

# 分布式多媒体系统服务质量管理机制的研究\*

王兴伟 张应辉 刘积仁 李华天

(东北大学软件中心研究部 沈阳 110006)

**摘要** 分布式多媒体系统的发展向服务质量 QoS (quality of service) 管理机制提出了新的挑战。本文从分析现有网络系统在 QoS 管理方面的不足出发,引出分布式多媒体系统中有关 QoS 管理的基本观点,讨论各层 QoS 参数和 QoS 管理机制的功能组成,并且给出一个实例,最后得出结论。

**关键词** 多媒体, 分布式多媒体系统, 服务质量, 服务质量管理, 多媒体通信。

**中图法分类号** TP393

近年来,随着高速网络技术和多媒体工作站技术的不断进步,出现了很多新型的分布式多媒体应用,如远程学习、计算机会议、远程医疗诊断等。用户对这些应用一般都有一定程度的服务质量 QoS (quality of service) 要求<sup>[1~4]</sup>,需要分布式多媒体系统提供强有力的 QoS 管理机制满足用户的要求。

服务质量 QoS 是指服务性能的聚集效应,它决定用户对特定服务的满意程度。<sup>[5]</sup>然而,现有网络对 QoS 概念的支持还比较薄弱。多数网络虽然允许服务用户说明对服务提供方的 QoS 要求,但是通常并不保证一定满足用户的 QoS 要求。例如,在 IP 协议中,只允许 IP 用户通过 IP 分组头部中的服务类型域进行简单的定性 QoS 说明,不能进行严格的定量 QoS 说明,而且下层网络机制也不提供 QoS 保证。实际上,现有的多数网络系统(如 Internet)在设计时并没有考虑对定量 QoS 的支持。其次,多数现有网络中的 QoS 一般都具有静态性。通信双方一经商定 QoS 参数值,在通信过程中,就不能对 QoS 参数值进行动态“再协商”。例如,用户不能选择将已有的一条视频连接的质量从彩色降级到单色,以便打开一条新的音频连接。此外,服务提供方必须承诺在整个连接生命期内提供协商好的 QoS。如果不能继续维持协商一致的 QoS,服务提供方往往就单方面地关闭连接。

因此,现有网络的 QoS 管理机制难以适应分布式多媒体系统及其应用的 QoS 要求,需要有新型的 QoS 管理机制来支持分布式多媒体系统的研究与发展。

## 1 分布式多媒体系统中的 QoS 管理基本观点

近年来,随着研究工作的不断深入,人们对分布式多媒体系统中的 QoS 概念及其管理机制的认识也不断深化,形成了如下基本观点。

### 1.1 QoS 管理机制面临的新的挑战主要来源于连续媒体

多媒体数据可以分成静态媒体 (Static Media) 和连续媒体 (Continuous Media) 两种类型。<sup>[1,2]</sup>静态媒体(如文本)是没有时间维的媒体,即其播放速度不会影响所含信息的再现。静态媒体也称离散媒体 (Discrete Media)。连续媒体(如视频和音频)是由媒体“量子”(如视频帧和音频采样)组成的,具有隐含的时间维,如果播放速度得不到满足,媒体信息的完整性就会受到影响。

连续媒体数据的录制、访问、传送与播放过程具有较强的实时性和等时性 (Isochronous), 会实时产生很大的数据量。因此,在网络上传送或在计算机上存储多媒体数据之前,一般都要在信息源对其进行压缩,而在目的地解压缩后播出。图 1 是多媒体数据在网络中流动的示意图,在处理连续媒体数据时,不仅需要保持同一媒体内的时间连续性,而且常常需要维持不同媒体间的同步关系。例如,播放一部影片,就不仅需要维持视频和音频信号本身的时间连续性,而且

\* 本文研究得到国家“九五”攻关项目基金资助。作者王兴伟,1968年生,在职博士生,讲师,主要研究领域为分布式多媒体信息处理技术,计算机网络。张应辉,1972年生,硕博连读生,主要研究领域为分布式多媒体信息处理技术。刘积仁,1955年生,博士,教授,博士生导师,主要研究领域为分布式多媒体信息处理技术与方法学,CSCW。李华天,1922年生,教授,博士生导师,主要研究领域为分布式多媒体信息处理技术与方法学。

本文通讯联系人:王兴伟,沈阳 110006,东北大学信息学院计算机系 129 号

本文 1996-10-15 收到原稿,1997-03-10 收到修改稿

需要在视频与音频信号之间维持严格的“对口型”关系,因此,多媒体信息处理的复杂性主要来源于连续媒体数据, QoS 管理机制面临的新的挑战也主要来源于连续媒体数据.

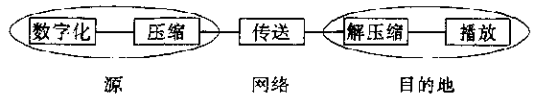


图1 多媒体数据在网络中的流动

### 1.2 QoS 管理应该是可配置的

分布式多媒体应用的 QoS 需求呈多样化趋势.<sup>[2~4]</sup>不同类型的媒体对 QoS 参数值的要求可能相距甚远.例如,一个未经压缩的 NTSC 制式的真彩色视频应用需要高达 27M 字节/s(30 帧/s \* 640 \* 480 像素/帧 \* 24 比特/像素)的带宽,而一个电话质量的音频应用只需要 64Kbps 的带宽.不同应用对 QoS 的具体要求通常也有所不同.例如,实时交互式计算机会议系统一般要求端到端通信延迟在 250ms 以下;多媒体电子邮件则无此严格要求.因此, QoS 管理机制应该是可配置(Configurable)的,允许用户对系统的 QoS 管理功能进行适当的剪裁,以便建立与应用相适应的 QoS 级别.

### 1.3 QoS 管理应该是完全端到端的

QoS 管理应该建立在从源地到目的地的完全端到端(Full End-to-End)基础之上.<sup>[1,2,4]</sup>例如,一个用于远程视频播放的端到端活动就可能由以下几部分组成:从媒体服务器获得视频,在源地进行压缩,将其传送到目的地,在目的地进行解压缩并根据播放窗口的大小对视频按比例进行调整,最后在视频窗口播放(参见图 1).在该端到端路径上,任何一个环节违背了 QoS 要求,都会损坏播放的完整性(除非其下游环节能够予以补偿).因此,用户应该从各个环节所支持的 QoS 中抽象出来,为完全的端到端会话配置 QoS.在分布式多媒体系统中,通常使用术语“流”(Flow)<sup>[1,4]</sup>来表示从媒体数据的产生、传送直至最终消费所经过的完整的端到端路径.流应该由一个 QoS 语句来控制.

### 1.4 QoS 管理应该是层次化的

目前,人们还没有就分布式多媒体系统的体系结构参考模型达成共识.但是,对于系统应该采用分层结构这一点已为人们普遍接受.如果将完全端到端的 QoS 管理观点映射到分布式多媒体系统的分层结构上,就形成了层次化 QoS 管理观点.

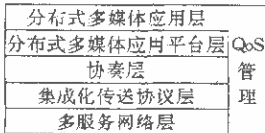


图2 分布式多媒体系统分层模型

综观国内外研究成果<sup>[1~6]</sup>,我们认为分布式多媒体系统采用如下层次模型比较合适,如图 2 所示.其中,多服务网络(Multiservice Network)层提供下层基础网络设施,既要为新型的分布式多媒体应用提供网络支持,也要支持传统网络服务. ATM 是一种比较好的多服务网络.<sup>[5]</sup>此外,快速 Ethernet, FDDI- I 等也可以提供多服务支持.集成化传送协议层既要提供支持连续媒体数据等时与实时传送的高速传送协议,支持媒体内时间连续性的实现,也要提供支持静态媒体数据异步传送的传统协议,因此该层应该支持协议矩阵(Protocol Matrix)和服务类(Service Class)概念<sup>[2]</sup>,允许高层根据应用的实际需要对该层协议与服务进行适当的剪裁和配置.协奏(Orchestration)层<sup>[2]</sup>支持媒体间同步(如“对口型”关系)的实现.平台层提供基于对象的分布式多媒体应用程序设计环境. RM-ODP 的计算模型和工程模型为该层提供了一种可资借鉴的模型,但是由于 RM-ODP 在制订时并没有充分考虑连续媒体的需要,因此,需要对该模型进行必要的扩充.<sup>[1]</sup>分布式多媒体应用层向端用户提供各种分布式多媒体应用. QoS 管理机制贯穿模型的所有层次,逐层管理,各司其职,相互协作,最终提供满足用户需要的 QoS.

### 1.5 QoS 应该是动态的

在分布式多媒体系统中,仅在会话开始时说明 QoS 参数值并且要求它们在会话期间都保持不变是不够的,在实际应用中也不易实现.因此,系统应该提供一种比较灵活的界面,允许用户根据实际情况在会话期间动态地变更 QoS 参数值<sup>[2~4]</sup>,提供动态 QoS 控制能力.

### 1.6 QoS 承诺

为了适应不同应用对 QoS 的不同要求以及动态 QoS 管理的需要,分布式多媒体系统应该提供多种不同的 QoS 承诺,可以分成 3 类<sup>[4]</sup>:

- 确定型(Deterministic)QoS 承诺 在通信过程中,提供 QoS“硬”保证,确保通信各方协商好的各 QoS 参数值,不允许有任何违背,否则可能会造成严重后果.这类服务一般用于硬实时应用(如远程医疗诊断中的患者 X 射线影像数据的实时无差错传送).

- 统计型(Statistical)QoS 承诺 在通信过程中,提供 QoS“软”保证,允许对通信各方协商好的各 QoS 参数值有一定比例的违约,适合于软实时应用(如远程影片点播).

• “尽全力”型(Best-Effort)QoS 承诺 不提供任何 QoS 保证. 由于带宽的限制, 广域网(如 Internet)中的现有分布式多媒体服务多属于这类服务.

## 2 QoS 参数集

不同应用的 QoS 要求通常不同, 因此应该允许用户使用 QoS 参数来定量和定性地说明各自的 QoS 要求. 下面, 我们就自底向上依次介绍图 2 各层所用的 QoS 参数集.

### 2.1 多服务网络层

ITU-T(原先的 CCITT)为 ATM 虚电路制订的 QoS 参数集是多服务网络层上的一组比较全面的 QoS 参数, 这些参数是: 信元峰值到达速率、峰值持续期、平均信元到达速率、阵发性、信元丢失率、信元插入率和比特出错率.

### 2.2 集成化传送协议层

该层使用的 QoS 参数主要有吞吐量、端到端延迟、端到端延迟抖动、出错率和优先级. 其中, 端到端延迟抖动(End-to-End Delay Jitter)是指在一条传送连接上 T-PDU 端到端延迟的最大变化量. 多媒体应用, 特别是实时交互式多媒体应用对端到端延迟有严格限制, 不能超过人所能容忍的极限, 否则会严重影响服务质量. 同样, 延迟抖动也必须维持在严格界限之内, 否则会造成语音的间断、视频的“痉挛”, 从而严重影响人对语音和视频信息的识别. 不同的分布式多媒体应用对 QoS 要求的严格程度不同, 因此其优先级应该有所不同, 如远程医疗诊断应该比多媒体电子邮件优先获得服务.

### 2.3 协奏层

该层的 QoS 参数应该是最大可容忍的媒体间失步程度, 维护媒体流之间所需的“对口型”关系. 还应指出, 用户对两个媒体流 A 和 B 之间的最大可容忍失步程度可能是不对称的. 例如, 由于人们已经习惯于“观景先于听音”, 因此在播放一部影片时, 可容忍的语音对视频的最大超前量比可容忍的视频对语音的最大超前量小. 实验结果表明, 对 NTSC 制式的视频应用来说, 前者的值在 120ms 左右, 而后的值约为 240ms. 因此, 更确切地说, 该层的 QoS 应该用如下两个参数来表示:  $ahead(A, B)$ : 最大可容忍的 A 对 B 超前量;  $lag(A, B)$ : 最大可容忍的 A 对 B 滞后量.

### 2.4 平台层

有关该层的 QoS 应该是完全端到端的. 该层的 QoS 参数应当是面向用户的, 而不应该是面向系统的, 低层方面的考虑应当对用户透明, 以方便应用的设计与实现. 此外, 该层的 QoS 参数应该是说明性的, 即用户应该说明的是他们需要什么而不是这些需要如何去实现. 因此, 我们可以考虑将常用的各种 QoS 水平抽象为各种标准流, 例如 MPEG 1 质量的视频流、CD 质量的音频流等, 作为 QoS 参数集, 供应用设计者们选择.

### 2.5 应用层

该层应向端用户提供各种播放片断, 如用来演示具有 HDTV 质量视频和 CD 质量音频的影片播放片断, 作为 QoS 参数集, 以直观、形象的方式向端用户演示不同的服务质量, 供用户选择.

注意, 应用可能要求 QoS 管理机制对同一层中各 QoS 参数提供不同程度的 QoS 承诺. 以传送层 QoS 参数为例. 文件传送应用一般要求 QoS 管理机制对出错率提供确定型承诺, 而对吞吐量只要求 QoS 管理机制提供“尽全力”型承诺; 计算机会议系统通常要求 QoS 管理机制对延迟、延迟抖动、吞吐量提供确定型承诺, 而对出错率一般只要求 QoS 管理机制提供统计型承诺. 还应指出的是, 对于各层所用的 QoS 参数集目前还没有完全达成共识. 我们相信, 随着研究的深入, 人们逐渐会对 QoS 参数的最优集合达成共识.

## 3 QoS 管理机制的功能

分布式多媒体系统中的 QoS 管理机制应该贯穿图 2 所示模型的各个层次, 每一层都要有相应的 QoS 管理器, 完成以下功能:

### 3.1 QoS 协商(Negotiation)

分布式多媒体应用的用户通常都希望系统提供一定程度的 QoS, 因此, 在使用服务之前, 用户应该将其特定的 QoS 要求通知系统, 进行必要的协商, 以便就用户可接受、系统可支持的 QoS 参数值达成一致, 成为用户和系统共同遵守的“合同”. 在图 2 所示的模型中, 用户和系统之间的协商包括双边对等协商、层间协商以及三边协商 3 种类型, 参见图 3.

双边对等协商是在两个服务用户(一为呼叫方, 一为被叫方)之间进行, 不允许服务提供者修改由服务用户提出的建议值. 双边层间协商在服务用户和服务提供者之间进行. 三边协商在两个服务用户(呼叫方和被叫方)以及服务提供

者之间进行,每个 QoS 参数用一个最小申请值和一个上限来表示。在三边协商中,呼叫方的最低目标是获得最小合同值,服务提供者可以朝着上限值方向修改最小值,被叫方作出最后决定,并将最终结果返回给呼叫方。三边协商可以看作是双边对等协商和层间协商的合成。

### 3.2 QoS 的监控、适配、再协商

当用户和系统就 QoS 达成一致之后,用户就开始使用分布式多媒体应用。在此期间,QoS 管理机制需要对 QoS 合同进行适当的维护,对用户和系统的行为进行必要的监控。

系统需要监视用户提交的数据量,使之符合 QoS 合同值的要求,称为通信量“定型”(Traffic Shaping)。当用户向系统提交的通信量超过 QoS 合同值时,系统将拒绝传递多出的通信量并且将有关情况通知用户;如果用户不将自己的通信量降至 QoS 合同值,则系统有权单方面撤回向用户提供的服务,称为通信量“管制”(Traffic Policing)。

如果由于网络负载增加等原因造成系统向用户提供的 QoS 实测水平低于合同值,则 QoS 管理机制可通过适当调整端系统和网络中间节点的 CPU 处理能力、网络带宽、缓冲区等资源的分配与调度算法进行细粒度调节,尽可能将 QoS 实测水平恢复到合同值,此即 QoS 适配(Adaptation)。如果通过 QoS 适配过程依然无法将 QoS 恢复到合同值,QoS 管理机制就把有关 QoS 降级的实际情况通知用户。用户可以重新与系统协商 QoS,根据当前实际情况就 QoS 合同值达成新的共识,此即 QoS 再协商(Renegotiation)。QoS 再协商的过程实际上是对 QoS 进行粗粒度调节的过程。当然,用户也可主动要求提高或降低 QoS 水平,即 QoS 再协商也可由用户主动发起。

### 3.3 QoS 映射(Mapping)

在如图 2 所示的模型中,不同层使用的 QoS 参数集是不同的。因此,在协商和再协商 QoS 合同值时,需要在层间进行适当的 QoS 参数集及各参数值的转换。通过映射,各层都将获得适合于本层使用的 QoS 参数,供协商和再协商之用。

### 3.4 计费管理

从向尽可能多的端用户提供服务的角度来看,我们应该将计费管理纳入 QoS 管理机制的范畴,将用户所申请的 QoS 水平和应付的费用有机地结合起来。如果没有计费概念,端用户往往就会选择系统所能提供的最佳 QoS 而不是可接受的 QoS。这通常会大大减少系统可同时服务的端用户数。因此,QoS 管理机制应该有计费管理功能,提供“QoS 减低、收费越少”的计费策略,从而鼓励用户选择性能价格比最优的可接受的 QoS。只有通过引入适当的计费策略,QoS 协商才有可能成为真正意义上的协商。

## 4 一个实例

IBM 欧洲网络中心的 HeiProject 项目开发了一种综合 QoS 管理模型<sup>[4]</sup>,在端系统和网络中提供 QoS 保证,其结构如图 4 所示。其中,数据链路层采用 HeiDL(heidelberg datalink)协议,提供数据链路层功能,建立与各种网络的界面。网络层使用 HeiDL 提供的服务来发送与接收数据,建立多点播送组。HeiDL 将收到的数据分组提交给相应的 ST-1 连接。HeiDL 根据所用网络的类型(令牌环、FDDI 等)及其实际能力,按照数据分组的时间关键性安排这些分组的发送。网络层使用 ST-1 协议,提供确定型和统计型服务。ST-1 是一种面向连接的网络层单工“流”协议,支持连续媒体传送。流是一个源和一个或多个目的地之间的单工连接,在建立连接期间,通过交换“流规范”协商资源需求。此外,网络层还通过 QoS Finder 算法支持基于 QoS 的路由选择和 QoS 过滤。传送层使用 HeiTP 协议,支持数据分组的分段与重组、基于速率的流控、用户可剪裁的差错控制机制,而且提供 QoS 映象和媒体调节功能,以便适应网络负载的变化。在 Heidelberg QoS 模型中,HeiRAT(resource administration technique)资源管理技术是提供端到端 QoS 保证的关键。实际上,HeiRAT 是一种综合 QoS 管理模式,包括 QoS 协商、QoS 计算、QoS 执行、接纳控制与资源调度。HeiRAT 调度策略采用速率单调模式,即执行协议处理功能的操作系统线程的优先级与该线程接收报文的速率成正比。

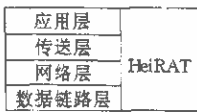


图4 Heidelberg QoS模型

对于确定型 QoS 承诺来说,HeiRAT 根据最坏情形下的资源需求为应用预约和分配资源;对于统计型 QoS 承诺来说,HeiRAT 根据平均资源需求为应用预约和分配资源。



图3 协商示意

### 5 结束语

QoS 管理是设计分布式多媒体系统时必须解决的主要问题之一. 由于不同分布式多媒体应用的 QoS 需求不同, 因而进一步加大了 QoS 管理机制设计与实现的难度. 如何设计并实现一种通用的分布式多媒体系统 QoS 管理机制, 仍然是一个没有得到很好解决的问题. 这就需要我们进一步研究与探索, 以推动分布式多媒体系统的发展.

### 参考文献

- 1 Coulson G *et al.* Supporting the real-time requirements of continuous media in open distributed processing. *Computer Networks and ISDN Systems*, 1995, 27(8):1231~1246
- 2 Blair G S *et al.* System support for multimedia applications; an assessment of the state of the art. *Information and Software Technology*, 1994, 36(2):203~212
- 3 Ferrari D *et al.* Network support for multimedia: a discussion of the tenet approach. *Computer Networks and ISDN Systems*, 1994, 26(10):1267~1280
- 4 Campbell A *et al.* A quality of service architecture. *Computer Communication Review of the ACM*, 1994, 24(2):6~27
- 5 Raif O O. *Asynchronous transfer mode networks, performance issues*. Boston, London: Artech House, 1994. 39
- 6 Ronald J V. ATM concepts, architectures, and protocols. *Communications of the ACM*, 1995, 38(2):30~38
- 7 Dong Xuan-ming, Xu Guang-you. Architecture reference model for distributed multimedia systems. *Mini-Micro Systems*, 1996, 17(1):1~6
- 8 Wolf L C *et al.* The system architecture of the heidelberg transport system. *ACM Operating Systems Review*, 1994, 28(2):51~64

## Research on Quality of Service Management Mechanisms in Distributed Multimedia Systems

WANG Xing-wei ZHANG Ying-hui LIU Ji-ren LI Hua-tian

(Research Division of Software Center Northeastern University Shenyang 110006)

**Abstract** The rapid advances in distributed multimedia systems pose new challenges to the QoS (quality of service) management mechanisms. In this paper, the drawbacks of the QoS management mechanisms in the existing computer networks are introduced, and then the basic viewpoints about the QoS in the distributed multimedia systems are presented. The QoS parameters at each layer and the QoS management functions are discussed in detail. An example is also described. Finally, some conclusions are reached.

**Key words** Multimedia, distributed multimedia systems, quality of service, quality of service management, multimedia communication.

**Class number** TP393