

视频编码与通信



视频通信系统主要组成：

1. 视频信号的摄取与表示 视频压缩编码 视频信道编码
2. 视频信道解码 视频压缩解码 视频信号的分解和提取 图像显示

Chap1 视频信号基础

1. 视频感知基础

光谱：

电磁波，波动性，微粒性

单色光：7种色（自然界中没有单色光，谱色光）

可见光：波长380nm-780nm

颜色

色彩来源：自身发出，反射

不同光源照射颜色不同，同样的颜色可以是不同的光谱（同一波长或者不同波长混合）

色温（K，光源波谱效应）

色温:绝对黑体的加热温度(光源光谱和绝对黑体辐射光谱相同)，与光源功率与色温无关，表征光源波谱特性。

绝对黑体：不反射光，不透射光。完全吸收入射光，温度高，辐射能量大，功率谱向短波移动，亮度增加，发光颜色改变

人眼

杆状细胞：边缘，灵敏度高，彩色不敏感，低照度时分辨明暗，夜晚

锥状细胞：中心，对彩色敏感，分辨力高，白天

视细胞在视网膜分布不均，视细胞在视网膜上分布不均匀，黄斑区锥状细胞最多，成像最清晰

光谱光效率(光度量，视敏函数， $V(\lambda)$)

同一波长不同强度，不同波长同一强度-->亮度不一样

$K(\lambda)$ 光效能，人眼对某波长光的灵敏度

$\lambda = 555nm$ （黄绿色）光效能最大，最亮最灵敏，最迟钝的是蓝色和紫色

对 $K(\lambda)$ 作归一化

$$V(\lambda) = \frac{K(\lambda)}{K(555)}$$

光的度量

光通量， 发光强度， 照度， 亮度

亮度（B）， 单位烛光， 图像的亮度取决于图像的平均直流成分

对比度： 最大亮度与最小亮度的比值

灰度： 亮度级差或亮度层次

色彩三要素

锥状光敏细胞对红黄绿敏感

色彩三要素： 色调， 色饱和度， 亮度

类别 深浅

色调和色饱和度统称为色度

三个颜色的混合比例决定了混合色的色调和饱和度， 亮度为三基色亮度之和

相加混色(光)： 时间， 空间， 生理混色（RGB加色坐标系）

相减混色（染料， 印刷）： 颜色染料吸色性质（CMY减色坐标系）

色彩坐标系：

HIS坐标系

Hue（色调）， 饱和度， 亮度（Intensity）

CIE色度图（1/3, 1/3）处代表 E 白

RGB： 穿过中心的直线相交两点相加为白色

XYZ记色制： Y表示亮度， XYZ比例表示色度， X=Y=Z时为白色

亮度方程： $Y = 0.3 R + 0.59 G + 0.11 B$

Y,R,G,B只需要传送三个， 一般传送Y, (R-Y)色差, (B-Y)色度。因为G-Y太小了

黑白图像： 色度为0， RGB相等

彩色图像， RGB中一个或两个为0， 饱和色

2. 人眼视觉特性HVS

亮度感觉

不同环境亮度下，人眼对同一亮度的主观感觉不同，视觉范围：百分之几 nit 到几百万 nit ($10^{-4} \sim 10^6 \text{cd/m}^2$) 不能同时感受这么大的亮度范围

主观亮度与亮度的对数成正比

对比度 $C = L_{\max}/L_{\min}$ 可见度阈值 ΔL_{\min}

L 大， ΔL 大，很大范围内 $\Delta L/L$ 为常数

(时间空间) 不均匀背景， ΔL 大

韦伯比

$$\text{韦伯比 } \xi = \frac{\Delta B_{\min} (\text{最小可分辨亮度差})}{B (\text{背景亮度})} \text{ 为一常数}$$

分辨率

人眼对被观察物体上能分辨的相邻最近的 两点视角 θ 的倒数

空间分辨率高，时间分辨率低，反之亦然

图片很清晰，亮度变化不敏感

黑白比彩色敏感

大面积着色原理

原理：人眼对彩色的分辨力比黑白低

方法：利用亮度信号取代颜色细节

色调分辨阈值：480~640色调分辨力最高

降低彩色分辨率：4:1:1, 4:2:2

亮度为4，色差和色度为1或者2

马赫带效应

亮区边缘暗，暗区边缘亮

亮度突变的地方会产生过冲作用，增强轮廓

原因：视神经的侧抑制现象，这一点受亮度刺激，会使邻近细胞灵敏度下降

视觉惰性和闪烁感觉

视觉滞留：0.05-2s 电视，电影的基础：断续放映，得到连续运动感觉

闪烁：当周期性光脉冲刺激人眼时，若脉冲占空比低，重复频率不够高，会有闪烁感。在屏中大面积亮区尤为严重

临界闪烁频率：

$$f_c = a \lg L_m + b$$

在临界闪烁频率以上，主观感受的亮度等于光脉冲亮度的平均值：

$$S = \frac{1}{T} \int_0^T L(t) dt$$

人眼视觉模型

彩色视觉掩蔽：亮度变化剧烈，对色彩变化的敏感度就降低了。（数字水印）

人眼对0，90度方向刺激最敏感，对角线变弱，45和135度时最低

JND：恰可察觉差

3. 数字视频格式

像素的亮度，色度和色饱和度即是空间的二维函数，同时也是时间函数

帧传送

一幅图像每个像素传送一遍成为一个帧处理

并行传输无法实现，根据视觉惰性，按一定顺序把每个像素的亮度转换成电信号传输，接收端按顺序把电信号转换成相应亮度的像素

收发方必须要同步，传送时间小于视觉暂留时间

扫描：

图像连续，帧频不低于32Hz，不闪烁，高于48Hz

频幕亮度提高，尺寸变大，观看距离缩进，帧频应相应提高

摄像与显像：

摄像管：光电转换，显像管：电光转换

分辨率与灰度级

计算机不能处理模拟信号-->离散信号

分辨率：512*512 Pixel多少个像素

灰度级：256级，28级，8bit

重建

数字图像重建成连续图像

时域：内差 频域：周期函数截取

采样：

频率越高，信号越逼近元信号，冗余信息多，工作量存储空间大。

若点数给定，采样时间短，一些信息会被排除在外。

采样频率低，间隔远，高频信息损失，无法复原，信号混淆。（棋盘效应）

一维取样分析：

$$\text{取样函数: } s(x) = \sum_{m=-\infty}^{+\infty} \delta(x - m\Delta x) \quad S(u) = \frac{1}{\Delta x} \sum_{m=-\infty}^{+\infty} \delta\left(u - m \frac{1}{\Delta x}\right)$$

$$\text{取样信号: } f_s(x) = s(x)f(x) = \sum_{m=-\infty}^{+\infty} f(x)\delta(x - m\Delta x)$$

频域：原信号频谱与S(u)的卷积

取样：频谱进行周期搬移

二维取样：

$$\delta(x, y) = \lim_{n \rightarrow \infty} [n^2 \text{rect}(nx) \text{rect}(ny)]$$

y 在 $[-\frac{1}{2n}, \frac{1}{2n}]$ 以及 x 在 $[-\frac{1}{2n}, \frac{1}{2n}]$ 范围内时

矩形函数有非零取值1。

矩形函数 $\text{rect}()$ ：

$$\text{rect}(\alpha) = \begin{cases} 1, & |\alpha| \leq \frac{1}{2} \\ 0, & |\alpha| > \frac{1}{2} \end{cases}$$

$$\begin{aligned}
F_s(u, v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\Delta x} \frac{1}{\Delta y} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta\left(\alpha - m \frac{1}{\Delta x}\right) \delta\left(\beta - n \frac{1}{\Delta y}\right) F(u - \alpha)(v - \beta) d\alpha d\beta \\
&= \frac{1}{\Delta x} \frac{1}{\Delta y} \sum_m \sum_n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta\left(\alpha - m \frac{1}{\Delta x}\right) \delta\left(\beta - n \frac{1}{\Delta y}\right) F(u - \alpha)(v - \beta) d\alpha d\beta \\
&= \frac{1}{\Delta x} \frac{1}{\Delta y} \sum_m \sum_n F\left(u - m \frac{1}{\Delta x}\right) F\left(v - n \frac{1}{\Delta y}\right)
\end{aligned}$$

二维取样的频谱是原始信号那个的频谱沿u轴和v轴周期重复的结果

取样实际问题：

噪音的带宽达，因此噪音欠采样，重建产生混叠

解决方案：采样前先滤波

二维采样定理

若 $\frac{1}{\Delta x} > 2W_u, \frac{1}{\Delta y} > 2W_v$ ，则没有混迭。（即：采样间隔足够小），
再用低通滤波器只滤出，就可完全恢复出原始信号。

——二维取样定理

只要间隔足够小，就可以完全恢复出原始信号。

重建

常用内插函数：sinc

0阶样本内插：方形函数

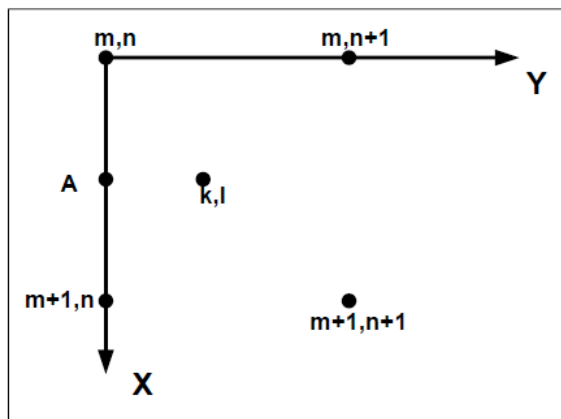
1阶样本内插：三角形函数 2个方形函数的卷积

2阶样本内插：钟形函数 3个方形函数的卷积

3阶样本内插：4个方形函数的卷积，两段具有连续性和平滑性（B样条）

双线性内插：

$$\begin{aligned}
f(k, l) &= (m+1-k)(n+1-l)f(m, n) + (k-m)(n+1-l)f(m+1, n) \\
&\quad + (m+1-k)(l-n)f(m, n+1) + (k-m)(l-n)f(m+1, n+1)
\end{aligned}$$



n个特殊点:

- “角点”上: 该点权为 1, 其他三点权为 0
- “中心”: 四点权均为 $\frac{1}{4}$, 即4点平均
- “边的中点”: 如 A点, 为上、下二个样点灰度值的平均

重建实际问题: 采样孔径可能过大, 造成反向

- 若 $\frac{1}{\tau} \gg \overline{W}_u$, 接近理想采样 (注意噪声问题)
 若 $\frac{1}{\tau} > \overline{W}_u$, 造成边缘模糊
 若 $\frac{1}{\tau} < \overline{W}_u$, 可能造成“反相”

量化

人眼对灰度误差有个敏感度阈值, 量化层次多到一定程度, 量化误差就可小于视觉阈值

量化误差过大, 会出现伪轮廓

最常见的是256级量化, 即用1Byte表示一个样本值。对高保真的图像有10/12bit量化的要求

量化过粗会导致有用信息丢失, 后续增强处理没有意义

1. 均匀量化

当随机变量的概率密度函数为常数时, 我们将取值范围均匀的分成k个子区间, 取每个子区间左右两个端点值的平均值作为这个子区间的值

$$\therefore \varepsilon^2 \propto \frac{1}{k^2}$$

均匀量化会引入量化失真 (非线性失真)

2. 非均匀量化

概率小的分的粗一点，量化区间长一点

每个区间的值应该为

$$q_i = \frac{\int_{Z_i}^{Z_{i+1}} zP(z)dz}{\int_{Z_i}^{Z_{i+1}} P(z)dz}$$

根据这个可以找到

$$\Rightarrow Z_i = \frac{1}{2}(q_i + q_{i-1})$$

迭代算法 (MAX)

①输入 $Z_0, Z_k, P(Z)$ ，分层总数 k ；

②任选一个 q_0 的值，取 $i=0$ ，由 $q_i = \frac{\int_{Z_i}^{Z_{i+1}} ZP(Z)dZ}{\int_{Z_i}^{Z_{i+1}} P(Z)dZ}$ 求出 Z_1 ；

③取 $i=1$ ，由 $Z_i = \frac{1}{2}(q_i + q_{i-1})$ ，求出 q_1 ；

④按照②③步做法，由面积左右相等，由 q_1 求出 Z_2 ；由 Z_2 是 q_1q_2 的中点，求出 q_2 ；直到求出 q_{k-1} ；

⑤若 q_{k-1} 是区间 $[Z_{k-1}, Z_k]$ 内 $P(Z)$ 的重心。即满足 $q_{k-1} = \frac{\int_{Z_{k-1}}^{Z_k} ZP(Z)dZ}{\int_{Z_{k-1}}^{Z_k} P(Z)dZ}$ ，

则所有 Z_i, q_i 均已确定，若 q_{k-1} 不满足上述条件，则重新选定 q_0 的值，重复②-⑤步，直到满足条件为止。

视频文件格式 (many)

视频信号ITU-BT.R.601

RGB在摄像机中变成Y, Cb, Cr

$$Y = (0.30 * R) + (0.59 * G) + (0.11 * B)$$

$$C_B = 0.56 * (B - Y)$$

$$C_R = 0.71 * (R - Y)$$

亮度Y是Cb或Cr分辨率的2倍，4:2:2

码流中的开始和结束由SAV和EAV标记

AV到SAV之间有水平空白间隔，传输辅助信息 SAV和EAV码字各自由4个8或10bit码字组成，第一个码字是全1，第二和第三个码字全0，第四个码字包含相应场或相应垂直空白间隔的信息

从0~9位，最高位第9位

Chap2 视频编码基础

1. 视频压缩理论基础（一堆信息论。。。）

2. 视频信号基本特性

图像分辨率：

垂直/水平分辨率：沿垂直和水平方向所能分解的像素数

图像信号直流分量

图像的最低频率是直流，景物的背景亮度

图像信号带宽

知道最高频率就能确定带宽

最高频率出现在全是细节的情况下

水平分解力：水平方向上有几个像素点

- 水平分解力为 N ，则每个像素的时间为 t_d
- 摄像输出的信号近似为正弦波，周期为 $2t_d$ ，因而图像信号的最高频率 f_{\max} 为：

$$f_{\max} = \frac{1}{2t_d} = \frac{N}{2T_{\text{HS}}} = \frac{583}{2 \times 52 \times 10^{-6}} \approx 5.6\text{MHz}$$

图像信号特征

相邻两行或相邻两帧信号之间有较强的相关性，差别极小，可近似认为周期信号

模拟信号的频谱

a. 静止图像

- 只在水平方向有亮度变化：线状谱
 - 位于行频 f_H 及其谐波 nf_H 上。 n 越大，能量越小
- 垂直方向也有亮度变化：离散谱线簇(谱线群)
 - 频谱成份为 $nf_H \pm mf_V$ ，主谱线为 nf_H
 - n 或 m 越大能量越小

b. 活动图像

- 以行频及其谐波为中心,形成梳齿状的离散频谱
- 黑白图像信号的主要能量分布在视频信号的低频

无论静止或活动图像，能量主要分布在以行频及其各次谐波频率为中心的 较窄范围内

3. 视频图像评价标准

平均亮度，对比度（最大亮度/最小亮度），清晰度，分解率（分辨率）

图像质量评价

图像质量可分成图像逼真度和图像可懂度

图像逼真度：图像处理前后的相似度（每个点在变换前后误差的太小）

图像可懂度：图像能被人理解的程度

客观评价

PSNR:
$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f(m,n) - g(m,n)]^2}$$

与主观评价一致性更好：

MPSNR（视觉掩蔽效应）：

$$MPSNR = 10 \log_{10} \frac{255^2}{E^2}$$

MSE

介绍了三种方法：

- 比较处理前后图片损伤（精度高，处理复杂），例如：ITU T J.144-2001,
- 在原始图像中加一些信息，比较重建后的图像中提取出来的信息，估计损伤，可以克服图像经过传输系统后接受的不一致性

- 根据码流，压缩算法估算损伤，不需要从源视频处获取信息，但与主观评价相关性较小

基于HVS的图像客观质量评价：

基于**视觉感知**的测量方法：**JND**（刚辨差), eg:PQA300,计算出的客观PQR（Picture Quality Rating）和人眼主管评价之间的一致性 可以达到0.88相关性

PQA500， 差值平均估计分DMOS (Differential Mean Opinion Score)， 不再将损伤图象与人眼所能接收的最差图象(training sequence)比较，用于分辨率低的场合，**实时性差**

基于**视觉兴趣**的测量方法：**ROI**（感兴趣区）

块效应(DVQ)

离散余弦傅立叶变换会导致块效应，通过计算相邻像素之间Y,Cb, Cr差值可以用来检测块效应。

对每个宏块，Y分量每行求得16个像素差值，共16行；然后每列求得16个像素差值，共16列；块

每行内所有宏块的像素差值求和，每行共得到16个值。一帧内每行的16个值再求和，每帧共得到16个值。这16个值最终作为所有宏块水平和垂直方向的0到15的平均像素差值

SA

需要考虑**空间活动性**以及时间活动性对块效应的干扰

精细结构的图像，具有很高的空间活动性

单色的平滑图片对应的SA为0

最大SA为水平和垂直方向均为白黑像素间隔

TA

总体反应连续帧的变化或运动

最大的TA对应于连续帧内所有像素都由白变黑，由黑变白

TA=0，图像序列没有运动

主观评价：

打MOS分（找人打分）

ITU-R B T.500 主观影响太大，无法适应编码以及显示技术的飞快发展

原理上有两种主观评价方法：

- DSCQS(double stimulus continual quality scale)：双刺激连续质量 分级法
- SSCQE(single stimulus continual quality evaluation)：单刺激连续 质量评价方法

有时候两者会冲突，我们偏向于主观评价，然而主观评价成本太高，我们希望得到与主观评价契合度比较好的客观评价方法

影响图像质量的因素

模拟电视：

1. 视频信号干扰：长距离传输时高频损失，噪声引入
2. 编解码过程：模拟视频转分量视频，磁带多版复制等

数字电视：

1. 编解码过程
2. 不同码流，编码格式，压缩率
3. 信道误码

DTV信号图像质量的影响因素与模拟电视信号不同

模拟电视信号噪声：雪花

数字电视噪声：

- 信道误码率增加，大部分误码可以被恢复。如果信道噪声太大，会造成信道突然中断
- 线性或非线性失真不会对声音和图像有直接影响，在极端情况会造成完全传输中断
- 不需要VITS(vertical insertion test signal)来检测线性和非线性失真，不需要黑电平信号来检测噪声，也不需要有关传输连接的测试信号

改善图像质量的方法

1. 低码流网络播出：H.264比MPEG-2质量好
2. 高码流，MPEG-2全I真编码质量好，提高制作效率
3. 播出时，长GOP的MPEG-2编码，不牺牲图像质量的情况下提高压缩率

视频编码框架

摄像机输出RGB通过矩阵变换得到Y,Cb,Cr 信号

降低信号带宽，信号通过低通滤波

模数转换,色度分辨率降为4:2:2

ITU-BT.R.601信号码率270Mbit/s

必须压缩到2~6(15)Mbit/s,压缩比高达130

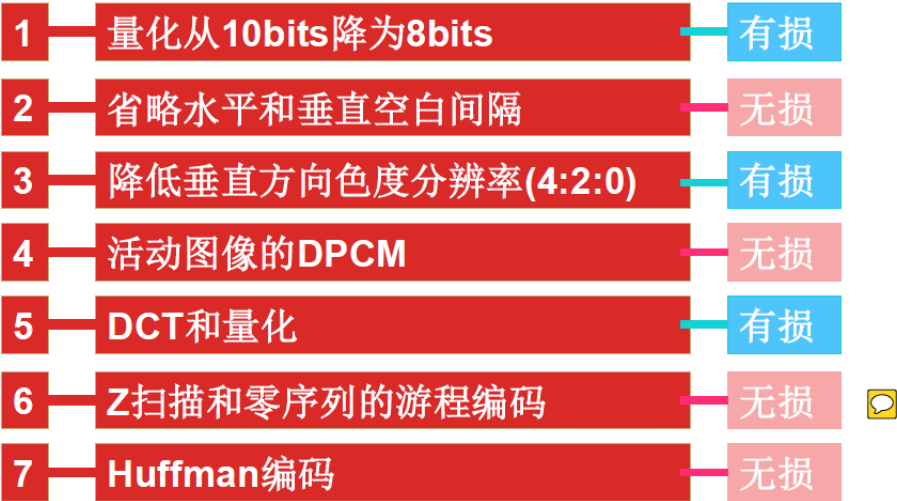
压缩：

无损压缩：冗余度去除 eg: VLC游程编码，霍夫曼编码

有损压缩：不相关性消除

基于假设：不必要的信息，色彩视觉细胞比亮度视觉细胞少很多，人眼不能分辨过于精细结构和粗结构

MPEG采用的视频压缩步骤



1、10bits量化降为8bits量化

模拟电视中，视频信号的信噪比大于48dB时，噪声恰好人眼不可见，从10bit到8bit，节约码率20%，量化噪声增加12dB

2. 省略水平和垂直空白间隔

水平：节约19%，垂直节约8%

总节约25%

3、降低垂直色度分辨率(4:2:0)

节约50%

Chapter3 基于块的变换编码

图像变换

图像的大部分能量都集中在低、中频段，高频分量很弱，仅仅体现了图像的某些细节

根据做变换的变换基和信号有没有关系分为固定核（DFT，DCT，W-H变换）或者可变核（K-L变换）

DCT变换的最大特点就是**图像能量主要集中于少数低频DCT系数上（左上角）**，为图像压缩打下基础（去掉右下角的系数，高频分量）。

正交变换会导致方块效应-->可用滤波器来平滑

图像变换应用

Chapter8 传输控制

1. 传输同步

媒体资源可通过对各传输通道进行传输，需要 相关的内容进行同步

时延偏差

不同发送端的媒体流具有不同网络时延，导致媒体间不同步

- 让媒体流在不同时刻进行发送
- 接收端缓存对时延偏移进行补偿

时延抖动

发送端到接收端网络时延的变化，多源于网络中队列

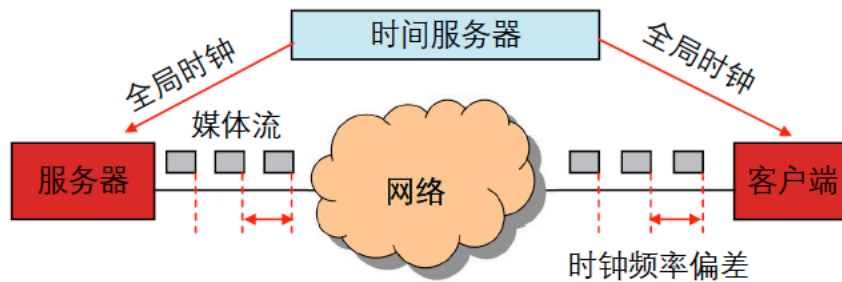
- 接收端缓存

时钟频率偏移

无全局时钟时收发端本地适中频率可能存在偏差

使用时钟同步协议

- 解决策略
 - 在网络中使用时钟同步协议（网络时间协议NTP）



连续媒体内部的同步

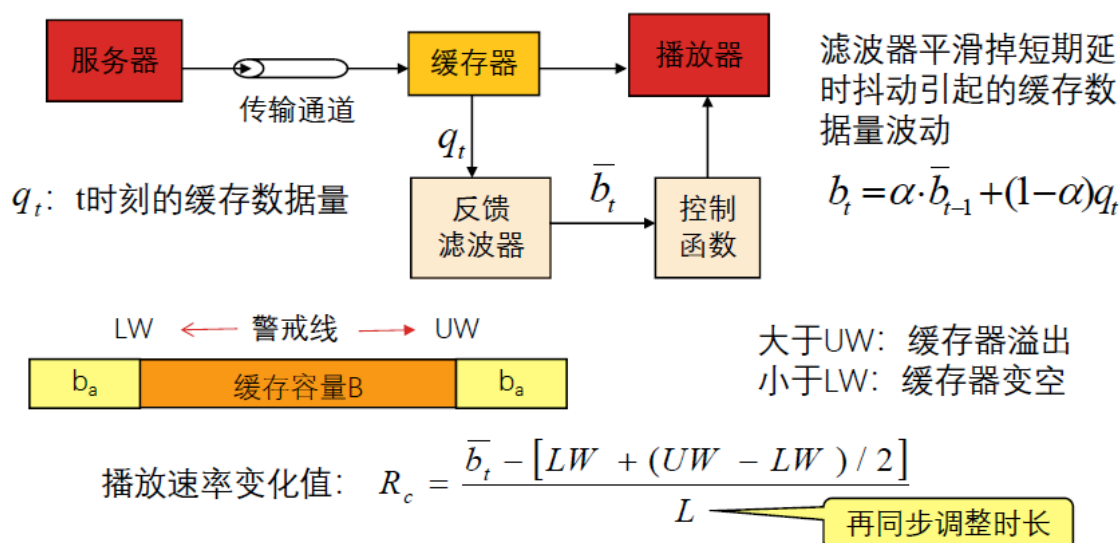
连续媒体是由若干个逻辑数据单元LDU构成的时间序列

网络抖动导致连续媒体内部LDU的相互时间间隔发生变化

处理方法：接收出设置缓存，延时抖动进行过滤

技术：

- 基于播放时限的同步方法
 - 接收端接收到第一个LDU之后，延迟一段时间再开始播放
 - 播放延迟时间：最大延时-最小延时
- 基于缓存数据量控制的同步方法
 - 周期性检查数据缓存量，如果缓存数据量超过预定警戒线，认为不同步



媒体流之间的同步

主流：选择一条媒体LDU流时间轴为基准

从流：其他媒体流为从流

保证主流播放速率不变，调整从流播放速度

一般选择音频为主流，视频为从流

1. 基于全局时钟的时间戳同步技术

各媒体按时间顺序分成LDU，每个单元由一个时间戳，所有信源（服务器）与接收端需要进行全局时钟同步

接收端具有相同时间戳的媒体但愿同时进行表现

2. 基于反馈的流间同步机制

相对时间戳RTS:

各媒体流LDU序列距离该序列第一个LDU其实是可的相对标记

不同媒体流需要同步播放的LDU的RTS应该一样，相同RTS的LDU被服务器同时发送

根据用户反馈的RTS，改变从流发送速率

拥塞控制

最早的网络拥塞：1986年10月 仅距离400码LBL(Lawrence Berkeley Laboratory)到UC Berkeley之间的数据吞吐量从32kbps降到了40bps

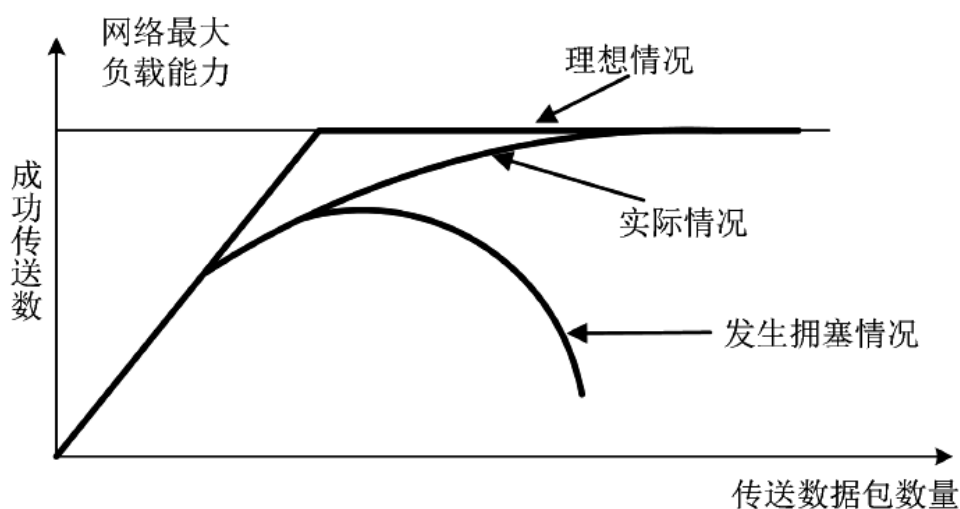
互联网中10%的数据包会被丢弃

95%数据流使用TCP/IP

拥塞控制的根本原因：

端系统提供给网络的负载超过了网络容量与处理能力

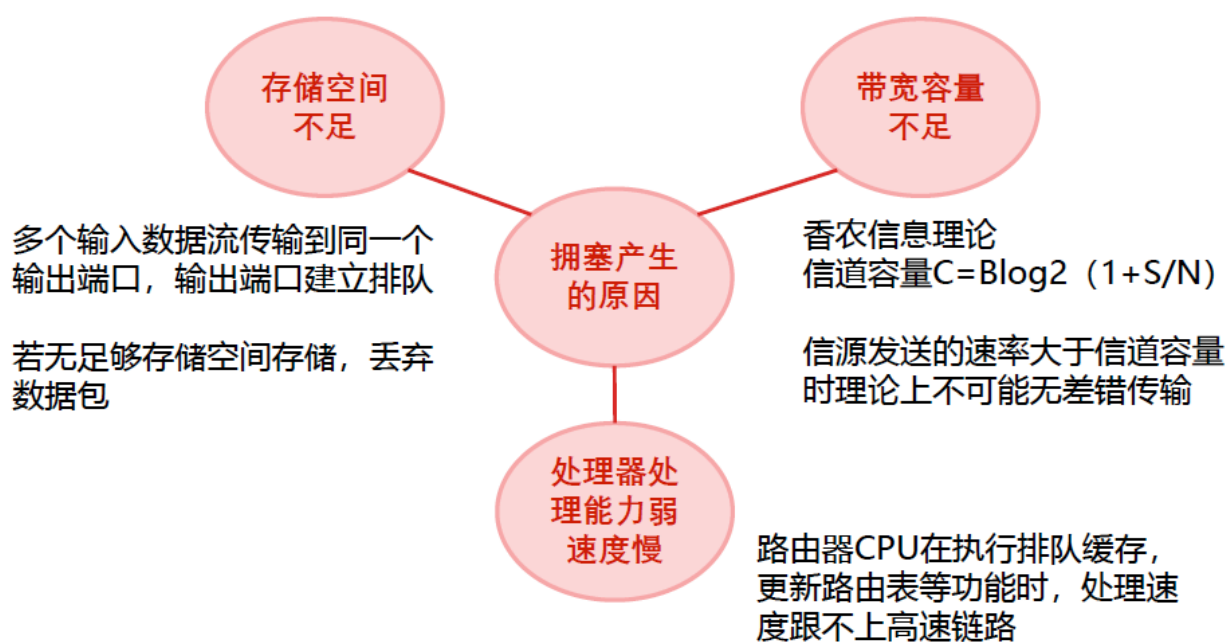
导致数据包时延增大，丢包率增大，上层应用系统性能下降



理想情况：buffer无限大

发送给路由器的包速率大于路由器转发的速率

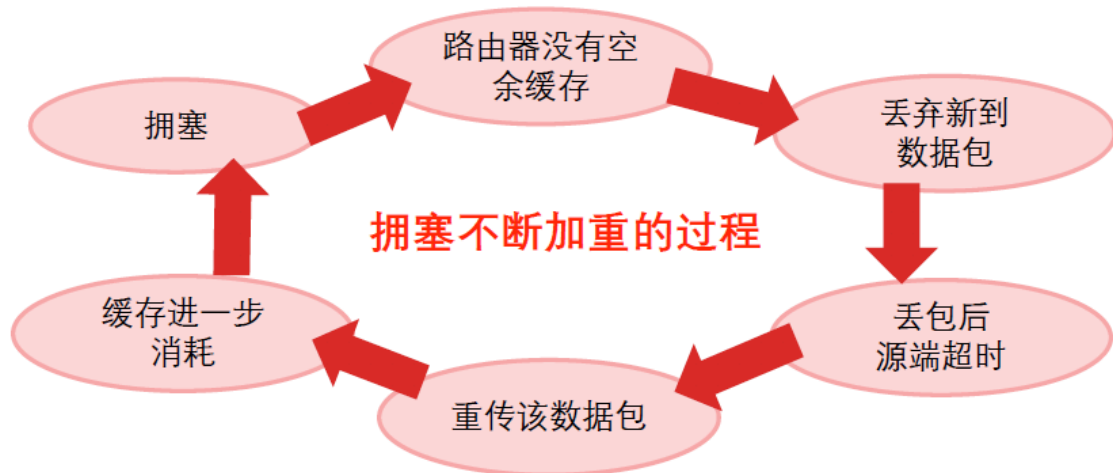
队列增大至超过缓存最大值，发生拥塞，无法缓存的数据包被丢弃



避免拥塞

- 需要综合考虑三点原因避免拥塞

- 提高链路速率 —————> 网络瓶颈转移到处理器上
- 增大缓存空间 —————> 排队越来越长，时延增大



拥塞检测

显示拥塞检测：IP层

路由器告诉发送端产生了阻塞，降低发送速率

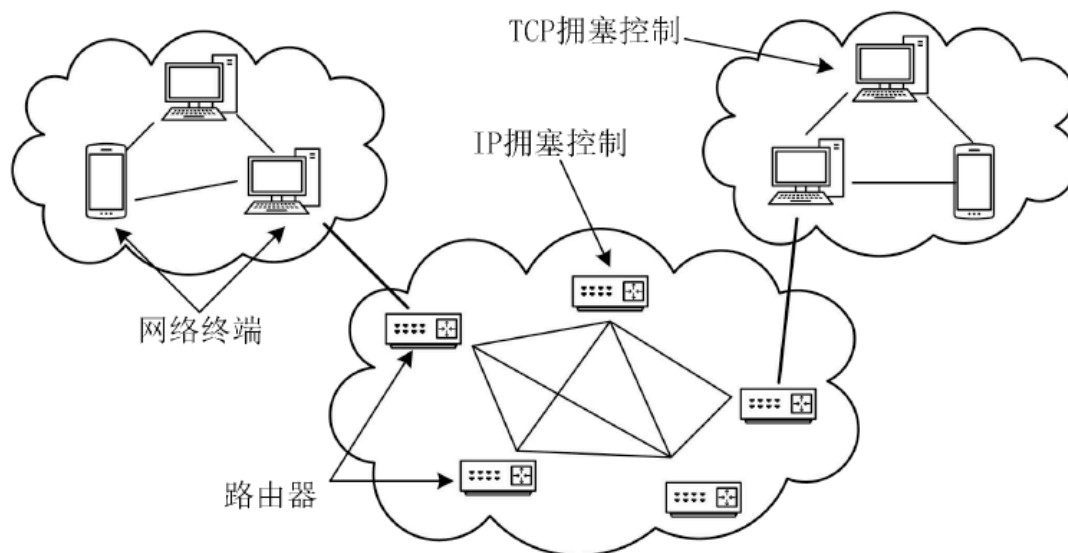
隐式拥塞指令：

用丢包、时延的信息作为发生拥塞的标志

拥塞种类：

端到端：传输层（收发段系统）

网络内部拥塞控制（网络层/IP层）：作用于网络路由器中



传输层拥塞控制

▪ TCP

- 有拥塞控制
- 基于窗口的拥塞控制
- 窗口：每当新的反馈收到时，发送端就将一定量的数据包（或比特数）进行发送
- 优点：“数据包守恒原则”
- 缺点：不适合传输流媒体

▪ UDP

- 无拥塞控制
- 给使用UDP的传输加上拥塞控制
- 通过应用层添加控制功能实现
- 基于速率的拥塞控制
- 优点：持续发送数据，根据反馈动态变化

TCP

tcp报文段首部由16位的窗口大小

Chapter10 典型视频编码标准

MPEG

1. MPEG-1

目标：1.5Mbit/s获得VHS质量（Video Home System）

应用于存储媒体（CD和硬盘），不应用于广播

分辨率：CIF（352*288）

视音频复用的PES包组成的PS流，只有一个节目

2. MPEG-2

存储应用的同时考虑传输

TS流针对传输层，提供更小的包结构，更扩展的复用和抗误码机制，最多支持20个节目

支持图像分辨率为SDTV，HDTV

MPEG-1和MPEG-2基本算法框架一致：I，B，P帧，前向 /后项预测，DCT变换

VCD和 DVD

1. VCD：

- 一个MPEG-1 PS流
- 总码率：1.5Mbit/s

2. SVCD

- MPEG-2数据流
- 总码率：2.4Mbit/s，质量近似于VHS质量，甚至更高

3. DVD：

- MPEG-2编码
- 码率：10.5Mbit/s
- 质量比VHS高得多，可以提供字幕，其他信息
- 应用广泛，视频，音频，数据
- 与CD（700M）相比，存储容量从4.7G到17.1G，可以有 1，2，或4层，每层4.7G

HD DVD与蓝光DVD

蓝光DVD使用波长更短的蓝色激光（405nm），减小了凹槽长度和道间距，能记录下DVD多出数倍的数据

HD-DVD也是用蓝色激光（405nm），单道间距更大，不及蓝光DVD存储容量大

MinDv与DVPro对比

MinDv： 家庭视频摄像机，总码率36MB/s，视频码率25MB/s

图像大小720*576，帧率： 25fps

可以在码流任何位置编辑，只有类似I帧组成

DVPro：

演播室标准，支持视频码率： 25~50Mbits

两者共同点：

- 都是用DCT编码技术，没有帧间编码，都是M-JPEG的特例
- 与MPEG相比，没有量化表，也没有宏块层控制的量化参数，只有一系列本地量化表，由编码器选择对每个宏块最合适的表
- 相对码率较高，质量好，易于处理

Standard	Video coding	Resolution	Video data rate [Mbit/s]	Total data rate [Mbit/s]
MPEG-1	MPEG-1	352X288 192X144 384X288	0.150- (1.150) -3.0	最大约3.5 1.4112
MPEG-2	MPEG-2	720X576 (SDTV,25fps) 最大HDTV	最大15	基本开放 最大270
VideoCD	MPEG-1	352X288	1.150	1.4112
SVCD	MPEG-2	480X576	2.4	2.624
VideoDVD	MPEG-2	720X576	最大9.8	10.5
MiniDV	M-JPEG特例	720X576	25	近似30
DVPro	M-JPEG特例	720X576	25/50	近似30/55

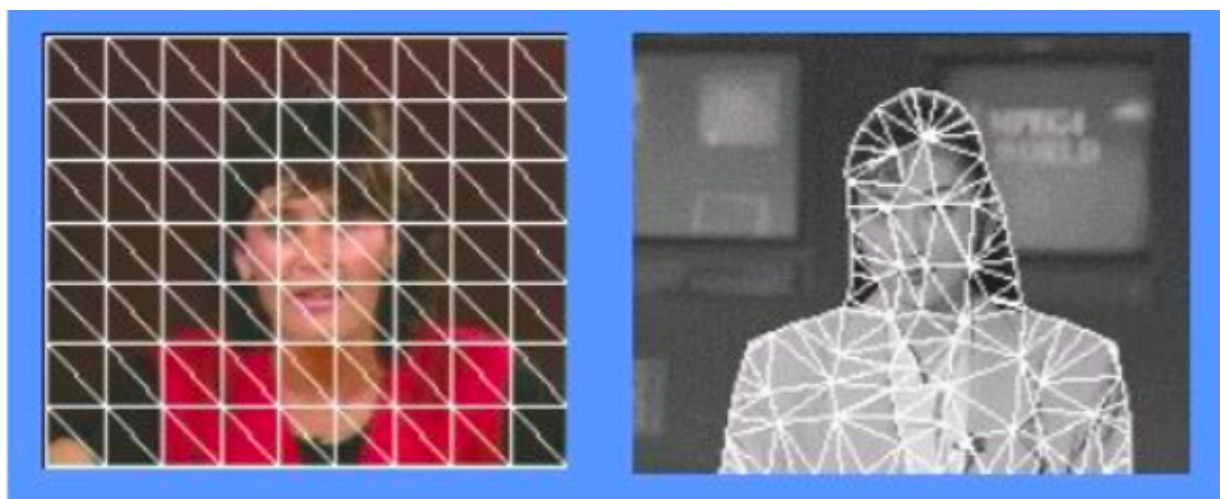
MPEG-4

1. 基于对象编码

把图像中的物体分割出来作为视频对象，一视频对象为单位进行编码

2. 基于模型编码

特殊的基于对象的编码，对象除了边界存在形状特调整外，对对象的纹理区域也存在一定结构，这样结构信息通过网格模型内的节点和变来表示



3. 支持Sprite编码

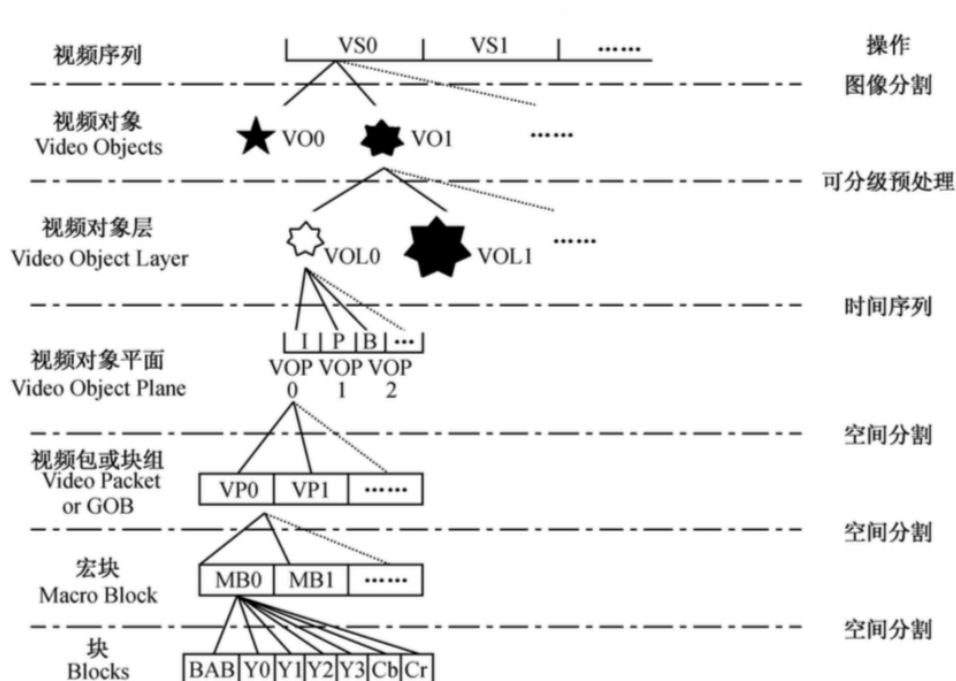
通过图像镶嵌技术把整个序列的背景图像拼接成一个大的完全背景变化

首先对Sprite图像编码传输一次并存储在解码端，随后图像只需传输摄像机运动参数就可以恢复背景

编码效率高，主要面向视频会议和视频电话等场景固定的应用

MEPG-4层次结构

MPEG-4视频的层次结构



MPEG-7 (1+2+4)

实现对多媒体内容的搜索

定义一种标准的多媒体内容的接口，规范对多媒体内容的描述

MPEG-21

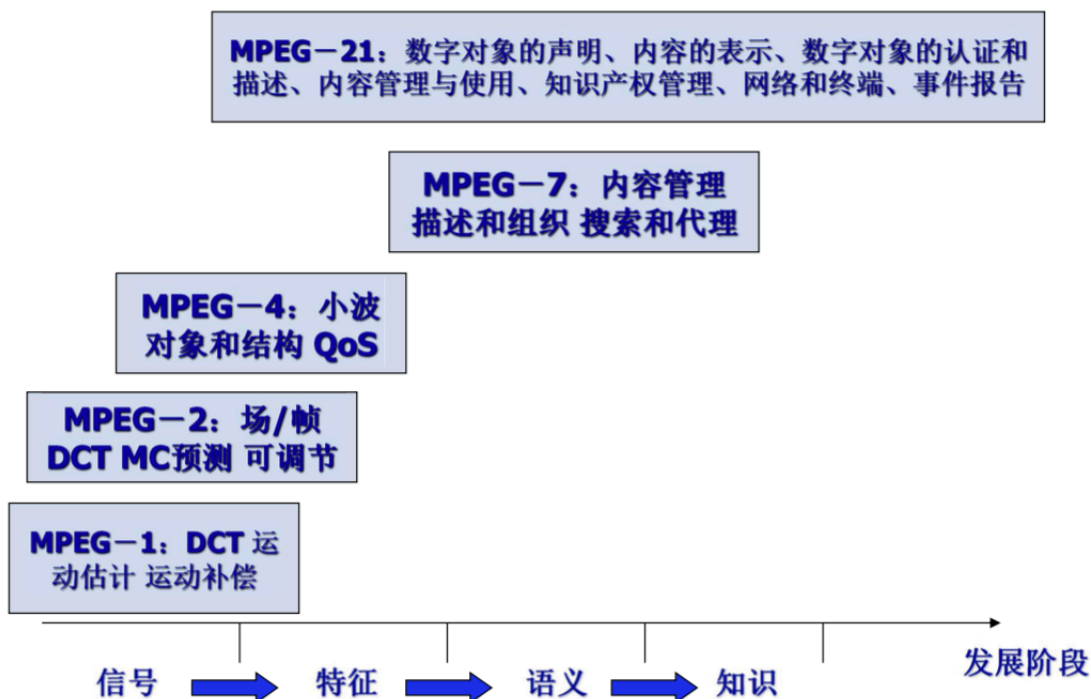
包含所有其他MPEG标准的补充工具和方法

如何将不同的技术和标准结合在一起，需要什么标准，完成不同标准的结合工作

广播，多媒体和Internet趋向融合

- 广播一点到多点的码率比Internet高得多
- Internet可以很好地利用有反馈新到的广播信道
- 在纯粹Internet信道中，码率低，视频质量一般，传输开销大，时常不稳定
- ADSL高码率只能在交换中心和用户之间，不能再整个网络

MPEG对比图



H.26X系列标准

H.261

第一个实用的数字视频编码标准

之后的视频编码国际标准都是基于H.261相同的设计框架（MEPG-1/2， H.262,H.263,H.264）

使用了混合编码框架，包括运动补偿的帧间预测，DCT，量化，z扫描，熵编码

基本单位：宏块，使用YCbCr颜色空间，4:2:0

是ITU-T在综合业务数字网ISDN上开展双向声像业务制定的

- 速率64kb整数倍
- CIF，QCIF（分辨率）

- 压缩编码：运动补偿的帧间预测，DCT，熵编码
- 每帧图像分为图像层，宏块层（GOB），宏块（MB），块（Block）

H.263

ITU-T的一个草案，为低于64kb/s的窄带通信信道定制的基地码率视频编码标准，在很多应用中可以取代H.261

编码算法和 H.261一样，提高了性能和纠错能力

H.261和H.263之间的区别

H.263在低码率下能够提供比H.261更好的图形效果，两者区别：

- H.263运动补偿：半像素精度，H.261:全像素精度以及环路滤波
- H.263可以选择数据流层次结构中的某次部分，可以编解码可以配制成更低的数据率或更好的纠错能力
- H.263有4个可协商选项可以改善性能
- H.263采用无限制的运动向量以及基于语法的算数编码
- 采用事先预测和与MPEG相同的PB模式
- H.263支持物种分辨率，除了H.261中锁支持的QCIF和CIF，还有SQCIF，4CIF，以及QUIF，SQCIF相当于QCIF的一半分辨率，4CIF，16CIF为CIF的4倍和16倍

H.264

或称MPEG-4第十部分（高级视频编码部分），有ITU-T和ISO/IEC第二次联手开发的视频编码标准

低码率的情况下32kbps的H.264相当于128kbps的MPEG-4图像质量

应用：网络电视、广播电视、数字影院、远程教育、会议电视

在设计标准和实现功能上做了进一步改进，同等视频质量条件下，节约50%的码率

H.265/HEVC

比起H.264/AVC相比，H.265提供了更多的编码工具

- 宏块大小从264中固定16*16变成最小8*8到64*64
- 帧内预测模式支持33种方向，H.265只支持8种
- 提供了更好的运动补偿和矢量预测方法

相同视频质量，265比264平均减少50%码流

H.265帧内预测模式

- skip模式，运动合并模式和先进运动矢量预测（AWVP）

- 基于DCT变换的变换编码和熵编码
- 去方块滤波，像素自适应补偿技术
- 并行处理（帧，片，瓦片，波前）

AVS（国家标准）

AVS1.0标准特点：

- 性能比H.264略高，MPEG-2高两倍
- 复杂度低，算法复杂度比H.264低
- 实现成本低,软件和硬件实现成本都比H.264低
- 专利授权费用低

核心：

8x8整数变换、量化，帧内预测、1/4精度像素插值、特殊的帧间预测运动补偿、二维熵编码、去块效应环内滤波等

编码工具集

AVS



- | | |
|-----------|----------|
| ▪ 双向预测 | ▪ 熵编码 |
| ▪ 隔行编码 | ▪ 环路滤波 |
| ▪ 运动矢量预测 | ▪ 缓冲区管理 |
| ▪ 子像素插值 | ▪ 其它 |
| ▪ 多参考帧预测 | ▪ 图像组头 |
| ▪ 可变块大小预测 | ▪ 防伪起始码 |
| ▪ 帧内预测 | ▪ 码流顺序 |
| ▪ 变换和量化 | ▪ 时间参考索引 |

编码工具集

AVS2

压缩效率比AVS+和H.264/AVC提高了一倍

支持：三维视频、多视角和虚拟现实视频的高效编码;立体声、多声道音频的 高效有损及无损编码;监控视频的高效编码;面向三网融合的新型媒体服务

VP8和VP9

VP8

开放视频压缩格式，最早由On2 Technologies 开发，随后由 Google 发布，同时 Google 也发布了VP8编码的实做库:libvpx

具体技术：

空间域帧内预测，时间域帧内预测技术

与H.264相似：变换，量化，熵编码，环路滤波，只是具体技术细节实现略有不同

VP9

Google在2013年6月份开发并投入使用，混合编码

Chapter11 典型视频广播标准

MPEG-2 TS传输协议

MPEG-2 Transport Stream协议是ISO/IEC协议 13818-1协议族的第一部分：系统部分

- 主要用于广播系统的多媒体节目封装&传输，支持多路节目复用
- 目前还被很多广播系统广泛应用

基本层次：

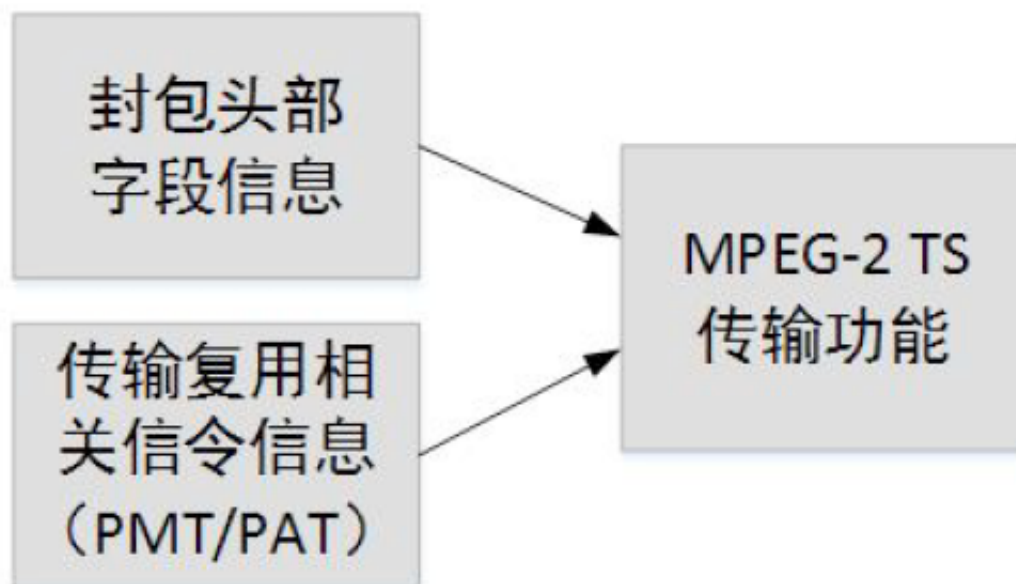
- 节目（program）
- 基本流（elementary stream, ES）
- 包化基本流（Packetized Elementary Stream, PES）
- 传输流（Transport Stream, TS）

上层由下层组成

内容封装&传输封包

TS协议传输功能：

- 内容封装：媒体内容的逻辑组织形式，包括节目，基本流等
- 传输封包：PES包和TS包的封装结构，包头中的字段规定了传输的顺序，优先级等多种信息



PES包结构

- 访问单元加上PES包头形成PES包
- AU一般是一个视频帧或者一段音频数据
- PES包=PES包头+PES负载（音视频帧/AU）

packet_start_code_prefix: 指示一个PES包的起始处

stream_id: 流标识符, 标志这个PES包属于基本流的类型的标号

PES_packet_length: 指示PES包的长度

PES_header_data_length: 指示PES包头的长度

PTS

$$PTS(k) = ((system_clock_frequency \times tp_n(k)) DIV 300) \% 2^{33}$$

DTS

$$DTS(k) = ((system_clock_frequency \times td_n(k)) DIV 300) \% 2^{33}$$

TS包

PES切割成TS包, TS包长度为188字节

封包格式: header+payload

具体自己看ppt吧111

信令&复用机制

信令: 节目特定信息 (PSI), 说明了节目复用, 消费, 网络配置等信息

- 表格形式
 - 节目关联表 (PAT)
 - 节目映射表 (PMT)
 - 条件访问表 (CAT)
 - 网络信息表 (NIT)
- 大的信令表格被分成多个较小的分段 (PSI section)
- PSI section可以被多个TS包携带

节目复用: 把多个基本流复合成一个单节目流

传输复用: 把包含单个节目的多个TS流复合成一个包含多个节目的复合TS流

TS技术小结

- ISO/IEC 13813协议族的系统部分 (主要用于广播系统中多媒体数据传输)
- 传输封装 (PES包+TS包) 和信令组成了TS协议的基本架构
- 复用机制将不同节目的数据交织起来, 一路数据多个节目
- 目前被广泛使用

FLUTE背景

ALC（异步分层编码）协议：内容传输协议

- 通过一个发端可同时向百万个收端传输内通，各手段可随时介入
- 大规模可扩展组播分发基础协议
- 缺点：仅传输对象，不知道也不存储对象属性相关信息（接收端需知道接受的对象是什么）
- 在ALC基础上定义了FLUTE

FLUTE协议定义了FDT，用于存储传输对象的属性信息（制定机构：IETF）

FLUTE简介

DVB-H：移动电视标准，定义了透过广播网络传送IP封包的方法

DVB-IPDC：一种IP层之上的移动电视服务平台标准的集合

FLUTE基于DVB-IPDC CDP标准中负责DVB-H广播网络上传送档案或其他非实时行的数据的通信协议

FDT

file delivery table：在这个表中收端能知道要接受哪些信息

FLUTE在ALC会话管理功能基础上引入的文件属性映射机制（精髓）

ALC只完成了任意二进制对象的传输，不关心传输对象是什么

FDT提供了文件传输会话内传送与文件有关的各种属性信息的机制

FDT实力被看成一个特殊的文件同其他文件对象一起在传输会话中传输

FLUTE	单向多点传送协议	传输“对象”（object）	传输对象不具有任何属性
ALC	单向多点传送协议	传输“档案”（file）	FDT中记录传输对象属性



- FDT中包含属性信息如下所示：
- 文件传输相关属性
 - 标示文件的TOI (Transport Object ID) 值
 - FEC-OTI (Object Transmission Information) 信息
 - 传输对象的大小
 - 发送数据包到所有通道的数据聚合率 (Aggregate rate)
- 文件自身相关属性
 - 名称，标识，文件位置 (URI)
 - 文件媒体类型
 - 文件大小
 - 文件编码方式
 - 文件信息摘要

RTP

RTP：实时传输系统，用于实时语音，视频数据传输的标准协议（针对多媒体数据流的传输协议）

RTCP：实时传输控制协议

传统的传输协议注重高可靠性，RTP注重数据传输的实时性

流媒体

- 使音频和视频形成稳定和连续的传输流和回放流的一系列技术、方法和协议的总称
- 从网络上获取音频和视频等多媒体数据的新方法，能够支持多媒体数据流的实时传输和实时播放
- 优点：
 - 启动延时短
 - 系统缓存容量需求降低
 - 可以传输不知道大小的媒体数据

RTP：

- 一对一或一对多
- 提供时间信息，实现流同步，实时数据传输
- 工作在UDP或TCP或ATM协议上

RTP只保证实时数据的传输，不可靠，不提供流量控制或拥塞控制（RTCP）

负责对流媒体数据进行封包，实现媒体流的实时传输

RTP位于传输层之上，应用程序之下

实时语音，视频数据经过处理送给RTP封装成RTP数据单元，RTP数据单元被封装成UDP包，UDP包再给网络层封装成IP包

RTCP：

- 流量控制
- 拥塞控制
- 会话源管理

向会话中所有成员周期发送控制包，对服务质量进行控制，对网络状况进行诊断

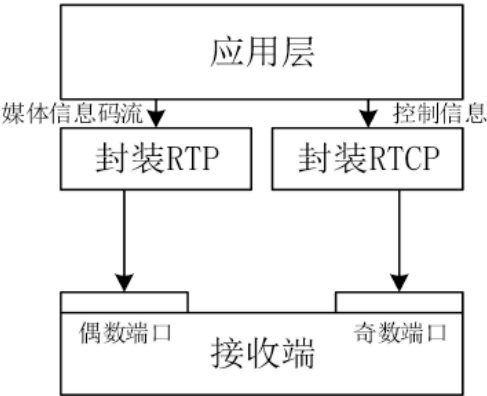
五种包：

RTCP的控制包信息包括5种数据包类型：

类型	缩写表示	用途
200	SR（Sender Report）	发送端报告
201	RR（Receiver Report）	接收端报告
202	SDES（Source Description Items）	源点描述
203	BYE	结束传输
204	APP	特定应用

RTP工作机制

- **RTP工作机制：**
- 当应用程序建立一个RTP会话时，应用程序将确定一对目的传输地址。目的传输地址由一个网络地址和一对端口组成，有两个端口：一个给RTP包，一个给RTCP包，使得RTP/RTCP数据能够正确发送。RTP数据发向偶数的UDP端口，而对应的控制信号RTCP数据发向相邻的奇数UDP端口（偶数的UDP端口 + 1），这样就构成一个UDP端口对。 RTP的发送过程如下，接收过程则相反。



广播

用户被动接受流

数据包的单独一个靠背发送给网络上的所有用户，用户端接受流，不能控制流（不管用户需不需要）

优点：简单

缺点：浪费网络带宽

单播

在客户端和媒体服务器之间建立一个单独的数据通道

优点：便于控制和管理

缺点：每个用户必须对媒体服务器发送单独的查询，媒体服务器必须向每个用户发送所申请数据包的拷贝，服务器负担沉重，响应时间很长

组播

- 允许路由器一次讲数据覆盖之到多个通道
- 一次只需要发送一个数据包
- 所有发出请求的客户端可同时收到连续数据流无延时
- 网络传输数据包总量变少

优点：减少网络上传户的信息包总量，网络利用效率大大提高，成本大为下降

缺点：不同用户同时点播同一个节目，后点播的用户不是从头播放，二是一招网络中同时点播节目的其他用户播放进度，当前用户极有可能从那个节目的中间开始看起，很难做到个性化

Chapter12 流视频OTT传输系统

1. 流媒体

渐进流：下载与播放并行，不具备快进的功能

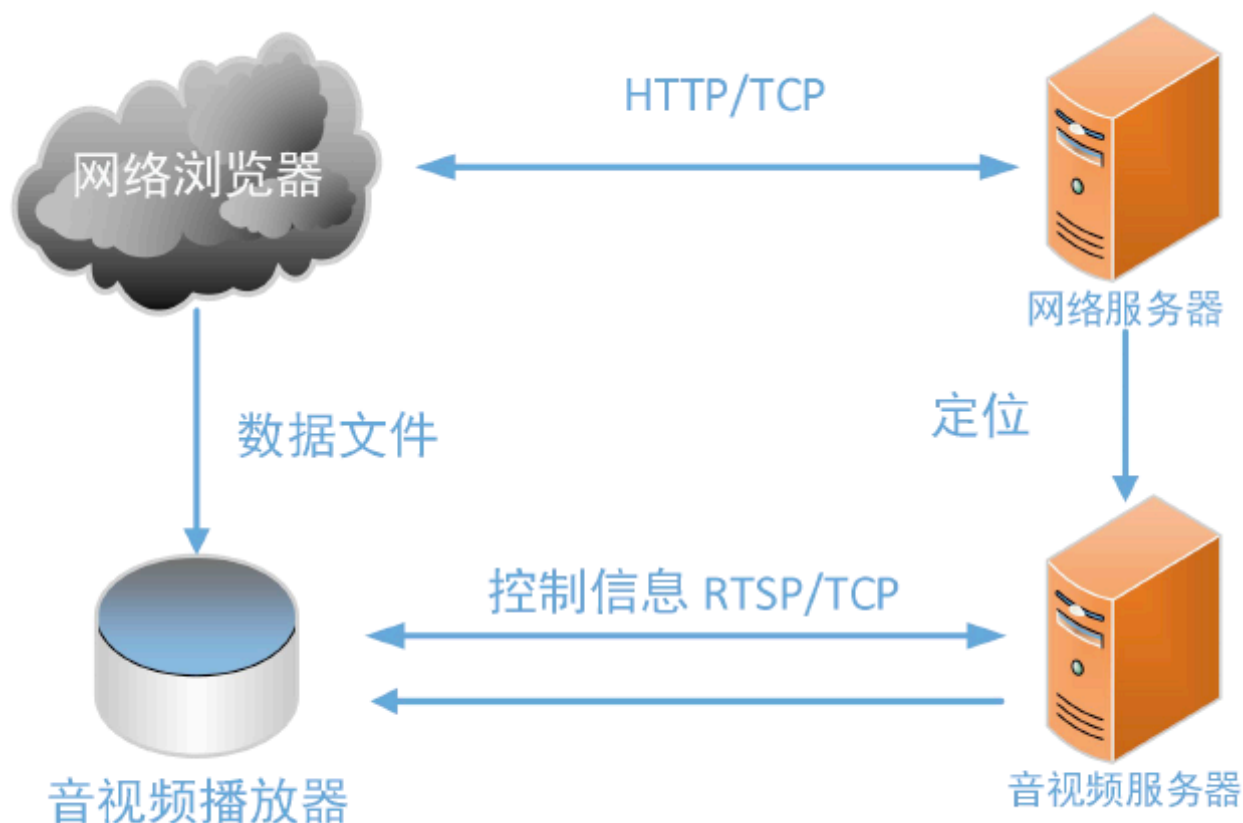
动态自适应流：先由一个片段进入缓存，根据网络决定后面传输视频的参数

流媒体三巨头：

微软：Media Player：支持.asf以及.wmv

Realnetworks: real alternative : 支持.rm

苹果：QuickTime：支持.mov



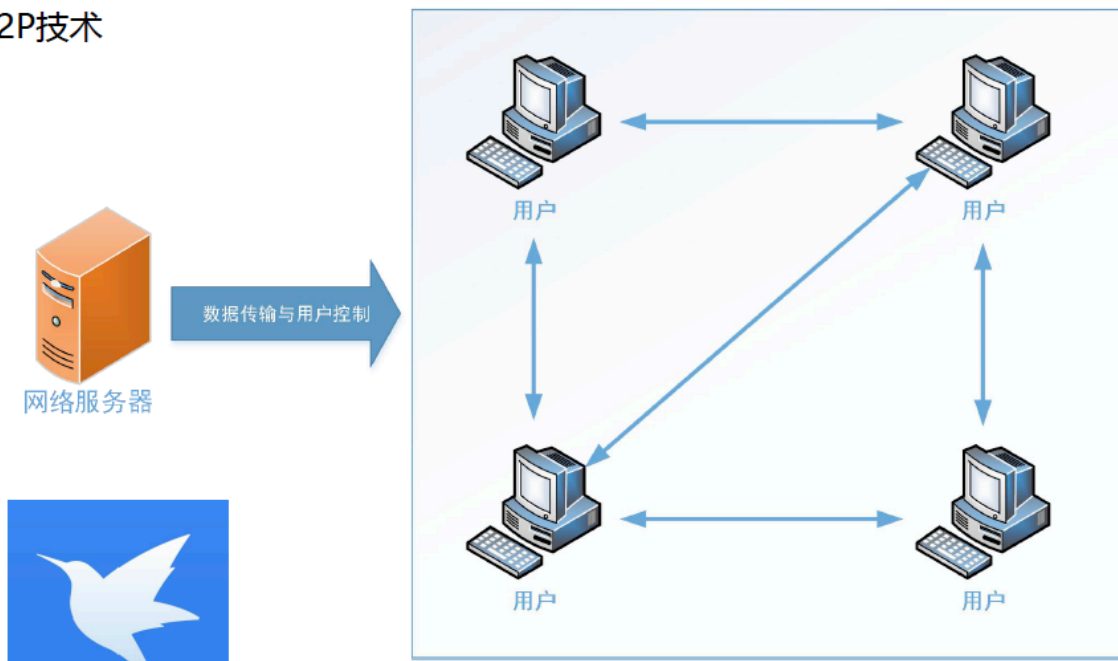
1. 浏览器以及服务器通过HTTP/TCP交换控制信息（编码类型以及相关服务器地址）
2. 运行RTSP，交换控制信息（控制播放，快进，快退，暂停等功能）
3. A/V Server使用RTP/UDP把流媒体数据传给播放器

步骤总结：1. 检索数据 2. 初始化 3. 交换控制信息 4. 播放

移动流媒体技术：

边下边播，点播，播放连续

· P2P技术



DASH---核心思想：动态自适应

服务器端提前存好同一内容的不同码率、不同分辨率的多个分片以及相应的描述文件MPD，客户端在播放时即可以根据自身性能以及网络环境选择最适宜的版本

使不同终端的用户同时使用不同质量的媒体内容

MPD文件

Period：标注了视频的时长信息，也可以看做是更新mpd文件的最长时长。AdaptationSet：包含了媒体呈现的形式，（视频/音频/字幕）。Representaiton：包含不同的码率、编码方式、帧率信息等。

Segment：每一个具体的片段。（1,2,4,6,10s ...）

DASH企业方案

Apple：HLS

Microsoft：Microsoft Smooth Streaming

Adobe：Adobe HTTP Dynamic Streaming

2. OTT业务与内容分发网络（CDN）

OTT：Over The Top

内容或服务建构在基础电信服务至上，不需要网络运营商额外的支持,早期特指音频和视频内容的分发，后来包含基于互联网的内容和服务。

OTT业务

- 传统电视服务、通信服务：服务和基础设施由同一方提供
- OTT电视服务、通信服务：服务和基础设施由不同方提供



三种：

- 美国：完全以互联网架构为基础的开放模式
- 欧洲：兼顾传统广播电视与OTT业务（HBBTV）
- 中国：以播控平台牌照为基础的互联网架构

国内

OTT业务：点播 IPTV：实时电视业务+点播业务（增值业务）

OTT优势：丰富的点播内容（原创，购买版权的电影，电视剧）

劣势：网络性能无保障

IPTV优势：电信运营商专网传输，网络性能保障，丰富的实时电视节目内容

劣势：点播内容以增值业务形式展开，内容较少

编码对比

项目	OTT TV	IPTV
编码方式	以H.264为主（其它有VP8, H.265）	H.264
码率控制	编码器工作在低码率状态下，采用多次编码的VBR方式	编码器工作中高码率状态下，采用一次编码的CBR方式
封装格式	TS、mp4文件（HTTP下载） TS、mp4切片（HTTP流）	TS文件
编码复杂度	多次编码，计算资源消耗较多，实时性差	计算资源消耗较少，实时性高
网络依赖性	传输采用HTTP下载方式时，通过增大终端缓存换取对网络QoS依赖的降低	传输采用RTP方式，通过以传输码率等于编码码率换取播放的实时性，代价是要求网络QoS稳定

传输对比

项目	OTT TV	IPTV
业务类型	以单播为主	直播、单播均支持
承载网络	互联网	运营商专网
视频传输形式	渐进式下载	实时流媒体
视频承载协议	HTTP承载信令和视频内容，通过TCP传输	RTSP承载信令，RTP承载视频内容，通过UDP传输
视频传输特征	客户端按网络带宽拉流	服务端按编码速率推送码流
播放时延	较大缓存后启动播放	较小缓存立即启动播放
网络适应能力	支持码率自适应，对网络分组丢失不敏感，对网络时延抖动敏感	不支持码率自适应，对网络分组丢失敏感，对网络时延不敏感

CDN

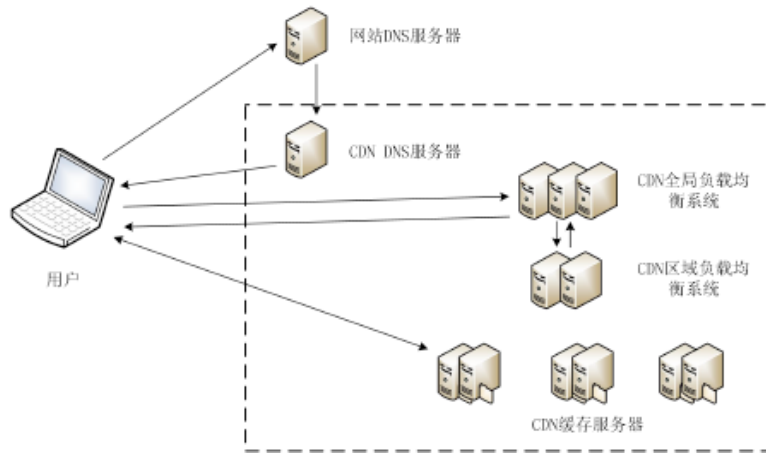
基本思路：把内容放到离用户更近的地方，以避开互联网上有可能影响数据传输速度和稳定性瓶颈的环节

实现方法：在网络各处放置节点服务器，构成智能虚拟网络，实时根据网络流量和各界点的情况将用户的请求重新导向离用户最近的服务器节点上

CDN发展史

发展阶段	时间	主要特点
第一阶段	1998-2001年	主要服务于Web，硬件以专用硬件为主
第二阶段	2001-2008年	开始提供视频、应用等加速，硬件转向通用服务器
第三阶段	2008年-至今	提供高清视频加速，以大容量通用服务器为主

请求流程



引入CDN后用户请求流程

CDN关键技术

1. 内容统一标识：

根据内容表示来定位内容，位用户提供服务（唯一）

- 表示方法：统一资源定位符（URL）
 - 优点：无需重新这是CDN，保证内容唯一性
 - 缺点：URL中无CRN可用信息，内容管理系统（CMS）未使用URL作为表示
- 自定义二元组，不实用URL
 - ProviderID: 每个CMS在CDN中唯一标识
 - ContentID: 每个内容在CMS中唯一标识

内容管理系统：是一种位于WEB 前端（Web 服务器）和后端办公系统或流程（内容创作、编辑）之间的软件系统

2. 负载均衡

处理节点的负载信息通过某代理软件传递给均衡器，由均衡器做出决策进行动态分配

负载均衡策略：

- 软/硬件负载均衡
- 全局/本体负载均衡
- 链路聚合技术

3. 缓存替换算法

- CDN节点存储空间有限，无法存储所有内容
- 缓存替换算法用来提高用户请求命中率，指导存储内容进行更新

意义：

- 缩短整体用户请求响应时延
- 提高网络高峰时承载能力

■ 缓存替换算法

- 基于新近性的缓存替换算法
 - 典型算法：最近最少使用（Least Recently Used, LRU）
 - 优点：适应网络负载变化，复杂度低，部署简单
 - 缺点：未考虑缓存对象大小与频率因素
- 基于频率的缓存替换算法
 - 典型算法：最不经常使用（Least Frequently Used, LFU）
 - 优点：适用于流行度变化较小的静态网络
 - 缺点：无法应对网络负载动态变化的环境
- 基于新近性+频率的缓存替换算法
 - 优点：参数选择恰当可避免单独考虑两种因素的缺点
 - 缺点：引入了额外的复杂度
- 随机缓存替换算法
 - 优点：缓存操作便捷，部署简单
 - 缺点：算法性能难以评估

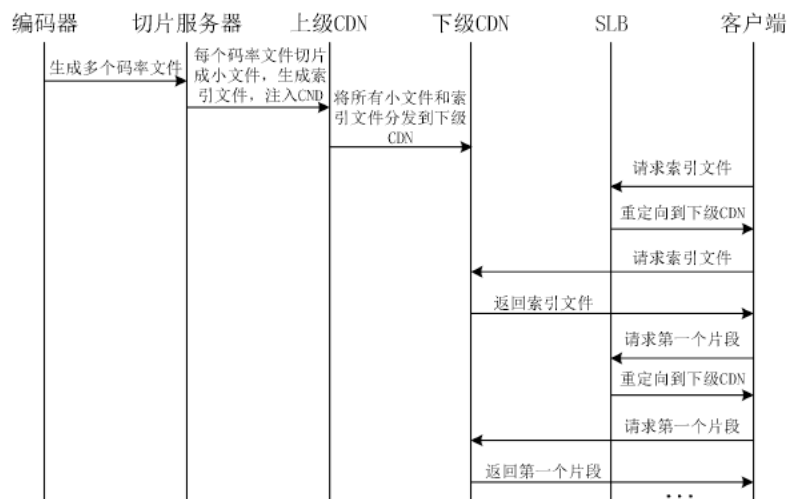
1. LRU：最近最少使用算法（访问时间排序）
2. LFU：最近不经常使用算法（访问次数排序）

基于DASH的内容分发网络架构



■ 基于DASH的CDN流程

- 原理：将同一视频源内容编码、切片生成的每个视频片段文件，都看作一个独立的节目内容，在服务器组件将它们注入到CDN时，每个片段文件都被CDN单独管理；当客户端获取索引文件后，根据索引文件中记录的每个片段内容的URL地址发起HTTP请求，由SLB逐个处理并重定向到相应的下级CDN节点



补充章节 ATSC 3.0标准

ATSC 3.0网络架构

1. ATSC非营利全球性组织，致力于数字电视标准的研发

在不同的通信媒介中协调电视标准的研发，主要关注数字电视、互动电视系统、宽带多媒体通信等

2. ATSC标准演进

1.0: 标/高清视频多播 移动数字电视 5.1声道环绕立体声

2.0: 后向兼容ATSC1.0，支持非实时传输，高级音/视频压缩，增增强交互性

3.0: 不向后兼容，无限三网融合，高清/超高清视频，3D

多声道音频

3. ATSC 3.0新的需求 ▪ 实现无线三网融合，把地面数字电视广播发送端“升级”为新添无线宽带 互联网网站功能（增加回传信道，把单向广播改造为双向交互的） ▪ 除大屏幕电视机用户群体外，优先面向平板PC和智能手机等新用户群体 ▪ 采用各种新技术，数字传输、音视频编码和传送等，以更有效利用地面电 视频谱 ▪ 可发送4K超高清视频和3D多声道音频 ▪ 可融合各种新技术和其他有关标准

4. 广播和宽带业务

文件封装

目标应用：ISO BMFF

支持时序媒体数据：音频，视频

非时序媒体数据：元数据

应用：

- 录制； ▪ 数据交换和下载，包括增量（Incremental）下载和播放； ▪ 本地播放； ▪ 编辑、组合和铺叠（lay-up）； ▪ 从流化服务器流化，并且捕捉流生成文件