A Model for Removing Skew in Network Multimedia Communication for Guaranteed QoS in Packet Network by Max- Packet Generation

Shyamalendu Kandar and C. T. Bhunia

***Abstract*—Multimedia data are human sensible. These types of data are delay intolerable but error tolerable to some extend. Skew and Jitter are two important parameters that degrade the quality of service (QoS) of multimedia services. Achieving guaranteed Quality of Service (QoS) of multimedia service is a great research challenge. Earlier investigations attempted to use buffer management and timestamp approach to meet the challenges of skew to achieve the guaranteed QoS .In this paper a new approach called Max-Packet generation approach is proposed and described to remove the effect of skew in network multimedia communication to achieve the guaranteed QoS. A model of generation of Max-Packet is also described in the paper.**

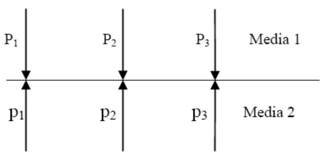
***Index Terms*—Combined packet, max-packet, only video packet, skew.**

I. Introduction

Multimedia data are integration of one or multiple media components like Text, Graphics, Audio, Video and Animation [1],[2]. In continuous bit rate (CBR) services in multimedia communication like audio and video, media synchronization is one of the key challenge. In order to achieve some guaranteed QoS, two issues that are paramount importance particularly for continuous bit rate (CBR) services [3] are: jitter and skew [4],[5]. Jitter is caused due to the variable delay that occurs during transmission through network between the packets of a particular service, say only for audio or only for video. Jitter and Skew may be grouped into media synchronization. Jitter is called intramedia synchronization. Jitter can be removed by providing a compensation buffer at the receiving side or by modified packet using clock synchronization [6].

Skew refers to the variable delay between the two (or more) corresponding packets of two (or more services) during transportation in the network. Skew can be called intermedia synchronization generally referred to as lip synchronization.

Transmission Side:



Manuscript received July 25, 2012; revised August 30, 2012.

Shyamalendu Kandar is with Computer Science & Engineering of Haldia Institute of Technology, India (e-mail: shyamalenduk@yahoo.com).

C. T. Bhunia is the Director of NIT, Itanagar, Arunachal Pradesh, India, and senior member of International Centre for Theoretical Physics, Italy(e-mail: ctbhunia@vsnl.com).

Receiver Side:

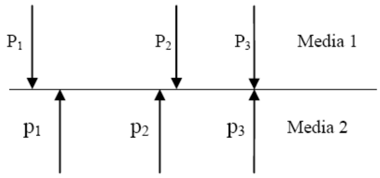


Fig. 1. Occurrence of skew

At the transmitter side say for two media components Audio & Video, Pi is the packet for Video and pi is the packet for Audio. At the time of transmission at the transmitter side say T is instant of transmission of K th sample of media1 (Video). So in the transmitter side T will also be the instant of transmission of K th sample of media2 (Audio). At the receiver side, say at $T^{/}$ instant of time K th sample of media1 (Video) is received. Say at $T^{//}$ instant of time the K th sample of media2 (Audio) is received. Skew will occur when $T^{/} \neq T^{//}$ [Fig. 1] [7].

If skew occurs, at the receiver side there will be a mismatch between audio & video. This will affect the user in realizing the multimedia data. This type of mismatch occurs due to different types of delay in the network. These delays are Propagation delay, Transmission Delay, Queuing Delay and Node Processing Delay. In order to remove skew different researchers investigated different techniques. From implementation point of view, these techniques are hard to provide guaranteed QoS. Basically they are supposed to provide a pre defined perceived Quality of Service.

Section 2 describes and finds the merits and demerits of the existing techniques. Section 3 describes the structure of the proposed Max-Packet. Section 4 elaborately discusses the Max-Packet generation process. Section 5 describes the model of the generation of the Max-Packet and model of the structure of the Max-Packet. Section 6 gives a calculation of the overhead bits. Section 7 draws the conclusion and Section 8 gives the future scope of the current work.

II. Already Existing Techniques

A. Media Synchronization Based on Buffer Control

There are several techniques to reduce the effect of the problem of Skew. One such technique is known as Media Synchronization Based on Buffer Control [8]. This technique is represented by a model where there is a computer with a display and several distributed databases interconnected through high speed digital network. Each medium is stored in the database and it is transmitted through network to the computer for display. All of the components are governed by a common clock.

DOI: 10.7763/IJMO.2012.V2.210

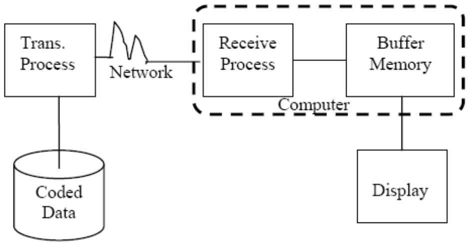


Fig. 2. Model for media synchronization based on buffer control.

Every frame has the sequential frame number starting with 0 and a timestamp indicating the time at which the frame should be displayed. The frame with frame number m should be displayed at m X T_s, where T_s is the frame interval. All the frame waits in the buffer memory and according to the common clock the computer starts display of the successive frames waiting in the buffer.

The model assumes the rate of motion video is 30 frames/ sec and if the computer is unable to display 30 frames/ sec it is allowed to discard some frames. But it does not discuss the number of frames to be discarded. The model also assumes that the coded data will be transmitted through a high speed network. But in real life a network (internet) is busy with transmitting so many data packets. It has not given the critical issues for a highly loaded network. The model has not given the size of the buffer memory. It also has not given details about the common clock and the implementation of the given model.

B. Real Life Audio-Video Synchronization Using Timestamp

The model for real life Audio-Video Synchronization using timestamp consists of two parts- sending end and receiving end. Sending end consists of four parts Data Acquisition, Data Compression, Data bale for both audio and video and Data Sending [9]. The main aim of data bale is making sending end's common work more convenient and subjoining essential additional information like timestamp.

At the receiving end there are a receiving device and Buffer, Decompression unit and Display unit for both audio and video.

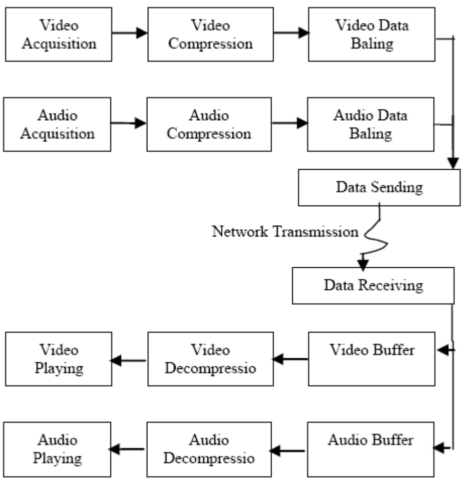


Fig. 3. Model for real life Audio-Video synchronization using Timestamp

According to the model the audio and video data are acquired from respective buffer and sent to respective player according to the timestamp of audio data. If T_a is the time stamp for audio and T_v is the timestamp for video data the model acts as the following.

If $|T_a - T_V| \leq 200$ms , audio and video data are synchronous.

If $|T_a - T_V| \geq 200$ms, audio and video data are asynchronous and the model has given two circumstances

If $T_a>T_V$, video need to quicken play speed until $|T_a - T_V| \leq 200$ms. If $T_a<T_V$, some current and later video frames can be played repeatedly until audio and video data are synchronous.

According to this model if $|T_a - T_V| \geq 200$ms , there appears skew whether it is $T_a> T_V$ or $T_V > T_a$. The process of quickening or slackening the video frames is not discussed elaborately in the model. The model does not give the size of audio and video buffer. The model has not given detail about the clock by which the timestamp is being placed to the data packets. Above all no experimental or implementation process is not given for the model.

III. STRUCTURE OF THE PROPOSED MAX-PACKET

Skew occurs due to the variable delay between the two (or more) corresponding packets of two (or more services) during transportation in the network. From the previously described techniques it is clear that the Skew is not totally removed by any of the techniques. The two techniques can remove Skew to a certain limit and only in preassumed conditions like high speed dedicated network. To remove skew, the concept of Max packet is proposed in this work. The goal is to provide guaranteed QoS in network multimedia communication by removing Skew totally.

If the samples of two or more media components generated with in a same time instant is grouped in a single packet then at the receiver side the variable delay between two or more corresponding packets of two or more services will not appear during transportation in the network. If two media components voice & video are considered then the corresponding data of two medias generated within a same time instant will be in the Max-Packet.

If a multimedia data consists of only audio & video then with in a same time instant it is seen that size of video data is much more than the size of audio data. For video the least frequency of a video signal is 33.4 MHz, where as for audio the maximum audible audio frequency is 20 KHz. For some instant of time the size of data for both of the components can vary. So the Max-Packet will be of variable length. In a Max-Packet there are more than one media components [In this case it is two, Audio & Video], so there must be some extra bits in between two media data in the Max-Packet to differentiate between the two media components at the receiver side. As in the Max-Packet data of more than one media exist there must be some bits to indicate each type of media data within the packet.

In multimedia there are 6 types of multimedia data [Text, image, Graphics, Audio, Video & Animation], but those can be grouped into 4types [Text, Image, Audio, Video]. So two bits are sufficient to identify the media components [00, 01, 10, 11]. Yet in this model 3 bits are considered means 8 combinations can be generated. Say for 000 for Audio, 001 for Video, 011 for Text, 111 for Image. This bit pattern will be known to both sender & receiver side.

Flag (1 byte)	Address (1 to several byte) [Depends on the number of nodes	Control bit (1 or two byte)	Bit to recognize audio	Audio (Variable Length)	Stuffed bit	Bit to recognize Video	Video (variable length)	Frame check sequence(2 to 4 byte)	Flag (1 byte)

Fig. 4. Variable length MAX-Packet

IV. MAX-PACKET GENERATION PROCESS

On a laboratory scale Audio signal of frequency 20 KHz and video signal of 40 MHz are taken. These signals are sampled separately just at Nyquist rate (40 KHz for Audio and 80MHz for video) followed by conventional process of PCM (Pulse code modulation).Then the samples are needed to be quantized. In order to maintain the synchronization between audio and video signals the quantization level (**Q**$_l$) must be kept same for both of the signals. After quantization, the signals will be digitized. These digitized values will be put in packets. To remove skew the digitized values of the samples of audio & video with in a given time instant will be put in a single packet. For developing the packet, one media is taken as Master and another is taken as Slave. In the current work Audio is taken as Master and Video is taken as Slave. In current research work there is a frequency difference of 2 X 10^3 times to video signal than that of audio signal. That means if two audio samples are taken at time instant t_1 & t_2 , then for these two audio samples there will be $(t_2 - t_1)$ X $(2 * 10^3)$ number of video samples provided $t_2 > t_1$. If a combined packet is produced by taking the audio and video samples between t_1 & t_2, the packet size will be huge. By using the large packet the benefit of packet switching will not be obtained.

This can be handled in another way. One combined packets will be made from the quantized values, obtained from the samples at time instant t_1, and in time instant t_2 another combined packet will be made. Between t_1 & t_2 there will be (n-1) number of quantized slave sample, i.e. only video samples. Where n= C_f video/C_f audio, C_f stands for carrier frequency. m number of video samples will be grouped to form a Only Video packet. For (n-1) video samples number of packets will be $\lceil (n-1)/m \rceil$

For each packet [Combined or only video] there is a header of 1 byte i.e. 8 bit. For combined packet the header is 00000000 & for only video packet the header is 11111111. Between two combined packets the only video packets are numbered to do synchronization at the receiver side. There is no packet containing only audio data.

V. MODEL OF THE GENERATION OF MAX-PACKET

From the discussion in the previous topic it is clear that the total process is divided into several stages. The model of generation of Max-Packet consists of four phases given below.

1) Sampling of Audio signal
2) Sampling of Video signal
3) Combined packet Formation
4) Only video packet Formation

The different phases are described in detail.

A. Sampling of Audio Signal

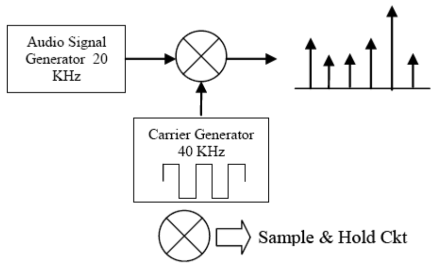


Fig. 5. Sampling process of audio signal

B. Sampling of Video Signal

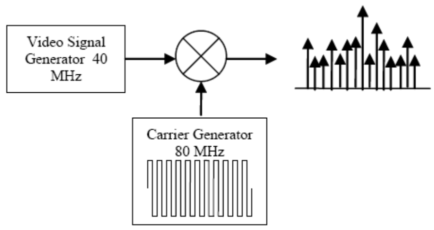


Fig. 6. Sampling process of video signal

C. Combined Packet Formation

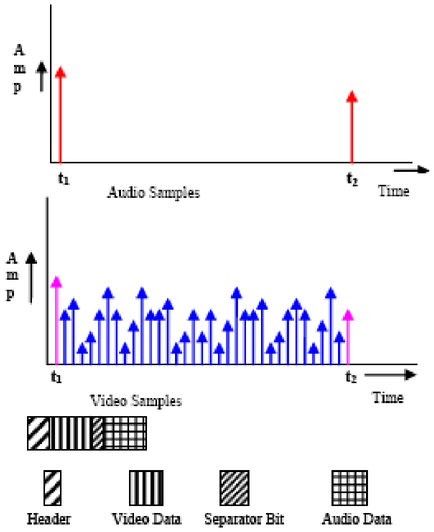


Fig. 7. Combined packet containing video & audio data

D. Only Video Packet Formation

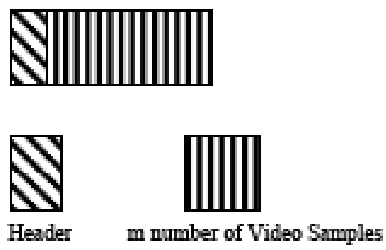


Fig. 8. Only video packet

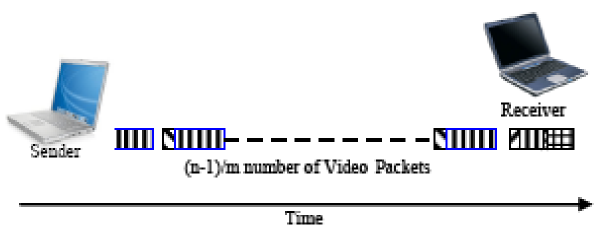


Fig. 9. Packet transmission from source to destination

VI. Calculation of Overhead Bits

Overhead bits means how many number extra bits are sent with the original data at the time of packets transmission from sender to receiver. For packet switching these overhead bits must be taken into consideration at the time of transmitting data from source to destination. Overhead bits are added in the form of header of the packet or any stuffed bits.[In this case in the form of separator bit]. This calculation is necessary because overhead bit increases size of the data which increases the transmission time from source to destination in packet network.

Let the multimedia data is transmitted within a time period T.

After a certain time period interval audio and video data are grouped and forms combined packets. Let such type of first packet is formed at the time instant t_1 and the second such type of packet is formed at the time instant t_2.

So total number of combined packets within the time period T, is $\lceil T/(t_2 - t_1) \rceil + 1$.

For each such type of combined packet the overhead bits are in the formed of Header (h) and Separator bit (s).

So number of overhead bits for combined packets is $(\lceil T/(t_2 - t_1) \rceil + 1) \times (h + s)$. Within the time period (t_2 - t_1) there are $\lceil [(n-1)/m] \rceil$ number of Only Video packets. Where n= quantization level and m= number of video samples in a single Only Video packet.

Total number of Only Video Packets within the time period T is $[T/(t_2 - t_1)] \times \lceil [(n-1)/m] \rceil$. For each Only Video packet overhead bit is header (h). So total number of overhead bits for only video packets is

$$h \times [\{T/(t_2 - t_1)\} \times \lceil [(n-1)/m] \rceil]$$

So, during transmission of the total multimedia data the overhead bits are

$$\left[(\lceil T/(t_2 - t_1) \rceil + 1) \times (h + s) \right] + h \times [\{T/(t_2 - t_1)\} \times \lceil [(n-1)/m] \rceil]$$

$$\approx [T/(t_2 - t_1)] \times [h + s + h \times \lceil [(n-1)/m] \rceil]$$

A future challenge of this work is to make a comparative gain of QoS with respect to overhead bits. The overhead bits provide a trade off with expected QoS.

VII. Conclusion

In this paper a concept of Max packet is proposed, to eliminate the effect of skew in network multimedia communication over packet network. In comparison with the other models which are able to remove skew to a certain extent, this model can remove skew totally due to the group of audio and video samples produced at same time instant in the proposed Max-Packet. Real time video communication like video conferencing through internet where there is a high chance of occurring skew, this technique can be applied to provide the QoS. Data compression to the video data by providing I, B and P frames can be done on the packets containing only video data. The proposed technique is under practical implementation & testing.

VIII. Future Scope

By this technique skew will be totally removed but due to network delay there may occur jitter of video packets. We believe if the technique is implemented with other techniques namely buffer management technique of removing skew & jitter; we may derive it as a sound technique which will successfully remove skew & jittering in network multimedia communication and will provide Quality multimedia service. A future challenge of this work is to make a comparative gain of QoS with respect to overhead bits. The overhead bits provide a trade off with expected QoS.

References

[1] R. Stteinmetz and K. Nahrstedt, *Multimedia: Computing, Communications and Applications*, Pearson Education Asia, 1995

[2] R. Parekh, *Principles of Multimedia*, Tata McGraw Hill, 2006

[3] C. T. Bhunia, "Multimedia Services with Guaranteed QoS over Packet Networks/ Internet," *Journal of Institution of Engineer, India,* C. P., vol. 87, pp. 38, November 2006

[4] J. D. Gibson, *Multimedia Communications*, Academic Press, 2001

[5] J. F. K. Buford, *Multimedia Systems*, Addison Wesley, 2000

[6] S. Kandar and C. T. Bhunia, "A New Protocol for Minimizing Jitter for Guaranteed QoS in Network Multimedia Communication," *ACE 2010, Bangalore,* pp. 15-18 *IEEE explore 2010* , June 2010

[7] C. T. Bhunia, S. Kandar, and M. Barman , "Technique in removing skew and analysis," in *Proc. National Conf. RAFTIT,* 2008, pp. 184-189.

[8] Y. Sakai el at. , "Media Synchronization Method for Motion Video Based on Buffer Control," *IEEE explore,* 1995, pp. 785-789

[9] Z. Jingfeng and L. Ying el at, "Using Timestamp to Realize Audio-Video Synchronization in Real-Time Streaming Media Transmission," *IEEE explore,* 2008, pp. 1073-1076.