

文章编号:1001-9081(2010)09-2509-03

基于次序编码的无线传感器网络数据融合算法

侯飞, 刘群

(重庆邮电大学 计算机科学与技术学院, 重庆 400065)

(cucacula@139.com)

摘要: 数据包压缩为无线传感器网络的数据融合提供了一种简单、有效的解决方案。针对一种基于次序编码的数据包压缩技术进行了研究, 分析了其处理过程过于集中和复杂这两个局限性, 提出一种数据融合新算法(DCOA)。算法建立了新的实现流程, 设计了新的映射表数据结构, 通过分布式动态处理, 分解降低了实现的复杂性, 平衡了节点的能量负载, 大幅提高了数据融合效率。通过仿真实验对算法的优越性进行了验证。

关键词: 无线传感器网络; 次序编码; 数据融合; 映射

中图分类号: TP393 **文献标志码:** A

Data aggregation algorithm based on order coding for wireless sensor network

HOU Fei, LIU Qun

(College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Data packet compression provides a simple and effective solution for the data aggregation of Wireless Sensor Network (WSN). Research was conducted on a data packet compression technology which was based on order coding. Two limitations were analyzed, which were too concentrated and too complex, and a new distributed coding by ordering algorithm (DCOA) was put forward. The new algorithm established a new implementation flow, designed a new mapping table data structure, reduced the complexity of implementation and balanced the energy load of the nodes through the distributed dynamic processing, and significantly increased the efficiency of data aggregation. The superiority of the algorithm is verified through simulation experiments.

Key words: Wireless Sensor Network (WSN); order coding; data aggregation; mapping

0 引言

无线传感器网络(Wireless Sensor Network, WSN)^[1]是一种由大量传感器节点通过无线通信方式构成的自组织、多跳式的网络系统, 它通过实时监测、感知和采集网络监测区域内的各种目标对象的信息, 将虚拟的信息世界与客观的物理世界联系在一起。数据融合^[2-3]是 WSN 的重要研究领域之一。数据融合技术通过合并多个数据源产生的数据, 去除冗余信息, 能够有效地减少网络中的数据传输量, 从而节省传感器节点的能量, 延长 WSN 的生命周期。

当前, 对 WSN 数据融合的研究主要从两方面展开: 一是基于网络层的路由级^[4]融合; 二是基于应用层的数据级融合。路由级融合是指在设计路由协议时要为数据融合提供支持, 为相关融合方法提供更高的融合效率和更好的融合实现。目前几个比较流行的 WSN 路由协议都有这样的设计。LEACH 协议^[5]对网络进行了分簇, 然后通过一定的选举算法选举每个簇的簇头, 在簇头对该簇成员的数据进行融合处理; PEGASIS 协议^[6]利用贪心算法将网络中各节点连接成一条单链, 每个节点在收到相邻节点传递的数据后要与自己的数据进行融合处理, 最终由选举的首领节点将数据传送给 sink 节点; Directed Diffusion 协议^[7]则通过缓存机制来抑制已经转发过的数据, 在兴趣扩散和数据传递两个过程中均可进行融合处理。

数据级融合是指对原始数据和数据包进行综合、分析, 找出其中的相关性, 建立具体的融合实现。数据级融合可分为基于查询的数据融合和基于压缩的数据融合。基于压缩的数据融合又可分为依赖于语义的有损融合和独立于语义的无损融合。文献[8]介绍了一种基于查询的数据融合方法, 它将整个 WSN 看成一个分布式的数据库, 通过一个 WSN 的通用数据融合接口来实现对数据库的查询, 查询操作的过程就是数据融合的过程; 文献[9]介绍了一种分布式数据压缩技术, 它设计了一个支持低运算开销的编码器, 编码器将原始数据 x 和 y 中的一方进行编码压缩, 再将压缩后的数据 $f(x)$ 和 y 发送给 sink 节点, 然后通过解码还原出原始数据。该方法是独立于语义的无损融合, 它的一个必要前提是数据 x 和 y 之间存在某种关联性; 文献[10]在一种数据漏斗的路由基础上, 提出一种次序编码(Coding by Ordering)的数据包压缩技术, 实现了与文献[9]中方法类似的、独立于语义的无损数据融合。本文将对该技术以及该技术的局限性进行分析, 并提出一种 DCOA(Distributed Coding by Ordering Algorithm)。

1 次序编码技术及其局限

1.1 次序编码技术

次序编码技术最初建立在文献[7]的数据漏斗路由协议上, 该协议将 WSN 分成若干个目标区域, 然后面向目标区域对兴趣包进行定向泛洪, 任一节点在收到兴趣包后检查自己

收稿日期:2010-03-29;修回日期:2010-06-03。 基金项目:重庆市自然科学基金资助项目(CSTC2007BB2386)。

作者简介:侯飞(1985-),男,安徽宿州人,硕士研究生,主要研究方向:计算机网络; 刘群(1969-),女,江西南昌人,教授,博士,主要研究方向:智能网络。

是否处于目标区域中。如果不是,则更新兴趣包的代价域,然后继续对兴趣包泛洪;如果是,则标记自己为边界节点,然后修改兴趣包,再在其所在区域内对兴趣包进行泛洪。为避免边界节点的高能耗,可以利用一定的选举算法对边界节点进行轮换。如图1所示,数据信息从源节点经边界节点到sink节点的路由路径像一个倒置的漏斗,故称之为数据漏斗。漏斗对数据的过滤过程就是数据融合的过程。它是一个网络层的路由级融合。

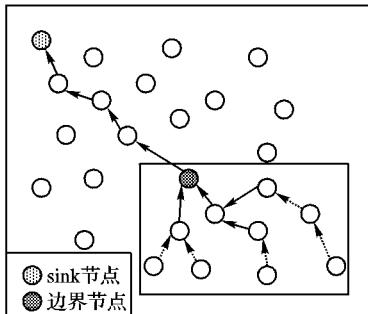


图1 数据漏斗算法的路由模型

假设各源节点的数据包信息的次序对于应用并不重要时,那么在传输的过程中,通过对数据包ID的次序进行重新组织,利用特定的次序向接收者传递其所蕴含表达的相应附加信息,实现部分数据包值信息的压缩,达到数据融合的目的。在数据漏斗算法中,选取边界节点作为进行数据包压缩的融合点。边界节点在将所在目标区域内的各源节点的初始数据包合并成一个超包的过程中进行压缩融合处理,此为编码阶段。编码阶段被压缩的值信息在sink节点进行解码时将会得以完全还原。编码和解码是完全对称的。

编码和解码都必须严格遵循一个统一的映射表,它反映了属性“未压缩节点ID的排列次序”到属性“压缩节点的值信息”的一个双射。表1反映了ID分别为1、2、3的数据包通过 $3! = 6$ 种可能的排列次序,映射出ID为4的数据包{0,1,...,5}共6个整数取值。ID为4的数据包从而被压缩。

表1 次序编码的映射

排列次序	整数值
1 2 3	0
1 3 2	1
2 1 3	2
2 3 1	3
3 1 2	4
3 2 1	5

为了确保有足够的排列次序来映射数据包的值信息,参与合并的节点数目需要满足一定的条件。设n为边界节点需要处理的数据包的数目,k为每个传感器节点数据域的值信息的取值范围的宽度,d为节点ID的取值范围的宽度,l为可被最大限度压缩的数据包数目,有:

$$(n-l)! \geq (d-n+l)^l \cdot k^l \quad (1)$$

又考虑到每个数据包的节点ID是独一无二的,因此有:

$$(n-l)! \geq C_{d-n+l}^l \cdot k^l \quad (2)$$

1.2 两个局限

该方法的局限主要有两点,即过于集中和过于复杂。算法集中在边界节点进行,直接导致编码和解码的复杂度过高,进而使边界节点产生较高的能量消耗和较高的处理时延。另外,该算法在初始化时即确定了编码和解码的规模,每一轮因

时延或死亡而导致未发送数据的节点均被冗余地计入了映射表,导致映射表过于庞大,同时大大增加了算法的复杂度。

2 DCOA

2.1 本文算法的实现过程

针对次序编码的两个局限性,本文提出一种分布式动态处理的解决方案,将编码工作由边界节点分布到广大中间节点来动态地处理实现,将解码工作集中在稳定、可靠、处理能力强的sink节点来实现。编码和解码工作均遵循事先制定好的统一的映射规则。

在编码阶段,局部区域内邻近的n个节点的数据包进行合并,合并后形成的新数据包将由映射规则,根据l个节点的值,对 $n-l$ 个节点的数据包进行排列,从而实现对l个数据包的值信息的压缩。算法的具体执行流程如下。

1) 由于不同节点的传输具有不同的时延,因此融合节点选取包括自己在内的时延最短的l个节点作为待压缩节点,根据它们的值信息来映射所需将其表达的节点排列,然后等待所需 $n-l$ 个节点的陆续到达。

2) 考虑到对所需 $n-l$ 个节点的等待会带来较长的时延,因此有必要设定一个时间戳上限。若在上限范围内,到达节点的数目大于等于 $n-l$,则转入步骤4)执行;否则,到达节点的数目小于 $n-l$,则启动向下兼容模式,进入步骤3)执行。

3) 根据实际到达的节点数目,缩减调整n为 n' ,l为 l' ,这里 $n' < n, l' < l$ 。当 $0 < l' < l$ 时,修改标记域信息,以说明该轮压缩是兼容模式下进行的,然后更新n和l的值,进入步骤4)执行;否则,当 $l' = 0$ 时,说明没有节点可被压缩,放弃并回滚已执行步骤,根据相应路由协议所选路径继续传输所有节点的全部初始数据包,直至下一融合节点,再进入本文算法流程。

4) 排列这些节点的相应顺序,然后合并成一个中间包(mid-packet)。对于新生成的中间包,增加一个新的mid-ID域作为其ID,取值为包内 $n-l$ 个节点ID的平均值。将中间包作为和初始包节点平行的对象,参与新一轮映射压缩。这时,中间包只能作为排列对象,而不能作为压缩对象。可压缩的对象只能是初始包节点。

在解码阶段,sink节点根据映射规则,将超包层层分解,还原出各节点的原始数据,实现数据信息的无损融合。

2.2 映射表的设计

映射表是算法的核心关键,是最重要的数据结构。在原算法中,它主要由排列次序和值信息两个属性组成,其中排列属性的值是数据包节点ID值的从小到大的次序。由于本文算法是分布式动态设计的,而且由于中间包的引入,原算法的以节点ID作为排列单元的方法不再可行。这里,本文在解码端设计了一个转换器。它的功能是将层层排列的初始包和中间包的节点ID转换为具有一般意义的特定字符序列,例如{a,b,c,...},再由它们的次序和值信息一一对应,从而实现整个映射过程。

在表2中,映射表增加了一个转换器属性,而在排列属性中,将ID为3的初始包换成了ID为x的中间包,表示由ID分别为若干值的初始包合并成的中间包。

新的解码映射方案摒弃了静态而死板的ID数字排列,采用排列的抽象顺序语义来映射节点值信息,从而使映射表更加动态灵活,更具适应性。

表2 改进后的次序编码的映射

排列次序	转换器	整数值
1 2 x	a b c	0
1 x 2	a c b	1
2 1 x	b a c	2
2 x 1	b c a	3
x 1 2	c a b	4
x 2 1	c b a	5

2.3 本文算法需满足的条件

以上对 DCOA 的实现过程和映射表的数据结构两方面进行了描述,下面将讨论在算法改变后,满足的条件将发生了怎样的变化。

1) 初始模型。

由于本文算法的解码和编码过程都是动态执行的,因此当轮未传输数据的节点将不会被计入算法的考虑范畴,有 $d = n$,于是 $C_{d-n+l}^l = C_l^l = 1$,可知:

$$(n - l)! > k^l \quad (3)$$

2) 对 n 和 l 的扩展。

在编码阶段的算法流程3)中,算法有可能进入兼容模式运行,即 n 更新为 n' , l 更新为 l' 。映射表将为不同的 $n - l$ 加入相关的映射行来实现扩充,以保证算法运行。这时

$$(n' - l')! > k^l \quad (4)$$

3) 对 k 的扩展。

截至目前,本文算法只描述了节点的数据域为一个属性值的情况,而实际应用中,传感器节点传输的节点信息常为多个属性值。例如,在森林气象传感器网络中,除了温度属性外,还可以有烟雾浓度属性、空气湿度属性等。设温度属性的域为 k_1 ,烟雾浓度属性的域为 k_2 ,则 $k = k_1 \cdot k_2$ 。则需满足:

$$(n - l)! > k_1^l \cdot k_2^l \quad (5)$$

3 仿真实验

DCOA 在 WSN 的应用层实现,算法的实现过程和相应的数据结构均封装在节点类中,通过接口与网络层的路由协议通信。因为不同的路由协议所选取的数据融合点不同,所以本文算法的性能也将随之发生较大的变化。本文选择 Directed Diffusion 路由协议来进行仿真实验。同时选取经典的路由级数据融合算法 LEACH 与本文算法做比较。实验平台选择 NS2,一些基本实验环境参数如表3。

表3 基本实验环境参数

参数	值
实验面积	500 m × 500 m
单个节点初始能量	1 J
数据包的大小	100 B
每轮测试节点数目	150, 300, 450, 600, 750
MAC 协议	802.11b
发射和接收功率	0.035 mW

为了简化实验时的算法实现,本文取消映射表的兼容模式,属性域的个数取为 1。仿真时间由初次实验时中间节点(非源节点)能量耗尽的用时来确定,后续实验均保持一致。主要测试两个指标,即数据包(初始包)的平均传输数目和网络的平均能耗。这里,数据包的平均传输数目是指每轮 sink 节点接收到的带有数据包值信息的节点数目的加权平均值,它体现了当前网络的数据融合率;而网络的平均能耗则体现

了网络的生命周期。

从图2可以看出,采用了本文算法后的节点平均能量消耗一直略低于未采用本文算法时的情况,而 LEACH 算法的能量消耗在节点数较少时低于本文算法,在节点数越来越多时高于本文算法,整体表现更平缓;从图3可以看出,随着节点数量的增加,采用了本文算法后的数据包平均传输数目同比有明显下降,数据融合率则相应明显提高,符合算法预期,而 LEACH 算法在融合效率上提升并不明显。

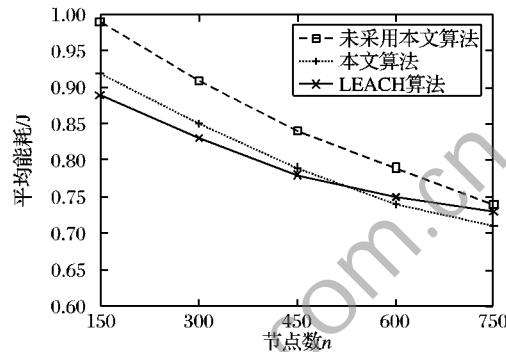


图2 能量的平均消耗

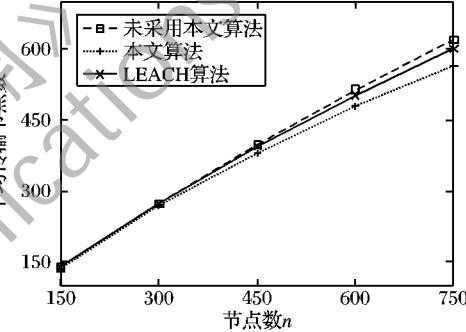


图3 数据包的平均传输数目

4 结语

本文首先介绍了一种次序编码的数据融合技术,然后针对它的局限性进行了研究分析,在算法的实现过程和映射表的数据结构两方面加以改进,提出了一种分布式动态处理的次序编码的数据融合算法。与原算法相比,本文算法的复杂性得以降低,灵活性得以增强,综合效能得以大幅提升。通过实验为本文算法寻找最佳的路由协议作为支撑以及如何更加高效地实现映射有待进一步研究。

参考文献:

- [1] 任丰原,黄海宁,林闯. 无线传感器网络[J]. 软件学报, 2003, 14 (7): 1282 – 1291.
- [2] YU BO, LI JIANZHONG, LI YINGSHU. Distributed data aggregation scheduling in wireless sensor networks [C]// Proceedings of IEEE International Conference on Computer Communications. Washington, DC: IEEE Computer Society, 2009: 2159 – 2167.
- [3] SHWE H Y, JIANG XIAOHONG, Horiguchi S. Energy saving in wireless sensor networks [J]. Journal of Communication and Computer, 2009, 6(5): 20 – 27, 45.
- [4] IQBAL M, GONDAL I, DOOLEY L. An energy-aware dynamic clustering algorithm for load balancing in wireless sensor networks [J]. Journal of Communications, 2006, 1(3): 10 – 20.
- [5] HEINZELMAN W R, CHANDRAKASAN A, BALAKRISHNAN H. Energy-efficient communication protocol for wireless microsensor networks [C]// Proceedings of the 33rd International Conference on System Sciences. Hawaii: IEEE Computer Society, 2000: 1 – 10.

(下转第 2525 页)

身需要消耗少量的能量。图5中算法的节能效果比图4要好,这是由于模拟的任务比较多,算法的优势越明显的原因。总体上算法的节能效果在40%~82%左右,较大限度上降低了系统能耗。

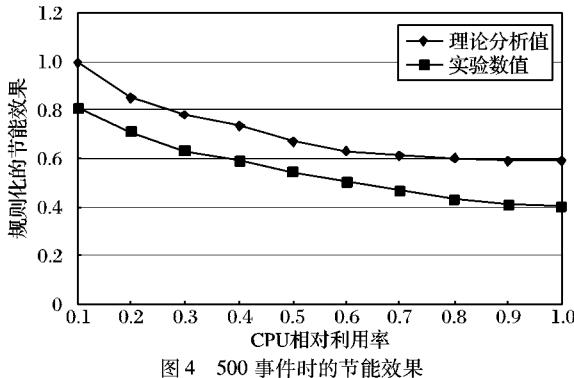


图4 500事件时的节能效果

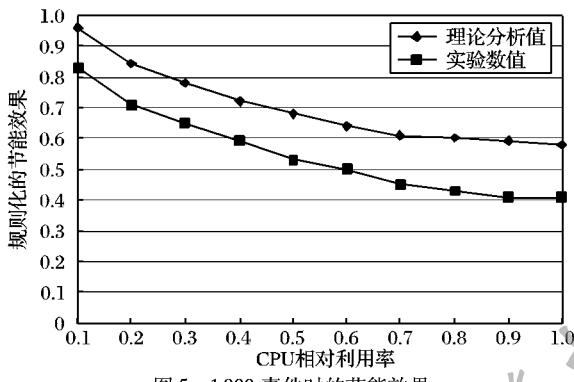


图5 1000事件时的节能效果

图6中模拟周期性的任务模式,验证了H-DVS算法的节能效果。

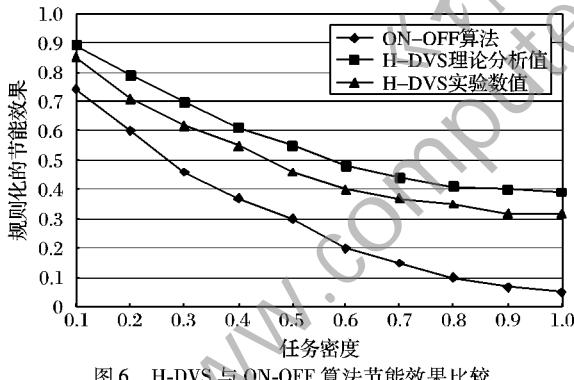


图6 H-DVS与ON-OFF算法节能效果比较

由图6可知,总体上H-DVS算法的性能要比ON-OFF电压调节算法表现得好,与理论分析一致。ON-OFF电压调节算法约比H-DVS算法要低0.2左右,且任务密度越大与

H-DVS算法的差距越大,这可能是由于频率的进行频率/电压切换所致。同时,由于实际的频率/电压调节等级有限且CPU频率调整本身需要消耗少量的能量,实际节能率比理论分析值要略低。

4 结语

本文针对计算复杂度较高的WSN应用背景和WSN应用普遍存在的任务模式,研究一种相对更通用的、基于混合任务模型的DVS算法H-DVS。H-DVS能在任务相对期限没有限制的情况下,与EDF调度算法结合,支持周期任务和零散任务同时存在的混合任务模式。H-DVS算法根据CPU的工作负载,由频率调节因子对CPU进行实时电压和频率调节。理论分析与实验数据表明H-DVS算法能在保证任务的实时性要求下降低能耗,与理想状态下的ON-OFF电压调节算法相比,仍然具有较好的节能效果。

参考文献:

- [1] GHARAVI H, KUMAR S P. Special issues on sensor networks and applications [J]. Proceedings of the IEEE, 2003, 91(8): 1151–1153.
- [2] 田丰民,陈向东,张传武.无线传感器网络动态功率管理方法[J].传感器技术,2005,24(11): 33–35.
- [3] ZHANG F, CHANGSON S T. Processor voltage scheduling for real-time tasks with non-preemptible sections [C]// Proceedings of the 23rd IEEE International Real-Time Systems Symposium. Washington, DC: IEEE Computer Society, 2002: 235–245.
- [4] SINHA A, CHANDRAKASAN A. Dynamic power management in wireless sensor networks [J]. IEEE Design and Test of Computers, 2001, 18(2): 62–74.
- [5] SIMUNIC T, BENINI L, ACQUAVIVA A, et al. Dynamic voltage scaling and power management for portable systems [C]// Proceedings of the 38th Annual Design Automation Conference. New York: ACM Press, 2001: 524–529.
- [6] 张承刚,徐成.适合无线传感器网络的自适应动态电压调节算法[J].计算机应用研究,2008,25(12): 3800–3803.
- [7] LIU C L, LAYLAND J W. Scheduling algorithms for multiprogramming in a hard real-time environment [J]. Journal of the ACM, 1973, 20(1): 46–61.
- [8] 李国徽,江德平.基于流预测的无线传感器网络动态功率管理[J].华中科技大学学报:自然科学版,2007,35(7): 27–30.
- [9] 陈欢,陈向东,胡黎黎,等.基于小波的DVS在无线传感器网络中的应用[J].计算机应用研究,2007,24(8): 262–263.
- [10] QUAN G, HU X. Energy efficient fixed-priority scheduling for real-time systems on variable voltage processors [C]// Proceedings of the 38th Annual Design Automation Conference. New York: ACM Press, 2001: 828–833.

(上接第2511页)

- [6] LINDSEY S, RAGHAVENDRA C S. PEGASIS: Power efficient gathering in sensor information systems [C]// Proceedings of IEEE Aerospace Conference. Washington, DC: IEEE Computer Society, 2002: 1125–1130.
- [7] INTANAGONWTWAT C, GOVINDAN R, ESTRIN D, et al. Directed diffusion for wireless sensor networking [J]. IEEE/ACM Transactions on Networking, 2003, 11(1): 2–16.
- [8] MADDEN S, SZEWCZYK R, FRANKLIN M J, et al. Supporting aggregate queries over Ad-Hoc wireless sensor networks [C]// Proceedings of 2002 IEEE Workshop on Mobile Computing Systems and

- Applications. New York: IEEE Press, 2002: 49–58.
- [9] CHOU J, PETROVLC D, KNANNA R. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks [C]// Proceedings of 2003 IEEE International Conference on Computer Communications. San Francisco: IEEE Press, 2003: 1054–1062.
- [10] PETROVIC D, SHAH R C, RAMCHANDRAN K, et al. Data funneling: Routing with aggregation and compression for wireless sensor networks [C]// Proceedings of the 1st IEEE International Workshop on Sensor Network Protocols and Applications. Anchorage: IEEE Press, 2003: 156–162.