

基于分簇的无线传感器网络数据汇聚传送协议*

杨军^{1,2+}, 张德运¹, 张云翼¹, 王毅¹

¹(西安交通大学 电子与信息工程学院, 陕西 西安 710049)

²(宁夏大学 数学计算机学院 计算机科学系, 宁夏 银川 750021)

Cluster-Based Data Aggregation and Transmission Protocol for Wireless Sensor Networks

YANG Jun^{1,2+}, ZHANG De-Yun¹, ZHANG Yun-Yi¹, WANG Yi¹

¹(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

²(Department of Computer Science, School of Mathematics and Computer Science, Ningxia University, Yinchuan 750021, China)

+ Corresponding author: E-mail: bigdragonone@gmail.com

Yang J, Zhang DY, Zhang YY, Wang Y. Cluster-Based data aggregation and transmission protocol for wireless sensor networks. Journal of Software, 2010,21(5):1127-1137. <http://www.jos.org.cn/1000-9825/3534.htm>

Abstract: A cluster-based data aggregation and transmission protocol (CDAT) for wireless sensor networks (WSNs) is proposed. CDAT achieves a good performance in terms of lifetime by a clustering method of balancing energy consumption and data prediction transmission strategy. In clustering phase, the initial probability of node for cluster head election is derived from mathematical relation between application's seamless coverage ratio and numbers of required cluster heads, and residual energy and node degree are also employed to elect cluster head. In data aggregation phase, Cluster heads broadcast message for node joining and aggregate sampling data after clustering. According to the temporal correlation of sampling data, cluster heads send data to base station using prediction transmission strategy while satisfying transmission precision in the data transmission phase, and the lifetime of WSNs is prolonged with this strategy. Theoretical analysis and simulation results show that CDAT outperforms LEACH (low-energy adaptive clustering hierarchy) and PEGASIS (power-efficient gathering in sensor information systems) in terms of network lifetime by balancing energy consumption and decrease of transmission while satisfying desired Qos (quality of service) of application.

Key words: wireless sensor networks; cluster-based; seamless coverage ratio; data aggregation; prediction transmission

摘要: 提出了一种基于分簇的无线传感器网络数据汇聚传送协议 CDAT(a cluster-based data aggregation and transmission protocol for wireless sensor networks). CDAT 通过均衡能耗的分簇方法及数据预测传送机制,可以有效延长网络的生命期.在簇头选取阶段,利用应用期望的无缝覆盖率与所需簇头数的数学关系,限制节点竞选簇头的初始概率,并联合节点的度和剩余能量来选取簇头;在数据聚合阶段,簇头广播消息,接收所有加入该簇的成员节点,然

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2005AA121130 (国家高技术研究发展计划(863))

Received 2008-04-17; Accepted 2008-12-01; Published online 2009-03-31

后对簇内数据进行聚合;在数据传送阶段,利用数据在时间上的相关性,簇头在满足传送精度的要求下,采用预测传送机制将数据传送给基站,通过该机制,网络有效地减少了数据传送的次数.理论分析和模拟实验结果表明,CDAT协议在满足应用期望的服务质量要求下,通过均衡能耗、减少数据传送次数,使得网络生命期优于 LEACH(low-energy adaptive clustering hierarchy),PEGASIS(power-efficient gathering in sensor information systems)等协议.

关键词: 无线传感器网络;基于分簇的;无缝覆盖率;数据聚合;预测传送

中图法分类号: TP393 **文献标识码:** A

无线传感器网络(wireless sensor networks,简称 WSNs)作为一种全新的信息获取和处理模式,构筑起信息空间和物理空间的桥梁,广泛应用于国防军事、国家安全、环境监测、交通管理、医疗卫生、制造业、反恐抗灾等领域^[1,2].

WSNs 本质上是一类特殊的自组织(ad-hoc)网络,具有以下特质:1) 网络规模大,节点密度高;2) 节点的能量、计算和存储能力严格受限;3) 节点易失效,网络拓扑结构频繁变化;4) 节点没有统一的标识;5) 节点部署后位置一般不发生变化.一方面,传感节点依靠电池供电且通常工作在无人值守的环境下,补充能量非常困难,因此,能量优化要贯穿于通信、工作模式、任务协调、数据管理等各个层面;另一方面,无线传感器网络的设计是面向具体应用的,不同的应用有不同期望的服务质量要求,网络设计要首先考虑应用的服务质量.同时,传感节点所采集的感知数据又有很强的相关性,这种相关性使得数据在时间和空间上都可能存在冗余,减少这种冗余成为设计无线传感器网络考虑的一个重点.

基于分簇的数据收集方法经常应用在无线传感器网络的路由协议中,其优点是方便管理、增强网络的可扩展性及易于实现数据融合.为此,本文提出了一种基于分簇的数据汇聚传送协议 CDAT(a cluster-based data aggregation and transmission protocol for wireless sensor networks),根据应用期望的服务质量来计算簇头数,使用轮换簇头的方法来平衡能耗,通过预测性传送机制减少数据的传送次数,进一步降低网络的传送能耗,延长网络的生命期.

本文第 1 节介绍相关工作.第 2 节对网络模型进行描述并提出了问题.第 3 节给出 CDAT 协议的详细设计.第 4 节对协议的能耗作出分析.第 5 节进行模拟验证.最后总结全文.

1 相关工作

分簇路由协议一般会根据某种规则把 WSNs 节点集划分为多个子集,每个子集为一个簇,每个簇由一个簇头节点负责全局路由,其他节点通过簇头接收或发送数据.LEACH(low-energy adaptive clustering hierarchy)^[3]就是一个典型的分簇协议,通过等概率地随机循环选择簇头,将整个网络的能耗负载均衡地分配到每个传感器节点,从而达到降低网络能耗、延长网络生命周期的目的.但是,LEACH 没有考虑以下问题:1) 簇头数的选择没有考虑应用期望的服务质量问题;2) 没有对数据传送阶段作很好的优化;3) 簇头是随机选取的,不能保证簇头在网络中的均匀分布.

HEED(a hybrid, energy-efficient, distributed clustering approach)^[4]协议在簇头选择中考虑了节点的剩余能量,并以主从关系引入了多个约束条件作用于簇头的选择过程.HEED 在簇头选择标准及簇头竞争机制上都与 LEACH 不同,实验结果表明,HEED 分簇速度更快,能够产生更加分布均匀的簇头、更合理的网络拓扑.针对 LEACH 协议的缺点提出的 PEGASIS(power-efficient gathering in sensor information systems)^[5],把所有的传感器节点视为一个簇,根据节点的地理位置形成一条相邻节点之间距离最短的链,所有传感器节点把数据传送到簇头,簇头执行聚集操作并把结果传输到基站.TEEN(threshold sensitive energy efficient sensor network protocol)^[6]协议在协议中设置了硬、软两个阈值,以减少发送数据的次数.通过设置软、硬门限可以在精度和能耗之间取得合理的平衡,但存在以下缺陷:1) 如果感知的数据达不到硬门限的要求,节点就不会传送数据,即使网络中所有的节点能量全部耗尽,用户也不会知道;2) 协议不能反映网络数据整体的变化情况,在达不到硬门限的情况下,感知数据即使发生剧烈的变化也不能反映出来.

此外,文献[7]是一个使用 MPEG-2 视频压缩算法作预测的数据收集协议,文献[8]是一个基于双预测的目标跟踪协议,文献[9,10]使用数据聚合的方法来减少数据的时空相关性,文献[11]交替使用预测和压缩两种方式来减少数据的发送.但这些预测协议大都是平面路由协议,没有考虑到网络期望的服务质量,也没有考虑模型的动态更新等问题.

2 网络模型和问题描述

2.1 网络模型

本文假设 N 个无线传感器节点随机、均匀地分布在一个 $L \times L$ 的二维方形区域 A 内,该传感器网络具有以下性质:

- 1) 无线传感器网络为静态网络,节点部署后不再移动.
- 2) 节点采用布尔感知模型,节点的感知半径为 r ,且忽略边界因素影响.
- 3) 唯一的基站部署在区域 A 以外的一个固定位置,具有强大的能量和计算能力.
- 4) 节点是同构的,具有相同的初始能量,且节点的无线发射功率可调.
- 5) 应用对于网络有服务质量的要求,这里主要是指网络的无缝覆盖率和感知数据传送的误差门限.

定义 1. 有效覆盖面积.

在无线传感器网络的监测区域内,一块区域有可能被几个节点所覆盖,如图 1 所示.

节点 C_1 的有效覆盖面积 S_{C_1} 为节点 C_1 的覆盖范围 πr^2 减去重复覆盖区域 C 面积 S_C 的一半,即

$$S_{C_1} = \pi r^2 - \frac{1}{2} S_C \tag{1}$$

定义 2. 文献[12]证明,在图 2 所示的相邻节点覆盖拓扑中,当节点 A 所覆盖的区域是一个边长为 r 的正六边形时,其所覆盖的无缝面积最大,即

$$S_A = 6 \times \left(\frac{1}{2} \times r \times \frac{\sqrt{3}}{2} \times r \right) = \frac{3\sqrt{3}}{2} r^2 \tag{2}$$

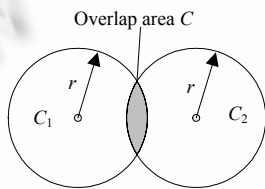


Fig.1 Illustration of effective coverage area

图 1 有效覆盖面积示意图

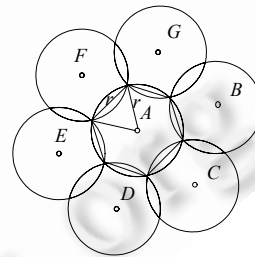


Fig.2 Illustration of maximal seamless coverage area

图 2 最大无缝覆盖面积示意图

2.2 问题描述

有效的分簇算法应该考虑以下问题:1) 应使用分布式的分簇算法,这样有利于节省网络能量和提高可扩展性;2) 为了实现节点能耗的均衡,簇头应尽可能均匀地分布在整个网络,否则会导致监测区域中某些节点能量消耗过快,造成感知空洞;3) 无线传感器网络大都与应用相关,应根据应用要求的服务质量来设计网络;4) 在满足误差要求的情况下,根据感知数据在时间或空间上的相关性进行优化,减少传送次数,进而减少传送能耗.目前,对于无线传感器网络分簇算法的研究大多集中在前两项,对于数据传送的优化涉及较少.为此,在满足一定网络服务质量(网络覆盖率、数据传送精度)的前提下,根据感知数据的时间相关性,本文提出了一种基于分簇的数据汇聚传送协议.

3 CDAT 协议设计

CDAT 协议和大部分分簇协议一样是按轮运行的,每轮分为簇头节点的选取、数据聚合和数据传送 3 个阶段.在簇头选取阶段,利用应用期望的无缝覆盖率为与所需要簇头数的数学关系限制节点竞选簇头的初始概率,以此获得期望的簇头数,并联合节点的剩余能量和最小度来选取簇头;在数据聚合阶段,簇头广播消息,接收所有加入该簇的成员节点,然后对簇内数据进行聚合;在数据传送阶段,利用数据的相关性,簇头在满足传送精度的要求下,采用预测传送机制进行数据传送.

3.1 计算簇头数

假定应用提出的网络无缝覆盖率为 η ,也就是说,每轮监测中网络需要若干个簇头(假定数量为 k),由这 k 个簇头形成的无缝覆盖面积占总的监测面积 $\|A\|$ 之比为 η .由公式(2)可知,单个簇头最大无缝覆盖的面积为 $3\sqrt{3}r^2/2$,监测区域 A 中任意一点 $a(x,y)$ 没有被随机选取的任一簇头 $C_i(1 \leq i \leq k)$ 覆盖的概率为

$$P_{a\text{-ncovered}} = 1 - (C_{i\text{-cover}} / \|A\|) = 1 - 3\sqrt{3}r^2/2L^2 \quad (3)$$

因此,在区域 A 中随机选取 k 个簇头,任一点 $a(x,y)$ 至少被其中一个簇头节点覆盖的概率 $P_{a\text{-covered}}$ 也就是网络无缝覆盖率,即

$$P_{a\text{-covered}} = \eta = 1 - P_{a\text{-ncovered}}^k = 1 - (1 - 3\sqrt{3}r^2/2L^2)^k \quad (4)$$

从而簇头数 k 为

$$k = \left\lceil \frac{\ln(1-\eta)}{\ln(1-3\sqrt{3}r^2/2L^2)} \right\rceil \quad (5)$$

3.2 簇头选择和簇内数据聚合

簇头选择机制采用分布式的算法,单个节点根据概率独自地选择是否成为簇头.为了使能耗均匀地分布到所有的节点,节点 i 成为簇头的概率由下式计算:

$$P_{i\text{-ch}} = \max \left(\frac{k}{N} \times \frac{E_{i\text{-current}}}{E_{origin}}, \frac{k}{N} \times \frac{E_{\min}}{E_{origin}} \right) \quad (6)$$

其中, k 由公式(5)确定, k/N 主要是为了限制初始簇头的数量, $E_{i\text{-current}}$ 表示节点当前的能量, E_{origin} 表示节点初始的能量, E_{\min} 表示簇头节点能量的最小阈值,当节点的能量小于 E_{\min} 时,该节点不再参与簇头的竞争. E_{\min} 值主要是由簇头参与一次数据的接收、聚合和发送能耗决定.本算法采用与 LEACH 协议相同的能量衰减模型,发送数据和接收的无线通信模型分别为

$$E_{tx}(h, d) = E_{tx\text{-elec}}(h) + E_{tx\text{-amp}}(h, d) = \begin{cases} hE_{elec} + h\epsilon_{fs}d^2, & d < d_0 \\ hE_{elec} + h\epsilon_{mp}d^4, & d \geq d_0 \end{cases} \quad (7)$$

$$E_{rx}(h) = E_{rx\text{-elec}}(h) = hE_{elec} \quad (8)$$

其中, E_{elec} 表示无线收发电路能耗, ϵ_{fs} 和 ϵ_{mp} 分别表示自由空间模型和多路衰减模型的放大器能耗, d_0 是常数, h 为要发送或接收的数据位数.因此, E_{\min} 可由下式计算:

$$E_{\min} = hE_{elec}(N/k - 1) + hE_{da}(N/k) + hE_{elec} + h\epsilon_{mp}d_{toBS}^4 \quad (9)$$

其中, E_{da} 为聚合单位数据所需能耗.对于任意节点 i ,在簇头选取阶段需要执行簇生成算法,主体是一段循环代码,循环的次数与节点的当前能量 $E_{i\text{-current}}$ 有关.为了更好地均衡网络能耗,在竞争簇头的过程中还考虑了节点的度,使得每一轮中具有最小度的节点成为簇头,算法如图 3 所示.

由上可知,节点 i 首先通过接收周围节点的广播消息来计算自己的度和 $P_{i\text{-ch}}$.循环开始时,节点 i 判断临时簇头集合 CH_set 是否为空,若不空,则选出度数最小的节点来,如果选出来的节点是自己且竞争簇头的概率为 1,则广播自己为正式簇头;如果竞争簇头的概率不为 1,则为临时簇头.如果临时簇头集合为空,且竞争簇头的概率为 1,则广播自己为正式簇头;否则,如果随机生成的数不大于竞争簇头的概率,则成为临时簇头.循环结束后,若

还没有成为簇头,且临时簇头集合不空,则加入以最小度节点为簇头的簇;否则认为自己没有被覆盖,广播自己为正式簇头。

分簇形成以后,簇头通过建立一个 TDMA 调度时隙表来控制本簇内的数据发送,并把这个调度广播给簇成员,这样就可以确保簇内各个成员无冲突地发送数据,并且节点在其他节点的发送时隙可以关闭电源以节省能量。簇头接收完簇成员数据后,对数据进行聚合,其输出的数据量取决于待聚合数据之间的相关性。

```

1.   broadcast(my_id, Ei-current)
2.   receive_msg (from neighbor in node i's cluster range)
3.   compute Pi-ch, degree
4.   Is_final_ch=FALSE
5.   While (Pi-ch-pre!=1){
6.     If (Ch_set!=null){
7.       My_ch=mindegree(Ch_set)
8.       If (My_ch==my_id){
9.         if (Pi-ch=1){
10.          Ch_msg(my_id,final_ch,degree)
11.          Is_final_ch=TRUE
12.        }else Ch_msg(my_id,tmp_ch,degree)
13.      }
14.    }
15.    else if (Pi-ch=1){
16.      Ch_msg(my_id,final_ch,degree)
17.      Is_final_ch=TRUE
18.    }
19.    else if (random(0,1)≤Pi-ch)
20.      Ch_msg(my_id,tmp_ch,degree)
21.      Pi-ch-pre=Pi-ch
22.      Pi-ch=min(2×Pi-ch,1)
23.    } //end of while
24.    If (Is_final_ch==FALSE){
25.      If (Ch_set!=null){
26.        My_ch=mindegree(Ch_set)
27.        Join_ch(Ch_id,my_id)
28.      }else Ch_msg(my_id,final_ch,degree)
29.      Ch_msg(my_id,final_ch,degree)
30.    } //end of if

```

Fig.3 Pseudo-Code for the cluster formation algorithm

图 3 簇生成算法伪码

3.3 预测数据传送机制

在数据传送阶段,根据网络服务质量选出的 k 个簇头,经过本簇内的数据聚合后要和基站进行数据通信,如何优化数据传送成为减少能耗的关键。

无线传感器网络中,节点的感知数据往往是时间上相关的,相互邻近的数据在数值上有一定的相关性,也就是时间序列数据。为了减少数据在时间上的冗余,在满足误差条件下,本文在数据传送阶段提出了预测数据传送机制。该机制的基本思路是,把预测数据与模型计算分离,基站根据历史数据建立模型,并把模型参数发送给各个簇头。单个簇头在与基站通信的过程中,通过预测和比较数据,只要预测数据和感知数据聚合值不大于根据应用要求设定的误差门限,就可以不用传送数据,而是在基站直接预测即可;为了适应感知数据的变化,传送机制还可以随着感知数据的变化作动态的调整。另外,在整个网络中只有监测对象所在簇的感知数据会发生显著变化,也只有这些簇头才会与基站发生数据传送,其他簇由于没有监测到对象,它们的感知数据变化不大,从而数据传送得到抑制。这样,通过把运算量较大的模型参数计算模块放在能量相对充足的基站,簇头只进行简单的预测和数据比较计算,从而极大地减少了簇头的计算和传输能耗,提高了网络的生存能力。

3.3.1 预测算法

一般的数据预测算法有很多,考虑到簇头节点的能量有限和数据在时间上的相关性,这里采用简单的中心化自回归(AR)模型^[13],并且假定感知数据序列是平稳的,即

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \tag{10}$$

其中, $X_i \in R$ 为簇头第 i 次的感知数据聚合值, ϕ_i 是回归系数, $\{a_t\}$ 是服从 $N(0, \sigma_a^2)$ 标准正态分布的白噪声序列, 公式(10)表明, 第 t 次感知数据聚合值可以用前 p 次数据的线性组合加上一个随机的白噪声来表示. 令

$$\begin{cases} Y = [X_{p+1}, X_{p+2}, \dots, X_N]^T \\ \phi = [\phi_1, \phi_2, \dots, \phi_p]^T \\ \varepsilon = [a_{p+1}, a_{p+2}, \dots, a_N]^T \end{cases} \tag{11}$$

$$X = \begin{pmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & & \vdots \\ X_{N-1} & X_{N-2} & \dots & X_{N-p} \end{pmatrix} \tag{12}$$

由公式(11)、公式(12)可得,

$$Y = X\phi + \varepsilon \tag{13}$$

由最小二乘法可得系数矩阵和方差的估计, 即

$$\hat{\phi} = (X^T X)^{-1} X^T Y \tag{14}$$

$$\hat{\sigma}_a^2 = \frac{1}{(N-p)} \sum_{t=p+1}^N (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} - \dots - \hat{\phi}_p X_{t-p})^2 \tag{15}$$

进而 t 时刻的 l 步前向预测值及方差分别为

$$\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \dots + \phi_p \hat{X}_t(l-p) \tag{16}$$

$$Var[e_t(l)] = (1 + G_1^2 + G_2^2 + \dots + G_{l-1}^2) \sigma_a^2 \tag{17}$$

其中,

$$\hat{X}_t(k) = \begin{cases} \hat{X}_t(k), & k \geq 1 \\ X_{t+k}, & k \leq 0 \end{cases}, G = \begin{cases} G_0 = 1 \\ G_j = \sum_{k=1}^j \phi_k' G_{j-k}, & j = 1, 2, \dots, \phi_k' = \begin{cases} \phi_k, & k \leq p \\ 0, & k > p \end{cases} \end{cases}$$

由上可知, 对于平稳时间序列的预测, 实质上就是根据所有已知的历史信息 X_t, X_{t-1}, \dots 对序列未来某个时期的发展水平 $X_{t+l} (l=1, 2, \dots)$ 做出估计.

3.3.2 数据传送的动态调整

为了使模型对于感知数据的变化具有自适应能力, 模型被设计成动态调整的. 这主要是通过基站和簇头的信息交互来完成, 即把模型计算与数据预测相分离, 由基站完成模型的计算和更新, 根据具体应用设定的比率决定是否把模型参数发给簇头节点; 由簇头节点完成感知数据的聚合和误差比较, 从而决定是否发送感知数据聚合值给基站. 具体算法如图 4、图 5 所示.

```

1. while (1){
2.     wait for data from CH_i;
3.     if (no data received){
4.         If (no_receiving data counts of CH_i > threshold α)
5.             continue; //No data sensed by CH_i and do nothing;
6.         else {
7.             predict l step data using predict model;
8.             continue;
9.         }
10.    }
11.    If (the ratio of number of incoming data > threshold β){
12.        update model and send parameter to CH_i;
13.        continue;
14.    } else save received data;
15.    } //end of while
    
```

Fig.4 Pseudo-Code for the algorithm of computing and updating of model on BS

图 4 基站模型计算和更新算法伪码

```

1. while (1){
2.     wait for model's parameters from BS;
3.     if (model's parameters received){
4.         update model;
5.     }
6.     do{
7.         gather sampling data and predict data;
8.     }while (prediction error ≤ threshold Errthreshold);
9.     send data to BS;
10. } //end of while
    
```

Fig.5 Pseudo-Code for the algorithm of aggregation and comparison on CH

图 5 簇头聚合和比较算法伪码

4 CDAT 协议能耗比较分析

本节对 CDAT 协议和未使用预测发送机制的一般分簇 GCP(general clustering protocol)协议进行能量消耗比较.假定传感器节点每次产生的单元数据长度为 h ,簇头每一轮聚合这些数据后,要发送给基站的数据量为 mh .

对于 CDAT 协议,簇头使用预测传送机制只需要发送 \tilde{m} 个数据给基站, \tilde{m} 为数据预测传送所需次数,根据第 3.3 节预测算法可知,若 $l=1$,则

$$\tilde{m} = \sum_{i=1}^m \delta(|\hat{X}_i(1) - X_{i+1}| - Err_{threshold}) \quad (18)$$

其中, $\delta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ 为示性函数, $Err_{threshold}$ 为网络服务质量给定的误差门限.现在计算每一轮中一个簇的能量消耗,理想情况下,簇的范围为一个覆盖半径 R 的圆, R 可由下式计算:

$$R = \sqrt{\|A\|/k\pi} = \sqrt{L^2 \ln(1 - 3\sqrt{3}r^2/2L^2)/\pi \ln(1 - \eta)} \quad (19)$$

k 由公式(5)计算,因此有

$$E[d_{toch}^2] = \int_0^{2\pi} \int_0^R \rho r^3 dr d\theta = \frac{L^2}{2\pi k} \quad (20)$$

则簇成员节点的能量花费为

$$E_{non-ch} = mh(E_{elec} + \varepsilon_{fs} d_{toch}^2) = mh \left(E_{elec} + \varepsilon_{fs} \frac{L^2}{2\pi k} \right) \quad (21)$$

对于簇头节点,能量消耗包括接收、聚合成员节点的感知数据和发送数据给基站,即

$$E_{ch} = mhE_{elec} \left(\frac{N}{k} - 1 \right) + mhE_{da} \frac{N}{k} + \tilde{m}h(E_{elec} + \varepsilon_{mp} d_{toBS}^4) \quad (22)$$

一个簇在一轮内消耗的能量为

$$E_{cdat-cluster} = E_{ch} + \left(\frac{N}{k} - 1 \right) E_{non-ch} = mh \left(\frac{N}{k} E_{elec} + \frac{N}{k} E_{da} + \varepsilon_{fs} \frac{L^2}{2\pi k} \right) + \tilde{m}h(E_{elec} + \varepsilon_{mp} d_{toBS}^4) \quad (23)$$

所以,总的能量消耗为

$$E_{cdat-total} = \tilde{k}E_{cluster} = \tilde{k} \left(E_{ch} + \left(\frac{N}{k} - 1 \right) E_{non-ch} \right) = mh\tilde{k} \left(\frac{N}{k} E_{elec} + \frac{N}{k} E_{da} + \varepsilon_{fs} \frac{L^2}{2\pi k} \right) + \tilde{m}h\tilde{k}(E_{elec} + \varepsilon_{mp} d_{toBS}^4) \quad (24)$$

其中, \tilde{k} 为没有受到数据传送抑制的簇头个数.对于 GCP 协议,一轮传送 m 个数据的能量消耗为

$$E_{general-clustering} = mh \left(\frac{N}{k} E_{elec} + \frac{N}{k} E_{da} + \varepsilon_{fs} \frac{L^2}{2\pi k} \right) + mh(E_{elec} + \varepsilon_{mp} d_{toBS}^4) \quad (25)$$

总的能量消耗为

$$E_{general-cluster} = kE_{clustering} = mh \left(NE_{elec} + NE_{da} + \epsilon_{fs} \frac{L^2}{2\pi} \right) + mhk(E_{elec} + \epsilon_{mp}d_{toBS}^4) \quad (26)$$

由于 $\tilde{k} < k, \tilde{m} < m$, 公式(24)和公式(26)相比较可得,

$$E_{general-total} - E_{cdat-total} = mh \left(NE_{elec} + NE_{da} + \epsilon_{fs} \frac{L^2}{2\pi} \right) \left(1 - \frac{\tilde{k}}{k} \right) + h(E_{elec} + \epsilon_{mp}d_{toBS}^4)(mk - \tilde{m}\tilde{k}) > 0 \quad (27)$$

数据传送能耗在 WSNs 的整体能耗中所占比重很大,因此使用预测机制对数据传送进行优化的 CDAT 协议,其能耗远小于未采用预测传送机制的一般分簇协议,后面的实验也证明了这一点.

5 模拟实验

模拟实验环境是监测区域大小为 200m×200m,随机部署节点总数为 200,基站的坐标为(100,300),监测区域的无缝覆盖率为 95%.为了验证数据预测传送的性能,选取赤道附近海洋表面空气温度^[14]作为感知数据,感知间隔为 10min.,模拟实验参数见表 1.

Table 1 Simulation parameters

表 1 模拟参数

Parameter	Value
Network size	200m×200m
Node number	200
Sensing range (m)	10, 15, 20
Transmission range (m)	25, 35, 40
BS position	(100,300)
Threshold distance (m)	75
E_{elec} (nJ/b)	50
ϵ_{fs} (pJ/b/m ²)	10
ϵ_{mp} (pJ/b/m ⁴)	0.001 3
E_{da} (nJ/b/m ²)	5
Data packet size (bits)	500
Broadcast packet size (bits)	200
Initial energy (J)	2
Qos ($\eta, Err_{threshold}$)	(0.95,0.1)
Sampling interval (min.)	10
Total of samples	4 000

5.1 应用期望的无缝覆盖率 η 与簇头数 k 的关系

为了验证公式(5)的正确性,对理论分析与模拟实验的结果进行比较.从图 6 可知,由公式(5)计算出来的簇头数与应用期望的无缝覆盖率的关系曲线与模拟实验所得结果比较吻合,这说明对于簇头数量的分析结论是正确的.

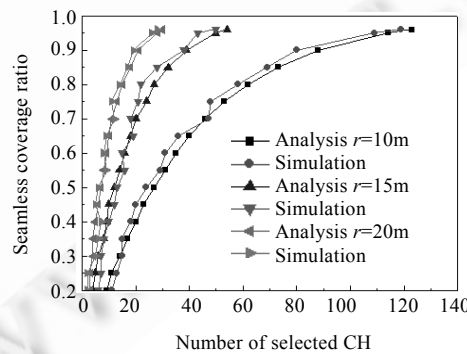


Fig.6 Relationship between numbers of CH and seamless coverage ratio: Analysis vs. simulation

图 6 簇头数与无缝覆盖率关系的理论分析和实验比较

5.2 网络生命期的比较分析

设置:监测区域面积为 200m×200m,部署的节点数量为 200,汇聚点的位置为(100,300),监测区域要求 95%被覆盖,误差门限为 0.1,采用文献[15]中关于网络生命期的定义,即网络初始运行到第 1 个节点死亡之间的时间长度 FND(first node die),在 LEACH 协议的基础上扩展了 CDAT, PEGASIS 和 DIRECT 协议.实验结果如图 7 所示.

从图 7 可以看出,CDAT 协议的网络生命期比 PEGASIS 和 LEACH 分别提高了 45%和 260%,这主要贡献于能耗均衡的分布式分簇算法和预测传送机制.通过限制簇头初始竞选概率,联合剩余能量和节点的度来选取簇头使得能耗负载均匀分布,预测传送机制大大减少了感知数据传送的次数,大部分长距离的数据传送能耗被避免(下节实验显示,80%左右的数据不用传送).PEGASIS 协议由于要获得所有节点的全局消息,在节点数量较多的情况下控制开销较大,所以性能受到一定的影响.不像 CDAT 协议,LEACH 不是通过应用的覆盖需求得到簇头数,而是簇头数量与节点数量成正比.因此,当节点数量较多时,其簇构造的开销也较大,同时也没有对数据传送进行优化,使得协议的性能下降.DIRECT 协议中所有的节点直接与基站通信,长距离的通信能耗导致网络生命期最短.

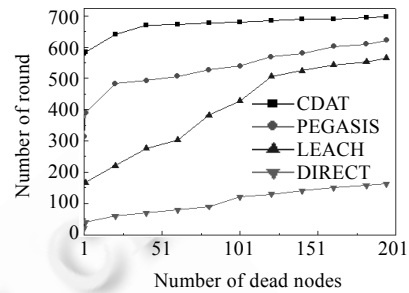


Fig.7 Lifetime of network

图 7 网络生命期

5.3 传送机制对性能的影响

预测数据传送机制能够有效地减少数据的传送次数,主要依赖于误差门限 $Err_{threshold}$ 及主动调节参数 β .为此,下面主要对这两个参数进行研究.

5.3.1 误差门限 $Err_{threshold}$ 的选择

由第 4 节能耗比较分析可知, $Err_{threshold}$ 的选择决定着感知聚合数据值传送次数的多少,对于预测传送机制的性能起到决定性作用.在考虑 $Err_{threshold}$ 时,不仅要考虑尽可能地少传数据,而且还要考察合适的预测误差范围,取得传送效率和预测精度的平衡.为此,使用数据发送率和平均相对误差两个指标来确定 $Err_{threshold}$ 的值.为了比较预测误差,我们在 GCP 协议上也加上了预测功能,使用跨距为 3 的移动平均方法来预测数据.数据发送率 R_{ds} 定义为因预测失败而发送的数据占全部要发送数据的比例,即

$$R_{ds} = \left(\left(\sum_{i=1}^n \delta(|X_i - \hat{X}_i| > Err_{threshold}) \right) / n \right) \times 100\% \quad (28)$$

平均相对误差 Err_{mrd} 定义为

$$Err_{mrd} = \frac{1}{n} \sum_{i=1}^n (|X_i - \hat{X}_i| / X_i) \quad (29)$$

由表 2 可知,当 $Err_{threshold}$ 为 0.05 时,GCP 仅能过滤 42%的冗余数据,而 CDAT 则能滤除 66%的冗余数据,CDAT 较之 GCP 的预测精度提高了 24%.当 $Err_{threshold}$ 为 0.1 时,CDAT 能够滤除 83%的冗余数据,较之 GCP 的预测精度提高了 22.8%.当 $Err_{threshold}$ 为 0.15 时,CDAT 能够过滤 89%的冗余数据,而 GCP 能够过滤 78%的冗余数据,但二者对应的 Err_{mrd} 也比较大.由此可知,当 $Err_{threshold}$ 较大时,虽然传送次数减少,但预测误差也较大,本文取 $Err_{threshold}$ 为 0.1,使得 CDAT 协议能够在传送次数和预测精度之间取得折衷.

Table 2 Transmission efficiency and prediction precision: CDAT vs. GCP

表 2 CDAT 和 GCP 的传送效率和预测精度比较

Algorithm	$Err_{threshold}$	Err_{mrd}	R_{ds}
GCP	0.05	0.000 73	0.58
CDAT	0.05	0.000 56	0.34
GCP	0.10	0.001 7	0.28
CDAT	0.10	0.001 3	0.17
GCP	0.15	0.002 7	0.22
CDAT	0.15	0.002 1	0.11

5.3.2 主动调节参数 β 对算法的影响

主动调节参数 β 为预测失败的次数与一轮数据总数之比,通过调整参数 β 使得能量消耗和预测精度达到平衡. β 与模型调整次数 MAT(model adjusting times)的关系如图 8 所示,从中可以看出,模型调整次数随着 β 的增大而减小. Err_{mrd} 随 β 发生变化的情况如图 9 所示,从中看出, β 在 0.01~0.10 之间, Err_{mrd} 随 β 的增大而减小,而当 $\beta>0.1$ 时, Err_{mrd} 随 β 的增大而增大.由此可见, β 在 0.1 左右,算法在调整次数和预测精度之间取得平衡,获得较好的性能.

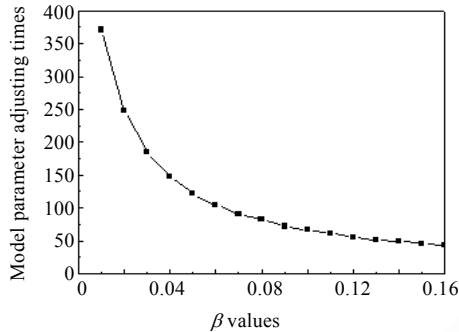


Fig.8 Impact of β to MAT

图 8 β 对模型调整的影响

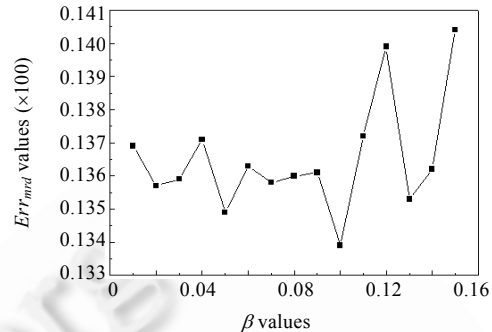


Fig.9 Impact of β to Err_{mrd}

图 9 β 对 Err_{mrd} 的影响

6 总 结

数据收集是无线传感器网络的主要功能,如何优化分簇以及如何优化数据传送成为影响网络生命期的关键.为此,本文基于应用期望的服务质量提出了一种基于分簇的数据汇聚传送协议,根据应用期望的无缝覆盖率来决定簇头数,使用轮换簇头的方法来平衡能量消耗,通过预测性传送机制减少数据传送次数,进一步降低了网络的传送能耗,延长了网络的生命期.本文对簇头数量、能量消耗和数据传送机制进行了详细的分析,通过大量的模拟实验证明了 CDAT 协议良好的性能.

下一步的工作就是研究如何抽象和分类应用的服务质量要求,在现实的环境下,根据 QoS 的分类考察 CDAT 协议的性能,以更好地优化网络的分簇和数据传送机制,使其更适合实际的应用需求.

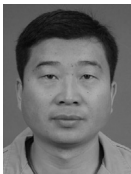
References:

- [1] Akyildiz IF, Su W, Sankarasubramanian Y, Cayirci E. Wireless sensor networks: A survey. *Computer Networks*, 2002,38(4): 393–402.
- [2] Li JZ, Gao H. Survey on sensor network research. *Journal of Computer Research and Development*, 2008,45(1):1–15 (in Chinese with English abstract).
- [3] 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.
- [4] Younis O, Fahmy S. HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Trans. on Mobile Computing*, 2004,3(4):366–379.
- [5] Lindsey S, Raghavendra CS. PEGASIS: Power-Efficient gathering in sensor information systems. In: Williamson DA, ed. *Proc. of the IEEE Aerospace Conf. Big Sky*: IEEE Computer Society, 2002. 1125–1130.
- [6] Manjeshwar A, Agrawal DP. TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. In: Werner B, ed. *Proc. of the 15th Int'l Parallel and Distributed Processing Symp.* San Francisco: IEEE Computer Society, 2001. 2009–2015.
- [7] Goel S, Imielinski T. Prediction based monitoring in sensor networks: Taking lessons from MPEG. *ACM Computer Communications Review*, 2001,31(5):82–95.

- [8] Xu Y, Winter J, Lee WC. Dual prediction-based reporting for object tracking sensor networks. In: Finin T, Ghidini C, Porta TL, Petrioli C, eds. Proc. of the 1st Annual Int'l Conf. on Mobile and Ubiquitous Systems: Networking and Services. Boston: IEEE Computer Society, 2004. 154–163.
- [9] He T, Blum BM, Stankovic JA, Abdelzaher T. AIDA: Adaptive application independent data aggregation in wireless sensor networks. ACM Trans. on Embedded Computing Systems, 2004,3(2):426–457.
- [10] Krishnamachari B, Estrin D, Wicker S. Modelling data-centric routing in wireless sensor networks. In: Lee D, Orda A, eds. Proc. of the IEEE INFOCOM. New York: IEEE Computer Society, 2002. 2–14.
- [11] Lazaridis L, Mehrotra S. Capturing sensor-generated time series with quality guarantees. In: Dayal U, Ramamritham K, Vijayaraman TM, eds. Proc. of the 19th Int'l Conf. on Data Engineering (ICDE 2003). Bangalore: IEEE Computer Society, 2003. 429–440.
- [12] Wang X, Yang Y, Zhang Z. A virtual rhomb grid-based movement-assisted sensor deployment algorithm in wireless sensor networks. In: Ni J, Dongarra J, Zheng Y, Guo GC, Wolfgang G, Jin H, eds. Proc. of the 1st Int'l Multi-Symp. on Computer and Computational Sciences. Hangzhou: IEEE Computer Society, 2006. 491–495.
- [13] Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis Forecasting and Control. 3rd ed., Upper Saddle River: Prentice Hall, 1994. 54–69.
- [14] McPhaden MJ. Tropical atmosphere ocean project. 2006. <http://www.pmel.noaa.gov/tao/index.shtml>
- [15] Chang JH, Tassiulas L. Maximum lifetime routing in wireless sensor networks. IEEE/ACM Trans. on Networking, 2004,12(4): 609–619.

附中文参考文献:

- [2] 李建中,高宏.无线传感器网络的研究进展.计算机研究与发展,2008,45(1):1–15.



杨军(1972—),男,宁夏吴忠人,博士生,副教授,主要研究领域为无线传感器网络.



张云翼(1975—),男,博士生,主要研究领域为无线传感器网络.



张德运(1941—),男,教授,博士生导师,主要研究领域为网络性能评价,网络安全.



王毅(1981—),男,博士生,主要研究领域为无线路由协议.