

基于 H.332 协议的纯软件视频会议系统的研究与实现

Research and Implementation of a Software Videoconferencing System Based on H.332

宋玲 陈燕 (广西大学 计算机与电子信息学院 南宁 530004)

摘要:按照 H.323 的扩展协议 H.332, 研究并开发了一个分布式视频会议系统。系统基于 IP 组播网络, 具有与会成员的身份认证机制, 会议全过程采用分布式控制和管理, 并基于 RTP/RTCP 协议, 对系统音视频实时发送和接收过程中的丢包、同步和时延抖动问题给出了有效的处理方法, 保证了一定的服务质量。系统主要功能由软件完成, 成本较低, 具有一定的实用价值。

关键词:IP 组播 分布式视频会议系统 RTP/RTCP 服务质量

1 引言

1998 年 9 月, ITU (国际电信同盟) 在 H.323 的基础上制定了分布式、松耦合扩展协议 H.332, 该协议在网络层使用 IP 组播技术, 在传输层使用 RTP 协议进行组包/解包, 采用分布式松耦合的系统结构, 较好的解决了视频会议系统的规模扩展问题。按照 H.332 协议规范实现的视频会议系统, 不存在集中控制所有终端和集中转发所有数据的 MCU, 取而代之的是用软件实现的多点会议管理器 (会议主持人), 通过网络组播基于将异地分布且动态变化的会议成员组织起来, 不仅可以大大降低系统成本, 提高系统的规模扩展性, 而且方便操作、维护和升级。

H.332 支持分布式的会议系统, 它利用组播技术来组织会议, 参加会议的终端之间以组播的方式传送视频和音频信息, 而无须使用 H.323 中规定的 MCU。

图 1 和图 2 是这两种会议系统的示意图。传统的基于 H.323 协议的多媒体通信系统采用的是一种集中式、紧耦合的体系结构。这种系统以 MCU 为中心, 呈星型结构, MCU 用来集中转发会议数据, 所有终端以点对点方式向其发送数据。会议参与者所产生的多媒体信息流汇集到中心, 经过 MCU 处理后, 以 IP 单播的方式转发给多个接收者。而基于 H.332 协议的分布

式会议系统, 不存在集中式方案中的 MCU, 通过引入会议主持人对与会终端进行有效的管理, 数据传送则利用了分布式网络的组播功能。很明显地, 分布式的系统比集中式系统结构简单, 实现容易, 具有较好的规模扩展性。

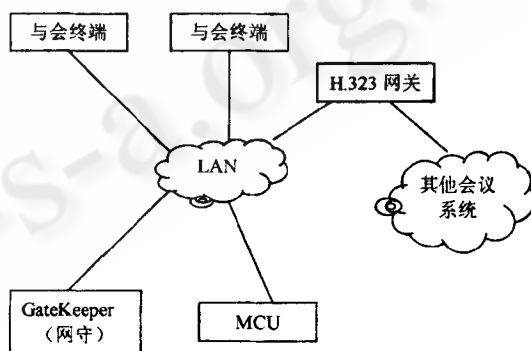


图 1 集中式会议系统结构

2 RTP/RTCP 协议

RTP/RTCP 协议是 IETF 在 1996 年为支持网络数据实时传输而制定的标准。RTP (实时传输协议) 用来提供端到端的实时数据传输服务, 包括载荷标识, 数据序

① 广西科学基金项目 (桂科基 0342011), 广西教育厅项目 (桂教科研 [2004] 20 号)

号,时戳和传输控制。RTP 数据通常采用 UDP/IP 封装,它利用 UDP 的多路复用及校验和服务,共同完成实时数据传输功能。在发送端,开发人员必须把执行 RTP 协议的程序写入到创建 RTP 信息包的应用程序中,然后应用程序把 RTP 信息包发送到 UDP 的套接接口;同样,在接收端,RTP 信息包通过 UDP 套接接口输入到应用程序,因此开发人员必须把执行 RTP 协议的程序写入到从 RTP 信息包中抽出媒体数据的应用程序。

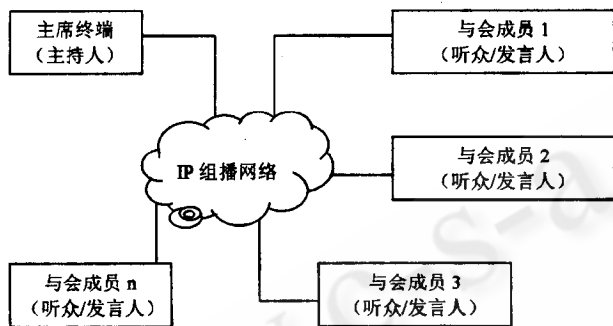


图 2 分布式会议系统结构

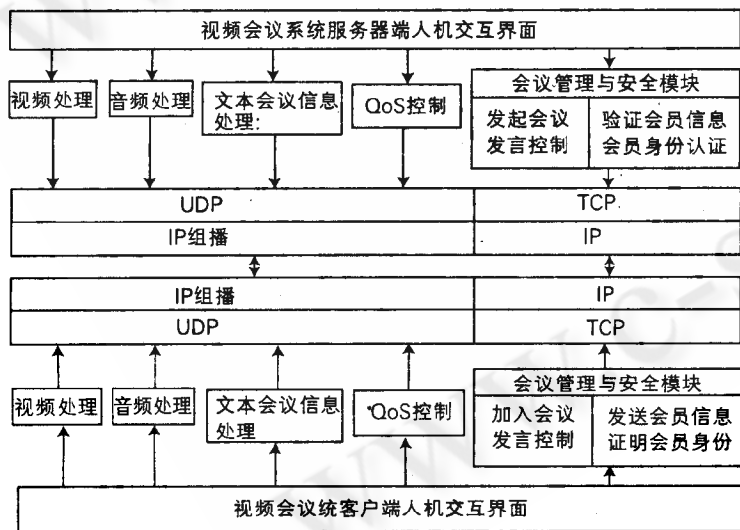


图 3 本视频会议系统功能和结构关系模型图

RTCP(实时传输控制协议),用来控制服务质量,并在正在进行的会话里传送各个参与站点的信息。在 RTP 会话期间,各站点周期性的传送 RTCP 包。RFC1889 定义了 RR、SR、SDS、BYE、APP 等五类 RTCP 控制报文。其中 RR(接收方报告)报文由实时数据接收方发送,RR

报文针对每个信源都提供报文丢失数、已收报文的最大序列号、到达时间抖动、接收最后一个 SR(发送方报告)的时间、接收最后一个 SR 的延迟等信息。

站点利用 RTCP 包中含有已发送的数据包的数量、丢失的数据包的数量等统计资料动态的改变传输速率,调整缓冲区大小和个数,甚至改变有效载荷类型。RTP 和 RTCP 配合使用,能以有效的反馈和最小的开销使传输效率最佳化,保证网上实时传输的 QoS(服务质量)。

3 系统的设计与实现

我们在 H. 332 规范的基础上开发了一套纯软件的分布式视频会议系统。本系统不存在 MCU,采用 PC 摄像头加软件的形式,通过引入会议主持人对与会终端进行有效的管理,同时各与会终端还可以向会议成员分发自己的信息,而数据传送则利用了 IP 网络的组播功能。本系统利用 Windows2000 操作系统内置的软件 CODEC 实现音视频数据的软件编解码,具体的发送/接收过滤器的设计采用微软公司的 DirectShow 组

件技术,开发工具为 Visual C++ 6.0。系统功能和结构关系如图 3 所示。

3.1 身份认证机制

分布式会议系统中会员信息的保密以及对会员身份的验证是至关重要的。会员信息的保密。包括:会员用户名、密码的保密;会员身份的认证,则确保参与会议的会员的身份的可靠性。为此,我们设计了一个基于证书的安全身份认证的方案,主席端(服务器)负责会员信息的解密与会员身份的验证,客户端负责会员信息加密与提供身份证明。

系统中,CA 管理服务器和主席端共用一台机。首先,客户端与服务器端必须事先获得对方的公共密钥并导入到本地的系统中;其次,服务器发起会议前必须指定服务器端使用的证书;第三,客户登录前必须知道并填写服务器及客户端使用的证书。

客户登录时,输入登录会议的用户名及密码,系统用服务器端的公钥对用户名进行加密,然后发送到服务器端,服务器端用自己的私钥对接收到的消息进行解密,如果无法解密,将要求客户检验使用的证书是否

正确,如果能解密,则用解密出来的用户名与密码作为关键字到服务器端预先建立好的会议用户数据库中查找相应的用户,如果无法找到,则登录的用户为非法用户,如果找到,则产生一串随机数据发回给客户,客户再用自己的私钥对该随机数据进行签名并发回给服务器,服务器再用该用户的公钥对该签名过的随机数据进行验证来对用户的身份进行认证,并根据认证结果的匹配与否来决定是否允许用户参加会议。

3.2 分布式控制和管理

(1) 服务器端。主席端(服务器)首先开启。主席在发起会议时,必须提供自己的证书名称。随后,主席端在指定的组播地址上发起会议,同时在指定的端口监听客户端的连接请求。当有客户请求与主席端建立连接时,主席端会检查客户端是否是合法用户,若是,主席端与该客户端就建立一个 TCP 连接,该连接用来传输会议系统中的管理信息和控制信息。

主席端具有优先发言权。当主席没有把发言令牌发放出去时,它可以选择是否发言。会议发言控制部分:所有参加会议的与会者,在通过主席对身份的验证并与会议服务器端成功建立连接以后,其用户名均会显示在主席端和其他与会成员的“与会者列表”里。当与会人员申请发言后,其名字就会显示在主席端“申请发言者列表”里。主席可以根据会议进展情况选定其中一个申请者,给予其发言令牌。当主席自身要发言时,它可以选择“禁止所有”选项而使所有申请发言者不能获得发言权限。

主席端通过对与会成员的入会、离会请求以及发言申请的处理,实现会议系统的分布式的动态管理。

(2) 客户端。客户端启动后,需要与主席端建立 TCP 连接,一旦客户端提供的用户名和密码在主席端得到验证,客户端就成为该会议系统中的一个列席者,可以接收会议组中的音频视频信息和文本信息,但还不可以发言。同时,该客户分发自己的个人信息(如用户名)到组播组的所有成员,并显示在各与会终端的“与会成员名单”处(包括主席端)。该客户同时接收其他与会成员发送来的用户信息,显示在其界面上。当与会成员离开会议时,它发送离开信息给所有终端,各终端把该成员信息从其系统删除。

作为列席者,可以根据会议进展情况举手向会议主席申请发言令牌。列席者发出申请后,其用户名将

会显示在主席端界面的“申请发言者列表”处,由主席决定是否允许其发言。当客户的发言申请被批准后,其就拥有发言令牌,此时,该客户就可以向会议系统发送其音视频信息了。当其发言时,将组播其个人信息以告知系统其正在发言,所有与会成员收到该信息后,将其显示出来。一旦会员发言完毕,停止音视频媒体数据的发送时,该信息也会被组播给所有会员。各与会终端根据收到的信息,在其界面上更改该发言人的角色。

所有与会成员加入和退出会议都要分发自己的个人信息给其他成员,同时接收其他成员发来的信息并显示出来;而当其获得发言令牌后,与会成员将直接组播其媒体数据给所有其他与会者,实现了视频会议系统的分布式控制以及数据传输。

3.3 QoS 控制

一个成熟的会议系统,必须能保证良好的会议服务质量。本系统对音视频的丢包控制、同步和时延抖动等问题做了处理,保证了一定的 QoS。

3.3.1 丢包率控制

在 RTP 协议会话期间,各参与者周期性地传送 RTCP 包。对于接收端,采取最多 5 秒发送一个 RTCP 反馈包到发送端,其中包含从发出上一个反馈包到发送此包期间内的 QoS 信息。而随着与会成员的增加,RTCP 报文的发送间隔会逐渐增大,以保证系统中传输的 RTCP 报文最多占用实时传输带宽的 5%。发送方可以利用 RTCP 包中的 RR 报文的信息动态地改变发送速率。

会议进行中,要优先保证音频数据的可靠传输。当丢包率超过 2% 时,减少视频数据的发送量,从而尽量缓解音频数据包的丢失。在视频发送过滤器中,实时接收由接收端反馈的 RTCP 包,并根据计算出的丢包率,决定是否只发送视频关键帧即 I 帧,而忽略非关键帧。在视频接收端的过滤器设计中,我们规定从视频信息的关键帧开始接收,如果首先到达的视频帧为非关键帧,将其丢弃直到关键帧到来为止,这样可以避免接收端视频画面出现绿屏或者马赛克现象,以提高视频回放效果。当丢包率超过 5% 时,为了保证会议顺利进行,取消视频发送。

除了在发送过滤器里处理是否发送视频非关键帧外,我们还根据丢包率的值在视频捕捉过滤器中对视

频发送帧速做一定限制。丢包率小于 2% 时,按正常速率发送,即 20F/s;当丢包率介于 2% ~ 5% 时,发送速率则在 5F/s ~ 20F/s 间动态变化,而当丢包率大于 5% 时,停止视频发送。

3.3.2 媒体数据间的同步

同步控制是音视频同传过程中要解决的技术关键和难点。同步分为流内同步和流间同步。流内同步是指音频流或视频流数据包的顺序播放,RTP 协议已经解决了这个问题,因为接收端表现过滤器可以根据 RTP 包头的“包序号”字段顺序播放数据包,从而实现流内同步。至于流间同步,解决起来相对难些。DirectShow 的解决方案就是为过滤器图选择一个公共的参考时钟,并且要求传送到表现过滤器的每个样本都打上时间戳;视频表现过滤器和音频表现过滤器根据样本的时间戳以及参考时钟当前的参考时间,来正确安排样本的播放时机。如果到达表现过滤器的样本晚了,则马上播放这个样本;如果样本到达的时间早了,则表现过滤器等待,直到样本时间戳的开始时间再开始播放。以下代码的作用是过滤器设计中给媒体数据打时间戳:

```
STDMETHODIMP // 给媒体样本设置时间戳
CMediaSample::SetTime ( IN REFERENCE_TIME *
pTimeStart, IN REFERENCE_TIME * pTimeEnd )
{ if ( pTimeStart == NULL ) return S_OK ;
REFERENCE_TIME m_rtStart = * pTimeStart ;
if ( pTimeEnd != NULL ) REFERENCE_TIME m_
rtEnd = ( * pTimeEnd ) ;
}
```

3.3.3 时延抖动控制

由于音视频数据包采用 RTP 传输(底层为 UDP 传输),所以数据包到达接收端是无序的,而且到达时间不均衡。合理的缓冲区设置可以既保证时延不超过 400ms,又能减轻音视频抖动。例如:声音采样缓冲区的大小直接影响声音延迟,对 8kHz 采样频率、8bits 采样位数,则采样 1 秒的数据量为 8000 字节,即每 8 字节产生 1ms 延迟,因此,缓冲区越大,延迟越大。本文通过分析反馈的时延信息,合理设置发送端采样缓冲区大小,并随着网络状态变化动态调整接收端缓冲区,从而较好的解决了时延抖动问题。

(1) 发送端缓冲区。发送端声音采样缓冲区的大

小直接影响延迟。DSP Group TrueSpeech (TM) 的帧长是 20 ms,每 8 字节产生 1ms 延迟,因此,每帧原始数据大小为 160 字节。根据上述分析,我们把声音采样缓冲区定为 480 字节,即三帧数据量,采样延迟为 60ms。为了能连续、无间隙地采样,我们采用两个缓冲区来循环存放采样的音频数据。视频缓冲区设置原理相同。

```
WAVEFORMATEX * pWF = ( WAVEFORMATEX * )
pmt -> pbFormat ;
pWF -> wFormatTag = WAVE_FORMAT_PCM ;
pWF -> nChannels = 1 ;
pWF -> nSamplesPerSec = 8000 ; pWF ->
nAvgBytesPerSec = 8000 ; pWF -> wBitsPerSample =
8 ;
```

AllocProp.cbBuffer = 480 ; // 音频缓冲区大小

AllocProp.cBuffers = 2 ; // 音频缓冲区个数

(2) 接收端缓冲区。设置发送端采样缓冲区是为了减少数据处理所带来的延迟并保证数据块的完整,而接收过滤器中缓冲区的合理设置则有助于减少数据处理时延,并减轻数据的抖动,保证音视频的流畅。如下代码则是在接收端过滤器中对音频缓冲区大小的动态处理。每个音频包原始数据量约为 50 字节(数据包头 40 字节,压缩的音频数据约 10 字节)。当 RR 报文反馈的时延值小于 200ms 时,则在接收缓冲区存放 4 个数据包,即缓冲区大小为 200 字节;当时延值介于 200ms 和 400ms 之间时,接收缓冲区存放 3 个数据包,即缓冲区大小为 150 字节;而当时延值大于 400ms 时,接收缓冲区存放 2 个数据包,即缓冲区大小为 100 字节。在接收端建立两个缓冲区,循环接收,可以起到优化接收效果的作用,从而较好地解决了抖动。试验结果表明,我们对缓冲区的处理比较有效,不仅控制了处理时延,还抑制了抖动,音视频都非常流畅。

4 结束语

分布式会议系统能将异地分布且动态变化的与会成员通过 IP 网络组织起来,完成会议全过程。我们开发的分步式视频会议系统符合 ITU 的 H. 332 规范,系统基于 IP 组播技术,采用 RTP/RTCP 协议,其主要功能由软件完成,使成本大大降低,且方便操作、升级和维

(下转第 84 页)

(上接第 80 页)

护。系统具有身份认证机制,保证会议成员的合法性。

参考文献

1 ITU – T Recommendation H. 332 – H. 323 extended for loosely coupled conference. 1998.

2 RFC 1889 (rfc1889) – RTP: A Transport Protocol for Real – Time Applications.

3 Serban Catrava, Forouzan Golshani. A Performance System for High Quality Desktop Video Conference IEEE 2003.