

基于最小覆盖集的 WSN 数据聚集算法

吴 旦, 王改云, 李小龙

(桂林电子科技大学计算机与控制学院, 广西 桂林 541004)

摘 要: 针对无线传感器网络(WSN)能量有限且网内聚集错误检测率较低的问题, 提出一种基于最小覆盖集的 WSN 数据聚集算法。构造一颗以汇聚节点为根并包含最少中间转发节点的树, 其中间转发节点为树的最小覆盖集。在中间转发节点中引入读向量的相似性判断, 以去除网内冗余及错误数据。实验结果表明, 该算法能减少网内通信能耗, 提高收集数据的准确性。

关键词: 无线传感器网络; 最小覆盖集; 错误检测; 数据聚集; 读向量

Data Aggregation Algorithm for WSN Based on Minimal Covering Set

WU Dan, WANG Gai-yun, LI Xiao-long

(Institute of Computer and Control, Guilin University of Electronic Science and Technology, Guilin 541004, China)

【Abstract】 For the limited energy in the Wireless Sensor Network(WSN), and the shortage of low error detection rate in traditional in-net aggregation. A data aggregation algorithm for WSN based on Minimal Covering Set(MCS) is proposed. It constructs a tree, which contains minimum intermediate forwarding nodes, and root at the sink node. The forwarding nodes are the minimal covering set of the tree. In order to remove the redundant and error data, intermediate nodes introduces similarity judgment of read vectors. Experimental results show the algorithm can reduce the energy consumption of communication within the network, and can improve the accuracy of data collection.

【Key words】 Wireless Sensor Network(WSN); Minimal Covering Set(MCS); error detection; data aggregation; read vector

DOI: 10.3969/j.issn.1000-3428.2012.02.031

1 概述

无线传感器网络(Wireless Sensor Network, WSN)由密集部署于监测区域的传感器节点组成, 自组织地形成网络, 协作地感知和处理数据, 并将有用的信息发送给用户, 尤其适用于无人看守、环境比较恶劣的监测区域。目前, 传感器网络已被广泛运用于交通管理、环境监测等领域^[1]。

然而, 由于传感器节点较小, 且依靠电池供电, 如何高效地利用有限的资源, 延长网络的生命周期是一个具有挑战性的课题。在传感器网络中, 由于节点计算的能量消耗要远远小于节点间的通信所消耗的能量, 因此网内的数据聚集成为延长网络生命周期的重要手段, 其核心思想是尽可能地减少节点间的通信消耗, 数据在转发之前进行预处理, 减少传输数量及数据包大小, 从而实现能量的高效利用。文献[2]提出了一种基于管道的网内数据聚集 TAG(Tiny Aggregation)算法, 该算法从叶子节点到根节点逐层进行聚集操作, 减少能量消耗。文献[3]提出了一种减少冗余广播的调度方案, 能避免广播风暴, 减少节点间的通信消耗, 独立地传输调度, 避免冲突, 减小传输延迟。文献[4]提出了一种基于剩余能量的预测模型, 通过减少节点同汇聚节点的汇报次数, 减少通信。复杂的环境及低廉的价格使得传感器节点容易受到外界环境的影响而产生噪音或者失效, 从而报告错误的读数。文献[5]认为空间关系临近的节点往往会产生相似的数据, 利用空间关系可以进行错误数据的判断, 通过邻居节点投票的方式(Voting)决定数据是否正确。

本文提出一种基于最小覆盖集的 WSN 数据聚集算法(MCSSDA), 该算法在考虑能量均衡的前提下, 寻求一个能覆盖整个网络的最少节点集合, 即最小覆盖集。

2 网络定义

网络被模拟为一个连通图 $G=(V, E)$, 其中, V 表示网络中的节点集; E 表示无线通信链路集。假定网络中只有一个汇聚节点 Sink, 传感器节点是静态的, 并以多跳的方式向 Sink 转发数据, 各感知节点具有的共同属性如下:

(1)每个节点都有一个 ID 进行唯一识别。

(2)每个节点维护一个一跳的邻居节点集, 其中包含了邻居节点的 ID, 剩余能量。

(3)每个节点同其一跳的邻居节点交换邻居信息, 从而每个节点获知其所有 2 跳的邻居信息。

(4)每个节点均维护一个冗余数据集和一个错误数据集, 用于存储聚集过程中的数据。

从以上属性可以看出, 每个节点根据其 1 跳和 2 跳的邻居信息, 可以获知其距离 Sink 的最短距离, 即最小跳数。

3 基于最小覆盖集的 WSN 数据聚集算法

3.1 基本定义

假设 x 是传感器网络中的任意节点, $hop(x)$ 表示 x 距离 Sink 最短路径的长度, A 表示 x 的一跳邻居节点集, B 表示 x 的 2 跳邻居节点集。设: $AL=\{i|i \in A, hop(i)=hop(x)+1\}$, $BL=\{i|i \in B, hop(i)=hop(x)+2\}$, 则 $AL \subseteq A$ 、 $BL \subseteq B$ 。节点 x 的最小覆盖集 $MCS(x)$ 是 AL 的最小子集。 BL 表示如下:

基金项目: 国家自然科学基金资助项目“基于信任管理的无线传感器网络高可靠技术研究”(61063040)

作者简介: 吴 旦(1987-), 男, 硕士研究生, 主研方向: 无线传感器网络; 王改云, 教授; 李小龙, 副教授、博士

收稿日期: 2011-07-08 E-mail: wugex123@sina.com

$$BL = \bigcup_{y \in MCS(x)} N(y)$$

其中, $N(y) = \{z | y \in AL, hop(z) = hop(x) + 2, e(y, z) \in E\}$ 。

3.2 最小覆盖集查找算法

最小覆盖集查找算法的基本思想是: 每个节点从其 AL 中按照 ID 升序的顺序挑选一个未被选取的节点, 判断该节点所覆盖的邻居节点数包含 BL 中未被覆盖的节点个数, 同 AL 中的其余节点进行对比, 寻求最多; 考虑到网络能量的均衡问题, 假定 E_x 表示节点 x 的当前剩余能量, $E_{threshold}$ 表示加入最小覆盖集的阈值, 即 $E_x > E_{threshold}$, 当且仅当同时满足上述 2 个条件的节点才能加入最小覆盖集中。随着时间的推移, $E_{threshold}$ 阈值应随着网络能量的减少而动态的发生变化, 以适应算法的需要。该算法流程如图 1 所示。

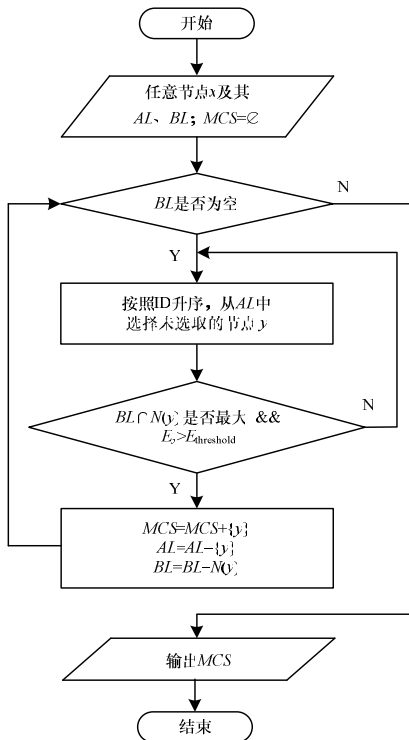


图 1 最小覆盖集查找算法的流程

图 2 给出了一个查找最小覆盖集的例子。假定节点 0 表示 Sink, Sink 的 $AL = \{1, 2, 3, 4, 5\}$, $BL = \{6, 7, 8, 9, 10, 11\}$, $MCS(0) = \{3, 1, 4\}$; 节点 3 的 $AL = \{8, 9, 10\}$, $BL = \{14, 15, 16, 17\}$, $MCS(3) = \{8, 9\}$ 。

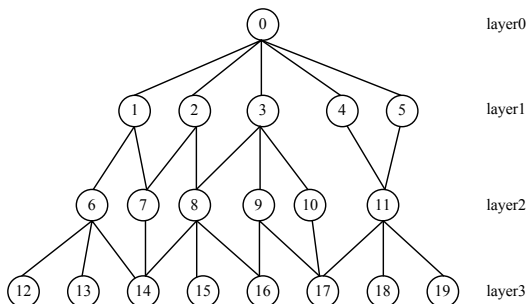


图 2 最小覆盖集查找示例

3.3 聚集树的构造

构造聚集树的基本思想: Sink 节点首先调用最小覆盖集查找算法, 获取 MCS , 然后向外广播聚集树建立消息 ATC , 该消息中包含了 MCS 的节点列表, 每个收到 ATC 消息的节点执行最小覆盖集转发算法, 图 3 给出了该算法的流程。

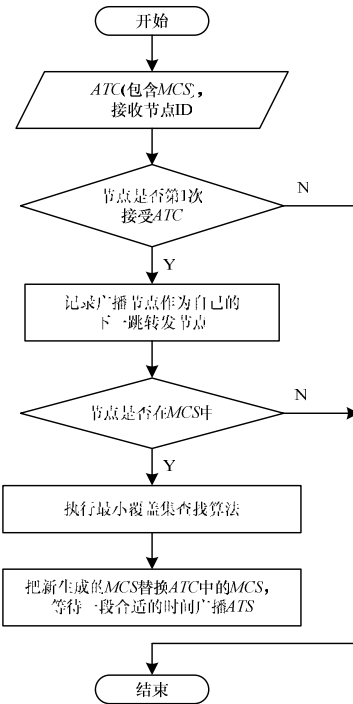


图 3 MCS 转发算法流程

首先判断该节点是否第 1 次接收 ATC , 如果不是, 则终止算法; 否则, 记录广播节点作为自己的下一跳转发节点, 接着判断该节点是否属于 MCS 中的节点, 如果属于, 计算自己的 MCS , 继续广播, 否则, 终止算法, 重复上述步骤, 直至在每个层都产生一个 MCS , 且以 Sink 为根的树。其中, MCS 中的节点广播次序按照查找算法中依次加入的顺序, 即覆盖 BL 中节点数多的, 排在前面。每广播一个 ATC , 等待一段合适的时间, 再广播下一个 ATC , 以避免有共同邻居的多个 MCS 节点同时广播所引发的碰撞, 造成“孤立”节点; 此外, 当节点收到相同的来自上层节点的 ATC 信息, 则选择较早的广播节点作为自己的下一跳转发节点。如图 2, 节点 3 的 MCS 中有一个共同的邻居节点 16, 执行算法后, 节点 16 选择了节点 8 作为下一跳转发节点。

3.4 数据聚集算法

本文采用读向量相似性的方法, 在中间节点去除冗余及错误信息, 把正确的数据以一个读向量的形式转发给下一个中间节点。传感器节点 N_i 的读向量是由一个缓存窗口 Δt 内的一系列读数组成, N_i 读数可以表示为 $R_i = \{x_i(t - \Delta t + 1), x_i(t - \Delta t + 2), \dots, x_i(t)\}$, 其中, $x_i(t)$ 表示 t 时刻; 节点 N_i 的采样值。节点 N_i 在一个缓存窗口 Δt 内的读数代表一个读向量。

正常数据是有趋势且渐变的, 错误数据会使当前的数据变化趋势发生改变或者会使当前的数据发生跳跃。为捕捉这种变化, 本文利用读向量相似系数 ρ 判断冗余及错误数据。

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sqrt{D(X)D(Y)}}$$

其中, $E(X)$ 、 $D(X)$ 分别表示读向量 X 的数学期望和方差; ρ_{XY} 的取值范围为 $[-1, 1]$, $|\rho_{XY}|$ 越大, 表示 X 与 Y 的误差越小, 冗余性越高, 反之, 则其中至少有一个为错误数据。因而, 中间节点进行数据聚集时首先设定一个阈值 θ , 当 $\rho_{XY} > \theta$ 时, 则把数据放入冗余数据集, 否则放入错误数据集, 算法结束后, 丢弃错误数据集, 对冗余数据集求平均值, 并将结果转发给下一个聚集节点。

3.5 聚集树的维护

聚集树建立之后, 每个节点均知道其下一跳转发节点, 在 MCS 中的每个节点也知道在其子树中的所有 MCS。当节点监测到异常数据, 如果它不是 MCS 中的节点, 则立即将数据传输给其上一层的 MCS 中的节点, 否则, 等待一段合适的时间聚集接收的数据, 再将聚集结果传输给上层的 MCS。

传感器网络的生命周期是以第 1 个节点的失效来衡量, 相比非中间节点, 中间聚集节点需要消耗更多的能量, 为了保持网络能量的均衡消耗, Sink 节点应该根据应用的需要, 周期性的广播 ATC, 重建 MCS 聚集树, 动态的变更阈值 $E_{threshold}$, 让能量低于阈值的节点退出中间节点, 延长网络的生命周期。

4 实验结果与分析

本文实验采用 Matlab 进行设计。100 个传感器节点被随机部署在一个单位面积的正方形区域内, 每个节点有 2 个~5 个邻居, 假定线路都是可靠的, 不考虑线路不连通的情况。该实验把 MCSDDA 算法同 GIT、SPT 及 AC^[6]3 种算法进行了对比, 分别比较了在节点通信半径 R 和源节点个数 K 变化的情况下, 3 种算法的平均通信次数的差异, R 的变化范围为 0.15~0.45, K 的变化范围为 1~15, 仿真过程同时进行, 结果如图 4 所示。可以看出, GIF、SPT 及 MCSDDA 算法相对 AC 算法, 在距离 Sink 跳数较多的情况下, 能节省 34%~45% 的能量。

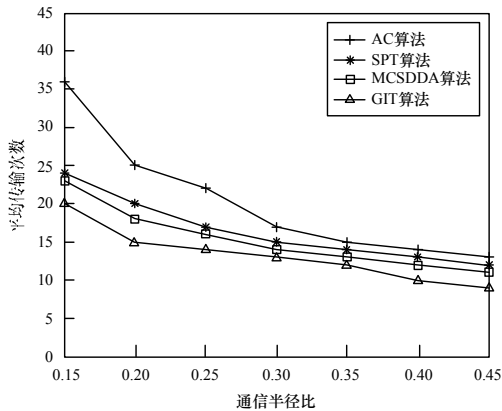


图 4 4 种算法的通信半径比较

4 种算法的源节点个数比较如图 5 所示, 随着源节点个数的增加, 传输消耗均呈上升趋势。其原因在于源节点的增加, 将有更多的数据需要传输, 增加了平均的传输次数。

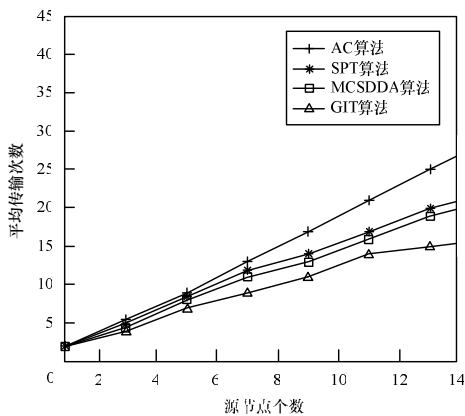


图 5 4 种算法的源节点个数比较

可以看出, MCSDDA 的平均传输次数比 SPT 和 AC 少,

比 GIT 多。虽然 GIT 的性能更好, 但是在不知道全网信息的情况下, 很难实现分布式算法, 仅通过局部信息, 一个源节点不可能知道距离 Sink 最近的转发节点。然而, MCSDDA 算法只需要了解其 1 跳和 2 跳的邻居信息就能很容易实现。MCSDDA 算法比 SPT 具有更好的性能, 因为在 MCSDDA 算法中, 树的每一层有更多距离 Sink 最短路径的节点。

同时, 与 TAGVoting 算法在错误数据检测进行了对比, 在每轮采样数据中随机地将一些正常数据改变为错误数据, 使错误数据达到一定的比例。2 种算法的错误率比较如图 6 所示。可以看出, MCSDDA 算法在错误检测率方面优于 TAGVoting 算法, 这是因为 Voting 算法仅仅以值的大小来判断节点读数的相似性, 而 MCSDDA 算法不仅仅考虑值的因素, 还考虑了趋势的变化。

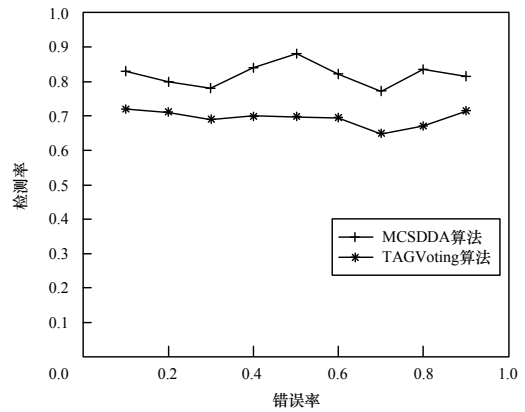


图 6 2 种算法的错误率比较

5 结束语

本文提出一种基于最小覆盖集的 WSN 数据聚集算法, 实验结果表明, 与同类算法相比, 本文算法性能较优, 实现的网络功耗低、融合度高。在进行数据聚集时, 如何选择较为合理的聚集时间是下一步的研究重点。

参考文献

- [1] 孙利民, 李建中, 陈渝, 等. 无线传感器网络[M]. 北京: 清华大学出版社, 2005.
- [2] Madden S, Franklin M, Hellerstein J, et al. TAG: A Tiny Aggregation Service for Ad Hoc Sensor Networks[C]//Proc. of the 5th Symp. on Operating Systems Design and Implementation. Boston, USA: ACM Press, 2002.
- [3] Sheu Jang-Ping, Hung Pei-Kai, Hsu Chih-Shun. Scheduling of Broadcasts in Multihop Wireless Networks[M]//The Handbook of Ad Hoc Wireless Networks. Boca Raton, USA: CRC Press, 2003: 483-495.
- [4] 周平, 张胜, 舒坚. 基于预测模型的 WSN 节点能量融合机制[J]. 计算机工程, 2010, 36(1): 110-111.
- [5] Krishnamachari B, Iyengar S. Distributed Bayesian Algorithms for Fault-tolerant Event Region Detection in Wireless Sensor Networks[J]. IEEE Trans. on Computers, 2004, 53(3): 241-250.
- [6] Wen Yean-Fu, Lin F Y S, Wang Hung-Shi. A TDMA-based Scheduling and Routing Algorithm for Data-centric Wireless Sensor Networks[C]//Proc. of Global Telecommunications Conference. Washington D. C., USA: [s. n.], 2007.

编辑 顾姣健