GSM 短信传送文件的方法的研究与实现

万 求，李 森，张 建，罗 毅
（中国科学院 合肥智能机械研究所，合肥 230031；
中国科学院 研究生院，北京 100039）
E-mail：qiu_wan@163.com

摘 要：在对 GSM 移动通讯系统的短信息(SMS)业务通信中 AT 指令的介绍，以及短信格式、编码与解码等分析的基础上，介绍了一种使用 GSM 短信传送任意格式文件的方法，将文件压缩并分割封装成适合短信传送的数据包，通过 GSM 短信发送给接收方，接收方接收之后将数据包还原、解压得到原始文件，从而实现了通过 GSM 网络在两台计算机之间传送文件。

关键词：GSM 短信；AT 指令；数据包

文章编号：1002-8331(2006)32-0122-04 **文献标识码**：A **中图分类号**：TP393

Study and Realization of Method to Transfer File by GSM Short Messages

WAN Qiu，LI Miao，ZHANG Jian，LUO Yi
（Hefei Institute of Intelligent Machines，Chinese Academy of Sciences，Anhui Hefei 230031，China；
Graduate School of the Chinese Academy of Sciences，Beijing 100039，China）

Abstract：On the base of introduction of AT Command in Short Message Service(SMS) of GSM，and analysis of short messages elements，encoding and decoding，this paper introduces a method to transfer file by GSM.Compress the file，then split it and package it to data packages according to the short message's specification.Transfer the data packages to the receiver by GSM.After receive all data packages，receiver combine data packages to one，then decompress it to get the original file.This method realize file's transfer between two computers by GSM which one or both of them can not access to Internet.

Key words：GSM Short Messages；AT Command；data package

1 引言

全球移动通讯系统 GSM (Global System for Mobile Communications)是由欧洲电信标准化协会(ETSI)开发的数字移动电话网络标准，其目的是让全球各地共同使用一个移动电话网络标准，让用户使用一部手机就能行遍全球；也是我国目前覆盖范围最广、功能最强、用户最多的移动通信系统，GSM 中的短信息业务 SMS(定义于 ETSI 制定的标准 GSM 03.40)提供的短信服务具有收费低廉、随时随地获取信息的便利。但现在绝大多数用户只是单纯的利用 GSM 短信发送简单且极短的文字文本信息。即使是现在一些基于 GSM 的数据采集监控系统，其本质仍然是利用 GSM 传送数字或文字信息。

2 问题的描述

在实验室课题项目示范应用的电脑农业示范区，专家系统的使用者大多数在农村。专家系统涉及知识库实时、快速更新的问题，但相当多的这部分地区，不能使用 Internet 获取资料，以邮寄等其他方式又存在时间性的问题；相反，移动通信网络已经覆盖到。所以，本文提出了一种利用 GSM 短信传送任意格式文件的方法，使不能连接到 Internet 网络的计算机之间通过 GSM 网络进行文件传送成为可能，从而也使知识库的实时、快速更新成为可能。

3 问题的解决

3.1 GSM 短信接收与发送的实现

3.1.1 计算机接入到 GSM 网络的方式

本文的最终目的是实现两台计算机之间通过 GSM 网络传送文件，首先要选择计算机连接到 GSM 网络的方式。目前接入方式主要有三种[3]：

(1)专线接入运营商短信网关。因为运营商对于设备和业务量有一定的要求，开展的业务须经过运营商的综合评测，如果用户涉及多个运营商网络(如中国移动、中国联通)，则需要分别接入，另外设备等价格昂贵，因此该方法适用于大型企业用户。

(2)虚拟运营商接入。用户只作为其中一个客户，利用虚拟运营商提供的客户端软件或二次开发接口发送短消息，这样的好处是设备投入比较少。但是依赖于虚拟运营商，业务内容和

基金项目：国家 863 高技术研究发展计划资助项目(2003AA118040)。

作者简介：万求(1978-)，男，硕士研究生，研究方向为模式识别与人工智能；李森(1955-)，女，研究员，研究方向为人工智能中的知识与信息处理等。

服务的质量将会收到其限制。另外,与第(1)种方式相同,也需要 Internet 网络的支持,依赖于互联网络。

(3)通过 GSM MODEM 或手机接入。所需设备是 GSM MODEM 或者支持与计算机通信的手机,该方式不受运营商的限制,也不依赖于 Internet 网络,缺点是短信的接收发送速度因硬件设备本身条件限制,但可以通过增加设备的方法解决。

因为本文的最终应用是为了解决无法连接到 Internet 网络的计算机之间的文件传送问题,所以只能选择第(3)种方式。实验设备选择的是 WAVECOM 的 GSM MODEM,支持 AT 指令。

3.1.2 GSM 短信数据格式分析

3.1.2.1 短信发送模式与相关 AT 指令

短信的发送和接收目前有两种模式[3,4]:基于 AT 指令(AT-Command)的文本模式(Text Mode)和基于 AT 指令的 PDU 模式(Protocol Description Unit)。AT 指令最初仅用于 Modem 的操作,后来几个大的手机生产商诺基亚、摩托罗拉、西门子等共同为 GSM 研制了一套 AT 指令集,用于 GSM 移动通信设备的控制,其中涉及到 GSM 短信的指令主要有[5]:

AT+CSMS:选择所支持的短信息服务。

AT+CNMA:新信息确认应答。

AT+CPMS:优先信息存储。这个命令定义用来读写信息的存储区域。

AT+CMGF:优先信息格式。执行格式有文本模式和 PDU 模式。

AT+CNMI:新信息到达指示。这个命令选择如何从网络上接收短信息。

AT+CMGR:读短信息。信息从+CPMS 命令设定的存储区域读取。

AT+CMGL:列出存储的短信息。

AT+CMGS:发送短信息。

AT+CMGW:写短信息并存储。

AT+CMSS:从存储器中发送信息。

AT+CSMP:设置文本模式的参数。

AT+CMGD:删除短信息。删除一条或多条短信息。

AT+CSCA:设置短信服务中心地址。

文本模式和 PDU 模式实质上是指短信息数据的编码格式,编码得到的字符串表面上由"0~9"、"A~F"这些字符组成。PDU 编码得到的字符串不仅包含可显示的短信息本身,还包含很多其它信息,如短信服务中心号码、目标号码、回复号码、编码方式和服务中心时间戳等。PDU 模式可采用的编码方式有三种:7-bit 编码、8-bit 编码和 UCS-2 编码。7-bit 编码只能发送 ASCII 字符,8-bit 编码通常用于传送数据信息,UCS-2 编码用于发送 Unicode 字符,例如中文。文本模式实际上也是由 PDU 模式显现的位串编码,主要区别在于文本模式仅支持 7-bit 编码,只能发送纯英文和数字信息。所以选用 PDU 模式。

3.1.2.2 PDU 字符串格式分析

发送方和接收方的 PDU 字符串不完全相同。下面以实例来分析 PDU 编码的格式。短信息为"测试 PDU 编码",发送号码为 13866709472,目标号码也为 13866709472,短信中心号码为 13800551500。

发送方 PDU 字符串为:

SCA	PDU Type	MR	DA	PID	DCS	VP	UDL	UD

0891683108501505F011000D91683168769074F20008000E6D4B8BD50050004400557F167801

接收方 PDU 字符串为:

SCA	PDU Type	OA	PID	DCS	SCTS	UDL	UD

0891683108501505F0040BA13168769074F2000850908041042180 0E6D4B8BD50050004400557F167801

(1)发送方 PDU 字符串分析

短信中心(SCA):0891683108501505F0。其中:

08 h(h 表示十六进制数,b 表示二进制数,下同)短信中心地址的长度,指 91 68 31 08 50 15 05 F0 的八位位组的数目;

91 h TON/NPI 遵守 International/E.164 标准,指电话号码,在号码前加'+'号,国际电话号码的表示形式;

683108501505F0 短信中心号码,实际表示 8613800551500。当号码位数为奇数时,会补一个字符 F。目标号码和回复号码的情况与此相同。

备注:如果短信中心地址的长度为 00 h,且后面没有跟号码,则使用 SIM 卡中存储的短信地址号码。

协议数据单元类型(PDU Type):11 h。是十六进制数表示的一个八位位组,11 h=0001 0001 b。包含了应答路径、状态报告、有效期格式等信息。

信息参考(MR):00 h。00 表示参考本身号码。短信发送时在短信中心的序号参考在 0-255 之间每成功发送一条序号累加 1,超过 255 时转为 0。

目标号码(DA):0D91683168769074F2。其中:

0D h:目标号码的实际长度,不包括 91,注意不同与短信中心号码长度的计算;

91 h:同短信中心号码;

683168769074F2:目标号码,实际表示 8613866709472,共 13 位,即长度是十六进制数 0D。

协议标识(PID):00 h。00 h=0000 0000 b。对于标准情况下的短信发送,设为 00。

数据编码方案(DCS):08 h。08 h=0000 1000 b,主要是第 2 位和第 3 位上的数据,00 h 表示 7-bit 编码,01 h 表示 8-bit 编码,10 h 表示 UCS-2 编码,11 h 预留。

信息有效期(VP):00 h。00 h 表示 5 min。VP 的范围是 00 h~FF h。

用户数据长度(UDL):0E h。是紧随其后的以八位位组的个数,而不是字符的个数。

用户数据(UD):6D4B8BD50050004400557F167801。由短信内容经过编码得到的字符串,采用压缩的 BCD 码表示,每两个字符组成一个八位位组。此字符串共 28 个字符,14 个八位位组,所以 UDL 是 14,即 0E h。

(2)接收方 PDU 字符串分析

短信中心(SCA):0891683108501505F0,分析同发送方。

协议数据单元类型(PDU Type):04 h。分析同发送方。

回复号码(OA):0BA13168769074F2。其中:

0B h:目标号码的实际长度,不包括 A1;

A1 h:TON/NPI 遵守 International/E.164 标准,表示是国内的电话号码;

3168769074F2:回复号码,表示 13866709472,共 11 位,即长度是十六进制数 0B。

协议标识(PID):00 h。分析同发送方。

数据编码方案(DCS):08 h。分析同发送方。

服务中心时间戳(SCTS):50908041042180。共7个八位位组,前6个表示时间,最后一个表示时区。表示时间是05年9月8日14时40分12秒,+8时区。

用户数据长度(UDL):0E h。分析同发送方。

用户数据(UD):6D4B8BD50050004400557F167801。分析同发送方。

3.1.3 程序实现

计算机通过串口与GSM MODEM通讯。程序实现上涉及两个方面,一是计算机与串口的通信,二是短信发送与接收的处理。采用两个工作线程,串口事件监视线程用于监听串口的通信事件[1,2],主要监听EV_RXCHAR通信事件,根据EV_RXCHAR事件从串口读取的内容进行相应的动作;短信处理线程负责短信编码、解码,向串口发送PDU字符串,从串口读取PDU字符串,发送队列、接收队列的管理。程序框图见图1和图2。

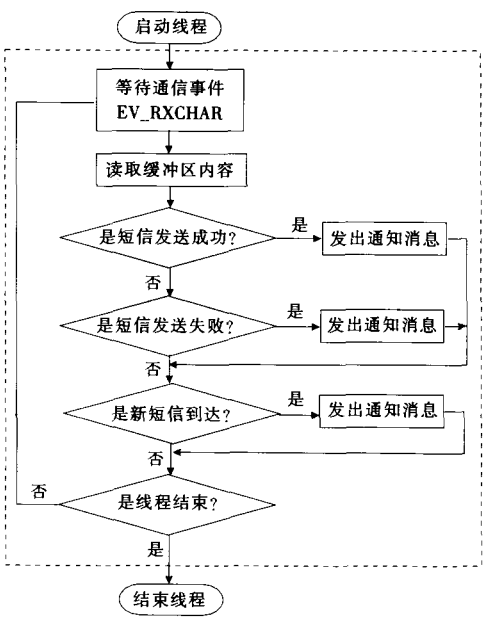

图1 串口事件监视线程流程图

对新短信到达、短信的发送报告,由通信事件EV_RXCHAR触发后面的动作。采用事件通知方法相比连续不断检测端口方法可节省CPU时间。

用户发送的短信首先加入到发送队列中,线程从发送队列中取出待发短信发送;接收到的短信也首先放入接收队列中。采用队列缓冲池方法处理发送和接收到的短信,可以提高程序运行效率。

3.2 GSM短信传送文件的实现

在上一章节的基础上,要实现GSM短信传送文件,首先需要将文件分割并封装成适合GSM短信数据长度的数据包,然后再制定发送方与接收方的传送策略,及具体实现步骤。

3.2.1 数据包的构成

在本文中,任意格式的文件指在计算机上所有以二进制形式存储的任何文档、图片、声音、影像等文件。每条GSM短信传送的数据长度有限,例如7-bit编码最多160个字节,8-bit编码最多140个字节。文件的字节数一般都远大于这个数值,所以需要首先将文件分割并封装成适合GSM短信传送大小的数

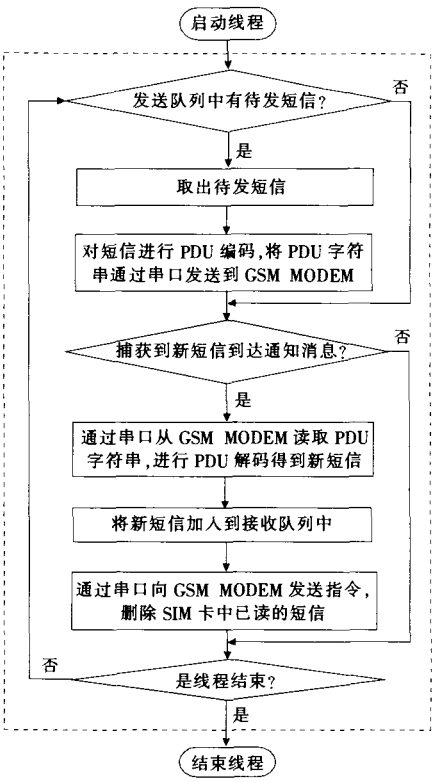

图2 短信处理线程流程图

据包。如3.1.2节所述,对于传送文件数据,GSM短信的编码方式选用8-bit编码,每条短信140个字节,数据包的格式定义为:

数据包序号	数据包总数	检验和	数据
4字节	4字节	1字节	131字节

各部分说明如下:

数据包序号 作为分割后每个数据包次序的标识,在数据还原时按照次序依次读取数据。序号从1开始,最大9 999。

数据包总数 文件分割为数据包的总数量,接收方验证数据包的总数,判断是否有数据包丢失

数据包总数=文件总字节数/数据的长度

检验和 采用bbc异或检验(block check character)。GSM短信通信质量较为可靠,所以采用这种既简单又相当准确的数据检验方法。在发送方,从文件中读取数据后,计算一次数据的检验值。在接收方,接收到数据后,也计算一次检验值与接收到的检验值比较。

数据 指有效数据,读取的是任意格式文件的二进制值。顺序依次读取文件的二进制值,每次131个字节,直到读至文件结束。

3.2.2 传送文件的步骤

发送方和接收方传送一个文件主要分为6个步骤:

(1)数据包的生成。按照数据包定义的格式,在发送方,将文件分割并封装成数据包。为了减小文件的大小,在分割之间,先将文件压缩,使用zip压缩算法,在程序中加入了Hans Dietrich的zip算法[6]的压缩程序进行文件压缩。

(2)GSM通信网络的通信质量相对可靠和稳定,所以发送、接收双方采用"两次握手"建立"连接",发送方首先发送一条询问接收方是否在线的短信,同时在该短信中包括发送方准备发送的短信总数。接收方收到该短信后,记下将要接收的短

信总数,并回复一条准备就绪可以发送的短信给发送方。这样通过两条短信建立"连接"。

(3)在发送过程中,发送方按照数据包序号递增的次序发送封装好的数据包,首先发送序号为1的数据包。对GSM MODEM或GSM网络原因发送失败的数据包最多重新发送三次,如果还不能成功发送,将该数据包的序号通知给发送方用户处理,然后继续下一个数据包。原则上只有在前一数据包发送成功的基础上再发送下一个数据包。

接收方接收到短信后首先校验数据是否损坏,如果数据完好则将该序号保存到序号数组中,然后存储该短信到计算机上,否则丢弃该短信。在接收到最大数据包序号的短信后,根据数据包总数和序号数据中存储的序号,计算出丢失的数据包或者损坏的数据包序号,发给发送方要求重新发送这些序号的数据包。

(4)如果有需要重新发送的数据包,发送方重新发送这些序号的数据包。

(5)接收方在确认所有的数据包都已经正确接收后,发送一条短信通知发送方传送过程结束。

(6)在接收方,将接收到的短信按照数据包序号将数据包还原,使用zip压缩算法[6]的解压程序解压得到源文件。至此,通过GSM短信传送文件的过程结束。

4 结束语

本文提出的GSM短信传送文件的方法,在Windwos2000环境下,已通过Visual C++6.0开发出应用程序实现,可以应用于实际。该方法使短信服务不再局限于传递文本信息,文档、图像、声音、影像等任意格式的文件都可以通过GSM短信从发送方传送给接收方。下一步打算将该方法与专家系统知识库更新算法结合,从而应用于专家系统知识库的异地远程实时快速更新。(收稿日期:2006年2月)

参考文献:

[1] 李现勇.Visual C++串口通信技术与工程实践[M].北京:人民邮电出版社,2002.

[2] 龚建伟,熊光明.Visual C++/Turbo C串口通信编程实践[M].北京:电子工业出版社,2004.

[3] 宣彩平,王皓,邹国良.利用GSM无线模块发送短消息[J].计算机应用,2004,24(5):148-150.

[4] ESTI.Digital cellular telecommunications system(Phase 2+):Technical realization of the Short Message Service(SMS)Point-to-Point(PP)(GSM03.40)[EB/OL].[1996].http://www.etsi.org.

[5] ESTI.Digital cellular telecommunications system(Phase 2+):AT command set for GSM Mobile Equipment(ME)(GSM 07.07)[EB/OL].[1996].http://www.etsi.org.

[6] DIETRICH H.XZip and XUnzip-Add zip and/or unzip to your app with no extra lib or dll[EB/OL].[2003].http://www.codeproject.com/cpp/xzipunzip.asp.

(上接107页)

然后,对本系统支持的并发用户数以及系统响应时间进行测试。

为了模拟真实的用户环境,这里采用JAVA编写了多线程程序对服务端的并发支持能力进行了测试。测试的方法是每个线程模拟一个用户,间隔一定的时间读或写一次审计策略或查询一次审计日志,测试用的安全监控与审计系统服务端的配置是:P4 2.4 G CPU,1 G DDR内存,80 G硬盘的工控机。测试结果如表1所示。

表1 系统服务端并发支持能力测试结果

并发用户数	50	150	1 000	2 000
响应时间/ms	10~25	20~60	250~900	300~1 600

因此,大规模多用户网络安全监控与审计系统的特点如下:

(1)高集成度和强扩展性。能融入各类新型安全技术研究成果;

(2)大规模多用户交互。各对象间的交互简单、明了,且实现了交互信息的可视化,同时可支持大约1 000人并发访问;

(3)灵活的策略制定方式。可针对机器、操作员、事件、处理方式及日期和时段等对象组合制定监控策略;

(4)实时可靠的审计信息采集方式。利用嵌入操作系统内核的审计客户端实时采集操作信息。

多用户网络安全监控与审计系统已经在本科生中得到了使用,支持了大约150个学生同时进行安全审计文件、网络、打印机和拨号的访问和禁止策略的制定和日志查询实验,而且实际响应时间也与测试结果基本相符合。学生在完成本系统设计的实验后,获得了很好的效果。

5 结束语

本文研究和实现了一种大规模多用户的网络安全监控与审计系统,适用于计算机网络应用的各个领域,特别是大型企业、科研院校等对于信息安全性要求很高的单位内部,并支持大规模的教学实践。设计实现的系统所具备的各个功能及其性能均已在实践中得到了验证,并且取得了很好的效果。因此,本系统不论是在信息安全实验教学,还是在实际应用中都具有较好的推广前景。(收稿日期:2006年7月)

参考文献:

[1] LO E C,MARCHAND M.Security audit:a case study[information systems][C]//Electrical and Computer Engineering:2004 Canadian Conference on,2004,1:193-196.

[2] RAO K N.Security audit for embedded avionics systems[C]//Computer Security Applications Conference,1989 Fifth Annual,1989-12:78-84.

[3] 许霆,袁萌,史美林.网络监控审计系统的设计与实现[J].计算机工程与应用,2002,38(18):149-150.

[4] 刘海峰,卿斯汉,刘文清.安全操作系统的实时报警[J].计算机学报,2003(3).

[5] 潘军,王桂森,李祥和.Windows平台下安全审计技术的探讨与实现[J].微计算机应用,2005(3).

[6] 张英,王景新.网络安全基础[M].WILLAMS S,译.北京:中国电力出版社,2004-06.

[7] 刘海峰,卿斯汉,刘文清.安全操作系统的实时报警[J].计算机学报,2003(3).