

传感器网络中基于域的分布式自动成簇算法研究

谢志军, 钱江波

(宁波大学信息科学与工程学院, 浙江宁波 315211)

摘要: 在传感器网络中,分簇是其他应用的基础,本文结合传感器网络的节点特性和位置信息,提出一种基于域的分布式自动成簇算法 DCAM(Distributed Clustering Auto Model),DCAM 把传感器网络按域划分来构建簇,簇之间是相互连通并且可以覆盖网络中所有的传感节点,簇头节点与簇中的网关节点就构建成网络的连通核,当传送数据时,传感节点只需在连通核中寻径,因而能明显减少寻径时间复杂度并且具有更好的分布性;然后在 DCAM 的基础上提出了簇的自愈和更新算法,更新和自愈算法可更大程度地延长整个网络的生命周期.

关键词: 传感器网络; 分簇; 分布式; 域

中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2010) 01-0218-04

Researches on Area Based Distributed Clustering Auto Model in Sensor Networks

XIE Zhi-jun, QIAN Jiang-bo

(College of Information and Engineer, Ningbo University, Ningbo, Zhejiang 315211, China)

Abstract: Clustering is the basic of other applications such as routing, query processing etc. Considering the characteristics and location information of nodes in sensor networks, a modified directed transfer model of sensor networks, and a novel clustering algorithm based on area DCAM(Distributed Clustering Auto Model) are proposed. Theoretical analysis and simulation show that the above new methods can reduce the energy costs of nodes in data transferring efficiently for sensor networks. So, it can prolong the lifetime of the whole networks to a greater degree.

Key words: sensor networks; clustering; distributed; domain

1 引言

分簇算法是根据无线传感器网络具体的应用需求按照某种规则将网络分成可以相互连通并覆盖所有节点的多个簇,并在网络结构发生变化时更新簇结构以维护网络的正常功能.分簇算法的目的就是通过某种规则,获得相互连通、覆盖所有节点的簇结构,并且随网络结构变化生成新的簇结构,确保感知信息正确地传递,使用较少的计算和通信开销来构造和维护一个能够覆盖整个网络、支持资源管理和路由协议的相互连接的簇集合.在层次结构网络中,分簇是其他应用的基础.

2 相关工作

文献[2]提出了 LEACH 算法,它将所有节点分为若干簇,每个簇选举一个簇头,其余的一般节点为簇内成员.其操作是分成轮进行的,每一轮包含簇建立和稳定运行两个阶段,文献[3]提出了 TEEN 协议,其重要特征是适应对时间敏感的数据采集应用.簇头节点给其簇内成员发送一个硬阈值(Hard Threshold)和一个软阈值

(Soft Threshold),只有当采集属性值处在令人感兴趣的范围内才允许传输数据.文献[4]提出了 TopDisc 算法,该算法利用颜色区分节点状态,网络中的一个节点启动发送用于发现邻居节点的查询消息,随着查询消息在网络中传播,算法依次为每个节点标记颜色.最后,按照节点颜色区分出簇头节点,并通过反向寻找查询消息的传播路径在簇头节点之间建立通信链路.文献[5]提出了高效采集协议 PEGASIS (Power Efficient Gathering in Sensor Information Systems),节点仅需与最邻近节点通信,用贪婪算法形成一条数据传输链,所有数据都沿该链逐跳传输到唯一的簇头节点.文献[6]提出了一种基于位置的路由协议 GPSR(Greedy Perimeter Stateless Routing)协议,节点利用贪心算法尽量沿直线转发数据,产生或收到数据的节点向以欧氏距离计算址靠近目的节点的邻节点转发数据,当遇到空洞时,利用右手法则沿空洞周围传输避免空洞.文献[7]提出了一种简单实用的满足 min-max 优化和近似优化的聚簇模式,当给定了一个节点域和聚簇的大小,地理位置相邻的节点形成一个簇,在簇内,数据沿着最短路径树传送到簇头节点,数

据在簇头节点进行聚集运算后,送到 Sink 节点.文献[8]提出了一种近似的最短路由算法,算法用图形标记来指派到一组分布式的标记向量中,通过数据源与目的节点的标记推断出最短路径.然后递归地把图形剖分成碎片,对每一个碎片 P,计算并储存 P 的最短路径树,每个节点将会有有一个标记,然后利用 unit disk graph 的延层面来构造上述图标记.文献[9]提出了一种传感器网络能量敏感的 Qos 路由算法,算法采用动态聚簇来提高首位能量消耗模型,基于这个动态簇,然后建立了一个多目标的规划模型来支持 Qos.

3 传感器网络中基于域的分簇模型

定义 1(核) 图 G 的节点集 $C \subseteq V$ 为核,当且仅当节点集 C 满足以下条件: $\forall p \in V \Rightarrow p \in C$ 或 p 为 C 中的某个节点 q 的邻节点.

定义 2(连通核) 给定图 $G = (V, E)$,若图 G 的节点集 $C \subseteq V$ 为满足如下条件的节点集合:由 C 导出的子图是连通图,且 C 是图 G 的一个核,则称 C 为连通核.

定义 3(域) 任意传感节点 p ,假定其对 Sink 节点 R 的相对坐标为 (x, y) ,Sink 节点 R 及各传感节点具有相同的有效通信半径 r ,则称 p 属于域 (m, n) ,当且仅当如下公式成立:

$$m = \lfloor x \rfloor \lfloor \frac{r}{\sqrt{2}} \rfloor, n = \lfloor y \rfloor \lfloor \frac{r}{\sqrt{2}} \rfloor$$

其中“ \lfloor ”为除法运算符;“ \lfloor ”为取大于等于的整数运算符.

引理 1 假定 Sink 节点 R 及各传感节点具有相同的有效通信半径 r ,则任意传感器网络存在满足定义 3 中公式的唯一域划分.

3.1 分布式成簇自动生成算法 DCAM

算法中我们假定传感器网络中各节点具有相同的有效通信距离,相邻节点间存在一条对称的通信链路,每个节点都能够通过 GPS 进行通信.对任意传感节点 p ,假定其地理坐标为 (x, y) .假定 Sink 节点 R 及各传感节点具有相同的有效通信半径 r ,则任意传感节点 p 可以依据定义 3 计算出自己所属的域,同时,根据引理 1 的域划分规则,我们可以给出以下分布式数据汇聚模型 DCAM(Distributed Cluster Auto Make)^[1]:

步骤 1:每个节点 p 利用 GPS 计算自己的地理坐标和剩余能量,并利用定义 3 计算出自己所属的域;

步骤 2:每个节点 p 周期性地和其所有邻节点交换如下信息:(i) p 的当前状态 S_p (分为簇头、网关或成员 3 种状态);(ii) p 的剩余能量 E_p ;(iii) p 所属的域 G_p ;(iv) p 的地理坐标.通过此操作,每个节点可以获知自己邻节点的状态、剩余能量、所属域及到邻节点的直线距离等信息;

步骤 3:初始时,图 G 中的 Sink 节点的状态为簇头,而所有传感器节点的状态均为成员.在每个周期内,每个节点 p 根据自己邻节点的状态 S_p 、能量 E_p 及所属域 G_p 的信息,按如下规则来计算自身新的状态:

(1) p 所属域 G_p 内无簇头,则 G_p 内的各节点依据剩余能量 E_p 的大小随机选举一个具有最大剩余能量的节点作为本域中的簇头;

(2) 否则,若 p 不为簇头或网关,且 p 与其他域中的节点相邻,则令 p 为网关;

步骤 4:任意某个节点产生传感数据后,将数据直接传送给自己所属域中的簇头.

文献[1]证明了引理 2 成立:

引理 2 若图 $G = (V, E)$ 是简单连通无向图,则由算法 DCAM 得到的节点集 $\Psi = \{p \mid p \text{ 为簇头或网关且 } p \in V\}$ 是图 G 的一个连通核.

由引理 2 可以得知:由 DCAM 得到的节点集(构成了一个连通核.所以,可以把传感器网络按域划分来构建连通核,传感节点只需要在连通核中进行路由,这样可以有效降低寻径的开销.

3.2 簇更新和维护

3.2.1 簇头节点轮换算法

簇头节点的死亡将会影响整个簇.我们采用簇头节点轮换的方式.在成员节点里记载本簇的簇头节点,当有数据要传送时,就把数据传送到簇头节点;在簇头节点里面要维护两个表,一张是路由信息表(T_R),用来记载该簇的路由信息,另外一张表是用来记载本簇内节点信息表(T_I),记载域内每个节点标识,状态和能量,基于这两张表,簇头节点轮换过程如下:

当簇头节点 H 发现自身的能量比域内任何节点的能量都少时,就挑出一个能量最大的节点 i 充当簇头节点.簇头节点 H 在簇内广播一个消息,宣布 i 充当簇头节点,同时修改自身的状态为成员节点状态.

步骤 1:各成员节点收到簇头节点 H 发来的消息以后,修改本簇的簇头节点信息为 i ,在下一个周期如果有数据要传送,则发送给 i .

步骤 2:节点 i 接收到簇头节点的信息后,修改自身状态为簇头状态.同时修改路由表(T_R).

步骤 3:相邻簇头和网关节点收到簇头节点 H 发送来的消息以后,修改自己的路由表,在下一个周期内就按新的路由表转发数据,即不再把数据转发到 H ,而是转发到节点 i .

3.2.2 簇的自愈算法

簇的自愈问题主要是旧节点消亡和新节点加入的情况.而旧节点消亡分为成员节点的死亡和簇头节点的死亡两种情况,其更新算法如下:

步骤 1: 当簇头节点 H 在 T_m 时间后没有收到某个成员节点 j 的数据包, 则认为 j 死亡. 由于在整个无线传感器网络中, 某个簇成员的死亡并不会对整个网络造成太大的影响, 故而对于这种情况只是在簇头节点的节点信息表(T_I)中简单地删除节点 j 的有关信息.

步骤 2: 若成员节点在 T_h 时间内没有收到簇头节点的信息, 则认为簇头节点死亡, 自身转入空闲状态, 启动域重组:

(1) 域内每个节点 q 和其所有邻节点交换如下信息: (i) q 的当前状态 S_q (分为簇头、网关或成员 3 种状态); (ii) q 的剩余能量 E_q ; (iii) q 所属的域 G_q ; (iv) q 的地理坐标. 通过此操作, 每个节点可以获知自己邻节点的状态、剩余能量、所属域及到邻节点的直线距离等信息;

(2) q 所属域 G_q 内无簇头, 则 G_q 内的各节点依据剩余能量 E_q 的大小随机选举一个具有最大剩余能量的节点 H 作为本域中的簇头;

(3) 否则, 若 q 不为簇头或网关, 且 q 与其他域中的节点相邻, 则令 q 为网关;

步骤 3: 簇头节点 H 生成本簇内节点信息的表(T_I), 同时向相邻的簇广播自己的 ID 号, 根据各相邻簇头节点和网关节点返回的消息生成一个路由信息表(T_R).

步骤 4: 各相邻簇头节点和网关节点根据簇头节点 H 广播的消息修改自身的路由表信息.

新节点加入算法:

步骤 1: 节点 X 新加入时根据其自身的地理位置计算自己属于哪个簇中, 其初始状态为成员.

步骤 2: 节点 X 把自己的 ID 号连同请求加入消息发送给簇头节点 H ;

步骤 3: 簇头节点 H 接收到节点 X 的请求加入消息后, 更改本簇内节点信息的表(T_I), 同时向节点 X 发送允许加入信息.

步骤 4: 节点 X 收到簇头节点的允许加入信息后, 把产生的数据发送给簇头节点进行数据的融合运算.

定理 1 若图 $G = (V, E)$ 是简单连通无向图, 在经过簇头轮换算法和簇的自愈算法后得到的节点集 $\Psi = \{p | p \text{ 为簇头或网关且 } p \in V\}$ 仍是图 G 的一个连通核.

证明 由引理 2 可知, 由算法 DCAM 得到的节点集 $\Psi = \{p | p \text{ 为簇头或网关且 } p \in V\}$ 是图 G 的一个连通核. 由定理 1 以及算法 DCAM 的步骤 3(1)可知: 图 G 中的每个节点要么为簇头, 要么它的邻节点中至少存在一个簇头, 即与某个簇头相邻. 因此, 集合 Ψ 是图 G 的一个核. 下面我们证明经过簇头节点轮换算法以后集合 Ψ 仍然是连通的.

设 p, q 为 Ψ 中的任意两个簇头节点, 即 $p, q \in \Psi$.

设节点 i 是簇头节点 p 所在域中的候选簇头节点, 由于假定传感节点具有相同的有效通信半径, 故为表述方便, 我们不妨设均为 1 个单位长度, 由引理 2 可知, 由于图 G 是连通的, p, q 之间的距离即为 p, q 之间的最短路径的长度, 记为 $d(p, q)$, 因此 $d(p, q) < R$ (R 为一实数), 也即 $d(p, q)$ 为有限整数. 经过轮换算法以后, 节点 i 为节点 p 所在域的簇头节点 (簇头节点), 只要证明 $d(i, q) < R$ (R 为一实数) 即可.

(1) 从引理 1 可知同一域中的各传感节点互为邻节点, i, p 相邻 $\Rightarrow d(i, p) = 1$;

(2) 由引理 2 可知, p, q 是连通的, 即存在 G 中的路径 $(p, r_1, r_2, \dots, r_m, q)$, 故有 $d(p, q) = m$;

(3) 由 (1) 和 (2) 可知, $d(i, q) = (1 + m) < R$; 所以 i, q 是连通的.

同理可证节点加入和簇头节点死亡时也是连通的.

4 算法性能分析及实验结果

为了具体分析比较新方法的节能性, 我们将新方法进行了模拟实验, 并将实验结果与典型的分簇算法 LEACH、TEEN 和 PEGASIS 算法的性能进行了比较.

网络能量消耗是衡量算法的重要指标, 我们首先根据给定的有效通信距离 CR (Communication Radius) 设定两种典型的场合来比较算法的性能: 场合 1 中在一个 100×100 区域内, 随机部署 100 个初始能量为 50 单位的传感节点, 并随机生成一个 Sink 节点; 场合 2 中在一个 400×400 的区域内, 随机部署 500 个初始能量为 50 单位的传感节点, 并随机生成一个 Sink 节点进行仿真实验. 在模拟实验中, 实验数据均为执行 LEACH 和 DCAM 算法各 100 次, 选取其平均值为最后结果. 具体实验结果如图 1 和图 2 所示.

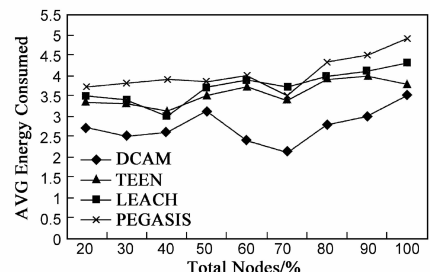


图1 场合1中DCAM和其他算法的节点耗能比较

在传感器网络中网络寿命也是评价算法性能的重要指标. 本文将节点寿命评价指标定义为三种形式: 第一个节点死亡、一半节点死亡和全部节点死亡. 在上述的两个典型的场合来比较 DCAM 和其他算法的性能图 3 和图 4 分别是两种场合中剩余节点百分比随时间变化图. 从图中可以看出, 无论是第一个节点死亡还是一半节点死亡和全部节点死亡的时间, DCAM 的时间都比

其他算法好.

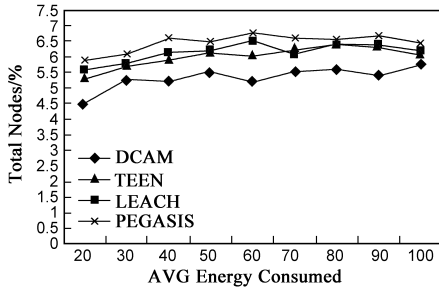


图2 场合2中DCAM和其他算法的节点耗能比较

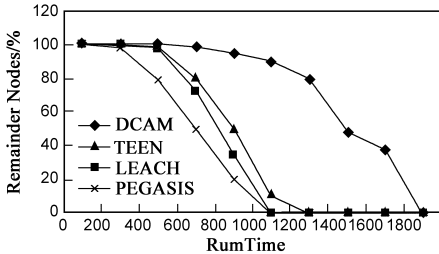


图3 在场景1中DCAM与LEACH算法的寿命

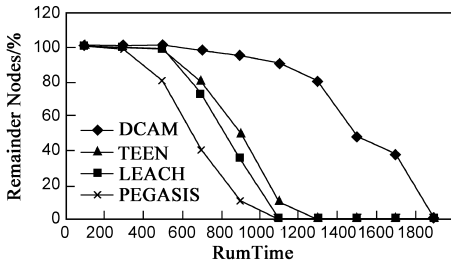


图4 在场景2中DCAM与其他算法的寿命

网络负载均衡程度是无线传感器网络的一个很重要的性能指标.当某个传感器节点负载很重时,很容易成为网络运行的瓶颈;由于每个簇中所包含节点数目的差异会导致不同簇之间数据的容量不同,因此簇的负载均衡程度是分簇式路由算法中衡量算法性能的重要标准之一.网络负载平衡性能通过负载平衡因子来表现,BF的数值越大说明簇的负载越平衡.图5给出了仿真结果,DCAM算法的负载均衡程度总体上高于其他算法.

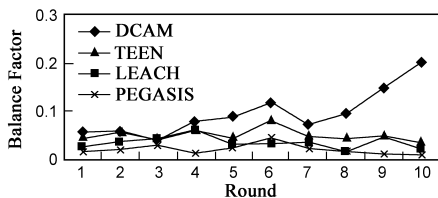


图5 平衡负载度比较

5 结论

本文结合传感器网络的节点特性和位置信息,提出了一种基于域的分簇算法 DCAM,DCAM 把传感器网络按域划分来构建簇,簇之间是相互连通并且可以覆

盖网络中所有的传感节点.簇头节点与簇中的网关节点构建成网络的联通核,当传送数据时,传感节点只需在连通核中寻径,因而能明显减少寻径时间复杂度并且具有更好的分布性;对比传统的分簇算法,新算法可更大程度地降低传感器节点数据传输的能耗,从而可进一步延长整个网络的生命周期.

参考文献:

- [1] 谢志军,王雷,等.无线传感器网络中基于数据压缩的汇聚算法研究[J].软件学报,2006,17(4):860-867.
- [2] W Heinzelman, A Chandrakasan, H Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks [A]. Proceedings of the 33rd Annual Hawaii International Conference on System Sciences [C]. Maui: IEEE Computer Society, 2000. 3005-3014.
- [3] A Manjeshwar, et al. TEEN: A protocol for enhanced efficiency in wireless sensor networks [A]. Proceedings of 15th International Parallel and Distributed Processing Symposium [C]. San Francisco: IEEE Computer Society, 2001. 2009-2015.
- [4] W R Heinzelman, A Chandrakasan, H Balakrishnan. An application-specific protocol architecture for wireless microsensor networks [J]. IEEE Transactions on Wireless Communications, 2002, 1(4): 660-670.
- [5] S Lindsey, C Raghavendra. PEGASIS: Power-efficient gathering in sensor information systems [A]. Proceedings of IEEE Aerospace Conference [C]. San Francisco: IEEE Computer Society, 2002. 125-1130.
- [6] B Karp, H Kung. GPCR: Greedy perimeter stateless routing for wireless networks [A]. Proceedings of the 6th Annual International Conference on Mobile Computing and Networking [C]. Boston: ACM Press, 2000. 243-254.
- [7] S Patten, B Krishnamachari, R Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks [J]. ACM Transactions on Sensor Networks, 2008, 4(4): 24:1-24:32.
- [8] J BRUCK, J GAO, A JIANG. Localization and routing in sensor networks by local angle information [J]. ACM Transactions on Sensor Networks, 2009, 5(1): 1-31.
- [9] K Zhou, L Meng, Z Xu. A dynamic clustering-based routing algorithm for wireless sensor networks [J]. Information Technology Journal, 2008, 7(4): 694-697.

作者简介:

谢志军 男,1974年出生,博士后,副教授.主要研究领域为传感器网络、数据库和数据流. E-mail: xiezhiyun@nbu.edu.cn

钱江波 男,1974年出生,博士,副教授,硕士生导师.主要研究领域为数据库和数据流管理技术、数据挖掘、逻辑电路设计等.