

多媒体数据传输策略

Multimedia Data Transmission Policy

彭富春 陈火旺 谭庆平

(国防科技大学计算机学院 长沙400073)

Abstract In this thesis, aiming at the characteristics of multimedia information transmission over dial-up network and LANs, we research the limitations and feasibility of multimedia information transmission over dial-up network and LANs. The article shows some ways to get over the limitations by using some advanced software technology. Finally, the authors research the dynamic multimedia data transmission policy.

Keywords Multimedia, Network, TCT/IP

随着 Internet 网络和多媒体技术的飞速发展,两者的有机结合产生了许多新事物,如远程教学等;同时新的矛盾也产生了,如 TCP/IP 网络上的实时多媒体数据传输存在时延、抖动、带宽变、分组丢失等问题,使基于网络和多媒体数据的服务质量难以保证。

本文针对 Internet 网络上远程教学服务质量难以保证的难点,采用动态多媒体数据传输策略,充分利用网络可用的带宽,面向用户,将时延控制在最低限度,从而提高服务质量。这种方案亦可用于其它多媒体数据传输过程中。

一、Internet 传输多媒体信息的分析

Internet 网络的优势在于异地计算机间的互连、通讯及数据传输,从而实现资源共享,但它能否满足多媒体信息的传输需求要从以下几个方面分析:

1) 多媒体信息对带宽要求。如 CD 质量的立体声为 1.4Mb/s,全运动视频的数据量更是巨大,640 * 480 像素真彩色图像,以 30 帧/S 的速率显示,则数据量将是 $640 * 480 * 24 * 30 = 221\text{Mb/s}$,经压缩后其数据量仍很巨大,而且,多媒体信息传输往往是交互多路的,其带宽需求可能因此增加几倍。所以,10BASE-T 局域网带宽虽已达 10Mb/s,但在传输某些多媒体信息时仍有不足,拨号网络则更加突出。

2) 多媒体数据传输对带宽连续性的要求。音频、视频等媒体是连续的,需要在网络中一直不断地传输,这就要求网络保证它们持续地独享一定的带宽,而拨号网络由于所采用的 TCP/IP 协议的限制,很难满足这一要求。

3) 多媒体信息对网络传输还有实时要求。局域网/拨号网络中,数据的传输是以帧形式传送的。在传输实时多媒体信息时,必须先采集一定时间的多媒体数据来填充帧,然后才能传输帧,这就造成“数据填充延时”。在加上无法预测的传输延时,增加了多媒体信息实时传输的难度。

从以上分析可以看出,Internet 网络传输多媒体信息有其优势,但也有一定的局限性,必须在软件上采取一定的策略,才可满足传输多媒体信息的要求。

二、多媒体传输系统中所存在问题的解决策略

1 媒体信息传输实时性的实现

(1) 消除传输瓶颈 拨号网络中两台计算机间的点对点数据传输过程如图 1 所示,设所需传输的数据已经存于计算机 A 的内存或硬盘。

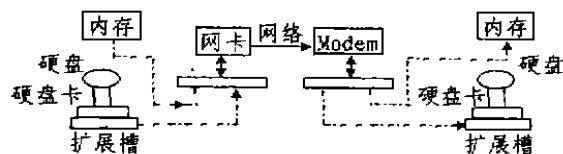


图 1 数据在拨号网络中的传输过程

对计算机硬件的性能进行分析后可以确定,调制解调器(即网络传输速率)是最窄瓶颈,其次为硬盘。在多媒体数据传输中,各种媒体信息基本上都经过采集-处理-发送接收-重现这一过程,如果系统设计科学,设备先进,完全可以用边采集边处理边传输数据(接收方边处理边重现)的方法减少数据的存储量。当数据的存

彭富春 硕士,研究方向为多媒体与网络技术,陈火旺 中国工程院院士,谭庆平 教授,研究生导师,

储量减少到一定程度,内存就可以承担起数据存储中转的任务,从而避开硬盘这一瓶颈。但是网络带宽这一瓶颈是数据传输的必由之路,无法消除,只能适当调配多媒体信息的数据组织方式,使之适应网络。

(2)以多线程编程方法来满足不同工作的时间需求 在多媒体系统中,许多工作要同时进行;如对声音的采样要持续不断,对已采集的数据要同时进行压缩,压缩过的数据要即时送出,而且在这三件事中,压缩的工作量较大,但又不能滞后于其他两项工作,因此它需要较多的时间片,优先级高的线程优先分配到较多的时间片,优先级低的线程在优先级高的线程执行后分配到时间片,利用 WIN98支持多线程及线程优先级的特性,我们可以从容地调度上面所举的多媒体处理中的各项工作,即可以把多媒体处理进程的优先级设得最高;在多媒体处理进程内部,为声音的采样压缩和发送各创建一个线程,使它们可以同时工作,分享 CPU 时间;声音数据压缩线程的工作量相对于声音采样和传送线程要大些,可以把它的优先级设得高些,以使它获得更多 CPU 时间片;声音采样线程一直保持活跃,而数据压缩和发送线程在无数据处理时,可以休眠或挂起;当采样线程采满了一个缓冲区的数据后,可以恢复或唤醒数据发送线程,让它把数据通过网络传输出去。用这种实现方法,不仅可以保证三项工作同时进行,而且由于相应的三个线程同处于一个进程中,共同进程的数据段、代码段和堆栈段节省了系统资源,而且线程间的数据交换也方便多了。

(3)原始数据包的大小 根据 OSI 参考模型的工作原理,网络是分层次的,在数据发送方,所要传送的数据以帧的形式从高层至低层被处理,直到通过物理层传输出去。在数据接收方,数据的处理顺序相反。在数据发送方,网络的每一层都给数据包加上一个包头和包尾,在接收网络的每一层都剥掉数据包中属于本层的包头和包尾。TCP/IP 协议集并没有专门针对多媒体传输的高层协议,因此需要利用操作系统提供的某些功能编制程序以实现数据传输的管理、对数据的解释及与用户进程的接口等,经 TCP/IP 协议处理的结果,数据帧的最终长度在 406 至 1018 字节之间,其中有大约 49 字节至 90 字节的包头包尾信息。数据帧越长,它所造成的“填充延时”越大,在所采样的数据经压缩处理时“填充延时”更大。但数据帧较短时,其中的帧头帧尾所占的比例较大,有效数据率较小,必然浪费有限的带宽。要根据网络传输效率和数据传输实时要求两方面的因素,确定原始数据包的大小。在实时要求较高时,尽可能取较小的包长。

(4)传输协议的确定 对应于 OSI 模型传输层的协议有 TCP 协议和 UDP 协议。TCP 是面向连接的协

议,它保证数据可靠有序地传输,即不会有位错误,也没有包的丢失、重复或顺序紊乱。UDP 是非连接协议,它不提供传输的可靠性保证,TCP 协议为了保证可靠性,要多做许多工作,如建立和拆除端对端连接;对数据进行顺序控制和流量控制;收方校验收到的数据,并与发方进行应答;数据传输出错重发等等。这些都加大了网络上传输的数据量和传输延时。再者,TCP 协议不支持广播方式发送,当数据发送方要用 TCP 协议送同一份数据给 N 台计算机时,只能和这些用户逐一建立连接,逐一地给它们发送数据,这与 UDP 协议采取的广播方式相比,所传输的数据量和传输时间之比要增加 N 倍以上,因此对于需广播发送的数据,或实时要求较强的数据,可考虑用 UDP 协议。如连续声音数据实时性较强,即使接收失败也不允许重发,而且需要广播发送,可采用 UDP 方式;而控制信息,因实时要求不高,数据较重要,应采用 TCP 协议。

(5)多媒体数据传输优先级的确定 多媒体数据具有数据量大、连接复杂的特性。在服务器收到用户请求后,所需传送的文件可能不止一个,这样在服务器端就出现了如何组织多个文件进入发送队列的问题。为减少时延,有必要对其传输队列进行优化组合,下面我们通过一个具体的例子加以说明。如某系统(如 CAI 系统)内部构造如图 2。此系统共有七个文件,它们之间的关系是一种树形结构。发送端首先应将 P_{00} 置入发送队列,下一个进入发送队列的是 P_{10} 还是 P_{20} ,就需要一种选择机制,在这里我们通过一个优先发送函数 R_n 来确定发送次序,函数 R_n 如下:

$$R_n = \left(\sum_{i=0}^n t_{ij} / \sum_{j=0}^i d_{ij} \right) \rho_i$$

其中 t 表示程序在用户端的运行时间, d 表示程序的数据量的大小, ρ 为该程序在用户端被申请执行的概率, ρ 的获取由系统通过统计而得,它的值是动态变化的。这样 R 的值确定了该程序进入发送的次序;值越大优先级越高。这种发送序列可以使用户端接收与播放同时进行,对用户来说是透明的,从而达到消除或减少时延的目的。

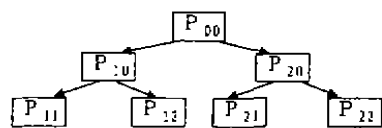


图2 欲传系统的内部文件构造关系

2 多媒体信息传输连续性的实现

动态媒体(如音频)是连续的媒体,要求对其采样也要连续,否则采样得到的数据在重放时将有停顿,影响效果。即使在将采样数据通过网络发送时,也同样不

能中断采样,因此,每当采样数据填满了一个缓冲区,必须立刻提供别的缓冲区来存放后继的采样数据,而已被填满的缓冲区中的数据要立刻被处理,被发送。为解决这一问题,我们设计了一种如图3所示的内存循环使用的方法:从系统申请两段内存,并把它们加入采样值存储区序列中;采样线程每填满一个内存块,便换一个内存块的控制权,并发送信息给传输线程;传输线程收到该消息后,把该内存块中的数据传掉,然后把该内存初始化,再加入到采样存储内存块序列的尾部,两段内存块一直循环使用,保证了采样的连续性,又避免了每次申请释放内存这种较耗时的操作。同样,在接收方动态媒体的再现也采用上述循环方法。但这种方法只解决了采样、重放环节的连续。由于网络传输延时不确定性,使这种方法仍不够完善——当网络延时突然增大时,有可能造成媒体信息接收重放方在放完一帧数据后,仍未收到新的数据帧的情况,难以完全做到媒体实时重放无停顿。而且以上讨论是对采样数据率不变的情况进行分析的,如果边采样边对数据进行压缩,然后填入内存块中,就象某些视频捕捉卡对视频数据所做的那样,那么其数据率就是变量,每个内存块所存数据对应的时间长度也在变化,这时接收方播放连续与否就不仅受传输延时变化的影响,而且也受内存块填充时间变化的影响,因而整个变化的幅度范围也增大了,“2内存块循环方式”就更暴露出了其不完善性。因此,在“2内存块循环方式”基础上又可以设计出类似的“N内存块循环方式”,在采样一方,当数据存满m块内存后才开始传输数据,这样,接收方就有了m块内存的数据缓冲,再佐以其他手段,就足以消除上述不稳定因素对多媒体信息重放连续性的影响了。

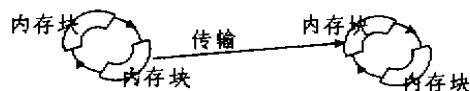


图3 2内存块循环方式

3 多媒体传输交互性的实现

在多媒体传输系统中媒体的交互可以采用单工、半双工、全双工三种方式。单工方式不能满足信息交互

的需要。全双工方式是一种理想的交互方式,真正做到了多媒体信息在各方如同聚在一起交流的效果。由于拨号网络中只有2路信息,可以采用双工方式来实现多媒体传输的交互性。而在局域网中,为了实现三方面的完全交互,任何一方必须接收并处理另两方的信息,而且还必须把本方的信息送给另两方,这样在网络中必须有6路信息,这对网络带宽及实时保证方式提出了较高的要求。而且,A,B,C三台计算机都要同时处理3路信息(接收两路,发送一路),其负荷很大。如果改三方交互为四方交互,则网络上要同时传输12路数据(见图4),无论网络还是计算机都难以承受。因此,我们采用一种合理的管理机制和半双工方式来实现多媒体传输的交互性。

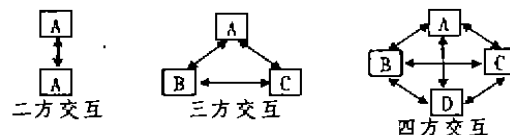


图4 全双工交互

结束语 本文针对 Internet 网络上远程教学质量难以保证的难点,采用动态多媒体数据传输策略,充分利用网络可用的带宽,面向用户,将时延控制在最低限度,从而提高服务质量。这种方案亦可用于其它多媒体数据传输过程中。

参考文献

- 1 冯丹,等. 主-从分布式控制系统通信机制与性能的关系. 小型计算机系统,1997(3)
- 2 Zhang H, Ferrari D, Verma D C. Delay Jitter Control for Real-Time Communications in a Packet Switching Network. Computer Communications, 1992, 15(6): 367~373
- 3 LeGall D J. A Video Compression Standard for Multimedia Applications. Communications of the ACM, 1991, 34(4): 46~58
- 4 Crawford M. A Method for the Transmission of IPv6 Packets over Ethernet Networks. RFC 1972, August 1996