

基于 RCFfile 的无线传感器网络数据存储策略研究

闵 林¹ 樊卫北² 郭拯危² 凡高娟²

(河南大学网络信息中心 开封 475004)¹ (河南大学计算机与信息工程学院 开封 475004)²

摘 要 随着无线传感器网络技术的发展,其应用价值遍布环境监测、工农业、抢险救灾、军事国防、生物医疗等许多领域。由于无线传感器网络特征与节点特点,其数据存储与查询策略研究成为热点。对已有的数据存储策略进行了详细介绍与研究,并分析它们的优缺点;其次结合大数据中一种高效的数据存储结构——RCFile,并将其应用到传感器网络的数据存储中,结合行列存储的优势,改变数据存储结构,提出了一种基于 RCFfile 的无线传感器数据存储算法(Wireless sensor network Data Storage based on RCFfile, WDSR),并给出了仿真结果分析。仿真结果表明,提出的算法在低能耗、高效率方面存在一定优势。最后指出了无线传感器网络数据存储算法的发展方向。

关键词 无线传感器网络,数据存储,RCFile,压缩感知

中图法分类号 TP393 文献标识码 A DOI 10.11896/j.issn.1002-137X.2015.4.014

Wireless Sensor Networks Data Storage Strategy Based on RCFfile

MIN Lin¹ FAN Wei-bei² GUO Zheng-wei² FAN Gao-juan²

(Network Information Center, Henan University, Kaifeng 475004, China)¹

(School of Computer and Information Engineering, Henan University, Kaifeng 475004, China)²

Abstract With the development of wireless sensor network technology, it is widely used in many fields, like the military and national defense, industry and agriculture, urban management, biomedical, environmental monitoring, disaster relief. Due to the wireless sensor network nodes characteristics, its data storage and query strategy have become a hot research. In this paper, the existing data storage strategies were described in detail and researched, and their strengths and weaknesses were analyzed. Following by the combination of big data, an efficient data storage structure-RCFile was applied to the data stored in the sensor network. The data structure was changed though combining the row-store and column-store. Based on RCFfile, a data storage algorithm was proposed named wireless sensor network data storage based on RCFfile (WDSR), and then the corresponding results of the analysis were given. The simulation results show that the proposed algorithm has certain advantages in low power and high efficiency. Finally, the trend of the future development of wireless sensor data storage and query strategy research networks was given.

Keywords Wireless sensor network, Data storage, RCFfile, Compressed sensing

1 引言

无线传感器网络(WSN)的特点是外形小、价格低,具有无线通信、传感、数据处理能力,已经应用于科学、军事、经济、生活等众多领域^[1]。WSN由大量节点组成,通过自组织的方式形成网络,节点内的传感器实时感知和收集周边环境中的数据,对数据进行处理,然后反馈给用户。WSN节点实时采集数据,而且每个节点的存储空间及能量都是有限的,如何有效地利用存储空间和快速查询所需数据,成为WSN数据管理和处理的重要研究内容之一。

目前传感器网络数据存储主要有3种方式:外部存储、本地存储和以数据为中心存储^[2]。外部存储中Sink节点主要存储数据,用户发出访问请求也是通过Sink节点。

由于Sink节点是一类特殊的存储节点,如图1所示,它的存储空间和能量不受限制,而且不需要消耗其它节点能量。但是其它节点都要将采集数据传送到Sink节点,这样会消耗大量能量;假如网络中所有节点都向Sink节点发送数据,就会造成网络堵塞,临近的Sink节点会失效。本地存储是将数据都存储在自身节点,这种存储方式消耗的能量很少,查询命令只需要以某种方式发送到各个节点,节点接收并处理后将结果再传送到Sink节点,查询消耗的时延较长,只有当数据的产生频率远远高于数据的查询频率的情况下,本地存储的方式才显得比较高效^[3]。以数据为中心存储的主要特点是以数据为中心,按照某种特定的信息中介机制存储于传感器网络中某些节点,例如特定路由规则、索引或者Hash映射等。以数据为中心存储可以弥补其它两种存储方式的不足,它的

到稿日期:2014-05-22 返修日期:2014-18 本文受国家自然科学基金(61300215, 61103195),中国博士后基金(2014M560439),江苏省博士后科研资助计划(1302084B)资助。

闵 林(1963-),男,教授,主要研究方向为网络管理, E-mail: mlin@henu.edu.cn; 樊卫北(1987-),男,硕士,主要研究方向为计算机网络; 郭拯危(1963-),女,教授,主要研究方向为WSN; 凡高娟(1983-),女,博士,讲师,主要研究方向为WSN, E-mail: fangaojuan@henu.edu.cn。

缺点就是所需要的信息中介机制繁琐且需要付出额外的代价。

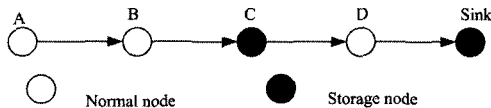


图1 WSN数据传输

本文对国内外传感器网络的数据存储研究问题与成果进行归纳,并与大数据中的存储算法相结合,从存储空间、能量有效性两个全局性的问题出发来讨论存储的有效性。最后,基于对传感器网络数据存储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RCFile的传感器网络数据存储算法。

2 WSN数据存储策略研究

2.1 外部存储

基于层次结构的LEACH协议^[4]用于从传感器网络中收集数据,从WSN中随机选择一部分节点作为簇首,其它节点通过计算不同簇之间的距离来加入。一个周期内,节点向簇首传递数据,簇首加工处理数据后再传送给Sink节点。PEGASIS协议^[5]对LEACH协议进行改进,传感器网络被组织成一个链状结构,每个节点与附近节点接收和转发信息,Sink节点在链上仅选出一个节点与之通信。数据在一个节点传递到另一个节点的过程中聚集,最终到Sink节点,与LEACH相比,消耗较少的能量。用于检测突变数据的数据采集协议TEEN^[6]是对LEACH协议的改进,该协议设立了软、硬阈值,可以动态调整、比对收集到的数据来减少不必要的数据传输。当节点数据高于硬阈值,数据被发送,并将该数据作为新的硬阈值;如果数据小于硬阈值,但与硬阈值差值大于软阈值,数据也被发送出去,并调整为二者的差值。

这种存储策略是面向于数据采集的协议,忽略了WSN的数据存储能力及节点对数据的需求。综上,外部存储更多地用于数据收集,存储过于依赖基站,而且与网络本身分离。

2.2 本地存储

定向扩散Directed Diffusion协议^[7,8]是将节点采集的数据存储在本地节点,Sink节点通过向网络广播一种“兴趣消息”,来获得自己所需的信息。收到“兴趣消息”的节点会在网络内建立一个梯度,指向汇聚节点。该节点就建立了一条或多条到达Sink节点的路径,泛洪查找并进行数据传输。地理和能量感知路由协议GEAR^[9]是基于DD协议的改进,GEAR协议向目标区域内发送查询消息时,考虑了地理位置信息,利用地理位置,使得“兴趣消息”的传播仅限于目标区域,而非整个网络,避免了泛洪,减少了路由的开销。

本地存储策略的存储过程简单,该算法的研究重点在于数据查询处理,但是缺少对信息的描述,查询过程中消耗大量能量。

2.3 以数据为中心存储

以数据为中心存储是近年来数据的热点研究方向,主要研究如何有效存储网络内传感器节点监测的感知数据,从而保证后期查询的高效性、稳定性及实时性。

以数据为中心存储DCS(Data Centric Storage)的概念被Ratnasamy等人提出,并设计了基于地理信息映射表的数据

存储算法GHT(Geographic Hash Table)^[10,11],它的核心思想是:根据属性来存储数据,将特定的数据定义为事件。传感器检测到数据,通过一个hash函数对事件进行散列,然后得到一个地理位置,基于地理位置的路由协议保存到最近节点。这种算法利于数据查询,只需要根据查询的事件属性,采用映射函数即可找到存储节点,避免泛洪。其缺点是缺少高效的存储热点处理机制,数据存储过载时,不能转移到其它节点;获取地理位置信息需要GPS定位,消耗系统的能量。基于自适应环形索引结构的ARI^[12]数据存储算法用来解决DCS算法中的热点问题。利用hash函数将某种类型的事件散列到事件存储节点,在事件存储节点周围建立一个环,事件被分散保存在索引节点。通常,难定义明确的无线传感器网络的界限,对热点问题处理并不理想。将基于两层数据存储结构的TTDD^[13]数据存储算法,用于移动的多汇聚节点的WSN模型中追踪移动目标。该算法是通过建立虚拟网格来传输和存储数据。网格保存节点采集的数据及查询时,只需在网格内泛洪请求即可,能够节省能量,动态事件的建立会使得虚拟网格消耗系统资源。另外有学者提出了分布式结构索引DIFS算法^[14],DIFS在TTDD的基础上,在空间分解技术和散列函数构造多层次二叉树的过程中,采用地理散列方式作为数据的索引。相应索引的节点通过hash函数保存观测到的数据,可以通过查询范围来确定覆盖范围的最小索引节点数。

目前传感器网络的数据存储算法有两个不足:第一,以数据为中心的存储算法在传输过程中会消耗大量能量,节点采集到大量数据,映射原因可能会导致源节点和存储节点距离过远,或者需要的不是单一监测信息,会造成大量能量的消耗;第二,热点问题,预先定义了时间类型,如果对同一时间类型存储或查询操作过多,会造成节点负载,消耗能量快,影响网络寿命。

3 大数据中的RCFile

近年来互联网、云计算、移动和物联网的迅猛发展,使大数据成为一个热点。无线传感器实时监测信息、普及化的移动设备、RFID、数以亿计用户的互联网服务每时每刻都在产生大量交互信息,这些信息数据量大,增长速度快,而业务需求和竞争压力对数据处理的实时性、有效性又提出了更高要求,传统的常规技术手段根本无法应付^[15]。

大数据具有如下4种典型特征:(1)海量。各种企业的数据量的大规模增长。(2)多样性。安装在火车、汽车和飞机上的传感器都增加了数据的多样性。(3)高速。数据被创建和移动的速度非常快。创建实时数据流已成为一种趋势,不仅需要了解如何快速创建数据,还必须知道如何快速处理、分析并返回给用户,以满足他们的实时需求。(4)大数据具有多层结构。与传统的业务数据相比,大数据具有模糊不清和不规则的特性,很难甚至无法使用传统的应用软件进行分析。

3.1 行存储

行存储(Row-store)是将数据放入连续的物理位置,类似于传统的记录和文件系统,然后由数据库引擎根据每个查询提取需要的列。

基于行的存储是将数据组织成多个行,如图2所示,这样就能在一个操作中找到所有的列,其优点是可以快速加载数

据和具有高动态负载适应能力,这是因为行存储保证了相同记录的所有域在同一个集群节点,即同一个 HDFS 块。但是,行存储的缺点也很明显,首先它不支持快速查询处理,因为当仅查询多列表中的少数几列时,它不能跳过不必要的列进行读取;其次,由于列混合着不同数据值,行存储不易获得一个极高的空间压缩比,即空间利用率不易大幅提高。

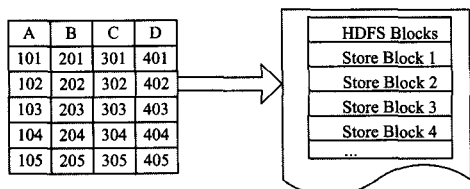


图2 行存储格式

3.2 列存储

列存储法(Column-store)是将数据按照列的格式存储到数据库中^[16],如图3中A列和B列在同一行组存储,而C列和D列分别在单独的列组存储。在查询时,列存储能够避免不必要的列读取,并且将一个列中的相似数据进行压缩,能够实现较高的压缩比。但是,由于元组重构需要较高开销,列存储不能保证同一集群节点存储相同记录的所有域,因此在记录的重构过程中,将导致通过集群节点网络的大量数据传输。再者由于列交叠存在于多个组之间,多余的列数据存储可能会被列组创建,导致存储利用率降低。

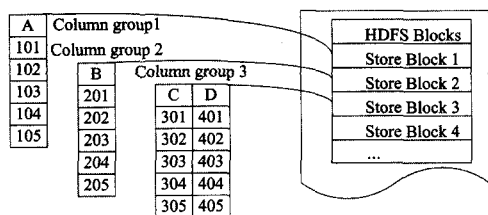


图3 列存储格式

3.3 RCFile 存储结构

RCFile(Record Columnar File)的基本思想是“先水平划分,再垂直划分”^[17]。它结合了行存储和列存储的优势:首先,RCFile 保证同一行的数据存储在同一节点,因此元组重构的开销很低;其次,按照列存储的格式,RCFile 可以利用列维度的数据压缩,并且在读取的过程中能跳过不必要的列。

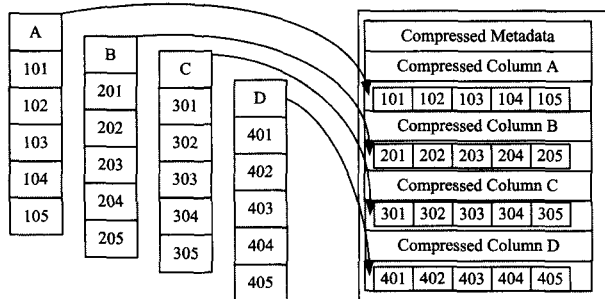


图4 RCFile 存储格式

4 基于 RCFile 的传感器网络数据存储算法

随着 WSN 技术的发展,其规模和应用领域也随之扩大,通常一个监测区域的节点数量众多,分布密集,而且每个节点产生连续的无限数据流,从而产生大量数据,但节点的存储空间

间和能量有限,不能够有效地存储大量信息,这就需要一个有效的存储方法。其次由于数据本身具有多样性和不可靠性的特点,节点在处理数据的过程中,无时无刻不在消耗能量,合理地存储数据能够降低数据转发与查询所消耗的能量,提高节点寿命,从而提高整个网络的性能。

4.1 存储策略评价指标

传感器网络中的数据存储要考虑多种指标,根据不同的指标有不同的分类标准,WSN 中数据存储与查询的评价指标有:

(1) 数据存储空间

目前,数据量变得越来越大,如何有效存储如此大的数据量是一个关键问题。不同的数据存储结构需要不同的数据存储空间。数据存储空间的大小对于评价性能是一个重要因素。

(2) 查询响应时间

查询响应时间对于数据存储结构来说也是一个关键的因素。查询响应时间是非常重要的,这要求底层存储结构能够随着查询数量的增加而保持高速的查询处理。

(3) 数据存储负载均衡性

负载均衡在数据存储中十分重要,特别是在以数据为中心的数据存储中。这样避免了节点瓶颈问题和消耗电量过早,可以延长网络生命周期。用所有节点负载的方差来定义负载均衡的衡量标准:

$$\varphi(S, N) = \frac{\sum_{i=1}^n (L_i - \bar{L})^2}{n}$$

其中, S 为存储算法, N 为传感器网络, $\bar{L}$ 代表网络负载的平均值, L_i 代表节点 i 的负载。

4.2 WDSR 存储过程

WDSR 算法的基本思想是:普通节点将收集到的信息,如温度、湿度、光照等,以相同的数据类型首先按照行的方式存储,保证同一类型数据放在同一节点;其次,按照列存储的优势,利用列维度的数据压缩,将每个列独立压缩,结合节点数据的特点,使用去重有损压缩(RCFile Data Compression, RDC),这样在读取数据时能够跳过不必要的列。

节点内部数据存储结构示意图如图5所示。

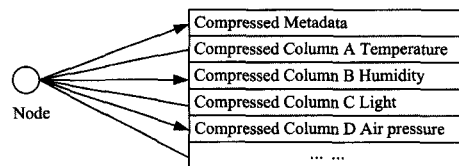


图5 节点内部存储

由于传感器数据的不可靠性,并且每个节点持续产生无线数据流,因此 WSN 有能量供应、通信带宽、处理速度和内存空间均有限的特点。基于查询技术的合理性和有效性等特点,将节点数据分成多个单独的列来存储,相同的数据放在一起排成一行,易得到一个较高的压缩比。如何最大限度利用传感器网络的有限资源成为当今的研究热点。

在 WDSR 算法中,当某个传感器节点感知到一个事件时,根据事件的类型把数据信息进行分类。

如图6所示,行存储首先保证温度、湿度和光照等信息的数据装载进同一节点,然后将相同重复的数据删除,进行有损

压缩,使得读取数据时能够跳过不必要的列读取,减少查询过程消耗的时间。

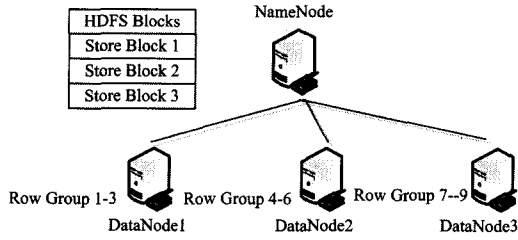


图6 RCFFile 存储格式

5 仿真结果及分析

仿真实验通过 MATLAB 7.0 完成 WDSR 算法的实现。以 $100 \times 100 \text{m}^2$ 的矩形区域为静态传感器网络的监测区域,在区域内随机部署 100 个传感器节点,传感器节点的感知半径设置为 4m,通信半径设置为 8m,随着网络中节点不断感知、转发数据信息,节点的存储空间逐渐下降,查询响应时间出现延迟。

(1) 存储空间

对于同一形状大小监测区域取相同节点得到的数据,使用行存储 (Row-store)、列存储 (Column-store) 与 RCFFile 存储,分别对数据进行存储压缩,得到 3 种存储算法占用的节点空间大小。

图 7 显示了 3 种存储算法所需要的存储空间大小。可以看出,当节点数量低于 120 左右时,采用 RD 存储策略表现的性能并不明显,随着节点数增加,信息量的增大,其性能效果显著。数据压缩可以明显地减少存储空间,不同的数据结构显示不同的数据压缩效果。行存储的压缩效率与列存储、RCFile 相比,其效率最差。如图 8 所示,按照列存储与按行存储的混合数据域相比,数据压缩效果更好。RCFile 与列存储相比可以减少更多的空间,由于行存储把元数据和实际数据存储在起,不能把它们单独压缩。RCFile 把每一列的实际数据和元数据分开存储,然后单独压缩。所以,与列存储相比,RCFile 有更好的压缩比。

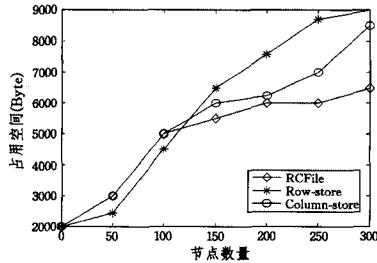


图7 存储策略与原始存储对比

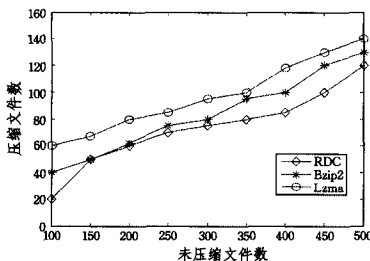


图8 3种不同的压缩算法

数据压缩对于存储空间的使用率起到决定性作用,一个好的压缩算法影响到存储算法的性能。为了验证 RDC 压缩算法的优势,将其与两种常用压缩算法 Bzip2^[18] 与 Lzma^[19] 相比,对相同数据进行压缩,得到相应的压缩比。

图 8 显示了在 RCFFile 的存储过程中,采用不同的压缩算法对相同文件进行压缩的结果。可以看出文中经过优化的压缩算法 RDC,与另外两种压缩算法 Bzip2 和 Lzma 相比,表现出更好的压缩性能。由于传感器节点具有空间有限、能量有限的特点,RDC 压缩算法更适合节点的数据压缩,在提高空间使用率方面更具优势。

对于某些实时监测的节点,用户要及时获得需要的数据,在数据的存储与转发过程中,要尽量降低转发过程中的查询时延。本文算法是改变节点内部数据存储结构,结合行列存储的优势,对于 Sink 节点发出的查询指令,能够快速找到感兴趣的信息。

比较节点内部查询的延迟结果如图 9 所示。当感知节点收到 Sink 节点发出的查询指令后,WDSR 结合行列存储的优势,在查询的过程中跳过不必要的列查询,所以在查询过程中表现出了强大的实时性。相反,TTDD 和 DIFS 由于是两种结构化的查询算法,需要通过映射关系把复本或者查询映射到一个较远的地理位置,如果它们要保证查询成功的高概率,付出的代价就是较长的分发和查询延迟。

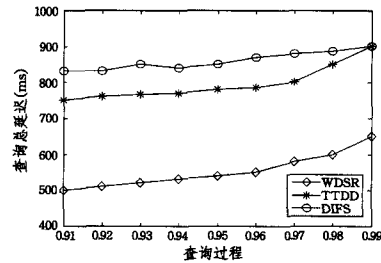


图9 3种存储策略响应时延

WDSR 算法在存储过程中结合了行列的存储优势,相同的数据保存在同一行组,在查询时可以跳过不必要的列,所以在查询的过程中表现了强大的实时性。相反,TTDD 和 DIFS 由于是两种结构化的查询算法,需要通过映射关系把查询映射到一个较远的地理位置,如果它们要保证查询成功的高概率,需付出的代价就是较长的分发和查询延迟,即保证高概率查询成功时,二者必须在延迟上有所折衷。与 DIFS 相比,WDSR 在查询过程延迟上平均缩短了约 30%。

(2) 能量消耗

将本文提出的 RCFFile 存储算法 WDSR,与 TTDD 算法及 DIFS 算法进行对比仿真。通信模型中,每个节点有两种通信模式:发送数据和接收数据。在存储数据和查询数据的过程中消耗总能量。通过能量模型来计算节点的能量消耗,定义发送一条信息消耗的能量为:

$$E_{\text{send}} = E_{\text{trans}} \times s + E_{\text{amp}} \times d^2$$

其中, E_{trans} 指发送一条信息消耗的能量, E_{amp} 指信号放大能量消耗, s 是数据的个数, d 指数据传送的距离。

接收一条信息通过下面公式表达:

$$E_{\text{recv}} = E_{\text{trans}} \times s$$

在仿真中,假设发送和接收一条信息能量消耗相同,每个传感器节点初始能量为 1J,存储空间初始值为 0,最大存储为

200 条数据;假定整个检测区域的事件是均匀分布的。仿真考察本文提出的 WDSR 算法与 TTDD 算法及 DIFS 算法在单个节点内部查询数据消耗的能量。

图 10 给出了对单个存储节点内部事件查询的能量消耗,为了与提出的存储方法(WDSR)进行性能比较,提出了另外两种存储方法:一种是自适应环形索引结构(TTDD),执行 $n=10$ 次随机选择存储节点集合,并选择能耗最少的。另一种是分布式结构索引算法(DIFS)。从图 10 可以看出,随着存储节点数目增加,能耗逐步下降,WDSR 曲线下降更加明显且较为平滑,特别是存储节点数目较大时,TTDD 曲线下降并不明显,甚至出现波动。DIFS 曲线刚开始与提出的 WDSR 类似,后来差异明显,这表明本文的存储方法明显需要更少的能耗,更能延长网络的生命周期,这与 WDSR 以能量均衡为目标之一相符合。

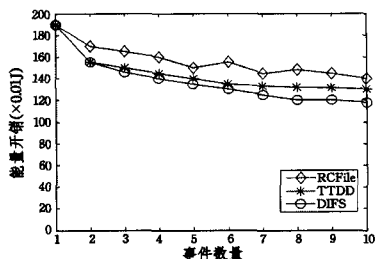


图 10 节点内部查询的能量开销

结束语 本文对无线传感器网络 3 种主要的存储方式做了介绍,并对 3 种存储方式相关的算法做了相关研究,外部存储和本地存储是两种极端的存储方式,以数据为中心存储是两种方式的结合,但仍然存在两个不足:1)数据传输过程中的能量浪费问题;2)存储热点问题。文中提出的基于 RCFile 的 WSN 数据存储算法,能够降低查询过程的能量消耗,并解决存储热点的问题,达到负载均衡,避免了节点过早死亡。

WSN 数据存储和大数据的处理是当今的研究热点,本文对这两方面的研究进行了综述,并提出了基于 RCFile 的 WSN 数据存储模型,为其他学者的更进一步研究提供借鉴作用。随着传感器网络更广泛和更深层的应用,数据存储与查询的问题会更加重要,虽然一些理论模型较为成熟,但具体到实践中,还有一段距离,相信随着科技的发展,会有更多高效节能的方法应用于传感器网络中。

参考文献

- [1] 庄庆德. 传感器网络的研究现状[J]. 国外电子测量技术, 2005 (4): 1-5
- [2] 方红萍, 方康玲. 无线传感器网络数据存储策略研究综述[J]. 计算机工程与设计, 2010, 31(11): 2445-2448, 2499
- [3] Coman A, Nascimnto M A, Sander J. A framework for spatio-temporal query processing over wireless sensor networks[C]// Proc. of the 1st Int'l Workshop on Data Management for Sensor Networks in Conjunction with VLDB. New York: ACM Press, 2004: 104-110
- [4] Heinzelman W R, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless micro sensor net-

- works[C]// The 33rd Annual Hawaii International Conference on System Sciences (HICSS-33). Maui, USA: IEEE, 2000: 1-10
- [5] Lindsey S C, Raghavendra S. PEGASIS: Power-efficient GATHERing in sensor information systems[C]// 2002 IEEE Aerospace Conference Proceedings, 2002: 1125-1130
- [6] Manjeshwar A, Agrawal D P. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks[C]// The 1st Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing. IEEE, 2001: 2009-2015
- [7] Intanagonwiwat C, Govindan R, Estrin D, et al. Directed diffusion for wireless sensor networking[J]. IEEE/ACM Transactions on Networking, 2003, 11(1): 2-16
- [8] Intanagonwiwat C, Govindan R, Estrin D, et al. Directed diffusion: a scalable and robust communication paradigm for sensor networks[C]// Proceedings of MobiCom 2000. Sixth Annual International Conference on Mobile Computing and Networking. New York, NY, USA: ACM, 2000: 56-67
- [9] Yu Y, Govindan R, Estrin D. Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks[R]. VCLA Computer Science Department, Tech Rep: VCLA/CSD-TR-01-0023, 2001
- [10] Ye F, Luo H, Cheng J, et al. A two-tier data dissemination model for large-scale wireless sensor networks[C]// ACM, 2002: 148-159
- [11] Li M, Liu Y. Rendered Path: Range-free Localization in Anisotropic Sensor Networks with Holes[C]// ACM MobiCom. Montreal, Canada, 2007
- [12] Ghosh Y W A, Krishnamachari B. Efficient Distributed Topology Control in 3-DIFS Sensorial Wireless Networks[C]// Proceeding IEEE Comm. Soc. Conf. Sensor, Mesh and Ad hoc Comm. and Networks, 2007: 91-100
- [13] Fan P, Zhi C, Wei C, et al. Asor: An Energy Efficient Multi-hop Opportunistic Routing Protocol for Wireless Sensor Networks Over Rayleigh Fading Channels[J]. IEEE Transactions on Wireless Communications, 2009, 8(5): 2452-2463
- [14] Liu Y H, Zhang Q, Ni L M. Opportunity-Based Topology Control in Wireless Sensor Networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2010, 21(3): 405-416
- [15] 刘正伟, 文中领, 张海涛. 云计算和云数据管理技术[J]. 计算机研究与发展, 2012, 40(Suppl.): 26-31
- [16] 李静, 孙莉, 王梅. 列存储数据查询中的连接策略选择方法[J]. 计算机科学与探索, 2010, 4(9): 858-859
- [17] He Yong-qiang, Lee Ru-bao, Huai Yin. RCFile: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems[C]// 2011 IEEE 27th International Conference on Data Engineering (ICDE), 2011
- [18] Marcelloni F, Vecchio M. A Simple Algorithm for Data Compression in Wireless Sensor Networks[J]. IEEE Communications Letters, 2008, 12(6): 411-413
- [19] Tavli B, Bagci I E, Ceylan O. Optimal Data Compression and Forwarding in Wireless Sensor Networks[J]. IEEE Communications Letters, 2010, 14(5): 408-410