

RFID



YZLI 0890092152

RFID SHISHI
ZHONGJIANJIAN JISHU
实时中间件技术

廖国琼 著



西南交通大学出版社
[Http://press.swjtu.edu.cn](http://press.swjtu.edu.cn)

RFID SHISHI ZHONGJIANJIAN JISHU

实时中间件技术

廖国琼 著



西南交通大学出版社

[Http://press.swjtu.edu.cn](http://press.swjtu.edu.cn)

图书在版编目 (C I P) 数据

RFID 实时中间件技术 / 廖国琼著. —成都: 西南
交通大学出版社, 2010.6

ISBN 978-7-5643-0727-1

I. ①R… II. ①廖… III. ①无线电信号—射频—信
号识别 IV. ①TN911.23

中国版本图书馆 CIP 数据核字 (2010) 第 123320 号

RFID 实时中间件技术

廖国琼 著

责任编辑 祁素玲
特邀编辑 杨丽坤
封面设计 袁 飞
出版发行 西南交通大学出版社
(成都二环路北一段 111 号)
发行部电话 028-87600564 028-87600533
邮 编 610031
网 址 <http://press.swjtu.edu.cn>
印 刷 成都蓉军广告印务有限责任公司
成品尺寸 145 mm×210 mm
印 张 5.062 5
字 数 135 千字
版 次 2010 年 6 月第 1 版
印 次 2010 年 6 月第 1 次
书 号 ISBN 978-7-5643-0727-1
定 价 18.00 元

版权所有 盗版必究 举报电话: 028-87600562

前 言

射频识别 (Radio Frequency Identification, RFID) 技术是一种利用射频通信实现的非接触式自动识别技术。近年来,随着大规模集成电路、射频通信、信息安全等技术的发展,RFID 已开始广泛应用于物流、制造、交通运输、医疗、防伪、跟踪、设备和资产管理等公共信息服务行业。由于具有高速移动物体识别、多目标识别和非接触识别等特点,RFID 技术被认为是 21 世纪最有发展前途的信息技术之一。研究 RFID 技术和发展 RFID 产业对提升信息化水平、促进经济可持续发展和增强公共及国防安全等方面将产生深远影响。

作为 RFID 系统的神经中枢,RFID 中间件 (Middleware) 主要是对阅读器传来的与标签相关的事件、数据进行过滤、汇集和计算,以减少从阅读器传往企业应用的巨量原始数据,并抽象出有意义的信息供上层应用使用。本书主要研究 RFID 实时中间件关键技术,共分 9 章:第 1 章简要介绍 RFID 中间件的作用及研究现状;第 2 章在分析 RFID 数据流特征和 RFID 实时中间件的功能需求基础上,设计 RFID 实时中间件系统架构;第 3 章研究基于 EPCIS 标准的 RFID 时态对象数据模型,并讨论如何实现时态数据查询;第 4 章研究集中式环境下 RFID 数据流清洗方法;第 5 章研究分布式 RFID 数据流孤立点检测技术,以降低中心节点的通信量和计算负荷;第 6 章设计一种根据新到达事务截止期与事务价值计算动态优先级的策略,以提高实时事件处理的成功率;第 7 章研制支持实时与非实时查询的

RFID 混合存储架构以及混合存储引擎；第 8 章采用事件驱动技术开发基于 RFID 的零售超市系统；最后是全书总结与展望。

本书研究内容具体完整，结构合理，重视设计方法与实现技术介绍，特别是能够理论联系实际，将研究成果应用于实际系统开发。因此本书可作为高等院校高年级本科学生、研究生以及相关领域科技工作者的参考书。

本书研究成果是在已鉴定完成江西省教育厅重点科技项目研究成果的基础上进行总结及扩充而成。本书是团队成员共同努力的结果，在此，谨向团队中的狄国强教授、邓琨、舒雷等老师等表示衷心感谢。同时，李晶、吴凌琴等硕士研究生参与了本书部分工作，在此一并表示衷心感谢。

本书在成稿过程中，得到了万常选教授等的热心帮助和鼓励，在此向他们表示衷心感谢。

本书的出版得到了国家自然科学基金项目（60863016）、江西省自然科学基金（2008GQS0019）、江西省教育厅重点科技项目（GJJ08343）的支持和江西财经大学信息管理学院的资助，在此一并表示感谢。

感谢西南交通大学出版社的大力帮助，以及成都一鸣文化传播有限公司杨丽坤女士为本书出版所付出的辛勤劳动。

本书内容是作者研究所得，相关问题仍在研究与探索之中。由于水平有限，书中疏漏、偏颇及谬误之处在所难免，恳请专家、同仁和广大读者不吝指正。

廖国琼

2010 年 6 月于南昌

目 录

第 1 章 绪 论	(1)
1.1 RFID 中间件概述	(1)
1.2 国内外研究现状	(4)
1.3 RFID 标准组织介绍	(13)
1.4 RFID 技术发展趋势	(15)
1.5 本书主要内容及组织	(16)
第 2 章 RFID 实时中间件体系结构	(22)
2.1 RFID 数据流特征	(22)
2.2 RFID 实时中间件功能需求	(24)
2.3 RFID 实时中间件系统结构	(26)
2.4 本章小结	(29)
第 3 章 RFID 时态对象数据模型及查询	(30)
3.1 引言	(30)
3.2 EPCIS 标准及数据分类	(31)
3.3 RFID 时态对象模型	(33)
3.4 时态对象模型在 TOODM 上的实现	(34)
3.5 时态查询实例	(37)
3.6 本章小结	(39)

第 4 章	RFID 数据流清洗策略	(41)
4.1	引言	(41)
4.2	问题描述及定义	(43)
4.3	基于核密度的孤立点检测策略	(47)
4.4	概率的使用	(49)
4.5	窗口自适应调整策略	(53)
4.6	性能测试与分析	(54)
4.7	本章小结	(66)
第 5 章	基于距离的 RFID 孤立点检测方法	(68)
5.1	引言	(68)
5.2	问题描述及定义	(70)
5.3	基于距离的局部孤立点检测	(73)
5.4	基于近似估计的全局孤立点检测	(78)
5.5	性能测试与分析	(81)
5.6	本章小结	(85)
第 6 章	RFID 实时事件调度策略	(87)
6.1	引言	(87)
6.2	基于优先级表的优先级分配策略	(88)
6.3	基于优先级表的调度策略设计	(89)
6.4	性能测试及分析	(100)
6.5	本章小结	(102)
第 7 章	RFID 混合存储引擎	(104)
7.1	引言	(104)
7.2	混合存储模式分析	(105)
7.3	RFID 混合存储架构	(108)
7.4	本章小结	(114)

第 8 章 事件驱动 RFID 零售超市系统开发	(116)
8.1 引言	(116)
8.2 相关应用分析	(119)
8.3 事件驱动机制概述	(122)
8.4 RFID 事件处理框架	(124)
8.5 系统功能及系统结构	(126)
8.6 事件定义及处理	(129)
8.7 历史数据模型及查询	(142)
8.8 本章小结	(149)
第 9 章 总结与展望	(151)
9.1 工作总结	(151)
9.2 未来工作	(153)

第1章 绪论

射频识别(Radio Frequency Identification, RFID)技术是一种利用射频通信实现的非接触式自动识别技术,其最早应用可追溯到第二次世界大战中飞机的敌我目标识别。但是由于技术和成本原因,一直没有得到广泛应用。但近年来,随着大规模集成电路、射频通信、信息安全等技术的不断发展,RFID已开始广泛应用于物流、制造、交通、运输、医疗、防伪、跟踪、设备和资产管理等公共信息服务行业。由于具有高速移动物体识别、多目标识别和非接触识别等特点,RFID技术正显示出巨大的发展潜力与应用空间,被认为是21世纪最有发展前途的信息技术之一。研究RFID技术和发展RFID产业对提升信息化水平、促进经济可持续发展和增强公共及国防安全等方面将产生深远影响。

1.1 RFID中间件概述

一个完整的RFID系统通常由电子标签(Tag)、读写器(Reader)、RFID中间件(Middleware)和应用程序四个模块组成。在实际应用中,电子标签中保存有约定格式的电子数据,附在待识别物体的表面,而读写器通过天线发送出一定频率的射频信号。当标签进入磁场时产生感应电流从而获得能量,发送出自身编码等信息,被读写器读取并解码后送至主机中的RFID中间件。RFID中间件对接收到的信息进

行相关处理后提供高层信息给上层企业应用。因此,RFID 中间件是 RFID 应用系统中介于读写器与企业应用之间的支撑软件,其主要任务是对读写器传来与标签相关的事件和数据进行过滤、清洗、汇集、计算和抽象,以保证将准确和无噪声的数据传送给企业应用服务器,以及减少从读写器传往企业应用的数据量。因此,中间件是 RFID 系统的神经中枢^[1]。

RFID 中间件通过提供通用应用程序接口(Application Program Interface, API)实现不同应用程序与 RFID 读写器的连接。其好处是,当存储 RFID 标签数据的数据库软件或后端应用程序发生修改或更换时,或增加新的 RFID 读写器时,企业应用程序不需做任何修改。

一般而言,RFID 中间件应具有下列特征:

- 独立于架构。RFID 中间件独立并介于 RFID 读写器与后端应用程序之间,并且能与多个 RFID 读写器以及多个后端应用程序连接,以减轻架构与维护的复杂性。

- 遵循标准。EPC Global 是一个受业界委托成立的非营利性 RFID 标准组织,它提出了产品全球唯一识别码通用标准,即产品电子编码(Electronic Product Code, EPC)。当存放在 RFID 标签中的 EPC 被 RFID 读写器读出后,即可提供对 EPC 所代表的物品名称及相关信息进行追踪。该组织于 2005 年 9 月推出了 RFID 中间件应用软件标准 ALE (Application Level Events),得到了 SUN 等许多公司的支持。

- 数据流处理。RFID 主要目的在于将实体对象转换为信息环境下的虚拟对象,因此数据处理是 RFID 中间件最重要功能。RFID 中间件应能进行数据收集、过滤、整合与传递,将正确的对象信息传给企业应用系统。而且,RFID 数据是以流的形式自动、快速地产生的,需要积累起来以支持追踪和监控等 RFID 应用,因此 RFID 中间件应能处理按时间顺序到达的数据流。

- 实时处理。EPC Global 出台的第二代 RFID 读写器标准对读写器的阅读速率做出了规定:平均每秒钟应能识别 1800 个电子标签^[2]。也就是说,在高峰期,一个拥有 20 台读写器的配送中心每秒钟能阅读

36000 个事件。因此,RFID 中间件应能在规定时间内对大量 RFID 数据进行实时处理。

- 海量数据处理。RFID 设备的大规模部署会产生空前的海量数据。据统计,沃尔玛三天内会产生超过美国国家图书馆所包含的总体数据量^[3]。即使是部署适量的 RFID 设备公司一天之内也会产生十几亿字节的数据。

- 过滤与清洗。现有的 RFID 读写器在读取标签时会产生误差,如漏读、多读和脏读等。据统计,现实世界中部署的 RFID 读写器的漏读率达到 30% ~ 40%^[4-5]。不准确的数据对于高层的应用是无用的,有时甚至会导致做出错误的决策。因此,在数据进入系统之前要对其进行预处理。

- 挖掘语义信息。被观察的 RFID 数据通常携带有与上下文状态和背景知识相关的信息,而这些信息通常是隐含的,且与上层应用逻辑之间存在密切关系。例如,从产品 EPC 码可查出它的型号、价格及产地等;从读写器的位置可得知物品所在的位置等。原始 RFID 数据是一种低层的基础数据,必须上升为高层的业务逻辑数据并与应用相集成,才能真正发挥作用。

目前,国内市场的 RFID 中间件主要由 IBM、SAP、BEA、Microsoft 和 Oracle 等国际公司提供,它们大多是将 RFID 技术融入现有中间件产品体系中去^[6-9]。例如,IBM 在 WebSphere 架构中增加了 RFID 选项,提出了一种轻型 RFID 中间件架构,并使用 RFID 总线来代替复杂的层次结构。SAP 在 NetWeaver 中增加了 RFID 功能。Sun 公司的 RFID 中间件基于 EPC 网络基本构架设计,并支持 EPC Global ALE 软件标准,可过滤、聚集与处理大量来自 RFID 网络边缘的数据,为企业的各类应用提供流式商务信息。EPC 网络是由 Auto-ID 中心提出,它包括 EPC 编码、Savant、对象名字服务(Object Name Service, ONS)、EPC 信息服务(Information Services)、物理标识语言(Physical Markup Language, PML)等关键技术,得到了部分大学和许多企业的支持。Microsoft 所提出的 RFID 系统结构包括设备层、数据收集和管理层、事件

和工作流管理层、基于开放标准的服务接口层和应用与解决方案层等。Oracle 设计了内嵌在其应用服务器 10g 中的传感器边缘服务器,具有数据收集、分组、规则过滤、数据打包发送等功能。Sybase 公司推出了自己的 RFID 中间件,包含物理设备层、RFID 网络层、处理层、维持层、整合层、表示层和一些建模和工具组件集,而且还包含一些探查、警报、通告和管理的部分组件。Siemens RFID 中间件由读写器、事件管理器 and 数据服务器组成。该中间件基于动态关系 ER 数据模型设计,并提出了基于规则的事件过滤、转换和聚合机制。

国内方面,由于 RFID 技术进入时间较短,各方面的工作还处于起步阶段。虽然我国政府在国家“十一五”规划和 863 计划中,对 RFID 应用提供了政策和资金的支持,但与国际技术的发展相比,在很多方面还存在明显的差距。依托国家 863 计划“无线射频关键技术与开发”课题,中科院自动化所开发了 RFID 公共服务体系基础架构软件和血液、食品、药品可追溯管理中间件;华中科技大学开发了支持多通信平台的 RFID 中间件产品 Smarti;上海交通大学开发了面向商业物流的数据管理与集成中间件平台。此外,国内产品还包括东方励格公司的 LYNKO - ALE 中间件、清华同方的 ezRFID 中间件、ezONEez-Framework 基础应用套件等。虽然目前已经有了一些初具规模的 RFID 中间件产品,但与国外的 RFID 中间件产品相比,还显得不够成熟。

虽然上述公司、企业和研究机构都开始了 RFID 中间件研究、设计与开发,但是,大多数 RFID 中间件产品只是从 RFID 硬件设备收集原始数据,并进行简单过滤处理。当企业应用系统集成这些 RFID 中间件后,仍需要自己开发软件对 RFID 数据做进一步处理,因此大大降低了 RFID 系统应用集成效率。更重要的问题是,这些中间件产品很少考虑实时处理要求和缺乏实时处理能力^[10]。

1.2 国内外研究现状

上述特征给 RFID 中间件研究带来许多挑战,包括描述标签的动

态信息、数据压缩、过滤与清洗、RFID 事件模型及处理等。以下简要介绍 RFID 中间件领域国内外研究现状。

1. RFID 时空信息模型

RFID 标签动态信息要求建立适当的数据模型来表达时空信息,并在此基础上实现快速有效查询。已有 RFID 数据模型主要有:

- 英国剑桥 Auto-ID 中心实验室提出了用三维坐标来描述时态数据^[11],其中一维是时态坐标,另外两维是描述两个实体的动态关系,并用关系表来实现时序数据存储和查询。静态属性数据则采用 XML 模式进行存储和查询。这种数据模型比较适用于简单时序数据查询,而对于复杂查询和商业上下文信息查询效率不高。

- 西门子提出的数据模型包括四个静态实体^[12]:对象实体、传感器/读写器实体、位置实体和交易实体。这些实体之间通过状态变化(如位置变化,包含关系变化)和事件变化(如观察变化、交易项目变化)来影响对方,这些变化被表示成和时序(事件变化)相关的二元关系。该数据模型强调状态的变化和商业处理的时序语义,但破坏了数据的完整性。由于中间观测数据在状态变化过程中被清除了,因此难以完成历史数据查询。

- 韩国的 Tuyen Nguyen 等认为数据可分为主数据(静态数据)和事件数据(动态数据)两大类^[13],其中主数据包括类数据、实例数据和交易上下文数据;事件数据包括对象事件、聚合事件、数量事件和交易事件四种类型。它采用对象-关系模型,即底层采用关系表存储数据,而在此之上创建对象视图,支持从对象视图中查询相关信息。但该模型较为复杂,且查询速度较慢。

- SUN 公司提出的数据模型将每个静态实体存放在三个不同的表里^[14]:一个表存放实体自身的数据,包括标识符和通用属性;一个表存放实体特有的属性;一个表存放父子包含关系。动态数据以不同的日志形式存储,如包含日志、观察日志、事务日志和标签分配日志(标签分配的历史信息)等,且每个日志项都包含时间信息。该模型较

适合定位和运输等商业处理行为。

微软、IBM 和 Sybase 等公司虽然也推出了相应的产品,但未见针对 RFID 数据模型方面的研究。在国内,虽然已开始开发 RFID 中间件,也未见数据模型方面研究。

已有 RFID 中间件研究存在的不足主要有:标准不统一、缺乏通用有效的数据模型、产品及原型缺乏实时处理能力、查询效率低下、功能不完善、不支持复杂查询,等等。

2. RFID 数据一致性控制保证机制

RFID 数据库中的事务在读取时空数据时,通常要求数据值能够反映现实世界的当前状态,即数据值不能过“老”。这就要求系统应提供时间一致性控制机制,以保证在事件处理过程中满足数据的时空一致性要求。通常的做法是根据事务自身的“截止期”和“数据有效期”来确定事务的优先级,并根据优先级高低进行并发控制。主要方法有基于锁的并发控制协议和乐观并发控制协议。

(1) 基于锁的并发控制协议。

- 基于优先级两段锁协议(2PL - HP)^[15]。2PL - HP 就是在传统的 2PL 的基础上,用高优先级优先的冲突解决机制来保证高优先级事务不被低优先级事务所阻塞。该算法不但能够避免死锁,而且解决了优先级倒置问题。但存在一个严重缺点就是可能导致重启,这将会严重影响系统的性能。

- 优先级继承协议(2PL - PI)^[16]。当优先级倒置发生时,持有锁的低优先级事务将以等待锁事务中的最高优先级执行,直到它释放锁。该策略解决了死锁和优先级倒置的问题,但是也阻止了中等优先级事务抢占锁,从而可能引起这些事务错失截止期。

- 优先级顶协议(2PL - PC)^[17]。该协议为每个关键段分配一个优先权顶,记录可能进入该关键段任务的最高优先级。然而这种机制需要每一事务所存取数据对象的优先级知识,比较保守。

(2) 乐观并发控制协议。

在乐观并发控制协议中,事务不受阻碍地执行,直到到达其提交点,再验证是否发生数据不一致。

● OCC - BC 协议^[18]。该协议是一个优先级无关协议,它要求所有与正在验证事务发生冲突的事务全部重启,从而保证正在验证事务的提交。若一个事务最终将被夭折,那么它不会导致其他事务重启,其性能优于两段锁协议。

● OPT - SACRIFICE 协议^[19]:是一种采用优先级牺牲策略来消解冲突的乐观并发控制协议,但是会引起事务牺牲问题,降低系统性能。

● OPT - WAIT^[19]:该方法采用优先级等待机制,大大提高了具有较高优先级事务满足截止期的机会,但也存在浪费的牺牲问题。

● WAIT - 50 协议^[20]:是一种优先级相关协议,使具有较高优先级的事务首先满足截止期并且没有无效重启。WAIT - 50 协议可以看作 OPT - WAIT 和 OCC - BC 两个协议的折中。

3. RFID 数据压缩

RFID 数据量通常是海量级的,且传输 RFID 数据需要更多的网络开销,一般网络环境很难满足。海量数据会导致系统性能下降,如降低存取速度、增加传输时间、降低查询效率和实时处理能力。因此,对海量 RFID 数据进行压缩显得尤为重要。

传统的压缩方法分为通用数据压缩和多媒体数据压缩两种。通用数据压缩又包括基于统计模型的压缩技术和基于字典模型的压缩技术。基于统计模型的压缩技术主要有 Huffman 编码和算术编码。基于字典模型的压缩技术有 LZ77, LZ78 和 LZW。多媒体数据压缩包括图像压缩和音频视频压缩,主要有二值图像 CCITT JBIG 等,灰度图像、FELICS JPEG 等,彩色图像 RLE 编码 JPEG 等,矢量图像 PostScript WMF CAD 等。

目前面向 RFID 的数据压缩方法主要有位置压缩(location compression)和包压缩(containment compression)^[21]。位置压缩是指当多个 RFID 对象在同一个位置停留时间超过一个周期,会被读写器重复

检测,但产生的事件信息与上次相同,故在这段时间内可以将这一系列的对象浓缩成一个对象,减少事件数据量。包压缩是指当多个 RFID 对象被放进同一个容器对象中以组的方式移动时,可只用容器的标签来同时表示容器和里面对象的位置。

4. RFID 数据清洗技术

在 RFID 系统中,由于射频信号干扰和标签周围环境等诸多原因,RFID 读写器产生的数据流通常是不可靠的,严重地影响了 RFID 技术的广泛应用。因此,系统必须增加对 RFID 数据的预处理环节,将脏数据过滤掉,使读写器读取来的数据,无论从时间上还是从数值上,尽可能地接近真实状态和值。

在众多的数据管理系统中,数据清洗都是大家关注的一个大问题。在数据仓库中,针对预定义的任务,可采用离线、集中、循环、交互式等处理方式^[21]。但由于 RFID 数据具有时序特性,其处理方式只能采用在线形式。RFID 数据管理中,脏数据通常包括三种形式:漏读、错读和数据冗余。

针对漏读和错读情况,大部分中间件系统采用平滑窗口进行清洗,即设定一个大小固定的窗口来填补漏读数据。但由于窗口大小固定不变,致使窗口过小会产生拒真数据,窗口过大会产生纳伪数据。因此,大小固定的窗口不能根据数据的特点精确地填补数据。美国加州大学 Berkeley 分校在文献[5]中提出了一种基于概率模型的 SMURF 自适应数据清洗方法。它从标签的完整性和动态变化性两个方面进行考虑,将 RFID 数据流看作统计学中的随机事件,然后根据概率来适应性地调整窗口大小,有效地保证了数据的准确性,因此相比于静态窗口填补算法,该方法具有明显的优势。但它只解决了单个读写器情形,没有考虑多个读写器。文献[23]基于 SQL 查询模型提出了一种 ESP(Extensible Sensor Stream Processing)机制,它根据数据具有时间相关性和空间相关性对数据进行清洗,考虑范围比较宽广,可用于清洗来自不同读写器的数据,并且能针对各类型错误数据进行清

洗。该机制提供了一个描述性的查询处理工具,便于处理数据和建模。但在处理的过程中,如何设置时间粒度和空间粒度、一旦设置是否可改变及如何改变等问题仍在研究中。文献[24]提出了以机器学习为背景的清洗算法,它针对大规模 RFID 数据集提出了一系列数据清洗策略,并分析了各个策略对应的清洗开销,最后给出了优化算法。文献[25]针对不同的数据清洗类型提出了相应的清洗方案。针对漏读现象,提出了通过设置高的读写频率来解决;针对多读现象,该算法提出预先设定一窗口大小,统计在该段时间内各个标签被读到的次数,如果某个标签的统计次数小于阈值,则认为相关数据是多读数据;对于重复数据,将相隔时间太短的不同或相似数据消除,只留下一个数据;解决时间戳乱序问题,主要是将接收到的数据进行缓存排序,预先设定一个最大延迟时间,在此时间之后到达的数据给予删除。该算法针对性地解决了多种数据清洗问题,但处理数据的延迟较大,导致效率不高,而且只考虑了单个读写器探测标签的情况,还有许多地方需要改进。针对数据冗余情况,文献[26]提出了一个 RRE 算法。它是一个随机、分散、局部近似的算法,可以消除读写器层次的冗余。但该算法是在读写器固定不动的假设下进行处理,这个假设在应用中通常是不能保证的,如供应链环境中,为了优化商业处理,读写器的位置通常是不断变化的。

5. RFID 事件处理模型

事件是“指示某种行为的信息”,包括系统产生的消息,系统状态的改变,任务的开始和结束等。事件在形式上类似于消息,如都包含数据,不同之处在于事件直接指示某些行为的发生。事件可以有多种分类方法。例如,根据事件语义的聚合程度不同,可分为简单事件和复杂事件;从分层的角度划分,可分为底层事件和高层事件:底层事件是系统产生的实际事件,而高层事件是由用户自定义的。从系统响应角度划分,事件可分为常规事件、异常事件等。因此,建立合适的 RFID 事件模型,提高处理效率显得非常重要。已有的事件处理模型