

无线传感网中基于能量平衡的数据收集树算法

李虹^{1,2}, 黄刘生^{1,2}, 徐宏力^{1,2}, 王继春^{1,2}

(1. 中国科学技术大学计算机科学与技术系, 安徽合肥 230027; 2. 高性能计算及应用省部共建重点实验室, 安徽合肥 230027)

摘要: 针对实际应用中传感器节点采样频率不同的特点, 提出了一种传感网中基于能量平衡的数据收集树算法(an energy balancing and tree-based data gathering algorithm, EBDGA). EBDGA 在基站集中式的构建是以本基站为根结点、基于能量平衡的生成树, 根据节点的采样频率定制生成树中子节点加入的策略. 仿真实验表明, 在这种特定的网络模式下, EBDGA 比传统的数据收集树算法 PEDAP 和 PEDAP-PA 具有更好的能量平衡特性, 提升了能量性能.

关键词: 无线传感网; 数据收集; 采样频率; 能量平衡; 能量有效

中图分类号: TN919.2 **文献标识码:** A

An energy balancing and tree-based data gathering algorithm in wireless sensor networks

LI Hong^{1,2}, HUANG Liu-sheng^{1,2}, XU Hong-li^{1,2}, WANG Ji-chun^{1,2}

(1. Department of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China;
2. Anhui Province—MOST Co-Key Laboratory of High Performance Computing and Its Application, Hefei 230027, China)

Abstract: In view of the nodes' sampling frequencies in real applications being different, an energy balancing and tree-based data gathering algorithm (EBDGA) in wireless sensor networks was proposed. EBDGA is executed centrally on the base station to create an energy balancing tree with this station as its root. The selection policy of each children node takes the node's sampling frequency into account. Simulation results show that under this special network mode, EBDGA achieves more notable energy balancing capability than traditional algorithms, PEDAP and PEDAP-PA, thus enhancing energy efficiency.

Key words: wireless sensor networks; data gathering; sampling frequency; energy balancing; energy efficiency

0 引言

无线传感网由大量微体积、低功耗、低成本的传感器节点组成, 这种传感器节点包含传感部件、处理部件、无线收发器及便携式电源^[1], 它被广泛应用于环境检测、军事监控、医疗警报及各类民生应用^[1~4]. 无线传感网是一类特殊的无线自组网, 它不

同于传统的无线网络. 无线传感网中的传感器节点分布稠密, 节点数目较大; 传感器节点的能量、存储空间及其计算能力有限, 不能有效或根本无法运行复杂或计算量较大的大型应用程序; 由于无线信道的不可靠性及节点能量受限等因素, 数据通信的失败率较高, 网络拓扑变化较快; 并且由于大量传感器节点的存在, 赋予每个节点全局性的标识几乎不可

收稿日期: 2006-09-06; 修回日期: 2007-02-01

基金项目: 国家发展和改革委员会项目(CNGI-04-10-2A), 国家重点基础研究发展(973)计划(2006CB303006)资助.

作者简介: 李虹, 女, 1982年生, 博士生. 研究方向: 无线传感器网络. E-mail: lihong@mail.ustc.edu.cn

通讯作者: 黄刘生, 教授. E-mail: lshuang@ustc.edu.cn

能. 在这些特性中, 能量有限成为提升传感网性能的重大障碍之一. 当网络中一些节点能量耗尽时, 网络可能陷入瘫痪, 无法提供给用户特定的服务数据. 考虑到在偏远地区和恶劣环境下重新布置或投入新的传感器节点较为困难、成本较高, 多数算法的设计都是基于能量有效的.

目前已有许多讨论数据收集的算法. 根据算法结构的不同, 可分为三类: 一是基于簇结构的数据收集算法, 如 LEACH^[5] 和 HIT^[6]. 整个网络都被分割成若干簇, 每个簇内的所有节点只向本簇的簇首节点发送数据包, 簇首节点对这些数据包进行聚集, 然后传送至基站. 为了避免冲突, 每个簇头都需支持相应的 TDMA MAC 协议; 二是基于链式结构的数据收集算法, 如 PEGASIS^[7] 和 DWZ^[8]. 所有节点形成一条链接, 并保证一次只有一个节点和周围的某个节点进行数据通信, 在数据包传输过程中进行数据的聚集以减少能耗. 文献[7]随机选择一个节点作为终结的数据聚集点, 由该节点将所有数据发往基站; 三是基于树结构的数据收集算法, 如 PEDAP^[9] 和 CTCCAA^[10]. 文献[9]构建以基站为根节点的最小生成树, 并最小化每轮数据传输的总能耗; 文献[10]为所有的节点分配不同的码以避免冲突. 由文献[9]可知, PEDAP 的能量有效性优于 LEACH 和 PEGASIS, 而 PEDAP-PA 在 PEDAP 的基础上进一步提升了能量有效性, 并在节能的基础上考虑了能量平衡特性, 保证节点间的能量差不会随着网络运行的时间延续而变得巨大, 也不会导致某些节点过早地失效, 从而提高了每个节点在整个传感网中的利用率, 且能延长大部分节点的生命期. 文献[5~10]提出的算法均为集中式算法.

根据应用环境的不同, 节点生成数据包的形式可大致分为三类: (I) 周期型. 节点每隔时间 t 对周围环境进行感知, 把得到的数据传向基站; (II) 主动型. 节点不间断地检测周围环境, 一旦发生特殊事件, 就往基站发送数据包; (III) 被动型. 只有当基站向传感节点发出询问请求时, 节点才生成特定的数据包. 在第 (I) 类节点周期性生成数据包的情况下, 以往的数据收集算法均假设网络中每个节点的采样频率 f ($f=1/t$) 一致, 而在实际应用中, 为了既防止在同一区域内产生过多冗余的数据又能保证在基站获得较高的数据精度, 小部分节点的采样频率会设置的低一些; 或者在某些非重要区域, 应用事件较少发生, 此区域内节点的采样频率也会设置的低一些.

EBDGA 针对存在多种不同采样频率的简单网络模型, 以树结构为基本的路由结构, 根据节点的采样频率进行子节点选取, 并在最小化生成树总能量的基础上考虑能量平衡性. EBDGA 较 PEDAP 和 PEDAP-PA, 在一定幅度上节省了节点的能耗, 并获得了明显的能量平衡特性.

1 模型描述

1.1 网络模型

本文研究的无线传感网满足以下几点假设:

(I) 无线传感网由一个基站 B (标号为 0) 和 n 个传感器节点 (标号为 $1, \dots, n$) 组成. 基站和节点的位置固定, 基站知道所有节点的地理位置, 这可以通过在基站中输入节点坐标得到, 也可以通过 GPS 全球定位系统或额外执行定位算法获得.

(II) 传感器节点周期性地监测周围环境信息并采集相关数据. 在每个时间单元内, 节点收集到的数据都将发往基站, 我们称这个时间单元为轮 (round), 它的时间间隔为 t . 在数据传输过程中, 采用数据聚集技术^[11,12]减少能耗.

(III) 传感器节点 i 的采样频率 $f_{T_0}(i) \in \{1/k\}$, ($k \in \mathbb{N}$). $f_{T_0}(i) = 1/k$ 表示节点 i 每隔 k 轮向基站发送一个数据包. 基站知道所有节点的采样频率.

(IV) 基站能量充足; 传感节点能量有限. 为保证能量的一致性, 设定节点的初始能量与采样频率成正比, 即节点 i 的初始能量 $E(i) = E_0 \times f_{T_0}(i)$, E_0 为常量. 每个节点既能与网络中任意一个节点进行通信, 也能和基站进行直接的数据通信^[7,9].

(V) 节点 i 到节点 j 的距离用 $d(i, j)$ 表示; 节点 i 到基站的距离用 $d(i, 0)$ 表示.

1.2 能量模型

本文采用的能量模型是文献[5]中给出的 first order radio 模型. 假定无线电收发电路装置需要 $E_{\text{elec}} = 50 \text{ nJ/bit}$ 来启动传输或接收, 传输放大器需要 $\epsilon_{\text{amp}} = 100 \text{ pJ}/(\text{bit} \times \text{m}^2)$ 来获得合适的信噪比. 在长度为 d 的信道中传输 h bit 的消息所需能量为

$$E_{T_x}(h, d) = E_{\text{elec}} \times h + \epsilon_{\text{amp}} \times h \times d^2 \quad (1)$$

接收 h bit 的信息所需能量为

$$E_{R_x}(k) = E_{\text{elec}} \times h \quad (2)$$

2 EBDGA 算法描述

2.1 算法概述

由 1.1 节网络模型可知, 基站知道所有节点的

参数及位置信息,且能量充足,因此可由基站集中计算一棵以基站为根结点、连通整个无线传感网的生成树. EBDGA 算法的执行可分为三个步骤:(I)算法初始化时,生成树只含有一个根节点,即基站;(II)每次向生成树中加入新的节点时,本文选取优先级最高的边加入;(III)重复步骤(II),不断向生成树中加入新的节点,直至网络中所有存活的节点都已在生成树中.完成后由基站广播树结构信息,每个接收到广播信息包的传感节点就能获悉其父节点的信息.数据包沿着这棵生成树进行传输,每个传感节点只接收来自其子节点的数据包,并将其转发给自身的父节点,直至数据包到达基站.在数据包的转发过程中,EBDGA 算法采用数据聚集技术:不管节点在每一轮内接收到多少个来自各子节点的数据包,都将它们压缩、融合为单个数据包,再进行转发^[9].在 EBDGA 算法中,假设已经存在有效的 MAC 机制,减少了数据通信之间的干扰和冲突.

2.2 节点加入策略

定义 2.1 节点集合 $N(T) = \{i, 0 \leq i \leq n\}$ 和边集合 $E(T) = \{(i, j), i \neq j, 0 \leq i, j \leq n\}$ 共同表示生成树 T , 树 T 的生命期 $L(T) = \min\{l(i), i \in N(T)\}$, 树 T 中所有节点的平均生命期 $A(T) = (\sum_{i \in N(T)} l(i)) / |N(T)|$. 其中, $l(i)$ 表示节点 i 的生命期.

定义 2.2 对于任意节点 $i \in N(T)$, i 到基站的路径 $p_T(i)$ 也是一种生成树.

如图 1 所示,对于任意节点 $k \notin N(T_m)$, $j \in N(T_m)$, 当 k 加入树 T_m 中并以 j 为其父节点后,形成新树 T_{m+1} . 节点 k 的频率只影响路径 $p_{T_{m+1}}(k)$ 上节点的能量,其他节点能量不受影响.因此可用 $L(p_{T_{m+1}}(k))$ 近似表示 $L(T_{m+1})$, 选择最大的 $L(T)$ 值可保证最大的网络生命期.对于 1.1 节所示的网络模型,网络生命期并不能以第一个失效节点的生

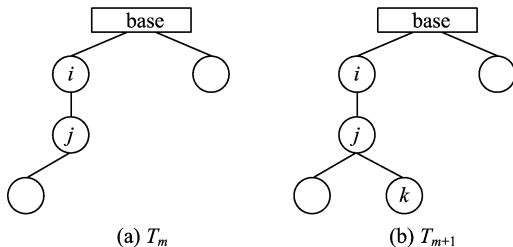


图 1 节点 k 加入树 T_m 中, 并以 j 为其父节点构成新树 T_{m+1}

命期来表示,大部分节点的生命期还需提高,因此要引入能量平衡性.这里,我们用 $A(p_{T_{m+1}}(k))$ 近似表示 $A(T_{m+1})$ 来作为能量平衡手段.要提高生成树 T_{m+1} 的能量有效和平衡性,就应该同时考虑 $L(T_{m+1})$ 和 $A(T_{m+1})$,也就是考虑 $L(p_{T_{m+1}}(k))$ 和 $A(p_{T_{m+1}}(k))$.

定义 2.3 对于任意节点 $k \notin N(T_m)$, $j \in N(T_m)$, k 加入树 T_m 并以 j 为父节点形成新树 T_{m+1} , 则 $C_{T_{m+1}}(k, j)$ 表示边 (k, j) 的优先级, $C_{T_{m+1}}(k, j) = L(p_{T_{m+1}}(k)) \times A(p_{T_{m+1}}(k))$.

若要使网络生命期增大, $L(p_{T_{m+1}}(k))$ 应该增大, $C_{T_{m+1}}(k, j)$ 值也对应增大;若要使网络更平衡, $A(p_{T_{m+1}}(k))$ 应该增大, $C_{T_{m+1}}(k, j)$ 值也对应增大.所以我们每次往生成树中新加入一个节点时,应该选取获得 $\max\{C_{T_{m+1}}(k, j), k \notin N(T_m), j \in N(T_m)\}$ 的边 (k, j) .

2.2.1 节点生命期的计算

定义 2.4 对于任意节点 $i \in N(T)$, i 在树 T 中每轮接收耗能为 $R_T(i)$, 每轮发送耗能为 $S_T(i)$, 则节点 i 的生命期为 $l_T(i) = E(i) / (R_T(i) + S_T(i))$. 其中, $E(i)$ 为节点 i 的能量.对于图 1(b) 中新加入的叶子节点 k , 有

$$R_{T_{m+1}}(k) = 0 \quad (3)$$

$$S_{T_{m+1}}(k) = E_{T_x}(h, d(k, j)) \times f_{T_{m+1}}(k) \quad (4)$$

对于任意节点 i 及其子节点 a , 父节点 b ; $i, b \in N(p_{T_{m+1}}(j))$ 且 $a \in N(p_{T_{m+1}}(k))$, 有

$$R_{T_{m+1}}(i) = R_{T_m}(i) + E_{R_x}(h) \times (f_{T_{m+1}}(a) - f_{T_m}(a)) \quad (5)$$

$$S_{T_{m+1}}(i) = E_{T_x}(h, d(i, b)) \times f_{T_{m+1}}(i) \quad (6)$$

当加入 k 后, $p_{T_{m+1}}(j)$ 上的所有节点都随 k 的频率变化相应地发生能耗的变化.但是不管怎样,每个节点 i 的接收能量都是原有的接收能量加上子节点 a 多转发数据包所需接收的能量,我们用 $f_{T_{m+1}}(a) - f_{T_m}(a)$ 表示子节点 a 实际增加的发送频率,其中 $f_{T_{m+1}}(a)$ 、 $f_{T_m}(a)$ 分别为节点 a 在树 T_m 和 T_{m+1} 中的实际发送频率.

2.2.2 节点的实际发送频率

为了计算每个节点在生成树中的实际发送频率,引入频率集合 $F(i)$.算法初始化时,对于任意节点 $i (1 \leq i \leq n)$, $F(i) = \{f_{T_0}(i)\}$.对于图 1(b) 中新加入的叶子节点 k ,它在树 T_{m+1} 中的实际发送频率等于初始频率,即 $f_{T_{m+1}}(k) = f_{T_0}(k)$.对于任意节点 i , $i \in N(p_{T_{m+1}}(j))$, 频率集合 $F(i)$ 的更新规则如下:

(I) 如果存在 $x \in F(i)$, $y \in Z$, 有 $x/f_{T_0}(k) =$

Fig. 1 Node k added into tree T_m with j as its parent, and forms a new tree T_{m+1}

y , 则 $F(i)$ 不变;

(II) 如果存在 $x \in F(i), y \in Z$, 有 $f_{T_0}(k)/x = y$, 则 $F(i) = F(i) \cup \{f_{T_0}(k)\} - \{x\}$;

(III) 其他情况, $F(i) = F(i) \cup \{f_{T_0}(k)\}$.

变量 $F(i) = \{x[1], x[2], \dots, x[|F(i)|]\}$, $x[d] (1 \leq d \leq |F(i)|)$ 按照从大到小排列

float temp=0, $f_{T_{m+1}}(i) = 0$;

算法 2.1 计算节点 i 在树 T_{m+1} 中实际发送频率的代码.

根据频率集合 $F(i)$ 计算 $f_{T_{m+1}}(i)$ 的算法描述如下:

```

for (int k=1; k<=|F(i)|; k++)
{
    if (temp==0)  $f_{T_{m+1}}(i) = f_{T_{m+1}}(i) + x[k]$ ;
    else  $f_{T_{m+1}}(i) = f_{T_{m+1}}(i) + x[k] - x[k] \times \text{temp}$ ;
    temp =  $f_{T_{m+1}}(i)$ ;
}

```

对于上述规则和算法描述, 若父节点发送频率为 0.25, 子节点发送频率为 0.5, 则子节点发送数据

包比父节点更为频繁, 所以父节点频率更新为 0.5, 如规则 (II); 若父节点频率为 0.5, 子节点频率为 0.25, 则父节点发送数据包比子节点更为频繁, 频率无需增加, 保持 0.5 不变, 如规则 (I); 若父节点频率为 0.25, 子节点频率是 1/3, 则父节点的频率集合应是 $\{0.25, 1/3\}$, 实际发送频率应为 $0.25 + 1/3 - 0.25 \times (1/3)$, 因为要减去它们公共发送的部分, 如规则 (III).

3 实验模拟及性能分析

本文对 EBDGA 算法进行了实验模拟及分析, 并与协议 PEDAP 和 PEDAP-PA 进行了性能比较.

n 个传感器节点随机分布在 $100 \text{ m} \times 100 \text{ m}$ 的正方形区域内. E_0 为 0.2 J, 数据包大小为 1 000 bit. 网络内有 5 种频率, 不失一般性, 它们分别取 1, 1/2, 1/3, 1/4 和 1/5. 图 2~5 分别显示了 n 为 150, 200, 基站坐标为 (0,0) 和 (50,50) 时, 前 80% 失效节点的生命期. 从图中可以看出, 不管基站位置处于中心还是远离节点群, EBDGA 中前 80% 失效节点的

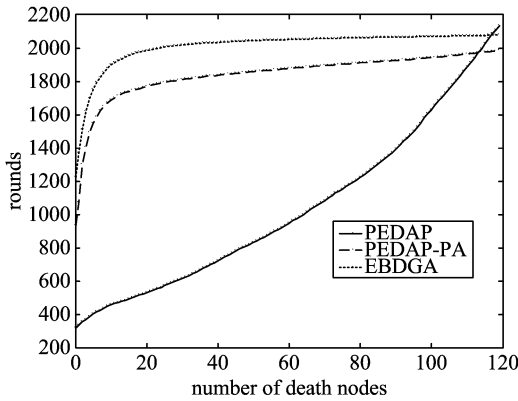


图 2 150 节点, 基站位置 (50,50)

Fig. 2 150 nodes, the base is (50,50)

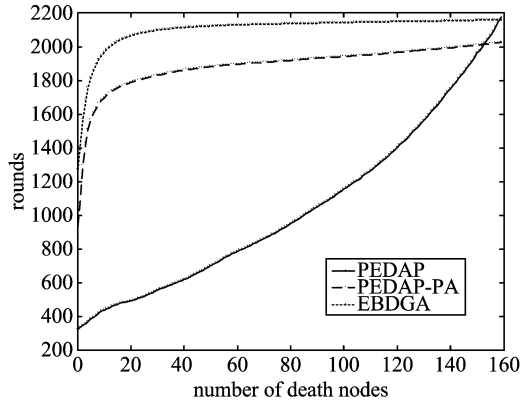


图 4 200 节点, 基站位置 (50,50)

Fig. 4 200 nodes, the base is (50,50)

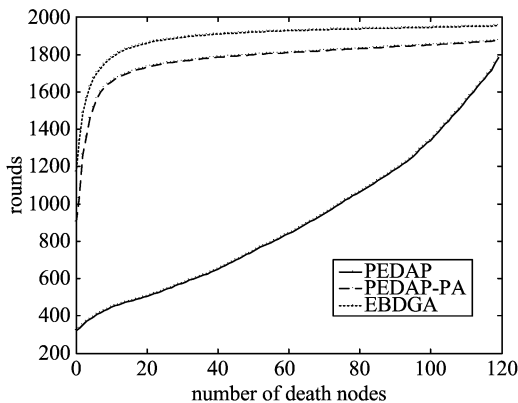


图 3 150 节点, 基站位置 (0,0)

Fig. 3 150 nodes, the base is (0,0)

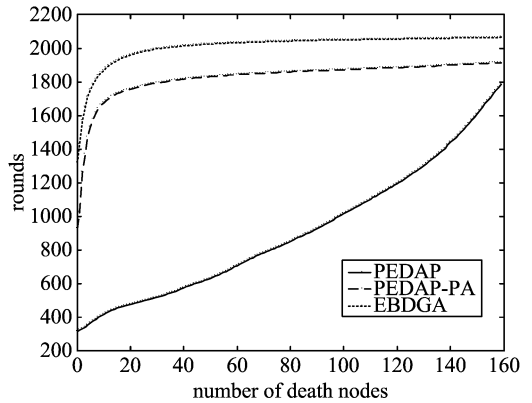


图 5 200 节点, 基站位置 (0,0)

Fig. 5 200 nodes, the base is (0,0)

生命期均大于 PEDAP-PA, 前 75% 失效的节点生命期远大于 PEDAP. 当节点数增加时, EBDGA 的性能略有提高.

为了量化能量平衡性能, 本文用 $\text{round}[i]$ 的标准方差 \sqrt{VR} 作为衡量标准, $\text{round}[i]$ 是节点 i 最终的生命期, 故有

$$\left. \begin{aligned} AVR &= \left(\sum_{i=1}^n \text{round}[i] \right) / n \\ \sqrt{VR} &= \sqrt{\sum_{i=1}^n (\text{round}[i] - AVR)^2 / n} \end{aligned} \right\} \quad (7)$$

\sqrt{VR} 越小, 说明算法的能量平衡性越好. 表 1 列出了对应于图 2~5 各种条件下三种算法的 \sqrt{VR} 值. 由表 1 可以看出, EBDGA 的 \sqrt{VR} 值远小于 PEDAP 和 PEDAP-PA 的 \sqrt{VR} 值. PEDAP 和 PEDAP-PA 的 \sqrt{VR} 值受基站位置影响较大, 当基站位置离传感器节点群较远时, \sqrt{VR} 值显著减小. 因此, EBDGA 的能量平衡性能最好, 且不易受基站位置的影响.

表 1 各种网络参数下的 \sqrt{VR} 值

Tab. 1 \sqrt{VR} values in different network parameters

节点数	基站位置	PEDAP	PEDAP-PA	EBDGA
150	(50,50)	982	573	119
150	(0,0)	714	275	107
200	(50,50)	1 004	571	110
200	(0,0)	753	300	98

4 结论

本文提出了在传感网具有多种不同采样频率下, 基于能量平衡的数据收集树算法 EBDGA. 在这种特定的网络模式下, EBDGA 比 PEDAP 和 PEDAP-PA 具有更好的能量平衡特性, 并延长了前 75% 失效节点的生命期. 在今后的研究中, 我们还将把节点发送数据包的数量作为参数来考虑分析 EBDGA 的性能, 并设计分布式的算法.

参考文献(References)

[1] Akyildiz I F, Su W, Sankarasubramaniam Y, et al. A survey on sensor networks[J]. IEEE Communications Magazine, 2002, 40(8):102-114.
 [2] Al-Karaki J N, Karnal A E. Routing techniques in wireless sensor networks; a survey[J]. IEEE Wireless Communications, 2004, 11(6):6-28.

[3] Estrin D, Govindan R, Heidemann J, et al. Next century challenges: scalable coordination in sensor networks [C]// ACM/IEEE International Conference on Mobile Computing and Networking, Seattle: ACM Press, 1999: 263-270.
 [4] LIN Zhi-ting, QU Yu-gui, ZHAI Yu-jia, et al. Algorithm for the placement of nodes for effective coverage[J]. Journal of University of Science and Technology of China, 2005, 35(3):411-416.
 蔺智挺, 屈玉贵, 翟羽佳, 等. 一种高效覆盖的节点放置算法[J]. 中国科学技术大学学报, 2005, 35(3): 411-416.
 [5] Heinzelman W R, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks [C] // Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. Maui Hawaii: IEEE Press, 2000, 2: 3 005-3 014.
 [6] Culpepper J, Dung L, Moh M. Hybrid indirect transmission (HIT) for data gathering in wireless micro sensor networks with biomedical applications [C]//Proceedings of 18th Annual Workshop on Computer Communications, 2003:124-133.
 [7] Lindsey S, Raghavendra C S. PEGASIS: power-Efficient gathering in sensor information systems[C]// Proceedings of the IEEE Aerospace Conference. Montana: IEEE Press, 2002, 3(9):1 125-1 130.
 [8] Du K, Wu J, Zhou D. Chain-based protocols for data broadcasting and gathering in the sensor networks [C]// Proceedings of the international Parallel and Distributed Processing Symposium. Nice, France, 2003: 260-267.
 [9] Tan H O, Korpeoglu I. Power efficient data gathering and aggregation in wireless sensor networks [J]. SIGMOD Record, 2003, 32(4):66-71.
 [10] Annamalai V, Gupta S K S, Schwiebert L. On tree-based convergecasting in wireless sensor networks [C]// Proceedings of the IEEE Wireless Communication and Networking Conference. 2003, 3: 1 942-1 947.
 [11] Dasgupta K, Kalpakis K, Namjoshi P. An efficient clustering-based heuristic for data gathering and aggregation in sensor networks [C]// Proceedings of the IEEE Wireless Communication and Networking. New Orleans: IEEE Press, 2003, 3: 1 948-1 953.
 [12] Petrovic D, Shah R C, Ramchandran K, et al. Data funnelling: Routing with aggregation and compression for wireless sensor networks [C]// Proceedings of the 1st IEEE Sensor Network Protocols and Applications. Anchorage: IEEE Press, 2003:156-162.