

一种分簇无线传感器网络中的分布式信源编码算法

秦智超* 周正 赵小川

(北京邮电大学泛网无线通信教育部重点实验室 北京 100876)

摘要: 针对密集型无线传感器网络中信息存在大量冗余的问题, 该文提出了一种适用于分簇无线传感网络的分布式信源编码算法。该算法以边信息作为初始参考信源, 利用信源间的相关性来决定各信源的编码顺序和参考信源, 然后由各信源相对于参考信源进行相关编码, 接收端则根据编码顺序和参考信源进行相关译码。该文同时针对模值编码的方式, 给出了一种低复杂度的译码算法。理论分析和仿真结果表明, 将该算法应用于分簇路由协议中可以有效地降低节点的发送比特数, 从而降低网络的能耗以延长网络寿命。

关键词: 无线传感器网络; 分布式信源编码; Slepian-Wolf 编码; 相关度

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2013)02-0328-07

DOI: 10.3724/SP.J.1146.2012.00723

A Distributed Source Coding Algorithm for Clustering Wireless Sensor Networks

Qin Zhi-chao Zhou Zheng Zhao Xiao-chuan

(Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In order to remove the vast information redundancy in dense Wireless Sensor Networks (WSN), a distributed source coding algorithm for clustering WSN is proposed. The algorithm uses the correlation among the sources to define the coding sequence and their reference source with the side information as the initial reference source and each source codes its data with respect to its reference source, and the receiver decodes the data according to the coding sequence and reference source. In addition, a decoding algorithm with low complexity for modulus coding is given. The analysis and simulation results show that the proposed algorithm applied to clustering routing protocol can effectively reduce the number of bits to be sent, and thereby decrease the energy consumption of network and prolong the network lifetime.

Key words: Wireless Sensor Networks (WSN); Distributed source coding; Slepian-Wolf coding; Correlation degree

1 引言

无线传感器网络(WSN)是由大量廉价的微型传感器节点组成的多跳自组织无线网络, 在医疗、农业、环境监测、智能仪表等领域发挥着越来越重要的作用^[1]。由于传感器节点通常由电池供电且在部署完成后难以进行更换, 所以如何让由有限能量的微型传感器节点所构成的无线传感器网络实现较长的生命周期, 成为大规模无线传感器网络发展所面临的重大挑战^[2]。

利用数据间的空间相关性减少数据发送量是使

网络高效运行的重要手段之一。为了实现精确感知的需求和对单个节点失效的鲁棒性, WSN往往需要进行密集部署, 这使得感知数据间存在丰富的空间相关性^[3]。例如在相似地理区域的温湿度监测和移动车辆的跟踪应用^[3]中以及保证服务质量(QoS)的视频监控应用^[4]中均存在大量的信息冗余。由于无线传感器网络的能耗主要集中在数据传输上且与数据发送量成正比, 所以这些冗余信息将会造成巨大的能量浪费。若能有效地利用数据相关性, 减少发包次数, 可以延长网络寿命^[5]。分布式信源编码(Distributed Source Coding, DSC)为消除这种相关性提供了很好的思路, 它可以使得各相关信源在不用相互通信的情况下进行独立编码, 并获得与相互通信同样的数据压缩性能, 目前在WSN领域中得到了广泛的研究^[6-10]。

2012-06-11 收到, 2012-11-28 改回

国家科技重大专项(2009ZX03006-009)和韩国知识经济部仁荷大学ITRC 基金资助课题

*通信作者: 秦智超 qinzichao@bupt.edu.cn

层次路由是使网络高效运行的另外一个重要手段。理论与实验证明,采用层次路由比平面路由更能满足能量有效性的要求^[11]。层次路由通过分簇算法对无线传感器网络中的节点进行层次划分,将数据采集和通信任务分配给角色不同的节点执行,有利于能耗在全网节点上的均匀、一致地分布,达到提高工作寿命的目的^[12],其中比较经典的协议有 LEACH^[13], HEED^[14]等。

将分布式信源编码和层次路由相结合,可以实现对采样数据进一步压缩的目的。文献[8]提出了一种将分布式信源编码应用于链状簇的方案,沿着信号路径进行相关信源编码。文献[9]提出了一种将分布式信源编码应用于多跳簇的方案,并给了两种数据融合方式(即沿着信号路径向前融合和向后融合)。但是这两种方案仅仅利用了两个相邻节点间的相关信息,压缩效率没有最大化。文献[10]虽然提出了一种能够达到 Slepian-Wolf 理论^[15]极限值的编码算法,每个节点根据自己与所有邻居节点间的相关信息来进行编码,但是仅仅停留在理论层面,如何利用这些邻居节点间的相关信息并没有给出。针对上述不足,本文尝试为分簇传感器网络设计一种编码算法,主要的贡献如下:(1)针对使用模值编码的方法,设计了一种简单的译码算法,降低了译码的复杂度;(2)提出了一种多信源相关编码算法(MSCC),充分利用了各信源间的相关性;(3)将 MSCC 算法应用到分簇无线网络中,延长网络的生命周期。

2 理论基础

1973年提出的 Slepian-Wolf 理论是分布式信源编码的理论基础,本节首先对该理论进行简要介绍,然后对基于该理论的模值编码进行阐述,为本文所提的 MSCC 算法提供理论依据。

2.1 Slepian-Wolf 编码理论

根据 Slepian-Wolf 编码理论,对两个互相关的信号源 X, Y 进行编码,如果 X 和 Y 知道彼此的互相关信息,比如联合概率分布 $p(x, y)$,那么 X 就可以在不需事先知道 Y 的情况下获得与事先知道 Y 一样的编码效率,也就是说图1所示的两种编码方案的效果是一样的。但是独立编解码方案不需要与 Y 进行通信,不但具有更大的灵活性,同时还可以减少通信能耗,这一点对于能量受限的无线传感器网络来说颇具吸引力。

在 Slepian-Wolf 编码理论中将 Y 称作边信息(side information)。对 X 和 Y 进行独立编码,并在译码端进行无失真地联合译码,需要满足以下条件:

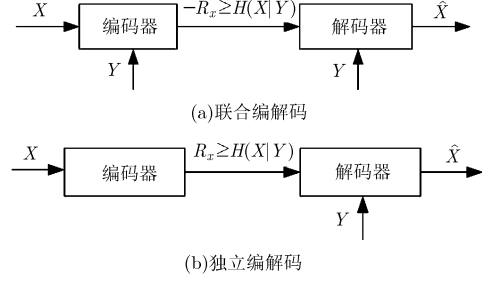


图1 两种编码方案

$$\left. \begin{aligned} R_x &\geq H(X|Y) \\ R_y &\geq H(Y|X) \\ R_x + R_y &\geq H(X, Y) \end{aligned} \right\} \quad (1)$$

其中 R_x, R_y 表示 X 和 Y 的编码速率, $H(X|Y), H(Y|X)$ 分别表示对应的条件熵, $H(X, Y)$ 表示两者的联合熵。Slepian-Wolf 编码理论针对两个相关信源,后来被拓展到多个相关信源^[16],直到1999年 DISCUS^[17] 算法被提出之前一直处于理论分析阶段。目前为止,已经有许多实际的 DSC 方法被提出,如使用伴随式的分布式信源编码^[17],使用 Turbo 码的方法^[18],使用 LDPC 码的方法^[19],使用模值的方法^[3]等。对于温度监控等应用场景来说,由于其相关性往往表现为其差值在一定范围内且监测值连续,采用模值的方法更具有优越性(即编码简单),下面对其进行介绍。

2.2 模值编码

考虑温度监控的场景,设其数据范围为 $[\text{mins}, \text{maxs}]$, 监测精度为 Δ (该值取决于实际应用的需求,比如 0.1°C), 数据比特数为 n , 则其可能的监测样本集合为 $\Omega = \{s_i | s_i = \text{mins} + i \times \Delta, i \in (0, 1, \dots, 2^n - 1)\}$ 。根据文献[3],如果信源 X 与边信息 Y 之间的差值小于 $2^{k-1} \Delta (0 \leq k \leq n)$, 则 X 可以以 k bit 进行编码以实现数据压缩的目的。具体采用的模值编码公式为

$$f(X) = \text{index}(X) \bmod 2^k \quad (2)$$

其中 $f(X)$ 表示 X 编码后的信息比特, $\text{index}(X)$ 表示 X 在 Ω 中的次序(从 0 开始), 计算公式如下:

$$\text{index}(X) = (X - \text{mins}) / \Delta \quad (3)$$

将 X 中编码后相同的数据归为一个集合(称之为陪集),并将 $f(X)$ 作为 X 所在陪集的索引,可以得到陪集共有 2^k 个,每个陪集内的元素个数为 2^{n-k} 。

在译码端则根据已知的边信息 Y 和接收到的 $f(X)$ 恢复出实际的信号 X , 采用的译码公式为

$$\hat{X} = \min_{r_i \in S} \|Y - r_i\| \quad (4)$$

其中 \hat{X} 表示译码后的结果, S 表示由 $f(X)$ 给出的陪集, r_i 表示陪集内的元素。

下面给出一个编译码的具体例子。设 $\Omega = \{0, 0.1, \dots, 1.5\}$, X 与 Y 之间的差值小于 2Δ , 则有 $n=4$, $\Delta = 0.1$, $k=2$, Ω 可以划分为 4 个陪集, 分别为: $S_{00} = \{0, 0.4, 0.8, 1.2\}$, $S_{01} = \{0.1, 0.5, 0.9, 1.3\}$, $S_{10} = \{0.2, 0.6, 1.0, 1.4\}$, $S_{11} = \{0.3, 0.7, 1.1, 1.5\}$ 。

其中每个陪集符号的下标表示对应的编码比特。如果 X 所发送的数据为 0.7, 边信息 Y 的数据为 0.6 则根据式(4)可以得到编码后的信息比特为 11。解码器接收到信息比特后, 寻找对应的陪集中与边信息 Y 距离最近的一个即可恢复出 X 。可以看出, 模值编码可以有效地降低发送比特数, 对本例而言, 压缩率有 $r = n/k = 2$ 。

以上编码过程比较简单, 仅仅为数值计算, 其复杂度为 $O(1)$, 但是译码过程需要在陪集中进行搜索, 在陪集中元素按序排列的情况下, 其平均复杂度与陪集的元素个数成正比, 即为 $O(2^{n-k})$, 证明略。

可见对于样本值比较大(即 n 值较大)及相关性较大(即 k 值较小)时, 其译码复杂度较高, 为了改善此问题, 本文提出一种新的译码方法, 其复杂度仅为 $O(1)$, 在 3.1 节中进行阐述。

3 算法设计

本节首先给出只有两个相关信源时的编码算法, 然后将其拓展到多个相关信源。

3.1 二信源相关编码

编码过程采用模值编码方式, 即式(2), 不再赘述。但对于译码过程, 则按照以下公式得出:

$$\hat{X} = Y + \varepsilon\Delta \quad (5)$$

其中

$$\varepsilon = \begin{cases} X' - Y', & |X' - Y'| < 2^{k-1} \\ (|X' - Y'| - 2^k) \times \text{sign}(X' - Y'), & |X' - Y'| \geq 2^{k-1} \end{cases} \quad (6)$$

$$X' = \text{index}(X) \bmod 2^k \quad (7)$$

$$Y' = \text{index}(Y) \bmod 2^k \quad (8)$$

$\text{sign}(\cdot)$ 表示符号函数。

下面通过逆推的方式(由式(5)推导出式(6))给

出具体的证明过程, 当实现理想译码时(即 $\hat{X} = X$) 有

$$\varepsilon = (X - Y) / \Delta \quad (9)$$

根据式(3), 式(9)有

$$\varepsilon = \text{index}(X) - \text{index}(Y) \quad (10)$$

由式(7), 式(8), 可将 $\text{index}(X)$ 和 $\text{index}(Y)$ 表示成如下形式:

$$\text{index}(X) = z_x 2^k + X' \quad (11)$$

$$\text{index}(Y) = z_y 2^k + Y' \quad (12)$$

其中 z_x, z_y 为非负整数。

同时根据式(10), 式(11), 式(12)则有

$$\varepsilon = (z_x - z_y) 2^k + X' - Y' \quad (13)$$

由于 X 与 Y 的差值小于 $2^{k-1}\Delta$, 根据式(9)有 $|\varepsilon| < 2^{k-1}$, 则必有 $|z_x - z_y| \leq 1$ (证明略)。所以在下面的推导过程中只会出现 $|z_x - z_y|$ 为 0 或者 1 的情况。

(1) $|X' - Y'| < 2^{k-1}$ 时: 如果 $|z_x - z_y| = 1$, 则有 $|\varepsilon| = |(z_x - z_y) 2^k + X' - Y'| \geq |(z_x - z_y) 2^k| - |X' - Y'| > |2^k - 2^{k-1}| = 2^{k-1}$, 与 $|\varepsilon| < 2^{k-1}$ 矛盾, 所以必有 $|z_x - z_y| = 0$, 根据式(13)有 $\varepsilon = X' - Y'$ 。

(2) $|X' - Y'| \geq 2^{k-1}$ 时: 如果 $|z_x - z_y| = 0$, 则有 $|\varepsilon| = |X' - Y'| \geq 2^{k-1}$, 与 $|\varepsilon| < 2^{k-1}$ 矛盾, 所以必有 $|z_x - z_y| = 1$ 。另外如果 $z_x - z_y$ 与 $X' - Y'$ 符号相反, 则有 $|\varepsilon| = 2^k + |X' - Y'| > 2^{k-1}$, 与 $|\varepsilon| < 2^{k-1}$ 矛盾, 所以 $z_x - z_y$ 与 $X' - Y'$ 必然符号相同, 则有 $\varepsilon = (|X' - Y'| - 2^k) \times \text{sign}(X' - Y')$ 。证毕

可见本译码算法仅仅利用数值计算的方式, 不受样本规模和相关性的影响, 有效地降低了译码复杂度和改善了编译码过程带来的时延问题。下面给出几个具体的译码例子, 仍采用 2.2 节的场景, 如表 1 所示。

3.2 多信源相关编码

为了更加清楚地描述该算法(MSCC), 给出以下定义。

相关度: 反映信源间相关性的指标。该指标越大, 相关性越大, 可以用越少的比特数进行编码。

相关矩阵: 以相关度为元素组成的矩阵 C , 其

表 1 编译码例子

相关信源	中间过程部分参数	译码结果
$X = 0.7, Y = 0.6$	$X' = 3, Y' = 2, k = 2, \varepsilon = 1$	$\hat{X} = 0.7$
$X = 0.8, Y = 0.7$	$X' = 0, Y' = 3, k = 2, \varepsilon = 1$	$\hat{X} = 0.8$
$X = 0.7, Y = 1.0$	$X' = 7, Y' = 2, k = 3, \varepsilon = -3$	$\hat{X} = 0.7$
$X = 0.8, Y = 0.6$	$X' = 0, Y' = 6, k = 3, \varepsilon = 2$	$\hat{X} = 0.8$

中 C_{ij} 表示信源 X_i 与信源 X_j 之间的相关度。

参考信源：如果信源 X_i 以信源 X_j 为边信息进行编码，则称信源 X_j 是信源 X_i 的参考信源。

假设 $(X_1, X_2, \dots, X_n, Y)$ 为多个相关信源，其中 Y 为边信息， (X_1, X_2, \dots, X_n) 为需要进行编码的信源。为表达式统一，令 $Y = X_{n+1}$ 。相关矩阵 C 已知，以 S 表示待编码信源集合， \bar{S} 表示已编码信源集合。 R 表示参考信源集合， E 表示编码结果，其编码过程如表 2 所示。

表 2 MSCC 算法编码过程

步骤 1	初始化 $R = \Phi, E = \Phi, S = \{X_1, X_2, \dots, X_n\}, \bar{S} = \{Y\}$;
步骤 2	分别从 S 中选取 X_i ，从 \bar{S} 中选取 X_j ，使其满足 $\operatorname{argmax}_{X_i \in S, X_j \in \bar{S}} C_{ij}$;
步骤 3	从 S 中移除 X_i 并将其放入 \bar{S} 中，并将 X_j 放入 R 中；
步骤 4	对 X_i 关于 X_j 进行二信源相关编码，并将译码结果 $f(X_i)$ 放入 E 中；
步骤 5	重复步骤 2 到步骤 4，直到 S 为空为止；
步骤 6	删除 \bar{S} 中的 Y 。

最终 \bar{S} 为各信源的编码顺序， E 为各信源按此顺序编码后的结果， R 则为对应的参考信源。下面给出一个比较形象的编码过程示意图，以信源间的距离来度量彼此间的相关度(距离越近相关性越大)，黑色圆圈表示边信息 Y ，圆圈中数字则表示编码顺序，箭线表示参考关系，即箭尾端的信源以箭头端的信源为参考信源进行相关编码，如图 2 所示。

以 D 表示译码结果，如果能够得到编码阶段产生的信息(即 \bar{S}, R, E)，则可以成功地进行译码。译码过程如表 3 所示。

4 算法应用

分布式数据压缩在无线传感器中具有先天的优越性^[8]，而同时簇状拓扑又非常有利于 DSC 算法的运行，所以本节将 MSCC 算法应用到分簇路由协议

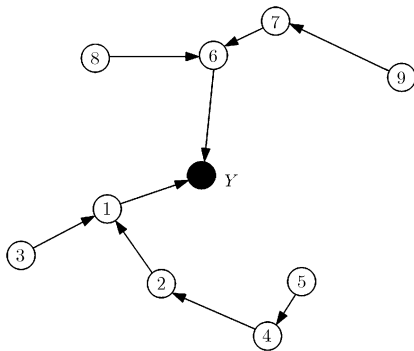


图 2 MSCC 算法过程示意图

表 3 MSCC 算法译码过程

步骤 1	初始化 $D = \{Y\}$;
步骤 2	取出 E 和 R 中第 1 个元素(分别记为 e 和 r)，将 e 相对于 D 中的参考信源(由 r 指定)进行二信源相关译码。将译码结果按序放入 D 中，然后删除 e 和 r ;
步骤 3	重复步骤 2，直到 E 和 R 为空时为止；
步骤 4	删除 D 中的 Y ，此时 D 为按照由 \bar{S} 指定编码顺序的译码结果，将 D 进行自然排序即为最终的译码结果。

中，以改善其网络性能。考虑到 LEACH 协议是目前多数分簇路由协议的基础，以其为例进行改进具有普遍意义。本节首先对 LEACH 协议进行简要介绍，然后给出 MSCC 算法在 LEACH 协议上的具体应用。

4.1 LEACH 协议

LEACH 协议是最早的专为无线传感器网络设计的经典路由协议。该协议将数据传输过程分为多轮，并在运行过程中不断地循环执行簇的重构过程，每次重构主要分为簇建立阶段(setup phase)和稳定运行阶段(ready phase)。其中簇建立阶段又可分为竞争簇首，簇首广播，成员加入和调度广播 4 部分。

4.2 具体应用

基于簇的传感器网络中，由于簇内节点在地理位置上相对较近，它们采集的数据具有很强的相关性，这种相关性可以根据彼此间的距离(可以通过一般的定位协议^[10]来获得)和专业领域知识来估计出(如温度在空间的梯度变化)，或者根据一段时间内数据的统计特性来得到(如在网络初始化阶段簇首通过收集足够的样本数据来估计相应的统计特性，并将结果在簇内进行广播)^[20]。在已知相关特性的条件下，就可以对 LEACH 协议进行少量改进以引入 MSCC 算法，具体步骤如下：

(1)在每轮数据传输的开始，各节点开始竞争簇首。竞争方式同 LEACH 协议，即每个节点独立的随机产生一个介于 0 到 1 之间的随机数，和阈值 $T(n)$ 进行比较。如果小于该阈值，则被选为新一轮的簇首并通告其邻节点，否则意味着没有竞争成功或者已经当选过簇首，只能成为成员节点。其中， $T(n)$ 表示为

$$T(n) = \begin{cases} \frac{p}{1 - p[r \bmod (1/p)]}, & n \in G \\ 0, & \text{其它} \end{cases} \quad (14)$$

其中 p 是簇首百分比， r 表示当前循环次数， G 是最近 $1/p$ 轮中没有当选过簇首的节点集合。

- (2)竞争成功的簇首广播自身成为簇首的消息。
- (3)没有竞争成功的节点，作为簇成员选择距离

自己最近的簇首申请加入簇。

(4)簇首根据簇内成员间的相关信息，通过 MSCC 算法计算出 \bar{S} 和 R ，以各信源在 \bar{S} 内的顺序作为时隙分配的次序，以 R 中的对应元素作为各信源的参考信源，作为调度信息在簇内进行广播。

(5)各节点相对于参考信源进行相关编码，并将编码后的数据在分配时隙内发送到簇首，簇首则进行相关译码。

(6)数据传输阶段结束后，重复步骤(1)–步骤(5)开始新一轮数据传输。

对 LEACH 协议的改进体现在步骤 4 和步骤 5，其余均保持不变。改进思路可以归纳为以下 3 点：首先，成员节点的时隙分配不再按照成员节点申请加入的顺序分配时隙，而是依据 MSCC 算法进行时隙分配。其次，簇首节点需要向成员节点广播其对应的参考信源信息。第三，各信源传递的不再是完整的信息，而是编码后的信息，需要在簇首进行相关译码才能恢复出所有的信息。由于 LEACH 协议是多数 WSN 分簇路由协议的基础，所以该算法可以很容易地移植到其他分簇路由协议中。

5 仿真分析

本节通过比较以下 3 种情形下的网络性能来对 MSCC 算法的有效性进行评价。

情形 1: 采用原始的 LEACH 协议;

情形 2: 在 LEACH 协议的基础上，仅在成员节点与簇首间采用二信源相关编码算法;

情形 3: 在 LEACH 协议的基础上引入 MSCC 算法。

由于编码和译码的复杂度均为 $O(1)$ ，这使得编译码过程与数据规模无关且仅需有限的指令即可完成。同时因为计算能耗一般很小(如 WINS NG2.0 传感器节点，发送 1 bit 信息的功耗与处理单条指令的功耗之比约为 1400^[21])，所以本文忽略编译码过程中的计算能耗，并采用与文献[13]相同的无线通信能耗模型，节点传输 l bit 数据到距离为 d 的目标节点处所需的能耗为

$$E_{Tx}(l, d) = \begin{cases} lE_{elec} + l\varepsilon_{fs}d^2, & d < d_0 \\ lE_{elec} + l\varepsilon_{mp}d^4, & d \geq d_0 \end{cases} \quad (15)$$

其中 E_{elec} 为无线收发电路消耗的能量， ε_{fs} 和 ε_{mp} 分别为当 $d < d_0$ 和 $d \geq d_0$ 时功率放大器的能耗比例系数， d_0 是区分自由空间模型和多径衰落模型的距离阈值，其值为

$$d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}} \quad (16)$$

接收方接收 l bit 数据消耗的能量为

$$E_{Rx}(l) = lE_{elec} \quad (17)$$

相关模型与文献[9]相同，设信源 X_i 与信源 X_j 之间的相关度为

$$C_{ij} = \begin{cases} 1 - (d/T_r)^2, & d \leq T_r \\ 0, & d > T_r \end{cases} \quad (18)$$

其中 T_r 表示簇成员节点的通信半径，超过此距离则认为两者间的数据没有相关性，否则相关度与距离的平方成反比关系。所需数据比特数 k 与相关度的关系为

$$k = \lceil n \times C_{ij} \rceil \quad (19)$$

其中 n 表示采用 MSCC 算法前的原始比特数， $\lceil \cdot \rceil$ 表示向上取整。

考虑 100 个传感器节点随机部署在 $100 \text{ m} \times 100 \text{ m}$ 的对温度进行监控的场景。温度的监控范围为 $[-40^\circ\text{C}, 60^\circ\text{C}]$ ，监测精度为 0.1°C ，则所需的原始数据比特数 $n=10 \text{ bit}$ 。基站位于区域中心，每个节点的初始能量为 1 mJ ， p 取 5%，其余仿真参数均与文献[9]，文献[13]相同，如表 4 所示。

表 4 仿真参数

E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/m ²
ε_{mp}	0.0013 pJ/bit/m ⁴
T_r	20 m

图 3 给出了节点的分布及运行过程中一轮簇建立时的情况。由图可见，在该轮数据传输过程中共生成 6 个簇，其中部分成员节点由于距基站比其他簇首节点更近，所以选择将数据直接发往基站，从而形成一个虚拟簇(其簇 ID 为 6)，簇首不断地进行轮换来保证能量的均匀消耗。

在 LEACH 协议中引入 MSCC 算法会改变原来的时隙分配，表 5 给出了引入 MSCC 算法前后的时

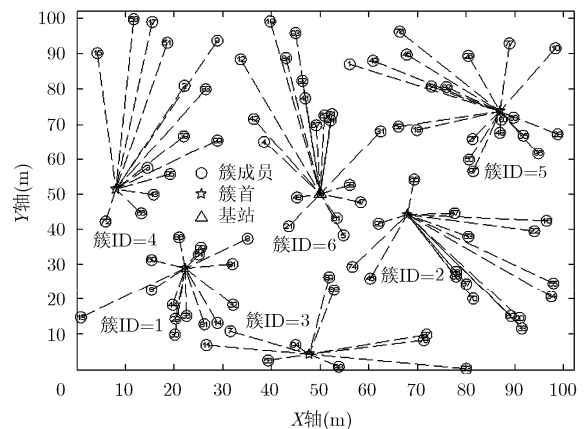


图 3 网络分簇结构

表 5 簇 ID 为 1 的簇引入 MSCC 算法前后的时隙分配对比

LEACH 协议	54	79	60	86	6	91	44	36	32	29	8	81	14	30	13
LEACH +MSCC 协议	54	79	86	60	91	8	6	44	29	36	81	14	30	32	13

隙分配对比。为节省篇幅，这里仅以簇 ID 为 1 的簇为例。这种改变是因为 MSCC 算法中译码存在一定的先后顺序，即参考节点必须先于被参考节点译码。

图 4 给出了各个簇的成员分别在 3 种情形下的发送比特数，簇内成员的 ID 号按照 MSCC 算法中决定的编码顺序进行排列。

由图 4 可以看出，在发送比特数方面，每个簇都遵循情形 3 小于等于情形 2，情形 2 小于等于情形 1 的规律。这是因为情形 3 充分利用了簇内成员间的相关性，情形 2 仅仅利用了成员节点与簇首间的相关性，而情形 1 没有利用任何相关性。情形 1 与情形 2 中出现相等的原因是因为个别节点与簇首间的距离达到或者超过了相关范围导致数据间不再具有相关性，而情形 2 与情形 3 出现相等的原因在于情形 3 中的一些节点将簇首作为参考节点。图 5 给出了不同簇中每个成员的平均发送比特数，该图明显的显示了 MSCC 算法应用于 LEACH 协议后的

优越性。

图 6 给出了 3 种情形在每轮循环中的剩余节点个数对比 (对随机生成的场景仿真 100 次取平均)。从图中可以看出，情形 3 相对于情形 1 和情形 2 而言，有效地延长了网络寿命。以第 1 个节点死亡时的轮数为例，情形 1 为 848 轮，情形 2 为 1103 轮，情形 3 为 2863 轮。相对于情形 1，情形 2 和情形 3 分别提高了 30%和 3.4 倍左右，性能改善明显，这充分验证了 MSCC 算法的有效性。

6 结束语

为了解决无线网络中的能量瓶颈问题，消除 WSN 中存在的大量信息冗余，本文提出了一种适用于分簇无线传感器网络的分布式信源编码算法，并以 LEACH 协议为例对算法的有效性进行了验证。本文关于 LEACH 协议的改进思路对于 MSCC 算法在其他分簇路由协议上的应用同样具有

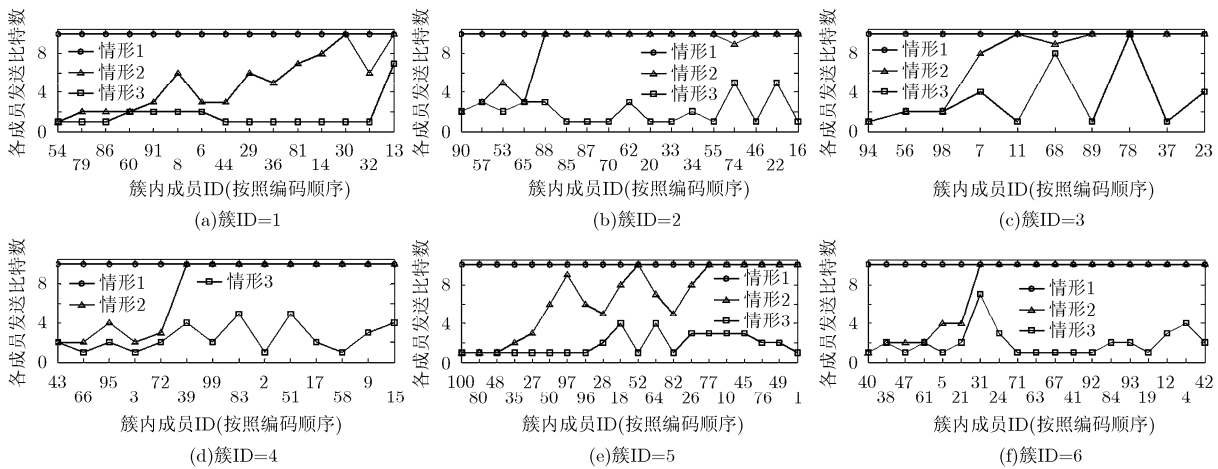


图 4 簇内成员发送比特数

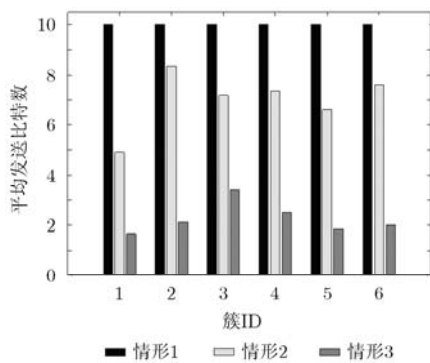


图 5 簇内成员平均发送比特数

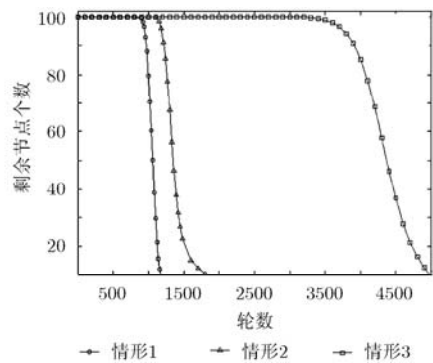


图 6 3 种情形在不同轮数时的剩余节点个数

借鉴意义。仿真结果表明,该算法有效地降低了发送比特数,从而达到了降低能耗,延长网络寿命的目的。但是该算法建立在成员间相关信息已知的基础上,因此如何获得精确的相关信息是算法能够成功应用的关键,这需要在未来的研究中进行考虑。另外,MSCC算法仅从物理层减少发送比特数出发,如何将其与其他层(如MAC层、网络层)进行联合优化也需要进一步探索。

参考文献

- [1] Mainetti L, Patrono L, and Vilei A. Evolution of wireless sensor networks towards the internet of things: a survey [C]. Proceedings of the 19th International Conference on Software, Telecommunications and Computer Networks, Split, Croatia, 2011: 1-6.
 - [2] 李文峰, 沈连丰, 胡静. 传感器网络簇间通信自适应节能路由优化方法[J]. 通信学报, 2012, 33(3): 10-19.
Li Wen-feng, Shen Lian-feng, and Hu Jing. Adaptive energy-saving routing algorithm for inter-clustering communications in WSN[J]. *Journal on Communications*, 2012, 33(3): 10-19.
 - [3] Chou J, Petrovic D, and Ramchandran K. A distributed and adaptive signal processing approach to exploiting correlation in sensor networks[J]. *Journal of Ad hoc Networks*, 2004, 2(4): 387-403.
 - [4] Wang Pu, Dai Rui, and Akyildiz I F. Collaborative data compression using clustered source coding for wireless multimedia sensor networks[C]. Proceedings of the 29th Conference on Computer Communications (INFOCOM), San Diego, USA, 2010: 1-9.
 - [5] 李燕君, 叶敬川, 朱艺华. 能量感知的传感器网络分布式时空相关数据收集方案[J]. 北京邮电大学学报, 2011, 34(5): 110-114.
Li Yan-jun, Ye Jing-chuan, and Zhu Yi-hua. An energy-aware distributed protocol for spatial-temporal correlated data gathering in wireless sensor networks[J]. *Journal of Beijing University of Posts and Telecommunications*, 2011, 34(5): 110-114.
 - [6] Wang Tian-qi, Vosoughi A, Heinzelman W, et al. Maximizing gathered samples in wireless sensor networks with slepian-wolf coding[J]. *IEEE Transactions on Wireless Communications*, 2012, 11(2): 751-761.
 - [7] Barcelo-Llado J E, Morell A, and Seco-Granados G. Distributed source coding in large wireless sensor networks[C]. 2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, 2010: 1535-1539.
 - [8] Hua Guo-gang and Chen Chang-wen. Distributed source coding in wireless sensor networks[C]. Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, Orlando, USA, 2005: 1-7.
 - [9] El-Shabrawy T O and Mounir N M. An energy-efficient framework for data aggregation in wireless sensor networks based on distributed source coding[C]. 6th International Computer Engineering Conference, Giza, Egypt, 2010: 56-61.
 - [10] Yuen K, Liang Ben, and Li Bao-chun. A distributed framework for correlated data gathering in sensor networks[J]. *IEEE Transactions on Vehicular Technology*, 2008, 57(1): 578-593.
 - [11] 李莉, 温向明. 无线传感器网络中分簇算法能量有效性分析[J]. 电子与信息学报, 2008, 30(4): 966-969.
Li Li and Wen Xiang-ming. Energy efficient optimization of clustering algorithm in wireless sensor network[J]. *Journal of Electronics & Information Technology*, 2008, 30(4): 966-969.
 - [12] Handy M J, Haase M, and Timmermann D. Low energy adaptive clustering hierarchy with deterministic cluster-head selection[C]. Proceedings of the 4th IEEE International Workshop on Mobile and Wireless Communications Network, Stockholm, Sweden, 2002: 368-372.
 - [13] Heinzelman W, Chandrakasan A, and Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks[C]. Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Hawaii, USA, 2000: 3005-3014.
 - [14] Younis O and Fahmy S. HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks[J]. *IEEE Transactions on Mobile Computing*, 2004, 3(4): 366-379.
 - [15] Slepian D and Wolf J. Noiseless coding of correlated information sources[J]. *IEEE Transactions on Information Theory*, 1973, 19(4): 471-480.
 - [16] Xiong Zi-xiang, Liveris A D, and Cheng S. Distributed source coding for sensor networks[J]. *IEEE Signal Processing Magazine*, 2004, 21(5): 80-94.
 - [17] Pradhan S S and Ramchandran K. Distributed source coding using syndromes (DISCUS): design and construction[C]. Proceedings of Data Compression Conference, Snowbird, USA, 1999: 158-167.
 - [18] Liveris A D, Xiong Zi-xiang, and Georgiades C N. A distributed source coding technique for correlated images using turbo-codes[J]. *IEEE Communications Letters*, 2002, 6(9): 379-381.
 - [19] Sartipi M and Fekri F. Lossy distributed source coding using LDPC codes[J]. *IEEE Communications Letters*, 2009, 13(2): 136-138.
 - [20] He Shi-bo, Chen Ji-ming, Yau D K Y, et al.. Cross-layer optimization of correlated data gathering in wireless sensor networks[J]. *IEEE Transactions on Mobile Computing*, 2012, 11(11): 1-14.
 - [21] Raghunathan V, Schurgers C, Sung P, et al.. Energy-aware wireless microsensor networks[J]. *IEEE Signal Processing Magazine*, 2002, 19(2): 40-50.
- 秦智超: 男, 1981年生, 博士生, 研究方向为无线传感器网络、认知无线电。
- 周正: 男, 1945年生, 教授, 博士生导师, 从事无线通信新技术、信号与信息处理方面研究。
- 赵小川: 男, 1984年生, 博士生, 研究方向为无线传感器网络、压缩感知。