

消耗速率的平均吞吐量。因为事件是直播的，尽管定时限制没有会话式语音那么严格，但时延也可能成为问题。从用户选择观看一个实况传输到播放开始，能够容忍的时延最多为 10 秒。我们在本书中将不涉及流式实况媒体，因为用于流式实况媒体的许多技术（如初始缓存时延、适应性带宽使用和 CDN 分发）都类似于流式存储媒体所使用的技术。

7.2 流式存储视频

对于流式视频应用，预先录制的视频放置在服务器上，用户向这些服务器发送请求按需观看这些视频。用户可能从开始到结束都在观看视频而没有中断它，也可能在视频结束前停止观看它，或者通过暂停、重新定位到后面或前面镜头来与视频交互。流式视频系统可分为三种类型：UDP 流（UDP streaming）、HTTP 流（HTTP streaming）和适应性 HTTP 流（adaptive HTTP streaming）。尽管在实践中所有这三种系统都在使用，但绝大多数今天的系统应用了 HTTP 流和适应性 HTTP 流。

所有这三种形式的视频流的共同特点是广泛使用了客户端应用缓存，以此来缓解变化的端到端时延和变化的服务器和客户之间可用带宽量的影响。对于流式视频（存储的和实况的），用户通常能够容忍在客户请求某视频与该流视频在客户端播放之间有几秒钟的初始小时延。所以，当视频开始到达客户时，客户不必立即开始播放，反而能够在应用程序缓存中建立该视频的储备。一旦该客户建立起几秒钟的“已缓存但尚未播放”的视频储备，客户就可以开始视频播放了。这种客户缓存（client buffering）具有两种重要的优点。第一，客户端缓存能够吸收服务器到客户时延中的波动。如果某特殊部分的视频数据延迟了，只要它在“接收到但尚未播放”的视频耗尽之前到达，这个长时延将不会被注意到。第二，如果服务器到客户带宽暂时低于视频消耗速率，用户能够继续享受连续的播放，只要客户应用缓存仍没有完全排尽。

图 7-1 显示了客户端的缓存。在这个例子中，假定视频以固定的比特率编码，因此每个视频块包含了能在相同固定时间量 Δ 区间播放的视频帧。服务器在 t_0 传输第一个视频块，在 $t_0 + \Delta$ 传输第二个视频块，在 $t_0 + 2\Delta$ 传输第三个视频块等等。一旦客户开始播放，为了重新产生初始录制视频的定时，每个块应当在前一个块之后播放 Δ 时间单元。第一个视频块于 t_1 时刻到达，第二个视频块于 t_2 时刻到达。第 i 块的网络时延是服务器传输该块

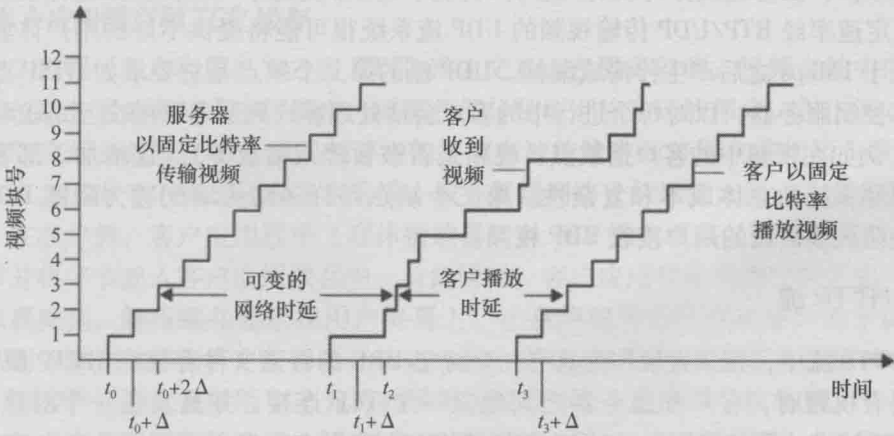


图 7-1 视频流中的客户播放时延