MPEG-2 over ATM: System Design Issues

Subir Varma
Communications Products Division
LSI Logic Corp.
1525 McCarthy Blvd., MS G-711
Milpitas, CA 95035
e-mail: subirv@lsil.com

ABSTRACT

The ISO/IEC Standard Committee recently adopted a packet format known as the Transport Stream (TS) for sending MPEG-2 encoded data over communication networks. During the past year the ATM Forum established a sub-working group to examine issues connected with the transmission of MPEG-2 TS packets over ATM networks. The objective of this paper is to summarize the present state of this work, and also to point out the design issues that had to be resolved. In particular we describe the issues of MPEG-2 System Clock recovery in the presence of network jitter and encapsulation of TS packets in AAL5 PDUs.

1.0 Introduction

The objective of this paper is to provide a survey of the issues involved in transmitting MPEG-2 Transport Stream (TS) packets over ATM networks. The MPEG-2 standard was developed by the Moving Pictures Expert Group (ISO/IEC JTC1/SC29/WG11) while ATM is being developed by the ITU SG-XVIII and the ATM Forum. The Forum is in the process of finalizing Phase 1 of an implementation agreement that addresses Video on Demand (VoD) using constant bit rate (CBR) encoded MPEG-2 Single Program Transport Streams (over ATM). We will summarize some of the technical issues that are behind the design choices that were made in this agreement.

The rest of this paper is organized as follows: In Section 2 we provide an overview of the MPEG-2 standard, in Section 3 we discuss the protocol layering structure for MPEG-2 over ATM and the choice of an appropriate AAL. In Section 4 we discuss the issue of clock recovery and the effect of network jitter on it, while Section 5 covers the issue of encapsulation of TS packets in AAL5 SDUs. Finally in Section 6 we discusss some of the problems involved in supporting VBR MPEG-2 TS streams over ATM networks.

2.0 MPEG-2

MPEG is an inter-frame codec specification for digital video and it currently is the de-facto standard for video compression. The MPEG-2 Standard consists of three parts:

MPEG-2 Video. This specifies the compression algorithm for high-quality digital video. In addition to being compatible with the MPEG-1 Video Standard (ISO/IEC IS 11172-2), MPEG-2 also supports interlaced video formats, increased image quality, several picture aspect ratios and a number of other advanced features, including those needed for HDTV.

MPEG-2 Audio. This supports up to five full bandwidth channels, plus an enhancement channel, and/or upto seven commentary/multilingual channels. The MPEG-2 Audio standard also provides improved quality coding of mono and conventional stereo signals for bit-rates at or below 64 kbps.

MPEG-2 Systems. This standard specifies how to combine multiple. audio, video and private data streams into a single multiplexed stream. It performs packetized stream control and synchronization and is designed to support a wide range of broadcast, tele-communications, computing and storage applications. It defines two kinds of streams: Program and Transport Streams. The Systems Layer processes the compressed video codec, audio codec and data streams in two steps:

1. First the raw codec and data streams are combined with system level information and packetized to produce Packetized Elementary Streams (PES).
2. The PES's are then combined to form either a Program Stream or a Transport Stream.

The Program Stream supports the creation of a single audio-visual program, which could have multiple views and multi-channel audio. It utilizes variable length packets and is designed for transmission in relatively error free environments.

The Transport Stream multiplexes several programs, comprised of video, audio and private data, for transmission and storage using a wide variety of media. It is designed for transmission in a lossy or noisy environment and utilizes a fixed size 188 byte (184 byte data + 4 byte header) packet. It can re-construct the program clock for each one of the multiplexed programs and also play the decoded elementary streams (in each pro-

1063-6390/96 $5.00 © 1996 IEEE
Proceedings of COMPCON '96

gram) with correct synchronization between the audio and video. Packets in a Transport Stream follow each other without gaps, when there is no data to send, a special null packet is sent.

Note that either raw elementary streams, PES, Program Streams or Transport Streams can be used for carrying MPEG-2 data across an ATM network. The fundamental drawback to using raw elementary or PES streams is that neither stream type carries explicit information to facilitate clock recovery at the decoder. Both Program and Transport Streams carry this information, but Transport Streams are preferable for the following reasons:

- Transport streams will dominate non-ATM forms of video distribution, such as DBS, CATV etc. Extensive Transport Stream hardware development for these applications is already underway
- Transport streams are more robust to errors than Program streams, since they use short fixed sized packets.

Hence the choice was made that MPEG-2 data should be transmitted across ATM networks by using the Transport Stream format.

3.0 Choice of AAL

The MPEG-2 standards do not specify how the MPEG-2 Transport Stream should be sent to the decoder. ATM's cell-based structure and large bandwidth make it an attractive candidate for delivering MPEG-2. The protocol reference model for MPEG-2 over ATM is

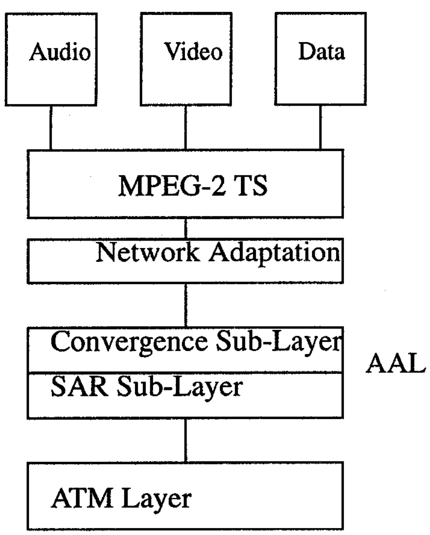


Fig. 1: Protocol Reference Model

given in Fig. 1. As shown, the ATM part of the stack consists of the ATM Adaptation Layer (AAL) and the ATM layer. The function of the AAL is to isolate the higher layers from the specific characteristics of the ATM layer, by mapping the higher layer PDUs into the information field of the ATM cell. The two main candidate AALs for the transport of MPEG-2 TS packets were AAL1 and AAL5.

AAL1 was designed to carry circuit mode, CBR type traffic over ATM. The two main services that AAL1 provides to the upper layers are source clock recovery and Forward Error Correction (FEC) at the cell level. By means of these services AAL1 provides a time-stamped, connection oriented, synchronized session between end systems.

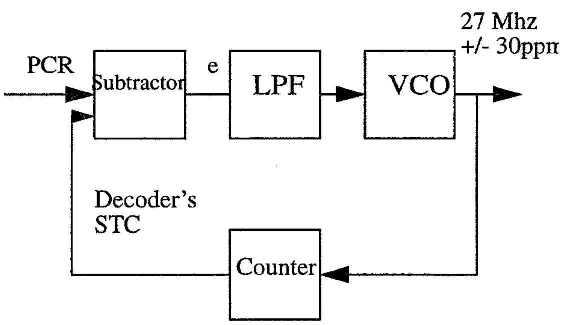
AAL5 was initially introduced as an efficient AAL for carrying data traffic. The main service it provides is detection of corrupted Convergence Sub-Layer PDUs (by means of CRC-32). Even though AAL1 by virtue of its synchronization mechanisms, seemed better suited for transporting video, the ATM Forum decided to use AAL5 for encapsulating MPEG-2 TS packets for the following reasons:

- Chipsets implementing AAL5 are most common among ATM vendors today, hence also the cheapest. Since AAL5 is also used for signalling, it is supported in most end-systems.
- AAL1 cannot be used for carrying VBR MPEG-2 TS packets.
- The AAL1 SRTS technique of source clock recovery synchronizes the source and destination clocks, by tying it to the common network clock. This has several problems: If the source and destination do not share the same network clock, then it won't work. Also if the MPEG-2 TS stream is being played from a disk, then the system clock is not physically present at the source, which makes SRTS impossible to use.
- The AAL1 Adaptive Buffering technique for clock recovery was thought to be un-economical since the decoder would require a PLL to recover the source clock and another PLL to recover the MPEG-2 system clock (assuming the two were locked). A single PLL solution was favored because of its lower cost.

An important consequence of the decision to use AAL5 is that the MPEG-2 system clock recovery becomes very expensive if there is a lot of jitter in the network (we will cover aspect in detail in Section 4). Since VBR MPEG-2 streams would tend to be jittered much more than CBR MPEG-2 streams, it was decided that Phase 1 of the Forum implementation agreement should focus on CBR MPEG-2 TS streams, for which it could be assumed that the network jitter is less than 1 ms.

4.0 Network Jitter and Clock Recovery

MPEG provides a precise mechanism for correctly timing the decoding and display of video and audio at the decoder. The System Time Clock (STC) that is used at the time of encoding has to be re-generated at the decoder, so that video and audio sample clocks can be derived from it. This control is provided by inserting time stamps, called Program Clock References (PCR) in the bit stream at the rate of at least 10 per second. The PCRs are generated at the encoder by sampling the STC, which runs at 27 Mhz ±30 ppm, and inserting the result into a special field in the TS packet. Since a decoder's free running system clock frequency will not match the encoder's STC, the decoder can be made to slave its timing by using the received PCRs. The typical technique for carrying this out is by means of a Phased Locked Loop (PLL) (Fig. 2).



LPF: Low Pass Filter
VCO: Voltage Controlled Oscillator

Fig. 2: Decoder PLL used for source clock recovery

When the Transport stream is first received at the decoder, its STC counter is set to the received PCR, and the PLL is subsequently operated as a closed loop. Subsequently the error term "e" reflects the difference between the received PCR and the decoder's STC. This difference is input into a LPF whose output control's the instantaneous frequency of the VCO. There are two important cases to consider for analyzing the behavior of this clock recovery technique:

Consider the case when the network delay is fixed, as in the case of a circuit switched network. In this case, after the PLL is started, there will be an initial transient since the encoder and decoder clocks can be offset from each other by upto 60 ppm. For example if the encoder clock is running faster and the decoder clock is running slower, then during the transient phase, the decoder buffers will begin to fill up. However, the two clocks soon converge to a common value as the loop goes into the locked state, and in this case the value of the error term "e" becomes zero. The time required for the loop to lock is controlled by the frequency response and gain of the LPF. For example a filter with a smaller cut-off frequency will lead to a slower lock.

If the network delay is not fixed, as is the case for an ATM network, then the variations tend to cause a difference between the values of the PCRs and the value they should have when they are actually received. This is referred to as PCR jitter. For example if the delay in delivering one PCR is greater than the delay experienced by other PCRs, then the decoder PLL would attribute this to a decrease in the encoder clock frequency, and correspondingly try to decrease its own frequency. Hence effectively the network jitter adds a second order noise term to the variations in source clock frequency, and the challenge is to detect this term and separate it from the slower longer term variations in source clock frequency.

One of the proposals to solve this problem, was to add a Service Specific Convergence Sub-Layer (SSCS) called the Video-Audio SSCS (VASSCS) to the Convergence Sub-Layer of AAL5. The function of the VASSCS would be to time-stamp AAL5 PDUs at the transmitter (perhaps with the same STC), and use a de-jittering buffer at the AAL layer of the receiver to make sure that the decoder does not see any network jitter. Hence the function of the de-jittering buffer would be to make sure that the sum of the network delay plus the buffer delay is fixed.

While this proposal was being debated, Perkins and Skelly [1], showed that under the assumption that the network jitter is limited to a maximum of 1 ms, the PLL scheme in Fig. 2 is still capable of performing source clock recovery, with the added expense of a modest amount of buffering at the decoder (about 50 KBytes), so that a VASSCS is not needed. Subsequently there were several studies of jitter induced in CBR streams in ATM networks [2], [3], and they concluded that the 1 ms figure was quite reasonable. Hence Phase 1 of the ATM Forum Implementation Agreement for MPEG-2 over ATM does not define a VASSCS, but specifies that the network jitter should not exceed 1 ms. One way of meeting this requirement would be negotiate a jitter contract of less than 1 ms with the network. Since most ATM networks today are not sophisticated enough to do this, an alternative technique is to insert a simple adaptive de-jittering mechanism in front of the PLL [4], which has been shown to work quite effectively. These schemes rely on the CBR nature of the TS stream, and will not work for VBR TS streams.

The technique proposed in [1] for clock recovery in the presence of jitter, increases the order of the LPF, until the variation in frequency of the decoder clock from the encoder's clock (see Amplitude in Fig. 3), is small enough to satisfy performance requirements. If the decoder clock is used to directly synthesize a chroma sub-carrier for use in NTSC video, then the frequency deviation should not exceed 10 Hz, which is quite stringent. In other cases the decoder's clock is only used to drive the pixel clock, and if the user is willing to tolerate an occasional frame slip or repetition, then a much larger variation in frequency can be tolerated.

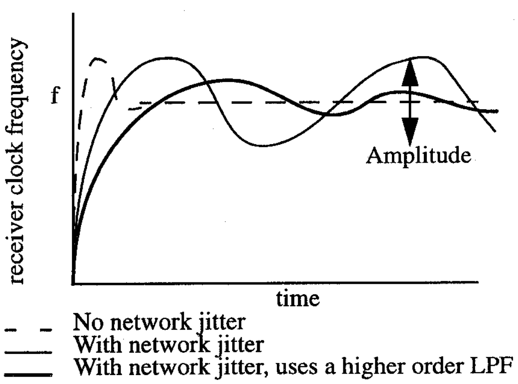


Fig. 3: Sample frequency adjustment behaviors

Fig. 3 illustrates some of the points we have made. It asumes that source clock frequence is fixed at f. In the absence of network jitter, there is an initial transient period after which the receiver clock locks on to the source clock and stays locked. With network jitter, there are variations in receiver clock frequency from that of the source, which can be reduced by increasing the order of the LPF. But as shown, this is done at the expense of increasing the time of the initial transient phase. The amount of buffering required depends directly on the duration of the transient phase, since during this time the sender and receiver clocks are running at different frequencies.

5.0 TS Packet Encapsulation

Once the decision was made to use AAL5 and focus on CBR TS streams, the next major design decision that had to be made was the set of rules to be used for encapsulating TS packets in AAL5 SDUs. The following issues are relevant to this problem:

From the point of view sending MPEG-2 TS streams across a computer data bus, it is preferable to have large AAL5 PDUs. This is due to the fact that usually every packet transfer results in an interrupt at the CPU, and very high rate of interrupts can cripple the system. Hence computer companies at the ATM Forum were in favor of packing a large number of TS packets in each AAL5 SDU. On the other hand companies manufacturing set-top boxes were not much concerned about large AAL5 PDUs since their hardware could incorporate proprietary mechanisms which are tailor-made for transferring short packets.

The problem with packing too many TS packets in a single AAL5 SDU is that in case one of the ATM cells from that SDU is dropped, then the entire SDU is corrupted (this can be detected by means of the CRC field in the AAL5 trailer). If it contains data from an MPEG-2 I frame, then this error has the potential of propagating for several frames into the future. Also larger AAL5 SDUs means that the ATM adaptor at the receiver needs to have a larger buffer to hold the SDU. Hence from these points of view, not more than one or two TS packets should be packed into an AAL5 SDU.

Another aspect of packing multiple TS packets into an AAL5 SDU is that of additional jitter that may be introduced into PCR bearing TS packets due to the following reason: Since the ATM adaptor at the receiver waits until all TS packets in an AAL5 PDU have arrived before passing them on to the decoder, if the PCR bearing packet is not the last packet in the PDU, then it will have to wait for an additional amount of time in the adaptor. Since this time will vary from one PCR packet to another (since their position within the PDU is not fixed), it will manifest itself as jitter. Larger the number of TS packets that are packed, larger is the magnitude of this jitter. For example if the CBR bit rate is 1.5 mbps and two TS packets are encapsulated in each AAL5 SDU, then the magnitude of this packing jitter is 1 ms. Such a large number was of serious concern, especially to set-top box manufacturers who wanted to keep the amount of memory required for de-jittering to a minimum to keep the cost down.

As a compromise on these conflicting requirements, the ATM Forum initially came up with the following rule for TS packet encapsulation: *The default number of TS packets per AAL5 SDU is two, but by means of SVC signalling, the end- systems are free to negotiate a larger value. Moreover a PCR bearing TS packet should always be the last packet in its SDU.*

This rule shifted the burden of taking care of the jitter situation from the receiver to the sender, since now the ATM adaptor at the sender has to be MPEG-2 aware, and make sure that if a TS packet has a PCR in it, then it is the last packet in the AAL5 SDU. When this rule was implemented by MPEG-2 server manufacturers, they observed a significant decrease in the effective throughput, and consequently in the number of MPEG-2 TS streams that the server could support simultaneously [5]. This had an impact on the cost of the MPEG-2 video content, and the server vendors successfully argued that the increase in the recurring costs of playing the content was much more than the fixed cost of additional memory in the set-top box to offset the packing jitter. Their case was buttressed by the fact that the amount of additional buffering to offset the packing jitter at the receiver was quite small. Consequently the ATM Forum modified the packing rule during its October'95 meeting, and removed the PCR awareness requirement.

6.0 Support for VBR MPEG-2

One of the more beneficial aspects of ATM over STM transport is the statistical multiplexing gain and the resulting saving in bandwidth when VBR streams are used. In order to exploit this advantage, several companies are exploring the support for VBR MPEG-2 over ATM [6].

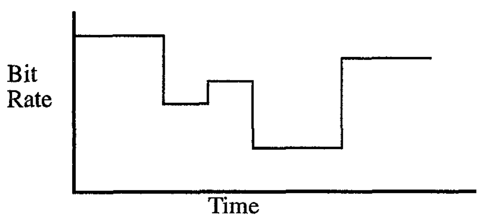


Fig. 4: VBR MPEG-2 TS Stream

As shown in Fig. 4, VBR MPEG-2 streams are in fact piecewise constant, since the bit rate is not allowed to change in-between PCRs. The main problem in supporting VBR streams is that of clock recovery in the presence of network jitter. This problem is excarbated by the fact that network jitter for VBR streams is much larger as compared to that for CBR streams. Also simple adaptive schemes that rely on buffer fillness to remove jitter are no longer applicable, since it is impossible to distinguish the change in buffer fill due to bit rate variability from that of network jitter. It was in light of these problems that Phase 1 of the ATM Forum MPEG-2 over ATM Specification focused in CBR streams only. The following is a summary of some of the proposals that have been made to solve this problem.

The simplest approach for handling the additional jitter in VBR streams is to seek explicit jitter guarantees from the network at the time when the connection is set up [8]. For example if the jitter guarantee is 1 ms or less, then the same techniques as that for CBR (see Section 4) can be used for VBR clock recovery. However most ATM networks today are not sophisticated enough to offer explicit jitter guarantees. Also it is not clear whether the additional network bandwidth required to provide such guarantees would be cost justified.

Another proposal is to add a VASSCS to the AAL5 CS, whose function would be to enable real-time communications over AAL5 [7]. Hence this VASSCS could be used by other real time services in addition to MPEG-2. The main function provided by it includes jitter removal and time base recovery. It would carry out its function by means of a 32 bit time stamp, added by the sender. The receiver would use this timestamp in conjunction with a holding buffer to de-jitter AAL5 PDUs and perform timebase recovery. The main problem with this proposal is that VASSCS based adaptors would have to gain substantial market acceptance, before they are comparable in price to plain AAL5 adaptors.

A third approach for VBR MPEG-2 streams tries to use an un-modified AAL5 layer as for CBR streams, by exploiting the piecewise constant nature of a VBR MPEG-2 TS stream [6]. Over each constant bit rate segment, the CDV can be filtered exactly as for a true CBR stream. At the time of transition from one rate to another, the proposal is to insert a special packet called the Rate Change Indicator (RCI) into the TS stream. The RCI would carry an estimate of the new rate to the receiver, which would use it to adjust its filtering algorithm.

References

[1] M. Perkins and P. Skelly, "A hardware MPEG clock recovery experiment in the presence of ATM jitter," ATM Forum Contribution ATM94-0434, May 1994.

[2] S.I.A. Shah,"Cell delay variation in CBR connections," ATM Forum Contribution ATM 94-0417R2, July 1994.

[3] M. Grossglauser and S. Keshav,"Performance of CBR traffic in WANs," Submitted to INFOCOM'96.

[4] P. Hodgins and E. Itakura,"The issue of transportation of MPEG over ATM," ATM Forum Contribution ATM94-0570, July 1994.

[5] J. Brenneman et. al.,"Comments on MPEG-2/AAL5 encapsulation," ATM Forum Contribution ATM95-1265, October 1995.

[6] P. Hodgins and E. Itakura,"VBR MPEG-2 over AAL5," ATM Forum Contribution ATM94-1052, December 1994.

[7] T. Brunner and M. Schwartz,"Essential features of a video and audio SSCS for AAL5," ATM Forum Contribution ATM94-0404, May 1994.

[8] D. Verma, D. Ferrari and H. Zhang,"Guaranteeing delay jitter bounds in packet switching networks," Tricomm'91, April, 1991.

[9] ISO/IEC JTC1/SC29/WG11,"MPEG-2 Systems ISO/IEC 13818-1,".

[10] ATM Forum SAA Group,"SAA audio-visual multimedia service implementation agreement," S.A. Wright Editor, ATM95-0012, December 1995.