

无线传感器网络数据管理技术研究进展

张少平^{1,2} 汪英华³ 李国徽¹

(华中科技大学计算机科学与技术学院 武汉 430074)¹

(江西农业大学计算机与信息工程学院 南昌 330027)² (湖南衡阳市教育科学研究所 衡阳 421000)³

摘要 无线传感器网络本质上是一个以数据为中心的网络,它处理的数据为传感器采集的连续不断的数据流。因此,现有的数据管理技术把无线传感器网络看作为来自物理世界的连续数据流组成的分布式数据库。由于传感器节点的计算能力、存储容量、通信能力以及电池能量有限,再加上 flash 存储器以及数据流本身的特性,给数据管理带来了传统分布式数据库系统没有的一些新挑战。从数据库系统的体系结构、数据存储与索引技术、数据模式、查询处理及优化技术等方面介绍了无线传感器网络的数据管理技术的研究现状。

关键词 无线传感器网络,数据管理,数据模式,数据存储,查询优化,数据挖掘

中图分类号 TP393 **文献标识码** A

Overview of Data Management in Wireless Sensor Networks

ZHANG Shao-ping^{1,2} WANG Ying-hua³ LI Guo-hui¹

(School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)¹

(School of Computer and Information Engineering, Jiangxi Agriculture University, Nanchang 330027, China)²

(Institute of Education Sciences, Hengyang Hunan, Hengyang 421000, China)³

Abstract Wireless sensor networks are data-centric networks in nature, they deal with continuous data streams collected by sensor nodes. Therefore, existing data management technologies view wireless sensor network as a distributed database system composed of continuous data streams which come from the physical world. Due to the limitations of Sensor nodes' computing power, storage capacity, communication capabilities and battery energy, combined with the characteristics of flash memory and data streams, these lead to some challenges that the traditional distributed database systems have not encountered for data managements. This paper introduced the database system architecture, data storage and index technologies, data models, and query processing and optimization technologies for wireless sensor networks.

Keywords Wireless sensor network, Data management, Data model, Data storage, Query optimization, Data mining

1 引言

无线传感器网络(Wireless Sensor Network, WSN)是由部署在监测区域内大量的微型传感器节点通过无线通信形成的一个多跳的自组织网络系统,其目的是协作地感知、采集和处理网络覆盖区域内被监测对象的信息,并发送给观察者。它综合了传感器技术、嵌入式计算技术、分布式信息处理技术、通信技术和微电机技术,在军事、工业、医疗、交通、环保等诸多方面有着巨大的应用价值。

WSN本质上是一个以数据为中心的网络^[1],传感器采集的数据称为感知数据,其特征是只有追加操作的连续数据流及近似的模糊数据,并且具有连续不断的查询。因此,现有的研究都把 WSN 数据库看作为来自物理世界的连续数据流组成的分布式数据库。由于 WSN 中节点计算能力、存储容量、通信能力都有限,且节点依靠电池供电,在很多场合下,电池

是不可更换的,直接影响网络的寿命。WSN 中的感知数据特性以及传感器节点自身特性给数据管理带来了传统分布式数据库技术没有的一些新挑战。具体表现在如下几个方面^[2-4]: (1)需要研究针对 WSN 数据特征的数据管理技术。由于传感器节点可以持续采集监测环境中的数据,WSN 中的数据往往是连续无限的数据流,其数据往往是近似的且数据分布的统计特征是未知的。而传统的分布式数据库中的数据往往是间断有限的,数据是确定的且数据分布的统计特征已知。(2)需要研究平衡能量消耗和响应时间的数据管理技术。能量消耗是 WSN 的一个重要技术指标,直接影响到网络的使用寿命;而响应时间是 WSN 的另一项重要指标,尤其是对实时监测应用。响应时间和能量消耗是一对相互冲突的技术指标。因此,如何平衡响应时间和能量消耗,确保在满足响应时间要求的情况下,尽可能降低能量消耗,是一个值得深入研究的课题。(3)需要研究针对 flash 存储器,以降低能耗为目标的数

到稿日期:2009-07-22 返修日期:2009-10-16 本文受 863 国家高技术研究发展计划(2007AA01Z309),国家自然科学基金(60873030)资助。
张少平(1968-),男,博士生,副教授,主要研究领域为无线传感器网络、数据挖掘,E-mail: spzhang@mail.jxau.edu.cn;汪英华 女,主要研究领域为数据管理和数据挖掘;李国徽(1973-),男,博士,教授,博士生导师,主要研究领域为传感器网络、移动实时数据库、数据挖掘。

据存储技术。由于 flash 具有高可靠性、高密度、低能耗等一些特点,传感器节点都是用 flash 作为永久性存储器。和硬盘技术相比,flash 具有许多独特特性,传统的数据库存储技术不再适用。

本文将从数据库系统的体系结构、数据存储与索引、数据模式、数据查询及优化和数据挖掘等方面阐述 WSN 的数据管理技术。

2 传感器网络数据库系统体系结构

典型的传感器网络的系统结构包括资源受限的传感器节点群组成的多跳自组织网络、资源丰富的 Sink 节点、互联网和用户界面等^[5]。映射到传感器网络的分布式数据库系统也采用两层体系结构,如图 1 所示,它是由运行在传感器节点上本地数据库和运行在 sink 节点上与局部数据库进行交互的分布式数据库管理层组成^[3]。

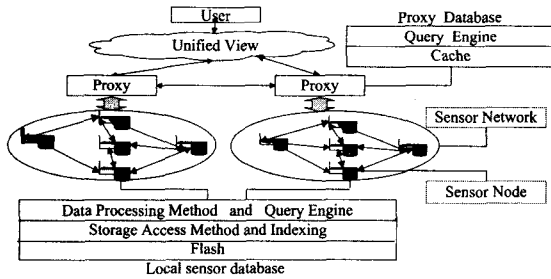


图 1 无线传感器网络数据库系统体系结构

本地数据库主要具有 3 个关键元素:(1)查询引擎。生成能量有效的查询计划用于执行查询,得到具有可信值的查询结果。(2)数据处理方法(如数据汇总及数据老化算法等)。为高效的数据查询提供多粒度的数据汇总,删除一些已过时的数据,以节约 flash 存储空间。(3)能量有效的存储管理。用于实现 flash 存储器分配与管理,为提高查询处理速度建立相关的索引等。这 3 个组件的实例依赖节点的能力。

位于 Sink 节点上的分布式数据管理层通常亦称为代理数据库,它包括两个关键组件:一个为数据缓存(cache),用于保存低层节点监测到的数据汇总及从传感器节点查询得到的结果;另一个为查询处理引擎,决定如何处理每个查询命令。查询可以使用 cache 中数据进行局部查询;或在做相应的优化操作后,把查询请求传送给相应的节点,从节点中提取更多的数据。

3 传感器节点的数据存储与索引技术

在通常情况下,传感器节点需要保存监测数据、系统可执行代码及系统设置信息等,永久性存储器是传感器节点的组成部分之一。出于抗震性、节点大小以及能量消耗等方面考虑,硬盘不适用于作为传感器节点的永久性存储器,flash 是目前的最佳选择。传感器节点的数据存储与索引技术要考虑 flash 特性、节点的能量消耗等因素。

3.1 flash 特性

根据 flash 存储单元的组织方式,flash 分为两类:NAND 和 NOR。NOR 采用随机访问方式,存储容量小,适用于作为程序存储器。而 NAND 以页(page)的方式进行访问,存储容量大,适用于存储大量数据的永久性存储器。NAND flash 的一个异常特性为读、写操作的不对称性。读、写是以页(Page)

为单位进行的,而且写数据之前必须先进行擦除。而擦除是以块为单位,每块由若干页组成。也就是说,若要更新某一页数据,必须先把该页所在的块读出,然后把该块进行擦除,再写入。因此,写一页的代价很高,而且 flash 块受擦除次数限制,典型值大约为 10 万次^[6,7]。flash 这些特性给基于 NAND flash 的存储技术带来了挑战。表 1 为一些 flash 存储器的性质^[7]。

表 1 一些 flash 芯片的性质

Type	Capacity	Page Size	Erase Block (Pages)
Atmel NOR	512kB	256B	1
Telos NOR	512kB	256B	256
Hitachi MMC	32MB	512B	16
Toshiba NAND	16MB	512B	32
Micron NAND	512MB	2kB	64

然而,NAND Flash 的显著优点是它的高容量和低功率消耗。和 DRAM 及 NOR flash 存储器相比,其容量大(单片容量目前可达 32GB),每 GB 的单价低;而且读写所消耗的能量也低。另外,和通信能耗相比,对同一数据,本地存储到 NAND Flash 上比把数据发送到邻居节点(MICA2 通信设备)的能耗要低 100 倍以上^[3]。因此,NAND flash 是传感器网络数据存储于传感器节点上的一种理想设备。表 2 为这些 flash 读写一个字节数据的能量消耗比较^[7]。

表 2 一些 flash 芯片的读写数据的能量消耗比较

Type	Energy per Bytes				
	Read	Write	Erase	Bulk Erase	Total
Atmel NOR	0.26	4.3	2.36	n/a	6.92
Telos NOR	0.056	0.127	n/a	0.185	0.368
Hitachi MMC	0.06	0.575	0.47	0.0033	1.108
Toshiba NAND	0.004	0.009	n/a	0.004	0.017
Micron NAND	0.027	0.034	n/a	0.001	0.062

3.2 基于 flash 的传感器节点数据存储与索引技术

根据 WSN 的数据流以及 flash 存储器的特征,目前传感器节点上的数据存储方式主要有两种方式:一种为不带索引基于日志结构的文件存储方式,另一种为基于索引结构的存储方式。

针对传感器节点设计的第一个文件系统是 matchbox^[8],已集成到 tinyOS 系统中。Matchbox 提供了基本的 flash 存储器的磨损平衡机制和远程文件访问方法,只允许对数据进行追加(append)操作,不允许随机访问文件中的数据,如修改(modify)操作。可以同时打开多个文件。其特点是程序代码很小,只有 10kB,运行时内存占用量也很少,最小约为 362 个字节。当打开的文件多时,内存占用量也会随着增加。

文献[9]是针对传感器节点设计的基于 flash 存储器的日志结构文件系统(Efficient Log Structured Flash File System,ELF)。ELF 考虑到传感器节点的资源限制,仅为系统的通用任务提供了一些基本的文件管理操作,如 open, create, modify, append, read, seek, delete, rename, truncation 等。与传统的基于 flash 的文件系统的不同之处主要在于对文件的写操作(包括 append 和 modify)。对于 append 操作,ELF 并不为每个 append 操作创建一个日志数据项,而是对每个文件,利用一个写缓冲区缓存追加到同一页的日志项,当缓冲区满时再写入到 flash 页上。这样可以减少对 flash 写的次数,以延长 flash 的使用寿命和降低能量消耗。而对于 modify 操作,ELF 申请一个新的 flash 页来存储修改的页面,而不是写

人原来的页中,以实现 flash 页的磨损平衡。另外,ELF 提供了碎片回收和故障恢复机制。碎片回收用于实现 flash 页的磨损平衡、可用空间的擦除和再分配,故障恢复用于当系统或 flash 页发生故障时利用检查点技术实现数据恢复。

受到日志结构文件系统^[10]的启发,文献[11]提出一种基于 B+ tree 索引的日志结构数据存储技术,其基本思想是把索引组织成事务日志。把对 B+ tree 树节点的写操作编码成一个日志记录,并存储在内存缓冲区中。当缓冲区包含的数据足够装满一页时,则写入 flash 中。另外,对每个 B+ tree 树节点,还保存一个页地址链接表,指向该 B+ tree 节点存储的日志记录的 flash 页的地址。

FlashDB^[12]根据 WSN 不同类型的工作负荷参数和 NAND flash 设备的一些特性参数,设计了一种自调节的数据存储方法。它综合了传统的基于磁盘的 B+ tree 索引和基于日志结构的 B+ tree 索引技术,采用自调节技术动态地调节它的存储结构,以适应不同的工作负荷和 flash 设备,灵活地以两种方式中的一种存储索引节点。它把索引的自调节性质形式化为一个双态系统并提出相应的算法,实现理论上的最优。

Microhash^[13,14]提出一种基于 hash 索引结构的数据存储方式。它把 flash 的数据存储区组织成堆(heap),监测数据按时间顺序以循环数组方式存储在 flash 的数据存储区上,这种方式直接解决了删除、写以及磨损平衡问题。Microhash 在把监测数据存储到 flash 时,同时建立索引。索引采用两层索引结构,即 index 层和 directory 层。Index 的每个索引记录格式为[idx,offset],其中 idx 为数据存储的 flash 页的地址,offset 为存储在该页相对起始地址的偏移值。directory 的每个记录项包括 index 层的某个 flash 页的地址以及索引数据值的上、下界。

值得进一步研究的是,在传感器节点数据库构建和维护索引时,必须考虑能量消耗问题。建立索引可以提高访问数据的速度,减少读取 flash 页的次数。但建立和维护索引除了增加额外的存储空间外,其读写也需要消耗能量,尤其是对索引的写操作。这些能量消耗要在查询处理中得到补偿,索引才有意义。因此,使用索引结构只有在数据访问操作非常频繁时才有效,否则通过顺序扫描来执行查询更节约能量。

4 以数据为中心的存储与分布式索引技术

4.1 以数据为中心的存储技术

在 WSN 中,传感器监测数据可以存储在本地节点,也可以根据数据的属性通过某种映射技术存储到网络中的一些指定节点上,即以数据为中心的存储技术。

文献[15]提出基于地理位置散列表(Geographic Hash Table,GHT)的以数据为中心的存储方法。其基本思想为:首先用一个 GHF 将数据映射到一个地理位置,然后采用地理路由协议——贪心周边无状态路由协议(Greedy Perimeter Stateless Routing,GPSR)^[16]将测量数据存储到距离该位置最近的传感器节点。当某个监测数据出现频率很高时,会导致很多数据映射到同一个节点,即出现“热点”(hot spot)现象,GHT 使用结构复制技术解决这种问题。

文献[17]针对目标跟踪应用提出了一种网内数据存储方案(Energy-conserving Approximate Storage Scheme,EASE)。

EASE 在网络内保持两个版本的目标跟踪数据,一个为高精度数据,一个是低精度近似数据。高精度数据保存在移动目标附近的节点上,以避免长距离的更新引起能量消耗。而与此相对应的低精度数据,则复制到一个指定节点上,以减少查询费用,指定节点对用户来说是透明的。存储在指定节点的不精确移动目标位置数据由一个近似半径来限定。也就是说,如果移动目标保持在近似半径范围内移动,它的精确位置数据更新只存储在最近的节点上。此时,其对应的不精确表示并不更新,即不发送到指定的节点。相应地,若查询的精度约束条件低于近似半径指定的值,则查询结果可直接通过指定存储节点获得,否则查询要转发到本地存储节点,以获得更精确的结果。EASE 还通过适当地设置近似半径来平衡数据更新传输流量和查询流量,以优化网络的性能。

WSN 在军事上得到了很大的应用。为了实现数据的安全存储和安全传输,防止敌方能够从捕获的节点中获取数据,文献[18]提出了一种漫游数据存储方法(Evasive Data Storage,EDS)。其基本思想为:网络中有价值的价值数据并不是存储在某个固定的节点上,而是以一种不可预知的方式在网络中漫游。敌方若已经捕获先前存储数据的节点,也不能很快地访问数据,因为数据已不再存储在该节点上。另外,EDS 采用位置限定法,对漫游的位置加以限制,降低通信负担;采用数据划分(Data Splitting)策略,把数据分成小片,每一片独立地存储在网络的不同节点上,以防止敌方的睡眠攻击。

4.2 分布式索引技术

以数据为中心的存储技术根据数据的属性把相关联的数据存储到指定的节点,可通过数据融合技术对数据进行处理,避免把大量的测量数据传输到网外,以达到降低数据传输能耗的目的。在数据查询中,为了能快速定位到数据的存储节点,避免在全网泛洪广播查询请求,需要对网络中的数据建立分布式索引技术。如 GHT 技术可以直接根据数据的属性,利用 Hash 函数,定位到数据的存储节点。

加州大学洛杉矶分校开发的 Dimensions^[19]系统采用空间分解技术对数据建立索引。其基本思想为:首先根据查询的空间范围确定层次级别数 d ,然后递归地对查询空间进行分解,每一级分成 4 个子区域。即第 0 级为整个监测区域,选择一个节点为簇头节点;第 1 级将整个监测区域划分为 4 个子区域,每个子区域都选择一个节点为簇头节点;第 2 级又把第 1 级的子区域划分为 4 个子区域,同样每个子区域都选择一个节点为簇头节点。以此类推,一直划分到第 d 级为止,下级簇头节点不能和上级簇头节点使用同一个节点。当查询指定空间范围的数据时,将查询结果从第 d 级开始逐级传送到顶点和用户。该索引技术适用于本地存储数据和指定空间范围的多分辨率查询要求,其缺陷是簇头节点能量消耗过快,并且易造成通信瓶颈问题。

文献[20]提出了一种分布式索引方法(Distributed Index for Features in Sensor Networks,DIFS),该方法综合了 GHT 技术和空间分解技术,利用 GHT 技术实现了以数据为中心的存储,利用空间分解技术实现对分布式数据的索引。区别于文献[19]所采用的空间分解技术,其构造的层次结构的每个非根节点具有多个父节点,以解决能量消耗和通信瓶颈问题。每个节点都存储特定地理范围内和特定监测数据值范围内的数据。上层节点存储的数据覆盖的地理范围大,但覆盖

的监测数据值的范围小。相反,下层节点的数据覆盖的地理范围小,但数据值的范围大。在查询数据时,首先选择最高父节点的集合,这些节点覆盖所有查询要求的数据名的范围。然后根据查询要求的空间范围逐层进行遍历,最后得到查询结果。DIFS 适用于指定空间范围以及指定数据值范围的单属性的查询要求。

上述索引技术只用于对单一属性数据建立索引,而在异构 WSN 中可具有多种类型传感器,能测量到不同属性的数据。文献[21]提出了一种支持多属性范围查询的分布式索引技术(Distributed Index for Multi-dimensional data, DIM)。DIM 方法依赖两种技术:局部保持(locality-preserving)地理散列和基于地理位置的贪婪周边路由协议 GPSR。它首先通过局部保持地理散列函数将一个多维数据映射到二维平面空间的一点,然后应用 GPSR 将该数据存储在离该点最近的节点上。同样,在查询数据时,根据查询请求的数据要求,通过局部保持地理散列函数获得测量数据所在的区域,并应用 GPSR 把查询传送到这些区域,从这些区域中提取相应的数据。

5 数据模式

现有的 WSN 数据库系统对传感器监测的数据流建模大多为对传统的数据模式进行扩展,主要有基于工作流模式、基于关系模式、基于对象模式。针对 WSN 的一些特殊应用,也可以建立特殊的数据模式。

Aurora 系统^[22]是一种面向时间工作流模式建模的系统,其查询建立在 Aurora 查询代数基础上,包括 3 个与顺序无关的操作(Filter, Map 和 Union)和 4 个对顺序敏感的操作(BSort, Aggregate, Join 和 Resample)。Aurora 的数据流采用统一的元组序列形式(TS, A_1, A_2, \dots, A_n),其中 TS 为时间戳, $A_i (1 \leq i \leq n)$ 为与应用相关的数据域,对数据流的操作只有添加操作。Aurora 系统针对于高效调度、服务质量和优化结构而设计,以统一处理方式支持连续查询、滑动窗口和其它的一些特别查询。Borealis^[23]对 Aurora 系统进行了扩展,其数据流为形如(TS, tuple-type, id, A_1, A_2, \dots, A_n)的元组序列,其中 tuple-type 支持插入、删除和替换操作, id 为元组的标识符。另外, Borealis 还设计了支持如下操作的数据流引擎:查询的动态更改、查询结果的动态更新。Aurora 和 Borealis 都实现了分布式数据流模型,可根据网络条件的变化动态配置。

TinyDB^[24]采用基于关系的数据模式,并对传统的关系模式进行了扩展。它把传感器节点的测量数据定义为一个单一的、无限长的、有两类属性的虚拟关系表:一类用来定义测量数据,如节点标识符、测量时间、测量数据类型、单位等;另一类用来描述测量数据本身,如温度、位置等。传感器产生的测量数据对应表的一行,对数据的查询就是对这个无限虚拟表的查询。美国斯坦福大学针对 WSN 开发的 STREAM^[25]系统采用的也是基于关系的数据模式。它把数据流建模为无边界的、只能进行添加操作的元组对(tuple, timestamp)组成的数据流,把关系作为支持更新、插入和删除操作、随时间变化的元组包。其语义建立在 3 组抽象操作上:关系-关系操作、数据流-关系操作和关系-数据流操作。

COUGAR^[26]是一个基于抽象数据类型(Abstract Data

Type)的数据流系统,它采用两种模式对数据进行建模,用对象关系模式来组织建模存储数据;引入一种时间序列模式建模组织传感器监测数据,并定义了相应的关系代数操作、时间序列操作以及关系及时间序列之间的操作。

针对传感器测量数据的不确定性,PSRA^[27]扩展传统的关系模型到概率数据流关系(Probabilistic Stream Relation)模型,并扩展传统关系模型的操作,在概率数据流模型上定义了 Stream Union, Stream Intersection, Stream Select, Stream project, Stream Join 等操作。PSRA 通过概率数据流模型有效地解决了 WSN 的数据不确定性以及数据的相互关系等一些特征,并提供了能量高效的操作。

传感器网络的有些应用并不需要精确的测量数据,如对森林防火监控,用温度传感器对周围的环境进行监控,对测量的数据并不需要它的精确值,只需把测量数据划分为低、较低、中、较高、高、极高几个等级。根据这一类应用的特征,文献[28]中提出一种基于粗糙集(Rough Set)理论的数据建模方法。利用粗糙集对数据建模,可以很好地实现数据融合操作,从而减小数据存储量及网络传输量,达到节约能量,延长网络寿命的目的。

6 数据查询处理与优化

6.1 数据查询

WSN 的数据查询应用可以分为两大类:查询动态数据和查询历史数据。在查询动态数据中,数据在传感器监测到的一个小的时间窗内有效,例如事件检测查询或一些特定查询(当前的温度是多少?)。而查询历史数据是指对检测到的历史数据进行数据挖掘,用于发现事件特殊模式,分析数据走向,形成特定事件的理想模型等。对这一类应用来说,每一个数据都是重要的,不能被抛弃。

WSN 数据库系统可理解为一个两层结构的分布式数据库系统:运行在 Sink 节点上的代理数据库服务器和运行在传感器节点上的局部数据库。数据查询的处理过程一般为:首先用户使用命令式查询接口把查询请求发送到网络,通过路由技术传送到运行在 Sink 节点的代理服务器。其次,代理服务器根据接收到的用户请求生成相应查询计划。然后,代理服务器把查询计划通过路由技术发送到相应的传感器节点。节点接收到查询后,执行查询,并把结果传送到代理服务器。最后,代理服务器对节点返回的结果进行处理,并把最终结果返回给相应的用户^[3,29]。

由于 SQL 语言在数据库领域广泛应用,其显著优点为:(1)方便用户。用户使用 SQL 语言只需定义他所想要的数
据,而不需要知道如何获取这些数据,并且 SQL 语言易学,也容易理解。(2)语言形式与实现分离。也就是说,查询系统内部修改了如何执行一个查询,而查询语言形式并没有改变。目前,WSN 的数据查询语言大多都延续了传统的 SQL 语言形式,并对 SQL 语言进行了扩展。TinyDB 的查询语言在 WSN 中具有一定的代表性,其语法结构表述如下:

```
SELECT select-list  
[FROM sensors]  
WHERE where-predicate  
[GROUP BY gb-list  
[HAVING having-predicate]]
```

```
[TRIGGER ACTION command-name[(param)]]
[EPOCH DURATION time]
```

其中,select-list 是数据属性或与属性相关的聚集函数,gb-list 是数据属性表,where-predicate 和 having-predicate 是谓词,方括号中的内容是可选项。

6.2 查询优化

WSN 中的查询优化策略大致可分为运行在 Sink 节点上的多查询优化策略和运行在网内节点上的单查询优化策略^[30]。这两种技术结合起来构造 WSN 的查询优化系统。优化的目标要在保证网络服务质量的前提下,尽可能降低能量消耗,以延长网络的寿命。

多查询优化策略建立在单查询优化策略之上,它把用户发送到 Sink 节点的查询集合 Q 优化成一个新的查询集合 Q',以尽可能地删除 Q 中不同查询中的冗余请求。优化的最佳情形为新查询集 Q'中的查询结果刚好能满足 Q 中的所有查询请求,并且 Q 中不同查询所需要的同样数据可根据 Q'中查询在传感器网络中仅获取一次。文献[30]利用贪婪查询插入算法把相似的查询集构造为一个新的优化后查询集,以尽可能地减少冗余的查询请求,优化后再把优化的查询发送到网络中。文献[31]也提出了一种基于 Sink 节点的查询优化策略,它把查询计划扩展到查询执行的各个方面,包括路由、传感器监测和数据/元数据的收集。其查询优化过程分为两个阶段:划分阶段和精化阶段。划分阶段通过评估查询集合的查询计划得到一个查询费用最小的查询集并决定是否要收集元数据。若不需要收集元数据,则把新的查询集发送到相关的节点,否则进入精化阶段。精化阶段收集元数据并重新评估查询计划,得到一个新的代价最小的查询集,并传送到相关的节点。每个节点收到查询请求,执行查询,并把查询结果传送到 Sink 节点。

目前,传感器网络广泛采用网内数据处理技术来降低数据传送量,以节约传输的能量消耗。大部分系统都是结合数据融合和路由技术,在数据传输的路由节点上把相关联的数据融合在一起,以降低数据传输量^[32]。这类技术对数据查询的汇总操作(如 max,sum 等)很有效,但没有对查询运算进行优化。文献[33]提出了一种针对查询运算(如 filter,join 等)的层次式网络查询优化策略。网络采取层次式组织,越在上层的节点,其计算能力和通信能力越强。数据由叶节点获取,查询请求由根节点向下传送到叶节点,查询结果从叶节点向上传送到根节点。为了降低网络通信量,查询运算一般在低层节点上执行,但这时需要较高的计算费用。该策略通过贪婪算法优化查询运算的执行层次来平衡计算费用与网络通信量,以达到降低整个网络能量消耗的目的。

现有的传感器网络数据库系统也都采用一些查询优化策略。TinyDB^[34]的查询优化目标是降低网络的总能量消耗。它采用基于代价的查询优化技术来产生能量消耗尽可能少的查询执行计划。查询代价由传感器节点采集数据和传输查询结果的能量消耗决定。其优化技术主要集中于数据采集和谓词操作的执行次序,并且确定可以共享的数据采集操作,删除不必要的数据采集操作。TinyDB 还通过优化基于事件的查询来降低冗余的数据采集操作。根据这一特点,TinyDB 采用基于重写的多查询的优化技术,把多个外部事件转化为一个事件流,使得不管事件以何种频率发生,只能同时有一个查询

在运行,这样就避免频繁地启动数据采集操作。在 STREAM^[25]系统中,一旦持续查询发布,就生成一个相应的查询计划。查询计划的执行由用于运行状态资源管理的全局调度器来控制,使得单数据流查询在运行时内存占用方面几乎是最佳的。STREAM 系统所采用的优化技术包括:在查询计划中重新分配窗口运算;使用数据流限制来减少窗口的大小;标识共享计算和共享内存的时机;当由于资源限制迫使降低查询的精度要求,得到近似的查询结果时,可以使用减少滑动窗口、降低采样频率等相关的技术来实现。

查询优化问题是传感器网络领域的研究难题之一。它必须设计一些高效的分布式处理和数据重用技术,既要降低全网的能量消耗,又要避免少量节点因负担过重,能量消耗过快而失效,从而影响到整个网络的使用寿命。

6.3 历史数据挖掘

传感器网络是一个以数据为中心的网络,传感器从监测环境收集到的大量数据可能存在某种内在联系。数据挖掘技术可用于从大量的数据中挖掘出人们感兴趣的数据关联规则或传感器节点间的关系。

文献[35]提出了一种用于挖掘传感器节点行为模式的算法,其主要目的是确定传感器节点行为模式的关联规则。这些规则可以用于资源管理或用于弥补网络通信的不利因素,以便改进网络服务质量。文中给出了传感器节点行为模式关联规则的形式化定义;提出了一种针对数据挖掘处理而收集数据的压缩表示结构——位置词典树(Positional Lexicographic Tree),该结构有助于快速而有效地挖掘节点的行为模式;设计了一个挖掘行为模式的分布式数据挖掘算法。

传感器网络的数据大多表现为一种数据流的形式,频度序列挖掘(Frequent sequential mining)可用于挖掘数据流的频度序列模式(frequent sequential patterns)。文献[36]中提出了一种在线挖掘算法——SSM(Sequential Stream Mining)来挖掘数据流的频度序列模式,可适用于传感器网络的数据流。SSM 算法采用 3 种数据结构(D-List, PLWAP-tree 和 FSP-tree)处理数据流中挖掘频度序列模式的复杂性。算法采用在线方式,即每监测到一项数据,使用 D-List 统计有关数据项及其频度数,D-List 实质上就是提取原数据库与挖掘有关的数据及其频度的一种压缩存储表示;然后,对 D-list 进行批挖掘,应用 PLWAP-tree 构造算法生成频度序列模式,并把频度序列模式保存在 FSP-tree 中,FSP-tree 是频度序列模式的一种压缩存储数据结构。

传感器网络的应用领域之一是目标跟踪,其数据是要反映目标的运动轨迹。文献[37]针对目标跟踪的传感器网络,提出了挖掘时间运动模式的 TMP-Mine(Temporal Moving Patterns Mine)算法。TMP-Mine 算法通过构造数据库的一种压缩表示 TMP-Tree,对 TMP-Tree 采用 TMP-Mine 算法挖掘出时间运动模式。时间运动模式和现有的预测模型结合起来可以降低预测的错误,从而达到节约能量的目的。

由于传感器节点资源有限,资源利用是实现数据挖掘算法要考虑的目标之一。文献[38]介绍了一种与资源有关的在线数据挖掘算法——ERA-Cluster(Extended Resource-Aware Cluster)。ERA-Cluster 基于簇路由协议,设计了一个资源监控组件监督并报告各种资源的使用情况,根据资源的使用率来调节算法的相关参数,以支持分布式数据处理技术。

结束语 无线传感器网络是一个以数据为中心的网络,它管理的传感器监测数据和传统数据库中的数据有很大的差别,再加上传感器节点自身的特性,给数据管理技术带来了巨大的挑战与机遇。目前,在 WSN 的数据管理技术方面已经取得了一些研究成果,但离实际应用还很远,仍具有很大的研究空间。本文主要从数据库的体系结构、数据在传感器节点上的存储与索引技术、以数据为中心的存储与分布式索引技术、数据模式、数据查询与优化以及数据挖掘等方面阐述传感器网络的数据管理技术。

参 考 文 献

- [1] 李建中,李金宝,石胜飞. 传感器网络及其数据管理的概念、问题与进展[J]. 软件学报, 2003, 14(10): 1717-1727
- [2] Zhuang L Q, Zhang J B, Zhang D H, et al. Data Management for Wireless Sensor Networks; Research Issues and challenges[C]// Proceeding of 2005 International Conference on Control and Automation. Budapest, Hungary, IEEE, 2005: 208-213
- [3] Diao Y, Ganesan D, Mathur G, et al. Re-thinking Data Management for Storage-centric Sensor Networks[C]// Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research (CIDR). 2007: 22-31
- [4] 王殊, 阎毓杰, 胡富平. 无线传感器网络的理论及应用[M]. 北京: 北京航空航天大学出版社, 2007: 248-260
- [5] 崔莉, 鞠海玲, 苗勇, 等. 无线传感器网络研究进展[J]. 计算机研究与发展, 2005, 42(1): 163-174
- [6] Leventhal A. Flash Storage Memory [J]. Communications of the ACM, 2008, 51(7): 47-51
- [7] Mathur G, Desnoyers P, Ganesan D, et al. Ultra-low Power Data Storage for Sensor Networks [C]// Proceedings of the 5th International Conference on Information Processing in Sensor Networks. Nashville, Tennessee, IEEE, 2006: 374-381
- [8] Hill J, Szewczyk R, Woo A, et al. System Architecture Directions for Networked Sensors [J]. ACM SIGPLAN Notice, 2000, 35(11): 93-104
- [9] Dai H, Neufeld M, Han R. ELF: an efficient log-structured flash file system for micro sensor nodes [C]// Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems. Baltimore, USA, New York, USA: ACM, 2004: 176-187
- [10] Rosenblum M, Ousterhout J K. The design and implementation of a log-structured file system [J]. ACM Transactions on Computer Systems, 1992, 10(1): 26-52
- [11] Wu C H, Chang L P, Kuo T W. An efficient b-tree layer for flash-memory storage systems [C]// RTCSA (2003). New Orleans, USA. New York, USA: ACM, 2003: 17-24
- [12] Suman N, Aman K. FlashDB: dynamic self-tuning database for NAND flash [C]// Proceedings of the 6th International Conference on Information Processing in Sensor Networks. Cambridge, USA. New York, USA: ACM, 2007: 410-419
- [13] Zeinalipour-Yazti D, Lin S, Kalogeraki V, et al. MicroHash: An Efficient Index Structure for Flash-based Sensor Devices [C]// Proceedings of The 4th USENIX Conference on File and Storage Technologies (FAST'05). Berkeley, USA: USENIX Association, 2005: 31-44
- [14] Lin S, Zeinalipour-Yazti D, Kalogeraki V, et al. Efficient Indexing Data Structures for Flash-based Sensor Devices[J]. ACM Transactions on Storage, 2006, 2(4): 468-503
- [15] Ratnasamy S, Karp B. GHT: Geographic Hash Table for Data-centric Storage [C]// Proceedings of 1st ACM WSNA. New York, USA: ACM, 2002: 94-103
- [16] Karp B, Kung H T. GPSR: Greedy Perimeter Stateless Routing for Wireless Networks[C]// Proceedings of the 6th International Conference on Mobile Computing and Networking. New York, USA: ACM, 2000: 243-254
- [17] Xu J L, Tang X Y, Lee W C. A New Storage Scheme for Approximate Location Queries in Object-tracking Sensor Networks [J]. IEEE Transactions on Parallel and Distributed Systems, 2008, 19(2): 262-275
- [18] Benenson Z, Freiling F C, Cholewinski P M. Advanced Evasive Data Storage in Sensor Networks [C]// Proceedings of International Conference on Mobile Data Management. IEEE, 2007: 146-151
- [19] Ganesan D, Estrin D, Heidemann J. DIMENSIONS: Why do We Need a New Data Handling Architecture for Sensor Networks [J]. ACM SIGCOMM Computer Communication Review, 2003, 33(1): 145-148
- [20] Greenstein B, Ratnasamy S, Shenker S, et al. DIFS: a distributed index for features in sensor networks [J]. Ad Hoc Networks, 2003, 1(2/3): 333-349
- [21] Young X L, Kim J, Govindan R, et al. Multi-dimensional Range Queries in Sensor Network [C]// Proceedings of the 1st International Conference on Embedded Networked Sensor Systems. New York, USA: ACM, 2003: 63-75
- [22] Abadi D, Carney D, Cetintemel U, et al. Aurora: A New Model and Architecture for Data Stream Management [J]. The VLDB Journal, 2003, 12(2): 120-139
- [23] Abadi D, Ahmad Y, Balazinska M, et al. The Design of the Borealis Stream Processing Engine [C]// Proceedings of the 2005 CIDR Conference. 2005: 277-289
- [24] TinyDB[OL]. <http://telegraph.cs.berkeley.edu/tinydb/tinydb.pdf>
- [25] Motwani R, Widom J, Arasu A, et al. Query processing, resource management, and approximation in a data stream management system [C]// Proceedings of the 2003 CIDR Conference. 2003: 245-256
- [26] Demers A, Gehrke J, Rajaraman R, et al. The Cougar Project: A Work-In-Progress Report [J]. ACM SIGMOD Record, 2003, 34(4): 53-59
- [27] Liu H, Hwang S Y, Srivastava J. PSRA: A Data Model for Managing Data in Sensor Networks [C]// Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing. IEEE, 2006: 1-8
- [28] Dai Z F, Li Y X, He G L, et al. Uncertain Data Management for Wireless Sensor Networks Using Rough Set Theory [C]// Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing. IEEE, 2006: 1-5
- [29] Gehrke J, Madden S. Query Processing in Sensor Networks [J]. IEEE Pervasive Computing, 2004, 3(1): 46-55
- [30] Xiang S, Lim H B, Tan K L. Impact of Multi-query Optimization in Sensor Networks [C]// Proceedings of the 3rd International Workshop on Data Management for Sensor Networks (DMSN'06). New York, USA: ACM, 2006: 7-12
- [31] Ross R, Lee W C, Bhuvan U. Optimizing Energy-efficient Query Processing in Wireless Sensor Networks [C]// Proceedings of the International Conference on Mobile Data Management. 2007: 24-29
- [32] Nakamura E F, Loureiro A A F, Frery A C. Information Fusion for Wireless Sensor Networks; Methods, Models, and Classifications [J]. ACM Computing Surveys, 2007, 39(3)

2)成功传送数据包的概率:只有将数据完整传送到汇聚节点才能体现数据传送的准确性,如式(6)所示, $Data_Success_Num$ 为一段时间内簇头节点成功传送到汇聚节点的数据包数, $Data_All_Num$ 为在一段时间内不考虑故障因素传输成功的数据包总数,利用比值体现该模型准确数据传送的能力。如图5所示,该模型数据成功率基本在90%左右。

$$P = \frac{Data_Success_Num}{Data_All_Num} \times 100\% \quad (6)$$

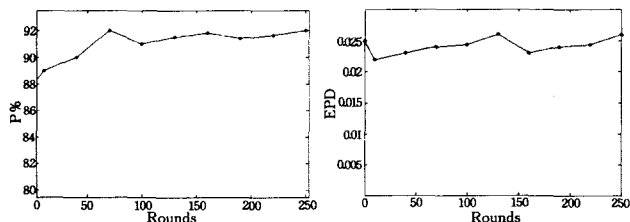


图5 成功传送数据包的概率 图6 成功传送所消耗能量代价

3)成功传送所消耗能量代价:反映将数据有效地传送到汇聚节点所消耗的能量情况,如式(7)所示。 $E(0)$ 为初始能量值。由图6可以看到,该模型构造的网络拓扑能耗均衡。

$$EPD = \frac{\sum_i E(0) - \sum_i E(i)}{Data_Success_Num} \quad (7)$$

可以看到,该模型演化的 WSN 簇间拓扑结构具有能量使用率和能量使用均衡性,能有效延缓网络寿命。在一定的簇头失效情况下,该模型仍具有较强的数据传输能力,可见其具有良好的容错性。

结束语 本文依据无线传感器网络的特点,给出了具有容错性能的簇间拓扑演化模型。该模型采用最小世界理论,簇头节点依据自身的局部视图构建簇间拓扑,实现容错和节能目标。仿真实验结果表明,该模型具有良好的容错性能和能量有效性,并能有效延长网络生存周期。

参考文献

- [1] Akyildiz IF, Su W, Sankarasubramanian Y, et al. A survey on sensor networks[J]. IEEE Communications Magazine, 2002, 40(8): 102-114
- [2] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-Efficient communication protocol for wireless micro sensor networks[C]//Proc. of the 33rd Annual Hawaii Int'l Conf on System Sciences, Maui; IEEE Computer Society, 2000; 3005-3014
- [3] Heinzelman W. Application-Specific protocol architectures for wireless networks [D]. Boston: Massachusetts Institute of Technology, 2000
- [4] Lindsey S, Raghavendra C S. PEGASIS: Power-efficient gathering in sensor information systems[C]//Proceedings of IEEE Aerospace Conference. 2002, 3: 1125-1130
- [5] Ozgur T H, et al. Power efficient data gathering and aggregation in wireless sensor networks[J]. SIGMOD Record, 2003, 32(4): 66-71
- [6] Watts D J, Strogatz S H. Collective dynamics of 'small-world' network[J]. Nature, 1998, 393(6884): 440-442
- [7] Kleinberg J. Navigation in a small world[J]. Nature, 2000; 406-845
- [8] Zheng Zeng-wei, Wu Zhao-hui, Lin-Huai-zong, et al. CRAM: an energy efficient routing algorithm for wireless sensor network [C]//International Symposium on Computer and Information Science. Antalya, Turke, 2004; 341-350
- [9] 江贺, 务志坤, 张宪超. 一种容错的无线传感器网络聚类路由协议[J]. 小型微型计算机系统, 2007, 8(8): 1362-1366
- [10] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks [C]// Proc. Sixth Annual International Conference on Mobile Computing and Networks. 2000
- [11] Nasipuri, Das S. On-Demand Multipath Routing for Mobile Ad Hoc Networks[C]//8th Intl. Conference on Computer Communications and Networks(IC3N 99). 1999
- [12] Dulman S, Nieberg T, Wu Jan. Trade-off between traffic overhead and reliability in multipath routing for wireless sensor networks [C] // Wireless Communication and Networking, WCNC2003. 2003; 1918-1922
- [13] Kashyap A, Khuller S, Shayman M. Relay placement for higher order connectivity in wireless sensor networks[C]//Proceedings of 25th IEEE International Conference on Computer Communications. Span, 2006; 1-12
- [14] Han Xiao-feng, Cao Xiang, Errol L, et al. Fault-tolerant relay node placement in heterogeneous wireless sensor networks[C]// Proceedings of the 26th IEEE International Conference on Computer Communications. Anchorage, AK, 2007; 1667-1675
- [15] Younis O, Fahmy S, Santi P. Robust communications for sensor networks in hostile environment [C]// Proceedings of the 12th IEEE International Workshop on Quality of Service. 2004; 10-19
- [16] Sharma G, Mazumdar R. Hybrid sensor networks: A small world [C]//MobiHoc 05. New York, USA; ACM, 2005; 366-377
- [17] 陈力军, 刘明, 陈道蓄, 等. 基于随机行走的无线传感器网络簇间拓扑演化[J]. 计算机学报, 2009(1): 69-76
- [18] 叶秀彩, 许力, 林力伟. 基于小世界现象的无线传感器网络拓扑优化[J]. 福建师范大学学报: 自然科学版, 2008(5)
- [19] Jing Weipeng, Liu Yaqiu. DECD: An energy-aware route protocol for wireless sensor networks [C]// 2nd International Symposium on Systems and Control in Aerospace and Astronautics. 2008
- [33] Srivastava U, Munagala K, Widom J. Operator placement for in-network stream query processing [C] // Proceedings of the Twenty-fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. New York, USA; ACM, 2005; 250-258
- [34] Sameul R M, Michael J F, Joseph M H. TinyDB: an acquisitional query processing system for sensor networks[J]. ACM Transactions on Database Systems, 2005, 30(1): 122-173
- [35] Boukerche A, Samarah S. A Novel Algorithm for Mining Association Rules in Wireless Ad Hoc Sensor Networks [J]. IEEE Transactions on Parallel and Distributed Systems, 2008, 19(7): 865-877
- [36] Ezeife C I, Monwar M. SSM: A Frequent Sequential Data Stream Patterns Miner [C] // Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining. IEEE, 2007; 120-126
- [37] Vincent S T, Kawuu W L. Mining Temporal Moving Patterns in Object Tracking Sensor Networks [C] // Proceedings of the 2005 International Workshop on Ubiquitous Data Management. IEEE, 2005; 105-112
- [38] Nhan D P, Mohamed M G, Uwe R. Resource-aware Online Data Mining in Wireless Sensor Networks [C] // Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining. IEEE, 2007; 139-146

(上接第16页)