

世界超高清视频产业联盟标准

T/UWA 012.10—2024

“百城千屏”超高清视音频传播系统视音频 编码：系统(SMT)

Coding of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”:
system(SMT)

(V1.0)

2024 - 11 - 04 发布

2024 - 11 - 04 实施

目 次

前 言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 传输流作为 SMT 的负载传输	2
5.1 概述	2
5.2 AVS2 视音频 SMT 传输	2
5.3 AVS3 音频 SMT 传输	2
5.4 AVS3 视频 SMT 传输	6
5.5 融合网络的切片封装与呈现	6
附录 A（规范性） 融合网络的切片封装与呈现	7

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由世界超高清视频产业联盟提出并归口。

本文件起草单位：上海数字电视国家工程研究中心有限公司，中央广播电视总台，上海交通大学，广播电视总局广播电视规划院，北京中视广信科技有限公司，中国移动通信集团有限公司，咪咕文化科技有限公司，杭州当虹科技股份有限公司，广东图盛超高清创新中心有限公司，海信视像科技股份有限公司，工业和信息化部电子第五研究所，北京数码视讯科技股份有限公司，京东方科技集团股份有限公司，北京流金岁月传媒科技股份有限公司，山东浪潮超高清智能科技有限公司，广州博冠智能科技有限公司，深圳市洲明科技股份有限公司，中联超清（北京）科技有限公司。

本文件主要起草人：徐异凌，王振中，张文军，管云峰，顾军，孙剑，高杨，常江，王一帆，杨开发，侯朴玥，金学骥，姚天赐，李英斌，麻书城，徐翔，马健，张仁宇，陈宇，陈丽丽，郭佩佩，叶建华，陈刚，于路，陈益军，韦胜钰，周骋，夏涛，曾泽君，董磊，房兰涛，曾德祥，李永杰，王付生。

“百城千屏”超高清视音频传播系统视音频编码：系统（SMT）

1 范围

本文件规定了“百城千屏”超高清视音频传播系统中编码的视频基本流、音频基本流在T/AI 114—2021定义的传输流中复用传输的要求，并规定了传输流在IP网络中的传输要求。

本文件适用于“百城千屏”超高清视音频传播系统中的视音频编码码流的复用与传输。本标准与《“百城千屏”超高清视音频编码：系统》都为“百城千屏”超高清视音频传播系统视音频编码码流的复用与传输标准，在系统实现时，二者只需实现其中一种。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB/T 20090.2—2013 信息技术 先进音视频编码 第2部分：视频
- GB/T 33475.2—2024 信息技术 高效多媒体编码 第2部分：视频
- T/AI 109.2—2021 信息技术 智能媒体编码 第2部分：视频
- GB/T 33475.6-2024 信息技术 高效多媒体编码 第6部分：智能媒体传输
- T/AI 109.6-2022 信息技术 智能媒体编码 第6部分：智能媒体格式
- T/UWA 012.2—2023 “百城千屏”超高清视音频传播系统视音频编码：系统
- ISO/IEC 14496-12 信息技术 音视频对象的编码 第12部分：ISO基本媒体文件格式（Information technology - Coding of audio-visual objects - Part 12: ISO base media file format）

3 术语和定义

下列术语和定义适用于本文件。

3.1

知识图像 library picture

T/AI 109.2中规定的知识位流中的图像，可以被其他位流中的图像参考

3.2

知识位流 library stream

T/AI 109.2中规定的包含知识图像的位流

3.3

主位流 sequence stream

T/AI 109.2中规定的可参考由该位流之外的信息提供的知识图像进行解码的位流

4 缩略语

下列缩略语适用于本文件。

AU 访问单元（Access Unit）

AVS2 信息技术 高效音视频编码 第2部分: 视频(Information Technology—High Efficiency Media Coding Part2: Video)

AVS3 信息技术 智能媒体编码 第2部分: 视频 (Information Technology Intelligent Media Coding Part2: Video)

bslbf 比特串, 即二进位串, 左位在先。(Bit string, left bit first, where "left" is the order in which bit strings are written in the specification.)

CEU 通用封装单元 (Common Encapsulation Unit)

HOA 高阶立体声场 (High Order Ambisonics)

PA 包访问 (Package Access)

SMT 智能媒体传输 (Smart Media Transport)

SMTp 智能媒体传输协议 (Smart Media Transport Protocol)

uimsbf 无符号整数, 最高有效位在先 (Unsigned integer, most significant bit first)

UTC 世界标准时间 (Coordinated Universal Time)

ISOBMFF ISO基媒体文件格式 (ISO Base Media File Format)

5 传输流作为 SMT 的负载传输

5.1 概述

SMT (智能媒体传输) 是应用层协议, 定义了涵盖媒体内容封装格式、传输协议及相关控制的智能媒体传输技术系统框架, 将包括多媒体内容组件的内在关系与时间同步关系进行统一封装, 实现高效的多元内容组织和传输。为适应在异构网络环境下媒体内容的分级、可靠传输, 采用媒体的流式传输与文件化兼容传输的方式, 自适应地对特定媒体内容进行差异化组织和个性化传输。针对伴音移动端同步呈现难的问题, 采用以UTC 时间戳为主, 兼容传统编码器时钟的方式, 建立全网时钟机制, 实现不同数据包之间的同步。

SMT协议的详细信息请遵循中关村视听产业技术创新联盟发布的T/AI 114—2021《信息技术 高效多媒体编码 第 6 部分: 智能媒体传输》, 相关内容见附录A。

SMT协议利用ISOBMFF标准来封装和组织媒体内容。根据ISOBMFF的规范, 媒体数据被分割为多个独立的盒子 (box), 每个盒子可以包含特定类型的媒体数据。

5.2 AVS2 视音频 SMT 传输

根据ISOBMFF标准将多种视音频编码封装成可用于SMT协议传输的媒体资源 (asset), SMT中使用asset_type来表示媒体资源的类型。该类型在MP4REG¹的四字符编码 (“4CC”) 类型中得到描述。其中AVS2视频使用 “avst”进行标识, AVS2音频使用 “cacs” 进行标识。

5.3 AVS3 音频 SMT 传输

5.3.1 AVS3 音频媒体资源描述符

5.3.1.1 定义

AVS3音频媒体资源描述符用于指示Audio Vivid编码位流的编码类别、编码档次、存储模式等信息。AVS3音频媒体资源描述符在SMT的MP表中进行扩展, 用于解决Audio Vivid在SMT协议下灵活传输与个性化消费的需求。使用 “av3a” 进行标识。

¹ <http://www.mp4ra.org>

5.3.1.2 语法

AVS3 音频媒体资源描述符语法见表 1。

表 1 AVS3 音频媒体资源描述符

语法	值	比特数	备注
Audio_info_descriptor() {			
descriptor_tag		16	uimsbf
descriptor_length		16	uimsbf
audio_format_type		1	uimsbf
audio_codec_id		4	uimsbf
coding_profile		3	uimsbf
average_bitrate_flag		1	bslbf
hoa_order_flag		1	bslbf
channel_number_flag		1	bslbf
object_info_flag		1	bslbf
reserved	'1111'	4	uimsbf
if(average_bitrate_flag==1)			
average_bitrate		16	uimsbf
else {			
max_bitrate		16	uimsbf
min_bitrate		16	uimsbf
}			
if(hoa_order_flag){			
max_hoa_order		8	uimsbf
}			
if(channel_number_flag){			
max_channel_number		8	uimsbf
}			
if(object_info_flag){			
max_object_channel_number		8	uimsbf
}			
bit_depth_resolution		8	uimsbf
sample_rate		24	uimsbf
}			

5.3.1.3 语义

descriptor_tag用于标识descriptor的类型。

descriptor_length指示标识符的长度，单位为字节。

audio_format_type指示AVS3音频编码位流的类别。该字段取值为0表示位流为Audio Vivid AASF存储格式的位流；该字段取值为1表示位流为Audio Vivid AATF传输格式的位流。

audio_codec_id指示音频媒体资源的编码类别。该字段取值为0时表示媒体资源为通用高码率音频编码数据；取值为1表示媒体资源为无损音频编码数据；该字段取值为2表示媒体资源为通用全码率音频编码数据；其余取值保留。

coding_profile指示音频媒体资源的编解码档次。该字段取值为0表示音频媒体资源的编解码遵循基本框架；该字段取值为1表示音频媒体资源的编解码遵循对象元数据编码框架；该字段取值为2表示音频媒体资源的编解码遵循HOA数据编码框架。

average_bitrate_flag取值为0时表示音频媒体资源不具备平均码率；取值为1时表示音频媒体资源具备平均码率。

max_bitrate、average_bitrate、min_bitrate分别指示音频媒体资源的最大码率、平均码率、最小码率，以kbps为单位。

hoa_order_flag取值为1时表示当前描述符中指示HOA阶数；取值为0时表示当前描述符中不指示HOA阶数。

channel_number_flag取值为1时表示当前描述符中指示声道数；取值为0时表示当前描述符中不指示声道数。

object_info_flag取值为1时表示当前描述符中指示声音对象信息；取值为0时表示当前描述符中不指示声音对象信息。

max_hoa_order指示当前媒体资源支持的最大HOA阶数。

max_channel_number指示当前媒体资源支持的最大声道数。

max_object_channel_number指示当前媒体资源包含的全部对象支持的最大声道数量。

bit_depth_resolution指示音频输入信号的量化比特数。

sample_rate指示音频输入信号的采样频率。

5.3.2 交互反馈信令表

5.3.2.1 定义

交互反馈消息提供沉浸式媒体消费时，服务器与客户端之间的交互反馈。当沉浸式媒体消费中的服务器与客户端之间需要发送交互反馈信息时，使用此消息进行会话。一个交互反馈消息信令中可包含一个或多个交互反馈信令表。交互反馈信令表中包含了服务器和客户端之间交互反馈的信息，不同类型的交互反馈信令表用于指示不同类型的交互反馈信息。

对于Audio Vivid编码位流的媒体资源，若其包含可交互的声音对象，则用户对于声音对象的交互操作可以通过交互反馈信令表进行反馈，其中声音对象的交互反馈信令表的字段取值应遵循如下约束：

- table_type应取值为3；
- asset_group_flag应取值为0。

5.3.2.2 语法

交互反馈信令表语法见表2。

表2 交互反馈信令表

语法	值	比特数	类型
interaction_feedback_table() {			
table_id		8	uimsbf
version		8	uimsbf
length		16	uimsbf
table_payload {			
table_type		8	uimsbf
timestamp			
message_source		1	
asset_group_flag		1	uimsbf
reserved		6	uimsbf
if(asset_group_flag){			
asset_group_id		8	
}			
else{			
asset_id()			

语法	值	比特数	类型
<pre> } if(table_type == 3){ coordinate_type if(coordinate_type == 0){ ClientPosition() } if(coordinate_type == 1){ azimuth elevation distance } } } } </pre>		8	uimsbf
		8	uimsbf
		8	uimsbf
		8	uimsbf

5.3.2.3 语义

table_type指示交互反馈信令表携带的信息类型。其取值含义如表3所示：

表3 交互反馈信令表类型

取值	描述
0	全景视频用户位置变动信息
1	容积视频用户位置变动信息
2	自由视角视频用户位置变动信息
3	音频声音对象交互信息
4到255	未定义

timestamp指示当前交互产生的时间，使用UTC时间。

message_source指示消息源，0表示交互反馈消息是客户端发往服务器，1表示交互反馈消息是服务器发往客户端。该值此处置0。

asset_group_flag指示当前消费内容是否属于一个媒体资源组。取值为1表示客户端当前消费内容属于一个媒体资源组；取值为0表示客户端当前消费内容不属于媒体资源组。

asset_group_id指示客户端当前消费内容的媒体资源组标识符

asset_id指示客户端当前消费内容的媒体资源标识符。

coordinate_type指示用户交互位置的坐标类型，该字段取值为0表示交互位置以笛卡尔坐标系指示；该字段取值为1表示交互位置以球面坐标系指示。

ClientPosition()指示全局坐标系下用户交互位置的x,y,z坐标，其具体定义如下。

```

aligned(8) class ClientPosition () {
    signed int(16) position_x;
    signed int(16) position_y;
    signed int(16) position_z;
}

```

其中，position_x指示用户实时位置相对起始位置沿着x轴位移，取值范围为 $(-2^{15}, 2^{15} - 1)$ ，以毫米为单位。

position_y指示用户实时位置相对起始位置沿着y轴位移，取值范围为 $(-2^{15}, 2^{15} - 1)$ ，以毫米为单位。

position_z指示用户实时位置相对起始位置沿着z轴位移，取值范围为 $(-2^{15}, 2^{15} - 1)$ ，以毫米为单位。

Azimuth、elevation、distance分别指示用户交互位置的方位角、高度和距离。

5.4 AVS3 视频 SMT 传输

在SMT中，PA消息用于指示媒体数据包消费，一个PA消息应当包含一个PA表、一个MP表和一个媒体呈现层信息表。其中，MP表包含一个或多个描述符，用于指示媒体资源的媒体数据包信息，通过读取描述符的标识符(descriptor_tag)确定该描述符的类型。其中包括缓存内容更新信令、包含非对齐MFU依赖关系的样本格式、Asset关系信息描述符、彩色信息描述符，详细信息请参见T/AI 109.6—2022第8章内容。AVS3视频使用“avs3”进行标识。

5.5 融合网络的切片封装与呈现

百城千屏通过融合网络进行媒体传输，使用SMT协议传输百城千屏超高清视音频时还需遵循融合网络下的切片封装与呈现方法，融合网络的切片封装与呈现方法见附录A。

附录 A（规范性） 融合网络的切片封装与呈现

融合网络的切片封装与呈现方法考虑了百城千屏应用的网络环境和传输需求，对 SMT 协议进一步定义和扩展，为百城千屏应用提供了适应融合网络环境的解决方案。

A.1 架构

SMT 协议从逻辑上可以分为三个功能区，分别是封装功能区、传送功能区和信令功能区。

封装功能区定义了内容和服务的逻辑组织结构，实现了数据内容和数据描述的分离，并进行了数据碎片化。这种分布式布置的媒体内容可以通过数据描述的关联来灵活组织和动态配置服务。

传送功能区则负责异构网络下多媒体内容的传输，包括多媒体数据包的封装和流式传送。传送功能区支持对媒体数据的重要性分级，能够快速进行数据存储格式与传输格式的转换，并提供数据内容的摘要和检索功能（例如视音频指纹和检索特征向量）。同时，传送功能区还引入了应用层纠错保护机制，建立了接收端缓冲模型来处理时延和抖动，并能够自适应网络状态的变化。

融合网络的切片封装方法通过以上三个功能区的协同工作，结合封装功能区的逻辑组织和数据碎片化，传送功能区的多媒体传输和数据处理能力，以及信令功能区的协调和控制机制，为融合网络环境下的百城千屏媒体传输提供了高效可靠的解决方案。

A.2 数据模型

SMT 协议提供了媒体数据的流式传输和存储式传输。在传输过程中，SMT 协议用信令消息保护数据模型。为满足 SMT 内容灵活组织的技术需求，SMT 服务中内容的关联关系由描述文件指定，服务的改变不需要对数据流层级进行更改，只需要更新描述文件。

媒体由多个媒体资源组成，一个媒体资源指的是建立多媒体呈现所用到的任意多媒体数据，是封装了编码媒体数据且具有相同媒体资源标识符的内容碎片的逻辑集合。内容碎片可以命名为通用封装单元（Common Encapsulation Unit, CEU），其中包含的编码媒体数据可以是时序的或非时序的。时序数据是指有内在时间轴的编码媒体数据，要求数据单元在指定时间同步解码并呈现。相对的，非时序数据指的是在解码并呈现媒体内容时，没有内在时间轴的数据类型。也就是说，非时序数据中每个数据单元的解码及呈现未必要和该数据中的其他数据单元相互依赖。

A.3 通用封装单元

通用封装单元（Common Encapsulation Unit, CEU）是一个符合 ISO BMFF 的文件。Asset 标识、CEU 序列号以及相关信息由‘cceu’盒子提供，以明确地标识出封装进 CEU 文件中的媒体数据。‘moov’盒子包含所有编码器配置信息，以解码和呈现媒体数据。

时序媒体数据作为 ISO BMFF 的轨道存储，CEU 中允许单媒体轨道。非时序媒体作为 ISO BMFF 的元数据的部分存储。图 1 描绘了两个封装的例子，一个是时序媒体，另一个是非时序媒体。对于封装化的 CEU 传送，SMT 提示轨道提供将封装的 CEU 转化成 SMTP 负载和 SMTP 包的信息。

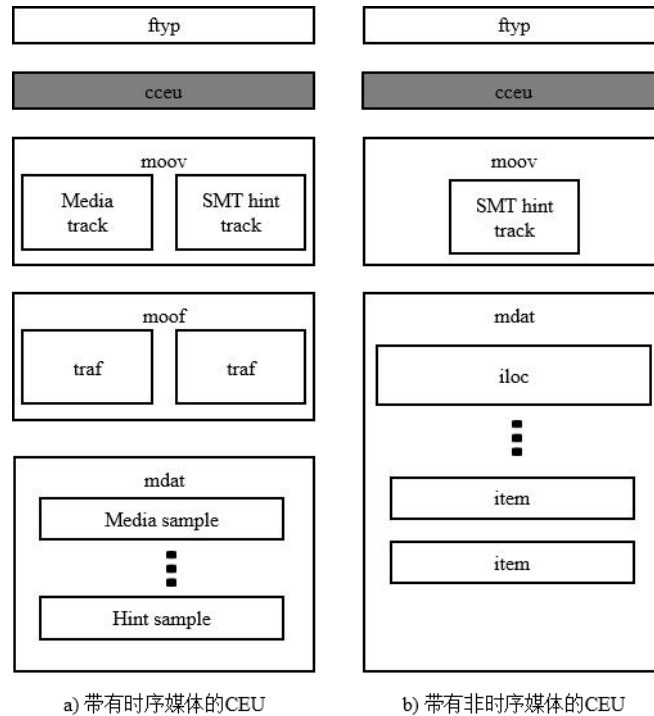


图 1 CEU 封装结构

CEU 标签‘ceuf’(CEU file)确定了遵从 CEU 封装规则的文件。‘ceuf’标签需要‘isom’标签的支持。

一个 CEU 文件由一组使 CEU 自包含的元数据盒子组成。一个 CEU 文件应包含一个‘ftyp’，‘cceu’，‘moov’盒子，一个可选‘sidx’盒子，以上所有都是 CEU 的元数据一部分。其他盒子也被允许，但如果分析程序不识别它们则会被忽略掉。

‘moov’盒子应最多包含一条媒体轨道，且可以包含用于标识传输格式中媒体最小分割单元的 SMT 提示轨道。‘moov’盒子中的轨道应不包含媒体帧以保证开销较小（就是说‘stts’，‘stsc’和‘stco’盒子中的 entry_count 应被设成 0）。存储时序媒体数据 CEU 的文件中，‘moov’盒子应包含‘mvex’盒子，以指示使用了 moof 盒子结构。‘mvex’盒子也设定了以后 moof 盒子的轨道和样本的默认值。

此外，一个‘cceu’盒子应出现于文件级别，且应用如下规则，包含如何决定盒子们的顺序：

- a) 如果出现，‘cceu’盒子应紧跟‘ftyp’盒子放置其后；
- b) 对于时序媒体数据，零个或更多‘sidx’盒子可以在文件中出现。如果出现，它们应索引构成当前 CEU 的 moof 盒子。

除了盒子顺序，如下限制也应用于‘ceuf’标签：

- a) 文件中独立媒体轨道的最大数量为 1（如，空的‘traf’盒子）。并且，非空‘traf’盒子的轨道（如提示轨道）也可用；
- b) 对时序媒体数据，文件应至少包含一个 moof 盒子；
- c) 对非时序媒体数据，‘meta’盒子应出现于文件级别且应包含 CEU 的非时序媒体元；
- d) 如果出现，编辑列表盒子（‘elst’）应仅提供初始偏移；
- e) 样本数据序列应以解码顺序放置于‘mdat’盒子，并且两者间无任何其他数据；
- f) 任何样本辅助数据，如‘saio’和‘saiz’描述，应放置在‘mdat’盒子开始，所有样本数据之前；
- g) 任何提示数据应放置于‘mdat’样本数据之后（或者样本数据之后的另一个 mdat），以使传输前后不改变样本偏移。

‘tfdt’盒子应在每个 moof 盒子的‘traf’盒子内部，以提供该片段按照解码顺序第一个样本的的解码时间。

如果任何‘elst’盒子可用，则它指示的偏移应适用于该 CEU 中依照呈现顺序的第一个样本的合成时间，和任何呈现信息提供的呈现时间。

时序媒体数据作为 ISO BMFF 的一条轨道存储，被‘moov’和‘moof’盒子索引，支持反相兼容。SMT 提示轨道指导 SMT 发送实体将封装的 CEU 转化成封包化的媒体流以采用诸如 SMT 的传输协议来传送。

非时序媒体数据作为元数据项目存储在‘meta’盒子里面。‘meta’盒子应出现在文件层级。每个非时序媒体数据文件应作为单独项目分别存储在‘meta’盒子里。非时序媒体的进入点应被标记为‘meta’盒子的主要项目。

A.4 CEU 盒子

A.4.1 概述

- 盒子类型：‘cceu’
- 容器：文件
- 强制性：是
- 数量：一或多个

通用封装单元(‘cceu’)盒子包含了当前 CEU 属于的 Asset 的标识和当前 CEU 的其他属性信息。Asset 标识可全局性无歧义地标识 Asset。CEU 信息包含该 CEU 在 Asset 中的序号以及相关属性信息。

A.4.2 语法

```
aligned(8) class CEUBox
  extends FullBox(‘cceu’, version, 0){
  unsigned int(1) is_complete;
  unsigned int(7) reserved;
  unsigned int(32) ceu_sequence_number;
  AssetIdentifierBox();
}

aligned(8) class AssetIdentifierBox {
  unsigned int(32) asset_id_scheme;
  unsigned int(32) asset_id_length;
  unsigned int(8) asset_id_value[asset_id_length];
}
```

A.4.3 语义

is_complete: 指示该 CEU 是否包含了媒体最小分割单元 (MFU) 结构中描述的所有 MFU。

ceu_sequence_number: 当前 CEU 的序列号。媒体资源 (Asset) 中的第一个 CEU 序列号为 0，之后的 CEU 序列号依次递增。此序列号在一个媒体资源中是唯一的。

asset_id_scheme: 区分用来表示 Asset ID 的策略，决定了 asset_id_value 的类型。有效的策略见表 4。

表 4 asset_id_scheme 取值列表

取值	描述
“UUID”	UUID (通用唯一标识符)
“URI”	URI (统一资源标识符)

asset_id_length: asset_id_value 的长度。

asset_id_value: 即媒体资源的标识符。其取值格式依赖于 asset_id_scheme 的类型。

A.5 SMT提示轨道

A.5.1 概述

出于传送的目的, SMT 提示轨道为 SMT 发送实体提供将 CEU 分解(或分割)为传输格式中媒体最小分割单元的信息。该最小分割单元是媒体已知的, 且被用来搭建 SMTP 的负载, 即 CEU 中的媒体数据在传送时被 SMT 发送实体提取进 SMTP 负载。

SMT 提示轨道也提供了从 SMTP 负载中提取和重建 CEU 的信息。SMTP 负载可以包含 CEU 元数据、分片元数据、或是一或多个传输格式中的最小分割单元。CEU 元数据可以包含‘ftyp’, ‘sidc’, ‘cceu’, 和‘moov’盒子。

A.5.2 提示轨道样本属性

A.5.2.1 语法

```
aligned(8) class SMTHintSampleEntry() extends SampleEntry('smth') {
    unsigned int(16) hinttrackversion = 1;
    unsigned int(16) highestcompatibleversion = 1;
    unsigned int(16) packet_id;
    unsigned int(1) is_fragment;
    unsigned int(1) is_timed;
    unsigned int(6) reserved;
}
```

A.5.2.2 语义

packet_id: 指示该提示轨道应用于哪一个 Asset 的唯一标识符。

is_fragment: 指示 CEU 是否切分为最小分割单元 (MFU) 的标识。若该标识位置‘0’, 该提示轨道应用于完整的 CEU, 即每个轨道片段 (track fragment) 对应于一个媒体片段; 否则, 每个提示样本应用于单个 MFU。

is_timed: 指示该轨道提示的媒体数据是否是时序的。

A.5.3 提示轨道样本格式

A.5.3.1 定义

每个媒体样本被划分为一个或多个 MFU。每个 SMT 提示轨道样本对应一个或多个 MFU。

A.5.3.2 语法

```
aligned(8) class SMTHSample {
    unsigned int(32) sequence_number;
    if (is_timed) {
        signed int(8) trackrefindex;
```

```

        unsigned int(32)  movie_fragment_sequence_number;
        unsigned int(32)  samplenumber;
        unsigned int(8)   priority;
        unsigned int(8)   dependency_counter;
        unsigned int(32)  offset;
        unsigned int(32)  length;
    }
    else {
        unsigned int(16)  item_ID;
    }
}

```

A.5.3.3 语义

sequence_number: 指示该 MFU 在 CEU 中序列号的整数值。CEU 中序列号的断续是允许的，用于指示特定 MFUs（其序列号在序列中缺失）在 CEU 组包后未被处理。由下层网络实体完成传送和缓冲是处理 MFU 的一个例子。

movie_fragment_sequence_numbe: 该 MFU 中媒体数据所属媒体片段的序列号。

trackrefindex: 所描述的媒体对应的 trackID。

samplenumber: 该 MFU 提取自媒体样本的编号。样本编号 n 指向当前媒体片段累计媒体样本中的第 n 个。媒体样本中第一个样本的样本编号置为‘1’。

item_ID: 对非时序媒体数据，这是该 MFU 包含该项目的标识。

priority: 指示 CEU 中该 MFU 相对于其它 MFU 的优先级。

dependency_counter: 指示依赖于该 MFU 解码的 MFU 的个数。该字段的值等于按 sequence_number 排列的可能无法正确解码的后续 MFU 的个数。例如，该字段的值为 n，则若无此 MFU 后续有 n 个 MFU 可能无法正确解码。

offset: 该 MFU 中媒体数据偏移。偏移基准为包含‘mdat’盒的开始。MFU 应放置在偏移指示的位置。

length: 该 MFU 中数据的长度，以字节为单位。

A.6 传输模型

SMT 传输模型能够解决多方面问题，如不同网络通道 QoS 参数各异情况下如何保证数据的可靠传输，异构网络条件下如何设计终端缓存模型使多源媒体数据同步呈现，以及当网络条件恶化时提供基于数据重要等级的保护机制差异化等。

为支持 SMT 传输需求，SMT 信令文件和传输包均根据媒体服务逻辑包进行组织，媒体组织和传输都与媒体服务相关，有利于根据不同服务类型差异化实现媒体内容组织和传输的自适应性。

SMT 逻辑包可以序列化为 SMT 文件，支持媒体文件式的存储、传输和下载；也可以打包为 SMT 传输包，以支持媒体的流化传输。由于文件格式和传输包格式内容上的高相关性，SMT 支持两者的便捷转换，便于中继转发，同时传输模型也可以高效响应服务内容的动态配置。传输过程中服务描述和传输控制等信令文件可以采用与媒体内容不同的传输模式，既能满足带外传输需求，又能对信令文件加以更高等级的保护措施。

媒体通用封装单元的碎片化、自包含性也决定了 SMT 传输模型能够应对传输过程中的包错误问题，实现弹性容错和错误隐藏，支持接收端快速恢复错误。每个通用封装单元及其切片单元都被唯一标识，头部信息与信令消息相结合，指定数据范围、优先级、QoS 要求、保护方式、呈现时间等不同层级的参

数设置，对于媒体内容的分级、异构、多通道、可靠性传输有重要参考意义。

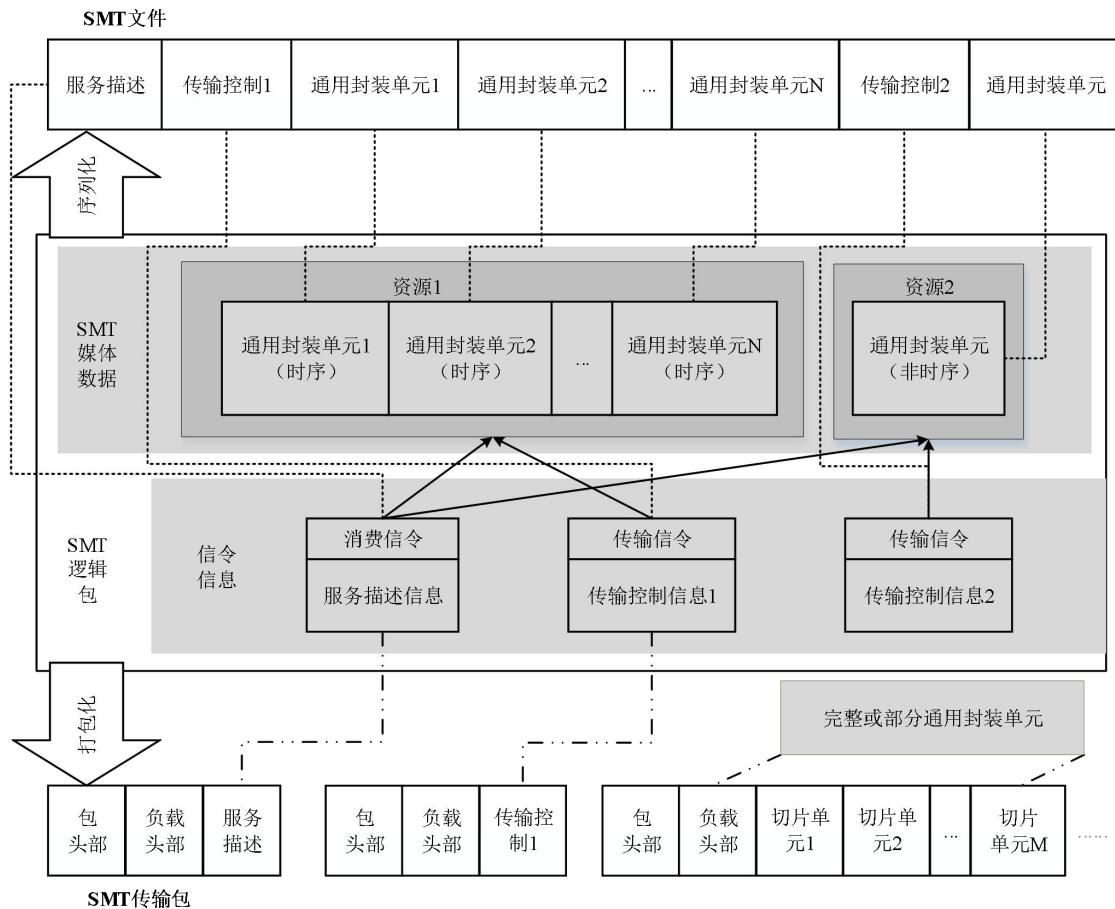


图 2 传输模型

A.7 信令

信令功能区中包含了一整套消息格式传达用于传输和消费 SMT 数据包的必要信令消息，分别称为传输信令和消费信令。本规范详细介绍了承载信令表、描述符或者传输相关信息的消息格式。信令表包含特定信令消息的元素与属性集，也可以包含描述符来携带更多细节信息。

SMT 协议中使用 MP 表来包含所有媒体资源的媒体数据包的信息，MP 表中的资源描述符（asset_descriptors）用于存放媒体数据的描述符，AVS3 视频描述符 AVS3_video_descriptor 与 AVS2 视频描述符 AVS2_video_descriptor 可包含在 MP 表的资源描述符中。

SMT 提供时钟关系信息描述符(CRI 描述符)用于时间同步功能,说明了NTP时间戳和MPEG-2 STC时间的映射关系。

A.8 呈现模型

SMT 的呈现模型支持多屏应用及个性化应用等多方面的呈现新需求，见图 2。SMT 数据经过协议解析引擎分析后，分为信令控制信息和媒体数据信息，该协议解析引擎可以配置专门的智能网关设备，为多屏终端提供接入和控制接口。媒体数据经过媒体处理模块后可分为时序媒体资源和非时序媒体资源，

它们将按照信令消息的描述组合为某种媒体服务，从而有序地呈现于不同终端之上。信令控制信息描述了媒体数据的相互关系及呈现信息，一方面指导媒体数据的选择性接收或者主动获取，同时提供媒体数据在终端正确解码处理的提示信息，另一方面在空间布置和时域更新上为不同终端提供媒体呈现策略。呈现引擎根据控制信息正确渲染并呈现所接收的媒体数据，并作为控制器完成多媒体资源、多屏设备之间的同步、互动。

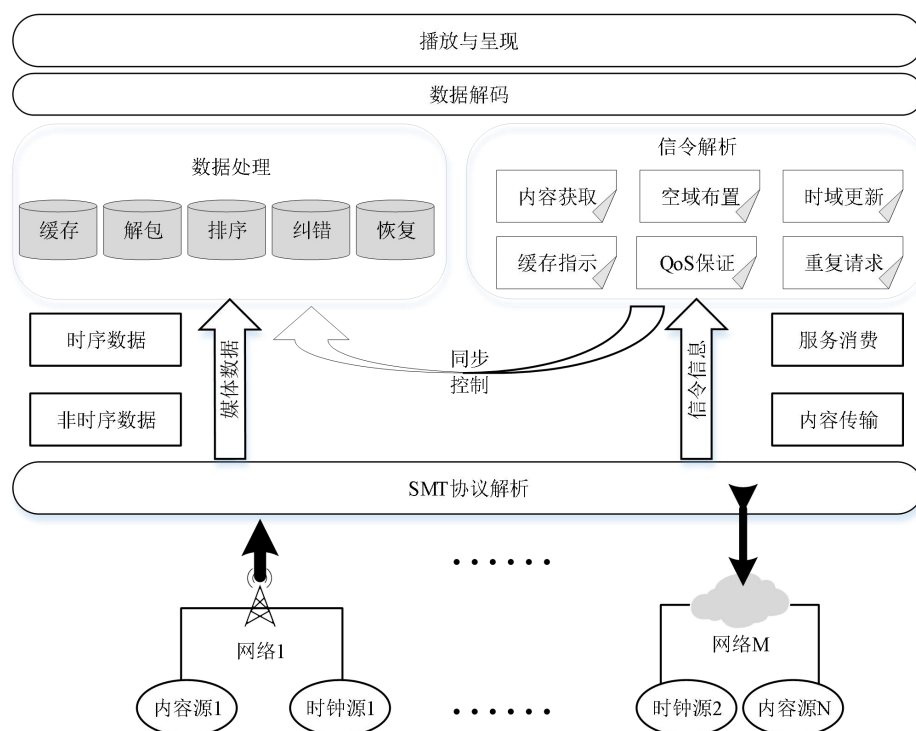


图 3 SMT 呈现模型

SMT 的呈现模型通过协调不同设备之间的数据传输和呈现，实现了多屏同步播放、画中画、四分屏等多种形式的多屏应用。无论是在大屏幕上的多个视角，还是在小屏幕上的不同视角，SMT 的呈现模型能够确保媒体内容的同步播放和一致性呈现，为用户提供更丰富、更沉浸的观看体验。

A.9 多设备同步呈现播放模型

SMT采用 UTC 时钟进行时间戳标注，以实现不同数据包之间的同步。通过SMT信令得到视频的绝对呈现时间，建立一个呈现时钟作为视音频播放的参考值，音频解码后，同步控制功能按呈现时钟进行呈现时间调节，这个过程是SMT的基本同步调节机制。SMT协议提供CEU时间戳描述符（CEU_timestamp_descriptor）提供了对应CEU的第一个AU的呈现时间。

SMT的多设备同步呈现模型可以实现多个设备之间的同步播放。这种模型百城千屏应用中具有广泛的应用场景，其中一种方式是在大屏上播放主视频，同时在个人终端上同步播放音频。这种组合可以为观众提供更加全面的观赏体验。当大屏幕展示精彩的图像和视觉效果时，个人终端可以通过播放相应的音频，使观众能够听到与所看内容相关的声音，从而加强对内容的理解和感受。

此外，多设备同步呈现模型还可以用于大屏幕播放主视角，而个人终端则播放其他视角或者关联数据。通过这种方式，观众可以在大屏上看到主要场景或角色的视角，同时在个人终端上切换到其他角度

的视角或者查看相关的数据。这样的呈现方式有助于提供更加全面、多角度的信息，使观众能够深入了解所呈现内容的细节和背景。

为了解决百城千屏中不同大屏之间传输的时间差异问题，大屏最终的呈现时间 T 应由大屏播放器自身决定，呈现时间 T 应在接收到媒体流、媒体解码与媒体缓存之后。各个百城千屏大屏需计算出呈现时间差值 ΔT 并上报到播控平台。

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

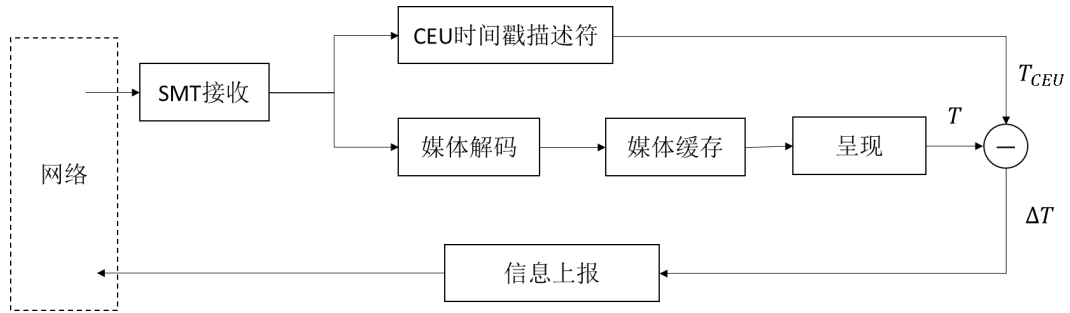


图 4 SMT 传输时大屏同步呈现示意图

调节量 ΔT 由各个百城千屏大屏决定，不同的大屏的呈现时间差值会不同。

小屏终端的媒体播放需要与百城千屏大屏同步，其呈现时间需要由CEU时间戳描述符与呈现时间差值 ΔT 相加得到。

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

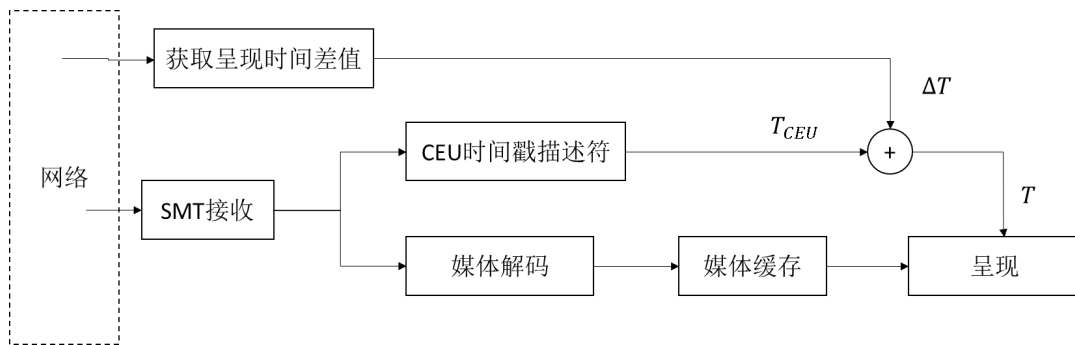


图 5 SMT 传输时小屏同步呈现示意图