

WSN 中基于压缩感知的高能效数据收集方案

李鹏^{1,2} 王建新¹ 丁长松²

摘要 可靠高效的数据收集是无线传感器网络 (Wireless sensor networks, WSN) 应用中的关键问题. 然而, 由于无线通信链路的高失效率、节点资源受限以及环境恶劣等原因, 网络容易发生丢包问题, 使得现有的数据收集方法无法同时满足高精度和低能耗的要求. 为此, 本文提出了一种基于压缩感知的高能效数据收集方案. 该方案主要分为节点上的数据处理和数据收集路径优化两个步骤. 首先设计了基于指数核函数的稀疏矩阵来对感知数据进行稀疏化处理, 然后综合考虑了数据的传输能耗和可靠性等因素, 采用分块矩阵的思路, 将单位矩阵和准循环低密度奇偶校验 (Low density parity check, LDPC) 码的校验矩阵相结合构造了测量矩阵, 并证明了它与稀疏矩阵之间满足限制等距性质 (Restricted isometry property, RIP). 最后, 将数据收集路径优化问题建模为哈密尔顿回路问题, 并提出了基于树分解的路径优化算法进行求解. 仿真结果表明, 在网络存在丢包的情况下, 本文方案仍然能够保证数据收集的高精确度, 相比于其他数据收集方案而言, 本文方案在数据重构误差和能耗方面的性能更优.

关键词 无线传感器网络, 数据收集, 压缩感知, 树分解, 重构误差, 能耗

引用格式 李鹏, 王建新, 丁长松. WSN 中基于压缩感知的高能效数据收集方案. 自动化学报, 2016, 42(11): 1648–1656

DOI 10.16383/j.aas.2016.c160258

Energy-efficient Data Gathering Scheme Based on Compressive Sensing in Wireless Sensor Networks

LI Peng^{1,2} WANG Jian-Xin¹ DING Chang-Song²

Abstract Reliable and energy-efficient data gathering is a key problem in the application of wireless sensor networks (WSN). However, due to the high failure rate of wireless communication link, limited resource and severe environment, the network easily generates the packet loss problem, which makes the existing data gathering methods fail to meet the requirements of high-accuracy and low-energy consumption at the same time. To solve this problem, an energy-efficient data gathering scheme based on compressive sensing is proposed in this paper. It is divided into the two steps: the data processing of nodes and the data gathering path optimization. The sparse matrix based on the exponential kernel function is firstly designed for sparse processing of sensed data. Then considering both the energy consumption and reliability of data transmission, a measurement matrix is constructed by using the idea of block matrix, which combines the unit matrix and the check matrix of quasi cyclic low density parity check (LDPC) code. It is proved that the restricted isometry property (RIP) is satisfied between the sparse matrix and the measurement matrix. Finally, the data gathering path-optimization problem is modeled as the Hamilton loop problem, and a path optimization algorithm based on the tree decomposition is proposed to solve this problem. Simulation results show that the proposed scheme can still guarantee the high-accuracy of data gathering in the case of packet losses. Compared with the other data gathering schemes, the proposed scheme has better performance in terms of the data reconstruction error and energy consumption.

Key words Wireless sensor networks (WSN), data gathering, compressive sensing, tree decomposition, reconstruction error, energy consumption

Citation Li Peng, Wang Jian-Xin, Ding Chang-Song. Energy-efficient data gathering scheme based on compressive sensing in wireless sensor networks. *Acta Automatica Sinica*, 2016, 42(11): 1648–1656

收稿日期 2016-03-15 录用日期 2016-05-16
Manuscript received March 15, 2016; accepted May 16, 2016
国家自然科学基金 (61472449, 61173169, 61402542) 资助
Supported by National Natural Science Foundation of China (61472449, 61173169, 61402542)
本文责任编辑 赵千川
Recommended by Associate Editor ZHAO Qian-Chuan
1. 中南大学信息科学与工程学院 长沙 410083 2. 湖南中医药大学
管理与信息工程学院 长沙 410208
1. School of Information Science and Engineering, Central
South University, Changsha 410083 2. School of Management
and Information Engineering, Hunan University of Chinese
Medicine, Changsha 410208

无线传感器网络 (Wireless sensor networks, WSN)^[1] 是由一组传感器节点构成的无线自组织网络, 它实时感知、采集各种被监控对象的数据加以分析处理^[2], 然后将处理结果发送到 Sink 上, Sink 再通过互联网或卫星网络等发送数据到达管理节点, 从而为网络决策提供支持. 对于 WSN 中的任何应用而言, 都要求各个节点准确无误地将自身的感知数据发送到 Sink 节点上, 即要保证数据收集的可靠性. 但由于 WSN 通常部署在露天环境中, 容易受到

外部因素的影响以及节点自身资源的限制,网络的运行是不稳定的,经常存在由于节点间通信链路中断或节点死亡导致的数据传输丢包现象,因此,如何保证数据收集的可靠性成为 WSN 面临的一项主要挑战. 已有研究表明^[3-4],为了应对网络通信不可靠而发生的多次重传将额外耗费节点的能量,从而使得网络的真实性能远远达不到预期. 现有的可靠传输方法主要包括:重传机制、前向纠错机制以及多路径传输等,这些方法存在着延时较高、能耗较大的缺点. 例如,传感器节点之间数据传输的丢包率最高可达 30%^[5],如果节点经过 5 跳能将自身的数据传输至 Sink,则传输的成功率大约为 16.8%. 此时,为了保证数据传至 Sink 的可靠性达到 90% 以上,则每跳节点至少需要重传 2 次以上,即节点的能耗要比预先估计至少增大 2 倍. 为此,本文基于压缩感知理论^[6],设计了一种高能效的数据收集方案,在保证数据收集可靠性的前提下,尽可能地延长网络寿命.

1 相关工作

无线传感器网络的数据收集可靠性问题是目前的研究热点. Wu 等^[7]提出一种基于冗余的方法用于实现数据收集的高可靠性和低延时,它采用网络编码的思想来改善数据包的冗余度,并能根据应用需求和链路丢失率自适应地变化数据包的冗余水平,以满足端到端的延时要求. 然而,该方法中节点需要传输的冗余数据量较大,导致了较高的通信能耗. Joo 等^[8]提出基于内网数据融合的无线广播策略来保证数据传输的可靠性,该文利用无线媒介的多样性来应对数据丢包现象,避免了采用重传策略所导致的高延时. 然而,该方法的一个主要问题是存在多节点随机竞争信道所导致的数据包冲突. 文献 [9] 针对网络中由于节点失效或故障导致的频繁丢包现象,综合考虑了节点能耗、数据收集延时和数据收集率来建模数据收集可靠性问题,然后提出了基于 Reed-solomon (RS) 编码的数据收集策略,它以较低的能耗获得了可使节点能耗最优的重传参数和数据包编码数目,可在保证数据收集质量的同时,使网络能耗最小. 然而,该方法在 RS 编码过程中需要多个节点间的信息交换,采用穷举法进行求解,使得算法的复杂性较高,不适用于大规模无线传感器网络. 此外,文献 [10] 提出一种基于压缩感知和纠错码的数据收集算法 (Compressive sensing erasure encoding, CSEC),首先基于伪随机采样策略构造得到初始的测量矩阵,然后将二进制删除信道建模为丢失概率为 p 的贝努利分布,最后基于信道丢失概率自适应地增加测量矩阵的行数,即采用过采样来实现对节点数据的投影操作,从而保证即使投影值的传输出现丢失也能在 Sink 处精确地恢复出原始

数据. 但是,随着网络规模的增加,CSEC 方案仍然需要采集较多的数据才能保证数据重构质量,导致了较大的能耗,使得利用压缩感知进行数据收集的优势无从体现. 文献 [11] 分析了 CDG 算法^[12]在网络中存在丢包情形下无法有效实现数据收集的不足,设计了一种基于最稀疏随机调度的数据收集方案 (Sparsest random scheduling, SRS),该方案不仅降低了数据传输开销,且在链路丢包率达到 15% 的情况下仍然能够保证精确的数据收集性能. 但是,SRS 方案还不够完善,它主要利用地理位置相近的节点间感知数据的相关性来保证数据收集精度,没有考虑节点感知数据的时间相关性对于数据收集质量的影响. 因此该方案容易受到网络拓扑结构变化的影响,其应用的局限性较大. 为了弥补以上方法的缺陷,本文提出了一种基于压缩感知的高能效数据收集方案,文中首先通过稀疏矩阵和测量矩阵的设计保证节点上数据采集的可靠性,然后通过和数据收集路径进行优化达到节省网络能量的目的.

2 问题建模

2.1 系统模型

设 1 个 Sink 节点和 N 个传感器节点被随机地部署在一个大小为 $N \times N$ 平方米的监测区域中. 所有节点之间构成一个动态连通的无向图 $G = (V, E)$,其中 V 是传感器节点集合, $V = \{V_1, V_2, \dots, V_N\}$, $|V| = N + 1$; E 是图中边的集合,如果任意的两个节点 V_i 和 V_j 能够相互通信,则有 $(V_i, V_j) \in E$. 节点无需知晓自身的地理位置,节点通过向其周围发送交换 1 个 Hello 消息来获取其邻居节点的信息. 此外,WSN 还具有如下的特性:

1) Sink 节点和其他所有的传感器节点在部署后保持静止不动,传感器节点的传输半径相同. 网络中存在由于节点间通信链路中断或节点失效所导致的传输丢包现象,且无法预测丢包发生在哪些传感器节点上.

2) 传感器节点之间的感知数据具有时空相关性. 采用 DEBUC 协议^[13]对网络进行分簇,簇内各节点基于压缩感知理论对感知数据进行采样并将其发送到各自的簇头.

3) 节点的初始能量是异构的,而且不能补充,网络中节点采用的能量消耗模型如下:

$$E = (E_{tr} + E_{re} + E_{amp} \times d^2) \times packet_{size} \quad (1)$$

其中, E_{tr} 和 E_{re} 分别为传输电路和接收电路消耗的能量, E_{amp} 为多路衰减模型的功率放大系数, d 表示源节点到目标节点的距离, $packet_{size}$ 为包的大小. 所有节点采用相同的发射功率和接收功率.

2.2 相关定义

为了方便描述, 给出本文用到的相关定义.

定义 1. 限制等距性质 (Restricted isometry property, RIP)^[6]. 对于任意的信号 $x \in \Sigma_k$ (Σ_k 表示 k 稀疏向量集合), 如果存在常数 $\delta_k \in (0, 1)$, 有下式成立:

$$(1 - \delta_k)\|x\|^2 \leq \|\phi x\|^2 \leq (1 + \delta_k)\|x\|^2 \quad (2)$$

则称矩阵 ϕ 满足 k 阶 RIP 性质.

定义 2. 图的树分解^[14]. 对于任意一个无向图 $G(V, E)$ 和树 T , 设 $\mathcal{U} = (\mathcal{U}_i)_{i \in V(T)}$ 表示顶点集 $(\mathcal{U}_i) \subseteq V(G)$ 的包, 图 G 的树分解可由树 T 和 T 的每一个节点关联的子集构成, 即 $\Gamma = (T, \mathcal{U})$. 当且仅当 T 和 \mathcal{U} 满足以下三个条件:

- 1) 顶点覆盖: $V(G) = \cup_{i \in V(T)} \mathcal{U}_i$;
- 2) 边覆盖: 对于任意的 $e \in E(G)$, 总是存在一个节点 $i \in V(T)$ 使得 e 的两端点都在 \mathcal{U}_i 中;
- 3) 一致性: 如果 i_2 是树 T 中连接节点 i_1 和 i_3 的路径中的节点, 则有 $\mathcal{U}_{i_1} \cap \mathcal{U}_{i_3} \subseteq \mathcal{U}_{i_2}$.

定义 3. 亚高斯分布^[15] 给定任意的随机变量 ω , 如果存在一个常数 $c > 0$, 使得对于任意的 $y \in \mathbf{R}$, 有下面的不等式成立:

$$E(\exp(\omega y)) \leq \exp\left(\frac{c^2 y^2}{2}\right) \quad (3)$$

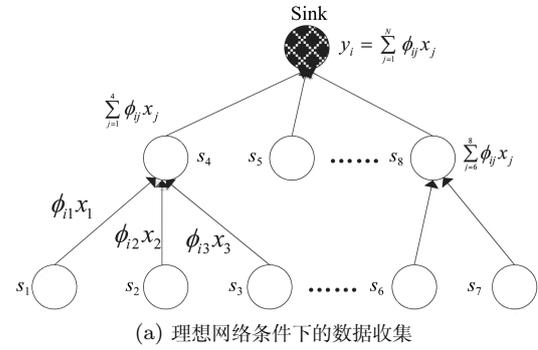
则称该随机变量 ω 服从亚高斯分布, 即 $\omega \sim \text{Sub}(c^2)$.

2.3 问题描述

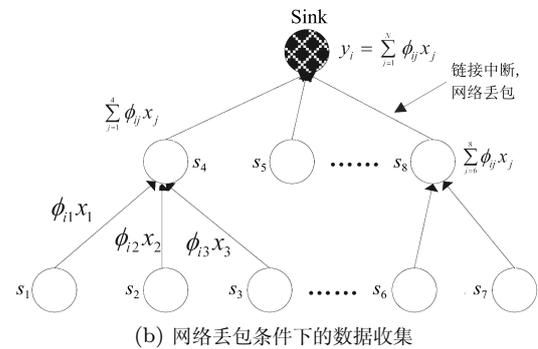
WSN 一般部署在比较危险或人类很难进入的区域, 这些区域中的无线传感器节点在部署后就无法进行更换, 因此如何让这些节点能够工作尽可能长的时间显得至关重要. 而采用压缩感知理论进行数据收集^[16-17] 能够降低节点上的数据采集量, 实现网络负载均衡, 进而延长网络生命周期. 其基本过程为: 首先将所有节点的原始数据表示为 $N \times 1$ 的列向量 $x(n) \in \mathbf{R}^N$. 由于网络数据的时空相关性, x 在某一变换基 Ψ 上可被稀疏化为 $x = \sum_{i=1}^N s_i \Psi_i = \Psi s$; 然后采用一个与 Ψ 不相关的测量矩阵 $\phi(M \times N)$ 对 x 进行测量, 得到 $M \times 1$ 的测量值 $y = \phi x$; 最后, 当 Sink 节点收到 M 个测量值后, 通过求解 l_1 范数最小化问题就能精确地重构出 x , 如图 1(a) 所示.

然而, 以上的基于压缩感知的数据收集过程只适用于理想条件下的 WSN, 当 WSN 中存在由于外部干扰或节点失效等原因导致的通信中断使得数据传输发生丢包 (见图 1(b)) 时, 现有的方法无法保证数据收集质量. 特别当丢包现象发生在网络中较为重要的节点 (例如度数较高的节点、离 Sink 较近的

节点等) 上时, 则会严重影响到数据收集应用的可靠性. 为此, 本文对网络丢包条件下的数据收集可靠性问题进行了研究, 依据图 1(b) 所示的网络模型, 研究在采用压缩感知理论进行数据收集的过程中如何应对传输丢包现象, 以及如何对数据收集路径进行优化, 从而在保证数据收集质量的同时, 尽可能地延长网络生命周期.



(a) Data gathering under the conditions of ideal network



(b) Data gathering under the conditions of network packet loss

图 1 基于压缩感知的数据收集

Fig. 1 Data gathering based on compressive sensing

3 节点上的数据处理

3.1 稀疏矩阵设计

在压缩感知理论中, 信号表示越稀疏, 则信号的恢复就越准确、越可靠. 文献 [18] 已经证明光滑信号的离散余弦变换、小波变换以及振荡信号的 Gabor 变换等都具有足够的稀疏性, 可以通过压缩感知恢复信号. 然而由于传感器节点的种类繁多, 可以采集具有复杂特征的多种信号 (例如图像、声音、温度或光照等), 固定的正交基有时不足以捕获信号的多种特征使信号在变换域足够稀疏, 从而影响了信号重构精度. 例如, 正交小波变换由于缺乏平移旋转不变性而不能有效压缩几何图像数据. 为了弥补现有方法的不足, 本文提出一种基于指数核函数的稀疏矩阵来对传感器网络中的感知数据进行稀疏化, 提高

了数据稀疏表示的可靠性。

根据第 2.1 节所示的节点间的感知数据具有时空相关性这一假设, 对于任意两个传感器节点 i 和 j 而言, 本文采用如下的指数核函数 $\kappa(x_i, x_j)$ 来衡量 i 和 j 之间感知数据的相关性:

$$\kappa(x_i, x_j) = \exp\left(\frac{-d_{ij}}{2\sigma^2}\right) \quad (4)$$

其中, d_{ij} 表示节点 i 和节点 j 之间的距离, σ 表示指数核函数的宽度参数, 主要用于控制该函数的径向作用范围. 它的取值可以采用交叉验证法、梯度下降法或最大似然法等估计得到. 指数核函数是高斯核函数的变种, 它相对于高斯核函数而言, 对于参数的依赖程度较低. 对式 (4) 进行推广可得, 无线传感器网络中所有 N 个节点之间的感知数据相关性可以用矩阵 R 进行表示:

$$R = \begin{pmatrix} e^{-\frac{d_{11}}{2\sigma^2}} & e^{-\frac{d_{12}}{2\sigma^2}} & \cdots & e^{-\frac{d_{1N}}{2\sigma^2}} \\ e^{-\frac{d_{21}}{2\sigma^2}} & e^{-\frac{d_{22}}{2\sigma^2}} & \cdots & e^{-\frac{d_{2N}}{2\sigma^2}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\frac{d_{N1}}{2\sigma^2}} & e^{-\frac{d_{N2}}{2\sigma^2}} & \cdots & e^{-\frac{d_{NN}}{2\sigma^2}} \end{pmatrix} \quad (5)$$

从式 (5) 可以看出, R 属于 T 型矩阵 (Toeplitz). 根据 T 型矩阵性质可知, 对 R 做对角化处理可得 $R = \Psi_R \Gamma \Psi_R^{-1}$. 其中, Γ 是对角矩阵, 其取值由 R 的特征值构成. Ψ_R 是一个正交基函数, 其取值由 R 的正交特征向量构成, 它能够满足压缩感知理论中的信号稀疏表示要求. 因此, 本文将 Ψ_R 作为稀疏矩阵来对 WSN 中的所有感知数据进行稀疏化处理, 即有 $x = \Psi_R s$.

3.2 测量矩阵设计

测量矩阵设计是基于压缩感知的数据收集方法的关键步骤. 现有的大多数方法^[11-12] 在测量矩阵的设计过程中关注的重点是降低数据重构误差和减少数据传输开销, 达到了节省节点能耗的目的, 但是当网络中存在传输丢包时, 这些方法不仅无法保证数据收集质量, 而且还会导致能量的浪费. 假设图 1(b) 为某一个簇内节点组成的数据收集树, 从图中可以看到, 如果叶子节点的感知数据丢失, 根据时空相关性可知, 这一部分数据可由与其物理位置相近的其他节点感知数据代替. 但如果靠近 Sink 或节点度较大的节点发生丢包, 则会严重损害数据重构质量, 因为这部分节点除了传输自身的数据外还需承载其他节点的数据. 为此, 本文综合考虑了数据的传输能耗和可靠性等因素, 对数据收集树中的叶子节点和非叶子节点分别进行不同的处理来设计测量矩阵, 具体过程如下:

1) 对于叶子节点而言, 它们的数据被直接收集上来并将其发往上游节点. 相当于采用一个单位矩阵 I 对数据收集树中的所有叶子节点做压缩采样.

2) 对于非叶子节点而言, 则将其信道矩阵看作是测量矩阵的一部分, 由于信道编码校验矩阵列向量之间满足线性无关性, 且通常可稀疏设计. 因此, 本文将信道编码中的校验矩阵用于压缩感知的测量矩阵设计中, 提出一种基于准循环低密度奇偶校验 (Low density parity check, LDPC 码)^[19] 的方法来构造测量矩阵 ϕ' .

综上所述, 本文要设计的测量矩阵为 $(I|\phi')$.

3.2.1 基于准循环 LDPC 码的测量矩阵构造方法

准循环 LDPC 码是一种具有稀疏校验矩阵的分组纠错码. 它的性能逼近香农限, 描述和实现简单, 且可并行操作, 适合硬件实现. 已有研究表明^[19], 对于一个任意给定的准循环 LDPC 码, 如果它的校验矩阵能够实现良好的信道编码或解码性能, 则将其的校验矩阵作为基于压缩感知的信号处理中的测量矩阵, 也能实现精确的数据重构. 为此, 本文根据准循环 LDPC 码校验矩阵的正交性和稀疏性来设计测量矩阵. 设 SM 是大小为 $S \times S$ 的单位矩阵的 1 次循环移位阵

$$SM = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad (6)$$

则 SM^i 是单位阵的第 i 次循环移位阵, $0 \leq i < S$. SM^∞ 为零方阵. 设 $M \times N$ 的矩阵 CM 为

$$CM = \begin{pmatrix} SM^{a_{11}} & SM^{a_{12}} & \cdots & SM^{a_{1N}} \\ SM^{a_{21}} & SM^{a_{22}} & \cdots & SM^{a_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ SM^{a_{M1}} & SM^{a_{M2}} & \cdots & SM^{a_{MN}} \end{pmatrix} \quad (7)$$

其中, $a_{ij} \in \{0, 1, \dots, S-1, \infty\}$, $1 \leq i \leq M$, $1 \leq j \leq N$, 则以 CM 为校验矩阵的码字 C 称为准循环 LDPC 码. 对于网络存在丢包情况下基于压缩感知的数据收集应用而言, 为了保证数据收集质量和节约网络能耗, 测量矩阵的设计必需考虑稀疏性、高纠错性等因素, 而采用准循环 LDPC 码来设计测量矩阵恰好能够满足这一要求. 因此, 本文设计了一种基于准循环 LDPC 码的测量矩阵构造算法.

算法 1. 基于准循环 LDPC 码的测量矩阵构造

步骤 1. 构造一个由 $M \times N$ 个大小为 $S \times S$ 的零方阵

组成矩阵 CM' .

步骤 2. 随机选择 a_{i1} 和 a_{1j} 构造移位矩阵 $SM^{a_{i1}}$ 和 $SM^{a_{1j}}$, $0 \leq a_{i1}, a_{1j} < S$, $0 < i \leq M$, $1 < j \leq N$, 并用 $SM^{a_{i1}}$ 和 $SM^{a_{1j}}$ 替代 CM' 中对应位置的零方阵.

步骤 3. 随机选择 a_{ij} 构造移位矩阵 $SM^{a_{ij}}$, $0 \leq a_{ij} < S$, 并用 $SM^{a_{ij}}$ 替代 CM' 中对应位置的子矩阵.

步骤 4. 令

$$\widetilde{CM}' = \begin{pmatrix} SM^{-a_{11}} & SM^{-a_{12}} & \cdots & SM^{-a_{1N}} \\ SM^{-a_{21}} & SM^{-a_{22}} & \cdots & SM^{-a_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ SM^{-a_{M1}} & SM^{-a_{M2}} & \cdots & SM^{-a_{MN}} \end{pmatrix}$$

计算 \widetilde{CM}'^T , CM' , 如果有 $a_{ij} \geq 2$, 则转步骤 3, 重新选择 a_{ij} ; 否则, 在矩阵 CM' 中自左至右逐列按照“下-右-上-右-下”的顺序构造下一个移位矩阵, 并用其替代 CM' 中对应位置的零方阵, 直到构造出合适的移位矩阵 $CM^{a_{mn}}$, 从而获得大小为 $M \times N$ 的二进制稀疏校验矩阵 CM .

步骤 5. 将校验矩阵 CM 列向量进行归一化处理, 得到标准正交基. 然后抽取 M 个线性无关向量填充压缩感知测量矩阵 ϕ' 的前 M 个列向量, 即: $[\phi'_1, \phi'_2, \dots, \phi'_M] = [CM_1, CM_2, \dots, CM_M]$, 最后通过前 M 列的线性组合来表示测量矩阵中的剩余 $N - M$ 个列向量, 从而得到 ϕ' .

在算法 1 的步骤 4 中, 当 $a_{ij} \geq 2$ 时重新选择 a_{ij} , 这是因为 LDPC 编码算法经过 2 次迭代后, LDPC 译码无法收敛或收敛速度太慢, 因此需要回溯. 在步骤 5 中, CM 的秩为 $|M|$, 因此总是可以保证至少存在 M 个线性无关向量. 此外, 算法 1 的主要开销在于构造准循环 LDPC 码的校验矩阵 CM , CM 属于线性分组码, 在编码过程中, CM 需采用高斯消去法转换为 $(CM''|I)$ 形式的矩阵. 其中, 单位矩阵 I 是 M 阶, CM'' 是 $(N - M) \times M$ 阶. 因为对 CM 进行高斯消去后, 其中的元素不再是稀疏的, 因此构造校验矩阵的复杂度为 $(N - M)M/2$, 即算法 1 的复杂度为多项式时间, 从节省传感器节点能耗来看, 算法 1 的开销是可以接受的.

3.2.2 RIP 性质

设 $\Theta = \Psi_R \phi$, 压缩感知理论告诉我们, 为了精确地重构出原始数据, Θ 必须满足 RIP 性质.

定理 1. 对于任意给定的 $\sigma \in (0, 1)$, 矩阵 Θ 的每行满足亚高斯分布, 如果 $M = O(k \log(N/k))$, 则对于任意 N 维的 k 稀疏信号 s 而言, Θ 能以较高概率满足 $(1 - \delta) \leq \frac{\|\Theta s\|_2^2}{\|s\|_2^2} \leq (1 + \delta)$, 即 Θ 满足 RIP 性质.

证明. 首先对 Θ 进行归一化处理, 有 $\Theta = \sqrt{N/M}[\lambda_1, \lambda_1, \dots, \lambda_M]^T$, $E(\lambda_{ij}) = 0$ 和 $\text{var}(\lambda_{ij}) = E(\lambda_{ij}^2) = 1/N$. 根据内积定义可得:

$$E\left(\left\langle \sqrt{\frac{N}{M}}\lambda_i, s \right\rangle\right) = E\left(\sqrt{\frac{N}{M}}\sum_{j=1}^N \lambda_{ij}s_j\right) = \sqrt{\frac{N}{M}}E(\lambda_{ij})s_j = 0 \quad (8)$$

$$\text{var}\left(\left\langle \sqrt{\frac{N}{M}}\lambda_i, s \right\rangle\right) = \text{var}\left(\sqrt{\frac{N}{M}}\sum_{j=1}^N \lambda_{ij}s_j\right) = \sqrt{\frac{N}{M}}\sum_{j=1}^N \text{var}(\lambda_{ij})s_j^2 = \frac{1}{M}\sum_{j=1}^N s_j^2 = \frac{\|s\|_2^2}{M} \quad (9)$$

$$E(\|\Theta s\|_2^2) = E\left(\sum_{i=1}^M \left(\left\langle \sqrt{\frac{N}{M}}\lambda_i, s \right\rangle\right)^2\right) = \sum_{i=1}^M E\left(\left\langle \sqrt{\frac{N}{M}}\lambda_i, s \right\rangle\right)^2 = \sum_{i=1}^M \text{var}\left(\left\langle \sqrt{\frac{N}{M}}\lambda_i, s \right\rangle\right) = \sum_{i=1}^M \frac{\|s\|_2^2}{M} = \|s\|_2^2 \quad (10)$$

根据定义 3 可知, 对于任意给定的一个 M 维随机向量 $l = [l_1, l_2, \dots, l_M]^T$, 如果其中的每一个 $l_i \sim \text{Sub}(c^2)$ 和 $E(l_i^2) = \sigma^2$, 则对于 $\forall \alpha, \beta \in (0, 1)$, 存在一个常数 Ω 使得:

$$P(\|l\|_2^2 \leq \alpha M \sigma^2) \leq e^{-\frac{M(1-\alpha)^2}{\Omega}}$$

$$P(\|l\|_2^2 \geq \beta M \sigma^2) \leq e^{-\frac{M(\beta-1)^2}{\Omega}}$$

设 $\alpha = 1 - \delta$ 和 $\beta = 1 + \delta$, 则有如下的不等式^[11]成立:

$$P\left(\frac{\|\Theta s\|_2^2}{\|s\|_2^2} \leq 1 - \delta\right) \leq e^{-\frac{M\delta^2}{\Omega}} \quad (11)$$

$$P\left(\frac{\|\Theta s\|_2^2}{\|s\|_2^2} \geq 1 + \delta\right) \leq e^{-\frac{M\delta^2}{\Omega}} \quad (12)$$

$$P\left(\left|\frac{\|\Theta s\|_2^2}{\|s\|_2^2} - 1\right| \geq \delta\right) \leq 2e^{-\frac{M\delta^2}{\Omega}} \quad (13)$$

对于本文设计的 Θ 而言, 它有 (N, k) 个 k 维子空间. 根据斯特林公式^[11]可知, $(N, k) \leq (eN/k)^k$, k 稀疏信号 s 满足 $\left|\frac{\|\Theta s\|_2^2}{\|s\|_2^2} - 1\right| \geq \delta$ 的概率为

$$P\left(\left|\frac{\|\Theta s\|_2^2}{\|s\|_2^2} - 1\right| \geq \delta\right) = \left(\frac{N}{k}\right)^k \times 2e^{-\frac{M\delta^2}{\Omega}} = 2e^{-\frac{M\delta^2}{\Omega}} + k \log\left(\frac{N}{k}\right) + 1 \quad (14)$$

因此, 当 $M = O(k \log(N/k))$ 时, 对于任意 N 维的 k 稀疏信号 s , Θ 以接近于 1 的概率满足

$$(1 - \delta) \leq \frac{\|\Theta s\|_2^2}{\|s\|_2^2} \leq (1 + \delta)$$

□

3.3 簇内数据收集

综上所述, 簇内数据处理过程为: 在采用 DE-BUC 协议对 WSN 进行分簇后, 各个传感器节点依据第 3.1 节设计的稀疏矩阵 Ψ_R 对感知数据进行稀疏化, 依据第 3.2 节设计的测量矩阵 ϕ 决定是直接采样还是压缩采样, 然后将自身的数据发往所属的簇头 (网络中各个簇都采用类似的处理). 最后, 整个网络的感知数据都被集中到簇头上.

4 数据收集路径优化

4.1 路径分析

在节点上的数据处理过程完毕后, 下一步则需要为各个簇头找一条到达 Sink 的最优路径, 以将各个簇头上的数据收集上来, 并采用压缩感知重构算法恢复出各个传感器节点的数据, 以完成数据收集过程. 能否找到一条最优的路径来实现簇头与 Sink 之间的数据传输, 对于节省网络传输开销, 延长网络生命周期至关重要. 考虑一个包含 6 个簇头 $CH1, CH2, CH3, CH4, CH5, CH6$ 和一个 Sink 的无线传感器网络, 如图 2 所示. 假设各个簇头上待传输的感知数据为 $x_i, i = 1, \dots, 6$. 各个簇头的投影向量为 $\phi = [0.1, 0.3, 0.15, 0.25, 0.05, 0.15]$, 则可以采用路径 $Sink - CH1 - CH4 - CH6 - CH5 - CH3 - CH2 - Sink$ 来收集整个网络中的测量值, 这可以通过由 Sink 向整个网络广播该条路径信息来实现, 即有 $y = 0.1x_1 + 0.3x_4 + 0.15x_6 + 0.25x_5 + 0.05x_3 + 0.15x_2$. 从图 2 可以看出, 可以采用多条路径来收集整个网络中簇头的数据到 Sink 上, 在这些路径中, 某些簇头可能被多次访问到, 额外增加了不必要的数据传输开销, 浪费了网络能量. 因此, 从节省网络能耗的角度出发, 最优的数据传输路径是: 从 Sink 节点出发, 只需访问各个簇头一次再回到 Sink 节点的路径. 如何找到一条这样的路径是一个典型的哈密顿回路问题^[20], 属于 NP (Non-deterministic polynomial) 困难问题. 但对于本文研究的数据收集问题而言, 由于待访问的簇头数目是有限的, 因此我们提出了一种基于树分解的数据收集路径优化算法进行解决.

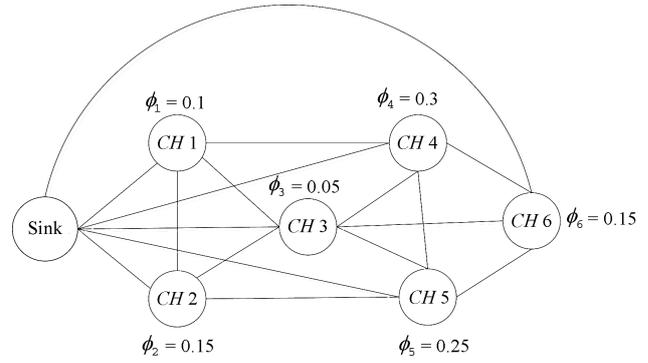


图 2 测量值传输

Fig. 2 Measurement transmission

4.2 优化算法

根据定义 2 可知, 采用树分解技术可以将任意一个图中的节点划分为相互关联的节点集合. 在求解某些较为复杂的图问题时, 只要给定的图满足有限树宽条件, 就可以先对图进行树分解, 然后在其分解得到的各部分节点集上单独求解, 再将各部分求解结果综合起来即可. 自然世界中用图表示的诸多问题虽然其规模十分庞大, 但是通常只有很小的树宽. 例如, 一个包含 1000 个传感器节点的 WSN 的树宽仅等于 $4^{[21]}$. 因此, 利用树分解可以很容易解决以该网络为背景的各种复杂问题. 在本节研究的问题中, 为了节省簇头上的测量值传输能耗, 首先对各个簇头构成的连通图进行树分解, 然后采用动态规划来为测量值的传输找到一条最优路径. 具体过程如算法 2.

算法 2. 基于树分解的数据收集路径优化

输入. 由簇头组成的无向连通图 $G' = (V', E')$.

输出. 测量值传输路径 P .

步骤 1. Sink 发送一个查询包到各个簇头, 各个簇头根据 Sink 的广播路径返回各自的测量值. 如果只存在唯一的路径 P 来传输各个簇头的测量值到 Sink 则返回 P ; 否则, 转步骤 2~4.

步骤 2. 在图 $G' = (V', E')$ 中删除一个节点 v' 和与 v' 相连的边, 并在余下的图中补充边, 使得 v' 原来的邻居节点构成一个团. 然后, 采用最大势搜索策略^[22] 逐个消去图 $G' = (V', E')$ 的所有节点, 可以得到图 $G' = (V', E')$ 的非平凡树分解 $\Gamma = (T, \mathcal{U})$, 树宽为 k . 对于每一个节点 $i \in \mathcal{U}$, $V_i = \cup \mathcal{U}_j, j = i$ 或 j 是 i 的孩子节点, $E_i = E' \cap (V_i \times V_i)$.

步骤 3. 对每个节点 $i \in \mathcal{U}$, 计算包含节点 i 参与的所有回路集合的哈希表 $HT_i(i, \theta), \theta \subseteq \mathcal{U}_i, |HT_i| \leq 2^{k+1}$.

步骤 4. 对于 $\theta \subseteq \mathcal{U}_i$, 在 $HT_i(i, \theta)$ 中找到一条最短的回路 P 使得 $\mathcal{U}_i \cap P = \theta$, 不存在回路则设置 $HT_i(i, \theta) = -\infty$; 返回 P .

步骤 5. 对于所有的节点 $i \in \mathcal{U}$, 按自下向上的顺序计算 $HT_i(i, \theta)$, 重复执行步骤 2~4, 直到 \mathcal{U} 中所有的节点都处理完毕.

5 仿真实验

为验证本文方案的性能,采用 CitySee 系统^[23]测得的温度数据进行仿真实验.采用 Matlab 2012 作为仿真工具,实验过程中的主要参数设定如表 1 所示.

表 1 仿真实验的主要参数
Table 1 The main parameters in simulation

参数	取值
数据包长度	1000 b
发送 1 bit 数据所需的能量	100 nJ
接收 1 bit 数据所需的能量	5 nJ
单个节点每次数据收集所需能量	1000 nJ
数据重构算法	OMP 算法

测试所使用的软硬件环境如下: Inter (R) Core (M) i3-3240 CPU 3.40 GHz, 500 GB 硬盘, 4.0 GB 内存, Microsoft Windows 7 Professional. 在本文的仿真中,节点间的同步通过物理层和数据链路层来实现,其中,数据链路层采用 IEEE 802.15.4 标准中的 CSMA/CA 机制进行空闲侦听和冲突避免,物理层采用 2.4 GHz 频段,其数据传输率为 250 kb/s. 以重构误差和能耗作为评价指标,主要对比分析本文方案与 SRS 方案和 CSEC 方案在不同场景下的实验结果.取本文方案 50 次仿真运行结果的平均值作为最终的实验结果.其中,数据重构误差采用下式衡量:

$$RE = \frac{\|x - \hat{x}\|_2}{\|x\|_2} \quad (15)$$

其中, x 表示原始数据, \hat{x} 是其重构值.

图 3 给出了在不同的丢包率 ($pl_{ratio} = 0, 0.05, 0.10, 0.15$) 条件下本文方案的重构误差情况.可以看到,随着测量次数的增加,无论丢包率如何,重构误差的总体趋势都在下降.其中 $pl_{ratio} = 0$ 表示理想状态下的传感器网络,此时的网络不发生丢包,在该情况下本文方案的重构性能最优.另外,仔细观察不同丢包率下的重构误差曲线可以发现,当测量次数增加到 600 次后,不同丢包率下的重构误差越来越接近,甚至当测量次数达到 1000 次时,四种情形下的重构误差基本达到一致,这充分表明了本文方案在网络发生丢包情况下进行数据收集的可靠性,此时就算网络发生较大规模的丢包,只要适当地增加测量次数,本文方案仍然能够保证精确地重构出网络中各节点的原始数据.

图 4 给出了本文方案与 SRS 方案和 CSEC 方案的重构误差比较情况.图 4(a) 首先给出了测量次

数为 600 次时,不同丢包率 pl_{ratio} 下三种方案的重构误差实验结果.从图 4(a) 可以看到,对于三种数据收集方案而言,其重构误差都随着丢包率的增加而增加,但本文方案的重构误差始终低于 SRS 方案和 CSEC 方案,且随着网络丢包率的不断增加,本文方案的性能优势越来越明显.特别地,当网络丢包率为 0.15 时,就重构误差而言,本文方案比 SRS 方案低 42.6%,比 CSEC 方案低 31.2%.表明本文方案在网络存在丢包的情形下仍然能够实现有效的数据收集,算法的鲁棒性较强.究其原因,主要是因为 SRS 方案采用稀疏随机调度应对网络丢包,当网络规模较大、丢包现象较为频繁时,该方案的稀疏矩阵会导致部分节点的数据丢失,影响了数据重构精度.CSEC 方案则基于信道丢失概率增加测量矩阵的行数应对网络丢包,并对不同信道模型下的丢包现象做了自适应处理,因此取得了比 SRS 方案更好的重构性能.本文方法主要针对 SRS 方案和 CSEC 方案的不足进行改进,以保证数据收集可靠性这一目标设计测量矩阵,并对数据收集路径做了一定的优化,因此进一步提高了数据重构精度.图 4(b) 给出了丢包率为 0.1 时,不同测量次数下三种方案的重构误差实验结果.从图 4(b) 可以看到,在丢包现象存在的前提下,本文方案的重构性能始终优于其他两种方案.特别地,为了达到数据重构精度为 0.03,本文方案需要花费约 600 次测量,而 SRS 方案和 CSEC 方案分别需要花费约 800 次测量和 900 次测量,表明本文方案在保证数据收集高可靠性的前提下,比 SRS 方案和 CSEC 方案的效率更高,成本更低.

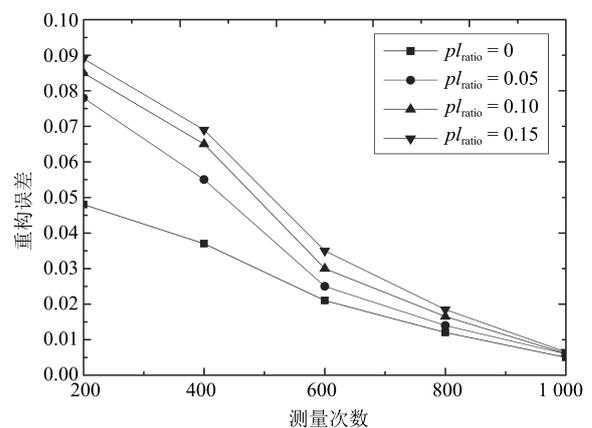
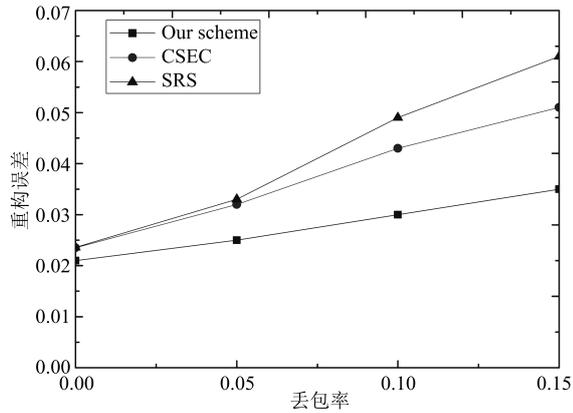


图 3 不同丢包率下的本文方法重构误差比较

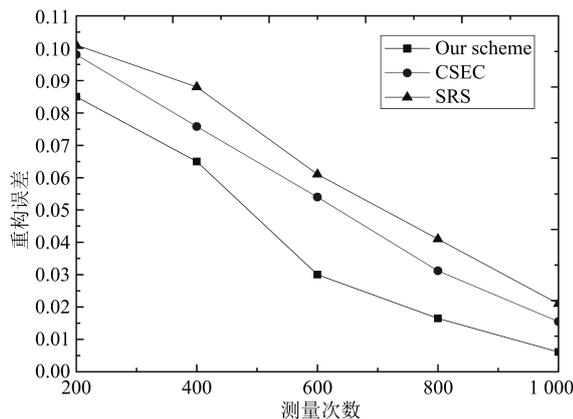
Fig. 3 Reconstruction error comparison of the proposed scheme with different packet loss rates

图 5 给出了本文方案与 SRS 方案和 CSEC 方案的能耗比较情况.从图 5 可以看出,网络能耗与重构误差成反比关系,因为如果要保证重构误差尽可能低,节点通常需要采集更多的数据,导致网络的传



(a) 不同丢包率下三种方案的数据重构误差比较

(a) Reconstruction error comparison of the three schemes with different packet loss



(b) 不同测量次数下三种方案的数据重构误差比较

(b) Reconstruction error comparison of the three schemes with different measurement times

图 4 不同方案的重构误差比较

Fig. 4 Reconstruction error comparison of different schemes

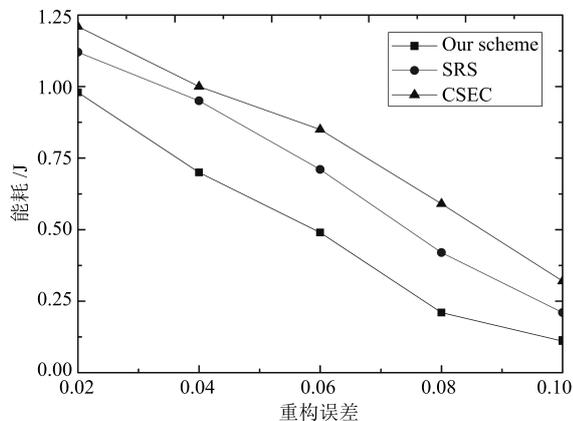


图 5 不同方案的能耗比较

Fig. 5 Energy consumption comparison of different schemes

输开销加大, 能耗增高. 但无论纵向对比还是横向对比, 本文方案的能耗或重构误差都低于 SRS 方案和 CSEC 方案. 就能耗而言, 当应用要求的重构误差为 0.1 时, 本文方案比 SRS 方案节能 41.6%, 比 CSEC 方案节能 56.1%. 究其原因, SRS 方案将每一个采样值看作一次测量, 始终保持测量矩阵的每一行只有一个非零元素, 这种稀疏随机调度的采样方式以牺牲一定的重构精度为代价来保持网络能量. CSEC 方案通过增加测量矩阵的采样次数保证数据收集可靠性, 一定程度上导致了冗余采集, 因此 SRS 方案比 CSEC 方案更加节能. 本文方案更进一步, 将信道矩阵看作是测量矩阵的一部分, 设计了基于准循环 LDPC 码的测量矩阵对网络内的数据进行压缩收集, 减少了节点的传输开销, 然后提出一种基于树分解的数据收集路径优化算法来收集各个簇头的的数据, 保证每个簇头的的数据只被收集一次, 因此节省了能量, 延长了网络寿命.

6 结束语

在无线传感器网络的诸多应用中, 均要求传感器节点以无差错形式将监测数据传输至 Sink 节点. 为了满足应用需求, 本文以压缩感知理论为基础, 提出了一种高能效的数据收集方案. 该方案首先通过设计的稀疏矩阵和测量矩阵完成节点上的数据处理, 然后采用树分解和动态规划的思路优化数据收集路径. 仿真结果表明, 在网络存在丢包的情况下, 本文方案仍然能够保证数据收集的高精确度, 相比于其他数据收集方案, 本文方案的性能更优. 在下一步工作中, 我们关注的重点是: 1) 对本文的网络场景假设做出改变, 以提高数据收集精度和降低网络能耗为目标, 研究移动 Sink 条件下基于压缩感知的数据收集方案; 2) 对真实无线传感器网络中节点感知数据的稀疏性进行分析, 研究基于压缩感知的自适应数据重构算法.

References

- Guo S, He L, Gu Y, Jiang B, He T. Opportunistic flooding in low-duty-cycle wireless sensor networks with unreliable links. *IEEE Transactions on Computers*, 2014, **63**(11): 2787–2802
- Li Jian-Zhong, Gao Hong. Survey on sensor network research. *Journal of Computer Research and Development*, 2015, **45**(1): 1–15
(李建中, 高宏. 无线传感器网络的研究进展. *计算机研究与发展*, 2015, **45**(1): 1–15)
- Luo H, Tao H X, Ma H D, Das S K. Data fusion with desired reliability in wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 2011, **22**(3): 501–513
- Guo W Z, Hong W, Zhang B, Chen Y Z, Xiong N X. Reliable adaptive data aggregation route strategy for a trade-off between energy and lifetime in WSNs. *Sensors*, 2014, **14**(9): 16972–16993

- 5 Rosberg Z, Liu R P, Le Dinh T, Dong Y F, Jha S. Statistical reliability for energy efficient data transport in wireless sensor networks. *Wireless Networks*, 2010, **16**(7): 1913–1927
- 6 Xie Zhi-Peng, Chen Song-Can. CSMP: compressive sensing matching pursuit based on restricted isometry property. *Journal of Computer Research and Development*, 2015, **49**(3): 579–588
(谢志鹏, 陈松灿. CSMP: 基于约束等距的压缩感知匹配追踪. 计算机研究与发展, 2015, **49**(3): 579–588)
- 7 Wu C, Ji Y S, Xu J, Ohzahata S, Kato T. Coded packets over lossy links: a redundancy-based mechanism for reliable and fast data collection in sensor networks. *Computer Networks*, 2014, **70**(10): 179–191
- 8 Joo C, Shroff N B. On the delay performance of in-network aggregation in lossy wireless sensor networks. *IEEE/ACM Transactions on Networking*, 2014, **22**(2): 662–673
- 9 Chou C T, Ignjatovic A, Hu W. Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults. *IEEE Transactions on Parallel and Distributed Systems*, 2013, **24**(8): 1525–1534
- 10 Charbiwala Z, Chakraborty S, Zahedi S, He T, Bisdikian C, Kim Y, Srivastava M B. Compressive oversampling for robust data transmission in sensor networks. In: Proceedings of the 29th IEEE Conference on Computer Communications (INFOCOM). San Diego, CA, USA: IEEE Press, 2010. 1–9
- 11 Wu X G, Yang P L, Jung T, Xiong Y, Zheng X. Compressive sensing meets unreliable link: sparsest random scheduling for compressive data gathering in lossy WSNs. In: Proceedings of the 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBICOM). Maui, HI, USA: ACM Press, 2014. 13–22
- 12 Luo C, Wu F, Sun J, Chen C W. Efficient measurement generation and pervasive sparsity for compressive data gathering. *IEEE Transactions on Wireless Communications*, 2010, **9**(12): 3728–3738
- 13 Jiang Chang-Jiang, Shi Wei-Ren, Tang Xian-Lun, Wang Ping, Xiang Min. Energy-balanced unequal clustering routing protocol for wireless sensor networks. *Journal of Software*, 2012, **23**(5): 1222–1232
(蒋畅江, 石为人, 唐贤伦, 王平, 向敏. 能量均衡的无线传感器网络非均匀分簇路由协议. 软件学报, 2012, **23**(5): 1222–1232)
- 14 Fafianie S, Bodlaender H L, Nederlof J. Speeding up dynamic programming with representative sets: an experimental evaluation of algorithms for Steiner tree on tree decompositions. *Algorithmica*, 2015, **71**(3): 636–660
- 15 Ai A, Lapanowski A, Plan Y, Vershynin R. One-bit compressed sensing with non-Gaussian measurements. *Linear Algebra and Its Applications*, 2014, **441**: 222–239
- 16 Zheng H F, Yang F, Tian X H, Gan X Y, Wang X B, Xiao S L. Data gathering with compressive sensing in wireless sensor networks: a random walk based approach. *IEEE Transactions on Parallel and Distributed Systems*, 2015, **26**(1): 35–44
- 17 Ebrahimi D, Assi C. Network coding-aware compressive data gathering for energy-efficient wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 2015, **11**(4): 1–24
- 18 Malloy M L, Nowak R D. Near-optimal adaptive compressed sensing. *IEEE Transactions on Information Theory*, 2014, **60**(7): 4001–4012
- 19 Liu Dong-Pei, Liu Heng-Zhu, Zhang Bo-Tao. Study on encoding algorithms for QC-LDPC codes with dual-diagonal parity check matrix. *Journal of National University of Defense Technology*, 2014, **36**(2): 156–160
(刘冬培, 刘衡竹, 张波涛. 基于准循环双对角阵的 LDPC 码编码算法. 国防科技大学学报, 2014, **36**(2): 156–160)
- 20 Pang S C, Ma T M, Liu T. An improved ant colony optimization with optimal search library for solving the traveling salesman problem. *Journal of Computational and Theoretical Nanoscience*, 2015, **12**(7): 1440–1444
- 21 Akiba T, Maehara T, Kawarabayashi K. Network structural analysis via core-tree-decomposition. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM Press, 2014. 1476–1485
- 22 Azad A, Buluç A, Pothén A. A parallel tree grafting algorithm for maximum cardinality matching in bipartite graphs. In: Proceedings of the 2015 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Hyderabad: IEEE Press, 2015. 1075–1084
- 23 Wu X G, Xiong Y, Yang P L, Wan S H, Huang W C. Sparsest random scheduling for compressive data gathering in wireless sensor networks. *IEEE Transactions on Wireless Communications*, 2014, **13**(10): 5867–5877



李鹏 中南大学信息科学与工程学院博士研究生. 主要研究方向为无线传感器网络, 压缩感知技术. 本文通信作者.

E-mail: lpchs617@csu.edu.cn

(LI Peng Ph.D. candidate at the School of Information Science and Engineering, Central South University. His research interest covers wireless sensor

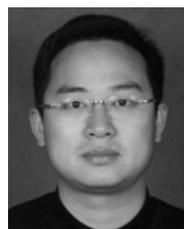
networks and compressive sensing technology. Corresponding author of this paper.)



王建新 中南大学信息科学与工程学院教授. 主要研究方向为网络优化理论, 参数计算和生物信息学.

E-mail: jxwang@mail.csu.edu.cn

(WANG Jian-Xin Professor at the School of Information Science and Engineering, Central South University. His research interest covers network optimization theory, parameter calculation, and bioinformatics.)



丁长松 湖南中医药大学管理与信息工程学院副教授. 主要研究方向为网络优化理论, 云计算.

E-mail: dinghongzhe@yeah.net

(DING Chang-Song Associate professor at the School of Management and Information Engineering, Hunan University of Chinese Medicine. His

research interest covers network optimization theory and cloud computing.)