

ICS 35.040
L71



中华人民共和国国家标准

GB/T 20090.8—YYYY

信息技术 先进音视频编码
第 8 部分：在 IP 网络上传输 AVS
第 2 子部分：AVS-P2 视频 RTP 净载格式
Information technology - Advanced coding
of audio and video
Part 8: AVS over IP
Sub-part 2: AVS-P2 RTP Payload Format
(送审稿)

2007 年 7 月

××××-××-××发布

××××-××-××实施

国家质量监督检验检疫总局 发布

目 次

目 次..... I

前 言..... II

引 言..... III

1 范围..... 1

2 规范性引用文件..... 1

3 术语和定义..... 1

3.1 定义..... 1

3.2 缩写..... 1

4 NAL 单元 2

5 RTP 净载格式..... 3

5.1 RTP 头的使用..... 3

5.2 RTP 净载格式的公共结构..... 4

5.3 打包模式..... 5

5.4 解码顺序号（DON） 5

5.5 单 NALU 包 6

5.6 复合包..... 7

5.7 分割单元..... 13

6 打包规则..... 15

6.1 公共打包规则..... 15

6.2 单 NALU 模式 15

6.3 非交织模式..... 15

6.4 交织模式..... 15

7 净载格式参数..... 16

7.1 媒体类型注册..... 16

7.2 SDP 参数..... 19

7.3 序列头考虑..... 22

8 安全考虑..... 22

9 拥塞控制..... 23

10 IANA 注册..... 23

11 解包过程（资料性）..... 23

11.1 单 NALU 和非交织模式..... 23

11.2 交织模式..... 23

11.3 额外的解包指导原则..... 24

前 言

GB/T 20090在《信息技术 先进音视频编码》的总标题下，包括以下九个部分：

第1部分：系统；

第2部分：视频；

第3部分：音频；

第4部分：一致性测试；

第5部分：参考软件；

第6部分：数字媒体版权管理；

第7部分：移动视频；

第8部分：在IP网络上传输AVS；

第9部分：文件格式。

本部分为GB/T 20090的第8部分的第2子部分：AVS-P2视频RTP净载格式。

本部分由中华人民共和国信息产业部提出。

本部分由全国信息技术标准化技术委员会归口。

本部分由信息产业部数字音视频编解码技术标准工作组组织起草。本部分起草单位包括XXX等。

本部分主要起草人：XXX等。

引 言

本部分定义了 IP 网络上传输 AVS-P2 视频时的 RTP 净载格式。该 RTP 净载格式应用范围宽广，可支持可视电话、视频会议、网络流媒体、IPTV 和视频点播等应用。

信息技术 先进音视频编码 第8部分:在IP网络上传输 AVS

第2子部分: AVS-P2 视频 RTP 净载格式

1 范围

本部分的应用范围包括可视电话、视频会议、网络流媒体、IPTV和视频点播等。

2 规范性引用文件

下列文件中的条款通过本部分的引用而成为本部分的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本部分,然而,鼓励根据本部分达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本部分。

GB/T 20090.2—2006 信息技术 先进音视频编码 第2部分: 视频 (以下简称AVS-P2)

IETF RFC 3550, RTP: a transport protocol for real-time applications

IETF RFC 2327, SDP: session description protocol

IETF RFC 3264, An offer/answer model with session description protocol (SDP)

IETF RFC 3548, The base16, base32, and base64 data encodings

3 术语和定义

AVS-P2视频标准中的定义和缩写适用于本部分。另外,下面的定义和缩写也适用于本部分。

3.1 定义

NAL单元 (Network abstract layer unit)

在使用本文所定义RTP净载格式对AVS-P2视频比特流进行打包之前,必须首先将比特流中每两个相邻的起始码前缀(0x000001)之间的数据转换为一个NAL单元。NAL单元的详细定义以及从AVS-P2视频比特流到NAL单元流的转换方法参见本文第4节。

NAL单元流 (NALU stream)

由一到多个NAL单元所组成的序列。

NALU解码顺序 (NALU decoding order)

NAL单元流中各NAL单元的顺序。

解码顺序号 (DON, Decoding order number)

RTP净载结构中的一个域,或一个导出变量,用于指定NALU解码顺序。取值范围为0..65535。当达到最大值后,下一个DON值为0。

传输顺序 (Transmission order)

按RTP序列号上升顺序(在模算术的情况下)的包顺序。

访问单元

组成一帧编码图像的一系列NAL单元。除编码图像头和条带数据外,一个访问单元中还可包含其它类型数据。在RTP打包时,同一个访问单元内的所有数据具有相同的时间戳值。

MANE (Media aware network element)

能够解析RTP净载头或净载的某些部分并针对其内容作出相应处理的网络单元,如应用层网关。

3.2 缩写

DON: 解码顺序号 (Decoding Order Number)

- DONB: 解码顺序号基数 (Decoding Order Number Base)
- DOND: 解码顺序号差 (Decoding Order Number Difference)
- FEC: 前向纠错 (Forward Error Correction)
- FU: 分割单元 (Fragmentation Unit)
- MTAP: 多时间复合包 (Multi-Time Aggregation Packet)
- MTAP16: 16位时间戳偏移的多时间复合包 (MTAP with 16-bit timestamp offset)
- MTAP24: 24位时间戳偏移的多时间复合包 (MTAP with 24-bit timestamp offset)
- MTU: 最大传输单元 (Maximum Transfer Unit)
- NAL: 网络抽象层 (Network Abstraction Layer)
- NALU: NAL单元 (NAL Unit)
- PSI: 净载结构指示符 (Payload Structure Indicator)
- RTP: 实时传输协议 (Real-time Transport Protocol)
- STAP: 单时间复合包 (Single-Time Aggregation Packet)
- STAP-A: A类单时间复合包 (STAP type A)
- STAP-B: B类单时间复合包 (STAP type B)

4 NAL 单元

NAL单元由一个字节的NAL单元头和一到多个字节的NAL单元数据两部分所组成。

从AVS-P2视频比特流到NAL单元流的转换过程如下：将AVS-P2视频比特流中每两个相邻的起始码前缀之间的数据（不包含起始码前缀但包括起始码值，必定是字节对齐的）直接映射成为NAL单元数据，并根据其中的起始码值及其上下文语义在该NAL单元数据之前插入一个NAL单元头。

NAL单元头的格式如图1所示。

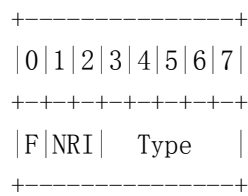


图1. NAL单元头格式

其中各字段的语义解释如下：

F: 1 bit

禁止位 (forbidden_zero_bit)。值应为 ‘0’ 。

NRI: 2 bit

NAL参考标号 (nal_ref_idc)。值不等于0说明该NAL单元包含的数据为序列头或参考帧数据；值等于0说明该NAL单元所包含的数据为非参考帧数据。对于序列头NAL单元，nal_ref_idc不应为0。对于某一帧，如果其某一NAL单元的nal_ref_idc为0，则属于该帧的所有NAL单元的nal_ref_idc均应为0。I帧的NAL单元的nal_ref_idc不应为0。

Type: 5 bit

NAL单元类型。其值参考随后NAL单元数据中的起始码值及其相应的图象头内容而定，具体取值依据见表1所示。

表1. NAL单元头中NAL单元类型取值依据

Type	类型说明	取值依据
0	保留	
1	序列头	起始码值为B0
2	视频扩展	起始码值为B5
3	用户数据	起始码值为B2
4	视频编辑	起始码值为B7
5	I帧图象头	起始码值为B3
6	P帧图象头	起始码值为B6，且图象头中编码方式为01
7	B帧图象头	起始码值为B6，且图象头中编码方式为10
8	I帧条带数据	起始码值为00～AF，所属图象的图象头起始码值为B3
9	P帧条带数据	起始码值为00～AF，所属图象的图象头起始码值为B6 且图象头中编码方式为01
10	B帧条带数据	起始码值为00～AF，所属图象的图象头起始码值为B6 且图象头中编码方式为10
11-23	保留	
24-31	未定义	

按照上述方式将AVS-P2比特流转换为NAL单元流后，才可以使用本文所定义的RTP打包方法。

解码器在收到NAL单元流后，必须首先丢弃每个NAL单元中的NAL单元头，并在相同位置插入一个起始码前缀（0x000001）将其还原成AVS-P2视频比特流之后才能进行解码。

5 RTP 净载格式

5.1 RTP 头的使用

IETF RFC 3550中定义了RTP头的格式，如图2所示。

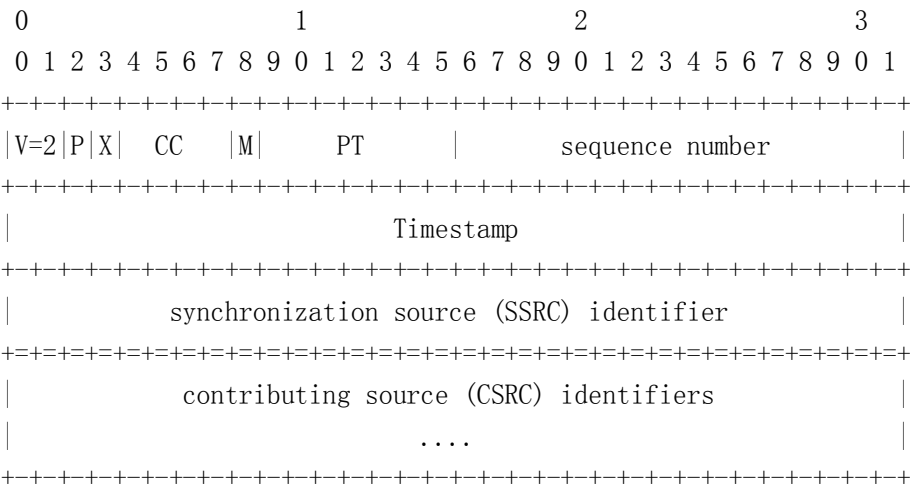


图2. RTP头

本文所述RTP净载格式对其中数据域的使用方法符合IETF RFC 3550的规定。其中对于M和Timestamp域的设定还需遵从如下规则：

标记位（M）：1bit

一个访问单元（由RTP时间戳确定）的最后一个包的标记位设为1，这个访问单元的其它包的标记位均设为0。对于复合包（STAP或MTAP），标记位的值应等于当该包中最后一个NALU单独打包时的标记位值。RTP包接收器可以通过这个比特得知是否已经收到访问单元的最后一个包，但是不能完全依靠这个信息。

时间戳（Timestamp）：32bit

RTP时间戳设为编码内容的采样时间戳，所用的时钟必须为90000赫兹，即RTP时间戳的时间单位为1/90000秒。如果NALU本身不具有时间属性，例如序列头NALU，其RTP时间戳设为紧跟该NALU之后的第一个编码图像的RTP时间戳。MTAP的RTP时间戳的设定在5.6.2小节规定。

5.2 RTP 净载格式的公共结构

本净载格式定义了三种不同的基本净载结构，每种又可分为不同的子类型：

- 单NALU包
一个RTP包净载中仅包含一个NAL单元。
- 复合包
一个RTP包净载中包含多于一个NAL单元。复合包又分为四类：A类单时间复合包（STAP-A）、B类单时间复合包（STAP-B）、16位时间戳偏移的多时间复合包（MTAP16）和24位时间戳偏移的多时间复合包（MTAP24）。
- 分割单元
用于将一个NALU分割到多个RTP包净载中。分割单元又分为两类，A类分割单元（FU-A）和B类分割单元（FU-B）。

不同的净载结构通过RTP净载中的第一个字节PSI（净载结构指示符）来进行区分。PSI具有与第4节所述NAL单元头完全相同的格式，如图3所示。

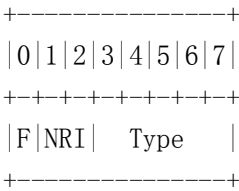


图3. PSI（净载结构指示符）格式

F：1 bit

值0表示RTP净载中不宜包含比特错误或其它语法错误。值1表示RTP净载中可能包含比特错误或其它语法错误。如果检测到RTP净载中有比特错误，MANE宜将F设为1。

NRI：2 bit

除了必须符合第4节中针对NALU的规定之外，NRI值表明RTP包的相对传输优先级。MANE可以利用这个信息更好地保护更重要的RTP包。优先级由高到低的NRI值依次为11b、10b、01b和00b。含有序列头和I帧图像的RTP净载中的NRI值宜为11b。

Type：5 bit

用于指示RTP净载的结构，语义见表2所示。

表2. PSI中Type字段所对应的净载结构

Type	净载结构	净载结构全称	章节
0	未规定		—
1-23	单NALU	单NAL单元包	5. 5
24	STAP-A	A类单时间复合包	5. 6. 1
25	STAP-B	B类单时间复合包	5. 6. 1
26	MTAP16	16位时间戳偏移的多时间复合包	5. 6. 2
27	MTAP24	24位时间戳偏移的多时间复合包	5. 6. 2
28	FU-A	A类分割单元	5. 7
29	FU-B	B类分割单元	5. 7
30-31	未规定		—

5.3 打包模式

本净载格式支持三种打包模式：单NALU模式、非交织模式和交织模式。在单NALU模式或非交织模式下，NALU按照解码顺序进行传输；交织模式允许NALU不按照解码顺序进行传输。

采用的打包模式决定了RTP净载中允许包含哪些净载结构，打包模式可以通过可选的媒体类型参数 packetization-mode 或外部方式来进行传送。表3总结了各打包模式下允许的净载结构，其中标为“未规定”的净载结构被保留以便将来扩展。发送端不宜发送这些净载结构，接收端应忽略这些净载结构。

表3. 各打包模式下允许的净载结构（yes = 允许，no = 不允许， ig = 忽略）

PSI Type	净载 结构	单NALU 模式	非交织 模式	交织 模式
0	未规定	ig	ig	ig
1-23	单NALU	yes	yes	no
24	STAP-A	no	yes	no
25	STAP-B	no	no	yes
26	MTAP16	no	no	yes
27	MTAP24	no	no	yes
28	FU-A	no	yes	yes
29	FU-B	no	no	yes
30-31	未规定	ig	ig	ig

5.4 解码顺序号（DON）

在交织打包模式下，NALU的传输顺序允许与解码顺序不同。解码顺序号（DON）是净载结构中的一个域或一个导出变量，用于指定NALU解码顺序。

通过可选的媒体类型参数 sprop-interleaving-depth 控制传输顺序和解码顺序之间的耦合。当 sprop-interleaving-depth 值为0，或按非解码顺序传输NALU被外部方式禁止时，NALU必须按照解码顺序进行传输。当 sprop-interleaving-depth 值大于0，或当外部方法允许NALU按非解码顺序传输时：

- MTAP16和MTAP24中的NALU顺序不必为解码顺序；
- 通过分解连续两个包得到的STAP-B、MTAP和FU中的NALU顺序不必为解码顺序。

单NALU包、STAP-A和FU-A的RTP净载结构中不包括DON；STAP-B和FU-B的RTP净载结构中包括DON；MTAP的RTP净载结构允许导出DON，见5.6.2小节。

在单NALU打包模式下，NALU的传输顺序（由RTP序列号决定）必须与解码顺序相同。在非交织打包模式下，单NALU包、STAP-A和FU-A中的NALU的传输顺序必须与解码顺序相同。包含在STAP中的NALU必须按解码顺序出现。这样，首先通过STAP中的暗含顺序，然后通过确定STAP、FU和单NALU包的RTP序列号来提供解码顺序。

得到包含在STAP-B、MTAP和由一个FU-B开始的一系列FU中的NALU的DON值的方法分别在5.6.1、5.6.2和5.7小节介绍。传输顺序中的第一个NALU的DON值可以任意设定。DON的取值范围为0..65535。当达到最大值后，下一个DON值为0。

确定包含在STAP-B、MTAP或由一个FU-B开始的一系列FU中的两个NALU的解码顺序的方法如下。

令DON(i)为传输顺序索引为i的NALU的DON值，规定函数don_diff(m, n)如下：

If $DON(m) == DON(n)$, $don_diff(m, n) = 0$

If $(DON(m) < DON(n) \text{ and } DON(n) - DON(m) < 32768)$,
 $don_diff(m, n) = DON(n) - DON(m)$

If $(DON(m) > DON(n) \text{ and } DON(m) - DON(n) \geq 32768)$,
 $don_diff(m, n) = 65536 - DON(m) + DON(n)$

If $(DON(m) < DON(n) \text{ and } DON(n) - DON(m) \geq 32768)$,
 $don_diff(m, n) = - (DON(m) + 65536 - DON(n))$

If $(DON(m) > DON(n) \text{ and } DON(m) - DON(n) < 32768)$,
 $don_diff(m, n) = - (DON(m) - DON(n))$

若don_diff(m, n)等于0则两个NALU的解码顺序可以为任意顺序，值为正表明传输顺序索引为m的NALU的解码顺序在传输顺序索引为n的NALU之前，值为负则与值为正的情形相反。

与DON有关的域（DON、DONB和DOND，见5.6小节）的值必须符合以下条件：由DON值按上述方法决定的解码顺序必须与NALU解码顺序相同。如果两个NALU的解码顺序互换后的新顺序与NALU解码顺序不一致，则它们必须具有不同的DON值。如果NAL单元流中两个连续的NALU的顺序交换后的新顺序仍然与NALU解码顺序一致，则它们可以具有相同的DON值。DON值相同的NALU可以以任意顺序解码，而DON值不同的两个NALU宜按上面规定的顺序送给解码器。如果两个解码顺序连续的NALU具有不同的DON值，则第二个NALU的DON值宜等于第一个NALU的DON值加1。

5.5 单 NALU 包

单NALU包必须包含且仅包含一个本文第4节中所定义的NAL单元。复合包或分割单元不能在单NALU包中使用。通过按照RTP序列号顺序解封装单NALU包所得到的NAL单元流必须符合NALU解码顺序。在单NALU包净载结构中，RTP净载中的第一个字节（即PSI）直接采用所封装NALU的NAL单元头，其后紧跟该NALU的NAL单元数据，如图4所示。

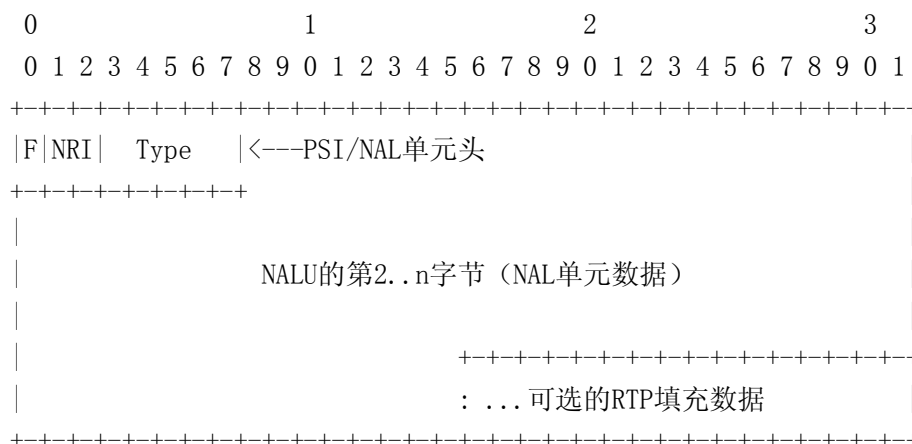


图4. 单NALU包的RTP净载格式

5.6 复合包

本净载格式定义了以下两类复合包：

- 单时间复合包 (STAP)：包含具有相同NALU时间的NALU。分为两个子类：STAP-A和STAP-B。STAP-A净载结构中不包含DON，而STAP-B净载结构中包含DON。
- 多时间复合包 (MTAP)：包含可能具有不同NALU时间的NALU。根据NALU时间戳偏移的长度分为两个子类：MTAP16和MTAP24。

其中术语“NALU时间”定义为如果一个NALU单独打包为一个RTP包时的RTP时间戳。

复合包中的每个NALU被封装成一个复合单元。后面给出了四种不同的复合单元和它们的特征。

复合包的RTP净载格式如图5所示。

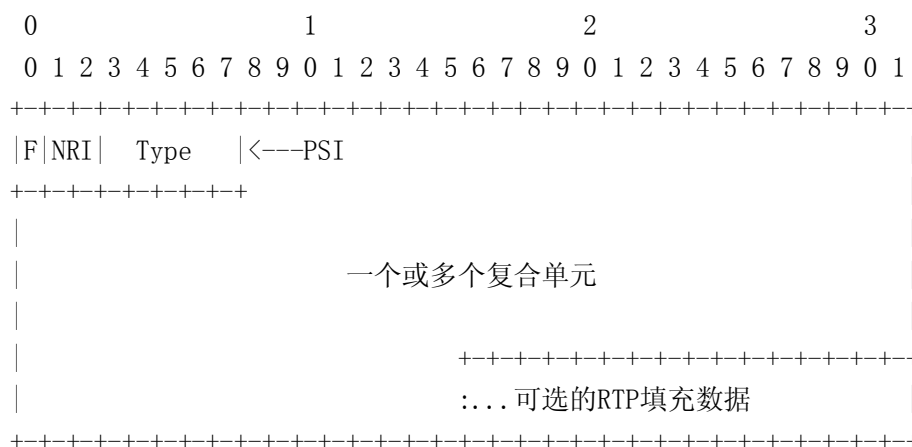


图5. 复合包的RTP净载格式

复合包中的RTP时间戳必须设为其中所有NALU的最早NALU时间。PSI中的Type域必须根据情况设为表4中所列出的4个值之一。如果所有被复合的NALU的F位都为0，则PSI中的F位设为0；否则PSI中的F位设为1。PSI中NRI的值必须为所有被复合的NALU的NRI最大值。

表4. 复合包（STAP或MTAP）中PSI的Type域

PSI Type	净载 结构	时间戳偏移 域的长度（位）	是否存在与DON相关的 域（DON、DONB、DOND）
24	STAP-A	0	否
25	STAP-B	0	是
26	MTAP16	16	是
27	MTAP24	24	是

RTP头中的标记位（M）设为如果该复合包中最后的NALU单独打包时的标记位的值。

每个复合包的净载中紧随PSI之后包含一个或多个复合单元。四类复合单元的定义见5.6.1和5.6.2小节。一个复合包可以根据需要包含任意数目的复合单元，但是复合包的总数据量必须适合装入一个IP包，该IP包的长度宜小于传输网络的MTU。复合包中不能包含5.7小节介绍的分割单元。复合包不允许相互嵌套，即一个复合包中不能包含另一个复合包。

5.6.1 单时间复合包

任何时候需要复合NALU时间相同的NAL单元时，宜采用单时间复合包（STAP）。STAP-A净载由至少一个单时间复合单元组成，并且不包含DON，如图6所示。STAP-B净载由一个16位的无符号的DON（网络字节顺序）和至少一个单时间复合单元组成，如图7所示。

STAP-B的DON域给出按传输顺序该包中第一个NALU的DON值。对于每个随后出现的NALU，其DON值等于前一NALU的DON值加1模65536。

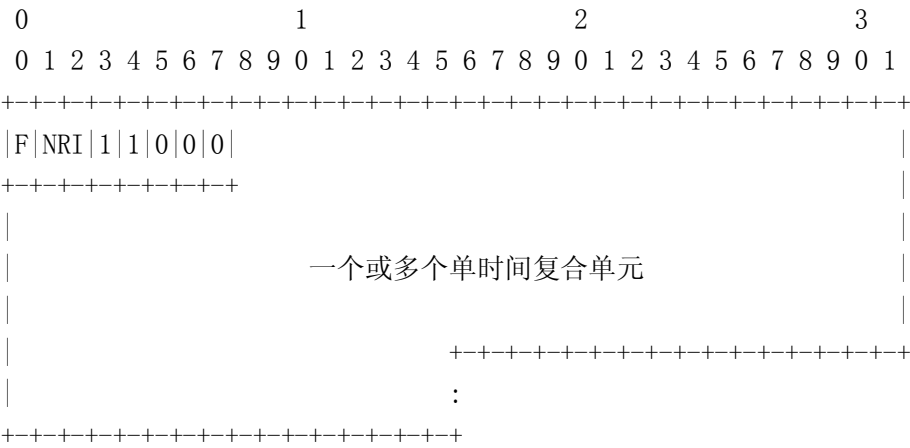


图6. STAP-A净载格式

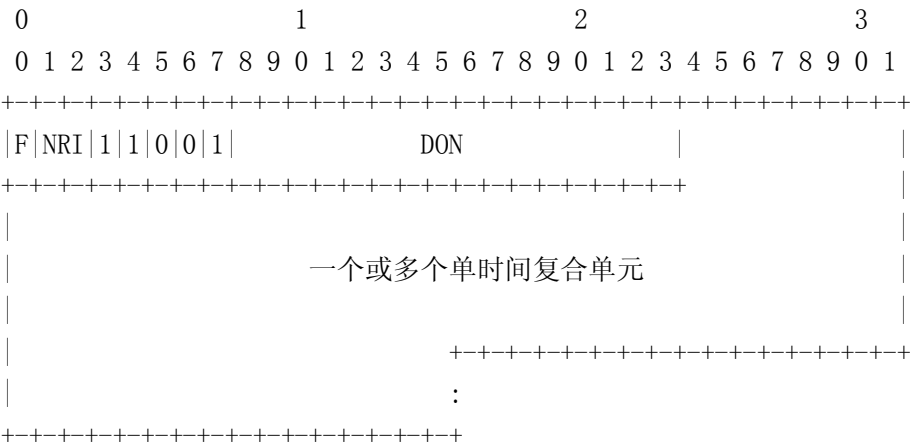


图7. STAP-B净载格式

一个单时间复合单元包含一个16位的无符号尺寸信息（网络字节顺序，表明随后的NALU的大小，单位为字节，不包括两字节的尺寸信息，但是包括NALU头）和NALU本身（包括NALU头）。单时间复合单元在RTP净载内是字节对齐的，但不一定是32位对齐的。图8示出了单时间复合单元的结构。

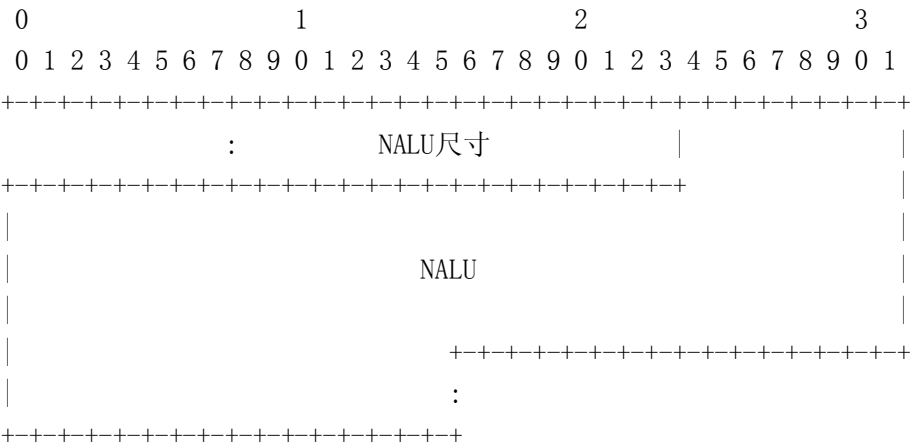


图8. 单时间复合单元的结构

图9给出了包含一个STAP-A的RTP包的例子。该STAP包含两个单时间复合单元，在图中标记为1和2。



图9. RTP包举例：包含一个由两个单时间复合单元组成的STAP-A

图10示出了包含一个STAP-B的RTP包的例子。该STAP包含两个单时间复合单元，在图中标记为1和2。



图10. RTP包举例：包含一个由两个单时间复合单元组成的STAP-B

5.6.2 多时间复合包

MTAP净载包含一个16位的无符号解码顺序号基数（DONB，网络字节顺序）和至少一个多时间复合单元，如图11所示。DONB必须包含MTAP中第一个NALU（解码顺序）的DON值。



图11. MTAP净载格式

本文定义了两种多时间复合单元。一个多时间复合单元中包含一个16位（网络字节顺序）的无符号尺寸信息（用于描述随后的NALU的大小）、一个8位的无符号解码顺序号差（DOND）、和一个16或24位（网络字节顺序）的时间戳偏移。不同MTAP类型（MTAP16或MTAP24）的选择取决于具体应用。时间戳偏移越大，MTAP的灵活性也就越高，但数据传输效率却会下降。

MTAP16和MTAP24的结构分别示于图12和图13。其中各多时间复合单元的起始或结束位置不必是32位对齐的。复合单元中的NALU的DON值等于DONB与DOND值求和再模65536。本部分不规定MTAP中NALU的顺序，但在大多数情况下宜使用NALU解码顺序。

时间戳偏移的设定如下：如果NALU时间大于等于RTP包的时间戳，则设为NALU时间减去RTP时间戳；否则设为NALU时间减去RTP时间戳再加上2³²。

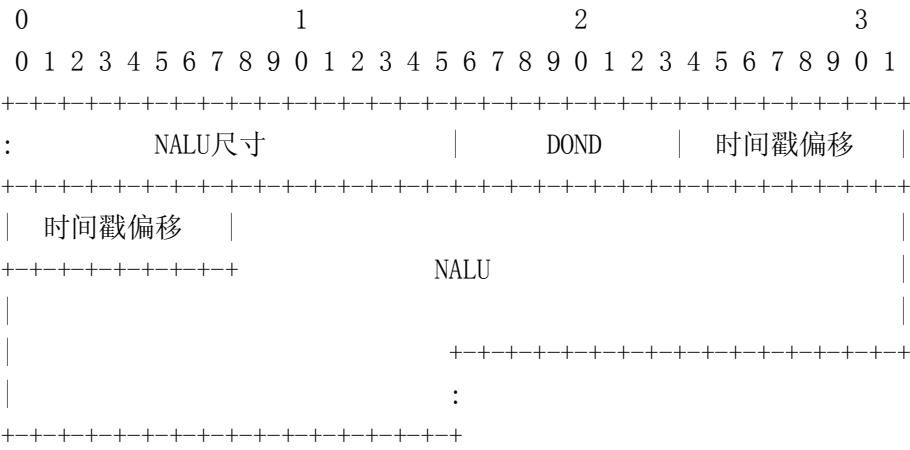


图12. 16位时间戳偏移的多时间复合单元结构

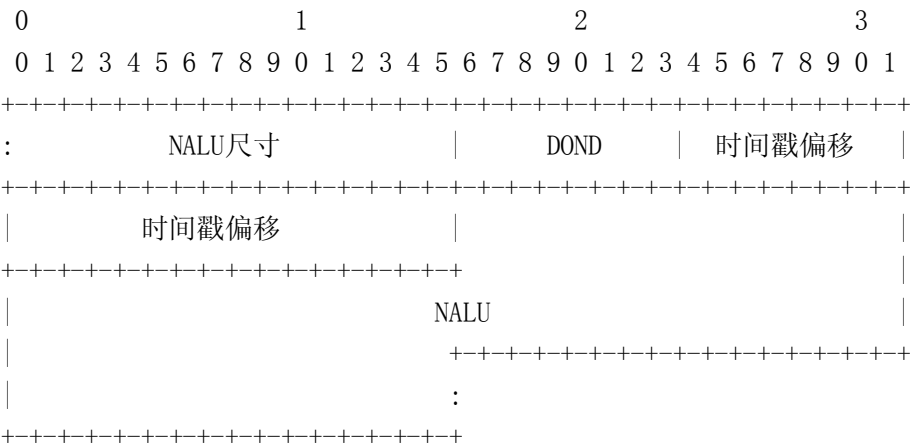


图13. 24位时间戳偏移的多时间复合单元结构

MTAP中“最早的”多时间复合单元的时间戳偏移值必须为0。因此，MTAP本身的RTP时间戳等于最早的NALU时间。其中“最早的”多时间复合单元是指当MTAP中包含的所有NALU都单独打包时，扩展RTP时间戳最小的那个包所包含的NALU对应的多时间复合单元。扩展RTP时间戳的长度大于32位。“最早的”多时间复合单元不一定是MTAP中按打包顺序的第一个，也不必是按NALU解码顺序的第一个。

图14是包含一个MTAP16的RTP包的例子。该MTAP16包含两个多时间复合单元，在图中标记为1和2。



图14. RTP包举例：包含一个由两个多时间复合单元组成的MTAP16

图15是包含一个MTAP24的RTP包的例子。该MTAP24包含两个多时间复合单元，在图中标记为1和2。



图15. RTP包举例：包含一个由两个多时间复合单元组成的MTAP24

5.7 分割单元

分割单元净载结构允许将一个NALU分割到多个RTP包中。在应用层进行分割而不是依赖于低层分割（如IP层），可以在IPv4网络中传输大于64K字节的NALU。大于64K字节的NALU可能存在于预先录制的视频流中，特别是在高分辨率格式下。

分割单元只适用于单个NALU而不适用于任何复合包。一个分割单元包含一个NAL单元数据中部分连续的整数个字节，而NAL单元数据中的每个字节必须完全包含在某一分割单元中。同一NALU的分割单元的发送必须以连续RTP序列号增长的顺序进行，即不能在这些分割单元中间插入发送其它RTP包。因此，一个NALU必须按照RTP序列号顺序组装。

一个被分割成分割单元的NALU叫做分割的NALU。STAP和MTAP不能被分割。分割单元不允许相互嵌套，即一个分割单元中不能包含另一个分割单元。

包含一个分割单元的RTP包的RTP时间戳设置为被分割的NALU的NALU时间。

图16所示为FU-A的RTP净载格式，其中包含一个PSI（1字节）、一个分割单元头（1字节）和一个分割单元净载。

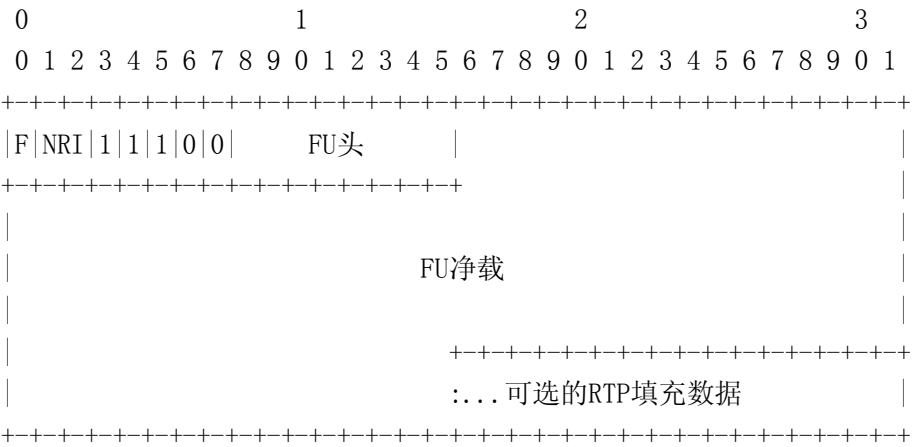


图16. FU-A的RTP净载格式

图17给出了FU-B的RTP净载格式。一个FU-B包含一个PSI（1字节）、一个分割单元头（1字节）、一个解码顺序号（DON，网络字节顺序）和一个分割单元净载。

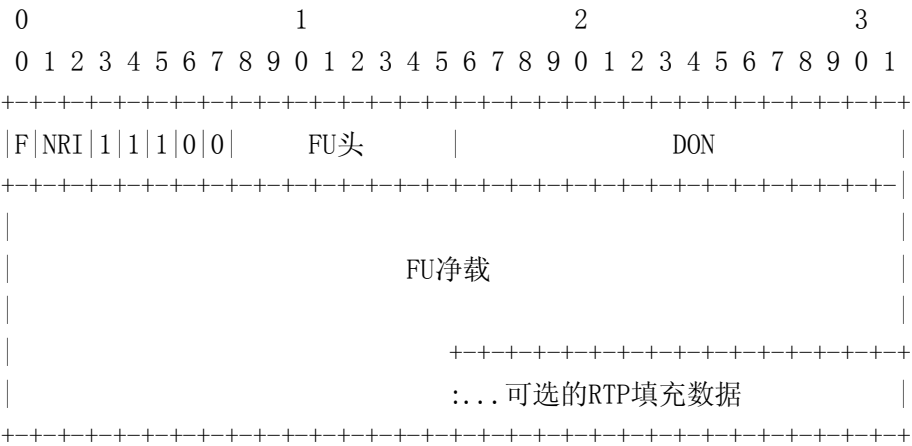


图17. FU-B的RTP净载格式

FU-B只能用于交织打包模式下作为一个分割的NALU的第一个分割单元，而不能用于任何其它场合。也就是说，在交织打包模式下，每个分割的NALU的第一个分割单元为FU-B，后跟一个或多个FU-A。PSI字节中的NRI域的设定与被分割的NALU的NAL单元头中的NRI域相同。紧随PSI之后的分割单元头（FU头）的格式定义如图18所示：

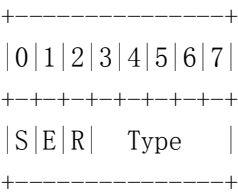


图18. 分割单元头格式

起始位 (S) : 1bit

为1表明一个分割的NALU的开始。如果随后的分割单元净载不是分割的NALU的开始，则设为0。

结束位 (E) : 1 bit

为1表明一个分割的NALU的结束，即FU净载的最后字节也是分割的NALU的最后字节。如果随后的分割单元净载不是分割的NALU的最后分割单元，则设为0。

保留位 (R) : 1 bit

保留位必须为0，接收器应将其忽略。

Type: 5 bit

本文第4节表1中所定义的NAL单元类型。

FU-B中的DON域的值的选择参见5.4小节。

一个分割的NALU不能在同一个分割单元中传输，即同一分割单元头中的S位和E位不能同时为1。

将连续的分割单元的净载连接起来可以重建被分割的NALU的净载。分割的NALU的NAL单元头并不直接包含在分割单元净载中，而是包含在分割单元的PSI字节的F和NRI域以及FU头的Type域中。一个分割单元净载可能包含任意个字节，也可能为空单元。

如果一个分割单元丢失，接收器宜丢弃按传输顺序随后所有对应于同一被分割NALU的分割单元。

终端或MANE中的接收器可以将一个NALU的前n-1个分割单元复合成一个不完全的NALU，即使该NALU的第n个分割单元尚未收到。在这种情况下，NALU头中的forbidden_zero_bit必须设为1以表明语法错误。

6 打包规则

6.1 公共打包规则

无论使用那种打包模式，所有的发送器都必须遵守以下打包规则：

- 同一编码图像的图像头或条带NALU（这些NALU有相同的RTP时间戳值）可能以AVS-P2视频标准相应档次所允许的任意顺序传送。然而，对于低延时系统，它们应按原始解码顺序传送以减小延时。
- 序列头按7.3小节给出的规则和推荐方法进行处理。
- 除了序列头或图像头NALU外，发送器（包括MANE）不能重复任何其它NALU。重复的序列头不能影响任何正在使用的序列头。重复的图像头后面必须紧跟该图像的条带NALU（不必为该图像的第一个条带）。重复宜在应用层进行而不是通过重复具有相同序列号的RTP包进行。

采用非交织模式和交织模式的发送器必须遵守以下打包规则：

- MANE可以在RTP翻译器（translator）中将多个单NALU包转换成一个复合包、将一个复合包转换成多个单NALU包或将二者混合。RTP翻译器至少宜考虑如下参数：路径MTU大小、非均匀保护机制、允许的系统延时和接收器的缓冲能力。

6.2 单 NALU 模式

当可选的媒体类型参数packetization-mode值等于0或不存在，并且没有外部方式确定使用其它打包模式时，使用单NALU打包模式。所有的接收器都必须支持这个模式。在该模式下只能使用单NALU包，而不能使用STAP、MTAP和分割单元。单NALU包的传输顺序必须与NALU解码顺序一致。

6.3 非交织模式

当可选的媒体类型参数packetization-mode值等于1或有外部方式确定使用非交织模式时，使用非交织打包模式。在这个模式下只能使用单NALU包、STAP-A和FU-A，而不能使用STAP-B、MTAP和FU-B。NALU的传输顺序必须与NALU解码顺序一致。

6.4 交织模式

当可选的媒体类型参数packetization-mode值等于2或有外部方式确定使用交织模式时，使用交织打包模式。某些接收器可以支持这个模式。在这个模式下只能使用STAP-B、MTAP、FU-A和FU-B，而不能使用单NALU包和STAP-A。包和NALU的传输顺序参加5.4小节。

7 净载格式参数

本节规定可以用来选择可选的净载格式特征和比特流的某些特征的参数。这些参数被指定为AVS-P2视频标准的媒体子类型注册的一部分。为了便于某些应用采用SDP协议（IETF RFC 2327），同时也给出了这些参数与SDP域之间的匹配。如果使用不涉及媒体类型参数或SDP的控制协议，则可以在其它地方定义等价的参数。

一些参数为接收器提供发送流的特性。这些参数的名字以“sprop”作为前缀。有些“sprop”参数受限于其它净载或解码配置参数。所有“sprop”参数的选择由媒体发送器而不是接收器完成。“sprop”参数的这个特点可能不同于一些信令协议概念（在涉及这些信令协议的情况下应避免使用这些参数）。

7.1 媒体类型注册

本注册使用RFC 4288中定义的模板并遵从RFC 3555中的格式。

AVS-P2视频的媒体子类型从IETF树中分配。接收器必须忽略任何未规定的参数。

媒体类型名: video

媒体子类型名: AVS1-P2

要求的参数: 无

可选的参数: profile-level-id, max-mbps, max-fs, max-dpb, max-br,
sprop-parameter-sets, parameter-add, packetization-mode,
sprop-interleaving-depth, sprop-deint-buf-req, deint-buf-cap,
sprop-init-buf-time, sprop-max-don-diff, max-rcmd-nalu-size

Encoding considerations: 该媒体类型是按帧分割的而且包含二进制数据

Security considerations: 见本文第8节

Interoperability considerations: 无

Public specification: 请参考本文及其第2节

Applications that use this media type: 可视电话、视频会议、网络流媒体、IPTV和视频点播等

Additional information: 无

Person & email address to contact for further information: lishuo@jdl.ac.cn

Intended usage: 普通 (COMMON)

Restrictions on usage: 该媒体类型依赖于RTP打包，因此该类型的定义仅仅为了RTP（IETF RFC 3550）传输。

File extensions: 无

Macintosh file type code: 无

Object identifier or OID: 无

Author: lishuo@jdl.ac.cn

Change controller: AVS Workgroup of China

7.1.1 可选媒体类型参数的语义

profile-level-id:

AVS-P2视频标准中规定的序列头中以下两个字节的base16表示: profile_id和level_id。

如果profile-level-id用于表明AVS-P2比特流的特性，它表明为了解码时符合AVS-P2视频标准一个解码器必须支持的档次和级别。

如果profile-level-id用于能力交换或会话建立过程,它表明该编解码器支持的档次和支持的对应该档次的最高级别。

如果profile-level-id不存在,缺省档次和级别为基准档次4.0级别。

max-mbps, max-fs, max-dpb, and max-br:

这些参数可以用来表明一个接收器的能力,而不能用作任何其它用途。包含这些参数中任何一个的能力描述必须同时包含profile-level-id参数。接收器必须完全支持profile-level-id表明的级别。如下所述,这四个参数可以用来表明接收器级别规定之外的扩展能力。

当这四个参数中多于一个参数存在时,接收器必须同时支持所有表明的能力。例如,如果max-mbps和max-br同时存在,则接收器支持由profile-level-id指定的相应的级别加上扩展的帧率和比特率。也就是说,接收器能够解码这样的AVS-P2比特流,其最大宏块处理速度(即每秒最大宏块个数)为max-mbps,最大比特率为max-br,编码图像缓冲区大小按照下面max-br的语义规定的方法导出,而其它的特性符合相应的级别。

接收器给出的这四个参数不能同时符合下列条件:符合一个比由profile-level-id指定的相应的级别更高的级别的要求。

max-mbps:

max-mbps的值是一个整数,指定每秒最大宏块个数,单位为宏块/秒。max-mbps表明接收器能够以高于profile-level-id指定的级别要求的速率进行解码。如果max-mbps存在,接收器必须能够解码这样的AVS-P2比特流:除了用max-mbps替代AVS-P2视频标准中表B.4和B.5中的每秒最大宏块个数之外,该比特流符合相应的级别。max-mbps的值必须大于等于AVS-P2视频标准中表B.4和B.5中给出的每秒最大宏块个数的值。发送器可能根据这个信息发送比相应级别指定的更高的帧率。

max-fs:

max-fs的值是一个整数,指定最大帧尺寸,单位为宏块。max-fs表明接收器能够解码大于profile-level-id指定的级别要求的图像尺寸。如果max-fs存在,接收器必须能够解码这样的AVS-P2比特流:除了用max-fs替代AVS-P2视频标准中表B.4和B.5中的每帧最大宏块个数之外,该比特流符合相应的级别。max-fs的值必须大于等于AVS-P2视频标准中表B.4和B.5中给出的每帧最大宏块个数的值。发送器可能根据这个信息发送比相应级别指定的尺寸更大的图像,此时帧率可能减小。

max-dpb:

max-dpb的值是一个整数,指定最大解码图像缓冲区大小,单位为1000比特。max-dpb表明接收器具有比profile-level-id指定的级别要求的最少解码图像缓冲区内更多的内存。如果max-dpb存在,接收器必须能够解码这样的AVS-P2比特流:除了用 $1000 \times (\text{max-dpb})$ 替代AVS-P2视频标准中表B.4和B.5中的BBV缓冲区大小(比特)等于之外,该比特流符合相应的级别。 $1000 \times (\text{max-dpb})$ 的值必须大于等于AVS-P2视频标准中表B.4和B.5中给出的BBV缓冲区大小(比特)的值。发送器可能根据这个信息生成比相应级别指定的BBV缓冲区大小能够达到的更高的压缩效率。

max-br:

max-br的值是一个整数,指定最大视频比特率,单位为1000比特。max-br表明接收器的视频解码器能够解码比profile-level-id指定的级别要求的更高的比特率。如果max-br存在,接收器的视频解码器必须能够解码这样的AVS-P2比特流:除了用 $1000 \times (\text{max-br})$ 替代AVS-P2视频标准中表B.4和B.5中的最大比特率中的最大比特率之外,该比特流符合相应的级别。 $1000 \times (\text{max-br})$ 的值必须大于等于AVS-P2视频标准中表B.4和B.5中给出的最大比特率的值。发送器可能根据这个信息发送比相应级别指定的更大的最大比特率,从而提高视频质量。

sprop-parameter-sets:

这个参数可以用来传输任何AVS-P2序列头。该参数不能用在任何能力交换过程中表明编解码器的能力。该参数的值为序列头码流的base64表示。这些序列头按解码顺序排列,每两个序列头中间插入一个半角逗号','。

parameter-add:

这个参数可以用来表明是否允许这个参数的接收者在它的信令响应中使用sprop-parameter-sets参数添加序列头，值为0或1。0表示不允许添加序列头；1表示允许添加序列头。如果该参数不存在，其缺省值为1。

packetization-mode:

这个参数可以用来表明一个RTP净载类型的特性或接收器的能力。由于只能指出一个配置点，当表明具有支持多种打包模式的能力时，必须使用多个配置点（多个RTP净载类型）。

值为0或不存在时，必须使用第6.2小节介绍的单NALU模式；值为1时必须使用第6.3小节介绍的非交织模式；值为2时必须使用第6.4小节介绍的交织模式。packetization-mode的取值范围为0..2。

sprop-interleaving-depth:

当packetization-mode不存在或packetization-mode值为0或1时，sprop-interleaving-depth必须不存在。当packetization-mode值为2时，sprop-interleaving-depth必须存在。

sprop-interleaving-depth表明NALU流的特性。它指定相对于任一NALU传输顺序在前而解码顺序在后的NALU数目的最大值。因此，可以保证接收器能够采用大于等于sprop-interleaving-depth+1个NALU的解码顺序恢复缓冲区来重建NALU解码顺序。

sprop-interleaving-depth的取值范围为0..32767。

sprop-deint-buf-req:

当packetization-mode不存在或packetization-mode值为0或1时，sprop-deint-buf-req必须不存在。当packetization-mode值为2时，sprop-deint-buf-req必须存在。

sprop-deint-buf-req表明NALU流需要的解交织缓冲区尺寸，单位为字节。该参数的值必须大于等于11.2 小节描述的解交织缓冲区需要的最大缓冲区充满度（单位为字节）。

sprop-deint-buf-req的取值范围为0..4294967295。

deint-buf-cap:

这个参数表明接收器的能力，指定接收器具有的用来重建NALU解码顺序的解交织缓冲区空间的大小（单位为字节），并表明接收器能够处理任何sprop-deint-buf-req值小于等于deint-buf-cap的流。

如果参数不存在，则deint-buf-cap的缺省值为0。deint-buf-cap的取值范围为0..4294967295。

sprop-init-buf-time:

这个参数可以用来表明NALU流的特性。如果packetization-mode值为0或1，sprop-init-buf-time必须不存在。

该参数表明接收器开始解码前从传输顺序恢复NALU解码顺序所必须经过的初始缓冲时间。假设传输是可靠的和瞬时的、传输和解码的时间线相同以及第一个包到达时解码开始，sprop-init-buf-time等于一个NALU的传输时间和解码时间之差的最大值。

下面给出了一个指定sprop-init-buf-time值的例子。NALU流的传输顺序为从左到右，数值表示解码时间：

0 2 1 3 5 4 6 8 7 ...

假设NALU传输速率恒定，则传输时间为：

0 1 2 3 4 5 6 7 8 ...

让传输时间和解码时间按列相减，得到：

0 -1 1 0 -1 1 0 -1 1 ...

因此，这个例子中sprop-init-buf-time的值为1，单位为NALU传输时间间隔。

sprop-init-buf-time为非负整数，单位为1/90000秒，取值范围为0..4294967295。如果该参数不存在，则初始缓冲时间未定义。

除了sprop-init-buf-time，接收器宜考虑各种网络单元产生的传输延时抖动。

sprop-max-don-diff:

这个参数可以用来表明NALU流的特性，但不能用来表明发送器、接收器或编解码器的能力。如果 packetization-mode 值为0或1，sprop-init-buf-time 必须不存在。sprop-max-don-diff 的取值范围为0..32767。该参数不存在表示参数值未规定。sprop-max-don-diff 的计算方法如下：

$$\text{sprop-max-don-diff} = \max\{\text{AbsDON}(i) - \text{AbsDON}(j)\}, \text{ 任意 } i \text{ 和任意 } j > i$$

其中*i*和*j*表示NALU的传输顺序索引，AbsDON表示NALU的绝对解码顺序号，即其值到达65535后继续增加而不重置为0。AbsDON的计算方法如下：

令*m*和*n*为传输顺序连续的两个NALU。对于传输顺序上的第一个NALU（其索引为0），AbsDON(0) = DON(0)；对于其它的NALU，

$$\text{If } \text{DON}(m) == \text{DON}(n), \text{ AbsDON}(n) = \text{AbsDON}(m)$$

$$\text{If } (\text{DON}(m) < \text{DON}(n) \text{ and } \text{DON}(n) - \text{DON}(m) < 32768), \\ \text{AbsDON}(n) = \text{AbsDON}(m) + \text{DON}(n) - \text{DON}(m)$$

$$\text{If } (\text{DON}(m) > \text{DON}(n) \text{ and } \text{DON}(m) - \text{DON}(n) \geq 32768), \\ \text{AbsDON}(n) = \text{AbsDON}(m) + 65536 - \text{DON}(m) + \text{DON}(n)$$

$$\text{If } (\text{DON}(m) < \text{DON}(n) \text{ and } \text{DON}(n) - \text{DON}(m) \geq 32768), \\ \text{AbsDON}(n) = \text{AbsDON}(m) - (\text{DON}(m) + 65536 - \text{DON}(n))$$

$$\text{If } (\text{DON}(m) > \text{DON}(n) \text{ and } \text{DON}(m) - \text{DON}(n) < 32768), \\ \text{AbsDON}(n) = \text{AbsDON}(m) - (\text{DON}(m) - \text{DON}(n))$$

其中DON(*i*)传输顺序索引为*i*的NALU的解码顺序号，见5.5小节。

max-rcmd-nalu-size:

这个参数可以用来表明接收器的能力，而不能用作任何其它目的。该参数值指定接收器能够有效处理的最大NALU尺寸，单位为字节。该参数值是个推荐值而并不是一个严格的上限。发送器可以产生更大的NALU但是必须知道其处理的代价可能更高。

max-rcmd-nalu-size必须为整数，取值范围为0..4294967295。如果该参数不存在，则其值为未知。发送器仍然必须考虑发送器和接收器之间的MTU尺寸。为此，发送器宜运行MTU检测手段。

7.2 SDP 参数

7.2.1 媒体类型参数与 SDP 域之间的匹配

媒体类型video/AVS1-P2字符串到SDP域的匹配方法为：

- SDP "m=" 行的媒体名必须为video（媒体类型名）。
- SDP "a=rtpmap" 行的编码名必须为AVS1-P2（媒体子类型名）。
- SDP "a=rtpmap" 行的时钟率（clock rate）必须为90000。
- 如果存在，可选参数profile-level-id、max-mbps、max-fs、max-dpb、max-br、sprop-parameter-sets、parameter-add、packetization-mode、sprop-interleaving-depth、sprop-deint-buf-req、deint-buf-cap、sprop-init-buf-time、sprop-max-don-diff和max-rcmd-nalu-size必须包括在SDP "a=fmtp" 行。这些参数的表示形式为一系列由半角分号';'分开的“参数=值”对。

下面给出了一个SDP媒体描述的例子（基准档次级别6.0）：

```
m=video 49170 RTP/AVP 98
a=rtpmap:98 AVS1-P2/90000
```

a=fmtp:98 profile-level-id=2040;sprop-parameter-sets=[SH#0]

其中[SH#0]表示一个序列头的base64表示。

7.2.2 在SDP提供/回答模型中的用法

当在单播情形下使用RTP来提供AVS-P2视频，并且在协商过程中使用SDP的提供/回答（Offer/Answer）模型时，应遵守以下规则和限制：

- 用于识别AVS-P2视频媒体格式的媒体类型参数为profile-level-id、packetization-mode和sprop-deint-buf-req（如果packetization-mode等于2）。这三个参数必须对称地使用：即回答者必须要么保留所有的配置参数，要么在无法支持一个或多个参数值的情况下完全去除该媒体格式（净载类型）。为了简化这些配置的处理和匹配，提供中使用的RTP净载类型值应同样在回答中使用。除非由上述三个参数决定的配置相同，回答中不能包含提供中用到的净载类型值。
- 媒体类型参数sprop-parameter-sets、sprop-deint-buf-req、sprop-interleaving-depth、sprop-max-don-diff和sprop-init-buf-time用于指示提供者或回答者发送的具有某一配置的NAL单元流的特性。这有别于提供/回答参数的一般用途——指示提供者或回答者能够接收的流的特性。对于AVS-P2视频，提供者假定回答者能够接收采用所提供配置编码的媒体。
- 能力参数max-mbps、max-fs、max-dpb、max-br和max-rcmd-nalu-size可以用来表明其它能力。对它们的解释取决于方向属性。如果方向属性为sendonly，则这些参数描述发送器能够产生的RTP包和NALU流的限制。如果方向属性为sendrecv或recvonly，则这些参数描述接收器能够接受的限制条件。
- 如前所述，提供者必须在提供中包含交织的AVS-P2视频流的解交织缓冲区的尺寸。为了使提供者和回答者都能够告诉对方自己在解交织缓冲方面的能力，推荐二者都包含deint-buf-cap。这个信息可以在第二轮提供和回答过程中选择了sprop-deint-buf-req的值之后使用。对于交织流，如果接收器的能力未知，推荐考虑提供多个缓冲需求不同的净载类型。
- 参数sprop-parameter-sets的使用如前所述。另外，回答者必须在它的回答中保留所有从提供中收到的序列头。如果parameter-add为0，回答中不能增加任何序列头；如果parameter-add为1，回答者可以在它的回答的sprop-parameter-sets中增加序列头。无论parameter-add的值为0或1，回答者必须答应接收采用包含在自己的回答中的sprop-parameter-sets的视频流。

对于通过组播方式发布的流，必须额外遵守以下规则：

- 回答者不能改变流特性参数sprop-parameter-sets、sprop-deint-buf-req、sprop-interleaving-depth、sprop-max-don-diff和sprop-init-buf-time。因此，一个提供的净载类型要么被原封不动地接受，要么被去除。
- 回答者必须对所有申明为sendrecv或recvonly的流支持接收器能力参数max-mbps、max-fs、max-dpb、max-br和max-rcmd-nalu-size。否则，回答者必须去除该媒体格式或拒绝会话。

下面完整地列出各参数在不同情形下的意义：

- 在“a=sendrecv”或未使用方向属性或“a=recvonly”的提供和回答中各参数的意义如下：
申明要接收的配置或特性：

- profile-level-id
- packetization-mode

申明要发送的流的特性（仅当“a=sendrecv”或未使用方向属性时适用）：

- sprop-deint-buf-req
- sprop-interleaving-depth
- sprop-parameter-sets
- sprop-max-don-diff
- sprop-init-buf-time

申明接收器实现的能力：

- max-mbps
- max-fs
- max-dpb
- max-br
- deint-buf-cap
- max-rcmd-nalu-size

申明如何进行提供/回答协商：

- parameter-add

- 在“a=sendonly”的提供和回答中各参数的意义如下：

申明提议要发送的流的配置和属性：

- profile-level-id
- packetization-mode
- sprop-deint-buf-reg
- sprop-max-don-diff
- sprop-init-buf-time
- sprop-parameter-sets
- sprop-interleaving-depth

申明发送器接收流时的能力：

- max-mbps
- max-fs
- max-dpb
- max-br
- deint-buf-cap
- max-rcmd-nalu-size

申明如何进行提供/回答协商：

- parameter-add

另外，还需要考虑下面的因素。

- 申明接收器能力的参数通常是可以降低等级的，即它们表达发送器可能的行为的上限。因此发送器可以选择设定它的编码器只使用这些参数的更小（低）或相等的值。sprop-parameter-sets不能用来申明发送器的能力。
- 除了profile-level-id的级别（level）部分以外，申明配置点的参数不可以降低等级。这个表明了接收器期望使用的值，发送器端必须原封不动地使用这些值。
- 如果申明了发送器的能力，并且申明中使用了不可降低等级的参数，则这些参数表明了一个可接受的配置。为了达到更好的互用性，可以提供多个可替换的配置，例如不同的打包模式。由于不可能采用同一个净载类型提供多个配置，如果提供多个配置，每个配置需要自己的RTP净载格式。
- 接收器应当理解一个净载格式中所有的媒体类型参数，即使它只支持该净载格式的一部分功能。这保证了接收器能够理解一个提供的流可以被降级到自己能够支持的程度。
- 回答者可以扩展额外的媒体格式配置到提供中。然而，为了能够使用这些额外的配置，大多数情形下需要提供者的第二个提供，以给出媒体发送器要使用的流特性参数。这个过程还有另外一个效果，就是提供者必须能够接收这个媒体格式配置，而不仅仅是发送。
- 如果提供者希望发送和接收的能力不对称，则它必须提供不同的RTP会话，即分别申明为recvonly和sendonly的不同SDP “m=” 行。

7.2.3 在申明性的会话描述中的用法

如果AVS-P2视频RTP流的SDP提供方式为申明性的（例如在RTSP中），则需要考虑下面的因素。

- 所有能够表明NAL单元流和接收器特性的参数都用来表明NAL单元流的特性。例如，在这种情况下，参数profile-level-id申明流使用的值，而不是发送器的能力。因此，参数的意义如下：

申明配置或特性：

- profile-level-id
- sprop-parameter-sets
- packetization-mode
- sprop-interleaving-depth
- sprop-deint-buf-reg
- sprop-max-don-diff
- sprop-init-buf-time

不可用的：

- max-mbps
- max-fs
- max-dpb
- max-br
- max-rcmd-nalu-size
- parameter-add
- deint-buf-cap

- SDP接收器需要支持提供的所有参数和参数值，否则接收器必须拒绝（RTSP）该会话。会话创建者应负责采用接收者期望能够支持的值。

7.3 序列头考虑

序列头对AVS-P2视频编解码器的操作起关键作用。鉴于其对解码过程的重要性，序列头的丢失或错误传输造成的影响很难在接收端进行隐藏。引用一个错误的序列头通常会对解码过程造成非常严重的后果。这种错误可能是由于错误传输或丢失一个序列头数据，也可能是由于未及时传输序列头更新。因此，下面为RTP发送器的实现推荐了一些指导原则。

序列头数据的传输可以一下通过三种不同的原则：

- 采用会话控制协议在实际RTP会话之前带外（out-of-band）传输。
- 采用会话控制协议在实际RTP会话过程中带外（out-of-band）传输。
- 在实际RTP会话过程中作为RTP流的一部分带内（in-band）传输。

A和B原则要在会话控制协议中实现。C原则由本部分定义的RTP净载格式支持。

A原则用于带外传输整个序列的初始序列头，B原则用于进行带外的序列头更新，C原则用于进行带内的序列头更新。

在会话过程中采用A原则带外传输初始序列头，用B或C原则更新序列头。初始序列头的带外传输能够保证解码器至少能有一个序列头可用，当新的序列头到来后再进行更新。如果使用B原则更新序列头，则无法确定该序列头与带内传输的编码视频帧之间的同步关系，容易在发送端和接收端引起混乱，因此此时推荐使用C原则进行序列头的更新。

8 安全考虑

采用本部分定义的净载格式的RTP包遵守IETF RFC 3550以及任何适当的RTP档次，如IETF RFC 3551，中讨论的安全考虑。这意味着媒体流的保密性通过加密得到，例如采用IETF RFC 3711。因为本净载格式对数据压缩的采用是端到端的，加密需要在压缩之后进行。

采用导致接收端计算量不一致的压缩技术的数据编码存在着遭受拒绝服务攻击的风险。攻击者可以向流中注入病态的数据，这些数据的解码复杂，可以导致接收器超负荷。AVS-P2对这种攻击尤其敏感，因为很容易产生一些NALU以影响后续很多其它NALU的解码过程。因此，推荐至少对RTP包进行数据源鉴别和数据完整性保护，例如采用IETF RFC 3711中的方法。

注意，保证RTP包及其净载的保密性和完整性的适当机制取决于具体应用和相应的传输与信令协议。因此，尽管上面的例子中给出了IETF RFC 3711，其它的选择也是可能的。

端到端的安全如鉴权、完整性或保密性保护使得MANE除丢弃完整的包以外不能进行其它基于媒体感知的操作。在保护保密性的情形下，甚至基于媒体感知丢弃包也不能进行。为了允许任何MANE进行操作，要求允许该MANE是包含在安全环境建立过程中的可以信任的实体。

9 拥塞控制

RTP拥塞控制的采用必须根据IETF RFC 3550以及任何适当的RTP档次，如IETF RFC 3551。如果采用尽力服务模式，一个额外的需求是：净载格式的使用者必须监视包丢失状况，以保证丢包率在可接受的参数范围内。在下面这种情况下认为丢包率可以接受：如果一个TCP流在相同的网络条件下通过同一网络路径在合理的时间范围内达到不低于RTP流达到的平均吞吐量。这个条件可以通过以下方式满足：实现调节传输率或分层多播会话订购层数的拥塞控制机制，或当丢包率过高时安排一个接收者离开会话。

遵守拥塞控制需要的比特率调节在实时编码情况下很容易实现。然而，在传输预先编码的内容时，带宽调节需要存在同一内容的不同比特率的编码描述，或比特流中存在非参考图像。不同描述之间的切换通常可以在同一RTP会话中进行，如在I帧处。仅当不可降低等级的参数，如档次级别中的档次，需要改变时，才需要结束再重新开始媒体流。这个可以通过使用不同的RTP净载类型实现。

10 IANA 注册

将请求IANA注册一个新的媒体类型，见7.1小节。

11 解包过程（资料性）

解包过程取决于具体实现。因此，下面的描述仅是一个合适的例子，其它方法也可以采用，也可以对所介绍的算法进行优化。11.1小节介绍单NALU和非交织打包模式的解包过程；11.2小节给出交织打包模式的解包过程；11.3小节介绍额外的接收器解包指导原则。

所有与缓冲区管理有关的常用RTP机制适用于此。尤其是，重复的和过时的RTP包（根据RTP序列号和时间戳）将被丢弃。为了确定确切的解码时间，必须考虑为了适当的媒体流之间的同步而故意允许的时延等因素。

11.1 单 NALU 和非交织模式

接收器包含一个接收缓冲区以补偿传输时延抖动。接收器按接收顺序将RTP包存入接收缓冲区。解包按RTP序列号顺序进行。对于单NALU包，解包得到的NALU直接传递给解码器。对于STAP-A，解包得到的NALU按它们在包中出现的顺序传递给解码器。对于FU-A，解包得到的分割单元与以前解包得到的同一NALU的分割单元（如果存在）进行拼接，如果拼接的结果为完整的NALU则传递给解码器。

11.2 交织模式

这些解包规则的概念实际上就是将NALU从传输顺序重排序到NALU解码顺序。

接收器包含一个接收缓冲区以补偿传输时延抖动。本节在假设没有传输时延抖动的情况下描述接收器操作。为了区别于实际中同时用于补偿传输时延抖动的接收缓冲区，这里称之为解交织缓冲区。

11.2.1 解交织缓冲区尺寸

如果在会话建立时采用了SDP提供/回答模型或任何其它能力交换过程，接收到的流的特性不宜超过接收器的能力。在SDP提供/回答模型中，接收器可以通过媒体类型参数deint-buf-cap表明其分配解交

织缓冲区的能力。发送器通过参数sprop-deint-buf-req表明对解交织缓冲区尺寸的需求。因此，推荐将解交织缓冲区的尺寸（单位为字节）设为大于等于sprop-deint-buf-req的值。

如果在会话建立时采用了申明性的会话描述，参数sprop-deint-buf-req仍然表明对解交织缓冲区尺寸的需求。因此，同样推荐将解交织缓冲区的尺寸（单位为字节）设为大于等于sprop-deint-buf-req的值。

11.2.2 解交织过程

接收器有两种缓冲状态：初始缓冲和回放缓冲。初始缓冲发生于RTP会话初始化时。初始缓冲后，解码和回放开始，此时的缓冲状态为回放缓冲。

无论缓冲状态如何，接收器按接收顺序将包存入解交织缓冲区的方法如下：复合包中的NALU逐个存入解交织缓冲区；每个NALU的DON值被计算并保存下来。

借助于下面的函数和常数，接收器的操作描述如下：

- 函数AbsDON的定义见7.1小节。
- 函数don_diff的定义见5.4小节。
- 常数N等于媒体类型参数sprop-deint-buf-req的值加1。

直到下列条件之一成立，初始缓冲结束：

- 解交织缓冲区中有N个NALU。
- 如果sprop-max-don-diff存在，则don_diff(m, n)大于sprop-max-don-diff的值，其中n表示接收的NALU中AbsDON值最大的NALU，而m表示接收的NALU中AbsDON值最小的NALU。
- 初始缓冲的持续时间等于或超过sprop-init-buf-time的值。

确定要从解交织缓冲区中移出的NALU的方法为：

- 如果解交织缓冲区中有至少N个NALU，则按下面的顺序移出NALU直到缓冲区中只剩下N-1个NALU。
- 如果sprop-max-don-diff存在，所有使得don_diff(m, n)大于sprop-max-don-diff的NALU按下面的顺序移出解交织缓冲区并传递给解码器，其中m表示待测NALU，n表示解交织缓冲区中AbsDON值最大的NALU。

上面提到的将NALU从解交织缓冲区中移出并传递给解码器的顺序为：

- 令变量PDON在RTP会话开始时初始化为0。
- 对每个具有DON值的NALU计算DON距离如下：如果该NALU的DON值大于PDON值，则DON距离等于(DON - PDON)；否则，DON距离等于(65535 - PDON + DON + 1)。
- NALU按DON距离增加的顺序从解交织缓冲区中移出并传递给解码器。如果几个NALU具有相同的DON距离，它们可以以任何顺序从解交织缓冲区中移出并传递给解码器。
- 当解交织缓冲区中只剩下N-1个NALU时，PDON值设为最后一个从解交织缓冲区中移出并传递给解码器的NALU的DON值。

11.3 额外的解包指导原则

可以采用下面的额外解包指导原则来实现实际的AVS视频解包器：

- RTP接收器（如网关）可能检测丢失的分割单元。如果发现了丢失的分割单元，网关可以决定不传送同一分割的NALU的后续分割单元，因为它们对AVS-P2视频解码器已经没有意义。这样，MANE可以无需解析复杂的比特流丢弃无用包，从而减小网络负载。
- 在必须丢弃RTP包或NALU的情况下，接收器应首先丢弃所有NALU头中NRI值为0的包或NALU。这样将减小对用户的影响并保留参考图像。如果需要丢弃更多的包，宜先丢弃PSI中NRI值较小的包，然后再丢弃NRI值较大的包。然而，丢弃任何NRI值大于0的包都可能引起解码错误传播，因此宜避免。