

RFID 复杂事件处理关键技术的研究与改进

王庆生, 魏晓伟

(太原理工大学 计算机科学与技术学院, 山西 太原 030024)

**摘 要:**RFID 复杂事件处理是一个新兴的技术领域,它用来处理大量的简单事件,并从中整理出有价值的事件。RFID 事件和传统的事件相比较具有海量性、空间性和时间性、数据不准确性等特征。文中在分析 RFID 数据特点的基础上,对 RFID 复杂事件处理的关键技术进行研究和改进,主要介绍 RFID 数据的清洗和事件检测技术。对于 RFID 数据清洗部分,提出了多层次过滤的方法使得到的数据更接近真实情况,而事件检测方面则提出了局部检测和全局检测相结合的方法对相关数据进行检测以得到更有意义的数据供上层应用使用。最后,对 RFID 复杂事件处理的发展趋势做出展望。

**关键词:**无线射频识别;复杂事件处理;数据清洗;事件检测

中图分类号:TP391 文献标识码:A 文章编号:1673-629X(2012)01-0045-04

Research and Improvement on Key Technology of RFID  
Complex Event Processing

WANG Qing-sheng, WEI Xiao-wei

(School of Computer Science and Technology, Taiyuan University of Technology,  
Taiyuan 030024, China)

**Abstract:** RFID complex event processing is an emerging technology to handle a large number of simple events and sort out the valuable events. Compared to the traditional events, RFID events have the following characteristics: large volume, temporality and spatiality, inaccuracy etc. Based on the analysis of the characteristics of RFID data, research and improve the key technology of RFID complex event processing, focus on RFID data cleaning technology and event detection technology. For the RFID data cleaning, a multi-level filtering approach, which can obtain data closer to the real situation, has been proposed. And for event detection, local detection and global detection has been combined to get more meaningful data for the upper application. Finally, give an outlook of development trend of RFID complex event processing.

**Key words:** RFID; complex event processing; data cleaning; event detection

0 引 言

无线射频识别 (Radio Frequency Identification, RFID) 技术是一种非接触式的自动识别和数据获取技术,利用无线射频信号实现物体识别和信息传递<sup>[1]</sup>。RFID 的工作原理:首先阅读器向标签发送能量,然后标签返回相应数据,最后阅读器收到数据后解码并将数据传送给主机。与其他识别技术相比,RFID 技术具有低成本、寿命长、体积小、容量大和可重复使用等特点,所以在很多领域(国防、医疗、零售、物流等)显示出了广阔的应用前景,因此被誉为 21 世纪最有潜力的技术之一。

典型的 RFID 应用系统结构<sup>[2]</sup>如图 1 所示。

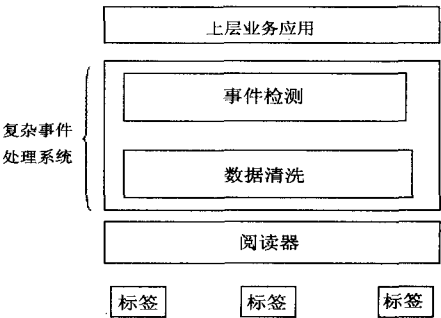


图 1 典型的 RFID 应用系统结构

复杂事件处理是 RFID 中间件的核心功能。通过对 RFID 事件层层过滤和归并,将底层的 RFID 原始数据汇聚成具有真实意义的抽象事件,从而满足上层业务的需要。虽然在 RFID 复杂事件处理方面已经有了不少研究,但是,目前 RFID 复杂事件处理仍然存在很多问题,主要是数据清洗和事件检测两方面存在问题。

收稿日期:2011-06-23;修回日期:2011-09-26  
基金项目:山西省科技攻关项目(2007031121-05-02)  
作者简介:王庆生(1962-),男,硕士,副教授,研究方向为 IPv6、网络安全、嵌入式系统;魏晓伟(1985-),女,河北石家庄人,硕士研究生,研究方向为网络安全。

文中将主要就 RFID 数据清洗和事件检测量方面的相关内容做一些详细介绍。

## 1 RFID 基本概念介绍

### 1.1 RFID 数据

对 RFID 数据进行归纳总结可以得出 RFID 数据具有:

(1) 实时性和关联性特点: RFID 源数据是标签的实时状态观测结果。源数据包含观测的具体时间、标签位置以及所处状态等数据。如货物出入库等信息。显而易见,任何一个 RFID 数据都不是独立存在的,而是相互关联的。时间关联指事件发生的前后顺序,空间关联则指事件的发展轨迹,两者结合清晰地表达了事件的变化过程。可见,传统的数据处理方法不能很好地管理这些数据。

(2) 语义丰富性特点: 通常标签都携带与背景知识及上下文状态相关的信息,它们与上层应用有密切的关系,但是这些信息是隐含的,所以,必须利用一些规则将这些信息的潜在含义表达出来以便被上层业务使用。如可以根据产品的 ID 号查出它的基本信息,如产地、型号、价格等。RFID 源数据是一种较低级的基础数据,所以必须依据一定的规则构建成语义丰富的业务逻辑数据才能真正地发挥作用。

(3) 不准确性和异构性特点: 由于射频干扰、外界环境影响等原因,现有 RFID 阅读器存在读取误差,其中漏读、多读和脏数据是最常见的三种。除了以上三种情况,还可能出现一个阅读器对同一个标签识别出的数据信息不相同的错误,这取决于阅读器的协议类型。例如,在某一单位入口,阅读器既能读取到人员信息也能识别人员携带物品的相关信息。显而易见,这些信息具有不同的性质,当然所对应的语义也不会相同。所以必须经过相应的处理才能得到上层应用所需数据信息。

(4) 流特性、海量和批量特点: RFID 数据以流的形式产生;因此需要积累以后用于跟踪和监控应用。除此之外,RFID 数据具有批量性特点,也就是说多个对象可能会被集中起来观察。例如,对货车进行登记时,会读取到大量数据。不难想象,在 RFID 设备大规模使用的同时,将生成空前的海量数据。目前,阅读器每秒可捕获 120 个到 400 个标签数据。对于一个部署有 100 个阅读器的中型仓储,每秒可产生 1.2~4 万条数据,若每条数据占 20 字节,则每天产生 1.6 GB~60 GB<sup>[3-5]</sup>。可见数据量非常庞大,因此,急需有一套切实可行的数据处理方法。

### 1.2 RFID 复杂事件处理基本概念

事件: 状态的一次有意义的改变,是指某种行为信

息,为硬件设备与软件系统间数据传送的单位。事件具有可嵌套性、可组合性和时序性特征。计算机系统的事件包括系统产生的消息、任务的开始和终止、系统状态改变等。RFID 系统事件分为原子事件(简单事件)和复杂事件。

原子事件(简单事件): 指读写器与标签间的一次数据交互。这些事件具有“原子性”,即事件在某时刻只有发生或不发生两种状态。

复杂事件: 原子事件或复杂事件按照某种运算规则组合而成的事件。

复杂事件处理是 RFID 中间件的核心功能之一,它主要是经过多层次的过滤和归并将底层 RFID 数据聚合成含有业务信息的高级事件,从而达到从大量的原子事件中获取有用信息的目的。

复杂事件处理技术主要实现的功能是:

1) 在大量的数据流中,快速找到用户需要的事件,并从中过滤出来。

2) 概括总结底层事件,并将它们抽象成更有意义的高层事件,以供上层业务应用使用。

3) 对分布不同的子系统及不同事件之间的因果关系进行查看。

4) 在系统操作层的任何抽象监控事件日志自动监测有违背关键请求的现象(如安全现象)<sup>[6]</sup>。

图 2 为复杂事件处理工作示意图。

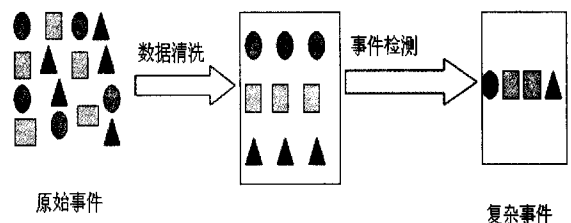


图 2 复杂事件处理工作示意图

## 2 数据清洗技术

当前限制 RFID 技术不能广泛采用的一个主要原因是 RFID 阅读器产生的数据不够精确。在 RFID 系统中,因为射频干扰、标签读取结构等原因的存在,读取到的标签数据可信度较低,漏读、不准确读问题在 RFID 应用中异常普遍,尤其是对于那些低成本、低计算能力的硬件以及无线通信的情况,漏读问题更为严重,部分有问题的阅读器更是会频繁导致读取数据丢失。不难想象,大量不可靠的数据对于高层处理应用毫无意义。因此,在系统中增加 RFID 数据的预处理是必不可少的。经过预处理后,无论从时间上还是数据的准确性上,读取的数据都更接近真实的数据。

阅读器从标签读取的大量数据中包含冗余信息,如果不经清洗直接传送给上层业务应用,不但增加

上层应用的负担,还可能淹没一些有用信息,更甚至造成网络传输瘫痪。因此,必须在事件信息源头进行清洗处理,以减少相应的负面影响。造成 RFID 源数据不准确的主要原因如下:

(1)漏读:标签在阅读范围,却没有被读取。在实际应用中,某个阅读器在同一时间段内同时读取多个标签,可能导致某些标签被漏读。此类错误也称为拒真(false negative)。

(2)脏数据:标签在阅读范围,阅读器感应到标签,但其读取到的标签数据不正确。

(3)多读:标签在阅读范围外,标签仍然被读取。在实际应用中,为了提高读取率,常常部署多个阅读器,可能出现同一标签同一时刻被多个阅读器读取并上报。除此之外也可能发生一个标签被同一阅读器多次读取的情况。此类错误也称为纳伪(false positive)。

针对 RFID 数据的特点,文中提出了可以完成 RFID 数据清洗功能的过滤模块。数据过滤模块可以有效地减少数据冗余、提高数据准确性。具体的解决方案是设置各种过滤器从而达到过滤掉冗余数据的效果。对大多数情况来说,一个 RFID 系统需要部署多个过滤器,从而满足多种不同的应用需求。过滤主要有两种类型:基于读写器过滤、基于标签和数据过滤<sup>[7,8]</sup>。根据现阶段需求分析,文中将采用分层设计思想,设置3层过滤器对输入的事件进行层层过滤,最终得到上层应用所需信息数据。

设置的三层过滤器是:

(1)冗余事件过滤器:阅读器将读取的海量的 EPC 标签数据传送给 RFID 中间件系统,这些数据存在大量多读、漏读和错读的信息。在传送数据给上层业务应用之前应该对标签数据过滤去重,从而减少系统的负担。所以冗余事件过滤器是取得有效事件的基础。

(2)EPC 码过滤器:该过滤器负责过滤出用户指定类型的标签,即对标签 EPC 编码中的 ID 号进行比较。当用户想知道某个产品的一些具体信息时就可以使用此种类型的过滤器进行筛选和过滤。

(3)时间过滤器:此类型过滤器可以过滤出一个具体时段的标签,该过滤器主要是根据时间记录进行过滤。如用户只想知道某个时间段的具体情况时,就可选此过滤器。

可以看出,在实际过滤过程中,这三种过滤器可以按顺序自由组合形成不同的过滤器组,这样原始标签数据就可以通过过滤器组进行过滤。经过过滤器组后原始的 RFID 数据已经基本上接近真实数据了,解决了上述提到的问题,为接下来的事件检测环节提供了

良好的数据基础。

### 3 事件检测

目前复杂事件检测方法分为两类:以数据为中心和以事件为中心<sup>[9]</sup>。

在 RFID 事件处理中最早被采用的是以数据为中心的复杂事件检测方法。此检测方法中,在数据库技术的基础上对数据进行建模,并且将相关数据保存到数据库中。然而,由于本身性能的限制,该方法的使用范围比较狭窄。它只能对存储在数据库中的数据进行查询或在线查询低速事件,却不能对大量的高速事件进行连续查询。

以事件为中心的复杂事件检测方法相对来说是一种比较新的 RFID 复杂事件检测技术。它直接对事件进行检测和处理,这样做的结果是获得了较高的效率,从而可以处理更多更复杂的事件。

#### 3.1 目前常用的复杂事件检测模型

目前常用的复杂事件检测模型有如下几种:

##### 1) 自动机复杂事件检测模型。

简单的复杂事件表达式与正则表达式相似。在自动机模型中,当参与复杂事件构成的原子事件出现时,自动机的状态将会发生改变,当自动机的状态可以被接受时,可以确定发生了复合事件。但是,在简单的自动机模型中存在这样一个问题,即某个基本事件匹配过之后便不能再被访问。因此可以构建扩展的自动机模型解决上述问题:引入可以保存基本事件间时间和数据关系的数据结构或者在自动机跃迁谓词中加入更加复杂的时间或数值条件限制。

##### 2) 匹配树复杂事件检测模型。

在此模型中通过树结构实现复合事件的聚合。匹配树的叶节点表示基本事件,中间节点则表示复合事件,根节点则为意义更复杂的事件。

##### 3) 有向图复杂事件检测模型。

有向图模型使用有向无环图来达到复杂事件检测的目的。在模型中,节点表示事件并且节点会对相关事件的引用进行标记,即节点本身也带有规则,当节点事件发生时,其所带规则被触发,边表示事件的合成规则。可以看出,基于有向图的检测模型与基于匹配树的检测模型类似。

##### 4) 基于 Petri 网的复杂事件检测模型。

通过构建 Petri 网进行复杂事件检测。在 Petri 网模型中,输入基本事件,输出复合事件。此模型中,跃迁守护函数成立将引发跃迁,与此同时会标记跃迁节点的位置。复合事件发生即序列中最后的节点被标记。

虽然每种模型都有了比较成熟的应用,如自动机

模型在事件流处理问题上比较成熟,而匹配树模型对事件的层次关系表示比较自然,然而在实际应用过程中上述几种模型仍有一定的局限性。如 Petri 网和自动机模型<sup>[10,11]</sup>只匹配顺序到达的简单事件,而匹配树和有向图模型又不考虑基本事件的顺序或时序距离,鉴于此,文中提出了一种具有更好性能和灵活性的分层复杂事件检测方法。

### 3.2 分层复杂事件检测

定义 1 ANY 运算符:从输入的事件中选择在  $t$  时刻发生的任意事件并输出。

定义 2 SEQ 运算符:从输入的事件中选择特定的事件并输出。

定义 3 AND 运算符:输入的 2 个事件在特定时间同时发生时,则在此运算符下 2 个事件将无序输出。

层次复合事件检测模型中,层次复合事件用树结构表示,叶节点表示输入的原子事件,上层节点表示相应的操作符。如,层次复杂事件  $SEQ(D, AND(SEQ(A, B), C))$ , 可表示为如图 3 所示的结构:

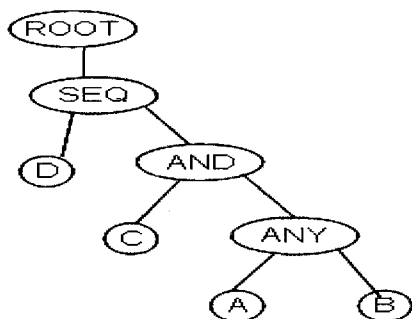


图 3 层次复杂事件的树结构图

在分层检测模型中采用局部检测和全局检测相结合的方式,这样可以有效地提高检测的性能指标。局部检测即按正常的树形结构逐步结合以得到意思表达更为精准的复杂事件。全局检测即将整个检测过程分为不同的分区,得到分区结果后再进行融合。多级别的并行分层事件检测有效地提高了系统的检测性能,为上层的业务应用提供了所需数据。

## 4 结束语

复杂事件处理技术可以用在众多领域中,尤其是随着计算机技术和网络技术的发展,大规模的复杂系统的应用越来越广泛,以前的数据处理方式已经不能

满足需要,这就更能显示出复杂事件处理的优越性。

RFID 事件处理是 RFID 技术的重要研究内容,涉及到数据库技术、人工智能技术、实时技术、嵌入式技术、普适计算技术、数理统计理论、现代信号处理与分析理论等领域<sup>[12]</sup>。文中对 RFID 复杂事件处理中的关键技术数据清洗方法和事件检测方法进行了归纳和总结。可以看出,RFID 的普及给 RFID 的数据管理方面的研究提供了新的机遇并且随着大规模通讯系统的发展,这种技术会有更广阔的发展前景,但是,与此同时也看到了研究支持复杂应用的 RFID 处理技术还面临很多挑战,所以必需整合各种可利用资源促进相关技术快速发展。

### 参考文献:

- [1] 彭小娟,刘世安,熊春如,等.复杂事件处理在大规模 RFID 数据通信中的应用研究[J].化工自动化及仪表,2009,36(4):76-79.
- [2] 赵黎.RFID 中间件事件管理系统的设计与实现[D].武汉:华中科技大学,2006.
- [3] 尹方鸣,康慕宁,陈群,等.基于内存受限的 RFID 复杂事件处理优化算法[J].计算机应用研究,2009,26(8):2864-2867.
- [4] 谷峪,于戈,张天成.RFID 复杂事件处理技术[J].计算机科学与探索,2007,1(3):255-267.
- [5] 张晓鹏,李建民,林振荣.复杂事件处理技术在 RFID 中间件中的研究[J].微计算机信息,2009,25(9-2):145-147.
- [6] Jeffery R, Garofalakis M, Franklin M J. Adaptive cleaning for RFID data streams[C]//Proc of VLDB. [s. l.]: [s. n.], 2006:163-174.
- [7] 蒋邵刚,谭杰.RFID 中间件数据处理与过滤方法的研究[J].计算机应用,2008,28(10):2613-2615.
- [8] Bai Y, Wang F S, Liu P Y. Efficiently filtering RFID data streams[C]//Proc of the 1st Int VLDB Workshop on Clean Databases. [s. l.]: [s. n.], 2006:50-57.
- [9] 金功联.基于复杂事件处理的 RFID 中间件的研究与设计[D].杭州:浙江工商大学,2009.
- [10] 王永恒,杨圣洪,郭波.高效的射频识别数据流层次复杂事件检测[J].计算机工程,2010,36(6):84-88.
- [11] Jin Xingyi, Lee Xiaodong, Kong Ning, et al. Efficient Complex Event Processing over RFID Data Stream[C]//Proc of ICIS'08. [s. l.]: IEEE Press, 2008.
- [12] 张乐.RFID 中间件数据处理研究与开发[D].上海:上海交通大学,2008.

## 计算机技术与发展友情提示:

本刊为中国科技核心期刊,中国科技论文统计源期刊。《中国核心期刊数据库收录期刊》、《中国学术期刊综合评价数据库统计源期刊》、《中国期刊全文数据库收录期刊》、《万方数据资源系统数字化期刊群上网期刊》、《中国学术期刊(光盘版)》。不愿意通过上述媒体发行者,请在来稿首页注明。