install a video gateway at the common node of the two sub-tree. Then the source sends a data stream in intraframe coding to the former, and send another data stream in interframe coding to the latter (In fact, one of the video streams is generated and sent by the gateway.)

In this way, we can also achieve better utilization of the bandwidth resource as layered coding. We refer to two different video coding schemes as dual-video-scheme. Further analysis and comparison of dual-video-scheme among the video coding schemes is interesting, but this is beyond the scope of this paper. Our focus will concentrate on transport of dual-video-scheme data streams.

Videoconferencing falls into two categories: interactive videoconferencing and non-interactive videoconferencing. Interactive videoconferencing allows every participant to send video (and audio data), while non-interactive videoconferencing only allows one participant to do so. That is, in the interactive case, there are multiple video sources, while in the non-interactive case, there is only one single. If we use dual-video-scheme in interactive videoconferencing, the overhead of sub-tree construction and video gateways may be too large to be accepted, so dual-video-scheme is only suited for non-interactive videoconferencing . Hence in the following, we will only investigate the case of non-interactive videoconferencing.

In the following section, we will address the issues related to the video gateway. Section 3 will discuss architectural considerations for dual-video-scheme videoconferencing. In Section 4, we present a dual-video-scheme videoconferencing system IVS-H261/JPEG. Section 5 concludes the paper.

2 Video gateway

In dual-video-scheme videoconferencing, two video data streams which carry the same video information are multicasted on two sub-trees of the multicast tree, respectively. In order to save bandwidth resources, the two sub-trees of the multicast tree are supposed to have no common link (but they have a common node). This is different from layered coding. In the case of layered coding, data streams are multicasted to overlapped sub-trees (the data stream of the enhanced layer is sent to a sub-tree of the tree where the data stream of the basic layer is sent).

The construction of the two sub-trees depends on bandwidth characteristics of network links. Suppose in a multicast tree T illustrated in Fig.1, the bandwidth of each link directly attached to the source is 10Mbps, and the others links have bandwidths of 64Kbps, 1.5Mbps or

10Mbps. The links directly attached to the source can form a broadband sub-tree T_1, and the others form another sub-tree T_2. T_1 has enough bandwidth to accommodate a video stream in some intraframe coding scheme, while *T2* can only accommodate a video stream in an interframe coding scheme such as H261. Therefore we can multicast a video stream in intraframe coding on T_1 and another video stream in interframe coding on T_2.

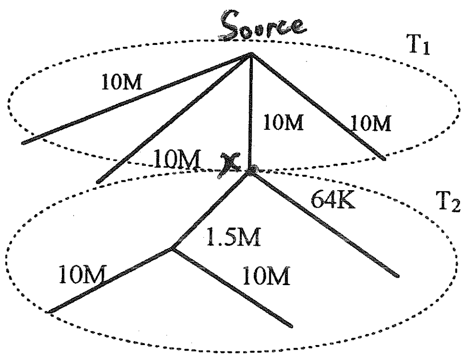

Fig.1 Multicast tree *T* for a videoconference

In fact, the multicast tree *T* has been divided into two multicast trees T_1 and T_2. We need to give each of T_1 and T_2 a host group address that will be different from the group address of *T*. Logically, *T*, T_1 and T_2 can be considered to be three parallel multicast trees. As we see later, according to the architecture proposed in this paper, a dual-video-scheme videoconferencing system will no longer multicast data on *T*.

Clearly, at the common node of T_1 and T_2, a video gateway is required to receive data from T_1, and translate to another data stream then multicast to T_2.

Translation from a video scheme S_1 to another one S_2 can be achieved through direct conversion. Direct conversion is an effective approach to translate video data. However in practice, conversion method is too hard to be found out for some schemes (or does not exist at all).

Another approach to translate video data is first decoding then encoding, that is, first decoding video data in S_1 to get original data, then encoding in S_2. Conversion may not be realized for some schemes. But this way can always be realized. Such translating processes can be described as a directional pair:

$$P(S_1\text{->}S_2): <S_1 \textit{ decoding}, \quad S_2 \textit{ encoding}>.$$

Of course $P(S_1\text{->}S_2)$ could also be considered to be of a type of special conversion. $P(S_1\text{->}S_2)$ may need certain time overhead, but its influence on non-interactive videoconference can be ignored. In our implementation of dual-video-scheme videoconferencing prototype IVS-H261/JPEG, we adopted the approach first decoding then encoding.

Direct

3 Architectural considerations

A dual-video-scheme videoconferencing system will operate in two different video schemes. It can be implemented as one single system with two video codecs, or two subsystems, each of which contains one video codec. We can find that the latter gives an open system structure for dual-video-scheme videoconferencing. It allows us to adopt some existing systems to implement a dual-video-scheme videoconferencing system. This is just the approach to implement our dual-video-scheme videoconferencing system IVS-H261/JPEG. Therefore in the following, we will consider that the dual-video-scheme videoconferencing system consists of two mono-video-scheme videoconferencing systems operating in different video schemes with a gateway.

The two subsystems in a dual-video-scheme system are expected to have the same architecture. Otherwise the system complexity increases unnecessarily, and it may be difficult for the subsystems to implement communications. The gateway should also have the same architecture as the subsystems. Thus both the subsystems are similar except for their different video codecs. This implies the optimized way to implement the two subsystems is first developing an one-video-scheme videoconferencing system or selecting an existing videoconferencing system then substituting the video codec in it by another one to obtain the other system. We will discuss further the architecture for dual-video-scheme videoconferencing systems in a particular example in the following section.

4 Dual-video-scheme system IVS-H261/JPEG

IVS-H261/JPEG is a dual-video-scheme videoconferencing system on the Internet. It consists of two videoconferencing systems IVS-H261 [18] and IVS-JPEG [19] with a gateway IVS-SERV [1]. IVS-H261 and IVS-JPEG had been developed successively, then we have developed IVS-SERV which allows them to become a dual-video-scheme videoconferencing system.

IVS-H261 and IVS-JPEG adopt the video scheme H261 and JPEG, respectively. IVS-H261 is also known as IVS on the Internet. We refer to it as IVS-H261 in this paper to explicitly point out the video scheme used. The data rate of video codec in IVS-JPEG can reach about 1.5Mbps, while that in IVS-H261 is relatively low, about 30Kbps. The difference is great. This is because, first, H261 is an interframe coding scheme, and JPEG is an intraframe coding scheme; second, the video codec in IVS-H261 is implemented in software while that in IVS-JPEG is hardware codec in XVideo Parallax card, and the speed difference of hardware codec and software codec makes the difference of those two data rates even greater. Clearly, IVS-H261 is suitable to narrow-band links of the Internet. And IVS-JPEG is suitable to broadband links of the Internet (We assume the hosts involved are equipped with Parallex card). Although IVS-JPEG requires a comparatively high bandwidth, it can tolerate data loss happening in the network.

Both IVS-H261 and IVS-JPEG have the same architecture, illustrated in Fig.2. They are based on the User Datagram Protocol (UDP). Compared with another protocol TCP on top of IP, although UDP doesn't ensure reliable data transport, it generates small transport delay that is essential for real time network applications such as videoconferencing.

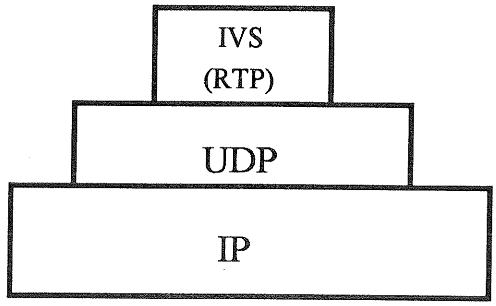

Fig. 2 Architecture of IVS-H261 and IVS-JPEG

IVS-H261 and IVS-JPEG use the Real-Time Transport Protocol (RTP) [20] to deliver video and audio data. RTP is proposed by the Internet Audio-Video Transport Working Group to implement real time video and audio data transport in packet-switched networks such as the Internet. RTP is comprised of two sub-protocols: RTP Data Transfer Protocol and RTP Control Protocol. The former provides data delivery service including data type identification, data packets sequence numbering, and data samples timestamping. The latter monitors the quality of service (QoS). In either IVS-H261 or IVS-JPEG, RTP doesn't emerge as a separate layer. It is integrated into the two systems, as illustrated in Fig.2.

Video transport and audio transport multiplex the same multicast tree. That is, they multiplex the same IP host group address. The two types of data streams are demultiplexed through different UDP ports.

IVS-H261 and IVS-JPEG are incorporated into the dual-video-scheme videoconferencing system IVS-H261/JPEG. The typical application case of IVS-H261/JPEG is where some participants are in the same local area network (LAN) as the conference speaker, and the others are in other LANs connected with that LAN through backbone links. The first set of participants can take advantage of the broadband links of the LAN. They use IVS-JPEG. The bandwidths of backbone links on the Internet are not too great (1.5Mbps, 384Kbps, or even 64Kbps). In addition, there are various traffic flows on backbone links. The bandwidth available will only accommodate IVS-H261 traffic. The two sets of participants form two host groups, and each group is given a unique IP group address. The two multicast trees corresponding to the two host groups are, in turn, created. In fact, these two trees are two sub-trees of the original multicast tree which contains all participants. At the common node of the two tree, the video gateway IVS-SERV will translate video data from JPEG to H261.

The gateway IVS-SERV is supposed to have the same architecture as IVS-H261 and IVS-JPEG. It is based on UDP and uses RTP to implement real time transport. Since the gateway is the common node of the two sub-trees, it needs to join the two host groups corresponding to the two sub-trees. The gateway receives video data in the host group which the speaker is in, translates and sends to the other host group. The method of video translation is first decoding then encoding, that is, through the directional pair

<JPEG decoding, H261 encoding>.

The JPEG codec and H261 codec in IVS-SERV are hardware codec and software codec, respectively, as in IVS-JPEG and IVS-H261.

The speaker also sends audio data. Since audio transport multiplexes the same multicast tree with video transport in IVS-JPEG, those who are in the host group using IVS-H261 are not able to receive audio data sent by the speaker. In order to enable all participants to hear the speaker, the gateway needs to forward the audio data, from the host group using IVS-JPEG to the other. IVS-JPEG and IVS-H261 can operate in the same audio scheme, say PCM or ADPCM, to have audio interoperability. Therefore audio translation is not required.

Videoconferencing should perform QoS control. The gateway IVS-SERV can be considered as a destination in one host group and the source in the other, so we can consider that dual-video-scheme videoconferencing creates two separate videoconferencing sessions, and QoS control will be performed within each session separately. In the IVS-JPEG session, as a destination, the gateway sends QoS monitoring data to the source, while as the source in the IVS-H261 session, the gateway receives QoS data from the participants and control its data rate through the same QoS control mechanism as in both IVS-H261 and IVS-JPEG[21].

5 Conclusion

We have investigated the use of dual video schemes in videoconferencing. Dual-video-scheme videoconferencing can produce two type video streams to adapt to heterogeneous network environments. The architecture for dual-video-scheme videoconferencing is proposed. A dual-video-scheme videoconferencing system consists of mono-video-scheme subsystems with a gateway. The gateway translates video data between the two subsystems. Translation for different video schemes can be achieved in two ways. One is direct conversion; the other is first decoding then encoding. Direct conversion is more effective but cannot always be achieved. The translation in the way first decoding then encoding is implemented in IVS-SERV, the gateway for IVS-H261 and IVS-JPEG. IVS-SERV allows two existing videoconferencing systems to become a dual-video-scheme videoconferencing system IVS-H261/JPEG. IVS-H261/JPEG can well adapt to heterogeneous network environments of the Internet.

References

[1] Bernard Cousin, Zhou Junhui, Translation gateway for videoconferencing systems, Proc. IMAGE'COM 96, May 1996, Bordeaux, France, pp. 133-138

[2] M. Aras, James F. Kurose, Douglas S.Reeves, H. Schulzrinne, Real-TimeCommunication in Packet-Switched Networks, Proceedings of the IEEE, Vol. 82, NO.1, January 1994,

[3] Mark J. Handley, Peter T. Kirstein, M. Angela Sasse, Multimedia integrated conferencing for European researchers (MICE): piloting activities and the conference management and multiplexing centre, Computer Networks and ISDN Systems, No. 26, 1993, pp. 275-290.

[4] Michael R. Macedonia, Donald P. Brutzman, Mbone provides audio and video across the Internet, IEEE COMPUTER magazine, April 1994, pp. 30-36.

[5] R. Aravind, G. L. Cash, D. L. Duttweller, H. M. Hang, B. G. Haskell, A. Puri, Image and Video Coding Standards, AT&T Technical Journal, January-February, 1993, pp. 66-89.

[6] Gregory K. Wallace, The JPEG still picture compression standard, Communications of the ACM, April 1991, Vol.34, No.4, pp.31-44.

[7] Ming Liou, Overview of the p times 64kbps video coding standard, Communications of the ACM, April 1991, Vol.34, No.4, pp.59-63.

[8] LeGall, D.J. , MPEG: A video compression standard for multimedia applications, Communications of the ACM, April 1991, Vol.34, No.4, pp.47-58.

[9] Baker, S., Multicasting for sound and video , Jnix Review, Feb. 1994, pp. 23-29.

[10] K. Jeffay, D.L. Stone, F. D. Smith, Transport and display mechanisms for multimedia conferencing across packet-switching networks, Computer Networks and ISDN Systems, July 1994, Vol.26, pp. 1281-1304.

[11] Shiro Sakata, Development and Evaluation of an In-House Multimedia Desktop Conference System, IEEE JSAC, Vol. 8, NO. 3, April. 1990, pp. 340-347.

[12] Cosmos Nicolaou, An Architecture for Real-Time Multimedia Communication Systems, IEEE JSAC, Vol. 8, NO. 3, April. 1990, pp. 391-400.

[13] T. Turletti, J. C. Bolot, Issues with multicast video distribution in heterogeneous packet networks, Proc. 6th. International Workshop on Packet Video, Porland, Oregon, Sept. 1994, pp. F3.1-3.4.

[14] J. Pasquale et al., Filter propagation in dissemination trees: Trading off bandwidth and processing in continuous media networks, Proc. NOSSDAV'93, Lancaster, UK, Oct. 1993, pp. 259-268.

[15] Don Hoffman, Michael Speer, Gerard Fernando, Network support for dynamically scaled multimedia data streams, Proc. NOSSDAV'93, Lancaster, UK, Oct. 1993, pp. 240-251.

[16] D. Ferrari, Client requirements for real-time communication service, IEEE Communications, November 1990, pp. 65-72.

[17] Xu Mengxia, Advances in Picture Coding, Journal of China Institute of Communications, March 1993, Vol.14, No. 2, pp 40-47.

[18] Thierry Turletti, H.261 software codec for videoconferencing over the Internet, Technical Report 1834, INRIA Sophia Antipolis, January 1993.

[19] Le Coq Patrice, et al., Analyse et modification d'un systeme multimedia, Rapport de projet DIIC 3 LSI, IFSIC, France, Feb. 1994.

[20] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, RTP: A Transport Protocol for Real-Time Applications, Internet-Draft, March 21, 1995.

[21] J-C. Bolot, T. Turletti, A rate control mechanism for packet video in the Internet, Proc. IEEE INFOCOM'94, June 1994, Toronto, pp 1216-1223.