Module 10

MULTIMEDIA SYNCHRONIZATION

Lesson 36
Packet architectures and audio-video interleaving

Instructional objectives

At the end of this lesson, the students should be able to:

1. Show the packet architecture
2. Define the elements of the packet header
3. State and significance of presentation time stamp (PTS) and decoding time stamps (DTS)
4. Show how audio and video packets can be interleaved
5. State the design considerations of pack size

36.0 Introduction

In the previous lesson, we had discussed pack architecture in details. Packs are composed of packets, which may be audio, video or any other media type. In this lesson, we are going to discuss in details about the packet architecture that is composed of packet header and the audio/ video data in the encoded form. We are also going to present how the audio and the video packets are interleaved.

36.1 Packet architecture

Fig.36.1 presents the composition of packets. Every packet consists of its individual header, followed by data. Packets can be composed of only one media type and multiple media types cannot be mixed in a packet. Packets are of fixed size, typically 188 Bytes. In contrast, the information content of audio and video frames may vary from one frame to the other and hence frame boundaries need not coincide with the packet boundaries. Each packet may include multiple number of frames, some partial and a frame may have to be accommodated in multiple packets, as shown in fig.36.1. Table-36.1 shows the elements of packet and pack headers.

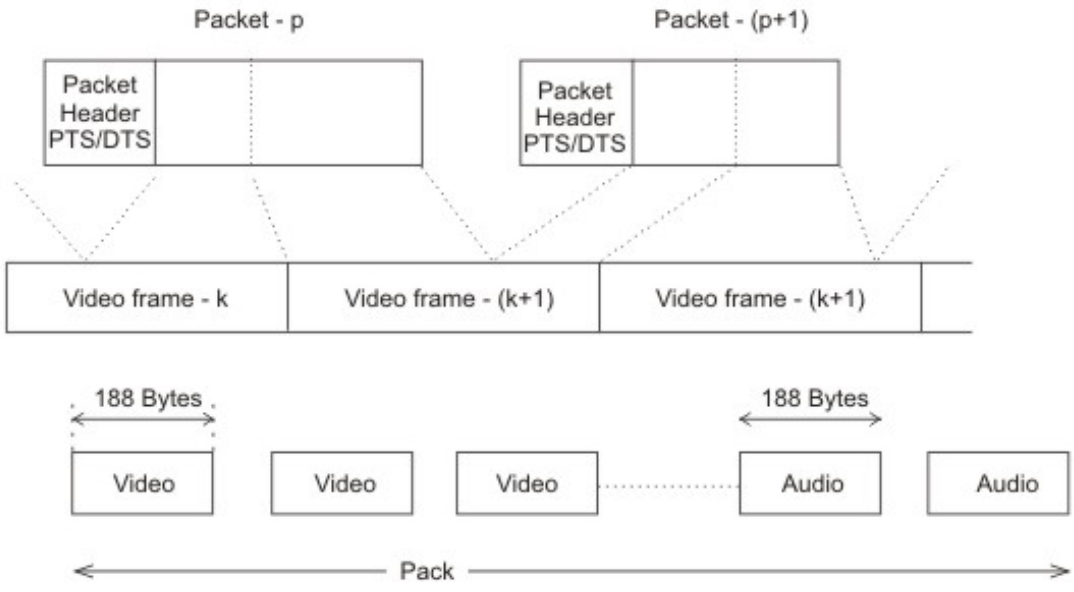

FIGURE 36.1 Composition of a packet

Packet header	Bits	Pack header	Bits
Start code	24	Start code	32
Steam ID	08	-------	
Packet Length	16	SCR	33
STD Buffer scale	01		
STD Buffer size	13	Mux rate	22
PTS	33		
DTS	33		

36.2 Presentation Time stamp (PTS) and Decoding Time Stamp (DTS)

The packet header may contain

a) *A decoding time stamp (DTS)*, defined as the absolute time at which the decoding of the media unit is to be begun

b) *A presentation time stamp (PTS)*, defined as the absolute time at which the first media unit present in the packet is to be presented.

In an idealized decoder, the decoding process is assumed to be instantaneous, and hence, the PTS and DTS may not differ at all.

In real decoder, the difference between PTS and DTS represents the decoding computational overhead. The time stamps are represented as ticks counted on the 90 kHz system lock (modulo 2^{33}) and therefore, can time video playbacks of greater than 24 h in length. In fact, the longest duration of a playback can be

$$\frac{2^{33}}{90 \times 10^3 \times 3600} = 26 \text{ hours } 30 \text{ mins}$$

A PTS need not be provided for each media presentation unit, but the separation between successive PTS's should not exceed 0.7 seconds.

The packet header also contains STD buffer size and scale fields, the product of which defines the max input buffer size needed at the STD for the packet.

The input buffering delay cannot exceed one second

$$\forall \text{ bytes } i \in \text{presentation unit } p : t_d(p) - t_a(p) \leq 1 \text{ second}$$

where, $t_d(p)$ is the decoding time of presentation unit p.
$t_a(i)$ is the arrival time of bytes i.

An encoder must construct its multiplexed stream so that the input buffers at STD neither overflow nor underflow.

ISO 11172 multiplexed stream specifications.

- The stream's buffer requirements at STD do not exceed 46 x 1024 bytes for video and 4096 bytes for audio.
- The stream conforms to the following spaces :
-

$$\text{Packet rate} \leq \begin{cases} 300 \text{ packets/s} & \text{if } mux-rate \leq 10^5 \\ 300 \times \dfrac{mux-rate}{10^5} \text{ packets/s} & \text{otherwise} \end{cases}$$

36.3 Audio and Video interleaving

The data part of a packet contains a variable number of contiguous bytes from an individual media stream. Packets need not start at media presentation unit boundaries.

MPEG Stream Examples

Video: 1.2 Mb/s

Audio: 192 Kb/s (encoded CD quality stereophonic sound)

Length of audio and video packets = 188 bytes (as per MPEG-2)

As Video bit rate: Audio bit rate = 6.25: 1, one audio packet is interleaved with every 6 or 7 video packets to match the bit rate ratio. The interleaving may be achieved as:
One audio packet for every 6 video packets- Repeated thrice

One audio packet for every 7 video packets- Repeated once.

36.4 Choice of pack size

Smaller pack size means frequent transmission of system headers. Too small a pack size would impose unnecessary additional packing overheads. The upper bound on the pack size is the 0.7 seconds of max separation between successive packs.
If the pack size is 8 packets, we have 10 ms separation between successive packs. At this pack size, the pack separation is 10 ms.

Video presentation unit : One frame
Audio presentation unit : 1584 Bytes

For example, the first pack has an SCR equal to 4008. The first video frame has a DTS equal to 22012.

The video decoder setup time = $\frac{22012 - 4008}{90,000}$ seconds = 200 ms.

Thus, the first I picture is decoded 200 ms after arrival. The PTS of the first video frame is 24000, which is extra 20 ms after decoding, but before displaying the first I frame. This represents the time needed to initialize the video display buffer.

36.5 Conclusion

In this lesson, we have presented the composition of a packet and showed how packets of different media may be interleaved. We have also studies the requirements for presentation time stamp and decoding time stamp. Decoding early and presenting later require buffering, which is also necessary for frames reordering in MPEG bit stream. Further, buffering ensures continuity in media playback. This aspect will be discussed in the next lesson.