动态无线传感网中低延迟高可靠的数据查询机制

梁俊斌"马方强"何宗键"

¹⁾(广西大学计算机与电子信息学院广西多媒体通信与网络技术重点实验室 南宁 530004)
 ²⁾(奧克兰大学网络研究中心 奧克兰 新西兰 1142)

摘 壷 查询处理作为大规模无线传感器网络中智能服务的一个重要操作,可以根据用户需求对网络中的感知数 据进行检索和回传.然而,部署在恶劣环境中的无线传感网络,节点容易遭受外力破坏,或者自身资源(能量、存储 等)有限,可能会导致节点发生位移和故障,从而造成网络拓扑不断改变以及部分节点的感知数据失效.同时,由于 节点感知数据容量大、传输带宽有限以及网络链路不可靠等情况,可能会造成网络通信时延大大增加,这些因素使 得快速、可靠的数据查询处理成为无线传感网中一个难题.为了解决这个难题,提出一种动态网络中低延迟高可靠 的数据查询机制.该机制是一种非聚合随机查询方式,通过将传感节点划分为源节点和查询节点来实现数据查询. 首先,根据监测事件将网络划分为若干个子区域,每个子区域中的源节点相互协作,并按照时间顺序依次轮流监听 该区域的事件信息;接着,源节点根据预估的平均节点故障概率,计算出一个合理的备份数量,并将源数据按照该 数量存储到邻居节点中,以降低源数据的失效概率;然后,为了加快数据查询速度,源节点定期对源数据块进行编 码压缩,并选取剩余能量和存储空间袋小的多个邻居节点作为下一跳接收节点.这些接收节点基于局部区域中节 点个数大小,决定是否接收存储该报文,重复上述过程,直至压缩数据均匀地分布在网络中.另一方面,查询节点接 收到查询请求时,也使用负载均衡多路分发方式将查询请求传输到部分节点上.为了避免目标数据的冗余回传,当 查询请求成功查询到目标数据时,目标节点先修改访问位,再选取与查询节点距离最近的邻居节点作为下一跳接 收节点,迭代执行上述操作,直到用户获得所需要的事件信息.在以上过程中,为了节省节点能量,在保证高成功查 询率的条件下,建立通信能耗最小化的优化模型,计算出最优的压缩数据副本数和查询消息副本数,之后,源节点 和查询节点分别按照该数量进行副本数据分发.最后,理论分析和实验结果表明,与其它四种查询算法相比,提出 的查询机制具有更高的查询成功率、更低的通信能耗和通信时延.

关键词 无线传感器网络;智能服务;编码压缩;优化模型;多路分发 中图法分类号 TP393 DOI号 10.11897/SP.J.1016.2020.00555√

Low Latency and High Reliable Data Query Mechanism in Dynamic Wireless Sensor Networks

LIANG Jun-Bin¹⁾ MA Fang-Qiang¹⁾ HE Zong-Jian²⁾

 ¹⁾ (Guangxi Key Laboratory of Multimedia Communications and Network Technology, School of Computer and Electronic Information Guangxi University, Nanning 530004)
 ²⁾ (Centre for eResearch University of Auckland, Auckland, New Zealand 1142)

Abstract As an important operation of intelligent services in large-scale wireless sensor networks, query processing can retrieve and return perceived data in the network according to user requirements. However, when wireless sensor networks are deployed in harsh environments, nodes are vulnerable to external forces, or its self- resources (energy, storage, etc.) are limited, which may lead to nodes displacement and failure, resulting in constant changes in the network topology and the failure of perceived data of some nodes. At the same time, due to the large capacity of sensor

收稿日期:2018-06-05;在线出版日期:2019-04-19.本课题得到国家自然科学基金(61562005,61762010)、广西自然科学基金(2018GXNSFBA281169,2019GXNSFAA185042)、广西高等学校千名中青年骨干教师培育计划基金(桂教人(2017)No.49)资助.梁俊斌,博士,教授,主要研究领域为无线传感器网络、智能服务.E-mail: liangjb2002@163.com.马方强,硕士研究生,主要研究方向为无线传感器网络、智能服务.fiangjb2002@163.com.马方强,硕士研究生,主要研究方向为无线传感器网络、智能服务.fiermail: liangjb2002@163.com.

data, limited transmission bandwidth and unreliable network links, the network communication delay may be greatly increased. These factors make fast and reliable data query processing a difficult problem in wireless sensor networks. In order to solve this problem, a low latency and high reliability data query mechanism in dynamic networks is proposed. The mechanism is a non-aggregated random query method, which can realize data query by dividing sensor nodes into source nodes and query nodes. Firstly, the network is divided into several sub-regions according to the monitoring events. The source nodes in each sub-region cooperate with each other, and listen to the event information of the region in turn according to the time sequence; then, according to the estimated average node failure probability, the source nodes calculate a reasonable number of backups, and stores the source data in the neighbor nodes according to the number, so as to reduce the failure probability of the source data; secondly, in order to speed up the data query speed, the source nodes periodically encode and compress the source data block, and select multiple neighbor nodes with less residual energy and storage space as the next hop receiving nodes. These receiving nodes decide whether to receive and store the message based on the number of nodes in the local area. The above process is repeated until the compressed data is uniformly distributed in the network. On the other hand, when the query node receives query request, it also uses the method of load balancing multi-path dissemination to transmit the query request to some nodes. In order to avoid the redundant return of the target data, when the query request succeeds in querying the target data, the target node first modifies the access bit, then selects a node with the smallest distance between its neighbor nodes and the query node as the next hop receiving node, and performs the above operation iteratively until the user obtains the required event information. In the above process, in order to save the energy of the nodes, under the condition of ensuring high successful query rate, an optimization model of minimizing communication energy consumption is established, and the optimal number copies of the compressed data and the query request are calculated, then the source nodes and query nodes disseminate its replica data according to the number respectively. Finally, theoretical analysis and experimental results show that compared with other four query algorithms, the proposed query mechanism has higher successful query rate, lower communication energy consumption and communication delay.

Keywords wireless sensor networks; intelligent services; encoding and compression; optimization model; multi-path dissemination

1 引 言

随着无线通信技术和微电子技术的快速发展^[1-2],WSNs(Wireless Sensor Networks)采集的数据容量更加庞大、数据类型更加多样及用户对数据的需求也更加个性化,使得基于智能化服务的数据查询技术已经成为WSNs的发展趋势^[3-4].无线传感器网络(Wireless Sensor Networks,WSNs)是一种由数量庞大、廉价且资源(电池能量、计算能力、内存大小等)有限的传感器节点通过无线多跳的通信方式自组织而成的网络,已经在各个领域有广泛的应用,如灾难监测^[1]、军事监视^[2]、工业气体泄漏检

测[5]等.

智能服务化的数据查询是一种个性化的非结构 化查询方式,可根据用户需求对WSNs采集的各种 数据(如温度、光照、适度、声音、图像等)进行快速、 可靠和低能耗的信息检索^[6-8].与传统的聚合查询方 式相比,它无需事先构建固定的网络拓扑结构及对 数据进行聚合过滤操作,从而有效地避免了单点失 效和节点负载不均等问题^[9-10].

然而,在恶劣的监测环境中,使用 WSNs 对事件信息进行查询往往面临以下几个难点:(1)感知数据容量大,节点存储空间和传输带宽有限;(2)节 点容易遭受外力破坏损毁或位置发生位移,导致网络拓扑改变和感知数据失效^[11-12];(3)网络连通性 差且通信链路不可靠,通信延迟高^[13-14].例如:在地 震救援中,节点容易遭受余震、暴雨、泥石流等外力 破坏而损毁,可能会出现网络拓扑动态变化而连通 差、节点的存储数据失效、通信链路不可靠等情况. 这使得搜救人员在救援定位查询中及时、准确地 获取所需的事件信息更加困难.因此,如何在动态 WSNs中,实现低延迟高可靠的数据查询是一项具 有挑战性的工作.

在本文,提出一种低延迟高可靠的数据查询 (Low Latency and High Available Data Query, LLHADQ)机制,它是一种非聚合随机查询方式.在 数据感知阶段,源节点使用分片布鲁姆过滤器^[15-16] (Division Bloom Filter)定期编码压缩源数据,将其 随机均匀地分发至部分节点上.考虑到恶劣环境可 能导致基站(Base Station,BS)严重损毁而失效,在 数据查询阶段,用户可以随机选取任一存活节点注 入查询消息进行数据查询.当查询消息与编码数据 在某节点匹配成功,则查询消息按照压缩数据分发 路径快速查询到目标数据,将其返回给用户.

为了避免感知数据的过度冗余存储,提出一种 动态时分协作感知存储协议.主要是将监测同一事 件的多个节点组成为一个集合,节点之间通过定期 互相通信,推选出一个头结点(Leader),Leader 节 点为成员节点分配监测时间和顺序.之后,所有节点 依次轮流记录事件数据,并对其进行同步备份存储, 以降低数据的失效率.

为了支持低能耗的数据查询,在用户期望查 询成功概率下,建立通信能耗最小化(Minimizing Communication Energy Consumption, MCEC)优化 模型,计算出最优压缩数据副本个数和查询消息副 本个数.在此基础上,提出一种完全分布式的负载均 衡多路均匀分发算法.发送节点采用定期 Hello 消 息机制^[17-18]获取有效的邻居节点信息,利用其剩余 能量和存储空间大小,选取多个邻居节点作为副本 的接收节点;接收节点则根据自身邻居节点的个数, 决定是否接收该数据副本.从而保证了数据分发的 及时性和均匀性,延长了网络寿命.

此外,还设计了一种低能耗、高可靠的剪枝容灾 路由回传协议,包括剪切冗余路由和路由回传两部 分.前者目标节点通过修改自身访问位剪切网络中 多个相同查询请求的回传路由,避免不必要的目标 数据传输能耗;后者发送节点根据邻居节点与查询 节点所在方向和之间的距离,选取与查询节点所在 方向上和两者之间距离最小的邻居节点作为新回传路由,从而提高了数据回传的速度和可靠性.

本文的主要贡献如下:

(1)提出了一种比较完善且实用性更强的 LLHADQ机制,它分别从数据存储、检索和回传三 个阶段进行优化设计,全面提升了数据查询时的通 信开销、时延及成功查询概率等性能指标;

(2) 建立了 MCEC 优化模型,在保证高查询成 功概率的条件下,计算出最优通信能耗下的压缩数 据副本个数和查询副本个数,并采用负载均衡多路 均匀分发算法加快数据报文的分发速度,从而延长 网络寿命,降低查询搜索时延;

(3)还设计了一种新的剪枝容灾回传路由协议,通过对查询结果回传路径的优化,降低了目标数据回传的通信能耗和时延,提高了查询回传的可靠性.

本文第2节介绍目前已有的相关研究工作;第 3节给出符号说明和相关定义、网络模型和问题描述;第4节详细设计LLHADQ机制;第5节理论分析LLHADQ机制的性能指标;第6节仿真实验及 对比分析;第7节工作总结.

2 相关工作

在大规模 WSNs 中,以数据为中心的分布式查 询方法一直是人们研究的重点.目前,在无线传感器 网络数据查询方面已经存在大量的研究工作,可根 据是否基于簇的网内数据聚合处理,将其分为聚合 查询和非聚合查询两种方式.

2.1 聚合查询

在 WSNs 中,聚合查询是一种利用簇头节点 (Cluster)使用符合用户查询请求的聚合函数,对来 自多个邻居节点的数据进行聚合、过滤操作,并按照 固定的网络拓扑结构将查询结果返回给用户的查询 方式.

Liu 等人^[19]提出一种交叉剪枝框架的聚合查询 算法,分为三个方面:(1)将网络划分为多个子区域, 每个子区的 Cluster 节点采用启发式算法构建一颗 以 Sink 为根的能量最优的生成树(MST);(2)成员 节点将感知数据汇集到 Cluster 节点,Cluster 节点 根据数据的相关性计算阈值范围,并将其广播给成 员节点;(3)用户将查询消息发送至所有 Cluster 节 点,通过局部阈值和查询阈值范围交叉剪枝冗余 数据,直至查询数据传输给 Sink 节点.该算法节省 了查询时间和能耗,但是构建和维护 MST 的通信 能耗较大,且不适合感知数据变化幅度较大的数 据查询.

基于上述问题, Cheng 等人^[20]提出一种新的聚 合查询算法.主要思想: Sink 节点和 Cluster 节点根 据之前 *k* 个阈值动态地更新查询阈值范围,确定最 优查询阈值.从而避免遗漏最近感知的新数据.该算 法需要频繁计算阈值范围,导致 Cluster 节点能耗 过大而死亡.

在以上两个算法基础上,Sarode 等人^[5]提出一 种自适应剪枝聚合查询(APDA)算法.它根据空间 位置理论^[21],认为位置相近的节点所感知的数据值 差异很小,子区域 Cluster 节点定期汇集成员节点 数据,计算新的阈值,并与查询阈值进行比较,剪切 掉不符合查询请求的子区域节点数据.APDA 算法 避免了不必要的数据传输,节省了通信能耗,加速了 查询时间.但是,需要用户对感知数据有所了解,并 能给出合理的查询阈值范围.

上述几个算法虽然在数据的查询时间和能耗性 能方面有所提升.但是,它们都存在构建和维护 MST 而能耗过大、单点失效、查询结果不精确及网 络负载不均衡等问题.

2.2 非聚合查询

在 WSNs 中,非聚合查询是一种无需对网内数 据进行聚合过滤操作,使用查询请求直接对节点的 本地数据进行查询比对的方法.最简单直接的方法 就是 Flooding^[22],查询节点通过向全网广播查询消 息来搜索目标数据.但是,该算法由于过大的通信能 耗和传输冲突而很少被直接使用.

Kolcun 等人^[23]提出一种基于分布式的静态属 性表(DSAT)查询算法. 它将全局的 DSAT 划分为 P 个数据量级相同的子表,采用 Netchange^[24]启发 算法,将 P-1 个子表广播存储在部分节点上. 在数 据查询时,查询消息只需要根据 Routing Table(RT) 快速查询所有子表,即可找到目标数据. 该算法需要 构建全局的 DSAT 和动态维护 RT 列表,导致能量 消耗过大.

针对上述问题, Chen 等人^[25]提出 SAoD 查询 算法.通过将查询消息和编码数据副本异步随机均 匀地分发至网络部分节点上.当编码数据与查询消 息在某一节点相遇时,查询请求根据查询列表找到 目标节点,再按照原路径将目标数据返回用户.该算 在一定程度上降低了通信时延,保证了较高的成功 查询概率.但是,SoAD算法不仅需要利用 gossip^[26] 算法预先构建副本分发路径,而且也未考虑节点故 障导致的感知数据失效等问题.

Tsou 等人^[27]采用一种新的基于压缩感知的自适应 top-k 查询算法(CSA top-k). 在数据感知阶段,源节点使用 CS 技术压缩数据并将其传输至本 地存储节点(SNs)上;在数据查询阶段,用户发送 top-k 请求至存储节点,SNs 根据查询请求从压缩 域上导出所需数据传输给用户.该算法在一定程度 解决了数据丢失的问题且避免数据的冗余传输等问题. 但是,它需要构建固定的回传路径,使得网络存 在单点失效的问题.

为了解决这个问题, Chen 等人^[28]设计一种自 适应容错的 QoS 控制查询(AFTQC)算法. 它从数 据的冗余存储和 Cluster 间构建多条通信路径两个 方面,保证了网络链路质量,延长了网络寿命. 但是, 该算法存在数据冗余度高,导致节点存储和通信负 载过重的问题. 因此, Gao 等人^[29]提出一种新的基 于数据属性的空间索引查询算法(SIQ). 它首先使 用局部聚类存储模型避免相同数据发送至同一个 SN,再以所有 SNs 为叶子节点,基于叶子节点中的 数据属性构建空间索引 R-tree. 从而使查询请求能 够快速访问到相关数据,并将其返回给用户. 由于 SIQ 算法存在单点失效的问题,使得数据查询可靠 性差.

3 网络模型及问题描述

本节对将需要用到的一些符号和基本概念进行 了说明,同时对网络模型及本文研究的问题进行了 详细描述.

3.1 符号说明

如表1所示,对后面章节使用到的几种网络节 点、接收和发送数据能耗进行了符号表述及说明.

表 1 符号说明

符号名称	说明
源节点(Source Node,SN)	监测感知事件信息
发送节点(Transmit Node, <i>TN</i>)	发送数据报文
接收节点(Receive Node,RN)	接收数据报文
邻居节点(Neighbor Node,NN)	与当前节点可以直接通信
查询节点(Query Node,QN)	首次注入查询消息的节点
E_t	发送 1b 数据的能耗
E_r	接收 1b 数据的能耗

3.2 相关定义

定义1. 空闲节点(Idle Node, IN). 在 WSNs

中,未对监测事件进行数据感知存储的传感器节点. 其中,在事件发生的整个过程中,始终无法感知、存储事件数据的传感器节点,称为孤立节点(Isolated Node,IsN).

定义 2. 查询时间(Query Time, QT). 在 WSNs 中,从用户在网络中注入一次查询并获得查询结果 经历的整个过程所消耗的时间.

定义 3. 节点使用率 η . 在 WSNs 中,节点使用 率等于总节点个数 n 与非孤立节点总数($n - n_{ISN}$)的 比值,即 $\eta = 1 - n_{ISN} / n$.

推论 1. 如果 WSNs 中的孤立节点数越少,则 节点的使用率越大. 当孤立节点数为零,那么网络中 节点的使用率为 100%.

引理 1^[25]. 在无向图 G(V,E)中,如果数据副本数为 l,查询消息副本数为 g,将其随机均匀地分布在网络中,那么至少存一个节点同时包含数据副本和查询消息副本的概率 $P_{\text{sque}} \ge 1 - e^{-t_{\text{sque}}}$.

3.3 网络模型

无线传感器网络被部署在一个边长为I 上的 矩形区域A 上,并长期执行事件监测任务.其中,随 机均匀地分布着N个监测事件和n 个传感器节点, 每一个传感器节点的通信半径为r.如果任意两个 节点之间的距离 $d \le r$,则这两个节点可以直接通 信.这里,可将 WSNs 视为连通无向图 G(V,E),其 中, $V = \{v_1, v_2, \dots, v_n\}$ 为网络传感器节点的有限集 合, v_i 为节点唯一的编号(id); $E = \{E_{ij} | v_i, v_j \in V\}$ 为 任意两个节点间可以直接通信边的有限集合, E_{ij} 为 节点i和j的直接通信边.同时,网络具备以下特性:

(1)节点故障或发生位移导致网络拓扑结构动态变化和存储数据失效;

(2)网络链路不可靠,通信延时大;

(3)节点能量有限且不可再生;

(4)节点根据定位技术或 GPS、SIN 获取自己的位置信息^[30].

3.4 问题描述

在 WSNs 中,数据查询主要是基于感知数据的 完整性和网络拓扑结构两方面.由于恶劣环境破坏 和软件故障,使得节点存在一定的故障概率.根据文 献[31],假定网络节点的平均故障率是 p_f ,源数据 的备份个数为 K,则它的丢失概率为 $P_{loss} = p_f^{(K+1)}$, 这意味着通过数据备份可以尽可能保证源数据的完 整性.为了解决网络通信不可靠造成的通信时延过 大的问题,可根据引理 1,将 l 个数据副本和 g 个查 询消息副本多路均匀分发至网络部分节点中,不仅 保证了查询消息与数据副本在同一节点相遇的高概 率 $P_{Sque} \ge 1 - e^{-lg/n} (P_{Sque} 为查询成功概率),而且加$ 快了查询速度.因此,这个问题可以归纳为,在动态网络拓扑结构和网络通信不可靠的情况下,实现数据成功查询时,通信时延T最小化和成功查询概率 $<math>P'_{Sque}$ 最大化,即

$$\{\min T, \max P'_{\text{Sque}}\}$$
(1)

其中 $T = t_{que} + t_{ret}$, t_{que} 为查询到目标数据的时间, t_{ret} 为将查询结果返还给用户的时间, 查询成功概率为 $P'_{Sque} = (1 - P_{loss}) \cdot P_{Sque}$.

经上述分析和总结,本文的研究目标是在动态 网络拓扑结构和不可靠通信的 WSNs 中,提出一种 新的低延迟高可靠的数据查询机制.当发生紧急事 件时,源节点能够同步地将感知数据进行备份存储, 并且能够快速地将压缩数据报文均匀分发至其它节 点上.用户可以在任意时刻选取任意节点处注入查 询消息,查询节点将查询消息副本快速地均匀转发 至网络部分节点.当查询成功时,查询消息根据压缩 数据分发的路由列表快速查找到目标数据,并将其 快速可靠地返回给用户.

4 算法设计

4.1 基本思想

LLHADQ 机制主要分为两个阶段:数据监测存储和数据查询.如图 1 所示,前一阶段,源节点首 先采用动态协作感知存储算法,对监测事件进行协作 记录和备份存储;接着源节点使用 Division Bloom Filter(DBF)定期对源数据块进行编码压缩;最后, 源节点根据 MCEC 优化模型,计算出最优的查询压



缩数据副本个数和查询消息副本个数,降低副本分 发的通信开销.后一阶段,查询节点和源节点异步地 将压缩数据和查询消息多路均匀分发至网络中部分 节点上.当查询到目标节点时,目标节点修改自身访 问位,剪切掉之后相同的回传查询结果请求,再将查 询结果快速可靠地返回给用户.

4.2 动态协作感知存储协议

协议主要分为动态协作感知数据和备份存储数 据两部分.前一部分,首先使用 Voronoi 划分将整个 网络分割成以事件为中心的若干个子区域(见图 2), 以保证每个子区域内所有节点到事件的距离最近; 然后,子区域内的节点相互通信,将能够感知事件的 多个节点组成一个集合 S_i. S_i 中所有节点采用时 分协作技术^[32],依次轮流记录事件数据;最后,考虑 到网络拓扑动态变化,每个集合中的节点定期互相 通信,对加入或死亡的节点进行添加或删除,并根据 剩余能量最大的原则推选新的 Leader 节点. Leader 节点重新为成员节点分配监测顺序和时间槽. 从而 增强了网络的可扩展性,节省了节点内存空间、



后一部分,考虑到节点故障导致存储数据失效 的情况,需要对其进行备份存储.当监测事件发生 时,源节点首先将在时间槽 τ上记录的数据 aⁱ_j(aⁱ_j为 节点 j 监测的第 i 事件数据)作为独立的数据块存 储在内存中.接着,根据节点预估故障概率计算出合 理的数据备份个数,再进行备份存储.

根据文献[31],假定网络中所有节点的故障概 率相同且为 *p*_f,每块数据的有效备份数为 *K*,则每 个数据块的平均可用备份数为

$$E_{\text{arg}} = \sum_{i=0}^{k} {\binom{K}{i}} \cdot i \cdot (1 - p_{f})^{i} \cdot p_{f}^{K-i} = K(1 - p_{f}) (2)$$

若每块数据至少拥有一个可用的数据备份,则有 $K(1-p_f) \ge 1$.因此,以数据备份存储的通信能耗最

小化为目标函数,建立优化模型:

min
$$Cost_r = \sum_{k=1}^{K} hop(i,k) \cdot S_{sdata}(i,k) \cdot (E_s + E_r)$$

s. t. $\begin{cases} K(1-p_f) \ge 1 \\ KNQ_{event} < nc \end{cases}$
(3)

其中 hop(i,j)为节点i 到节点j 的通信跳, $S_{sdata}(i,k)$ 为节点i的第k个数据副本, Q_{event} 为每个监测事件 的数据总容量,c为节点存储空间.

假定 hop(i,j)为常数,求解模型(3)可得,当数 据备份数和节点故障概率分别为

$$\binom{K = \lceil 1/(1-p_f) \rceil}{p_f < 1 - NQ_{\text{event}}/n}$$

时,有数据备份时的总能耗 Cost,最小.

此外,这里允许节点在两跳范围内选取空闲邻 居节点进行数据备份存储,以降低数据备份的时延 和能耗.

4.3 数据编码压缩

布鲁姆过滤器(Bloom Filter)是一种空间高效的随机二进制数据结构,利用其对数据进行编码压缩处理,可以极大地降低数据传输的能耗和通信时延^[16,33].然而,Bloom Filter存在假阳性误判,可能会将原本不属于某集合的数据对象误判为该集合成员.

为了降低 Bloom Filter 假阳性误判概率,将其 改进为一种新的分片布鲁姆过滤器(Division Bloom Filter, DBF). 它通过将 Bloom Filter 中的二进制位 向量 V_{BF} 划分为 $x \land k \not g \not b m$ 的二进制位向量集 合,即 $V_{DBF} = \{V_1, V_2, \dots, V_x\},$ 并采用一组 $k+1 \land$ 一致性哈希函数 $\{h_0, h_1, \dots, h_k\}$ 建立数据对象与位 向量之间的映射关系. 其中, h_0 的哈希地址范围为 $\{1, 2, \dots, x\},$ 剩余 $k \land h_i$ 的哈希地址范围由一组 0 或 1 的 m 位二进制向量组成.

Division Bloom Filter 的 x 个位向量 V_i 的初始 化值都为 0 向量,将集合 $A = \{a_1, a_2, \dots, a_a\}$ 中的 α 个数据对象装入 DBF,每个向量 V_i 中仍然存在 0 bit 的概率为

$$\varphi = \left(1 - \frac{k}{m}\right)^{a/x} \tag{4}$$

之后,当检测某一数据对象 b 是否属于集合 A 中的元 素时,先使用 $h_0(b) = i$ 找到对应的位向量 V_i ,再将 数据对象 b 插入其余 k 个哈希函数中.如果 $h_i(b) =$ $1(i=1,2,\dots,k+1),则认为 b$ 是集合 A 中的成员, 否则不属于集合 A.

然而,Division Bloom Filter 存在局部假阳性误 判,即使所有的 *h_i*(*b*)=1,也无法直接判定数据对象 *b* 是集合 *A* 中的成员.所以,结合等式(4),可以得到 DBF 可能发生局部假阳性误判的概率为

$$f = \frac{(1-\varphi)^{k}}{x} = \frac{\left(1-\left(1-\frac{k}{m}\right)^{a/x}\right)^{k}}{x} \approx \frac{(1-e^{-ka/mx})^{k}}{x}$$
(5)

对等式(5)两边取对数,并对k求导. 解得当 $k = mx \ln 2/\alpha$ 时,Division Bloom Filter 有最小局部假阳 性误判概率为 $f_{min} = 0.6185^{mx/\alpha}/x$.

由 4.1 节可知,每个源数据块存在 K 个备份而 被多次编码压缩.因此,在编码压缩之前,源节点(父 节点)与备份节点(孩子节点)通过心跳通知机制^[7], 检测父节点是否存活.如果存活,父节点编码压缩源 数据,否则选取最先查找到父节点所在集合中的 Leader 节点的孩子节点,并让其编码压缩源数据. 若 Leader 节点(监听编号为 1)也失效,则选取监听 编号为 2 的成员作为集合 Leader 节点,依次类推, 以保证孩子节点能够找到 Leader 节点.算法的详细 描述如下:

- 算法1. 源数据编码算法.
- 输入:the subset of S_{i1} ;

 S^i_j is a child-nodes set of parent-node No. $j \in S_{i1}$

- 输出:coded data of No.j node
- mutual communication between the child-nodes and the parent-node;
- 2. IF parent-node is existing
- source data is inserted into Division Bloom Filter by the parent-node;
- 4. ELSE
- 5. IF Leader node is existing
- 6. child-nodes find Leader node of S_{i1} ;
- 7. select the first child-node $u (u \in s_j^i)$ to reach the Leader node;
- source data is inserted into Division Bloom Filter by the child-node No. u;
- 9. ELSE
- select the node with a monitoring number of 2 as the leader node;
- 11. repeat $6 \sim 8$ steps;
- 12. END
- 13. END

如图 3 所示,每个数据文件块 a_i^i 由感知数据和 起始感知时间两部分组成,即 $a_j^i = \langle data, t_{start} \rangle$.源 节点定期将存储的多个数据块 a_j^i 插入到 Division Bloom Filter 中,将其编码压缩为一个长度为 xm bits 的数据包.因此,Division Bloom Filter 编码每个数 据块时的时间复杂度为 O(k+1),压缩率为 $\gamma = xm/aS_{a_i^i}$ (a为数据块个数, $S_{a_i^i}$ 为数据块 a_j^i 容量大小).



4.4 MCEC 优化模型

由于压缩数据副本个数和查询消息个数直接影 响着网络的通信能耗,而源数据的完整性是保证可 靠性查询的前提.因此,根据引理1和4.2小节,可 以得到实际的成功查询概率为

 $P'_{\text{Sque}} = (1 - e^{-lg/n}) \cdot (1 - f_{\min}) \cdot (1 - p_f^{K+1}) \quad (6)$

这里,假定用户期望的查询成功概率为 B,数据 查询时的网络查询通信开销为 Cost_Q,建立 MCEC 优化模型:

$$\min Cost_Q = (lS_{\text{DBF}} + gS_q) \cdot (E_r + E_s)$$
s. t. $P'_{\text{Soure}} \ge B$
(7)

其中, S_{DBF} 是压缩数据容量大小, S_q 是查询消息容量大小.

由 4.2 小节可知,若 p_f 给定,则 $P_{ava} = 1 - p_f^{K+1}$ 为常数.结合等式(6),将优化模型(7)可简化为

$$\begin{cases} \min Cost_Q = lS_{\text{DBF}} + gS_q \\ \text{s. t.} (1 - e^{-lg/n}) \cdot (1 - f_{\min}) \cdot P_{\text{ava}} \ge E \end{cases}$$

这里,令 $a=lS_{DBF}$, $b=gS_q$,则有 $a+b\geq 2\sqrt{ab}$,当且 仅当 $lS_{DBF}=gS_q$,网络查询通信开销最小.从而得到 的不等式方程组为

$$\begin{cases}
lS_{\text{DBF}} = gS_q \\
lg \ge n \ln \left(1 - \frac{B}{(1 - f_{\min}) \cdot P_{\text{ava}}}\right)^{-1} \quad (8)
\end{cases}$$

对不等式组(8)求解得压缩数据副本和查询副本的 取值范围分别为

$$\begin{cases} l \ge \sqrt{\frac{S_q \cdot n \cdot \ln\left(1 - \frac{B}{(1 - f_{\min}) \cdot P_{ava}}\right)^{-1}}{S_{\text{DBF}}}}\\ g \ge \sqrt{\frac{S_{\text{DBF}} \cdot n \cdot \ln\left(1 - \frac{B}{(1 - f_{\min}) \cdot P_{ava}}\right)^{-1}}{S_q}} \end{cases}$$
(9)

当不等式组(9)取等号时,压缩数据副本个数和查询 消息副本个数有最优值,且通信开销 Cost_o最小.

4.5 负载均衡多路均匀分发算法

在 MCEC 优化模型的基础上,根据文献[34-35], 提出一种新的动态均匀分发算法.在数据查询阶段, *TN*_i首先根据定期 Hello 消息机制获取邻居节点剩 余能量 *E*_{Rj}和存储空间 *C*_{Rj},分别计算出剩余能量和 存储空间的平均值为(*n*_i为 *TN*_i的邻居节点个数)

$$egin{cases} \hat{E}_{Ri} = ig(\sum_{j=1}^{n_i} E_{Rj}ig)/n_i \ \hat{C}_{Ri} = ig(\sum_{j=1}^{n_i} E_{Rj}ig)/n_i \end{cases}$$

接着, TN_i 选取 $E_{Rj} \ge \hat{E}_{Ri} \perp C_{Rj} \ge \hat{C}_{Ri}$ 的所有邻居节 点作为 RN, 计算下一跳每个 RN的需要分发的压 缩数据副本个数 $\omega_j = (\omega_i - 1)/SF(i)$.其中, ω_i 为 TN_i 的副本个数, ω_j 为下一跳 RN_j 的分发副本个数, SF(i)为 TN_i 下一跳所有 RN的总个数, 最后, 迭代 执行压缩数据副本的多路分发, 直到 $\omega_i - 1 = 0$ 停 止.为了保证压缩数据副本尽可能地均匀分布在网 络中, 需要满足以下两个条件:

(1) 在副本分发过程中,每个节点只能接收且 存储一次相同的压缩数据副本;

(2) 网络中的节点会定期更新自己的邻居表, 更新局部邻居节点的平均值 \hat{n}_i ,从而可以在一段时间内可以较好地保证每个节点的接收概率 $P_{re} = \min\{n_j, \hat{n}_j\}/\hat{n}_i$ 近似相等.

当 $P_{re} < 1$ 时, RN_j 以概率 $P_{re1} = n_j / \hat{n}_i$ 接收存储 该副本数据; 当 $P_{re} = 1$ 时, RN_j 以相等的概率 $P_{re2} = \hat{n}_i / n_{max}^i$ 来接收该副本数据 (n_{max}^i 为 RN_j 对应子区域 中所有节点的最大邻居节点个数). 如图 3 所示, 压 缩数据副本个数为 l = 20 且随机均匀地分布在网络 中,其中空心圆点为接收存储 DBF 副本数据的节 点. 具体的算法描述如下:

算法 2. 负载均衡多路均匀分发算法.

输入:The number of DBFs' $\omega_i = l$;

RT: Routing Table

- 输出:DBFs are evenly distributed in the network;
- 1. SN acquires neighbor nodes' information;
- TN initializes its the number of DBFs is ω_i = l and RT_i = {location_i};
- 3. TN computes its neighbor nodes' \hat{E}_{Ri} and \hat{C}_{Ri} ;
- 4. TN selects $RN_j = \{NN_j | E_{R_j} \ge \hat{E}_{R_i} \& \& C_{R_j} \ge \hat{C}_{R_i} \}$;
- 5. TN Computes ω_j and transmits DBF, RT_i and ω_j ;

6. WHILE $\omega_j > 0$

- 7. RN_j computes itself the probability of receiving P_{re} ;
- 8. IF $P_{\rm re} < 1$
- *RN_j* stores the DBF and *RT_j* = *RT_i* ∪ {*location_j*}, then repeat 3~8 steps;
- 10. ELSE
- 11. IF RN_j receives the DBFs with a probability of P_{rel} ;
- 12. Implement 9 steps;
- 13. ELSE
- 14. RN_j implements $3 \sim 4$ steps and computes $\omega_{j+1} = \omega_j / SF(j)$,
- 15. And directly forwarding the DBF and $RT_j = RT_i \bigcup \{location_j\},\$
- 16. END
- 17. END
- 18. END
- Regular Hello message mechanism is used to update NNs' information of SNs.

当用户进行数据查询时,可以在网络中选择任一 节点注入查询消息,QN使用算法2将查询消息均匀 地转发到网络部分节点上进行查询匹配.如图4所 示,空心圆点表示存储压缩数据副本的节点,星号表 示转发过查询消息的节点.当星号与空心圆点在同 一位置时,表示查询消息与压缩数据匹配成功.显 然,在保证较高成功查询率的条件下,使用该算法进 行压缩数据副本分发的时间复杂度为 $O(\log_{SF} l)$,其中SF为每一轮平均接收副本的节点个数.



4.6 剪枝容灾路由回传协议

在数据查询过程中,网络可能存在多个相同 的查询成功节点(见图 5),导致冗余的查询结果返 回给用户.因此,根据文献[36-37],设计了一种延 迟小、可靠性高的剪枝容灾路由回传协议,主要分 为剪枝路由回传和容灾路由回传两部分.

2020年



前一部分,首先,匹配成功的节点根据压缩数据 分发路由列表快速寻找到目标节点,将预先设置的访 问位(*mark*access)从0修改为1.如果访问位*mark*access = 1,则目标节点剪切掉之后到达的所有同类查询请 求,否则接收该查询消息并修改访问位,接着,目标 节点按照监测顺序收集最近一段时间其它节点中的 数据块,并将完整的事件数据返回给用户;最后,在 每次查询完成后,目标节点将访问位重新置为零,以 保证用户能够继续查询到之后感知最新的数据.

后一部分,考虑到网络拓扑结构的动态变化, 导致目标数据回传时延大.因此,在回传过程中, TN_i 首先以自己位置为坐标原点判断 QN 位置 (x_{NN}, y_{NN}) 的方向,并发送 (x_{NN}, y_{NN}) 给对应方向 上的所有邻居节点 $i \in S_{NN}^{j}(S_{NN}^{j})$ 为 TN_{j} 在对应方向 上的邻居节点集合);接着,每个邻居节点i以自己 的位置为坐标原点,确定新的回传方向及在此方向 上的邻居节点集合 S_{NN}^{i} ;最后, TN_{j} 选取 $|S_{NN}^{i}| > 0$ 的邻居节点,计算节点i与QN 之间的距离 $d_{QN}^{i} = \sqrt{(x_{i} - x_{NN})^{2} + (y_{i} - y_{NN})^{2}}((x_{i}, y_{i}))$ 为节点i的位 置坐标),并选取 $\min\{d_{QN}^{i} | i \in S_{NN}^{j}\}$ 的邻居节点作 为新的回传路由节点.迭代上述过程,直到目标数据



点所在方向区域,选取距离查询节点最近的邻居节 点作为新的回传路由,减少数据回传的路由个数.算 法的具体描述如下:

- 算法 3. 剪枝容灾路由回传协议.
- 输入:*TN*;*RN*;*QN*; *GN*:Goal node;

QM: Query message

- 输出:QR:Query Result
- 1. GN receives the earliest arrival of QM and modify $mark_{access} = 1;$
- 2. $TN_j \leftarrow GN;$
- 3. TN_j computes the location and orientation of QN;
- 4. IF $(x_{QN} x_j) = 0$ & & $(y_{QN} y_j) = 0$
- 5. TN transmits a dictation packet to the QN;
- 6. ELSE
- 7. IF $(x_{QN} x_j) > 0$ & & $(y_{QN} y_j) > 0$
- 8. TN_j selects neighbor nodes of $(x_{QN} x_i) > 0$ & & $(y_{QN} y_i) > 0$ ($i \in S_{NN}^j$) to form a set S_{NN}^j ;
- 9. TN_j sends $location_{QN}$ to all nodes in the S_{NN}^j ;
- 10. all nodes $i \in S_{NN}^{i}$ compute the direction of QN, and determine the neighbor nodes set S_{NN}^{i} in this direction;
- 11. TN_j selects neighbor nodes of $\{i \mid i \in S_{_{NN}}^j \&$ $\mid S_{_{NN}}^i \mid > 0\};$
- 12. TN_j computes d_{NN}^i and selects neighbor node of

 $\min\{d_{QN}^{i} \mid i \in S_{NN}^{j} \& \mid S_{NN}^{i} \mid > 0\} \text{ as } RN;$

3. ELSE IF $(x_{QN} - x_j) \le 0$ & $(y_{QN} - y_j) > 0$

- 14. TN_j selects neighbor nodes of $(x_{QN} x_i) \le 0$ & &. $(y_{QN} - y_i) > 0 (i \in S_{NN}^j)$ to form a set S_{NN}^j ;
- 15. repeat steps $9 \sim 12$;
- 16. ELSE IF $(x_{QN} x_j) > 0$ & & $(y_{QN} y_j) \le 0$
- 17. TN_j selects neighbor nodes of $(x_{QN} x_i) > 0$ & & $(y_{QN} y_i) \le 0$ ($i \in S_{NN}^j$) to form a set S_{NN}^j ;
- 18. repeat steps $9 \sim 12$;
- 19. ELSE IF $(x_{QN} x_j) < 0$ & & $(y_{QN} y_j) < 0$
- 20. TN_j selects neighbor nodes of $(x_{QN} x_i) < 0$ & & $(y_{QN} y_i) < 0(i \in S_{NN}^j)$ to form a set S_{NN}^j ;
- 21. repeat steps $9 \sim 12$;
- 22. IF RN is not equal to QN
- 23. $TN_i \leftarrow RN;$
- 24. repeat steps $3 \sim 23$;
- 25. ELSE
- 26. TN_j transmits a dictation packet to the RN and ends the return;
- 27. END
- 28. END
- 29. END
- 30. END
- 31. END
- 32. END

在查询回传过程中,如果发送节点发现对应方向上没有符合要求的邻居节点,则它以逆时针为方向选新方位上的邻居节点,按照算法3继续寻找距离查询节点最近的节点作为新的回传路由. 假定在每一跳回传过程中,都存在符合要求的邻居节点,则目标数据回传的时间复杂度为 $O(d_{GN_QN}/\hat{r})$,其中, d_{GN_QN} 为目标节点与查询节点之间的距离, \hat{r} 为节点的平均传输距离.

5 理论分析

在这部分,首先理论分析了副本均匀分发时局部阈值(平均邻居数)的选取;接着,定量分析了 LLHADQ机制在通信开销、通信延迟以及查询成 功率三个方面的性能.

5.1 局部阈值的选取

根据引理1可知,让压缩数据副本和查询消息 副本随机均匀地分布在网络中是保证高成功查询概 率的前提.这意味着每个节点需要获取全网的拓扑 结构和位置信息,才能保证副本的随机均匀分发.但 是,在实际情况中,网络拓扑结构动态变化、节点分 布不均匀及节点能量有限,使得节点获取全网节点 信息是不可能的.

定义 4. 近似均匀分发.如果网络中所有节点 均能够以近似的概率接收副本数据,则每个节点的 接收概率需要满足以下两个条件:

(1)存在一个任意小的正数 ε_1 ,使得网络中任 意子区域 $v(v \in N)$ 都有 max $P_{re}(v)$ 一min $P_{re}(v) \le \varepsilon_1(\max P_{re}(v)$ 和 min $P_{re}(v)$ 分别为子区域 v 中最大 接收概率和最小接收概率);

(2)存在一个任意小的正数 ε_2 ,使得网络中任意两个子区域 $v, u(v, u \in N)$ 都有 $|\hat{P}_{re}(v) - \hat{P}_{re}(u)| \le \varepsilon_2(\hat{P}_{re}(v))$ 为子区域v中的平均接收概率).

因此,在数据查询过程中,采用一种完全分布式 的动态多路均匀分发算法,通过定期更新每个子区 域的平均邻居数 *ĥ*;来保证网络中所有节点都能够以 近似的概率接收副本数据.

假定子区域v中所有节点的邻居节点数为 n_j ($j \in S_v, S_v$ 为子区域节点的集合),局部平均阈值为 \hat{n}_v ,则每个节点接收副本数据的概率为

$$P_{\rm re}(v) = \begin{cases} \frac{n_j}{\hat{n}_v}, & n_j < \hat{n}_v \\ \\ \frac{\hat{n}_v}{n_{v_{\rm max}}}, & n_j \ge \hat{n}_v \end{cases}$$
(10)

根据定义 4 中的条件(1)可知,在任意子区域 *i* 中, 存在一个任意小的正数 ε,使得

$$\max P_{\rm re}(v) - \min P_{\rm re}(v) \le \varepsilon \tag{11}$$

由不等式组(10)和不等式(11),可得到以下三 种情况:

(1) 当子区域中所有节点的邻居数 $n_j > \hat{n}_v$ 时, 则有 $\hat{n}_v \leq n_{v_{\min}} - 1$,且使得所有节点具有相等的接收 概率 $P_{re}(i) = \hat{n}_v / n_{v_{\max}}$;

(2) 当子区域中所有节点的邻居数 $n_j < \hat{n}_v$ 时, 则有 $\frac{\hat{n}_v - 1}{\hat{n}_v} - \frac{n_{v_{\min}}}{\hat{n}_v} \le \varepsilon_1^v$;

(3) 当子区域中所有节点的邻居数存在 $n_j \leq \hat{n}_v$ 和 $n_i \geq \hat{n}_v$ 时,则有

$$\begin{cases}
\frac{\hat{n}_{v}-1}{\hat{n}_{v}} - \frac{\hat{n}_{v}}{n_{v_{\max}}} \leq \varepsilon_{1}^{v} \\
\frac{\hat{n}_{v}}{n_{v_{\max}}} - \frac{n_{v_{\min}}}{\hat{n}_{v}} \leq \varepsilon_{2}^{i}
\end{cases}$$
(12)

在根据定义4中的条件(2)可得,针对情况(1), 网络中所有子区域的之间接收概率的总误差为

$$\Delta_{1} = \sum_{v=1}^{N} \sum_{u=v+1}^{N} \left| \frac{\hat{n}_{v}}{n_{v_{\text{max}}}} - \frac{\hat{n}_{u}}{n_{u_{\text{max}}}} \right|$$
(13)

针对情况(2),网络中所有子区域的之间接收概率的 总误差为

$$\Delta_{2} = \sum_{i=v+1}^{N} \sum_{j=1}^{N} \left| \frac{\sum_{j=1}^{|S_{v}|} n_{j} / \hat{n}_{v}}{|S_{v}|} - \frac{\sum_{j=1}^{|S_{u}|} n_{j} / \hat{n}_{u}}{|S_{u}|} \right| \quad (14)$$

针对情况(3),网络中所有子区域的之间接收概率的 总误差为

$$\Delta_{3} = \sum_{v=1}^{N} \sum_{u=v+1}^{N} \left| \frac{\sum_{j=1}^{|S_{v}|} P_{re}(j)}{|S_{v}|} - \frac{\sum_{j=1}^{|S_{u}|} P_{re}(j)}{|S_{u}|} \right| \quad (15)$$

假定 $1 \le n_{v_{min}} < n_{v_{max}}$ 且 $n_{v_{max}} \ge 3$ 时,当 $|S_v| = |S_u|$ 时,可由等式(13)~(15)分别获得三种情况下的所有节点之间接收概率的总误差的最大值为

$$\max \Delta_1 = \sum_{v=1}^{N} \sum_{u=v+1}^{N} \left| 1 - \frac{1}{n_{v_{\max}}} - \frac{1}{n_{u_{\max}}} \right|$$
(16)

$$\max \Delta_2 = \sum_{v=1}^{N} \sum_{u=v+1}^{N} \left| 1 - \frac{1}{n_{v_{\max}} + 1} - \frac{1}{n_{u_{\max}} + 1} - \frac{1}{n_{u_{\max}} + 1} - \frac{1}{|S_v|} \cdot \left(2 - \frac{2}{n_{v_{\max}} + 1} - \frac{2}{n_{u_{\max}} + 1} \right) \right| \quad (17)$$

$$\max \Delta_{3} = \sum_{v=1}^{N} \sum_{u=v+1}^{N} \left| 1 - \frac{1}{\hat{n}_{v}+1} - \frac{1}{\hat{n}_{u}} - \frac{1}{|S_{v}|} \cdot \left(2 + \frac{1}{\hat{n}_{v} \cdot (\hat{n}_{v}+1)} + \frac{1}{\hat{n}_{u} \cdot (\hat{n}_{u}+1)} \right) \right|$$
(18)

由于 $\hat{n}_v < \hat{n}_v + 1 \le n_{v_{max}} < n_{v_{max}} + 1$,则由等式(16)~(18)可知,max $\Delta_3 < \max\Delta_1$ 且 max $\Delta_3 < \max\Delta_2$.因此,只需要对情况(3)进行分析来确定阈值.

这里,令 $\varepsilon_1^i \approx 0$ 和 $\varepsilon_2^i \approx 0$ 且 $n_{v_{\min}} < \hat{n}_v < n_{v_{\max}}$,由 不等式组(12)解得该区域的平均邻居节点数为

$$\hat{n}_{v} = \lfloor \sqrt{n_{v_{\min}} \cdot n_{v_{\max}}} \rfloor \tag{19}$$

特别,当子区域 v 中所有节点的邻居节点数 n_i 都相 等时,取 $n_{v_{min}} = n_j/2$ 及 $n_{v_{max}} = n_j$,则由式(19)可获得 该区域的平均邻居节点数为 $\hat{n}_v = |\sqrt{2}n_i/2|$.

这样,每个 Leader 节点只需要从定期更新的对 应子区域所有节点的邻居数列表中,选取邻居个数 最小和最大的两个数,根据式(19)快速计算出子区 域的局部阈值,并发送给所有成员节点,以保证副本 分发的均匀性.

5.2 性能分析

5.2.1 通信开销

引理 2^[38]. 网络连通性. 假定网络节点总数 为 n,将其部署在一个 $L \times L$ 的方形区域 A 中. 根据 节点个数将网络划分成一个 $\sqrt{n} \times \sqrt{n}$ 的二维网络,其 中每个方格的边长为 $\beta = L/\sqrt{n}$. 如果该网络满足以 下两个条件:

(1)每个方格至少随机分布一个节点;

(2)每个节点的通信半径 r≥√5L/√n,那么,该 网络中的所有节点之间都是连通的.

考虑到网络节点之间的传输距离不相等,为了 简化模型,假定网络具有良好的连通性且每个节点 总能以固定的传输距离发送数据,那么数据查询的 总传输能耗为数据传输的总次数与每一跳的平均通 信开销的乘积.

在本文,LLHADQ 机制的查询通信总开销查 询搜索前的节点编码源数据能耗(Cost_{code})、查询搜 索时的数据分发能耗(Cost_{dis})、目标数据回传能耗 (Cost_{re})和节点间信息交互能耗(Cost_{exinf})四部分, 则该机制的通信总能耗为

 $Cost_{total} = Cost_{code} + Cost_{dis} + Cost_{re} + Cost_{ex-inf} \quad (20)$

由 4.5 小节可知,每个接收查询请求的节点都 会使用 DBF 对其进行编码比对. 假定每个数据块进 行一次哈希映射(即 $h_i(b) = 1$)的能耗为 E_1 . 又由 4.5 小节可知,每个节点接收存储数据的平均概率 为 \hat{P}_{re} ,则每个副本实际的转发次数为 $1/\hat{P}_{re}^{[36]}$. 因 此,在查询搜索过程中,查询请求数据包编码所需的 能量为 $Cost_1 = \gamma S_{query} E_1 g/\hat{P}_{re}$,源数据包编码所需的 能量为 $Cost_2 = \gamma S_{data} E_1$ (即每次查询时源数据只编 码一次),则一次查询过程中的数据编码的总能耗为

$$Cost_{code} = \frac{\gamma S_{query} E_1 g}{\hat{P}_{re}} + \gamma S_{data} E_1 \qquad (21)$$

从式(21)可以看出,当 \hat{P}_{re} , E_1 和γ为固定常数时, 数据编码的能耗完全取决于源数据和查询请求的数 据量大小.

根据 MCEC 优化模型可知,当数据搜索成功时,压缩数据的副本个数和查询消息分发个数分别为(*l*,*g*),则在查询搜索过程中,数据分发的总通信 开销为

 $Cost_{se} = E_2 (l\gamma S_{data} + gS_{query})/\hat{P}_{re}$ (22) 其中, γ 为 DBF 的压缩概率, S_{data} 为源数据容量, S_{query} 为查询消息数据容量, E_2 为节点发送和接收 1 bit 的总能耗.

根据 4.6 小节可知,目标数据回传的平均传输 轮数为 $\sqrt{2} d_{GN-QN}/\hat{r}$,则查询回传的平均通信开销为

$$Cost_{\rm re} = \sqrt{2} S_{\rm data} d_{GN_QN} E_2 / \hat{r}$$
 (23)

由 4.6 小节可知,在结果回传过程中,发送节点 与部分邻居节点先进行信息交互确定下一跳接收节 点,再回传目标数据,则整个数据回传过程中节点信 息交互最大能耗为

 $Cost_{exinf} = E_2 (\pi r^2 S_1 / 2\beta^2) (\sqrt{2} d_{GN_QN} / \hat{r})$ (24) 其中, $S_1 (S_1 \ll S_{data})$ 为结果回传时节点间互相通信 一次的数据包大小, $\pi r^2 / 2\beta^2$ 为发送节点通信范围内 对应查询节点方向上某一侧的平均邻居节点个数.

综上分析,结合式(20)~(24)可得,一次数据查 询查询成功时的总通信能耗为

$$Cost_{total} = E_2 \left[\left(\frac{l\gamma}{\hat{P}_{re}} + \frac{\sqrt{2}d_{GN_QN}}{\hat{r}} \right) S_{data} + \frac{gS_{query}}{\hat{P}_{re}} + \frac{\pi r_2 S_1 \sqrt{2} d_{GN_QN}}{2\beta^2 \hat{r}} \right] + E_1 \left[\frac{\gamma S_{query}g}{\hat{P}_{re}} + \gamma S_{data} \right].$$

5.2.2 通信延迟

在 LLHADQ 机制中,数据查询的总通信时延 分别由副本分发时延 T_1 、数据回传时延 T_2 、源数据 编码计算时延 T_3 和信息交互时延 T_4 四部分组成, 即 $T_{total} = T_1 + T_2 + T_3 + T_4$.在数据回传过程中,由 于节点间信息交互时延很小(即 $S/B \ll S_{data}/B$),且 计算复杂度低,因此可以忽略这一阶段的通信延迟, 则总通信时延为

$$T_{\rm total} = T_1 + T_2 + T_3 \tag{25}$$

假定网络具有良好的连通性且每个节点的传输 速率均为*B*,每块源数据编码压缩一次的计算时间 为 t_3 ,与 5.2.1小节的数据编码能耗同理,当 \hat{P}_{re} , t_3 和 γ 为固定常数时,数据编码的时延完全取决于源 数据和查询请求的数据量大小,即

$$T_{3} = \frac{\gamma S_{\text{query}} t_{3} g}{\hat{P}_{\text{re}}} + \gamma S_{\text{data}} t_{3}$$
(26)

在查询搜索过程中,压缩数据报文分发一轮的 时延为 $t_1 = \gamma S_{data} / \mathcal{B}$,查询请求数据报文分发一轮的 时延为 $t_2 = S_{query} / \mathcal{B}$.根据4.5小节可知,压缩数据 副本分发的时延为 $\log_{\hat{sF}}(l) \cdot t_1$,查询消息副本转发 的时延为 $\log_{\hat{sF}}(g) \cdot t_2$,在一次查询搜索中成功寻找 到目标数据的最坏时延为

$$T_1 = \max\{\log_{\widehat{SF}}(l) \cdot t_1, \log_{\widehat{SF}}(g) \cdot t_2\} + \log_{\widehat{SF}}(l) \cdot t_2$$
(27)

在结果回传过程中,目标数据传输一轮的时延为 $t_3 = S_{data}/B$,根据4.6小节可得,结果回传的最坏回传时延为

 $T_2 = \sqrt{2} t_3 d_{GN_QN} / \hat{r}$ (28) 综上分析,结合等式(25)~(28)可得,完成一次 数据查询的总通信时延为

$$T_{\text{total}} = \max\{\log_{\widehat{SF}}(l) \cdot t_1, \log_{\widehat{SF}}(g) \cdot t_2\} + \log_{\widehat{SF}}(l) \cdot t_2\}$$

$$t_2 + rac{\sqrt{2t_3} d_{GN_QN}}{\hat{r}} + rac{\gamma S_{ ext{query}} t_1 g}{\hat{P}_{ ext{re}}} + \gamma S_{ ext{data}} t_1.$$

5.2.3 查询成功率

在这部分,主要从数据效率、数据查询成功率和 数据回传三个方面对 LLHADQ 机制的高可靠性进 行分析.

(1)提高数据的可用率.为了避免因节点故障 导致部分源数据失效,在数据感知存储阶段,根据预 估网络节点故障率计算出合理的备份个数 K,并按 照这个备份数选取空闲邻居节点进行备份存储,使 得源数据的可用率提高了μ倍,即

$$\mu = \frac{1 - p_f^{K+1}}{1 - p_f} = 1 + p_f + \dots + p_f^{K-1} + p_f^K;$$

(2) 高成功查询率. 该机制首先采用了 DBF 对 源数据进行编码,使得局部假阳性误判概率最小;接 着,引入引理 1,让随机均匀分发的查询消息和压缩 数据副本能够以高概率相遇. 从而保证了该机制能 够以高概率 $P_{Se} = (1 - e^{-lg/n}) \cdot (1 - f_{min})$ 查询到目标 数据;

(3)可靠的数据回传协议.4.4小节采用一种动态路由回传方式,在每一次数据回传过程中,回传节 点先与邻居节点互相通信,确定下一跳最佳的传输 路由节点再进行数据回传,以保证传输的可靠性.

6 实验模拟

在这部分,首先详细介绍了实验环境的设置;接着,对 Division Bloom Filter 的实验参数进行了设置;最后,与已有的方案的实验结果进行了比较,同时将理论分析结果和实验模拟结果进行了对比验证.

6.1 实验环境

本文在 NS2 搭建了一个动态 WSNs 网络环境, 接着,在该环境下分别实现了 LDHADQ 机制与 Flooding、SoAD、CSA top-*k* 和 SIQ 等几个对比算 法仿真.在仿真中,假定在一个 1000 m×1000 m 的 矩形区域中随机均匀地分布着 *n*=10000 个传感节 点和 10 \sqrt{n} 个高频监测事件.根据 4.2 小节的动态协 作感知存储协议,可将网络划分为 1000 个子区域,每 个子区域节点集合为 $S_i = S_i^1 \cup S_i^2 \amalg S_i^1 \cap S_i^2 = \emptyset$.其 中 S_i^1 为子区域源节点有限集合, S_i^2 为子区域完全空 闲节点有限集合,且 $\sum |S_i^1| \gg \sum |S_i^2|$.由引理2可 知,当每个节点的通信半径 $r \ge 20$ m 时,网络具有连 通性.

如图 7 所示,随着传感器节点通信半径的增大, 网络节点的使用率越高. 从图中可以看出,当 r =30m时,网络节点的使用率 η 为 98. 69%,即网络中 的完全空闲节点数为 $\sum |S_i^2| = 131$,远远小于网络 的节点总数. 所以,这里取 r = 30 m 为传感器节点的 通信半径.



文献[39]提出了新的 MOTE 原型可使得传感 器内存达到 8GB. 这里设置 LLHADQ 算法中数据 回传时每一次信息交互的数据容量大小为 40 bits, 每个节点的监听时间槽 $\tau = 7 \text{ ms}$,传感器节点接收和 发送数据的平均速率均为 256 kbit/s,且每隔 700 ms 对源数据块进行一次编码分发,用户每隔 3.5 s 对网 络进行一次数据查询,则该段时间内的总源数据容 量大小为 $S_{data} = 819.2$ kbits,每个节点的存储的压 缩数据大小为 $S_{DBF} = \gamma S_{data}/10, 查询消息的数据容$ $量大小为 <math>S_{query} = S_{data}^{r} = 1835$ bits(S_{data}^{r} 为一个时间片 内的感知数据).根据路径损耗模型^[37],在距离 d上 计算发送 1 bit 数据的能量损耗为 $E_t = \epsilon d^{\lambda} + E_{ele}$,接 收 1 bit 数据的能量损耗为 $E_t = E_{ele}$.其中, $\epsilon = 10^{-11}$, $\lambda = 2 \ Z E_{ele} = 50 \times 10^{-9}$ J.

如图 8 所示,当网络节点的平均故障概率 0 $\leq p_f < 1/2$ 时,感知数据的备份个数 $K \ge 2$ ($K = \lceil 1/(1-p_f) \rceil$).



如图 9 所示,当每个感知数据块的备份数 K= 0,1,2 时,测定了感知数据的丢失概率.可以看出, 随着节点故障概率的增大,不同备份数下的感知数 据丢失率均有所增加.其中,当备份数为 K=2 时, 感知数据的丢失率增长最缓慢,即可取感知数据的 备份个数为 2.



6.2 Division Bloom Filter 参数设置

由 4.3 小节可知,当 $k = mx \ln 2/\alpha$ 时,Division Bloom Filter 的局部假阳性误判概率最小,且为 $f = (1-e^{-\ln 2})^k$.

如图 10 所示,在取 k=1,...,10 的范围内,测定 了不同的分片数下的 Division Bloom Filter 检测成 功概率.可以看出,随着 Division Bloom Filter 中哈 希函数个数的不断增大,不同分片数下的检测正确 率均有所增加.其中,当 k=9, x=10 时,Division Bloom Filter 检测数据对象的正确率为 99.98%.



如图 11 所示,测定了相同检测正确率下的 Division Bloom Filter 与 Bloom Filter 的数据量的 大小. 当插入两个过滤器中的感知数据块均为 500 (α =700 ms/(140 μ s×10))时,有 S_{DBF} Subjection Bloom Filter 的平均数据量为 S_{DBF} = 813. 17 bits,平均压缩率为 γ =4.43% (γ = S_{DBF} / S_{data} , S_{data} =140 μ s×256 kbit/s×500); Bloom Filter 的平均数据量为 S_{BF} = 3209.4 bits.平均压缩率为 17.49%. 这表明,在相同的检测正确率下, Division



图 11 DBF 与 BF 数据容量

Bloom Filter 的数据量比 Bloom Filter 的数据量小 得多.

因此, Division Bloom Filter 中各项参数的设置 如表 2 所示.

表 2 Division Bloom Filter 参数值

参数	设定值	
α	500	
k	9	
x	10	
	**	

6.3 性能表现

本小节主要从数据查询的通信开销、通信延迟、 查询成功率等几个方面的性能进行了实验测试,并 与其它几种算法进行了对比分析.

6.3.1 通信开销

在这部分,首先设定每块源数据进行一次哈希 映射的能耗等于节点接收1bit数据的能耗,即 $E_1 = E_{ele}$.然后,根据 6.2 小节,取 DBF 和 BF 的检测精确 率为99%,它们的数据压缩率分别为 $\gamma_{DBF} = 1.31\%$, $\gamma_{BF} = 2.72\%$,则查询请求被压缩后的数据量分别为 24 bits 和 50 bits.最后,通过改变节点的传输半径和 网络规模来测定五种算法的总通信能耗.

如图 12 所示,在网络节点数不变的情况下,随着节点的通信范围越大,五种算法的通信开销呈逐渐下降趋势.其中,Flooding 算法的通信开销始终高于其它算法的通信开销,CSA top-k 算法的查询通信开销最小.LLHADQ 机制与 CSA top-k 算法之间的平均通信开销的绝对误差为 0.063309 J,相比较 CSA top-k 与其它算法之间的平均通信开销的绝对误差值,它的通信能耗是最接近 CSA top-k 算法的通信能耗.



如图 13 所示,在不同网络规模下,测定的几种算 法查询通信开销.可以看出,随着网络节点数 sqrt(n) 的不断增加,所有算法的通信能耗都有所增大.当 *sqrt(n)*>250时,LLHADQ机制的通信能耗将低于 CSA top-*k*算法的通信能耗.这表明,该机制比其它 几种算法更适合大规模 WSNs 中的数据查询.



实验结果表明,与 Flooding 算法相比,该机制 通过 MCEC 优化模型能够有效地避免查询消息全 网广播;与 SIQ、CSA top-k 和 SoAD 算法相比,该 机制使用 Division Bloom Filter 编码源数据块,减 少了数据传输能量.同时,LLHADQ 机制采用剪枝 容灾路由回传协议,通过优化传输路径来减少回传 跳数.

6.3.2 通信延迟

在这部分,首先设定每块源数据进行一次哈希 映射的时间等于节点接收 1 bit 数据所需的时间(即 $t_3 = (1/256 \text{ kbit})s)$,再对几种查询算法的通信时延 进行了测定.

如图 14 所示,五种查询算法的通信时延随着节点 通信半径的增大而减小.从图中可以发现,Flooding 算法的平均通信时延为 $T_f=22.68$ s, LLHADQ 机



制的平均通信时延为 $T_i = 21.65685$ s, SIQ 算法的 平均通信时延为 $T_s = 27.99182$ s, CSA top-k 算法 的平均通信时延为 $T_c = 22.38745$ s, SoAD 算法的 平均通信时延为 $T_c = 24.10527$ s,则有 $T_l < T_c < T_f < T_c < T_s$.显然,LLHADQ 机制的平均通信时 延是五种查询算法中最小的.

如图 15 所示,随着网络规模的增大,五种查询 算法的通信延迟呈近似线性增长.从图中可以看出, Flooding、LLHADQ、CSA top-*k* 及 SoAD 算法的通 信延迟增长缓慢,SIQ 算法的通信时延增长速度快. 其中 LLHADQ 机制的通信时延始终低于其它四种 算法的通信时延.这表明,在通信时延性能方面, LLHADQ 机制优于另外四种查询算法.



图 15 不同网络规模下查询时延

6.3.3 成功查询率

在这部分,首先对节点故障率和成功查询率的 实验结果与理论分析结果进行了对比;接着,分别测 定了在无节点故障和存在节点故障两种情况下的五 种查询算法的查询成功概率.



如图 16 所示,在感知数据备份数为 K=2,节点

故障率为 $p_f \in [0, 0.25]$ 时,测定了感知数据丢失率 模拟值和理论值.从中可以清楚地看出,感知数据的 丢失率的模拟值曲线与理论值曲线近似相等.

如图 17 所示, LLHADQ 机制的查询成功率进行了 1000 轮模拟查询.实验发现,随着查询次数地不断增大,模拟的数据查询成功率将会在理论查询成功概率的上下波动,且幅度很小.这表明 LLHADQ 机制的理论分析与实验模拟相符合,具有良好的可行性和真实性.



图 17 查询成功概率的理论值与实验值的比较

由于在无节点故障的情况下,SIQ和CSA top-k 算法一定能够查询到目标数据,即它们的成功查询 概率为100%.因此,在这部分,分别对LLHADQ、 SoAD及Flooding三种查询算法进行了500轮的数据 查询成功率的测定.由图18(a)可以发现,LLHADQ 机制的数据查询成功概率始终高于Flooding算法 的成功查询率,而且Flooding算法的平均成功查询 概率为76.65%,LLHADQ机制的平均成功查询率 为99.12%.由图18(b)可以看出,LLHADQ机制



图 18 无节点故障的成功查询率

和 SoAD 算法的成功查询概率近似相等,且 SoAD 算法的平均成功查询概率为 98.97%.上述对比结果表明,在无节点故障的情况下,LLHADQ 机制和 SoAD 算法在成功查询概率方面均优于 Flooding 算法.

如图 19 所示,当节点的故障率取值范围为 p_f ∈



[0,0.25]时,五种查询算法的成功查询概率均有所 下降.其中,LLHADQ 机制的成功查询概率下降不 明显,剩下四种算法的成功查询概率下滑明显.特别 是 Flooding、SIQ 及 CSA top-k 算法的成功查询概 率下滑快速.这表明,LLHADQ 在网络存在一定节 点故障率的情况下,相比其它几种查询算法,它依然 能够保证较高的成功查询概率.

6.4 实验总结

本文通过仿真实验,分别从通信开销、通信时延和成功查询率三个方面,将 LLHADQ 机制与 SIQ、 SoAD、CSA top-*k* 及 Flooding 等四种查询算法进行 对比分析(分析结果见表 3).从表 3 中可以发现,大 规模动态 WSNs 中,LLHADQ 机制在数据查询的 通信延迟和成功查询概率方面都优于其它算法.同 时,该算法在查询通信开销性能方面也低于 SIQ、 SoAD、Flooding 三种查询算法.此外,6.3.1 小节的 实验结果表明,CSA top-*k* 算法与该机制的通信开 销近似相同.

		400	并没任能的论	
算法	平均通信延迟	平均通信能耗	▼均成功查询概率(p _f =0)	成功查询概率(0 <p<sub>f≤0.25)</p<sub>
LLHADQ	最小	偏小	较高	较高
Flooding	偏小	最大	偏低	最低
SoAD	中	中	父	偏高
CSA top- k	偏大	最小	100%	偏低
SIQ	最大	偏大	100 %	偏低

7 总 结

在大规模动态无线传感器网络中进行数据查 询,不仅需要考虑节省节点的能量消耗和内存占用, 而且需要考虑降低数据查询的延迟及提高数据查询 的可靠性.如何在保证数据查询的低延迟和高成功 查询概率的前提下,使得节点的能量消耗尽可能均 衡是一个难点问题.通过对网络拓扑动态性和数据 查询算法的分析,本文提出一种新的低延迟高可靠 数据查询机制来解决这个问题.仿真实验结果表明, 与Flooding、SoAD、SIQ和CSA top-*k*四种查询算 法相比,无论是在节点网络规模、节点的通信半径发 生变化,还是在节点网络规模、节点的通信半径发 机制在查询通信延迟和成功查询概率方面均有较大 提升.另外,在查询通信开销方面,该机制相比 Flooding、SoAD、SIQ 三种算法也有明显优势.

LLHADQ 机制是一种动态非聚合查询方式. 在数据查询过程中,节点需要预先获取邻居节点信 息,判断本地网络的连通性是否良好.然而,由于节 点意外损毁或死亡导致网络部分区域连通性差而影 响数据查询的通信时延和可靠性.因此,在下一步工 作中,我们将对移动传感器网络(如 VANET)场景 下的低延迟高可靠数据查询方式进行研究.

致 谢 衷心感谢各位评审专家的审稿意见!

参考文献

- [1] Choudharyl R, Kumar S, Deepak A, et al. Data aggregation in wireless sensor network: An integer linear programming formulation for energy optimization//Proceedings of the 2016 13th International Conference on Wireless and Optical Communications Networks. Hyderabad, India, 2016: 1-6
- [2] Zhang H, Shen H. Balancing energy consumption to maximize network lifetime in data-gathering sensor networks. IEEE Transactions on Parallel and Distributed Systems, 2009, 20(10): 1526-1539
- [3] Gil D, Ferrández A, Moramora H, et al. Internet of Things: A review of surveys based on context aware intelligent services.

Sensors, 2016, 16(7): 1069

- [4] Divya M, Shunmuganathan K L. A multi-agent based intelligent query processing system for Hadoop with FIPA-OS using cooperating agent in cloud environment//Proceedings of the 2016 10th International Conference on Intelligent Systems and Control. Coimbatore, India, 2016: 1-6
- [5] Sarode P, Nandhini R. APDA: Adaptive pruning & data aggregation algorithms for query based wireless sensor networks //Proceedings of the 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication. Jalgaon, India, 2016: 219-224
- [6] Geetha R, Shunmuganathan K L. Intelligent query processing from biotechnological database using co-operating agents based on FIPA standards and Hadoop, in a secure cloud environment//Proceedings of the 2017 4th International Conference on Advanced Computing and Communication Systems. Coimbatore, India, 2017: 1-4
- [7] Yu S, Zhiping L, Jili X. A formal model of intelligent query answering systems//Proceedings of the 2015 10th International Conference on Computer Science and Education. Cambridge, UK, 2015: 460-464
- [8] Chen Y, Cheng L, Chen C, et al. Wireless sensor network for data sensing in intelligent transportation system// Proceedings of the VTC Spring 2009-IEEE 69th Venetular Technology Conference. Barcelona, Spain, 2009; 1-5
- [9] Siddiqui S, Khan A A, Ghani S. A survey on data aggregation mechanisms in wireless sensor networks//Proceedings of the 2015 International Conference on Information and Communication Technologies. Karachi, Pakistan, 2015; 1-7
- [10] Boukerche A, Mostefaoui A, Melkemi M. Efficient and robust serial query processing approach for large-scale wireless sensor networks. Ad Hoc Networks, 2016, 47: 82-98
- [11] Liu M, Gong H, Wen Y, et al. The last minute: Efficient data evacuation strategy for sensor networks in post-disaster applications//Proceedings of the 2011 IEEE INFOCOM. Shanghai, China, 2011; 291-295
- [12] Liu R, Liang Y, Zhong X. Monitoring routing topology in dynamic wireless sensor network systems//Proceedings of the 2015 IEEE 23rd International Conference on Network Protocols. San Francisco, USA, 2015, 406-416
- [13] Oh H, Chi T N. A slotted sense multiple access protocol for timely and reliable data transmission in dynamic wireless sensor networks. IEEE Sensors Journal, 2018, 18(5): 2184-2194
- [14] Ammari H M. Investigating the energy sink-hole problem in connected-covered wireless sensor networks. IEEE Transactions on Computers, 2014, 63(11): 2729-2742
- [15] Xie Kun, Wen Ji-Gang, Zhang Da-Fang, et al. Bloom filter query algorithm. Journal of Software, 2009, 20(1): 96-108 (in Chinese)

(谢鲲,文吉刚,张大方等.布鲁姆过滤器查询算法.软件学报,2009,20(1):96-108)

[16] Fan Jun-Mei, Wang Bin, Wang Guo-Ren, et al. Improved Bloom Filter filtering technology in distributed environment. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2005, 33(s1): 205-208(in Chinese) (范俊梅,王斌,王国仁等.分布式环境下改进的 Bloom Filter 过滤技术.华中科技大学学报(自然科学版), 2005, 33(s1): 205-208)

- [17] Ahmed R K, Ibrahim D M, Sarhan A M. A communication load balanced dynamic wireless sensor network topology management algorithm based on AODV (CLB-AODV)// Proceedings of the 2016 11th International Conference on Computer Engineering and Systems. Cairo, Egypt, 2016: 315-322
- [18] Abedi O, Fathy M, Taghiloo J. Enhancing AODV routing protocol using mobility parameters in VANET//Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications. Doha, Qatar, 2008: 229-235
- [19] Liu X, Xu J, Lee W C. A cross pruning framework for top-k data collection in wireless sensor networks//Proceedings of the 2010 11th International Conference on Mobile Data Management. Kansas, USA, 2010: 157-166
- [20] Cheng J, Jiang H, Liu J, et al. On efficient processing of continuous historical top- k queries in sensor networks. IEEE Transactions on Vehicular Technology, 2011, 60(5): 2363-2367
- [21] Zhu R. Efficient fault-tolerant event query algorithm in distributed wireless sensor networks. International Journal of Distributed Sensor Networks, 2010, 6(1): 593849
- [22] Sun Li-Min. Wireless Sensor Networks. Beijing: Tsinghua University Press, 2005(in Chinese)
 - (孙利民.无线传感网络.北京:清华大学出版社,2005)
- [23] Kolcun R, Boyle D E, Mccann J A. Efficient distributed query processing. IEEE Transactions on Automation Science and Engineering, 2016, 13(3): 1230-1246
- [24] Tajibnapis W D. A correctness proof of a topology information maintenance protocol for a distributed computer network. Communications of the ACM, 1977, 20(7): 477-485
- [25] Chen H, Jin H, Guo L. Sink-free audio-on-demand over wireless sensor networks. IEEE Transactions on Computers, 2016, 65(5): 1606-1618
- [26] Dimakis A G, Sarwate A D, Wainwright M J. Geographic gossip: Efficient aggregation for sensor networks//Proceedings of the 5th International Conference on Information Processing in Sensor Networks. New York, USA, 2006: 69-76
- [27] Chen Y S, Tsou Y T. Compressive sensing-based adaptive top-k query over compression domain in wireless sensor networks//Proceedings of the 2017 IEEE Wireless Communications and Networking Conference. San Francisco, USA, 2017: 1-6
- [28] Chen I R, Speer A P, Eltoweissy M. Adaptive fault-tolerant QoS control algorithms for maximizing system lifetime of query-based wireless sensor networks. IEEE Transactions on Dependable and Secure Computing, 2011, 8(2): 161-176
- [29] Gao Z, Zheng J. Queries based on data attribute spatial index in wireless sensor networks//Proceedings of the 2018 International Conference on Intelligent Transportation, Big Data and Smart City. Xiamen, China, 2018, 183-186

- [30] Koutsonikolas D, Das S M, Hu Y C, et al. Hierarchical geographic multicast routing for wireless sensor networks.
 Wireless Networks, 2010, 16(2): 449-466
- [31] Tang B, Jaggi N, Takahashi M. Achieving data k-availability in intermittently connected sensor networks//Proceedings of the 2014 23rd International Conference on Computer Communication and Networks. Shanghai, China, 2014: 1-8
- [32] Luo L, Cao Q, Huang C, et al. Enviromic: Towards cooperative storage and retrieval in audio sensor networks//Proceedings of the 27th International Conference on Distributed Computing Systems. Toronto, Canada, 2007: 34
- [33] Mun J H, Lee J, Lim H. A new Bloom filter structure for identifying true positiveness of a Bloom filter//Proceedings of the 2017 IEEE 18th International Conference on High Performance Switching and Routing. Campinas, Brazil, 2017: 1-5
- [34] Bash B A, Byers J W, Considine J. Approximately uniform random sampling in sensor networks//Proceedings of the 1st International Workshop on Data Management for Sensor Networks: In Conjunction with VLDB 2004. Toronto, Canada, 2004; 32-39
- [35] Funke S, Kesselman A, Meyer U, et al. A simple improved distributed algorithm for minimum CDS in unit disk graphs.



LIANG Jun-Bin, Ph. D., professor His research interests include wireless sensor networks and intelligent services. ACM Transactions on Sensor Networks, 2006, 2(3): 444-453

- [36] Intanagonwiwat C, Govindan R, Estrin D, et al. Directed diffusion for wireless sensor networking. IEEE/ACM Transactions on Networking, 2003, 11(1): 2-16
- Liang Jun-Bin, Zou Shao-Jun, Li Tao-Shen. Data collection based on directional angle routing in mobile sensor networks. China Science: Information Science, 2015, 45(1): 111-128 (in Chinese)

(梁俊斌,邹绍军,李陶深.移动传感网中基于定向角度路由的数据收集.中国科学:信息科学,2015,45(1):111-128)

- [38] Peng Shao-Liang, Li Shan-Shan, Peng Yu-Xing, et al. A real-time and efficient data storage and query method in wireless sensor networks. Journal of Communications, 2008, 29(11): 128-138(in Chinese)
 (彭绍亮,李姗姗,彭宇行等.无线传感器网络中一种实时高效的数据存储和查询方法.通信学报, 2008, 29(11): 128-
- [39] Zhang J, Iyer S, Schaumont P, et al. Simulating power/ energy consumption of sensor nodes with flexible hardware in wireless networks//Proceedings of the 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks. Seoul, South Korea, 2012; 112-120

MA Fang-Qiang, M. S. candidate. His research interests include wireless sensor networks and intelligent services.

HE Zong-Jian, Ph. D. , professor. His research interests include ubiquitous computing and intelligent services.



138)

Background

This paper mainly studies the problem of data query processing in wireless sensor networks (WSNs). Since WSNs need to perform monitoring tasks for a long time and the battery energy of each sensor node is limited, most of the current research works are focused on low energy consumption of data query, and seldom pay attention to reliability and low latency of the query. Therefore, there are many open issues that still need to be further studied. In this paper, considering applications of disaster relief, military surveillance and other scenarios, users have higher requirements for reliability and real-time performance of data queries. We proposed a practical low latency and high reliable data query mechanism. Due to dynamic nature of network topology, unreliable communications, and node failures, the mechanism considers the following aspects: the integrity of the sensing data, the high probability of successful query, the minimization of the communication energy consumption and delay. Using effective data storage mode, an optimization model of minimizing communication energy consumption based on successful data query is established, and a distributed and load-balanced multi-path data dissemination algorithm and a reliable disaster tolerance routing protocol are proposed. Thereby, higher probability of successful data query, and lower communication delay and energy consumption of network communication are achieved.

This work is supported by the National Natural Science Foundation of China (Nos. 61562005 and 61762010), the Natural Science Foundation of Guangxi Zhuang Autonomous Region (Nos. 2018GXNSFBA281169 and 2019GXNSFAA-185042), the Cultivation Plan for Thousands of Young and Middle-Aged Backbone Teachers in Guangxi Higher Education School (Guangxi Education People (2017) No. 49).