

无线传感器网络中的数据查询

潘群华, 李明禄, 张重庆, 张文哲, 伍民友

(上海交通大学 计算机科学与工程系, 上海 200030)

E-mail: qunhuapan@gmail.com

摘要: 无线传感器网络以数据为中心, 提供数据采集、处理和查询功能。数据查询是无线传感器网络研究中的重点和热点问题。本文从查询过程的角度出发, 对目前传感器网络数据查询研究现状进行了系统总结。在此基础上, 考虑传感器网络的异构性, 提出了一种基于感知数据之间的数据信息的查询模型, 并设计了感知数据交互必须的数据标签, 给出了这种数据查询的应用实例。分析表明, 该种数据查询能够大大提高传感器网络的查询效率, 同时增强网络的健壮性, 延长传感器网络使用寿命。

关键词: 传感器网络; 数据标签; 数据查询

中图分类号: TP311

文献标识码: A

文章编号: 1000-1220(2007)08-1357-05

Query Processing in Wireless Sensor Networks

PAN Qun-hua, LI Ming-lu, ZHANG Chong-qing, ZHANG Wen-zhe, WU Min-you

(Department of Computer Science and Technology, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: Wireless sensor networks are data centric, which provide data collection, processing and query services. Query processing is one of the hot spot in the research of wireless sensor networks. These paper overviews recent researches of data query in wireless sensor networks. Then a data collaboration based query model in heterogeneous sensor networks is proposed. The data tag is designed which is necessary for sensory data collaboration. By analyzing a mixed sensor networks application which uses the data tag, the query efficiency and network robust is improved and the lifetime of the sensor network is prolonged.

Key words: sensor networks; data tag; data query

1 引言

随着无线通信、集成电路、嵌入式计算及机电系统等技术的飞速发展和日益成熟, 具有感知能力、计算能力和通信能力的微型无线传感器已经在世界范围内广泛出现。这些传感器具有低成本、低功耗、多功能等特点, 以及无线通信、数据采集、信息处理、协同合作等功能。由这些微型传感器节点构成的传感器网络, 能够协作地实时监测、感知和采集网络分布区域内的各种环境或监测对象的信息, 并对这些数据进行处理, 从而获得详尽而准确的信息, 将其传送到需要这些信息的用户。作为信息感知和采集技术的一场革命, 传感器网络必将被广泛应用于军事斗争、国家安全、环境监测、交通管理、医疗卫生、制造业和反恐抗灾等领域^[28]。

传感器网络是以数据为中心的网络, 传感器网络应用的最终目的是在对目标进行监控并为用户提供有效的信息。一个传感器网络可能包含成千上万个能量严重受限的传感器节点, 这些节点产生大量的数据, 这些数据往往需要传送到基站或发送到用户作进一步的处理。如何快速有效地获取这些数据, 并与此同时最大限度地节省数据收集所耗费的能量, 尽可

能地延长传感器网络的寿命就成为传感器网络研究中的重要课题^[29,30]。许多文献提出了卓有成效的传感器网络数据查询技术, 一些数据查询原型系统比如 TinyDB、Cougar 也已经开发出来。目前, 数据查询已经成为解决传感器网络数据收集的有效技术, 并成为传感器网络研究中的一个热点领域。

本文对国内外对传感器网络的数据分发和查询研究问题和成果进行归纳, 并从查询语言、查询执行、查询结果处理三个过程分别进行描述, 还将从体系结构、能量有效性两个全局性的问题出发讨论查询有效性。最后, 基于对传感器网络数据查询未解决问题的分析和评述, 指出该领域的研究方向, 并提出一种基于数据标签的传感器网络查询解决方案。

2 无线传感器网络中的数据查询过程

传感器网络的数据查询是根据用户的需求, 通过一定的执行策略, 将查询语句或者查询数据包发送到数据源, 取得数据结果, 进行处理后返回给用户的过程。传感器网络的数据查询包括两个阶段: 查询注入传感器网络的阶段与收集数据并返回给用户的阶段。如图1(见下页)所示, 查询注入传感器网

网络的阶段包括如下过程: 用户提交查询请求、查询解析、查询优化、查询分发: 当这个阶段执行完毕, 用户提交的查询请求被解析与优化成有效的查询计划, 并发送到传感器节点, 为下一步数据收集工作做好了准备。

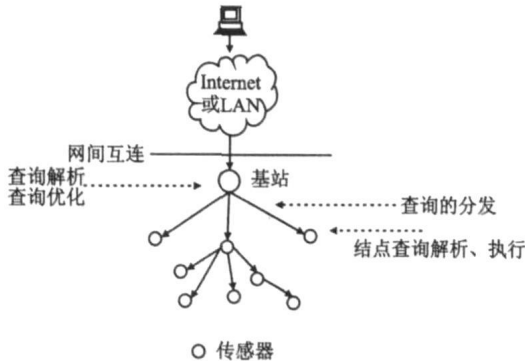


图1 传感器网络查询过程

Fig. 1 Query processing in sensor network

数据查询的第二个阶段是收集数据与返回查询数据给用户. 图2 给出了这个阶段的典型描述. 需要指出的是, 为了有效利用传感器的有效能量, 这个过程往往要对所收集到的数据将进行网内处理, 比如数据聚集、数据压缩等操作。

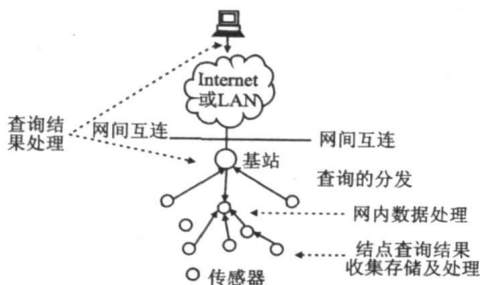


图2 查询数据结果处理过程

Fig. 2 Query result processing

根据用户所感兴趣的数据的类型, 传感器网络的查询可以分成以下几类:

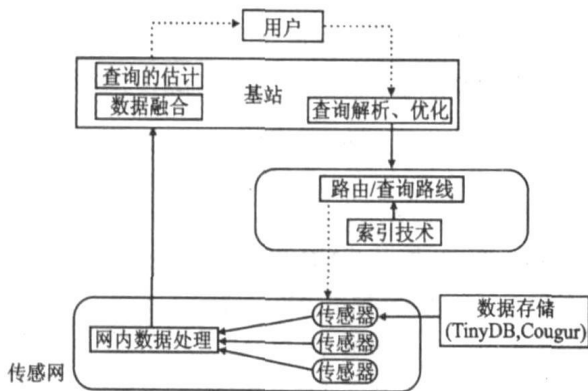


图3 传感网数据查询技术

Fig. 3 Query technologies in sensor network

Y 检测到的酸性污染的平均值”;

2. 非聚集查询, 指点查询或者面查询, 例如查询某个区域是否存在测量值大于等于 a 的传感数据, 并将数据返回;

3. 简单查询, 简单查询是指查询的单一属性的数据值, 例如查询温度属性;

4. 复杂查询, 例如对于温湿度, 污染监控网, “请在位置为 X 的传感器报告当前检测到的温湿度值以及酸性污染值”;

5. 连续查询, 例如 “在未来 10 分钟查询区域 Y 的车辆通过信息, 每隔 1 分钟探测并往回送一次数据”。

实际情况还可能是这几种查询的混合。

每种类型的查询都需要不同的查询技术, 一种类型的查询技术并不见得适用于另一种类型的查询. 总之, 传感器网络的查询涉及到方方面面的技术, 可以把这些技术按查询过程组织成如图 3 所示的图形。

3 传感器数据查询技术

无线传感器网络具有通信能力有限, 电源能量有限, 网络动态性强等特点, 现有的查询都是围绕这些基本特征展开研究, 此外还有对大规模传感器网络以及大量感知数据流等特性展开研究. 图 1 和图 2 是传感网查询的基本结构, 在这两个过程中的每个环节, 带来了许多研究课题和研究成果. 结合图 3, 以下部分将介绍当前的研究成果。

3.1 查询解析

在传统数据库中的查询语言, 并不适合在传感器中执行, 用户提交的查询进入传感器网络时, 需要传感器能够解析和执行的查询语言. 无线传感器网络查询语言方面, Jaikaeo, Chaiporn 提出了一种 Sensor Querying and Tasking Language (SCTL) 的传感器网络查询语言^[1]. 在 Cougar 系统^[2]中, 提出的声明式 (dedare) 查询语言对 SQL 语言稍做修改; 在 TinyDB 中, 介绍了一种获得式查询语言 (Acquisitional Query Language)^[3]; 它们都是类 SQL 的查询语言. 斯坦福的 Shivanath Babu and Jennifer Widom 在对连续数据流的查询的解决方法中, 提出了一种 CQL (Continuous Query Language,) 语言, 支持对流数据的管理. 传感器网络中的数据作为流数据的一种, 也适用这种语言. 传感器网络的查询, 并不一定需要象传统数据库中那样的规范化的语言, 有的只是触发传感器的感知过程, 并将结果返回。

3.2 查询的优化

为了快速的执行查询过程和获得查询数据结果, 索引技术是非常有效的方法, 传感器网络的索引是在最小化利用网络和计算资源的条件下, 通过数据聚集和网络链接, 设计的分布式索引模式. 由于传感器网络中是流数据, 而且网络的拓扑变化比较大, 设计的索引必须能够自适应聚集数据节点的变化以及网络链路发生的变化. 如果数据存储存储在多个数据中心, 多个查询可能将涌入一个数据中心, 引起该节点成为网络瓶颈, Ahmet Bulut 在^[4]中提出一种将数据流在本地存储, 而且将概要内容路由到其它数据中心的方法, 避免泛洪查询整个网络引起网络拥塞的情况. GHI^[5]方法提出了一种基于键

/值的分布式索引方法,每个事件赋予一个键,通过这些键可以恢复事件属性,在GHT中,这些键哈希(Hash)到一个节点的地理位置传感器将获得的数据存储在其周围最近的键节点处。根据这些键对应的地理位置进行查询,路由到这个位置附近的节点,获得查询结果。DIFS^[6]在GHT的基础上,提出了一种基于四分树的索引方法,在所有的子节点基础上构建一个直方图,但是会带来根节点访问过多的问题,其采用四分树对网络中存储的数据进行分层的组织形式,父节点存储4个子节点的信息,依次类推进行存储。Soon-Young Park在[7]中提出了一种基于最小包围框的空间数据索引方法,将传感器组织成R树结构。在每个最小包围框中的传感器有预先聚集的数据集,当执行带有空间范围的查询时,从根节点出发根据最小包围框到达目标传感器,将传感器中聚集的数据发回。和泛洪方法相比,减少了响应时间,节约能量。此外还有支持多维查询的分布式索引结构(Distributed Index for Multidimensional data, DIM)^[26],基于小波变换的DIMENSION系统^[27]等。

查询的优化在数据库系统中是研究比较深的问题,但是在传感器网络中,查询的解析和执行不同,优化的着重点也不同,[11]提出了结果共享和编码的解决方法,对完成例如sum和count、avg的查询功能需要的网络通信消耗进行了优化,指出最小化通信开销是NP问题,同时描述了优化的分布式执行过程,提出了一个在传感器上实现的协议,并验证在不同的查询时进行编码需要的时间和存储空间。[12]提出了一种由客户端,传感器和数据过滤/融合服务器的框架,多个用户通过过滤器或者融合器对传感器数据流处理,满足用户提交的查询。[13]提出一种将MAC协议的设计与作者提出的基于波纹区域的路由结构整合的方法,从能量消耗的角度优化了查询的执行。

3.3 查询的路由方法

基于静态的传感器建立查询路由虽然可以达到比较好的效果,但是在动态的情况下,由于传感器失效,尤其是靠近根部的节点和簇头更容易能量消耗完而死亡,原有的路由将不能使用,需要重新建立路由,会消耗更多的能量,Chen Avin在[8]中提出了一种随机行走的查询方法,不考虑节点失效,在考虑部分覆盖时间的情况下,查询按照随机的方式在传感器网络中行走,到达满足查询的节点后,发回查询结果,在考虑整体节点失效概率和部分区域节点失效概率情况下,效果较好。与此相关的工作是Narayanan Sadagopan在[9]中提出了一种主动查询的方法,主动的查询在传感器网络中流动,在每个停留节点上,收集d跳远的数据,然后该查询随机再向前d跳,到下一个停留节点上,继续收集d跳远的数据,这种方法适用于非聚集类的查询结果。该文对这种方法的能量消耗和查询效率进行建模。直接扩散方法[10]提出了一种限制泛洪查询的方法,在查询的路径上建立一个反向的梯度将数据传回,并对梯度的建立和路由的优化提出了解决方法。

3.4 查询结果处理

以数据为中心的传感器网络,对由查询产生的数据结果往往采用数据聚集的方法,然后进行数据传输。由于传感器的

感知数据是流数据,对流数据的一些研究方法也可以应用于传感器网络的结果处理,例如,对感知数据流进行推理和预测,可以得到未来时刻的查询数据值。在[14]中,对基站上现有的流数据,采用金字塔时间模型维护数据流上不同时刻的聚集值,并利用这些聚集值生成模糊预测系统,通过该预测系统做出预测近似及时地响应用户的查询,采用这种方法,传感器新感知的数据可以不必立即向sink传输,节省了传感器网络节点的能量。[15]对传感器的流数据表示和这些表示与数据之间的映射进行了研究,设计了一种有效内核能够表征这些顺序流数据的相似性的距离函数,提出了一个超级内核函数融合模式优化数据与数据之间的匹配。Apoorva Jindal在[16-18]对传感器网络的数据在空间上是相关连性进行了研究,用统计分析的方法,对经验数据进行分析,提出了一个对空间相关的感知数据简单而准确的模型,并推断了模型的一些参数,这些数据的相关性可以对数据合成与目标追踪非常有用。

3.5 查询的体系结构

由于传感器网络中的查询与传统的关系数据库和分布式数据库的查询有极大的不同,从查询的体系结构出发,也产生了许多有趣的研究结果。[19]定义了一个支持传感器网络中分布查询的通信体系结构,用以评估传感器网络的时空查询其在传感器节点通过查询树和分布查询执行器表示时空查询,由于查询执行需要比较大的计算能力,其提出了在传感器网络中将计算分配至计算能力强的节点上的方法,设计了一个自适应、分布式、低通信能耗的算法来定位资源强的节点,设计了一种支持网内查询出力的能量有效的路由结构,并提出了一种分布式算法TPDA。[21]提出了一种传感器网络系统的中间件结构,用基于命名的属性和节点的分层簇结构,同时也提出了一种SQLT查询任务语言,为传感器应用和SINA中间件之间提供了编程接口。

3.6 跨层的能量有效查询过程

几乎所有的传感器网络查询的研究都涉及能量有效的问题,这里对几个能量有效查询进行总结,当前的研究集中在以数据路由,或者以数据存储方面,[22]基于DCS研究了能量有效的数据组织方法(属性与传感器的逻辑物理匹配),设计了纯本地查询的优化技术。[23]提出了一种“动态能量平衡查询树”的协议,动态调整树结构,最小化广播开销,提出的算法是分布式的,无需全局信息。[13]提出了一种基于区域割裂树区内/区外的路由方法,减少在无线传感器网络中查询时的通信开销。其将路由层和MAC层的设计结合在一起,通过路由层和MAC层的参数交换扩展了传感器节点睡眠时期,延长了网络的寿命。

由于要考虑能量消耗情况以及处理器处理能力有限,传感器网络的数据查询与传统的数据库查询和分布式查询很不相同,当前人们研究了从查询发送开始到结果收集返回之间的整个过程,这些研究还远没有达到实用的程度,而且查询是和实际应用相关的,仍然有许多未解决的问题,此外,结合传统的数据库技术、网络技术,应用到新型的无线传感网中,也是一个值得关注的研究方向。

4 基于数据标签的无线传感器网络数据查询

无线传感器网络数据查询涉及无线传感网的各个方面,从节点系统中的嵌入式软件,通信协议中的路由,体系结构中的拓扑控制,时间同步,定位等方面的问题.当前解决方法都是基于同构的传感器网络的解决方法,而现实中的传感器网络应用越来越多在基于异构的传感器网络间进行协同工作,当前这些查询解决方法虽然可以解决部分的查询问题,节约能量,但是对异构,海量的传感器网络部署,往往达不到预期的效果,在此基础上,我们提出一种基于数据标签的无线传感器网络的网间协作数据查询方法.

异构传感器网络是由不同的传感器以随机或者预先设定的方法部署在关心区域的传感器网络,这些传感器可以感知关心区域的温度、声、光等不同的物理量,传感器本身也由于设计不同,具有不同的通信能力、处理能力和存储能力.例如图4.在关心区域有两种不同的传感器网络.

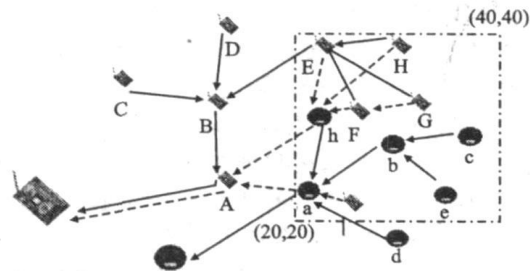


图4 异构传感器网络

Fig.4 Heterogeneous sensor network

按照一般查询方法,在两个传感器网络中分别构建两个查询路由树,查询沿着该路由树到达网内的各个节点,取得结果后,对网内数据结果边传输,边聚集,节约能量.我们可以利用这种混合传感器网络的特性,统一构建路由树,这样可以减少数据传输的跳数,而且使温度传感器网络一些无法通信的远离的孤立节点通过振动传感器网络,获得感知信息.减少数据传输量,延长网络工作寿命,提高网络的鲁棒性.

为了使异构的传感器网络之间的进行信息交互,我们设计一种数据标签,这些数据标签具有统一的标准和格式,不同的传感器网络收集的数据描述信息,传感器本身的描述数据,都可以封装在一个数据标签,作为数据包的头,一个数据包还包括感知数据.格式如图5所示.

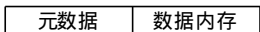


图5 数据标签

Fig.5 Semantic tag

数据标签包括元数据,数据内容,元数据描述传感器本身的特性例如通信距离、剩余能量、剩余存储空间等,以及定义感知数据的类型、精度等.数据内容以压缩的方式存储感知数据.

我们以温度感知数据为例,说明数据标签在两种温度传感器网络数据中的应用.其中的传感器测量的数值分别以华

氏和摄氏为单位.

```

< metadata>                < metadata>
  < location> 20, 20< /location>  < location> 20, 40< /location>
  < type> Temperatur< /type>      < type> Temperature< /type>
  ...
< /metadata>                < /metadata>
< content>                  < content>
  < unit> 摄氏< /unit>          < unit> 华氏< /unit>
  < value> 49< /value>         < value> 200< /value>
< /content>                 < /content>

```

图6 感知数据的语义表示

Fig.6 Semantic expression of sensory data

我们可以从以上的两个数据格式中发现,两个温度传感器网络虽然感知数据的单位不同,但是,通过统一的数据标签,可以进行标示,通过这个标示,为后续的应用程序提供接口.

以下我们说明数据标签在传感器网络中的简单应用.借助数据标签,可以使不同的传感器网络的节点配合进行辅助定位,也可以将不同种类的传感器数据统一在一个节点上进行聚集.

对应图4,我们可以得到逻辑图7,实线表示单个传感器网的查询树,虚线表示查询结果沿着网络协作构成的新的查询树返回sink,从图中可以看到,数据传输的跳数(Hop)减少,而且原来孤立的节点I也能通过新的连接线路工作.

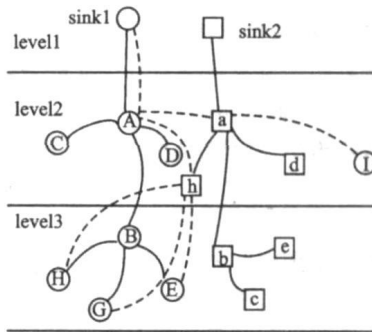


图7 异构传感器网络节点协作逻辑图

Fig.7 Logical view of collaboration in heterogeneous sensor network

数据标签的方法虽然可以提高网络的协调性,但是带来的问题是数据标签需要在传感器中占较大的存储空间,而且对于数据标签的解析,也需要较强的处理能力和内存.

5 总结

本文对传感器网络数据查询的研究现状进行了系统总结,根据当前的研究成果和实际应用,提出了一种基于数据标签的异构传感器网中协作数据查询模型,在该模型基础上,研究了数据标签,异构传感器网络之间进行查询路由的混合构建,和异构数据的聚集问题.通过这种网间的协作机制,提高查询效率,节约了节点的传输能耗,延长了传感器网络的使用寿命.另一方面,孤立节点可以获得重新连接,从而提高了网

络的健壮性. 分析表明, 在这种模型下, 现有数据查询效率可以获得进一步提高. 但是由于数据标签占据一定的存储空间, 数据解析也需要计算功能较强的处理器, 并且占用较多的处理时间, 本模型进一步改进之处在于设计更加小巧的数据标签和快速解析机制.

References:

- [1] Chaiporn Jaikaeo. Querying and tasking in sensor networks[J]. Proc. SPIE 2000, 4037: 184-194.
- [2] Gerhke JI. COUGAR design and implementation [EB/OL]. <http://www.cs.cornell.edu/database/cougar/>.
- [3] Samuel Madden, Michael J Franklin. The design of an acquisitional query processor for sensor networks[J]. SIGMOD June, 2003, 491-502.
- [4] Ahmet Bulut, Ambuj K Singh. Distributed data streams indexing using content-based routing paradigm[C]. IPDPS2005, 94.
- [5] Ratnasamy S B, Karp L, Yin F, et al. GHT: a geographic hash table for data-centric storage[C]. In: Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications, September 2002, 78-87.
- [6] Benjamin Greenstein, Deborah Estrin. DIFS: a distributed index for features in sensor networks[C]. In: Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, 2003, 163-173.
- [7] Soon-Young Park, Hae-Young Bae. A distributed spatial index for time-efficient aggregation query processing in sensor networks[C]. ICCS 2005, 405-410.
- [8] Chen Avin, Carlos Brito. Efficient and robust query processing in dynamic environments using random walk techniques[C]. In: Proceedings of the Third International Symposium on Information Processing in Sensor Networks, 2004, 277-286.
- [9] Narayanan Sadagopan, Bhaskar Krishnamachari, Ahmed Helmy. The acquire mechanism for efficient querying in sensor networks[C]. In: Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, 2003, 149-155.
- [10] Chalermek Intanagonwiwat, Ramesh Govindan, Deborah Estrin, et al. Directed diffusion for wireless sensor networks[J]. ACM/IEEE Transactions on Networking, February, 2002, 11(1): 2-16.
- [11] Niki Trigoni, Yong Yao, Alan Demers, et al. Multi-query optimization in sensor networks[C]. International Conference on Distributed Computing in Sensor Systems (DCOSS2005), 2005, 307-321.
- [12] Guofei Jiang, George Cybenko. Query routing optimization in sensor communication networks[C]. In: Proceedings of the 41st IEEE Conference on Decision and Control, December 2002, 1999-2000.
- [13] Hu Fei, Ankur Teredesai, Wu Hong-yi. Timing-controlled, low-energy data query in wireless sensor networks: towards a cross-layer optimization approach[C]. IEEE International Conference On Networking, Sensing and Control, ICNSC 05, March 2005, 1031-1036.
- [14] Guo Long-jiang, Ren Mei-rui, Li Jing-bao. Study on fuzzy prediction systems over sensor data streams[J]. Journal of Harbin University of Commerce(Natural Sciences Edition), 2005, 21(4): 88-91.
- [15] Y Wu, Edward Y Chang. Distance-function design and fusion for sequence data[C]. In: Proceeding of the 2004 Conference on Information and Knowledge Management, 324-333.
- [16] Apoorva Jindal, Konstantinos Psounis. Modeling spatially-correlated data of sensor network with irregular topologies[C]. IEEE SECON 2005.
- [17] Apoorva Jindal, Konstantinos Psounis. Modeling spatially-correlated sensor network data[C]. IEEE SECON'04, October 2004, 162-171.
- [18] Apoorva Jindal, Konstantinos psounis. A clustering method that uses lossy aggregation of data[C]. Extended Abstract in the Proceedings of ACM Sensys'04, November 2004, 269-270.
- [19] Neha Jain, Ratnabali Biswas, Nagesh Nandiraju, et al. Agrawal energy aware routing for spatio-temporal queries in sensor networks[C]. In: the Proceeding of IEEE Wireless Communications & Networking Conference, WCNC05, March 2005, 1680-1688.
- [20] Bilstrup Urban, Wiberg Per-Arne. An architecture comparison between a wireless sensor network and an active RFID system[C]. LCN 2004, 2004, 583-584.
- [21] Srisathapornphat C, Jaikaeo C, Chien-chung shen. Sensor information networking architecture[C]. International Workshops on Parallel Processing, 2000, 23-30.
- [22] Ramakrishna Gummedi, Xin Li, Ramesh Govindan, et al. Energy-efficient data organization and query processing in sensor networks[C]. In: Proceedings of the 21st International Conference on Data Engineering (ICDE 2005), April 2005, 157-158.
- [23] Yang H, Ye F, Sikdar B. A dynamic query-tree energy balancing protocol for sensor networks[C]. In: Proceeding of IEEE Wireless Communications and Networking Conference 2004, March 2004, 1715-1720.
- [24] Shakkottai, Sanjay. Asymptotics of query strategies over a sensor network[C]. IEEE INFOCOM04, 2004, 548-557.
- [25] Swapnil Patil. IDEA: an iterative deepening algorithm for energy-efficient querying in Ad Hoc sensor networks[C]. ADHOC-NOW 2003, LNCS 2865, 2003, 199-210.
- [26] Xin Li, Young Jin Kim, Ramesh Govindan, et al. Multi-dimensional range queries in sensor networks[C]. Sensys'03, Sept 2003, 63-75.
- [27] Ganesan D, Estrin D, Heidemann J. DIMENSIONS: why do we need a new data handling architecture for sensor networks [C]. In: Proceedings of the First Workshop on Hot Topics In Networks (HotNets-I), Princeton, NJ, 2002, 143-148.
- [28] Akyildiz I F, Su W, Sankarsubramaniam Y, et al. Wireless sensor networks: a survey[J]. Computer Networks, 2002, 38: 393-422.
- [29] Intanagonwiwat C, Estrin D, Govindan, R. Impact of network density on data aggregation in wireless sensor networks[C]. In: Proc. Int'l Conf. Distributed Computing Systems (ICDCS), 2002, 456-458.
- [30] Krishnamachari B, Estrin D, Wicker S. The impact of data aggregation in wireless sensor networks [C]. In: Proc. Int'l Workshop Distributed Event Based Systems (DEBS), July 2002.