

# 一种大规模传感器网络节能数据发布协议\*

石高涛<sup>+</sup>, 廖明宏

(哈尔滨工业大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

## An Energy-Efficient Data Dissemination Protocol in Large-Scale Sensor Networks

SHI Gao-Tao<sup>+</sup>, LIAO Ming-Hong

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: +86-451-86413213, E-mail: kyle\_shi@hit.edu.cn, <http://www.hit.edu.cn>

Shi GT, Liao MH. An energy-efficient data dissemination protocol in large-scale sensor networks. *Journal of Software*, 2006,17(8):1785-1795. <http://www.jos.org.cn/1000-9825/17/1785.htm>

**Abstract:** A critical issue in wireless sensor networks is energy-efficiently disseminating data from nodes to multiple sinks. Many data dissemination techniques have been proposed for sensor networks. However these protocols are based on flooding mechanism which wastes much precious energy. Some recent protocols have been proposed to reduce the flooding cost at different levels but can not guarantee the query success rate. In this paper, an energy-efficient data dissemination scheme DCS (diameter-chord scheme) is proposed which can reduce the energy consumption and improve the success rate as well. DCS exploits the fact that any chord in circularity vertically intersects with a diameter. Further more, Two-Phase protocol is proposed based on this dissemination scheme. This protocol works in two phases, and only when it is not successful in the first phase, the second is triggered. Two solutions are also proposed to deal with the delay of the Two-Phase. Extensive simulations and analysis are conducted to evaluate these protocols. The results show that the proposed protocols perform better than the peers.

**Key words:** sensor network; data dissemination; energy-efficient protocol; query scheme; disk network; success ratio

**摘要:** 无线传感器网络的主要功能是实现数据发布,在接收到信息查询时能够以有效的方式传输给查询者。目前的数据发布方式通常基于洪泛机制查询信息,浪费了有限的能源。虽然一些最近的数据发布协议从不同程度上解决了这一问题,但不能保证查询成功率。基于圆形节点分布网络模型提出了一种既能减少能源消耗,又能提高成功率的数据发布模式——直径-弦模式 DCS(diameter-chord scheme)。该模式利用了每个圆的弦都存在一个直径与之垂直相交的性质。在此基础上提出了 Two-Phase 协议。它在信息查询时分成两个工作方式不同的阶段进行,在第 1 段查询不中时触发第 2 段查找。为了减少 Two-Phase 的延迟,提出了基于优先级和概率转发两种解决方法。理论分析和实验结果都表明,所提出的数据发布协议具有较好的性能。

**关键词:** 传感器网络;数据发布;节能协议;查询模式;圆形网络;成功率

中图法分类号: TP393 文献标识码: A

\* Supported by the National Natural Science Foundation of China under Grant No.60533110 (国家自然科学基金); the Harbin Institute of Technology Foundation of China under Grant No.HIT 2002.74 (哈尔滨工业大学校基金)

Received 2005-07-09; Accepted 2005-11-08

传感器网络的主要目的是从环境中搜集信息,并把它送到汇聚点 sink 进行处理.因此,向 sink 节点发布信息(数据)是传感器网络的一个基本功能.根据应用类型的不同,数据发布的形式也各异.在持续管理应用中,sink 节点收到来自环境连续感知的数据;在事件驱动的应用中,传感器节点在感知到环境变化时沿着预定路径或者以按需的方式寻找路径向 sink 节点传输数据;在观测者触发的应用中,观测人向网络发出查询,符合条件的节点响应,并将满足条件的数据发给查询者.前两种方式具有一定的局限性,未来的应用将大部分是基于观测者触发的,比如单兵作战或环境信息查询等.在这一类型的应用中,多个观测者(sink)随机地在网络覆盖的区域内游走,需要信息时向网络发出查询.本文将以此一类应用作为研究背景.

由于传感器网络节点能源非常有限,因此,在数据发布时应当尽可能地节能.传感器网络数据发布的主要难点在于如何有效地传输查询和数据信息.基于洪泛的方法需要将查询信息转发到整个网络<sup>[1]</sup>,因此这一方法最为耗能,特别是当有多个 sink 需要查询时,洪泛的代价更高.文献[2-6]从不同程度上减少了查询和数据返回代价.然而,这些协议有些需要移动数据,有些查询成功率不高.

本文的工作是受文献[4,7]的启发.文献[4]提出的谣言(rumor)路由将查询和事件信息通过随机游走的方式在整个网络内传播,当事件和查询“偶遇”时,事件沿着查询路径返回给查询者.虽然这种方式可以减少查询代价,但查询的成功率非常低<sup>[8]</sup>.文献[7]在研究传感器网络节点重定位中,提出了基于法定人(quorum)思想的冗余节点发现方法,这种方法只有在网络节点覆盖区域呈方形的网络中才具有良好性能.

针对上述缺点,本文提出一个新的数据发布模式——直径-弦模式 DCS(diameter-chord scheme),通过将事件信息和查询有规律地进行缓存移动,不仅能够减少查询代价,而且可以提高查询的成功率.基于该模式提出了 Two-Phase 数据发布协议,能够有效地完成信息的发布和查询.

本文首先给出相关工作.然后,在第 2 节给出本文的网络模型和相关定义,并基于网格网络描述 DCS 的基本思想. Two-Phase 数据发布协议的详细工作过程在第 3 节给出.第 4 节进行性能分析.第 5 节通过仿真实验验证比较各种数据发布协议的性能.最后给出结论.

## 1 相关工作

传感器网络是近年来发展起来的技术,文献[9-11]对其相关问题研究进行了综述.数据发布是传感器网络的基本功能,其中的问题是如何将查询信息以有效的方式转发至目的节点或区域.目前,已有很多研究工作可以解决这一问题.由于查询者和数据源的数量不固定,随着场景和应用的不同而经常发生变化,因此,如何在数据的 push-pull 之间找到平衡点是一个主要的难点.

DD(directed diffusion)<sup>[11]</sup>和 GRAB(gradient broadcast)<sup>[12]</sup>采用了基于洪泛的方式转发查询消息.由于洪泛会因为网络风暴而产生较大的网络负载,网络性能将由于大量查询信息的洪泛而急剧下降<sup>[2]</sup>.这两个协议完全依靠查询信息的 push 过程.文献[2]将数据发布方式与具体应用相结合,但仍然是基于洪泛的.

文献[4]提出了谣言路由来查询信息,每个查询消息随机选择下一跳,每个事件广播到周围一个小范围内的节点上.文献[8]研究了 3 个不同的查询方式,分别是:(1) 查询消息随机游走;(2) 查询消息和事件消息都随机游走;(3) 查询信息随机游走,而事件信息在若干区域被缓冲.虽然这些方式不需要洪泛信息,但随机性会导致查询成功率不高.

文献[6]提出了一个梳子-针模型(comb-needle),每个数据源在周围邻居节点之间缓存,形成一个针状缓存区,而查询消息则形成类似梳子状的查询转发区,在推-送数据上作了很好的折衷.但是,这个协议的性能受到梳子柄长短的制约,也就是查询者的位置会影响到查询命中率.其原因是:该协议假定网络呈矩形,当梳子柄受到网络节点部署的限制时会比较短,此时,梳子(查询)就无法得到针(事件)了.Wang<sup>[7]</sup>在研究传感器网络节点重定位过程中,提出了基于法定人(quorum)思想的冗余节点发现方法.这种方法能够有效地减少查询代价,但这种方法只有在网络节点覆盖区域呈方形的网络中才具有良好性能,与文献[6]类似,当事件和查询者都位于网络边缘时就无法查找到事件.

文献[3]提出了一个两层的数据发布协议 TTDD(two-tier data dissemination),感知到信息的节点,主动创建

一个虚拟的网格进行元数据发布,每个格点在原数据发布的过程中维护着到源数据节点的路径,当用户查询数据时,查询信息能立即沿着该路径到达数据源.TTDD 引入了维护代价,并且由于事件信息需要发布到所有网格点上,因此,该性能受到网格大小的影响.文献[13]对 DD,TTDD 和 GRAB 这 3 个协议进行了比较.

文献[14]提出了一个称为要道(artery)的数据发布协议,将节点感知到的数据存放到要道节点上,由它建立其他节点到要道节点的路径.节点将探测到的事件发布给要道节点,查询在要道节点内进行.这种方法能够支持多点多个查询,但由于数据的大量移动会带来不必要的代价,因为有些缓存的数据可能不会被查询.SEAD(scalable energy-efficient asynchronous dissemination)<sup>[15]</sup>根据每个 sink 查询数据的频率为每个数据源建立一个  $d$ -树,将数据副本复制到位于 sink 和数据源之间的副本节点上.这种方式除了引入移动数据负载以外,还需要经常维护  $d$ -树结构,因此,在数据源较多和 sink 经常移动时,这种方式的性能会成为疑问.

文献[5]使用代理思想为数据源和 sink 节点各设置一个代理节点,所有代理节点组成代理树,数据源将数据沿着代理树以多播的方式发布给每个 sink 节点的代理,sink 直接查询代理.此外,文献[16]采用以数据为中心(data-centric)的存储方式,一些索引节点维护着数据源的位置,sink 节点查询索引节点得到数据源的位置,进而得到这些数据.由于部分区域需要维护着索引并且接受查询,这样会产生热点区域.文献[17]研究了基于事件触发的向移动的 sink 节点路由数据的方法,主要解决如何以节能的方式维护最初的路径.

## 2 网络模型和基本思想

### 2.1 网络模型

类似于文献[18,19],本文假设  $N$  个节点随机分布在半径为  $R$  的圆形区域内,如图 1 所示.在该区域内,有随时需要查询信息的若干随机游走的观测者,要求网络将满足一定条件的数据发送给查询者.圆形区域不会限制模型的一般化,因为当网络为其他多边形时,可以将该多边形的外接圆作为网络区域.因此,本文的协议虽然基于圆形网络,但同样可以应用于其他网络模型中.此外,本文作以下假设:

(1) 每个节点可以获得自己的位置信息,这可以通过定位协议或者 GPS 得到.

(2) 每个观测者能够将查询信息准确地发送到指定区域.目前已有很多应用需要依赖于这一假设<sup>[20,21]</sup>,也有很多定位协议能够满足这一要求<sup>[22,23]</sup>.

(3) 观测者了解网络部署区域情况,包括该区域的范围、面积等情况.这些信息可以在网络部署时获得.

(4) 查询者的移动速度不会在查询结果返回前离开发出查询时所处位置的通信范围之外.这一假设保证查询结果能够顺利返回给查询者.文献[3,15,17]给出了当查询者离开查询位置时的数据转发方案.

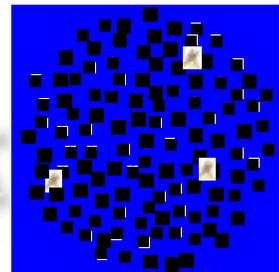


Fig.1 Node distribution model

图 1 节点分布模型

后面将讨论如何放宽假设(2)和假设(3).这样,即使观测者不了解网络部署区域的情况,协议也能正常工作.于是,只剩下假设(1)和假设(4).由于本文主要集中研究如何提高查询成功率,故作假设(4).基于以上模型,给出以下定义:

**定义 1(事发点,事件缓存区).** 事发点是探测到事件的节点.所有保存事件信息的节点形成一个虚拟的空间内存,称为事件缓存区,其大小用缓存事件的节点个数表示.

对事件  $E_i$ ,其缓存区表示为  $Cache(E_i)=\{node|node \text{ has } E_i\}$ ,其大小为  $|Cache(E_i)|$ ,在节点均匀分布时,可以使用这些节点占据的面积(长度)来近似表示.事件缓存区的平均大小可以反映数据发布系统的事件代价.

**定义 2(查询转发区,查询转发轴).** 所有转发查询信息的节点形成查询转发区,若查询转发区恰好和圆直径重合,则称该转发区为查询转发轴.

对于每个事件,可以构造一个元信息<sup>[1]</sup>(meta-data)与之对应,事件的元信息包含事件的发生地及与其相关的事件抽象信息.在不引起混淆的情况下,事件的元信息和事件信息在下文将交替使用,这里不再另作说明.

## 2.2 直径-弦模式(DCS)的基本思想

直径-弦 DCS 是利用了圆的一个基本性质:圆的任何弦都存在一个直径与它垂直相交,并且所有与该弦平行的其他弦也都和该直径垂直相交.根据这一性质,通过简单的调整事件和查询流向就能够使查询高效地找到感兴趣的事件信息,同时避免查询信息的洪泛带来的代价.

为便于解释,本节先以网格网络为例来阐述 DCS.如图 2 所示,如果事件信息沿着水平(垂直)方向从事发点向两端扩展,并且查询消息沿着该圆形区域的垂直(水平)方向的直径转发,那么,最终查询消息就能够找到事件信息(在以下的讨论中假设事件信息水平扩散).DCS 的整个数据发布的功能包含以下 3 个部分:

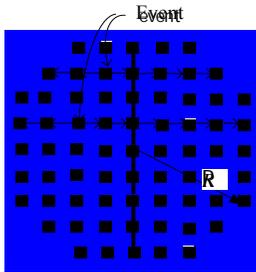


Fig.2 DCS in grid network

图2 网格网络中的直径-弦模式

(1) 事件缓存区的构建:每个探测到事件的节点将元信息传输给位于水平方向上的邻居节点,这些邻居节点记录下该信息后再转发给自己的其他位于水平方向上的邻居节点.依次下去,最后与探测到事件的节点所在同一个水平方向的节点都收到了该事件的信息,从而形成一个事件缓存区.

(2) 查询转发区的构建:查询者在需要信息时发出查询消息,该消息被一个节点沿水平方向更靠近垂直直径的节点转发,直到该消息被位于垂直直径的节点收到,然后再传输给该垂直直径上的所有节点.这样,所有位于直径上的节点形成了查询转发轴.

(3) 数据传输:当事件信息都能和查询信息相遇时,位于相遇点的节点判断查询信息和自己缓存的事件元信息是否一致,如果一致就沿着收到事件信息的路径逆方向传输给事发点,当事发点接收到查询消息后,就将数据沿着查询消息经过的路径的逆方向传输给查询点.

由以上描述过程可以看出:在理想情况下,只要查询信息能够最终通过网络覆盖区域的垂直直径,该网络中的所有事件信息就都可以成功查询到.基于该模式,本文提出 Two-Phase 数据发布协议.

## 3 Two-Phase 数据发布协议

Two-Phase 数据发布协议在查找信息时分成两个不同的阶段进行:第 1 段称为 Quorum,其基本思想与文献[7]类似;第 2 段称为 One-Step,是基于 CDS 模式的.在第 1 段查询不中时触发第 2 段.将两段相结合的原因参见第 3.4 节中的讨论.

本文使用一个基于地理信息的受限洪泛协议 CGF(constrained geographical flooding)<sup>[24]</sup>来完成路由.在 CGF 中,每当一个节点接到数据包时,它根据目的节点的位置来决定是否广播这一数据包.本文对 CGF 作了改动,加入贪心和随机算法,选择所有满足条件的距离目标节点(区域)最近的节点转发数据包.这一改动的效果与 CGF 一样,但却减少了能源消耗.详细的实现将结合数据发布的过程来描述.

在实际网络中,节点之间的位置关系不像网格网络那样完全水平或垂直.因此,类似于文献[6],为了保证查询信息和事件信息能够相遇,协议将查询转发轴和事件缓存区宽度设置为单位距离  $w$ .假设通信半径为  $r$ ,那么,实际物理宽度为  $w \times r$ .协议要求只有位于该宽度之内的节点才有资格转发查询或者事件信息.

另外,假定圆心为整个坐标系统的原点.Two-Phase 本质上是 Pull-Push 相结合的发布方式,因此,这一节首先描述事件信息的发布过程,然后再给出查询的两段过程.

### 3.1 事件数据发布

每当一个节点探测到一个事件时,由它负责事件缓存区的创建.假设探测到事件的节点坐标为  $(m,n)$ ,事件缓存区应当是所有节点坐标  $(x,y)$  满足下述条件的节点组成的区域  $E = \{(x,y) | m-w/2 \leq y \leq m+w/2\}$ .因此,当事件元信息转发时,只有位于这个区域的节点负责缓存并转发.每个位于区域  $E$  中的节点接收到事件信息时,根据它和数据源(上一跳节点)的距离  $d$  设置一个随机延迟,该延迟与  $d$  成反比.当延迟到时的时侯,节点继续转发该事件元信息,其他在延迟时间内的节点听到有其他节点转发该数据时,取消转发企图.算法如图 3(a)所示.

*m* is event message, *q* is query message, *d* is the distance between the receiving node and sending node

| Event caching algorithmic   | Query forwarding algorithmic   |
|---|--|
| <pre> <i>n</i> broadcast <i>m</i> when detecting event For every node <i>v</i> receiving <i>m</i> do   if <i>v</i> in <i>E</i>(<i>n</i>) &amp;&amp; <i>m</i> is new then     store <i>m</i> and the up-hop node     <i>T</i>=delay(<i>d</i>)     if not listen <i>m</i> in <i>T</i> then       broadcast(<i>m</i>)     endif   endif endif enddo                     </pre> | <pre> query node <i>g</i> broadcast <i>q</i> For every node <i>u</i> receiving <i>q</i> do   if <i>u</i> in <i>B</i>(<i>g</i>) ∪ <i>A</i><sub>II</sub> &amp;&amp; <i>q</i> is new then     record the source of <i>q</i>     <i>T</i>=delay(<i>d</i>)     if not listen <i>q</i> in <i>T</i> then       broadcast(<i>q</i>)     endif   endif endif enddo                     </pre> |
| (a)   | (b)  |

Fig.3 Algorithms of query forwarding of Phase II and events caching

图 3 Phase II 查询转发和事件缓存算法

### 3.2 Phase I: Quorum

假设查询者的坐标为(*m*,*n*),那么,Phase I 的查询转发区为  $A_I = \{(x,y) | m-w/2 \leq x \leq m+w/2\}$ ,所有事件缓存区呈水平方向,如图 4(a)所示.其中竖轴是转发区,横轴为事件缓存区.这样,整个事件查询分成两段,首先,每当查询者查询信息时,只有位置位于区域  $A_I$  的节点负责转发.每个位于区域  $E$  中的节点接收到事件信息时,根据它和数据源(上一跳节点)的距离 *d* 设置一个随机延迟,该延迟和 *d* 成反比,当延迟到时的时侯,节点继续转发该事件元信息,其他在延迟时间内的节点听到有其他节点转发该数据时取消转发企图.仅靠这一过程,在查询边缘事件时成功率比较低.Two-Phase 通过转移查询转发路径提高查询边缘事件的成功率.所以,如果这次查询不中,节点将按照第 3.3 节中描述的过程转发查询信息.

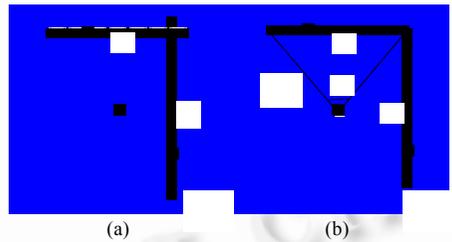


Fig.4 Query model in Phase I

图 4 Phase I 查询模型

第 1 次查询不中的判断可以根据时间延迟  $T_d=T(R)$ 来决定,它是网络半径的函数.如果在这个时间延迟内接收不到事件信息,理想情况下可以认为查询不中.此外,在实际网络中,查询不中的原因有很多,比如网络状态的影响.但只要在第 1 次查询不中后,Two-Phase 就触发第 2 阶段的查询.

### 3.3 Phase II: One-Step

One-Step 是基于第 2.2 节中的 DCS 数据发布模式来工作的.下面详细描述其工作过程.

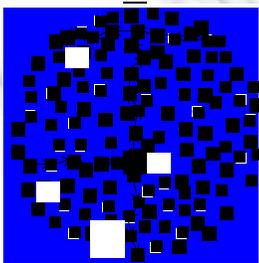


Fig.5 The dissemination process of Phase II

图 5 Phase II 发布过程

虽然节点知道自己的坐标,但并不了解整个区域的情况.因此,它无法确定本身是否就处于转发轴上,但查询者可以了解到整个区域的范围(第 3.4 节中将给出这一假设不成立时的解决方法).为此,查询者在发送查询消息时,根据自身的位置以及网络区域信息计算出转发轴的位置和区域,从而可以得到位于查询转发轴上的节点位置信息.根据这一信息,当查询者发送查询信息时,借助于节点位置信息就能够将查询信息转发到位于查询转发轴上的节点.注意:查询消息中包含查询点的位置信息.

如果查询者的坐标为(*m*,*n*),那么,查询消息需要先转发到区域  $C$  中的节点,其中  $C = \{(x,y) | -w/2 \leq x \leq w/2, n-w/2 \leq y \leq n+w/2\}$ ,如图 5 所示的斜线区域.因此,只有位于区域  $B = \{(x,y) | n-w/2 \leq y \leq n+w/2\}$  中并且更靠近区域  $C$  的节点才有资格转发并缓存消息.每当一个节点接收到一个查

询消息时,它就根据消息源的位置信息、区域  $C$  的位置信息(可以从消息中获取)以及自己的位置信息来判断是否位于区域  $B$  中.所有有资格转发的节点根据与上一跳节点的距离  $d$  来设置一个随机延迟,该延迟与  $d$  成反比,当时间延迟到时的时,该节点就转发查询消息.如果在这一时间延迟内收到其他节点的转发消息,该节点则取消发送.区域  $C$  中的节点在接到查询消息后,它同样按照上述过程将查询消息转发到区域  $A_{II}$  中.其中  $A_{II}=\{(x,y)|-w/2\leq x\leq w/2\}$ ,为如图 5 所示的垂直长方形.算法如图 3(b).

每一个探测到事件的节点都要负责创建事件缓存区.这样,当查询消息在查询转发轴中转发时,位于事件缓存区中的节点就有机会收到该查询消息.当事件缓存区中的节点接收与之匹配的查询时,可以认为事件查询成功.于是,它就将查询消息沿着事件消息转发的路径逆向传输给事发点.当事发点接收到查询消息以后,再沿着查询消息所经过的路径的逆方向将数据传输给查询点.可以看出,该协议的主要优点是能够以简单的方式查询到发生在网络边缘的事件信息.下一节将详细讨论两段发布的动机及其他相关问题.

### 3.4 讨论

#### 3.4.1 Two-Phase 发布的动机

Two-Phase 数据发布协议在最坏情况下需要两步完成信息查询.采用两步查询的动机如下:

由于事件信息的分布是随机的,因此,如果像文献[7]那样只进行 Phase I,则当事发点位于上下区域边沿,而查询者恰好位于左右区域边沿时,就无法查询到这些信息.另外,虽然单独进行 Phase II 完全可以解决 Phase I 的缺陷,而且过程简单.但由于 Phase II 每次都要经过一个固定的区域(转发轴),从而造成了热点(hot-spot)问题,导致热点区域的节点快速死亡,从而降低了网络寿命.增加 Phase I 后对一些事件信息可以不再进行 Phase II.

Two-Phase 数据发布协议不仅能够减少信息查询时的能源消耗,而且能够以简单的方式查询到发生在网络边缘的事件信息.这是其他协议所未考虑到的.总之,Two-Phase 数据发布通过将 Quorum 和 One-Step 相结合,不但提高了查询成功率,而且避免了热点问题.

#### 3.4.2 查询延迟

通过上文的讨论可以看出,Two-Phase 能够减少热点地区的负载.然而,这种方法也带来一定的查询延迟.文献[4]中提到在一次查询不中时再一次查询,但并没有考虑延迟问题.下面给出两种方法来降低查询延迟:第一,由于 One-Step 可以独立工作,为此,查询者可以对任务紧迫的查询赋予高优先级,直接按照 One-Step 转发,降低查询延迟,而任务不紧迫的节点采用 Two-Phase 查询.这样,高优先级的任务时间延迟就能够得到保证.第二,若任务紧迫程度相同,则可以采用概率转发:每个查询消息以  $p$  的概率按照 Two-Phase 协议查询,以  $1-p$  的概率采用 One-Step 来查询.下面分析  $p$  的取值.

如图 4(b)所示,首先固定事发点,查询者在圆内随机游走.若进行 Phase I 查找,那么,由于只有当查询缓存区  $A_I$  和事件缓存区相交查询才能成功.因此,Phase I 查询成功概率为  $(a \times b + 2S_{shadow})/\pi R^2$ .令  $f(\theta) = (a \times b + 2S_{shadow})/\pi R^2$ ,根据图 4(b)有  $f(\theta) = ab + \theta R^2/2 - (R^2 \sin \theta)/2$ .  $f(\theta)$  在  $\theta = \arccos(-1/3)$  时取得最大值  $\max f(\theta) = R^2(2\sqrt{2} + \arccos(-1/3))/2$ .因此,Phase I 对一个事件信息查询成功最大概率为  $(\max f(\theta))/\pi R^2 \approx 0.7$ .这也表明在进行 Two-Phase 查询时,最大会有 70%的查询能够在 Phase I 查找成功.因此,概率性查询可以对每一个查询采用 0.7 的概率、按照 Two-Phase 查询,而采用 0.3 的概率、按照 One-Step 协议进行查询,从而在一定程度上减轻了查询轴的负载.

#### 3.4.3 网络定位

第 2.1 节中假设观测者能够获得网络覆盖区域的全局情况,当不满足假设时,可以通过以下方法来近似估计.

1. 基于跳数(hopnum-based)估计.这种方法是借鉴文献[25]中的借助节点跳数来建立虚拟坐标的思想.查询者发送查询消息之前,首先发出一个探寻消息,该消息依然在水平查询区域内进行转发.所不同的是,该消息中包含一个跳数,并且消息向左、右两个方向转发.每个节点接收到该消息时记录下消息经过的跳数,在将跳数加 1 后继续转发,直到消息到达了最边沿.这样,在这个水平转发区域中,每个节点都记录着距离查询点的跳数.此时,两端的节点  $n_1$  和  $n_2$  可以将自己的跳数(假设分别为  $h_1$  和  $h_2$ )回送给查询点,查询点根据两个端点的跳数就可以计算出查询消息应当向哪个方向转发、转发多少跳能够到达近似的中间位置.如果  $h_1 > h_2$ ,那么,查询消

息应当向  $n1$  方向转发;否则,向  $n2$  方向转发. 查询消息由查询点再走  $|h1-h2|/2$  跳就到达了区域的“中间位置”.

2. 基于坐标(coordinate-based)估计. 其实现方法与基于跳数的方法类似,只不过探寻消息不包含跳数. 边缘节点将自己的坐标  $x1$  和  $x2$  回送给查询点,查询点据此来估计区域的中间位置. 估计方法与基于跳数的方法类似. 图 6 给出了 3 种方式的归一化误差比较. 可以看出,第 2 种方法估计值比较接近基于网络区域范围的方法.

#### 4 性能分析

本节将分析 3 种类别的数据发布方式的能源消耗,它们分别是 One-Step, Two-Phase 和 Quorum.

代价模型:所有数据发布协议都包含以下代价:查询代价  $Q$ 、数据缓存代价  $H$  和数据传输代价  $F$ . 查询代价是指网络为了将查询信息转发到事件信息发生地所需要的代价;数据缓存代价是指网络为了缓存探测到的数据所需要的代价;数据传输代价是事件信息传输到查询者所需要的代价. 当数据传输时,可以沿着查询信息的逆路径返回,也可以直接依靠基于地理信息的路由协议传输给查询者. 当采用后一种方法时,这一部分的代价可以认为是一样的,而且几乎所有的数据发布协议都可以依赖这种方式回传数据. 因此,下面的分析只讨论了查询和事件缓存的代价.

假设  $N$  个传感器节点分布在半径为  $R$  的圆形区域内,不考虑边界效应(edge effect),代价可以用查询转发区和事件缓存区的长度来近似表示. 假设事件的发生频率为  $f_e$ ,并且在整个网络内均匀分布,查询频率为  $f_q$ ,查询者在该区域随机游走.

1. One-Step 代价. 如图 7 所示,事件缓存区平均长度为

$$M = \int_0^R 2\sqrt{R^2 - x^2} \times \left( \frac{2\sqrt{R^2 - x^2} dx}{\pi R^2 / 2} \right) = \frac{8}{\pi R^2} \int_0^R (R^2 - x^2) dx = \frac{16R}{3\pi}.$$

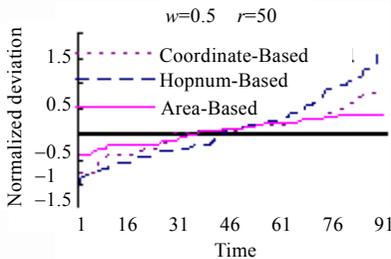


Fig.6 Normalized deviation comparisons

图 6 归一化误差比较

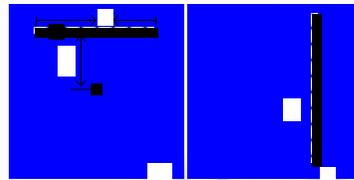


Fig.7 Position of event and query

图 7 事件和查询的位置

因此,缓存区代价平均为  $H_1 = f_e M = 16Rf_e/3\pi$ .

查询代价可以分成两部分:一部分是从查询点到直径的代价;另一部分是查询消息沿直径传输的代价. 采用与缓存代价同样的计算方法,可以得到第 1 部分为  $E(x) = \int_0^R x \frac{2\sqrt{R^2 - x^2} dx}{\pi R^2 / 2} = \frac{4}{\pi R^2} \int_0^R x\sqrt{R^2 - x^2} dx = \frac{4R}{3\pi}$ , 第 2 部分的长度为  $2R$ . 因此,查询代价为  $Q_1 = f_q E(x) = f_q (2R + 4R/3\pi)$ .

因此,查询和缓存代价为  $C_{One-Step} = H_1 + Q_1 = f_e M + f_q E(x)$ , 整理得到  $4R(4f_e + f_q(6\pi + 1))/3\pi$ .

2. Two-Phase 代价. 两段式查询需要考虑两个查询转发区的长度,其中第 1 个的平均长度与事件缓存区的平均长度一样. 那么,第 1 段查询的代价为  $Q_2 = 16Rf_q/3\pi$ . 因此,两段式查询的平均代价为  $C_{Two-Phase} = f_e M + Q_2 + q \times Q_1$ . 其中  $Q_2$  是 Two-Phase 必须完成的一步,如果这一步没有成功则进行第 2 步,而没有成功的概率为  $q$ .

3. Quorum 代价. Quorum 的代价为  $C_{Quorum} = f_e M + Q_2$ .

从以上分析可以看出,整个代价  $C_{Quorum} < C_{One-Step}$ ,  $C_{Quorum} < C_{Two-Phase}$ . 而  $C_{Two-Phase}$  和  $C_{One-Step}$  的代价则随着  $p$  的变化而变化.  $C_{Two-Phase} - C_{One-Step} \approx f_e(16 - 22.6p)R/3\pi$ . 因此,若  $C_{Two-Phase} > C_{One-Step}$ , 则有  $p < 16/22.6 \approx 0.71$ . 所以,大部分情况下,两段式查询要比 One-Step 耗能;当第 1 段的查询成功率比较高时(比如事发点比较靠近区域中心),两段

式查询的代价要小于 One-Step;一种极端的情况是  $q=0$ ,此时,Two-Phase 的代价相当于 Quorum.图 8 给出了当  $R=250, f_e=1, f_q=0.1$  时,3 种信息查询方式的代价比较.

### 5 仿真实验

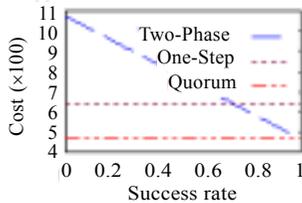


Fig.8 Protocol cost comparison  
图 8 协议代价比较

实验主要验证信息查询成功率、路由负载、能源消耗和两段式查询的性能.为了更清楚地说明协议的优点,根据事件缓存区的位置定义 3 种事件类型,分别是位于网络覆盖区域边缘地带的事件——边缘事件(Event#1)、位于网络覆盖区域中心地带的事件——平分半径事件(Event#2)和位于边缘地带和中部地带之间的事件——中心事件(Event#3).以上 3 种事件是一个模糊的概念,并没有严格的量化.

对网络能源消耗以数据包的个数作为代替.如前所述,由于考察的协议在查询到事件信息回送数据时的能源消耗可以认为是一样的,因此,实验统计的数据包负载只包括缓存查询数据包、广播数据包和缓存事件数据包.所有实验都是在 5 个随机产生的网络拓扑上运行平均结果,每种类型的事件随机产生 5 个,对每个事件依次查询 50 次.实验中的主要默认参数见表 1.

Table 1 The default parameters in experiment

表 1 实验默认参数

| Parameter                    | Value          | Parameter                    | Value     |
|------------------------------|----------------|------------------------------|-----------|
| R                            | 250m           | Initial energy               | 500 units |
| Event#1                      | $  y -200 <50$ | Buffer width $w$             | 0.5       |
| Event#2                      | $  y -100 <50$ | Needle length $L^{[12]}$     | 2         |
| Event#3                      | $ y <50$       | Comb width $S^{[12]}$        | 5         |
| Number of nodes $N$          | 600            | Radio range $r$              | 50m       |
| Energy consumed by broadcast | 5 units        | Energy consumed by receiving | 1 unit    |

#### 5.1 参数 $w$ 的影响

实验考察当  $w$  在取不同值时,Two-Phase 协议性能的变化情况.由图 9(a)可以看出:随着  $w$  的增大,查询成功率也相应地增大,但这是以查询能源消耗的增大为代价的.此外,当  $w$  增大时,第 1 步查询的成功率也相应地增大,从而进行第 2 步的概率也降低了,此时节省了第 2 步查询的能源消耗.因此,能源的消耗并不完全随着  $w$  的增加而绝对增加.如图 9(b)所示,对于不同的事件,能源的消耗变化趋势也不尽相同.对于第 1 类事件,第 1 步的查询成功率比较低,因此,第 2 步的可能性增加,从而能源消耗随着  $w$  的增加而增加.但是,对于第 2 类和第 3 类事件的查询则恰好相反,第 1 步的成功率增加,从而减少了第 2 步的能源消耗.进一步的实验取  $w=0.5$ .

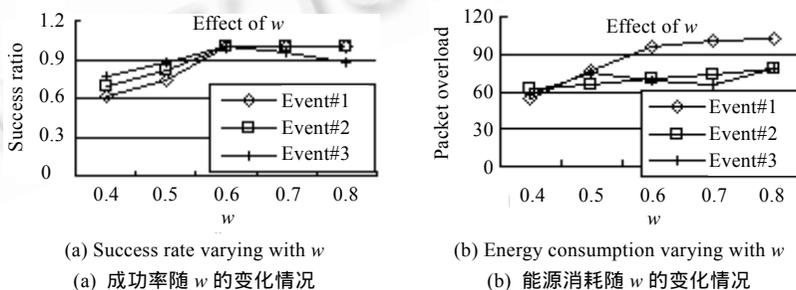


Fig.9 The effect of  $w$  on performance

图 9  $w$  对性能的影响

#### 5.2 One-Step vs. Two-Phase

图 10 给出了 One-Step 和 Two-Phase 方式的性能比较.从图 10(a)可以看出,Two-Phase 式查询在成功率方面

优于 One-Step,其原因是多引入了 1 次查询.但这种方式是以耗费更多的能源为代价的,由图 10(b)可以看出这一点.

热点区域会导致这一区域的节点过早死亡,从而影响查询的成功率.由图 10(c)可以看出,随着仿真的进行,One-Step 方式的查询成功率逐渐降低,其中的主要原因就是热点区域的节点不断死亡.而 Two-Phase 方式的查询成功率则比较稳定,只在经历很长一段时间(850s),成功率才开始降低.虽然 Quorum 能源消耗比较少,并且也没有热点问题,但其成功率则极其不稳定.

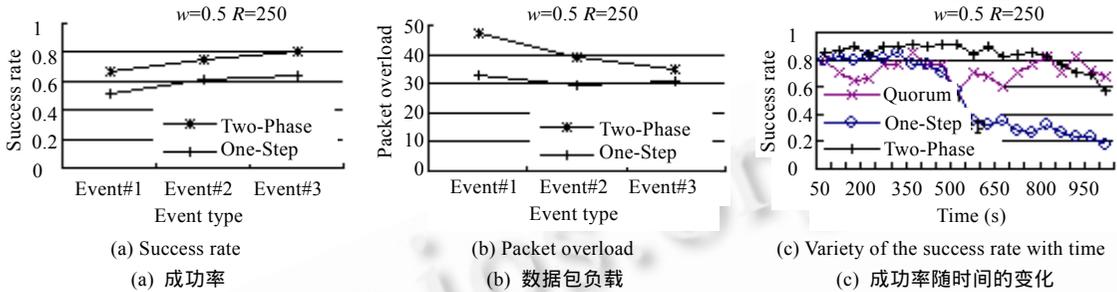


Fig.10 One-Step vs. Two-Phase  
图 10 One-Step 与 Two-Phase 比较

### 5.3 与其他协议比较

本节对 Two-Phase, Quorum<sup>[6,23]</sup>和 Comb-Needle<sup>[12]</sup>这 3 种不同的数据发布协议性能进行比较.图 11 给出了 3 种协议在查询不同类型的事件时的成功率.可以看出,Two-Phase 能够有效地查询到各种类型的事件,特别是对于边缘事件,Two-Phase 查询方式的成功率要明显高于其他两种方式;而对于其他两种类型的事件,3 种事件发布方式比较接近,这和我们的预期一致.

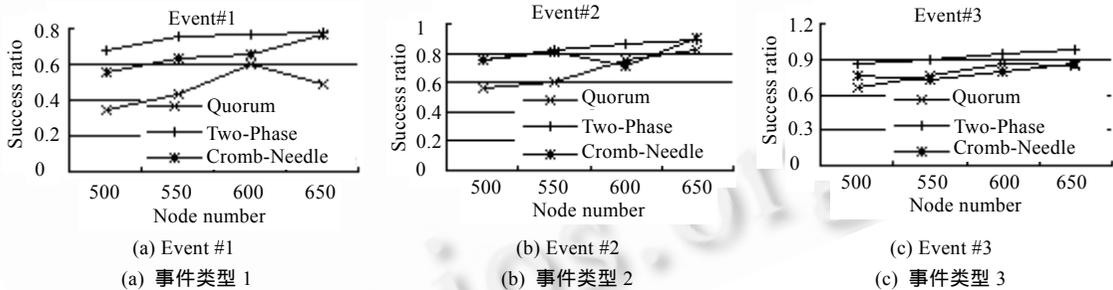


Fig.11 Success rate when querying different event  
图 11 查询不同类型事件的成功率

图 12 给出了 3 种不同协议的平均性能.由图 12(a)可以看出,在成功率方面,Two-Phase 仍然是最高的.此外,Quorum 的成功率维持在 70%左右,从另一方面验证了我们的证明.为了能够更清楚地观察到 Two-Phase 在成功率和能源节省方面的优越性,我们考察了 3 个协议每单位能源查询到的事件.为此,我们以查询到的事件数/数据包负载作为不同协议的指标.图 12(b)给出了 3 种协议的性能情况.可以看出,Two-Phase 能够以较少的能源获得较高的查询成功率.因此,综合查询成功率和协议负载,本文提出的数据发布方式性能较好.

### 6 结论及将来的工作

本文研究了无线传感器网络中的数据发布问题.在分析了现有发布协议缺点的基础上,基于圆形节点分布网络模型提出了一个新的数据发布模式 DCS.该模式通过将事件信息和查询有规律地进行缓存移动,不仅能够减少查询代价,而且可以提高查询的成功率.然后提出了 Two-Phase 协议,协议综合了 DCS 和 Quorum 的优点,

同时考虑到 Two-Phase 带来的延迟,给出了两个解决方案,分别是基于优先级转发和概率转发.实验和理论分析都表明:本文提出的数据发布协议具有较好的性能.

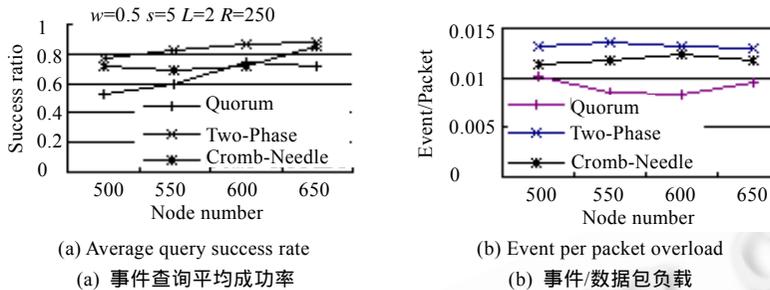


Fig.12 Average performance comparison

图 12 平均性能比较

进一步的工作是:将数据发布和节能机制结合在一起,此时节点密度相对稀疏,并且节点周期睡眠.如何在这种情况下有效地完成数据发布是我们将来的工作.

#### References:

- [1] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In: Pickholtz R, ed. Proc. of the 6th Annual Int'l Conf. on Mobile Computing and Networking. New York: ACM Press, 2000. 56-67.
- [2] Heidemann J, Silva F, Estrin D. Matching data dissemination algorithms to application requirements. In: Akyildiz IF, ed. Proc. of the 1st Int'l Conf. on Embedded Networked Sensor Systems. New York: ACM Press, 2003. 218-229.
- [3] Luo YH, Cheng J, Lu S, Zhang L. A two-tier data dissemination model for large-scale wireless sensor networks. In: Akyildiz IF, ed. Proc. of the 8th Annual Int'l Conf. on Mobile Computing and Networking. New York: ACM Press, 2002. 148-159.
- [4] Braginsky D, Estrin D. Rumor routing algorithm for sensor networks. In: Raghavendra CS, ed. Proc. of the 1st Workshop on Sensor Networks and Applications. New York: ACM Press, 2002. 1-12.
- [5] Zhang WH, Cao GH, Porta TL. Dynamic proxy tree-based data dissemination schemes for wireless sensor networks. In: Agrawal DP, ed. Proc. of the IEEE Int'l Conf. on Mobile Ad-hoc and Sensor Systems. Seattle: IEEE CS Press, 2004. 21-30.
- [6] Liu X, Huang Q, Zhang Y. Combs, needles, haystacks: Balancing push and pull for discovery in large scale sensor networks. In: Stankovic JA, ed. Proc. of the 2nd ACM Conf. on Embedded Networked Sensor Systems. New York: ACM Press, 2004. 122-133.
- [7] Wang GL, Cao GH, La Porta T, Zhang WS. Sensor relocation in mobile sensor networks. In: Znati T, ed. Proc. of the 24th Annual Joint Conf. of the IEEE Computer and Communications Societies. Seattle: IEEE CS Press, 2005. 2302-2312.
- [8] Shakkottai S. Asymptotics of query strategies over a sensor network. IEEE Trans. on Automatic Control, 2005,50(5):594-606.
- [9] Ren FY, Huang HN, Lin C. Wireless sensor networks. Journal of Software, 2003,14(2):1148-1157 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1148.htm>
- [10] Akyildiz IF, Su W, Sankarasubramaniam Y, Cayiri E. Wireless sensor networks: A survey. Computer Networks, 2002,38(4): 393-422.
- [11] 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>
- [12] Ye F, Zhong G, Lu S, Zhang L. Gradient broadcast: A robust data delivery protocol for large scale sensor networks. ACM Wireless Networks, 2005,11(3):285-298.
- [13] Bokareva BT, Bulusu N, Jha S. A performance comparison of data dissemination protocols for wireless sensor networks. In: Shah RR, ed. Proc. of the IEEE Globecom. Seattle: IEEE CS Press, 2004. 35-89.
- [14] Lin L, Wu HL. Artery: A data-centric architecture for wireless sensor networks. In: Ferrer J, ed. Proc. of the Int'l Conf. on Computer, Communication and Control Technologies. 2003. 123-128.

- [15] Kim HS, Abdelzaher TF, Kwon WH. Minimum-Energy asynchronous dissemination to mobile sinks in wireless sensor networks. In: Akyildiz IF, ed. Proc. of the 1st Int'l Conf. on Embedded Networked Sensor Systems. New York: ACM Press, 2003. 193–204.
- [16] Zhang W, Cao G, La Porta T. Data dissemination with ring-based index for sensor networks. In: Kim K, ed. Proc. of the 11th IEEE Int'l Conf. on Network Protocols. Seattle: IEEE CS Press, 2003. 305–314.
- [17] Akkaya K, Younis M. Energy-Aware routing to a mobile gateway in wireless sensor networks. In: Shah RR, ed. Proc. of the IEEE Globecom. Seattle: IEEE CS Press, 2004. 16–21.
- [18] Chiasserini CF, Garetto M. Modeling the performance of wireless sensor networks. In: Li VOK, ed. Proc. of the 24th Annual Joint Conf. of the IEEE Computer and Communications Societies. Seattle: IEEE CS Press, 2004. 220–231.
- [19] Luo J, Hubaux JP. Joint mobility and routing for lifetime elongation in wireless sensor networks. In: Znati T, ed. Proc. of the 24th Annual Conf. of the IEEE Communications Societies. Seattle: IEEE CS Press, 2005. 819–830.
- [20] Bonnet P, Gehrke J, Seshadri P. Querying the physical world. IEEE Personal Communications, 2000,7(5):10–15.
- [21] Karp B, Kung HT. GPSR: Greedy perimeter stateless routing for wireless networks. In: Pickholtz R, ed. Proc. of the 6th Annual Int'l Conf. on Mobile Computing and Networking. New York: ACM Press, 2000. 243–254.
- [22] Savvides A, Han CC, Strivastava MB. Dynamic fine-grained localization in ad-hoc networks of sensors. In: Rose C, ed. Proc. of the Annual Int'l Conf. on Mobile Computing and Networking. New York: ACM Press, 2001.
- [23] Bulusu N, Heidemann J, Estrin D. GPS-Less low cost outdoor localization for very small devices. IEEE Personal Communications Magazine, 2000,7(5):28–34.
- [24] Zhang Y, Fromherz MPJ. Message-Initiated constraint-based routing for wireless ad-hoc sensor networks. In: Bove VM, ed. Proc. of the IEEE Consumer Communications and Networking Conf. Seattle: IEEE CS Press, 2004.
- [25] Caruso A, Urpi A, Chessa S, De S. GPS free coordinate assignment and routing in wireless sensor networks. In: Johnson DB, ed. Proc. of the 24th Conf. of the IEEE Communication Society. Seattle: IEEE CS Press, 2005. 150–160.

#### 附中文参考文献:

- [9] 任丰原,黄海宁,林闯.无线传感器网络.软件学报,2003,14(2):1148–1157. <http://www.jos.org.cn/1000-9825/14/1148.htm>
- [11] 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>



石高涛(1978 - ),男,河南鹿邑人,博士生,主要研究领域为无线传感器网络,网络嵌入式系统.



廖明宏(1966 - ),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为新型操作系统,实时与嵌入式计算,对等计算.