

基于能耗均衡的无线传感网络自适应数据存取算法

胡爱娜

(黄河科技学院物联网传感技术及其应用重点实验室, 智能图像处理与识别重点实验室 郑州 450006)

【摘要】提出了一种能耗均衡的自适应数据存储算法, 将数据尽量存储在适合查询的位置。首先建立基于网格的网络模型, 然后分析采集节点与初始存储节点、查询节点之间的位置关系, 存储模式可在集中存储和分布存储之间切换。算法引入了虚拟扩展网格和能量分级消耗的机制, 可有效均衡网络能耗。实验结果表明, 该算法的节点能耗、生命周期和丢包率相比于GHT算法有一定的改善。

关键词 自适应算法; 均衡能耗; 数据存取; 生命周期; 无线传感器网络

中图分类号 TP393

文献标志码 A

doi:10.3969/j.issn.1001-0548.2014.02.015

Adaptive Data Brokerage Method Based on Energy-Balanced for Wireless Sensor Networks

HU Ai-na

(Key Laboratory of sensor technology and application of IOT, Key Laboratory of Intelligent Image Processing and Recognition,

Huanghe Science and Technology College of China Zhengzhou 450006)

Abstract In order to store data in the most suitable location, an adaptive data brokerage method is proposed in wireless sensor networks. First, a grid-based network model is established, and the relationship of location between acquiring node, sink and initial storage node is analyzed, the storage method can switch between the local storage and data-centric storage. The method is based on virtual extended grid and it can balance the total energy consumption. Compared with geographic Hash table (GHT), the simulation results show that the method can obtain better performance in terms of total energy consumption and network lifetime and packet loss rate.

Key words adaptive algorithm; balancing energy consumption; data brokerage; lifetime; wireless sensor networks

无线传感器网络是以数据为中心的网络, 所以其传感节点的资源受限, 如何均衡能耗、选择合适的数据存储和查询策略是无线传感器网络的研究重点。

由于网络的使用者最关心的问题是否能通过简单的描述快速查询到感兴趣的数据, 所以无线传感器网络中数据存取机制应该考虑的问题是为感知的数据选择合适的存储位置, 使数据便于查询、数据存储和查询能耗最小。这里的能耗主要考虑传感器节点之间信息交互消耗的能量, 它与数据的存储方法有直接的关系^[1]。

目前, 无线传感器网络常见的数据存取方法分为3种^[2-4]: 1) 本地存储, 节点采集的数据存储在节点自身的存储器中。2) 集中存储, 节点将采集到的数据存储在汇聚节点(sink节点)中。3) 分布式存储(data-centric storage), 节点采用分布技术将采集到的

数据存储在任意节点上。3种存储方式各有优缺点, 应该根据不同的应用来选择不同的数据存储方法。

本文针对大规模监测网络中的结构化数据存取算法进行了初步研究。

当数据的采集率远远大于数据的查询率时, 分布式数据存储算法相对合适^[5-6]。GHT(geographic hash table)^[7]是分布式数据存储的典型例子, 它通过哈希函数将采集到的“事件”映射到网络中不同的存储位置。整个网络均通过sink节点来查询数据。当sink节点频繁查询某类数据时, GHT会因资源消耗不均衡而产生严重的“热点(hot spot)”问题。文献[6]针对上述问题提出了一种热点缓解和替代机制, 它通过节点的剩余能量判断某节点趋于热点时, 寻找一个新的节点来替代原存储节点。该算法在一定程度上提高了网络的生命周期, 但是算法的复杂度过

收稿日期: 2012-11-21; 修回日期: 2013-12-17

基金项目: 河南省科技攻关项目(122102310412, 112102310582)

作者简介: 胡爱娜(1979-), 女, 副教授, 主要从事无线传感网络路由算法及网络安全方面的研究。

高。为了均衡网络能耗,文献[8]针对智能电网中大规模数据存储进行了研究,它采用网格内划分环、能耗分级的方法来均衡能耗。除此之外,最近的相关研究见文献[9-12]。

目前已有的分布式数据存储算法中,更多关注的是如何均衡网络的能耗,而对数据存储算法优化的研究相对较少。虽然分布式数据存储算法更符合无线传感器网络的特点,但也存在缺陷:如采集节点和查询节点本身相距很近,但通过哈希映射或构建索引可确定一个较远的存储位置。这样,不仅浪费能量,而且查询的实时性较差。

在上述思想的启发下,本文以“数据存放在最合适的地方”为目标,兼顾数据存取能耗和数据查询效率,将多种模式的数据存储算法融合,提出一种自适应数据存储算法。同时在选择存储节点时采用网络能耗均衡机制,以延长网络的生命周期。

1 本文提出的算法

本文提出了一种自适应数据存储算法,它的主要思想基于GHT算法,依据映射出的哈希存储位置和数据查询的类型(实时查询、非实时查询),确定不同的数据存储方法(集中存储或分布存储)。为了达到能耗均衡的目的,本文还采用了虚拟网格扩展和能耗分级的策略。

为了简化问题模型和分析过程,侧重数据存储算法和能耗均衡性研究,给出如下假设:

1) 查询节点和感知节点采用共同的事件-位置映射表; 2) 节点的位置已知; 3) 除sink节点外,其余节点的初始能量均等; 4) 整个网络均通过sink节点来查询数据,且sink节点的能量不受限。

常见变量声明如表1所示。

表1 常见变量声明

符号	描述
(x_i, y_i)	节点 <i>i</i> 的坐标
dist_{MN}	$\text{dist}_{MN} = \sqrt{(x_M - x_N)^2 + (y_M - y_N)^2}$ (节点 <i>M</i> 、 <i>N</i> 之间的欧几里德距离)
$\text{Hotspot}_{\text{threshold}}$	热点阈值
α	权值
E_R	节点的剩余能量

1.1 网络模型

在大规模的监测网络中,为了节省能耗,尽量缩短存储路径,监测区域被均等划分为若干个二维逻辑网格,每个网格ID记为 (X, Y) ,其计算公式为^[7]:

$$X = \left\lfloor \frac{x_i - x_0}{r} \right\rfloor, Y = \left\lfloor \frac{y_i - y_0}{r} \right\rfloor \quad (1)$$

式中, r 是网格的边长。为了保证每个网格中的节点都可以直接通信,网格的边长 $r \leq (R\sqrt{2}/4)$,其中 R 是节点的通信半径。本文选取的网格边长 $r = (R\sqrt{2}/4)$ 。

如果两点之间的距离超出了节点的通信半径,则由距发送节点最近的网格的网关进行转发,其他节点不转发数据,其中距离网格中心点最近的节点定义为网格的网关。

每个节点包含一个信息索引表,记录了同一个网格内相邻节点的信息。它包括7部分: 1) ID: 节点标识; 2) G_r : 节点所在的实际网格; 3) G_v : 节点所在的虚拟网格; 4) L_r : 节点的位置坐标; 5) L_v : 节点的虚拟坐标; 6) T_i : 节点存储的事件类型; 7) $\text{grid}_{\text{extend}}$: 虚拟扩展网格。

初始化时,节点的虚拟网格和实际网格是一致的,节点的虚拟坐标和实际坐标也是一致的,网格的扩展层 $\text{grid}_{\text{extend}}$ 设置为0, G_v 可以根据实际情况来扩展。

系统模型中有3类数据包: 1) 当节点监测到事件时,将给存储节点发送一个put-packet: (T_i, Data, G_v) 。2) 当汇聚节点查询某事件数据时,将给存储节点发送get-packet: (T_i, G_v) 。3) 当网格中的所有节点的剩余能量达到了阈值,网格中的最后一个达到阈值的节点将给所有的相邻网格发送 $\text{grid}_{\text{extend}}$ -packet: $(G_r, G_v, \text{grid}_{\text{extend}}, T_i)$ 。

1.2 数据存储算法选择

在分布式存储算法有可能出现采集节点和查询节点本身相距很近,却要在一个较远的存储位置交汇的情况,如图1所示。这时,查询时延和数据的存储能耗都会增大。为了将数据能够存储在合适的位置,本文考虑采用自适应数据存储算法。具体算法见算法1。

算法1 自适应数据存储策略。

输入: 采集节点*N*、哈希位置*L*、查询节点(即原点) $S(0,0)$; 以采集节点*N*与查询节点*S*的距离(即 dist_{SN})中点为圆心, $\text{dist}_{SN}/2$ 为半径的圆区域,记为*A*区; 以采集节点*N*为圆心, dist_{SN} 为半径的圆区域,记为*B*区; 除*A*、*B*之外的监测范围,记为*C*区。

输出: 最合适的存储节点。

1) 根据查询类型确定存储算法。

如果要求数据实时查询,则存储位置距离查询节点尽量近,且兼顾采集节点节能的要求。

当哈希位置 $L \in A$,即 $\text{dist}_{LS} \leq \text{dist}_{NS}$,采用分布存储,转入步骤2)。否则,采用集中存储算法。

如果要求数据非实时查询, 则主要考虑采集节点节能的要求。

当 $L \in B$, 即 $\text{dist}_{NL} \leq \text{dist}_{NS}$ 采用分布存储, 转入步骤2)。否则采用集中存储算法。

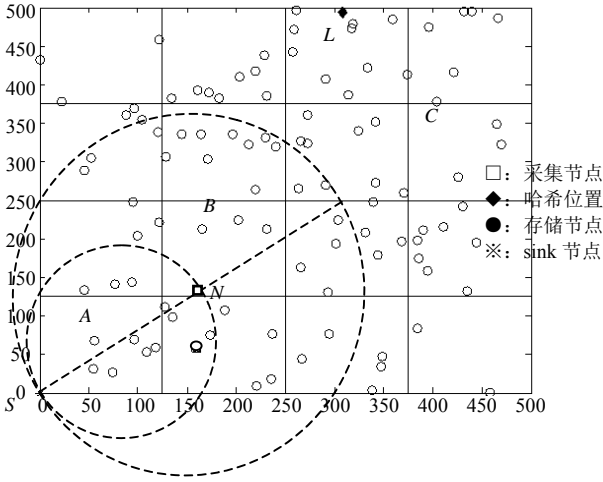


图1 数据存储算法的选择

2) 依据式(1)计算哈希位置所在的网格ID, 并计算该网格中节点的剩余能量是否满足式(2)。如果是, 说明节点趋于某级能耗的“热点”状态, 则将节点的虚拟坐标修改为 (∞, ∞) , 使其进入保护状态。否则, 转入步骤3)。

$$E_R \leq \text{Hotspot}_{\text{threshold}} \overline{E_i} \quad (2)$$

式中, $\overline{E_i}$ 是第 i 级剩余能量的平均值; $\text{Hotspot}_{\text{threshold}}$ 是热点阈值, 它的取值很关键。在实验中, $\text{Hotspot}_{\text{threshold}}$ 取 $[0,1]$ 中任意值对算法能耗均衡性能进行分析。经多次实验发现, 如果该值设置太小, 则节点的负载分配虽均匀, 但分配负载的成本较高; 反之, 节点的负载分配不均衡。故本文算法中, $\text{Hotspot}_{\text{threshold}}$ 取经验值 0.2。

3) 计算上述网格中各节点的影响因子为:

$$Q = \alpha E_R + (1 - \alpha) / \text{dist}(i, N) \quad (3)$$

式中, $\text{dist}(i, N)$ 是网格中任意节点 i 与采集节点 N 之间的欧几里德距离, $\alpha \in (0,1)$ 。 α 取值的不同, 说明了节点 i 的剩余能量与采集节点之间的距离对选择新的存储节点所产生的影响不一样。算法中要求剩余能量和距离采集节点均衡, 即 Q 值最大优先成为存储节点。经多次验证, 本文算法中 α 取值 0.5。

4) 最终找到合适的存储节点, 即本网格中影响因子 Q 最大的节点。

如图2所示, 当节点1监测到类型“Car”的事件后, 通过哈希映射到网 G_v 格 $(2,1)$ 中的位置 L 。然后节点1在本网格中发送 put 包 ($G_v = (2,1)$, $T_i = \text{“Car”}$)。当节点3收到该 put 包后, 就检查自己的索引表中的

G_v 、 T_i 和 put 包中的数据是否一致。如果一致, 节点3将在本网格中广播该 put 包, 那么该网格中的所有节点利用自己信息索引表中的 L_v 来计算自己与采集节点1之间的距离, 并通过式(3)来计算衡量指标。最终衡量指标最大的节点2成为存储节点。

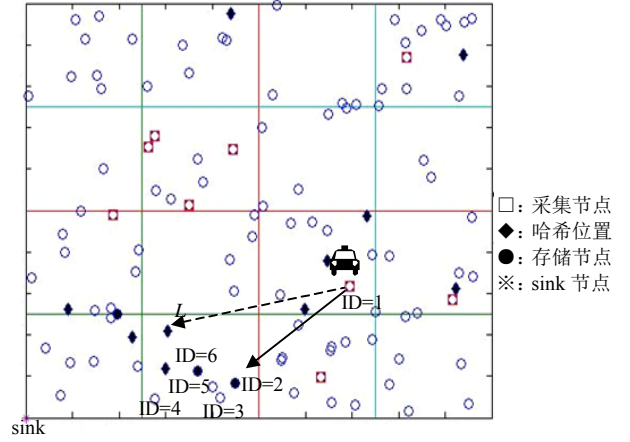


图2 数据存储算法示意图

当图2中的节点2经多次存储数据后, 其剩余能量满足式(2)时, 节点2的虚拟坐标 X 、 Y 将修改为 (∞, ∞) , 进入保护状态。如果再有事件需存储在网格 $(2,1)$ 中, 节点2就不可能成为存储节点, 这种机制可有效均衡节点的能耗。

1.3 能耗均衡机制

当某网格中的所有节点都因剩余能量满足式(2), 坐标均被修改为 (∞, ∞) 时, 如果有数据要保存, 必然导致数据丢失。此时网格中最后一个达到阈值的节点, 将发送一个虚拟网格扩展包给所有的相邻网格, 构成一个稍大的虚拟网格, 如图3所示。具体算法见算法2。

算法2 虚拟网格扩展及能量分级存储算法。

输入: 节点的总能量 E , 将能耗均匀划分为 n 级, 发出网格扩展包的节点所在的虚拟网格 (X, Y) 。

输出: 信息索引表的更新。

- 1) 将节点的能量消耗分为 n 级, 本文取 $n=2$;
- 2) 判断某网格中所有节点的剩余能量是否均处于第 i 级, $1 \leq i \leq n$;
- 3) 如果是, 则任意满足式(4)、式(5)中任一条件, 均可进入扩展虚拟网格中。

$$X_{\text{extended grid}} = X - \text{grid}_{\text{extend}} \quad (4)$$

$$Y_{\text{extended grid}} = Y \pm \text{grid}_{\text{extend}} \quad (5)$$

4) 当网络中所有节点的剩余能量均处于第 i 级, 节点的能耗转入第 $i+1$ 级, 并转入步骤2) 继续执行, 直至节点的能耗进入第 n 级。

如图3中网格 $(4,3)$ 中节点2、3、4均因能量消耗

后剩余能量满足式(4), 而将虚拟坐标修改为 (∞, ∞) 。当节点5最后接收一个包后, 剩余能量也达到了第2个阈值, 这时网格(4,3)就没有合适的节点来存储数据, 那么节点5将发送一个网格扩展包给相邻网格, 网格扩展包的内容为 $(G_r, G_v, \text{grid}_{\text{extend}}=1, T_i)$ 。通过式(7)、式(8)可以计算出网格(4,4)、(3,3)、(3,2)、(4,2)、(3,4)、(5,1)、(5,2)、(5,3)进入了扩展后的虚拟网格(4,3)中, 那些位于扩展网格中的节点现在需要将信息索引表中的 G_v 的值修改为(4,3), $\text{grid}_{\text{extend}}=1$, 能量消耗进入第3级。

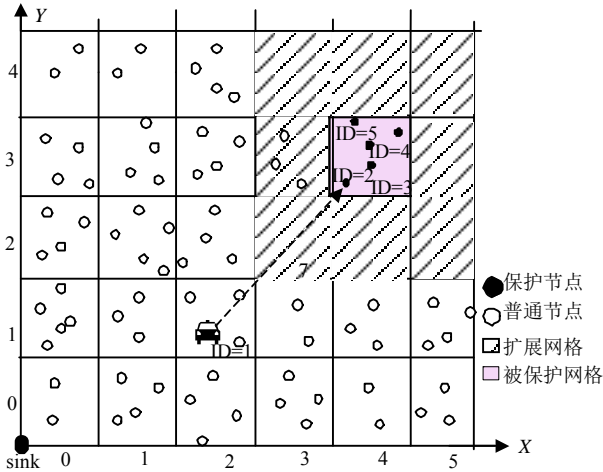


图3 虚拟网格扩展示意图

如经过一段时间的运行, 扩展后的虚拟网格中也没有合适的存储点, 那么该网格将继续往外扩展, 直至遍历整个网络。当整个网络的能耗进入第 n 级(最高级), 所有节点的虚拟坐标将重新变成实际坐标, 接收下一轮的存取任务, 直至网络寿命终止。

虚拟网格扩展及能量分级存储算法在一定程度上均衡了整个网络节点的能量消耗, 避免了存储和查询过于集中时某些节点过早失效的问题。

1.4 数据查询

当sink节点需要查询某事件数据时, 首先查找该数据是否存储在sink节点上。如果不在, 可利用事件类型去计算存储节点所在的网格。sink节点发出get包 (T_i, G_v) 。收到get包的节点将检查自己的信息索引表中的事件类型、虚拟网格是否一致。如果不一致, 继续向其他网格转发该包, 直到找到存储数据的节点所在的虚拟网格, 保存有相关数据的节点将数据传送给汇聚节点。

2 实验结果及分析

为了评估本文算法的性能, 采用Matlab进行仿真, 针对无线传感器网络的生命周期、整体能耗、

负载均衡性和查询时延等方面与GHT算法进行对比分析。在本文的系统模型中, 传感节点最多可有两种工作模式: 采集模式和存储模式。

为了能够和GHT算法进行对比分析, 本文在设置参数时参考了GHT算法的实验参数, 如表2所示。

表2 实验参数值

参数	描述	设置
N	传感节点的个数	2 000
$X*Y/m^2$	感知区域的大小	500*500
R/m	传输范围	50
$E_{\text{trans}}/nJ\cdot\text{bit}^{-1}$	发送一位信息的能量消耗	50
$E_{\text{rec}}/nJ\cdot\text{bit}^{-1}$	接收一位信息的能量消耗	50
$E_{\text{amp}}/nJ\cdot\text{bit}^{-1}\cdot\text{m}^{-2}$	放大器的能量消耗	0.1
S_{event}/B	事件信息的大小	8
S_{query}/B	查询包的大小	4

2.1 能耗均衡性

在分布存储的传感器网络中, 能耗均衡是一个重要的衡量指标^[14]。能耗均衡性的衡量标准可以用节点剩余能量的方差来描述:

$$\varphi = \frac{\sum_{i=1}^n (E_i - \bar{E})^2}{n} \quad (6)$$

式中, E_i 为节点 i 的负载; \bar{E} 是网络负载的平均值; n 为网络中所有节点的个数。 φ 值越大说明负载均衡性越差, 反之, 说明负载均衡性越好。

实验结果表明, 本文采用了网格扩展和多能量分级机制, 能耗均衡性比GHT算法有所改善, 如图4所示。

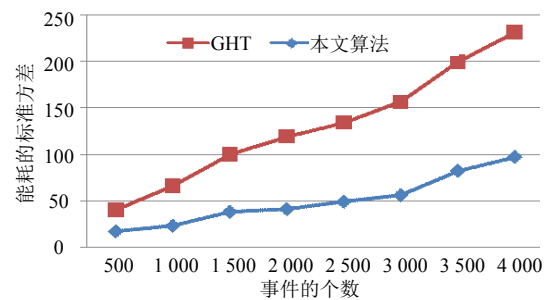


图4 网络的能耗均衡性

2.2 网络的整体能耗

本文的能耗计算公式分别如下:

发送信息能耗:

$$E_{\text{translation}} = E_{\text{trans}}s + E_{\text{amp}}d^2 \quad (7)$$

接收信息的能耗:

$$E_{\text{receive}} = E_{\text{rec}}s \quad (8)$$

查询处理的能耗与查询次数和节点上存储的数据量成正比。

$$E_{\text{query}} = e_{\text{query}}sn_{\text{query}} \quad (9)$$

式中, n_{query} 表示查询的次数; e_{query} 表示一次处理单位大小的数据的能耗; s 表示数据的大小; d 表示传输的距离。

实验结果如图5所示, 表明本文算法的总消耗要低于GHT算法, 分析原因如下:

1) 数据存储阶段的区别。GHT需要在整个网络中寻找存储点, 最终将信息存储在距离哈希位置最近的节点上。而本文算法在保证数据存储位置相对合适的情况下, 兼顾查询能耗和时延较小。具体特点是: ① 将监测网络均等划分为若干个网格, 可减少寻找存储节点的范围; ② 数据存储在哈希位置所在的网格中距离采集节点最近的节点上; ③ 将集中存储和分布存储有机结合, 可以将数据尽量存储在适合的位置。

2) 数据查询阶段的区别。GHT中所有的信息都分布存储在节点上, 而本文算法中部分数据直接存储在sink节点上, 无需查询, 这样查询时消耗的能量可降低很多。

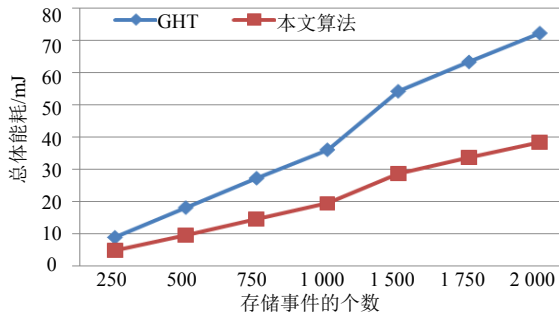


图5 节点的整体能耗

2.3 网络的查询成功率和生命周期

查询成功率是查询成功次数与查询次数之比, 是无线传感器网络性能的根本体现。

1) 节点密度对查询成功率的影响。随着节点数量的增加, 算法的查询成功率并没有明显的变化。当节点稀疏时, 采集节点若找不到合适节点来存储信息, 将会把信息直接交给sink节点, 即采用集中存储算法。

2) 存储事件与查询成功率之间的关系。本文在节点密度一定的情况下, 随着网络中存储事件的不断增加, 查询的成功率会明显下降。特别是某一类事件频繁发生时, 产生大量的同类型感知数据, 导致某片监测区域内节点的能耗急剧下降, 甚至死亡。这也是本文算法需要进一步改进的地方。

图6、图7表明了两种算法的查询成功率、生存周期的不同。这是由于两种算法在数据存储和能耗均衡机制有很大的区别。本文算法由于降低了整体

能量消耗、采用了网格扩展、能耗分级的通信负载均衡机制, 在一定程度上延长了网络的生命周期。

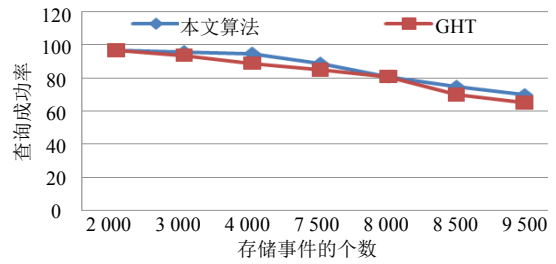


图6 网络的查询成功率

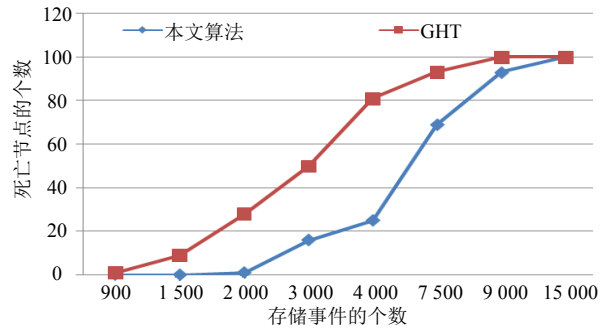


图7 网络的生命周期

但是, 本文提出的算法在事件个数从4 000开始增长的过程中, 相对于GHT算法, 死亡节点的个数急剧增加。可能由于某类事件频繁发生时, 对应的存储节点能耗过快而死亡, 这需要进行进一步优化改进。

3 结束语

本文针对大规模监测网络中的结构化数据存储方法进行了改进, 改进的算法可依据映射出的哈希存储位置的不同而选择更适合查询的自适应存储方法。实验结果显示, 本文为了避免局部能耗过快而采用虚拟网格扩展和能耗分级的策略, 在一定程度上均衡了网络能耗, 延长了网络的生命周期。

本文下一步的工作包括:

1) 对自适应数据存储算法针对某类数据频繁查询时的性能进一步优化; 当某类事件频繁产生时, 节点能耗的均衡性有待改善。2) 为了突出研究的重点, 本文算法忽略了一些影响因素, 如扩展虚拟网格的能耗对算法整体性能的影响, 数据产生率、数据查询率的变化、外界干扰等, 这些都是客观存在的实际问题。下一步将本文算法进一步改进后应用于实际的无线传感器网络项目中, 进行验证和优化。

参 考 文 献

[1] 蔚赵春, 周水庚, 肖斌. 无线传感器网络中自适应数据存取[J]. 软件学报, 2008, 19(1): 103-115.

- YU Zhao-chun, ZHOU Shui-geng, XIAO bin. Adaptive information brokerage in wireless sensor networks[J]. Journal of Software, 2008, 19(1): 103-115.
- [2] YU G J. Adaptive storage policy switching for wireless sensor networks[J]. Wireless Pers Commun, 2009(48): 327-346.
- [3] 蔚赵春, 周水庚, 关信红. 无线传感器网络中数据存储与访问研究进展[J]. 电子学报, 2008, 10(36): 2001-2015.
YU Zhao-chun, ZHOU Shui-geng, GUAN Ji-hong. Data storage and access in wireless sensor networks: a survey[J]. Acta Electronica Sinica, 2008, 10(36): 2001-2015.
- [4] LIAO W H, CHEN C C. Data storage and range query mechanism for multi-dimensional attributes in wireless sensor networks[J]. IET Communications, 2010, 4(15): 1799-1808.
- [5] 马礼, 唐长茂. 一种基于分簇的无线传感器网络数据存储管理系统[J]. 计算机研究与发展, 2011(48): 89-93.
MA Li, TANG Chang-mao. A clustering based data storage management system in wireless sensor networks[J]. Journal of Computer Research and Development, 2011(48): 89-93.
- [6] TAHANI M, SABAEI M. A distributed data-centric storage method for hot spot mitigation in wireless sensor networks[C]//2010 5th International Symposium on Telecommunications (IST'2010). Tehran: IEEE, 2010: 401-408.
- [7] RATNASAMY S, KARP B, SHENKER S, et al. Data-centric storage in sensornets with GHT, a geographic Hash table[J]. Mobile Networks and Applications, 2003, 8(4): 427-442.
- [8] 宋宝燕, 张洪梅, 王研, 等. 支持大规模智能电网的数据存储算法[J]. 计算机应用, 2012, 32(9): 2496-2499.
- SONG Bao-yan, ZHANG Hong-mei, WANG Yan, et al. Data storage method supporting large-scale smart grid[J]. Journal of Computer Applications, 2012, 32(9): 2496-2499.
- [9] LIAO Wen-hwa, SHIH Kuei-ping, WU Wan-chi. A grid-based dynamic load balancing approach for data-centric storage in wireless sensor networks[J]. Computers and Electrical Engineering, 2010(36): 19-30.
- [10] LIAO Wen-hwa, HUANG Chong-hao. An efficient data storage scheme for top-k query in wireless sensor networks[C]//2012 IEEE Network Operations and Management Symposium. Maui, HI: IEEE, 2012: 554-557.
- [11] LIAO Wen-hwa, YANG Huang-chun. An energy-efficient data storage scheme in wireless sensor networks[C]//2012 IEEE Network Operations and Management Symposium. Maui, HI: IEEE, 2012: 478-481.
- [12] TALARI A, RAHANAVARD N. Cstorage: Distributed data storage in wireless sensor networks employing compressive sensing[C]//IEEE Communication Society. Houston, TX, USA: IEEE, 2011: 1-5.
- [13] HEINZELMAN W R, CHANDRAKASAN A, BALAKRISHNAN H. Energy-efficient communication protocol for wireless micro-sensor networks[C]//Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS). [S.l.]: IEEE, 2000: 6530466.
- [14] 李志刚, 肖依. 无线传感器网络中对等数据存储策略研究[J]. 计算机工程与科学, 2012, 34(5): 24-30.
LI Zhi-gang, XIAO Nong. Research on the P2P data storage strategy in wireless sensor networks [J]. Computer Engineering & Science, 2012, 34(5): 24-30.

编辑 漆蓉