

浪涌气象数据采集系统及其关键技术

伍光胜¹ 李建勇¹ 刘艳中¹ 雷卫延¹

摘要

针对无人值守大规模自动气象站的数据采集特点和要求,基于无线网络,设计了一种浪涌气象数据采集系统,并对其关键技术进行了深入讨论.首先,分析了系统数据传输的协议结构及存在的问题,设计了基于UDP协议的可靠应用协议,并就其工作过程及可靠性传输进行了详细的讨论,解决了无线数据传输的不稳定性问题;然后,利用多线程技术和大规模通信事件响应处理技术,提出了实现数据接收线程池的方法,通过引入“数据分区管理”的概念,论述了数据并发处理缓冲区的设计,解决了大并发系统容易丢失数据、堵塞和低效率等问题;最后,通过对系统测试及其在业务中应用的分析,表明该系统实时、高效、可靠,能满足目前业务的需求,并对未来业务的发展具有同样的适应性.

关键词

无线通信;浪涌气象数据;可靠应用协议;线程池;数据缓冲区;并发处理

中图分类号 TP311

文献标志码 A

收稿日期 2012-07-02

资助项目 广东省科技计划项目(2010B060100055);广东省气象局科技项目(201021)

作者简介

伍光胜,男,高级工程师,研究方向为网络通信与软件开发技术. wgs@gmcc.gov.cn

¹ 广东省大气探测技术中心,广州,510080

0 引言

在气象自动观测数据采集集中,数据采集或控制等终端设备都分布在不同的地点,具有分散性和流动性.由于有线通信费用较高,通信线路易受干扰和雷电入侵,给自动气象站资料的正常采集带来很大的负面影响.同时,在一些条件恶劣、地域偏僻的地区,有线通信根本无法到达,使自动气象站的合理布点有一定的局限性.所以,在目前移动通信业务成熟、通信费用低、信号覆盖面广的情况下,自动气象站通信组网一般都基于无线通信业务,采用客户端与服务器的通信模式,所有采集到的数据或交互的信息都通过客户端通信终端定时地发送到数据服务中心^[1].这些数据通信客户端数量大,例如我国已建成的自动气象监测网络中,无人值守的自动气象观测站已达到了3万多个.在这样庞大的数据传输网络系统中,所要传输处理的数据信息具有数据量小、规模大,传输突发、频繁等特点,同时还要求数据传输处理实时高效、数据采集同步以及数据探测密度能随业务不同要求而变化.所以,对这种大规模突发气象监测数据(即浪涌气象数据)的同步传输和实时处理,系统必须解决无线通信的不稳定性及大量同步数据的堵塞、丢失、处理时效等问题.在此,将结合广东省自动气象站数据采集系统的设计,对如何设计可靠应用协议来保证数据传输的可靠性,如何利用线程池技术及缓冲区技术以实现数据的快速处理和存储作详细的讨论.

1 系统设计

1.1 系统设计的要求

目前,广东省已建成并在线业务运行的自动气象站数量为1 800多个.按气象业务化要求,各个自动气象站在常规天气每5 min形成1组观测数据,在诸如台风、暴雨、冰雹、雷暴、低温冰冻等特殊天气和公共突发事件气象应急过程中,自动气象站的观测周期加密为1 min,即每分钟产生1组观测资料,所有自动气象站观测资料都要实时同步地发送到省数据中心.随着气象业务发展和精细化预报的需要,自动气象站的布点还要继续增加,所以,系统设计的要求既能实时接收目前所有自动气象站周期性上传的大批量的浪涌观测数据,又要能满足未来日益增长的自动气象站的数据传输.同时,为了避免上传观测数据的积压,要求在每分钟周期内将所有报文处理入库.

1.2 系统设计方案

按业务的需求,系统采用 GPRS 无线网络作为通信构架,通过 DTU(无线通信终端)将自动气象站与数据采集中心连接起来,其结构如图 1 所示。

系统由数据采集、数据传输和采集中心 3 个组成部分。数据采集由自动气象站自动采集风向、风速、雨量、温度、湿度、气压等气象要素的实时数据并计算形成分钟数据,存储该数据并通过 RS232 转发到 DTU。DTU 通过 GPRS 无线网络将数据传送到数据采集中心,采集中心对所收到的数据进行处理并形成产品入库供业务使用^[2]。

1.3 系统设计的关键点

在自动气象站数据采集系统中,自动气象站业务已应用多年,技术成熟、稳定、可靠,一般都提供标准的 RS-232 接口,以方便通信的接入,不存在需要解决的技术问题。但采用 GPRS 无线网络作为系统的数据传输,也存在通信带宽窄、通信链路容易被抢占而出现抖动等不稳定现象,所以,如何在使用该通信架构的优势下保证数据传输的实时性和可靠性,是系统设计需要解决的关键点之一。同时,对于这种大规模自动气象站网监测数据的接收和信息的交互,虽然每个自动气象站的每份报文数据量不大,但站点数目大,采集发送数据时间需要同步,观测周期短,在同一时间点采集中心需要同时接收全部自动气象站的数据,并在短周期内完成数据处理及上传业务,所以,如何解决这些问题、需要使用哪些技术以及确保所开发的系统安全、可靠、高效是系统设计

需要解决的关键点之二。

2 数据传输设计

2.1 数据传输协议分析

在 GPRS 数据传输中,无线移动终端(DTU)通过 BSS(基站子系统)发起对 SGSN(GPRS 业务支持节点)的“PDP(分组数据协议)上下文激活”请求,SGSN 进行 PDP 激活有效性判别、APN(接入点名称)选择以及 APN 到 GGSN(网关支持节点)的地址解析,然后产生一个“创建 PDP 上下文请求”的消息发送到相应的 GGSN,SGSN 获得 GGSN 的回应后将 PDP 报文返回到 DTU,此时,DTU 完成 PDP 上下文激活流程并获得 IP 地址,从而 DTU 就可以通过 TCP/UDP 协议与采集中心(Acquiring Center)通信。其数据传输的协议结构如图 2 所示(图 2 中的各具体环节可参考文献[3-4])。

在基于 TCP/UDP 协议的通信中,采用面向非连接的 UDP 协议传输策略,可以避免如 TCP 协议产生大量的网络状态确认和数据确认而消耗系统资源,大大提高传输速度和效率,同时也不存在如 TCP 协议的海量并发连接问题。特别在于 GPRS 通信带宽的限制下,对于消息包传递的应用程序来说,基于帧的通信比基于流的通信更为直接和有效^[5]。所以,在大规模自动气象站数据突发传输中,为了简化系统的开发,提高网络通信系统的性能,在充分利用 GPRS 网络的优势的前提下,采用 UDP 协议作为数据传输协议。然而 UDP 协议不提供数据传送保证机

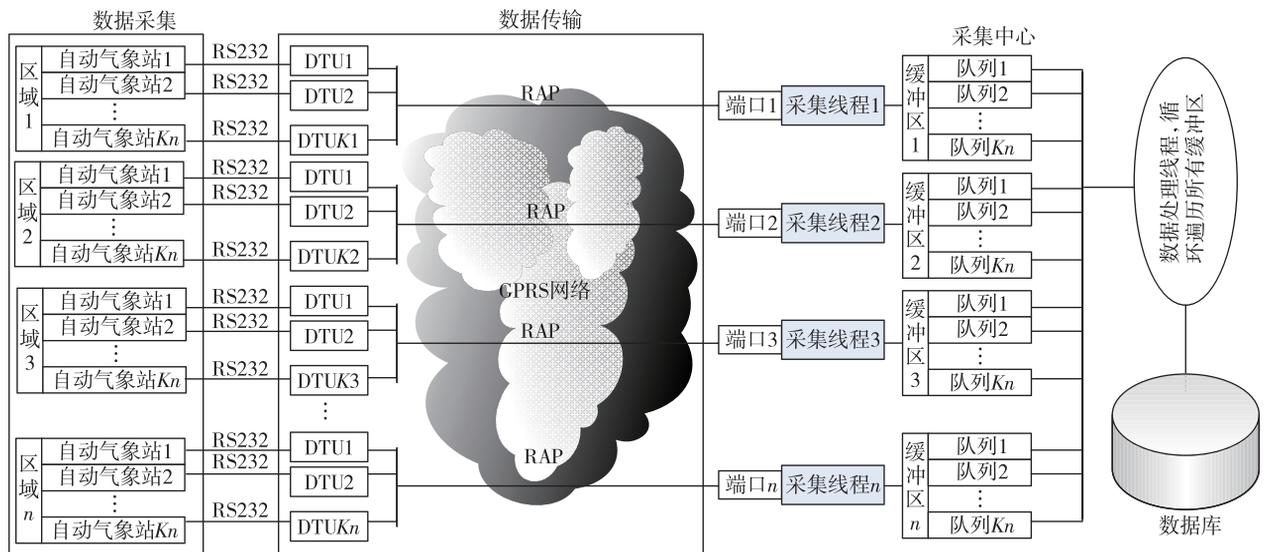


图 1 自动气象站数据采集系统

Fig. 1 Structure diagram of automatic weather station acquiring system

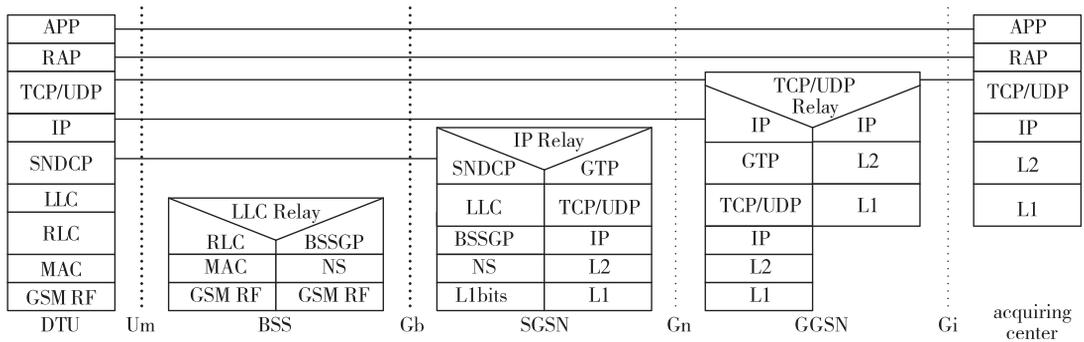


图2 数据传输协议结构

Fig.2 Structure diagram of data transmission protocol

制协议,存在可靠性差、传输功能单一的缺点,为确保自动气象站资料传输的可靠性,通过在 UDP 协议与业务应用层之间设计一个可靠应用协议(RAP),以解决 UDP 协议的无连接性、多个数据包突发传输时可能出现的数据包无序性及流量控制等问题。

2.2 可靠应用协议结构

RAP 通过对 UDP 协议结构的扩展,在原 8 B 报头的基础上增加 5 B,用于对自动气象站资料进行有序报文封装、站号标识及报文识别,其结构如表 1 所示^[6]。

表 1 RAP 协议报文数据格式

Table 1 RAP protocol packet data format

1~8	9	10	11	12	13	14~65 536
UDP 报文	命令	ID 高位	ID 低位	数据包 序列号 高位	数据包 序列号 低位	用户数据(可变量长度) 一般小于 1 kB

1) 第 9 个字节“命令”指自动气象站数据通信终端与数据采集中心通信的握手标识,不同的标识表示 UDP 包是注册包(数据通信终端上线时发出)、心跳包(数据通信终端在线定时测试通信链路)和用户数据包等。

2) 第 10 个字节“ID 高位”和第 11 个字节“ID 低位”指通信终端标识 0 至 65 535,用于区别不同安装地点的自动气象站发送的信息。

3) 第 12 个字节“数据包序列号高位”和第 13 个字节“数据包序列号低位”指每个自动气象站发送到数据采集中心的数据包流水号,从 0 到 65 535 重复执行,用于使 UDP 报文从无序变为有序。

4) 从第 14 个字节开始到第 65 536 个字节为自动气象站的监测数据,根据自动气象站每次观测的数据量、无线传输网络的特性及 UDP 协议的特点,

用户数据长度一般可设计小于 1 kB。

2.3 基于 RAP 的 DTU 数据传输

在植入 RAP 协议的 DTU 完成与无线网络有关 PPP 通信链路的操作协商后,IP 通信链路被创建,DTU 向通信服务系统发出 13 B 的 RAP 注册包,完成通信链路的建立登记,就可以进行基于 RAP 报文的自动气象站资料传输,同时通过定时器的注册包连续无回应次数 K 来判断通信链路是否建立成功,其工作流程如图 3 所示。

为了避免通信链路不稳定或断链时进行无效的数据传输,在每次完成缓冲队列的数据传输前,都要进行链路 RAP 心跳包(13 B)测试.在数据等待队列为空时,通过启动定时器的次数 M 决定是否发送心跳包来保持自动气象站通信终端的 PPP 链路持续连接.当永久连接的链路出现错误时,通过定时器的数据包连续无回应次数 N 来决定是否关闭通信链路。

在基于 GPRS 业务的自动气象站资料传输网络中,从服务器到通信终端的 PING 时间最大延时为 3 s,则对于小数据量注册包和心跳包的定时器 1 及定时器 2 的等待时间为可设定为 6 s 左右,连续没有收到 ACK 包的 K, N 值可设定为 3 次.对于每分钟产生 300 B 的气象观测数据包,定时器 3 及定时器 4 的等待时间可设定为 12 s 左右, M 可在定时器 3 的连续累计时间 3 ~ 5 min 内计算取值,视网络情况及需要,若为 3 min,则 M 取值 15(180/12)。

2.4 数据传输的可靠性分析

在应用系统中,传输保障是由应用协议与网络协议共同完成的. RAP 协议实质上是根据无线网络的特点在客户端的应用层对用户数据进行自定义用户协议封装,再采用 UDP 协议进行传输的协议.其设计在 UDP 协议的基础上充分利用 TCP 数据传输

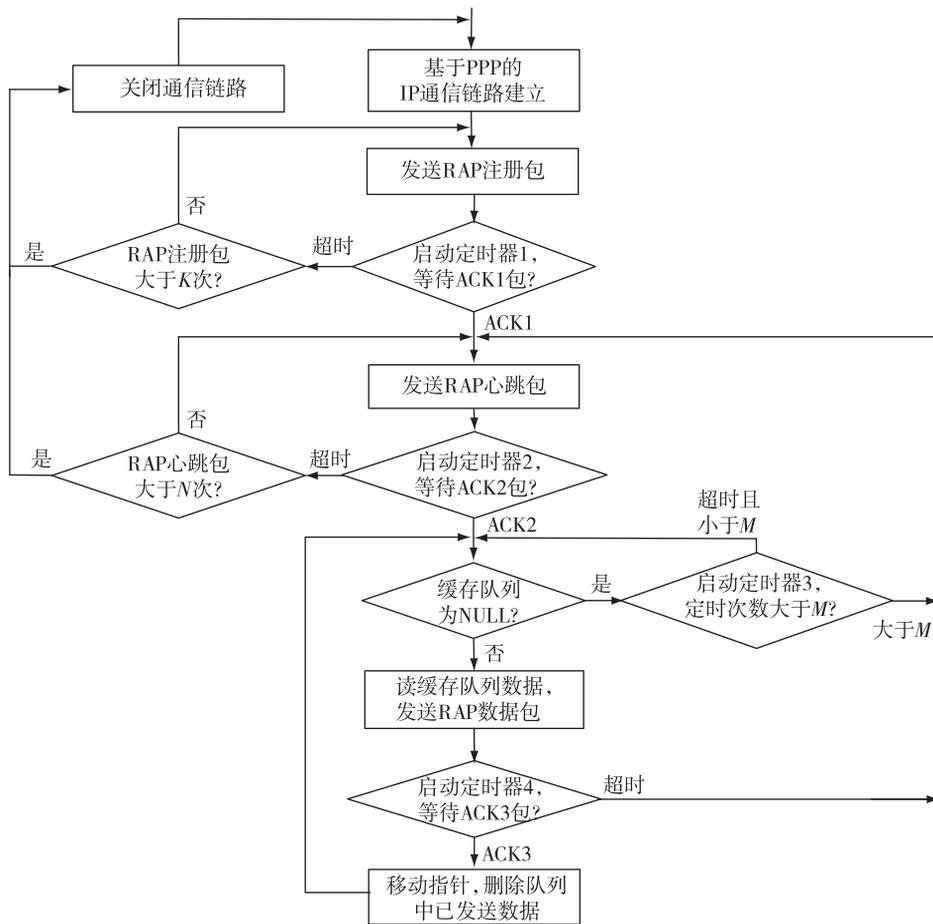


图3 基于RAP协议DTU工作流程

Fig. 3 Workflow chart of DTU based on RAP protocol

协议的特点,在自定义用户协议上建立类似TCP协议的严格按序数据传递及流量控制,又不像TCP发生错误时的信息重传导致不必要的阻塞和传输时延增大。

在基于RAP的自动气象站资料传输中,采用消息收发同步以及流量控制,探测数据传递的序号采用2B无符号整数表示,取值范围为0~65535,形成一个大的循环队列,发送数据大小相同且可设为一次探测数据的大小。当自动气象站通信终端向通信服务中心发送数据包后,都要等待服务器的ACK包回复,当ACK包在设定的时间内到达超时或丢失时,启动重传功能,在重传自动站数据包前,通过小数据量的心跳包对通信链路进行测试,防止在大流量时出现网络过度堵塞的现象。在GPRS网络中,按照UDP丢包的概率,重传概率在1%左右。这样通过数据包发送方与接收方的握手,有效地控制了发送数据的速度,同时,通信服务中心对收到的数据包序号检查而判断是否是重复包而作相应处理,从而

解决UDP传输丢包和失序的问题。

无线网络出现的异常,通常是带宽不足而出现拥塞或通信终端出现假在线情况。使用UDP连接,当网络拥塞时,部分数据包被丢弃,但可以改善接收数据严重滞后的情况,通过重传策略,又能保证数据发送成功。当网络抖动而通信终端出现假在线时,由于其无法收到注册包或心跳包的ACK包而复位重新向无线网络发起PPP连接,建立新的传输链路,实现数据传输的自恢复。

3 采集中心设计

由于系统的主要任务就是实时接收处理来自全省自动气象站每5min乃至1min同步发送的“浪涌”数据,为了便于对各地区自动气象站的管理和大规模资料同步接收处理,所以,系统的设计采用“分区处理技术”,即以每个地区所管辖的自动气象站为一个管理单位分别给予分配一个对应的服务器SOCKET端口和数据缓冲区。通过引入多线程技术

和 I/O 通信模型,构建多线程通信线程池,实现所有的通信端口由通信服务系统的线程池统一管理,即线程池的每一条线程对应管理一个端口,由线程通过被分配端口负责接收来自对应地区的自动气象站资料.采用“数据分区管理”技术,设计数据并发处理缓存区,实现对接收数据的实时存储和快速处理^[7].

3.1 线程池设计

3.1.1 线程池结构

为了避免在接收大规模突发数据时频繁地创建、分配、销毁线程而大量消耗系统资源和系统频繁在各个线程进行上下文切换而浪费大量 CPU 时间,通过使用多线程技术创建自动站资料接收线程池,根据系统当前的业务需求和负载状态决定创建线程的数量,并对预设接收线程进行有效管理,发挥多线程并发处理能力,提高系统的性能^[8].线程池工作示意如图 4 所示.

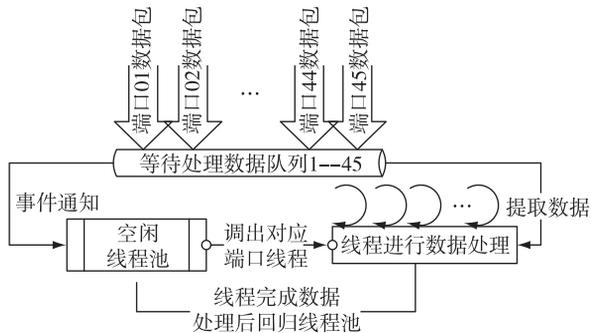


图 4 线程池工作示意

Fig. 4 Schematic diagram of the thread pool work

根据广东省地级市数量及对一些特殊应用的自动气象站归类(如海洋自动观测站、气象地质灾害自动监测站、城市热岛自动观测站、立体气候自动观测站、土壤湿度自动观测站、农业生态自动观测站、突发公共事件与灾害气象应急等)及未来业务拓展对备用端口的需求,共建立 45 路数据接收端口.创建一个用于存储待接收处理的自动气象站数据的 45 个信息队列,每个队列对应一个端口.创建用于管理自动气象站数据接收线程组的一个结构对象(线程池),根据端口的开启或关闭,该对象能动态地创建、销毁和管理对应端口的线程,如已启用了 30 个端口,则创建 30 条接收线程.当某一端口信息队列为空的时候该组线程阻塞,并置于空闲线程池中.当数据到达该队列时,采用 Windows 事件通知唤醒线程池中相应的线程,对其信息进行读取和处理.当线程

将数据存入缓冲区后,返回空闲线程池.

3.1.2 线程池的实现^[9]

线程池的实现主要是对线程的设计,主要采用 Windows 操作系统提供的网络编程套接字重叠 I/O 事件模型来实现.面向无连接 UDP 协议的大规模突发数据传输,其具有以下特点:避免使用完成端口模型而使处理线程的互斥消耗大大影响通信系统的性能;使用重叠的数据结构,可一次投递一个或多个 I/O 请求,对比 Select、WSAAsyncSelect 以及 WSAEventSelect 模型,使应用程序能达到更佳的系统性能,并且可以运行在支持 Winsock2 的所有 Windows 平台;和 WSAEventSelect 模型在实现上非常相似,能减少开发系统时编程的复杂性,同时也达到了与完成例程同样的性能和效果.

线程的编程通过创建 Windows 的 WSAEVENT 事件对象和 WSAOVERLAPPED 重叠结构,并由 WSAOVERLAPPED 结构中专门对应的参数将事件对象和重叠结构进行关联.通过 WSA Sendto, WSARecvFrom 函数中所包含的一个 Overlapped 参数将这些函数“绑定”到 WSAOVERLAPPED 重叠结构上,然后提交这些函数,所有请求就交给重叠结构去操作,等到重叠操作完成以后,系统就通过对对应的事件来通知操作已完成.例如:接收来自某个端口的自动气象站数据,在创建相应的事件对象和重叠结构并关联后,就必须提交一个 WSARecvFrom 请求,然后通过 WSAWaitForMultipleEvents 函数等待事件对象的触发,当重叠操作完成,通过调用 WSAGetOverlappedResult 函数得到本次 I/O 传送的字节数等相关信息,这样就可以根据重叠操作的结果取得自动气象站数据.线程池的线程实现如图 5 所示.

3.2 数据缓存区设计

在大规模突发气象数据的接收处理中,如果频繁地申请和释放内存,对于不间断运行的服务器而言,系统经历了长时间的运行后,必将造成内存碎片而使内存管理变得杂乱,对内存数据的操作将会消耗更多的 CPU 时间.当物理内存无法满足系统需要时,系统将会把物理内存中的数据交换到虚拟内存中,并且重新整合杂乱的内存区域,以获得需要的内存空间.系统堆的杂乱无章和物理内存与虚拟内存的频繁交换,导致服务器长时间运行之后变得越来越慢甚至可能崩溃^[10].所以,为了避免频繁申请和释放内存,同时要避免内存管理的不当造成内存泄漏,在数据并发处理缓存区的设计中引入了“数据分

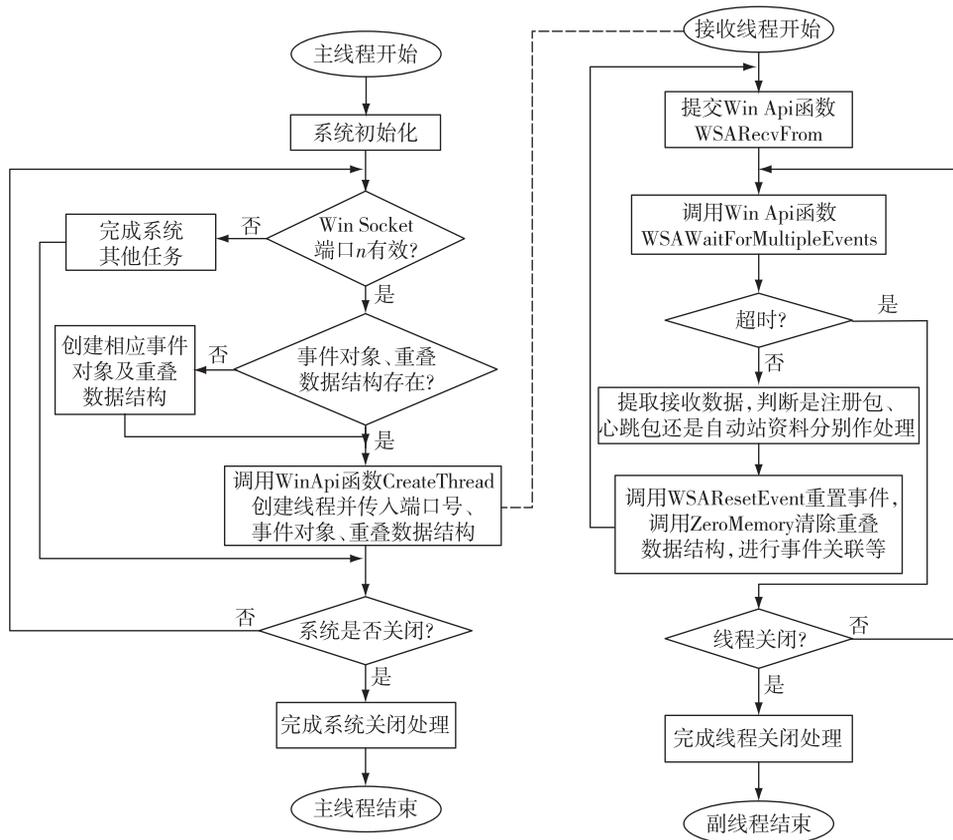


图5 线程池实现流程

Fig. 5 The realization flow chart of thread pool

区管理”的概念,采用内存池技术和高度的内存重用来达到节省内存空间和提高内存管理的性能,其设计可分为2部分,即并发存储结构的设计及缓冲区的并发访问技术。

3.2.1 并发存储结构

首先,根据业务对缓存的需求一次性向系统申请足够大且日后可扩充的内存块(Memory pool),将该内存块划分为大小均等的子内存块,所有子内存块由一个指针表(Free_List)来维护管理,每个子内存块由一个循环队列(Line[1 to n])组成。由于在实际系统中数据通信终端发出的UDP数据包大小小于512 B,如自动气象站的每份报文为300 B,所以,每个循环队列中的每一列大小可设为512 B。自动气象站的每份报文周期为5 min,特殊天气时周期加密为1 min,如果按12 h每分钟周期报文的缓冲能力计算,每个队列长度需要 $1 \times 60 \times 12 = 720$ 列,需要内存量为 $720 \times 512 = 360$ kB,按业务计划2 010个自动气象站计算,一次性申请Memory pool共需要 $2 010 \times 360 \approx 707$ MB内存。当然,在其他应用中,队列的长度和每列的大小都可根据实际业务需要而调

整,然后申请所需的内存。

其次,创建分区指针列表(Subarea_List),即为每个Socket端口对应分区创建一个指针,如共有45个端口,需要创建45个指针(Subarea_Point[1 to 45])。创建每个分区所管辖的自动气象站指针表(Station_List),如该管理区域有200个自动气象站,需要创建200个指针(Station_Point[1 to 200]),Subarea_List中的每个指针指向相应分区的Station_List。创建每个循环队列数据存取指针,即写指针Write_Point和读Read_Point,Station_List中的每个指针指向各自的队列存取指针。通过对指针的操作,就可以快速寻找到每个分区中各个自动气象站所对应循环队列的读写区。

3.2.2 缓冲区的并发访问技术

当某个地区的自动气象站发出的报文到达相应的端口时,数据到达事件通知线程池相应的空闲线程处理,通过端口号查找Subarea_List,获取Subarea_Point。由Subarea_Point取得Station_List入口地址,通过RUDP协议中的ID查找Station_List中的Station_Point。若Station_Point存在,取得Write_

Point,将数据写入队列;若 Station_Point 不存在,则向 Memory pool 申请,Free_List 释放子内存块队列指针,并将该指针赋予 Station_Point,加入到 Station_List 中,于是就可通过 Write_Point 将数据写入队列.当 Station_List 中某个队列很久没有发生存储操作时,释放该队列到 Memory pool 中,Free_List 回收该队列指针.

对队列数据的读取,采用单线程对每个分区中的每个队列进行遍历的方式读取数据,即先访问第一分区的所有队列后逐一访问下一分区的所有队列,访问完毕又重新开始.对队列读写,采用对 Write_Point 和 Read_Point 的同步加锁和比较,实现数据接收线程与数据发送线程对队列数据的并发处理.各线程访问指针列表和内存池关系如图 6 所示.

3.2.3 缓存结构及访问技术的优势

在多线程数据并发访问中,影响共享数据访问效率的主要因素是锁竞争.在单核多线程编程中,采用细粒度锁可以取得很好的分时效果;在多核编程时,细粒度锁更为重要,它不仅取得分时效果,还可以使加速比性能得到大大提升^[11].所以,在对数据缓存结构的设计中,采用线程读写缓冲区时只对某一队列操作,加锁机制只对该队列有效的方法.实际上是采用双指针对队列的操作,对指针变量加锁,从而实现了细粒度的锁机制,提高了线程对缓冲区的并发处理能力,同时避免造成内存伪共享问题.

采用多生产者与单消费者模式,对于由许多队列组成的内存操作,即在线程池接收线程组(Thread [1 to 45])分别对各自管理的队列(Line[1 to n],n 为分区分管的自动气象站数)写操作和数据发送线程遍历所有队列读操作时,发生读写线程同步等待对某队列的读写指针解锁几率极小,大大提高了对内存的访问速度.若采用多线程读缓冲队列(每读一条线程对应一条写线程),则需要的同步锁为原来的 45 倍.按参考文献[12],在一个 1.83 GB 的双核 CPU 上运行一个线程加锁解锁 1 000 万次所耗费的时间及 2 个线程并行运行 1 000 万次加锁解锁(每个线程运行 500 万次加锁解锁)时间进行比较,发现使用 2 个线程后,所耗费的时间不仅没有减少为单个线程时的一半,反而增加了 1 872 - 577 = 1 295 ms.可见锁竞争导致串行执行时,所耗费的时间里还增加了线程切换等其他开销,并且线程切换等开销占的比例还非常大.锁竞争导致的实际串行化执行情况是很不乐观的,如果程序中存在大量的加锁解锁操作的代码,其性能是可想而知的.

对于读写线程的同步锁设计,采用临界区 CRITICAL_SECTION 技术.在 Windows 系统提供的锁机制中,临界区是通过对多个线程的串行化来访问公共资源或一段代码.与其他同步对象如互斥相比,临界区相对较快,比较适合于控制数据的访问.

由于自动气象站的数据处理是比较复杂费时

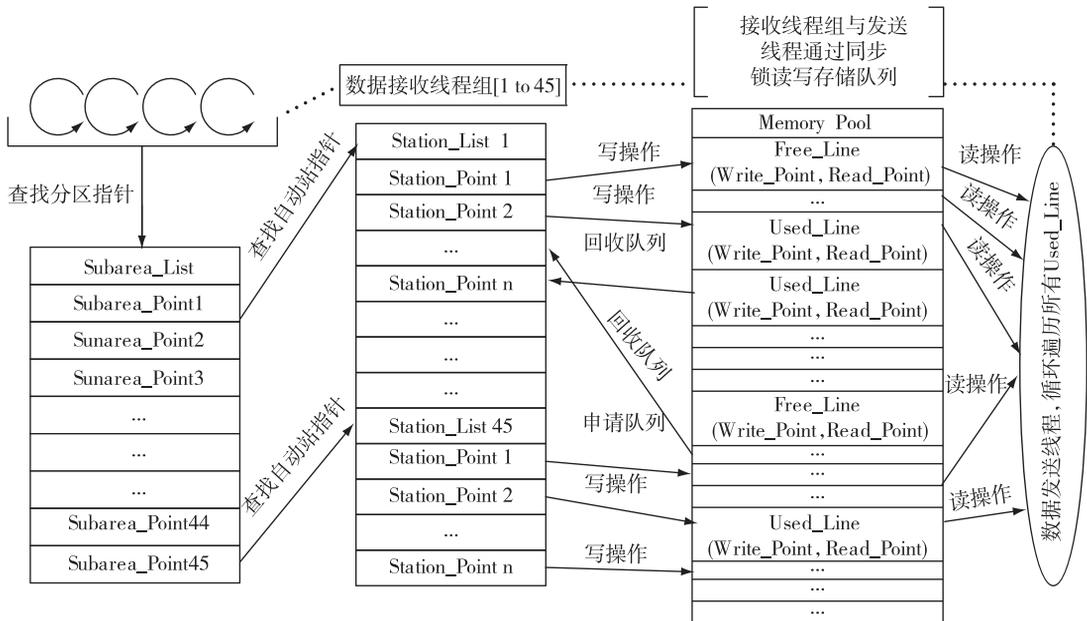


图 6 线程访问指针列表和内存池关系

Fig. 6 Threads access concurrently pointer list and memory pool

的,所以,为避免影响接收线程对浪涌数据的实时接收和存储效率,将数据处理任务分离出来,在其他服务器上处理.通过维护线程与多个处理进程的多TCP连接,采用单消费线程快速从缓冲队列读取数据并发送处理进程,实现数据的分布式处理^[13].

总之,在设计本系统时,为了简化编程任务和模块化,提高处理器的效率,减少任务的切换时间,提高吞吐量及并发度,采用了分区管理技术,将任务较为均衡地分配到每个接收线程;通过对读写指针的加锁,实现了细粒度的锁机制;采用多生产者与单消费者模式,减少了同步锁等待几率;利用临界区加锁,提高了数据读写速度;采用数据分布式处理方法,提高了系统的性能.

4 试验和应用

为了证明系统的性能优势及在业务应用中的出色表现,在服务器为 DELL PowerEdge 2950 (CPU 为 1.6 GHz,内存为 8 GB)、操作系统为 Windows 2003 Server 的环境下,对系统的性能进行以下测试和考核.

1) 通信容量处理能力测试. 在 100 MB 的局域网上,以 45 路端口接收模拟自动气象站每秒发送的 300 B 的气象数据包,为每个模拟自动气象站开设 500 条缓存队列,每条队列的长度为 1 kB,每次以 20 个模拟自动站为增量给各个端口分配模拟自动气象站.由于从客户端到服务器端的 PING 时间小于 1 ms,所以 RUDP 协议中定时器 1、2、3、4 值分别设置为 10、20、50 和 50 ms, K 、 N 、 M 的值设置为 3. 从 Windows 任务管理器得到服务器 CPU 及内存的使用情况如表 2 所示.当模拟自动气象站达到 6 000 个时,系统运行及资料处理能力正常,但 CPU 及内存使用的峰值已超 50%,CPU 峰值上升较快可能与可用物理内存较少有关.所以,在此环境下系统至少能达到每秒 6 000 个资料的处理能力,若要使系统有更大的通信处理能力,须要更高级的服务器支持.

表 2 通信容量处理能力测试表

Table 2 Test for communication and processing capacity

模拟自动气象站数/个	CPU 使用峰值/%	内存使用峰值/%
900	5	8
1 800	9	14
2 700	10	21
3 600	20	29
4 500	28	36
5 400	35	45
6 300	51	58

2) 系统缓存能力测试. 关闭数据处理服务器,让通信服务器将接收到的数据寄存在数据缓冲区.按 Windows 系统要求,为了防止出现未分页内存池被耗尽,当程序锁定内存时,不要超出系统规定的内存分页极限,应用程序总共可以锁定的内存最佳不要超过物理内存的 1/3.按 6 000 个模拟自动气象站每分钟来报计,每小时使用的内存为 $6\ 000 \times 60 \times 300 \approx 103$ MB.经测试,系统连续缓存数据能达 1 d,远远超出业务对故障处理要求的响应时间.同时,采用分布式数据处理,系统不会因数据处理服务进程故障而影响数据的正常接收.

3) 业务考核.在目前广东省 1 800 多个的自动气象站业务应用以来,系统运行高效、稳定、可靠.所有自动气象站每分钟同时向接收中心发送数据包,系统对于每个周期的“浪涌”数据接收、处理、上传过程耗时小于 1 min,系统大部分时间仍然处于等待空闲状态.系统(DataCenter.exe)内存及 CPU 使用情况如图 7 所示,系统界面如图 8 所示.

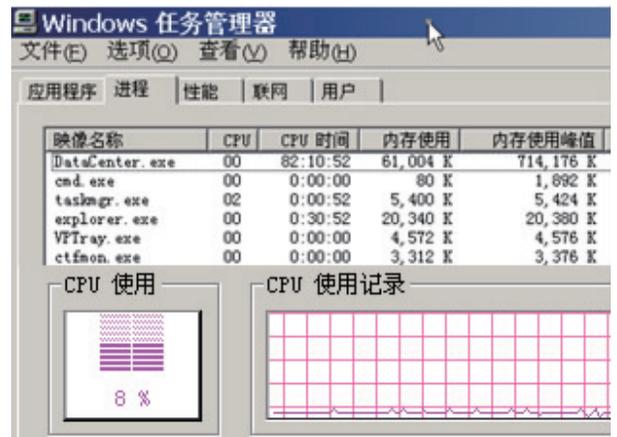


图 7 系统内存及 CPU 使用情况

Fig. 7 CPU and memory usage of system (DataCenter.exe)

从图 7 可看出,系统的 CPU 使用及占用的内存是非常小的.从图 8 系统的业务运行界面显示,目前共有注册上线的自动气象站有 1 842 个,系统的启动时间是 2012-09-07T9:16:24,界面的截图时间为 2012-09-24T15:17.当天的 183 个来包刚好是每 5 min 一个自动气象站数据,当天的重发包数极少,可见基于 RAP 协议、线程池及缓冲区技术的浪涌气象数据传输处理系统具有强的可靠性、高效性和健壮性.

5 结语

通过对大规模突发气象监测数据传输处理系统

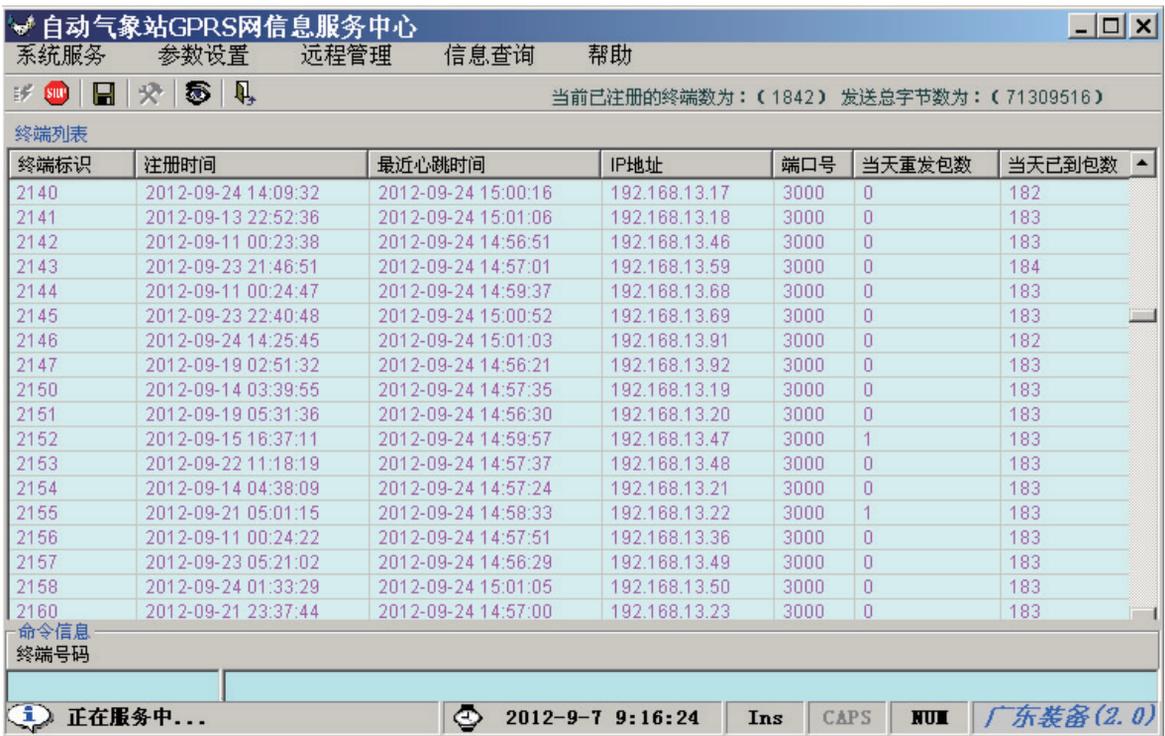


图8 系统(DataCenter.exe)界面

Fig. 8 System (DataCenter.exe) interface

的设计,研究了基于UDP协议的可靠应用传输协议、大规模通信响应处理技术及数据并发处理存取缓冲区。测试和应用表明,该技术解决了基于无线网络的无人值守、实时性、突发性UDP数据传输的不稳定性、无序性,实现了大规模突发“浪涌”气象数据的实时传输和处理,保证了系统在高性能运行的基础上极大地降低了系统软硬件成本、系统复杂性以及维护成本,同时满足了未来业务的发展对系统的需求。

参考文献

References

[1] 伍光胜,敖振浪,李源鸿,等. 大型自动气象监测网及数据采集中心的设计及应用[J]. 气象,2010,36(3):128-135
WU Guangsheng, AO Zhenlang, LI Yuanhong, et al. The design and application on large automatic weather monitoring network and data acquisition center[J]. Meteorological Monthly, 2010, 36(3): 128-135

[2] 谭鉴荣,伍光胜,林金田. 基于GPRS网络的远程气象数据采集系统设计和实现[J]. 气象科技,2009,37(5):587-592
TAN Jianrong, WU Guangsheng, LIN Jintian. Design and implementation of remote meteorological data acquisition system based on GPRS network[J]. Meteorological Science and Technology, 2009, 37(5): 587-592

[3] 孙剑俊. GPRS的APN激活过程分析[J]. 广东通信技术,2003,23(4):9-12
SUN Jianjun. Analysis of GPRS procedure in APN activation[J]. Guangdong Communication Technology, 2003, 23(4): 9-12

[4] 乌欣,胡捍英. GPRS的网络结构和协议分析[J]. 通信技术,2002(3):59-61
WU Xin, HU Hanying. Network structure and protocol analysis of GPRS[J]. Communications Technology, 2002(3): 59-61

[5] 施伟年,凌海宏. GPRS网络上的两种数据传输协议[J]. 电力系统通信,2004(8):20-22
SHI Weinian, LING Haihong. Discussion about data transmission protocols for GPRS network[J]. Telecommunication for Electric Power System, 2004(8): 20-22

[6] 伍光胜,敖振浪,谭鉴荣,等. 基于气象探测网应用的移动通信终端的实现[J]. 计算机应用研究,2007,24(增刊1):1708-1709
WU Guangsheng, AO Zhenlang, TAN Jianrong, et al. The Implementation of mobile communication unit applied to weather probe net[J]. Application Research of Computers, 2007, 24(sup1): 1708-1709

[7] 伍光胜,敖振浪,林建勇,等. 实时接收和分布式处理系统用于大规模突发数据[C]//Proceedings of the 2nd International Conference on Networking and Distributed Computing, 2011:111-115

[8] 伍光胜,宋信忠,郑明辉. 多线程技术及其应用的研究[J]. 计算机应用研究,2001,18(1):33-36
WU Guangsheng, SONG Xinzong, ZHENG Minghui. Re-

- search and application of multi-thread technology [J]. Application Research of Computers,2001,18(1):33-36
- [9] 黄超. Windows 网络编程[M]. 北京:人民邮电出版社,2003:264-301
HUANG Chao. Network programming for Microsoft Windows[M]. Beijing:Posts& Telecom Press,2003:264-301
- [10] 冯林,彭尉,滕弘飞. 车辆监控系统中通信服务器性能优化设计与实现[J]. 计算机工程与应用,2005,41(6):199-201
FENG Lin,PENG Wei,TENG Hongfei. Design & implementation for performance optimizing of communication server in vehicle supervision system[J]. Computer Engineering and Applications,2005,41(6):199-201
- [11] Lipman J, Stout Q F. A Performance Analysis of Local Synchronization [C] // Proceedings of the 18th ACM Symposium on Parallelism in Algorithms and Architectures,2006:254-260
- [12] 周伟明. 多核计算与程序设计[M]. 武汉:华中科技大学出版社,2009:28-31,55-67
ZHOU Weiming. Multi-core computing and programming [M]. Wuhan:Huazhong University of Science and Technology Press,2009:28-31,55-67
- [13] 周钦强,谭鉴荣,伍光胜,等. 基于 TCP 多连接通信实时并发数据处理技术研究[J]. 计算机工程与应用,2007,43(18):246-248
ZHOU Qinqiang,TAN Jianrong,WU Guangsheng, et al. Research on real-time and concurrent DP system based on TCP multi-connections[J]. Computer Engineering and Applications,2007,43(18):246-248

A surge meteorological data acquiring system and its key technologies

WU Guangsheng¹ LI Jianyong¹ LIU Yanzhong¹ LEI Weiyan¹

¹ Guangdong Province Atmospheric Observation Technology Centre,Guangzhou 510080

Abstract For characteristics and requirements of data acquiring in large-scale unattended automatic weather monitoring, the design on acquiring system of surge meteorological data are given, and the key technologies are discussed in depth based on wireless communication network. By analyzing the structure of data transmission protocol and its disadvantages in the system, a reliable application protocol is designed based on UDP and its working principle and reliability are discussed in detail, and the instability of the wireless data transmission is solved. For real time receiving and concurrent processing of surge meteorological data, by making use of multi-threading technology and large-scale communication events response technology, the data reception thread pool is constructed and realized. Through the introduction of data partition management concept, a concurrent processing data buffer is planned and excogitated, which can solve the problems such as data loss, data blocking and low efficiency in data processing. The system is tested and applied to real weather operation, and its performance shows that this system can acquire real time surge meteorological data efficiently and reliably. The system meets the current needs of the weather operation and is adaptable to the development of future weather operation.

Key words wireless communication; surge meteorological data; reliable application protocol; thread pool; data buffer; concurrent processing