

基于分簇的传感器网络数据聚集估算机制*

谢磊^{1,2}, 陈力军^{1,2+}, 陈道蓄^{1,2}, 谢立^{1,2}

¹(南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210093)

²(南京大学/香港理工大学 无线与移动传感器网络联合实验室, 江苏 南京 210093)

Clustering-Based Approximate Scheme for Data Aggregation over Sensor Networks

XIE Lei^{1,2}, CHEN Li-Jun^{1,2+}, CHEN Dao-Xu^{1,2}, XIE Li^{1,2}

¹(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

²(NJU-POLYU Cooperative Laboratory for Wireless Sensor Network, Nanjing University, Nanjing 210093, China)

+ Corresponding author: E-mail: chelj@nju.edu.cn, http://dislab.nju.edu.cn

Xie L, Chen LJ, Chen DX, Xie L. Clustering-Based approximate scheme for data aggregation over sensor networks. Journal of Software, 2009,20(4):1023–1037. http://www.jos.org.cn/1000-9825/3282.htm

Abstract: In this paper, a hierarchical clustering-based approximate scheme for in-network aggregation over sensor networks called CASA (clustering-based approximate scheme for data aggregation) is proposed, CASA achieves a good performance in terms of lifetime by minimizing overall energy consumption for communications and balancing energy load among all nodes, while maintaining user-specified quality of data requirement. CASA adopts an optimal parameter for the cluster size, which can well handle the hierarchical approximate work for in-network aggregation with the minimized communication cost. Furthermore, an adaptive tolerance-reallocating scheme is leveraged to further reduce the overall communication cost and maintain a load balance according to various data change rates over deployment regions. Experimental results indicate that significant benefits can be achieved through this CASA approximate scheme.

Key words: sensor network; data aggregation; approximation; cluster

摘要: 提出一种基于簇结构的传感器网络数据聚集估算机制 CASA(clustering-based approximate scheme for data aggregation).在保证用户对数据精确度需求的前提下,CASA 通过最小化网络通信开销以及协调节点间的负载均衡,有效地提高了估算机制的节能性能.CASA 采用最优的分簇规模参数,在基于分簇的网内聚集估算架构中能够最小化网络节点的总体通信开销.此外,CASA 考虑到部署区域感知数据变化率的差异性,采用自适应的误差分配方案来进一步降低网络节点的通信开销,维护节点间的负载均衡.模拟实验结果表明,CASA 估算机制能够显著地提升传感器网络网内数据聚集机制的节能性能,同时保证聚集数据的精确程度.

关键词: 传感器网络;数据聚集;估算;分簇

中图法分类号: TP393 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant Nos.60573132, 60873026, 60573106 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z199 (国家高技术研究发展计划(863)); the National Basic Research Program of China under Grant No.2006CB303000 (国家重点基础研究发展计划(973))

Received 2007-09-24; Accepted 2008-02-21

随着传感器技术、嵌入式计算技术以及低功耗无线通信技术的发展,大量具备感知、计算和通信能力的传感器节点共同组织成大规模的无线传感器网络,已经广泛出现在多种相关的应用领域中.这些传感器节点^[1]能够收集不同的感知数据,例如光照和温度,同时能够对这些数据进行一些更为复杂的操作,如缓存、聚集和过滤等.这种由大量传感器节点组成的网络往往在通信能力、计算能力尤其是能量消耗上存在严重的资源限制,因此,设计一种高效、节能的数据收集策略对大规模部署的传感器网络便至为重要.考虑到在传感器网络中节点的通信能量开销远大于节点计算能量开销,因此,通过合理利用节点的本地计算能力去降低网络的通信开销能够有效节省大量能量开销.近年来,基于数据库的处理方案^[2-3]被引入到传感器网络的研究与实现中,以有效地利用网内处理(in network processing)的特性来降低传感器网络的能量消耗.在对传感器网络的数据进行收集和處理时,将整个传感器网络看成一个大型的分布式数据库,采用类似于 SQL(structured query language)的声明性查询语言对传感器网络进行查询的方式来获取感知信息.通过这种方式,传感器网络中每个节点可以根据用户需求,有效地对相关数据在网内进行过滤、聚集、缓存等操作,最后传至基站终端用户.

网内聚集机制(in-network aggregation)^[4,5]作为一种高效、节能的数据聚集方式,能够充分利用传感器节点的本地处理能力,在将原始数据发送到基站之前,在中间节点就进行数据的聚集合并.目前对于传感器网络数据收集估算机制的研究^[6-9]已经得到越来越广泛的关注,通过合理权衡获取数据的精确度和能量消耗,在保证获取的数据在指定的精确度范围内这一前提下,能够有效地降低网络的能量开销.在本文中,我们针对网内聚集机制提出了一种基于分簇的估算机制 CASA(clustering-based approximate scheme for data aggregation),不同于以前的研究工作,我们提出了一种基于层次结构的估算机制,能够根据部署的环境信息计算获取最优的分簇规模参数,以最小化网络节点的总体通信开销.此外,针对部署区域感知数据变化率的差异性,CASA 采用自适应的误差分配方案进一步降低网络节点的通信开销,维护节点间的负载均衡,并通过理论证明和模拟实验证实,自适应的误差分配结果收敛于理论最优方案.

本文第 1 节介绍相关工作.第 2 节对本文的动因以及采用的系统模型进行描述.第 3 节给出 CASA 的详细设计.第 4 节进行模拟实验,对 CASA 性能进行分析.最后总结全文.

1 相关工作

对于网内聚集工作的研究,Krishnamachari^[4]等人对传感器网络中数据聚集机制的节能效果做出了全面分析.Madden 等人提出了 TAG(tiny aggregation)^[5],一种基于管道(pipeline)的网内数据聚集方案,文中对不同的聚集函数的属性进行了详细的分类,并利用管道技术对网内数据聚集执行有效的调度,以保证子节点能够在指定时间段内将数据聚集并传输给父节点.Demers 等人提出了一种基于推拉(push and pull)技术的聚集节点选择方案和聚集路由树的构建机制^[10].对于网内聚集机制的路由构建,相比于传统树形结构的路由机制,分簇路由具有拓扑管理方便、能量利用高效、数据融合简单等优点,成为当前重点研究的路由技术.Heinzelman 等人提出一种基于簇的数据收集协议 LEACH(low energy adaptive clustering hierarchy)^[11],其基本思想是构建一个两跳的分簇路由结构,每个簇内成员节点向簇头发送感知数据,由簇头将处理后的数据直接发送给基站.Liu 等人提出了一种分布式的传感器网络数据收集和聚合协议 DEEG(distributed energy-efficient data gathering)^[12],与 LEACH 有所不同,簇头之间以多跳方式将收集到的数据发送到指定簇头节点,然后通过该节点将整个网络收集的数据发送到基站.在文献[13]中,沈波等人对目前无线传感器网络簇算法进行了详细的分析和比较.

对于数据收集估算机制的研究,当前的研究工作主要集中于权衡数据精确度和能量消耗的关系,在满足用户对数据精确度需求的前提下,采用估算机制允许适量误差以有效降低网络的整体能量开销.Sharaf 等人提出了一套基于树形路由结构的网内聚集估算机制 TiNA^[6],利用误差率 TCT(temporal coherency tolerance)对聚集数据的精确度进行控制.Deshpande 等人提出了一种基于统计模型的传感器网络估算机制^[8],在基站通过机器学习的方法维护一个数据模型,用户通过查询的方式与该模型交互获取估算结果,该模型按照特定需求定期更新,通过这种方式显著减少了网络传输量,从而达到节能的效果.Tulone 等人继而提出了 PAQ(probabilistic adaptable query system)估算机制^[9],通过在节点本地维护一个轻量级的自回归模型(autoregressive model)来达到基于时间

序列进行数据估算的目的.Hellerstein 等人提出了一套基于小波压缩技术的聚集数据估算机制^[7].Olston 等人针对分布式数据源的聚集查询提出了基于缓存的估算机制^[14],能够动态地自适应调整各数据源的精度需求以最小化数据更新的开销.Goel 等人提出了 PREMON(prediction-based monitoring)^[15],利用感知数据的时空相关性来对移动目标的属性进行估算.上述研究工作大都从设计性能高效的估算模型入手来解决传感器网络的节能问题,不同于以前的研究工作,本文主要结合传感器网络多跳路由结构特性,从设计良好的支持估算机制的网络架构入手,有效地解决传感器网络的节能问题,本文是在我们前期研究工作^[16]的基础上进一步深入和扩展的研究成果.

2 系统模型与问题描述

2.1 系统模型

本文假设 N 个传感器节点随机、均匀分布在一个 $M \times M$ 的二维正方形区域 A 内,并假设该传感器网络具有如下性质:

- 传感器网络为静态网络,即节点部署后不再移动.
- 基站部署在区域 A 以外的一个固定位置,且基站是唯一的.
- 节点可以获知其位置信息.节点依靠 GPS、有向天线或定位算法等辅助设施或算法获取位置信息.
- 节点的无线发射功率可控,即节点可以根据接收者的距离来调整其发射功率.

对于节点通信能量消耗的计算,我们使用文献[11]中 LEACH 提出的能耗模型,该能耗模型给出一个阈值 $d_{crossover}$,当发送节点与接收节点的距离小于 $d_{crossover}$ 时,发送方发送数据的能量损耗与距离的平方成正比,否则与距离的 4 次方成正比.上述的两种能量衰减模型分别称为自由空间(free space)模型和多路衰减(multipath fading)模型.因此,每发送 l 比特的数据到 d 距离处的节点,本地节点需要能耗:

$$E_{Tx}(l, d) = E_{Tx-elec}(l) + E_{Tx-amp}(l, d) = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2, & d < d_{crossover} \\ lE_{elec} + l\epsilon_{mp}d^4, & d \geq d_{crossover} \end{cases} \quad (1)$$

每接收 l 比特的数据,本地节点需要能耗:

$$E_{Rx}(l) = E_{Rx-elec}(l) = lE_{elec} \quad (2)$$

在上面的公式中, E_{elec} 表示无线收发电路所消耗的能量. ϵ_{fs} 和 ϵ_{mp} 分别表示自由空间模型和多路衰减模型中放大器消耗的能量,其大小取决于发送节点与接收节点间的距离以及可接受的位错误率.

对于网内聚集模型,结合估算机制,我们采用了文献[6]中 TiNA 提出的指定精确度的聚集查询方式,如图 1 所示,查询语句的前 3 项 SELECT...FROM...WHERE 与通常的 SQL 语句相同,指定参与聚集查询的节点范围. EPOCH DURATION i 用来指定连续查询中聚集结果收集的时间间隔. VALUES WITHIN tct 用来指定聚集查询的误差范围.对于误差 tct 的定义,我们使用图 2 来说明.通常聚集查询对应的感知属性,如平均温度 $AVG(temp)$,往往在一个固定的范围之内(B_{low}, B_{up})变化, ZeroBound 则表示该属性的一个理论最低值,基站和每个传感器节点会存储每个感知属性的常规变化范围.我们使用 tct 来表示对于这个变化范围的相对误差,则整个聚集结果的绝对误差 Tol 即为 $Tol = (B_{low}, B_{up}) \times tct$.例如:对于某地平均温度 $AVG(temp)$,其常规变化范围为 $(-10^\circ\text{C}, 40^\circ\text{C})$,其理论最低值为 -273°C .若设定 $tct=10\%$,则绝对误差 $Tol = [40^\circ\text{C} - (-10^\circ\text{C})] \times 10\% = 5^\circ\text{C}$.对于其他聚集函数,如 $SUM(attribute)$,由于可以获知参与聚集计算的节点数 $count$, $SUM(attribute) = AVG(attribute) \times count$,其估算处理方案可以按照 $AVG(attribute)$ 进行处理,此时,整个聚集结果的绝对误差 $Tol = (B_{low} - B_{up}) \times tct \times count$.

```

SELECT AggregateFunction (attributes)
FROM sensors s
WHERE s.loc in Range R
EPOCH DURATION i
VALUES WITHIN tct

```

Fig.1 Aggregate query with precision specification

图1 指定精确度的聚集查询

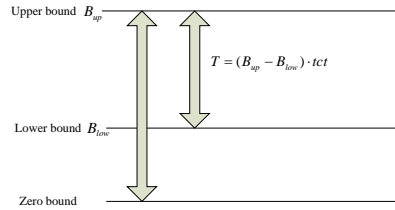


Fig.2 Definition of tolerance tct

图2 误差 tct 定义

2.2 问题描述与分析

在传感器网络上实现数据收集的估算机制的主要策略是:在路由结构中的父子节点上都维护一个估算值 V_{appr} ,当子节点产生新的感知数据 V 时,便计算是否有 $|V - V_{appr}| > (B_{low}, B_{up}) \times tct$,如果成立,则发送消息更新数据 V ,否则,父节点利用本地的估算值 V_{appr} 来估算 V ,通过这种方式可以有效地避免过多的通信开销,从而达到节能的目的.为了在传感器网络多跳路由结构上实现上述网内聚集的估算机制,需要考虑合理的路由结构来支持聚集估算机制的实现.如图3所示,传统的树形路由方案对于数据聚集的估算存在这样的问题:由于估算机制只是针对原始感知数据进行估算操作,对于传统的树形路由结构,仅有少部分分布在边界的叶节点能够产生原始感知数据,而大部分中间节点通过聚集产生部分聚集数据 **PSR**(partial state record),不能使用估算机制来减少网络传输量.因此传统的树形路由方案并不能充分利用网内聚集的估算机制来达到高效节能的目的.如图4所示,基于分簇的路由结构由于感知数据仅在少数簇头节点进行聚集,对于簇内大部分的节点都产生原始的感知数据,因此能够有效支持和实现估算机制的节能效果.

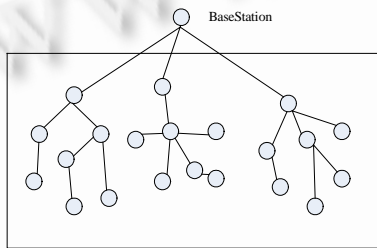


Fig.3 Tree-Based architecture

图3 树形路由方案

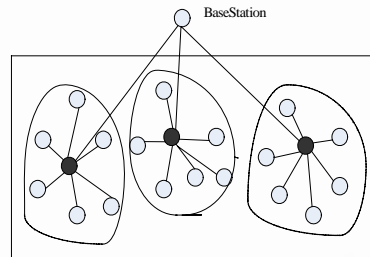


Fig.4 Cluster-Based architecture

图4 分簇路由方案

根据以上分析,在基于分簇的路由结构下如何有效地建立层次结构的估算机制,如何决定分簇的大小以使网络节点总体通信能耗接近最低,以及考虑到地区数据变化率的差异,如何在分布式环境中自适应地合理分配误差区间,在达到负载均衡目的的同时使网络节点总体能耗接近最低,便成为迫切需要解决的一些问题,本文提出的 **CASA** 估算机制旨在解决上述问题,在下一节中我们对 **CASA** 数据聚集估算机制进行具体的描述.

3 CASA 数据聚集估算机制

3.1 估算机制的架构

对于节点间估算值的计算,文献[6-9]提出了多种构建估算模型方案,本文中我们考虑采用基于缓存窗口的估算模型:在路由结构中的父子节点之间,双方节点都维护一个缓存窗口(window),窗口大小设为 m ,用来缓存最近 m 轮之内的数据,则估算值 V_c 取值最近 m 轮数据的平均数,计算如下:

$$V_c = \frac{1}{m} \sum_{i=1}^m V_i \quad (3)$$

当初始时窗口中缓存数据少于 m 个缓存数据时,估算值 V_c 则取窗口中所有数据的平均值.这种基于缓存窗口的估算模型相比于文献[6–9]中提出的估算模型更为简单,有效降低了节点的计算复杂度.同时,通过窗口机制维护估算数据的稳定,减少了感知数据剧烈动荡或接收到错误数据(outlier data)对估算数据产生的影响.基于缓存窗口估算模型,我们结合分簇的路由结构建立一套具有层次结构的估算机制,如图 5 所示.

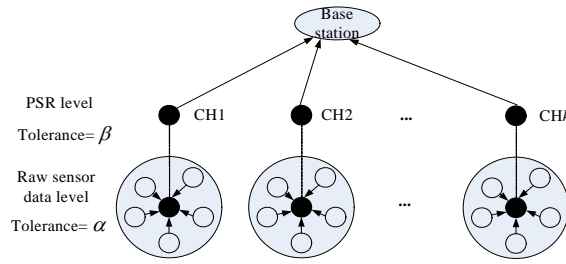


Fig.5 Cluster-Based hierarchical approximate architecture

图 5 基于分簇的层次估算架构

图 5 所示的基于分簇的层次估算架构针对网内聚集的实现,分为两个处理层次:第 1 个层次在每个簇内部实现,簇头 CH_j 对簇内 N_j 个成员中每个节点产生的感知数据 V_i 进行聚集,获得部分聚集数据 PSR_j ,第 2 个层次在所有 k 个簇头和基站之间实现,在基站对每个簇头产生的 PSR 进行聚集:

$$PSR_j = \sum_{i=1}^{N_j} V_i \tag{4}$$

$$AggValue = \sum_{j=1}^k PSR_j \tag{5}$$

因此,CASA 估算机制针对这两个聚集层次,分别提出了两个层次上的估算方案:在簇内聚集层次,对簇内原始感知数据进行估算,维护估算误差 α ;在簇间聚集层次,对部分聚集数据 PSR 进行估算,维护估算误差 β .在每个层次内均一的误差分配机制下,存在下面的误差分配关系:

$$|B_{up} - B_{low}| \cdot \alpha = \frac{1}{N_j} \cdot \sum_{i=1}^{N_j} (|B_{up} - B_{low}| \cdot \alpha) \tag{6}$$

$$|B_{up} - B_{low}| \cdot \beta = \frac{1}{k} \cdot \sum_{j=1}^k (|B_{up} - B_{low}| \cdot \beta) \tag{7}$$

这种具有层次结构的聚集数据估算机制,分别在原始感知数据和部分聚集数据层次进行估算操作:在原始感知数据层次进行估算操作,有效地降低了大量簇内成员节点向簇头节点传输数据的能量开销;在部分聚集数据层次进行估算操作,由于簇头节点相对基站的距离往往远大于簇内通信的距离,其通信开销相对簇内传输要高很多,通过估算机制降低了簇头与基站通信的概率,从而达到高效节能的目的.CASA 估算机制不仅适用于类似于 LEACH^[11]的两跳分簇路由结构,也适用于类似于 DEEG^[12]的簇头之间以多跳方式传输数据的分簇路由结构,其差别仅在于对前者簇间聚集层次估算适用于所有簇头节点,对后者簇间聚集层次估算适用于簇头节点组成多跳路由的所有叶节点,因此,对于不同分簇路由结构具有良好的可扩展性.下文中为了便于讨论 CASA 估算机制的性质,以 LEACH 分簇路由结构为模型进行分析,但其相关性质同样适用于其他分簇路由结构.

下面我们通过定理证明,这种层次结构的聚集数据估算机制的总体误差 tct 接近于两层误差之和 $\alpha+\beta$.

定理 1. 当 $\alpha + \beta \leq C$ (C 为常数)时,有总体误差 $tct \approx (\alpha + \beta)$, 且有 $|tct - (\alpha + \beta)| \leq C^2/4$.

证明:根据相对误差 α, β 的定义,有 $0 < \alpha < 1, 0 < \beta < 1$,原始感知数据估算层次的误差范围是 $(1-\alpha, 1+\alpha)$,部分聚集数据估算层次的误差范围是 $(1-\beta, 1+\beta)$,因此,最终聚集结果误差范围是 $((1-\alpha)(1-\beta), (1+\alpha)(1+\beta))$,因而有:

$$(1 + \alpha)(1 + \beta) = 1 + (\alpha + \beta) + \alpha\beta \tag{8}$$

$$(1 - \alpha)(1 - \beta) = 1 - (\alpha + \beta) + \alpha\beta \tag{9}$$

最终聚集结果的误差 $tct = \max(\alpha + \beta - \alpha\beta, \alpha + \beta + \alpha\beta) = \alpha + \beta + \alpha\beta$, 当 $\alpha + \beta \leq C$ 时,存在 $\alpha\beta \leq (\alpha + \beta)^2 /$

$4 \leq C^2/4$, 故有 $\alpha\beta \leq C^2/4$, 所以, $|tct - (\alpha + \beta)| = |(\alpha + \beta + \alpha\beta) - (\alpha + \beta)| \leq C^2/4$. \square

根据定理 1, 当 $\alpha + \beta < 20\%$ 时(通常用户指定的精确度在这个范围之内), 总体查询误差 tct 将十分接近于 $\alpha + \beta$, 其实际偏差小于 1%, 因此, 给定用户指定的误差 tct , 可以近似地分解为簇内聚集估算误差 α 和簇间聚集估算误差 $tct - \alpha$. 下面, 我们将利用该定理的结果计算最优分簇规模.

3.2 最优分簇规模的计算

第 3.1 节已经介绍了具有层次结构的聚集数据估算机制的原理, 该估算机制依赖于在传感器网络上构建一个基于分簇的数据收集结构, 考虑这种分簇结构的规模, 如何决定簇的大小以充分发挥层次结构估算机制的特性, 使网络节点总体通信能耗接近最低, 便成为一个至关重要的问题. 因为当分簇规模过大时, 成员节点往往需要通过较远的距离向簇头发送感知数据, 消耗了过多的能量; 当分簇规模过小时, 整个网络便退化为接近于传统的树形路由结构, 根据第 2.2 节的论述, 这种结构不能高效、节能地支持估算机制. 下面我们通过定理来证明: 根据基站获取的信息, 可以计算出最优分簇规模, 从而最小化传感器网络的总体通信能耗.

定理 2. 基于层次结构的聚集数据估算机制, 其最优分簇规模 k_{opt} 可以通过计算获得, 且有:

$$k_{opt} = \frac{M\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{fs} P(\alpha_{opt})}{E_{elec}[P'(tct - \alpha_{opt}) - P(\alpha_{opt})] + \varepsilon_{mp} d_{toBS}^4 \cdot P'(tct - \alpha_{opt})}}.$$

证明: 假设当前网络中存在 k 个分簇, 每个簇的大小基本相同, 根据第 2.1 节给出的能耗模型, 在簇内对每一个成员节点, 其能耗为

$$E_{non-CH} = (IE_{elec} + l\varepsilon_{fs} d_{toCH}^2) P(\alpha) \quad (10)$$

其中, d_{toCH} 为成员节点到簇头节点的距离, $P(\alpha)$ 为当误差为 α 时成员节点向簇头更新数据的概率. 对于每一个簇头, 其能耗为

$$E_{CH} = IE_{elec} \left(\frac{N}{k} - 1 \right) \cdot P(\alpha) + lE_{DA} \frac{N}{k} + (IE_{elec} + l\varepsilon_{mp} d_{toBS}^4) \cdot P'(\beta) \quad (11)$$

其中, d_{toBS} 为簇头到基站的距离, $P'(\beta)$ 为当误差为 β 时簇头向基站更新部分聚集数据 PSR 的概率, 根据定理 1, 有 $\beta \approx tct - \alpha$. 每个簇的能量消耗为

$$E_{cluster} = E_{CH} + \left(\frac{N}{k} - 1 \right) E_{non-CH} \approx E_{CH} + \frac{N}{k} E_{non-CH} \quad (12)$$

由文献[11]我们知道, 有 d_{toCH}^2 的期望值为

$$E[d_{toCH}^2] = \frac{1}{2\pi} \frac{M^2}{k} \quad (13)$$

因此, 网络中所有簇的总体能量消耗为

$$\begin{aligned} E_{total}(k, \alpha) &= kE_{cluster} \\ &= lE_{elec} [(2N - k) \cdot P(\alpha) + k \cdot P'(tct - \alpha)] + lE_{DA} N + lk\varepsilon_{mp} d_{toBS}^4 P'(tct - \alpha) + l\varepsilon_{fs} \frac{1}{2\pi} \frac{M^2}{k} N \cdot P(\alpha) \end{aligned} \quad (14)$$

为了获得该优化问题中 E_{total} 的最小值, 根据对最优点的求解方案, 我们对变量 k 和 α 求偏导数:

$$\begin{cases} \frac{\partial}{\partial k} E_{total}(k, \alpha) = 0 \\ \frac{\partial}{\partial \alpha} E_{total}(k, \alpha) = 0 \end{cases} \quad (15)$$

由此通过数值计算求出 α_{opt} , 从而得到最优的分簇规模 k_{opt} :

$$k_{opt} = \frac{M\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{fs} P(\alpha_{opt})}{E_{elec}[P'(tct - \alpha_{opt}) - P(\alpha_{opt})] + \varepsilon_{mp} d_{toBS}^4 \cdot P'(tct - \alpha_{opt})}} \quad (16)$$

当仅使用簇内聚集层次的估算机制, 即 $\alpha = tct, \beta = 0$ 时,

$$k_{opt} = \frac{M\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{\beta} P(tct)}{E_{elec}(1-P(tct)) + \varepsilon_{mp} d_{toBS}^4}} \quad (17)$$

证明完毕. \square

根据定理 2,对于刷新概率函数 $P(\alpha)$ 和 $P'(\beta)$,我们根据随机过程的原理假设其符合下面的形式: $P(\alpha) = Z_1 \cdot \alpha^{-p_1}$, $P'(\beta) = Z_2 \cdot \beta^{-p_2}$,其中 Z_i 为常数, p_i 为正数.对于原始感知数据估算层次的刷新概率 $P(\alpha)$,根据文献[14]的论证结果, $P(\alpha)$ 符合随机行走模型(random walk model),因而有:

$$P(\alpha) \approx \frac{1}{(2 \cdot s)^2} \cdot \alpha^{-2} \quad (18)$$

其中, s 为随机行走的步长信息. $P'(\beta)$ 由于其表示簇间聚集估算的刷新概率,其结果与实际分簇规模 k_{opt} 有关,但可以验证:当分簇大小 N/k 到达一定数量时,刷新概率随 k 变化相对稳定,为了简化计算,我们用平均分簇规模 k_{avg} 来估算 $P'(\beta)$ 的相关参数.有时,为了使运算结果更加精确,可以采用迭代的方法求取 k_{opt} .即采用 k_{avg} 为 k_{opt} 的初始估算值,通过多次迭代地将修正的 k_{opt} 估算值对应的 $P'(\beta)$ 代入式(16)中求得最终充分逼近 k_{opt} 的估算值.因此,根据以上分析,只需在传感器网络构建分簇结构之前,通过一个学习阶段(learning phase)变化相对误差,收集刷新信息,即可估算出 $P(\alpha)$ 和 $P'(\beta)$ 刷新概率函数中的相关参数,以指导基站对最优分簇规模的计算.

3.3 分簇结构的生成与维护

分簇结构的生成和维护主要分为 3 个阶段:学习阶段、构建分簇阶段和维护阶段.在学习阶段,当所有传感器节点都被部署在观测区域中时,每个节点周期性地获取感知数据,并向基站发送当前数据 V_i ,在基站为每一个节点根据缓存窗口机制维护一个估算值 V_{c_i} ,同时维护误差值 α ,在当前数据 V_i 与估算值 V_{c_i} 之差超过 α 指定的误差范围时,基站为每个节点维护的刷新次数 U_i 递增.基站在一定范围内调整 α ,从而获取相应的刷新次数 U_i ,利用计算出的平均刷新率估算出 $P(\alpha)$ 的参数.同理,基站在一定范围内调整簇间聚集误差 β ,利用节点的位置分布信息和平均分簇规模 k_{avg} ,结合 PSR 的平均刷新率估算出 $P'(\beta)$ 的参数.结合以上参数基站根据定理 2 计算出最优分簇规模 k_{opt} 以及误差参数 α_{opt} 和 β_{opt} ,发送给每个节点.在构建分簇阶段,每个节点以 $P=k_{opt}/N$ 的概率选择自己成为簇头节点,并向邻居节点广播消息构建分簇,成员节点根据到候选簇头的距离参数加入相应的分簇.在维护阶段,每个节点根据获得的误差参数 α_{opt} 和 β_{opt} 实现估算机制,定期地更新数据,并周期性地交替更换簇头以保证节点的负载均衡.学习阶段和构建分簇阶段往往会带来通信能量的额外开销,但可以有效地减少维护阶段的长时间周期的数据聚集通信能耗,其节能效果是显著的.分簇结构的生成算法流程如图 6 所示.

3.4 自适应的误差分配方案

当传感器网络被部署到观测区域中后,由于地区性数据变化率存在差异,会出现某些区域感知数据变化较快而某些区域变化较慢的现象,根据均一的误差分配机制,每个节点拥有相同的误差,这样会导致感知数据变化较快区域的节点频繁传输数据从而过早地耗尽能量,而变化较慢区域的节点则无法利用富余的能量.基于上述分析,可以考虑通过调整不同区域节点的误差,对于变化较慢区域的节点赋予较少的误差量,把更多的误差量赋予数据变化较快区域的节点,在分布式环境中根据不同的数据变化率采取自适应的方案来分配误差.通过这种非均一的误差分配机制^[14],可以使网络节点达到负载均衡的效果.同时,通过把富余的误差量分配给变化较快区域的节点,可以实现在不显著增加变化较慢区域节点通信能耗的同时,能够有效降低这些节点的通信能耗,从而降低传感器网络的整体能量消耗,达到高效节能的目的.

这样,结合地区数据变化率差异性,如何在分布式环境中自适应地合理分配误差量,在达到负载均衡目的的同时使网络节点总体能耗接近最低,便成为一个有待解决的问题.本文提出的自适应误差分配方案旨在解决上述问题,我们在簇内聚集估算层次使用自适应的、非均一的误差分配方案来有效降低网络的整体通信能耗,在簇间通信层次,由于簇头数量较少,其能耗并非占总体通信能耗的主要部分,同时,由于部分聚集结果变化相对稳定,所以仍考虑使用均一的误差分配方案.

<p>Sensor Node Side:</p> <pre> while (the timer(T) for learning phase is not expired) $V_i \leftarrow$ getLocalSensorData() send V_i to the base station endwhile receive [$k_{opt}, \alpha_{opt}, \beta_{opt}$] from the base station $P = k_{opt}/N$ generate a random number n, where $0 < n < 1$ if ($n > P$) elect itself to be the cluster head, broadcast HEAD_AD_MSG to neighbour nodes else wait for HEAD_AD_MSG and join the first-heard cluster head endif </pre>	<p>Base Station Side:</p> <pre> while (the timer(T) for learning phase is not expired) receive data V_i from each sensor node i if ($V_i - V_{c_i} > \alpha \times B_{up} - B_{low}$) $V_{c_i} = [V_{c_i} \times (m-1) + V_i]/m$ $V'_i = V_i$ $U_i = U_i + 1$ else $V_{c_i} = [V_{c_i} \times (m-1) + V'_i]/m$ endif endwhile Calculate parameters for probability function $P(\alpha)$ and $P'(\beta)$ $[k_{opt}, \alpha_{opt}] = \text{calculateOptimalClusterSize}(P(\alpha), P'(\beta))$ Send [$k_{opt}, \alpha_{opt}, \beta_{opt}$] to each sensor node i </pre>
--	--

Fig.6 Pseudo-Code for constructing clusters

图6 分簇的构造算法伪码

对于簇内聚集估算层次非均一的误差分配方案,存在以下的误差约束,所有节点的误差均值应等于聚集结果的误差值:

$$|B_{up} - B_{low}| \cdot \alpha = \frac{1}{N_j} \cdot \sum_{i=1}^{N_j} (|B_{up} - B_{low}| \cdot \alpha_i) \quad (19)$$

式(19)可以进一步简化为

$$\alpha = \frac{1}{N_j} \cdot \sum_{i=1}^{N_j} (\alpha_i) \quad (20)$$

我们定义负荷指数来指导自适应误差分配方案的误差分配.

定义 1(负荷指数 burden factor). 负荷指数刻画传感器节点平均每轮通信能量开销,其形式如下:

$$B_i = \frac{C_i \cdot P(\alpha_i)}{\alpha_i} \quad (21)$$

其中, C_i 表示向簇头更新数据包的能量开销, $P(\alpha_i)$ 表示簇内聚集误差设为 α_i 时的刷新概率, α_i 为节点的当前设定误差. 下面我们具体描述自适应误差分配方案.

在构建分簇步骤完成之后,每个成员节点 i 首先向簇头发送簇内通信开销 C_i ,在簇内聚集层次每一轮,每个成员节点产生感知数据 V_i ,成员节点根据估算机制判断是否有 $|V_i - V_{c_i}| > (B_{low}, B_{up}) \times \alpha_i$,如果成立,则向簇头发送数据 V_i ,簇头递增刷新次数 U_i ,成员节点和簇头同时更新 V_{c_i} ,维持误差 α_i 不变;否则,成员节点无须向簇头发送数据 V_i ,成员节点和簇头同时利用上次估算值更新 V_{c_i} ,并缩小误差 α_i ,有: $\alpha_i = \alpha_i \times (1 - \theta)$, θ 为设定的缩小比例, $0 < \theta < 1$. 每 $T-1$ 轮,成员节点需向簇头发送剩余能量值 Er_i .

对于簇头节点 j ,除了为首轮向基站节点发送簇内每轮平均通信开销 $Avg_j(C)$ 以及根据估算机制向基站节点更新部分聚集数据 PSR_j 外,在簇头上为每个成员节点 i 维护一个负荷指数 B_i ,在每 $T-1$ 轮还需向基站发送簇内平均剩余能量 $Avg_j(E_r)$ 和平均更新概率 $Avg_j(P)$,有:

$$Avg_j(E_r) = \frac{1}{N_j} \sum_{i=1}^{N_j} Er_i \quad (22)$$

$$Avg_j(P) = \frac{1}{N_j} \sum_{i=1}^{N_j} U_i \quad (23)$$

其中, Er_i 和 U_i 分别为 T 轮内节点 i 的剩余能量与刷新次数, N_j 为簇的大小. 同时,簇头统计当前簇内误差之和 $SUM_j(\alpha)$,作为误差调整的下界(lower bound),并发送给基站,有:

$$SUM_j(\alpha) = \sum_{i=1}^{N_j} (\alpha_i) = Avg(\alpha) \cdot N_j \quad (24)$$

对于基站节点,基站为每个簇 j 维护一个负荷指数 W_j ,用来刻画每个簇的总体负荷:

$$W_j = \frac{Avg(C) \cdot Avg(P) \cdot N_j}{Avg(\alpha) \cdot Avg(E_r)} \quad (25)$$

基站在每 T 轮根据 W_j 为各簇重新分配误差区间,首先保证分配各簇的误差下界值 $SUM_j(\alpha)$,对剩下的误差区间 ΔD 根据权重按比例给每个簇 j 分配 Δd_j ,即有:

$$\Delta D = \alpha \cdot N - \sum_{j=1}^k SUM_j(\alpha) \quad (26)$$

$$\Delta d_j = \frac{W_j}{\sum_{j=1}^k W_j} \Delta D = \frac{W_j}{\sum_{j=1}^k W_j} \left[\alpha \cdot N - \sum_{j=1}^k SUM_j(\alpha) \right] \quad (27)$$

在完成分配计算后,基站将为各簇分配的误差区间 Δd_j 发送给相应的簇头.在每 T 轮,簇头根据基站分配出的误差区间 Δd_j ,在簇内为每个成员节点分配误差量.簇头以每个成员节点的负荷指数 B_i 为权重,为每个节点再次按权重比例分配富余的误差值 $\Delta \alpha_i$,对分配到富余误差值的成员节点 i 更新其误差值 α_i .

$$\Delta \alpha_i = \frac{B_i}{\sum_{j=1}^{N_j} B_i} \Delta d_j = \frac{B_i \cdot W_j}{\sum_{j=1}^{N_j} B_i \cdot \sum_{j=1}^k W_j} \left[\alpha \cdot N - \sum_{j=1}^k SUM_j(\alpha) \right] \quad (28)$$

通过对每个节点以负荷指数为权重分配富余误差值的策略,使得当前簇内通信能耗较高并且误差分配量相对较小的那些节点能够获得较大的富余误差量补给,能够有效地均衡各节点的负荷并且降低整体网络通信开销.自适应误差分配的算法流程如图 7 所示.

3.5 自适应误差分配机制分析

第 3.4 节提出了自适应的误差分配方案,其目的是为了均衡各节点间的通信负载,同时有效降低整体的网络通信能耗.下面我们来证明在簇内聚集估算层次自适应误差分配方案的节能效果是接近理论最优的.

定理 3. 在簇内聚集估算层次自适应误差分配方案的节能效果收敛于非均一误差分配的理论最优方案.

证明:考虑在簇内聚集估算层次采用非均一的误差分配方案,如何在固定的总体误差范围内合理分配误差以使簇内聚集层次总体通信能耗最低,便成为一个优化问题.

我们使用效用函数(utility function) $C(\alpha)$ 来表示簇内聚集层次总体通信能耗:

$$C(\alpha) = \sum_{i=1}^{N-k} C_i P(\alpha_i) \quad (29)$$

$C(\alpha)$ 包括 $N-k$ 个成员节点的通信开销, C_i 为节点 i 的通信开销, $P(\alpha_i)$ 为节点 i 的通信概率, α_i 为分配给节点 i 的误差量,根据式(20)可以推导出对于所有的成员节点误差,其满足约束关系:

$$\sum_{i=1}^{N-k} \alpha_i = \alpha \cdot N = (tct - \beta) \cdot N \quad (30)$$

其中, $tct - \beta$ 即为簇内聚集估算分配的误差量 α ,在这里为常数.我们的优化问题便为如何在满足约束关系(32)的前提下调整 α_i 以最小化总体能量开销 $C(\alpha)$ 的数值.针对这样的优化问题,我们可以利用拉格朗日乘法^[17] (Lagrange multiplier)计算最优解,存在目标函数 $f(\alpha) = C(\alpha)$:

$$f(\alpha) = \sum_{i=1}^{N-k} C_i P(\alpha_i) \quad (31)$$

以及约束函数 $g(\alpha)$:

$$g(\alpha) = \sum_{i=1}^{N-k} \alpha_i - (tct - \beta) \cdot N \quad (32)$$

Algorithm for Cluster Head.

```

round=0
if (round==0)
receive  $C_i$  from each node, calculate
 $Avg(C_i)$  and send  $Avg(C_i)$  to base station
round=round+1
endif
for each round>0
if (receive  $V_i$  from member node  $i$ )
 $V_c=[V_c \times (m-1)+V_i]/m$ 
 $V'_i=V_i$     $U_i=U_i+1$ 
else
 $\alpha_i = \alpha_i \cdot (1-\theta)$ 
 $V_c=[V_c \times (m-1)+V'_i]/m$ 
endif
calculate  $PSR_j$  and send  $PSR_j$  to the base
station with approximate scheme
if (round% $T==(T-1)$ )
calculate  $Avg_j(E_r), Avg_j(P)$  and  $Avg_j(\alpha)$ 
and send to base station
endif
if (round% $T==0$ )
receive allocated tolerance  $\Delta d_j$  from base
station, calculate  $B_i$  and new allocated
 $\Delta \alpha_i$  for each node
send  $\Delta \alpha_i$  to node  $i$ 
endif
round=round+1
endfor

```

Algorithm for Cluster Member.

```

round=0
if (round==0)
send  $C_i$  to its cluster head
round=round+1
endif
for each round>0
 $V_i \leftarrow \text{getLocalSensorData}()$ 
if ( $|V_i - V_c| > (B_{low}, B_{up}) \times \alpha_i$ )
 $V_c=[V_c \times (m-1)+V_i]/m$ 
 $V'_i=V_i$ 
send  $V_i$  to its cluster head
else
 $\alpha_i = \alpha_i \cdot (1-\theta)$ 
 $V_c=[V_c \times (m-1)+V'_i]/m$ 
endif
if (round% $T==(T-1)$ )
send  $E_r$  to the cluster head
endif
if (round% $T==0$ )
receive new allocated  $\Delta \alpha_i$ 
 $\alpha_i = \alpha_i + \Delta \alpha_i$ 
endif
round=round+1
endfor

```

Algorithm for Base Station.

```

round=0
if (round==0)
receive  $Avg(C_i)$  from each cluster
round=round+1
endif
for each round>0
receive  $PSR_j$  from each cluster  $j$ 
with approximate scheme
if (round% $T==(T-1)$ )
receive  $Avg_j(E_r), Avg_j(P), Avg_j(\alpha)$ 
from each cluster  $j$ 
endif
if (round% $T==0$ )
calculate  $W_j$  and new allocated  $\Delta d_j$ 
for each cluster  $j$ 
send  $\Delta d_j$  to each cluster head  $j$ 
endif
round=round+1
endfor

```

Fig.7 Pseudo-Code for adaptive tolerance allocate algorithm

图7 自适应误差分配算法伪码

构造辅助函数 $F(\alpha) = f(\alpha) - \lambda \cdot g(\alpha)$, λ 为某一常数, 则根据拉格朗日乘子法有 $\frac{\partial}{\partial \alpha} F(\alpha) = 0$, 我们得到:

$$C_i \cdot \sum_{i=1}^{N-k} \frac{\partial}{\partial \alpha_i} P(\alpha_i) = \lambda \cdot (N-k) \quad (33)$$

根据随机行走的模型, 对 $P(\alpha_i)$ 有: $P(\alpha_i) \approx \frac{1}{(2 \cdot s)^2} \cdot \alpha_i^{-2}$, 根据上式我们得到下面的约束关系:

$$\frac{C_i \cdot P(\alpha_i)}{\alpha_i} = D \cdot \lambda \quad (34)$$

其中, $D \cdot \lambda$ 为常数, $D = -1/2$. 由负荷指数 B_i 的定义, 存在下面的等式关系:

$$B_i = \frac{C_i \cdot P(\alpha_i)}{\alpha_i} = D \cdot \lambda \quad (35)$$

上式表明, 当所有节点的负荷指数趋向于一个相等的常数时, 总体通信能耗最低, 而根据我们的自适应误差分配方案, 通过在数据变化快, 即 $P(\alpha_i)$ 高的节点上逐步提高误差 α_i 的取值, 以及在数据变化慢, 即 $P(\alpha_i)$ 低的节点上逐步降低误差 α_i 取值的方式, 最终能够使所有节点的负荷指数 B_i 收敛于一个相等的数值, 使得整体能耗最低。 □

由定理 3 我们得知, 通过自适应误差分配方案, 其节能效果最终可以收敛于非均一误差分配的理论最优方案. 其收敛速度受到自适应调整的时间间隔 T 的影响, 考虑到调整的时间间隔 T 越小, 其收敛速度越快, 但同时带

来的控制信息传输开销也会越大,因此,选择合适的自适应调整的时间间隔 T 对提高自适应误差分配方案的节能性能有很重要的影响,这将成为我们下一步研究工作需要考虑的问题。

4 性能评价

4.1 模拟实验环境设置和参数

本节通过模拟实验来对 CASA 估算机制的性能进行评估,我们基于 Glomosim 模拟实验平台实现了本文提出的估算算法.实验设备是一台运行 WinXP,具有 P4 处理器、1G 内存的 PC.模拟实验环境中节点随机分布在大小为 $100 \times 100(\text{m}^2)$ 的监测区域内,部署节点总数为 100.实验中,我们使用随机行走模型(random walk model)来模拟生成感知数据,其中使用随机度(randomness degree)来描述感知数据相比前一轮样本发生变化的概率,使用随机步长(random step size)来描述每次数据变化的幅度,用百分比表示.具体的实验参数设置见表 1.我们通过以下方面来评估 CASA 的性能:首先通过性能对比来验证估算架构的节能效果,然后通过实验结果验证最优分簇规模算法的有效性,最后通过性能对比来对自适应误差分配方案的节能效果进行验证.

Table 1 Parameters of simulation

表 1 模拟实验参数列表

Parameter	Value	Parameter	Value
Area (m^2)	100×100	Random degree	0.0~1.0, 0.5 default
Number of nodes	100	Random step size	0.0~1.0, 0.5 default
Base station position	(50,100), (50,150), (50,200)	(B_{low}, B_{up})	(-10,40)
Packet size (bits)	4 000	E_{elec} (nJ/bit)	50
Number of rounds	1 000	ϵ_{fs} (pJ/bit/ m^2)	13
T_{ct} reallocate interval T	10	ϵ_{mp} (pJ/bit/ m^4)	0.001 3
t_{ct}	0%~30%, 10% (default)	$d_{crossover}$ (m)	75
Initial energy (J)	2	Standard deviation α	0.05
Window size	1~10, 5 default		

4.2 估算架构的性能验证

图 8 给出了在不同误差 t_{ct} 情况下各种数据收集机制的性能比较.我们针对随机生成的同一节点分布分别构造出传统的树形路由结构和分簇的路由结构,在这两种不同的路由结构上,我们分别实现带有估算机制的传输方案和不带有估算机制的传输方案.可以看到,在没有估算机制的支持下,传统的树形路由和分簇路由方案需要大量的能量开销,而使用估算机制的树形路由方案在 $t_{ct}=0.2$ 时,可以减少 40% 的能量开销,相比于前 3 种方案,可以看到,基于分簇路由的 CASA 估算机制能够进一步有效地降低网络的能量开销,在 $t_{ct}=0.2$ 时,相比树形结构的估算方案能够进一步减少 38% 的能量开销.图 9 给出了具有不同误差的分簇估算机制的性能受数据变化率的影响,随着随机度 RD 的增加,数据的变化率进一步加快.可以看到,不使用估算机制的分簇方案不受任何影响,但一直处于一个高能耗开销的水平 42J,使用估算机制的分簇方案的能耗开销随之增长,此时具有更大误差量的估算方案($t_{ct}=0.3$)相比于小误差量的估算方案($t_{ct}=0.1$),面对较高的数据变化率有更好的节能效果.在 $RD=0.9$ 时,前者比后者可以节省 53% 的能量开销,但即使在这种情况下, $t_{ct}=0.1$ 的估算方案相比不使用估算机制的分簇方案仍能减少 31% 的能量开销.图 8 和图 9 的实验结果表明,基于分簇的估算架构在不同数据变化率的观测环境下都能够有效降低网络的总体能耗.图 10 描述了缓存窗口大小面对不同数据变化幅度时对估算机制节能效果的影响,实验中,我们利用数据变化的随机步长 RSS 来调整数据的变化幅度.可以看到,对于较小的数据变化幅度,不同缓存窗口大小的估算机制具有基本相同的能量开销,随着变化幅度的进一步增大,当 $RSS=0.9$ 时,具有较小窗口的估算机制(window size=1)能量开销高达 43J,而同时具有较大窗口的估算机制(window size=10)其能量开销也有 38J.相比而言,窗口大小适中的估算机制(window size=5)则仅有 31J 的能量开销.这是因为,对于较小窗口的估算机制,由于只缓存为数很少的数据,因此估算值受数据波动的影响非常大,总是受制于最新接收的波动数据,估算效果受到极大影响.而对于较大窗口的估算机制,由于其缓存过多的历史数据,估算值受历史数据影响,其变化非常缓慢,不能及时反映当前数据的变化状况,估算效果也受到很大影响.窗口大小适中的估算机制,由于既能及时反映当前数据的变化状况,又能不受数据波动的太大影响,能够达到很好的节能效果.因此,图 10

的实验结果表明,选择合理的估算窗口大小对估算机制的节能效果有很重要的影响.

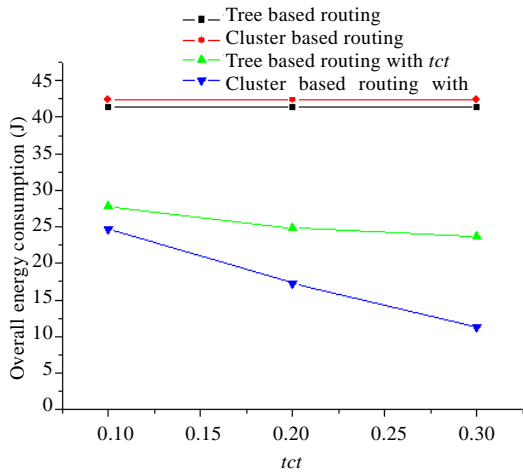


Fig.8 Energy consumption vs. tct

图8 不同估算机制的能量消耗

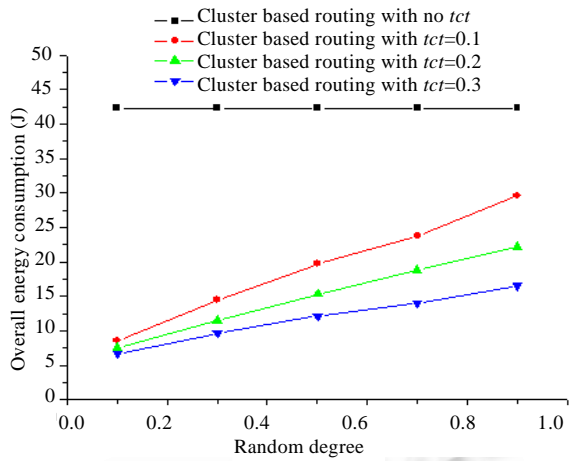


Fig.9 Energy consumption vs. random degree

图9 能量消耗与随机度的关系

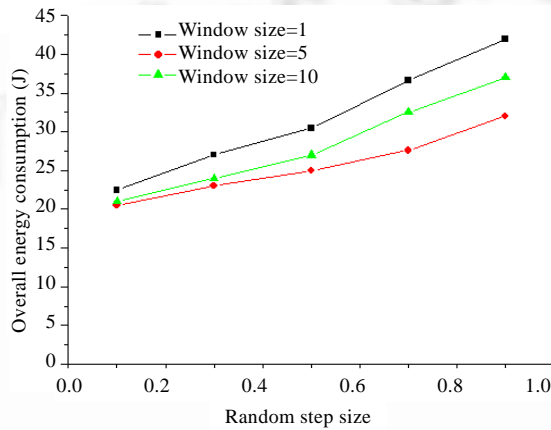


Fig.10 Energy consumption vs. cache window size

图10 能量消耗与缓存窗口大小关系

4.3 最优分簇算法的验证

图 11 给出了分簇规模对网络总体能耗的影响.我们将分簇数量在 1~50 间加以变化,可以看到,当分簇数量在 6~10 之间时能够使传感器网络达到整体最低能量消耗约 24J,而我们在基站通过最优分簇规模计算方法计算出的 $k_{opt}=8$ 正好处在这个范围之内,通过实验数据有效验证了最优分簇规模计算方法的有效性.图 12 进一步给出了基站位置对 k_{opt} 和 α 取值的影响.可以看到,随着基站位置进一步远离观测区域, k_{opt} 和 α 的取值都相对减少.这是因为,随着基站位置的进一步远离,簇头与基站之间的通信开销逐步增大,而根据 CASA 基于层次结构的估算机制,进一步降低 k_{opt} 有助于减少簇头数量,从而降低簇头与基站通信的整体开销,进一步降低簇内聚集估算误差 α ,从而提升簇间聚集估算误差 β ,有助于充分利用 PSR 的估算机制有效降低簇头与基站通信的概率,达到降低总体通信能耗的效果.图 11 和图 12 的实验结果充分验证了最优分簇规模算法的有效性.

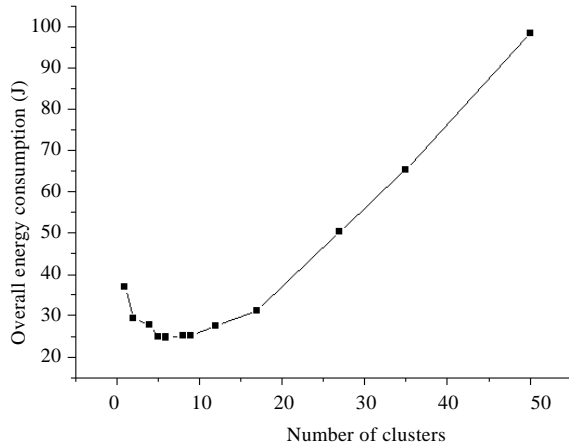


Fig.11 Energy consumption vs. cluster size

图 11 能量消耗与分簇规模的关系

BS location	k_{opt}	α
(50,100)	10	0.95
(50,150)	8	0.83
(50,200)	5	0.62

Fig.12 BS location vs. k_{opt} and α

图 12 基站位置与 k_{opt} 和 α 的关系

4.4 自适应误差分配方案的性能验证

在本节性能验证实验中均设定 $tct=0.1$,感知数据变化率均方差 $\alpha=0.05$,自适应调整的时间间隔 $T=10$.图 13 给出不同误差分配方案的能量消耗对比.实验中,我们构造出 3 种估算传输方案,分别为均一的误差分配方案、自适应的误差分配方案和理论最优的非均一误差分配方案,对于理论最优的非均一误差分配方案,其参数是根据当前的各节点区域的变化数值计算而得.可以看到,自适应的误差分配方案其能量消耗在 $round=0$ 时与均一的误差分配方案有相同的能量开销,随着轮数的递增,自适应的误差分配方案能够逐步调整误差分配,降低能量开销,逐渐收敛于非均一方案最优配置.可以看到,在 $T=10$ 时,其收敛速度还是很快的,在 $round=1000$ 时,其能量开销已经十分接近非均一方案最优配置的能量开销,此时,自适应的误差分配方案比均一的误差分配方案能够减少 24%的能量开销.图 14 给出不同误差分配方案网络寿命的对比.这里,我们分别构造出不使用估算机制的分簇路由方案、使用均一误差的分配方案以及使用自适应的误差分配方案.我们定义网络寿命为网络中第 1 个节点死亡的时间.可以看到,对于不使用估算机制的分簇路由方案,其网络寿命为 372,使用均一误差分配方案的网络寿命为 596,而使用自适应的误差分配方案网络寿命为 674,图 13 和图 14 的实验表明,自适应的误差分配方案不仅能够迅速收敛于最小化整体通信能耗的方案,同时能够通过均衡各节点的负载,有效地延长网络的寿命.

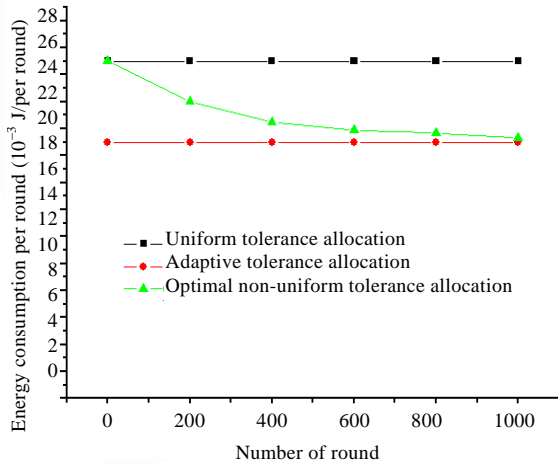


Fig.13 Energy consumption for different tolerance allocation strategies

图 13 不同误差分配方案的能量消耗

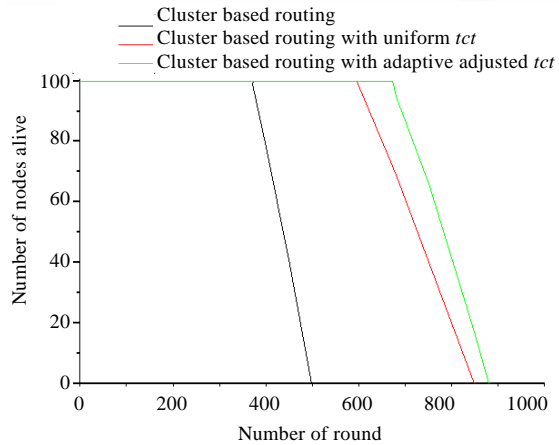


Fig.14 Measurement of network lifetime

图 14 网络寿命的比较

图 15 分析了地区性数据变化差异程度对不同误差分配方案性能的影响,实验通过调整感知数据变化率均方差 σ 来改变数据变化差异程度.可以看到,均一的误差分配方案不受任何数据变化差异程度的影响,始终维持一个较高的能耗水平 25.2J,而自适应的误差分配方案随着地区数据的变化,其差异程度逐渐增加,能够更为有效地利用非均一的误差分配方案来更合理地分配误差以降低网络通信开销,在具有较大的地区数据变化差异程度 $\alpha=0.1$ 时,自适应的误差分配方案能够比均一的误差分配方案节省 33% 的能量开销.

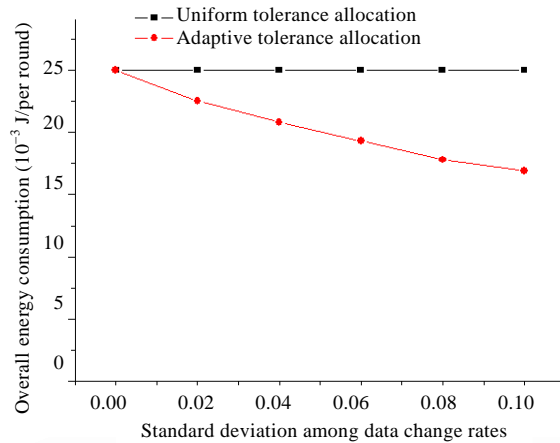


Fig.15 Impact of deviation for data change rates

图 15 数据变化差异对能耗的影响

5 总结和下一步工作

本文针对传感器网络的网内聚集问题提出了一种基于分簇层次结构的估算机制 CASA. CASA 采用最优分簇规模参数,在基于分簇的网内聚集估算架构中能够最小化网络节点的总体通信开销.此外, CASA 考虑到部署区域感知数据变化率的差异性,采用自适应的误差分配方案进一步降低网络节点的通信开销,维护节点间的负载均衡.实验结果表明, CASA 估算机制能够显著提升传感器网络网内聚集机制的节能性能,同时保证聚集数据的精确程度.在下一步的研究工作中,我们考虑将该估算机制由基于 LEACH 的两跳分簇路由结构进一步应用到更为普遍的多跳分簇路由结构中,研究如何在多个层次结构间以及层次结构内分配估算误差,决定最优的 α, β 误差分配参数;同时对自适应调整机制,将进行对自适应调整时间间隔 T 最优取值的研究,以充分提高节能效果.

References:

- [1] Hill J, Culler D. Mica: A wireless platform for deeply embedded networks. *IEEE Micro*, 2002, 22(6): 12–24.
- [2] Gehrke J, Madden S. Query processing in sensor networks. *IEEE Pervasive Computing*, 2004, 3(1): 46–55.
- [3] Li JZ, Li JB, Shi SF. Concepts, issues and advance of sensor networks and data management of sensor networks. *Journal of Software*, 2003, 14(10): 1717–1727 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1717.htm>
- [4] Krishnamachari B, Estrin D, Wicker SB. The impact of data aggregation in wireless sensor networks. In: *Proc. of the 22nd Int'l Conf. on Distributed Computing Systems*. Washington: IEEE Computer Society, 2002. 575–578.
- [5] Madden S, Franklin MJ, Hellerstein JM, Hong W. TAG: A tiny aggregation service for ad-hoc sensor networks. In: *Proc. of the OSDI. 2002*. <http://doi.acm.org/10.1145/844128.844142>
- [6] Sharaf MA, Beaver J, Labrinidis A, Chrysanthis PK. TiNA: A scheme for temporal coherency-aware in-network aggregation. In: *Proc. of the MobiDE. 2003*. 69–76. <http://doi.acm.org/10.1145/940923.940937>
- [7] Hellerstein JM, Hong W, Madden S, Stanek K. Beyond average: Toward sophisticated sensing with queries. In: *Proc. of the IPSN. 2003*. 63–79.

- [8] Deshpande A, Guestrin C, Madden S, Hellerstein JM, Hong W. Model-Based approximate querying in sensor networks. *VLDB Journal*, 2005,14(4):417-443.
- [9] Tulone D, Madden S. PAQ: Time series forecasting for approximate query answering in sensor networks. In: *Proc. of the EWSN*. 2006. 21-37.
- [10] Demers AJ, Gehrke J, Rajaraman R, Trigoni A, Yao Y. The cougar project: A work-in-progress report. *SIGMOD Record*, 2003,32(4):53-59.
- [11] Heinzelman WB, Chandrakasan AP, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. on Wireless Communications*, 2002,1(4):660-670.
- [12] Liu M, Gong HG, Mao YC, Chen LJ, Xie L. A distributed energy-efficient data gathering and aggregation protocol for wireless sensor networks. *Journal of Software*, 2005,16(12):2106-2116 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/2106.htm>
- [13] Shen B, Zhang SY, Zhong YP. Cluster-Based routing protocols for wireless sensor networks. *Journal of Software*, 2006,17(7):1588-1600 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/1588.htm>
- [14] Olston C, Widom J. Approximate caching for continuous queries over distributed data sources. Technical Report, Computer Science Department, Stanford University, 2002. <http://dbpubs.stanford.edu/pub/2002-8>
- [15] Goel S, Imielinski T. Prediction-Based monitoring in sensor networks: Taking lessons from MPEG. *ACM SIGCOMM Computer Communication Review*, 2001,31(5):82-98.
- [16] Xie L, Chen LJ, Chen DX, Xie L. A clustering-based approximation scheme for in-network aggregation over sensor networks. In: *Proc. of the UIC*. 2007. 503-513.
- [17] Stewart J. *Calculus: Early Transcendentals*. 2nd ed., Brooks/Cole, 1991.

附中文参考文献:

- [3] 李建中,李金宝,石胜飞.传感器网络及其数据管理的概念、问题与进展. *软件学报*,2003,14(10):1717-1727. <http://www.jos.org.cn/1000-9825/14/1717.htm>
- [12] 刘明,龚海刚,毛莺池,陈力军,谢立.高效节能的传感器网络数据收集和聚合协议. *软件学报*,2005,16(12):2106-2116. <http://www.jos.org.cn/1000-9825/16/2106.htm>
- [13] 沈波,张世永,钟亦平.无线传感器网络分簇路由协议. *软件学报*,2006,17(7):1588-1600. <http://www.jos.org.cn/1000-9825/17/1588.htm>



谢磊(1982—),男,江苏南通人,博士生,主要研究领域为分布式计算,传感器网络技术.



陈道蓄(1947—),男,教授,博士生导师,CCF高级会员,主要研究领域为分布式计算.



陈力军(1961—),男,博士,教授,主要研究领域为移动网络,分布式计算.



谢立(1942—),男,教授,博士生导师,CCF高级会员,主要研究领域为分布式计算,网络安全.