

无线传感器网络多频率查询的节能优化

陈颖文, 徐 明, 虞万荣
(国防科技大学计算机学院, 湖南长沙 410073)

摘 要: 在多频率查询的无线传感器网络中, 多个接收节点以不同的查询频率请求同一个数据源节点的数据. 由于查询频率的不同, 如何共享传输链路以降低能耗成为一个新的研究问题. 考虑数据序列之间的数据相关性, 本文提出了一种节能优化方法, 通过在共享链路上广播整合频率的数据序列, 再根据需要重构出与查询请求相对应的数据序列. 理论分析和模拟实验均表明, 在可容忍的平均相对误差下, 该方法能够节省能量的消耗.

关键词: 传感器网络; 数据查询; 多频率

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2008) 04-0701-08

Energy-Efficient Optimization for Multi-Rate Query in Wireless Sensor Networks

CHEN Ying-wen, XU Ming, YU Wan-rong

(School of Computer, National University of Defense Technology, Changsha, Hunan 410073, China)

Abstract: Multi-rate query in wireless sensor networks is that multiple sink nodes query the same data source with different frequencies. Because the data are requested at different frequencies, how to share the bandwidth for transmission becomes a new problem. Based on the data correlation between different data series, this paper proposes an energy-efficient optimization scheme, in which we can broadcast consolidated data series along the shared path and reconstruct the requested data series from these consolidated data series. Theoretical analysis results and simulation results indicate that we can save the energy consumption with a reasonable level of tolerance of the average relative error.

Key words: sensor networks; data query; multiple rates

1 引言

无线传感器网络是由一组在地理上广泛分布, 相互之间能够利用无线信道进行通信的传感器节点组成的, 每个节点都可以与一个或者多个其他节点进行协作, 将测量的环境参数或者收集的实时数据发送给接收节点 (sink node). 由于传感器节点往往利用不可再生的电池进行供电, 并且无线通信的能耗占据所有能耗的主体, 因此当前的许多研究主要围绕如何降低传感器节点的无线传输能耗而展开^[1], 如数据汇聚技术 (data aggregation) 和数据副本技术 (data replication) 等.

数据汇聚技术将数据的处理融入数据的传输过程, 通过减少数据的传输量来降低能耗. 对于数据汇聚而言, 汇聚节点的选取决定着数据查询的能耗, 因此主要的研究工作有: 讨论汇聚节点的最优选取策略^[2~4], 以及构造最优的汇聚路由树^[5~8].

文献[2]针对传感器节点具有不同的计算能力, 形式化定义了汇聚节点的选择问题, 并提出了近似最优选取算法, 使得传输的能耗降至最低. 文献[3]针对数据流的不对称性, 提出了一种分布式的自适应汇聚节点选取算法, 然而当节点密度较小的时候, 该算法的近似解与最优解的差距比较大. 文献[4]针对文献[3]的不足, 提出了一种不需要全局拓扑结构的解决方法, 适用于节点密度较小的情况.

文献[5]证明了, 构造最优的路由树用以支持数据汇聚等同于构造最小代价的 Steiner 树问题, 属于 NP 难问题, 而贪婪递增树 (GIT) 的构造不需要全网的拓扑结构, 具有较好的近似效果. 考虑不同数据源节点的数据相关性, 文献[6]提出了一种新的低能耗、可扩展的分布式近似算法. 文献[7]引入了能量核的概念, 进而提出一种分布式层次路由算法, 更好地降低了汇聚路由所需的传输能耗. 文献[8]结合传感器节点的位置信息, 提出了

一种基于连通支配集的定向传播模型,能很好的支持数据压缩.

现有数据汇聚技术的研究仅讨论了单一接收节点的情况,没有考虑多接收节点的查询优化.

数据副本技术通过在网络中设置一定的数据拷贝,减少数据源节点到不同接收节点之间的数据转发次数,从而达到降低能耗的目的.文献[10]讨论了一种存在多个可移动接收节点的无线传感器网络应用场景,为了降低数据查询的能耗,作者采用了数据副本,并提出了副本网关搜索算法及副本放置算法.在事件检测型传感器网络中,文献[11]讨论了最优副本数量的问题,并且推导了最小能耗的副本设置策略.

总的说来,数据副本技术考虑了多接收节点的能耗优化,但其应用仅限于多个接收节点采集相同数据的情况,仍然具有一定的局限性.

在无线传感器网络的应用系统中,存在这样的情形:即多个接收节点分别以不同的查询频率请求同一个数据源节点的数据.一种直接的查询处理方法是,将这些具有不同频率的查询请求分别进行处理,数据源节点在每个被请求的时间点上将数据发送给需要的接收节点.这种方法显然是低效的,因为不同的接收节点所需的数据产生于不同的时刻,因此数据在传输的过程中不能共享链路,从而导致传输的能耗开销较大.那么针对无线传感器网络中存在的多频率查询问题,我们该如何来降低查询的能耗呢?这一新问题既不同于当前所研究的数据汇聚问题,也不同于数据副本问题.由于发送给不同接收节点的数据构成具有不同采集频率的数据序列,考虑这些数据序列之间的数据相关性,在给定的误差范围内,可以利用一个整合后的数据序列来描述这些具有不同采集频率的数据序列.

本文的基本思想就是利用数据序列之间的数据相关性,让数据在传输的过程共享链路,发生交汇的节点广播整合后的数据序列,各链路的其他中间节点再根据需要,将整合的数据序列转换成具有对应查询频率的数据序列,从而降低传输的数据量,进而达到节能的目的.模拟实验结果表明,采用我们所提出的节能优化方法,平均可节省 50%~55% 的能耗,引入的平均相对误差控制在 5% 以内.

2 多频率查询模型

在无线传感器网络中,根据具体应用的不同,接收节点以不同的查询频率请求同一数据源节点的数据,因此形成了如图 1 所示的存在多种频率的查询系统.设 m 个接收节点 $s_i (i = 1, \dots, m)$ 均向节点 g 发出查询请求,要求节点 g 分别以 f_i 的频率将数据报告给各接收节点(不失一般性,总可以找到合适的单位时间,使得查询请

求的频率均为正整数).

例 1 假设无线传感器网络用于收集环境的温度,接收节点 s_1 需要节点 g 每隔 2 分钟报告一次温度值,接收节点 s_2 需要节点 g 每隔 3 分钟报告一次温度值,而接收节点 s_3 需要节点 g 每隔 6 分钟报告一次温度.如果所有的这些查询请求均从 0 时刻开始执行,那么该无线传感器网络应用构成一个多频率查询系统.该系统包含 3 条查询,要求传感器节点 g 分别在 2, 3, 4, 6, 8, 9, 10, 12, .. 时刻报告数据,设单位时间为 6 分钟,则有 $f_{r1} = 3, f_{r2} = 2, f_{r3} = 1$.

在多频率查询系统中,最直接的数据传输方法就是将多个查询分别进行处理,即在上述例子中,节点 g 分别以 f_{r1}, f_{r2}, f_{r3} 的频率将数据沿各自的路径发送给 s_1, s_2, s_3 节点.为了方便叙述,将这种最直接的查询处理方法记为 N-Strategy (Native Strategy). 定理 1-3 给出了 N-Strategy 所具有的一些性质.

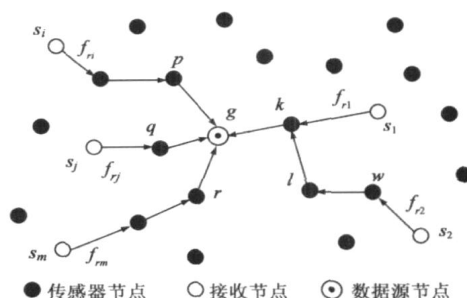


图 1 无线传感器网络中多个查询以不同频率向节点 g 请求数据

定理 1 采用 N-Strategy, 若 $f_i (i = 1, \dots, m)$ 分别表示所有接收节点的查询频率,那么数据源节点发送数据的频率上界 f_{up} 为 $\prod_{i=1}^m f_i$, 当且仅当对于各接收节点所采集的数据序列,在时间轴上没有交点.

例 2 假如例 1 中的三个接收节点分别在 0, 0.5, 1.2 时刻发出查询请求,显然各接收节点所采集的数据序列在时间轴上不存在交点,数据源节点需要在 2, 3.5, 4, 6, 6.5, 7.2, 8, 9.5, 10, 12, 12.5, 13.2, .. 时刻向各接收节点发送数据,因此在单位时间为 6 的时间间隔里,数据源节点发送数据的次数为: $1 + 2 + 3 = 6$, 发送数据的频率上界 $f_{up} = 6$.

定理 2 采用 N-Strategy, 数据源节点发送数据的频率达到下界 f_{low} , 当且仅当对于任意两个接收节点的数据序列,在时间轴上均存在交点,同时 f_{low} 可以由下式进行计算:

$$\prod_{k=1}^m [(-1)^{k-1} \gcd(\{F_j\}_{j=1}^k)] \quad (1)$$

$(F_j)_{j=1}^k \subseteq \{f_j\}_{j=1}^m$

其中 $\{F_j\}_{j=1}^k$ 表示在 m 个频率中选择 k 个频率所构成的

不同组合.

显然例 1 满足数据发送频率的下界条件,因此数据源节点在单位时间里发送数据的次数为: $1 + 2 + 3 - \gcd(1, 2) - \gcd(1, 3) - \gcd(2, 3) + \gcd(1, 2, 3) = 4$, 发送数据的频率下界 $f_{\text{low}} = 4$.

定理 3 给定 m 个频率 $f_{r1} \ f_{r2} \ \dots \ f_{rm}$, N-Strategy 发送数据的频率下界 f_{low} 满足 $f_{\text{low}} = \max\{f_{ri}\}_{i=1}^m$. 等号成立当且仅当 $\forall j \ i, f_{ri} | f_{rj}, 1 \leq i < j \leq m$, 记号“ $a | b$ ”表示 b 被 a 整除.

例 3 假设在图 1 所示的多频率查询模型中, 接收节点 s_1 每隔 8 分钟需要一次新数据, 接收节点 s_2 每隔 4 分钟需要一次新数据, 接收节点 s_3 每隔 2 分钟需要一次新数据, 三个接收节点的查询请求均从 0 时刻开始执行, 那么数据源节点 g 发送数据的时刻为 2, 4, 6, 8, 10, 12, 14, 16, ... 将单位时间设为 8 分钟, 则有 $f_{r1} = 1, f_{r2} = 2, f_{r3} = 4$. 因为 $f_{r1} | f_{r2}$ 且 $f_{r2} | f_{r3}$, 所以有 $f_{\text{low}} = \max\{f_{r1}, f_{r2}, f_{r3}\} = 4$.

一般而言, 信号采样的频率越大, 所得的信号失真度越低^[12], 可以利用不同频率数据序列之间的数据相关性, 把具有低频率的数据序列从高频数据序列中重构出来. N-Strategy 在能耗方面之所以不够高效, 就是在传输数据的时候没有考虑数据序列之间的相关性. 由于无线通信具有天然的广播特性, 在共享链路的时候, 广播一个整合的数据序列与分别发送各组数据序列相比, 能够大大减少数据的发送量, 进而降低数据的传输能耗. 如图 1 所示, 假设具有频率 f_{r2} 的数据序列在可接受的误差范围内能够从频率为 f_{r1} 的数据序列中重构出来, 那么数据源节点 g 仅需要以 f_{r1} 为频率向 k 节点发送数据. 当节点 k 以 f_{r1} 为频率向 s_1 节点转发数据的时候, 节点 l 也能获取到以 f_{r1} 为频率的数据序列, 然后节点 l 从接收到的数据序列中重构出具有频率为 f_{r2} 的数据序列, 最后将频率为 f_{r2} 的数据序列转发到接收节点 s_2 . 在上述过程中, 源节点 g 避免了分别向 s_1 和 s_2 发送不同频率的数据序列, 节省了能量的消耗. 除了源节点, 在路径发生交汇的中间节点上, 也能执行相类似的操作, 用以减少数据的发送次数. 为了与 N-Strategy 区别开来, 我们把节能优化的方法记作 E-Strategy (Efficient-Strategy). 在 E-Strategy 中, 如果不同频率的数据序列共享同一链路, 那么只需要传输具有最大频率的数据序列, 另外的数据序列可以由最大频率的数据序列重构而得. 定理 4 给出了 E-Strategy 的性质.

定理 4 给定 m 个频率 $f_{r1} \ f_{r2} \ \dots \ f_{rm}$, E-Strategy 发送数据的频率 f_E 和 N-Strategy 发送数据的频率 f_N 满足 $f_N \leq f_E$.

定理 4 说明 E-Strategy 传输数据的频率不高于 N-

Strategy, 因此具有较低的传输能耗.

在多频率查询系统中利用 E-Strategy 来降低能耗, 有三个问题需要解决: 1) 如何寻找合适的查询路由, 使得数据在传输给各接收节点的时候能尽可能多地共享链路; 2) 如何组织共享链路上的节点, 实现在共享链路上仅传输一个整合频率的数据序列; 3) 如何执行数据序列的频率转换, 从整合频率的数据序列中重构出所需频率的数据序列. 我们将在文章的第三节中分别给出具体的解决方案.

3 节能优化方法

无线传感器网络的多频率查询节能优化方法主要包含四个方面的内容: 查询请求的扩散、链路共享路由树的建立、数据序列的整合及传输、以及数据序列的频率转换.

3.1 查询请求的扩散

当接收节点产生一个查询请求的时候, 接收节点将该请求广播给它的邻居节点, 对于任意一个邻居节点 n 而言, 如果这是第一次接收到该查询请求, 那么节点 n 将决定是否要继续广播该查询请求. 如果该查询请求来自于节点 n 唯一的邻居节点, 那么节点 n 将不再广播该查询请求, 否则广播该查询请求. 如果对于节点 n 而言, 该查询请求是重复的请求, 那么节点 n 保持沉默, 不做任何动作, 上述过程不断重复, 直至查询请求最终到达数据源节点.

显然, 在上述过程中, 每一个节点最多广播查询请求一次. 如果每一个节点在广播查询请求的过程中, 均将自身的信息添加到查询请求中, 那么当查询请求到达任意一节点的时候, 该节点可以根据查询请求中所添加的节点信息, 找到通往接收节点的路径. 假设任意两相邻节点传输查询请求的时间均相同, 那么查询请求的扩散过程与广度优先搜索相似, 因此, 对于任意一个节点而言, 第一次接收到查询请求时, 查询请求中包含了该节点到接收节点的最短路径 (距离均用最少跳步数衡量). 对于拥有多个接收节点的情况, 为了在建立路由树的时候兼顾最小化路径长度与最大化共享链路, 网络中的每个节点均缓存自身到各个接收节点的最短路径. 链路共享路由树的构造过程见下一节.

3.2 链路共享路由树的建立

E-Strategy 最基本的想法就是通过共享尽可能多的链路, 来减少数据的传输量, 以达到节能的目的. 然而, 最大化共享链路的问题等同于寻找最小代价 Steiner 树的问题, 即:

给定带权无向图 $G = \langle V, E, W \rangle$ 以及节点的集合 $U \subseteq V$, 其中 $W = \{w(e) | e \in E\}$ 表示各边的权值, 寻找最小代价的边集合 T , 即 $\min_e \tau_w(e)$, 使得 $T \subseteq E$, 且满

足对于任意的 $s, t \in U$, 子图 $\langle V(T), T \rangle$ 均包含一条从 s 到 t 的路径. 集合 T 称为集合 U 在图 G 中的最小代价树. 符号 $V(T)$ 表示 T 中所包含的所有顶点.

已知最小代价 Steiner 树的问题是 NP 难问题, 因此我们在构造链路共享树的时候采用启发式的近似方法, 以期达到近似最优的链路共享树. 此外, 对于无线传感器而言, 任何一个节点都没有网络的全局拓扑信息, 因此路由树的构造还必须是一个分布式的过程. 总体上来说, 链路共享路由树的构造由数据源节点开始, 各接收节点按照查询频率由高至低、分布式逐个按最短路径接入到已生成的路由树上, 最终形成近似最优的链路共享路由树. 同时, 为了利用各数据序列之间的相关性, 需要在传输数据序列的中间节点上记录数据序列的频率信息. 因此, 无线传感器网络的所有节点均维护一个链表 RequestList, 用于维护所有经历该节点查询请求的数据频率.

当数据源节点收到来自不同接收节点的查询请求时, 分别将各接收节点查询请求的数据频率添加到 RequestList 中, 并按照数据频率由高至低对接收节点进行排序, 根据 3.1 节所述, 网络中每个节点均缓存了自身至各接收节点的最短路径, 因此数据源节点把到第一个接收节点的最短路径选择为初始的路由树 T_1 . 为了将第 $i (i > 1)$ 个接收节点 s_i 按最短路径接入已构造的路由树 T_{i-1} 中, 数据源节点需要在 T_{i-1} 中向接收节点的方向发送一个记为 M_{i-1} 的探测消息. 接收了 M_{i-1} 的节点 u 通过与邻居节点交换信息, 将 M_{i-1} 转发给距离 s_i 最近的邻居节点; 若最近的邻居存在多个, 则转发给具有更高数据频率的邻居节点; 若所有的邻居节点到 s_i 的距离均大于节点 u 自身到 s_i 的距离, 则节点 u 被选为 s_i 的接入点, 设 u 到 s_i 的最短路径为 $P(u, s_i)$, 那么有 $T_i = T_{i-1} \cup P(u, s_i)^*$, 同时在 $P(u, s_i)$ 上所有节点的 RequestList 中添加 s_i 节点查询请求的数据频率. 因为探测消息 M_{i-1} 需要沿着 T_{i-1} 进行传输, 因此两个探测消息之间必须插入一个时间间隔 T . 由于所有的探测消息均由数据源节点发出, 探测消息 M_{i-1} 始终先于 M_i 发出, 故 T 不需要设为太大的数值, 路由树的构造不会引入太大的延迟. 从上述各步骤可以看出, 链路共享路由树的建立是一个分布式的过程.

3.3 数据序列的整合及传输

根据 3.2 节所述, 在连接数据源节点及各接收节点的路由树上, 每个中间节点都保存了经过自身的查询请求的数据频率, 因此, 中间节点很容易判断是否存在链路共享. 实际上存在链路共享的节点, 其 RequestList 链表中包含至少两个频率. 如果存在链路共享的节点将 RequestList 中最大的频率作为整合后的数据频率, 那么通过共享链路能够减少数据的传输量, 从而节省能耗.

图 2 描述了数据序列的整合及传输算法, 从中可以发现, 数据源节点仅广播具有最高查询频率的数据序列, 因此对于中间节点来说, 其接收的数据序列的数据频率 $ReceivedF$ 大于 RequestList 中最高的频率 $RequestF$, 这意味着, 输入的数据量超过了所需要的数据量. 如何保证接收节点能够按照既定的频率接收到数据, 那么需要将数据频率进行转换, 我们将在下一节讨论如何从频率为 $ReceivedF$ 的数据序列中重构出频率为 $RequestF$ 的数据序列.

```

DataDissemination ( MyID)
begin
    RequestF FindMax( RequestList );
    if ( MyID = SourceID ) then // broadcast at the requested frequency
        broadcast ( Data , RequestF );
    else
        receive ( Data );
        ReceivedF GetFrequency ( Data );
        if ( RequestF < ReceivedF ) then // do downr sampling
            converFrequency ( Data , ReceivedF , RequestF );
            SendF RequestF ;
        else SendF ReceivedF ;
        if ( myID = SinkID ) then
            to Application ( Data );
        else broadcast ( Data , SendF );
    end if ;
end ;

```

图 2 数据的整合及传输算法

3.4 数据序列的频率转换

数据序列的频率转换问题可以定义为: 给定一个数据频率为 f_1 的数据序列 X , 如何获取数据频率为 f_2 的数据序列 Y 各个时间点的数据? 数据序列的频率转换问题与插值问题相类似, 都是根据一组离散数据点, 获取另一组数据点, 因此我们采用插值技术来实现数据频率的转换. 插值算法有很多种, 如线性插值、二次插值、三次样条插值等等. 本文中我们选择线性插值, 原因有两种: 首先, 线性插值是最简单的插值算法, 计算复杂度低; 其次, 模拟实验表明线性插值也具有较好的精度 (平均相对误差低于 5%).

在线性插值中, 待插入的数值位于任意两个连续的数据样本值之间的直线上, 因此我们可以用下式计算数据序列 Y 的估计数据序列 \hat{y} .

$$\hat{y}[i] = (x[\lfloor z_i \rfloor + 1] - x[\lfloor z_i \rfloor]) \cdot (z_i - \lfloor z_i \rfloor) + x[\lfloor z_i \rfloor] \quad (2)$$

其中 $z_i = i \cdot f_1 / f_2$, $\lfloor z_i \rfloor$ 表示求不大于 z 的最大整数.

* 实际上在寻找接入点 s_i 的时候, 并没有全局的信息, 因此找到的接入点 u 仅仅是局部最优值, $P(u, s_i)$ 并不是全局意义上 T_{i-1} 到 s_i 的最短路径

如果已知数据序列 Y 的真值, 我们可以利用平均相对误差(ARE) 来衡量插值的精度. 对于一个长度为 len 的数据序列, ARE 可以用下式进行计算.

$$ARE(Y, \hat{Y}) = \left(\frac{\sum_{i=0}^{len} |y[i] - \hat{y}[i]|}{\sum_{i=0}^{len} y[i]} \right) / (len + 1) \quad (3)$$

其中 $y[i] > 0$, 若 $y[i] = 0$, 那么 $|y[i] - \hat{y}[i]| / y[i]$ 用 $|y[i] - \hat{y}[i]| / (\max_{0 \leq i < len} y[i] - \min_{0 \leq i < len} y[i])$ 代替.

由式(2)可知, 如果欲求 \hat{Y} 第 i 个时间点的数值, 需要数据序列 X 的第 L_{z_i} 个以及 $(L_{z_i} + 1)$ 个时间点的数值. 实际上, $L_{z_i} \cdot f_1 - i / f_2 < (L_{z_i} + 1) / f_1$, 这说明计算 \hat{Y} 当前时间点数值的时候需要数据序列 X 将来时间点的数值. 解决这个问题的一种方法就是利用一定的历史信息, 对数据序列 X 的下一个时间点的数值进行预测, 我们采用下面的式子进行预测:

$$x[L_{z_i} + 1] = \alpha \cdot x[L_{z_i}] + (1 - \alpha) \cdot x[L_{z_i} - 1] \quad (4)$$

在图 2 所示的算法中, 当 $ReceivedF$ 大于 $RequestF$ 时, 那么以 $ReceivedF$ 为频率的已知数据序列就是 X , 以 $RequestF$ 为频率的未知数据序列就是 Y , 结合式(4)和式(2), 对于每一个自然数 i , 可以计算出 $\hat{y}[i]$ 的数值. 由于执行频率转换时, $ReceivedF$ 总是大于 $RequestF$, 即每一次都是将高频率的数据序列转换成低频率的数据序列, 因此具有较小的误差.

4 性能分析

首先, 我们分别给出极限情况下 N-Strategy 和 E-Strategy 理论上的能耗上界, 比较两种方法的优劣; 然后通过模拟实验比较一般情况下的能耗. E-Strategy 方法在性能上的优势得益于传输数据序列的过程中, 能够充分地共享传输链路, 进而节省了数据传输所需要的能量.

4.1 理论分析

定理 5 在极限情况下, 假设传感器网络中除数据源自身之外, 其他所有的节点均向该数据源节点发送查询请求, 那么单位时间里整个网络采用 N-Strategy 方法, 能耗的上界是 $O(D \cdot (N - 1))$, 而 E-Strategy 能耗的上界是 $O(N - 1)$, 其中 D 表示传感器网络的直径(采用跳步数衡量), N 表示传感器网络的节点数目.

证明 由定理 1, 采用 N-Strategy 方法, 能耗的上界为 $\sum_{i=1}^m f_i \cdot d_i$, 其中, d_i 表示接收节点到数据源节点的跳步数. 由于 $d_i \leq D$, 那么有

$$\sum_{i=1}^m f_i \cdot d_i \leq \sum_{i=1}^m f_i \cdot D = f_{\max} \cdot D \cdot (N - 1) \cong O(D \cdot (N - 1)) \quad (5)$$

采用 E-Strategy 方法, 因为每次传输的都是整合频率后的数据序列, 因此能耗的上界为每个节点均按照查询请求频率的最大值 f_{\max} 进行数据传输, 传输距离的上

界为网络的直径 D , 因此总能耗的上界为 $f_{\max} \cdot (N - 1)$, 即 $O(N - 1)$.

显然, E-Strategy 方法在能耗方面优于 N-Strategy 方法, 对于不同频率的查询请求而言, 数据在传输的过程中共享越多的链路, 那么 E-Strategy 就能比 N-Strategy 节省更多的能量. 定理 5 给出了在完全共享链路的极限情况下, E-Strategy 相对于 N-Strategy 在理论上的优势.

4.2 模拟实验

4.1 节在理论上分析了极限情况下 E-Strategy 相对于 N-Strategy 的优势, 为了更充分地评价 E-Strategy 的性能, 我们利用 Matlab 分别模拟了 E-Strategy 和 N-Strategy 的能耗及数据查询的精度. 此外, 我们还研究了网络参数与查询参数对两种方法性能所产生的影响.

表 1 传感器网络及查询的参数

参数说明	符号	缺省值
无线传感器网络覆盖面积		300 × 300m ²
节点个数	N	400
传输半径		30m
接收节点个数	m	6
查询频率	f	1 - 10
查询距离	H	3 - 10hops

在模拟实验中, 所有的传感器结点均匀分布于某一矩形区域, 传感器结点的通信半径均为, 其他网络参数、查询参数以及相应的缺省值见表 1.

为了确保模拟实验的可重复性, 我们采用人工合成的方法产生数据源的数据. 数据源节点的时间序列函数是一个随机行走序列^[13]的函数变换, 可以用下面的函数表示:

$$x[i] = 100 \cdot (\sin(0.1 \cdot RandonWalk[i]) + 1 + i / R) \quad (6)$$

其中 $i = 0, \dots, R - 1$; $RandonWalk[0, \dots, R - 1]$ 是随机行走序列; R 是行走的范围, 设为 100000. 单位时间间隔设为所有接收节点查询请求的数据频率的最小公倍数, 以确保所有数据序列的时间间隔均为整数.

每一次实验中, 随机选择一个数据源节点以及多个数据接收节点, 每一个接收节点在 1 - 10 范围内任意选择一个整数作为查询的频率. 为了验证 3.2 节中链路共享路由协议的效率, 我们将其与定向扩散路由协议^[5]进行比较. 由于定向扩散路由协议发现的路径是最短路径路由树, 我们将其记为 SPT(Shortest Path routing Tree), 链路共享路由协议发现的是链路共享路由树, 我们将其记为 PST(Path-Sharing routing Tree). 通信的能耗用单位时间里平均发送的报文数来衡量, 值越小表明性能越好; 查询的精度用各接收节点 ARE 的均值来衡量, 值越小表明性能越好.

由于无线传感器网络的随机性, 模拟实验所统计的是平均性能指标(单位时间平均发送报文数、ARE 均

值),其计算方法如下:设平均性能指标为 \bar{C} ,假设随机生成 u 个连通的网络拓扑结构,每个拓扑结构下随机的进行 v 次实验,第 i 个拓扑结构下的第 j 次实验的性能指标记作 C_{ij} ,那么有 $\bar{C} = \frac{1}{uv} \sum_{i=1}^u \sum_{j=1}^v C_{ij}$. 实验中, $u = 30, v = 100$.

4.2.1 不同查询距离的影响

第一组实验主要研究不同查询距离的情况下,N-Strategy 策略和 E-Strategy 策略所需的查询能耗,以及 E-Strategy 策略的查询精度. 查询距离主要用于反映数据源节点相对于接收节点的远近程度,其大小可以用接收节点到数据源节点的平均跳步数来计算. 在该组实验中,我们指定传感器节点的总数 $N = 400$,模拟结果如图 3 和图 4 所示. 由于 N-Strategy 本身采用最短路径进行路由,因此只有一组曲线图;对于 E-Strategy,采用最短路径进行路由的结果用 E-Strategy-SPT 表示;采用链路共享进行路由的结果用 E-Strategy-PST 表示.

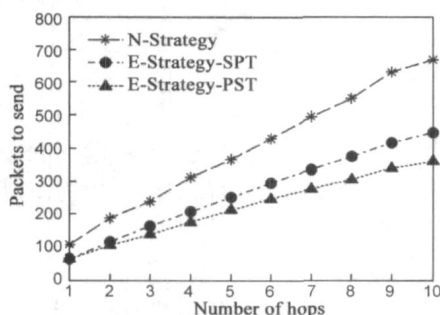


图 3 查询距离对能耗的影响

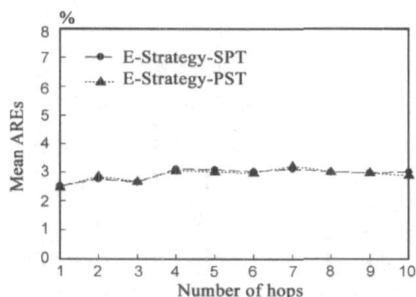


图 4 查询距离对精度的影响

从图 3 不难看出,采用 E-Strategy 可以较显著地减少传输报文的数量,进而降低传输的能耗,尤其是采用链路共享路由由树进行通信时,节能特性更加明显. 当查询距离 H 增加时,能耗呈线性增长,并且 N-Strategy 的能耗比 E-Strategy 的能耗增长要快,这是因为 N-Strategy 体现的是各条查询能耗的累加和,而 E-Strategy 具有共享链路的特性,从而具有较少的能耗. 基于链路共享路由树的 E-Strategy 能耗低于基于最短路径路由树的 E-Strategy,这是因为后者产生共享链路的概率低于前者. 由图所示,当平均查询距离为 10 跳时,基于链路共享树的 E-Strategy 能够比 N-Strategy 节省 50% 左右的能耗.

图 4 显示了 E-Strategy 引入的平均相对误差,采用线性插值对数据序列的频率进行转换,仍然能够保持较好的精度,平均相对误差控制在 3% 左右,并且与路由方式无关,与查询距离无关.

4.2.2 不同节点密度的影响

由于网络的拓扑结构与网络结点的密度相关,本组实验主要研究不同结点密度情况下,不同方法的能耗. 实验中,查询距离 H 等于 6,结点密度的变化通过改变结点的总数量 N 来实现,实验结果如图 5 和图 6 所示.

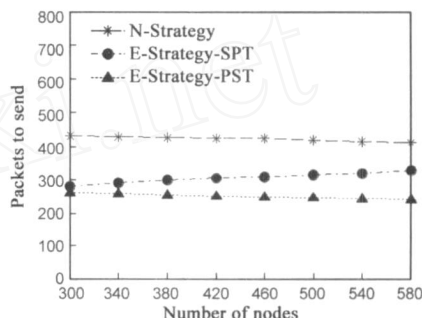


图 5 结点密度对能耗的影响

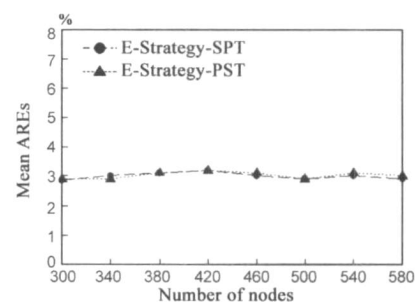


图 6 结点密度对精度的影响

从图 5 同样不难看出,采用 E-Strategy 可以较显著地减少传输的能耗. 随着节点密度的增加,N-Strategy 的能耗以及 E-Strategy-PST 的能耗均有轻微的降低,这是因为密度增加时,每一个节点的邻居数量增多,从而更有利于减少数据源节点到接收节点的跳步数,从而降低传输的能耗. 但是,随着密度的增加,E-Strategy-SPT 的能耗却有轻微的上升,这是因为数据源节点到接收节点的跳步数减小的同时,最短路径路由由树共享链路的可能性也随之减小,反而引起总能耗的增加,这一现象也进一步说明,对于 E-Strategy 来说,共享链路的特性比短路径的特性更重要.

考虑精度问题,图 6 显示 E-Strategy-SPT 和 E-Strategy-PST 均能将 ARE 的均值保持在 3% 左右,并且误差与路由方式无关,与节点密度无关.

4.2.3 不同接收节点数目的影响

网络的通信开销与接收节点的数目,也即查询请求的数目紧密相关,因此本组实验主要研究不同接收节点数目情况下,各种查询方法的能耗. 实验中,节点的总数 N 等于 400,查询距离 H 等于 6,接收节点的数目由 1 递

增至 10, 实验结果如图 7 和图 8 所示。

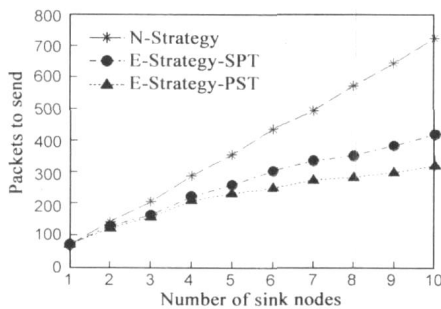


图 7 接收节点数量对能耗的影响

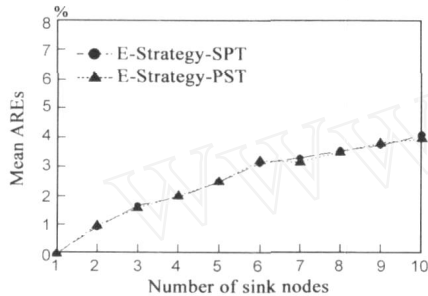


图 8 接收节点数量对精度的影响

图 7 同样表明 E-Strategy 能够显著的降低传输的能耗。当接收节点数目 m 增大的时候, N-Strategy 的能耗以及 E-Strategy 的能耗均近似线性的增长, 并且 N-Strategy 能耗增长的速度大于 E-Strategy 能耗的增长速度, E-Strategy-SPT 能耗增长的速度大于 E-Strategy-PST 能耗的增长速度。这是因为, 更多的接收节点意味着更多的查询请求, 从而导致更多的通信开销, 而基于链路共享路由树的 E-Strategy, 可以通过尽可能多地共享传输链路达到降低能耗的目的。当接收节点达到 10 的时候, E-Strategy-PST 与 N-Strategy 相比, 可以节省约 55% 的能耗。

与查询距离和节点密度不同, 接收节点的数目对数据序列频率的转换精度具有一定的影响。如图 8 所示, ARE 的均值随着接收节点数目的增加而增大, 这是因为更多的接收节点意味着更多不同的查询频率, 从而导致频率间所需的转换次数增加, 进而增大 ARE。尽管如此, 当接收节点增大到 10 时, ARE 的均值仍然低于 5%, 换句话说, 对于相当数量接收节点的情况下, ARE 的均值仍然可以接受。根据 E-Strategy-SPT 和 E-Strategy-PST 的曲线所示, ARE 的均值与路由方式无关。

5 结论

能耗是影响无线传感器网络性能的关键因素之一, 本文针对无线传感器网络中存在的多种查询频率共存的问题, 提出了一种节能的数据查询方法, 理论分析和模拟实验均表明, 在一定误差的允许范围内, 我们提出的方法能够有效的减少数据的传输量, 从而降低网络的传输能耗。在后续的研究中, 扩展性的工作主要有三个

方面: 首先, 链路共享路由树的节能需要节点 MAC 层多播操作的配合, 设计节能可靠的多播 MAC 层协议是将来研究的内容之一; 其次, 考虑无线传感器网络的动态特性, 需要研究节点移动及部分失效对系统带来的影响; 最后, 文章中的频率转换机制仅限于感知数据为实数值的情况, 如何将多频率查询机制扩展到感知数据为声音、图像等情形, 需要进一步的研究。

参考文献:

- [1] 郭龙江, 李建中, 李贵林. 无线传感器网络环境下时-空查询处理方法[J]. 软件学报, 2006, 17(4): 794 - 805.
Guo LJ, Li JZ, Li GL. Spatio-Temporal query processing method in wireless sensor networks [J]. Journal of Software, 2006, 17(4): 794 - 805. (in Chinese)
- [2] U Srivastava, K Munagala, J Widom. Operator placement for in-network stream query processing [A]. In Proc ACM SIGMOD [C]. ACM Press, 2005. 250 - 258.
- [3] BJ Bonfils, P Bonnet. Adaptive and decentralized operator placement for in-network query processing [J]. Telecommunication Systems, 2004, 26(2 - 4): 389 - 409.
- [4] YW Chen, HV Leong, M Xu, J Cao, KCC Chan, ATS Chan. In-network data processing for wireless sensor networks [A]. In Proc IEEE MDM [C]. IEEE Computer Society, 2006. 26 - 29.
- [5] B Krishnamachari, D Estrin, S Wicker. Modelling data-centric routing in wireless sensor networks [A]. In Proc. IEEE INFOCOM [C]. IEEE Computer Society, 2002. 2 - 14.
- [6] C Razvan, B Baltasar, V Martin, W Roger. Network correlated data gathering with explicit communication: NP-Completeness and algorithms [J]. IEEE Trans Netw, 2006, 14(1): 41 - 54.
- [7] 林亚平, 王雷, 陈宇, 张锦, 陈治平, 童调生. 传感器网络中的分布式数据汇聚层次路由算法 [J]. 电子学报, 2004, 32(11): 1883 - 1889.
Lin YP, Wang L, Chen Y, Zhang J, Chen ZP, Tong TS. A distributed data-centric clustering hierarchical routing algorithm for sensor networks [J]. Acta Electronica Sinica, 2004, 32(11): 1801 - 1805. (in Chinese)
- [8] 谢志军, 王雷, 林亚平, 陈红, 刘永和. 传感器网络中基于数据压缩的汇聚算法 [J]. 软件学报, 2006, 17(4): 860 - 867.
Xie ZJ, Wang L, Lin YP, Chen H, Liu YH. An algorithm of data aggregation based on data compression for sensor networks [J]. Journal of Software, 2006, 17(4): 860 - 867. (in Chinese)
- [9] HS Kim, TF Abdelzaher, WH Kwon. Minimum-energy asynchronous dissemination to mobile sinks in wireless sensor networks [A]. In Proc. ACM SenSys [C]. ACM Press, 2003. 193 - 204.
- [10] B Krishnamachari, J Ahn. Optimizing data replication for expanding ring-based queries in wireless sensor networks [A]. In

Proc. 4th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks [C]. IEEE Press, 2006. 1 - 10.

[11] J Lesurf. Information and Measurement[M]. London: Institute

of Physics Publishing, 2002.

[12] L Gao, XS Wang. Continually evaluating similarity-based pattern queries on a streaming time series[A]. In Proc ACM SIGMOD[C]. ACM Press, 2002. 370 - 381.

作者简介:



陈颖文 男,1979 年出生于湖南株洲,博士研究生,分别于 2001 年和 2003 年在国防科技大学管理科学与工程学院获得学士学位和硕士学位,2004 年起在国防科技大学计算机学院攻读博士学位,主要研究方向为无线传感器网络的数据路由、数据查询等。

E-mail: ywch_nudt@hotmail.com



徐明 男,1964 年出生于湖南长沙,博士,教授,博士生导师。现为国防科技大学计算机学院网络工程系副主任,主要研究领域为计算机体系结构、无线网络与移动计算技术等。

虞万荣 男,1976 年出生于江苏丹阳,博士,助理研究员,2006 年于国防科技大学计算机学院获得博士学位。研究领域为 Ad hoc 网络、无线传感器网络通信协议等。

www.cnki.net