

# 基于 GPRS 网络的无线 IP 语音通信系统

程 伟, 龙昭华, 蒋贵全

(重庆邮电大学移动通信技术重点实验室, 重庆 400065)

**摘 要:** 提出并实现一种基于通用分组无线业务网络的无线 IP 语音通信系统。该系统结合 S3C2410 处理器的低成本、高性能与嵌入式 Linux 软件系统高可靠性等优点, 采用独特的双缓冲四线程语音数据处理方法和延迟可控技术, 使实时语音播放更加平滑, 并将总体时延控制在可接受范围内。性能分析结果表明, 模块化的系统设计方法使该系统只需更换通信模块即可接入 3G 网络。

**关键词:** IP 语音; G.729A 算法; 双缓冲; 实时传输; 通用分组无线业务

## Wireless IP Voice Communication System Based on GPRS Network

CHENG Wei, LONG Zhao-hua, JIANG Gui-quan

(Key Laboratory of Mobile Communications, Chongqing University of Post & Telecommunications, Chongqing 400065, China)

**【Abstract】** This paper presents and realizes an IP voice system, which combines the S3C2410 processor's low-cost, high performance features with embedded Linux software system's high reliability advantages. It uses a unique way called four-threaded double-buffered voice data processing and delay-control playback technology to make real-time voice playback smoother, and time delay is controlled within an acceptable range. Performance analysis shows that the modular system design makes the system easier to access 3G networks by only replacing the communication module.

**【Key words】** IP voice; G.729A algorithm; double-buffer; real-time transmission; General Packet Radio Service(GPRS)

DOI: 10.3969/j.issn.1000-3428.2011.14.026

### 1 概述

通用分组无线业务(General Packet Radio Service, GPRS)是在现有 GSM 系统上发展出来的一种新承载业务, 具有数据传输不受地域限制、建设成本低、数据传输速率高以及永久在线等特点<sup>[1]</sup>。VoIP 是一种在 IP 网络上使用 IP 协议以数据包的方式传输语音的技术<sup>[2]</sup>。将这 2 种技术相结合, 利用现有 GPRS 网络传输 VoIP 语音数据, 相比传统的移动语音通信服务和基于 WLAN 与有线相结合的 VoIP 业务, 可大幅度降低通信费用, 扩大覆盖范围。本文通过借鉴已有研究成果, 使用成熟技术并进行了应用创新, 设计一种基于 GPRS 网络的无线 IP 语音通信系统。

### 2 国内外研究现状

国内一些研究团队对此领域开展了一些理论性研究工作。研究人员通过理论分析和实验证明了在 GPRS 网络窄带宽下进行实时语音传输是可行的<sup>[3]</sup>。另外, 针对无线网络和 Internet 的业务特点, 分析了在信道带宽有限、传输误码率高的无线网络中传输基于 IP 的实时业务所面临的问题, 提出了一种实时交互式语音业务的新方案<sup>[4]</sup>。比较而言, 国外在此领域研究起步较早, 部分成果已应用于移动通信和网络通信。例如, 设计人员提出了一种结合 GPRS 和 WLAN 技术的全网覆盖的语音数据传输技术<sup>[5]</sup>。

### 3 无线 IP 语音通信系统应用模型

本系统以应用为主要目的, 主要传输实时音频数据。图 1 为系统的应用模型。在图 1 中简化了 GPRS 内部网络结构。由图可知, 系统主要分为 3 个部分: 手持终端, 网络传输媒介和媒体网关。手持终端类似于普通手机, 可以采集音频数据并处理后发送。网络传输媒介主要指 GPRS 网络, 包括 SIM300 无线传输模块。媒体网关也可称为媒体服务器, 它使用 SIP 协议, 主要提供注册、位置服务等, 同时具备媒

体穿越功能。但在点对点通信模式下, 2 个终端可直接通信, 无需媒体网关的支持。

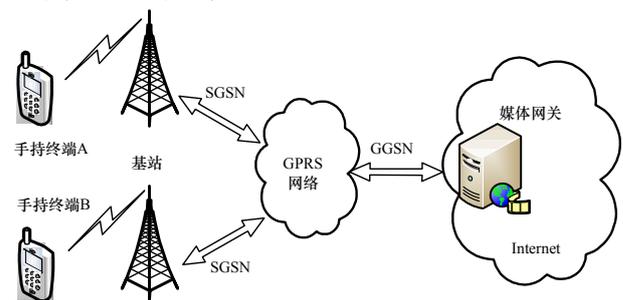


图 1 系统应用模型

### 4 系统硬件设计

在本系统中, 手持终端是需要进行硬件设计的部分。手持终端硬件平台以 S3C2410 处理器为核心。它使用 ARM 公司的 ARM920T 处理器核, 采用 0.18 μm 制造工艺, 拥有独立的 16 KB 指令 Cache 和 16 KB 数据 Cache, 同时具备非常丰富的外设接口, 可方便进行系统扩展。它具有最高 203 MHz 主频, 特别适合低成本、低功耗的应用领域。本系统主要处理音频数据, 计算量不大, 因此, 采用此款处理器可完全满足需要, 同时也兼顾了成本控制的要求。

SIM300 是一款体积小, 采用板对板连接器的三频/四频 GSM/GPRS 模块, 可提供人机界面和键盘/LCD 接口。它的最大优点是内嵌 TCP/IP 协议栈, 结合 AT 命令可方便进行网络产品实用开发, 目前已被广泛用于手持设备、监测设备

**基金项目:** 国家质检总局公益性行业科研专项经费基金资助项目(10-226); 重庆市科技攻关计划基金资助项目(CSTC, 2007AC2053)

**作者简介:** 程 伟(1984—), 男, 硕士, 主研方向: GPRS 网络, 嵌入式系统; 龙昭华, 教授; 蒋贵全, 副教授

**收稿日期:** 2010-12-01 E-mail: chw312@163.com

等需要进行大范围数据传输的领域。S3C2410 处理器和 SIM300 模块之间用 RS232 接口连接。当系统需要升级时, 例如需要由 GPRS 网络升级至 3G 网络, 仅仅更换 SIM300 通信模块为相应某种 3G 模块即可, 无需改变接口。图 2 显示了本系统手持终端硬件系统结构。

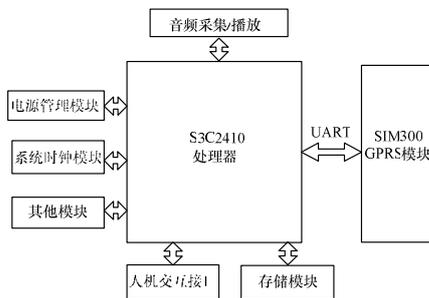


图2 手持终端硬件系统结构

由图 2 可知, 除 S3C2410 和 SIM300 核心组成部分外, 手持终端硬件还包括人机交互接口、存储系统等重要模块。其中人机交互接口主要指用于接收终端命令输入, 包括微型键盘/LCD、触摸屏等设备。为提高可操作性, 本系统设计手持终端时采用 9 键键盘加四线电阻式触摸屏相结合的方式。存储系统主要指 SDRAM 和 FLASH 设备。音频采集/播放设备主要包括 MIC 和 PHONE。上述各模块是一个语音终端基本组成部分, 可在此基础上进行更多版本的扩展性开发。

## 5 系统软件设计

软件是整个系统的灵魂。在本系统中, 手持终端和媒体网关需要进行软件开发。为加快软件研发进度, 部分软件采取移植现有开源软件的方式完成, 其他核心软件则需要自主开发。

### 5.1 软件平台

由于 Linux 具有强大的网络功能、丰富的软件资源和良好的移植性等优点, 因此本系统采用 Linux 作为手持终端与媒体网关的操作系统, 从而构成了 ARM-Linux 系统结构。同样, 宿主机也同样采用 Linux 系统。媒体网关中运行的主要软件包括 2 个部分: SIP 服务器 MINI SIPSERVER 和穿越服务器 STUN。前者主要用于提供 SIP 协议支持, 而后者主要实现媒体穿越功能, 这样即使 2 个终端位于各自的局域网内部, 仍然可实现通信。手持终端软件设计是本系统的重点, 下面分别介绍它的各个组成部分。

### 5.2 软件总体架构

手持终端软件是由多个软件模块构成的高效协作的软件系统。软件架构主要包括 5 个部分, 图 3 方框部分显示了各个模块的组成关系。

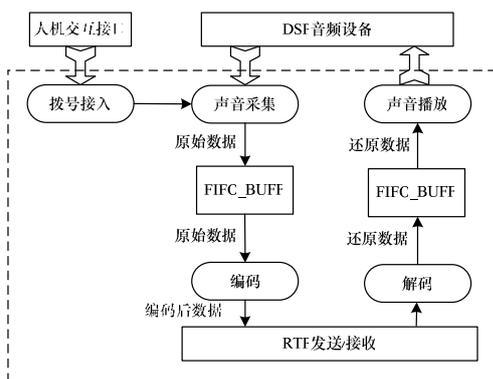


图3 手持终端软件总体架构

由图 3 可知, 软件系统 5 个模块分别是: 拨号接入与注册, 声音采集/播放, 音频数据缓冲, 音频编/解码, 数据无线实时传输。上述 5 个部分并不是每个类似终端软件系统都需要具备的, 在有些系统设计开发过程中, 可以音频数据缓冲部分。这样做, 一方面可以减少开发难度, 加快研发进度; 另一方面减少因中间过程处理带来的数据时延。但这时系统也丧失了部分系统性能, 如数据平滑性等, 降低了声音质量。

### 5.3 拨号接入与注册

拨号接入是在系统启动阶段通过向 GPRS 模块发送一系列 AT 命令使终端接入 Internet 的过程。GPRS 采用 PPP 协议作为上层协议。PPP 协议被设计成一种在同等单元之间进行数据传输的简单链路层协议。首先须配置 Linux 内核支持 PPP 协议, 然后采用编写 SHELL 脚本程序进行拨号连接。本系统使用文献[8]中类似 SHELL 脚本连接 Internet 网络。当 GPRS 模块与网络拨通后, 手持终端软件系统使用 SIP 协议向媒体网关注册自己, 以便其他终端可以发现自己并与自己通信。当服务器返回 200 OK 后, 表明手持终端已经介入到互联网。

### 5.4 声音采集/播放

在 2.6 内核中, Linux 默认采用 ALSA 为系统提供音频接口。ALSA 是一种开放源代码的音频驱动程序集, 它简化了应用程序的编写难度, 增加了灵活性。但是实验证明, ALSA 与 S3C2410 内部寄存器定义之间存在某些冲突。因此, 手持终端系统修改了内核配置, 重新启用 OSS 作为音频接口。OSS 是用于在 Unix/Linux 环境下访问声卡及其他音频设备的驱动程序。它直接将音频设备看为一个文件, 应用程序通过读写这个文件实现音频的采集和播放, 相比 ALSA 具有简易性和可控性的优点。

本系统音频采集/播放相关代码如下:

```
//打开音频设备
fd=open("/dev/dsp", O_WRONLY|O_RDONLY, 0);
//设置采样格式
int format = AFMT_S16_LE;
ioctl(fd, SNDCTL_DSP_SETFMT, &format);
//设置通道数
int channels = 2;//立体声
ioctl(fd, SNDCTL_DSP_CHANNELS, &channels);
//设置采样率
int speed = 44 100;//采样频率为 44.1 KHz
ioctl(fd, SNDCTL_DSP_SPEED, &speed);
//录音
read(fd, buffer, count); //buffer 为保存音频数据的缓冲区
//放音
write(fd, buffer, count);
//关闭音频设备
close(fd);
```

### 5.5 音频数据缓冲

手持终端软件系统设计了一种独特的双缓冲四线程语音数据处理方法。它包括 2 个基于 FIFO 的双端循环缓冲区 FIFO\_BUFF 和 4 个处理线程。缓冲区每段内存设为 1 024 个字节, 与每次采集的音频数据大小一致, 可设置一个缓冲区内内存段个数。FIFO\_BUFF 具有头指针 FIFO\_HEAD、尾指针 FIFO\_TAIL, 指向数据写入位置和读出位置。手持终端软件系统中共使用 2 个 FIFO\_BUFF, 图 4 为这个双缓冲区操作示意图。

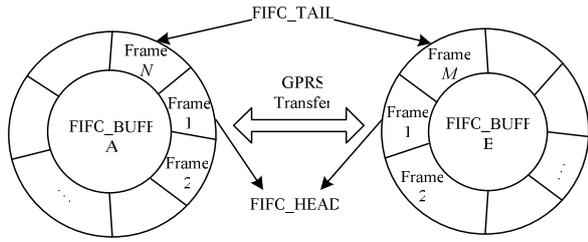


图4 双缓冲区操作示意图

手持终端软件系统共有4个共2对线程操作上述2个缓冲区。CAP\_THREAD负责采集音频数据,待完成1024个字节采集工作后将其写入CAP\_SEND\_BUFF;SEND\_THREAD负责每次取出1024个字节数据,然后完成编码、加密处理后,使用网络发送给对端;RECV\_THREAD负责从网络接收数据,然后执行解密、解码动作,处理完后将数据写入RECV\_PLAY\_BUFF;每隔一定时间,PLAY\_THREAD就会从RECV\_PLAY\_BUFF中取出一个数据块中数据,然后写入音频设备,实现音频播放功能。由于对单个缓冲区的写入和读出操作都只有一个线程,因此无需加锁,从而提高了存取效率。

双缓冲四线程语音数据处理方法虽然需增加部分内部存储空间,但它将接收和发送数据分开存储,利于后期处理与控制,同时可使整个处理过程显得极为简洁和高效。

### 5.6 音频数据编解码处理

手持终端系统采用双声道、44.1 KHz采样频率和16位采样格式,按照式(1),每秒钟数据量达到1411.2 Kb/s,如此巨大的数据量在极短时间内可使缓冲区溢出。同时窄带GPRS网络短时间内无法发送这些数据,因此必须将原始音频数据做压缩处理。

$$\text{数据量} = \text{采样率值} \times \text{采样大小值} \times \text{声道数 bps} \quad (1)$$

手持终端软件系统采用G.729A作为音频编解码器。G.729A是基于共轭结构的代数码激励线性预测(CS-ACELP)算法的8 kbit/s的低速率G.729语音编解码器的简化版本。它大幅度降低了算法复杂度,其基本编码原理如图5所示。

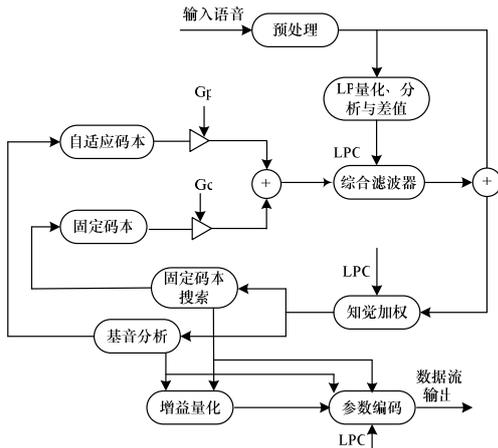


图5 G.729A 编码器原理示意图

G.729A需要输入16位格式8KHz的线性PCM语音数据,处理时延为15ms。在本系统中,SEND\_THREAD线程从FIFO\_BUF取出原始数据,然后调用G.729A编码器开始编码。ITU-T提供了2种浮点和定点和2种G.729A的C语言源码。手持终端软件系统移植了定点G.729A C代码,并从C编译器和内联函数2种方法对其进行部分优化,以提高其效率<sup>[6]</sup>。

### 5.7 无线数据实时传输

语音属于实时数据,较大的延迟会严重地影响通话质量,

因此,必须快速将处理后的音频数据发送到网络。不间断地发送数据又会产生拥塞,所以必须在发送2个数据包之间设置一个适当的发送间隔,因此本文采用一种延迟可控技术来有效控制处理时延,其核心思想是根据网络综合状况,设定合理的发送语音数据和播放接收数据的时间间隔,从而控制语音整体延迟时间。在手持终端软件系统中,采用的G.729A编解码算法的编码时延为15ms,经过反复实验,发送间隔取90ms时实际通信效果最好。另外,修改RTP头部,将发送语音IP数据包的固定附加长度从原来的50Byte减少到39Byte<sup>[3]</sup>。系统每个发送间隔内使用RTP发送一个数据包,其所需实际带宽可用式(2)计算获得。

$$BW = 8 \times (\text{Rate} + 39) / \text{Interval} \quad (2)$$

其中,BW为所占带宽(Kb/s);Rate为语音编码速率即每10ms编码输出的字节数(10ms编码时间不包括对下一帧信号进行前视线性预测编码所需的5ms时间);Interval为发送语音包的周期,应取总编码时延的整数倍。

## 6 系统测试与性能分析

为验证上述设计方案的实际性能,本文开发了手持终端的原型系统,图6为原型系统的实物图。同时,配合部署在公网上的媒体网关使用中国移动GPRS网络进行测试。测试方法为对2个原型系统进行点对点语音通信。每次测试进行10次通话,共进行5次测试,取平均数据,结果见表1。



图6 手持终端原型系统实物照片

表1 原型系统性能测试结果对照表

次数	使用带宽/(Kb·s <sup>-1</sup> )	间隔/ms	平均时延/s	音频 MOS 级别
1	20.6	30	2.7	3
2	13.3	60	3.1	4
3	11.5	90	3.4	5
4	10.7	120	3.9	4
5	10.1	150	4.5	3

在表1中,音频MOS级别指ITU-TP.800建议中规定的5种采用平均意见得分法(MOS),它采用5级评分标准对语音质量进行评定,分别为5(优)、4(良)、3(中)、2(差)和1(劣)。测试结果表明,随着间隔的不断增大,语音通信系统占用带宽不断减少,但平均时延增加,在测试网络环境下,当间隔取90ms时,音频质量最好。分析发现,发送间隔延迟和网络延迟是其延迟较大的主要原因。由于网络延迟无法控制,因此可通过调节发送间隔控制总体时延。因此,当使用系统时,适当减少发送间隔,可减少延迟时间。但是部分GPRS网络无法提供足够的上行带宽。解决此类矛盾的一个可行方法是根据系统使用时的网络情况,动态选择一个较为合适的发送间隔,使系统获得最好通信质量。

## 7 结束语

采用GPRS网络传输语音是本系统的一个特点,它利用GPRS网络覆盖范围大、成本低、可随时接入的优点。通过采用与传统类似语音系统不同的语音数据处理方法,提高了语音数据的通信质量和系统的整体性能。

(下转第87页)