

公路交通量调查数据处理系统

胡立辉¹, 田跃², 盘善荣³

(1. 长沙理工大学计算机与通信工程学院, 长沙 410076; 2. 湖南省交通厅信息中心, 长沙 410015; 3. 长沙理工大学图书馆, 长沙 410076)

摘要: 分析了一个公路交通量调查数据处理系统的需求特点、总体结构、处理流程、数据库设计、系统实现。该系统包含5个相关但独立运行的子系统: 实时数据包接收器, 数据包转发器, 运行监控系统, 查询与统计客户端, 数据在线发布子系统。介绍了系统的主要数据库表及存储过程, 给出了基于时序区间的判重算法, 推广了加权平均车速算法, 讨论了多线程处理、Web服务及Excel报表制作等实现技术。实际运行表明, 该系统有良好的稳定性、健壮性与可伸缩性, 满足了行业管理需求。

关键词: 公路; 交通量; Socket; Web服务

Processing System for Highway Traffic Flux Investigation Data

HU Li-hui¹, TIAN Yue², PAN Shan-rong³

(1. College of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410076;

2. Information Center, Communication Department of Hunan Province, Changsha 410015;

3. Library of Changsha University of Science and Technology, Changsha 410076)

【Abstract】 This paper analyses a processing system for highway traffic flux investigation data, which includes requirement characteristics, total structure, processing flow, database design, system realization. The system consists of five subsystems: real-time datagram receiver, datagram forwarder, subsystem of monitor for running, client for querying and statistics, which are related each other but running alone. It introduces the main tables and stored procedures in database. An algorithm for judging duplicate datagram based on time order interval is given, and an algorithm for weighted average car speed is extended, and several key realized techniques, including multi-thread processing, Web services and report with Excel are discussed. Application shows that the system is stable, robust and scalable, and it meets the requirements of highway vocation administration.

【Key words】 highway; traffic flux; Socket; Web service

1 概述

公路交通量调查数据主要指公路的某个断面或一段路的通行车辆数及平均车速^[1], 是反映公路运营状况的重要信息, 是公路建设与规划、决策、养护与管理等方面的重要依据。当前, 公路交通量调查的方法主要有自动数据采集设备(自动观测站)和人工数据采集设备(人工检测仪)2种, 图1显示了有IP设备的自动观测站的工作流程^[2]。数据服务中心(data service center, DSC)则汇集了省(市)全部交通量数据, 工作模式见图2。

2006年3月交通部规划研究院发布了“公路交通量调查设备与数据服务中心基础交通数据通讯协议”^[2], 规定了基于Internet的TCP/IP及Socket数据传输协议、交通量数据包格式、错误信息码等, 为建立国家级与省级DSC提供了基础规范协议。

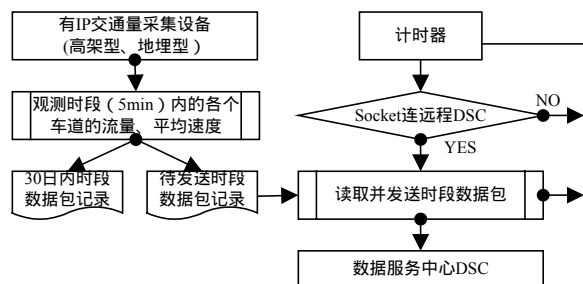


图1 自动观测设备工作流程

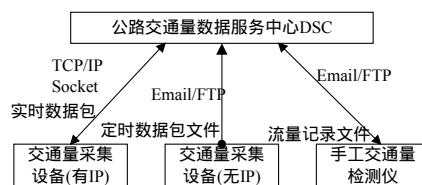


图2 数据服务中心工作模式

2 系统总体设计

2.1 需求特点

本系统的需求特点主要体现在实时包的处理速度、数据库的存储容量2个方面。

(1) 实时包的处理速度

一个观测站每5min上传一个数据包^[2]。这样, 有100个观测站时DSC将每3s处理1个数据包。按交通部“十一五”发展规划的要求^[3], 主要公路上每隔30km~50km要建立一个交通量观测站。湖南省将有600个以上的观测站, DSC将每秒钟处理2个以上的数据包。

(2) 数据库的存储容量

一个数据包有 $6+61*N$ 个字段数据(N 为公路的车道数)^[2]。按100个观测站、平均4个车道、每个车道0.5KB数据估计,

作者简介: 胡立辉(1965-), 男, 硕士、副研究员, 主研方向: 泛数据库理论与应用; 田跃, 高级工程师; 盘善荣, 硕士研究生、助理馆员

收稿日期: 2006-11-24 E-mail: ehulh@163.com

一年的记录数约 4 327 万条，需要存储空间 2.7GB。

2.2 总体结构与流程

系统由 5 个相互联系、独立运行的子系统组成，如图 3 所示。4 个独立子系统的工作流程见图 4。

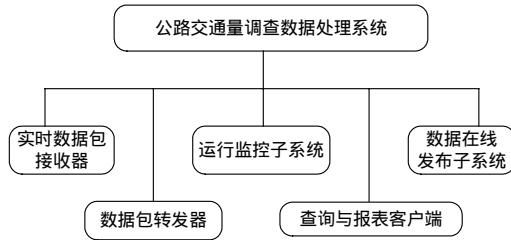


图 3 系统总体功能

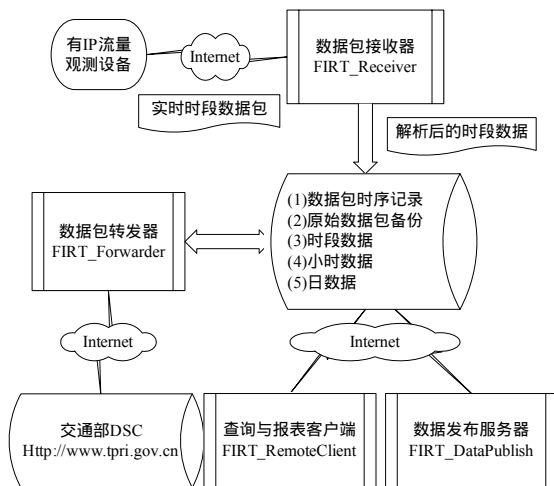


图 4 系统处理流程

各个子系统的功能简述如下：

(1) 实时数据包接收器

实时接收观测站上传的数据包。基本功能有：建立 Socket 连接，异步接收数据包，判断数据包的合法性，判断是否为重复包，解析数据包，保存原始数据包，保存(5min)时段数据，同步计算与保存小时与日数据，返回处理结果给观测站。

(2) 数据包转发器

转发接收到的数据包到交通部 DSC 中。为防止交通部 DSC 或网络异常等情况，在存储数据包后，通过独立运行的数据转发器转发数据包，增强了系统的健壮性。

(3) 运行监控子系统

基本功能有：定时检查数据包接收器、数据包转发器的运行情况，重装入已经停止的服务器进程，在系统崩溃时重启计算机或关闭计算机。本子系统部分代替了系统管理员的角色。

(4) 查询与报表客户端

基本功能有：查询指定观测站的时段、小时、日、月的交通量数据(包括时段的车道数据、小时的方向数据)，汇总指定观测站或一组观测站的时段、小时、日、月的交通量数据，统计与输出 Excel 报表文件，并提供时段、小时、日与月数据的下载。

(5) 数据在线发布子系统

为全社会提供公路交通量数据查询服务的 ASP.NET 网页。基本功能有：按时间、观测站查询 5min 时段的车道、方向、断面的交通量数据，按时间、观测站查询小时的方向数据、日及月的交通量数据。

2.3 开发与应用环境

数据库服务器为 Windows 2000 Server+SQL Server 2000(SP4)，Web服务器为 Windows 2000 Server+IIS6.0。服务器与客户端均需要安装 .NET 2.0 Framework，客户端需要安装 Office 的 Excel。开发工具选用免费的 Visual C# 2005 Express^[4-5]，报表制作与输出使用 Excel。

3 数据库设计

3.1 主要数据库表

系统主要的数据表包括：时序区间表 DatagramTime Order，存储已接收的数据包时序区间；原始数据包表 DatagramText，存储观测站的数据包，用于转发和备份；时段数据表 DatagramData，存储 5min 分车道的观测数据；小时数据表 HourData 存储 1h 分方向的观测数据；日数据表 DayData，存储 1 日的观测数据。

3.2 主要存储过程

系统主要的存储过程包括^[6]：存储数据包文本 Store_DatagramText，存储包解析后的数据 Store_Datagram Data，检查包时序并判重 Check_DatagramTimeOrder，客户端读取交通量数据 DataProvider_ResearchData。需要指出，在存储过程 Store_DatagramData 中不仅要保存 5min 时段的车道数据，还要计算小时的方向数据以及日数据，并累加到小时、日数据表中。

4 系统实现

4.1 判重与均速算法

4.1.1 时序区间的更新与判重算法

时序区间有 2 种更新方式：区间边界±1 扩展，增加时序区间。假定有按截止时序自大到小排序的 $N(>=0)$ 个不相交区间 Q_N, \dots, Q_2, Q_1 ，其中 $Q_i=[A_i, B_i]$ ，满足： $A_i \leq B_i, B_i < B_{i+1}$ (截止时序)。显然，这个区间序列对应了一维坐标上自右到左的 N 个不相交区间。现给定合法时序整数 K ，区间更新及判 K 重复的算法如下：

- (1) 若 $N=0$ 转(7)，否则令 $i=N$ ；
- (2) 若 $K=B_i+1$ ，则扩展边界 $B_{i+1} \rightarrow B_i$ ，转(8)；
- (3) 若 $K \in [A_i, B_i]$ ，则时序 K 重复，转(8)；
- (4) 若 $K > B_i+1$ ，则：
 - 1) 若 $K=A_{i+1}-1$ ，则扩展边界 $A_{i+1}-1 \rightarrow A_{i+1}$ ，转(8)；
 - 2) 否则，增加区间记录 $[K, K]$ ，转(8)；
 - (5) $i=i-1$ ，如果 $i \geq 1$ ，转(2)；
 - (6) 若 $K=A_1-1$ ，则扩展边界 $A_1-1 \rightarrow A_1$ ，转(8)；
 - (7) 增加区间记录 $[K, K]$ ；
 - (8) 结束。

容易证明，上述算法产生的新区间序列仍然是不相交的。

从时序区间表 DatagramTimeOrder 中按截止时序排序读出某个观测站的时序区间记录后，上述算法一趟扫描即可判重或更新区间记录。实际运行时只需要保存最近 T 天的时序区间记录，则 100 个观测站只有 $100 * T$ 条左右的记录数。这样，数据包判重及区间更新算法的时间不会随着 5min 时段记录数的增大而增大，显示了系统良好的可伸缩性。

4.1.2 加权平均速度算法

本文将文献[7]的算法公式推广到单个车型的时段断面平均速度、小时断面平均速度如下：

(1) 时段断面平均速度

$$V_t = \frac{\sum(N_{tp} * V_{tp})}{\sum N_{tp}}$$

其中， N_{tp} 为时段 t 第 p 车道的交通量； V_{tp} 为时段 t 第 p 车道的

平均速度。

(2)小时断面平均速度

$$V_h = \sum(N_i * V_i) / \sum N_i = \sum \sum(N_{ip} * V_{ip}) / \sum \sum N_{ip}$$

其中, N_{ip} 为该小时内的时段 i 第 p 车道的交通量, V_{ip} 为该小时内的时段 i 第 p 个车道的平均速度。

事实上, 这种基于流量加权的平均速度计算方法, 还可以推广到观测站所观测的路段平均速度、线路的平均速度等计算中。

4.2 数据包接收器

实时数据包接收器是整个系统的核心子系统, 其关键技术是一个数据结构、3 个并行线程。

(1)数据结构

类 Session 用于异步处理每个观测站的 Socket 连接及接收数据, 主要成员有: 客户端 Socket 对象 _clientSocket, 用于接收器与观测站之间的通信; RecieveBuffer 数组用于接收 Socket 的字节流, _lastDataRecievedTime 则用于判断 Session 对象的活跃状态。

(2)并行处理线程

系统的 3 个并行线程均由 .NET 线程池(thread pool)统一管理^[8], 其功能及处理流程介绍如下:

1)Socket 侦听、连接与包接收线程: 循环侦听远程观测站的 Socket 连接请求, 为每个 Socket 连接建立 Session 对象实例, 调用 Socket 的异步方法读取数据, 并将格式合法的数据包保存到 DatagramQueue 队列中。

2)数据包解析与存储线程: 解析数据包队列中的数据包, 验证其合法性, 判断是否为重复包, 并存储分车道的时段数据、分方向的小时数据、日数据。同时, 将验证结果返回给观测站。

3)Socket 资源清理线程: 在一定的时间段(如 3min)如果某个 Session 对象没有接收到数据包, 则认为该连接已经断开, 系统强制关闭并释放 Socket 资源。

在 .NET 2.0 下, 并行线程之间的互斥与同步通过 Monitor 类和 ManualResetEvent 类实现^[8]。

4.3 运行监控子系统

该子系统与数据包接收器、数据包转发器运行在同一台计算机上, 工作原理如下:

(1)转发器定时通过 Socket 向接收器发送状态消息, 监控子系统定时通过 Socket 向接收器发送询问指令;

(2)接收器保存转发器最近的消息时间, 在接到询问指令时, 发送自己的状态消息及转发器的消息时间给监控系统;

(3)监控系统根据询问指令返回的结果判断接收器与转发器运行是否正常, 如果不正常, 则重新装载系统、重启计算机或关闭计算机。

4.4 查询与统计客户端

查询与统计客户端是一个 .NET 平台下基于 Web 服务^[9]的富客户端(RichClient), Web 服务验证服务请求者身份、返回查询请求数据, 富客户端显示数据、制作与输出报表。关键技术有:

(1)使用 DataSet 返回 Web 服务的查询结果集, 包括时段、小时、日、月或年的交通量数据;

(2)通过封装 Excel 的 COM 接口的程序集 Interop.Excel.dll, 富客户端可以方便地操作 Excel 工作簿(WorkBooks)和工作表(WorkSheets), 制作并输出 Excel 报表。

5 结束语

本文探讨的交通量调查数据处理系统由 5 个相互关联但独立运行的子系统组成, 系统的技术特点主要有: (1)基于 .NET Framework 2.0 并使用 Visual C# 2005 Express 开发工具; (2)多线程异步处理; (3)Socket 资源清理; (4)系统运行监控; (5)基于时序区间的判重算法; (6)推广的加权平均车速算法。

整个系统已于 2006 年 5 月 1 日前投入使用, 连续不停地运行表明, 系统具有良好的稳定性、健壮性与可伸缩性, 满足了交通部与湖南省公路管理部门的需求。下一步的工作将集中在数据库安全性、数据分布性、网络安全性、基于 GIS 的在线数据与图片发布等方面。

参考文献

- 1 交通部. 公路交通情况调查统计报表制度[EB/OL]. (2005-09). <http://jd.tpri.gov.cn/gzjl/bzgf.jsp>.
- 2 交通部规划研究院. 公路交通量调查设备与数据服务中心基础交通数据通讯协议(2006 年修订稿)[EB/OL]. (2006-03). <http://jd.tpri.gov.cn/gzjl/bzgf.jsp>.
- 3 于胜英. 进一步完善交通量和运输量统计体系建设为公路水路交通发展提供有力的信息支撑[EB/OL]. (2006-04). <http://jd.tpri.gov.cn/gzjl/2006jtljh.htm>.
- 4 Microsoft Inc. Visual C# 2005 Express[EB/OL]. (2005-11). <http://msdn.microsoft.com/vstudio/express/visualsharp>.
- 5 Sharp J. Visual C# 2005 从入门到精通[M]. 周靖, 译. 北京: 清华大学出版社, 2006-06.
- 6 Henderson K. SQL Server 存储过程、XML 和 HTML 高级指南[M]. 康博, 译. 北京: 清华大学出版社, 2002-10.
- 7 沈永木. 交通状况分析[EB/OL]. (2005-11). <http://jd.tpri.gov.cn/gzjl/jsjl.jsp>.
- 8 Archer T, Whitechapel A. C#技术揭秘[M]. 马朝晖, 译. 北京: 机械工业出版社, 2003-07.
- 9 Ferrara A, MacDonald M. NET Web 服务编程[M]. 天宏工作室, 译. 北京: 清华大学出版社, 2003-09.

(上接第 279 页)

参考文献

- 1 吴继华, 王诚. Altera FPGA/CPLD 设计[M]. 北京: 人民邮电出版社, 2005.
- 2 任爱锋, 初秀琴, 常存, 等. 基于 FPGA 的嵌入式系统设计[M].

西安: 西安电子科技大学出版社, 2004.

- 3 Altera. NiosII Processor Reference Handbook[EB/OL]. (2006-05). <http://www.altera.com>.
- 4 Altera. Altera Embedded Peripherals Handbook[EB/OL]. (2006-05). <http://www.altera.com>.