

基于 RTP 协议和 MPEG-4 可扩展编码的非均匀 QoS 控制机制

Unequal QoS Control Mechanism Based on RTP and MPEG-4 Scalable Coding

史 轶 林守勋

(中国科学院计算技术研究所数字化技术研究室 北京 100080)

Abstract Since the significance of each layer in scalable coding is different, making use of this characteristic can reasonably use such resources as bandwidth, processing and caching by deploying varied QoS control strategies for different layers. This paper presents an idea of unequal QoS control and the implementation mechanism at two ends of transmission based on RTP and MPEG-4 scalable coding, redefines the sequence number field of RTP header and extends the RTP header. These can not only realize unequal bandwidth pre-allocation and unequal error protection, but also rapidly and precisely estimate the delay jitter and packet loss ratio.

Keywords QoS control, RTP, Scalable coding

1 引言

随着 IP 网络、ATM 网络和无线网上传输的多媒体实时数据量不断增加,网络资源增长的相对滞后和需求之间的矛盾日益突出,尤其是局部拥塞的频繁出现,而传统的路由器通常不提供拥塞控制^[1],因而对基于传输两端进行 QoS 控制的要求也越来越迫切。和 QoS 控制同样重要的是流媒体协议与编码^[2],而且两者是 QoS 控制的基础。本文将 MPEG-4 可扩展编码^[3]的自适应能力和灵活性作为提高 QoS 控制能力的前提,选择当前实时传输领域最重要的 RTP 协议^[4]作为协议基础,并在两者的基础上研究了传输两端对带宽、丢包、延迟、传输错误等 QoS 的非均匀控制机制,而且不涉及底层硬件。

当前许多基于可扩展编码的 QoS 控制机制并没有充分体现可扩展的独特性,既然基本层和每层增强层的重要性各不相同,而且基本层内不同部分(比如不同帧、不同 VOP)的重要性也不尽相同,如果对重要性不同的各层分别安排不同的 QoS 策略,必能使带宽、处理、缓存等资源得到更为合理的利用,因此本文提出了非均匀 QoS 控制的思路,以及基于

RTP 协议和 MPEG-4 扩展编码的实现方法。

另外,如果接收端能够尽早区分出接收的 RTP 包是基本层包还是增强层包,就能根据其重要性的不同有区别地处理,为此本文第 3 节提出了一种对 RTP 包头的重新定义和扩展方法,从而使接收端在分析 RTP 包头后就能准确区分出不同重要性的包(而当前的方法要在分析载荷后才能区分出基本层包或是增强层包),这对于实现非均匀 QoS 控制有着积极的意义。

QoS 控制总是在一定的框架下实现,本文描述了一个基于 RTP 协议和可扩展编码的流服务基本框架,以后讨论都是基于该框架进行,提出了对 RTP 包头的重新定义和扩展方法,阐述了非均匀 QoS 控制思想和实现方法,最后总结全文。

2 基于 RTP 协议和可扩展编码的流服务基本框架

一个基本的流服务框架主要包括编码器、流媒体服务器、传输协议、控制协议、客户端同步控制、解码器六部分,根据选择的编码方式、传输协议等不同,基本框架也会不同。

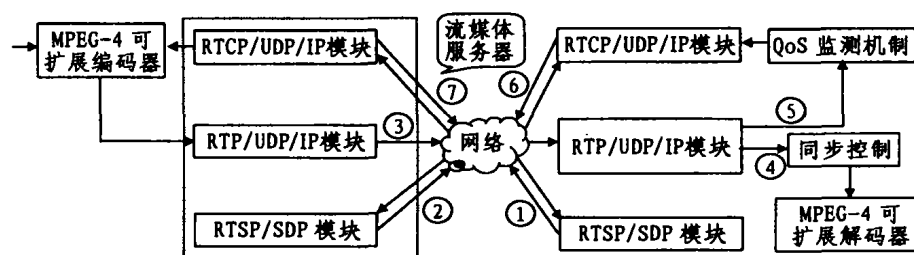


图 1 基于 RTP 协议和 MPEG-4 可扩展编码的流服务基本框架

图 1 所示的是基于 RTP 协议和 MPEG-4 可扩展编码的一种流服务基本实现框架,可扩展编码选择基于 MPEG-4,而不是 MPEG-2 或其他,是因为 MPEG-4 压缩率更大使得传输量更少,而基于对象的特性使得非均匀传输控制和错误保护更为有效。接收端和发送端的交互过程大致如下:

①接收端通过 RTSP 协议与发送端建立会话连接;②发送端通过 SDP 协议将会话的基本信息通知接收端;③发送端将流化后的 RTP/UDP/IP 数据包向外推送;④接收端将收到

的包解包、重组并送到同步控制模块和 MPEG-4 可扩展解码器进行同步、解码,最后播放;⑤同时,根据包发送时间、到达时间和 RTP 包头中序列号等信息对 QoS 进行监测;⑥接收端有间隔地发送 RR/RTCP/UDP/IP 包向发送端反馈拥塞信息,或发送 BYE/RTCP/UDP/IP 包结束会话;⑦发送端有间隔发送 SR/RTCP/UDP/IP 包,以报告发送信息和成员信息。

本文余下部分的讨论都是基于上述流服务框架进行的。

3 RTP 包头的重新定义和扩展

本节首先提出一种利用 RTP 包头的序列号域来区分基本层包和增强层包的方法。根据 RTP 协议^[4]规定, RTP 包头分配有 16 位的序列号域。本文用序列号域的前 n 位作为“层标志位域”以区分包所属的层,其余的 $16-n$ 位为“层内序列号域”。例如,若编码后分成 6 层就用序列号域的前 3 位作为层标志位域,前 3 位为 000 的包表示基本层包,为 101 的包表示增强层 5 的包。每层分别选择一随机数作为该层层内序列号域的初始值,即该层第一个 RTP 包的层内序列号值,该层其他包的层内序列号值在该初始值基础上依次递增,选择随机数作为初始值是为了安全的目的。发送端在会话连接初始化阶段通过 SDP 将可扩展编码后的分层层数通知接收端,接收端据此计算 n 为多少,并随后根据 n 位层标志位区分接收的 RTP 包所属何层,然后作出相应处理。

另外,根据 RTP 协议的描述,接收端根据 RTP 包的到达时间和该包包头的时间戳来估算传输延迟,但是时间戳记录的是包内载荷的采样时刻,和传输时刻并没有对应关系,当载荷为 MPEG 码流时更是如此^[4]。精确判断传输延迟的变化是 QoS 控制的必要条件,为此本节通过 RTP 包头的扩展,在扩展包头中记录下打包时刻(与传输时刻有固定对应关系)的 NTP 时间戳 P_i ,包到达接收端的 NTP 时间 T_i 与该包 P_i 的差值 $D_i = T_i - P_i$,近似等于传输延迟但和传输延迟有一定误差。接收端通过前后包的 D_i 和 D_j 的比较消除误差,从而得到相应时间段内的传输延迟抖动 $J = J + (|D_i - D_j| - J)/16$,可以说,这样计算出的传输延迟抖动既削减了噪声,又能很好地反映网络的波动情况。但是,我们没有必要对每个 RTP 包头都进行扩展,这样会降低包的利用率,只要选择一定时间间隔(比如 1 秒或 2 秒)对 RTP 包头扩展就足以提高接收端对传输延迟变化的判断准确程度。

4 非均匀 QoS 控制机制

从广义上说,QoS 控制大致包括传输前的参数初始化,传输中的接收端监测、反馈和发送端传输控制,传输后的错误控制三部分。本节讨论的非均匀 QoS 控制机制是根据可扩展编码中各层重要性、优先级的不同,分情况、有针对、不对称地对以上三部分进行控制处理,既有传输前对带宽的非均匀初始分配,也有传输中对传输延迟抖动、丢包率等的监测,接收端的反馈和发送端的流控制,还有传输后的非均匀检错、容错和重发措施。

4.1 非均匀带宽资源分配

在带宽资源始终不能满足需求的情况下,根据各层的重要性不同分配各层的带宽 B_i 而保持分配的总带宽不变,这就是非均匀带宽资源分配的思想,将能有效提高带宽资源的利用率。分配给基本层的带宽应最多,以保证基本数据传输的实时性,而随着增强层层数的增加,重要性也逐渐降低,因而分配的带宽也应逐渐减少。本文通过在 SDP 协议中加入带宽限定信息来实现带宽的非均匀分配,比如在基本层和增强层的各个媒体流描述中分别加入“ $b=AS$;带宽限定数”^[5]。另外,可以采用 RSVP 协议为基本层预留足够带宽,RSVP 协议的内容可参考文[6]。

在封装成 IP 包时,根据所属层不同在 IP 包头的 TOS 域^[7]中置入相应的优先级值,I-VOP 的基本层包被分配可获得最高优先级,P-VOP、B-VOP 的基本层包被分配和 TCP

流一样的优先级别,其他包分配的优先级别都应低于 TCP 流的优先级,这样在区分服务范围内既能方便地实现非均匀 QoS 控制,又能很好地保证 TCP 友好性。

4.2 QoS 监测

当前多数基于分层的传输体系都是将各层的包通过不同的端口连接进行传输,接收端只有在分析了 RTP 包的载荷后才能确定该包是属于哪一层的包,这就对 QoS 控制提出了以下问题:如果发送端本应将第 n 层的包从连接 l_n 传输,却错误地从传输第 m 层包的连接 l_m 发送了,接收端仅仅从序列号不一定能确定第 n 层是否丢包,尤其在发送端进行流控制动态增加或减少层数时。

有了第 3 节对 RTP 包头中序列号域的重新定义,不仅接收端在分析了 RTP 包头的“层标志位域”后能够迅速知道该包属于何层,而且即使发生了第 n 层的包从连接 l_m 传输的问题,通过“层内序列号”值的连续情况也可以准确判断每一层的丢包情况,这有利于增强基于分层 QoS 监测的实时性和准确性。

同时,因为 RTCP 包占用的带宽要求保持在总带宽 5% 的范围内,SR 包的时间间隔一般都在 5 秒以上^[4],因而仅仅依靠 SR/RTCP 包来判断网络状况并不能达到及时和实时的要求。通过第 3 节介绍的 RTP 包头的扩展方式,接收端每隔一定时间间隔(1 秒或 2 秒)接收到带 NTP 时间戳扩展的 RTP 包,在将到达时刻与 NTP 时间戳相减后得到近似传输延迟,通过一段时间内近似传输延迟的变化并参考多层的丢包率,可以非常准确、及时地发现网络异常。可以说,QoS 监测的准确性和实时性是整个 QoS 控制的关键。

4.3 接收端的反馈和发送端的流控制

发现网络异常或异常消失后,接收端通过 RR/RTCP 包向发送端反馈。另外,由于基本层的重要性,我们对基本层的丢包采用有限制的重发,即如果接收端通过以下计算认为重传不影响实时播放,就可以通过 RTCP 包向发送端请求某个包的重发或多个丢失包的同时重发:

如果 $T_c + 2 * L_c + D_c < T_d(n)$,可以请求重发。

其中 T_c 表示当前时刻, L_c 表示当前的平均传输延迟, D_c 表示其他不确定因素,包括发送端对重发请求的响应时间和传输过程的不确定损耗时间, $T_d(n)$ 表示第 n 个包被安排解码的时刻。

发送端的流传输控制采用逐层增加多层减少(或称线性递增指数骤减)的策略。即当 RR/RTCP 包反馈网络异常时,发送端一次性暂停多层增强层包的发送,即骤然减少增强层的传输速率,而暂停的层数由发送端根据网络异常的性质和程度决定。当 RR/RTCP 包反馈网络异常消失时,发送端逐层恢复增强层包的发送,即试探性地谨慎增加传输速率。

4.4 非均匀传输错误保护

网络上的传输错误是不可避免的,尤其在无线通信环境下,这里所说的传输错误不仅仅指丢包,即使接收端收到了包,但接收的包与发送端发出的包存在较大差异,也都属于传输错误。对于可扩展编码,基本层的传输错误往往是后果严重的,所以对基本层包的错误我们采用有限制的重发或是 FEC(前向纠错编码)来保护,而增强层可以容许部分错误,甚至允许以错误来换取传输效率,因而采用容错、隐藏错误等手段进行处理,或者干脆不理睬传输错误。基于重要性不同对各层的传输错误进行不同的处理,可以有针对性地处理错误并更加

(下转第 48 页)

触发更新中,各种触发条件的参数对网络负载和路由算法性能的影响也不可忽视。随着 th 和 B 的减小,更新策略变得更加“敏感”,这样无疑会增加信息更新分组从而增加网络负载,但同时也会提高路由算法的性能,因为此时路由算法获得的链路状态信息与当前链路状态信息相同或相近。为了在不增加网络负载的前提下尽量提高路由算法的效率,可以针对更新策略所带来的不确定性,在路由算法中引入概率的思想。例如在触发更新策略中,根据每条链路最近更新的 QoS 信息和当前的 QoS 要求为它们分配一个 safety 值,代表它们能够满足该要求的概率,再根据此 safety 值进行路由选择^[21]。为了弥补更新周期过长带来的信息过时的缺点,可采取 Randomized routing^[10],随机地从多条满足最低要求的路径中找出一条路径,这样可以有效地减少 Call Failure Rate。

结论与今后的工作 本文讨论了几种已有的层次网络中的拓扑压缩算法,通过对它们在不同拓扑结构和更新策略下性能的比较,可以看出大部分情况下全相连算法具有最优的性能,但是在某些特定的条件下,其他压缩算法可以获得与全相连算法相近甚至更优的性能。例如在 edge-core 网络中对 edge 网采用简单点算法,在更新周期长的情况下采用对称点算法。但是全相连算法需要扩散的信息过多,星型算法和生成树算法不但能有效地减少需要扩散的信息量,并且在性能上与全相连网络相近。

影响压缩算法性能的因素有很多,有拓扑结构的划分方法,更新策略的选取和路由算法等等。此外,网络负载的分布对压缩算法也有影响,直观地想,在网络负载比较均匀的情况下,采取对称点算法应有不错的性价比。在路由算法中如何设计一个带宽的函数,使得 Call Failure Rate 最小。在区间更新策略中,等值更新与指数更新各适用于哪些情况?这些都是今后的研究中需要考虑的问题。

参考文献

- 1 Lee W C. Spanning Tree Method for Link State Aggregation in Large Communication Networks. IEEE INFOCOM, 1995. 297~302
- 2 Private network-network interface specification version 1.0 (PNNI): [Technical report]. The ATM Forum technical committee, March 1996, af-pnni-0055.000
- 3 Lui K S, Nahrstedt K. Hierarchical QoS Routing in Delay-Bandwidth Sensitive Networks. In: Global Telecommunications Conf. 2000, GLOBECOM '00. IEEE, Volume: 1, 2000. 410~414
- 4 Van Mieghem P. Topology information condensation in hierarchical networks. Computer Networks, 1999, 31: 2115~2137

- 5 Awerbuch B, Du Yi, Khan B. Routing Through Networks with Hierarchical Topology Aggregation. Computers and Communications, 1998, ISCC '98 Proc. Third IEEE Symposium on, 1998. 406~412
- 6 Hao Fang, Zegura E W. On Scalable QoS Routing: Performance Evaluation of Topology Aggregation. IEEE INFOCOM 2000. 147
- 7 Lee W C. Topology aggregation for hierarchical routing in ATM networks. ACM Sigcomm, 1995, 25(2): 82~92
- 8 Dhandapani G. A Performance Evaluation Architecture for Hierarchical PNNI and Performance Evaluation of Different Aggregation Algorithms in Large ATM Networks: [Technical Report]
- 9 Hao F, Zegura E W. Scalability Techniques in QoS Routing. Networking and Telecommunications Group, College of Computing Georgia Tech, Atlanta
- 10 Apostolopoulos G, Guerin R, Tripathi S K. Quality of Service Based Routing: A Performance Perspective. In: Proc. of ACM SIGCOMM'98 Conf. Sep. 1998
- 11 Shaikh A, Rexford J, Shin Kang G. Dynamics of Quality of Service Routing with Inaccurate Link State Information: [Technical Report CSE-TR-350-97]. Computer Science and Engineering Division, Department of Electrical Engineering and Computer Science, University of Michigan, Nov. 1997
- 12 Guo L, Matta I. On State Aggregation for Scalable QoS Routing. ATM Workshop Proceedings, 1998 IEEE, 1998. 306~314
- 13 Hao F, Zegura E W, Bhatt S. Performance of the PNNI Protocol in Large Networks. ATM Workshop Proceedings, 1998 IEEE, 1998. 315~323
- 14 Awerbuch B, Shavitt Y. Topology Aggregation for Directed Graph. In: Third IEEE Symposium on Computers and Communications, June 1998
- 15 Guerin R, Orda A. QoS Based Routing in Networks with Inaccurate Information: Theory and Algorithm. In: Proceedings of INFOCOM, 1997
- 16 刘进. 大规模层次化网络中保证服务质量的路由算法: [博士学位论文]. 清华大学电子工程系, 2000
- 17 Iwata A, Suzuki H. QoS Aggregation Algorithms in Hierarchical ATM Networks. Communications, 1998. ICC 98. Conf. Record. 1998 IEEE Intl. Conf. on, Volume: 1, 1998. 243~248
- 18 Van Mieghem P. Dividing A Network into Peer Groups to Build A Hierarchical Structure
- 19 Tangmunarunkit H, Govindan R. Network Topologies, Power Laws, and Hierarchy: [Technical Report, ISI]. University of Southern California
- 20 Awerbuch B, Du Yi, Shavitt Y. The Effect of Network Hierarchy Structure on Performance of ATM PNNI Hierarchical Routing. In: Proc. of the Intl. Conf. on Computer Communications and Networks, 1998
- 21 Apostolopoulos G, Guerin R. Improving QoS Routing Performance Under Inaccurate Link State Information. In: Proc. of ITC'16, Edinburgh, Scotland, June 1999
- 22 Zegura E W, Calvert K L. How to Model an Internetwork, INFOCOM '96, Fifteenth Annual Joint Conf. of the IEEE Computer Societies. Networking the Next Generation, Proceedings IEEE, Volume: 2, 1996. 594~602

(上接第 52 页)

合理地使用处理、缓存等资源。

结论 本文针对 IP 网络的实时多媒体应用阐述了在传输两端进行非均匀 QoS 控制的思想及其基于 RTP 协议和 MPEG-4 扩展编码的实现方法。传输前的非均匀带宽资源分配使得带宽得以更加合理有效的利用。通过对 RTP 包头进行重新定义,使得接收端能够快速区分基本层包和增强层包,并准确判断各层的丢包情况和进行非均匀的传输错误处理。通过对部分 RTP 包头的扩展,又使得接收端对传输延迟变化的监测更为准确和及时。非均匀 QoS 控制能够保证流媒体应用的自适应性和实时性,及更合理地使用带宽、处理、缓存等资源。

参考文献

- 1 Braden B, et al. Recommendations on Queue Management and Congestion Avoidance in the Internet. RFC 2309, April 1998
- 2 Wu Dapeng, et al. Streaming Video over the Internet: Approaches and Directions
- 3 ISO/IEC 14496: MPEG-4, Oct. 2000
- 4 Schulzrinne H, et al. RTP: a transport protocol for real-time application. Internet Engineering Task Force, RFC 1889, Jan. 1996
- 5 Handley J. SDP: Session Description Protocol. RFC 2327, Internet Engineering Task Force, April 1998
- 6 Braden R, Ed. Zhang L, Berson S, Herzog S, Jamin S. Resource Reservation Protocol (RSVP). Sep. 1997
- 7 DARPA Internet Protocol. RFC 791, Sep. 1981