

传感器网络中一种基于估计代价的数据聚合树生成算法

叶 宁^{1,2}, 王汝传^{1,3}

(1. 南京邮电大学计算机学院, 江苏南京 210003; 2. 南京人口管理干部学院, 江苏南京 210042;
3. 南京大学计算机软件新技术国家重点实验室, 江苏南京 210093)

摘 要: 无线传感器网络是一种全新的技术,能够广泛应用于恶劣环境和军事领域.传感器网络在数据收集过程中,为减少冗余数据的传输耗能,降低延迟,需要采用数据聚合技术.本文采用定向传输方式,在消息路由机制基础上提出了一种基于估计代价的数据聚合树生成算法.该算法主要思想在于将节点能耗、传输距离与聚合收益三方面作为估计代价,优化聚合路径,实现数据聚合在能量与时延上的折中.

关键词: 无线传感器网络; 数据聚合树; 估计代价

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2007) 05-0806-05

A Tree Formation Algorithm for Data Aggregation Based on Estimate Cost in Sensor Networks

YE Ning^{1,2}, WANG Ruchuan^{1,3}

1. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210003, China;
2. Department of Information Science, Nanjing College for Population Programme Management, Nanjing, Jiangsu 210042, China;
3. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu 210093, China)

Abstract: Wireless sensor network is a novel technology, and can be applied to both abominable and military environments. During gathering the data, wireless sensor network should decrease the power costs of redundancy information and delay time. The technology of data aggregation can be adopted. A tree formation algorithm for data aggregation based on estimate cost is presented. The paper applies directed transmission model and message route mechanism. The idea of this algorithm is optimization of data aggregation route by estimate cost about energy, distant and aggregation gain, then realizes data aggregation trade-off in energy and delay.

Key words: wireless sensor network; data aggregation tree; estimate cost

1 引言

无线传感器网络由大量微型传感器通过无线链路形成,综合了传感器、嵌入式计算、分布式信息处理和无线通信等技术,具有可快速部署、可自组织和容错性高的特点.目前无线传感器网络已广泛应用在军事、环境、健康、家庭和商业等领域.

散布在网络中的大量传感器节点实时监测感知分布区域内的各种目标信息,并通过彼此之间协同工作,将采集的数据传送到汇聚节点(sink).与传统无线移动自组织网不同,无线传感器网络一般具有较大的节点密度,受成本及体积等因素的限制,传感器节点在存储容

量、计算能力和电池能量等方面都非常有限,且由于部署区域环境复杂,节点的能量无法得到补充.因此如何最大化网络的工作寿命成为无线传感器网络研究的一个关键问题.数据聚合技术的提出旨在通过对来自其他路由数据的合并处理,减少网络传输的数据量,降低能耗,延长网络的生存期.目前许多无线传感器研究领域,都采用了数据聚合技术. TinyDB 和 COUGAR^[1,2] 系统利用分布式数据库技术,通过执行 MIN、MAX、COUNT 等融合查询操作,达到聚合的效果.路由协议的研究与数据聚合技术结合,可以在一定程度上实现数据分组沿优化的聚合路径转发,减少通信量.其中定向扩散(directed diffusion, DD)路由协议^[3]采用面向应用的网内数据处

收稿日期:2006-08-07;修回日期:2006-10-20

基金项目:国家自然科学基金(No. 60573141, No. 70271050);江苏省自然科学基金(No. BK2005146);江苏省高技术研究计划(No. BG2005038, No. BG2006001);国家高科技 863 项目(No. 2005AA775050);南京市高科技项目(2006 软资助 105);现代通信国家重点实验室基金(No. 9140C1101010603);江苏省计算机信息处理技术重点实验室基金(No. kjs050001, No. kjs0606);南京人口干部管理学院科研项目(No. 2006C14)

理,实现在传输数据的同时计算聚集. LEACH^[4]路由协议使用分簇的方法进行数据聚合. 通过随机选择聚类首领,平均分担中继通信业务来实现. PEGASIS^[5]路由协议则通过发送能量递减的测试信号来选择聚类. 聚类首领参照位置关系优化出到 sink 节点的最佳链路. 与 LEACH 协议相比,PEGASIS 协议数据聚合能力更强,能量损耗更低. 目前虽然大多数研究人员认为数据聚合技术提高了无线传感器网络的传输性能,但由于传感器节点易失效性,所以数据聚合性能受多方因素影响,包括节点的定位、覆盖密度及拓扑结构等^[6]. Bhaskar 在文献[6]中提出,以数据为中心的聚合对于传感器网络能量与时延性能影响是折中的. 本文从 DD 协议出发,分析了影响数据聚合的主要方面. 在此基础上提出一种将节点能耗、传输距离与聚合收益 (Aggregation Gain, AG) 三方面作为估计代价的聚合树生成算法. 目的在于优化选择网内最佳的数据聚合处理节点,实现数据聚合在能量与时延上的折中. 最后通过仿真实验加以验证.

2 系统模型和问题描述

2.1 网络模型

本文研究基于无线传感器网络定向传输方式,网络拓扑由一个固定位置的 sink 节点和一定数量随机分布的传感器节点组成. 其中 sink 节点负责数据收集任务的发布以及将收集的数据转发到外部网络. 根据数据处理位置,聚合模式一般分为集中和分布(网内)两种^[7]. 前者模式中 sink 节点作为唯一的数据聚合中心. 虽然保证信息损失较小,但大量的数据传输耗能很大. 本文的研究采用分布式数据聚合方式,将传感器节点分为源节点和聚合节点两类. 源节点负责环境的监测,感知数据的生成与发送;聚合节点则具有数据转发和聚集双重功能.

无线传感器网络是以数据为中心的^[8],即网络中的节点不以地址作为标识 ID,而以节点可以提供的数据作为寻址依据. sink 节点在网络中广播以某种数据格式构成的消息(称兴趣)询问它所感兴趣的监测数据,与这种兴趣匹配的节点(称源节点)响应这种查询(称事件),并回送监测数据给 sink 节点,如图 1 所示.

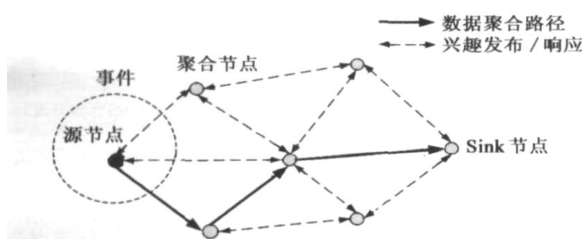


图 1 无线传感器网络数据聚合

2.2 通信模型

从节省能量考虑,无线传感器网络的 MAC 协议通常采用“侦听/休眠(listen/sleep)”交替的无线信道使用策略. 本文采用文献[9]中提到的 MAC 协议优化方法,节点完成数据聚合任务后,立即进入侦听/休眠状态.

2.2.1 通信时延

设定一个区域中分布 n 个传感器节点 $1, 2, \dots, n$ 和一个 sink 节点. 假设在相同的信道宽度 B 下,每个传感器单位时间产生一个数据包发送到 sink 节点. 为简化,我们将每单元时间称为一轮(round),每轮离散为 m 个时间槽(slot),且假设所有数据包的大小均为 K 比特. 源自不同数据源的信息需要在每轮收集,经聚合,发送到 sink 节点,其中数据源与 sink 节点间距离用跳数(hop)表示.

假设在侦听/休眠模式下,由多个节点发出的数据包在传感器 i 上发生聚合. 若 S 表示所有传感器节点组成的集合,则传感器 i 上产生的时延 D_i 为:

$$D_i = T_s(i) + T_l(i) + T_{tr}(i) + T_{AG}(i), \quad i \in S \quad (1)$$

其中 T_s 、 T_l 分别表示传感器节点 i 的睡眠时延与侦听时延; T_{tr} 是数据的发送时延,其计算公式如下:

$$T_{tr}(i) = \frac{K}{B}, \quad i \in S \quad (2)$$

T_{AG} 是由数据聚合处理产生的时延,假设在当前研究的网络模型下,各传感器采用相同的数据聚合算法,则:

$$T_{AG}(i) = \frac{K_{total}}{B}, \quad i \in S \quad (3)$$

其中 K_{total} 是该节点收到的数据包数量,是节点单位时间可聚合的数据量.

传感器网络每轮产生的全部数据总时延为:

$$D = \sum_{i \in S} \{ T_s(i) + T_l(i) + T_{tr}(i) + T_{AG}(i) \} \quad (4)$$

2.2.2 消息路由机制

在定向传输模式下,当一个 sink 节点需要从传感器区域收集数据时,它以数据包形式发布一个请求消息,该消息包括以下内容:

- (1) 所需收集的数据类型,一般为特定属性值,如温度、压力、湿度、光照等;
- (2) 与消息所匹配的信息传送时间间隔;
- (3) 消息的生存期;
- (4) 消息发布的区域;

源节点携带所需的信息通过响应消息来应答请求. 响应消息数据包中除了通信应用层数据外,还根据不同的路由策略包括用于建立路由的服务数据,如本节点标识 ID,下一跳目的节点标识 ID 等.

通过 sink 节点与源节点间的请求响应过程,建立起数据源到处理节点之间的路由,如图 2 所示.

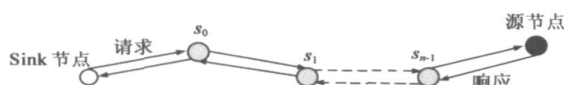


图2 消息路由机制

2.3 问题描述

基于上述消息路由机制,数据从源节点发送到 sink 节点的聚合路由过程类似于反向组播聚合树的建立过程^[6].考虑通信时延,分析得出以下几方面的研究问题:

(1) 节点 s_i 能否成为聚合节点,依赖于其他数据包到达该节点的可能性.

(2) 在传感器节点覆盖密集条件下,源节点感知的数据存在较高的相似性.在同一聚合节点上,已达的数据包,与后达数据包进行聚合的可能性随停留时间增大而增大.

(3) 若考虑节点在状态转换间的能耗,随着传输时延的增加,在一定程度上将消耗能量.

3 基于估计代价的聚合树生成算法

3.1 预备知识

设定 N 个源节点与一个 sink 节点通过请求/响应消息建立数据源至数据处理中心的聚合路由.在每个时槽内,一个消息到达聚合节点 A_n 的概率依赖于路径该节点的发送数据包节点数 W_n .设 M_n 是在聚合时延 T_n 内区域内产生的平均消息数.在 M_n 个时槽内一个到达聚合节点 A_n 的消息概率 P_n 可以通过以下公式计算:

$$P_n(A_n) = 1 - \left(\frac{N - W_n}{N} \right)^{m_n} \quad (5)$$

其中 M_n 必须满足条件: $\sum_{n=1}^h m_n = m$, 其中 h 是消息从源节点传至 sink 节点经历的跳数.

对于每个时槽,即 $M_n = 1$ 时,一个消息到达的概率为:

$$P_n(A_n) = 1 - \left(\frac{N - W_n}{N} \right) = \frac{W_n}{N} \quad (6)$$

定义1(聚合收益) 是一种测度,用于反映数据聚合在降低传输通信量方面的收益.

设 M 是实现给定应用层任务所需传输的数据量,采用聚合方法执行相同任务所需传送的数据量为 M_a ,则网络系统的聚合收益为:

$$= 1 - M_a / M \quad (7)$$

对于同一个聚合事件,由于受能量约束及邻近节点分布影响,各聚合节点的数据聚合能力存在差异.因此考虑研究单个聚合节点的估计聚合效益,为优化选择聚合节点提供依据.

定义2 (节点度 d) 所有距离该节点一跳的邻居节点的数目.

假设在每个时槽内各节点数据传输速率相同,各聚合节点采用相同的数据聚合算法条件下,给定聚合节点 A_n 所需传输的可能最大数据量 M 为所有源节点发送数据量总和.不失一般性,节点最大可传输的数据量与该节点度 d_n 成正比.聚合节点执行聚合后,可发送的数据量与可达该节点消息概率成正比.因此,改写式(7),聚合节点的估计聚合收益为:

$$= 1 - d_n / N \quad (8)$$

3.2 算法描述

本文采用 GEAR 路由机制^[10]中的路径代价方法,考虑节点间数据传输距离、节点剩余能量以及节点自身的估计聚合收益三方面因素.在建立从事件区域源节点到 sink 节点数据聚合路径时,中间节点使用估计代价来决定下一个聚合节点.假设 S 表示由所有源节点组成的集合, A 表示所有具有数据转发与聚合功能的中间节点组成的集合,则节点计算自己到事件区域源节点的路径代价公式如下:

$$C(i, j) = (d(i, j) + (1 - \alpha)e(i)) / \beta, \quad i \in A, j \in S \quad (9)$$

式(9)中 $d(i, j)$ 为节点 i 到节点 j 的距离(跳数), $e(i)$ 为节点 i 的剩余能量, α 为比例参数, β 为节点 i 的估计聚合收益值.

本算法分为如下3个阶段来实现:探测阶段、反馈阶段和反向聚合树生成阶段.

(1) 探测阶段:本阶段主要任务是 sink 节点周期性向邻居节点广播兴趣消息,消息中增加一个到 sink 节点的跳数域(初始值为0)以及表征消息请求/响应不同阶段的标记 TAG.在探测阶段, TAG = 0.与 sink 节点相邻节点收到该消息后,将 sink 节点作为自己父节点,并设置自己到 sink 节点的跳数为1.若该节点不是源节点,则将消息中跳数域值加1后继续广播.若一个节点同时收到多个广播消息,则选择消息信号更强的节点作为父节点.该过程一直扩展到整个网络,从而形成以 sink 节点为根的树型结构.源节点收到兴趣消息后,不再转发消息,而作为叶节点加入生成树中.

(2) 反馈阶段:各源节点收到 TAG = 0 的兴趣消息后,将其节点度数置为0,并连同自己至 sink 节点的跳数值,以消息形式广播.其父节点根据收到的消息计算源节点数,更新自己的度数,并形成消息沿生成树向 sink 节点转发. sink 节点收到所有子节点消息后,计算获得事件区域源节点数目.各中间节点收到消息后,保存源节点跳数. sink 节点将 TAG 置为1,与源节点数目一起通过消息机制发送到所有中间节点.各中间节点通过本地计算保存估计代价.

(3) 反向聚合树生成阶段:各源节点收到 $TAG = 1$ 的兴趣消息后,沿兴趣消息的反方向传输监测数据.数据传输聚合过程中采用贪婪算法,在邻居节点中选择估计代价最小的节点作为数据聚合节点.这个过程持续下去,一直到达 sink 节点为止.

4 仿真实验

下面将用仿真分析的方法来对该聚合树形成算法进行性能评估,仿真工具采用 NRL (Naval Research Laboratory) 开发的无线传感器网络仿真环境^[11],它基于 NS2 仿真模拟器^[12],可对大规模的网络进行仿真.根据节点有效通信半径,在一个 500m ×500m 的区域内利用坐标数据随机生成具有 50、100、150、200、250 和 300 个节点的拓扑结构,如图 3 所示.

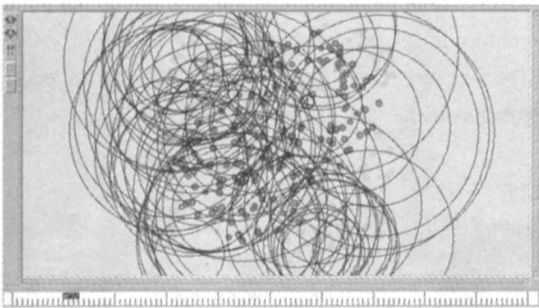


图 3 仿真拓扑结构

节点的具体的仿真参数配置如表 1 所示.

表 1 仿真参数配置表

区域范围	500m ×500m
环境事件生成源 (PHENOME)	2 个 (不移动)
sink 节点	1 个 (位置固定预置)
传感器节点	50、100、150、200、250 和 300 (随机散布)
MAC 协议	802.11
队列方式	Drop Tail/ PriQueue
天线类型	OmniAntenna
无线传输	双向无差错
数据包有效载荷	50Byte
通信带宽	250kbps
节点发送/接收数据能耗	173mW
节点初始能量	0.2J
仿真时间	20 秒

仿真采用的数据聚合模型分别基于定向扩散 (DD) 与本文提出的基于估计代价的聚合树生成算法.算法的性能主要从平均节点能耗与时延两个方面进行评估.在 50 ~ 300 个不同节点规模,上述每种性能仿真都随机运行 20 次,然后取其平均值.

图 4 反映了在同一个节点规模下,传感器网络在运行时间内所有聚合节点剩余能量的变化趋势.仿真场景选取 300 个节点,时间截取前 10 秒.从仿真结果分析:定向扩散与估计代价的聚合算法,由于均采用定向

传输方式,所以前期聚合路由能耗较大.相比较而言,在聚合路由稳定情况下,采用估计代价的聚合算法整个能耗要小于定向扩散方法.

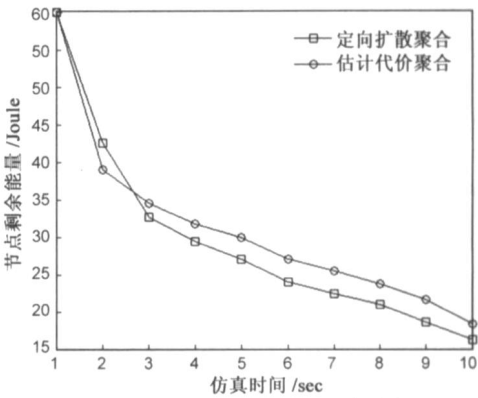


图 4 网络节点剩余能量 (300 个节点)

为进一步比较两种算法对网络能量的影响,我们在不同节点规模下,仿真计算出网内聚合节点的平均能耗,如图 5 所示.仿真结果表明:与定向扩散相比,通过优化数据网内处理中聚合节点的选择,使得网络节点的平均能耗降低 25 % 左右.

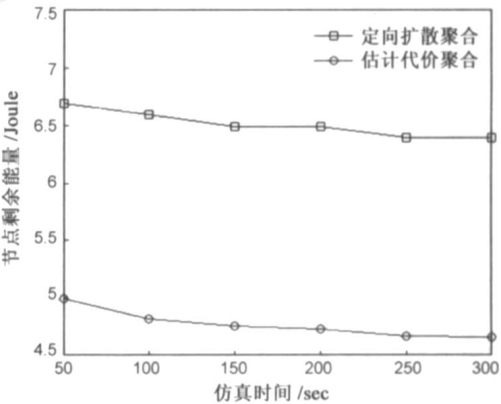


图 5 节点平均能耗

图 6 反映了在不同节点规模,网内数据平均时延.比较两种算法的仿真结果,采用估计代价的数据聚合方法,数据时延有所降低,且随网络规模增大,时延的

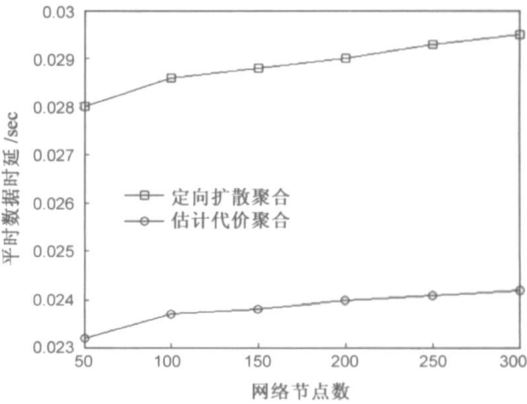


图 6 网络平均数据时延

增加趋于平缓。

5 结论

无线传感器网络是以数据为中心,基于感知事件的分布式数据处理系统。聚合技术通过减少传输数据量,降低了传感器节点的能耗,延长整个网络的生存期。本文提出基于估计代价的聚合树生成算法,将剩余能量以及估计可聚合度作为聚合节点选择的关键因素,优化数据聚合树的生成。通过仿真模拟进一步证明了该算法在聚合能耗与时延两方面提供较折中的解决方法。

参考文献:

- [1] Madden Sam, Hellerstein Joe, Hong Wei. TinyDB: In-Network Query Processing in TinyOS[R]. USA: Computer Department, University of California Berkeley, 2003.
- [2] Yao Y, Gehrke J. The cougar approach to in-network query processing in sensor networks[J]. SIGMOD Record, 2002, 31(3): 9 - 18.
- [3] C Intanagonwiwat, R Govindan, D Estrin. Directed diffusion: A scalable and robust communication paradigm for sensor networks[A]. ACM/IEEE International Conference on Mobile Computing and Networks (MobiCom 2000) [C]. USA, 2000. 56 - 67.
- [4] W R Heinzelman, A Chandrakasan, H Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks[A]. The Proceedings of the Hawaii International Conference on System Sciences [C]. Maui, Hawaii, 2000. 3005 - 3014.
- [5] Lindsey S, Raghavendra C S. Pegasus: Power-efficient gathering in sensor information systems[A]. Proc of the IEEE Aerospace Conf [C]. Montana: IEEE Computer Society, 2002. 1 - 6.
- [6] B Krishnamachari, D Estrin, S Wicker. The impact of data aggregation in wireless sensor networks[A]. Proceedings of the 22nd International Workshop on Distributed Event-Based Systems [C]. Vienna: IEEE Computer Society, 2002. 575 - 578.
- [7] 孙利民, 李建中, 等. 无线传感器网络[M]. 北京: 清华大学出版社, 2005.
- [8] Krishnamachari B, Estrin D, Wicker S. Modelling data-centric routing in wireless sensor networks[EB/OL]. <http://lecs.cs.ucla.edu/Publications/papers/Bhaskar-DataCentric.pdf>, 2002.
- [9] P Popovski, F Fitzek, H Yomo, T Madsen, R Prasad. MAC-layer approach for cluster-based aggregation in sensor networks[A]. International Workshop on Wireless Ad-hoc Networks (IW-WAN) [C]. Finland: 2004. 89 - 93.
- [10] Yu Y, Govindan R, Estrin D. Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks[R]. USA: UCLA Computer Science Department Technical Report UCLA/CSD-TR-01-023, 2001.
- [11] NRL's Sensor Network Extension to ns-2[EB/OL]. <http://nrlsensorsim.pf.itd.nrl.navy.mil/>, 2004.
- [12] The Network Simulator ns-2[EB/OL]. <http://www.isi.edu/nsnam/ns.html>, 2005.

作者简介:



叶宁女, 1971 生于江苏南京, 南京邮电大学计算机学院博士研究生, 南京人口管理干部学院讲师, 主要研究方向为计算机软件在通信中的应用、无线传感器网络等。



王汝传男, 1943 年生于安徽合肥, 南京邮电大学教授、博士生导师, 主要研究方向为计算机软件、计算机网络、信息安全、无线传感器网络、移动代理和虚拟现实技术等。
E-mail: wangrc@njupt.edu.cn