

流媒体传输中的 QoS 研究及其实现

王丕光, 李小平, 沙 金

(北京理工大学 信息科学技术学院 计算机工程系, 北京 100081)

摘 要: 文中提出的对视频会议的服务质量控制策略是为解决以下问题: 视频会议对涉及到的实时数据流传输对网络带宽、时延和丢包率有较高要求, 但是, 目前得到广泛应用的 IP 网络是一种尽力而为的网络, 它并没有对实时多媒体提供任何服务质量保证。该策略从两个方面对服务质量加以控制: 在发送端控制发送流量; 在数据再现端通过缓冲机制控制媒体同步。

关键词: 服务质量; 流量控制; 丢包率; 网络拥塞

中图分类号: TP393

文献标识码: A

文章编号: 1005-3751(2005)07-0096-04

Research and Realization of QoS in Streaming Multimedia Transmitting

WANG Pi-guang, LI Xiao-ping, SHA Jin

(Dept. of Computer Eng., Sch. of Info. Sci. and Techn., Beijing Inst. of Techn., Beijing 100081, China)

Abstract: As is known, the real-time data stream involved in video conference is strict to network bandwidth, time prolong and the rate of data package losing. Meanwhile, the current best-effort internet does not offer any quality of service(QoS) guarantees to real-time multimedia streaming, that is why we should provide QoS guarantees which is based on present network. This thesis has offered QoS control strategy to the video conference which provides QoS guarantees from two respects that is controlling the sending rate and the synchronization of multimedia through cushion mechanism.

Key words: QoS; control of data-streaming; rate of data package losing; blocking of network

随着网络技术和多媒体技术的发展,流媒体传输的应用越来越普遍,但现在的 Internet 都没有对多媒体传输提供质量保证,同时流媒体的传输对网络带宽、时延、丢包率都有一定的要求。解决上述问题可以通过增加带宽来解决,但这样代价比较高,于是在现行网络上的 QoS 研究就变得越来越重要。

1 QoS 的指标

ITU 将服务质量定义为决定用户对服务的满意程度的一组服务性能指标。从另一角度来说, QoS 参数也是多媒体信息传输网络的性能指标。

常用的 QoS 参数有如下几项。

1) 传输速率是指网络传送二进制信息的速率,也称比特率或带宽,在这里是指数据传输速率。

2) 传输延时定义为从信源发送出第 1 个比特的数据到信宿接收到第 1 个比特的数据之间的时间差,它是用来描述网络延时的一个指标。

3) 延时抖动定义为网络传输延时的变化率。

在传输系统中产生的错误由以下几种方式度量:

a. 误码率 BER(Bit Error Rate), 指在从一点到另一点的传输过程中比特级的错误率。BER 主要衡量的是链路层和物理层介质的质量。

b. 包错误率(Packet Error Rate), 是指包丢失、同一个包两次接收,或包的次序颠倒而引起的包的错误。

文中研究的流媒体传输系统是在应用层上实现的,而比特级的错误一般只在链路层和物理层产生,并且错误在这两层当中已被处理,因此可以认为应用层没有比特级的错误,包错误是应用层错误的主要原因,其中尤其以包丢失最常见。包丢失的主要原因往往是由于网络拥塞,造成包的传输延时过长,超过了应该到达的时限而被接收端舍弃,或网路节点来不及处理而被节点丢掉。

2 QoS 控制机制

流调度被定义为在终端、以及网络节点上如何传送数据包的策略。流调度与资源(如带宽、服务时间、优先级、缓存器容量等网络性能)是密切相关的。通常,基于速率的流调度策略有公平队列(Fair Queueing)、虚拟时钟(Vir-

收稿日期: 2004-10-19

作者简介: 王丕光(1978-), 男, 山东菏泽人, 硕士研究生, 研究方向为视频会议在远程教育中的应用。

tual Clock)、等级轮巡(HRR, Hierarchical Round Robin)等方法。文中所讨论的流媒体传输系统采用的是公平队列调度方法。

流量控制分为开环和闭环两种类型。开环控制是指信源严格按规定的速率向网络发送数据。而在闭环控制中, 传输策略是由发送端根据接收端或网络反馈回来的信息动态调整来实现的。本方案采用的是闭环控制。

3 QoS 控制的设计与实现

3.1 方案概述

QoS 控制由两个闭环反馈回路构成: 外反馈由服务器发送端、RTP 前向通道、客户端和 RTCP 反馈通道构成回路, 用以控制 RTP 数据包的网络传输质量^[1]; 内反馈通过实时调节播放器的播放速度来控制客户端数据缓冲区的数据量, 防止缓冲区数据上/下溢出, 从而提高播放质量(见图 1)。

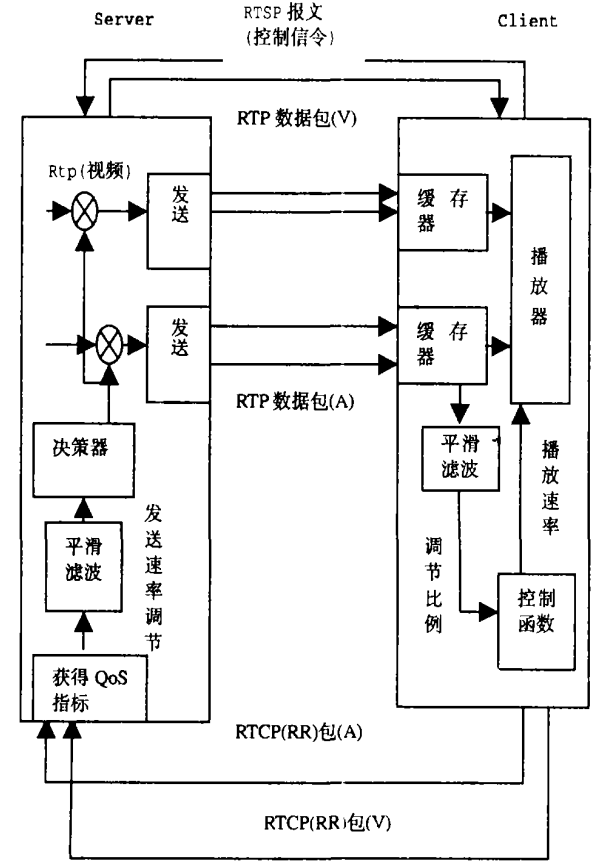


图 1 QoS 系统框图

在外反馈控制系统中, 客户端接收 RTP 数据包, 对 RTP 包进行 QoS 参数统计, 并将统计结果通过 RTCP 包发送到流服务器。服务器取其中最差的反馈信息(如果有多个客户端的话)作为控制依据, 先对反馈信息进行平滑滤波, 再进行决策。决策的结果用以调整 RTP 数据包的发送速率, 以达到控制网络传输流量的目的^[2,3]。由于 RTCP 包的发送周期较长, 所以外反馈回路的控制延迟较

大, 而且网络中不可预料的突发性干扰较多, 因此控制精度不高, 无法满足播放器的精度要求, 因此还需要靠内反馈进一步提高播放精度。

客户端接收到的视/音频数据不断送入数据缓冲区, 与此同时播放器不断从缓冲区提取数据以供解码回放, 要想获得较高的播放质量, 必须防止缓冲区上/下溢出, 因此一定要使缓冲区的输入与输出保持动态平衡。内反馈控制系统对缓冲区中的数据量进行实时监控, 通过调节播放器的播放速率, 来控制缓冲区数据的消耗量, 从而避免缓冲区数据上/下溢出, 进一步提高播放质量。

3.2 服务端外反馈控制的实现

3.2.1 QoS 参数的获得

QoS 参数是计算 QoS 指标的直接依据, QoS 参数包括序列号和时间戳两大类, 这些参数主要通过 RTP 和 RTCP 包携带, 接收者利用这些参数计算出 QoS 指标作为控制和决策的依据。

1) 参数定义。

SBASE—RTP 包的起始序列号; SWRAP—连续循环的序列号计数; SMAX—接收到的 RTP 包中的最大序列号。

2) RTP 包头中携带的时间戳(tTxj)和序列号(Seq)。

tTxj—32 位, 第 j 个 RTP 包发送的时间; Seq—16 位, 序列号。

3) RTCP 包头中携带的时间戳(tTx)和到目前为止发送的 RTP 包的个数 NTx。

tTx—RTCP 包发送的时间; NTx—目前为止发送的 RTP 包的个数。

3.2.2 QoS 指标的计算

QoS 指标包括丢包数、丢包率、延迟抖动、LSR 和 DL-SR, 接收者利用 QoS 参数计算出这些指标, 并写入 RTCP 包的相应字段中, 反馈给发送者。

1) 丢包数(NiLOST, 接收到第 i 个 RTP 包时的丢包数)。

$$NiLOST = NiE - NiRx = SMAX + SWRAP - SBASE + 1 - NiRx$$

NiRx—接收到第 i 个 RTP 包时实际收到的 RTP 包数; NiE—接收到第 i 个 RTP 包时已经发送的 RTP 包数。

2) 丢包率(FrijLOST, 在第 i 个和第 j 个 RTP 包之间发生的丢包率)。

$$FrijLOST = (NiLOST - NjLOST) / (NiE - NjE)$$

3) 延迟抖动(Ji, 接收到第 i 个 RTP 包时的抖动)。

$$Ji = Ji - 1 + (|DiI - 1| - Ji - 1) / 16$$
$$Dij = \text{delta-}tiRTP - \text{delta-}tjRTP = (tRxi - tRxj) - (tTxi - tTxj);$$

$$\text{delta-}tiRTP = tRxi - tTxi$$

$$\text{delta-}tjRTP = tRxj - tTxj$$

tRxi—接收到第 i 个 RTP 包的时刻(与发送端打时间

戳的记数单位相同)；

$t_{Tx i}$ —第 i 个 RTP 包的发送时刻(RTP 包上的时间戳)；

16—RTCP 协议中建议使用的平滑滤波系数。

4) LSR(最近一次收到的 RTCP(SR) 包中的 NTP 时间戳)。

$t_{LSR}=t_{TxSR}$

5) DLSR(在接收端对最近一次 RTCP(SR) 包的处理延时)。

$DLSR = \Delta t_{LSR} = t_{TxRR} - t_{RxSR}$

t_{TxRR}/SR —发送 RTCP(RR)/(SR) 包的时刻；

t_{RxSR}/RR —接收到 RTCP(SR)/(RR) 包的时刻。

3.2.3 服务端 QoS 决策

服务端在 QoS 决策之前,必须先对接收到的 QoS 指标进行平滑滤波。平滑滤波器一般采用如下公式。

$$\lambda_i = (1-a)\lambda_{i-1} + ab$$

其中 a 为平滑因子,一般取 $0.7 \sim 0.9$, b 为服务端接收到的最新指标(如丢包率), λ_{i-1} 为前一次的平滑滤波结果, λ_i 为平滑滤波后的新结果。 λ_i 被送入 QoS 决策器。

QoS 决策的原理是通过分析 RTCP 反馈的丢包率,判断网络的工作状态。本系统将网络状态分为正常、轻度拥塞、中度拥塞、重度拥塞、极度拥塞五段: $\lambda < \alpha$ 视为网络正常; $\alpha < \lambda < \beta_1$ 视为网络轻度拥塞; $\beta_1 < \lambda < \beta_2$ 视为网络中度拥塞; $\beta_2 < \lambda < \beta_3$ 视为网络重度拥塞; $\beta_3 < \lambda$ 视为网络极度拥塞。

在实际使用中,需要根据业务需求确定分段标准。在 MPEG 标准的音/视频传输业务中音频的优先级高于视频,视频中 I 帧的优先级最高,其次是 P 帧, B 帧的优先级最低,因此必须合理地估计 $\alpha, \beta_1, \beta_2, \beta_3$ 参数。 α 为网络正常情况下因偶然因素而导致的丢包率,一般取 0.1% ; β_1 根据人的感官可以容忍的降速下限来确定,一般取 $3\% \sim 5\%$; β_2 根据 B 帧所占的数据量比例确定,一般取 49% 左右; β_3 可根据 B/P 帧所占的数据量比例确定,一般取 83% 左右。根据对网络拥塞状况的判断,决策器将实施不同级别的流量控制。

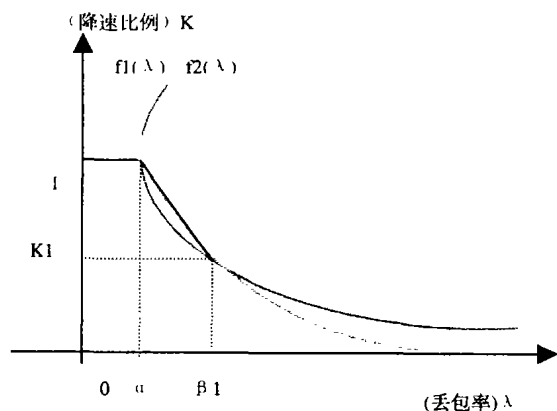


图 2 降速比例曲线

网络轻度拥塞时,决策器通过改变 RTP 包的时间戳来降低 RTP 包的发送速度,从而降低网络数据流量。如图 2 所示。

降速的方式采用负指数曲线或二次曲线作为参考,最终可由实验作出优化选择, K_1 一般取人的感官可以容忍的降速下限 95% 左右。

$$\begin{cases} K \text{ 取负指数函数:} \\ K = f_1(\lambda) = \exp[(\ln(K_1)(\lambda - \alpha))/(\beta_1 - \alpha)] \\ K \text{ 取二次函数:} \\ K = f_2(\lambda) = 1 + (K_1 - 1)[(\lambda - \alpha)/(\beta_1 - \alpha)]^{1/2} \end{cases}$$

正常发送速率时, $K = 1$

$t_{Tx i} = t_{Tx i-1} + (PTS_i - PTS_{i-1})/K$

$t_{Tx i-1}$ —第 $i-1$ 个 RTP 包打的时间戳;

$t_{Tx i}$ —第 i 个 RTP 包打的时间戳;

t_0 —发送第一个 RTP 包的时间戳,为随机数;

PTS_i —第 i 个 RTP 包的 PTS。

在网络中度、重度和极度拥塞时,本系统采用 QoS 过滤波实现 QoS 缩放(分级弃帧)。决策器根据网络的丢包率实施分段控制,在网络中度拥塞时,服务器首先抛弃优先级最低的 B 帧,重度拥塞时抛弃 B 和 P 帧,极度拥塞时 B、P 和 I 帧全部抛弃,音频数据始终不抛弃。此方法用牺牲图像质量的代价以降低数据的发送量,以减轻网络压力,保证网络的整体传输性能。

3.3 客户端内反馈控制的实现

连续媒体数据是由若干存在固定时间关系的 LDU (逻辑数据单元)构成的时间序列,若网络传输存在延时抖动,LDU 到达接收端时,连续媒体数据内部的时间间隔则发生变化,会影响解码播放的效果。这就需要在接收端采取措施,将失序的 LDU 重新整形。将从网络接收到的 LDU 先进入客户端缓存器,对延时抖动进行平滑、过滤,再将连续的 LDU 时间序列输出到播放器。内反馈控制算法包括三个部分:缓存器容量估计算法,平滑滤波算法,控制算法^[4,9]。

3.3.1 缓存器容量估计

如果对缓存器容量大小估计适当,既能消除延时抖动的影响,又不会过分加大播放延时。

1) 符号定义。

$t(i)$ —第 i 个 LDU 的发送时刻; $a(i)$ —第 i 个 LDU 到达信宿端的到达时刻; $d(i)$ —第 i 个 LDU 在传输线路上的传输延时; d_{\max} —最大传输延时; d_{\min} —最小传输延时; $p(i)$ —第 i 个 LDU 的播放时刻; B —缓存器容量; Bt —媒体单元的最大缓存时间; r —播放速率。

2) 条件限制。

$$\begin{cases} d_{\max} > d(i) > d_{\min} & (1) \\ p(i) \geq a(i) & (2) \\ p(i) - p(i-1) = t(i) - t(i-1) & (3) \\ a(i) = t(i) + d(i) & (4) \\ (i = 1, 2, \dots) \end{cases}$$

3) 估计算法。

由(3) 式得: $p(i) - p(1) = t(i) - t(1)$ (5)

由(4) 和(5) 式可得:

$$p(i) - a(i) = p(1) - a(1) - (d(i) - d(1))$$

再根据(2) 式, 可得: $p(1) - a(1) \geq (d(i) - d(1))$

按最坏打算, 取:

$$p(1) - a(1) = \max(d(i) - d(1)) = d_{\max} - d_{\min}$$

$$\begin{aligned} \text{又 } Bt &= \max[(p(i) - a(i)) \mid (i \in (2, 3, \dots))] \\ &= (d_{\max} - d_{\min}) - \min[d(i) - d(1)] \end{aligned}$$

所以, $B = [Bt \times r] = [2(d_{\max} - d_{\min}) \times r]$

3.3.2 平滑滤波算法

通过设置滤波器参数, 对由短期延时抖动引起的缓存数据量变化将被滤掉, 只有由因网络拥塞等因素导致的长期缓存量变化才能触发控制函数的工作。

1) 符号定义。

$qt-t$ 时刻的缓存数据量; $S(qt)$ — 平滑滤波函数; bt — 平滑后的缓存数据量; $a \in [0, 1]$ — 平滑因子, 需实验校正。

2) 公式: $bt = S(qt) = a \times bt - 1 + (1 - a) \times qt$

3.3.3 控制算法

控制算法以 L 时长为周期不断调整, 调整期结束后检测 bt 是否回到正常水平, 如果是, 则将 R' 恢复到 R ; 否则, 再启动一个新的调整期。为了防止控制滞后, 可附加一块缓存器容量 ba (以时间计算), 它至少应该容纳整个反馈时间段内的数据。

1) 符号定义。

Rc — 调整速率; R — 正常播放速率; R' — 调整后的播放速率; L — 调整期; uw 和 lw — 缓存器出现上下溢的警界线。

2) 公式, 控制函数 $C(bt)$ 为:

$$\begin{cases} \text{当 } uw > bt > lw \text{ 时, 播放速率不动。} \\ \text{当 } bt < lw \text{ 或 } bt > uw \text{ 时, 调整播放速率如下:} \\ \quad \begin{cases} R' = R + Rc \\ Rc = \{bt - [lw + (uw - lw)/2]\} / L \end{cases} \end{cases}$$

3.3.4 重要参数的计算和估计

1) 确定起始播放时间。当第一个 $LDU > lw$ 时, $p(1) = a(1) + d_{\max} - d_{\min}$, 所以 ba 的大小也影响到播放起始延时的长短。

2) 平滑因子 a : 初步估计为 $0.7 \sim 0.9$, a 越大, 同步机制启动越缓慢, 可能导致缓存器变空或溢出; a 越小, 缓存量的微小变化也会启动再同步机制, 导致系统不稳定。 a 的合理取值, 最终应由实验决定。

3) 最大(小) 传输延迟 d_{\max} (d_{\min}) 限制了缓存器的大小, 从而反映了缓存器的抗延时、抗抖动能力。也就是说, 延时 d 在 $[d_{\max}, d_{\min}]$ 范围内时, 通过缓存器抖动能够得到补偿, 媒体内同步能得到保证; 若超出范围, 就需要启动再同步反馈机制。当然, 在调整期, 播放质量受到一定影响。

4) 调整期 L 与 ba 有关, 应取合理的值, 以快速而稳定地实现再同步。

5) 附加缓存器容量 ba 涵盖从失步的产生 ($qt > uw$ or $qt < lw$)、检测 ($bt > uw$ or $bt < lw$, 与 a 有关) 到重新再同步的整个时间段, 才可以将失步对播放质量的影响减至最小。

4 方案评价

对网络的分析是基于包丢失和包延时的, 并且还需要实验来确定其中的一些参数。这实际上是统计方法, 一方面应该对包交换网络的情况做更深入的分析, 如果可以及时地反映网络的拥塞, 那么该策略会起到很好的效果^[3]。

QoS 控制只是解决了问题的一个方面, 由于 IP, UDP, RTP 这三个层面都没有提供可靠协议, QoS 控制策略充其量只是“尽力而为”地减少包丢失, 而不能杜绝, 所以需要研究新的适合实时多媒体传输的纠错机制。

参考文献:

- [1] Panagiotis. Interparticipant Synchronization in Real-Time Multimedia Conferencing Using Feedback[J]. IEEE/ACM Trans. Networking, 1996, 4(2): 73-78.
- [2] 罗明宇, 卢锡城, 韩亚欣. Internet 多媒体实时传输技术[J]. 计算机工程与应用, 2000(9): 44-50.
- [3] 张兵, 赵跃龙. 视频会议数据在 IP 网上的传输[J]. 多媒体园地, 2001(7): 51-55.
- [4] 李鹏, 朱光喜. 视频会议中可靠实时音频的传输[J]. 电子技术, 1999(5): 30-34.
- [5] Schulzrinne H. RTP Profile for Audio and Video Conference with Minimal Control(RFC 1890)[S]. 1996.

(上接第 95 页)

- Coding[A]. Published in the proceeding of ITC-CSCC'98 [C].[S. L.]:[S. n.], 1998. 245-248.
- [10] Huang Chung-ming, Hsu TZ-Heng. A User-Aware Prefetching Mechanism for Video Streaming World Wide Web; Internet and Web Information Systems[M]. The Netherlands; Kluwer Academic Publishers, 2003. 353-374.
- [11] Zhang zhi-Li, Wang Yuewei, Du D, et al. Video staging;

- A proxy-server-based approach to end-to-end video delivery over wide-area networks[J]. IEEE/ACM Transaction Networking, 2000, 8(1): 429-442.
- [12] Rejaie R, Yu Haobo, Handley M, et al. Multimedia proxy caching mechanism for quality adaptive streaming applications in the Internet[A]. in Proceedings of IEEE INFOCOM [C]. Tel-Aviv, Israel:[S. n.], 2000. 980-989.