

# 基于 Internet

## 的多点多媒体实时监控系统的

本文介绍了在开发基于 Internet 网络环境的下监控系统过程中, 利用 windows socket 规范和 IP Multicast 技术实现多媒体 (音频和视频) 实时多点传输的原理、方法以及实现过程中涉及的关键技术问题的解决方案。

山东工业大学 计算机系 胡毅  
山东工业大学 自动化系 胡咏梅  
山东工业大学 产业处 黄静

### 引言

多媒体监控系统广泛应用于工业生产、交通、电力和智能办公大楼, 现在正逐步由传统的模拟技术网络向数字化技术监控网络发展。随着监控范围不断的扩大, 计算机网络技术的发展, 特别是 Internet 网络在我国的发展十分迅速, 各种应用应运而生。音频和视频作为人类交流最自然的一种方式地 Internet 上的应用, 已普遍受到人们的关注。基于 Internet 的多点多媒体实时监控系统 (IMMRMS) 是结合 Internet 网络技术、多媒体信息的综合性和监控的实时性的系统。它使网络中分布于不同地点的多个监控者能实时地看到一点或多点被监控对象的多媒体信息, 从而实现同步交互、同步协商的协同管理环境。

### IMMRMS 的设计与实现

IMMRMS 由集中监控中心、协同监控端及网络传输系统组成。实现远程监视和远程控制, 其主要分为监控管理子系统和数据采集子系统。其监控管理子系统又分为集中监控中心子系统和协同监控子系统。集中监控中心子系统主要负责监控的日常工作, 如有需要请示或需协助解决的问题, 则联系协同监控子系统的工作人员。协同监控工作人员通过分组网络中桌面多媒体计算机可以实时获得被监控现场、集中监控中心和其他协同监控人员的多媒体信息, 实施协同控制。数据采集子系统由现场端监控主机及现场监控器组成, 设置在少人维护或无人值守的被监控的现场, 通过网络与监控管理系统进行通信。各子系统都被设计成局域网, 子系统之间通过互连网络进行联系。

该系统主要包括四个方面的功能: 1. 多路音频信号

的处理。2. 多路视频信号的处理。3. 用户信息的管理。4. 监控信息的处理。

#### 1. 网络传输协议

IP 网已被广泛使用在各部门中, 其 TCP / IP 协议是异种计算机和异种网络操作系统的互连和互通的工业标准 [1]。该系统建构于 TCP / IP 之上, 从而拓宽了该系统的应用范围, 但是该协议难以适应多媒体通信, 特别是连续的媒体流通信的要求。H.323 协议终端提供了实时的双向音频、视频和数据通信功能, 并规定了两类多播形式: 一是基于多点控制单元 (MCU) 的紧耦合方式, 其设计的复杂性, 开发难度较大、成本高; 二是基于 IP Multicast [2] 的可扩展的松耦合方式, 它可将异地分布且动态变化的成员组织在主机组中, 实现分布协同管理的功能。H.323 协议集中的 H.225 协议中使用实时传输和实时控制协议 (RTP / RTCP [4]) 描述了媒体 (音频和视频) 流的封装、媒体流的同步、控制流封装及控制信息的格式, 进行媒体流的打包与同步。它提供了端到端的实时媒体 (如交互式视频和音频) 传输协议。由于 RTP 本身并不具有独立传输数据的能力, 它必须与低层网络协议结合才能完成数据传输服务, 由于视频和音频的前后时间的相关性不高, 而数据的实时性要高于其可靠性, 所以在 UDP 上利用 RTP 协议传输视频和音频信号时使时延最小。

#### 2. 多播通信机制及 windows 9x 中对其支持

多播 (IP Mulcast), 是指在互联网上对一组 IP 站点进行数据传送, 这一组 IP 站点是动态形成的, 每一站点都可以动态地加入或退出这个组。多点播送的地址是 D 类地址, 它是动态分配和恢复的瞬态地址。当某个 IP 站点向互联网中的多个 IP 站点发送同一数据时, 多播可以减少不必要的重叠发送, 与使用多次点对点的单播及广播相



比,减轻了网络负载,提高了网络带宽的使用率及数据传输的实时性。其参与通信的各主机不论是源站还是目的站均使用同一程序,无客户机和服务器之分,从而具有对等性。

WINSOCK 2 系统中提供了对 IP 多播的支持,它是一个与协议无关的、具有紧急网络传输能力的网络接口 [8]。WINSOCK 多播模型中定义了数据平面和控制平面的概念。控制平面决定一个多播组建立通信的方式,数据平面决定通信成员间数据传输的方式。两平面中均采用“无”根模式,任一用户所发送的数据都将被传送到组中所有其他成员。数据报的发送者需指定一个多播地址作为目的地址,然后调用函数 `WSASendto()` 发送。数据的接收者可以通过分组源地址端口来区别发送源(多播地址不能作为源地址),因此用目的地址、目的端口、源地址、源端口组成的特定值来表示会话中某发送者的信息流。当接收端调用函数 `WSARecvFrom()` 从相应的套接字中读取数据时,应用程序首先应加入指定的 IP 多播组。它主要是通过调用一组函数 `WSASocket()`、`WSAIoctl()` 和 `WSAJoinleaf()` 来实现:

多播套接字的设置: `WSASocket(AF_INET, SOCK_DGRAM, IPPROTO_UDP, NULL, 0, WSA_FLAG_OVERLAPPED|WSA_FLAG_MULTICAST_C_LEAF|WSA_FLAG_MULTICAST_D_LEAF);`

`WSAIoctl()` 函数的 `SIO_MULTICAST_LOOP` 命令码来允许或禁多播通信的通信流量的内部回送。

通过对 IP 数据报头数据域的 TTL (生存时间) 的设置来限定 Multicast 流的范围,防止数据流从一网络到达另一网络,引起网络的数据阻塞。(TTL=1 将数据报限制在子网内) 用 WINSOCK2 中提供的函数 `WSAIoctl()` 的 `SIO_MULTICAST_SCOPE` 来实现 TTL 的设定。

`WSAJoinleaf()` 函数用来加入一个叶子节点到多播组。

### 音频和视频数据在网络中实时传输

视频、音频流在网络中实时传输是系统实现的基础。音频信号传输速率较低,可接受的位错率和组错率要求相对较低,但其实时性、连续性要求较高,最大可接受的延迟小于 250MS,采用分组交换方式传输音频的最大时延(相邻主机间的延迟)应小于 10MS,端到端的延迟应在 300MS 以下。ITU 标准规定,视频信号在交互视频中端到端的延迟应小于 150MS。视频和压缩视频信号都需要很高的传输率,占用较大的带宽,实时视频为 10-30 帧/S,因而压缩和解压缩时间不能超过 30-40MS。由此可知

影响视频和音频传输的关键除了带宽外还有一个十分重要的因素就是数据的处理速度,尤其是纯软件系统。

采用 G.723 协议为 5.3 / 6.4KBPS 的音频信号的编解码器,H.263 协议规定的低带宽网络下的视频信号的编解码器,它提供了时域和信噪比的可调性,能够与错误控制方案相结合进行压缩。在现有的低档奔腾 MMX233 机器上,每帧音频数据的压缩和解压缩时间仅需几毫秒,而视频信号每秒可以处理 11 帧以上,从而保证视频和音频信号实时传输的基本要求。

### 1. 实时音频数据通信

为使音频数据的采集满足电信部门话音质量的要求和网络中低码率的要求,系统中采用 8KHz 的采样频率、8 位分辨率的 PCM 编码方式。为了进一步提高速度,降低通信的数据量,减少网络负担和节省内存资源,一可将音频数据头数据格式结构进行简化去掉 WAVEHDR 中的两个保留项 `lpNext, reserved`,二采用消除静音和由语音采集设备引入的噪声的方法。静音在音频数据中占有较大的比例,它是通过声音的平均强度与某个阈值的比较来判断的。在发送端,采用静音压缩技术去掉静音,而对于接收端,为保证语音质量,即便源端没有讲话,也需不停使用一个滤波器来屏障静音和高频噪声数据。

音频应用的开发是通过低层音频函数,面向“流”式的音频数据,直接访问音频驱动程序,直接管理语音的采集和播放,才能保证其实时性。由于音频采集与发送、播放与接收是并行操作。应注意声音的录制和播放是一个连续的过程,而音频设备自己能够完成对声音的录制和缓冲区数据的不间断填充,程序必须不断地给音频设备发送空缓冲区的指令。程序设计中不仅为各种操作各分配了一个线程而且必须分配多个缓冲区,实验表明,程序只须设置两个缓冲区就可以实现声音的录制/播放和发送/接收并行执行。而缓冲区的大小和采样频率对声音的延迟有很大的影响。若帧缓冲区设置较大(4K 字节以上),通信时语音有较好的连续性和较大的延迟,反之若较小(512K 字节)则有较小的延迟,但连续性差且有丢包现象;将帧缓冲区设置为 1K 字节,其延迟控制在 150 毫秒,通信时语音有较好的连续性。

### 2. 实时视频数据通信

视频数据通信与音频数据通信的过程相似,在程序的控制下视频帧经数字化采样、压缩、网络传输到接收方后被解压缩,以及显示再现。每个站点都有两个发送缓冲区和多个视频接收缓冲区形成一个接收队列,多个视频接



收缓冲区并行工作,按照不同的视频源同时播放多个站点的视频图像。

针对视频传输,需要定义存放一帧压缩的图像帧数据缓冲区。当每一帧的图像由于内容不同而使其数据量大小不等时,则需将其分成等长的数据包进行传输。在接收端帧位元数据必须在正确的格式下才能正确回放,但网络传输过程中数据包的丢失或延迟破坏了数据格式的完整性,使还原时发生帧数据格式错误。所以在接收方需要检测数据包是否全部被接收,若出现丢包现象,则丢弃该帧。

### 3. 多媒体数据的实时传送

RTP 协议对视频和音频数据进行封装,使当某些包丢失的情况下,仍能根据包的序号对其他包进行重排序并进行解码。并用 RTP 头部时间戳域使视频和音频信号在解码时的同步。视频流常用频率为 65.536HZ 时间戳,而音频流则按采样速率打上时间戳。若后续的多个 RTP 组块属于同一个应用数据单元(即一个视频帧),那么它们就可以使用同一个时间戳,时间戳是接收方用来维持实时或与其他多媒体数据流的同步性。对接收方的本地时间和发送方 RTP 时间戳是分别计算时间间隔的,所以无需把接收方和发送方的时钟校准一致。

在此音/视频信号、呼叫信令和控制信号不会在同一条信道传输,也不使用相同的消息结构,即用于媒体数据传输的 RTP 协议和用于数据控制传输的 RTCP 协议分别使用不同的端口号。每种媒体向不同的传输地址分别发送视频和音频信号,保证了各种媒体信号有各自的帧序列号和服务质量。

### 用户信息的管理

在分布式多媒体远程监控系统中,监控人员的加入/退出、当前网络活动的状况、监控人员所在站点的信息以及各种媒体流都要进行用户管理。虽然 IGMP 协议的声明可用于控制 IP 多播组成员的加入和退出,但 IGMP 发送的信息中只包含 IP 多播的群组地址,而 IGMP 协议在 IP 协议层,一般程序不去处理它,因此要想建立正常的通信过程,必须在 IP 多播成员之间交换一些必要的信息。该系统基于 RTCP 协议对各种媒体流分布式地进行监控人员信息的传送和接收管理。监控人员必须注册所约定的 Multicast 地址,并在所有的用户需要系统初启时,向约定的 Multicast 地址自动发送一个 SDES 类型的 RTCP 包,将本端点信息通报给所有参与通信的站点, SDES 包中包含通信人姓名、IP 地址、编码标志、端口号和类名等与

站点相关的信息。所有站点需周期性(实验证明,周期  $T=10S$  时较为合理)地向约定的 Multicast 地址发送一个 SDES 类型的 RTCP 包并当有媒体状态变化时也需立即发送。一方面通报自己仍在通信过程中,另一方面把参与 IP 多播的人员信息及时告知新加入的站点,如果超过  $2T$  时间内还未收到某个站点的 SDES 包,则作超时失效处理,如果以后又收到该站信息,则重置相应的记录有效。当参与的人员较多时,  $T$  应增加一随机延时,以避免阻塞发生,影响媒体的通信质量。而退出以前以“BYE”包的形式通报给系统,各站点收到 BYE 包后置相应站点的退出标记。

### 监控信息处理

系统中对分布于不同区域的各个监控前端机把当前的运行状态信息向监控中心进行传输,监控中心根据前端的状况发送控制信息。监控人员可随时启动摄像设备和音频设备对现场进一步的监听、监视。对多个终端设备的控制和状态的查询,监控中心层的控制人员对设备有控制权(起主持人的角色),但必要时可以变更,把控制权授予协同监控人员。网络通信中为确保控制和状态信息数据的可靠传送,在系统采用 TCP 协议。并且为了保证监控信息处理中实时性的要求,必须设置超时机制使系统拥有可预测的系统响应时间。

### 结束语

本文介绍了基于 Internet 的多点多媒体实时监控系统的实现方法,并对其中的关键技术进行了一些探讨,由于采用纯软件的方式完成,计算机 CPU 性能和内存的容量对视频和音频的编解码处理有着重要的影响。在局域网中音频传送点不多的情况下,不必进行编/解码处理,也能有较理想的效果。对由网络分组的延迟抖动和丢失所造成的 QOS 的问题,可望通过基于 RTP 的 QOS 控制策略进一步的解决。■

### 参考文献

- 1 周明天 TCP / IP 网络编程原理与络技术 清华大学出版社
- 2 S.E.Deering. Host Extentions for IP Multicasting. REF1112 1998.8
- 3 H Schulzrinne , S Casner , R Frederick, V Jacobson(1996). RTP: A Transport Protocol for Real-time Applications. Resquest for comment1889