

RFID 中间件数据过滤方法研究

Research on Data Filtering for RFID Middleware

尚明¹, 蒋泰², 李立宪¹

SHANG Ming¹, JIANG Tai², LI Li-xian¹

(1. 桂林电子科技大学计算机科学与工程学院, 广西桂林 541004; 2. 广西瀚特信息产业股份有限公司, 广西桂林 541004)

(1. School of Computer Science and Engineering of Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China; 2. Guangxi Hunter Information Industry Co., Ltd., Guilin, Guangxi, 541004, China)

摘要:【目的】探讨对 RFID 数据冗余、脏读以及漏读数据进行过滤, 整理出上层应用所需要的有效事件的方法。【方法】运用基于时间和次数阈值、数据校验、TDS 数据解析以及基于伯努利概率的动态时间窗等方法进行研究设计。【结果】提出 3 种数据过滤器: 冗余过滤器用于处理多读数据, 平滑过滤器用于处理脏读数据, 基于伯努利的动态自适应平滑算法的漏读数据处理器处理漏读数据, 将这 3 种过滤器依次组合, 组成数据过滤模型, 最终构成数据过滤方法。【结论】运用这 3 种过滤器组成的数据过滤模型, 能够很好的对 RFID 原始数据进行过滤, 获得上层应用所需要的有效事件。

关键词: RFID 中间件 多读数据处理 脏读数据处理 自适应漏读处理

中图分类号: TP39 **文献标识码:** A **文章编号:** 1002-7378(2014)01-0047-04

Abstract:【Objective】This paper illustrates a data filtering method to filter redundancy data, dirty read and missed readings of RFID middleware and sort out the effective incident for the upper layer system.【Method】A time-based and number-based filter, a CRC and TDS data analysis filter, and a dynamic time window filter based on Bernoulli probability are used to established the method.【Result】Three filters was constructed, which are redundant data filter for more readings, smoothing filter for dirty readings and dynamic time window filter based on Bernoulli model for missed readings. Finally, the three filters were combined and the RFID data filtering method was proposed.【Conclusion】This data filtering is effective in filtering of RFID raw data and obtaining effective incident for the upper layer system.

Key words: RFID middleware, multiread data processing, dirty read data processing, adaptive leakage reading process

【研究意义】射频识别(Radio Frequency Identification, RFID)系统是物联网的应用形式之一。一个完整的 RFID 系统由 RFID 标签、读写器、RFID 中间件和应用软件组成。其中, RFID 中间件承担数据过滤、聚合、传输等重要任务, 它的研发对消除 RFID

应用瓶颈有重要意义。【前人研究进展】现有的对 RFID 中间件数据处理的方法主要分为以下两类^[1]: 1) 建立事件列表类。对每一个新到标签数据进行检测, 假如是新标签加进相应列表中, 且该标签在列表中已存在, 则仅更新对应标签的时间等状态数据, 而不新建标签数据记录, 以达到去除重复信息的作用。2) 事件编码类。对标签状态的改变进行编码, 标签出现编码为 0, 标签消失编码为 1。然后加进计时器机制, 对计时器有效时间内的同一标签的状态跳变进行忽略, 从而在状态定义和时间维度 2 个方

收稿日期: 2013-11-21

修回日期: 2013-12-12

作者简介: 尚明(1987-), 男, 硕士研究生, 主要从事基于 RFID 的企业信息管理系统研究。

面对数据进行去重化。【本研究切入点】结合以上两种过滤方法,从时间维度、次数维度以及建立事件列表的方面着手建立 RFID 中间件数据过滤方法。【拟解决的关键问题】通过对 RFID 数据冗余、错误以及漏读数据进行过滤,整理出上层应用所需要的有效事件。

1 RFID 中间件数据特点分析

RFID 数据具有实时性、不确定性以及大数据性等特点^[2],其中的不确定性主要是由多读、脏读以及漏读引起的^[3]。因此要降低上面3种原因造成的数据不确定性,需要针对这3种因素设计过滤方法。首先,设置标签数据的基本数据结构为:

```
Struct TagData1 {
String TagID; //标签 ID
String ReaderID; //读写器 ID
Date ReadTime; //读取 Tag 时间
}
```

再针对图1模型的过滤模型,设计数据过滤方法。

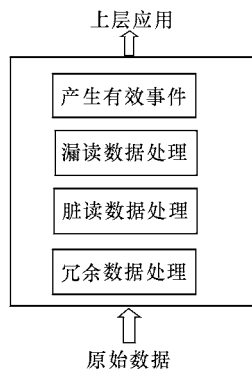


图1 RFID 数据过滤模型

2 RFID 中间件数据过滤器设计

2.1 多读数据处理——冗余过滤器

RFID 读写器读取速度快,会产生大量的冗余数据,增加上层应用系统的负担,因此需进行多读数据的清理。本文利用哈希表来进行数据的冗余过滤^[4]。

由于 RFID 系统对实时性要求很高,所以,冗余过滤器需要满足两点要求:(1)在冗余周期内没有该 EPC 出现,那么必须立即输出该事件;(2)对于 EPC 标签 TagID,过滤后,所有的 EPC 事件的生成时间,再加上冗余周期,所形成的时间段,必须包括该 TagID 的可见范围。即不允许标签 TagID 在数据源中的某时刻 t 是可见的,但在过滤后的时间段中没有包括该时刻,在时刻 t ,标签 TagID 变成了不可见

的。具体描述如下:

对于 EPC 标签 TagID,数据源中 TagID 的可见范围为 Tag_i ,输出事件为 $EPC_{i,j}$,事件的生成时间为 $TEPC_{i,j}$,冗余过滤的周期为 T_c ,则必须满足下面的关系:

$$Tag_i \subseteq \bigcup_j [TEPC_{i,j}, TEPC_{i,j} + T_c]$$

为了满足以上两点要求,本文设计了哈希实时冗余过滤算法:

(1)利用一个哈希表 EvCurrentHash 缓存当前的数据,另一个哈希表 EvDirtyHash 缓存进入到脏读数据处理器中的数据。在此哈希表中,TagID 为哈希表的关键字,哈希表 EvCurrentHash 中用来存储 $[T_n - T_c, T_n]$ 时间段内的事件,其中的 T_n 代表当前时间;

(2)判断读取的 TagID 是否存在于 EvCurrentHash 中;如果存在则判断 ReadTime,跳转至(3)或者(4);不存在则转到(5);

(3)如果 ReadTime 小于 $T_n - T_c$,即表示是一个新的标签读取,则生成 Tag 事件,并输出该事件到 EvCurrentHash,返回(1),执行下一个标签的读取;

(4)如果 ReadTime 在 $[T_n - T_c, T_n]$ 范围内,比较该标签与 EvCurrentHash 中相同标签的读取时间;如果该标签的读取时间较早,则忽略该 Tag 事件,否则,更新 EvCurrentHash 中该标签的读取时间;

(5)轮询 EvCurrentHash 表,将该表中所有 TagID 不同的 Tag 事件输出到 EvDirtyHash 中,等待脏读数据处理。

2.2 脏读数据处理——平滑过滤器

经过上面的冗余过滤器处理的数据存储在 EvDirtyHash 中,接下来进行脏读数据处理^[5]。

对同一个读写器数据流中的 Tag 事件进行脏读数据处理,主要通过两个通道进行^[6]:

首先,通过验证码校验和 TDS 解析的方式处理。通过校验码来检查读写器在读取的过程或者在数据传输的过程中造成的位错误,然后通过对 EPC 标签按照 EPCGlobal TDS 标准 (<http://www.gs1.org/epcglobal>) 进行解析,清洗其中无法解析和编码异常的数据,解决脏数据中的错误数据。

其次,运用设置阈值的方法。设置时间阈值 time,次数阈值 num;在 time 时间内出现次数达到 num 的 Tag 事件即为系统需要的标签,以此去除脏数据中偶尔被读到但不需要的数据。由于需要得到 Tag 事件的次数,因此修改原来的标签数据结构,添加读取次数:

```

Struct TagData2 {
Struct TagData1 ;
Int Num; //读取次数
}

```

在通过上述两个通道进行脏读数据处理的过程中,一共需要 3 个表:输入表、EvDirtyHash 存储表和输出列表。其中输入表保存原始数据,EvDirtyHash 表中存储的是经过多读数据处理的数据,输出列表存储的是经过平滑后的数据。先将 EvDirtyHash 和输入列表中所有 ReaderID 不是系统所需要的数据去除,然后将输入列表中的原始数据与 EvDirtyHash 存储表中的数据逐一进行比较,描述如下:

1) 比较两个表中 TagID: 相同,则进入到 2); 否则,进入 3);

2) 将存储列表 EvDirtyHash 中该标签的 Num 值加 1, 然后执行 4);

3) 将输入列表中的该标签的信息加入到 EvDirtyHash 中, 然后进入 4);

4) 遍历 EvDirtyHash, 将 TagID 相同的标签的 Num 累加到最后一项, 计算读取时间差 $t_{dur} = tTagID_i - tTagID_{last}$, 并删除前面的标签, 然后进入 5);

5) 从存储列表 EvDirtyHash 中取出该标签信息, 如果读取时间差 t_{dur} 不大于时间阈值 time, 并且出现的次数 Num 不小于次数阈值 num, 那么将该标签放入到输出列表中, 并清除原始数据和 EvDirtyHash 中该标签的信息, 转回 1), 执行下一个标签; 否则, 直接跳转到 1), 执行下一个标签。

2.3 漏读数据处理——基于伯努利的动态自适应平滑算法

RFID 的数据漏读主要是由标签的识别率、读写器的读写频率和周围环境这三者产生的相对恒定的漏读比造成的。本文采用基于伯努利概率模型的自适应窗口平滑算法进行漏读数据处理^[7]。

在自适应窗口^[8]的设计中, 定义 avgDura 为一段时间内 Tag 事件的平均读取时间间隔; evTime 为 Tag 事件的时间间隔; 通过一个窗口缩放因子 winFactor 来决定 evTime 向 avgDura 靠拢的速度, 通过一个精度因子 preFactor 来决定窗口大小 evTimeOut 的值。

定义这些值:

$$avgDura = \frac{durationTime \times sum(Tag)}{\sum event}$$

durationTime 表示需要分析的时间段长度;

sum(Tag) 表示这段时间内读取到的标签总数, 即有多少个不同的 TagID; $\sum event$ 表示事件的总数。通过上式计算得出 Tag 事件的平均读取时间间隔。

通过计算平均读取时间来降低由读写器所处环境、读写器读写频率、标签识别率这三方引起的恒定漏读比。

$$evTime = evTime + winFactor \times (avgDura - evTime)。$$

上式得出 Tag 事件的时间间隔, 用窗口缩放因子决定 evTime 向 avgDura 靠拢的速度。通过计算, 降低环境引起的漏读比。

$$evTimeOut = evTime \times preFactor。$$

窗口大小由精度因子 preFactor 和 evTime 共同决定。

自适应窗口的平滑算法, 在数据密度不变的情况下, 它所能弥补的漏读空隙完全由两个因子决定。相对固定窗口的平滑算法, 它在一定程度上提高了数据准确性, 但这只是静态的运用之前的值来计算平均窗口大小, 而如果标签在读取过程中动态发生改变的话, 就不能满足要求, 因此引出基于伯努利概率模型的动态自适应算法。

对于标签 i , 在平滑窗口内读到的次数 N 满足伯努利二项分布—— $B(\varepsilon_i, p_i^{avg})$, 其中, ε_i 表示读写器读取次数, 即实验次数, 表示窗口大小; p_i^{avg} 表示标签的平均读取频率, 即事件发生的概率。设置信度为 δ , 由公式 $\ln(1/\delta) = \varepsilon_i p_i^{avg}$ 可知, 要保证数据完整性, 窗口大小 ε_i 需满足 $\varepsilon_i \geq \frac{\ln(1/\delta)}{p_i^{avg}}$, 当窗口小于 ε_i 时, 适当增加窗口大小; 当标签发生动态变化时, 探测到标签发生变化的条件为: $|N_i| - \varepsilon_i p_i^{avg} > 2 \times \sqrt{\varepsilon_i p_i^{avg}(1 - p_i^{avg})}$, 其中 $|N_i|$ 为标签 i 在平滑窗口内被读到的次数, 当 $|N_i|$ 不在上式区间时, 应适当减小窗口大小。根据以上方法, 动态调整窗口大小, 可以保证较准确的数据填充。

3 实现 RFID 过滤的方法

经过对 RFID 冗余数据、脏读数据和漏读数据处理的过滤器设计之后, 将这 3 种过滤器依次组合, 构成本文所阐述的数据过滤方法:

RFID 原始数据被读卡器读取后, 读卡器将数据传入到如图 1 所示的数据处理模型中, 经过时间阈值和次数阈值过滤器后, 去除冗余数据, 然后进入下一层的脏读数据处理器中; 经过数据校验和 TDS 数

据分析,清除原始数据中的脏读数据,并进入到漏读数据处理器中;根据伯努利概率模型建立的动态时间窗漏读数据处理器,对原始数据进行漏读数据的处理,并最终输出到生成有效事件层中,即完成对RFID原始数据的过滤过程。

4 结束语

本文针对RFID数据特点,主要对冗余、脏读以及漏读数据进行处理,提出数据过滤方法。RFID原始数据先经过冗余过滤器去除冗余数据,再进入平滑过滤器清除原始数据中的脏读数据,最后采用基于伯努利的动态自适应平滑算法处理,对原始数据进行漏读数据的处理,最终生成系统需要的有效事件。

参考文献:

- [1] IBM Company. IBM Websphere RFID solution [R]. New York: IBM Company, 2008.
- [2] 白红霞,吴文华. RFID数据采集中间件研究[J]. 微计

算机信息, 2009, 25(4-2): 255-257.

- [3] 阴躲芬,龚华明. RFID中间件数据处理模型研究与实现[J]. 计算机与现代化, 2012(9): 200-202.
- [4] Klaus Finkenzeller. 射频识别技术[M]. 吴晓峰,陈大才,译. 第3版. 北京:电子工业出版社, 2006.
- [5] 张明哲,张强,袁魏,等. 嵌入式RFID中间件数据过滤模型研究[J]. 计算机工程与设计, 2010, 31(17): 3743-3746.
- [6] 张文新,邓毅华,谢胜利. 基于嵌入式RFID中间件的标签数据处理[J]. 微计算机信息, 2009, 25(14): 182-184.
- [7] 王文闯,郭凤宇. 基于动态时间窗的射频识别中间件数据过滤算法[J]. 信息与电子工程, 2009, 7(3): 177-179, 183.
- [8] Wang Wei, McFarlane D, Brusey J. Timing analysis of real-time networked RFID systems [R]. Cambridge: Cambridge AutoID Lab, 2005.

(责任编辑:陆雁)

广西科学院成为全国科学院联盟文献情报分会理事单位

新闻时间:2013-11-07

为进一步落实广西区政府与中国科学院科技合作协议,广西科学院在文献情报工作中取得新突破,成为全国科学院联盟文献情报分会理事成员之一。

中科院为落实与各省科学院协同创新、服务区域发展等举措,11月5日,由中科院国家科学图书馆发起的“全国科学院联盟文献情报分会”在京成立,会议正式通过《全国科学院联盟文献情报分会章程》,推选国科图馆长张晓林担任“全国科学院联盟文献情报分会”理事长、国科图副馆长刘细文担任分会秘书长。广西科学院副院长刘书龙和院办公室主任兼广西科学院文献信息共享服务站站长毛卫华同志代表广西科学院参加了分会成立仪式。

“全国科学院联盟文献情报分会”致力于推动全国科学院联盟成员机构实现文献情报服务共享,将本着“资源共享、优势互补、交流合作、共同发展”原则,依托国科图,联合各省科学院开展文献情报服务共享,提高文献信息保障效率,促进成员单位科技创新能力提升。

国科图针对各个省科学院文献情报需求,采取开通特色文献信息资源、培训文献情报服务队伍、提升省科学院研究所科研人员信息获取能力、开通运行省院文献情报服务门户、协助各省科学院提升情报跟踪研究能力等五大有效措施,拓展对省科学院文献情报服务取得成效,形成了共享文献情报服务机制,提升了各个省科学院文献信息获取能力。

分会成立后,国科图将继续探索巩固全国科学院协同共享的服务机制,把省级科学院文献情报能力建设纳入国科图发展力量的一部分,将国科图已有的优势资源共享给省级科学院,努力打造一流的文献情报服务体系、共享体系,共建覆盖全国地方科学院的科技和产业发展战略支撑平台。

(摘自广西科学院网页)