

# 传感器网络中一种分布式数据汇聚层次路由算法

林亚平<sup>1</sup>, 王 雷<sup>1</sup>, 陈 宇<sup>1</sup>, 张 锦<sup>1</sup>, 陈治平<sup>1</sup>, 童调生<sup>2</sup>

(1. 湖南大学计算机与通信学院, 湖南长沙 410082; 2. 湖南大学电气与信息工程学院, 湖南长沙 410082)

**摘 要:** 由于传感器网络具有能量约束, 低速率冗余数据和多对一传输等特点, 传统的端到端集中式路由算法一般不适合传感器网络. 提出了一种分布式数据汇聚层次路由算法, 该算法利用能量核的思想汇聚数据和减少传输到目的节点的信息. 模拟结果表明: 比较传统的端到端集中式路由算法, 该算法可以显著减少能量消耗; 与一般的数据汇聚算法相比, 该算法在保证能量消耗少的条件下, 具有复杂度低和可扩展性好的特点.

**关键词:** 传感器网络; 能量核; 数据汇聚; 路由算法

**中图分类号:** TP393 **文献标识码:** A **文章编号:** 0372-2112 (2004) 11-1801-05

## A Distributed Data-Centric Clustering Hierarchical Routing Algorithm for Sensor Networks

LIN Ya-ping<sup>1</sup>, WANG Lei<sup>1</sup>, CHEN Yu<sup>1</sup>, ZHONG Jin<sup>1</sup>, CHEN Zhi-ping<sup>1</sup>, TONG Tiao-sheng<sup>2</sup>

(1. College of Computer and Communication, Hunan University, Changsha, Hunan 410082, China;

2. College of Electric and Information Engineering, Hunan University, Changsha, Hunan 410082, China)

**Abstract:** Sensor networks have severe energy constraints, redundant low-rate data, and many-to-one flows. The traditional end-to-end routing algorithms are usually central and are not adapted for sensor networks. A distributed data-centric clustering hierarchical algorithm is proposed, which uses the energy-core concept to perform aggregation of data in the network and to reduce the amount of information that must be transmitted to the sink. Simulations show that, compared with the traditional end-to-end algorithm, this algorithm can significantly reduce energy dissipation. Compared with the data-centric algorithm without clustering, the data-centric clustering hierarchical algorithm offers better performance gains in complexity and scalability, while still being energy efficient.

**Key words:** sensor networks; energy-core; data-centric clustering; routing algorithm

### 1 引言

集数据采集、处理及通信功能于一体的无线集成网络传感器 (Wireless integrated network sensors)<sup>[1]</sup> 具有体积小、价格低等良好特性, 在环境与军事监控、地下、深水以及外层空间探索等许多方面都具有广泛的应用前景. 但外界环境的不确定性要求布置成百上千的传感器协同工作, 因此对由大规模无线集成网络传感器构成的传感器网络的研究正逐渐引起关注, 并被认为是本世纪的一项挑战性的研究课题<sup>[2~4]</sup>.

与传统网络相比, 传感器网络具有以下特性: (1) 节点分布稠密, 一般含有成百上千个节点; (2) 节点能量、存储空间及计算能力等资源十分有限; (3) 传输速率低, 数据是多对一的传输方式, 因此存在冗余. 因此, 传统网络中的路由算法不适合传感器网络, 必须针对其特性来研究新的路由算法. 目前提出的传感器网络路由协议主要有两类: 路由协议平面和层次路由协议. 典型的平面路由有 SAR (Sequential Assignment Routing)<sup>[5]</sup>, SPIN (Sensor Protocol for Information Negotiation)<sup>[6]</sup> 和定向传播路由 DD (Directed Diffusion)<sup>[7]</sup> 等. SAR 依据每条路径上的能量资源和 QoS 要求来决策路由, 为了避免因节点故障重新计算路由所带来的开销, SAR 采用多路径路由方案. SPIN 的主要思想是减少因采用洪泛广播而引入的开销. 由于传感器网

络主要目的是收集传感器采集的数据, 因此 DD 路由方案以数据为中心考虑路由, 和传统网络中基于地址的路由协议 (Address-Centric, 简称 AC) 不同, DD 采用基于数据的路由协议 (Data-Centric, 简称 DC). 文献 [8] 从节能的角度比较了 AC 和几种基于 DC 的路由算法, 分析和实验结果表明基于 DC 的路由算法具有更好的性能. 但这些算法均属于集中式算法, 要求每个传感器节点维护全局性的网络状态信息, 因此不具备良好的可扩展性和可维护性<sup>[6]</sup>. 平面路由协议要求所有传感器节点均具有路由功能, 导致节点因能量消耗过快而失效, 因此使得网络拓扑结构经常发生变化, 路由性能较低. 为了延长节点的生存期限, 提高网络的稳定性与路由性能, 研究人员提出了层次路由的思想. 在层次路由中只有一部分节点负责路由信息的转发, 因此可以达到延长大部分节点生存期限的目的. 典型的层次路由协议包括 LEACH (Low Energy Adaptive Clustering Hierarchy)<sup>[9]</sup>, TEEN (Threshold Sensitive Energy Efficient Sensor Network Protocol)<sup>[10]</sup> 和 PEGASIS (Power Efficient Gathering in Sensor Information Systems)<sup>[11]</sup> 等. LEACH 是一种基于簇的路由协议, 即本簇内所辖的节点以 TDMA 的方式分时向本簇的簇头传输数据, 数据经簇头汇聚和压缩后, 再向目的节点 (Sink) 发送. 通过这种方式减少大多数节点的能量消耗, 延长

收稿日期: 2003-06-18; 修回日期: 2004-06-06

基金项目: 国家自然科学基金 (No. 60272051); 湖南省自然科学基金 (No. 03JJ Y3098)

节点的生存期限. 其他的层次路由协议基本上属于 LEACH 的改进协议. 文献[12]给出了一种利用连通支配集中的支配点来构造 ad hoc 网络的虚拟骨干网络的分布式算法, 在该算法中, 当节点基于洪泛广播时, 消息只通过虚拟骨干网络进行扩散, 由此降低广播风暴出现的可能性, 达到节能的目的.

本文将数据汇聚和层次路由思想相结合, 在文献[12]中给出的连通支配集算法的基础上, 提出了“能量核”的概念, 并给出一种分布式数据汇聚层次路由算法. 算法的基本思想如下: 首先, 网络内的邻节点之间进行周期性的信息交换, 构造一个虚拟“能量核”, 使得“能量核”构成一个连通支配集; 然后, 当源节点向目的节点(Sink)发送数据时, 各源节点先在“能量核”中找一条到 Sink 的局部最短路径; 然后再将数据汇集至该路径中与其相邻的核心节点; 最后, 该核心在对其搜集到的所有数据进行合并处理后, 再沿找到的局部最短路径, 将汇聚数据发送到 Sink 节点. 算法具有如下良好性质: (1) “能量核”的构造使得算法能尽量利用具有较高剩余能量的节点进行数据汇聚, 提高了算法的路由性能; (2) 路由限制在“能量核”中进行, 与在整个规模网络中进行寻径相比减低了寻径时间复杂度; (3) 数据经过汇聚处理之后再转发到 Sink, 由此节约节点的能量; (4) 算法是分布式的, 和文献[8]中提到的算法相比, 该算法具有更好的可扩展性和可维护性.

## 2 传感器网络中的定向传播模型

### 2.1 数据汇聚

数据汇聚是指将从不同源节点接收到的数据进行合并压缩的处理过程, 主要包括以下几种方法: (1) 简单汇聚: 即将从不同源节点接收到的数据进行简单汇总; (2) 压缩汇聚: 即将从不同源节点接收到的数据进行同类项合并, 从而实现数据压缩; (3) 复杂汇聚: 即利用多输入单输出函数对从不同源节点接收到的数据进行聚合处理.

### 2.2 定向传播模型

图 1 所示为传感器网络通信结构. 传感器节点利用传感部件采集被监测对象的原始数据, 经过处理器部件处理后, 通过无线网络传输到目的节点 Sink, Sink 再通过因特网或卫星传输到用户数据管理中心. 定向传播模型中, 传感器网络中的节点不以地址作为标识 ID, 而是以节点可以提供的数据作为寻址依据. 即 Sink 在网络中广播以某种数据格式构成的消息询问它所感兴趣的监测数据, 这种消息简称为兴趣. 与这种兴趣匹配的节点(称为源节点)响应这种查询(称为事件), 并回送监测数据给 Sink. 因为节点分布的稠密性, 存在有多个节点匹配兴趣, 部分节点可能向 Sink 回送同样的监测数据. 为了节约节点的能量, 提高节点的生存期限, 这种数据可以在传输路径上先进行汇聚, 去掉冗余数据, 然后再传输给 Sink.

基于 AC 的算法中, 各传感器节点独立地计算到 Sink 的最短路径, 然后沿最短路径传输数据给 Sink, 如图 2(a) 所示. 若匹配兴趣的源节点分别为 Source1 和 Source2, 则在知道网络

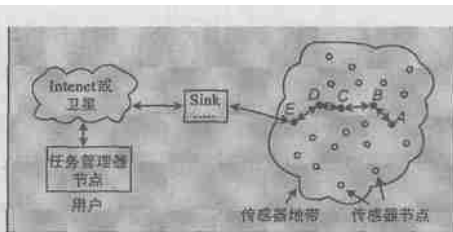


图 1 传感器网络的拓扑结构

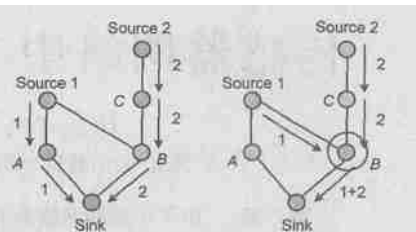


图 2(a) AC 路由算法; (b) DC 路由算法;

全局信息的前提下, Source1 和 Source2 可以分别利用 Dijkstra 算法计算到 Sink 的最短路径, 然后沿这些路径传输数据到 Sink, 在图 2(a) 的情况下, Source1 和 Source2 所经历的 hop 数之和为 5. 图 2(b) 描述了采用 DD 模型的 DC 算法, 其中 Source1 和 Source2 传输的数据首先在节点 B 处汇聚, 然后 B 将汇聚后的数据再统一传输给 Sink, 显然这种方法可以减小路径的 hop 数, 进而节约节点的能量. 对于同样的网络和源节点, DC 算法所经历的所有路径的 hop 数只有 4. 文献[8]给出了多种 DC 试探算法, 并证明了这些算法传输路径经历的 hop 数较 AC 算法要少, 而且通过模拟实验分析了这些算法的性能. 其中性能最优的是一种贪婪树算法 GIIDC (Greedy Incremental Tree Data Centric Protocol). 无论是 AC 算法还是以 GIIDC 为代表的 DC 算法, 都要求每个节点知道整个网络的拓扑结构, 属于集中式算法, 不具备良好的可扩展性. 由于传感器网络中节点的能量、计算和存储空间限制, 要求研究可扩展性好的分布式路由算法.

## 3 基于能量核的分布式数据汇聚层次路由算法

### 3.1 预备知识

定义 1 设图  $G = (V, E)$ ,  $G$  称为简单连通无向图, 当且仅当图  $G$  满足以下两个条件: (1)  $G$  为无自圈的、连通的无向图; (2)  $G$  中任意两个节点之间最多有一条边.

假定在传感器网络中各节点具有相同的有效通信距离. 称两个节点是相邻的即存在一条通信链路, 当且仅当这两个节点在彼此有效通信距离之内. 假定相邻节点之间的链路是对称的, 则传感器网络的拓扑结构可以看作是一个简单连通无向图  $G = (V, E)$ , 其中  $V$  为所有节点构成的顶点集合,  $E$  为所有链路构成的边集合.

定义 2(核) 图  $G$  的节点集  $C \subseteq V$  为核, 当且仅当节点集  $C$  满足以下条件:  $\forall p \in V \Rightarrow p \in C$  或  $p$  为  $C$  中的某个节点  $q$  的邻节点.  $C$  中的节点称为核心, 图  $G$  中不属于  $C$  的节点称为成员.

定义 3(连通核) 给定一个图  $G = (V, E)$ , 若图  $G$  的节点集  $C \subseteq V$  为满足如下条件的节点集合: 由  $C$  导出的子图是连通图, 且  $C$  是图  $G$  的一个核, 则称  $C$  为连通核.

定义 4(节点的能量) 假定传感器网络  $G$  中各节点均具有相同的初始能量  $E_0$ , 对  $G$  中的任意节点  $A$ , 我们用  $E_c(A)$  表示节点  $A$  的当前剩余能量.

由于传感器网络  $G$  中各节点的有效通信距离和初始能量相同, 而且  $G$  中的节点稠密分布, 因此可以假定  $G$  中任意相邻节点之间的距离相等, 即  $G$  中任意相邻节点之间转发一次相同大小的数据包时所需消耗的能量相同.

定义 5(节点的能耗) 假定在  $G$  中任意相邻节点之间转发一次大小为  $P$  的数据包时所需消耗的能量为  $P$ , 则从源节点  $A$  发送大小为  $P$  的数据包到  $Sink$  节点所需要的能耗  $EC(A, Sink) = P * d(A, Sink)$ , 其中  $d(A, Sink)$  为源节点  $A$  到  $Sink$  节点的  $hop$  数.

### 3.2 分布式“能量核”自动生成算法 DECA(Distributed Energy-Core Algorithm)

DECA 操作于定义 1 中的简单连通无向图  $G$ :

(1) 节点周期性地和邻节点交换以下信息:(i) 本节点是否为核心;(ii) 本节点之邻节点的相邻核心集合;(iii) 本节点的剩余能量. 通过此操作, 每个节点均可获知所有与它距离在 2 步 (2 hops) 以内的核心信息及本节点之邻节点的剩余能量信息.

(2) 初始时, 图  $G$  中的  $Sink$  节点的状态为核心, 而任意传感器节点  $p$  的状态均为成员. 在每个周期, 任意传感器节点  $p$  根据与本节点相距 2 步以内的核心信息及本节点之邻节点的剩余能量信息计算新的状态, 计算规则如下:

(a) 若  $p$  的所有邻节点  $p_1, p_2, \dots, p_n$  中无核心, 令  $E_c(p) = \text{Max}\{E_c(p_1), E_c(p_2), \dots, E_c(p_n)\}$ : (i) 若  $p$  的剩余能量大于等于任意  $p_k(k \in [1, n])$  的剩余能量, 即  $E_c(p) \geq E_c(p_k)$ , 则令  $p$  为核心; (ii) 否则, 即  $E_c(p) < E_c(p_k)$ , 则在满足条件  $E_c(p_r) = E_c(p_k)$  的节点中任选一节点  $p_r(r \in [1, n])$  为核心;

(b) 若节点  $p$  的 2 步以内的核心 (不包括节点  $p$ ) 不能构成一个连通集, 则令  $p$  为核心.

(c) 若节点  $p$  与某一个相邻节点构成“桥”(即没有一个连通的核心集可以互连它们), 且此邻节点不是核心, 则令  $p$  为核心.

(d) 若上述规则均不成立, 则  $p$  为成员.

图 3 具体给出算法 DECA 中规则 (2). (a). (i)、(2). (a). (ii)、(2). (b)、(2). (c) 的图例:

其中, 图 3(a1) 表示算法 DECA 中步骤 (2). (a). (1) 的情形; 图 3(a2) 表示算法 DECA 中步骤 (2). (a). (2) 的情形; 图 3(b) 表示算法 DECA 中步骤 (2). (b) 的情形; 图 3(c) 表示算法 DECA 中步骤 (2). (c) 的情形.

定理 1 若图  $G = (V, E)$  是简单连通无向图, 且每个节点维护的 2 步以内的节点状态信息是一致的, 则根据上述 DECA 算法得到的节点集  $C = \{p | p \text{ 为核心且 } p \in V\}$  是图  $G$  的一个连通核. 由于该连通核是基于局部最大能量而自动形成的, 因此称为“能量核”.

证明 首先, 由规则 (a) 可知, 图  $G$  中的每个节点要么为核心, 要么它的邻节点中至少存在一个核心, 即与某个核心相邻, 因此集合  $C$  是图  $G$  的一个核. 下面用归纳法证明  $C$  是连通的.

设  $p, q$  为  $C$  中的任意两个节点, 即  $p, q \in C$ . 由于  $E$  中的每条边的长度相等, 故为表述方便, 我们不妨设均为 1 个单位长度, 因此  $p, q$  之间的距离即为  $p, q$  之间的最短路

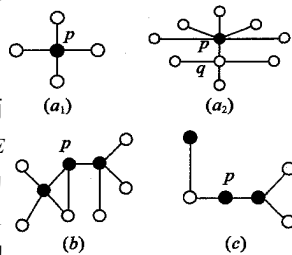


图 3 DECA 算法中的规则

径的长度, 记为  $d(p, q)$ . 由于图  $G$  是连通的, 因此  $d(p, q)$  为有限整数.

(1) (i) 若  $d(p, q) = 1 \Rightarrow p, q$  相邻, 故它们直接可达  $\Rightarrow$  是连通的; (ii) 若  $d(p, q) = 2$ , 即存在  $G$  中的路径  $(p, r, q)$ , 若  $p, q$  通过  $r$  不可达, 则由算法中的步骤 (b) 可保证节点  $r$  使得  $p, q$  通过  $r$  可达  $\Rightarrow$  是连通的; (iii) 若  $d(p, q) = 3$ , 即存在  $G$  中的路径  $(p, r_1, r_2, q)$ , 若  $p, q$  通过  $r_1, r_2$  不可达, 则由算法中的步骤 (c) 可保证节点  $r_1$  或  $r_2$ , 不妨设  $r_1$  对  $(r_1, r_2, q)$ , 由算法中的步骤 (b) 可保证节点  $r_2 \Rightarrow p, q$  通过  $r_1, r_2$  可达  $\Rightarrow$  是连通的.

(2) 假设  $d(p, q) = m(m > 3)$  时  $C$  是连通的.

(3) 对  $d(p, q) = m + 1$ , 即存在  $G$  中的路径  $(p, r_1, r_2, \dots, r_m, q)$ , 若  $p, q$  通过  $r_1, r_2, \dots, r_m$  不可达, 则由  $C$  是图  $G$  的一个核  $\Rightarrow r_2$  或  $r_3$  与  $r_1$  中的某个核心  $r$  相邻  $\Rightarrow d(p, r) \leq 3$ , 由证明的步骤 (1) 可知节点  $p, r$  通过  $r$  可达, 而  $d(r, q) \leq m$ , 由归纳假设可知  $r, q$  通过  $r$  可达  $\Rightarrow$  是连通的.

综合以上证明步骤可知  $C$  是连通的.

### 3.3 层次路由算法 DDCH

假定算法 DECA 形成的能量核  $C$  中, 任意核心节点通过邻节点之间的信息交换, 维护有到  $Sink$  节点  $R$  的距离信息, 由此核心节点之间可利用类似距离向量路由算法<sup>[13,14]</sup>的方法在核心节点之间进行路由, 寻找到  $Sink$  节点的最佳路径. 但和距离向量路由算法所不同的是相邻核心节点之间交换的信息只有到  $Sink$  节点的路由信息, 因此这种路由寻径算法不但保留了距离向量算法利用局部信息分布执行的特点, 而且可少传输的信息量. 设传感器网络中有  $n$  个源节点  $S_1, S_2, \dots, S_n$  发送数据给  $Sink$  节点  $R$ , 算法 DDCH 的具体步骤如下:

(1) 源节点  $S_1, S_2, \dots, S_n$  首先通过上述能量核中相邻核心节点之间的信息交换, 在能量核  $C$  中分布式地计算其到  $R$  的最优路径  $L_1, L_2, \dots, L_n$ ; 具体方法如下: (i) 若  $S_j$  不是核心, 则  $S_j$  通过其相邻核心节点, 利用上述能量核中相邻核心节点之间的信息交换, 在能量核  $C$  中分布式地计算其到  $R$  的最优路径  $L_j$ , 其中  $j \in [1, n]$ ; (ii) 若  $S_j$  是核心, 则  $S_j$  直接通过上述能量核中相邻核心节点之间的信息交换, 在能量核  $C$  中分布式地计算其到  $R$  的最优路径  $L_j$ , 其中  $j \in [1, n]$ ;

(2) 源节点  $S_1, S_2, \dots, S_n$  在找到其到  $R$  的最优路径  $L_1, L_2, \dots, L_n$  之后, 再利用如下方法进行数据发送: (i) 若  $S_j$  不是核心, 则  $S_j$  首先将数据发送给路径  $L_j$  中其所属的相邻核心  $N_j$ ; 若  $S_j$  是核心, 则  $S_j$  即为  $N_j$ ; 其中  $j \in [1, n]$ ; (ii)  $N_j$  对收到的从源  $S_1, S_2, \dots, S_m$  发送来的数据信息进行汇聚处理后, 将汇聚后的数据沿路径  $L_1$  发送到  $R$ , 其中  $t_k \in [1, n], k \in [1, m]$ .

由上述 DECA 和 DDCH 算法的具体描述可知: 上述“能量核”的构造算法 DECA 显然是一种分布式的算法. 另外, 和传统的 Ad hoc 网络不同, 在传感器网络中节点的移动性相对较少, 其拓扑变化主要源自于节点因能量损耗而失效以及  $Sink$  节点的动态移动性, 因此 DDCH 利用由具有局部最大剩余能量的节点所构成的“能量核”来进行路由, 适合于传感器网络的拓扑结构和节点动态变化的性质.

#### 4 算法性能分析与模拟

**引理 1** 算法 DDCH 所需的从  $n$  个 Source 节点到 Sink 节点的最短路径 (hop 数) 之和要小于等于算法 AC 所需的 hop 数之和。

**证明** 对任意 Source 节点  $S$ , 假设由算法 AC 求得的其到 Sink 节点  $R$  的最短路径 (hop 数) 为  $L_A(S, R)$ , 由算法 DDCH 求得的其到  $R$  的最短路径 (hop 数) 为  $L_D(S, R)$ , 下面首先用归纳法证明  $L_D(S, R) = L_A(S, R)$ :

(1) 若  $L_A(S, R) = 1 \Rightarrow S$  到  $R$  的距离为 1, 即  $S$  与  $R$  相邻, 又  $R$  为核心, 故  $L_D(S, R) = 1 \Rightarrow L_D(S, R) = L_A(S, R)$ ;

若  $L_A(S, R) = 2 \Rightarrow S$  到  $R$  的距离为 2, 由于  $R$  为核心, 故由能量核的自动生成算法 DECA 的步骤 (2) 可知:  $L_D(S, R) = 2 \Rightarrow L_D(S, R) = L_A(S, R)$ ;

(2) 假设对  $L_A(S, R) = m \Rightarrow L_D(S, R) = L_A(S, R)$ ;

(3) 对  $L_A(S, R) = m + 1 \Rightarrow S$  到  $R$  的距离为  $m + 1 \Rightarrow$  存在  $S$  的邻节点  $B$ , 使得在算法 AC 中节点  $B$  到  $R$  的最短路径  $L_A(B, R) = m$ , 由证明中步骤 (2) 的归纳假设可知:  $L_D(B, R) = m$ . 再由能量核的自动生成算法 DECA 可知: 若  $B$  不为核心, 则存在节点  $B$  的相邻核心  $C$ , 使得  $L_D(C, R) = m - 1$ ; 若  $B$  为核心, 则  $L_D(B, R) = m$ .

(1) 若  $B$  不为核心, 则对节点  $S, B, C$ , 由能量核的自动生成算法 DECA 的步骤 (2) 可知: 必存在  $S, C$  的共同邻节点  $D$ , 使得  $D$  为核心且  $S$  经过  $D$  与  $C$  连通, 故  $L_D(S, R) = L_D(S, C) + L_D(C, R) = m + 1 \Rightarrow L_D(S, R) = L_A(S, R)$ ;

(2) 若  $B$  为核心, 则  $L_D(S, B) = 1 \Rightarrow L_D(S, R) = L_D(S, B) + L_D(B, R) = m + 1 \Rightarrow L_D(S, R) = L_A(S, R)$ ;

综上所述可知  $L_D(S, R) = L_A(S, R)$  成立. 因此, 当 DDCH 中  $n$  个 Source 节点经最短路径到 Sink 节点  $R$  时它们所需经过的相邻核心节点两两互异, 则显然算法 DDCH 中节点  $A_1, \dots, A_m$  到 Sink 节点  $R$  的最短距离之和等于算法 AC 中节点  $A_1, \dots, A_m$  到 Sink 节点  $R$  的最短距离之和。

假设算法 DDCH 中有  $m$  ( $1 < m \leq n$ ) 个非核心的 Source 节点  $A_1, \dots, A_m$  均与核心  $C$  相邻, 且它们到 Sink 节点  $R$  的最短路径均需经过核心  $C$  时, 则由上面已经证明过程可知算法 AC 中节点  $A_1, \dots, A_m$  到 Sink 节点  $R$  的最短距离均为  $T + 1$ , 其中  $T$  为算法 DDCH 中核心  $C$  到 Sink 节点  $R$  的最短距离. 显然算法 DDCH 中节点  $A_1, \dots, A_m$  到 Sink 节点  $R$  的最短距离为  $T + m$ , 而算法 AC 中节点  $A_1, \dots, A_m$  到 Sink 节点  $R$  的最短距离为  $m^*(T + 1)$ , 易证  $T + m < m^*(T + 1)$ , 因此在该情形下算法 DDCH 中节点  $A_1, \dots, A_m$  到 Sink 节点  $R$  的最短距离之和小于算法 AC 中节点  $A_1, \dots, A_m$  到 Sink 节点  $R$  的最短距离之和。

综上所述, 从而可知引理 1 的结论成立。

**定理 2** 算法 DDCH 的路由性能优于算法 AC 的路由性能。

**证明** 由于算法 DDCH 只需在能量核中进行寻径,

而由能量核的自动生成算法 DECA 可知: 能量核的大小要远小于整个网络的大小, 因此算法 DDCH 的寻径时间复杂度要远小于算法 AC 的寻径时间复杂度; 再结合引理 1 可知: 算法 DDCH 所需的最少跳数之和要小于算法 AC 所需的最少跳数之和; 另外算法 AC 是集中式的, 而算法 DDCH 是分布式的; 因此算法 DDCH 的路由性能要优于算法 AC 的路由性能。

**定理 3** 假设算法 DDCH 中  $n$  个非核心的 Source 节点均与核心  $C$  相邻, 且它们到 Sink 节点  $R$  的最短路径均需经过核心  $C$  时, 则算法 DDCH 所需的从这  $n$  个 Source 节点到 Sink 节点的最短路径 (hop 数) 之和等于算法 GTDC 所需的 hop 数之和。

**证明** 假设算法 DDCH 中核心  $C$  到 Sink 节点  $R$  的最短距离为  $L_D(C, R)$ ; 则算法 DDCH 所需的从这  $n$  个 Source 节点到 Sink 节点的最短路径 (hop 数) 之和  $L_D = L_D(C, R) + n$ . 而这  $n$  个 Source 节点中到  $R$  距离最短的节点  $A$  到  $R$  的最短距离  $L(A, R) = L_D(C, R) + 1$ ; 而其余  $n - 1$  个节点到  $A, R$  构成的树的最短距离均为 1, 因此算法 GTDC 所需的 hop 数之和  $L_G = L(A, R) + n - 1 = L_D(C, R) + 1 + n - 1 = L_D(C, R) + n \Rightarrow L_G = L_D$ ; 故定理 3 的结论成立。

为了比较算法 DDCH 的路由性能, 我们依据文献 [8] 进行了模拟实验: 首先根据给定的有效通信距离 CR (Communication Radius) 在  $100 * 100$  的区域内随机生成一个含有 100 个坐标不同的节点的连通图, 并对每个节点随机生成 50 个单位以内的能量, 然后对传感器网络中的两种典型情况进行了模拟: 1) 情况 1: 随机抽取 10% 的节点 (1 个 Sink 节点, 9 个 Source 节点) 进行模拟实验; 2) 情况 2: 随机抽取 1 个 Sink 节点, 再随机选取占整个区域 10% 大小的子区域 (即一个  $10 * 10$  的区域) 中的节点作为 Source 节点进行模拟实验. 实验数据均为执行算法 100 次, 选取其平均值为最后结果。

图 4 和 5 分别给出了上述两种条件下, DDCH 算法中核心节点数占总节点数比例的实验数据; 由图 4, 5 可知: 能量核的大小和 CR、网络的节点稠密程度成反比, 其大小仅为网络节点总数的 20% ~ 40%, 因此能量核的构造可较大程度地减低算法路由的时间复杂度, 同时由于能量核的大小相比整个网络而言要小得多, 因此能量核的构造也使得新算法较算法 AC 和 GTDC 具有更好的可扩展性. 图 6 和 7 分别给出了 DDCH、AC 和 GTDC 算法从所有 Source 节点到 Sink 节点在情况 1、2 下时间复杂度的对比数据, 其中各算法的时间复杂度计算为各算法所需的执行时间。

由图 6, 7 可知 DDCH 的时间复杂度小于 AC 和 GTDC 的时间复杂度, 并且是随 CR 增大而收敛的. 通过实验过程分析可知, 对算法 AC 和 GTDC 而言, 由于其是集中式的, 需要各节点维护全局信息, 而网络中各节点的邻节点数是随 CR 增

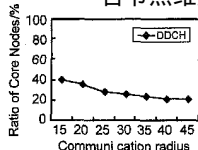


图 4 情况 1 中 DDCH 中核心节点数比例

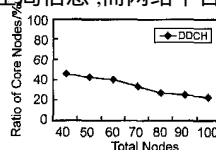


图 5 情况 2 中 DDCH 中核心节点数比例

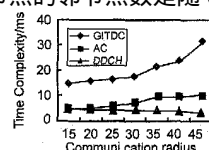


图 6 情况 1 中 DDCH, AC, GTDC 的时间复杂度

大而增多的,从而使得算法在寻找从所有 Source 节点到 Sink 节点最短路径的时间开销也随之增大,因此算法 AC 和 GITDC 的时间复杂度随  $CR$  增大而增大.对算法 DDCH 而言,由图 4、5 已知能量核的大小和  $CR$ 、网络的节点稠密程度成反比,而 DDCH 只需要在能量核中进行寻径,从而使得算法 DDCH 的时间复杂度是随  $CR$  增大而收敛的.

图 8 和 9 分别给出了 DDCH、AC 和 GITDC 算法从所有 Source 节点到 Sink 节点在情况 1、2 下经历的跳数 (hops) 的对比数据.由图 8、9 可知:DDCH 性能优于 AC;当  $CR$  较大时 DDCH 和 GITDC 所需的性能接近.特别是对情况 2,由图 9 可知 DDCH 所需的 hop 数与 GITDC 所需的 hop 数基本相同.这主要是因为情况 2 的情形将使得选取的 Source 节点接近定理 3 中的条件,从而使得算法 DDCH 所需的 hop 数接近算法 GITDC 所需的 hop 数.

由图 8、图 9 以及引理 1、定理 3 可知:算法 DDCH 所需的 hop 数要小于 AC 所需的 hop 数;且接近算法 GITDC 所需的 hop 数;再结合定义 5,显然有:算法 DDCH 所需的能耗小于 AC 所需的能耗;另外,由图 9、定理 3、以及定义 5 还可知,当 Source 节点相对集中时,算法 DDCH 所需的能耗接近于 GITDC 所需的能耗.因此算法 DDCH 具有良好的节能性.

最后,基于可扩展性和可维护性方面考虑,由于 DDCH 是分布式的,只需要能量核中的节点维护到 Sink 节点的距离信息;而算法 GITDC 和 AC 均是集中式的,需要所有节点维护整个网络的全局信息,因此新算法具有更好的可扩展性和可维护性.综上所述可知,DDCH 是一种具有较好性能的适合传感器网络特性的路由算法.

## 5 结束语

综合考虑传感器网络的节点分布的稠密性和能量等资源的有限性的特点,提出了一种分布式数据汇聚层次路由算法,并对其进行了模拟实验.算法利用能量核的思想,较好地解决了传感器网络因节点稠密分布和节点能量有限而带来的路由问题.实验结果表明,和算法 AC 和 GITDC 相比,新算法具有较好的路由性能.对传感器网络的研究刚刚兴起,在很多方面都需要做一些富有开创性的工作.如传输层的协议的研究;网络层中组播和 QoS 路由算法的研究等.这些也是我们正在进一步深入研究的课题.

## 参考文献:

- [1] Agre J, Clare L. An integrated architecture for cooperative sensing networks[J]. IEEE Trans On Computers, 2000, 33(5): 106 - 108.
- [2] 任丰原, 黄海宁, 林闯. 无线传感器网络[J]. 软件学报, 2003, 14(7): 1282 - 1291.
- [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[J]. 软件学报, 2003, 14(10): 1717 - 1727.
- [4] Estrin D, Gvindan R, Heideman J, Kumar S. Next century challenges: Scalable coordination in sensor networks[A]. Victor Bahl. 1999 5th ACM/ IEEE Annual International Conference on Mobile Computing and Networking Proceedings [C]. Seattle, Washington, USA: ACM, 1999.

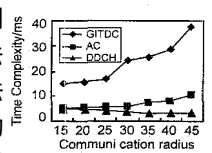


图 7 情况 2 中 DDCH, AC, GITDC 的时间复杂度

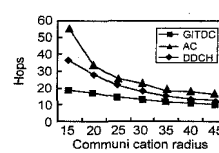


图 8 情况 1 中 DDCH 与 AC, GITDC 的跳数比较

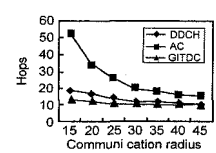


图 9 情况 2 中 DDCH 与 AC, GITDC 的跳数比较

263 - 270.

- [5] Akyildiz I F, Su W, Sankarabramanian Y, Cayirci E. A survey on sensor networks [J]. IEEE Communications Magazine, 2002, 40(8): 102 - 114.
- [6] Wendi Rabiner Heinzelman, Joanna Kulik, Hari Balakrishnan. Adaptive protocols for information dissemination in wireless sensor networks[A]. Victor Bahl. 1999 5th ACM/ IEEE Annual International Conference on Mobile Computing and Networking Proceedings [C]. Seattle, WA, USA: ACM, 1999. 174 - 185.
- [7] C Intanagonwiwat, R Govindan, D Estrin. Directed diffusion: A scalable and robust communication paradigm for sensor networks[A]. Raymond Pichholtz. 2000 6th ACM/ IEEE Annual International Conference on Mobile Computing and Networking Proceedings [C]. Boston, MA, USA: ACM, 2000. 56 - 67.
- [8] B Krishnamachari, D Estrin, S Wicker. Modelling data-centric routing in wireless sensor networks [A]. Parviz Kermani. 2002 IEEE Infocom Proceedings [C]. New York, NY, USA: IEEE Computer Society, 2002. 42 - 49.
- [9] W Heinzelman, J Kulik, H Balakrishnan. Negotiation based protocols for disseminating information in wireless sensor networks [J]. ACM Wireless Networks, 2002(8): 169 - 185.
- [10] Manjeshwar A, Agrawal D P. TEEN: A routing protocol for enhanced efficiency in wireless sensor networks [A]. Afonso Ferreira. 2001 15th Parallel and Distributed Processing Symposium Proceedings [C]. San Francisco, CA, USA: IEEE Computer Society, 2001. 2009 - 2015.
- [11] Lindsey S, Raghavendra C S. PEGASIS: Powerful-efficient gathering in sensor information systems [A]. Greg Richardson. 2002 IEEE Aerospace Conference Proceedings [C]. Big Sky, MT, USA: IEEE Computer Society, 2002. 9 - 16.
- [12] 彭伟, 卢锡诚. 一个新的分布式最小连通支配集近似算法 [J]. 计算机学报, 2001, 24(3): 254 - 258.
- [13] Bellman R E. Dynamic Programming [M]. Princeton, NJ: Princeton University Press, 1957.
- [14] Ford L R Jr, Fulkerson D R. Flows in Networks [M]. Princeton, NJ: Princeton University Press, 1962.

## 作者简介:



林亚平 男, 1955 年生于湖南邵阳, 湖南大学计算机与通信学院教授, 博士生导师, 主要研究方向为计算机通信网络, 机器学习.