

无线传感器网络能量高效数据收集的分簇算法^{*}

郝聚涛¹, 赵晶晶², 霍欢¹

(1. 上海理工大学 光电信息与计算机工程学院, 上海 200093; 2. 重庆大学 输配电装备及系统安全与新技术国家重点实验室, 重庆 400044)

摘要: 传感器网络所具有的集中式数据收集、多跳数据传输、多对一流量模式这三种特征会造成漏斗效应的出现, 这会导致严重的包碰撞、网络拥塞、包丢失, 甚至拥塞崩溃, 还会导致能量消耗的热点出现, 使某些节点甚至整个网络过早死亡。利用感知数据的空间相关性, 将相邻的节点分组, 每组选出一个簇首作为整个组的代表, 将其传递给基站来缓解漏斗效应问题。在 CAG 算法的基础上利用感知数据的空间特性提出了一个改进的节点分簇算法, 该算法可以有效减少簇首数量, 从而降低传输能量消耗。实验结果也验证了算法的有效性。

关键词: 无线传感器网络; 数据收集; 空间相关数据收集; 分簇算法

中图分类号: TP393 **文献标志码:** A **文章编号:** 1001-3695(2010)01-0295-03

doi:10.3969/j.issn.1001-3695.2010.01.087

Energy-efficient clustering algorithm for data gathering in WSN

HAO Ju-tao¹, ZHAO Jing-jing², HUO Huan¹

(1. School of Optical-Electrical & Computer Engineering, University of Shanghai for Science & Technology, Shanghai 200093, China; 2. State Key Laboratory of Power Transmission Equipment & System Security & New Technology, Chongqing University, Chongqing 400044, China)

Abstract: Wireless sensor networks are characterized by centralized data gathering, multi-hop communication and many-to-one traffic pattern. These three characteristics may give rise to funneling effects that can lead to severe packet collision, network congestion, packet loss and even congestion collapse. This can also result in hotspots of energy consumption that may cause premature death of sensor nodes and even premature death of entire network. Exploiting spatial correlation of nodes to form clusters of nodes sensing similar values, and only cluster head sensor reading was transmit to sink, such could efficiently alleviate the funneling effects. Based on the CAG algorithm, proposed a novelty clustering algorithm which could greatly reduce the number of cluster heads. Experimental results validate the effectiveness of this approach.

Key words: wireless sensor network (WSN); data collection; spatial correlated data collection; clustering algorithm

0 引言

随着微电子机械系统技术的发展, 低成本和低功耗无线微处理器节点的应用成为现实。通过部署传感器节点并组成网络, 可以获得远程监测区域内的被监测对象的行为、位置等信息^[1]。节点感知数据并将数据以无线通信的方式发送给基站, 基站综合收集的数据并将它们呈现给用户。

传感器节点由所携带的电池进行供电, 节点感知周围环境、接收和发送数据都会消耗能量, 因此设计一个能量高效的数据收集算法来延长整个网络的生存周期是必不可少的。目前已设计出很多算法、应用和网络协议来减少传感器节点的能量消耗, 如能量感知路由^[2]、能量高效的信息处理^[3]和能量优化的拓扑控制^[4]。

传感器通常以冗余的方式随机部署在监测区域。这种冗余部署一方面可以在较细的空间粒度上进行取样, 同时提高了容错和噪声免疫能力, 另一方面也带来了一系列的问题, 最突出的便是由于所有节点由不可更换电池供电, 电能成为非常珍

贵的资源, 传输每个节点的数据将会使电能很快耗尽。此外, 由于带宽有限, 向基站传输所有节点的读取数据, 会导致包丢失现象严重, 最终影响收集的数据质量。对于此类问题的一种切实可行的方法是, 利用相邻节点之间的空间相关性, 减少数据的传输量。具体做法各有不同, 但共同思想是选取部分节点作为代表, 只有这些代表所感知的数据才会被传送到基站, 而非所有节点的感知数据, 以此来减轻网络传输负载。

Gupta 等人^[5]通过选择传感器节点的一个称为连通关联支配子集 (connected correlation-dominating set) 来代表整个网络, 从而达到最小化通信能量消耗。在数据收集过程中仅子集中的节点参与通信, 为此这个子集必须是连通的, 并且基站可以通过这个子集的数据重构整个网络的数据。Chatterjea 利用节点感知数据的时间和空间关联性, 通过允许特定节点进行数据聚合来消除冗余数据的传输问题^[6]。Chitnis 给出了一个在出现错误的情形下计算聚合的协议, 该协议结合了树聚合和流言 (gossip) 聚合体系结构的优点来最小化错误^[7]。在 LEACH 方法中^[8], 一部分预先设定的节点将成为簇首, 簇首节点广播它的候选资格, 其余节点收到广播后, 会选择最近的簇加入。

收稿日期: 2009-03-29; **修回日期:** 2009-04-30 **基金项目:** 上海理工博士科研启动基金资助项目 (10D710); 上海优秀青年专项基金资助项目 (slg08014); 上海市教育委员会科研创新基金资助项目 (10YZ102)

作者简介: 郝聚涛 (1976-), 男, 山东临沂人, 博士, 主要研究方向为移动无线网络、生物信息处理等 (jutaohao@sjtu.edu.cn); 赵晶晶 (1980-), 女, 重庆人, 博士研究生, 主要研究方向为无线传感器网络、电力系统; 霍欢 (1979-), 女, 辽宁沈阳人, 博士, 主要研究方向为移动无线网络、分布式数据库。

该类方法的一个重要缺陷是簇首分布不均匀且簇首数量需预先确定。Sunhee 等人^[9]提出了一个分簇聚合算法 (clustered aggregation, CAG) 来减少传输的分组数, 对于用户的聚合查询返回一个近似的查询结果, 且该结果在满足用户所预先设定的误差允许范围之内。

本文在 CAG 算法的基础上, 提出了一个改进的分簇算法, 该算法较之 CAG 算法可以大大降低簇首数量。

1 CAG: 有损分簇聚合技术

CAG 由 TAG (tiny aggregation)^[10] 方法衍生而来, TAG 是一种基于树的路由机制, 通常与用户进行交互的一个节点被指定为根节点 (root)。根节点广播一个消息, 要求所有节点组织成一棵路由树; 该广播消息包含自己的 ID 和层次 (level), 任何未加入树的节点收到消息后, 将发送消息的节点作为父节点, 并将自己对应的层次加 1 后加入路由树; 然后该节点在消息中插入自己的 ID 和层次后再次进行广播。广播消息以这种方式在网络中泛洪, 直到所有节点都分配了一个父节点和层次。CAG 通过只允许代表节点参与聚合, 在节能方面 CAG 较之 TAG 更进一步。CAG 技术充分利用了监测环境数据的空间相关性来忽略冗余数据, 并迅速产生一个响应结果。

CAG 算法操作由查询和响应两个阶段组成。在查询阶段, CAG 采用 TAG 类似的算法来构造树, 与此同时利用用户指定的错误阈值完成分簇过程。在响应阶段, CAG 每个簇只发送一个响应值。因此 CAG 是一个只有簇首节点参与的有损聚合算法。

为了便于理解分簇过程, 本文用 τ 表示用户提供的错误阈值, CR 表示簇首传感器的读取值, MR 表示自己的传感器读取值。

CAG 算法的簇划分规则为: 如果 $MR < CR \pm CR \times \tau$, 那么该传感器与其父节点属于同一簇; 否则当前节点便成为簇首。利用 CAG 算法分簇实例如图 1 所示。

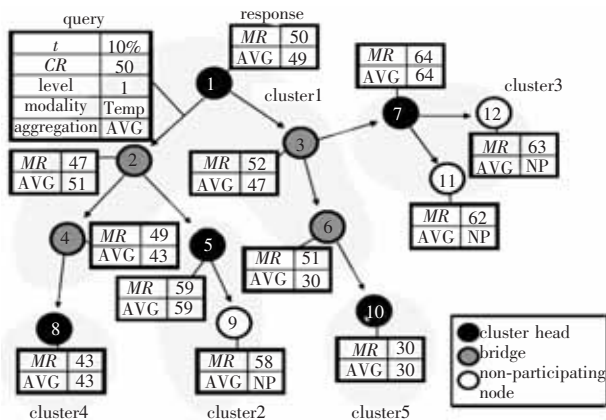


图 1 CAG 分簇实例

2 改进的分簇算法

CAG 算法实现了路由构造和分簇的同步完成。在分簇时, 采用了用户给定的错误阈值, 充分利用了数据内容的相关性和数据空间相关性。这个方案似乎非常完善, 但仔细研究仍存在问题, 在本文中只针对分簇方案存在的问题进行讨论。

文中分簇算法的最终目的是用簇首作为整个簇的代表, 对查询做出快速响应, 因此一方面要求簇首可以代表整个簇, 另

一方面簇首的数量尽可能少。为说明 CAG 算法存在问题, 以图 2 为例, 图中整个监测区域被分成三类, 每类采用不同颜色显示。假设处于同一类区域中的传感器节点读取的数据值相同, 按照 CAG 算法的思想, 同一区域内每个相互连通的子图选出一个节点作为簇首即可。然而结果如图 2 所示, 每一个父节点在另一区域的子节点都把自己标记为簇首, 因此这类分簇算法的缺点显而易见, 对于相同区域簇首个数太多。为此, 本文提出改进的分簇算法。

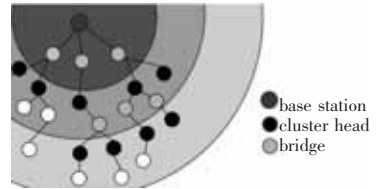


图 2 CAG 算法存在的问题

首先, 每个节点内部保存有两个路由项: 一个为目的节点是根节点的路由项, 节点利用该路由为其他簇向根节点转发数据; 另一个为目的节点是所在簇簇首的路由项, 节点利用该路由项向簇首发送数据, 便于簇首进行数据融合。

开始时节点发送 HELLO 消息, 获得周围邻居列表。当 sink 节点准备收集数据时, 把自己标记为根节点, 填写并发送网络初始化消息 NET_INIT, 它是一个七元组 $\langle \text{QueryID}, \text{Attribute}, \tau, \text{ParentID}, \text{MyID}, \text{level}, \text{CR} \rangle$ 。其中: QueryID 标记查询的编号; Attribute 是对携带多种传感器的节点而言, 指定要查询的属性; τ 为用户对当前属性所指定的阈值; level 为节点所在前向树中的深度。

当中间节点收到 NET_INIT 消息后, 根据 QueryID 查看是否收到这个广播消息, 如果第一次收到这个消息, 则在自己的路由表中添加该路由项, 否则丢弃该消息。

对于第一次收到广播的情况, 依据分簇规则 $MR < CR \pm \tau$ 判断是否与父节点属于同一簇:

情况 1: 如果与父节点属于同一簇, 将自己标记为成员节点并加入簇。

情况 2: 如果与父节点不属于同一簇, 则查找邻居列表是否有邻居已加入簇:

情况 2.1: 若无邻居节点已加入簇则把自己标记为簇首。

情况 2.2: 若有邻居节点已加入簇, 则依据分簇规则进行判断是否加入邻居所在的簇或自己成为簇首。

这个过程一直持续, 直至所有节点都加入路由树。

通过上述成簇算法可以发现, 在改进的算法中, 节点收到 NET_INIT 消息后不会主动成为簇首, 而是先要加入邻居所在的簇, 在条件都不满足的情况下自己才成为簇首。因此这个成簇方法可以克服位于同一层次的兄弟节点全部为簇首的情况。

3 实验结果

本章通过仿真实验验证本文所提分簇算法。实验所用的仿真程序是用 Java 开发的, 所用的传感器网络包含 500 个节点, 均匀分布在 $1000 \times 1000 \text{ m}^2$ 的正方形平面上, 所有节点的通信距离设定为 100 m。

为了验证算法在不同情况下的性能, 本文对仿真参数阈值 τ 、基站位置和传感器感知数据分别进行了实验。

实验 1 参数设置为: 阈值 $\tau = 5$; 基站位置位于传感器部署区域中心如图 3(a) 中黑色标记点所示。传感器仿真数据生

成方案为:网络分布区域中心点的读取值为 50,其他节点的读取值为 $50 \times (1 - \text{dist}/500)$, dist 表示当前节点与中心点的距离。

图 3(b)给出了 TAG 算法所构造的路由树,(c)和(d)分别表示采用 CAG 和本文所提算法产生的分簇结果。为便于观察,簇首用深黑色标记,簇首与父节点之间的连接也用粗线条表示。在本文中为形象显示每个簇,每个成员节点都通过一条边连接到簇首节点,但这条连接并不代表实际的路由,当某个成员节点作为其他簇首节点的桥接节点时,该节点会直接将数据转发给父节点而并非其所在簇的簇首节点。其中采用 CAG 算法产生的分簇数为 227,本文所提算法产生的分簇数为 71。

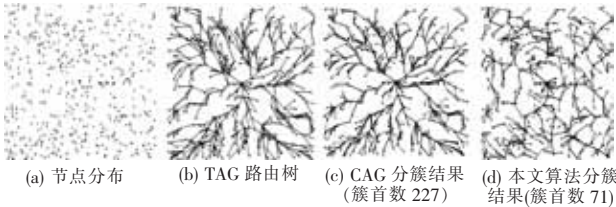


图 3 实验 1

实验 2 本实验与实验 1 的区别在于传感器读取值的设置上,在该实验中,在部署区域内随机选择 10 个数据中心点,每个数据中心之间的数值差为 10 个单位。节点的读取值为 $V \times (1 - \text{dist}/500)$, V 为距离最近的数据中心点的读取值, dist 表示两点之间的距离。

图 4(a)和(b)给出了采用 CAG 和本文算法的分簇结果,簇首数分别为 191 和 66。

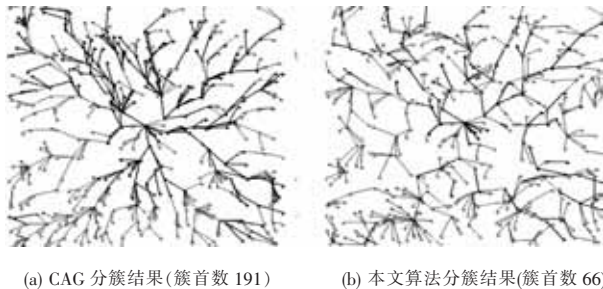


图 4 实验 2

实验 3 参数设置与实验 1 基本相同,区别为基站位于区域的左上角。分簇结果如图 5 所示,CAG 分簇数和本文算法分簇数分别为 176 和 50。

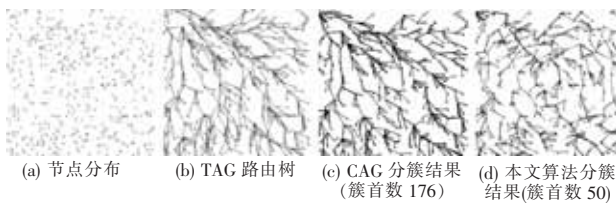


图 5 实验 3

实验 4 基站位置位于左上角,节点读取值产生方案与实验 2 相同。分簇结果如图 6 所示,CAG 分簇数和本文算法分簇数分别为 166 和 50。

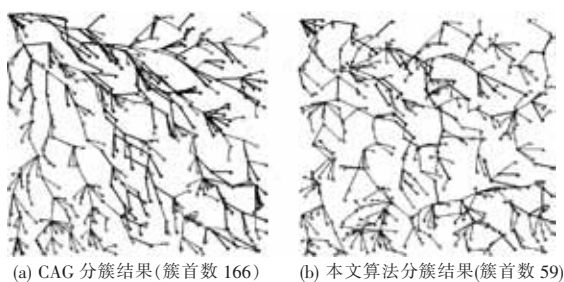


图 6 实验 4

实验 5 在该实验中主要用来测试阈值 τ 对分簇结果的影响。其中 τ 的取值分别为 2 和 10,其他设置与实验 4 相同。实验结果如图 7 所示。

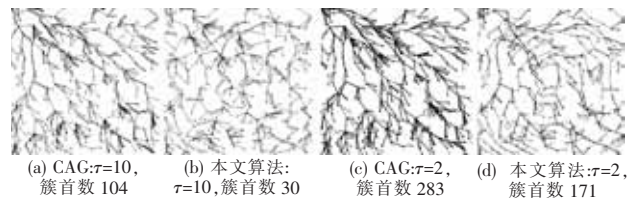


图 7 实验 5

4 结束语

随着微电子机械系统技术的发展,无线传感器网络得到越来越广泛的应用和研究。传感器网络所具有的集中式数据收集、多跳数据传输、多对一流量模式这三种特征会造成漏斗效应的出现,这会导致严重的包碰撞、网络拥塞、包丢失,甚至拥塞崩溃,还会导致能量消耗的热点出现,使某些节点甚至整个网络过早死亡。本文利用 CAG 算法在路由构造和分簇同步完成的优点,结合数据的空间相关性,提出了一个传感器节点分簇算法,该算法与传统 CAG 算法相比可以有效减少簇首数量,从而降低传输能量消耗。实验结果也验证了算法的有效性。

参考文献:

- [1] AKYILDIZ I F, SU W, SANKARASUBRAMANIAM Y, et al. Wireless sensor networks: a survey [J]. *Computer Networks*, 2002, 38(4): 393-422.
- [2] INTANAGONWIWAT C, GOVINDAN R, ESTRIN D. Directed diffusion: a scalable and robust communication paradigm for sensor networks [C]//Proc of International Conference on Mobile Computing and Networking (MobiCom). 2000:56-67.
- [3] CHU M, HAUSSECKER H, ZHAO Feng. Scalable information-driven sensor querying and routing for Ad hoc heterogeneous sensor networks [J]. *The International Journal of High Performance Computing Applications*, 2002, 16(3): 293-313.
- [4] WATTENHOFER R, LI Li, BAHL P, et al. Distributed topology control for wireless multihop Ad hoc networks [C]//Proc of IEEE INFOCOM. 2001:1388-1397.
- [5] GUPTA H, NAVDA V, DAS S R, et al. Efficient gathering of correlated data in sensor networks [J]. *ACM Trans on Sensor Networks*, 2008, 4(1): 1-31.
- [6] SUPRIYO C, TIM N, NIRVANA M, et al. A distributed and selforganizing scheduling algorithm for energy-efficient data aggregation in wireless sensor networks [J]. *ACM Trans on Sensor Networks*, 2008, 4(4): 1-41.
- [7] LAUKIK C S, ALIN D, SANJAY R. Aggregation methods for large-scale sensor networks [J]. *ACM Trans on Sensor Networks*, 2008, 4(2): 1-36.
- [8] HEINZELMAN W R, CHANDRAKASAN A, BALAKRISHNAN H. Energy-efficient communication protocol for wireless microsensor networks [C]//Proc of the 33rd Annual Hawaii International Conference on System Sciences. Maui: IEEE Computer Society, 2000:3005-3014.
- [9] SUNHEE Y, CYRUS S. The clustered aggregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks [J]. *ACM Trans on Sensor Networks*, 2007, 3(1).
- [10] MADDEN S, FRANKLIN M J. TAG: a tiny aggregation service for Ad hoc sensor networks [C]//Proc of the 5th Annual Symposium on Operating Systems Design and Implementation. [S. l.]: IEEE Press, 2002:131-146.