Mechanisms of MPEG Stream Synchronization

G J Lu, H K Pung and T S Chua
Department of Information Systems and Computer Science
National University of Singapore
Singapore 0511

Abstract

Media synchronization is an important issue in developing multimedia applications. MPEG is an international standard for coding moving pictures and associated audio for multimedia applications. Coded audio, video and other data streams are multiplexed into an MPEG stream. We introduce the syntax of the multiplexed MPEG stream and explain the mechanisms used to maintain media synchronization in a hypothetical model, system target decoder, in which it is assumed that data transfer and decoding are carried out instantaneously. Then we propose extensions to these mechnisms to achieve MPEG stream synchronization in a practical decoder.

1. Introduction

MPEG (Motion Picture Expert Group) is an international standard for coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s, prepared by ISO/IEC JTC 1/SC 29/WG11[1]. The draft standard [1] was published as ISO/IEC 11172. When referring to this standard, MPEG and ISO 11172 are used interchangeably. The standard consists of two parts. Part 1 is divided into three sections. Section 1, Systems, specifies the system coding layer. It defines a multiplexed structure for combining elementary streams, including coded audio, video and other data streams, and specifies means of representing the timing information needed to replay synchronized sequences in real-time. Section 2, Video, specifies the coded representation of video data and the decoding process required to reconstruct pictures. Section 3, Audio, specifies the coded representation of audio data. Part 2 , conformance testing, specifies the procedures for determining the characteristics of coded bit streams and for testing compliance with the requirements stated in Part 1.

MPEG Video is a very important aspect of the ISO 11172 standard and has been discussed by Gall [2]. This paper looks into the MPEG Systems, particularly the aspects related to media synchronization. Since coded MPEG stream consists of continuous media streams, i.e. video and audio, it is necessary to synchronize these media for real-time playback. There are two aspects of continuous media synchronization: intra-medium synchronization and inter-media synchronization. While intra-medium synchronization ensures the continuity for smooth playback of each medium, intermedia synchronization ensures synchronization between associated media. Thus in this paper media synchronization is defined as occurence in which each medium is played out at its fixed rate determined by the type of medium and/or the application concerned and the specified/required temporal relationships among the associated media are maintained.

The ISO 11172 standard specifies syntax and semantics based on a conceptual decoder model, system target decoder (STD). This model assumes that the multiplexed MPEG stream is stored on a constant latency digital storage medium, and data transfer and decoding within the decoder are instantaneous. ISO 11172 stream is designed such that the STD will be able to decode and display elementary streams synchronously.

However, in a practical decoder, data transfer and decoding cannot take place instantaneously. Since ISO 11172 stream is specified based on STD model, it is the responsibility of decoder to compensate for these data transfer and decoding delays to ensure synchronization.

In ISO 11172, the storage medium has a broad meaning, including CD-ROM, magnetic harddisk, digital audio tape and computer networks etc. These storage media are of indeterministic nature in terms of delay and transmission rate instead of constant latency as assumed in STD model. In order to feed the ISO11172 decoder with a stream of constant latency, a medium specific decoder is required. An appropriate buffer must be used in the medium specific decoder to smooth out the transmission jitter in order to provide data to the ISO11172 decoder at the required rate.

The rest of the paper is organized as follows. Section 2 discusses the MPEG Systems in general. Section 3 presents the principle and mechanisms used in the ISO11172 Systems to support media synchronization in STD. Section 4 discusses compensation of decoding delays in practical decoders required in order to maintain synchronization. Section 5 discusses buffer requirement to overcome the bursty nature of digital storage medium.

2. ISO 11172 Stream

While MPEG Video and Audio specify the coding of each individual stream, MPEG Systems specifies the syntax and semantics of information that is necessary in order to reproduce one or more MPEG audio or video compressed data streams in a system. An ISO 11172 stream consists of one or more elementary streams multiplexed together. A elementary stream consists of a number of access units (AU). In the case of compressed audio an access unit is defined as the smallest part of the encoded bitstream which can be decoded by itself. In the case of compressed video an access unit is the coded representation of a picture. A decoded audio access unit or decoded picture is called a presentation unit (PU). In a coded video stream, there are three types of access units: I-pictures, P-pictures and B-pictures. I-pictures are coded without referring to

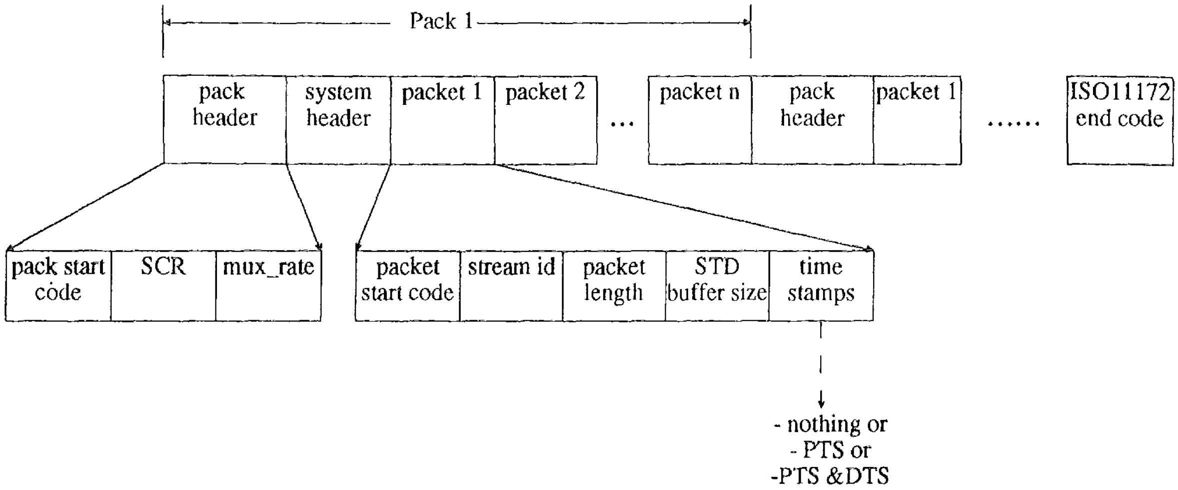

Fig.1 ISO11172 Stream Syntax

other pictures. P-pictures are coded using forward prediction and B-pictures are coded using both forward and backward predictions. Due to the way video stream is coded, picture order in the coded stream may differ from the display order. The decoder must carry out re-ordering if needed [1, 2].

An ISO 11172 stream is organized into two layers: the pack layer and the packet layer. The pack layer is for system operations and the packet layer is for stream specific operations. Fig.1 shows the syntax of an ISO 11172 stream.

An ISO 11172 stream consists of one or more packs. A pack commences with a pack header and is followed by zero or more packets. The pack header begins with a 32-bit start-code. The pack header is used to store system clock reference SCR and bitrate information, mux_rate. The SCR is a 33-bit number, indicating the intended time of arrival of the last byte of the SCR field at the input of the system target decoder. Mux_rate is a positive integer specifying the rate at which the system target decoder receives the ISO 11172 multiplexed stream during the pack in which it is included. The value of mux_rate is measured in units of 50 bytes/second rounded upwards. The value zero is forbidden. The value encoded in mux_rate field may vary from pack to pack in an ISO 11172 multiplexed stream. The mux_rate value together with SCR value defines the arrival time of each byte at the input to the system target decoder.

Data from elementary streams is stored in packets. A packet consists of a packet header followed by packet data. The packet header begins with a 32-bit start code that also identifies the stream to which the packet data belongs. The packet header defines the buffer size required at each elementary decoder for smooth decoding and playback of the elementary stream. The packet header may also contain decoding and/or presentation time-stamps (DTS and PTS) that refer to the first access unit in the packet. The purposes of DTS and PTS are discussed in next section.

The packet data contains a variable number of contiguous bytes from the same elementary stream. A data packet never contains data from more than one elementary stream and byte ordering is preserved. Thus, after removing the packet headers, packet data from all packets with a common stream identifiers are concatenated to recover a single elementary stream. The multiplex of different elementary streams is constructed in such a way (in terms of packet size and the relative placement of packets from different streams) as to ensure that the specified STD buffers do not overflow or underflow.

The system header is a special packet that contains no elementary stream data. Instead it indicates decoding requirements for each of the elementary streams. It indicates a number of limits that apply to the entire ISO 11172 stream, such as data rate, the number of audio and video streams, and the STD buffer size limits for the individual elementary streams. A decoding system may use these limits to establish its ability to play the stream. The system header also indicates whether the stream is encoded for constant rate delivery to the STD. The system header must be the first pack of the ISO 11172 stream. It may be repeated within the stream as often as necessary. Repeat of the system header will facilitate random access. Real-time encoding systems must calculate suitable limits for the values in the header before starting to encode. Non-real-time encoders may make two passes over the data to find suitable values.

Up to 32 ISO 11172 audio and 16 ISO 11172 video streams may be multiplexed simultaneously. Two private data streams of different types are provided. One type is completely private and the other follows the same syntax as audio and video streams. It may contain stuffing bytes, a buffer size field, and PTS and DTS fields. The use of these fields is not specified in ISO11172.

The system specification does not specify the architecture or implementation of encoder or decoders. However, bitstream properties do impose functional and performance requirements on encoders and decoders.

The MPEG Systems coding specification provides data fields and semantic constraints on the data stream to support the necessary system functions. These include the synchronized presentation of decoded information, the management of buffers for coded data, and random access. Random access is made possible by repeated appearance of the information needed to start decoding, such as SCR, PTS and system headers, and use of I-pictures (pictures coded without referring to other pictures). Other functions are all related to the smooth and synchrony playback of coded streams and are discussed in the next section.

3. Synchronization in the System Target Decoder

An ISO 11172 stream is constructed such that elementary streams will be synchronously decoded and presented by the STD and elementary input buffers in STD will never overflow or underflow. This section explains the mechnisms used to achieve this.

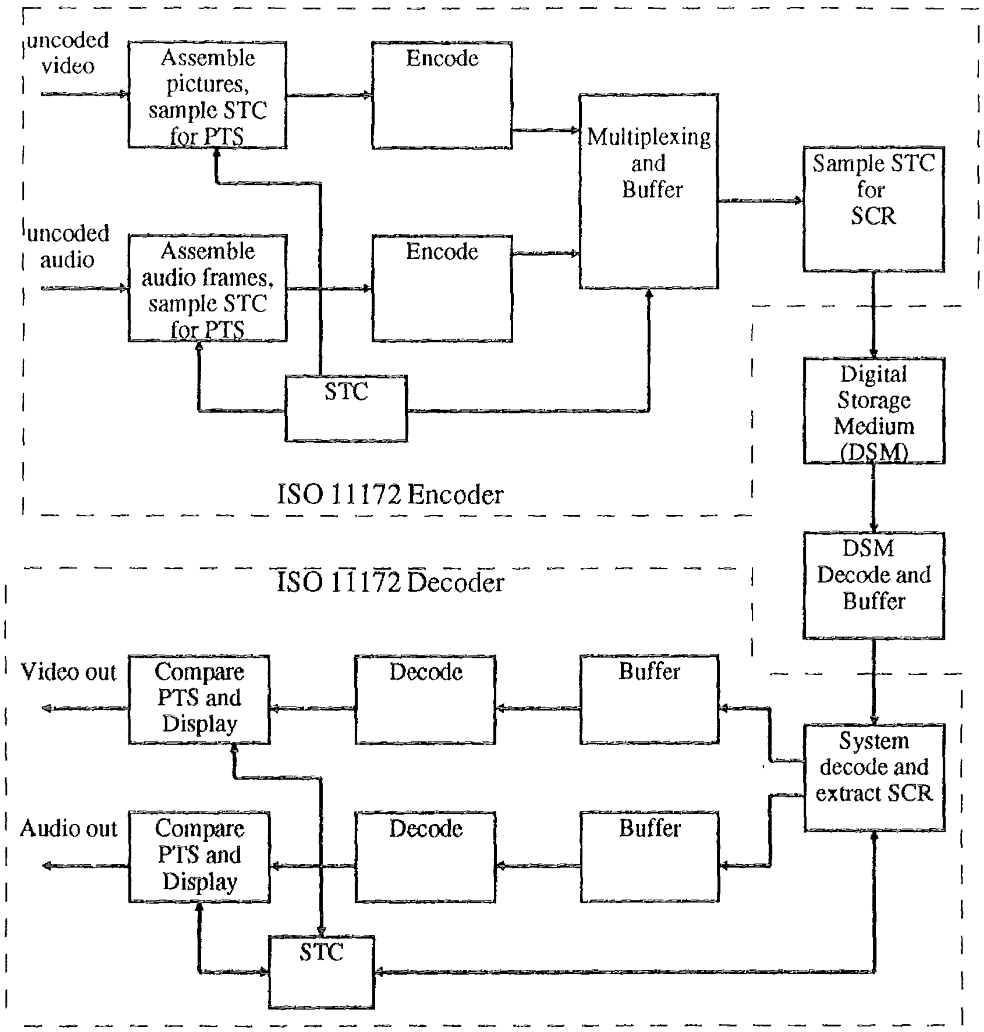

Fig.2 Protypical Encoder and Decoder

In STD, playback of N streams is synchronized by adjusting the playback of all streams to a master time base rather than by adjusting the playback of one stream to match that of another. The master time base may be one of the N decoders' clock, the DSM or the channel clock, or it may be some external clock. The similar synchronization principle has been used in a number of synchronization schemes, such as synchronization marker[3], logical time system [4] and relative time system [5]. In these scheme, each presentation unit has a time stamp and presentation units with the same time stamps are displayed at the same time to achieve synchronization.

MPEG Systems provide for end-to-end synchronization of complete encoding and decoding process. This is achieved by use of time stamps, including system clock reference (SCR), presentation time stamp (PTS), decoding time stamp (DTS). This end-to-end synchronization is illustrated in Fig.2, which includes a protypical encoder (upper part) and a protypical decoder

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.