

## Research on the Scheme of Data Aggregation Based on Clustering for Wireless Sensor Network

ZHANG Qiang, LU Xiao\*, CUI Xiaochen

(Telecommunication Engineering Institute of the Air Force Engineering University, Xi'an 710077, China)

**Abstract:** Data aggregation is hotspot researching and an important energy saving techniques in wireless sensor network at present. Based on clustering, the paper proposes a new scheme of data aggregation, which has different aggregation strategy on nodes of cluster member and cluster head. The cluster member uses comparatively entropy to reduce the sending data, and the cluster head maintains a feed back value, which uses to estimate whether transmit the data packet when it receives from the cluster member or from itself's sensor module. By the contrast experiment with LEACH, simulation results show that the new scheme can decrease the number of data transmission in network, reduce more power consumption, observably prolong the lifetime of network.

**Key words:** wireless sensor network (WSN); data aggregation; data fusion; clustering

EEACC: 6150P

doi: 10.3969/j.issn.1004-1699.2010.12.022

## 基于分簇的无线传感器网络数据聚合方案研究

张 强, 卢 潇\*, 崔晓臣

(空军工程大学电讯工程学院, 西安 710077)

**摘 要:** 数据聚合技术是目前无线传感器网络中的研究热点, 同时也是一种重要的节能技术之一。在基于分簇网络拓扑结构的基础上, 提出了一种新的数据聚合方案。分别对簇内成员节点和簇头节点进行数据聚合处理, 簇内节点引入相对信息熵减少数据量的发送, 而簇头节点维持一个反馈比较值, 当接收到簇内成员节点发送的数据或得到自身传感器模块的数据时, 该值可以用来判断是否转发接收到的数据。通过与 LEACH 协议的仿真对比实验, 结果表明新方案能有效减少网络中的数据包传送数目, 降低节点能耗, 并显著地延长了网络寿命。

**关键词:** 无线传感器网络; 数据聚合; 数据融合; 分簇

中图分类号: TP393

文献标识码: A

文章编号: 1004-1699(2010)12-1778-05

无线传感器网络是由部署在监测区域内的大量传感器节点组成, 通过无线通信形成一个多跳的自组织的网络系统, 其目的是协作地感知、采集和处理网络覆盖区域中被感知对象的信息, 并发送给观察者。如图 1 所示。其出色的实时检测能力使得人们可以从更为细微的角度了解周围环境或个体现象所出现的现象、所处的状态, 并根据物理环境变化采取措施实行反向控制, 为人类与客观物理世界的交互提供了一种新的有效手段<sup>[1]</sup>。随着研究的深入, WSN 能够广泛应用于军事、环境监测和预报、健康管理、智能家居、建筑物状态监控等领域。

大量节点密集部署的特点导致大部分数据是冗余的, 数据的重复传输必然消耗节点能量, 同时也必

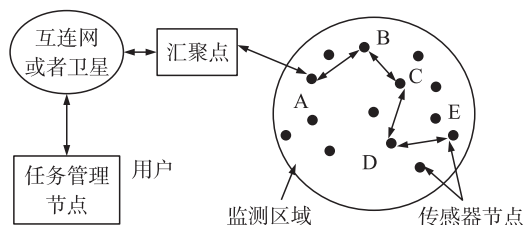


图 1 无线传感器网络体系结构

将引起无线通信的带宽拥挤, 增加信道协商或竞争过程造成的能量开销。而且传感器节点的绝大部分能量消耗在无线通信模块, 实验表明传输 1 bit 信息 100 m 需要消耗的能量大约相当于执行 3 000 条计算指令消耗的能量。因此在无线传感器网络中采用数据聚合机制, 删除冗余数据, 同时将来自不同节点

的信息进行聚合处理,减少网络数据传输的数据量,就可以减少网络能量消耗。数据聚合利用的是节点的计算资源和存储资源,只要将增加计算量的能量消耗控制在小于降低通信量的能量消耗,就可以达到节省能量延长网络寿命的目的。

数据聚合可以减少 WSN 网络中传输的分组数量,减少分组的冲突概率,减小 Sink 节点接收到重复分组的概率以及提高监测结果的准确性。它是目前 WSN 的一个研究热点,其中包括聚合点的选择、聚合时机的选择和数据聚合策略等三个问题<sup>[2]</sup>。目前已提出了很多数据聚合策略算法。文献[3]提出一种数据融合树算法,即汇聚节点作为树的根节点,通过反向组播树的形式从分散的传感器节点逐步将监测数据收集,树上的每个中间节点都对收到的数据进行融合处理。而树的建立时间复杂度为  $O(n^3)$ ,  $n$  为传感器节点个数。文献[4]中提出了一种基于移动代理的数据融合方法,监测数据保留在本地节点,当移动代理迁移到数据处进行融合处理。文献[5]中提出自适应的数据融合算法,汇聚节点根据节点的可靠度建立控制信息,网络中数据的传输则建立在控制信息的可靠度  $P$  基础上,可靠度  $P$  则反映了汇聚节点收到节点数据包个数。而大量节点所带来的控制信息将造成网络性能降低。文献[6]提出一种使用模糊逻辑推论方法用于分簇网络拓扑结构以达到数据融合的算法,算法分为四个步骤:模糊化、规则评定、规则融合和去模糊化。文献[7]提出了基于主元分析的神经数据融合算法。文献[8]提出了一种基于信息熵的分簇网络数据融合方案。在数据聚合的聚合时机研究方面,文献[9]对周期性检测应用中的周期性简单数据聚合时机模型、周期性逐跳数据聚合时机模型和周期性逐跳调整数据聚合时机模型等三种数据聚合时机模型进行了比较研究。文献[10]为事件触发类数据回传提出了一种确定数据时机的协议——MFS,当 WSN 中发生某个事件时,事件附近区域的多个传感器节点将采集数据并报告给汇聚节点。在本文中,笔者主要在分簇网络拓扑结构的基础上讨论数据聚合技术,并结合分簇路由协议提出了一种新的数据聚合方案。

## 1 数据聚合概念

数据聚合或数据融合是指利用传感器节点的本地处理能力,先对采集到的或接收到的其他传感器节点发送的多个数据进行网内处理,消除冗余信息,然后再传输处理后的数据。如图 2 所示。这与传统的多传感器数据融合有所不同,同样是在军事应用背景中,后者

是指对空间分布的多源信息,对所关心的目标进行检测、关联、跟踪、估计和综合等多级多功能处理,以更高的概率或置信度得到所需要的目标状态和身份估计以及完整的及时的态势和威胁评估,为指挥员提供有用的决策信息。它的数据源主要是具有不同体制和功能的雷达。而 WSN 数据融合或数据聚合主要是指单种传感器为了减少网络内的数据传输量,达到减少能源消耗、延长网络寿命的目的。它的数据源主要是同类的传感器节点。为区别传统的多传感器数据融合,本文使用数据聚合一词。这里的数据聚合也不同于信息融合,信息融合是对来自不同来源、不同模式、不同介质、不同时间、不同地点、不同表示形式的信息进行综合处理,从而获得更为准确、可靠的结论。融合的信息来源可能是数据库或知识库中的先验信息;也可能是人机交互中的人工输入信息。它是对人脑综合处理复杂问题能力的一种较全面的、高水平的模仿。而 WSN 数据聚合指中间节点对采集或接收到的多个数据进行合并,具体处理有:几个数据任选一个,计算数据的平均值、最大值或最小值等<sup>[2]</sup>。

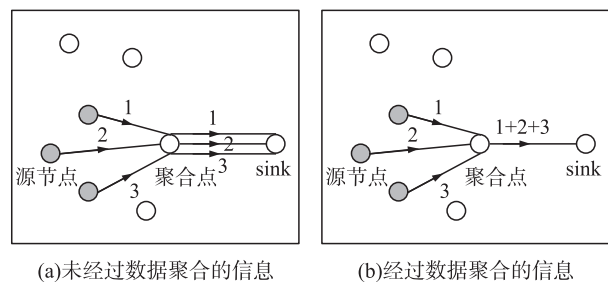


图 2 数据聚合处理图

## 2 数据聚合方案

本文中数据聚合的研究建立在 WSN 分簇的网络拓扑结构上,聚合点将选择簇头节点,并且结合分簇路由协议建立的路由回传数据;聚合时机主要考虑周期性数据回传应用模型。如果数据聚合只在监测节点进行,数据的准确性将受到较大影响;同样如果数据聚合只在汇聚节点进行,监测节点在发送数据时能量消耗较大,同时将造成无线信道拥挤。在分簇网络中,共有三类节点,分别是簇内节点、簇头节点和同时担任中转其它簇头数据的簇头节点。因此,笔者考虑在三类节点上各自采用不同的数据聚合策略,确保数据准确性和能量高效的平衡。

### 2.1 网络模型

本文中 WSN 采用分簇网络拓扑结构,分簇算法将整个传感器网络划分为若干个簇,每个簇包含一个簇头和多个簇成员节点。簇成员节点只需将采集到的数据发送至簇头,簇头负责管理簇内成员节点,

收集簇成员的监测数据,并将收集到的数据进行聚合处理后通过多跳方式报告至汇聚节点。LEACH 协议中所做的假设将同样适用本文中的研究,另外还有如下假设:

(1)网络中监测区域内部署着同种类型的大量传感器节点;

(2)相邻节点间的监测数据具有相似性,节点分布冗余并且检测得到的数据存在冗余信息,可进行数据聚合处理;

(3)各个节点周期性收集监测区域的数据,周期结束节点处理监测数据并决定是否转发;汇聚节点同样采用周期性地收集各个节点的数据。

(4)WSN 所要监测的物理量可以由部分传感器节点的监测数据聚合得到,而并非必须网络中全部传感器节点。

## 2.2 簇内节点聚合策略

WSN 内节点部署高度冗余,监测数据也具有冗余性,通过减少数据发送量的方法可以达到数据聚合的目的。簇内节点减少监测数据发送的方法可以采用相应的 MAC 协议根据时间片轮转来关闭部分节点,也可以降低节点收集数据频率。在本文中,笔者采用相对信息熵的方法,各簇内节点都将保留前一次向簇头节点发送的监测数据,当节点再次得到新的监测数据时,计算两个数据的相对信息熵值,当比较值具有明显差异时才向簇头节点发送,同时更新保留的数据。否则不发送新的检测数据。

信息熵是 Shannon 在 1948 年提出,并作为一个随机变量的不确定性或信息量的度量。如果某事物  $X$  具有的独立可能的结果为  $x_1, x_2, \dots, x_n$ , 并且每一种可能出现的概率分别为  $p(x_1), p(x_2), \dots, p(x_n)$ , 则该事物的信息熵为:

$$H(x) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i),$$

$$\sum_{i=1}^n p(x_i) = 1. \text{ 信息熵的物理意义是在 } X \text{ 中所包含的信息的量。}$$

假设  $P$  和  $Q$  是 2 个概率分布函数,  $P$  相对于  $Q$  的信息距离即相对信息熵<sup>[11]</sup> 为:  $D(P \parallel Q) = \sum_{i=1}^n p_i \log_2 \left( \frac{p_i}{q_i} \right)$ 。相对信息熵的物理意义是信息获取量,表征了 2 组概率分布之间的差异性程度,因而对于 2 组不同的概率分布  $P$  和  $Q$ , 计算其相对信息熵  $D(P \parallel Q)$  以及  $D(Q \parallel P)$ , 如果这 2 个值越小,表明 2 组概率分布越近似,数据冗余则越大。因此,对于簇内节点  $n_i$  新旧得到的监测数据  $y_i$  和  $y_i'$ , 如果其相对信息熵  $D(y_i \parallel y_i')$  和  $D(y_i' \parallel y_i)$  满足条件

$\max(D(y_i \parallel y_i'), D(y_i' \parallel y_i)) \leq \varepsilon$ , 其中  $\varepsilon$  根据实际应用设定为足够小的常量,则认为两组数据间存在较大冗余,也即所监测的物理量没有显著变化,所以节点可以不发送最新收集到的监测数据。否则按照该簇内节点的路由表,将新的监测数据发送给该簇的簇头节点,同时更新该簇内节点的监测数据值。

## 2.3 簇头节点聚合策略

在本文的分簇网络拓扑结构中,基于能量节省的角度考虑,簇头节点与汇聚节点间采用多跳通信方式。因此在聚合策略上,普通簇头节点只负责聚合本地监测数据和其簇内成员节点发送的数据,而担当转发簇头数据的簇头节点不仅负责普通簇头节点聚合的数据,同时还需将其他簇头节点发送过来的数据进行聚合。两种簇头节点的聚合策略为此有所不同。下面先介绍普通簇头节点的聚合策略。

普通簇头节点处理的监测数据包括自身传感器模块监测的数据和簇内各成员节点所发送的数据。每个簇头节点保存前一次转发的监测数据作为反馈比较值  $E$ , 初始值为零。当簇头节点接收到其成员节点的新数据或是当其自身传感器模块得到新数据时,该值的作用是决定新数据是否转发出去。实际监测环境中存在噪声误差,并且聚合多个传感器节点的数据存在着不确定性。在本文中节点传感器模块得到的测量值采用对称三角模糊数来表示,节点  $n_k$  的自身监测数据经过模糊化处理为  $V_k = (v_k, vs_k)$ , 其中  $v_k$  为对称三角模糊数  $V_k$  的中心值,  $vs_k$  为  $V_k$  的模糊宽度,该值与测量误差相关。同样估计值  $E$  也为模糊数,表示为  $E_i = (e_i, es_i)$ , 而这些模糊数值都将在汇聚节点处去模糊化。簇头包含多个成员节点,分别采用 TDMA 方式通信。当成员节点  $n_i$  发送的数据  $V_i$  到来时,  $n_k$  进行数据聚合。输入由三部分组成,分别是自身传感器得到的  $V_k, V_i$  和  $n_k$  维持的回馈比较值  $E_k$ 。聚合输出得到新值  $nE_k$  取代旧的  $E_k$ 。新值  $nE_k$  将与聚合表中的值进行比较,两者的绝对差值如果大于预先设定的阈值,阈值取聚合表中所获得两组数据间之差的最大值,则说明监测区域出现异常变化需立即转发  $nE_k$ , 则依据路由表中的路由发送数据到下一簇头节点。否则将新值  $nE_k$  存入聚合表中。簇内通信一轮完成后,簇头将对聚合表中所有的比较值  $nE_k$  进行平均值计算,同时将结果转发。节点的传感器模块能耗较低,因此可以将其数据采集周期设为与簇内成员节点的工作周期一致,这样将确保数据聚合处理不用等待某一输入数据的到来。簇头节点  $n_k$  的聚合算法流程图如图 3。

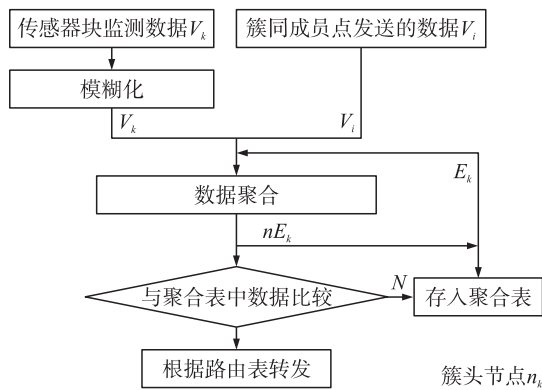


图 3 簇头节点  $n_k$  聚合流程图

2.3.1 数据聚合模块

数据聚合在簇头节点进行, 聚合的数据包括  $V_k(v_k, vs_k)$ 、 $V_i(v_i, vs_i)$  和  $E_k(e_k, es_k)$ , 得到新值  $nE_k(ne_k, nes_k)$ 。下面的计算中用  $I(i, is)$  表示  $V_k(v_k, vs_k)$  和  $V_i(v_i, vs_i)$ 。  $nE_k$  的取值同时还取决于 WSN 所监测数据的物理量, 分为下面情况讨论<sup>[14]</sup>。

(1)、 $I \cap E_k = \phi$ 。如果监测数据取最小值, 则  $nE_k = \min(I, E_k)$ ; 而如果监测数据取最大值, 则  $nE_k = \max(I, E_k)$ ;

(2)、 $I \cap E_k \neq \phi$ 。  $nE_k$  的计算如下, 引入加权和计算。

$$ne_k = e_k \cdot w_{E_k} + i \cdot w_I \tag{1}$$

$$nes_k = es_k \cdot w_{E_k} + is \cdot w_I \tag{2}$$

其中,  $w_{E_k} = \frac{is}{es_k + is}$ ,  $w_I = \frac{es_k}{es_k + is}$ 。

从上述计算可以看出, 如果  $V_k$  和  $V_i$  与估计值  $E_k$  不相关, 新估计值则与监测数据值的实际应用有关, 否则引入加权和来计算新估计值  $nE_k$ , 其权值与两个对称三角模糊数的模糊宽度(幅度)成反比。因为模糊数的不确定性越少, 聚合处理越相关, 则说明数据之间的冗余性也越大。

2.3.2 聚合表

簇头节点除了建立路由表, 在进行数据聚合处理时还需建立聚合表, 表的更新也随着簇的变化簇头的转换而动态建立。当簇头接收到簇内成员节点的数据后进行数据聚合处理, 表中记录下该次聚合的输出值  $nE_i(ne_i, nes_i)$  的记录项, 其中包括节点标识、簇内节点号、监测数据值、计数值和有效标志。其中节点标识为该簇内成员节点的节点号, 监测数据值是由簇头每进行一次聚合处理的输出值, 与成员节点标识一一对应; 计数值初始值都为 0, 簇头节点每向其下一跳节点发送一次数据时, 该值加 1, 而当簇头收到成员节点  $n_i$  的数据时,  $c_i$  清 0; 有效标志  $b_i$  初始值都为 1, 当  $c_i$  大于预定值  $c_{max}$  时  $b_i$  清 0,  $c_{max}$

取该簇的成员节点数。当节点死亡或是无法与簇头通信时,  $c_i$  的值将大于  $c_{max}$ , 所以  $b_i$  和  $c_i$  可以用来作为判断成员节点能否与簇头正常通信的标志位。如表 1 所示。

表 1 簇头节点聚合表

节点标识	簇内节点号	监测数据值	计数值	有效标志
$n_1$	1	$nE_1(ne_1, nes_1)$	$c_1$	$b_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n_i$	$i$	$nE_i(ne_i, nes_i)$	$c_i$	$b_i$

担任中转数据的簇头节点的聚合策略算法同上, 不同之处在于此类簇头节点还要接收其它簇头的的数据, 将其存入聚合表并将参与该簇头的转发决定。为了区分簇内成员节点和簇头发送的消息, 规定消息如表 2。

表 2 消息结构

节点标识 $n$	簇号	簇内标识	数据
----------	----	------	----

当中转簇头接收到其簇内成员节点的数据, 处理过程同普通簇头节点; 当接收到其它簇头节点的数据时, 根据消息中的簇号和簇内标识即可判定是否是簇头转发数据。聚合表与前不同, 将增加一列簇头数据标志位, 初始都为 0, 当接收到簇头数据变为 1, 簇内节点号将存入该簇头的簇号, 其它列数据不变。当中转簇头每完成一次数据聚合处理, 得出的新估计值不仅与聚合表中所有的簇内成员节点对应的值进行比较, 还将与其它簇头的的数据值比较, 同理, 在中转簇头完成一轮簇内通信时, 其它簇头的的数据也将参与周期转发数据的平均值计算。

3 仿真实验

本文采用 NS2 + NRL-SENSORSIM 作为仿真平台, 通过仿真环境来分析本文提出的数据聚合机制的性能。网络拓扑为 100 个节点随机部署在  $100\text{ m} \times 100\text{ m}$  的范围内, 汇聚节点位于  $(50, 150)$ , 采用 IEEE802.1b 作为 MAC 协议, 节点初始能量为 1 J,  $\epsilon$  分别取值 0.5 和 1。每 10 s 计算一次网络平均能耗(测量节点剩余能量)。在 NRL-SENSORSIM 中, 传感器网络事件由大量的 PHENOM 节点来模拟, 每秒 8 个事件在监测范围内随机产生, 且事件持续时间为 1 s, 假定用户对监测数据的最大值感兴趣。图 4 和图 5 分别为网络节点平均能量消耗图及网络寿命图。从图中可以看出本文提出的基于分簇数据聚合机制的网络平均能耗较 LEACH 协议有明显的改进, 当所监测区域数据没有明显变化时, 节点减少了



转发的数据量,网络平均能耗也有显著减少。随着 $\varepsilon$ 的增大,节点转发的数据量增加,能量消耗也随着增大。节点死亡率随时间增加有效降低,网络寿命得到明显延长。

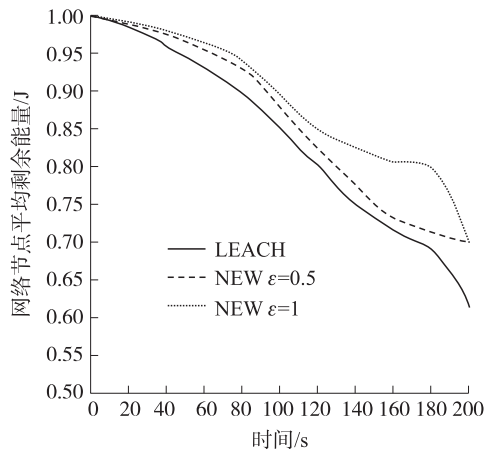


图4 网络节点平均能量消耗

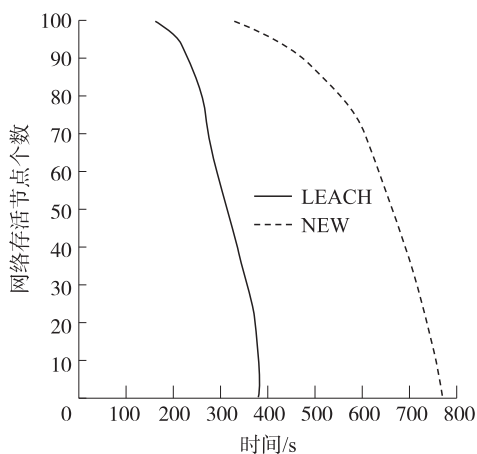


图5 网络寿命

## 4 结束语

分簇算法和数据聚合是 WSN 中节省能量的两种有效方法,本文在分簇网络拓扑的基础上,提出了一种新的数据聚合机制。在仿真实验中将新方法 with LEACH 协议进行对比分析,结果表明新方法在降低节点的数据传输量,节省能量消耗和延长网络寿命等方面都有明显改进。而数据聚合带来的延迟代价和鲁棒性问题,则是笔者下一步研究的重点。

### 参考文献:

- [1] 孙利民,李建中,陈渝,等. 无线传感器网络[M]. 北京:清华大学出版社,2005:1-10.
- [2] 于宏毅,李鸥,张效义,等. 无线传感器网络理论、技术与实现[M]. 北京:,国防工业出版社,2008:190-193.
- [3] Kalpakis K, Dasgupta K, Namjoshi P. An Efficient Clustering - based Heuristic for Data Gathering and Aggregation in Sensor Networks [C]//IEEE Wireless Communications and Networking, 2003: 1948-1953.
- [4] 王殊,阎毓杰,陈帅,等. 传感器网络中基于移动代理的数据融合方法研究[J]. 传感技术学报,2006,19:926-932.
- [5] So J, Kim J, Gupta I. Cushion: Autonomously Adaptive Data Fusion in Wireless Sensor Networks [C]//IEEE International Conference on Mobile Adhoc and Sensor Systems, 2005: 1-3.
- [6] Weilian S U, Theodoros C. Bougiouklis. Data Fusion Algorithms in Cluster - Based Wireless Sensor Networks Using Fuzzy Logic Theory [C]//Proceedings of the 11th Conference on 11th WSEAS International Conference on Communications, 2007: 291-299.
- [7] 冯秀芳,高战强. 无线传感器网络中基于 PCA 神经网络的数据融合[J]. 计算机科学,2007,34:372-375.
- [8] 唐晨,王汝传,黄海平,等. 基于信息熵的无线传感器网络数据融合方案[J]. 东南大学学报,2008,38:276-279.
- [9] Ignacio Solis, Katia Obraczka. The Impact of Timing in Data Aggregation for Sensor Networks [C]//IEEE International Conference on Communications, 2004: 3640-3645.
- [10] Yuan W, Krishnamachari S V, Tripathi S K. Synchronization of Multiple Levels of Data Fusion in Wireless Sensor Networks [C]//IEEE Global Telecommunications Conference, 2003: 221-225.
- [11] Yager R R. On the Entropy of Fuzzy Measures [J]. IEEE Transactions on Fuzzy Systems, 2000, 8(4): 453-461.
- [12] Silvio Croce, Francesco Marcelloni, Massimo Vecchio. Reducing Power Consumption in Wireless Sensor Networks Using a Novel Approach to Data Aggregation [J]. The Computer Journal, 2008, 51(2): 226-239.



张强(1985-),男,四川绵阳人,硕士研究生,主要研究方向为无线传感器网络, zq09284275@sina.com;



卢潇(1959-),女,上海人,副教授,主要研究方向为网络信息系统;



崔晓臣(1983-),男,山东栖霞人,硕士研究生,主要研究方向为无线传感器网络。