

一种分布式 SCADA 系统设计与实现

舒大有¹ 冯 暄¹ 卢 军² 郭本俊²

(四川省计算机研究院 成都 610041)¹ (成都信息工程学院软件工程学院 成都 610225)²

摘 要 传统 SCADA 是基于铜线构建的,一般用于小规模数据采集。随着工业信息化的进步,新型的 SCADA 系统正逐渐演变为基于 IP 的面向海量数据采集。在这样的变化中,SCADA 面临两个问题:(1)IP 化数据采集协议的自适应问题;(2)海量采集数据的分布式处理问题。针对上述问题,基于分布式数据管道的软总线技术和两级分布式调度机制,设计并实现了一种分布式 SCADA 系统,它可以有效地适用于 IP 化的海量 SCADA 系统。

关键词 SCADA,分布式,海量数据采集,协议自适应

中图分类号 TP393 **文献标识码** A

Design and Implementation of Distributed SCADA System

SHU Da-you¹ FENG Xuan¹ LU Jun² GUO Ben-jun²

(Sichuan Institute of Computer Sciences, Chengdu 610041, China)¹

(Software Engineering College, Chengdu University of Information Technology, Chengdu 610225, China)²

Abstract Traditional SCADA is based on the copper wire construction, commonly used in small-scale data collection. With the advancement of the industry informatization, the new SCADA system is gradually evolving into one kind of IP-based and facing massive data acquisition. In this change, SCADA is facing with two problems. One of them is self-adaptive problem of IP data acquisition agreement, and the other is distributed processing problem of mass data acquisition. According to above problems, this paper, based on the soft bus technology for distributed data pipeline and secondary distributed scheduling mechanism, designed and implemented a distributed SCADA, which can be effectively applied to IP-based mass SCADA system.

Keywords SCADA, Distributed, Large-scale data acquisition, Self-adaptive

1 引言

随着中国日益获得“世界工厂”的美誉,我国的制造业快速发展,一大批制造企业日益成长壮大,生产技术水平不断提升,竞争力与日俱增。

但我国制造业水平总体上还不高,特别是在工业生产的信息化应用方面,离工业化发达国家还有很大的距离。在工业生产领域,SCADA 是一种广泛应用的工业信息化系统。SCADA 简称工业数据采集与监视控制系统,可以将各类自动化采集和控制设备进行联网,把各类实时数据采集、加工、整理后送入大型实时数据库系统中,实现整个企业中各装置的生产运行监控、操作参数优化、调度指挥、在线故障诊断等,对提高企业生产管理水平、保障装置的安全稳定运行具有重要的意义。

目前,传统 SCADA 系统是基于铜线的,在向基于 IP 化的海量数据采集 SCADA 转变过程中,其面临两个突出的问题:

(1)数据采集协议的自适应^[8,10]问题。导致这一问题的原因是,基于 IP 的数据采集与传输协议数量非常多,协议之

间差别很大。在传感器的 IP 数据传输协议上,也没有统一的行业标准。因此,各种 IP 传感器可以使用 HTTP、FTP、SNMP、SSH、TELNET、MODBUS 等各种协议传输数据,同时传输数据的报文格式也不相同。这种差异性使得 SCADA 系统很难将各种各样的传感器接入到系统中,导致基于 IP 的 SCADA 系统面临巨大的推广压力。

(2)海量采集数据的分布式处理问题。SCADA 系统 IP 化的最主要的优点之一就是整个系统的容量几乎可以不受限制,可以轻松实现 10 万点计的大型数据采集与监视控制系统。这种大型系统同时带来以下问题:海量的采集数据如何在 SCADA 系统中高效地处理?如果不能解决这个问题,那么 SCADA 系统在大型企业中的应用将面临巨大的问题。

这两个问题制约了 SCADA 系统满足现代工业企业生产信息化的需求,越来越成为许多应用 SCADA 的企业进行高效、安全生产的信息化瓶颈。

本文针对 IP 化^[2]海量数据采集^[3,5] SCADA 软件框架进行研究,在研究过程中发现各种传输协议存在差异,传感器设备是多样的。本文提出:采用快速可配置方式,使用二级分布式调度、分布式数据管道作为 SCADA 系统的解决方案,增强

到稿日期:2012-10-11 返修日期:2013-03-01 本文受国家自然科学基金(61102076)资助。

舒大有(1956—),男,工程师,主要研究方向为计算机应用,E-mail:448938223@qq.com;冯 暄(1977—),男,高级工程师,主要研究方向为数据挖掘与软件工程;卢 军(1974—),男,教授,主要研究方向为普适计算、嵌入式等;郭本俊(1975—),男,讲师,主要研究方向为移动计算。

了协议的自适应性,为面向 SCADA 系统的通用软件系统设计提供了一种有效的解决方法。

2 IP 化海量数据采集框架^[4]

SCADA 系统中数据采集的最简单模型如图 1 所示。

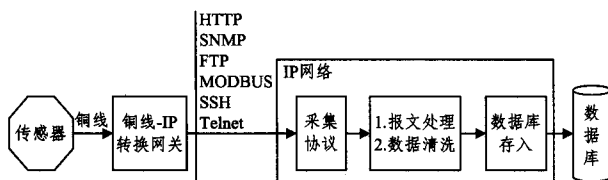


图 1 IP 化数据采集示意图

在图 1 所示模型中,可以看到传感器的值经过铜线—IP 转换网关后被转换为基于 IP 的上层协议的报文。由于基于 IP 的信息传输协议的多样性,在这个接口上的协议至少包括 HTTP、SNMP、FTP、MODBUS、SSH、TELNET 等。这个功能由图 1 中的“采集协议”模块完成。

另外,基于 IP 的传输协议,在上层报文上,格式也不尽相同。典型的报文格式包括:

- (1) 文字型格式;
- (2) 基于 XML 的报文。

对报文的解析与对数据的清理(例如单位转换、错误异常数据的订正)由图 1 中的“报文处理与数据清洗”模块完成。对已经经过了“报文处理与数据清洗”的数据,由“数据库存入”模块存放到数据库中。

图 1 所示的简化模型是仅用于展示 SCADA 系统中数据采集的最简单情况。本文提出 IP 化海量数据采集框架 IP—LSDMF(IP Large-Acale Data Mining Frame)在实际应用中必须考虑如下的问题:

(1) 由于不同的传感器采集模块使用的传输协议不一样,因此,采集协议模块必须能支持多种协议,并且还要能做到完全可配置,便于在不同现场灵活使用。

(2) 不同种类的传感器,不仅数据传输协议不一样,报文的格式也不一样,因此“报文处理与数据清洗”对于不同种类的传感器都是不一样的。所以,在系统中,需要为一类传感器配置一种“报文处理与数据清洗”程序。这需要整个采集框架灵活可定制,能方便地针对不同类型的传感器加入不同的“报文处理与数据清洗”模块。

(3) 为了支持海量的传感器信息采集,上述“采集协议”模块和“报文处理与数据清洗”都必须支持完整的分布式架构^[4,6,9],可以将负荷线性地分布式到不同的处理节点上,而对系统的整体功能没有影响。

针对上述对 IP 化海量数据采集框架的需求,本文设计与实现如图 2 所示的系统架构。在这个系统架构中:

(1) 采集协议模块支持多种协议,可以部署到不同的机器上。

(2) “报文处理与数据清洗^[6]”模块使用了软总线技术实现,可以灵活地使用任何语言编写,灵活地嵌入到系统,而不影响整个系统的功能。在实现软总线的时候,系统为了保证采集数据的传输性能,使用了高性能的分布式数据管道^[1]技术来实现,使任何支持文件访问操作的语言都可以编写“报文处理与数据清洗”功能模块,并可以自由地嵌入到整个系统上。

(3) “报文处理与数据清洗”模块可以分布到不同的计算机上。

(4) 系统中各个节点,完全使用统一的部署文件实现灵活的可配置功能。

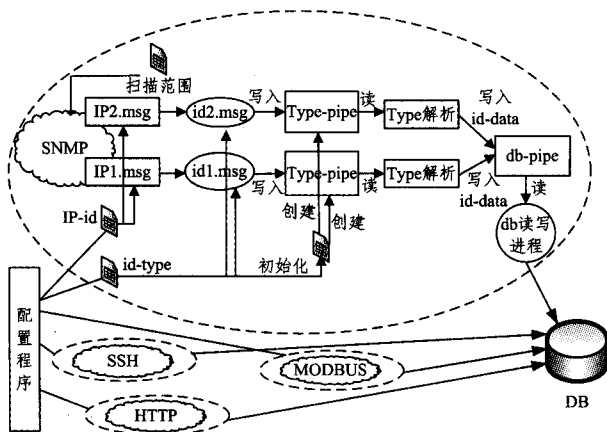


图 2 IP 化海量数据采集框架

3 报文解析模块嵌入

报文解析模块的嵌入是本系统研究的一个重要方面。

由于在一个大型 SCADA 系统中,传感器的种类可能达到上百种,因此系统支持使用不同语言、灵活多样的嵌入“报文处理与数据清洗^[6,9]”模块是系统设计的突出需求。

如图 3 所示,采集协议将采集的报文写入到分布式数据管道 1 中。对于一种类型的传感器,分布式数据管道 1 在系统中其名称(名称对应着访问途径)是固定的,因此,“报文处理与数据清洗”模块可以使用某类型传感器对应的分布式数据管道的名称去访问它。而在系统中,设计巧妙之处还在于,“报文处理与数据清洗”模块对于分布式数据管道的访问,使用了最简单的文件访问接口,例如 C 语言中的 open()、close()、read()、write() 功能。当“报文处理与数据清洗”模块通过文件访问接口获得了传感器报文后,就可以执行报文处理与数据清洗功能,然后再把最后得到的经过了处理的传感器的值使用文件访问接口写入到实现设定的分布式数据管道 2 中。在传感器的值被写入到分布式传感器 2 中后,系统中另外一个数据库写入模块,就可以从分布式数据管道 2 中读出传感器的值,写入到数据库中。

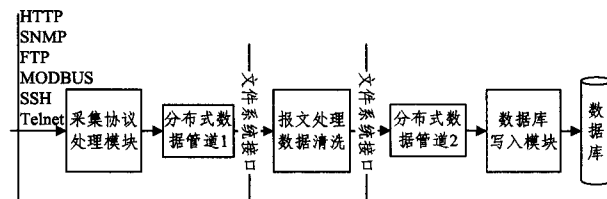


图 3 基于分布式数据管道技术的软总线技术示意图

分布式数据管道的设计与引入,使得本系统可以让用户使用任何支持文件访问接口的程序设计语言来设计“报文处理与数据清洗”模块,从而大大地减少了对用户开发“报文处理与数据清洗”模块的限制,提高了系统的灵活性。

4 海量数据采集

海量数据采集是 SCADA 系统的核心,其模型如图 4 所示。

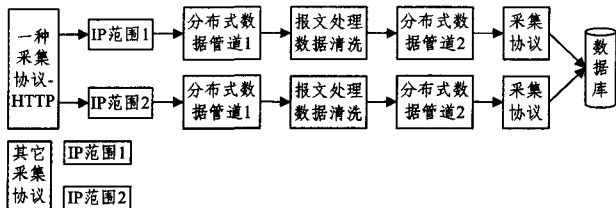


图4 二级分布式调度模式示意图

为了支持海量数据采集,本文需要设计如图4所示的二级分布式调度模式^[7],包括采集协议一级调度和IP范围二级调度。

(1)采集协议一级调度

SCADA系统实现了HTTP、SNMP、FTP、MODBUS、SSH、TELNET等常用协议,以适应传感器在IP网络上传输采集信息的不同类型协议。为保证系统透明性,采集协议通过POLL(轮询)系统配置协议—IP地址表格触发一级调度引擎。调度引擎根据系统配置将按照一种类型采集协议对应IP地址进行一级调度分配。

一级调度分配的设计与引入,将传感器按照大区域划分,为海量数据采集提供分布式实现基础。

(2)IP范围二级调度

IP范围二次调度是在采集协议一级调度基础上进行又一次任务分发,同样使用POLL(轮询)系统配置IP范围表格触发二级调度引擎。按照IP网络地址范围对传感器进行分组划分,部署到不同执行服务器节点,获取传感器报文信息,将报文信息写入到分布式数据管道中,通过“报文处理与数据清洗”模块获得了传感器报文后,执行报文处理与数据清洗功能,继而写入远端数据库。

二级分布式调度模式实现过程中,服务器节点承担传感器数据解析和数据入库任务,从而实现海量数据按照传感器类型与IP地址分布式采集,系统数据容量不受任何限制。

5 实验验证

本文所提出的基于IP化海量数据采集系统框架,通过了实验验证。在本实验中,构建一个SCADA系统数据采集实验环境,整个环境由多个不同类型IP传感器、传统传感器、服务器节点、数据库服务器、业务管理终端计算机构成。

在上述SCADA实验环境中,本文进行了如下实验验证:

(1)由业务管理终端计算机根据IP传感器类型和数量,快速配置部署SCADA系统,验证SCADA系统快速部署和数据采集订正能力。

(2)通过配置管理发送二级调度指令和规则,进行大批量数据采集、解析和传输,验证分布式数据采集处理能力。

(3)通过动态加入传统传感器和IP传感器采集处理数据,验证协议自适应和报文解析、传统传感器、各种类型新型IP传感器的全兼容。

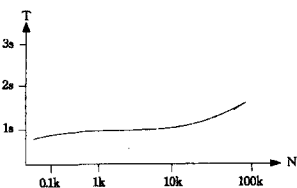


图5 响应时间与节点负荷线性对比图

图5示出SCADA系统海量数据采集处理能力、仿真响

应时间与节点负荷的线性对比。点数在0.1k(k为1000)到10k之间时响应时间在1s左右,点数在10k到100k海量处理时响应时间平滑靠近1.7s。

为进一步论证报文解析嵌入特性,表1给出以编程算法为依据在Linux Red Hat AS 5环境下的不同嵌入模块性能差异的结论。

表1 不同编程算法报文解析特性

编程算法	适应度	执行效率	部署难易度
PHP	高	高	易
C	高	高	易
Perl	较高	高	较易
Ruby	高	高	较易
Shell	高	高	高

通过上述实验,验证了本文提出的IP化海量数据采集系统可以稳定实现分布式海量数据采集处理、报文处理,以及数据清洗可采用不同编程语言(算法)灵活嵌入到系统。

结束语 本文提出的分布式SCADA系统具有如下的特点:

(1)全IP化的SCADA系统良好的拓展性、更高的系统容量,可以轻松地支持高达数十万节点,具有更高的拓展性与系统性能。

(2)基于分布式管道技术的协议解析模块嵌入技术,可快速部署SCADA系统,不需要停机。

在后续工作中,还有若干重要的问题需要进一步研究,主要包括:(1)传感器设备数量高速扩张,两级调度模式负载均衡和系统整体性能优化,在项目的推广过程中将重点研究。(2)现有协议匹配、自适应和解析是针对当前网络传输协议,对要使用的新型网络传输协议需自适应解析等问题,在后续工作中需要进一步研究。

参考文献

- [1] Papadopoulos S, Bakiras S, Papadias D. pCloud: A Distributed System for Practical PIR[J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(1): 115-127
- [2] 郭本俊,黄健,卢军,等.基于计算区域网的物联网普适计算框架模型[J].四川大学学报:自然科学版,2012,49(1):69
- [3] 宋欣,王翠荣.基于线性回归的无线传感器网络分布式数据采集优化策略[J].计算机学报,2007,35(3)
- [4] 梁梅,刘永强,袁明军.具有开放性的电能质量SCADA系统的设计与实现[J].电力系统保护与控制,2010,38(11)
- [5] 余卓煌,林辉,刘品.基于SCADA系统的面向海量数据的电力预测系统[J].计算机工程与设计,2010,31(5)
- [6] 曹建军,刁兴春,陈爽,等.数据清洗及其一般性系统框架[J].计算机科学,2012,39(z3)
- [7] 谢廷婷.一种基于分布式系统的队列多级调度算法[J].贵阳学院学报:自然科学版,2012,7(3)
- [8] 吴斌,毛新军,董孟高,等.自适应系统软件传感器设计与实现[J].计算机科学,2010,37(8)
- [9] Fan Wen-fei, Geerts F, Li Jian-zhong, et al. Discovering Conditional Functional Dependencies [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(5): 683-698
- [10] Xiang Xiao-jing, Wang Xin, Zhou Ze-hua. Self-Adaptive On-Demand Geographic Routing for Mobile Ad Hoc Networks[J]. IEEE Transactions on Mobile Computing, 2012, 11(9): 1572-1586