

# 视频用户体验评估——函数关系的确立

2017 年 5 月 27 日

## 1 摘要

## 2 问题重述

### 2.1 问题的提出

随着科技的进步——无线宽带网络的升级、智能用户端的便携化、各类视频直播软件的普及，愈来愈多的用户选择在移动智能终端上使用应用客户端（APP）观看网络视频。网络视频是一种基于传输控制协议(TCP)的视频传送及播放。

其中影响用户体验的两个关键指标——初始缓冲等待时间和视频播放过程中的卡顿缓冲时间，主要使用初始缓冲时延(sLoading)和卡顿占比(sStalling)（卡顿时长的占比=卡顿时长/视频播放时长）来定量地评价。研究数据表明，影响两个关键指标的主要因素有四——初始缓冲峰值速率、播放阶段平均下载速率、端到端环回时间（E2E RTT）以及视频参数。但目前以上四个因素与两个关键指标之间的关系未知。

请依据题目附件所提供的实验数据建立用户体验变量（初始缓冲时延，卡顿时长占比）与网络侧变量（初始缓冲峰值速率，播放阶段平均下载速率，E2E RTT)之间的函数关系。

## 3 模型假设及符号说明

### 3.1 模型的假设

- 1.当客户端请求视频时，服务器使用TCP将视频流传输给客户端。
- 2.平均TCP吞吐量不小于视频的宽带，满足流式传输视频的要求。
- 3.视频为恒定比特率(CBR)视频，视频播放速率为每秒的数据包，数据包的大小相同。
- 4.视频在客户端连续播放，不会停止并等待延迟的数据包，保持速率不变。
- 5.不同回合中的数据包丢失是独立的。
- 6.同一轮中的数据包丢失是相关的：若一个数据包丢失，则其所有剩余的数据包及其本身全部丢失。
- 7.实时流媒体的传输受到应用级视频的生成速率的约束。
- 8.第一个数据包的生成时间点为零点。

9. 视频的长度大于RTT。

### 3.2 符号说明

1.  $v_{max}$ : 初始缓冲峰值速率。
2.  $t_R$ : E2E RTT, 往返时延。
3.  $\bar{v}$ : 播放阶段平均速率。
4.  $\tau$ : 初始缓冲时延。
5.  $s_0$ : 初始缓冲下载数据量。
6.  $\mu$ : 视频码率。
7.  $t_{all}$ : 播放阶段总时长。
8.  $t_p$ : 播放时长。
9.  $t_s$ : 卡顿时长。
10.  $X_i$ : TCP在第i回合的状态。
11.  $S_i$ : TCP在第i回合成功传输的数据包的数量。
12.  $L$ : 视频的长度, 单位为回合。
13.  $N_i$ : 第i回合早到的数据包的数量。
14.  $N_i^l$ : 第i回合晚到的数据包的数量。
15.  $Y_i^c$ : 第i回合实时流媒体模型的状态。
16.  $Y_i^u$ : 第i回合存储流媒体模型的状态。
17.  $P_i$ : 第i回合至少有一个迟到的数据包的可能性。
18.  $f$ : 卡顿占比。
19.  $MSS$ : TCP模型中最大分段的大小。
20.  $p$ : 丢包率。

## 4 问题的分析

由题意可知, 本题的目的是为了建立一个多元函数模型来确定影响视频用户体验的两个关键指标与视频相关参数的关系。

由于本题目数据较多, 为了便于计算机进行统计和分析, 我们采用两种不同的策略进行建模: 第一种是白箱(Class Box)方法, 即传统意义上的数学建模方法——通过查找文献获取问题相关知识, 分析其机理并建立数学模型, 最终通过数据确定参数, 并分析模型准确度和误差; 另一种则是黑箱(Black Box)方法, 将数据内部关系视为黑箱, 仅仅从输入和输出数据的角度进行观测, 并通过统计学方法拟合出变量间函数关系。

已知黑箱方法的基础为大量观测数据, 而在该问题中, 我们得到了数量可观的实际数据, 这是促使我们采用这一方法的主要原因。同时, 值得一提的是, 以神经网络为代表的机器学习技术近年来在工程和科研计算中得到了越来越广泛的应用, 其对任意非线性多元函数进行拟合的能力相较于传统方法更具优势, 这是我们采取这一方法的另一原因。

## 5 模型的建立

### 5.1 模型的准备

若想对视频传输系统建立数学模型，我们首先要理解的就是基于TCP和移动网络的在线视频业务模型。在线视频技术就是我们常说的流媒体技术。

#### 5.1.1 传输控制协议(TCP)

传输控制协议(Transmission Control Protocol)是一种面向连接的、相对IP而言可靠的、基于字节流的传输层通信协议，其完成传输层所指定的功能。传输控制协议(以下简称TCP)层在因特网协议族(Internet Protocol Suite)中处于应用层之下，IP层之上的中间层。TCP有多种机制来响应网络堵塞，是被定义为友好的，提供了可靠的传输，从而消除了在较高级别的丢失恢复的需要。

TCP使用了简单的策略——停止与等待播放来处理延迟的数据包(Package)。TCP先将数据流进行分区为适当长度的报文段，装入一个包并为了防止丢失将其编号。而后TCP将包传给IP层，再通过网络将包传送给客户端。对于发出的每一个数据包，TCP将会启动重传计时器以等待接收器的确认(ACK)。若客户端对已经成功接受，则发回一个ACK；若客户端未接受到，既发送端在合理的往返时延(RTT)未收到三重重复的ACK时，则认定计时器超时、相应的数据包已丢失，而后进行重传，此时的窗口大小将会减为1。此外，在下一次重传时的定时器值将会被设置为先前定时器值的两倍，这是一种指数型退避行为，将会持续到重新发送的数据包被确认为止。

尽管当前利用TCP进行视频流的研究较少，但我们不得不承认其在商业视频流系统中的广泛应用。

#### 5.1.2 流媒体

实时流媒体是随着Internet与多媒体技术的发展而逐渐兴起的一种多媒体网络传输技术，其典型传输方式为：连续采样的多媒体信息经编码压缩后，从源端连续、实时地发送到网络，接收端从网络收到部分数据后就可以开始解码播放，而后续接收的数据则持续不断地存入本地缓存，形成可连续播放的媒体流。其具有连续性、实时性和时序性。

流媒体技术和业务是宽带多媒体技术和业务的一个重要分支，分为顺序流式传输(Progressive Streaming)和实时流式传输(Real-time Streaming)。顾名思义，顺序流式传输即为按顺序下载，下载文件和观看同时进行，在传输过程中，用户只能观看已下载的部分，想要观看未下载的部分时就会出现卡顿；而实时流式传输则可以在传输期间根据用户的连接速度进行相应的调整，保证媒体信号带宽与网络连接匹配，使用户可以实时观看媒体。

#### 5.1.3 相关因素的物理关系

1. 视频数据量 = 视频码率  $\times$  播放时长
2. 初始缓冲时延 = 初始缓冲下载数据量  $\div$  缓冲时段平均速率
3. 初始缓冲下载数据量 + 播放阶段总时长  $\times$  播放阶段平均速率  $\geq$  视频数据量

4. 不出现卡顿的条件: (视频数据量-初始缓冲下载数据量)  $\div$  平均时间

5. 卡顿时间= (视频数据量-初始缓冲下载数据量)  $\div$  播放阶段平均速率-视频播放时间

## 5.2 模型的建立

### 5.2.1 黑箱模型

### 5.2.2 白箱模型

在此, 我们探究两种形式的流媒体——实时流媒体和存储(顺序)流媒体。

对于本模型而言, 每个时间单位对应一个回合的长度, 一个回合的长度被定义为RTT。一个回合开始的标志为 $W$ 包(TCP拥塞窗口当前的大小)的回环传输。视频的长度为 $L$ (其单位为rounds), 视频的播放速率为每回合 $\mu R$ 个数据包。当拥塞窗口中所有的数据包被发出后, 在收到ACKs之前, TCP将不会再发出任何包。ACK标志着本轮传输的结束与下轮传输的开始。

设视频的播放起始时间为 $\tau$ (播放延迟)。对于实际流媒体而言, 其在数据的数据包上存在一个上限, 而存储流媒体没有, 我们令这个上限为 $G(t)$ 表示服务器根据时间生成的数据包的数量, 则有:  $G(t) = \mu t$ 。令 $A(t)$ 表示客户端根据时间收到的数据包的数量。已知TCP的传输受限于服务器的生成速率, 于是有 $G(t) \geq A(t)$ 。令 $B(t)$ 表示客户端播放数据包与时间的关系, 则有:  $B(t) = \mu(t - \tau), t \geq \tau$ , 于是 $G(t) - B(t) = \mu\tau$ 。令 $N(t)$ 表示早于播放时间到达的数据包的数量, 则有:  $N(t) = A(t) - B(t)$ 。其负值表示比播放时间晚到的包的数量, 此时会造成卡顿。对于流媒体, 我们有 $N(t) \leq G(t) - B(t) = \mu\tau$ , 然而对于存储流媒体,  $G(t)$ 可以大于 $\mu\tau$ 。

我们将离散时间马尔可夫模型简化为 $X_i|_{i=1}^{\infty}$ ,  $X_i$ 是一个数组,  $X_i = (W_i, C_i, L_i, E_i, R_i)$ , 其中 $C_i = 0$ 与 $C_i = 1$ 表示的是无丢包成功发送的次数,  $L_i$ 表示第 $(i-1)$ 轮中丢失的数据包的数量,  $E_i$ 表示连接是否处于超时状态以及退避指数的大小,  $R_i = 1$ 和 $R_i = 0$ 分别表示在超时阶段传输的数据包为重传数据包或新的数据包。而 $S_i$ 由 $X_i$ 和 $X_{i+1}$ 决定, 则令 $E[S_i]$ 表示 $S_i$ 的期望。于是,  $E[S_i] = E[S_i|X_i = (w, c, l, e, r), X_{i+1} = (w', c', l', e', r')]$ 。

我们将 $f$ 作为各种参数(损失率、RTT、重传计时器、视频播放速率)的函数, 而 $N_i$ 是 $N(t)$ 的离散版本:  $N_i = N(iR)$ 。令一个回合中返回的数据包的数量为 $\mu R$ 。在第 $i$ 回合,

$$E[N_i^l] = \sum_{k=1}^{\mu R} k P(N_i^l = k), N_i^l \in \{0, 1, \dots, \mu R\}$$

其中 $P(N_i^l = k)$ 表示第 $i$ 回合有 $k$ 个迟到的数据包的可能性。迟到的数据包的占比, 即为视频的卡顿占比为

$$f = \frac{\sum_{i=1}^L E[N_i^l]}{\mu R L} \quad (1)$$

其中分子分母分别表示在视频播放中迟到的数据包的期望值和视频所含的所有数据包。

令 $N_i^b$ 为在视频播放之后到达的数据包的数量, 并借其来获得 $P(N_i^l = k)$ , 根据定义我们有

$$N_i^b = \begin{cases} 0, & N_i \geq 0 \\ -N_i, & N_i < 0 \end{cases} \quad (2)$$

$$P(N_i^l = k) = \begin{cases} P(N_i^b = k), & k < \mu R \\ P(N_i^b \geq \mu R), & k = \mu R \end{cases} \quad (3)$$

需要注意的是，在第*i*回合迟到的数据包的数量 $N_i^l$ 的最大值为 $\mu R$ ，然而 $N_i^b$ 的值可以比 $\mu R$ 大。当 $N_i^b \geq \mu R$ 时， $N_i^l = \mu R$ ，即有 $P(N_i^l = \mu R) = P(N_i^b \geq \mu R)$ 。

我们已知传输速率的上限为： $Rate \leq \frac{MSS}{RTT} \frac{1}{\sqrt{\rho}}$ 。设 $w$ 为窗口的大小， $w_{max}$ 为最大拥塞窗口的大小， $T_0$ 为初始的重传超时、为常量。于是我们有

$$Rate = \begin{cases} \frac{MSS\{\frac{1-p}{p} + w(p) + \frac{Q(p, w(p))}{1-p}\}}{RTT[w(p)+1] + \frac{Q(p, w(p))G(p)T_0}{(1-p)}}, & w(p) < w_{max} \\ \frac{MSS\{\frac{(1-p)}{p} + w_{max} + \frac{Q(p, w_{max})}{(1-p)}\}}{RTT(0.25w_{max} + \frac{(1-p)}{pw_{max}} + 2) + \frac{Q(p, w_{max})G(p)T_0}{(1-p)}}, & otherwise \end{cases}$$

其中，

$$w(p) = \frac{2}{3} + \sqrt{\frac{4(1-p)}{3p} + \frac{4}{9}} \quad (1)$$

$$Q(p, w) = \min\{1, \frac{[1 - (1-p)^5][1 + (1-p)^3(1 - (1-p)^{w-3})]}{1 - (1-p)^w}\} \quad (2)$$

$$G(p) = 1 + p + 2p^2 + 4p^3 + 8p^4 + 16p^5 + 32p^6 \quad (3)$$

对于实际流媒体，我们用 $Y_i^c|_{i=1}^L$ 表示模型，其中 $Y_i^c$ 是有关于 $(X_i, N_i)$ 的数组。其中我们有 $N_{max} = \mu\tau$ ,

$$N_{i+1} = \min(N_{max}, N_i + S_i - \mu R), N_i \leq N_{max}, i = 1, 2, \dots, L$$

实际上，当 $N_i = N_{max}$ 时，TCP并不会在第 $(i+1)$ 回合发送任何的数据包。因为视频长度长于一个RTT，所以迟到的数据包的占比可以近似的利用视频长度来衡量

$$\lim_{L \rightarrow \infty} \frac{\sum_{i=1}^L E[N_i^L]}{\mu RL} = \lim_{i \rightarrow \infty} \frac{E[N_i^L]}{\mu R}$$

对于存储流媒体，我们用 $Y_i^u|_{i=1}^L$ 表示模型，存储流媒体并没有数据分包的上限，简化之后就有 $Y_i^u = X_i$ 。在此 $N_i$ 代表这冲量奖励(Impulse Reward)，我们将 $\rho_{yy'}$ 和 $Y_i^u = y$ 与 $Y_{i+1}^u = y'$ 的转换相联合，去定义收到的数据包数量与重播的数据包数量的区别。令 $N_i'$ 为第*i*回合累积的冲量奖励。当传输与重播同时在0点开始时， $N_i'$ 为第*i*回合早到的数据包的总数；当重播在 $\tau$ 开始时， $N_i = N_i' + \mu\tau$ 。在此我们有

$$P_i = P(N_i < 0) = P(N_i' < -\mu\tau) \quad (4)$$

定义 $\beta$ 为视频重播期间至少有一个迟到的数据包存在的可能性，我们有

$$\beta = 1 - P(N_1 \geq 0, N_2 \geq 2, \dots, N_L \geq 0)$$

$$\beta \leq 1 - \prod_{i=1}^L (1 - P_i) \quad (5)$$

## 6 模型的求解及结果分析

### 6.1 模型的求解方法

### 6.2 结果分析

## 7 模型的检验

## 8 模型的优缺点及改进方向

### 8.1 模型的优点

### 8.2 模型的缺点

### 8.3 模型的改进方向

## 9 参考文献

## 10 附录