

无线传感器网络随机性与结构化折衷的双盲数据发现方法*

杨金峰, 石高涛⁺, 赵增华

(天津大学 计算机科学与技术学院, 天津 300072)

Trade Randomness for Structure in Double Blind Data Discovery in Wireless Sensor Networks

YANG Jin-Feng, SHI Gao-Tao⁺, ZHAO Zeng-Hua

(School of Computer Science and Technology, Tianjin University, Tianjin 300072, China)

+ Corresponding author: E-mail: shgt@tju.edu.cn

Yang JF, Shi GT, Zhao ZH. Trade randomness for structure in double blind data discovery in wireless sensor networks. Journal of Software, 2011, 22(Suppl. (1)): 93-101. <http://www.jos.org.cn/1000-9825/11010.htm>

Abstract: Enabling anytime and anywhere random data access for the mobile sinks is an important solution to the problem in the large scale wireless sensor networks. However, the data discovery in large scale sensor networks faces a non-trivial challenge because of the double blindness between the data and the sinks. The existing schemes are not able to provide the high success rate and consume less energy at the same time. In this paper, the tradeoff between randomness and structure is explored. Based on random line walk scheme, three scalable and efficient data discovery methods are presented to solve the double-blindness. These methods utilize simple geometric property to construct the path of data cache and query and can work well without location. The extensive simulations, using ns-2 and theoretical analysis, have proven the efficient trade off between randomness and structure.

Key words: wireless sensor networks; double blind data discovery; randomness and structure; geometric probability

摘要: 为移动的 Sink 提供随时随地的随意数据获取是无线传感器网络在普适环境中的重要问题。然而,由于 Sink 与数据源的位置随机性分布特点,导致 Sink 和数据源之间存在着双盲性。在数据发现中引入随机性和结构化相折衷的思想,使用随机直线行走的消息转发模式;以平面上线与线的几何性质作为指导构建数据的缓存和查询消息的转发路径,给出 3 种不依赖节点地理位置信息的双盲数据发现方法,并对其性能进行理论分析。通过在 ns-2 上的大量仿真和理论分析表明,在随机性与结构化之间的折衷能够在不引入较多能耗的同时提供较高的数据发现成功率。

关键词: 无线传感器网络;双盲数据发现;随机性与结构化;几何概率

作为传感器网络的一个重要功能,数据获取长期以来都是人们关注的热点问题。然而,这些研究工作主要集中在只有单个静止 Sink 情况下的传感器网络,如文献[1,2]。当前越来越多的应用需要传感器网络能够满足具有大量位置随意变动的 Sink 的数据获取,比如在灾难营救或者单兵作战中,每个营救人员或者士兵都是一个 Sink,这些 Sink 在传感器网络覆盖区域内随处游走,并且会随时随地发出查询获取感兴趣的数据。我们称这类传感器网络为多移动 Sink 传感器网络 MMS(mobile multi-sink sensor networks)。由于 Sink 无法预先知道探测到数据的位置,同时,数据探测节点也无法预先掌握 Sink 当前的位置,从而数据源和 Sink 之间都无法预先知道对方的信

* 基金项目: 国家自然科学基金(60903193, 61072063); 教育部博士点新教师基金(200800561042)

收稿时间: 2011-05-02; 定稿时间: 2011-07-29

息.我们称之为双盲数据发现问题.

针对这一问题,一种最简单方法就是基于 flooding 的查询洪泛,Sink 将查询 flood 到整个网络,每个接收到查询的节点判断是否有满足查询的数据^[3,4].这种方法查询成功率高,但引入较大的消息负载.

为了降低 flooding 的代价,文献[5]提出了 rumor routing 的模式,使数据和查询进行随机转发.随机化的方法降低了消息负载,但是查询成功率无法保证.文献[6,7]通过有目的的控制数据和查询的缓存或转发轨迹,通过将控制数据(查询)沿水平(垂直)方向缓存(转发),提高了查询成功率.GHTs^[8]的方法是为数据类型建立映射,将具有相同关键字的数据根据 Hash 结果存储到相应的区域内.这些方法借鉴了结构化的思想,利用其可控的特点提高查询成功率.但结构化的方法在有效控制数据存储和转发方向的同时又需要依赖精确的位置信息,从而引入额外的负载.

由此可见,如何在消耗较少能源的同时提高查询成功率是 MMS 型传感器网络中的难题.本文针对这一问题,探讨随机性和结构化相折衷的机制在双盲查询中应用.提出了随机直线行走的消息转发模式,结合了随机性和结构化各自的优点.进而,在此基础上,利用平面几何规律,探讨了 3 种可用于双盲数据发现的方法.与已有的工作相比,本文的方法可在不需要节点地理位置信息的同时依然保持结构化的特点.

本文第 1 节对与本文相关的工作进行简要介绍.在第 2 节对本文研究的问题进行描述.在第 3 节给出 Random Line Walk 的基本思想,并对其进行理论分析.第 4 节将根据平面上线与线的关系探讨 3 种数据发现方法.第 5 节给出不同数据发现方法的仿真评价结果,最后是本文的结论和进一步的工作.

1 相关工作

传统 Sink 静止的无线传感器网络中为了查询到感兴趣的数据一般采用洪泛的策略^[1,2],但由于洪泛会因为网络风暴而产生较大的网络负载,同时洪泛会导致信号的冲突,因此网络性能将急剧下降.

文献[5]提出了谣言路由来查询信息,该方法初步运用了平面上的线与线的性质来控制数据的转发.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.

结构化的数据发现方式则是受到分布式网络和 Peer-to-peer 系统的启发,通过 Hash 表或者索引等机制来对数据和查询进行映射存储.在结构化的工作中,GHT^[8]是最典型的基于 Hash 表的数据查询和收集方法,它将数据按照关键字和地理位置一一对应.当 Sink 查询信息时,根据相同的映射规则得到特定类型事件的存储位置,然后使用基于地理位置的路由协议将查询转发到存储点上,从而获取事件信息.

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.但是,该工作的性能受到梳子柄长短的制约.文献[7]提出了一个两层的数据发布协议,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.该方法引入了维护代价,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[11]根据平面上几何位置和球体的映射关系提出了一个 Double Rulings 数据查询和收集模式,数据缓存位置使用 GHT 进行映射到位置 L 上,并根据位置映射关系沿着特定路线 C 转移到 L ,并在在 C 上也进行缓存数据.当查询进行转发时,过程与数据缓存类似,其转发路线为 C' ,使用特定的映射机制可以保证 C 和 C' 肯定相交,进而发现感兴趣的数据.文献[9]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,进而查询消息在要道节点内进行 flooding.

文献[13]提出了一个直径-弦模式 DCS(diameter-chord scheme),利用圆直径和弦的关系进行转发查询和缓存事件信息,每个事件信息水平缓存,而查询沿直径方向垂直转发,这样两者肯定相交,从而成功查找事件信息.GLIDER^[12]协议基于 Landmark 的思想将网络区域划分成很多具有以 Landmark 为中心的 Voronoi 分片,感知到的数据采用 GHT 映射到某一个 Landmark,并在转发过程中缓存到经过的 Voronoi 边上.查询消息使用同样的 GHT 映射到 Landmark,并在转发查询时察看经过的 Voronoi 边是否有满足条件的数据存在.但这一方法需要具

有全局网络信息.

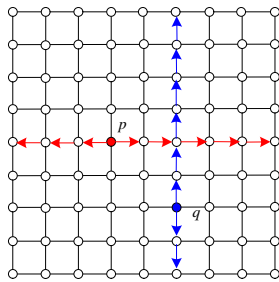


Fig.1 Quorum system
图 1 Quorum 系统

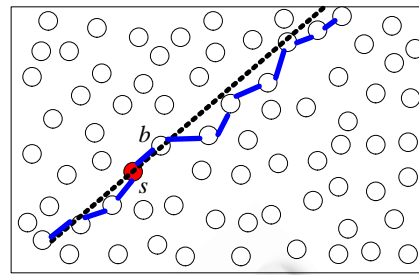


Fig.2 Random line walk
图 2 随机直线行走

2 网络模型与问题描述

本文假定有 n 个无线传感器节点 $V=\{v_1, v_2, v_3, \dots, v_n\}$ 均匀地分布在某一区域内,每个传感器节点的 *radio* 都是全向天线,节点的工作半径为 r .假定在该网络区域内存在 m 个 Sink 节点 $S=\{S_1, S_2, \dots, S_m\}$ 分别按照某一移动模型 M 移动,移动模型无法预先获得,因此,每个 Sink 的位置无法预先得到.

节点 u 和 w 之间的距离表示为 d_{uw} ,如果 $d_{uw} < r$,则 $uw \in E$.对任一节点 v ,距离它为 d 的邻居节点集表示为 $N_v(d)$,其中 $d \leq r$.同时假定节点不具有自身的位置信息,只能通过诸如信号强度的估计来获得邻居之间的相对距离.

在 Sink 移动环境下的传感器网络中,Sink 的位置不固定,因此,每个传感器节点无法预先建立到达 Sink 的路径.即使能够预先建立路径,为节点创建到达每个 Sink 的路径也会引入较高维护代价.同时,Sink 对自己感兴趣的数据分布位置可能也一无所知,因此,它无法使用诸如区域查询的方法来获取数据.这样在数据源和 Sink 之间就出现了类似于现实生活中的双盲性.针对以上介绍,本文对双盲数据获取问题做以下定义:

定义 1(双盲数据发现问题). 给定一个无线传感器网络 $G=(V,E)$ 和 m 个按照某一特定移动模型 M 移动的 Sink 节点 S ,其中任一 Sink 节点 S_p 感兴趣的某一数据 D_i 随机地由某一个传感器节点 v_j 产生,设所有拥有数据 D_i 副本的节点集为 C , S_p 发出的查询消息 Q 经过的节点集为 Q ,双盲数据发现问题就是构造 C 和 Q 使 $\emptyset \neq C \cap Q$.

对双盲数据发现问题,存在一个平凡数据发现方法,即 Sink 通过 flooding 查询消息的方法获取感兴趣的数据.但是其查询消息负载为 $O(n^2)$.直观上来看,在一个矩形平面内,如果将数据(查询)消息沿水平方向而查询(数据)消息沿垂直方向进行转发,理论上数据消息和查询消息肯定会相遇.如图 1 所示.这一思想称为 Quorum System.然而,可以清楚地看到,这一方法需要节点具有精确的地理位置信息,只有每个节点能够得到自身的地理位置信息(仅有相对位置亦不能工作),上述模式才能有效工作.文献[6,7,13]都是基于该思想的扩展,基本思路是通过控制消息按照特定的几何路线进行缓存和转发,通过设计不同的几何曲线进而能提供不同的相交概率.这些方法也是属于结构化的思想,在一定程度上能够保证数据查询的成功率,但都需要依赖节点的地理位置信息.

3 随机直线行走消息转发

基于 Random Walk 的消息转发主要的优点在于方法简单,代价小.但其随机不可控性导致性能具有严重的不确定性.本节给出一个随机直线行走 RLW(random line walk)的消息转发模式^[21],在随机性中引入结构化的特点,以便增加其可控能力.该模式由文献[21]首次提出,为行文完整性,本节对此思想和特性进行详细介绍.

随机直线行走的基本思想如图 2 所示,消息源 b 在转发时会随机选择一个邻居节点 s 作为转发节点,节点 s 和 b 将确定一个方向,称 s 和 b 的连线为基准线,该方向将作为其后消息转发的参考方向,尽量选择沿该方向的邻居节点进行转发.与 Random Walk 不同,RLW 除了第一次转发消息具有随机性之外,其后的转发节点选择都受到方向的约束.与 Random Walk 相比,由于消息转发过程尽量向远处拓展,因此,消息可以传播到更远的距离.

尽管 RLW 考虑消息转发的方向性,这并不意味着 RLW 需要依赖节点的位置信息.事实上,只要节点能够获得邻居节点之间的相对位置即可,而这可以通过建立局部坐标系得到.文献[18]给出了一个无需节点精确位置建立局部坐标系的方法.因此,该方法除了具有系统简单的优点之外,还具有相对较好的可控性.由于 RLW 的消息转发过程将覆盖的几何区域进行了扩展,因此通过组合消息和发现的过程将有更大的发现成功率.

为分析消息转发偏移情况,我们定义:

定义 2(角偏差和幅偏差). 假定转发基准线 L 已确定,对转发路径上任意节点 v 其下一跳为 w .假定每一个节点都有一个极坐标系,其极轴与 L 平行.令 w 在节点 v 极坐标中的极角为 θ_{vw}^L 称 $\pi - \theta_{vw}^L$ 为角偏差 AND(Angle Deflection),表示为 Δ_{vw}^L ,称 $|vw| \times \sin(\Delta_{vw}^L)$ 为 w 相对于 v 幅偏差 AMD(amplitude deflection),记为 h_v^w .

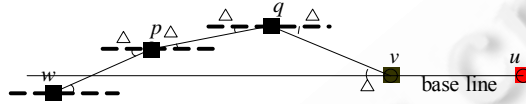


Fig.3 Illustration of AND and AMD

图3 AND和AMD示意图

图 3 给出了示意图.每个节点上的两个 Δ 分别代表自身的角偏差和下一个节点的角偏差.面朝基准线的负方向,当节点的角偏差小于 0 时,表明该节点转向左侧,反之转向右侧.幅偏差表明该节点相对于上一个节点的上升高度,当 $h_v^w < 0$ 表示节点正在下沉,否则表示正在上升.因此,偏差的绝对值 $|\Delta_{vw}^L|$ 和 $|h_v^w|$ 代表偏差的程度,正负性表示偏差的方向.

根据以上定义,可以分析经过若干跳之后的转发节点已经偏离了最初预定的基准线的程度.为了简单起见,假定每一跳的 AMD 为绝对值 1 的常数,在节点密集且均匀分布的情况下,整个转发过程可以模型化为 1 和-1 的随机序列.定义变量 X_i :

$$X_i = \begin{cases} -1, & \text{if AMD is negative} \\ 1, & \text{if AMD is positive} \end{cases} \quad i=1,2,\dots,n \quad (1)$$

此处 n 表示路径的长度.

为计算 $\sum X_i$ 的概率分布,令 S_i 表示经过第 i 步之后的 AMD,那么

$$S_n = X_1 + X_2 + \dots + X_n, S_0 = 0 \quad (2)$$

对幅偏差,我们有以下定理:

定理 1. 经过 n 步转发后距离基准线的 AMD 期望值是 $2^{-n} \sum_{h=1}^n \binom{n}{n+h} h$.

证明:首先,根据文献[17],经过 n 步转发之后整个 AMD 也就是 S_n 是 h 的概率为

$$P\{S_n = h\} = \binom{n}{n+h} h \quad (3)$$

这里 h 是从 $-n$ 到 n 的整数.由于对称性,根据期望值的定义,AMD 的期望值为 $\sum_{h=-n}^n |h| P(S_n = h)/2$.将式(3)代入可

得,可得经过 n 步转发之后,平均偏移值是 $2^{-n} \sum_{h=1}^n \binom{n}{n+h} h$.

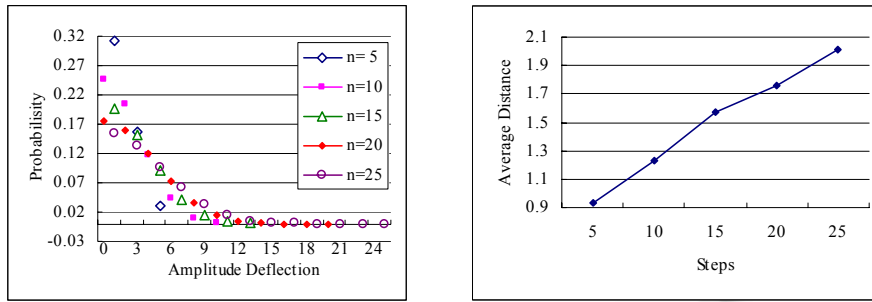


Fig.4 Probability Distribution of AMD and the average distance from the base line

图4 AMD的概率分布和平均偏移距离

一般来说,在 n 固定时, P 会随着 $|h|$ 的增加而降低.图 4 画出了当 h 和 n 变化时 P 的分布数值结果.由于 h 的对称性,图 4 只画出了当 h 是正整数时的情况.从图中可以看出,绝大部分的偏差都集中在一个很小的距离之内,平均偏差的范围更是有限,从右图可以看到,转发经过 10 跳平均偏移才达到 1.2 倍的传输半径.

4 随机性与结构化折衷的双盲发现方法

双盲数据发现的关键在于如何构造数据缓存路径和查询消息转发路径.本节将平面上的 3 种简单几何规律应用于双盲数据发现中,给出了 3 种不同的双盲数据发现方法,并对其性能作简要分析.此 3 种方法的在构造数据和消息路径时都采用了随机性与结构化相折衷的思想,使用随机直线行走的消息转发.

4.1 Line Routing

Line Routing 是基于随机直线行走模式最直接的应用,也是 Rumor routing 的一种变形.其基本思想与 Rumor Routing 类似,数据源节点和查询节点分别创建相应的数据缓冲节点集和查询消息转发节点集.但区别在于节点集的选择不同.在 Rumor Routing 中,每个节点的选择完全随机,而在 Line Routing 中,节点集的创建则是按照随机直线行走的模式进行.

Line Routing 具体的工作过程如下:

(1) 数据缓存阶段:对每一个监测到事件的节点 s ,它将随机选择一个邻居节点 u 作为一个方向上的第 1 跳节点,并以 \overline{su} 为基准选择节点 v 使 $|\pi - \angle u, s, v|$ 最小化,作为另一个方向上的第 1 跳的节点.之后,每个被选中的节点以上一跳节点作为参考,继续选择满足角度偏差最小化的下一跳节点.同时,每个被选中的转发节点将数据进行缓存.

(2) 数据发送阶段:每当 Sink 进行数据发现时,选择其附近的一个点 q 作为查询接入点,与节点 s 相同,节点 q 也会随机选择满足角度关系的两个邻居节点作为第一跳节点将查询消息向两个方向进行转发,此后的转发节点选择过程与数据缓存过程类似,不同的是每个转发节点都会将查询消息广播给所有邻居节点,询问是否缓存满足条件的数据,并且之后被选中的转发节点才会转发查询消息.如果邻居节点恰好存有数据,表明数据查中,消息即停止继续转发.

为了分析 Line Routing 的成功率,可以将其模型化为圆区域内的 Random Angle Model^[14].对从单位圆形内某点以随机方向引出的弦的内接角的概率密度函数是 $f(\theta) = (4\sin^2 \theta) / \pi$ ^[14],其中 θ 是该弦的内接角.根据这一结果,文献[15]给出了当其中一条弦长度为 l 时,从某点引出随机直线与之相交的概率 $p(l) = 2(\pi^2 l - \pi^2 l^2 + \sin^2(\pi l)) / \pi^2$.

图 5 画出了当弦的长度变化时发现成功率的变化情况.显然,当其中一条弦和直径重合时,此时的发现成功率达到最大值.但是,尽管如此,理论上其最大成功率也只有 70%.

另外, Line Routing 的负载也具有上界,理论上,消息转发路径和数据缓存路径之和不超过网络直径的 2 倍.

事实上,文献[19]证明在边长为 l 的正方向内随机引出的一条弦的平均长度为 $0.9464l$. 对于一个半径为 R 的圆上的弦来说,其平均长度可以根据下图 6 来计算,当两点固定时,图上的弦长为 $\sqrt{(R-R\cos\theta)^2 + (R\sin\theta)^2} = R\sqrt{2-2\cos\theta}$, 因此,平均长度可计算为 $\frac{1}{\pi} \int_0^\pi R \sqrt{2-2\cos\theta} d\theta = 4R/\pi$.

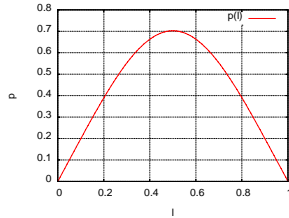


Fig.5 Discovery Probability

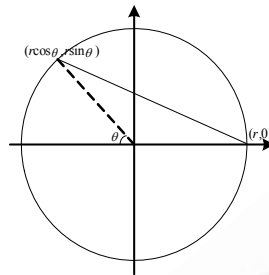


Fig.6 Average of a chord in a circle

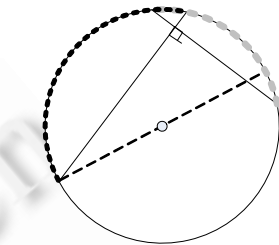


Fig.7 Effective length of orthogonal chord

图 5 发现概率随弦的长度变化情况

图 6 计算圆的弦的平均长度

图 7 正交弦的有效长度

4.2 X Routing

Line Routing 虽然在一定提高数据发现成功率,但是直观上,如果继续增加缓存和查询消息的路线数目将进一步提高发现性能.然而,如果增加过多的线路,将增加系统消息负载.X Routing 在构造数据缓存和查询转发路径时分别沿两条线路进行,虽然在一定程度上增加了节点负载,但能够提高数据的查询成功率.文献[20]研究了使用多条缓存和消息转发路径的方法,虽然增加转发路径在一定程度上提高发现概率,但是以提高消息负载为代价的.该工作并没有对这一问题进行研究.

与 Line Routing 不同,X Routing 在数据源节点或者查询点转发消息时,将按照 Line Routing 的消息转发方式选择两个方向进行转发,从而形成两条呈一定夹角的轨迹 L_1 和 L_2 ,并且这两个方向的夹角在 $[\alpha, \beta]$ 内等概率选取. α 和 β 的确定需要仔细安排.如果 α 太小或者 β 太大, L_1 和 L_2 将近乎成一条线路.本文在仿真中,将 α 设置为 $\pi/6$, β 设置为 $5\pi/6$.

4.3 Cross Routing

X Routing 通过增加额外的消息行走轨迹来增加发现成功几率,但如果夹角选择不当,性能将会受到影响.直观上来看,如果夹角接近 $\pi/2$ 度,那么两条相交的线路占据的有效区域相对其他角度要大,特别是当夹角为 $\pi/2$ 时,有效区域达到最大.因此,这启发我们设计 Cross Routing 模式,该模式与 X Routing 唯一不同的是两条转发轨迹的夹角不是随机变动,而是默认为 $\pi/2$.

事实上,在一个圆形区域内,有以下结论保证 Cross Routing 能够以极高的概率查询到数据.

结论.在圆内任两点随机引出两对正交的直线,它们相交于该圆形区域内的概率为 1.

为了方便论述该结论,首先给出如下定义:

定义 3(弦的有效长度). 对于位于圆形区域内的任意一条弦 l ,它将整个圆的边界分成长度为 δ_1 和 δ_2 的两部分,定义该弦的有效长度为 $l_j = \min(\delta_1, \delta_2)$.

同样可以定义多条弦(称之为弦集)的有效长度.

定义 4(弦集的有效长度). 对于位于圆形区域内的任意 n 条弦 l_1, l_2, \dots, l_n ,其有效长度分别为 $l_{j1}, l_{j2}, \dots, l_{jn}$,其对应的弧分别为 $\partial_1, \partial_2, \dots, \partial_n$,其中 $|\partial_i| = l_i, i=1 \dots n$.定义操作 $l_i + l_j = |\partial_i \cup \partial_j|$,因此 n 个弦组成的弦集 Λ 的有效长度为 $l_\Lambda = |\partial_1 \cup \partial_2 \cup \dots \cup \partial_n|$.

在以上定义指出上,我们给出有关弦集相交的结论.

定理 2. 对任意两个弦集 Λ_1, Λ_2 ,它们的有效长度分别记为 l_{Λ_1} 和 l_{Λ_2} ,如果 $l_{\Lambda_1} + l_{\Lambda_2} > 2\pi R$,则这两个弦集必

相交,其中 R 是圆形区域的半径.

证明:根据弦集的有效长度的定义,可推出,对任意两个弦集 ℓ_{A1} 和 ℓ_{A2} ,它们之和 $\ell_{A1} + \ell_{A2} = \left| \bigcup_{\partial_i \in A1} \partial_i \cup \bigcup_{\partial_j \in A2} \partial_j \right| = \left| \bigcup_{\partial_i \in A1 \cup A2} \partial_i \right|$,当 $\ell_{A1} + \ell_{A2} > 2\pi R$ 时,这表明这两个弦集的有效长度对应的弧已完全把圆周覆盖.只有这两个弦集相交,上述事实才能发生.因此定理得证. \square

根据以上初步结论,可以看出,在圆内随机选择一点引出两条正交弦,它将圆周划分成 4 份,假设该圆周的周长为 x ,则其中任一圆弧的长度都不大于 $1/2x$,但该正交弦的有效长度则大于 $1/2x$.图 7 给出了这一示意图.这样,当从另一随机点引入另外一对正交弦时,由于其有效长度不小于 $1/2x$,因此,这意味着这两对正交弦的有效长度之和大于 x .因此,根据定理 2 可知,这两对对正交弦必定相交,从而结论得证.

文献[16]使用数字模拟的方法得出该概率为 99.7%.这一几何规律在文献[16]中被首次使用,作者以此为依据利用定向天线的有向性设计了一个 Mesh 网络中的路由协议.

5 性能评价与比较

本小节对上文所给出的 3 种数据发现方法通过仿真的方法与一种纯粹随机的方法 Rumor Routing 进行比较,仿真工具选用 NS-2,分别在 $100 \times 100m^2$ 和 $300 \times 300m^2$ 的区域内变化节点个数进行性能比较,节点的通讯范围设定为 30m,每个数据点都是 100 次运行结果的平均.

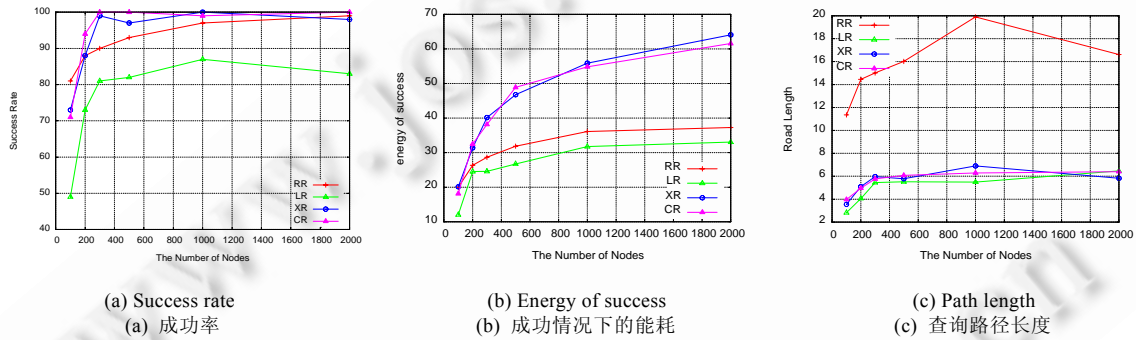


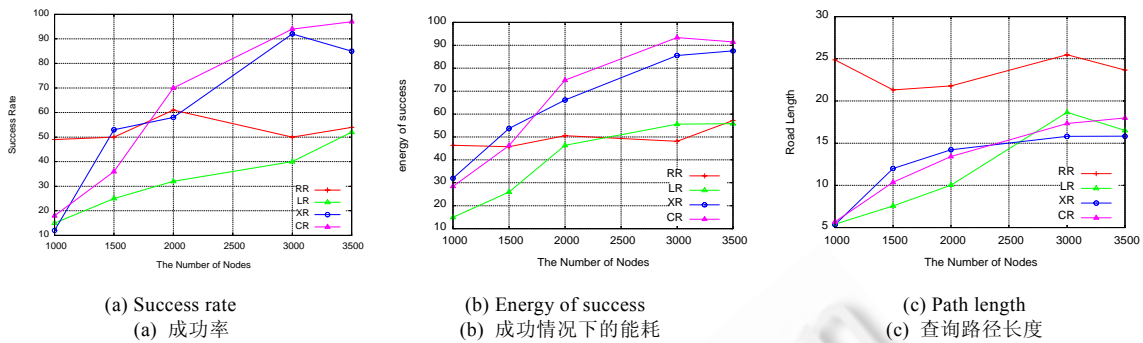
Fig.8 Performance when the area size is $100 \times 100m^2$

图 8 网络大小为 $100 \times 100m^2$ 的性能情况

仿真主要评价发现成功率、能源消耗和路径长度.其中能源消耗以跳数来表示,指在数据发现过程中所有缓存和查询转发涉及到的所有跳数;路径长度是指从查询点发出的查询消息在遇到缓存数据之前经过的跳数,该性能参数可以反映查询的延迟.

图 8 给出了当网络区域大小是 $100 \times 100m^2$ 时的性能比较结果.从图 8(a)中可以看出,所有方法的成功率基本都会随着节点密度的增加而增加,而 Line Routing 的成功率是最低的,其他 3 种方法的成功率相差不多.其中的原因在于消息沿着直线行走虽然拓长了传播空间,但是也限制了传播宽度的范围.而 Rumor Routing 虽然完全随机,但是由于其传播宽度的到一定的扩张,所以相遇的可能性较 Line Routing 高,特别是在网络区域较小的情况下.X Routing 和 Cross Routing 成功率较高的原因在于其同时在宽度和长度上都拓展了传播线路.但与此同时,其能耗也比 Line Routing 和 Rumor Routing 较高,图 8(b)验证了这一点.由于 X routing 和 Cross Routing 相对其他两种方法需要多出一倍的转发路线,因此能耗相对要高.

图 8(c)从查询延迟的角度考察了 4 种查询方法的性能,该仿真统计了所有成功情况下的查询消息转发跳数.从中可以看出,Rumor Routing 的方法在查询延迟上性能最差,这也是随机性转发的重要缺陷,而本文所提的 3 种方法在延迟方面差别不大.综合图 8(a)和图 8(c)可以看出,Line Routing 虽然成功率不高,但是一旦某个查询成功,其查询延迟也相对较小.

Fig.9 Performance when the area size is $300 \times 300 \text{m}^2$ 图9 网络大小为 $300 \times 300 \text{m}^2$ 的性能情况

进而,本文在更大的网络区域内对所有方法进行比较.图9给出了网络大小为 $300 \times 300 \text{m}^2$ 的结果.从图9(a)可以看出,4种方法的相对性能趋势与 $100 \times 100 \text{m}^2$ 条件下的趋势类似,Cross Routing 与 X Routing 成功率相对依然是最好的.但由于区域面积的增加,在网络密度较小时,消息在转发过程中会遇到没有转发节点的情况,从而相对于小区域内的成功率有所下降.但当网络密度增加时,Cross Routing 的成功率依然能够达到90%以上.

图9(b)同样对这一场景下的能耗进行比较,趋势与小区域场景一致.图9(c)给出了查询路径的长度比较情况,Rumor Routing 的路径长度依然是最长的,而本文所给出的3种方法基本上依然维持在一个相对比较稳定的范围内.这也说明通过随机性与结构化相结合,能够提供更加稳定的性能.

Rumor Routing 由于采用完全随机的方式,因此引入最少的负载.本文所给出的3种方法在负载方面与Rumor Routing 接近,但是在成功率方面,X Routing 和 Cross Routing 相对较好,特别是在网络规模较大时.综合来看,Cross Routing 能够在引入较少能耗的情况下提供较高的成功率.同时,本文探讨的3种方法不需要信息严格按照水平或垂直方向转发,因此可在无节点地理位置情况下工作.

6 结论

在移动 Sink 的传感器网络中,由于 Sink 的位置和数据源的位置之间存在的双盲性,给数据发现带来挑战.本文针对现有方法无法兼顾发现成功率和能耗的缺点,提出一个集成随机性与结构化的无需节点精确位置信息的信息转发模式.基于该模式,借助平面内线与线的关系,本文探讨了3种不同的数据发现方法.本文使用仿真的方法,在不同场景下对其性能进行评价和比较,结果表明,通过在随机化与结构化之间进行折衷可在提供较高成功率的同时而不引入过多的负载.

References:

- [1] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In: Johnson DB, ed. Proc. of the ACM MobiCom. Boston: ACM Press, 2000. 56–67.
- [2] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-Efficient communication protocol for wireless microsensor networks. In: Proc. of the 33rd Annual Hawaii Int'l Conf. on System Sciences. Maui: IEEE Computer Society, 2000. 3005–3014.
- [3] Chang N, Liu M. Revisiting the TTL-based controlled flooding search: Optimality and randomization. In: Haas ZJ, ed. Proc. of the 10th Annual Int'l Conf. on Mobile Computing and Networks (MobiCom 2004). New York: ACM Press, 2004. 85–99.
- [4] Krishnamachari B, Ahn J. Optimizing data replication for expanding ring-based queries in wireless sensor networks. In: Noubir G, ed. Proc. of the Int'l Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 2006). Washington: IEEE CS Press, 2006. 361–370.
- [5] Braginsky D, Estrin D. Rumor routing algorithm for sensor networks. In: Proc. of the 1st workshop on sensor networks and applications. Atlanta: ACM Press, 2002. 22–31.

- [6] Liu X, Huang Q, Zhang Y. Combs, needles, haystacks: Balancing push and pull for discovery in large-scale sensor networks. In Proceedings of the 2nd international conference on embedded networked sensor systems. ACM Press, 2004. 122–133.
- [7] 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.
- [8] 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 Networks and Applications (MONET), 2003,8(4):427–442.
- [9] 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.
- [10] Kim SK, Abdelzaher TF, Kwon WH. Minimum-Energy asynchronous dissemination to mobile sinks in wireless sensor networks In: Proc. of the 1st Int'l Conf. on Embedded Networked Sensor Systems. 2003. 193–204.
- [11] Sarkar R, Zhu X, Gao J. Double rulings for information brokerage in sensor networks. In: Petrioli C, Ramjee R, eds. Proc. of the 12th Int'l Annual Conf. on Mobile Computing and Networking. 2006. 286–297.
- [12] 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.
- [13] Shi GT, Liao MH. An energy-efficient data dissemination protocol in large-scale sensor networks. Journal of Software, 2006,17(8):1785–1795 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/1785.htm> [doi:10.1360/jos 171785]
- [14] Solomon H. Geometric probability. In: Proc. of the CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics. 1987.
- [15] Zhou DP, Gao J. Opportunistic processing and query of motion trajectories in wireless sensor networks. In: Proc. of the 28th Annual IEEE Conf. on Computer Communications (INFOCOM 2009). 2009. 1197–1205.
- [16] Cheng BN, Yuksel M, Kalyanaraman S. Orthogonal rendezvous routing protocol for wireless mesh networks. IEEE/ACM Trans. on Networking, 2009,17(2):542–555.
- [17] Feller W. An Introduction to Probability Theory and its Applications. Vol I, 3ed., Wiley, 1968.
- [18] Bejerano Y. Simple and efficient k -coverage verification without location information. In: Proc. of the 27th Conf. on Computer Communications. IEEE (INFOCOM 2008), 2008. 291–295.
- [19] Kuchel PW, Vaughan RJ. Average lengths of chords in a square. Mathematics Magazine, 1981,54(5):261–269.
- [20] Shokrzadeh H, Haghighat AT, Nayebe A. New routing framework base on rumor routing in wireless sensor networks. Computer Communications, 2009,32(1):86–93.
- [21] Shi GT, Zheng J, Yang JF, Zhao ZH. Double cross: A double-blind data discovery scheme for large-scale wireless sensor networks. In: Proc. of the 2011 IEEE Int'l Conf. on Communications (ICC). 2011.

附中文参考文献:

- [13] 石高涛,廖明宏.一种大规模传感器网络节能数据发布协议.软件学报,2006,17(8):1785–1795.<http://www.jos.org.cn/1000-9825/17/1785.htm> [doi: 10.1360/jos171785]



杨金峰(1987—),男,河北衡水人,主要研究领域为无线传感器网络.



赵增华(1974—),女,博士,副教授,主要研究领域为无线网络.



石高涛(1978—),男,博士,讲师,主要研究领域为无线和传感器网络.