ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 31,No. 6,Nov. 2016,pp. 1192－1198
DOI:10.16337/j.1004-9037.2016.06.015

http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:＋86-025-84892742

传感器网络中基于分布式压缩的数据聚合方法

杜淑颖[1,2]　丁世飞[1]

(1. 中国矿业大学计算机学院,徐州,221000;2. 徐州生物工程职业技术学院信息工程系,徐州,221000)

摘　要: 针对基于压缩数据采集的数据聚合需要高效的路由转发树协议,以便更好地采集到从传感器节点到 sink 节点的编码数据,提出了一种新型高效节能的分布式压缩数据收集方法。该方法中每个传感器节点均可独自寻找父节点,并构建一部分路由树,无需利用中心节点来构建所有转发树,从而允许每个传感器节点对转发树的构建和维护做出局部决策。仿真实验结果表明,相比传统的压缩数据收集方法,新方法的复杂性较低并且开销降低近 50%。

关键词: 传感器网络;压缩数据;数据聚合; 分布处理

中图分类号: TP393　　**文献标志码:** A

Aggregation Scheme Based on Distributed Data Compression in Sensor Networks

Du Shuying[1,2], Ding Shifei [1]

(1. School of Computer Science and Technoloy,China University of Mining and Technology, Xuzhou, 221000, China;
2. Department of Information Engineering,Xuzhou Vocational College of Bioengineering, Xuzhou, 221000, China)

Abstract: Data aggregation based on compressed data collection need efficient routing forwarding tree protocol to gather the coded data corresponding to the sensor nodes to the sink node effectively. A new distributed compressed data collection method with high efficiency and energy saving is presented. Each sensor node can be found alone in the part of its parent and able to build a routing tree, without using the center node to build all forwarding tree, which allows each sensor node to make local decisions on the forwarding tree construction and maintenance. Simulation results show that the complexity of the new method is lower compared with that of the traditional method, and the cost reduces nearly 50%.

Key words: WSN; compressed data; data aggregation; distributed processing

引　　言

无线传感器的通用性使其备受人们的关注,并且广泛用于健康监测、环境、交通和重要基础设施等监测多个领域。在这些应用中,无线传感器通常需通过多跳路径将感知数据定期发送给中心节点(如 sink 节点),然后中节点对这些感知数据进行处理。在部署了这些能量有限的传感器后,一般情况下它们得到的维护很少或根本得不到维护。因此以最高效节能的方式采集数据对无线传感器网络的使用寿命起着非常重要的作用,因此这种方式引起了人们的广泛研究。其中,将读出数据采集传输到 sink 节点时,采用的最有效的方法是进行数据聚合[1],数据聚合指对传感器采集的总数据进行计算得出统计度

基金项目: 国家自然科学基金(61379101)资助项目。
收稿日期: 2015-11-27;**修订日期:** 2016-01-20

量(例如平均数和总数),避免传输所有读出数据时产生的昂贵费用(通常是将这些数据从大量的传感器传输到远处的 sink 节点)。以压缩感知理论为基础的数据压缩是数据聚合[2]的另一种形式,由于这种形式具有能够降低全球通信成本同时不会产生大量计算或传输开销的能力,因此近来已成为人们关注的焦点。借助于压缩数据采集,sink 节点只需接收到所有读出数据中少数的加权(编码的)总和(m),而不是接收到所有来自网络的读出数据(例如 n)。此外,只要利用某种稀疏或正交变换域方法[3]能够对这些读出数据进行转换或压缩,那么 sink 节点就可将这些读出数据恢复(解码)成原始数据,此时,$m=O(k\log n)$,式中,k 为变换域中这种方法得到的数据稀疏度表征。因此,最近压缩数据采集已经引起了研究者的关注,由于这种技术可节省大量的能源,因此,沿着一条给定的路径将通信工作开销分摊到所有的传感器上可延长网络的使用寿命并实现负载平衡。本文研究由 n 个传感器构成的网络中的数据聚合问题。针对该问题,文献[4]针对单一聚合树提出了高效的启发式算法和近似算法,这两个方法的数据聚合效果较好,但是均局限于单一树状结构,算法适用性较差。文献[5]以最高效利用能源的方法解决了优化构建转发树的问题,以传送加权总和。考虑到问题的复杂度,这些作者提出了可获得近似最优解的算法,该算法通过增加加权系数来优化性能,但是其复杂度明显增加,网络的整体能耗会增加,网络的生存时间会降低。文献[6]将读出数据预处理到 sink 节点;且原始数据是通过收集稀疏映射恢复的,在这些稀疏映射中,每个映射都对应一个数据聚合。通过建立转发树采集映射,每个转发树对应一个映射,这种映射是一种包含(从节点处参与映射的)其他数据包的编码(或压缩)数据包。该算法通过构建有效的映射或采集树来收集加权总和,但是没有解决传输成本这个重要问题。本文将采取类似文献[6]的方法,在收集到所有的映射之后,sink 节点就能够通过解决凸优化问题来恢复原始数据。然而,上述文献的工作均假定树集中式构建,这类集中式方法需要得到网络的全部拓扑信息,而获得这些信息的费用可能很高,并且如果拓扑信息发生改变(由于节点移动、电池耗竭和信道损坏等),就很难维护转发树。因此,本文针对构建转发树提出了一种分布式处理方法。与先前的方法相比,这种方法不会产生过多的开销,且在路由过程中每个节点会做出局部决策。

1 系统原理和网络模型

1.1 网络模型

将无线传感网(Wireless sensor network, WSN)描绘成一个连通图 $G=\langle V,E\rangle$,V 为 n 个节点的集合,E 为任意在彼此的通信半径内的两个节点间的链接集合。假定每个传感器均有一个将要传输到 sink 节点的读出数据 x_j(例如温度或湿度),由于并非所有的节点都可直接与 sink 节点相连接,因此节点可能需要通过多跳路径将读出数据转发给 sink 节点。总之,sink 节点需要将每个循环中的所有节点恢复成有关 n 个节点的数据向量 $\boldsymbol{X}$,有

$$\boldsymbol{X}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}} \tag{1}$$

根据压缩传感原理可知,sink 节点可能从节点,而不是 n 原始数据处接收到 m 个映射,其中 $m\ll n$。sink 节点能够通过接收到的 $\boldsymbol{Z}$ 完全恢复原始数据向量 $\boldsymbol{X}$。

$$\boldsymbol{Z}=[z_1,z_2,\cdots,z_m]^{\mathrm{T}} \tag{2}$$

$$z_i=\sum_{j=1}^{n}\phi_{ij}x_j \tag{3}$$

式中:$i=1,2,\cdots,n$。z_i 为在随机样本矩阵 $\boldsymbol{\Phi}_{m\times n}$ 中从节点处得到的传感器读出数据的非零系数加权总和,即矩阵 $\boldsymbol{\Phi}$ 为测量矩阵。矩阵 $\boldsymbol{\Phi}$ 由 m 行(每行对应一个映射)和 n 列(每列对应一个传感器节点)组成。矩阵 $\boldsymbol{\Phi}$ 的列向量与傅里叶变换矩阵 $\boldsymbol{\Psi}$(利用其可获得数据的稀疏表示)[7]从常态分布或高斯分布中随机挑选出。为了使非零系数在矩阵 $\boldsymbol{\Phi}$ 中分布得更加均匀并且每个映射尽可能稀疏,需将矩阵 $\boldsymbol{\Phi}$ 每行中非零的数目选择为$[n/m]$,这样在矩阵 $\boldsymbol{\Phi}$ 的每列就不会出现全零条目[8]。通过特定的算法求解如下的优化(Non-deterministic polynomial, NP)问题可完美重构数据 $\boldsymbol{x}$,则

$$\underset{z}{\arg\min} \| \boldsymbol{\Phi} \boldsymbol{G} \boldsymbol{z} \| = \boldsymbol{\Phi} \boldsymbol{x} \tag{4}$$

1.2 压缩数据采集原理

假设已有网络 G,矩阵 $\boldsymbol{\Phi}$ 和 sink 节点,此时解决基于投影的压缩数据采集问题需要找到 m 个转发树,每个转发树 T_i 对应一个映射并只能从相应节点处采集一个加权总和 z_i,这样可将每个转发树上的链接数目最小化。这一问题不同于寻找最小能耗的聚合树[7]。因为后者是将树根与网络中的所有节点相连接。然而,与利用中继节点[7]构建聚合树相类似,不同之处在于本文提出的问题允许网络内数据编码到达 sink 节点。考虑到本文提出问题[7]的 NP 完整性质,很多先前的研究借助于已验证的近似值提出了很多方法。文献[6]旨在将每个转发树中的链接数目最小化,因此可利用整数线性规划建模构建最优树,这样可将传输每个加权总和的能耗最小化。于是人们针对集中式压缩数据采集问题提出了一种接近性能优化的启发式方法。这类方法从本质上来讲都属于集中式方法,也就是说,在最初时,sink 节点必须通过部署一个多对多泛洪来检索网络中的广泛信息才能完成拓扑搜索,然后利用算法构建转发树。然后,对于每个转发树,sink 节点需向网络中的所有节点发出消息来告知每个节点的父节点和子节点。但是,此类集中式方法所产生的开销使这类方法难以应用于实践。而且,这类集中式方法无法对拓扑的改变作出好的回应。因此,本文为构建转发树提出了一种分布式方法,在这种转发树中,每个节点都可以局部决定其父节点应将编码数据传输给哪个节点。

2 分布式压缩数据收集算法及其近似性分析

为构建 m 个转发树提出一种分布式方法。每个转发树将一个特殊映射的相关节点与 sink 节点相连接并允许 sink 节点以最低通信成本收集与映射相对应的加权总和。一旦收集到 m 个编码总和,sink 节点就会通过解决一个凸优化问题来恢复/解码原始数据。第 1 阶段,sink 节点向其邻近的节点发送搜索信息。每个节点在接收到搜索信息之后就会播报此搜索信息并允许不接近 sink 节点的其他节点接收这类搜索信息。因此,每个节点 v 都需要知道自己到 sink 节点的最短路径 P_{vs} 及在这条路径上的跳数 h_v。此外,节点 v 还需知道其邻节点 $N(v)$。节点 v 可根据对存储器中的矩阵 $\boldsymbol{\Phi}$ 检查的结果,确定节点 $u \in N(v)$, $\forall u$ 是否属于转发树 t 的利益节点 I_t 集合。在第 2 阶段,每个节点确定其父节点到 sink 节点上行链路的路径。

2.1 分布式压缩数据收集算法

第 1 阶段包含向网络发送探索信息,这样每个节点就可知道自己到 sink 节点的最短路径及跳数。在第 2 阶段开始构建转发树。本文描述了转发树 t 的构建过程,类似过程重复用于其他树的构建。网络中的节点一旦接收到搜索信息,每个节点 j 可根据接收到的搜索信息确定自身是否属于利益节点($j \in I_t$)。对每个利益节点分配一个属性,以此来指定其父利益节点 π_j,父利益节点可以是节点 j 的一个邻节点或通过其他的中继节点得到的父利益节点,并分配一个决定标识 F_j 表明节点 j 的父利益节点是否已设定($F_j=1$ 表示父利益节点已设定)。具体过程如下。

(1)在 sink 节点处开始使用宽度优先搜索(Breadth search, BFS)向所有节点传播搜索信息,每个节点($v \in G$)了解自己到树根的最短路径(P_{vs}, $s=$root)和跳数。

(2)对于每个树 T_t, $t=1,2,\cdots,m$,确定利益节点的集合 $T_t \subseteq V$,首先进行判断执行是(与转发树的树根邻近),每个利益节点选择树根作为其父节点。如果利益节点 j 与树根不相邻近,但具有一个邻近利益节点 b 且 $F_b=1$,那么,节点 j 就会选择利益节点 b 作为其父利益节点并分配一个决定标识 $F_j=1$。在节点 j 的所有邻近利益节点 b 都没有决定标识集(例如,$F_b=0$)的情况下,节点 j 将选择从邻近利益节点到 sink 节点具有最少跳数的邻近利益节点作为其父利益节点。

(3)现在,当只能找到 sink 节点的跳数与节点 j 到 sink 节点的跳数相同利益节点的邻接点时,节点 j 将选择其中一个邻近利益节点 b,这个邻近利益节点 b 需符合标准:邻近利益节点 b 相继的父利益节点

到达一个利益节点的跳数最少或决定标识 $F=1$ 或无父节点且未到达节点 j 的位置(为了避免循环)。当一个利益节点选择一个具有相同跳数的父节点时,它必须记录与自己相继的父节点,并在节点或链接出现故障时,这个节点可告知其子节点有关最新路径的决策信息。

(4)如果上述条件都不能实现,节点 j 就会使用 BFS 在通信半径 h_j-1 范围内寻找一个到 sink 节点跳数最少的利益节点 b,否则就会寻找一个决定标识 $F_b=1$ 的利益节点。在此过程中,为了避免此过程再次循环,节点 j 需避开选择利益节点 b(父利益节点 $\pi_b=j$)。最后,如果找不到利益节点,节点 j 就会通过最短的路径将自己直接与 sink 节点连接(可从搜索阶段了解这种路径)。$F_j=0$ 的节点需重复以上过程直到其标识属性 F_j 不再变化。图 1 中给出的是分布式算法。

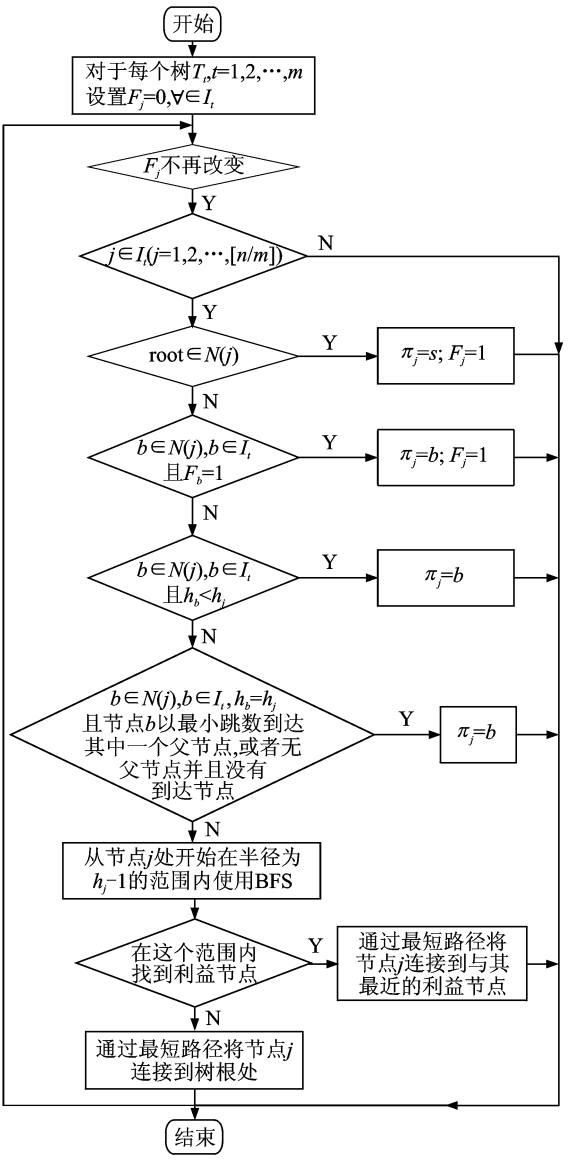

图 1 分布式压缩数据采集算法

Fig. 1 Distributed compression data acquisition algorithm

2.2 算法的近似性分析

考虑到网络拓扑、最差和最佳转发树/聚合树之间的能耗比值(或传输次数)以及可确定拓扑的性能界限/比率 δ,在所有可能的拓扑中,将最大性能界限 $\delta_{\max}$ 称为该算法[9]的近似比率。因此,通过分析网络中树拓扑结构(可给出最差和最佳转发树间的传输率最大值),可确定分布式压缩数据收集算法的近似比率。图 2 为 2 级连接树的两种最差情况的拓扑示例,而用实线/虚线表示的最佳/最差转发树需要 3/4 次传输,其性能比 $\delta=4/3$。如果在 2 级连接树中增加或减少利益节点的数量,其性能比不会超过 4/3。因此,2 级连接树的近似比率 $\delta_{\max}=4/3$。当可将整个网络分成 k 部分时,如图 2 显示的 2 级连接树,就会出现最差的情况。此时可根据最差情况的拓扑,见图 2 虚线,在 k 子图中选择聚合路径。在这种情况下,近似比值不会超过 4/3。然而需注意的是获得这一近似比率(最差情况下的性能)的可能性很小,正如仿真结果显示,分布式压缩数据收集算法总能达到与集中式压缩数据收集算法相接近的性能。

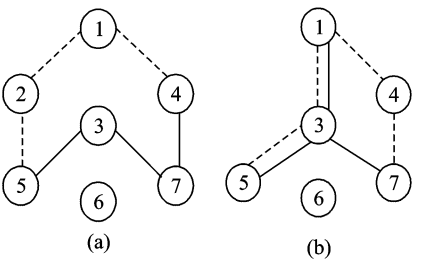

图 2 两种最差的情况链接图

Fig. 2 Two worst case link diagrams

3 示例验证与仿真

3.1 示例验证

在进行例证验证之前,需注意本算法的第 1 阶段一旦完成,为了配合构建转发树,每个传感器节点需局部进行所提的分布

式处理算法。如图 3 所示，假定每个节点都通过一个标准化距离（正距离）的链接与其所有的邻近节点相连接，且利用跳数计算从源节点到目的地的路径长度。灰色节点指需要通过高效的转发树才能与 sink 节点（黑色区域 S）相连接的利益节点。

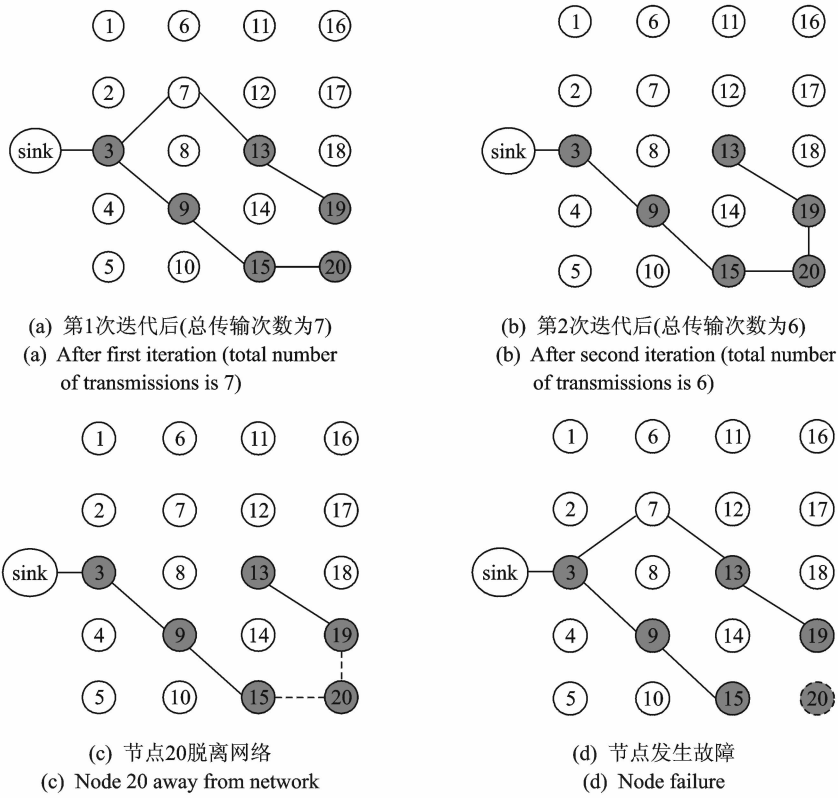

图 3 本文算法在样本网络中的运行过程

Fig. 3 Running process in sample network in the proposed algorithm

当搜索阶段完成后，每个利益节点将局部确定其在转发树中的父节点，所有的利益节点同时进行第 2 阶段。作为 sink 节点的邻节点，对节点 3 进行设置：$\pi_3 = s$ 且 $F_3 = 1$，然后通过节点 3，从 sink 节点接收到告知信息，节点 9 将其父节点设置为节点 3，$\pi_9 = 3$。此后节点 9 向其父节点（节点 3）发送告知信息，告知节点 3 是其父节点的决定）及（节点 9）将通过父节点（节点 3）向 sink 节点传输数据，此处，在向 sink 节点转发数据之前，节点 3 将独自对从节点 9 接收到的数据进行编码。节点 3 在接收到告知信息后会向其子节点（节点 9）发回一条信息并告知节点 3 的决定标识为 $F_3 = 1$，同时节点 9 将设置自己的决定标识为 $F_9 = 1$。同样可按照类似过程设定节点 15 和 20 的决定标识。在第 1 次迭代中，节点 19（图 3(a)）从节点 13 和 20 处接收到搜索信息，但由于节点 13 到 sink 节点的跳数最少，因此选择节点 13 作为其父节点（$\pi_{19} = 13$）。然而在第 2 次迭代（图 3(b)）中，当节点 20 告知其邻节点 19 自己已设定决定标识为 $F_{20} = 1$ 后，节点 19 就会将其父节点转换为节点 20（$\pi_{19} = 20$）并设定决定标识 $F_{19} = 1$。如上所述，决定标识是根据数据采集利益这一思路而设定的。还需注意每个节点只能使用一次分布式压缩数据收集算法，除非在网络触发路由维护过程中，节点能够从其邻节点处接收到告知信息（例如，标识值改变）或此节点后面的一个节点或链接由于节点移动或信道损坏造成的故障。在两种情况下，节点将通过分布式压缩数据收集算法来选择一个新的父节点以决定到达 sink 节点的新路径。图 3(c)给出了路由维护。

一旦节点 20 发生故障,节点 19 和节点 13 就会与转发树断开连接,此时节点 19 将使用分布式压缩数据收集算法告知利益节点 13,已将其选择作为其父利益节点并设定决定标识 $F_{19}=0$。节点 13 在接收到节点 19 的告知信息并了解这种变化后将局部使用分布式压缩数据采集(Compressing data gathering, CDG)算法选择节点 3 作为其父利益节点并设定决定标识 $F_{13}=0$,在节点故障恢复后,新的转发树显示在图 3(d)中。

3.2 仿真结果

为了测试本文算法的有效性及优越性,在 P4 双核 2.85 GHz CPU, 4 GB RAM,Windows XP 操作系统,Matlab 2012 平台下进行仿真实验。根据文献[10],假定具有一个 700×700 区域的 WSN 网络,节点按照均匀分布方式随机分布在该网络内,此外还假定所有的节点均具有一个通信范围。该 WSN 网络,节点需采集的原始数据包长度为 $N=256$,每个数据大小为 8 bit;数据包转发的信道容量为 $C=4$ Mbit/s,压缩数据后所得观测数据包长度的范围为 $0<M\leqslant N$。模拟中,改变节点的密度并利用不同的随机样本矩阵 $\boldsymbol{\Phi}$ 计算超过 10 次得到平均值。利用给出的仿真模拟结果对分布式算法的性能进行评估,并使用传输耗能和信息量与混合式压缩数据采集[5]、基于映射的方法[6]进行比较。传统压缩方案下,节点每上传一次数据包平均耗能为

$$E=k\times E_{\text{elec}}+\varepsilon_{\text{amp}}\times k\times D^2 \tag{5}$$

式中:$\varepsilon_{\text{amp}}=100$ pJ/(bit·m^{-2}),$E_{\text{elec}}=50$ nJ/bit,k 为发送数据位数,D 为发送数据距离。图 4 和图 5 分别为总传输耗能在不同映射次数 m 和节点密度下的变化情况,其中 m 表示节点密度,n 表示节点数量。从图 4 中可以清楚地看到,非 CS 情况的能耗最高,其次是混合式 CS。这是由于非 CS 这种方法没有利用基于映射的采集方法,而是仅利用构建的转发树收集所有的加权总和。利用基于映射的压缩数据采集方法得到的传输能耗最低,且提出的分布式算法与其非常接近,在图 4 中两种方法相差 2.7%,在图 5 中相差 3.1%。在图 4 和图 5 中,这种方法与基于映射的压缩数据采集方法的最大差距是6.2%。从图 4 中可发现压缩率(m 值较高时)较低时产生的传输能耗,比压缩率(m 值较低)较高时产生的传输能耗高。这是由于 m 值较高意味着映射越多,因此在网络中收集加权总和的传输次数就越多。最后,对利用分布式压缩数据采集方法构建转发树产生的开销与使用基于映射的压缩数据采集方法构建转发树产生的开销进行了对比,在不同规模网络中的比较结果加图 6 所示。可以清楚地看出,使用分布式 CDG 方法构建转发树产生的通信开销比使用 PB-CDG 低,使用分布式 CDG 方法产生的开销要比使用 PB-CDG 少 53%~65%。

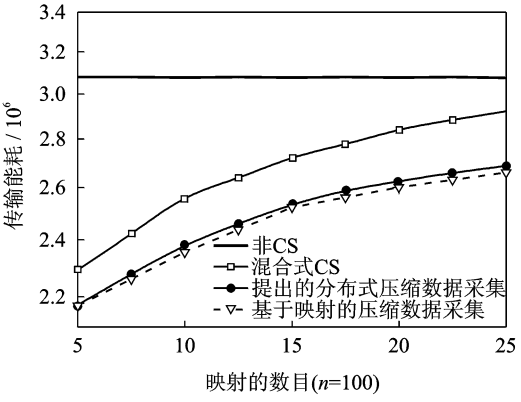

图 4 不同的映射数目的传输能耗

Fig. 4 Transmission energy consumption in different mapping number

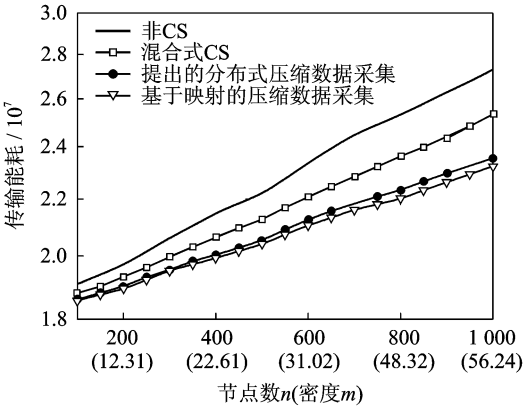

图 5 不同网络密度的传输能耗

Fig. 5 Different network transmission energy density

4 结束语

本文提出了一种高效的分布式压缩数据收集方法，在这种方法中，每个传感器节点可独自找到其父节点并能够构建路由树的一部分，无需利用中央单元来构建所有转发树，因此每个传感器节点对转发树的构建和维护做出局部决策。与先前的方法相比，这种方法不会产生过多开销，且在路由过程中每个节点会做出局部决策。通过模拟得出分布式压缩数据收集方法在传输能耗方面非常接近于最佳集中式方法，并在通信开销方面超过其他所有方法。

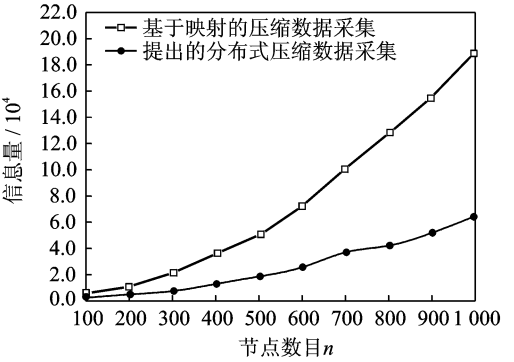

图 6 不同节点数时的信息流量

Fig. 6 Information flow with different nodes count

参考文献：

[1] Turkanović M, Brumen B, Hölbl M. A novel user authentication and key agreement scheme for heterogeneous ad Hoc wireless sensor networks, based on the Internet of things notion[J]. Ad Hoc Networks, 2014, 20(2):96-112.

[2] Kuo T, Tsai M. On the construction of data aggregation tree with minimum energy cost in wireless sensor networks: Np-completeness and approximation algorithms[C]// Proc 2012 IEEE INFOCOM. [S. l.]: IEEE, 2012: 2591-2595.

[3] Ebrahimi D, Assi C. Optimal and efficient algorithms for projection based compressive data gathering[J]. IEEE Commun Lett, 2013, 17(8):1572-1575.

[4] 欧庆波，宋荣方. 无线传感网中基于压缩感知的高效数据传输方案[J]. 南京邮电大学学报：自然科学版，2012，32(2)：44-51.
Ou Qingbo, Song Rongfang. High efficiency data transmission scheme based on compressed sensing in wireless sensor networks[J]. Journal of Nanjing University of Posts and Telecommunications: Natural Science Edition, 2012, 32(2): 44-51.

[5] 朱林，张海. 数据集成技术在树型 WSN 中的应用[J]. 数据采集与处理，2013，28(6)：818-822.
Zhu Lin, Zhang Hai. Application of data integration technology in tree WSN[J]. Journal of Data Acquisition and Processing, 2013, 28(6): 818-822.

[6] 杨海蓉，张成，丁大为，等. 压缩传感理论与重构算法[J]. 电子学报，2011，39(1)：142-148.
Yang Hairong, Zhang Cheng, Ding Dawei, et al. Compression sensing theory and reconstruction algorithm[J]. Acta Electronica Sinica, 2011, 39(1): 142-148.

[7] 李协，张效义，汪子嘉. 无线传感网中面向到达时差估计的数据压缩方法研究[J]. 信号处理，2012，28(9)：1226-1234.
Li Xie, Zhang Xiaoyi, Wang Zijia. Wireless sensor network for TDOA estimation data compression method[J]. Signal Processing, 2012, 28(9): 1226-1234.

[8] 罗永健，丁小勇，罗相根，等. 一种有效的无线传感器网络数据复原汇聚方法[J]. 数据采集与处理，2011，26(1)：90-94.
Luo Yongjian, Ding Xiaoyong, Luo Xianggen. An effective resilient data aggregation in wireless sensor networks[J]. Journal of Data Acquisition and Processing, 2011, 26(1): 90-94.

[9] 汪鲁才，赵延昇，林海军，等. 基于分布式压缩感知的能量收集 WSNs[J]. 传感器与微系统，2014，33(7)：45-48.
Wang Lucai, Zhao Yansheng, Lin Haijun, et al. Distributed compressed sensing energy collection system based on WSNs [J]. Sensor and Micro, 2014, 33(7): 45-48.

[10] 康莉，谢维信，黄建军，等. 无线传感器网络中的分布式压缩感知技术[J]. 信号处理，2013，29(11)：1560-1567.
Kang Li, Xie Weixin, Huang Jianjun, et al. Wireless sensor networks in distributed compressed sensing technology[J]. Signal Processing, 2013, 29(11): 1560-1567.

作者简介：

杜淑颖(1981-)，女，讲师，研究方向：智能信息处理，E-mail：13775883477@139.com。

丁世飞(1963-)，男，教授，博士生导师，研究方向：机器学习与数据挖掘、人工智能与模式识别等。