

无线传感器网络网内数据处理节点的优化选取*

陈颖文⁺, 徐明, 吴一

(国防科学技术大学 计算机学院, 湖南 长沙 410073)

Optimization of the Proxy Node Selection for In-Network Data Processing in Wireless Sensor Networks

CHEN Ying-Wen⁺, XU Ming, WU Yi

(School of Computer, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: Phn: +86-731-4574616, E-mail: ywch_nudt@hotmail.com

Chen YW, Xu M, Wu Y. Optimization of the proxy node selection for in-network data processing in wireless sensor networks. *Journal of Software*, 2007,18(12):3104-3114. <http://www.jos.org.cn/1000-9825/18/3104.htm>

Abstract: Energy is identified as the most crucial resource of the wireless sensor networks. A significant amount of energy is often consumed by message passing. Many researches have focused on the issues about how to minimize the communication overhead. In-network data processing is a commonly adopted technique, in which an intermediate node called proxy node is chosen to process the source data streams and forward the result to the sink node, thereby reducing the transmission energy consumption. As a result, the optimal selection of the proxy node can minimize the transmission overhead. This paper formulizes the transmission energy consumption of the proxy node selection and proposes an energy efficient selection strategy (EESS), which can greatly reduce the transmission energy consumption of the query without knowing the whole topology of the network. Compared with the method proposed by others, EESS uses fewer control messages and achieves better performance. Simulation results show that EESS performs well even in low-density network and long-distance query, which potentially improves the lifetime of the sensor networks.

Key words: wireless; sensor network; in-network data processing; proxy node selection; energy efficient

摘要: 能量是无线传感器网络至关重要的资源,数据传输占据着能耗的主体,当前,大多数研究围绕最小化传输能耗而展开。网内数据处理是选择数据传输的某一中继节点作为处理节点,利用该节点所具备的计算能力对原始数据进行处理,再将处理结果返回给接收节点,从而达到降低传输能耗的目的。网内数据处理节点的最优选取,可以最小化数据查询的传输能耗。通过建立数学模型来描述传输能耗与处理节点选取策略的定量关系,提出一种不需要全局网络拓扑信息的低能耗的处理节点选取策略(energy efficient selection strategy,简称 EESS)。与现有方法相比,该策略使用较少的控制开销并能显著降低数据的传输能耗。模拟实验结果表明,EESS 在低密度的网络结构以及长距离的查询操作下具有良好的性能,更有利于延长无线传感器网络的寿命。

关键词: 无线;传感器网络;网内数据处理;处理节点选取;节能

* Supported by the Research Foundation for the Doctoral Program of Department of the Ministry of Education of China under Grant No.20059998022 (国家教育部高等学校博士学科点专项科研基金)

Received 2006-01-09; Accepted 2007-01-24

中图分类号: TP393 文献标识码: A

无线传感器网络是由一组在地理上广泛分布、相互之间能够利用无线信道进行通信的传感器节点组成的,每个节点都可以作为接收节点(sink node)与1个或者多个其他节点进行协作,从而测量环境参数、辨别系统状态、收集实时数据、提取增值信息等.这种特殊的网络系统覆盖范围广、组网灵活、即抛即用,能够广泛应用于军事、环境、健康、家庭、空间探索和灾难拯救等各种领域^[1].

一般来说,无线传感器网络中的传感器节点由电池供电,因此,能量成为无线传感器网络至关重要的资源.有研究表明,数据传输在无线传感器网络的能耗中占据主体地位,传输1比特数据所消耗的能量大约相当于执行1000条CPU指令^[2],因此,大量的研究工作围绕着最小化传输能耗开展.由于无线传感器网络节点自身具有一定的计算能力,在进行数据采集或者数据查询的过程中,一部分计算工作可以从接收节点转移到中继节点上,通过中继节点对数据的处理来减少传输的数据量,从而达到降低网络能耗的目的.我们可以把所有在传输中继节点上执行的数据操作均定义为网内数据处理(in-network data processing),执行数据操作的节点则称为处理节点(proxy node).实验证明,采用网内数据处理,与原有仅在接收节点进行所有数据操作的方式进行比较,能够有效地减少传输能耗并延长网络寿命^[3].

在进行网内数据处理时,处理节点位置的选取决定着数据传输的能耗^[4]:处理节点越靠近数据源,越有可能节省能耗;反之,处理节点越靠近接收节点,则节省的能耗越有限.本文主要研究处理节点的最优选取问题,即选择在哪个传输中继节点上进行网内数据处理,使得数据传输的总能耗最小.通过对传输能耗进行分析,可以建立数学模型来描述传输能耗与处理节点选取策略的定量关系.进一步分析可知,该模型的最小值在全局网络拓扑信息已知的情况下多项式时间可解.然而,要在分布式的无线传感器网络中维护全局的网络拓扑信息,代价太大.因此,本文提出了一种近似最优的处理节点选取方法,即不需要全局网络拓扑信息的低能耗的处理节点选取策略(energy efficient selection strategy,简称EESS).该策略首先通过一定程度的近似,建立描述传输能耗与处理节点位置坐标之间关系的定量模型,通过求解该模型,得出最优“虚拟处理节点”的位置坐标,然后将“虚拟处理节点”映射到网络中的真实节点并利用路由发现机制建立路由树,进而实现网内数据处理.

本文的主要创新点在于:

- 1) 建立了可以描述传输能耗与处理节点选取策略之间定量关系的数学模型;
- 2) 提出了一种不需要全局网络拓扑信息的低能耗的处理节点选取策略,与其他方法相比,该策略除了路由发现消息以外,不需要额外的消息开销;
- 3) 利用模拟实验对提出的方法进行验证,实验结果表明,EESS在低密度的网络结构以及长距离的查询操作下具有良好的性能,更有利于延长无线传感器网络的寿命.

1 相关研究

网内数据处理的概念源自于互连网络,Heidemann等人首次将其引入传感器网络,并描述了相应的软件框架^[5],在他们的框架中,提出了优化的数据聚合方法,并且指出,网内数据处理能够显著减少网络传输开销. Bhardwaj等人将网内数据处理划分为层次化与非层次化两大类^[6]:前者允许数据流在不同的网络节点进行聚合,如查询某属性的最大值或最小值;而后者仅在唯一的网络节点进行数据聚合操作,如多属性的联合操作.

针对层次化的网内数据处理,Yao等人提出了用于局部数据聚合的同步机制,并对AODV路由协议进行改进,增强了数据传输的可靠性^[7].此外,Intanagonwiwat等人指出,为层次化的网内数据处理构造最优聚合树等同于构造Steiner树问题,两者都是NP难问题.因此,他们提出了一种启发式的数据聚合方法,该方法逐步增加聚合树的分支,并最大限度地增大路径的复用程度,以减少传输的能耗,从而达到优化网内数据处理的目的^[8].

对于非层次化的网内数据处理而言,处理节点的选取是节省网络传输能耗的关键,Bonfils等人讨论了处理节点的选取对传输能耗的影响,并且提出了一种自适应的分布式节点选取策略(adaptive and decentralized operator placement strategy,简称ADOPS)^[4].ADOPS方法包含邻居探测策略,因此可以在最小化传输能耗的方向

上,持续地更新处理节点的位置.然而,ADOPS 方法在网络节点密度较小时性能下降得很快,这是因为,当传感器节点仅含有较少的邻居时,邻居探测过程容易陷入远离全局最优值的局部最优值.除此之外,ADOPS 为了更新处理节点,候选节点需要将能耗的估计值发送给活跃节点进行比较,这些控制消息的交换需要消耗一定的能量.本文针对 ADOPS 的不足,提出了适用于非层次化网内数据处理的 EESS 方法.

2 系统模型与问题描述

本节主要介绍无线传感器网络模型以及非层次化网内数据处理模型,通过选择减小传输能耗作为优化目标,给出处理节点优化选取的数学模型.为了便于建模,首先给出如下假设:

- 1) 无线传感器网络各节点均匀分布于 X - Y 平面,各传感器节点可以通过 GPS 或者其他方法获取自身的地理位置信息,并保持静止;
- 2) 所有的传感器节点具有相同的通信半径,传输任意单位比特数据的平均能耗均相同,存在于彼此通信半径之内的两节点互为邻居节点;
- 3) 所有的查询均持续一定的时间,查询结果由非层次化网内数据处理方式获得,即查询结果由网络中唯一的处理节点处理后得到,数据源节点(source node)的位置及相应数据流产生的比特速率可以通过 Directed Diffusion^[9]等方法得到;
- 4) 每个传感器节点均具有一定的计算能力,可以在必要的时候执行数据处理操作.

无线传感器网络可以用无向图 $G=(V,E)$ 来表示,其中,每个顶点 $v \in V$ 代表一个传感器节点,每条边 $e \in E$ 代表两节点之间存在通信链路.对于处理节点而言,输入量是来自于不同数据源节点的原始数据流,输出量是经过处理后的结果数据流,因此,数据处理的实质是一种函数变换.不同的应用可以抽象出不同的数据处理函数,如何根据应用抽象处理函数不在本文的讨论范围之内,本文假设处理函数已知,并可以在任意的传感器节点上执行.

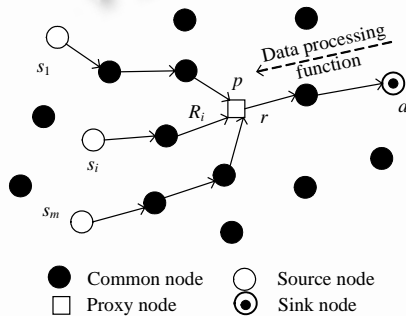


Fig.1 Non-Hierarchical in-network processing

图1 非层次化的网内数据处理

非层次化的网内数据处理如图 1 所示,接收节点 $a \in V$ 发出生命周期为 T_Q 的查询 Q ,即该查询在 T_Q 时间内持续地获取数据源节点的数据,假设数据源节点有 m 个,分别为 s_i ,各数据源节点产生的数据流平均比特速率为 R_i bps ($i=1, \dots, m$).为了执行网内数据处理,接收节点 a 将查询 Q 的处理函数 f_Q 发送到适当的处理节点 p 上,节点 p 则利用处理函数 f_Q 将来自 m 个数据源节点的原始数据流转换成平均比特速率为 r bps 的结果数据流.一般而言,转换函数 f_Q 可以减少传输的数据量,即满足 $r < \sum_{i=1}^m R_i$. 处理节点 p 执行完网内数据处理操作之后,将结果数据流发送给接收节点 a .

为了方便描述与建模,现引入如下记号. $h(s,d)$ 表示从节点 s 到节点 d 所经过的跳步(hop)数; ε 表示传输单位比特数据所需要的能耗; R_L 表示数据流 L 的平均比特速率; T_L 表示数据流 L 的持续时间;那么,从节点 s 到节点 d 传输数据流所需要的总能耗 $C(L)$ 可以用下列式子进行计算:

$$C(L) = \varepsilon \cdot R_L \cdot T_L \cdot h(s,d) \quad (1)$$

如图 1 所示,数据由各源节点 s_i 传输到接收节点 a 的路径可以看作树状结构,其中,接收节点 a 是树的根节点,源节点 s_i 是树的叶子节点.由于在无线传感器网络中,CPU 的能耗远远小于数据传输的能耗^[2],因此,查询 Q 的总能耗可以仅用传输能耗来度量,设处理函数 f_Q 的大小为 τ_Q bit,那么,查询 Q 选定节点 p 作为处理节点的总能耗可以用下式来计算:

$$C_Q(p) = \varepsilon \cdot \tau_Q \cdot h(a,p) + \sum_{i=1}^m \varepsilon \cdot R_i \cdot T_Q \cdot h(s_i,p) + \varepsilon \cdot r \cdot T_Q \cdot h(p,a) \quad (2)$$

记 $\hat{h}(s,d)$ 为节点 s 与节点 d 之间的最少跳步数,那么,选定节点 p 作为处理节点的最小查询能耗可以表示为

$$C_{\min}(p) = \varepsilon \cdot \tau_Q \cdot \hat{h}(a, p) + \sum_{i=1}^m \varepsilon \cdot R_i \cdot T_Q \cdot \hat{h}(s_i, p) + \varepsilon \cdot r \cdot T_Q \cdot \hat{h}(p, a) \quad (3)$$

基于以上结论,关于处理节点的最优选取问题,可以用式(4)来描述.即在网络中找到节点 p^* ,由该节点执行网内数据处理的操作,从而最小化查询的总能耗.

$$p^* = \arg_{p \in V} \min C_{\min}(p) \quad (4)$$

由式(4)可得,如果网络的全局拓扑结构已知,该优化问题可以分两步进行求解:首先,利用 Dijkstra 或 Floyd 算法求解任意两节点之间的最少跳步数,算法复杂度为 $O(|V|^3)$;接下来,对于所有的节点 $v \in V$ 搜索解空间,找出最小能耗所对应的处理节点选取方案,算法复杂度为 $O(|V|)$.因此,求解式(4)总的复杂度为 $O(|V|^3)$,这意味着优化问题(4)能够在多项式时间内求解.然而,该方法需要在每个节点动态维护网络的全局拓扑结构,开销过大,不适用于无线传感器网络.本文提出的 EESS 方法则仅需在每个节点上维护邻居节点的信息,并具有良好的节能特性.

3 EESS 方法

EESS 方法的基本思路是,放宽最优化问题(4)的约束条件,建立描述传输能耗与处理节点位置坐标之间关系的定量模型,通过求解近似模型,得出最优“虚拟处理节点”的位置坐标,然后利用节点映射机制,将“虚拟处理节点”映射到网络中的真实节点,并通过路由发现机制建立查询路由树,实现网内数据处理.下面主要对 EESS 方法的近似数值解、节点映射机制、路由发现机制以及自适应策略等进行详细介绍.

3.1 近似数值解

在不知道网络全局拓扑信息的情况下,获取任意两节点之间准确的最少跳步数是十分困难的.文献[10]指出,在均匀分布的无线传感器网络中,单位比特数据在节点之间的多跳传输能耗近似地正比于节点之间的欧氏距离,文献[11]在引理 1 中也证明了相同的结论.基于此,在 X - Y 平面上,记 $d(n_1, n_2)$ 表示节点 n_1 和 n_2 之间的欧氏距离, z 表示处理节点的位置坐标.如果处理节点可以存在于平面的任意位置,设传输能耗与欧氏距离的比例因子为 $1^{[10]}$,那么,查询 Q 的能耗与处理节点坐标之间的关系可以近似用下式表示:

$$C(z) = \tau_Q \cdot d(a, z) + \sum_{i=1}^m R_i \cdot T_Q \cdot d(s_i, z) + r \cdot T_Q \cdot d(z, a) \quad (5)$$

已知查询的总能耗函数 $C(z)$ 可由式(5)进行计算,由欧氏距离及不同欧氏距离的线性加权和性质可知,能耗函数 $C(z)$ 是下凸函数,存在最小值.进一步地,式(4)的最优化问题可以近似转化为式(6)的无约束优化问题,即在 X - Y 平面中寻找最优的节点 z^* ,使得能耗函数 $C(z)$ 取值最小,其中, z^* 对应于“虚拟处理节点”的坐标,是真实节点的近似解.

$$z^* = \arg_{z \in R^2} \min C(z) \quad (6)$$

求解形如式(6)的无约束最优化问题的常用方法是迭代法:首先给出函数 f 最小值点 x^* 的初始迭代点 $x^{(0)}$,然后构造迭代点的序列 $\{x^{(k+1)} = x^{(k)} + \alpha_k \cdot d_k\}$,使得在 f 和 $x^{(0)}$ 满足某些条件的情况下,迭代点序列收敛于 x^* .迭代法是一种数值求解方法,一般而言,求解精度越高,则计算复杂度越大.如何构造高精度低计算复杂度的迭代算法并不是本文研究的目标,本文仅讨论如何利用现有的迭代算法,在可接受的复杂度下,求解足够精度的解.

最速下降法(steepest descent algorithm,简称 SDA)是求解无约束最优化问题最常用的方法之一,算法描述如图 2 所示.SDA 使用负梯度向量作为最小化方向 d_k ,并且采用线性搜索算法计算步长参数 α_k .可用来搜索最佳步长 α_k 的线性搜索算法有很多,如二分法和黄金分割法,而黄金分割法的收敛比大于二分法,具有较快的收敛速度,故本文采用黄金分割法搜索 α_k ,算法描述如图 3 所示.

直接应用 SDA 求解式(6)很难证明其收敛性,因此不能保证算法在有限步内终止.然而数值实验表明,SDA 目标函数值在迭代点远离最小值点的时候下降得很快,而当迭代点逼近最小值点的时候,目标函数值的变化不明显^[12].因此,可以将 SDA 的迭代终止条件由 $\|\nabla C(z_k)\| \leq \varepsilon$ 放宽为 $|C(z_{k+1}) - C(z_k)| \leq \varepsilon$,即当目标函数值变化不明显时终止迭代过程.不难验证,放宽 SDA 的迭代终止条件后,算法是收敛的,即算法能够在有限步内终止.虽然放宽 SDA 迭代终止条件求解式(6)得出的数值解只是真实节点的近似坐标,但该近似解能够满足 EESS 方法的精度

要求,第4节的模拟实验结果说明了近似结果的有效性.

SteepestDescent(z_0, ε):
 Input: Initial estimate z_0 , error bound $\varepsilon(0 \leq \varepsilon \leq 1)$;
 Output: Optimal solution z^* .
 1. $k \leftarrow 0$;
 2. $d_k \leftarrow -\nabla C(z_k)$;
 3. **if** $\|\nabla C(z_k)\| \leq \varepsilon$, **then return** z_k ;
 4. $\alpha_k \leftarrow \arg \min_{\alpha > 0} C(z_k + \alpha d_k)$;
 //compute α_k using **GoldSection** on $C()$;
 5. $z_{k+1} \leftarrow z_k + \alpha_k d_k$;
 6. $k \leftarrow k+1$, **goto** step 2;

Fig.2 Steepest descent algorithm

图2 最速下降算法

GoldSection(φ, I_0, δ):
 Input: Initial interval $I_0=[a_0, b_0]$, error bound $\delta > 0$;
 Output: α minimizing φ ; // $\varphi(\alpha) = C(z_k + \alpha d_k)$.
 1. $k \leftarrow 0$;
 2. $\lambda_k \leftarrow a_k + 0.382(b_k - a_k)$; $\mu_k \leftarrow a_k + 0.618(b_k - a_k)$;
 3. **if** $\varphi(\lambda_k) > \varphi(\mu_k)$, **then goto** step 4 **else goto** step 5;
 4. **if** $b_k - \lambda_k \leq \delta$, **then return** μ_k
 else $a_{k+1} \leftarrow \lambda_k$, $b_{k+1} \leftarrow b_k$, $\lambda_{k+1} \leftarrow \mu_k$, $\mu_{k+1} \leftarrow a_{k+1} + 0.618(b_{k+1} - a_{k+1})$,
 $k \leftarrow k+1$; **goto** step 3;
 5. **if** $\mu_k - a_k \leq \delta$, **then return** λ_k
 else $a_{k+1} \leftarrow a_k$, $b_{k+1} \leftarrow \mu_k$, $\mu_{k+1} \leftarrow \lambda_k$, $\lambda_{k+1} \leftarrow a_{k+1} + 0.382(b_{k+1} - a_{k+1})$,
 $k \leftarrow k+1$; **goto** step 3;

Fig.3 Gold section algorithm

图3 黄金分割搜索算法

3.2 节点映射

根据假设3),接收节点已知所有数据源节点的位置和数据产生的比特速率,接收节点可以通过执行SDA算法求出最优“虚拟处理节点” p' 的位置坐标.由于 p' 在实际的网络中并不存在,可以将网络中距离 p' 点最近的真实节点选定为近似最优的处理节点 p .下面介绍如何在无线传感器网络中寻找处理节点 p .

接收节点 a 采用GPSR(greedy perimeter stateless routing)路由协议^[13],以 p' 的坐标为目的地,发送一个探测报文.根据GPSR路由协议,如果 p' 的坐标与某一真实节点的坐标相同,那么探测报文最终将到达该节点,该节点即所要寻找的近似最优处理节点 p ;如果 p' 的坐标与网络中所有真实节点的坐标均不相同,那么有两种情况发生: p' 要么位于某一内表面(interior face)的内部,要么位于某一外表面(exterior face)的外部^[14].如图4所示, z_1 位于内表面 $b-c-d-e-f-i-b$ 的内部, z_2 位于外表面 $a-b-c-d-e-f-g-h-a$ 的外部.无论发生哪种情况,GPSR协议最终能够将探测报文发送到相应的某一表面上,并且遍历表面上所有节点一次.通过在探测报文中添加报文所经过的节点信息,转发节点可以判断该报文是否已经完整地遍历了某表面一次.显然,距离 p' 点最近的真实节点必然是该表面上的节点,由于探测报文中保存了表面节点的信息,中继节点最终可以将探测报文发送到距离 p' 点最近的节点上,从而该点被选为处理节点 p .如果距离 p' 点最近的节点存在多个,那么探测报文最先经过的节点被选为处理节点 p .假设图4中的节点 a 是接收节点,求出的最优“虚拟处理节点” p' 的坐标位于 z_1 点,采用GPSR路由协议,探测报文的路由过程为 $a \rightarrow b \rightarrow i \rightarrow f \rightarrow e \rightarrow d \rightarrow c \rightarrow b$.遍历完表面节点 b, i, f, e, d, c 之后,表面节点的坐标均保存在探测报文中.当探测报文第2次到达节点 b 的时候,节点 b 可以根据已有信息判断出距离 z_1

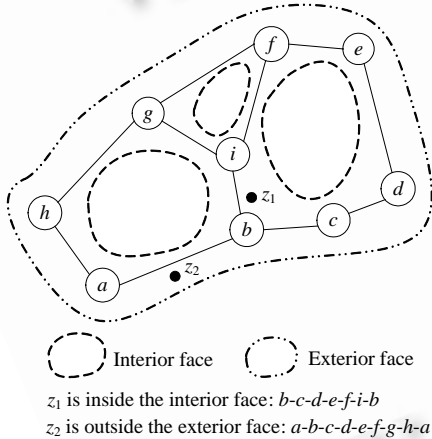


Fig.4 Interior face and exterior face

图4 内表面与外表面示意图

点最近的节点,并将探测报文发送给该节点.本例中,距离 z_1 最近的节点有两个:节点 b 和节点 i ,由于节点 b 出现在节点 i 之前,故节点 b 被选为处理节点 p .处理节点 p 被选定之后,沿反向路径回复一个确认报文给接收节点 a ,接收节点再将查询请求 Q 以及数据处理函数 f_Q 发送给处理节点 p ,从而完成节点映射过程.

3.3 路由发现

完成节点映射之后,网络中的某一真实节点被选定为处理节点 p ,要执行网内数据处理还需要建立以接收节点 a 为根节点、以处理节点 p 为中间节点、以数据源节点 s_i 为叶节点的路由树(如图1所示).由于查询请求

Q 可以包含各数据源节点和接收节点的坐标,处理节点 p 可采用 GPSR 路由协议分别建立到各数据源节点和接收节点的路由.然而,采用 GPSR 路由协议得到的路径并不具有最少跳步数,这不利于较长时间的查询.为了建立最小传输能耗的路由树,可采用 Directed Diffusion 协议^[9]寻找处理节点到各数据源节点以及处理节点到接收节点的最短路径.

处理节点 p 发起路由请求过程,向邻居节点广播一个路由请求报文,请求建立到接收节点 a 和各数据源节点 s_i 的路由.当网络中的任意节点 n 第 1 次接收到该路由请求报文时,按照下面的规则决定是否转发该路由请求报文:如果该报文来自于节点 n 的唯一邻居,则不转发报文;否则,向其他邻居广播该报文.如果节点 n 已经接收过该路由请求报文,则不再转发该报文.按照上述原则,网络中的任意节点最多转发路由请求报文 1 次.当路由请求报文到达接收节点 a 时,节点 a 将沿最短的反向路径将最短路径发送给处理节点;类似地,当路由请求报文到达数据源节点 s_i 时,节点 s_i 在沿反向路径发送最短路径的同时将原始数据送出.假设路由请求报文在任意相邻节点之间传输的时间相同,那么分布式的路由发现过程与广度优先搜索相似,路由发现所得到的最短路径具有全局最少的跳步数.

3.4 自适应策略

对于较长时间的查询而言,数据产生的平均比特速率可能发生较大的变化,继续采用当前的处理节点进行网内数据处理,可能并不能达到节省能耗的目的.因此,处理节点的选取应当具有一定的自适应能力,能够根据原始数据产生的比特速率的变化,自动调整所选择的处理节点.

在执行网内数据处理的同时,处理节点 p 周期性地监控输入/输出数据流的比特速率,如果原始数据产生的比特速率发生显著的变化,则利用 SDA 算法重新计算“虚拟处理节点” z^* 的坐标,如果 z^* 到当前处理节点 p 的距离大于一定的阈值,则进行处理节点的切换,即在网络中寻找新的处理节点 p^* ,并将处理函数迁移到新的处理节点 p^* 上.

处理节点的切换过程分为两个阶段,与两阶段提交协议^[15]相类似.首先,当前处理节点 p 以 z^* 坐标为目的发送一个“prepare-to-transfer”的消息报文,该消息报文包含接收节点 a 、各数据源节点 s_i 以及处理函数 f_Q 等相关信息,利用第 3.2 节的节点映射机制,该消息报文最终能够到达新处理节点 p^* .节点 p^* 接收到该消息报文之后,利用第 3.3 节的路由发现机制建立路由树,各源节点 s_i 记录下到节点 p^* 的新路径,但暂时并不将原始数据发往节点 p^* .当节点 p^* 完成准备工作之后,发送一个 ok 报文给节点 p ,并等待原始数据流的输入.节点 p 接收到该 ok 报文后,向各数据源节点 s_i 发送 switch 报文,并继续执行网内数据处理直到不再有数据流输入.各数据源节点 s_i 接收到 switch 报文后,则利用新的路由将原始数据发送给新处理节点 p^* .新处理节点 p^* 接收到原始数据流,则开始执行网内数据处理,并将处理结果发送给接收节点 a .至此,整个处理节点的切换过程结束.在上述过程中,如果找到的新处理节点 p^* 因为某些原因(如能量不够)不能完成网内数据处理的准备工作,那么,节点 p^* 发送一个 no 报文给节点 p .这种情况下,节点 p 将在节点映射过程中的表面节点上重新寻找另一个节点作为新的处理节点 p^* ,继续处理节点的切换过程直到成功.

尽管处理节点的选取具有自适应能力,可以降低能耗,但是,处理节点的切换需要消耗一定的传输能量,并且会引入一定的延迟.因此,过于频繁地处理节点切换会降低整个系统的性能.如何优化处理节点选取的自适应行为,是下一步研究工作的主要内容.

4 性能分析

我们利用模拟实验来验证 EESS 方法的有效性,通过实验,比较 EESS 方法的查询能耗与现有 ADOPS 方法的查询能耗,讨论网络参数与查询参数对网内数据处理能耗的影响,分析处理节点选取策略的自适应行为对能耗的影响.在实验中,所有的传感器节点均匀分布于某一矩形区域 Ω ,传感器节点的通信半径均为 ρ ,其他网络参数、查询参数以及相应的缺省值见表 1.

Table 1 Parameters of query and sensor network**表 1** 传感器网络及查询的参数

Parameter	Symbol	Default value
Distributed range of sensor network	Ω	$300 \times 300 \text{m}^2$
Number of sensor nodes	N	250~450
Transmission range	ρ	25m
Number of data sources	m	4
Bit-size of transformation function	τ_Q	100B
Lifetime of query	T_Q	100s
Source data rate	R_i	10bps
Consolidated data rate	r	10bps
Energy consumption per bit	ε	1
Query distance	D	3~10hops

为使模拟结果更具有一般性,我们模拟各种不同的查询请求.如果各数据源节点产生数据的比特速率均相同,即 $\forall i,j \in \{1, \dots, m\}, R_i = R_j$,则称该种查询为均匀(uniform)查询.实验中,取 $R_1 = R_2 = R_3 = R_4 = 10$;如果各数据源节点产生数据的比特速率是算术级增长的序列,即 $\forall i \in \{2, \dots, m\}, R_i = R_{i-1} + c$,其中, c 是不等于0的常数,则称该种查询为算术(arithmetic)查询.实验中,取 $R_1 = 4, R_2 = 8, R_3 = 12, R_4 = 16$;如果各数据源节点产生数据的比特速率是几何级增长的序列,即 $\forall i \in \{2, \dots, m\}, R_i = k \cdot R_{i-1}$,其中, k 是大于1的常数,则称该种查询为几何(geometric)查询.实验中,取 $R_1 = 8/3, R_2 = 16/3, R_3 = 32/3, R_4 = 64/3$.为了公平比较,各组实验保持相同的数据减少率,即 $r / \sum_{i=1}^m R_i$ 保持不变.除了考虑原始数据产生的不同比特速率以外,我们还考虑数据源节点的不同分布情况.如果所有的数据源节点都分布在远小于整个网络覆盖范围的某一区域 δ 内,则称该查询为区域受限的(constrained)查询.实验中, $\delta = 50 \times 50 \text{m}^2$,仅为整个网络覆盖范围 Ω 的3%;如果各数据源节点可以任意地分布在整个网络覆盖范围 Ω 中,则称该查询为区域不受限的(unconstrained)查询.实验中,为了排除边界效应的影响,对于区域不受限的查询,我们仅选择接收节点和数据源节点均位于 Ω 中央 $250 \times 250 \text{m}^2$ 区域内的查询.

实验模拟了4种不同的处理节点选取策略,分别记为 BASELINE, OPTIMUM, EESS 以及 ADOPS,各种策略下的查询能耗可以由第2节的式(3)进行计算.在 BASELINE 策略中,接收节点就是处理节点,这相当于不执行网内数据处理的情况;在 OPTIMUM 策略中,处理节点利用全局最优的方法选出,事实上,全局最优的方法在实际的无线传感器网络中难以实现,OPTIMUM 策略下的查询能耗仅用于比较.处理节点选取策略的性能用能耗的平均值 \bar{C} 来衡量,能耗的平均值越小,说明性能越好.假设随机生成 u 个连通的网络拓扑结构,每个拓扑结构下随机的进行 v 次查询,第 i 个拓扑结构下的第 j 个查询能耗记作 C_{ij} ,则有 $\bar{C} = \frac{1}{uv} \sum_{i=1}^u \sum_{j=1}^v C_{ij}$.实验中, $u=30, v=100$.

4.1 查询距离的影响

第1组实验主要研究不同查询距离的情况下,采用各种处理节点选取策略所需的查询能耗.查询距离主要用于反映数据源节点相对接收节点的远近程度,其大小可以用接收节点到数据源节点的平均跳步数来计算.在该组实验中,我们指定传感器节点的总数 $N=300$,模拟结果如图5所示.

从图5不难看出,采用网内数据处理,能够显著地减少查询的传输能耗.EESS方法仅在短距离查询情况下性能稍逊于ADOPS,当查询距离大于6跳时,性能均优于ADOPS.这是因为当接收节点和数据源节点距离较近时,ADOPS的邻居探测过程能够充分利用网络的局部拓扑信息,从而能够选取到接近全局最优的处理节点;而EESS依靠近似计算和节点映射来选取处理节点,无法利用网络的局部拓扑信息,容易产生偏差.随着查询距离的增大,过分依赖于网络局部拓扑信息的ADOPS方法容易陷入远离全局最优的局部最优解;而不依赖于网络局部拓扑信息的EESS方法,则能够保证近似最优解与全局最优解相差不会太大,从而优于ADOPS.例如,对于不同种类的查询,当查询距离增加到10跳时,EESS比ADOPS节能18%~30%.

由图5还可以看出,数据源节点的分布对网内数据处理性能的影响较大,而数据产生的比特速率对网内数据处理性能的影响较小.例如,图5(a)~图5(c)所示的区域受限的查询,尽管它们各自的数据产生速率不相同,但仍具有相似的曲线规律;而图5(a)、图5(d)所示的查询尽管具有相同的数据产生速率,但数据源节点的分布不同,

从而具有不同的曲线规律.实验结果表明,相同情况下,利用网内数据处理执行区域受限的查询比执行区域不受限的查询能耗少.这是因为当各数据源节点受限于一区域时,容易找到一个处理节点,使之靠近所有的数据源节点;而当数据源节点散布于整个网络覆盖区域时,很难找到一个处理节点,使之靠近所有的数据源节点.

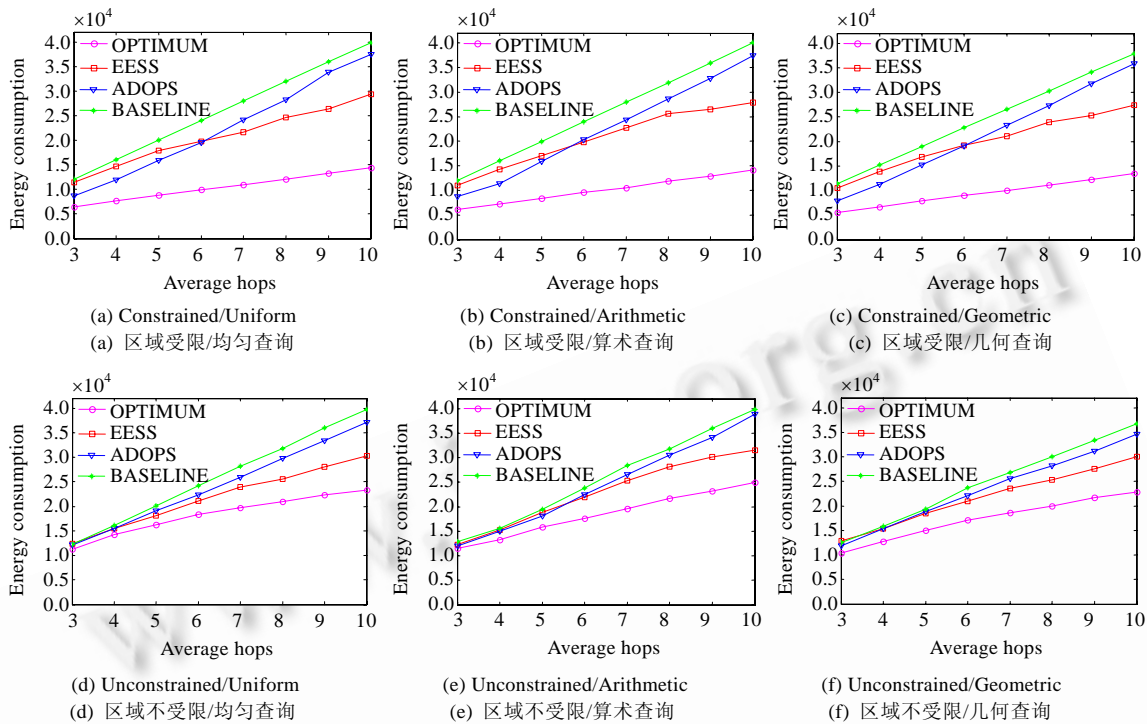


Fig.5 Energy consumption versus query distance

图5 查询距离对能耗的影响

4.2 节点密度的影响

由于网络的拓扑结构与网络节点的密度相关,本组实验主要研究不同节点密度情况下,各种处理节点选取策略所需的查询能耗.实验中,查询距离=8,节点密度的变化通过改变节点的总数量 N 来实现,实验结果如图 6 所示.

从图 6 可以看出,采用网内数据处理可以很明显地减少网络传输能耗,并且,EESS 方法的性能优于 ADOPS 的性能.虽然根据能耗曲线的变化趋势可以看出,随着网络节点密度的增加,ADOPS 性能的增长速度比 EESS、OPTIMUM 要快(这是因为对 ADOPS 而言,节点密度的增加除了可以减小处理节点到源节点和接收节点之间的跳步数以外,还可以使邻居探测过程朝着更接近全局最优值的方向推进),但是在实际应用中,无线传感器网络的密度越小,成本越低廉,我们期望用最少的网络节点实现最低的查询能耗.而 ADOPS 要达到与 EESS 同样的性能,需要增加一定数目的网络节点,如图 6(a)所示,当网络节点数 $N=280$ 时,ADOPS 要达到 EESS 同样的性能,需要将节点数增加约 50%.因此,EESS 比 ADOPS 更适用于无线传感器网络.

4.3 自适应分析

如前所述,当数据源节点数据产生的比特速率发生变化时,处理节点选取策略的自适应可以降低传输能耗,本组实验主要研究处理节点选取策略的自适应对能耗的影响.实验中的查询请求为几何查询,在查询的前 1/4 生命周期之内,各数据源节点数据产生的比特速率分别为 $R_1=8/3, R_2=16/3, R_3=32/3, R_4=64/3$,在此后的每四分之一生命周期之内,各数据源节点数据产生的比特速率按照循环右移一位的规律发生变化,即 $R'_1=64/3$,

$R_2' = 8/3, R_3' = 16/3, R_4' = 32/3; R_1'' = 32/3, R_2'' = 64/3, R_3'' = 8/3, R_4'' = 16/3; R_1''' = 16/3, R_2''' = 32/3, R_3''' = 64/3, R_4''' = 8/3$. 通过改变查询距离及节点密度,各处理节点选取策略的能耗与自适应能耗如图 7 所示.

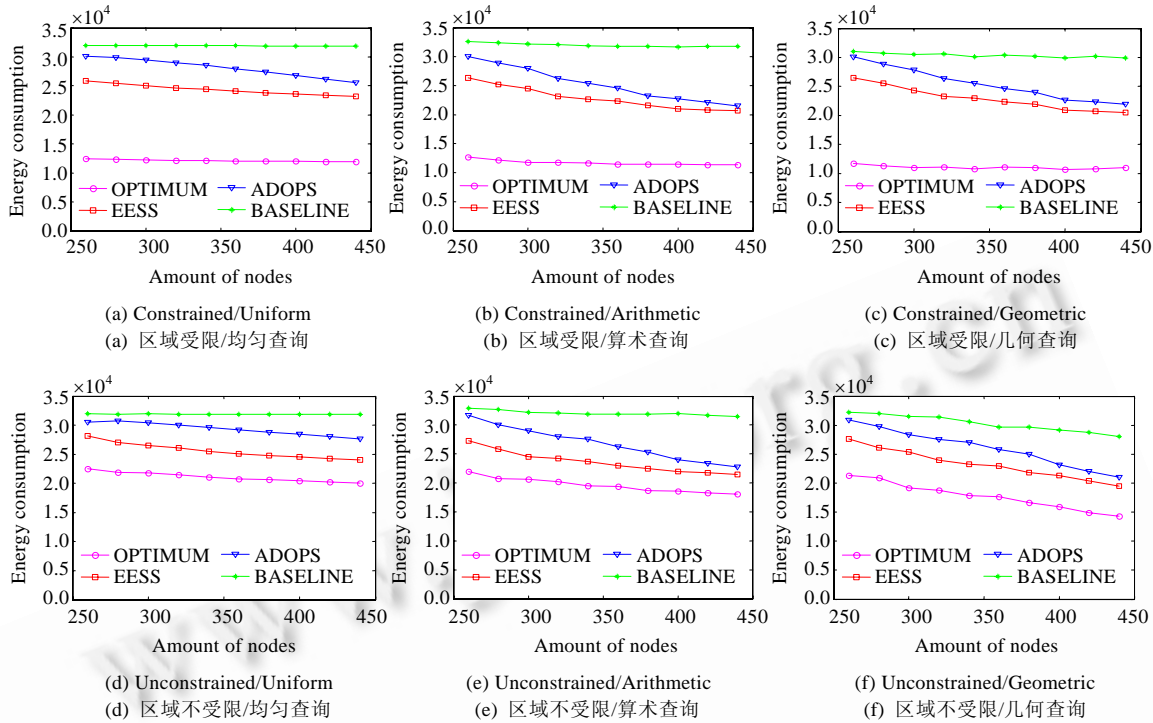


Fig.6 Energy consumption versus node density

图 6 结点密度对能耗的影响

与前两组实验的结果相似,在低密度网络与长距离查询情况下,EESS(adapt)的性能均优于 ADOPS (adapt). 此外,由图 7(a)、图 7(c)可知,对于区域受限的查询,EESS 的能耗曲线与 EESS 的自适应能耗曲线 EESS (adapt) 几乎相同,这是因为如果数据源节点受限于一区域,即使数据产生的比特速率发生变化,根据式(5)所计算的新处理节点与原处理节点相距不远,两者的能耗也相差不大,这说明 EESS 的自适应策略不适用区域受限的查询. 然而由图 7(b)、图 7(d)可知,对于区域不受限的查询,EESS(adapt)所示的能耗比 EESS 所示的能耗有一定程度的减小,这说明 EESS 的自适应策略适用于区域不受限的查询.如何改进 EESS 的自适应策略使之适用于各种不同的查询,是我们下一步研究的内容.

5 结论

能耗是影响无线传感器网络性能的关键因素之一,网内数据处理的应用可以显著地减少数据查询所需的能耗.本文提出的 EESS 方法可以优化处理节点的选取.与其他现有方法相比,EESS 更适用于低密度的网络以及远距离的查询,从而能够更有效地延长网络寿命.进一步的研究工作将考虑网络的动态拓扑结构、节点的可移动性以及自适应与容错等特性.

致谢 审稿人对原稿提出了中肯的意见,在此向他们表示感谢.此外,感谢国防科学技术大学电子科学与工程学院的博士生张雄明、张增辉给予本文的建议和帮助.

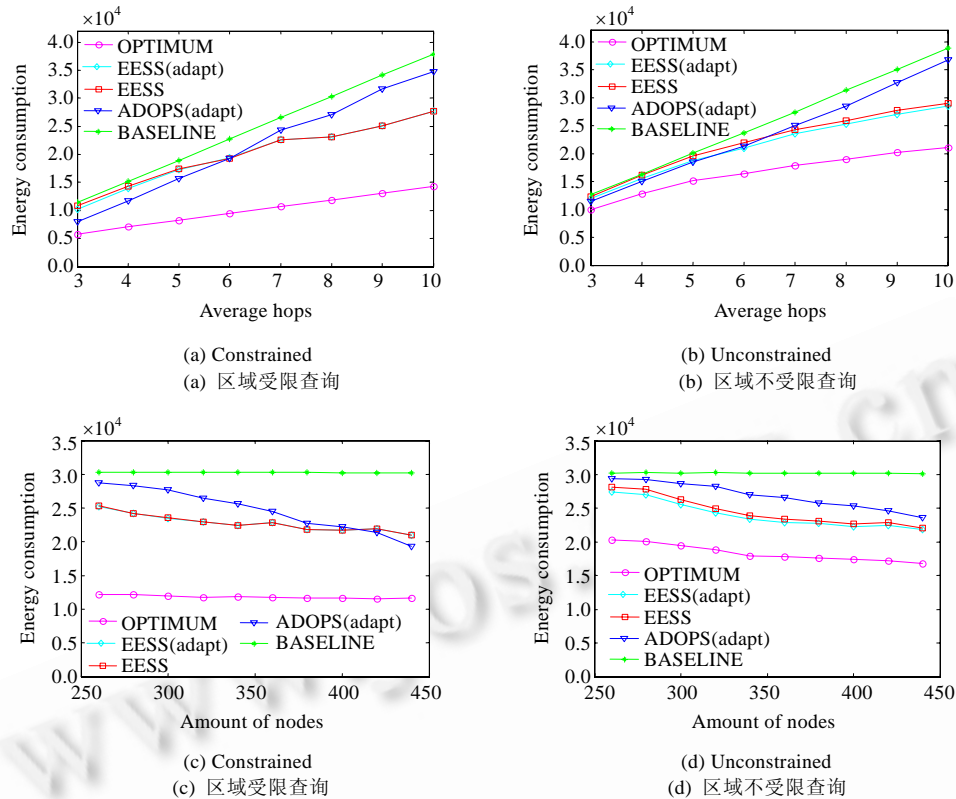


Fig.7 Energy consumption versus adaptation

图 7 自适应对能耗的影响

References:

- [1] Ren F, Huang H, Lin C. Wireless sensor networks. Journal of Software, 2003,14(7):1282-1291 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1282.htm>
- [2] Akyildiz I, Su W, Sankarasubramanian Y, Cayirci E. A survey on sensor networks. IEEE Communications Magazine, 2002,40(8): 102-114.
- [3] Yao Y, Gehrke J. The cougar approach to in-network query processing in sensor networks. ACM SIGMOD Record, 2002,31(3): 9-18.
- [4] Bonfils B, Bonnet P. Adaptive and decentralized operator placement for in-network query processing. Telecommunication Systems, 2004,26(2-4):389-409.
- [5] Heidemann J, Silva F, Intanagonwiwat C, Govindan R, Estrin D, Ganesan D. Building efficient wireless sensor networks with low-level naming. In: Marzullo K, ed. Proc. of the 18th ACM Symp. on Operating System Principles. New York: ACM, 2001. 146-159.
- [6] Bhardwaj M, Chandrakasan A. Bounding the lifetime of sensor networks via optimal role assignments. In: Proc. of the IEEE INFOCOM 2002. New York: IEEE, 2002. 1587-1596.
- [7] Yao Y, Gehrke J. Query processing in sensor networks. In: Proc. of the 1st Biennial Conf. on Innovative Data Systems Research (CIDR). 2003. <http://citeseer.ist.psu.edu/yao03query.html>
- [8] Intanagonwiwat C, Estrin D, Govindan R, Heidemann J. Impact of network density on data aggregation in wireless sensor networks. Technical Report, 01-750, University of Southern California, 2001.

- [9] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In: Proc. of the MobiCom 2000. Boston: ACM, 2000. 56–67.
- [10] Baek S, Veciana G, Su X. Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation. IEEE Journal on Selected Areas in Communications, 2004,22(6):1130–1140.
- [11] Guo L, Li J, Li G. Spatio-Temporal query processing method in wireless sensor networks. Journal of Software, 2006,17(4):794–805 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/794.htm>
- [12] Yuan Y, Sun W. Optimization Theory and Method. Beijing: Science Press, 1997. 108–110 (in Chinese).
- [13] Karp B, Kung HT. GPSR: Greedy perimeter stateless routing for wireless networks. In: Proc. of the MobiCom 2000. Boston: ACM, 2000. 243–254.
- [14] Huang Q, Lu C, Roman GC. Reliable mobicast via face-aware routing. In: Proc. of the IEEE INFOCOM 2004. Hong Kong: IEEE, 2004. 2108–2118.
- [15] Tanenbaum AS. Distributed Operating Systems. Prentice Hall, 1995. 153–154.

附中文参考文献:

- [1] 任丰原,黄海宁,林闯.无线传感器网络.软件学报,2003,14(7):1282–1291. <http://www.jos.org.cn/1000-9825/14/1282.htm>
- [11] 郭龙江,李建中,李贵林.无线传感器网络环境下时-空查询处理方法.软件学报,2006,17(4):794–805. <http://www.jos.org.cn/1000-9825/17/794.htm>
- [12] 袁亚湘,孙文瑜.最优化理论与方法.北京:科学出版社,1997.108–110.



陈颖文(1979—),男,湖南株洲人,博士生,主要研究领域为无线传感器网络的路由,数据查询.



吴一(1982—),男,硕士生,主要研究领域为无线传感器网络的路由,数据查询.



徐明(1964—),男,教授,博士生导师,主要研究领域为计算机体系结构,无线网络,移动计算技术.