

## 基于环结构的传感器网络多分辨率数据存储机制\*

谢磊<sup>1,2</sup>, 陈力军<sup>1,2+</sup>, 陈道蓄<sup>1,2</sup>, 谢立<sup>1,2</sup>

<sup>1</sup>(南京大学 计算机软件新技术国家重点实验室,江苏 南京 210093)

<sup>2</sup>(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)

### Ring-Based Multi-Resolution Data Storage for Sensor Networks

XIE Lei<sup>1,2</sup>, CHEN Li-Jun<sup>1,2+</sup>, CHEN Dao-Xu<sup>1,2</sup>, XIE Li<sup>1,2</sup>

<sup>1</sup>(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

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

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

Xie L, Chen LJ, Chen DX, Xie L. Ring-Based multi-resolution data storage for sensor networks. *Journal of Software*, 2009,20(12):3163-3178. <http://www.jos.org.cn/1000-9825/3474.htm>

**Abstract:** In this paper, a ring-based multi-resolution storage scheme for sensor networks is proposed. Combined with the hierarchical scheme for data storage and query, it leverages characters of the ring-based structure to support multi-resolution data storage and query operations, achieving a good performance in terms of energy consumption. It adopts optimal parameters for ring-based storage structures, which can well handle the hierarchical data storage and query scheme with the minimized overall communication cost. Furthermore, the authors do theoretical analysis on the performance of the ring-based multi-resolution storage scheme, including energy efficiency, load balance. The experimental results indicate that significant benefits can be achieved with this ring-based multi-resolution storage scheme.

**Key words:** sensor network; data storage; event; query; ring; multi-resolution

**摘要:** 提出了一套基于环结构的传感器网络多分辨率数据存储机制,结合层次结构的存储查询方案,有效地利用了环结构的特性高效、节能地支持事件信息不同分辨率的存储和查询操作,并采用优化的环结构参数,在基于环的层次结构数据存储架构中能够最小化网络节点的总体通信能耗。同时,对环结构多分辨率数据存储机制的相关性能从节能性、负载均衡性等多个角度进行了具体理论分析。模拟实验结果表明,基于环的层次结构存储机制能够高效、节能地支持传感器网络事件数据的多分辨率存储和查询操作。

**关键词:** 传感器网络;数据存储;事件;查询;环结构;多分辨率

中图法分类号: TP393 文献标识码: A

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,无线传感器网络已经广泛地出现在环

\* 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 Basic Research Program of China under Grant No.2006CB303000 (国家重点基础研究发展计划(973)); 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))

Received 2007-12-06; Accepted 2008-10-09; Published online 2009-04-07

境监测、目标追踪、突发事件报警等多种相关应用领域中.组成无线传感器网络的传感器节点不仅能够接收物理世界的感知数据,而且还能够有效地存储、计算和转发这些数据,以实现部署区域感知信息的有效收集.本文主要关注对传感器网络部署区域内相关事件进行监测的一类应用,在这类应用中,用户所关心的并不是大量原始的感知数据,而是由这些底层原始数据所综合而成的事件信息.例如,对于火灾这个事件,信息往往包括火灾发生的地点、时间以及周围的温度、湿度等诸多原始感知数据.在这种应用场景下,传感器网络的用户往往分布在传感器网络的部署区域之中,随时通过传感器网络对当前所感兴趣的事件信息进行查询.对于这种大规模的事件探测和查询应用,传感器节点可能会在不同的观测区域探测到多种不同种类的事件信息,而同时用户发出的查询往往具有很高的选择性,仅仅对其中某些区域某些种类的事件感兴趣,加之事件产生和查询请求是随机出现的,分布在传感器网络的部署区域内,为了能够有效地处理用户的查询请求,需要充分利用传感器网络的网内存储特性来支持信息的分发与收集.由于组成传感器网络的传感器节点往往具有苛刻的能耗限制,因此,设计一套高效节能的数据分发与收集机制,对大规模部署的传感器网络应用更为重要<sup>[1]</sup>.

针对传感器网络上不同查询用户所关心事件的粒度不同,某些用户只会关心特定地点某些特定事件的精确信息,而其他一些用户则关心局部区域或整个区域内相关事件信息的概要数据,如平均值、最大值等,本文旨在利用传感器网络以数据为中心的网内存储机制,提供一套多分辨率的事件存储机制以有效支持精确事件查询、不同粒度的聚集查询以及由粗到细的混合查询请求,充分满足传感器网络中多种用户不同种类的应用需求.针对当前已有的传感器网络多分辨率存储机制在优化存储部署问题上的不足,本文提出了一套基于环的层次结构的存储查询机制,采用优化的存储结构参数和存储查询策略来高效、节能地支持不同分辨率的存储查询操作.实验结果表明,基于环结构的多分辨率存储机制能够有效地降低传感器网络的整体存储查询能耗.

本文第 1 节介绍以数据为中心的存储查询机制的相关工作.第 2 节对本文研究的问题进行描述与分析,并给出系统模型描述.第 3 节具体介绍基于环的层次结构的事件存储和查询机制,对存储查询策略和优化的环结构参数进行分析.第 4 节对环结构存储查询机制的各项性能进行具体理论分析.第 5 节通过实验对相关性能进行验证.最后总结全文.

## 1 相关工作

传感器网络的数据存储机制主要分为外部存储、本地存储和以数据为中心的存储机制.文献[2-4]对这 3 种机制的相关性能进行了具体的分析.其中,以数据为中心的存储机制(data centric storage)<sup>[2]</sup>基于感知数据的属性来存储和查询数据,通过映射算法将事件信息路由到对应的网内节点存储,同时,查询请求也通过相应的映射算法路由至存储节点获取数据,这种机制又进一步分为结构化的存储机制和无结构的存储机制.

结构化的存储机制主要依赖于采用固定的映射算法,如地理散列函数将固定类型的事件信息映射到确定的存储位置,查询节点能够使用与事件产生节点相同的地理散列函数寻找到特定的事件存储节点,并采用地理路由协议(GPSR)<sup>[5]</sup>路由到存储节点获取数据.这种类型的存储方案主要包括GHT<sup>[5-9]</sup>一类的方法,文献[10]基于GHT提出了优化的结构化存储机制.无结构的存储机制主要依赖于采用push-pull技术来实现<sup>[11,12]</sup>.Push是指事件节点在网络中的某些路径上分发事件信息,pull是指查询节点在网络中的某些路径上扩散查询信息.当查询路径与事件分发路径相交时,即实现了相关事件信息的收集.因此,当前的无结构存储方案主要在于研究如何在事件数据的push和pull之间实现更好的平衡,以充分达到高效、节能的效果.这种类型的存储方案主要包括Double Ruling<sup>[13-15]</sup>,Rumor Routing<sup>[16]</sup>等.文献[17]提出了一个GHT与Double Ruling混合的数据分发方案,将整个部署区域分为多片,在片间实现GHT分发方案,在片内实现Double Ruling的分发方案.文献[18]对这两类结构化和无结构的数据存储机制进行了基本的扩展性分析.

由于传感器网络上不同查询用户所关心事件的粒度不同,因此需要提供多分辨率层次结构的存储查询机制来有效地支持不同粒度由粗到细的查询请求.文献[6,7]提出了一个称为维(DIMENTION)的层次体系存储结构,这种层次结构适合于下钻(drill down)操作,文献[8]给出了一种称为 DIFS 的分布式索引方法,可以有效地处理区域查询,文献[9]提出了一个称为 DIM 的能够支持多维属性区域查询的空间索引结构.文献[19]提出了一个

基于环结构的索引机制,在环上提供多个复制的索引节点来有效支持数据的分发与查询。

上述这些支持多分辨率查询的层次结构的存储机制都采用基于GHT<sup>[5]</sup>的存储方案,具体结构如图 1 所示. 通过将部署区域递归地划分为多个层次(L<sub>1</sub>,...,L<sub>d</sub>),每个层次内包括一个或多个子区域,第 1 层只有 1 个子区域,即传感器网络覆盖的地理区域,第 i 层具有 4<sup>i-1</sup>个子区域,第 i 层每个子区域的大小是第 i+1 层子区域大小的 4 倍. 在每个层次内采用地理散列函数把事件信息按关键字散列(hash)到某个地理点,并把数据存储到离该点最近的传感器节点上.这种基于 GHT 的存储方案使用地理散列函数将存储节点均匀地分布在部署区域内,结合使用结构复制(structured replication)<sup>[5]</sup>的方法,能够有效地分摊事件存储和查询的负载.但是,这种方案并不能够充分考虑到存储节点的优化部署问题,对于事件的存储需要更新每一层的相关存储节点,并且对于多分辨率查询中的全集查询(all-type query)和聚集查询(aggregate query),需要同时访问多个相关的存储节点,而基于 GHT 的存储查询方案由于利用地理散列函数过分地分散存储节点,需要在网络中来回往复地访问多个存储节点,其存储查询的总体能耗开销是相当大的.针对这个问题,基于 Double Ruling 的存储机制<sup>[13-15]</sup>则使用了一种 push-pull 结合的存储查询方式,如图 2 所示,事件生成节点 p 与查询节点 q 分别沿部署区域的水平方向和垂直方向产生事件发布的 push 路径和事件查询的 pull 路径,当两条路径相交时即完成了对事件的查询.两条路径的正交性保证了对于任意查询必然能够找到对应的事件存储节点获取数据.Double Ruling 的存储查询机制通过合理部署冗余存储节点,保证了在不需精确定位信息情况下也能够实现事件存储查询任务,同时,通过平衡 push-pull 路径有效地降低了邻近节点间的事件存储查询开销,但由于事件和查询节点的随机分布性导致其存储节点的部署结构也并非充分优化的,同时不能提供合理的多分辨率层次结构的检索机制.

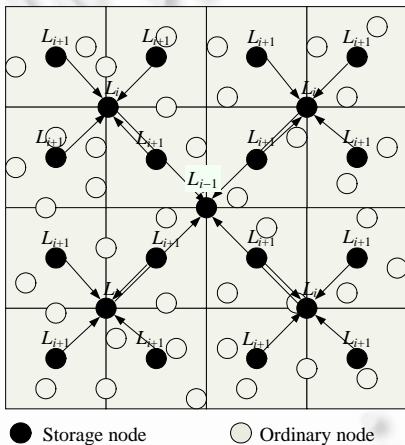


Fig.1 GHT based data storage  
图 1 基于 GHT 的存储机制

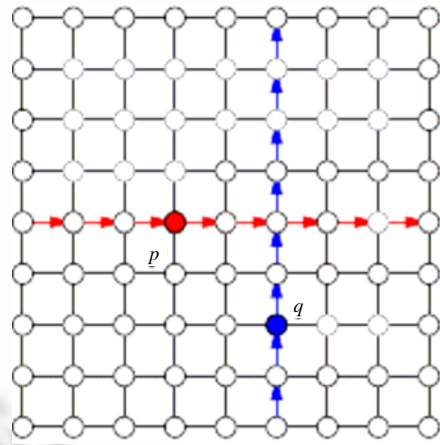


Fig.2 Double ruling based data storage  
图 2 基于 Double Ruling 的存储机制

## 2 问题描述与系统模型

### 2.1 问题描述

基于前面对支持多分辨率查询的层次结构存储机制的分析,为了能够实现高效节能的传感器网络多分辨率数据存储机制,我们需要考虑如何部署层次结构中的存储节点,充分减少在多个层次结构上的数据存储和查询的路由开销,从而最小化存储查询的整体能耗.由此,我们需要构建优化的存储查询结构,并定制相应的优化存储查询策略,从网络整体节能性、数据存储查询的有效便捷性、节点间的负载均衡性这几个方面来指导构建优化的实现机制.基于这样的思路,我们综合借鉴 GHT 和 Double Ruling 的设计思想,提出半结构化的基于环的层次结构的存储查询机制(如图 3 所示),采用优化的存储结构参数和存储查询策略来高效、节能地支持不同分辨率的存储查询操作.

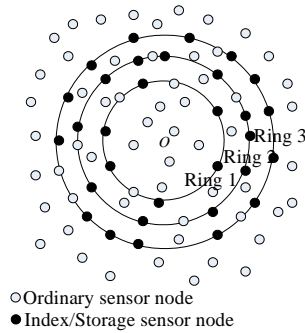


Fig.3 Ring based data storage

图3 基于环结构的存储机制

## 2.2 系统模型

本文假设 $N$ 个传感器节点随机均匀地分布在半径为 $L$ 的圆形部署区域内,节点间的平均距离为 $r$ .根据多分辨率存储查询机制的需求,层次结构的维度即环的个数为 $d$ ,这里设默认值 $d=3$ ,从内到外各环半径分别为 $R_1, R_2, \dots, R_d$ .所有层次上复制节点的冗余度为 $n$ .存在 $k$ 个事件类型 $E_1, E_2, \dots, E_k$ ,事件发生频率为 $f_1, f_2, \dots, f_k$ ,其总体更新频率为 $f$ ,有 $f = \sum_{i=1}^k f_i$ ,对应 $k$ 个事件类型每层的存储节点数目(不考虑复制节点)分别为 $k_1, k_2, \dots, k_d$ .用户对 $d$ 种不同分辨率数据的查询频率分别为 $q_1, q_2, \dots, q_d$ ,其中, $q_i (i=1, 2, \dots, d)$ 为相对查询频率.我们假设网络中的节点密度足够大,从而保证在每个方向都有很好的连通性,这样,通过统计数据在网络中传输路径的长度即可衡量对应的通信能耗开销.对于传感器网络的应用属性,本文假设具有如下性质:(1) 传感器网络为静态网络,即节点部署后不再移动;(2) 节点可以获知其位置信息.节点依靠GPS或定位算法等辅助设施或算法获取位置信息;(3) 查询用户分散在传感器网络部署区域内,事件信息与查询请求的生成符合随机、均匀分布的性质.

## 3 基于环结构的多分辨率事件存储和查询方案

对于传感器网络存储节点的优化部署问题,我们需要找出将存储节点放置在何处才能够使得整个传感器网络事件存储和查询总能量开销最低.对于这个问题的研究,存在定理 1.

**定理 1.** 若传感器节点均匀地分布在观察区域内,指定距离网络中心最近的节点作为事件存储节点,可使传输事件所消耗的能量最少.

对于定理 1 的证明,文献[10]有详细的证明过程.因此在这里略去具体证明过程.通过定理 1 我们得知,对于层次结构的多分辨率存储机制,为保证事件存储和查询的高效节能性,我们需要尽可能地将存储节点放置在网络中心区域中,同时,考虑到这样可能会造成网络热点(hot spot)问题,需要利用复制存储节点方案(replicated storage)来分摊存储和查询的传输量,由此,我们需要一个接近网络中心同时结合复制存储节点方案的网内存储结构来高效、节能地处理多分辨率的事件存储和查询请求.基于以上分析,我们提出一个基于环的层次结构的存储机制,如图 3 所示,各层的存储节点成环状的形式以网络中心位置 $O$ 为圆心聚集在其周围,这里,我们假设根据应用需求设定层次结构维度 $d=3$ .环 1~环 3 上存储着不同分辨率的数据信息,各环之上包括同一层次的数据存储节点(storage node)和复制存储节点(replicated storage node),以及普通的转发节点(forwarding node).

### 3.1 基于环的层次式索引结构存储机制

对于层次式的索引结构,与DIMENSION, DIM相同,我们采用索引树的存储机制来构建层次式的存储结构.考虑索引树结构的存储,将整个索引树放到单个节点上的方案是不明智的,这样容易出现单点失效(single point of failure)的问题,不利于存储节点的负载均衡性和高可用性需求.因此,我们考虑将索引树中的各个节点信息存放在多个分布式的传感器节点上,如图 3 所示,以一个 3 层的索引树为例,我们将每层的索引节点分别存放在不同的环结构上,按照由高到低的分辨率,环 1 存放着子区域 $L_1, L_2, \dots, L_m$ 上对事件 $E_1, E_2, \dots, E_k$ 的精确数据,环 2 存放

着每个子区域对事件 $E_1, E_2, \dots, E_k$ 的概要数据,环3存放着对所有子区域相关事件的综合的概要数据,概要数据的类型可以包括平均值(average)、最小值(min)、最大值(max)等类型.这些索引节点的存储信息分别适用于不同的查询类型,包括精确数据查询、不同粒度的聚集查询、交互式的由粗到细的查询.各环的半径大小关系 $R_1, R_2, R_3$ 可以根据用户的实际需求决定,也可以根据计算最优环半径参数结果而得,我们将在第3.4节讨论该问题.各索引树存储节点之间通过转发节点连接,通过这种方式在传感器网络的分布式结构上构建出一个层次式的索引树结构.在传感器网络的监测过程中,随着事件的产生将对索引树结构的相应节点进行更新,不同类型的查询请求也根据其具体需求对相应层次的索引存储节点进行查找.考虑到负载均衡和高可用性的因素,可以在层次式的环结构上提供索引树结构的冗余复制,复制的存储节点之间通过环结构进行同步和更新操作.假设环结构上提供的信息冗余度为 $n$ ,即存在 $n$ 个复制索引树结构,我们基于角度来衡量环结构的信息冗余程度,即环上每隔角度 $\theta$ 部署一个索引树结构,其中有 $\theta=2\pi/n$ .

### 3.2 基于环结构的事件存储和查询路由方案

通过第3.1节介绍了基于环的层次式索引结构的数据存储机制,可以看到,该存储机制是对传统以数据为中心多维存储机制的扩展.对于这样的层次式索引结构,需要考虑如何构建高效节能的路由策略来支持事件的存储和查询.利用环和索引树的结构特性,我们提出基于环结构的事件存储和查询路由方案.图4展示了基于环结构的事件存储和查询路由方案,可以看到,基于环结构的路由路径包括两种:(1) 向心路径:指向环的圆心的路由路径,可以同时沿向心和离心两个方向延伸;(2) 绕环路径:沿环结构作顺时针或逆时针的路由路径.对于事件存储或查询的路由路径,会同时由这两种路径组成.下面,我们分别介绍事件存储的路由策略与查询的路由策略.

对于事件存储的路由方案,事件节点 $S_e$ 感知到相关事件 $E_i$ 后,根据 $S_e$ 在多个环结构中的位置,首先会沿向心路径的单个/两个方向路由传输事件信息,路由到相应的环结构后,沿环绕路径顺时针或逆时针方向传输事件信息,当到达环上索引结构中对应的存储节点时,存储或更新该节点上的事件信息.若网络中存在复制的索引树结构,则在绕环路由时需要遍历更新所有的复制存储节点.

对于事件查询的路由方案,查询节点 $S_q$ 发出查询 $Q_i$ 后,根据查询 $Q_i$ 的类型,首先会沿向心路径向对应层次的环结构路由传输查询请求,到达对应环结构后,沿绕环路径顺时针或逆时针方向传输查询信息,寻找对应的事件信息,当到达环上索引结构中对应的存储节点时,获取相应的查询结果.完成查询后,存在两种路由策略:(1) 沿原有路径返回;(2) 寻找到查询节点 $S_q$ 最近的路由路径返回.

在整个路由过程中,所有节点使用地理路由协议GPSR根据节点的位置信息路由数据包,文献[20]对于在曲线上的路由方案给出了详细的介绍,本文在设计绕环路由的具体实现时参考了文献[20]的曲线路由方案.

### 3.3 基于环的层次索引结构的构建与维护

本节介绍如何在传感器网络分布式的环境中构建和维护层次式的多环索引结构,具体包括环的构建以及索引树的构建.

#### (1) 环的构建

对于环结构的构建,我们以部署区域中心处存在基站和不存在基站两种场景分别考虑:如果在部署区域中心处存在基站,则在构建环结构之前由基站发送事先计算好的环结构参数 $(R_1, R_2, \dots, R_d)$ 的广播报文,每个节点 $S$ 接收到广播报文后,分别统计到基站的跳数 $h_i$ ,同时将跳数 $h_i$ 递增( $h_i=h_i+1$ ),向外层节点转发报文.假设节点间平均距离为 $r$ ,则跳数为 $\lceil R_1/r \rceil, \lceil R_2/r \rceil, \dots, \lceil R_d/r \rceil$ 的节点分别为环1, 2, ...,  $d$ 上的节点.对于部署区域中心处不存在基站

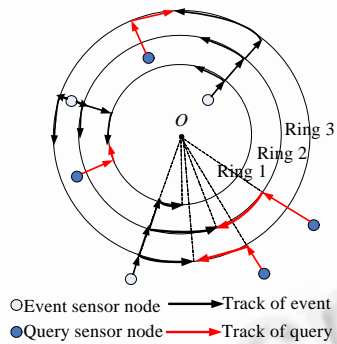


Fig.4 Routing strategy for event storage and query  
图4 事件存储和查询的路由策略

的情况,可以事先利用部署区域的地理位置信息计算出中心位置坐标 $(X_0, Y_0)$ ,在构建环结构之前,每个节点根据其位置 $(X, Y)$ 计算到中心位置 $(X_0, Y_0)$ 的距离.这里,我们定义距离函数 $D((X, Y), (X_0, Y_0))$ :

$$D((X, Y), (X_0, Y_0)) = \sqrt{(X - X_0)^2 + (Y - Y_0)^2} \quad (1)$$

若对于节点 $S$ 有 $|D((X, Y), (X_0, Y_0)) - R_i| \leq R_i \cdot \alpha\%$ (其中, $\alpha\%$ 为预设的误差参数,默认值 $\alpha\% = 5\%$ ),则当前节点设为环 $i$ 上的存储节点.对于上述两种场景,该阶段完成之后,环上每个节点向一跳内的周围邻居节点发送广播寻找同一环上的邻居节点,收到该广播的环上节点进行反馈构建环上邻居关系,由此,同一环上所有节点可以采用基于GPSR的曲线路由机制<sup>[20]</sup>围绕环结构顺时针或逆时针方向建立路由联系.

(2) 索引树的构建

完成环结构的构建之后,需要考虑如何在环结构上部署索引树结构来有效地支持多分辨率的事件存储和查询任务.我们根据用户对 $d$ 个不同层次存储结构的相对查询频率 $q_1, q_2, \dots, q_d$ 来将不同层次的索引存储节点部署到不同半径的环结构上,根据查询频率递增地排序,我们依次将对

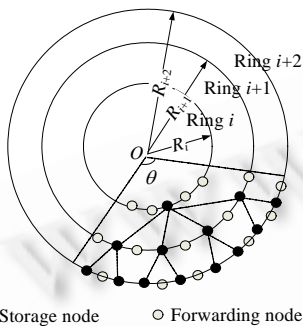


Fig.5 Construct index tree over the rings  
图5 在环结构上构建索引树

对应层次存储结构部署在由内至外半径依次递增的环结构上.为了便于后文分析,这里假设存在 $q_1 < q_2 < \dots < q_d$ ,从而有对应的 $R_1 < R_2 < \dots < R_d$ .通过后面第3.4节的论述可以验证这种部署策略能够有助于使用优化的部署参数,是高效节能的.考虑到负载均衡和高可用性的需求,假设存在 $n$ 个复制索引树结构,则需要在环上每隔 $\theta$ 角度部署一个索引树结构,其中有 $\theta = 2\pi/n$ .如图5所示,传感器网络在每个以 $\theta$ 角度扇形的区域内分别部署索引树结构,在每个扇形区域的多个环结构上,传感器网络随机均匀地选择节点成为该区域的索引存储节点.对于该部署机制的分布式实现,在最内环上每个节点 $S_i$ 产生一个随机数 $N_i$ ,通过绕环路路由传播并比较出数值最大的随机数 $N_{max}$ ,此时将产生 $N_{max}$ 的节点 $S_k$ 设为地标

(landmark)节点 $S_{L_1}$ .  $S_{L_1}$ 以向心路径逆方向路由由接触到多个外环对应的地标节点,分别将其设为 $S_{L_2}, S_{L_3}, \dots, S_{L_d}$ ,每个环上的地标节点沿绕环路路由由 $\theta$ 角度,均匀地选择环上节点成为索引树的存储节点,每隔 $\theta$ 角度完成对一个索引树结构的部署.如图5所示,环结构上部署完成的存储节点间由中转节点衔接保持互联关系.

3.4 选择优化的环结构参数

衡量传感器网络以数据为中心的存储查询方案的一个重要性能指标便是其节能特性,对于基于环结构的多分辨率数据存储机制,为达到最小化的节能效果,我们需要对环结构的相关参数的优化问题进行研究.

为了计算环结构的优化参数,我们首先需要考察事件存储和查询对每个环结构的访问频率.考虑层次结构每个环上事件存储的访问频率.由于每个事件需要更新索引结构上每层相关节点,其对每个环的访问频率都是固定的,因此将每个环的总体访问频率设为 $f$ ,根据第2.2节系统模型描述,有 $f = \sum_{i=1}^k f_i$ ,即为每种事件发生频率之和.对于事件查询的访问频率,针对查询类型的不同,对每个环的访问频率也不同.通常情况下,不同类型的查询只对相应层次的环结构进行访问.由此,我们根据第2.2节的系统模型定义对 $d$ 种不同层次环结构的查询访问频率分别为 $q_1 \cdot f, q_2 \cdot f, \dots, q_d \cdot f$ ,其中, $q_i$ 为相对查询频率.根据第2.2节系统模型描述,我们用事件存储查询路由路径的长度来表示其能量开销,下面我们从小量化传感器网络总体能耗的角度来对环的最优参数进行分析.

我们来计算部署区域内部节点到环结构的平均距离 $D$ ,如图6所示.

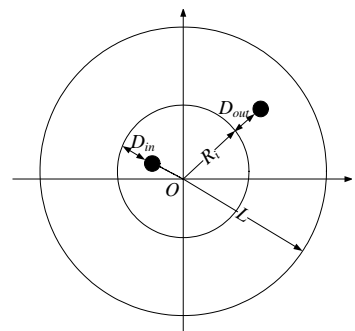


Fig.6 Model of the ring-based structure  
图6 环结构的模型

(1) 环内节点到环*i*的平均距离 $D_{in}$ :

$$D_{in}(R_i) = \frac{\int_0^{R_i} (R_i - r) \left( \int_0^{2\pi} r d\theta \right) dr}{\int_0^{2\pi} \frac{1}{2} R^2 d\theta} = \frac{1}{3} R_i \quad (2)$$

(2) 环外节点到环*i*的平均距离 $D_{out}$ :

$$D_{out}(R_i) = \frac{\int_{R_i}^L (r - R_i) \left( \int_0^{2\pi} r d\theta \right) dr}{\int_0^{2\pi} \frac{1}{2} L^2 d\theta - \int_0^{2\pi} \frac{1}{2} R^2 d\theta} = \frac{2}{3} \left( \frac{L^2 + R_i^2 + LR_i}{L + R_i} \right) - R_i \quad (3)$$

因此,环内、环外所有部署节点到环*i*的平均距离*D*为

$$D(R_i) = \frac{\int_0^{2\pi} \frac{1}{2} R_i^2 d\theta}{\int_0^{2\pi} \frac{1}{2} L^2 d\theta} \cdot D_{in}(R_i) + \frac{\int_0^{2\pi} \frac{1}{2} L^2 d\theta - \int_0^{2\pi} \frac{1}{2} R_i^2 d\theta}{\int_0^{2\pi} \frac{1}{2} L^2 d\theta} \cdot D_{out}(R_i) = \frac{2}{3} \cdot \frac{R_i^3}{L^2} + \frac{2}{3} L - R_i \quad (4)$$

考虑基于环结构的多分辨率存储查询模型整体能量消耗 $E_{overall}$ ,由于在各环上存储事件信息时共享向心路由的路径,在这里我们单独将共享的向心路由能耗 $E_{centri}$ 抽取出来计算,对于各环上的路由能耗 $E(R_i)$ 只包括事件存储的绕环路由能耗 $E_{producer\_round}(R_i)$ 和事件查询能耗 $E_{consumer}(R_i)$ :

$$E_{overall} = E_{centri} + \sum_{i=1}^d E(R_i) = E_{centri} + \sum_{i=1}^d (E_{producer\_round}(R_i) + E_{consumer}(R_i)) \quad (5)$$

下面,我们分析环的半径 $R_i$ 以及冗余复制数*n*这两个环结构参数对总体性能的影响.

事件存储绕环路由由平均每次能量开销为  $2\pi R_i \cdot P(n)$ ,其中 $P(n)$ 为绕环路由的比例参数,这是由于复制存储机制的存在,对应事件信息需要更新到所有的复制存储节点,结合*n*个复制节点的均匀分布性,可以验证,在平均情况下:

$$P(n) = 1 - \frac{1}{2n} = \frac{2n-1}{2n} \quad (6)$$

根据第 2.2 节系统模型描述,假设事件存储的对环*i*的访问频率为*f*,因此有

$$E_{producer\_round}(R_i) = 2\pi R_i \cdot \frac{2n-1}{2n} \cdot f \quad (7)$$

对于事件查询在环结构上的平均能量开销,主要包括向心路由的能耗开销 $D(R_i)$ 和绕环路由的能耗开销 $2\pi R_i \cdot P'(n)$ , $P'(n)$ 为绕环路由的比例参数,由于复制存储机制的存在,查询只需要访问到任意一个复制存储节点即可,可以验证,在平均情况下:

$$P'(n) = \frac{1}{2n} \quad (8)$$

这里我们假设查询完成后采用沿原路返回的路由策略,结合事件查询对环*i*的相对访问频率 $q_i$ ,因此有

$$E_{consumer}(R_i) = 2q_i \left[ D(R_i) + \frac{2\pi R_i}{2n} \right] \cdot f \quad (9)$$

根据上述分析结果,环*i*上事件存储与查询的整体平均能耗开销(除去事件存储共享向心路由的能耗):

$$E(R_i) = E_{producer\_round}(R_i) + E_{consumer}(R_i) = 2\pi R_i \cdot \frac{2n-1}{2n} \cdot f + 2q_i \left[ D(R_i) + \frac{2\pi R_i}{2n} \right] \cdot f \quad (10)$$

带入公式(4)中 $D(R_i)$ 的结果,我们有,

$$E(R_i) = \frac{4q_i}{3L^2} \cdot f \cdot R_i^3 + \left[ \frac{\pi(2n-1+2q_i)}{n} - 2q_i \right] \cdot f \cdot R_i + \frac{4Lq_i}{3} \cdot f \quad (11)$$

对于事件存储共享向心路由的能耗 $E_{centri}$ ,考虑到事件节点相对多环结构的位置可能存在 3 种情况:多环之内、多环之间以及多环之外,由此综合考虑向心路由能耗的期望值有

$$\begin{aligned}
 E_{centri} &= [D_{in}(R_1) + (R_d - R_1)] \times \frac{R_1^2}{L^2} \cdot f + (R_d - R_1) \times \frac{R_d^2 - R_1^2}{L^2} \cdot f + [D_{out}(R_d) + (R_d - R_1)] \times \frac{L^2 - R_d^2}{L^2} \cdot f \\
 &= \frac{1}{3L^2} \cdot f \cdot (R_1^3 + R_d^3) + \frac{2}{3} L \cdot f - f \cdot R_1
 \end{aligned} \quad (12)$$

由公式(12)可以看到,  $E_{centri}$  仅与最内环半径  $R_1$  和最外环半径  $R_d$  相关, 因此, 我们将整体的优化问题拆分为  $R_1, R_d$  和  $R_i (i=2, \dots, d-1) (R_1 < R_i < R_d)$  的优化问题, 其中, 对于  $R_i$  参数 ( $i=2, \dots, d-1$ ) 的优化问题, 由于各环结构的能耗仅与其半径  $R_i$  相关,  $R_i$  对其他环的能耗并不造成影响, 为使整体能耗  $\sum_{i=2}^{d-1} E(R_i)$  最低, 我们可以逐个单独地考虑使每个环结构上能耗最小化的参数  $R_i$ ; 对于  $R_1, R_d$  的参数优化问题, 我们在满足  $R_1 \leq R_2, R_d \geq R_{d-1}$  前提下考虑  $R_1, R_d$  的优化取值以使其整体相关能耗  $E'$  取值最小。

我们将  $E(R_1), E(R_d), E_{centri}$  综合相加计算  $E'$ , 有

$$E' = E(R_1) + E(R_d) + E_{centri} \quad (13)$$

将  $E(R_1), E(R_d)$  根据公式(11)的计算结果以及  $E_{centri}$  根据公式(12)的计算结果代入公式(13), 得到  $E'$ , 我们有

$$\begin{aligned}
 E' &= \frac{4q_1 + 1}{3L^2} \cdot f \cdot R_1^3 + \left[ \frac{\pi(2n-1+2q_1)}{n} - 2q_1 - 1 \right] \cdot f \cdot R_1 + \\
 &\quad \frac{4q_d + 1}{3L^2} \cdot f \cdot R_d^3 + \left[ \frac{\pi(2n-1+2q_d)}{n} - 2q_d \right] \cdot f \cdot R_d + \frac{2L}{3} \cdot f \cdot (2q_1 + 2q_d + 1)
 \end{aligned} \quad (14)$$

根据公式(11)和公式(14)的表述, 我们首先分析冗余复制数  $n (n \geq 1, n$  为整数) 对性能的影响, 为求得整体最低能耗, 将  $E(R_i)$  或  $E'$  对  $n$  求偏导数得:

$$\frac{\partial E(R_i)}{\partial n} = \frac{\partial \left( \frac{\pi R_i f \cdot (2n-1+2q_i)}{n} \right)}{\partial n} = 0 \Rightarrow \frac{\pi R_i f \cdot (1-2q_i)}{n^2} = 0 \quad (15)$$

从公式(15)可以看出, 如果  $1-2q > 0$ , 即当  $q < 1/2$  时,  $E$  随  $n$  的增大而增大, 此时, 当  $n=1$  时效果最优; 如果  $1-2q < 0$ , 即  $q > 1/2$ , 则  $E$  随  $n$  的增大而减小, 此时, 当  $n \rightarrow \infty$  时节能效果最优, 但考虑到传感器节点的存储能力是有限的, 因此  $n \rightarrow \infty$  是不可行的, 我们只有在满足节点存储能力的前提下尽可能地提高  $n$ . 下面, 我们将分别考虑最优半径参数的取值。

对于  $R_i (i=2, \dots, d-1)$  参数 ( $R_1 < R_i < R_d$ ) 的优化问题, 我们根据冗余复制数目  $n (n \geq 1, n$  为整数) 结合  $q_i$  的不同取值情况 ( $q_i > 0$ ) 分别加以讨论:

(1) 当  $n$  满足  $1 \leq n < \pi$  时, 存在  $\frac{\pi(2n-1)}{2(n-\pi)} < 0$ , 此时  $q_i > \frac{\pi(2n-1)}{2(n-\pi)}$  恒成立, 公式(11)中  $R_i$  项的参数为非负数,  $E(R_i)$

与  $R_i$  呈单调递增关系, 为保证总体能耗最低, 此时,  $R_i=0$  时使  $E(R_i)$  取得最小值, 因此, 在这种条件下  $R_i$  应越小越好。

(2) 当  $n$  满足  $n > \pi$  时, 存在  $\frac{\pi(2n-1)}{2(n-\pi)} > 0$ , 此时有:

当  $0 < q_i \leq \frac{\pi(2n-1)}{2(n-\pi)}$  时, 公式(11)中  $R_i$  项的参数为非负数,  $E(R_i)$  与  $R_i$  呈单调递增关系, 为保证总体能耗最低,

此时,  $R_i=0$  时使  $E(R_i)$  取得最小值。

当  $q_i > \frac{\pi(2n-1)}{2(n-\pi)}$  时, 公式(11)中  $R_i$  项的参数为负数,  $E(R_i)$  随  $R_i$  递减后又递增. 因此, 我们需要通过求取极值的方式

来求取最优解  $R_i^*$ , 我们将  $E(R_i)$  对  $R_i$  求偏导数得:

$$\frac{\partial E(R_i)}{\partial R_i} = \frac{4q_i}{3L^2} \cdot 3R_i^2 + \left[ \frac{\pi(2n-1+2q_i)}{n} - 2q_i \right] = 0 \quad (16)$$

通过公式(17)求得最优解  $R_i^*$ :

$$R_i^* = \sqrt{\frac{1}{2} - \frac{2n-1+2q_i}{2n} \cdot \frac{\pi}{2q_i}} \cdot L \quad (17)$$



从公式(17)可以看出,此时最优解  $R_i^*$  的取值与  $q_i$  呈单调递增关系,由于我们在部署层次索引结构时实施了按照访问频率  $q_i$  递增关系由内至外部署环结构的策略,即存在  $q_2 \leq q_3 \leq \dots \leq q_{d-1}$ ,因此,我们可以确保  $R_2^* \leq R_3^* \leq \dots \leq R_{d-1}^*$  的结果.

对于环结构的半径  $R_1, R_d$  的优化问题,我们根据冗余复制数目  $n(n \geq 1, n$  为整数) 结合  $q_i$  的不同取值情况 ( $q_i > 0$ ) 分别加以讨论:

(1) 当  $n$  满足  $1 \leq n < \pi$  时,对于  $R_1$  存在  $\frac{\pi(2n-1)-n}{2(n-\pi)} < 0, q_1 > \frac{\pi(2n-1)-n}{2(n-\pi)}$  恒成立;对于  $R_d$  存在  $\frac{\pi(2n-1)}{2(n-\pi)} < 0, q_d > \frac{\pi(2n-1)}{2(n-\pi)}$  恒成立.这样,公式(14)中  $R_i$  项 ( $i=1, d$ ) 的参数为非负数,此时,  $E(R_i)$  与  $R_i$  呈单调递增关系,为保证总体能耗最低,当  $R_i=0$  时,使  $E(R_i)$  取得最小值.

(2) 当  $n$  满足  $n > \pi$  时,对于  $R_1$  存在  $\frac{\pi(2n-1)-n}{2(n-\pi)} > 0$ ,对于  $R_d$  存在  $\frac{\pi(2n-1)}{2(n-\pi)} > 0$ ,此时有:

当  $0 < q_1 \leq \frac{\pi(2n-1)-n}{2(n-\pi)}$  或  $0 < q_d \leq \frac{\pi(2n-1)}{2(n-\pi)}$  时,公式(14)中  $R_i$  项 ( $i=1, d$ ) 的参数为非负数,  $E(R_i)$  与  $R_i$  呈单调递增关系,为保证总体能耗最低,当  $R_i=0$  时,使  $E(R_i)$  取得最小值.

当  $q_1 > \frac{\pi(2n-1)-n}{2(n-\pi)}$  或  $q_d > \frac{\pi(2n-1)}{2(n-\pi)}$  时,公式(14)中  $R_i$  项 ( $i=1, d$ ) 的参数为负数,  $E(R_i)$  随  $R_i$  递减后又递增.因此,我们将  $E'$  对  $R_i$  求偏导数来求取极值  $R_i'$ , 得到

$$\begin{cases} \frac{\partial E'}{\partial R_1} = 0 \\ \frac{\partial E'}{\partial R_d} = 0 \end{cases} \Rightarrow \begin{cases} R_1' = \sqrt{\frac{2(n-\pi)q_1 + n - 2\pi n + \pi}{4nq_1 + n}} \cdot L \\ R_d' = \sqrt{\frac{2(n-\pi)q_d - 2\pi n + \pi}{4nq_d + n}} \cdot L \end{cases} \quad (18)$$

为了确保  $R_1 \leq R_2, R_d \geq R_{d-1}$  的约束关系,对于  $R_1, R_d$  的最优值  $R_1^*, R_d^*$  的取值有,

$$R_1^* = \min(R_1', R_2^*), R_d^* = \max(R_d', R_{d-1}^*).$$

通过上述的环结构半径最优值求解,我们最终可以确保  $R_1^* \leq R_2^* \leq \dots \leq R_d^*$  的结果.当存在一个或多个  $R_i^* \rightarrow 0 (i=1, 2, \dots, k, 1 \leq k \leq d)$  的情况时,此时考虑到有效分摊各区域节点负载的需求,我们定义一个最小环间隔的参数  $\Delta d$ , 存在当  $k < d$  时有  $0 < \Delta d < R_{k+1}^*/k$ , 当  $k=d$  时有  $0 < \Delta d < L/d$ . 此时可以设定  $R_1 = \Delta d, R_{i+1} = R_i + \Delta d (i=1, 2, \dots, k, 1 \leq k \leq d)$ . 可以验证该机制对多个  $q_i$  相等或逼近相等导致  $R_i^*$  充分接近的情形同样适用.  $\Delta d$  的设定可以根据具体应用的节能需求来设定,通常  $\Delta d$  越小,其节能效果越逼近最优值,但其负载均衡性能越差;反之,  $\Delta d$  越大,其负载均衡性能越优,但同时节能效果越偏离最优值.

## 4 性能分析

本节我们对基于环结构的传感器网络多分辨率存储查询机制的相关性能进行理论分析,分析过程中,我们从下列几个方面进行性能评价:(1) 事件存储和事件查询的能耗开销以及网络总体能耗开销;(2) 聚集查询的性能;(3) 网络能耗负载均衡性能.

### 4.1 环结构存储查询机制的性能

根据第 3.4 节的分析,在选择优化的环结构参数后,我们最终得到的整体网络的能耗为  $E_{overall}$ , 有

$$E_{overall} = E_{centri}(R_1^*, R_d^*) + \sum_{i=2}^{d-1} E(R_i^*) = E_{centri}(R_1^*, R_d^*) + \sum_{i=2}^{d-1} (E_{producer-round}(R_i^*) + E_{consumer}(R_i^*)) \quad (19)$$

对于同样的参数,如果使用基于 GHT 结构化的层次结构存储机制,其整体网络的能耗  $E_{overall}$  包括事件存储开销和查询开销,在最坏情况下,所有的存储和查询开销都为整个部署区域的直径长度,因此,其能耗为

$$E_{overall} = f \cdot \sum_{i=1}^d (2L + 4L \cdot q_i) \quad (20)$$

考虑结构化层次结构的存储机制在平均情况下的能耗开销,图 7 展示了层次结构存储机制中某一层的存储节点部署结构.可以看到,在结构化的存储机制中,其存储节点是均匀分散在部署区域中的,因此,对不同方位存储节点的平均存储查询路由开销是不同的.我们可以根据其距离中心位置的不同划分为多个近似的环结构,利用公式(4)的运算结果来近似计算其存储和查询开销.假设 1-d层的层次结构中,第*i*层存在*k<sub>i</sub>*种类型的存储节点,每种类型存储节点存在 *n* 个复制节点,则存储节点间距离  $D_i = 2L\sqrt{1/(n \cdot k_i)}$ ,存在的环结构数目为  $n_r = \lceil \log_2(n \cdot k_i / 2) \rceil$ ,每个环的半径近似等于  $0.5D_i, 1.5D_i, \dots, (n_r - 0.5)D_i$ ,则根据式(4)的运算结果,*i*层的能耗开销  $E_i$  可以表示为

$$E_i = f \cdot (1 + 2q_i) \sum_{j=1}^{n_r} \left[ \frac{2}{3} \cdot \frac{((j-0.5) \cdot D_i)^3}{L^2} + \frac{2}{3} L - ((j-0.5) \cdot D_i) \right] \quad (21)$$

因此,平均情况下结构化层次结构的存储机制总体能耗为

$$E_{overall} = \sum_{i=1}^d E_i = f \cdot \sum_{i=1}^d \left[ (1 + 2q_i) \sum_{j=1}^{n_r} \left[ \frac{2}{3} \cdot \frac{((j-0.5) \cdot D_i)^3}{L^2} + \frac{2}{3} L - ((j-0.5) \cdot D_i) \right] \right] \quad (22)$$

通过分析比较上述环结构存储查询方案和基于GHT的结构化存储查询方案我们可以看到,环结构存储查询方案的特点决定了其性能上的优点:(1) 环结构事件

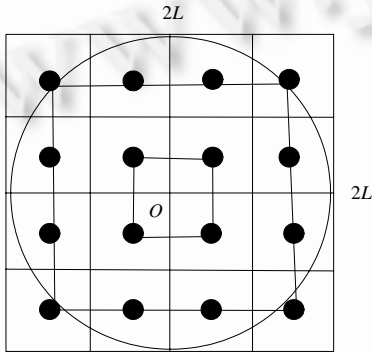


Fig.7 GHT based deployment of the *i*th level storage

图 7 基于 GHT 的第 *i* 层存储节点部署

存储节点放置到节能效果最优化的位置上;(3) 环结构存储查询方案在最坏情况下的事件存储路径长度、事件查询路径长度分别为  $D_s$  和  $D_q$ ,有

$$D_s = \max(R_d, L - R_1) + \sum_{i=1}^d (2\pi R_i) \quad (23)$$

$$D_q = \max(R_d, L - R_1) + \left( \frac{1}{n} \cdot 2\pi R_i \right) \quad (24)$$

而对于结构化的存储查询方案,在最坏情况下,事件存储路径长度和事件查询路径长度分别为  $D'_s$  和  $D'_q$ ,我们有  $D'_s = 2Ldn$  以及  $D'_q = 2L$ .通常情况下,当复制节点冗余度 *n* 足够大时存在  $D_s \ll D'_s, D_q \ll D'_q$ ,可以看到,环结构的存储查询方案相比结构化的存储查询方案能够有效地确保低延迟性和低能耗性.

## 4.2 聚集查询的性能

传感器网络上的聚集查询需要获取区域内部所有事件信息的聚集值,由于环结构提供了多个层次的索引机制,因此存在多种可行的查询方案.如图 8 所示,对于某一区域的聚集查询,存在两条路由路径:路径 1 为直接查询所有相关的原始事件存储节点进行聚集,路径 2 为查询相关的部分聚集结果存储节点进行聚集.我们分析比较了两条路径方案的性能,设查询节点的位置半径为 *R*,到  $R_1$  环的向心路由距离为  $D_{R_1} = |R_1 - R|$ ,到  $R_2$  环的向心路

由距离  $D_{R_2} = |R_2 - R_1|, R_1 < R_2$ , 设单个复制索引结构覆盖的角度为  $\theta$ , 因此有  $E_1 = D_{R_2} + R_2 \times \theta, E_2 = D_{R_1} + R_1 \times \theta$ , 我们可以计算出  $E_1 - E_2 = (R_2 - R_1) \times \theta + (D_{R_2} - D_{R_1})$ , 所以, 当存在  $D_{R_1} - D_{R_2} > (R_2 - R_1) \times \theta$  时, 方案 1 的节能性更优, 否则, 方案 2 的节能性更优, 节点可以根据当前所处的位置决定选择优化的聚集路由方案. 可以看到, 两种查询方案都能够将聚集查询的能耗限制在常数范围之内, 同时, 通过对聚集查询提供多种查询路由方案, 针对传感器网络的节点失效性和链路失效性问题还提供了有效的冗余查询路径, 提高了系统的可用性.

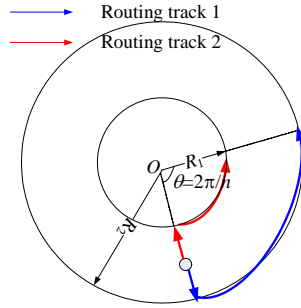


Fig.8 Strategy for aggregate query routing

图 8 聚集查询的路由策略

与传统结构化的存储查询机制相比, 假设层  $i$  存在  $k_i$  个事件存储节点, 提供冗余度为  $n$  的复制节点, 如图 7 所示, 存储节点间距离  $D_i = 2L\sqrt{1/(n \cdot k_i)}$ , 则对于聚集查询而言需要遍历所有这些存储节点, 因此遍历这  $k_i$  个事件存储节点至少需要能耗  $E$ , 有

$$E = D_0 + D \cdot k_i = D_0 + 2L\sqrt{k_i/n} \tag{25}$$

其中,  $D_0$  为查询节点到第 1 个存储节点的距离. 可以看到, 在结构化存储机制中, 聚集查询的能耗  $E$  与存储节点的数目  $k$  密切相关, 而基于环的聚集查询, 其能耗与存储节点的数目没有直接联系, 能够将其约束在常数范围之内, 具有很好的可扩展性.

### 4.3 负载均衡性

考察基于环的层次结构的存储查询机制负载均衡的性能, 我们从以下几个方面对其进行评价: 环上复制存储节点间的负载均衡性; 各环节点的负载均衡性; 环上节点与环外节点的负载均衡性.

对于环上复制存储节点间的负载均衡性, 我们假设事件节点与查询节点是随机均匀分布的, 由于我们在对基于环结构的索引树构建过程中对复制节点的部署也是均匀的, 并且采用的是向心路由与绕环路由相结合的路由策略, 对环结构上的各复制节点的访问频率是相同的, 因此复制节点间的负载是均衡的. 假设事件发生的频率为  $f$ , 查询的频率为  $q_i f$ , 转发一个信息的能耗为  $\varepsilon_m$ , 复制冗余度为  $n$ , 环上节点数为  $N$ , 我们设环  $i$  上各节点 (包括存储节点和普通转发节点) 的平均负载为  $E_{avg}$ , 则有

$$E_{avg} = \frac{f \cdot \left( N \cdot \frac{2n-1}{2n} \cdot \varepsilon_m \right) + 2q_i \cdot f \cdot \left( N \cdot \frac{1}{2n} \cdot \varepsilon_m \right)}{N} = \frac{f \cdot \varepsilon_m}{2n} (2n-1+2q_i) \tag{26}$$

对于各环节点的负载均衡性, 各环  $i$  上节点的负载与查询的相对频率  $q_i$  相关, 由公式 (26) 可知,  $q_i$  越大, 节点能耗  $E_{avg}$  越大. 之前我们假设复制冗余度  $n$  在每个环上都是相同的, 如果我们允许在每个环  $i$  上部署具有差异性的复制冗余度  $n_i$ , 那么通过适当部署冗余值  $n_i$ , 可以有效地平衡各环节点间的负载. 根据公式 (26) 的能耗计算结果计算出均衡各环负载的  $n_i$  值, 其中  $C$  为常数:

$$\frac{1}{2n_i} (2n_i - 1 + 2q_i) = C \Rightarrow n_i = \frac{2q_i - 1}{2C - 2} \tag{27}$$

为了确保总体的存储量不变, 即有  $n \cdot d = \sum_{i=1}^n n_i$ , 将公式 (27) 的结果代入, 得到常数  $C$  的取值:

$$C = \frac{1}{2dn} \sum_{i=1}^d (2q_i - 1) + 1 \quad (28)$$

对于环上节点与环外节点的负载均衡性,由于环外节点只承担向心路由的负载,而环上节点既承担向心路由的负载又承担绕环路由的负载,其负载显然会高于环外节点.但由于复制节点的存在,环上负载能够被有效地降低,因此在适量的事件和查询量的情况下,负载差距并不会很大.同时,我们可以考虑定期地调整环结构的半径来均衡传感器网络内各部分节点的负载.

## 5 性能评价

### 5.1 模拟实验环境设置和参数

本节通过模拟实验来对基于环结构的传感器网络多分辨率数据存储机制的性能进行评估,在模拟实验环境中,节点随机、均匀地分布在半径大小为 $L=50\text{m}$ 的圆形监测区域内,部署节点总数为 $N=1\,000$ ,节点的通信半径为 $r=3\text{m}$ .我们以部署区域的圆心为中心部署一个3层的层次存储结构为例来验证不同方案的性能,在每个层次,我们分别部署 $k_1=1, k_2=4, k_3=16$ 个数据存储节点,并设置节点的冗余度为 $n$ .假设事件的存储频率为 $f=1\,000$ ,对每个层次结构的相对查询频率分别为 $q_1, q_2, q_3$ ,各环的半径为 $R_1, R_2, R_3$ ,具体参数见表1,其相应的默认参数值用default标出.实验中,我们以计算数据包路由路径的长度来统计网络的能量消耗,最终的实验结果是根据随机运行1 000次所得的平均结果计算而得.我们通过以下方面来评估基于环结构存储查询方案的性能:首先,我们通过实验性能对比从总体能耗、平均存储能耗、平均查询能耗几方面来验证环结构存储查询机制的节能效果;然后,通过实验验证环结构最优参数算法的有效性;最后,通过实验数据分析来对聚集查询的处理性能以及负载均衡性进行验证.

Table 1 Parameters of simulation

表1 模拟实验参数

Parameter	Value	Parameter	Value
Radius of network ( $L$ )	50 (m)	Communication radius ( $r$ )	3 (m)
Number of nodes ( $N$ )	1 000	Dimension of storages ( $d$ )	3
Number of replications ( $n$ )	1, 4 (default), 8, 10, 16	$(k_1, k_2, k_3)$	(1, 4, 16)
Event frequency ( $f$ )	1 000	$(q_1, q_2, q_3)$	(5, 10, 20) (default)
Number of local areas ( $m$ )	4 (default), 8, 16	$(R_1, R_2, R_3)$	(5, 15, 25) (default)

### 5.2 环结构存储查询机制的性能验证

图9给出了基于GHT的层次结构存储查询机制与基于环的层次结构存储查询机制在不同冗余度 $n$ 下的总体能耗比较:当 $n=1$ ,即不提供复制存储节点的情况下,GHT方案相比环结构方案能够减少33%的能耗.随着冗余复制程度的增加,GHT方案在 $n=4$ 时达到较低能耗水平后,其能耗随着冗余度的增加而递增,而环结构方案的能耗随着冗余度的增加而一直地逐步递减,在 $n=16$ 时能够比GHT方案减少53%的能耗.这是因为,在冗余度 $n$ 较小的情况下,环结构方案在将数据通过向心路径传输到达环结构后,需要沿绕环路经行走较长的距离才能够到达对应存储节点,该绕环路由开销随着环半径的增大而增加.而GHT方案可以通过直线路径到达指定的存储节点,相比绕环路由具有较小的开销.随着冗余度的增加,GHT方案在事件存储时需要沿着不同路由路径逐一更新复制存储节点,需要大量的路由能耗.相比之下,环结构方案在更新不同复制存储节点时可以共享同一段向心路由路径到达对应环结构,在完成绕环路由过程中可以逐一实现对多个复制节点的更新操作,因此,环结构方案的事件存储能耗并不随冗余度 $n$ 的增加而显著增大.同时,由于两种方案的查询路由能耗都会随冗余度的增加而减少,因此在冗余度较大的情况下,环结构方案能够比GHT方案达到更好的节能效果.为了验证上述分析结果,后文图12、图13进一步地对两种方案的平均存储能耗和平均查询能耗进行了对比.图10显示了GHT和环结构方案平均存储能耗对比,随着冗余度的增加,由于事件存储需要更新多个复制节点,GHT方案的存储能耗开销以很快的速度递增,而环结构方案则能够以很小的幅度递增,在 $n=16$ 时,其存储能耗仅相当于GHT方案存储能耗的

12%.图 11 显示了GHT和环结构方案平均查询能耗对比,随着冗余度的增加,两种方案的查询能耗都随之减少,但环结构方案比GHT方案具有更大的减少幅度,在 $n=16$  时,其平均查询能耗仅为 36,此时,GHT方案平均查询能耗为 24,但相比于其存储能耗 2 480,环结构方案仅为 300.因此,GHT方案较低的查询能耗是建立在大量的存储能耗基础之上的.相比而言,环结构方案具有更好的整体节能效果.这里,我们使用了环结构半径 $R_1, R_2, R_3$ 分别为 5,15,25 的默认值,通过选择优化的半径参数,我们能够得到更为优化的节能效果.

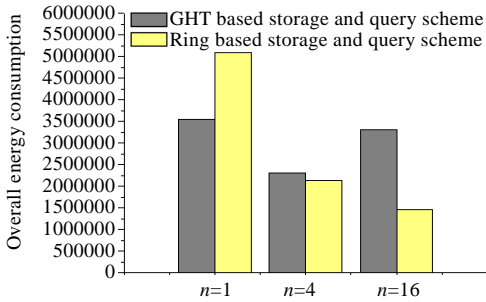


Fig.9 Overall energy consumption comparison

图 9 总体能耗对比

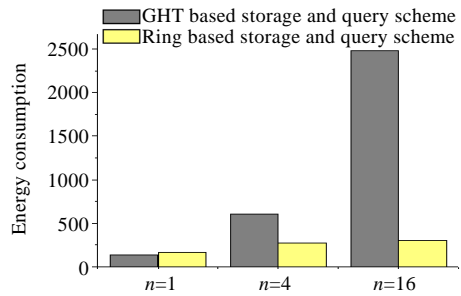


Fig.10 Energy consumption for event storage

图 10 事件存储能耗

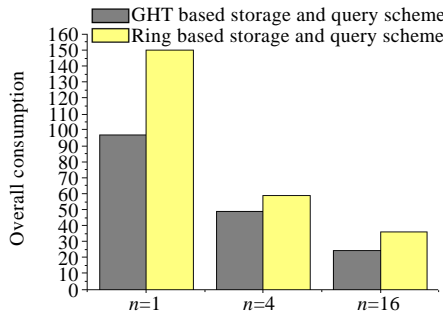


Fig.11 Energy consumption for event query

图 11 事件查询能耗

### 5.3 环结构最优参数的验证

本节将通过实验验证在不同参数 $(n, q_i)$ 情况下,环结构半径 $R_i$ 的最优取值.为了能够展示环结构最优半径的取值效果,我们使用一个两层的索引树结构( $d=2$ )来展示在不同半径取值情况下的能耗情况.如图 12 所示,我们在 $n=16, q_1=10, q_2=20$  的参数环境下对不同半径取值的能耗情况进行观察,可以看到,整体能耗在 $R_1=0.25 \times L, R_2=0.35 \times L$ 的情况下能够取得最低值.以此为最低点,整体能耗向其他半径范围逐步提升.根据第 3.4 节提出的环结构优化参数算法的分析结果,此时符合 $n > \pi, q_i > \frac{\pi(2n-1)}{2(n-\pi)}$ 的情况,有  $0 < R^* < L$ .通过公式(16)和公式(18)的计算

结果,可以验证该最优半径参数序列 $(R_1^*, R_2^*)$ 的取值与环结构优化参数算法的结果正好符合.如图 13 所示,我们在 $n=8, q_1=2, q_2=4$  的参数环境下对不同半径取值的能耗情况进行观察,可以看到,此时整体能耗在接近 $(0,0)$ 点处取得最低值,以 $(0,0)$ 点为最低点向其他半径范围逐步提升.该结果正好验证了第 3.4 节环结构优化参数算法的分析结果,此时符合 $n > \pi, q_i < \frac{\pi(2n-1)}{2(n-\pi)}$ 的情况,有 $R^* \rightarrow 0$ .通过上述两个实验结果的验证,表明本文提出的环结构优

化参数算法能够准确地分析并计算出环结构的最优参数,从而辅助环结构在传感器网络内部的部署,以充分实现最小化的节能效果.

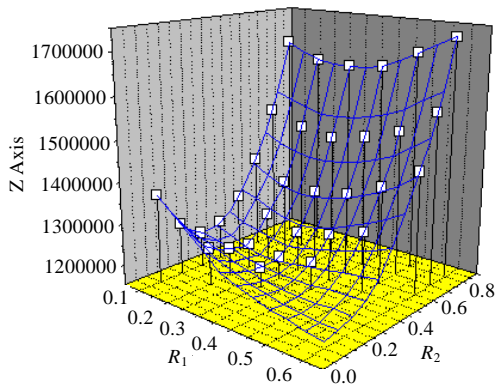


Fig.12 Optimal parameters for minimized energy consumption with ring based scheme ( $0 < R^* < L$ )  
图 12 最小化整体能耗的环结构优化参数选择 ( $0 < R^* < L$ )

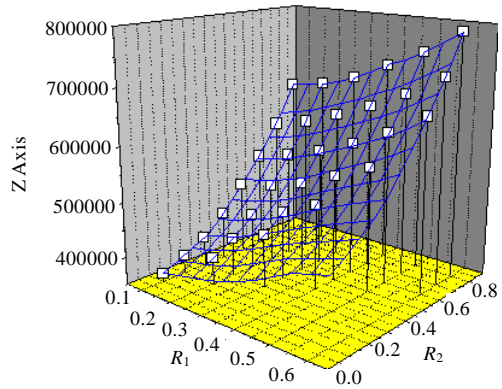


Fig.13 Optimal parameters for minimized energy consumption with ring based scheme ( $R^* \rightarrow 0$ )  
图 13 最小化整体能耗的环结构优化参数选择 ( $R^* \rightarrow 0$ )

5.4 聚集查询性能以及负载均衡性的验证

图 14 给出了 GHT 和环结构方案及其优化方案的聚集查询能耗对比.随着冗余度的增加,3 种方案的聚集查询能耗都随之减少,这是由于复制存储节点的存在使得聚集查询能够在局部区域完成.但相对于 GHT 方案,环结构方案能够显著减少聚集查询能耗,其优化方案能够进一步降低其能耗,在  $n=16$  时,环结构方案能够比 GHT 方案减少 85% 的能耗,同时,环结构优化方案能够比普通环结构方案进一步减少 23% 的能耗,这主要是由于事件存储过程中向心路径的共享与聚集查询过程绕路径的共享机制使得环结构方案能够达到更好的节能效果.而环结构优化方案能够根据查询节点的实际位置决定其聚集查询的路由策略,从而使用一条更为节能的路由路径来收集聚集查询数据.

图 15 中,我们对环结构的负载均衡性进行验证.在实验中,由于  $q_1, q_2, q_3$  的取值 5, 10, 20 存在差异,导致各环上的负载产生不均衡的现象.当各环冗余度取均一值  $n=8$  时,各环上的负载分别为 1 281, 1 593, 2 218.此时,使用第 4.3 节提出的非均一冗余度机制可以有效地均衡各环的负载,在确保整体冗余存储量不变 ( $3n=24$ ) 的前提下,我们根据公式(28)重新分配各环的冗余度,得到  $n_1=3, n_2=7, n_3=14$ ,此时,通过实验结果我们可以看到,各环的负载被有效均衡,我们得到接近均一的负载 1 697.

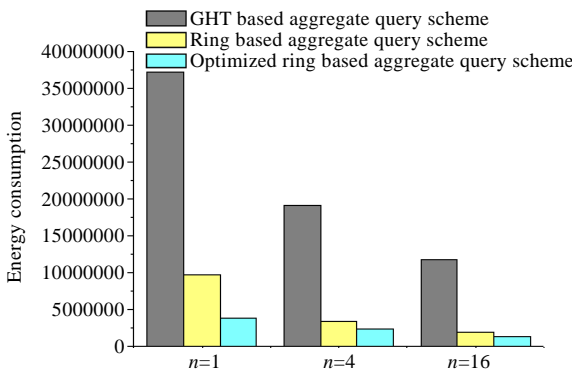


Fig.14 Energy consumption for aggregate query in GHT based scheme vs ring based scheme  
图 14 GHT 和环结构方案聚集查询能耗对比

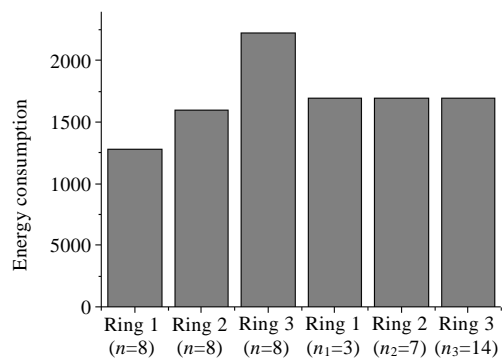


Fig.15 Energy consumption comparison for different rings  
图 15 不同环结构上的能耗对比

## 6 总 结

本文提出了一套基于环结构的传感器网络多分辨率数据存储机制,结合层次结构的存储查询方案,有效地利用了环结构的特性来高效节能地支持事件信息的不同分辨率的存储和查询操作,并采用优化的环结构参数以最小化传感器网络的总体能量开销.同时,本文对环结构多分辨率数据存储机制的相关性能进行了理论分析,从节能性、负载均衡性等多个角度对其进行了具体的分析.模拟实验表明,基于环结构的传感器网络多分辨率数据存储机制能够达到显著的节能效果.对于下一步研究工作,我们将考虑采用实际的传感器网络实验环境来进一步验证该存储机制的性能与特性,结合更多的实际因素进行研究,以更全面地提出一套性能高效的传感器网络多分辨率数据存储机制.

### References:

- [1] 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>
- [2] Shenker S, Ratnasamy S, Karp B, Govindan R, Estrin D. Data-Centric storage in sensornets. *SIGCOMM Computer Communication Review*, 2003,33(1):137-142.
- [3] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In: *Proc. of the 6th Annual Int'l Conf. on Mobile Computing and Networking (MobiCom 2000)*. New York: ACM Press, 2000. 56-67. <http://doi.acm.org/10.1145/345910.345920>
- [4] Ye F, Luo H, Cheng J, Lu S, Zhang L. A two-tier data dissemination model for large-scale wireless sensor networks. In: *Proc. of the 8th Annual Int'l Conf. on Mobile Computing and Networking (MobiCom 2002)*. New York: ACM Press, 2002. 148-159. <http://doi.acm.org/10.1145/570645.570664>
- [5] Ratnasamy S, Karp B, Shenker S, Estrin D, Govindan R, Yin L, Yu F. Data-Centric storage in sensornets with GHT, a geographic hash table. *Mobile Network Application*, 2003,8(4):427-442. <http://dx.doi.org/10.1023/A:1024591915518>
- [6] Ganesan D, Estrin D, Heidemann J. Dimensions: Why do we need a new data handling architecture for sensor networks? *SIGCOMM Computer Communication Review*, 2003,33(1):143-148. <http://doi.acm.org/10.1145/774763.774786>
- [7] Ganesan D, Greenstein B, Perelyubskiy D, Estrin D, Heidemann J. An evaluation of multi-resolution storage for sensor networks. In: *Proc. of the 1st Int'l Conf. on Embedded Networked Sensor Systems (SenSys 2003)*. New York: ACM, 2003. 89-102. <http://doi.acm.org/10.1145/958491.958502>
- [8] Greenstein B, Estrin D, Govindan R, Ratnasamy S, Shenker S. DIFS: A distributed index for features in sensor networks. *Ad Hoc Networks (ADHOC)*, 2003,1(2-3):333-349.
- [9] Li X, Kim YJ, Govindan R., Hong W. Multi-Dimensional range queries in sensor networks. In: *Proc. of the 1st Int'l Conf. on Embedded Networked Sensor Systems (SenSys 2003)*. New York: ACM Press, 63-75. <http://doi.acm.org/10.1145/958491.958500>
- [10] Guo LJ, Li JZ, Li GL. Spatio-Temporal query processing method in wireless sensor networks. *Journal of Software*, 2006,17(4): 794-805 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/794.htm>
- [11] Trigoni N, Yao Y, Demers AJ, Gehrke J, Rajaraman R. Hybrid push-pull query processing for sensor networks. In: *Proc. of the GI Jahrestagung, Vol.2*. 2004. 370-374.
- [12] Kapadia S, Krishnamachari B. Comparative analysis of push-pull query strategies for wireless sensor networks. In: *Proc. of the Int'l Conf. on Distributed Computing in Sensor Systems (DCOSS)*. 2006.
- [13] Sarkar R, Zhu X, Gao J. Double rulings for information brokerage in sensor networks. In: *Proc. of the 12th Annual Int'l Conf. on Mobile Computing and Networking (MobiCom 2006)*. New York: ACM Press, 2006. 286-297. <http://doi.acm.org/10.1145/1161089.1161122>
- [14] Gao J, Guibas L, Milosavljevic N, Hershberger J. Sparse data aggregation in sensor networks. In: *Proc. of the 6th Int'l Conf. on Information Processing in Sensor Networks (IPSN 2007)*. New York: ACM, 2007. 430-439. <http://doi.acm.org/10.1145/1236360.1236414>

- [15] Liu X, Huang Q, Zhang Y. Combs, needles, haystacks: Balancing push and pull for discovery in large-scale sensor networks. In: Proc. of the 2nd Int'l Conf. on Embedded Networked Sensor Systems (SenSys 2004). New York: ACM, 122-133. <http://doi.acm.org/10.1145/1031495.1031510>
- [16] Shakkottai S. Asymptotics of query strategies over a sensor network. In: Proc. of the 23th Conf. of the IEEE Communication Society (INFOCOM 2004). 2004.
- [17] Fang Q, Gao J, Guibas LJ. Landmark-Based information storage and retrieval in sensor networks. In: Proc. of the 25th Conf. of the IEEE Communication Society (INFOCOM 2006). 2006.
- [18] Ahn J, Krishnamachari B. Fundamental scaling laws for energy-efficient storage and querying in wireless sensor networks. In: Proc. of the 7th ACM Int'l Symp. on Mobile Ad Hoc Networking and Computing (MobiHoc 2006). New York: ACM Press, 2006. 334-343. <http://doi.acm.org/10.1145/1132905.1132942>
- [19] Zhang WS, Cao GH, La Porta T. Data dissemination with ring-based index for wireless sensor networks. In: Proc. of the 11th IEEE Int'l Conf. on Network Protocols (ICNP). Washington: IEEE Computer Society, 2003. 304-314.
- [20] Nath B, Niculescu D. Routing on a curve. SIGCOMM SIGCOMM Computer Communication Review, 2003,33(1):155-160.

#### 附中文参考文献:

- [1] 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>
- [10] 郭龙江,李建中,李贵林.无线传感器网络环境下时-空查询处理方法.软件学报,2006,17(4):797-805. <http://www.jos.org.cn/1000-9825/17/794.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.



陈力军(1961—),男,博士,教授,主要研究领域为移动网络,分布式计算.



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.



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.