ITUPublications

International Telecommunication Union
Radiocommunication Sector

Report ITU-R BT.2557-0

(09/2025)

BT Series: Broadcasting service (television)

Application of Smart Media Transport based on MPEG Media Transport under different service requirements

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from https://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at https://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
TF	Time signals and frequency standards emissions

***Note**: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2025

REPORT ITU-R BT.2557-0

Application of smart media transport based on MPEG media transport under different service requirements

(2025)

TABLE OF CONTENTS

Abbreviations/Glossary

ADC Asset delivery characteristic

AU Access unit

CEUs Content encoding units

CMG China Media Group

MMT MPEG Media transport

PA Programme Association

QoS Quality of service

QoE Quality of experience

SMTP Simple mail transfer protocol

1 Overview

This Report describes application scenarios and the general technical approach descriptions to implement these applications based on Recommendation ITU-R BT.2074, which specifies MPEG Media Transport (MMT) and the extension of MMT by Smart Media Transport (SMT) for MMT-based broadcasting systems. Recommendation ITU-R BT.2074 specifies an encapsulation format of media components, delivery protocol, signalling information, media presentation, and adaptive forward error correction mechanism for various applications, including broadcasting applications.

The operators need to consider different network environments and business needs when implementing the applications. The Report provides some reference applications for future operators using Recommendation ITU-R BT.2074.

SMT has become a Chinese national standard for broadcast and broadband audiovisual distribution. China Media Group (CMG) and China Mobile Migu (Migu) deployed the standard during multiple sports events such as the Paris Olympics, Beijing Winter Olympics, and Hangzhou Asian Games to provide high-quality, high-precision, and high flexibility synchronized immersive services through various angles, multiple devices, and cross networks. The experiments demonstrate the potential of SMT in applications such as live sports broadcasting, concerts, shows, and other large-scale events.

2 SMT data reception and presentation

The SMT media playback process involves several stages that allow the receiver terminal to process and render media content. This flow ensures that the correct media is delivered, decoded, and presented at the appropriate time. The process begins with receiving the IP data flow, handling signalling messages, identifying and decoding the media components, and finally rendering them according to the specified presentation time and display configurations. The specific steps are as follows:

1) Receiving IP data flow

   The receiver terminal begins by receiving the IP data flow, which contains IP packets encapsulating Simple Mail Transfer Protocol (SMTP) packets.

2) Processing the Programme Association (PA) Message

   The receiver first processes the signalling messages, starting with the PA Message. SMTP packets with type=0x01 in the SMTP header are identified, as the type field specifies the

payload data type. These packets are parsed to extract the PA Message, from which the PA Table, Mapping (MP) Table, and Layer Display Table are retrieved.

3) Identifying required content encoding units (CEUs)

Using the MP Table, the receiver identifies the IP data flow and the packet_id of SMTP packets carrying the required CEUs. The packet_id, which is included in the SMTP header, is associated with an asset_id. The mapping between packet_id and asset_id is provided by the MP Table, which is part of the signalling messages.

4) Determining Presentation Timing

The receiver retrieves the CEU Timestamp Descriptor from the MP Table. This descriptor specifies the presentation time for the CEUs.

5) Receiving and Decoding Media Components

The receiver collects the SMTP packets identified earlier, which carry media components as MFUs. These MFUs are decoded into their respective media streams.

6) Rendering the Media

Decoded media components are rendered at the specified presentation time based on the CEU Timestamp. The Layer Display Table provides information about each presentation layer and its corresponding media content layer, enabling flexible and personalized display configurations.

3 Application scenarios and key technologies

3.1 "Bai Cheng Qian Ping": Synchronized playback of 8K video on large screens and audio on mobile devices

In CMG's "Bai Cheng Qian Ping" project, ultra-high-definition video signals originating from events such as the Paris Olympics, Beijing Winter Olympics, and Spring Festival Gala are transmitted through various network channels. By utilizing the SMT protocol, precise and synchronized presentation across different platforms and devices is achieved, ensuring the simultaneous display of contents on televisions, computers, mobile devices, and other terminals. This solution significantly enhances the consistency of various angles, multiple devices, and cross networks, providing a diversified user experience with a reliable technical foundation for real-time synchronized playback in complex network environments.

One notable use case in the "Bai Cheng Qian Ping" application is the synchronized presentation of the main video on a large screen. At the same time, the audio is played on personal devices. This combination offers audiences a more immersive experience, allowing them to enjoy high-quality visuals on the big screen while simultaneously hearing relevant audio on their personal devices, which enhances their understanding and appreciation of the content.

Furthermore, the multi-device synchronized presentation model can also be utilized to play the primary perspective on the large screen while secondary perspectives or associated data are displayed on personal devices. Through this approach, viewers can watch the main scene or character on the big screen while switching to alternative angles or accessing related data on their personal devices. This method provides a more comprehensive, various-angle view, helping audiences delve deeper into the content's details and background, offering a richer, more interactive viewing experience.

For broadcasters, such scenarios are suitable for 8K/4K ultra-high-definition programs on outdoor large screens, digital signage, collective viewing and other application scenarios, the common video content is used for large screens, and the audio content that is personalized or not suitable for outdoor play is used for personal receivers, realizing the public service expansion.

FIGURE 1

Multi-device synchronized playback system architecture

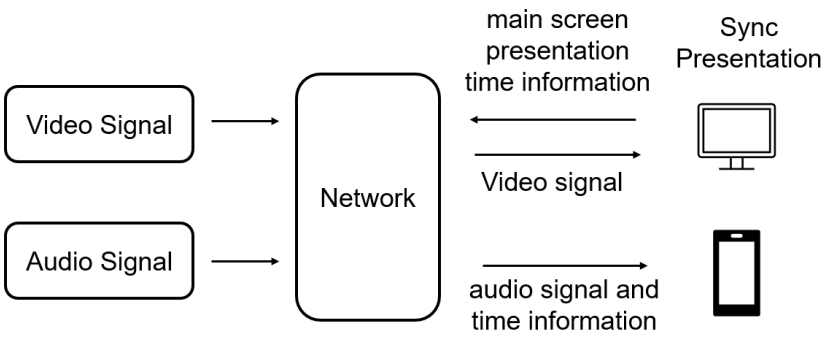
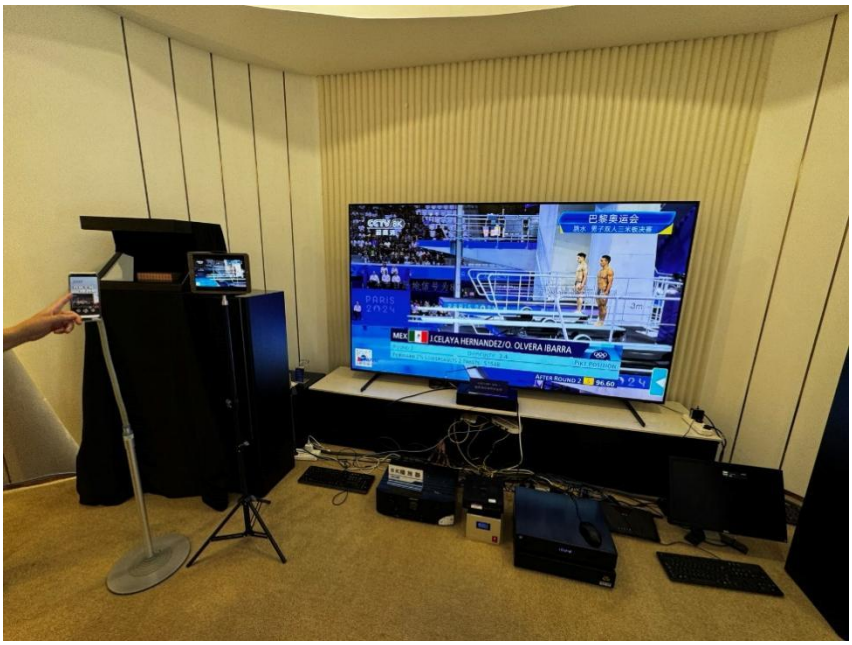


FIGURE 2

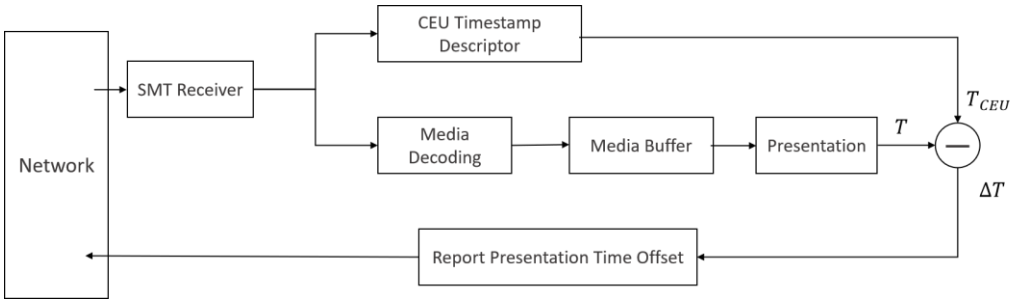
Multi-device synchronized playback demo



SMT employs UTC timestamps to achieve synchronization across different data packets. The SMT protocol provides a CEU Timestamp Descriptor (CEU_timestamp_descriptor), specifying the first Access Unit (AU) presentation time in the corresponding CEU.

FIGURE 3

Large screens presentation in the "Bai Cheng Qian Ping" workflow



Large screens are distributed across different regions, leading to significant differences in transmission delays. These regional variances in latency can affect the synchronization of content, especially when it is being broadcasted across multiple locations simultaneously. To ensure seamless playback and consistent user experience, it is essential to account for these delay differences and implement mechanisms for adjusting synchronization across various screens in real time. Because of this issue, each large screen determines its final presentation time locally. This ensures that the presentation time reflects the specific conditions of media stream reception, decoding, and buffering at each location. Each large screen calculates its presentation time offset and reports the result to the control platform. The presentation time offset is calculated as follows.

$$\Delta T = T - T_{CEU}$$

Here ΔT represents the offset value, T is the actual presentation time of a large screen.T_{CEU} is the presentation time specified in the CEU Timestamp Descriptor from the SMT signalling.

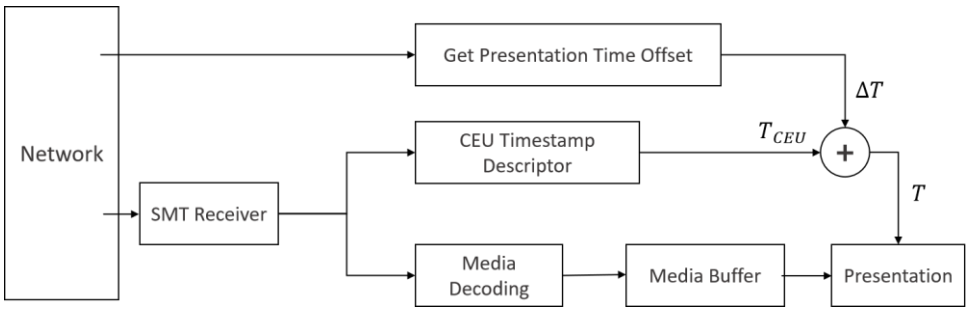
The offset value ΔT is used to align the presentation times across different devices. The actual presentation time T is determined locally by each large screen, meaning that the presentation time offset between different large screens can vary.

For personal receivers, their media playback must be synchronized with the large screen. Therefore, their presentation time is calculated by adding the CEU Timestamp Descriptor's time value T_{CEU} to the presentation time offset ΔT. This ensures that personal receivers are synchronized with the corresponding large screen for consistent and synchronized playback.

$$T = T_{CEU} + \Delta T$$

FIGURE 4

Personal receivers presentation in the "Bai Cheng Qian Ping" workflow



3.2 Multi-source media synchronized presentation

For the Hangzhou Asian Games, China Mobile Migu provided live coverage using multi-view cameras capturing highlight signals in 1080P HD quality. The service offered multi-view synchronized playback, allowing users to select their preferred camera angles for an enhanced viewing experience. Utilizing the SMT protocol, precise synchronized playback of multi-view videos was achieved, greatly improving the viewing experience for the audience. This demonstration showcases the high precision and flexibility of the SMT standard in multi-source synchronization control and offers an innovative solution for multi-source interaction in scenarios such as sports event broadcasting and live concerts.

For broadcasters, such scenarios are suitable for sports and artistic performance programs that are collected and produced by multiple cameras. The same sports competition or artistic performance scene is captured synchronously by the camera to achieve multi-dimensional live broadcast and all-round viewing experience, which can give users more abundant games and performance information. This application will give users a more immersive viewing experience and personalized perspective selection.

FIGURE 5

Multi-view video transmission architecture

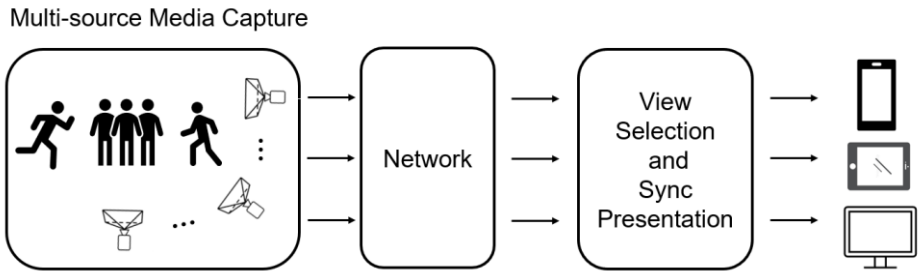


FIGURE 6

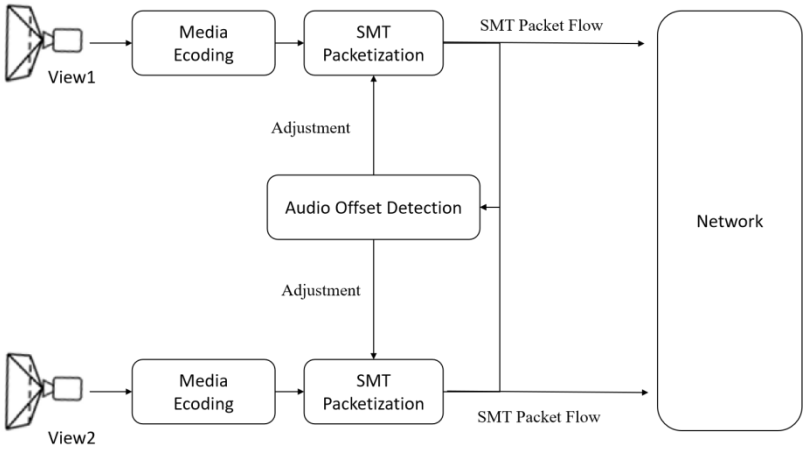
Multi-view video demo

In the Multi-view Video Demo, each view consists of two assets (audio and video), and each asset is made up of a series of CEUs. Each asset includes asset location and a set of asset descriptors. The general location information of an asset defines the data type carried in the data stream and the transmission method, such as unidirectional broadcast networks, the internet, and corresponding port numbers. In the asset descriptors, the CEU Timestamp Descriptor indicates the sequence number of the CEU corresponding to the descriptor and specifies the presentation time of the first AU of the designated CEU. The Available Time (AT) descriptor specifies the accessibility time of the current media resource. These CEUs are synchronized based on the CEU Timestamp Descriptor, ensuring accurate presentation timing when the media contents are displayed on the terminal.

During live broadcasts, the synchronization of each CEU's presentation time can be adjusted according to the current network conditions. The SMT sender entity can calculate the media resource's accessible time and communicate this information to the receiver. This ensures optimal receiver buffer settings while maintaining transmission quality, thereby enabling a dynamic time window mechanism. Network condition information is transmitted through Asset Delivery Characteristic (ADC) messages, which include Quality of Service (QoS) requirements related to media resource transmission and associated Quality of Experience (QoE) information. To provide more accurate information, the SMT sender entity can update the parameters in the ADC message and send it periodically or as needed.

FIGURE 7

Dynamic synchronization correction workflow



If synchronization deviations occur in the video from different camera angles or if there are synchronization issues with the view during the live broadcast, the system allows for the correction of the presentation times for subsequent CEUs to align the content properly. In this demo, synchronization deviations between different views are detected using audio offset detection. By identifying these deviations, the system can adjust the synchronization (CEU timestamp descriptor) for subsequent CEUs, ensuring that all views remain in synchronization.

This approach enables dynamic synchronization corrections, ensuring that different viewpoints and media streams stay perfectly aligned throughout the live broadcast.

3.3 Multi-variant media synchronized presentation

Based on CMG's World Cup broadcast signals, advanced AI technology is used to track the real-time position of players on the large screen, intelligently cropping the player-specific area and adapting it to the mobile screen size. Through the SMT protocol, video contents for both large and small screens are transmitted separately, achieving seamless integration and synchronized playback. This demonstrates the advantages of the SMT protocol in multi-version content synchronization, offering an innovative solution for seamless transitions between different screen sizes.

For broadcasters, such scenes are suitable for achieving cost-effective UHD live broadcast scenes. The main game scene adopts UHD shooting, tracks and cuts UHD images through AI technology, generates customized shots of players or actors in big games or performance scenes, and meets users' application scenarios for close-up real-time tracking of the performance of designated players and actors. To provide users with more rich and personalized competition details.

FIGURE 8

Multi-variant media creation and transmission

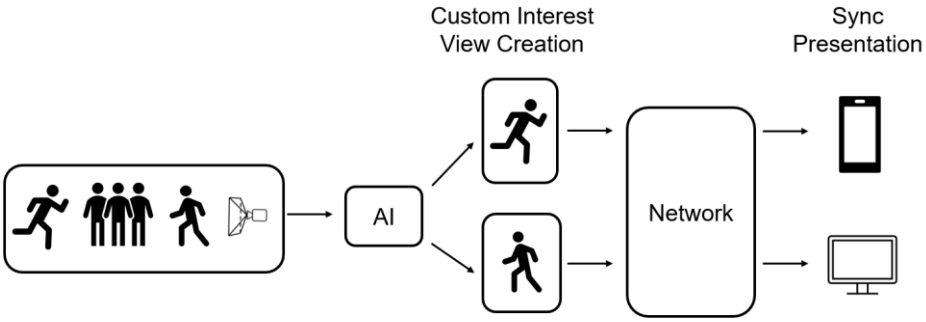


FIGURE 9

Full-field view and single-person view

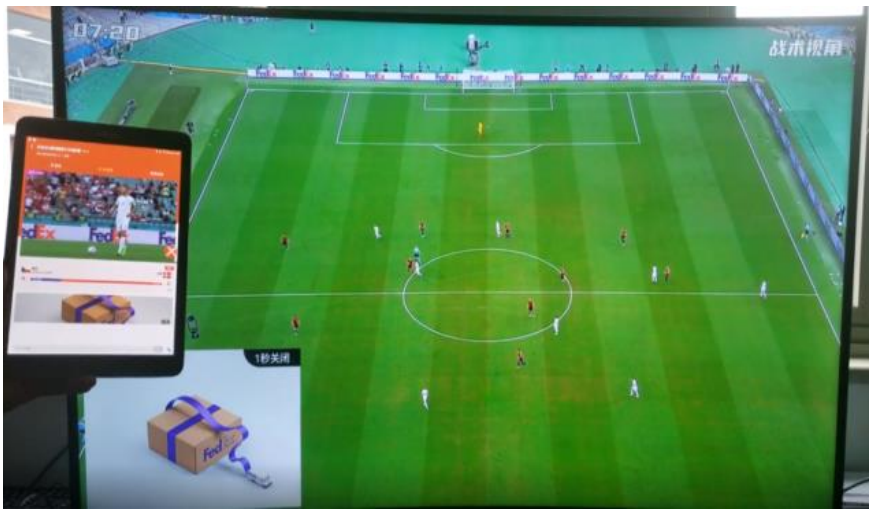
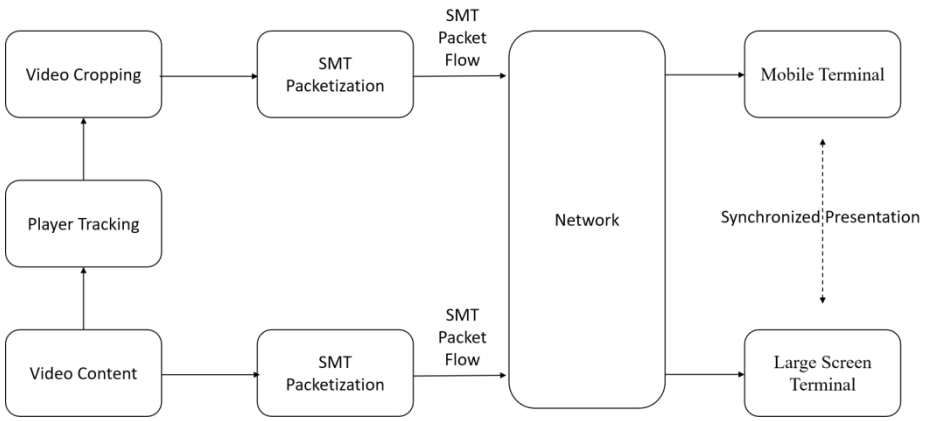


FIGURE 10

Multi-variant media creation and transmission workflow



This approach enables intelligent content recognition across multiple media sources, facilitating dynamic content structuring and association. Video content is recognized through player tracking, and then video cropping is used to extract specific segments. The result is the creation of customized shots optimized for multi-network collaborative distribution and multi-terminal presentation. The generated videos are encapsulated in CEUs, with metadata such as position and synchronization time

embedded into SMT signalling. These CEUs and SMT signalling are then packaged into an SMT Packet Flow, ready for transmission to terminals. Different terminals request the corresponding SMT Packet Flow using the Asset Relationship Information Descriptor, which describes relationships such as dependency, combination, substitution, and supplementation among created media resources, ensuring synchronized and customized presentation across all devices.

These approaches ensure the seamless creation, organization and delivery of content across various platforms and devices, providing efficient and synchronized multimedia experiences.