

# 基于连通支配集的无线传感网 Top- $k$ 查询最优支撑树研究

邬海琴, 王良民

(江苏大学计算机科学与通信工程学院, 江苏镇江 212013)

**摘 要:** 构建底层逻辑树能有效降低集中式 top- $k$  查询带来的巨大通信开销, 针对现有逻辑树都以固定汇聚节点为根节点, 导致其附近节点能耗太大、过早死亡的问题, 本文在无固定汇聚节点的网络背景下, 基于连通支配集, 提出一种能耗均衡的 top- $k$  查询最优支撑树构建方法, 综合节点能量、度数以及与邻节点通信开销, 选取能量代价小的作为支配节点负责查询中间数据处理, 在每次查询中, 节点基于地理位置 ID 轮流作为根节点, 有效均衡节点的能耗. 仿真实验表明, 与其他逻辑拓扑树相比, 基于最优支撑树的 top- $k$  查询具有相近的查询时间, 但其平均每轮查询能耗更小, 多次查询后各节点能耗达到均衡, 有效延长了网络生命周期.

**关键词:** 无线传感器网络; top- $k$  查询; 连通支配集; 能耗均衡; 最优支撑树

**中图分类号:** TP393      **文献标识码:** A      **文章编号:** 0372-2112 (2017)01-0119-09

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2017.01.017

## Connected Dominating Set Based Support-Tree for Top- $k$ Query in Wireless Sensor Networks

WU Hai-qin, WANG Liang-min

(School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China)

**Abstract:** Constructing underlying logic trees can effectively reduce the large communication overhead for centralized top- $k$  query. In this paper, an energy-balanced optimal support-tree is proposed to solve the problem of earlier death for nodes near sink in sensor networks without fixed sink. This method aggregates node energy, degree and communication overhead, and nodes with low energy cost are chosen to be dominating nodes for data processing based on connected dominating set. In each query, nodes take turns to be root by location-based ID, which effectively balances the energy cost. The simulation experiment shows that top- $k$  query based on support-tree outperforms the existing trees by similar query time, but lower average energy cost per round and balanced consumption among all sensor nodes after continuous top- $k$  queries, which effectively prolong the network lifetime.

**Key words:** wireless sensor networks; top- $k$  query; connected dominating set; energy-balanced; optimal support-tree

### 1 引言

无线传感器网络(Wireless Sensor Network, WSN)是由部署在监测区域内的大量传感器节点,通过无线通信方式形成的特殊 ad hoc 网络,节点不具有移动性且能量较少,计算、通信能力较弱.近年来,WSN 在多个领域得到广泛应用,如环境污染监测<sup>[1]</sup>、动物保护以及目标跟踪<sup>[2,3]</sup>等,而 top- $k$  查询作为传感网查询应用的重要内容,用户可根据自身需要查询监测区前  $k$  个最大(或

最小)监测值,如温度、污染物指数等,从而起到森林预警、判断污染程度较高地区的作用.由于无线传感器节点通常部署在无人看守的恶劣环境中,因此对能量严格受限的传感器网络,降低网络能耗、延长网络寿命成为 top- $k$  查询的关键问题.

Top- $k$  查询主要从两方面减少网络能耗,一是构建底层逻辑拓扑树,主要以数据查询节点(sink)作为根节点,在节点间建立父亲—孩子关系,先由 sink 广播用户查询至叶子节点,各叶子节点将自身感知值发送至其

父节点汇聚,若数据量大于  $k$ ,则先在本地进行 top- $k$  查询,再将查询结果发送给父节点,直至 sink. 与集中式查询相比,大大减少了通信开销,但也由此带来 sink 附近节点负载过多,能耗太大过早死亡的问题,这些“瓶颈节点”决定了网络寿命;另一方面是研究高效的 top- $k$  查询算法,如文献[4]提出的基于 filter 的算法,其主要思想是将下一轮最有可能成为 top- $k$  结果的数据发送给 sink,减少无效的数据传输,从而降低通信开销.

延长网络寿命,不仅要减少网络的整体能耗,更应使网络中各节点能耗达到均衡. 在查询过程中,能量消耗快的节点将成为整个网络的“瓶颈节点”,节点能耗不均会导致某些节点提前死亡,因此,构建能耗均衡的逻辑树对延长网络寿命起着至关重要的作用.

本文在连通支配集理论基础上,提出了能耗均衡的最优支撑树构建方法,其主要特点在于:

(1)综合考虑了节点度数、能量和通信代价三个因素选择低能耗连通支配节点,负责其被支配节点的数据查询处理和转发,而高能耗节点作为叶子节点参与较少通信,有效均衡了节点能耗,避免“瓶颈节点”过早死亡.

(2)所选择的低能耗连通支配节点虽不是最小连通支配集,增加了单轮查询能耗,但通过协调节点间的通信量,延长了单个节点寿命,最终延长整个网络寿命,从而降低了平均每轮查询的能耗.

(3)每轮 top- $k$  查询由节点基于地理位置 ID 轮流作为 sink,打破了根节点唯一的传统,有效延长了网络寿命.

## 2 相关工作

Top- $k$  查询在多种数据环境中的应用已有广泛研究,如多媒体数据库<sup>[5]</sup>,关系数据库<sup>[6]</sup>以及传感器网络<sup>[7,8]</sup>等. 在分布式传感网环境中,top- $k$  查询主要分为两类:一类称为多维 top- $k$  查询<sup>[9]</sup>,旨在找到  $k$  个排名最高的数据点,每个数据点有多个属性值,根据用户查询函数  $f$  计算排名;另一类则是找出  $k$  个最高感知值,假设每个传感器节点每次只产生一个感知值,又称单一 top- $k$  查询.

对于单一 top- $k$  查询,目前已有一些底层逻辑树构建算法<sup>[10-12]</sup>和 top- $k$  查询算法<sup>[13-16]</sup>,Wu<sup>[4]</sup>等人提出的基于范围的 FILA 算法以 SPT 作为逻辑树,SPT 以 sink 为根节点,其余节点均有到 sink 的最短路径,大大减少了集中式查询带来的巨大通信开销,但其查询结果不准确.

Balijeet<sup>[11]</sup>等人提出了基于 DST 的 top- $k$  查询算法 EXTOK,通过为每个节点设置过滤阈值来决定其下一轮是否向 sink 发送更新值. 该算法保证了 top- $k$  查询的

准确性,能耗也远小于 FILA. 其中 DST 的构建主要依赖于最小连通支配集,选择邻居节点多(度数大)的作为支配节点,使中间传输数据的非叶子节点的个数达到最少.

Cho<sup>[17]</sup>等提出了利用空间相关的感知数据建立偏序树(POT)进行 top- $k$  查询,其主要思想是在网络中选择“热点”(较高感知值的节点)建立逻辑拓扑来减少不必要的更新. 此方法适用于高度非均匀的数据分布,但当热点较少或没有热点时,其性能会迅速降低.

最近,层次索引树的构建也成为 top- $k$  查询的一个研究重点. Tang<sup>[18]</sup>等构建的层次网格索引树(HGIT)将监测区域划分为  $2^n \times 2^n$  个网格单元,建立  $n+1$  层索引树,简化了逻辑拓扑;Wang<sup>[19]</sup>等人也将 WSN 分为两层拓扑:簇头节点层和普通节点层,有效降低了数据查询更新代价.

以上逻辑树都以固定 sink 为根节点,每轮查询根节点唯一,且簇头选取时未考虑节点初始能量,当查询次数、 $k$  值增大时,会导致 sink 附近节点能耗不均而死亡,表 1 比较了不同逻辑拓扑树的构建方法.

表 1 Top- $k$  查询底层逻辑树构建方法对比

方法	构建特点	网络环境	平均每轮查询能耗	是否考虑节点能量
SPT	基于最短路径	sink 唯一	较高	否
DST	基于最小连通支配集	sink 唯一	较高	否
POT	基于空间相关性	sink 唯一	较低	否
HGIT	基于层次索引	sink 唯一	较低	否

针对这些问题,本文综合考虑节点能量、度数以及通信代价,并以此定义最优函数选取支配节点,构建最优支撑树(Optimal Support-Tree, OST). 树中的节点按 ID 号轮流作为根节点,即每次 top- $k$  查询的 sink 均不同,此方法打破了传统传感器网络 sink 固定带来的其附近节点过早死亡的弊端,有效均衡了节点的能耗,从而延长整个网络寿命.

## 3 网络模型及相关定义

假设无线传感器网络中  $n$  个节点随机分布在二维平面上,每个传感器节点按其地理位置形成唯一的 ID 号,即  $i=1,2,\dots,n$ ,ID 号相邻的节点地理位置也相邻,无线传感器网络的物理拓扑结构可抽象为一个简单的无向连通图  $G(V,E)$ ,其中  $V$  表示节点集合, $E$  表示边集合,节点  $u,v$  之间存在边当且仅当  $d(u,v) \leq r, (u,v) \in E$  (其中  $u,v \in V$ ) 即节点  $u,v$  是相邻的,表 2 给出了本文中用到的相关符号和参数说明.

**定义 1** 独立集和极大独立集(MIS):给定图  $G(V,E)$ ,若  $M \subseteq V$  且对于  $\forall u,v \in M$ ,都有  $(u,v) \notin E$ ,则称  $M$

为独立集;若增加任意节点  $w$  至  $M$  中,  $\exists v \in M$ , 使得  $d(v, w) \leq r$ , 即增加任一节点都会破坏  $M$  的独立性, 则称  $M$  为极大独立集.

表 2 符号及参数说明

符号/参数	说明
$r$	传感器节点通信半径
$d(u, v)$	节点 $u, v$ 间距离
$NB_i(v)$	节点 $v$ 的第 $i$ 个邻节点
$n(v)$	节点 $v$ 邻节点的个数
$E_{\text{initial}}(v)$	节点 $v$ 的初始剩余能量
$a$	发射电路消耗的能量
$b$	功率放大消耗的能量
$c$	接收电路消耗的能量
$d$	数据传输距离
$k_1$	一个数据包的比特数
$P(v_i)$	节点 $v_i$ 的未支配邻节点集

**定义 2** 连通支配集和最小连通支配集: 图  $G(V, E)$  的一个节点子集  $V' \subseteq V$  若满足  $\forall u \in V - V'$ , 至少  $\exists v \in V'$ , 使得  $(u, v) \in E$ , 属于  $V'$  的顶点称为支配节点, 其他称为被支配节点, 且由  $V'$  导出的子图是连通的, 则称  $V'$  为连通支配集(CDS); 图  $G(V, E)$  的所有连通支配集中, 节点个数最少的称为最小连通支配集.

## 4 最优支撑树构建

### 4.1 基本思想

目前已有不少研究采用最小连通支配集<sup>[20]</sup>构建网络逻辑树进行高效数据查询, 但最小连通支配集没有考虑节点能量, 并不能有效均衡节点能耗. 本文在现有连通支配集基础上, 充分考虑节点能量, 通信开销以及度数, 以此定义最优函数选取支配节点, 组成能耗均衡的连通支配集, 节点按 ID 号轮流作为根节点, 构建最优支撑树, 虽然一轮查询能耗大于 DST, 但由于节点能耗均衡, 不会过早死亡, 因此延长了整个网络寿命, 从而降低平均每轮的能耗.

### 4.2 最优函数确定

Top- $k$  查询中, 中间数据聚合节点(支配节点)起到了两大作用: 本地 top- $k$  数据查询和向父节点传输数据. 因此其能量消耗远大于其他节点. 考虑网络的能量开销, 应选择较少的支配节点, 即支配节点的度数应比较大; 从节点自身的角度考虑, 应选择剩余能量高, 与邻节点通信开销小的作为支配节点, 这样可以均衡节点能耗. 充分考虑节点度数、剩余能量和通信开销, 定义节点  $v_i$  的能量代价(最优)函数如式(1)所示.

$$f(v_i) = E_{\text{Cost}}(v_i) = \frac{\sum_{j=1}^{n(v_i)} \text{Cost}(v_i, NB_j(v_i))}{n(v_i) \times E_{\text{initial}}(v_i)} \quad (1)$$

其中  $\sum_{j=1}^{n(v_i)} \text{Cost}(v_i, NB_j(v_i))$  为  $v_i$  与其邻节点的通信开销, 包括发送和接收数据能量消耗, 本文采用文献[21]中的能量模型:

$$E_t(k_1, d) = a \times k_1 + b \times k_1 \times d^n \quad (2)$$

$$E_r(k_1, d) = c \times k_1 \quad (3)$$

其中  $E_t(k_1, d)$ ,  $E_r(k_1, d)$  分别为发送和接收  $k_1$  bit 数据包的能量消耗, 则

$$\begin{aligned} & \sum_{j=1}^{n(v_i)} \text{Cost}(v_i, NB_j(v_i)) \\ &= \sum_{j=1}^{n(v_i)} (E_t(k_1, d(v_i, NB_j(v_i))) + E_r(k_1, d(v_i, NB_j(v_i)))) \\ &= \sum_{j=1}^{n(v_i)} (a + c) \times k_1 + b \times k_1 \times d(v_i, NB_j(v_i))^n \end{aligned} \quad (4)$$

由式(1)可以看出, 节点  $v_i$  能量、邻节点越多, 通信代价越小, 则其能量代价值越小, 选择其作为支配节点参与较多通信, 其余能量代价值大的节点参与较少通信, 能有效均衡节点的能耗.

### 4.3 构建算法

最优支撑树的构建分为三个阶段:(1)根据最优函数产生支配节点;(2)选择合适的连接节点, 形成能耗均衡的连通支配集;(3)基于地理位置 ID 轮流作为根节点, 构造最优支撑树. 假设网络中各节点均有一个属性 color (white: 未处理节点, black: 支配节点, grey: 被支配节点), 初始时, 所有节点均为白色未处理节点, 且假设每个节点都知道其邻居节点的能量代价信息.

**阶段 1** 描述如算法 1, 算法 1 结束, 支配节点构成一个 MIS  $D$ .

#### 算法 1 支配节点选择

```

Input:  $G(V, E)$ ,  $f(v_i), i = 1, 2, \dots, n$ 
Output:  $G(V, E)$  的 MIS  $D$ 
1: while ( $v_i$ .color = white) do
2:   if ( $f(v_i) < \min(f(NB(v_i)))$  &&  $NB(v_i)$ .color = white) then
3:      $D \leftarrow v_i$ ,  $v_i$ .color = black
4:      $v_i$  broadcasts the message "dominator( $v_i$ )" to  $NB(v_i)$ 
5:   end if
6:   if ( $v_k$  received the "dominator( $v_i$ )" &&  $v_k$ .color = white) then
7:      $v_k$ .color = grey
8:      $v_k$  broadcasts the message "dominatee( $v_k$ )" to  $NB(v_k)$ 
9:   end if
10: if ( $v_l$  received the "dominatee( $v_k$ )" &&  $v_l$ .color = white) then
11:   remove  $v_k$  from  $P(v_l)$ 

```

```

12:   if ( $f(v_i) < f(P(v_i)) \vee P(v_i).color \neq \text{white}$ ) then
13:      $D \leftarrow v_i, v_i.color = \text{black}$ 
14:      $v_i$  broadcasts the message "dominator( $v_i$ )" to  $NB(v_i)$ 
15:   end if
16: end if
17: end while

```

**定理 1** 算法 1 所选择的支配节点集合是能量代价最小的最大独立集。

**证明** 若节点  $v_i$  成为支配节点, 则其邻节点为被支配节点, 故任意两个支配节点不相邻, 所有支配节点集为独立集; 对任意一个被支配节点, 其邻居中至少存在一个支配节点, 若此被支配节点成为支配节点, 则会破坏当前支配节点独立集, 故支配节点集为最大独立集, 且每个支配节点都具有最小能量代价, 得证。

**定理 2** 算法 1 的消息复杂度和平均时间复杂度分别均为  $O(n)$ 。

**证明** 阶段 1 除了初始发送 hello 消息包、能量代价消息包外, 在执行过程中, 各节点只需广播一条自身状态至邻节点, 因此消息复杂度为  $O(n)$ ; 每个节点需与邻节点中能量代价最小的节点比较, 以此确定其是否成为支配节点, 设  $\Delta$  为节点的最大度数, 最坏情况下, 邻节点找出能量代价最小值需比较  $\Delta$  次, 节点与代价最小节点比较 1 次, 且最后一个节点需等待前  $n-1$

个节点确定状态后才能确定自身状态, 故最坏时间复杂度为  $O(n(\Delta+1))$ , 即  $O(n\Delta)$ , 平均时间复杂度为  $O(n)$ , 得证。

文献[22]中是由 sink 指定的第一个支配节点依次选取下一支配节点, 除要发送相关状态外, 潜在的支配节点还需发送其信息至上一支配节点进行选择, 选择结果再以消息通知相应节点, 其消息复杂度为  $O(n\Delta)$ , 大于算法 1。

以图 1 为例, 说明算法 1 的执行流程。图 1(a) 为传感器节点的物理拓扑图, 初始所有节点均为白色未处理节点, 假设每个节点已根据式(1)计算出能量代价  $f(v_i)$ , 按 ID 号依次为 10, 15, 21, 19, 17, 14, 18, 20, 18, 15, 24, 23, 21, 18, 20。

由 line2 可知, 节点 1, 10, 14 成为黑色支配节点, 同时向邻节点广播支配信息。邻节点收到后, 若其为未处理节点, 则成为被支配节点, 同时向邻节点广播被支配信息, 如图 1(b) 所示。此时剩下三个未处理节点 5, 8, 9, 节点 5 收到被支配信息, 且其没有未处理邻节点, 则自身成为支配节点并广播支配信息; 节点 8 收到被支配信息, 与邻节点 9 比较得到  $f(v_9) < f(v_8)$ , 故节点 9 成为支配节点, 节点 8 为被支配节点。此时图中无未处理节点, 如图 1(c) 所示, 算法结束, 输出支配节点集  $D = \{1, 5, 9, 10, 14\}$ 。

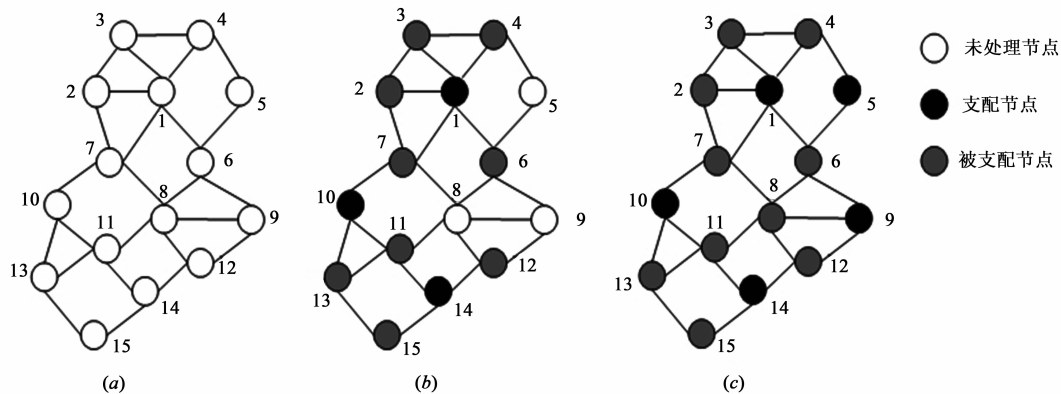


图1 支配节点选择

**阶段 2** 支配节点根据邻节点能量代价和被支配情况选择连接点, 构成连通支配集, 描述如算法 2。

#### 算法 2 构建连通支配集 C

```

Input:  $G(V, E), D$ 
Output:  $G(V, E)$  的 CDS  $C$ 
1:  $C \leftarrow D$ 
2: if ( $v_i.color = \text{grey}$  &&  $v_k$  is  $v_i$ 's the one-hop dominator)
then
3:    $v_i$  broadcasts the message  $\text{dominatee1}(v_i, v_k)$  to  $NB(v_i)$ 
4: end if

```

```

5: if ( $v_l$  received  $\text{dominatee1}(v_i, v_k)$  &&  $v_k$  isn't  $v_l$ 's 1-hop dominator)
then
6:   2-hop dominator of  $v_l \leftarrow v_k$ 
7:   if ( $v_l.color = \text{black}$ ) then
8:      $\text{list1} \leftarrow (v_i, v_k)$ 
9:   else  $v_l$  broadcasts the message  $\text{dominatee2}(v_l, v_i, v_k)$  to  $NB(v_l)$ 
10:  end if
11: end if
12: if ( $v_m$  received the  $\text{dominatee2}(v_l, v_i, v_k)$  &&  $v_k$  isn't  $v_m$ 's 2-hop
dominator) then
13:   3-hop dominator of  $v_m \leftarrow v_k$ 
14:   if ( $v_m.color = \text{black}$ ) then

```

```

15: list2 ← (v_l, v_i, v_k)
16: end if
17: end if
18: if (f(v_i) < f(v)) then
19: list1 C ← v_i
20: end if
21: if (f(v_l) + f(v_i) < f(v) + f(v')) in list2 then
22: C ← v_l
23: end if

```

**定理 3** 算法 2 的消息复杂度和时间复杂度分别为  $O(2n')$  和  $O(2(n-n')\Delta)$ .

**证明** 阶段 2 中被支配节点需向邻节点广播其一跳远、两跳远支配节点的信息,故其消息复杂度为  $O(2n')$ ,其中  $n'$  为被支配节点的个数;每个支配节点需  $O(\Delta)$  时间建立其 2 跳(或 3 跳)远支配节点信息列表,在选择连接节点时,最坏情况下需  $O(\Delta)$  时间进行比较,故其时间复杂度为  $O(2(n-n')\Delta)$ ,其中  $n-n'$  为支配节点个数. 得证.

文献[22]虽然时间复杂度为  $O(D)$  ( $D$  为网络直径),优于算法 2,但其没有存储 2 跳(或 3 跳)远节点信息,需向邻节点发送消息请求,故其消息复杂度为  $O(n\Delta)$ ,在能量受限的无线传感器网络中,本文算法 2 具有较小的通信开销.

算法 2 结束后,每个节点都有一条到达其最近支配节点的代价最小路径,如图 2 所示,节点 6,7,12 成为连接节点,与支配节点 1,5,9,10,14 构成连通支配集  $C = \{1, 5, 6, 7, 9, 10, 12, 14\}$ .

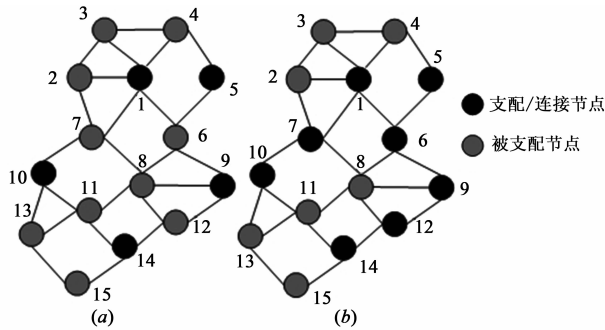


图2 构造连通支配集

**阶段 3** 构建最优支撑树 (OST), 算法伪代码如算法 3 所示.

算法 3 构建 OST

```

Input: G(V, E), C
Output: OST(V, E)
1: while ( (u, v) ∈ E ) do
2: if ( both u, v are dominatees or u, v are dominatee and connector node )
then
3: (u, v) ∉ E

```

```

4: end if
5: if ( u is dominated by dominator v and u is not dominated by other domi-
nators ) then
6: (u, v) ∈ E'
7: u. parent = v
8: end if
9: if ( u is dominated by both v, w && d(u, v) > d(u, w) || w has least
dominatees ) then
10: (u, w) ∈ E'
11: u. parent = w
12: end if
13: if ( u, v are dominator and connector node ) then
14: (u, v) ∈ E'
15: end if
16: end while
17: each node is chosen to be the root followed by ID

```

在阶段 2 的基础上,保留支配节点与被支配节点间的边,其余边均去除. 若一个被支配节点同时被两个支配节点支配,则先选择具有较少被支配节点的节点作为支配节点,如节点 4,同时被节点 1 与 5 支配,节点 5 只有一个被支配节点 4,故将 5 作为支配节点. 如图 3 所示,每次 top-k 查询时,节点按 ID 号轮流作为根节点.

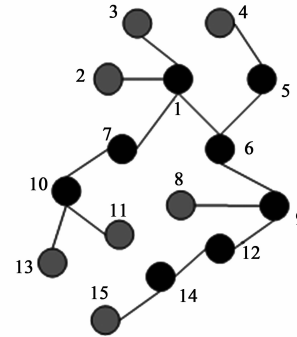


图3 构建最优支撑树

## 5 仿真实验与性能评估

为了分析和验证 OST 在 top-k 查询中的性能,本文利用 Matlab 仿真建树和 top-k 查询过程,其中 top-k 查询方法采用 EXTOK 算法<sup>[14]</sup>,并分别以最短路径树 (SPT)、支配集树 (DST) 及本文构建的 OST 为逻辑树做仿真对比实验.

### 5.1 实验场景及参数设定

本实验采用生成的数据集,假设传感器节点随机分布在  $100\text{m} \times 100\text{m}$  的监测区域,节点通信半径均为  $10\text{m}$ ,初始选择离区域中心最近的节点作为 sink 节点. 节点初始剩余能量各不相同,其中有 5% 的低能节点,初始剩余能量小于  $50\text{mJ}$ ,5% 的高能节点,能量大于  $150\text{mJ}$ ,其余节点能量为  $80\text{mJ} \sim 100\text{mJ}$ ;假设每轮查询每个传感器节点只产生一个感知值,实验随机产生介于  $20 \sim 40$  间的数值

作为每轮传感器节点的“温度”感知. 为了使分析更加合理, 取 10 次运行的平均值作为最终实验结果, 每次模拟包含 400 轮查询, 具体参数设置如表 3 所示.

表 3 实验参数设置

主要参数	参考值
网络大小	100m × 100m
节点个数 $N$	200 ~ 1000
通信半径 $r$	10 m
发射电路能量消耗 $a$	50nJ/bit
功率放大能量消耗 $b$	0. 1nJ/m <sup>2</sup> /bit
接收电路能量消耗 $c$	50nJ/bit
能量模型中的 $n$	2
消息包大小	8Byte
top- $k$ 值	10, 20, 30, 40, 50

## 5.2 实验结果与分析

实验首先比较了不同逻辑树的建立结果, 如图 4

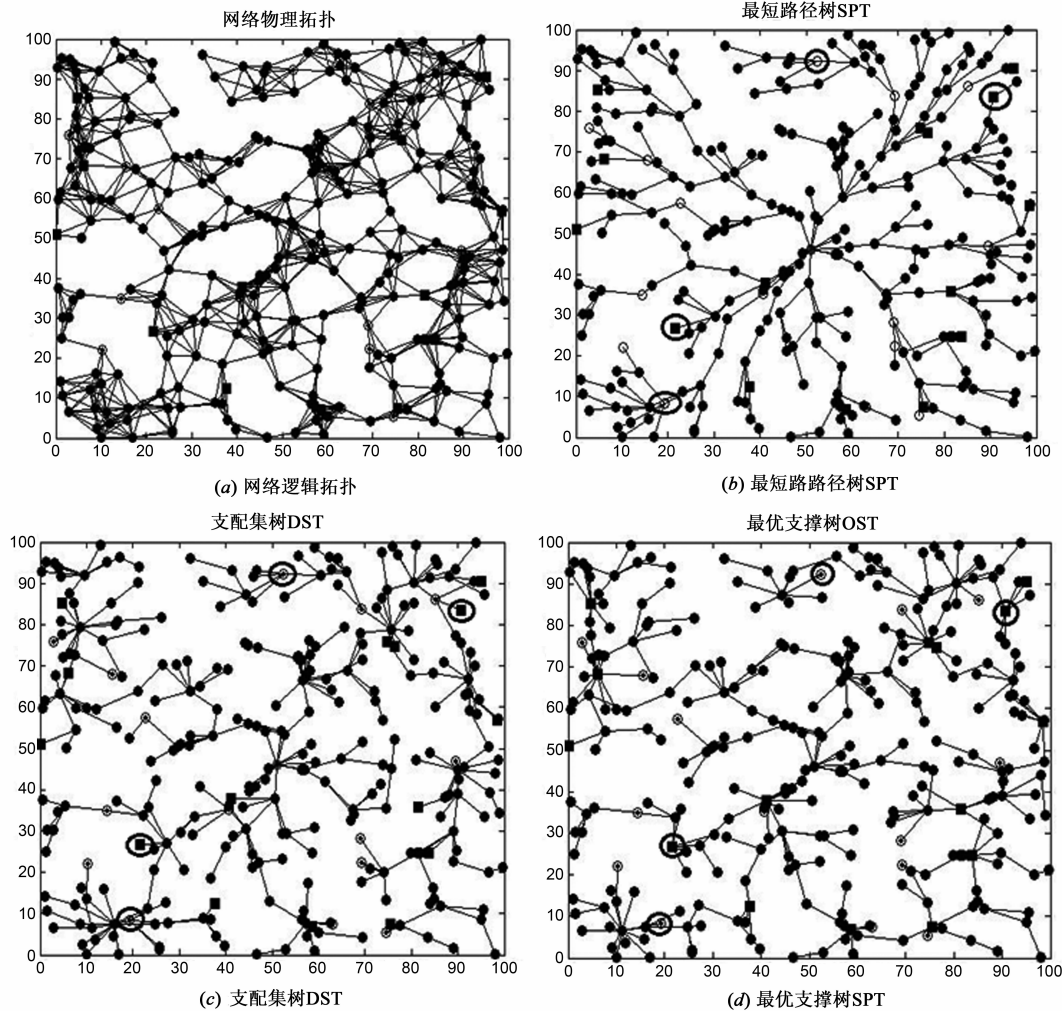


图4 不同算法形成的逻辑树对比图

### 5.2.1 能量消耗

据文献[9]所述, 传感网中节点执行一条指令需耗

所示, 图 4(a) 为初始网络物理拓扑, 其中空心圆表示低能节点, 正方形为高能节点, 其余为普通节点. 图 4(b)、图 4(c)、图 4(d) 分别为 SPT、DST 以及本文所构建的 OST, 从图中可以看出, SPT、DST 建立过程中 sink 固定 (区域中心节点), 且均未考虑节点剩余能量, 故存在低能节点作为聚合节点, 而高能节点成为叶子节点 (如图中圈出的点) 的情况, 在查询过程中, 会带来低能节点能耗过多、提前死亡的问题, 当查询轮数增多时, sink 附近节点也会不堪重负. 本文构建的 OST 有效均衡了节点的能耗, 从图 4(d) 可以看出, 原先作为聚合节点的低能节点成为叶子节点, 而高能节点作为父亲节点.

下面分别从能量消耗、查询时间及网络寿命三个方面对本文提出的 OST 进行 top- $k$  查询性能对比分析.

能 0. 84nJ, 但发送一个数据包大约需消耗 0. 685mJ, 远大于计算开销, 因此本文忽略节点感知和计算能耗, 主要考

考虑 top- $k$  查询节点间的通信开销. 图 5 给出了不同生成树 top- $k$  查询能耗, 图 5(a)、图 5(b) 分别是能量消耗随  $k$  值和网络节点总数  $N$  的变化, 从图中可以看出, 网络每轮查询能耗均随  $k$  值、节点数目  $N$  的增加而增加, 这是由于当  $k$  增大时, 每轮会有更多节点向 sink 发送更新值;  $N$  增大

时, sink 将向更多节点发送过滤值, 因此通信开销增大. 但对比发现, 本文建立的最优支撑树 (OST) 平均每轮能耗最小, 主要因为 OST 从三方面定义了节点能量代价, 提供更高效率的拓扑结构, 有效均衡了高、低能耗节点的通信, 延长了网络寿命, 从而减少了平均每轮能耗.

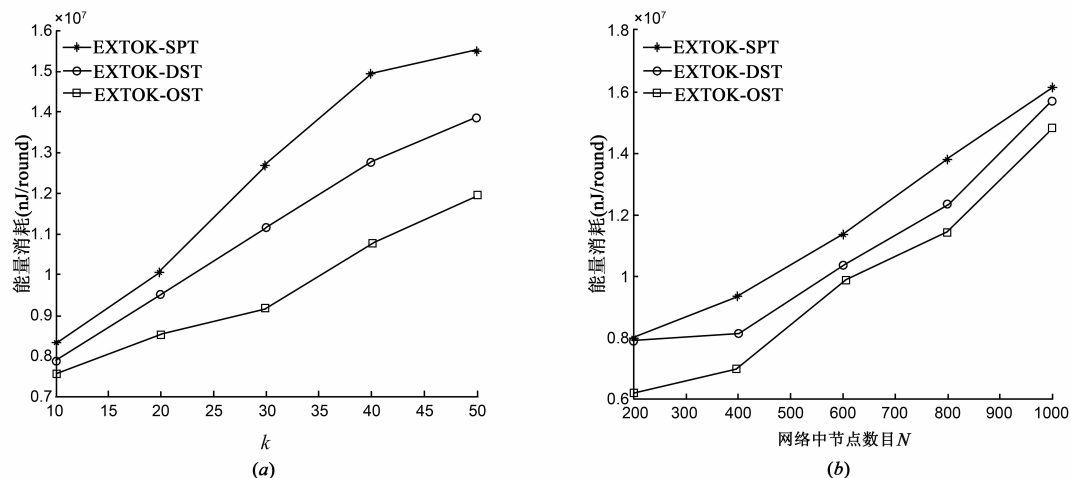


图5 不同逻辑树top- $k$ 查询能量消耗

### 5.2.2 查询时间

图 6 比较了不同逻辑树的平均每轮 top- $k$  查询时间, 从图 6(a)、图 6(b) 可以看出, 随着  $k$  值和节点数目  $N$  的增加, 平均每轮查询时间也相应增大, 原因与上面能耗增大相似; 而 SPT 的平均查询时间远大于 DST 和

OST, 这是由于 SPT 中间聚合节点较多, 查询时聚合节点间通信耗费时间长; 而 OST 平均查询时间略小于 DST, 原因在于两者都是基于连通支配集, 中间支配节点数目相差不多.

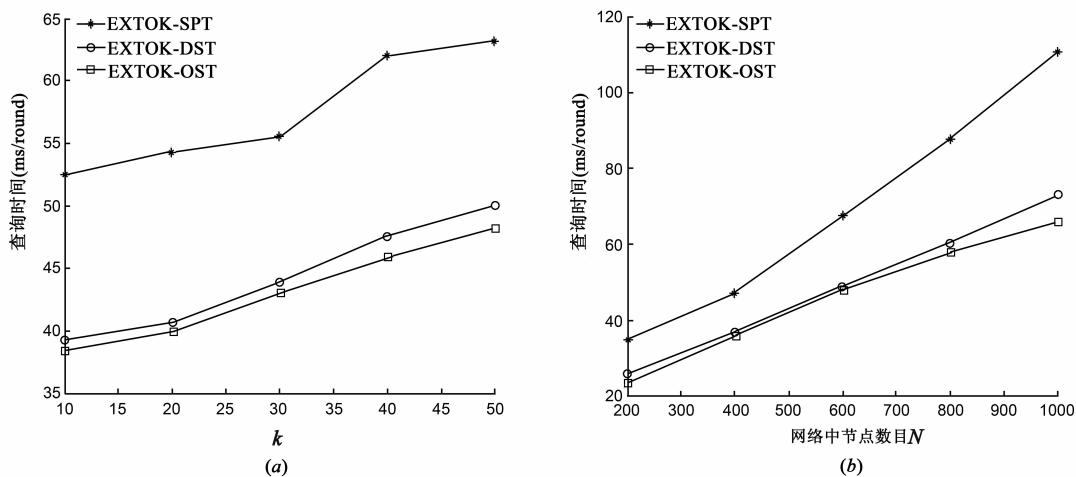
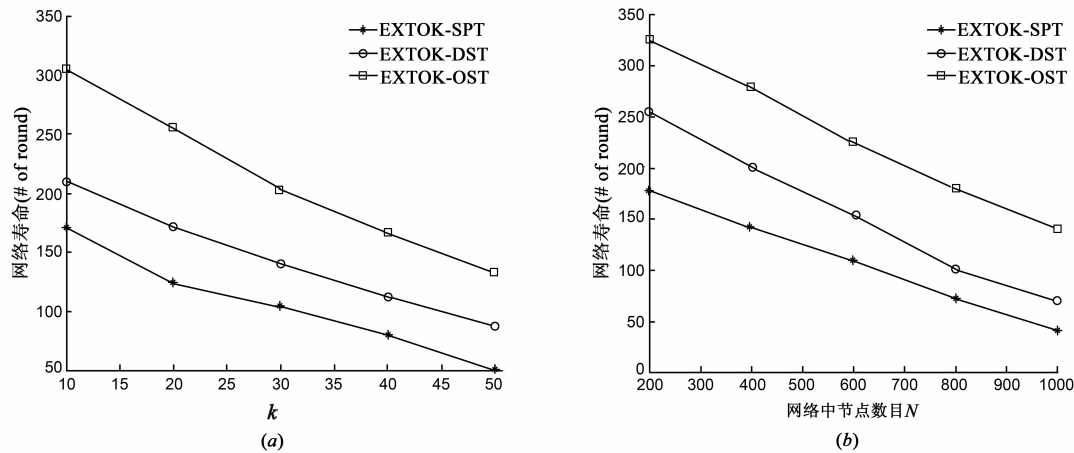


图6 不同逻辑树top- $k$ 平均查询时间

### 5.2.3 网络寿命

本文认为当网络中第一个节点能量耗尽时网络死亡, 因此将网络死亡时查询轮数定义为网络寿命. 图 7(a)、图 7(b) 分别比较了三种逻辑拓扑结构下  $k$  值和节点数目  $N$  变化对网络寿命的影响, 可以看出, 随着  $k$  值、节点数目的增加, 三种逻辑结构下的网络寿命均缩短, 这与图 5 能耗增加相一致. 而本文提出的 OST 有效

延长了网络寿命, 在节点数为 200 时, OST 达 325 轮, 与 SPT 相比, 延长了约 82% 的网络寿命. 这是由于 OST 有效均衡了节点能耗, 将能耗小的节点作为支配节点负责更多通信, 而能耗大的节点作为叶子节点参与较少通信, 且每次查询由不同节点作为根节点, 避免其附近节点过早死亡, 延长单个节点的寿命来延长整个网络寿命.

图7 不同逻辑树top- $k$ 查询网络寿命

## 6 结束语

本文在连通支配集理论基础上,提出了一种能耗均衡的 top- $k$  查询最优支撑树构建方法,该方法建立在无固定 sink 的网络背景下,充分考虑了节点能量、度数以及与邻节点通信开销,选择能量代价较小的节点作为支配节点负责较多通信与本地查询,而能量代价大的节点作为叶子节点参与较少通信,从而使节点能耗得到均衡,延长了节点寿命;同时在每轮 top- $k$  查询中,节点基于地理位置 ID 轮流作为根节点,打破了传统根节点唯一带来的其附近节点过早死亡的弊端。

仿真实验主要从能量消耗、查询时间及网络寿命三方面进行评估,分析表明,与其它逻辑树相比,基于 OST 进行 top- $k$  查询能有效减少平均每轮查询的能耗,从而延长整个网络寿命,同时在查询时间上也略有优势,下一步工作将研究基于 OST 的高效新颖的 top- $k$  查询方法。

### 参考文献

- [1] Dyo V, Ellwood S A, Macdonald D W, et al. WILDSENSING: design and deployment of a sustainable sensor network for wildlife monitoring [J]. ACM Transactions on Sensor Networks (TOSN), 2012, 8(4): 29.
- [2] Hsieh M H, Lin K W, Tseng V S. A hybrid scheme for energy-efficient object tracking in sensor networks [J]. Knowledge and Information Systems, 2013, 36(2): 359 - 384.
- [3] Sathyan T, Hedley M. Fast and accurate cooperative tracking in wireless networks [J]. IEEE Transactions on Mobile Computing, 2013, 12(9): 1801 - 1813.
- [4] Wu M, Xu J, Tang X, et al. top- $k$  monitoring in wireless sensor networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(7): 962 - 976.
- [5] Chaudhuri S, Gravano L, Marian A. Optimizing top- $k$  selec-

tion queries over multimedia repositories [J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(8): 992 - 1009.

- [6] Meng X, Shao J. Finding top- $k$  semantically related terms from relational keyword search [A]. Proceedings of IEEE International Conference on Data Science and Advanced Analytics (DSAA) [C]. Shanghai: IEEE, 2014. 505 - 511.
- [7] 莫尚丰, 陈丁洁, 陈红, 等. 无线传感器网络中 top- $k$  连接查询处理 [J]. 计算机学报, 2013, 36(3): 557 - 570.  
Mo Shang-feng, Chen Ding-jie, Chen Hong. Algorithms for top- $k$  join queries in wireless sensor networks [J]. Chinese Journal of Computers, 2013, 36(3): 557 - 570. (in Chinese)
- [8] Ye M, Liu X, Lee W C, et al. Probabilistic top- $k$  query processing in distributed sensor networks [A]. Proceedings of IEEE 26th International Conference on Data Engineering (ICDE) [C]. Long Beach, CA: IEEE, 2010. 585 - 588.
- [9] Jiang H, Cheng J, Wang D, et al. Continuous multi-dimensional top- $k$  query processing in sensor networks [A]. IEEE Proceedings on INFOCOM [C]. Shanghai: IEEE, 2011. 793 - 801.
- [10] 贾杰, 代恩亮, 陈剑, 等. 无线传感器网络中联合路由优化的能效链路调度 [J]. 电子学报, 2013, 42(6): 1118 - 1124.  
JIA Jie, DAI En-liang, CHEN Jian, et al. Energy-efficient link scheduling combined with routing optimization in wireless sensor network [J]. Acta Electronica Sinica, 2013, 42(6): 1118 - 1124. (in Chinese)
- [11] Malhotra B, Nascimento M A, Nikolaidis I. Exact top- $k$  queries in wireless sensor networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(10): 1513 - 1525.
- [12] Tabus V, Astola J. Maximizing the lifetime of energy constrained wireless sensor networks having tree topology [A]. Proceedings of IEEE 6th International Symposium



- on Communications, Control and Signal Processing (ISCCSP) [C]. Athens:IEEE,2014. 388 – 391.
- [13] Zhu C, Yang L, Shu L, et al. Insights of top-k query in duty-cycled wireless sensor networks[J]. IEEE Transactions on Industrial Electronics,2014,62(2):1317 – 1328.
- [14] Liang W F, Chen B, Yu J X. top-k query evaluation in sensor networks under query response time constraint[J]. Information Sciences,2011,181(4):869 – 882.
- [15] Wang H, Guan Z T, Yang T T, et al. top-k query framework in wireless sensor networks for smart grid[J]. China Communications,2014,11(6):89 – 98.
- [16] 梁俊斌, 马行坡, 奎晓燕. 查询驱动模式下两层传感器网络 top-k 查询汇聚算法研究[J]. 电子学报,2014,42(10):2075 – 2080.  
Liang Jun-bin, Ma Xing-po, Kui Xiao-yan. Research on data aggregation algorithms for top-k queries in query-driven-based two tiered sensor networks[J]. Acta Electronica Sinica,2014,42(10):2075 – 2080. (in Chinese)
- [17] Cho Y H, Son J, Chung Y D. POT: an efficient top-k monitoring method for spatially correlated sensor readings [A]. Proceedings of the ACM 5th Workshop on Data Management for Sensor Networks [C]. Auckland, New Zealand: Association for Computing Machinery, 2008. 8 – 13.
- [18] Tang J, Wang Z, Sun Y, et al. Top-k queries in wireless sensor networks leveraging hierarchical grid index [A]. Proceedings of IEEE Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS) [C]. Birmingham: IEEE, 2014. 381 – 386.
- [19] Kim D, Wu Y, Li Y, et al. Constructing minimum connected dominating sets with bounded diameters in wireless networks[J]. IEEE Transactions on Parallel and Distributed Systems,2009,20(2):147 – 157.
- [20] Tan H O, Korpeoglu I, Stojmenovic I. Computing localized power-efficient data aggregation trees for sensor networks[J]. IEEE Transactions on Parallel and Distributed Systems,2011,22(3):489 – 500.
- [21] Mohanty J P, Mandal C. A distributed greedy algorithm for construction of minimum connected dominating set in wireless sensor network [A]. Applications and Innovations in Mobile Computing (AIMoC) [C]. Kolkata: IEEE,2014. 104 – 110.
- [22] Das A, Mandal C, Reade C, et al. An improved greedy construction of minimum connected dominating sets in wireless networks [A]. Wireless Communications and Networking Conference (WCNC) [C]. Cancun, Quintana:IEEE,2011. 790 – 795.

#### 作者简介



**邬海琴** 女,1992年3月出生,江苏如皋人.2014年毕业于江苏大学计算机系,现为硕士研究生,从事无线传感器网络与大数据安全方面的有关研究.

E-mail: whq92\_ujs@126.com



**王良民** 男,1977年9月出生,安徽潜山人,教授、博士生导师.1999年、2007年分别在吉林大学、西安电子科技大学获理学学士、工学博士学位.主要从事无线传感器网络、物联网及其大数据处理等方面的研究工作.

E-mail: wanglm@ujs.edu.cn