

一种基于网格地理哈希表的无线传感器网络数据分发算法^{*})陶孜谨¹ 罗明宇² 欧阳一星¹ 龚正虎¹(国防科技大学计算机学院 长沙 410073)¹ (解放军体育学院 广州 510500)²

摘要 无线传感器网络是一种新型数据监测网络,其重要特性是传感器节点的能量有限,一般依靠电池驱动,能量效率是传感器网络设计最重要的考虑因素。GHT-DCS 是一种新型能量高效的数据分发方式,能在数据的查询和存储之间取得一种平衡。但是,采用 GHT-DCS 机制的能量效率仍然有改进的空间。本文提出了一种基于网格 GHT 的数据分发算法,也是一种以数据为中心的存储,并在此算法的基础上提出了一种基于索引存储的网格 GHT 数据分发算法,能更进一步提高能量效率。本文对这两种算法进行了性能分析,与原有的 GHT 算法进行了性能对比。分析表明,这两种算法在性能上都比原 GHT 算法有很大改进,而复杂度增加较少,是能量更加高效的数据分发算法。

关键词 传感器网络, GPSR, 网格 GHT, 索引-网格 GHT, 数据分发, 节点, 能量高效

A Grid-based Geographic Hash Table Data Dissemination Algorithm of Wireless Sensor Networks

TAO Zi-Jin¹ LUO Ming-Yu² OUYANG Yi-Xing¹ GONG Zheng-Hu¹(School of Computer, National University of Defense Technology, Changsha 410073)¹(PLA Institute of Physical Education, Guangzhou 510500)²

Abstract The Wireless Sensor Network (WSN or WSNs) is a new type of data monitoring network. One important characteristic of the WSNs is it is highly energy-constraint and powered by batteries. The most important design elements is the energy-efficiency. The GHT-DCS is an energy-efficient data storage method for the WSNs and achieves the balance between data querying and storage. But the DCS scheme based on the GHT still has some space to improve. In this paper a Grid-based GHT (GGHT) data dissemination algorithm is proposed and is also a type of DCS data dissemination method, and again on this basis a new index-based GGHT (I-GGHT) data dissemination algorithm is proposed. The performance analysis of the two algorithms are given and compared to the original GHT algorithm they can improve the energy-efficiency and the complexity of the algorithm has not been increased much, so the GGHT and I-GGHT are two energy-efficient data dissemination algorithms for the WSNs.

Keywords Wireless sensor networks, GPSR, Grid-based GHT (GGHT), Index GGHT, Data dissemination, Energy-efficient

1 引言

无线传感器网络由具有感应的、相互之间可以通信的、可计算的节点组成。许多路由、能量管理以及数据分发的协议都是用于无线传感网的,其中能量问题是最主要的。这些协议的侧重点根据应用以及网络结构而不同。作为信息时代的主流技术,传感器网络的出现给研究者带来了许多独特的挑战。这些网络大部分是由数以百计甚至是上千个微小的传感器节点、功能单元组成,在大部分情况下,这些节点都只有有限的能源^[1]。

以数据为中心的路由采用基于元数据标识的、以数据为中心的通信方式,在传感器网络上行之有效。在无线传感器网络中,将这种建立在分组路由之上实现监测范围内的分布式数据的收集和处理的办法,称为数据分发算法^[8~10]。

对于不同的应用传感器网络可采用不同方式组网,因此不同应用所适用的数据分发方案也不同。按照三种不同的数据存储方式,可以将目前常用的适用于无线传感器网络的数据分发方案分为三类^[2]:基于外部存储的数据分发方案、基于局部存储的数据分发方案以及基于以数据为中心存储的数据

分发方案。

对于一般无线传感器网络,其典型应用如森林火灾监控、交通监控、道路桥梁监控、环境监控以及战场监控等都是一类基于查询的数据分发模式。例如对于森林火灾监控,管理员向布置在一片区域的传感器发出“温度高于 50 度的传感器的位置、具体温度值”的查询,对于战场监控,作战人员发出“地面抖动 $> x$ 、持续时间大小 $> y$ ”的查询,收到查询后,只有满足条件或存储有相关数据的传感器才返回应答。因此,一般而言查询频率不高,满足条件的传感器节点也不多,传回查询节点的数据量并不大。但是由于传感器持续工作,采样数据量很大,所以应该尽量降低原始数据的存储开销。当查询时,也要尽量避免采用广播而消耗大量能量,并且能尽可能快地返回数据。

近年来提出了基于地理位置的路由^[4~6]、以数据为中心的存储 (DCS)^[2,11] 正好适应了这种需求,它们假定节点都通过某种带外机制(节点通过某种定位算法获知自身的位置,当节点定位后通过广播告知其他节点自己的坐标)已知网络区域内部署的各节点坐标,节点之间的消息传递基于地理路由, GPSR^[5] 能有效避免地理路由中出现的“空洞”,顺利到达目的

^{*}) 基金项目:国家“九七三”重点基础研究发展规划项目:“新一代互联网路由与交换理论”(2003CB314802)。

节点。

GHT^[2]数据分发算法是一种基于以数据为中心存储(DCS)^[2,10,11,14]方案的数据分发算法,沿用了以数据为中心的路由算法^[11~13]的基本思想,即查询分组和监测数据都采用元数据标识,通过分配不同的属性值来表示不同的任务或监测事件。基于DCS方案的数据分发算法中,特定监测事件存取点的选择至关重要。GHT算法根据元数据中特定的属性值来确定存取点,使同一类型监测事件的监测数据保存至同一存取点,而不再是直接在监测节点上保存。

GHT算法的核心步骤是建立哈希函数 $H(K)$ ^[3,7~9],将数据分组和查询分组元数据标识中的特定属性值关键字 K (如:事件类型属性值)哈希映射到一定的地理位置,并选择该地理位置附近的节点为存取点。GHT算法支持存和取两种相关操作:

Put(k, v)基于监测数据属性值的关键字为 K 的监测数据 V 。

Get(k)获取属性值关键字为 K 的监测数据。

特定的 K 值通过Put()操作或Get()操作都将映射到相同的地理位置,从而使对 K - V 对存或取都在同一存储点上进行。

GHT算法建立在GPSR^[4~6]路由选择算法之上,由存取点选择、数据分组存储、数据分组获取三个步骤组成。

对于这种数据分发机制,用户感兴趣的事件被分为几种类型,每种类型有一个唯一的关键词。通过特定的哈希函数 $H(\cdot)$,关键词被映射到探测区域的一个或多个位置。即有 $(\forall e \in E)H(e) = L_e \in R, E$ 是事件类型集合, R 是探测区域。

对于GHT算法,我们发现发送大量监测信息至存储节点,需要很大的通信开销,所以针对减小转发监测数据这个过程带来的通信开销,本文提出了一种基于网格的地理位置哈希表数据分发算法GGHT(Grid-based Geographic Hash Table),它也是基于以数据为中心存储的数据分发方案。该算法有效地减小了监测数据在网络中存储时传输的距离,减少了系统通信量。

2 网格地理哈希表数据分发算法(GGHT)

无线传感器网络用于监测一个区域内多个移动的目标,网络中的传感节点监测各个目标的状态,定期地产生感知数据。用户可以通过sink节点发送一个关于目标当前状态或某时间段内的所有活动的查询。

我们假设传感器节点是静态的,并且通过GPS^[1,15,16]来感知各自的地理位置,为了节省能量,缩短存储路径,把传感器网络等分成 n^2 个网格。当某个网格中的一个传感器节点监测到信息时,通过Hash函数映射为该传感节点所在网格中的某个传感节点,并把监测信息存储至该节点。一个区域划分网格的数目与应用类型密切相关,最优的网格数目在后面的性能分析部分给出。

GGHT算法与GHT算法一样,主要工作过程也由存取点选择、数据分组存储、数据分组获取三个步骤组成。但是,在系统的初始阶段,要根据历史经验数据对系统区域进行网格划分;在系统工作一段时间后,要根据现有应用的典型数据值,对网格进行调整,并在全网范围通知调整的结果;使得数据的存储和查询一致,并且总的能量消耗尽可能的小。

GGHT数据分发算法是建立在著名的GPSR^[5]路由算法之上的,当sink节点需要查询时,同时通过GPSR向所有网

格发出查询消息,当网格中的数据存储节点收到查询后,各网格中目的节点把匹配数据通过GPSR返回给sink节点。

2.1 GGHT初始化

把一个网络覆盖区域(把网络覆盖区域看成一个正方形)分为 n^2 等份,即 n^2 个网格。该算法与GHT算法的不同点在于,哈希函数的参数不再只有 k (即数据分组和查询分组元数据标识中的特定属性值关键字),而是加入了另一关于网格号 z 为参数,变成两个参数。

这样设计的主要原因是考虑到把数据的存储限制在各个网格内部,这样就降低了存储开销。而后面的分析表明,系统的其他开销,例如查询和数据刷新和备份导致的开销并没有随着网格的划分而明显增加,这样就降低了系统的总开销。

2.2 GGHT存储点的选择

在每个网格内部,GGHT算法与GHT算法一样,建立哈希函数 $H(k, z)$ 。通过Put()或Get()操作将分组的属性值关键字 k 映射为相应的地理位置 $H(k)$,此位置即为分组的目的存取位置。因为 $H(k)$ 位置上很可能没有分布有效节点,便利用GPSR路由选择算法贪婪模式和周界模式相互转变的特点,在 $H(k)$ 位置附近选择稳定有效的存取点——家乡节点^[5],作为Put()和Get()对同一关键字 k 操作的集合点。

家乡节点即距离 $H(k)$ 最近的节点。设分组 p 属性值的关键字为 k ,则 p 目的存取位置为 $H(k)$,用 N 表示 $H(k)$ 附近的节点集合,Distance(x, D)表示节点 x 到 D 的距离,若存在节点 y 满足下式: $\forall w \in N, \exists y \in N: \text{Distance}(y, H(k)) \leq \text{Distance}(w, H(k))$,则 y 为关键字 k 的家乡节点HOME _{k} 。

比较家乡节点的定义和GPSR算法从贪婪模式转为周界模式的条件可知,要么家乡节点位于 $H(k)$,要么GPSR算法令分组 p 在其家乡节点上进入周界模式。GGHT算法利用GPSR算法的此特性选择家乡节点。分组 p 在家乡节点上进入周界模式,随后将以周界模式遍历 $H(k)$ 位置的周围节点,直至返回家乡节点。遍历过程结束后形成的封闭周界圈为家乡周界,家乡周界上的非家乡节点为备份节点。

2.3 GGHT数据分组存储

通过 $H(k, z)$ 映射出属性值为在网格号为 z 属性值为 k 的监测信息的存储位置,监测节点根据自己的位置信息可以计算出自己所在网格号。当一个网格中的某个节点监测到数据时,这个哈希函数映射地理位置也在该节点所在网格中,而不会出现在其它网格中。当sink节点需要查询时,通过 $H(k, z)$ 找到各个网格的存储点的位置,再通过get(k)把查询分组发送到各个网格的存储点,从而获得监测数据。相对于GHT算法中的PRP更新协议,在我们算法中也要分别对四个网格中进行。

下面以一种简单哈希方案为例:一个网络覆盖区域为 $R * R$,通过计算出最佳 n 后,网络被分为 n^2 个网格,一个网格的大小为 $a * a (a = R/n)$,整个网络中当所有节点监测到 K 类型数据时,都先映射到地理位置 $G(x_G, y_G)$,某网格中监测节点 A 的位置为 $A(x_A, y_A)$,若 $(|x_A - x_G| < a) \wedge (|y_A - y_G| < a)$ 成立,则 A 的存储点为距 G 点位置最近的节点,若横坐标或纵坐标不满足,则通过 $x_G \pm a, y_G \pm a$ 来找到 G' 使上式成立,距 G' 最近的节点为 A 对应的数据类型为 K 的数据的存储点。

为了避免节点移动、节点失效等网络拓扑变化引起的存取点的失效以及存取点和备份点数据不同步造成数据丢失,GGHT算法引入周界刷新协议(Perimeter Refresh Protocol

简称 PRP)^[2] 实现 K-V 对备份和家乡节点的动态更新。PRP 令关键字 k 的家乡节点 HOME_k 以周期 T_h 产生更新分组 RP_k , 该 RP_k 分组的目的位置为 $H(k)$, 负载数据是其保存的所有 DATA-AN 信息。同进 $H(k)$ 采用 GPSR 算法将 RP_k 向 $H(k)$ 发送, RP_k 将遍历 k 的家乡周界 PERIMETER_k , 直至返回更新了家乡节点 HOME_k 。

在节点移动的情况下, HOME_k 的周界节点位置会发生变化, 这可能导致 RP_k 在周界上无休止地循环直至 TTL 为 0。为避免这种循环带来的大量开销, GGHT 算法把 RP_k 的 TTL 限制为 10 跳(hops), 中等规模的网络中家乡周界长度一般都小于 10, 因此将 TTL 固定设置为 10, 这已经足够完成周界刷新。

PRP 机制保证在网络拓扑发生变化后, 数据(如果对于后面提出的 I-GGHT 算法, 则为数据的索引, 下同)存储点仍然是离 $H(k)$ 最近的节点, 同时也保证了数据(索引)点和备份数据(索引)点所保存的数据(索引)信息是同步的。

PRP 中数据(索引)备份点设置刷新周期为 T_i , 每当其接收到 RP_k 时重设 T_i , 若 T_i 时间段之外仍未收到 RP_k , 则数据(索引)备份点产生 PR_k 分组并触发 PRP 周界刷新过程。这种机制保证了节点失效时数据(索引)信息的连贯性(persistence); 若家乡节点失效, 数据(索引)备份点在一段较长时间内未接到 PR_k 分组, 则触发 PRP 周界刷新过程, 选择新的家乡节点并使周界上节点的数据同步。数据(索引)点和数据(索引)备份点设置其存储信息的有效期为 T_d , 若节点收到新的 PR_k 时重设 T_d , 过期未接收到 PR_k 分组的节点将清空其数据(索引)信息缓存。为避免网络拥塞等原因造成的分组暂时性丢失, 算法设置 $T_d > T_i > T_h$ 。

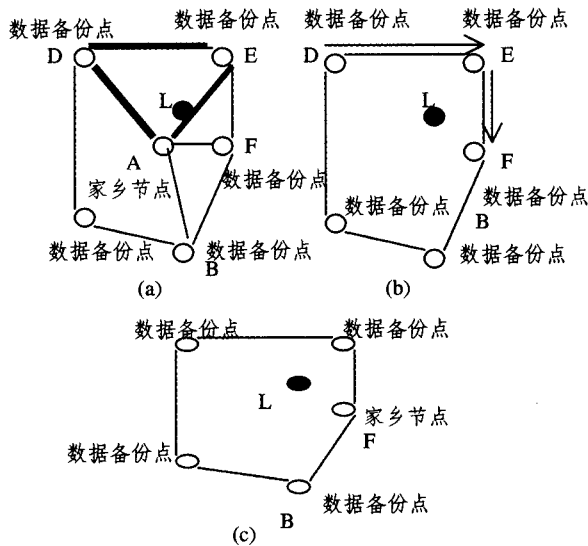


图1 PRP 周界刷新过程示例

图1的(a)到(c)为 PRP 周界刷新过程的一个示例。图1(a)中, L 代表 DATA 中属性关键字 K 的映射位置 $H(K)$, 节点 A 离位置 $H(K)$ 最近, 因此成为 K 的家乡节点, A 触发 PRP 周界刷新过程后形成由家乡节点 A , 数据(索引)备份点 D, E 构成的家乡周界。若节点 A 失效, 则时间 T_i 后, 由于节点 D 没有收到从 A 发出的更新分组, D 将触发周界刷新, 如图1(b)所示。 D 发出的更新分组将到达 F , 由于此时 F 离 $H(K)$ 最近, F 成为新的家乡节点。在 F 触发的周界刷新过程后形成由家乡节点 F , 数据(索引)备份点 B, C, D, E 构成的

家乡周界, 如图1(c)。

2.4 GGHT 数据分组获取

当终端节点发生一个查询请求时, 因为 sink 节点事先不知道哪些网格监测到了其所需信息, 必须向所有网格发送该查询请求。每次终端节点进行一次查询时, 把该查询报文发送 n^2 次, 每一次向一个网格发送(网格数为 n^2)。在网格内部, 采用前面提到的哈希算法把查询发送到数据(对于 I-GGHT, 则是索引)的存储点。当报文到达网格时, 该网格的存储节点查看自己是否有与查询相匹配的信息, 若有, 则该存储节点把匹配信息通过 GPSR 直接返回给终端节点; 如无匹配信息, 则丢弃查询, 不返回任何信息。

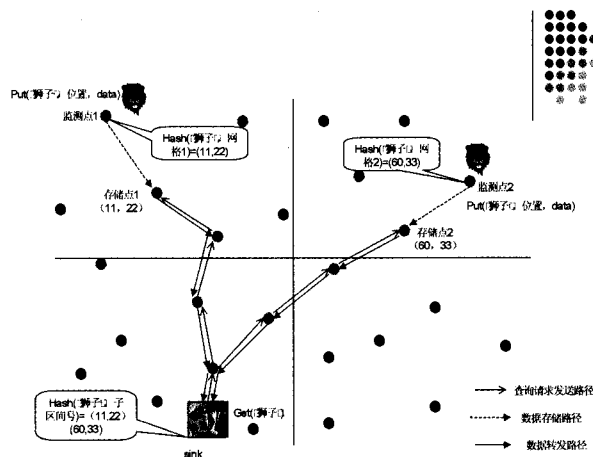


图2 四个网格的 GGHT 算法数据存取原理

图2描述了一个被分成四个网格的网络区域, 位于第一个网格内的监测节点1, 监测到属性值关键字为“狮子”的信息时, 把属性值关键字和网格号做为哈希函数的参数, 映射到同一网格内的一个存储点(11, 22), 并把监测数据发送至该点进行存储。网格2同理映射到存储点(60, 33), 该点在位于网格2内。当 sink 节点想要获取属性值关键字为“狮子”的信息时, 通过哈希函数找到四个网格关于“狮子”的存储点, 然后分别向四个网格发送查询信息, 由于只有1和2两个网格中存储得有匹配信息, 则通过 Greedy 算法返回存储信息至 sink, 网格3和4无匹配信息, 则把查询分组丢弃。

2.5 基于索引的 GGHT 算法 - IGGHT (Index-based GGHT)

在 GGHT 算法的基础上为了更进一步减少系统的通信量, 提出了一种基于索引存储的数据分发算法。在 GGHT 中, 传感器探测到的数据依然存储在非感知节点上, 当传感器采集到的数据很多时, 这一部分的数据通信量将大量消耗系统的能量, 而这些数据只有很少的一部分被查询。我们提出的基于索引的 IGGHT 算法就是为了解决这个问题。在该算法中, 事件的感知数据被存储在探测节点本身或与探测节点邻近的节点(这些节点叫做存储节点)。当存储节点接收到来自 sink 节点的查询时把数据发送到 sink。为了让存储节点的位置信息能够被 sink 或其他查询节点获知, 我们把事件数据的相应信息(例如位置、类型、时间等)构建成索引存储在事件类型关键字相关的地理哈希节点上。哈希函数把事件类型映射到网格内的一个或多个特定位置。所以即使是索引信息的存储也局限在一个网格内, 极大的节省了通讯所消耗的能量。

当 sink 节点要获取某种类型的数据时,先把查询发送到 hash 函数所获得的位置最近的节点,该节点上存储有数据的索引信息。该节点收到查询后,根据索引信息获得数据的实际存储位置,把查询转发到数据的实际存储节点。数据存储节点收到查询后,把数据内容发回 sink。

由于数据的存储在本地进行,而在哈希节点上存储的只是较小的数据索引信息,而查询报文一般不是很大,所以在监测时间长、数据存储量大、查询报文数量不多的情况下,I-GGHT 算法能更进一步降低系统通讯量,改进算法性能。算法的具体工作如图 3 所示。

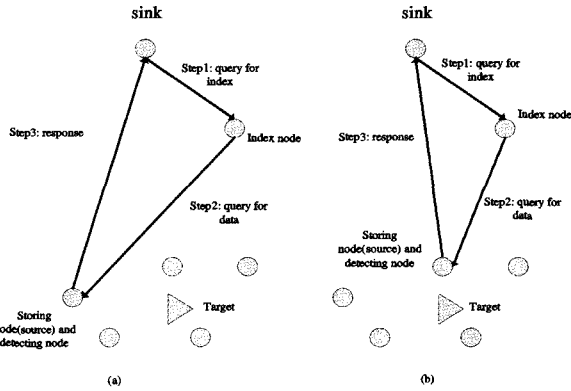


图 3 基于索引的网络 GHT 数据分发算法

3 GGHT、I-GGHT 算法性能分析

下面根据 GPSR 地理路由的特点对 GGHT 和 I-GGHT 算法的性能做一个分析。

假设网络覆盖区域 A , N 个节点平均分布在 A 中,节点分布密度为 $M(M=N/A)$,网络的通信范围为 $R \times R$,每个网格的通信范围为 $R \times R/4$,网络区域中监测到的总的数据量为 D_{total} ,数据大小 d ,终端节点共发送查询分组 Q ,大小为 d_l ,终端节点从区域中节点接收到 d_q 个数据分组,大小为 d ,RP 更新分组大小为 b 。相应地,如果采用索引方式,总的索引数据量为 $D_{total-an}$,索引大小为 d_{an} ,RP 更新分组大小为 b_{an} 。

在 GGHT 中,Put(k, v) 发送 D_{total} 个数据分组发送至监测节点所在网格存储点的通信开销为 $c\sqrt{\frac{N}{n^2}}D_{total}d$ 。 $c\sqrt{\frac{N}{n^2}}$ 是源节点发送数据至该源节点所在网格存储点的路径上的平均节点数($0 < c < \sqrt{2}$)。相应地,对于 I-GGHT,发送 $D_{total-an}$ 个索引分组至网格存储点的通信开销为:

$$c\sqrt{\frac{N}{n^2}}D_{total-an}d_{an}。$$

GGHT、I-GGHT 算法的 PRP 更新协议,该协议在 T 时间内发出更新分组 T/T_h 次,每次更新分组遍历整个家乡周界。家乡周界面积 S 最大为 $(\frac{R}{n})$ 。 f 个不同类型关键字 K 值,对于一个具有最大 f 值的网格来说,PRP 更新协议的最大通信开销为 $fM(\frac{R}{n})^2 b \frac{T}{T_h}$,那么对于整个网络覆盖区域,一共有 n^2 个网格,PRP 更新协议的最大通信开销为 $n^2 fM(\frac{R}{n})^2 b \frac{T}{T_h}$ 。对于 I-GGHT 算法,该值为 $n^2 fM(\frac{R}{n})^2 b_{an} \frac{T}{T_h}$ 。

对于终端节点,Get(k)操作向一个网格发送 Q 个查询消息(每个消息的大小为 d_l)的通信开销为: $c\sqrt{N}Qd_l$,对于 n^2

个网格,只需要发送 n^2 个查询分组,则总的通信开销为($n^2 c\sqrt{N}Qd_l$),终端节点共获取 d_q 个数据分组的通信开销为: $c\sqrt{N}d_q d$ 。对于 I-GGHT,发送查询消息的开销为 $2n^2 c\sqrt{N}Qd_l-an$,获取数据分组的通信开销与 GGHT 算法相同。

表 1 GGHT 参数列表(I-GGHT 类似,略)

参数	描述
N	区域 A 中节点数
M	节点分布密度
F	一个网格中不同类型关键字的个数
D_{total}	网络区域中监测到的总的数据量
d_q	网络中所有节点向终端节点返回的元数据总个数
$R \times R$	网络通信范围
$R \times R/n^2$	一个网格的平均通信范围
D	网络节点监测信息的大小,即:终端节点接收到的数据分组的大小
d_l	终端节点共发送查询分组的大小
T_k	每次发生 PRP 更新的时间间隔
F	不同数据类型的种类数
B	PRP 更新信息的大小
Q	终端节点共发送查询分组的个数

对于 GHT 算法,其总通信量为:

$$\lceil CO_{GHT} \rceil = c\sqrt{N}D_{total}d + fMR^2 b \frac{T}{T_h} + c\sqrt{N}Qd_l + c\sqrt{N}d_q d$$

以 4 个网格的 GGHT 算法为例,其总通信量为:

$$\lceil CO_{GGHT_4} \rceil = (c\sqrt{\frac{N}{4}})D_{total}d + 4fM(\frac{R}{2})b \frac{T}{T_h} + 4c\sqrt{N}Qd_l + c\sqrt{N}d_q d \quad (1)$$

n^2 个网格的 GGHT:

$$\lceil CO_{GGHT_{n^2}} \rceil = (c\sqrt{\frac{N}{n^2}})D_{total}d + n^2 fM(\frac{R}{n})^2 b \frac{T}{T_h} + n^2 c\sqrt{N}Qd_l + c\sqrt{N}d_q d \quad (2)$$

对于 n^2 个网格的 I-GGHT:

$$\lceil CO_{I-GGHT_{n^2}} \rceil = (c\sqrt{\frac{N}{n^2}})D_{total-an}d_{an} + n^2 fM(\frac{R}{n})^2 b_{an} \frac{T}{T_h} + 2n^2 c\sqrt{N}Qd_l + c\sqrt{N}d_q d \quad (3)$$

通过式(2)来找到 n 的最佳值,即使得 $\lceil CO_{GGHT_{n^2}} \rceil$ 最小化,那么,对式(2)求导,并使之为零,得 $n = \sqrt[3]{\frac{D_{total}d}{2Qd_l}}$ 为最佳 n 值,可让 GGHT 算法总的通信量达到最小。对于 $\lceil CO_{I-GGHT_{n^2}} \rceil$,当 $n = \sqrt[3]{\frac{D_{total-an}d_{an}}{4Qd_l}}$,I-GGHT 算法的通信开销最小。

GHT 算法的最大总通信开销(CO)为

$$\lceil CO_{GHT} \rceil = (c\sqrt{N})D_{total}d + fMR^2 b \frac{T}{T_h} + c\sqrt{N}Qd_l + c\sqrt{N}d_q d \quad (4)$$

$$\lceil CO_{GHT} \rceil - \lceil CO_{GGHT_{n^2}} \rceil = (1 - \frac{1}{n})(c\sqrt{N})D_{total}d + (1 - n^2)c\sqrt{2}Qd_l \quad (5)$$

对于 I-GGHT 有:

$$\lceil CO_{GHT} \rceil - \lceil CO_{I-GGHT_{n^2}} \rceil = c\sqrt{N}(D_{total}d - \frac{1}{n}D_{total-an}d_{an})$$

$$+f\frac{T}{T_h}MR^2(b-b_{an})+(1-2n^2)c\sqrt{NQd_l} \quad (6)$$

由式(5)可知,当监测数据分组大小 d 一定的情况下,GHT算法和划分 n^2 个网格的GGHT算法的网络通信开销的差值 $\lceil CO_{GHT} \rceil - \lceil CO_{GGHT,n^2} \rceil$,取决于 D_{total} 、 d_q 和 n (这里应该做一个定性的分析,例如,如果 D_{total} 比较大,而 Q 和 d_l 比较小,那么GHT算法的开销将大于GGHT算法,而且随着存储的数据量 D_{total} 不断增多,这一差距将不断扩大。对于I-GGHT算法,由于数据总量 $D_{total}=D_{total}-a_n$,而 $d \gg d_{an}$, $b \gg b_{an}$,所以当查询量不是很多时, $\lceil CO_{GHT} \rceil > \lceil CO_{I-GGHT,n^2} \rceil$ 。

4 GGHT 算法的具体数据分析

考虑200s时间内,假设参数值如表2。

表2 GGHT 参数值

参数	参考取值
网络区域中节点个数 N	200
一个监测信息的大小 d	500B
查询分组大小 d_l	25B
返回数据分组个数 d_q	160
不同数据类型 f (即事件类型)	20种
已监测到的数据类型 D_{total}	10/数据类型
节点密度 M	1个节点/16*16m ²
网络覆盖区域 $R \times R$	227m*227m
PRP更新报文的大小 b	25B
在时间段 T 内	200s
PRP周期更新时间 T_h	10s
查询开始时间	42s
查询产生的概率(由此算出 Q)	1/2s
查询分组 Q 的个数	79次

由上述参数,且假设 $c=1$,可得出 $n=\sqrt[3]{\frac{D_{total}d}{2Qd_l}} \approx 3$,最佳网格个数为 $n^2=9$ 。

在上述参数情况下,根据式(3)来验证一下(表3)。

表3 不同个数网格的GGHT通信开销

网格个数 n^2	$\lceil CO_{GGHT,n^2} \rceil$
1	4.35MB
4	3.76MB
9	3.67MB
16	3.74MB
25	3.91MB
36	4.16MB
49	4.47MB

由表3可知,把网络区域分为9个网格时,总通信开销最小,且划分的网格个数越接近9则网络总的通信开销越小。

对网络使用GHT数据分发算法,相当于把这个网络区域划分为一个网格,所以 $\lceil CO_{GHT} \rceil = \lceil CO_{GGHT,1} \rceil$,那么,由(5)式 $\lceil CO_{GHT} \rceil - \lceil CO_{GGHT,n^2} \rceil$ 可得,该网络划分的网格个数为 n^2 ($n \leq 6$)时,从网络总通信开销来看,GGHT优于GHT;当网格个数为 n^2 ($n \geq 7$)时,GHT优于GGHT。

为了让网络的总通信量可以尽量的减少,我们可以每隔一段时间,把前段时间网络的监测信息量和sink的查询通信量等信息作经验值,通过这些经验值对 $n=\sqrt[3]{\frac{D_{total}d}{2Qd_l}}$ 来求一个

最佳 n 值,然后把网络区域重新分成 n^2 个网格,以便实时对网络中网格的个数进行调整,把网络通信量减少到最小。

如果调整了网格数,sink根据当前数据计算出新的 n 值,形成新的哈希函数,该哈希函数使监测节点通过其计算,使映射出的存储位置与监测节点的同一网格内,然后在网络内广播此哈希函数,当网络内节点监测到目标数据时,按照新的哈希函数计算此监测信息的存储位置,以此来适应网络的调整。在实际应用,由于数据大小、节点数目、查询分组的大小在系统运行后基本保持不变,网格的调整只是在系统初始阶段进行,长时间运行时可以保持不变,因此网格调整次数不多,带来的开销不大。

总结及下一步工作 本文以无线传感器网络中数据分发算法的研究与改进为出发点,研究如何解决无线传感器网络中数据分发的能源有效性和负载均衡性的问题,在分析了无线传感器网络工作特点和现有数据分发算法的基础上,提出了两种新的算法GGHT和I-GGHT,对其性能进行了理论分析,证明其性能优于已有的GHT算法,能降低系统的能量消耗,延长系统的生存时间。下一步的工作主要是根据无线通讯的理论(能量)模型,实际分析这两种算法,在特定应用背景下(例如楼宇监控)采集各种数据信息,计算其能量消耗的具体值,与GHT算法进行对比。并将此理论应用到实际的传感器节点上(例如MicaZ)。

参考文献

- 1 US Naval Observatory (USNO) GPS Operations. <http://tycho.usno.navy.mil/gps.html>, April 2001
- 2 RatNasamy S, Karp B, Yin L, et al. GHT: A Geographic Hash Table for Data-Centric Storage[C]. In: ACM International Workshop on Wireless Sensor Networks and Applications, September 2002
- 3 Druschel P, Rowstron A. Pastry: scalable, distributed object location and routing for large-scale peer-to-peer systems[C]. In: Proceedings of the 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001), Nov. 2001
- 4 Karp B. Greedy perimeter state routing[C]. invited seminar at USC/ISI, Arlington, VA, USA, July 1998
- 5 Karp B, Kung H T. GPSR: greedy perimeter stateless routing for wireless networks[C]. In: Proceedings of the Sixth Annual ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom 2000), Boston, MA, USA, Aug. 2000
- 6 Karp B. Geographic Routing for Wireless Networks[D]: [Ph D Dissertation]. Division of Engineering and Applied Sciences, Harvard University, Oct. 2000
- 7 Ghose A, Grobklags J, Chuang J. Resilient data-centric storage in wireless ad-hoc sensor networks[C]. In: Proceedings the 4th International Conference on Mobile Data Management(MDM'03), 2003. 45~62
- 8 Intanagonwiwat C, Govinda R, Estrin D. Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks[C]. In: the Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking. 2000
- 9 Estrin D, et al. Directed Diffusion for Wireless Sensor Networking[J]. IEEE/ACM Transactions on Networking, February 2003
- 10 Kulik J, Rabiner W, Balakrishnan H. Adaptive Protocols for Information Dissemination in Wireless Sensor Networks[R]. Massachusetts Institute of Technology, 1999
- 11 Shenker S, Ratnasamy S, Karp B, et al. Data-centric storage in sensor networks[C]. ACM SIGCOMM HotNets, July 2002
- 12 孙利民,李建中,陈渝,等. 无线传感器网络[M]. 清华大学出版社, 2005
- 13 Chong C Y, Kumar S P. Sensor networks: Evolution, opportunities, and challenges[J]. Proceedings of the IEEE, 2003, 91(8): 1247~1256
- 14 Finn G G. Routing and addressing problems in large metropolitan-scale internetworks[R]: [Tech. Rep]. ISI/RR-87-180. Information Sciences Institute, Mar. 1987