

异构无线传感器网络中簇首的优化部署策略

刘琳¹, 于海斌²

(1. 国家卫星海洋应用中心, 北京 100081; 2. 中国科学院 沈阳自动化所, 辽宁 沈阳 110016)

摘 要: 分簇对于环境监测的无线传感器网络具有较好的适应性和节能性, 由高能力节点担任簇首可以更好地实现节能并改善网络性能。在大规模网络中, 高能力簇首部署问题是一类 NP-hard 问题。同时考虑了影响网络性能的簇半径和簇首的容量限制, 把簇首的优化部署问题形式化为一个整数规划问题。针对该问题, 提出了启发式的基于 K -平均和模拟退火混合算法的 KMSA 算法, 对簇首节点进行有策略的部署。仿真结果表明, 在不同的网络规模和簇个数情况下, KMSA 可以改善网络性能。

关键词: 异构无线传感器网络; 簇首; K -平均; 模拟退火

中图分类号: TP393

文献标识码: A

文章编号: 1000-436X(2010)10-0229-09

Optimal cluster head deployment in heterogeneous wireless sensor networks

LIU Lin¹, YU Hai-bin²

(1. National Satellite Ocean Application Service, Beijing 100081, China;

2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China)

Abstract: Cluster-based method has better adaptability and energy-efficiency to wireless sensor networks (WSN) used for environmental monitoring. If the cluster head is served by more powerful node, the performance of WSN will be improved greatly. In large scale WSN, high-power cluster head deployment is a kind of NP-hard problem. The optimal problem of cluster head deployment was formulated as an integer programming with the condition of restrictions of cluster head capacity and the maximal cluster radius. To satisfy the time effectiveness, a heuristic algorithm called KMSA was proposed which was a hybrid algorithm of K -mean and simulated annealing. The simulation results show that the KMSA can improve the performance of WSN on varieties of network size and cluster number.

Key words: heterogeneous wireless sensor networks; cluster head; K -mean; simulated annealing

1 引言

无线传感器网络 (WSN, wireless sensor networks) 是由微型智能传感器节点 (SN, sensor node) 自组形成的多跳网络。WSN 领域的早期研究多集中在同构网络, 即网络内所有 SN 节点软硬件能力

相同, 节点被随机部署在监测区域内, 地位平等, 网络中无中心管理节点。由于 SN 节点由电池供电, 不易更换, 节点软硬件能力有限等因素使得网络协议简单, 节点通信能力差, 网络生命期短, 导致同构网络不可靠, 难以投入实际应用。近年来, 人们在研究和实验中更多采用异构 WSN^[1~3], 也就是在

收稿日期: 2010-05-19; 修回日期: 2010-09-20

基金项目: 国家自然科学基金资助项目 (60434030); 国家杰出青年科学基金资助项目 (60725312)

Foundation Items: The National Natural Science Foundation of China (60434030); The National Science Fund for Distinguished Yong Scholars (60725312)

WSN 中部署一些高能力的节点来完成复杂繁重的任务,从而减轻 SN 节点的负担,达到延长网络生命期的目的。高能力节点具有可持续的能量资源,不需要对其进行频繁更换,其通信能力和数据处理能力也远高于 SN 节点。因此,相对于同构 WSN,异构的 WSN 是一种更现实的网络模型^[4]。

分簇问题是 WSN 研究领域的热点问题,通过分簇可以提高网络的可管理性,延长网络生命期。该问题的研究大多以同构网络为背景,网络中的节点都有可能成为簇首,典型的有 LEACH^[5,6]协议、DEEUC^[7]协议等。基于同构 WSN 的簇首选择适用于节点随机部署的网络,节点都由低能力的传感器节点构成。簇首是局部网络的处理中心,管理的 SN 节点越多,消耗的能量越大。为保证簇首间的能量均衡,需要频繁更新簇首节点,而簇首更新需要在网络内进行频繁的信息交换,带来巨大的通信开销,因此簇维护成本在一定程度上抵消了簇结构带来的好处。

若用高能力的节点担任簇首则会大大减少簇维护带来的开销,加入了高能力节点的 WSN 是一种异构的 WSN。但由于高能力簇首节点成本较高,不可能在网络中大量部署,甚至不能保证 SN 节点和高能力簇首之间能够一跳到达,因此需要在网络成本和网络性能之间做出均衡,在满足性能的基础上使部署的簇首个数尽量少。簇首个数越少,则单个簇首管理的 SN 节点数越多,若簇首位置布放不合理,可能会使某些簇首节点负载过重。因此除了簇首个数,簇首位置也会对网络性能产生影响。因此,在异构传感器网络中,为使网络性能最优,高能力的簇首节点需放置多少、安装在哪里是网络投入使用前首先要解决的问题。而对于节点数目众多的 WSN 而言,从网络部署区域计算出最佳的簇首节点位置是一个 NP-hard 问题^[8],需要对该问题进行简化,计算其次优解。

文献[8~10]首先筛选出有限的候选位置,然后从这些候选位置中选出最佳的几个来放置簇首节点。文献[8,9]中的候选位置为所有传感器节点位置,文献[8]选择最佳位置的目标是最大化网络生命期,并把该问题形式化为整数线性规划问题,利用禁忌搜索方法得到次优解。文献[9]选择最佳位置的目标是缩短报文平均传输路径,提出一种分布式的 1-hop 方法。文献[10]把簇首部署问题归结为最小集合覆盖问题,这也是一类 NP-hard 问题。文章提出

一个迭代算法得到了次优解,每次迭代都从簇首的候选位置中找出那些覆盖 SN 节点最多的位置,在该位置部署实际的簇首节点。上述研究都需要事先确定簇首的候选位置,当网络中节点数较多时,算法计算复杂。文献[11]提出一种简单算法,称之为 BSL (best sink locations) 方法,该方法从网络中随机选择 K 个 SN 节点的位置作为初始聚类中心,通过 K -mean 算法计算出最佳的聚类中心放置簇首。 K -mean 方法计算量小,但是算法具有初始值敏感性,若初始值选择不当,得不到较好的次优解。

综上所述,文献[8~11]研究的问题与本文相似,都是解决簇首节点在 WSN 中的优化部署问题。不同的是上述研究中,簇首部署的目标主要考虑降低总能耗,延长网络生命期,却没有考虑网络成本、簇首间的均衡特性和数据收集实时性等问题。本文以用于实时环境监测的 WSN 为背景,讨论了网络中高能力簇首节点的部署问题,对于大规模网络,该问题是一类 NP-hard 问题,为此本文提出了一种启发式的基于 K -mean 和模拟退火混合算法的簇首部署 (cluster head deployment based on hybrid of K -mean and simulated annealing) 策略,算法充分考虑了数据收集实时性和簇首的容量限制,根据当前网络状态信息,在不增加网络成本的基础上使网络性能达到最优。

2 簇首部署问题分析描述

2.1 网络模型定义

本文讨论的 WSN 网络结构如图 1 所示, SN 节点随机分布在监测区域内,被划分为 3 个簇,每个簇内部署了一个高能力的簇首节点 CH。SN 节点负责实时感知和采集周边环境的相关数据,并周期性地采样数据通过多跳传送到所属簇的 CH, CH 收集到的数据经处理后发送到网络中唯一的基站 BS。CH 之间以及 CH 和基站(BS)之间采用高可靠的通信协议,如 802.11,形成骨干网络。上述网络模型满足以下假设条件:

- 1) SN 节点随机密集部署且位置可知,节点能够自组形成一个静态连通网络,链路具有对称性;
- 2) SN 节点具有相同的软硬件,相同的通信能力,由电池供电且初始能量相同;
- 3) SN 节点周期性把采集到的数据发送至簇首,所有节点的数据产生速率相同,每个周期产生一个固定长度的报文;

4) SN 节点之间以及 SN 和簇首之间的通信满足自由空间模型;

5) 簇首节点是高能力节点, 无能量限制, 负责收集和处理簇成员发送的信息, 处理后的结果发送到网络中唯一的基站;

6) 簇首节点和 SN 节点均有容量限制, 节点容量与单位时间内节点能够收发和处理的报文数有关;

7) 簇内 SN 节点到簇首之间采用利用 Floyd-Warshall 方法计算出的最短路径进行数据传输。

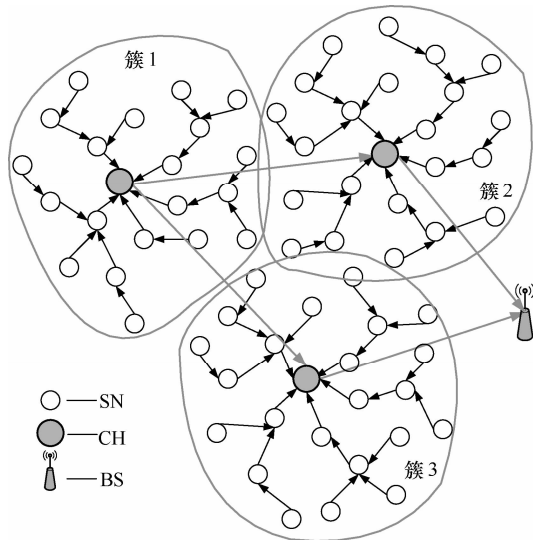


图 1 异构 WSN 网络结构

图 1 中 SN 节点集合用 V_{SN} 表示, 节点数为 n 个, $V_{SN} = \{1, 2, \dots, n\}$ 。假设所有可能放置簇首节点的 m 个位置 (m 可能为无穷) 都存在一个虚拟簇首节点, 用 V_{CH} 表示此类节点集, $V_{CH} = \{n+1, n+2, \dots, n+m\}$ 。 V_{SN} 和 V_{CH} 中的节点共同构成一个连通的网络, 可抽象化为一个网络图 $G=(V, E, \omega)$, 其中, V 为节点集, $V = V_{SN} \cup V_{CH}$; E 为边集, 也是网络中的通信链路集合。

$$E = \{(i, j) \mid d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} < R, \forall i, j \in V \wedge i \neq j\}$$

其中, R 为 SN 节点的无线通信半径, (x_i, y_i) 为 V 中任意节点 i 的位置坐标; 每条边对应一个权值 e_{ij} , ω 为对应 E 中每条边的权值集合, $\omega = \{e_{ij} \mid (i, j) \in E\}$ 。下面给出用到的几个定义。

定义 1 边的权 (weight), 用 e_{ij} 表示, $(i, j) \in E$, 权值可以定义为两点间的距离或链路信号强度。

定义 2 路径, G 内任意两节点 i_0, s 之间的路

径 P^{i_0s} 为 G 的一个子图, $P^{i_0s} = (V^{i_0s}, E^{i_0s})$, $V^{i_0s} = \{i_0, i_1, i_2, \dots, i_m, s\} \subseteq V$, $E^{i_0s} = \{(i_0, i_1), (i_1, i_2), \dots, (i_m, s)\} \subseteq E$ 。

定义 3 路径长度, 为路径上各链路的权重之和, 表示为 $e(P^{i_0s}) = \sum_{(v_k, v_j) \in E^{i_0s}} e_{kj}$ 。

定义 4 最短路径, $P_{min}^{i_0s}$ 为 G 内任意两节点 i_0, s 之间所有路径中长度最小的路径, $P_{min}^{i_0s} = (V_{min}^{i_0s}, E_{min}^{i_0s})$; $P_{min}^{i_0s} = \arg \min_{P^{i_0s}} \{e(P^{i_0s})\}$ 。

定义 5 簇大小, 为簇内 SN 节点的个数。

定义 6 簇半径, 定义为簇内最长最短路径长度。

定义 7 节点容量, 单位时间内节点能够收发和处理的报文最大数。

簇首节点的容量和簇内节点数目密切相关, 簇内节点数越多, 簇首的负载越大, 因此定义簇首的容量为簇内可容纳的最多簇成员数量。

2.2 簇首部署问题的形式化描述

由于簇首的容量限制使得簇内成员个数不可能无限多, 否则可能造成报文丢失或影响簇首的处理速度, 从而影响系统可靠性。因此, 在簇首数量一定的情况下, 簇间的负载均衡是满足簇首容量限制的前提。另外, 在某些应用环境中, 数据收集的实时性是一个重要的评估指标。实时性要求 SN 节点在规定的时间内把数据发送到簇首, 由于报文在每一跳的数据处理时间是延迟产生的主要原因, 因此最小化传输时延要求簇内传输路径的最小化。报文传输的跳数除对时延产生影响外, 还对网络多个性能指标产生影响: 首先, 跳数越多, 报文丢失的可能性就越大, 报文传输可靠性降低; 其次, 在多跳网络中, 报文发送经历的跳数越多, 参与转发该报文的节点数越多, 因此节点能耗越大; 最后, 跳数越大, 通信受到干扰的节点数也越多。因此, 簇首的部署以及簇划分除了需要考虑到 SN 节点的节能性外, 还要考虑簇首的容量限制和簇内的跳数限制。

本文规定网络中所有边的权值为 1, 因此两节点间的通信路径为最小跳数路径。为表述方便, 下文中的路径均指最短路径。下面给出用到的参数定义。

常数参数:

K: 簇首节点最大个数 ($K \ll n$);

CM: 簇首的容量限制;

SM: SN 节点的容量限制;

maxH: 簇半径限制, 也就是簇内最长路径的跳数限制。

变量参数:

$$z_j = \begin{cases} 1, & \text{如果选择位置 } j \text{ 放置簇首节点} \\ 0, & \text{其他} \end{cases} \quad \forall j \in V_{CH}$$

$$x_{ij} = \begin{cases} 1, & \text{如果节点 } i \text{ 选择 } j \text{ 做为簇首} \\ 0, & \text{其他} \end{cases} \quad \forall i \in V_{SN}, \forall j \in V_{CH}$$

$$p_v^{i_0s} = \begin{cases} 1, & v \in V_{min}^{i_0s} - \{i_0, s\} \\ 0, & \text{其他} \end{cases} \quad \text{表示路径 } P_{min}^{i_0s} \text{ 是否经过点 } v$$

$\forall v \in V_{SN}$, 经过节点 v 的所有路径数

$$p_v = \sum_{j \in V_{CH}} \sum_{i \in V_{SN}} x_{ij} z_j p_v^{ij}$$

如果网络被划分成 k 个簇 ($k \leq K$), 分别表示为 C_1, C_2, \dots, C_k , $C_1 \cup C_2 \cup \dots \cup C_k = V_{SN}$, $C_1 \cap C_2 \cap \dots \cap C_k = \Phi$. $C_o = \{o_1, o_2, \dots, o_k\}$ 为簇首节点集合, 也表示簇首节点的位置集合, $C_o \subseteq V_{CH}$. c_1, c_2, \dots, c_k 分别为对应各个簇大小, 那么

$$c_j = \sum_{i=1}^n x_{ij} z_{o_j}, j=1, \dots, k$$

平均簇大小为 $\mu = \frac{N}{k}$. 式(1)为各个簇大小的均方差, 用它来表示每个簇大小与平均值的偏离程度, σ_c 越小表示簇划分越均衡, 同时也意味着簇间的负载均衡。

$$\sigma_c = \sqrt{\frac{1}{K} \sum_{j=1}^K |\mu - c_j|^2} \quad (1)$$

第 i 个簇的簇半径可以用式(2)表示:

$$r_i = \max_{v \in C_i} \{e(P_{min}^{v o_i})\} \quad (2)$$

用 r_c 表示网络的最大簇半径, r_c 通过式(3)来计算, 那么 r_c 越小, 表示网络内节点到各自簇首的最长路径越短, 数据收集的实时性越好。

$$r_c = \max_{i=1, \dots, k} \{r_i\} \quad (3)$$

当网络簇划分的个数一定时, 各个簇半径 r_c 和簇间的均衡特性 σ_c 是衡量网络数据收集实时性和可靠性的重要指标。而网络中 SN 节点位置的随机特性难以保证上述 2 项指标同时达到最优, 因此需要对两者做出均衡, 经过大量的仿真发现利用两者的乘积 $r_c \sigma_c$ 作为簇个数时, 簇首部署的目标优化函数能够得到较好的仿真结果。下面给出异构 WSN 中簇首部署问题的形式化描述:

目标函数: $f = \min\{r_c \sigma_c\}$

约束条件:

- ① $\sum_{j \in CH} z_j \leq K$;
- ② $\sum_{j \in CH} x_{ij} z_j = 1, \forall i \in V_{SN}$;
- ③ $\sum_{i \in SN} x_{ij} z_j \leq CM, \forall j \in V_{CH}$;
- ④ $\sum_{j \in CH} \sum_{i \in SN} x_{ij} z_j p_v^{ij} \leq SM, \forall v \in V_{SN}$;
- ⑤ $e(P_{min}^{ij}) x_{ij} \leq \max H, \forall i \in V_{SN}, \forall j \in V_{CH}$;
- ⑥ $x_{ij} \leq z_j, \forall i \in V_{SN}, \forall j \in V_{CH}$;
- ⑦ $x_{ij}, z_j, p_v^{ij} \in \{0, 1\}, i, v \in V_{SN}, j \in V_{CH}$ 。

簇首的部署问题被形式化为一个整数规划 (IP) 问题, 目标函数中 σ_c, r_c 可利用式(1)和式(3)计算得到, 约束条件①表示网络中最多有 K 个簇首; 条件②是指每个 SN 节点有且只有一个簇首; ③表示每个簇的成员个数不能超过簇首的容量限制 CM ; ④指经过每个 SN 节点的转发负载不得超过节点的容量限制 SM ; ⑤限制了节点到簇首的路径长度不得超过 $\max H$; ⑥表示簇首节点被部署后, SN 节点才可以向其发送报文; ⑦为 x_{ij} 和 z_j 的取值范围。当节点数很少时 (几十个), 整数规划问题可以用现有的软件 CPLEX 和 Xpress-MP 来解决。如果节点数很多 (几百个以上), 则该问题是 NP 问题, 难以计算最优解, 因此需要一个近似算法来求解。

3 启发式的簇首部署策略 KMSA

本文启发式的簇首部署策略是在 LEACH-C^[6] 算法和 BSL^[11] 算法基础上的改进, 考虑了实际应用中的节点容量限制和数据收集实时性需求, 为了解决容量和跳数约束下大规模网络的簇首部署问题提出了一种启发式的基于 K -平均(KM, K -mean)和模拟退火(SA, simulated annealing)混合算法的簇首部署(KMSA, K -mean and simulated annealing)策略。

3.1 KM 算法和 SA 算法简介

KM 算法是一种常用于数据挖掘的基于划分的聚类算法, 所谓聚类就是把数据集中相似度较高的数据对象划分成一个簇, 而簇与簇之间数据对象的差别较大。相似度是指簇内成员与聚类中心的接近程度, 越接近, 相似度越小^[12]。KM 算法思想简单、容易实现、收敛快、内存占用小、能有效地处理大数据集。但是 KM 算法只能保证收敛到局部最优, 导致聚类结果对初始聚类中心的选择非常敏感^[13]。若把上述定义中的数据集看

作 SN 节点的位置集合, 则可以利用 KM 算法对 WSN 分簇。但是对于分布不均匀的 WSN, 由于局部的节点密度不同, 加之 KM 算法的初始值敏感性, 经 KM 计算后得到的簇大小或簇半径不均衡。本文前期对 KM 方法的初始值选择对系统性能的影响进行了研究分析, 研究结果表明, 无论何种初始值选择方法, 得到的分簇后的网络性能变化不大。对于局部节点密度不相同的 WSN, 无法用该方法实现簇间的均衡性^[14]。

模拟退火算法是一种适合解决大规模组合的优化问题, 特别是对 NP-hard 问题通用的有效全局优化方法^[15]。其基本思想是利用随机化问题求解过程与统计力学中热平衡问题的相似性, 通过设定初温、初态以及降温率控制温度 t 的不断下降, 利用具有概率突跳特性的 Metropolis 准则在解空间进行随机搜索, 并以 Metropolis 准则接受新解。该准则描述如下: 若 S' 和 S 分别表示新解和当前解, $f(S')$ 和 $f(S)$ 为对应解的目标函数值, $\Delta f = f(S') - f(S)$, 若 $\Delta f < 0$, 接受 S' , $S \leftarrow S'$; 否则以概率 p 接受 S' , p 由式(4)计算出, 其中, t_k 为当前温度:

$$p = e^{(-\Delta f / t_k)} \quad (4)$$

优化的目标函数相当于金属的内能, 优化问题的解空间相当于金属内能状态空间, 问题的求解过程就是寻找一个解, 使目标函数值最小。SA 算法由于采用随机搜索策略, 只能在概率意义上求得全局最优解, 但是可以证明, SA 算法最终能以概率 1 收敛于全局最优解^[16]。SA 算法具有初始解顽健性强、通用性好以及算法简单易实现等优点。由于 SA 算法的搜索过程是随机的, 且当 t 值较大时可以接受部分恶化解, 但也可能从最优解跳出, 随着 t 值的衰减, 恶化解被接受的概率逐渐减小直至为 0, 而某些当前解要达到最优解时必须经过暂时恶化的“低谷”, 因此, 停止准则无法保证最终解正好是整个搜索过程中曾经达到的最优解。为解决该问题, 采用了一种“有记忆”的 SA 算法, 算法思想将在 3.2 节中详细描述。

3.2 KMSA 算法设计

KMSA 利用 KM 和 SA 2 种方法的特点来解决大规模网络的簇首部署问题, 利用 SA 算法的初始值顽健性来弥补 KM 算法初始值敏感性的不足, 同时利用 KM 算法的局部最优特性来避免因 SA 算法的“突跳”特性错过最优解。除 KMSA 算法主体外, 簇首节点优化部署主要分为以下几步:

- 1) 基站收集 SN 节点位置信息, 以及各节点的邻居信息, 根据这些信息计算出网络拓扑;
- 2) 利用 KMSA 算法计算满足需求的簇首节点位置, 部署簇首;
- 3) 部署好簇首节点后, 簇首节点向周围 maxH 跳范围内发送广播信息, 其中包含簇首的位置信息等;
- 4) SN 节点选择距离自己最近的簇首并请求加入;
- 5) 建立簇内通信路径开始数据收集。

要对网络进行合理部署需要收集当前网络状态信息。由于 SN 节点部署在复杂环境内, 环境中的障碍物、噪声干扰影响通信, 因此在彼此通信范围内的 2 个节点间不一定存在通信链路。为使簇首的部署更适合实际应用环境, 需要收集网络实际状态信息, 网络状态信息收集不是本文讨论的重点, 感兴趣的可参考文献[17]。以下为第 2) 至 5) 步的详细设计。

KMSA 采用 SN 节点的位置作为簇首节点的候选位置集合 V_{CH} , 算法的主要思想如下: 利用 SA 算法选择初始聚类中心, 利用基于 KM 的 KMCD^[14] 算法对节点进行簇划分, 并计算出簇首的位置, 利用分簇后得到的网络计算出目标函数值, 如果该值可接受或满足其他终止条件, 则算法终止。由于 SA 有可能跳过已经查找到的最优解, 因此采用了“有记忆”的 SA 算法, 也就是定义一个存储器 optSolution 存放曾经搜索到的最优解, 当算法终止时若得到的最优解 nowSolution < optSolution, 则 nowSolution 为最终结果, 否则, optSolution 为最终结果。算法描述如下, 其中, 函数 Fobjfunc 中新解的产生方法采用了 Metropolis 抽样策略, 通过对当前解 nowCH 做随机扰动产生一个新解 newCH, 算法具体实现如下: 从节点集合 $\{V_{SN} - \text{nowCH}\}$ 中任选 2 个节点 v_1 、 v_2 替换当前解 nowCH 中任选的 2 个节点 v_1 、 v_2 , 构成新解 newCH, 也就是与 nowCH 相比, newCH 中只 2 个节点被更新。

//KMSA 主程序

//常量定义:

InitTemp=100; // 初始温度

StopTemp=1e-8; //终止温度

StopOF=-inf; //目标函数的可接受值, 小于该值时程序终止, 这里设为负无穷

MaxTries=200; // 内循环的最大迭代次数

MaxNewSolutionRej=50; //新解未被更新次数, 达到该值时程序停止运行

```

MaxOptSolutionRej=1 000; //最优解未被更新次数, 达到该值时程序停止运行
arfa=0.95; // 降温率
NumberofCH=k; //网络分簇个数
//输入量
V; //网络 SN 节点 ID 及对应位置
r; //SN 节点通信半径
//参数初始化:
V0=V; T=InitTemp; L= MaxTries; //
NoUpdateNew= MaxNewSolutionRej; //新解未被更新计数器
NoUpdateOpt= MaxOptSolutionRej; //最优解未被更新计数器
termination=false; //程序终止标记, 为 true 时终止
nowCH=1: NumofCH; //初始簇首位置, 用节点 ID 表示
nowSolution=Fobjfunc(V0,nowCH); // 求对应 nowCH 的目标函数值
optCH= FsaveOptCH(nowCH); //保存当前最优解
optSolution=nowSolution; //
updated=false; //解更新标记
do while (termination==false)
{   for i=1:L //内循环
    {
        updated=false;
        newCH=Fgenerate(nowCH, NumofCH);
        //产生新解的初始簇首位置
        newSolution=Fobjfunc(V0, newCH);
        // 求对应 newCH 的目标函数值
        diff=newSolution-nowSolution;
        if (diff<0) or (exp(-(diff/T))> random(0,1)) //利用 Metropolis 准则接受新解
        {   nowCH=newCH;
            nowSolution=newSolution;
            NoUpdateNew= MaxNewSolutionRej; //重置新解未被更新计数器
            updated=true;
        }
        else //利用 Metropolis 准则拒绝新解
            NoUpdateNew= NoUpdateNew-1;
            //更新新解未被更新计数器
        if updated
            if nowSolution<optSolution

```

```

{
    NoUpdateOpt=MaxOptSolutionRej;
    //重置最优解未被更新计数器
    optSolution=nowSolution; //
    optCH=FsaveOptCH(nowCH);
    //保存当前最优解
}
else
{
    NoUpdateOpt=NoUpdateOpt-1;
    //更新最优解未被更新计数器
    if nowSolution==obtSolution
        obtCH=FaddObtCH(nowCH);
        //添加新的最优解
}
if (NoUpdateNew ==0) or (NoUpdateOpt==0)//检验程序终止条件是否被满足
    termination=true;
} //end of for
T=T*arfa; //更新当前温度
if (T<StopTemp) //是否满足程序终止条件
    termination=true;
} // end of while cycle
return;

```

KMSA 可以计算出满足约束条件的簇首部署位置。簇首节点部署完成后向周围 $\max H$ 跳范围内发送广播报文, 报文内容包括自身 ID、位置信息、以及经过的跳数。SN 节点记录所有接收到的来自不同簇首的、跳数最小的广播信息, 并记录簇首 ID 及到达该簇首的跳数。根据广播信息中的位置信息选择距离自己最近的簇首作为所属的簇首, 并发送一个请求加入报文, 报文中包含自身位置信息、节点 ID 等。当簇首节点接收到请求加入报文会判断已加入的节点是否超出自身容量限制, 若没有, 回送允许加入报文, 否则回送拒绝加入报文。若 SN 节点的加入请求被拒绝, 它会在其余簇首节点中选择一个最近的, 重复上述过程直至加入请求被允许。

分簇完成后, SN 节点以一定的周期向簇首发送数据。数据报文采用一种负载均衡的路由协议, 每个节点转发报文的同时记录本周期开始后已经转发的报文数, 并在发送 ACK 报文时通过捎带的方式发送给上游邻居节点, 因此每个节点要维护一个上游

邻居、下游邻居及其转发负载的列表。当上游邻居节点再次往上游节点发送报文时可以选择负载最小的下游邻居发送，实现了节点间的负载均衡，关于负载均衡路由协议的详细设计参考文献[18]。

4 仿真分析

KMSA 算法的仿真主要针对 3 个性能指标：

1) 可靠性，主要由簇大小均方差 σ_c 来衡量；2) 实时性，主要由簇半径 r_c 来衡量；3) 网络生命期，定义为网络从开始运行到网络内第一个传感器节点因能量耗尽而死亡的时间。采用 MATLAB 对算法进行了仿真，为验证 KMSA 方法的性能，用 LEACH-C 算法和 KMCD 算法与之相比较。由于 LEACH-C 以完全由低能力 SN 节点构成的同构 WSN 为背景，这里采用一个改进的 LEACH-C(ILC, improved LEACH-C)方法，簇首的选择利用 LEACH-C 的思想，而簇首由高能力节点来担任，只需选择一次，而且簇成员可通过多跳到达簇首。仿真比较了上述 3 种方法在簇个数相同情况下簇大小均方差 σ_c 、簇半径 r_c 和网络生命期 3 个性能指标。由于 ILC 和 KMCD 的仿真结果都具有随机性，因此采用算法运行 50 次得到的结果的平均值。

4.1 仿真场景和参数设置

假设 SN 节点随机部署在一个方形平面区域内，节点在区域内的密度 ρ 一定。SN 节点的初始能量相同，为 E_0 ；SN 节点，簇首节点的通信半径相同，均为 R 。相邻节点之间的通信满足自由空间模型。参数定义见表 1。

参数	值
SN 节点个数 n	100 200 300 400 500
SN 节点密度 ρ	0.04/m ²
分簇个数	2~7
通信半径 R	10m
初始能量 E_0	5J
报文长度	1 024bit
E_{elec}	50nJ/bit
ϵ_{fs}	10pJ·(bit·m ²) ⁻¹
初始温度 t_0	100
同一温度下的迭代次数	200
退温率 α	0.95
终止温度	10 ⁻⁸
新解未被更新最大次数	50
最优解未被更新最大次数	1 000
目标函数可接受值	—∞

表 1 中最后 4 项参数为 KMSA 算法的终止条件，满足 4 个中的任何一个条件，算法都会终止运行。最后一项设置为负无穷表示对目标函数的可接受值没有明确的要求，希望得到的结果能接近全局最优。

4.2 数据收集可靠性分析

图 2 为不同节点数下，簇划分个数为 5 时，网络内簇大小均方差的比较。图中 KMSA 的均方差远远小于 ILC 和 KMCD，说明 KMSA 的簇间负载均衡特性较好，这是因为 KMSA 方法中考虑了簇大小的限制因素，簇的划分没有超过簇首的容量限制，一定程度上保证了数据的可靠传输。而 ILC 和 KMCD 只考虑了数据收集能耗最小，没有考虑簇首的容量限制。图 2 中显示，随着节点数的增加，3 种方法簇大小均方差均呈现上升的趋势，说明网络中 SN 节点数越多，簇的划分越不均衡。因此，当 SN 节点数增加时，可通过相应增加簇首个数来进一步改善簇划分不均衡的问题。另外，对于 KMSA 方法的仿真结果还受算法迭代次数、初始温度以及退温率等参数的影响，需要经大量仿真来搜索上述参数的最佳值。

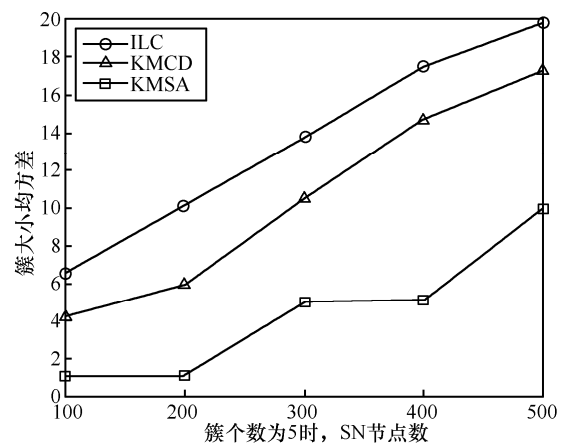


图 2 簇大小均方差

4.3 数据收集实时性分析

簇半径将会影响簇内数据收集的实时性。图 3 为当网络划分为 5 个簇时，不同网络规模下的最大簇半径的比较。图 3 表明 KMSA 方法得到的最大簇半径明显小于与之相比较的上述 2 种方法，说明了 KMSA 数据收集的实时性较好。当节点数增加时，KMSA 方法得到的最大簇半径远小于 KMCD 方法，这是因为 KMCD 方法把密度相似的节点都划分为一个簇。当局部节点密度相似的区域覆盖范围很大

时，由于初始聚类中心选择不合理，该区域内的节点很可能被划分为一个簇，导致簇半径很大。KMSA 和 ILC 方法的仿真曲线较接近，是由于 ILC 方法中簇首选择的目标函数是总能耗最小，而能耗和距离平方有关，因此 SN 节点总选择距离最近的节点做簇首。但由于网络中节点分布的随机性，局部节点密度差异较大，因此总能耗最小并不能保证各个簇最大簇半径最小。而 KMSA 算法把最小化最大簇半径作为一个优化目标，使得 KMSA 的仿真结果好于 ILC。

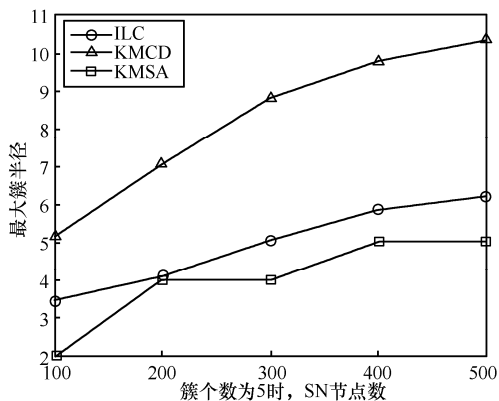


图 3 最大簇半径比较

4.4 数据收集能耗和网络生命期比较

由于 WSN 的网络生命期与网络能量消耗密切相关，因此对随机分布的 100 个 SN 节点构成的网络，比较了 3 种方法下的数据收集能耗和网络生命期。

数据收集能耗的比较如图 4 所示。节点通过 2 种途径消耗能量：发送自身产生报文和转发其他节点报文。如图 4 所示，网络中 SN 节点数一定的情况下，随着簇个数的增加，数据收集能耗呈现减小的趋势。这是由于随着簇个数的增加，节点到簇首的平均跳数减小，进行数据收集发生的总跳数也会减少，因此总能耗有所降低。图 4 中 KMSA 的能耗

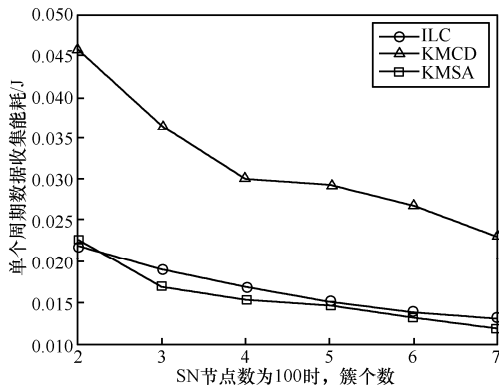


图 4 能耗情况比较

远远小于 KMCD 方法，和 ILC 的仿真结果接近，当簇个数多于 2 个时，能量消耗比 ILC 方法稍低。说明 KMSA 方法的节能性比以能量最小化为目标的 ILC 方法略好。因此这种考虑容量限制和簇半径限制的 KMSA 方法节能效果更好。

3 种方法得到的网络生命期比较如图 5 所示。ILC 和 KMCD 的生命期曲线为不考虑簇首容量限制情况下得到的。可以看出，不同网络规模的情况下 KMSA 的网络生存时间更长，当分簇个数较少时更明显。可见同时考虑簇半径和簇容量等约束条件的分簇网络，与 ILC 和 KMCD 相比，网络内 SN 节点间的负载趋于均衡，促使网络生命期得到了延长，由于仿真时迭代次数较少，KMSA 方法计算出的解可能与最优解差距较大，因此 KMSA 的曲线有波动性。可以通过提高初始温度，减小退温率或增加算法运行次数获取接近最优解的次优解，作为最终的簇首部署方案。

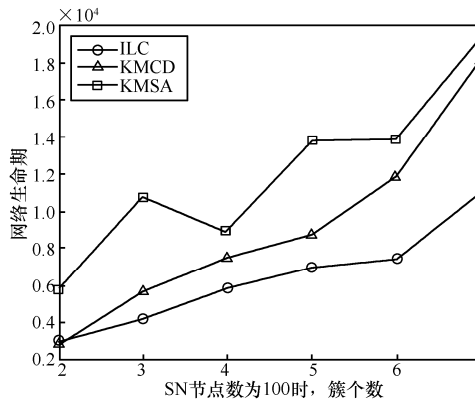


图 5 网络生命期比较

5 结束语

在 WSN 中，除普通的 SN 节点外，通过有策略地部署少量高能力节点作为簇首可以使网络性能大大提高。高能力簇首节点的部署问题是 NP-hard 问题，本文首先把该问题形式化为整数规划问题。针对现有簇首部署方法研究存在的问题，考虑了数据收集实时性和簇首间的负载均衡等因素，提出了一种利用 KM 和 SA 混合算法的簇首部署策略。该方法充分利用了 KM 算法简单、实时性好的特点，对簇首节点进行快速初选，然后利用 SA 算法的初始值顽健性来弥补 KM 算法初始值敏感性的不足，同时利用 KM 算法的局部最优特性来避免因 SA 算法的“突跳”特性错过最优解。通过仿真分析比较了簇大小标准差、簇内最大跳数和网络生

命期等 3 项指标, 验证了 KMSA 方法的有效性。目前的研究没有考虑无线干扰因素, 无线干扰会影响网络的吞吐量和可靠性, 下一步将以此为重点展开深入研究。

参考文献:

- [1] SZEWCZYK R, MAINWARING A, POLASTRE J, *et al.* An analysis of a large scale habitat monitoring application[A]. Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems[C]. New York, 2004. 214-226.
- [2] 俞黎阳, 王能, 张卫. 异构无线传感器网络中异构节点的部署与优化[J]. 计算机科学, 2008, 35(9): 48-51.
YU L Y, WANG N, ZHANG W. Placement of heterogenous nodes in wireless sensor networks[J]. Computer Science, 2008, 35(9): 48-51.
- [3] DESNOYERS P, GANESAN D, SHENOY P. TSAR: a two tier storage architecture using interval skip graphs[A]. Proceedings of the Third ACM Conference on Embedded Networked Sensor Systems[C]. San Diego, CA, 2005. 39-50.
- [4] YARVIS M, KUSHALNAGAR N, SINGH H, *et al.* Exploiting heterogeneity in sensor networks[A]. 24th Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM 2005)[C]. 2005. 878-890.
- [5] HEINZELMAN W R, CHANDRAKASAN A P, BALAKRISHNAN H. An application-specific protocol architecture for wireless microsensor networks[J]. IEEE Transactions on Wireless Communications, 2002, 1(4): 660-670.
- [6] HEINZELMAN W R. Application-specific Protocol Architectures for Wireless Networks[D]. Boston: Massachusetts Institute of Technology, 2000.
- [7] 尚凤军, ABOLHASAN M, WYSOCKI T. 无线传感器网络的分布式能量有效非均匀成簇算法[J]. 通信学报, 2009, 30(10): 34-43.
SHANG F J, ABOLHASAN M, WYSOCKI T. Distributed energy efficient unequal clustering algorithm for wireless sensor networks[J]. Journal on Communications, 2009, 30(10): 34-43.
- [8] HU W, CHOU C T, JHA S, *et al.* Deploying long-lived and cost-effective hybrid sensor networks[A]. First Workshop on Broadband Advanced Sensor Networks[C]. Santa Diego, USA, 2004.
- [9] VINCZE Z, VIDA R, VIDACS A. Deploying multiple sinks in multi-hop wireless sensor networks[A]. IEEE International Conference on Pervasive Services[C]. Istanbul, Turkey, 2007.55-63.
- [10] XU K, WANG Q, HASSANEIN H S, *et al.* Optimal wireless sensor networks(WSN) deployment: minimum cost with lifetime constraint[A]. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'05)[C]. Montreal, Canada, 2005.
- [11] OYMAN E I, ERSOY C. Multiple sink network design problem in large scale wireless sensor networks[A]. IEEE International Conference on Communications[C]. Paris, France, 2004. 3663- 3667.
- [12] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48-61.
SUN J G, LIU J, ZHAO L Y. Clustering algorithms[J]. Research Journal of Software, 2008, 19(1): 48-61.
- [13] 雷小锋, 谢昆青, 林帆等. 一种基于 K-Means 局部最优性的高效聚类算法[J]. 软件学报, 2008, 19(7): 1683-1692.
LEI X F, XIE K Q, LIN F, *et al.* An efficient clustering algorithm based on local optimality of K-Means[J]. Journal of Software, 2008, 19(7): 1683-1692.
- [14] 刘琳. 面向数据收集的无线传感器网络节点部署问题研究[D]. 沈阳: 中国科学院沈阳自动化研究所, 2009.
LIU L. Research on node deployment problem in data collection oriented wireless sensor networks. Shenyang: Shenyang Institute of Automation, Chinese Academy of Sciences, 2009.
- [15] 闫利军, 李宗斌, 卫军胡. 模拟退火算法的一种参数设定方法研究[J]. 系统仿真学报, 2008, 20(1):245-247.
YAN L J, LI Z B, WEI J H. Study on parameter setting method for simulated annealing algorithm[J]. Journal of System Simulation, 2008, 20(1): 245-247.
- [16] 王凌. 智能优化算法及其应用[M]. 北京: 清华大学出版社, 2001.
WANG L. Intelligent optimization algorithms with applications. Beijing: Tsinghua University Press, 2001.
- [17] LINDSEY S, RAGHAVENDRA C, SIVALINGAM K. Data gathering in sensor networks using the energy-delay metric[J]. IEEE Trans on Parallel and Distributive Systems, Special Issue on Mobile Computing, 2002, 13(9): 924-935.
- [18] 吴小兵, 陈贵海. 无线传感器网络中节点非均匀分布的能量空洞问题[J]. 计算机学报, 2008, 31(2):253-261.
WU X B, CHEN G H. The energy hole problem of nonuniform node distribution in wireless sensor networks[J]. Chinese Journal of Computers, 2008, 31(2): 253-261.

作者简介:



刘琳 (1973-), 女, 山东郓城人, 博士, 国家卫星海洋应用中心工程师, 主要研究方向为无线传感器网络管理、系统可靠性等。



于海斌 (1964-), 男, 黑龙江密山人, 博士, 中国科学院研究员、博士生导师, 主要研究方向为现场总线、工业无线通信、传感器网络及嵌入式系统。