

边缘计算中可靠数据收集方法^{*}

张德干¹, 董文淼¹, 张捷², 张婷³, 赵彭真¹, 朱浩丽¹

¹(计算机视觉与系统省部共建教育部重点实验室(天津理工大学), 天津 300384)

²(北京交通大学 电子信息工程学院, 北京 100044)

³(天津体育学院 体育经济与管理学院, 天津 301617)

通信作者: 董文淼, E-mail: 15620193392@163.com



摘要: 针对在实际环境下无线传感器网络出现随机丢包、能量消耗快的问题, 结合传感器网络的特点和压缩感知的优势设计了一个边缘计算场景下的可靠的数据收集方法, 首先对网络进行分簇, 在数据采集阶段设计基于实际链路状态的测量矩阵并构造适合该传感器数据的稀疏基, 在数据传输阶段即从簇头传输到汇聚节点, 采取最优最差蚁群算法对链路质量进行评估, 然后进行基于链路质量的多路径传输, 最后将数据重构任务卸载到边缘节点执行. 实验结果证明所提数据收集方法与其他方法对比, 在链路出现随机丢包的情况下, 数据传输的可靠性与网络的能耗都表现出较好的效果.

关键词: 压缩感知; 可靠传输; 边缘计算; 任务卸载; 多路径

中图法分类号: TP393

中文引用格式: 张德干, 董文淼, 张捷, 张婷, 赵彭真, 朱浩丽. 边缘计算中可靠数据收集方法. 软件学报. <http://www.jos.org.cn/1000-9825/6427.htm>

英文引用格式: Zhang DG, Dong WM, Zhang J, Zhang T, Zhao PZ, Zhu HL. Method of Reliable Data Collection Based on Compressed Sensing for Edge Computing. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/6427.htm>

Method of Reliable Data Collection Based on Compressed Sensing for Edge Computing

ZHANG De-Gan¹, DONG Wen-Miao¹, ZHANG Jie², ZHANG Ting³, ZHAO Peng-Zhen¹, ZHU Hao-Li¹

¹(Key Laboratory of Computer Vision and System (Ministry of Education)(Tianjin University of Technology), Tianjin 300384, China)

²(School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)

³(School of Sports Economics and Management, Tianjin University of Sport, Tianjin 301617, China)

Abstract: Aiming at the problem of random packet loss and fast energy consumption in actual wireless sensor networks, a reliable collection method is designed based on the characteristics of sensor network and the advantages of compressed sensing. Firstly, the network is clustered. In the data acquisition phase, the measurement matrix based on the actual link state is designed and sparse base suitable for the sensor data is constructed. In the data transmission phase, the data is transmitted from the cluster head to the aggregation node. The best-worst ant system is adopted to evaluate the link quality. Then, the multi-path transmission based on the link quality is carried out. Finally, the data reconstruction task is unloaded to the edge node Implementation. The experimental results show that in the scenario of random packet loss in the network, the data collection method proposed in this study, compared with other methods, the reliability of data transmission and energy consumption of the network show better results.

Key words: compressed sensing; reliable transmission; edge computing; task unloading; multi-path

人工智能, 大数据, 云计算以及 5G 网络的发展极大促进了物联网的广泛应用^[1]. 然而在万物互联的背景下, 物联网底层设备的计算和存储能力一直制约其进一步发展. 针对于此催生了边缘计算模型, 边缘计算是指在网络的

* 基金项目: 国家自然科学基金 (61571328); 天津市自然科学基金 (18JCZDJC96800); 天津市重大科技专项 (17YFZCGX00360)

收稿时间: 2021-01-12; 修改时间: 2021-03-21, 2021-06-12; 采用时间: 2021-08-04; jos 在线出版时间: 2022-09-16

边缘提供计算和存储能力的一种新型模型^[2,3]. 其基本思想是将原有的云计算模型的计算和存储能力迁移到边缘设备如基站、网关、路由器等, 从而减缓端到端的时延, 挖掘底层感知网络内在能力, 提高底层设备智能化^[4,5]. 计算卸载作为边缘计算的关键技术^[6,7], 是指底层终端设备将计算任务卸载到资源丰富的边缘服务器.

无线传感器网络 (wireless sensor network, WSN) 作为支持物联网发展的关键数据收集技术一直是研究热点^[8-11]. WSN 是由大量传感器节点组成的一种自组织网络, 集数据的采集、处理和传输于一身, 通常用于特定区域监测并通过基站将数据接入网络, 已广泛应用于环境监测、道路监控、军事侦察等场景^[11]. 但由于在实际的网络环境中存在许多不可预见的因素, 传感器节点在进行传输时经常面临数据错误和丢包的问题, 从而影响感知数据的可靠性和数据重构的准确性^[12-14]. 目前, 针对不可靠的传输的解决办法主要有数据重传和数据纠错, 但是这是以消耗能量为代价, 大大影响网络的生命周期^[15-17]. 本文结合传感器网络的特点和压缩感知的优势设计了一个边缘计算场景下的可靠的数据收集方法.

1 相关研究工作

将 CS 理论应用于边缘计算框架下的 WSN 数据收集, 一方面要考虑 WSN 的结构, 另一方面就是 CS 理论的适用性^[18]. 目前已经有大量的学者提出了很多的基于压缩感知的无线传感器网络的收集方法. 文献^[19]的作者提出了一个基于 CS 的 WSN 能耗分析模型, 对优化基于 CS 的 WSN 的设计有非常大的帮助.

现在的国内外研究学者已经提出了众多的解决方案^[20-23], 仿真结果也证明压缩感知在 WSN 中数据收集的优势, 但是绝大部分的研究均假定 WSN 的链路是相对理想的, 即链路上无误码. 文献^[24]就提出了一种基于半方差的压缩感知方法, 基于测量值之间的时空相关性, 测量气候数据的传感器数据样本, 但其并未考虑到气候和环境的变化对实际效果的影响. 而实际中, 受复杂部署环境的影响及无线收发功耗和节点硬件成本的约束, 传感网的无线链路上分组丢失、出错十分常见. 目前, 针对有损链路下的压缩感知数据收集算法的研究已经提上议程^[25-30]. 文献^[31]采用过采样的方式来保证丢失数据的恢复, 文献^[32]采取丢包预测和丢失重传的机制进行数据可靠传输的保障. 文献^[33]提出了一种基于压缩感知丢包匹配数据收集算法 (CS-MDGA), 实现了不可靠链路下的多路径可靠交付, 但是可以改进的一点是增加链路状态感知机制, 避免一定数量的丢包的发生. 因为现有工作假定 WSN 的链路相对理想, 默认在链路无损情况下进行研究, 问题模型不能很好地表示实际情况, 所以本文考虑了更多实际上可能发生的状况, 针对的是有损链路下的压缩感知数据收集方法的研究, 通过考虑链路的状况对于感知数据和测量矩阵的影响, 并采用多路径路由传输策略来共同实现可靠的数据收集.

如今随着各项技术的发展, 通过压缩感知和其他算法的结合, 我们可以做到在无线传感器网络监测的区域内进行信号目标源的准确探测, 并且应用在煤矿井等实际场景中, 精确的定位目标位置. 还可以应用在水下数据采集方面, 解决水下数据传输不准确, 耗能大等问题.

2 问题的分析和建模

2.1 问题分析

根据压缩感知的理论, 当选择合适的测量矩阵以及与测量矩阵不相关的稀疏基, 即满足 RIP 限制条件, 则可以精确恢复出每个传感器采集的原始数据, 然而在实际的应用中由于传感器网络的无线链路不够稳定, 我们应当考虑因链路不稳定导致局部数据丢失以及测量矩阵不能准确描述参与节点的信息从而导数据重构困难和能耗增加的问题.

利用压缩感知进行无线传感器网络数据收集的基本过程 (图 1) 可以描述为: 假设网络中部署一个汇聚节点和 N 个感知节点在感兴趣的区域进行数据的收集, 用 $X = [x_1 \dots x_N]^T \in \mathbb{R}^{N \times 1}$ 表示 N 个传感器节点收集到的数据, 其中 x_i 表示节点 i 的读数. 由于传感器网络的时空相关性, X 在某一稀疏基 $\Psi \in \mathbb{R}^{N \times N}$ 下为 K 稀疏信号 s , $K \ll N$, 即 $X = \Psi s, s \in \mathbb{R}^{N \times 1}$, 然后采用观测矩阵 $\Phi \in \mathbb{R}^{M \times N}$ 对 X 进行测量, 可以通过网络中所有的节点产生一个随机的种子并利用各个传感器节点的加权运算来构建, 即 S_1 节点收集完数据后乘上自己的权值向后发送给 S_4 , S_4 将自己收集

的数据乘上权值加上接收的 S_1 数据往后传输, 这样一直传输到汇聚节点, 最后得到 M 个测量值. 在汇聚节点利用 CS 恢复算法, X 可以从 M 维线性测量向量 Y 被准确地重建.

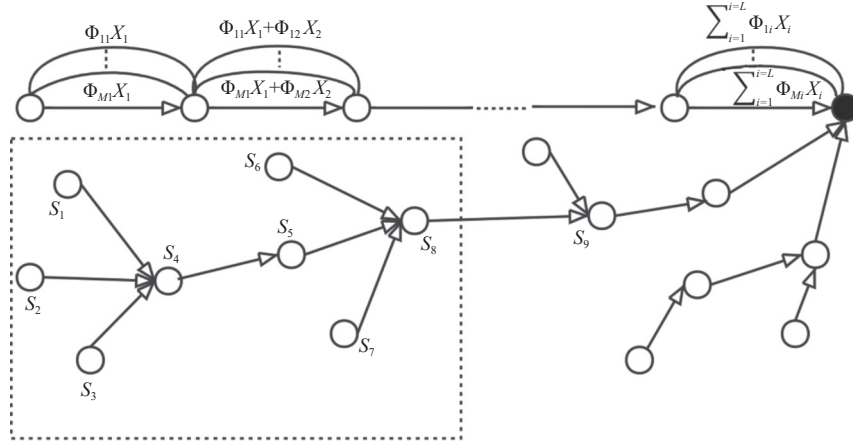


图 1 数据传输的过程

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{M1} & \phi_{M2} & \dots & \phi_{MN} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \quad (1)$$

然而这是在理想的情况下, 在实际的应用中由于传感器网络部署环境的复杂性以及能量有限和其他不确定的因素, 传感器网络的无线链路不够稳定, 大概率会出现丢包, 如果丢包发生在叶子节点处这样的影响是比较小的, 一旦丢包发生在距离汇聚节点较近的位置, 会对数据的恢复造成极大的影响. 如: 若在第 i 次采集中丢包发生在 $S_1 \rightarrow S_4$ 影响的只有 S_1 的数据, 若丢包发生在 $S_7 \rightarrow S_8$, 丢失将是 $S_1, S_2, S_3, \dots, S_7$, 由上边的式子可知, 数据的丢失则对数据的恢复产生严重影响.

$$Y_i = \begin{bmatrix} \phi_{i1} & \phi_{i2} & \dots & \phi_{iN} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ X_8 \\ \vdots \\ X_N \end{bmatrix} \quad (2)$$

2.2 模型建立与分析

为了更好地表述, 我们对环境做如下描述: 位于设备层的无线传感器节点均匀分布于一片感知区域中. 远离边缘节点的传感器节点不能在传输范围内直接与其进行交流, 所有的传感器节点可以通过多跳传输到达边缘节点, 并且具有以下特点.

- (1) 初始条件下不会出现孤立节点;
- (2) 基站是唯一的, 拥有边缘计算节点的能力;
- (3) 传感器节点的初始能量固定, 且能量不能增加;
- (4) 传感器节点具有唯一的标识, 且具有定位能力;
- (5) 传输范围内的传感器节点均可以和基站进行通信.

具体的模型建立如下: 假设网络中 N 个传感器节点, 被分为了 H 个簇, 首先根据每个簇内实际的链路情况充

分利用压缩感知的优势构建子测量矩阵, 而且每个簇头只存储自己的测量矩阵 Φ_i , 本文设计的测量矩阵将链路的实际状况考虑在内, 最后 sink 节点根据各个簇头节点传来的子观测矩阵 Φ_i , 构成整个网络的观测矩阵 Φ .

$$\Phi = \begin{bmatrix} \Phi_1 & 0 & \dots & 0 \\ 0 & \Phi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Phi_h \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_h \end{bmatrix} \quad (3)$$

则其数据的观测过程可以描述为公式 (4). 其中 Y_i 表示的是第 i 个簇内收集到的数据 (采样得到的数据), 等式左边是网络中压缩采样得到的数据.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_h \end{bmatrix} = \begin{bmatrix} \Phi_1 & 0 & \dots & 0 \\ 0 & \Phi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Phi_h \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_h \end{bmatrix} \quad (4)$$

边缘任务卸载通常需要 6 个步骤: 1) 节点发现: 寻找可用的计算节点; 2) 程序切割: 将需要处理的任务进行分割, 确保各部分功能的完整性; 3) 卸载决策: 决定程序的哪一部分卸载到边缘层; 4) 程序传输: 将决定卸载的任务以某种方式传递到边缘层; 5) 执行计算: 边缘节点接受到任务后启动一个进程分配一定的资源进行计算处理; 6) 计算结果回传: 将计算结果回传至终端设备. 相比一些其他物联网与边缘计算结合的场景, 比如车联网, 车联网节点是移动的节点需要进行边缘节点的发现, 而 WSN 是固定的节点并且边缘节点也是固定的. 并且车联网节点与边缘节点之间需要实时可靠的双向交互, 而 WSN 的交互较为简单仅为单向没有回传任务. 结合 WSN 的场景本文任务卸载主要包括: 簇内压缩采样构建计算任务, 簇间多跳传输完成任务卸载.

簇内数据收集模型: 设某个簇中有 N_i 个成员节点, 其中有 M_i 个节点采集数据, 簇内数据传输采取直接将采集数据传输至簇头节点, 不采集数据的传感器节点进入休眠状态, 并且采集的数据在稀疏基 Ψ 下是稀疏的, 簇内数据具有很强的相关性.

簇间传输模型: 当簇内收集的数据传递到簇头节点的时候, 首先利用最优最差蚁群算法对网络中的路径进行信息素的标记, 然后根据链路的质量来构建基于负载均衡的多路径传输, 然后在 sink 节点进行数据的重组与重构.

整体数据的收集过程描述如下.

(1) 初始化, 将整个感知区域内的节点进行分簇, 得到若干个簇, 并选取簇头节点. 以簇为单位进行数据的收集 (各个簇都采用类似的处理), 簇内的各个传感器节点依据提前设计好的采样率 p 进行随机采样, 然后将自身的数据发往所属的簇头, 根据感知节点传输到簇头节点的实际情况构建测量矩阵. 最后, 整个网络的感知数据都被集中到簇头上.

(2) 簇头传输每个簇的测量值, 首先利用改进的蚁群算法评估链路的质量, 再依据链路质量进行负载均衡多路径传输, 当 sink 节点接收到所有簇头节点的区域测量值, 则形成一次完整的测量, 最后在 sink 节点进行数据重构.

2.3 网络的能耗模型

本文提出的数据的收集方法主要是分两个阶段进行, 并且没有过于复杂的计算, 复杂的计算任务卸载到具有持续供能的基站, 其能量消耗可以不进行考虑, 本文将节点能耗分为簇内能耗 E_{intra} 与簇间能耗 E_{inter} , 即:

$$E_{\text{total}} = E_{\text{intra}} + E_{\text{inter}} \quad (5)$$

其中两个阶段中的数据传输的能耗分析按照已有的通信能耗模型进行计算, 即:

$$E_{\text{Tx}}(L', d) = E_{\text{elec}} \times L' + \varepsilon_{\text{amp}} \times L' \times d^2 \quad (6)$$

$$E_{\text{Rx}}(L') = L' \times E_{\text{elec}} \quad (7)$$

其中, $E_{\text{Tx}}(L', d)$ 表示数据的发送节点将 1 个 L' 比特的数据传输距离为 d 时所消耗的能量, $E_{\text{Rx}}(L')$ 表示数据的接受节点接收 L' 比特所消耗的能量, E_{elec} 表示节点发送或者接受单位比特所消耗的能量, ε_{amp} 表示节点功率放大的系数, 由公式 (7) 可知, 节点接受或者发送数据的能耗与数据包的大小有关.

3 算法的设计与分析

3.1 簇内数据的采集

无线传感器网络的数据通常具有时空相关的特性, 为了衡量相邻节点间的相关性, 不失一般性我们定义了如下核函数 $\kappa(x_i, x_j)$:

$$\kappa(x_i, x_j) = e^{-\frac{d_{ij}}{2\tau^2}} \quad (8)$$

其中, d_{ij} 表示节点 ij 之间的欧式距离, τ 表示核函数的宽度参数可以根据收集的感知数据训练得到, 核函数的宽度参数采用交叉验证方式进行训练, 即将数据集分成 k 等份, 每次将 $k-1$ 份数据作为训练数据, 剩下的数据作为测试数据, 重复 k 次, 将 k 次的平均值作为最终参数, 对于不同类型的数据只需要一次训练即可, 不需要更新. 宽度参数控制着函数的径向范围, 对实验结果有一定影响. N 个节点的相关矩阵 $\mathbf{A}$ 可以表示为:

$$\mathbf{A} = \begin{bmatrix} e^{-\frac{d_{11}}{2\tau^2}} & e^{-\frac{d_{12}}{2\tau^2}} & \dots & e^{-\frac{d_{1N}}{2\tau^2}} \\ e^{-\frac{d_{21}}{2\tau^2}} & e^{-\frac{d_{22}}{2\tau^2}} & \dots & e^{-\frac{d_{2N}}{2\tau^2}} \\ \dots & \dots & \dots & \dots \\ e^{-\frac{d_{N1}}{2\tau^2}} & e^{-\frac{d_{N2}}{2\tau^2}} & \dots & e^{-\frac{d_{NN}}{2\tau^2}} \end{bmatrix} \quad (9)$$

观察可得矩阵 $\mathbf{A}$ 为 Toeplitz 矩阵, 则可对角化 $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, 其中 $\mathbf{\Lambda}$ 为对角矩阵, $\mathbf{U}$ 为标准正交基. 我们将 $\mathbf{U}$ 用于稀疏表示基, 令 $\Psi_G = \mathbf{U}$, 则 X 可以表示为:

$$X = \Psi_G s \quad (10)$$

当 $N = 200$, $\tau^2 = 1$ 时, Ψ_G 中的每一行元素的均值和方差 的值相在一个很小的范围内波动, 当 n 的值逐渐变大均值和方差逐渐趋近于一个稳定的值 0. 当传感节点较大时该稀疏基中的每一行元素可以看作是同一随机变量产生的随机序列, 即稀疏基可以看成由 N 个随机变量 $\xi_1, \xi_2, \dots, \xi_M$ 产生的随机序列. 这些随机变量有相同的数值特征:

$$E(\xi_i) = 0, \text{Var}(\xi_i) = E(\xi_i^2) = \frac{1}{N} \quad (11)$$

为了方便描述我们用 $I_i = \{r_i^1 \dots r_i^{M_i}\}$ 表示第 i 个簇中发送感知数据的节点的序号的索引, 即 $I_i = |M_i|$. 在某一采集时刻 t , 我们设置每个簇内的成员节点以概率 p 随机向簇头节点发送自己感知的数据.

由于在实际部署的环境中无线传感器网络的链路状态不太稳定存在一定的丢包率, 我们设置网络的丢包率为 p_1 , 则可以计算出每一个簇中采集的数据量的期望值为 $M_i = N_i p(1 - p_1)$, 簇首根据收集到的数据生成测量矩阵 $\Phi_i = (\phi_1^T \dots \phi_{M_i}^T)$, 其定义为:

$$\Phi_i(k, j) = \begin{cases} 1, & j = r_i^k \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

其中, $\Phi_i \in R^{1 \times M_i}$ 为 Φ_i 中的一行, $k = 1, \dots, M_i$. 矩阵中的每一行有且仅有一个非零的值, 可以用来反映节点之间的空间关系.

我们以一个簇内有 15 个感知节点为例来进行说明, 假设在某一采集任务时刻, 有 10 个节点参与数据的采集, 即 $N_i = 15$, $M_i = 10$. 假设参与采集的节点编号为 $I_i = \{1, 2, 4, 5, 6, 8, 10, 12, 14, 15\}$, 其余几个感知节点处于休眠状态, 在休眠状态不耗费能量. 簇头节点在收到采集节点的数据之后, 根据公式 (13) 进行判断生成 Φ_i :

$$\Phi_{i10 \times 15} = I_{15} \quad (13)$$

若在簇内传输的过程中如果出现包的丢失则此时生成的测量矩阵就是一个小于 10 行的一个矩阵, 在实际的传输过程中测量矩阵可以简单不耗费太多能量自适应的构建, 在控制好采样率的情况下可以完成数据的精确恢复.

定义 1. 给定任意随机变量 ξ , 如果存在一个常数 $c > 0$, 使得对于任意 $\lambda \in R$. 有如下不等式成立, 则称随机变量 ξ 满足亚高斯随机分布, 记作 $\xi \sim \text{Sub}(c^2)$.

$$E(e^{\lambda \xi}) \leq e^{\frac{c^2 \lambda^2}{2}} \quad (14)$$

推论 1. 当 N 较大时, $\Theta = \Phi \Psi_G$ 的任意一行可以看做是由随机变量 ξ_i 产生的随机序列并且满足 $\xi_i \sim \text{Sub}(2)$.

证明: 因为 Φ 是分块对角矩阵, 并且 Φ_i 中任意行仅有的非零元素的位置是随机分布的, 任意行互不相关, 所以

Θ 可以看作是 M 个独立的随机变量 $\xi_1, \xi_2, \dots, \xi_M$ 产生的随机序列构成的.

$$E(e^{\lambda \xi_i}) = E\left(\sum_{n=0}^{\infty} \frac{\lambda^n \xi_i^n}{n!}\right) = 1 + \sum_{n=2}^{\infty} \frac{\lambda^n E(\xi_i^n)}{n!} \leq 1 + \sum_{n=2}^{\infty} \frac{\lambda^n}{n!} \leq e^{|\lambda|} - |\lambda| \quad (15)$$

$$\frac{\ln(e^{|\lambda|} - |\lambda|)}{\lambda^2} \leq 1 \Leftrightarrow \ln(e^{|\lambda|} - |\lambda|) \leq \lambda^2 \Leftrightarrow e^{|\lambda|} - |\lambda| \leq e^{\lambda^2} \quad (16)$$

$$E(e^{\lambda \xi_i}) \leq e^{\lambda^2} \Leftrightarrow E(e^{\lambda \xi_i}) \leq e^{\frac{\lambda^2}{2}} \quad (17)$$

则 $\xi_i \sim \text{Sub}(2)$, 证毕.

引理 1^[17]. 假设 $\xi = [\xi_1, \xi_2, \dots, \xi_N]^T$, 其中 ξ_i 是独立同分布的, 并且满足 $\xi_i \sim \text{Sub}(c^2)$, 记 $E(\xi_i^2) = \sigma^2$ 则有对于任意的 $\alpha \in (0, 1)$, $\beta \in \{c^2/\sigma^2, \beta_{\max}\}$ 存在一个常数 c^* 使得下式成立:

$$\begin{cases} P(\|\xi\|_2^2 \leq \alpha N \sigma^2) \leq e^{(-N(1-\alpha)^2/c^*)} \\ P(\|\xi\|_2^2 \leq \beta N \sigma^2) \leq e^{(-N(\beta-1)^2/c^*)} \end{cases} \quad (18)$$

推论 2. 给定任意 $\sigma \in (0, 1)$, 如果 $M = O(k \log(N/k))$, 则对于任意的 N 维 k 稀疏信号 s 而言, 当 Θ 满足等式 20, 则原始信号 N 能以高概率从 M 中恢复.

$$(1-\sigma) \leq \frac{\|\Theta s\|_2^2}{\|s\|_2^2} \leq (1+\sigma) \quad (19)$$

证明: 我们首先对 Θ 做归一化处理: $\Theta = \sqrt{N/M}[\theta_1, \theta_2, \dots, \theta_M]^T$.

由 $E(\theta_{ij}) = 0$, 得:

$$E\left(\left\langle \sqrt{\frac{N}{M}} \cdot \theta_i, s \right\rangle\right) = E\left(\sqrt{\frac{N}{M}} \cdot \sum_{j=1}^N \theta_{ij} \cdot s_j\right) = \sqrt{\frac{N}{M}} \cdot \sum_{j=1}^N E(\theta_{ij}) \cdot s_j = 0 \quad (20)$$

由 $\text{Var}(\theta_{ij}) = E(\theta_{ij}^2) = 1/N$ 得:

$$\text{Var}\left(\left\langle \sqrt{\frac{N}{M}} \cdot \theta_i, s \right\rangle\right) = \text{Var}\left(\sqrt{\frac{N}{M}} \cdot \sum_{j=1}^N \theta_{ij} \cdot s_j\right) = \frac{N}{M} \cdot \sum_{j=1}^N \text{Var}(\theta_{ij}) \cdot s_j^2 = \frac{\sum_{j=1}^N s_j^2}{M} \cdot \frac{\|s\|_2^2}{M} \quad (21)$$

$$E(\|\Theta s\|_2^2) = E\left(\sum_{i=1}^M \left(\left\langle \sqrt{\frac{N}{M}} \cdot \theta_i, s \right\rangle\right)^2\right) = \sum_{i=1}^M E\left(\left\langle \sqrt{\frac{N}{M}} \cdot \theta_i, s \right\rangle^2\right) = \sum_{i=1}^M \text{Var}\left(\left\langle \sqrt{\frac{N}{M}} \cdot \theta_i, s \right\rangle\right) = \sum_{i=1}^M \frac{\|s\|_2^2}{M} = \|s\|_2^2 \quad (22)$$

由引理 1 可知: $P(\|\xi\|_2^2 \leq \alpha M \sigma^2) \leq e^{(-M(1-\alpha)^2/c^*)}$, 我们假设 $\alpha = 1 - \delta$ 可得:

$$P(\|\xi\|_2^2 \leq \alpha M \sigma^2) \leq e^{(-M(1-\alpha)^2/c^*)} P\left(\frac{\|\xi\|_2^2}{M \sigma^2} \leq 1 - \delta\right) \leq e^{(-M(1-\alpha)^2/c^*)} P\left(\frac{\|\Theta s\|_2^2}{\|s\|_2^2} \leq 1 - \delta\right) \leq e^{(-M(1-\alpha)^2/c^*)} \quad (23)$$

由斯特林公式可知对于 N 维的 k 稀疏信号: $(N, k) \leq \left(\frac{eN}{k}\right)^K$ 则对于 k 稀疏信号 s 有:

$$\left(\frac{eN}{k}\right)^K \cdot 2e^{\frac{-M\delta^2}{c^*}} = 2e^{\frac{-M\delta^2}{c^*}} + k \log\left(\frac{N}{k}\right) + 1 \quad (24)$$

则当 $M \geq O(k \log(N/k))$ 时, 对于任意的 N 维的 k 稀疏信号 s , Θ 以趋近于 1 的概率满足 RIP 特性.

簇内的数据收集算法流程如算法 1 所示.

算法 1. 簇内数据收集算法.

输入: 数据的采样率 p ;

输出: 观测矩阵, 采集数据向量.

步骤:

1) M_i 个节点以概率 p 随机地 v 发送数据包至簇头节点完成感知数据的采样;

```

2) for  $i = 1 : 1 : N$ 
3) 节点  $i$  随机生成一个 0 到 1 的随机数  $p_d$ ;
4) if  $p < p_d$ 
5) 节点  $i$  发送感知数据至簇头节点;
6) else  $sleep()$ ;
7) end if;
8) end for;
9) for  $i = 1 : 1 : h$ 
10) 第  $i$  簇头生成的已接受节点的节点号索引  $I_i$ ;
11)  $l = 1$ ;
12) for  $i = 1 : 1 : M_i$ 
13)  $\Phi_i = (k, I(l, i)) = 1$ ;
14)  $l++$ ;
15) end for;
16) 准备发送  $y_i$  和  $\Phi_i$  至 sink
17) end for;

```

算法的复杂度分析: 假设一个簇中有 h 个节点, 节点在数据收集时随机的选择 M_i 个节点进行数据的收集, 如果整个网络中有 M 个节点参与数据的收集与传输, 则整个网络在簇内最多传输 M 条数据, 并且这是在理想的情况下, 即是 $O(n)$. 在数据传输到簇头节点之后是进行传输节点的统计生成测量矩阵, 也是 $O(n)$. 则簇内收集算法的时间复杂度是 $O(n)$.

3.2 簇间传输算法

簇内的数据处理完毕以后, 接下来则需要为各个簇头节点到汇聚节点找到一条稳定可靠的路径, 将各个感知节点的数据传递到汇聚节点进行恢复. 簇头数据的丢失将严重影响压缩感知的精度, 本章将采用最优最差蚁群算法对各个簇头节点到达汇聚节点的最优路径进行搜索, 根据链路的质量采取多路径并发传输, 并且在数据传输中增加纠错码和负载均衡机制, 确保簇头到汇聚节点的传输精度和准确率.

3.2.1 链路质量评估

首先是进行网络中簇头节点的等级划分, 初始化数据包并且采用最优最差蚂蚁系统 (best-worst ant system, BWAS) 进行信息素值的确定, 根据归一化后信息素的值来进行链路质量的评估, 为建立从簇头节点到汇聚节点的不交叉的多条传输路径做好准备.

定义 2. 簇的等级. 根据簇距离汇聚节点的远近, 将网络划分为不同等级的簇, 对簇头节点进行等级标定. 每个等级区域的节点具有相同的等级度, 每个簇内的节点具有相同的等级度, 即:

$$G_{ci} = m, \text{ if } cdist(ci, s) \in ((m-1) \cdot r, m \cdot r) \quad (25)$$

其中, $cdist(ci, s)$ 表示簇头节点到汇聚节点的欧式距离, G 表示簇的等级度, m 为等级值, r 为初始设定的距离值.

定义 3. 信息素. 蚁群算法的信息素是蚁群在搜寻最优路径过程中在各路径上留下的信息素浓度值, 是反映最优路径的一个重要参数, 本文对信息素的计算是综合考虑了路径寻优过程中节点的能量、路径距离和簇的等级.

算法描述如算法 2, 算法参数如表 1 所示.

算法 2. 信息素计算方法.

输入: 簇头节点的信息;

输出: 节点信息素的值.

步骤:

1) 如果选择最优最差蚂蚁系统进行信息素的确定, 则进行步骤 2), 如果选择强化学习方法进行信息素的确定, 则进行步骤 8);

2) 初始化蚂蚁 (数据包) 的参数, 即在每一个簇头节点设置一定的数据包, 数据包按照公式 (26) 中的概率去选择下一跳节点;

$$p_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t) n_{ij}^\beta(t)}{\sum_{s \in allowed_k} \tau_{ik}^\alpha(t) n_{is}^\beta(t)}, & j \in allowed_k \\ 0, & otherwise \end{cases} \quad (26)$$

$$\tau_{ij}(t + ch_n) = \rho_1 \tau_{ij}(t) + \Delta \tau_{ij}(t, t + ch_n) \quad (27)$$

3) 当蚂蚁生成路径后, 对其信息素的值按照公式 (28) 进行局部更新;

$$\tau_{rs} \leftarrow (1 - \rho) \cdot \tau_{rs} + \alpha \cdot \Delta \tau_{rs} \quad (28)$$

$$\Delta \tau_{rs} = ch_{n(L_{mn})^{-1}} \quad (29)$$

4) 循环执行第 1) 步和第 2) 步直至簇头节点中的每只蚂蚁都生成了一条路径, 并根据每只蚂蚁经历路径的长度评选出最优、最差蚂蚁;

5) 对最优蚂蚁按照公式 (28) 进行全局的更新;

$$\tau_{rs} \leftarrow (1 - \alpha) \cdot \tau_{rs} + \alpha \cdot \Delta \tau_{rs} \quad (30)$$

其中,

$$\Delta \tau_{rs} = \begin{cases} (L_{gb})^{-1}, & \text{if } (r, s) \in globalbest \\ 0, & otherwise \end{cases} \quad (31)$$

6) 对最差蚂蚁按照公式 (30) 进行全局的更新;

$$\tau(r, s) = (1 - \rho) \cdot \tau(r, s) - \varepsilon \cdot \frac{L_{worst}}{L_{best}} \quad (32)$$

7) 分别对其他簇头中的蚂蚁进行第 2) 步到第 5) 步, 当有信息包从源节点到达目的节点后停止搜寻, 对路由表中生成的链路的信息素进行更新, 直到在每个节点上的信息包到传输到目的节点. 在生成的路由表里计算信息包的路径长度并记录最优最差路径值, 并将最优最差路径上的信息素更新;

8) 初始化状态集 S , 动作集 A , 路径上信息素的总值 π ;

9) 初始化马尔可夫过程的状态, 从 $t = 0$ 时刻开始;

10) 根据当前状态 s_t 按照定义的执行动作 a_t , 根据公式 (28) 计算 $Q(s, a)$, 选择最优的执行动作 a_t , 更新状态为 s_{t+1} , 令 $t = t + 1$;

11) 重复步骤 10), 直到生成一条路径, 计算并保存累计的信息素值 π ;

12) 回到步骤 9), 直到 π 值不再变化, 得到最优的路径和信息素的值.

表 1 算法中参数定义

参数	值
节点初始能量	0.3 J
发送 1 bit 数据所需能量	100 nJ
接收 1 bit 数据所需能量	5 nJ
最大通信半径	70 m
距离阈值	50 m
休眠的功耗	0.000 5 W
数据收集所需的能量	1 000 nJ
CS 重构算法	OMP 算法

3.2.2 传输路径的构建

假设某一簇头节点准备发送 bM 字节的数据包到汇聚节点. 数据包被分解为 $(M+k)$ 个 b 字节的数据片. 汇聚节点对接收到的至少 M 的数据片进行解码重组源数据包. 设 $path_i$ 路径一次接收 z_i 个数据片, 则汇聚节点能重构的概率为 p_{succ} 为:

$$p_{succ} = p \left[\sum_{i=1}^n z_i \geq M \right] \quad (33)$$

其中, p_{succ} 成功传输的概率, 我们用 $\mathbf{q} = [q_1, q_2, \dots, q_n]$ 表示每条路径数据传输的成功率, 并且每一条路径是相互独立的, 则有:

$$p_{succ} \geq Q(x, \mathbf{q}, k) = \sum_{j=0}^k \frac{e^{-\lambda(x)} [\lambda(x)]^j}{j!} \quad (34)$$

其中, $\lambda(x) = \sum_{i=1}^n \ln(q_i^{-x_i}) = -\sum_{i=1}^n x_i^T \ln q_i$.

簇间传输模型数据传输的成功率, 即簇间传输的可靠性为:

$$\eta_{trans} = \frac{M}{M+K} \quad (35)$$

对于从簇头到汇聚节点的某一条路径 $path_i$ 来说, 从源节点到目的节点的成功率为:

$$q_i = \prod_{l=1}^j q_i^l \quad (36)$$

对于 $path_i$, 汇聚节点接收到的数据片的数量为:

$$x_i' = x_i \cdot \prod_{l=1}^j q_i^l \quad (37)$$

则对于从簇头节点到汇聚节点的一次完整的传输的数据的片数为:

$$x = \sum_{i=1}^n x_i' = \sum_{i=1}^n x_i \cdot \prod_{l=1}^j q_i^l \quad (38)$$

簇间传输模型数据实际传输的成功率, 即簇间实际传输的可靠性为:

$$\eta'_{trans} = \frac{x}{M+K} = \frac{\sum_{i=1}^n x_i \cdot \prod_{l=1}^j q_i^l}{\sum_{i=1}^n x_i} \quad (39)$$

多路径的建立原则如下.

- 1) 源节点是在通信范围以内选择其他簇头节点作为下一跳候选节点;
- 2) 同级簇头节点之间不建立传输路径, 优先选择质量最优路径作为下一跳数据的传输节点, 直至到达目的汇聚节点, 从而建立一条到目的节点的最优传输链路;
- 3) 源节点选择当前候选节点中 (不包含已经选取建立传输路径的节点) 选择质量最优路径作为下一跳节点数据的传输节点, 依次类推直至汇聚节点, 从而建立第 2 条从源到目的地的次优传输链路;
- 4) 路径质量是按照信息素的值进行衡量的, 我们已经通过最优最差蚂蚁系统在两个不同等级之间已经生成了反映链路质量的信息素的值.

负载均衡机制: 因为传感器节点的计算资源和能量受限, 如果将源节点需要传输的数据在每条路径上都进行传输的话将造成能量的过度消耗, 因此在传输的过程中需要引入多路径负载均衡机制, 根据链路的质量来进行对数据包的分配, 如图 2 所示.

定理 1. 使用 s 负载均衡机制汇聚节点接受到的数据量 x_{lb} 大于未使用负载均衡机制的数据量 x_{no} , 即 $x_{lb} \geq x_{no}$.

证明: 首先对所有链路的成功传输率 $q_i = \prod_{l=1}^j q_i^l$, $i \in (1, 2, \dots, n)$ 进行归一化有:

$$q_i' = \frac{\prod_{l=1}^j q_i^l}{\sum_{i=1}^n \prod_{l=1}^j q_i^l} \quad (40)$$

其中, $\sum_{i=1}^n q_i' = 1$. 则 $path_i$ 分配的数据片为 $q_i'(M+K)$.

$$x_{lb} = \sum_{i=1}^n q_i'(M+K) \cdot \prod_{l=1}^j q_i^l, \quad x_{no} = \sum_{i=1}^n x_i q_i \quad (41)$$

(1) 当 $q_1 = q_2 = q_i = \dots = q_n$ 时,

$$x_{no} = \sum_{i=1}^n x_i q_i = q_i (M+k) \quad (42)$$

$$x_{lb} = \sum_{i=1}^n q_i' (M+K) \cdot \prod_{l=1}^j q_i^l = \frac{\sum_{i=1}^n (M+k) q_i^2}{\sum_{i=1}^n q_i} = q_i (M+k) x_{lb} = x_{no} \quad (43)$$

(2) 当 $q_1, q_2, \dots, q_n$ 不全相等时:

① 当 $x_1 = x_2 = x_i = \dots = x_n$ 时:

$$x_{no} = \sum_{i=1}^n x_i q_i = \frac{(M+K)}{n} \sum_{i=1}^n q_i \quad (44)$$

$$x_{lb} = \sum_{i=1}^n q_i' (M+K) \cdot \prod_{l=1}^j q_i^l = \frac{\sum_{i=1}^n (M+k) q_i^2}{\sum_{i=1}^n q_i} \quad (45)$$

$$\frac{x_{lb}}{x_{no}} = \frac{n \sum_{i=1}^n q_i^2}{(\sum_{i=1}^n q_i)^2} = n \cdot \sum_{i=1}^n \left(\frac{q_i}{\sum_{i=1}^n q_i} \right)^2 = n \cdot \sum_{i=1}^n (q_i')^2 \geq 1 \quad (46)$$

令 $f(q_1', q_2', \dots, q_n', \lambda) = \sum_{i=1}^n (q_i')^2 - \lambda (\sum_{i=1}^n q_i' - 1) = 0$, 当且仅当 $q_1' = q_2' = \dots = q_n'$, 有 $\min(f(q_1', q_2', \dots, q_n', \lambda))$.

② 当 $x_1, x_2, \dots, x_n$ 不全相等时:

令 $f(q_1', q_2', \dots, q_n', x_1', x_2', \dots, x_n', \lambda_1, \lambda_2) = \sum_{i=1}^n (x_i q_i') - \lambda_1 (\sum_{i=1}^n x_i - (M+K)) - \lambda_2 (\sum_{i=1}^n q_i - 1) = 0$.

(3) 当 $x_1 = x_2 = x_i = \dots = x_n, q_1' = q_2' = \dots = q_n'$ 时,

$$\frac{M+K}{n} \leq f(q_1', q_2', \dots, q_n', x_1', x_2', \dots, x_n', \lambda_1, \lambda_2) < (M+k),$$

则有 $E(x_{no}) < E(x_{lb})$, 证毕.

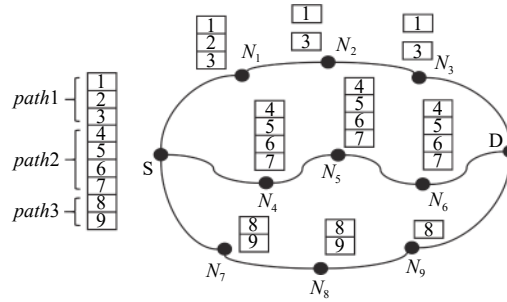


图2 负载均衡机制

因为最优最差蚁群算法产生的信息素综合了路径寻优过程中节点的能量、距离、链路质量的优势, 在进行分配的时候根据链路质量, 即信息素 τ 的值的高低来判断链路质量, τ 值越大链路的质量越高, 采用负责均衡分配的数据片就越多. 如图2所示, 源节点 s 经过编码后生成了9个子块, 并且从源节点 s 到目标节点 d 按照链路质量 $path_2 > path_1 > path_3$ 有3个路径, 分别分配一定数量的数据子包.

簇间数据传输算法描述如下.

步骤1. 初始化参数和数据包, 并计算每个簇头节点的等级;

步骤2. 执行 BWAS 算法, 输入簇头节点坐标和能量, 输出各簇头节点的信息素的值, 并在本地路由表中建立路由信息;

步骤3. 按照多路径的构造规则在传输范围内为源节点选取最优的簇头节点作为下跳中继节点建立传输路径, 并依此建立到目的节点的第一质量路径;

步骤 4. 按照多路径的构造规则在传输范围内为源节点选取不含已选取节点的最优簇头节点作为下跳中继节点建立无交叉传输路径;

步骤 5. 每建立完成一条路径, sink 节点会对建立的路径质量进行评估, 并将评估结果信息包返回给源节点, 如果链路数等于 3 或者不满足条件执行下一步骤, 否则重复步骤 4;

步骤 6. 源节点根据链路的质量对传输包进行子包的划分, 按照能者多劳的原则进行负载均衡传输;

步骤 7. 数据子包纷纷到达 sink 节点后首先进行的是包的组装;

步骤 8. 构建网络的测量矩阵, 利用重构算法进行数据的恢复.

算法的复杂度分析: 无线传感器网络共有节点数为 n , 簇头节点数为 h . 因此使用最优最差蚁群算法在簇头之间进行信息素值的标记的时间复杂度为 $O(IK(h)^2)$, 其中 I 为算法迭代的次数, 即计算的复杂度由网络的规模以及迭代的次数共同决定.

4 实验与分析

本文为了验证所提出方法的性能与可靠性, 分别进行了仿真分析与场景的测试. 采用 Matlab 2018a 作为数据处理的工具, 我们首先在理想状态下进行和高斯随机矩阵配合 DCT 转换域的对比, 通过数值的模拟进行理论分析验证, 然后在随机丢包的环境中对本文提出的收集算法与 CSEC (CS erasure encoding, 压缩编码擦除编码)^[31]、SPT_CDG (shortest path tree compressed data gathering, 基于最短路径树的压缩感知算法)^[34]、CS_LEACH (combined with compressive sensing and LEACH, 压缩感知和 LEACH 协议结合数据采集方案)^[35]、SRS (sparsest random scheduling, 最稀疏随机调度)^[25]以及 CS_RTSC (CS based on retransmission and temporal sequence correlation prediction, 基于重传和时间序列相关预测的压缩感知)^[32]进行对比. CSEC 是将压缩感知进一步用作恢复丢失数据的一种应用层擦除编码策略; SPT_CDG 针对普通压缩感知所存在的密集投影问题, 对现有稀疏投影机制进行改进, 将距离选择策略加入到数据收集过程; CS_LEACH 将 LEACH 协议和压缩感知融合在一起, 先根据 LEACH 协议对网络节点分簇, 再在 sink 节点用压缩感知重构算法进行信息重构; SRS 将最稀疏的测量矩阵用于基于压缩感知的数据收集, 而不增加压缩感知测量的数量, 降低了传输开销; CS_RTSC 在数据采集过程中出现丢包时, 进行有限次重传来恢复数据包. 如果恢复不成功, 则采用时间序列相关预测. 最后在实际的环境中对本文的方法进行验证.

4.1 测量矩阵和稀疏基的性能的评估

为了验证测量矩阵与稀疏基的性能, 我们采用绿野千传 (GreenOrbs) 系统收集到的 330 个数据作为数据集进行实验, 图 3 为使用 OMP 重构算法之后的恢复的数据和原始数据的对比, 通过这些我们可以看到, 本文设计的测量矩阵和稀疏基可以完成数据的压缩感知.

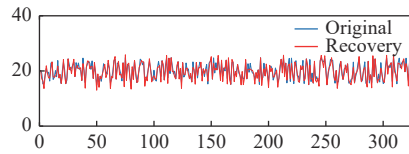


图 3 原始数据与恢复数据对比

高斯随机测量矩阵和 DCT 域是公认的满足 RIP 特性的而且较为通用的一组感知矩阵, 在这里我们使用本文提出的方案与之进行对比, 采用 OMP 数据重构算法进行数据的恢复, 分别对 GreenOrbs 系统真实收集的 5 轮数据进行实验, 设置不同的采样率对比重构的误差, 不设置丢包率 (理想状态), 采用 5 次实验的平均值作为实验结果进行对比, 其实验结果如图 4 所示.

实验结果证明, 满足 rip 特性的测试的矩阵和稀疏基能以高概率进行恢复, 并且重构的效果还是比较明显的. 本文使用的测量矩阵和稀疏基在采样率为 0.6 的时候其恢复误差与高斯 0.8 的采样率相比基本持平. 这也证明了本文的测量矩阵和稀疏基能够满足无线传感器网络的真实数据的恢复并且其重构误差相比高斯随机矩阵与 DCT

域在采样率一样的情况下更低. 这是因为高斯随机矩阵和 DCT 这种固定的正交基对于大部分数据都是能够进行稀疏变换的, 但是不是专门针对某种类型的数据特定设计的, 而本文采用的稀疏基是专门针对无线传感器网络中收集的数据进行稀疏化的, 提高了数据稀疏表示的可靠性.

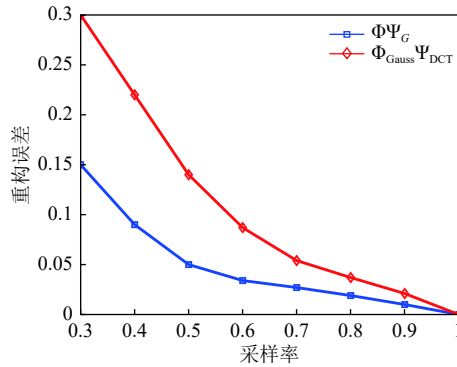


图 4 理想状态下的重构误差的对比

4.2 仿真实验测试与分析

为了进一步衡量本文提出的方案的性能, 我们进行了一系列的仿真实验, 采用 GreenOrbs 系统测得的数据进行仿真实验, (GreenOrbs 系统在一片森林区域部署了 330 多个传感器节点采集温湿度等相关的数据) 与已有的使用压缩感知的收集方法进行比较, 在相同的测试环境下进行多次实验并对比多次实验的平均结果以减少实验误差, 然后分析不同方法的网络的能耗、网络的生命周期、数据重构的精度. 实验过程中的主要参数设定如表 2 所示.

表 2 实验参数

参数	值
节点初始能量	0.3 J
发送 1 bit 数据所需能量	100 nJ
接收 1 bit 数据所需能量	5 nJ
最大通信半径	70 m
距离阈值	50 m
休眠的功耗	0.0005 W
数据收集所需的能量	1000 nJ
CS 重构算法	OMP 算法

我们首先对本文提出的数据收集的方法在不同的丢包率和不同的采样率下进行了实验, 图 5 展示了我们提出的方法在丢包率 $p_1=0, 0.05, 0.1, 0.15$ 下, 设置了不同的采样率 ($p=0.3-1$) 的数据重构误差的实验结果, 其中 $p_1=0$ 表示的是理想不丢包的状态, 通过实验结果我们可以看出, 随着采样率的增加, 无论是在何种的丢包率下都表现出了数据重构误差呈下降的趋势; 从图中我们还可以看出当丢包率为 0.05 的时候, 数据重构的误差和理想的情况下相差的不是很大, 这是因为我们使用的测量矩阵是根据实际的链路状态进行设置的, 测量矩阵比较准确, 并且在簇间是采用的基于链路状况的路由; 而且从实验结果可以看出本文提出的收集方法即使是在网络中存在大规模的丢包情况下, 只要适当的增加采样率仍然可以保持较高的数据的恢复的准确率, 并且不会随着丢包率的增加出现较大的波动.

图 6 展示了本方案与 SPT_CDG、CSEC、CS_LEACH、SRS、CS_RTSC 在采样率 $p=0.7$ 时, 不同丢包率下的 4 种方案的重构误差的实验结果对比. 图 7 是设置丢包率为 0.15 设置不同的采样率下的数据恢复误差的对比.

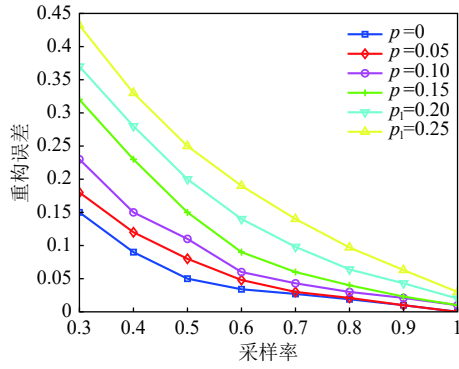


图5 不同的丢包率下本文的重构误差的比较

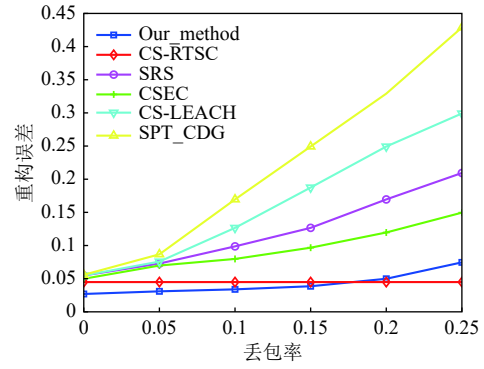


图6 不同的丢包率下各个方案重构误差的比较

通过实验结果我们可以看到, 当采样率一定时, 重构误差随着丢包率的增加而增加, 本文提出的方案的重构误差始终小于 CSEC、CS_LEACH、SRS 和 SPT_CDG, 并且随着丢包率的增加, 本文方案的优势已经表现的越来越明显, 尤其是当丢包率超过了 0.15 以后, 而 CS_RTSC 方案一直与本文提出的方案持平, 并且在丢包率逐渐增大的时候单从重构误差上要优于本文的方案; 当丢包率为 0.1 时不同的采样率下, 本文提出的方案的重构误差同样小于 CSEC 和 SPT_CDG、CS_LEACH、SRS, 与 CS_RTSC 持平. 究其原因, SPT_CDG 的每一个节点将自己收集完的数据沿着最短路径传输至汇聚节点, 当链路出现故障丢包时, 容易出现局部影响整体的情况; CS_LEACH 没有对丢包进行特殊处理; 而 CSEC 方案是通过增加采样的次数来保证数据的可靠性, 采用过采样机制来应对网络的丢包, 但是由于出现丢包导致测量矩阵的所描述的参与数据恢复的节点不够准确导致恢复存在误差; CS_RTSC 即使在丢包率较为严重的时候仍能保持较低的恢复误差是因为该方案对链路丢包进行检测预测, 采用丢包重传和恢复机制进行丢包的应对. 本文的方法是在收集过程中保证测量矩阵的准确性并且在路径传输上采取基于链路质量的传输, 保证参与重构节点的准确性.

图 8 给出了这 4 种方案在丢包率 $p_1=0.15$ 的环境下数据重构的误差与网络能耗的对比, 由图中的结果可以看出, 网络的能耗与重构的误差呈现反比的关系, 这与实际的情况是相符合的, 因为要保证的数据重构的误差尽可能低, 要么设置更高的采样率或者采取一些重传机制, 这些都将导致网络能耗的增加. 单从数据重构的方向我们可以发现 CS_RTSC 无论在何种环境下都能保证较好的重构的误差, 但是通过性能的对比我们可以发现 CS_RTSC 的高精度是以能耗的消耗为代价, 根据无线传感器网络的特点, 我们的方法更加适合丢包率不超过 0.25 的情况下, 并且对于能源的消耗是可以接受的, 相对于单纯采用过采样的 CSEC 算法我们还考虑了测量矩阵所表示的准确率以及链路传输的质量, 是能耗与准确率的折中. 在网络丢包率不超过 0.25 的网络中, 当网络重构的误差在 0.1, 我们的算法始终是最优的.

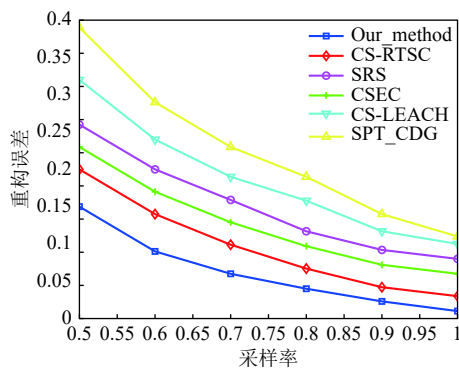


图7 不同的采样率下各个方案重构误差的比较

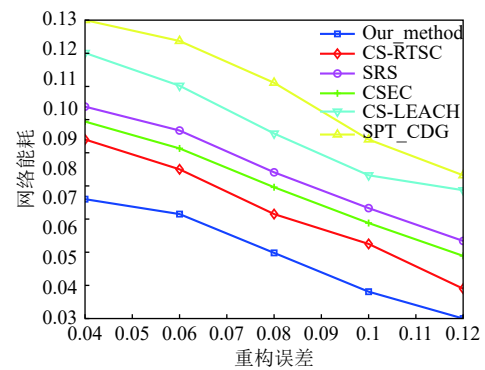


图8 不同的误差下各个方案能耗的比较

图9、图10分别是6种方案下在丢包率为0.1下重构误差在0.05时的网络节点剩余能耗的平均值和网络生命周期,从图中我们可以看到本文方案的能耗相对更低,网络的生命周期更长.

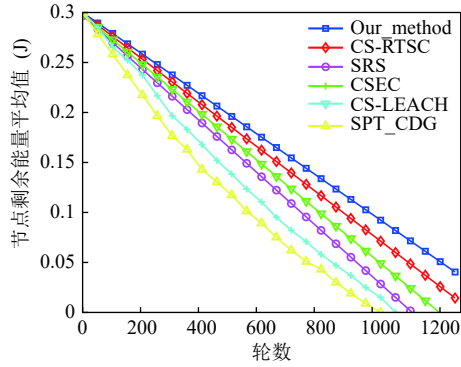


图9 不同轮数下各个方案节点剩余能量的平均值

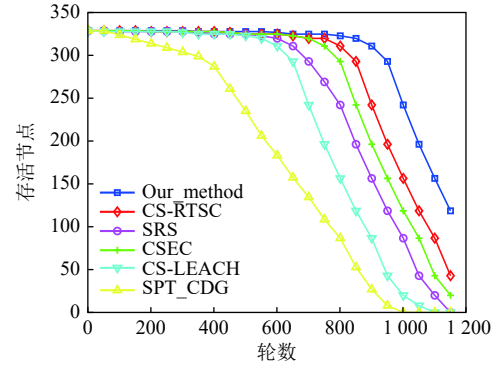


图10 不同方法网络的生命周期

图11是本文提出的方案与CS_RTSC和CSEC这3种方案在不同丢包率下规定重构误差在0.05时的网络能耗的对比,从图中我们可以直观地看到本方案的能耗始终低于另外两种方案.图12是本文提出的方案与文献[25]、文献[33]、文献[36]、OPABRL^[37]及HARL算法所消耗的时间的对比.

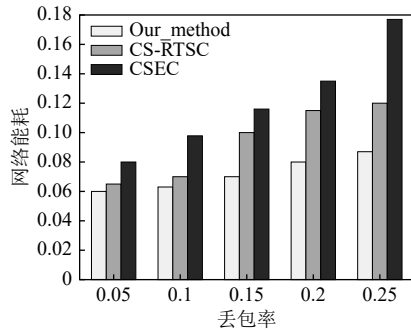


图11 不同丢包率相同误差下各个方案能耗的比较

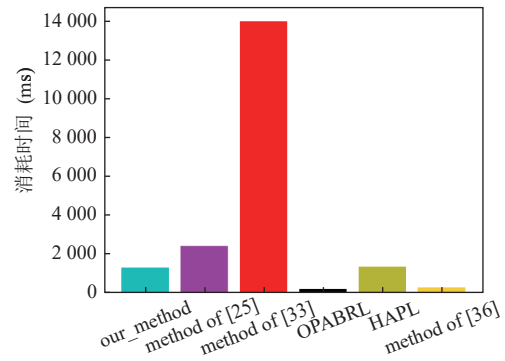


图12 不同方法的时间消耗

4.3 实际应用场景测试与分析

为了验证本文提出的算法在实际存在丢包情况下的性能,我们对实验过程中的主要参数设定如表3所示.

表3 实验参数

参数	值
节点初始能量	0.3 J
发送1bit数据所需能量	100 nJ
接收1bit数据所需能量	5 nJ
最大通信半径	70 m
距离阈值	50 m
休眠的功耗	0.0005 W
数据收集所需的能量	1000 nJ
cs重构算法	OMP算法

为了验证仿真情况与实际的状况的区别,我们做了一系列统计,图13表示的是采样率为0.7时死亡节点在第250轮,500轮,750轮和1000轮的分布情况.图14展示了采样率为0.6,0.7,0.8的真实情况的重构精度.

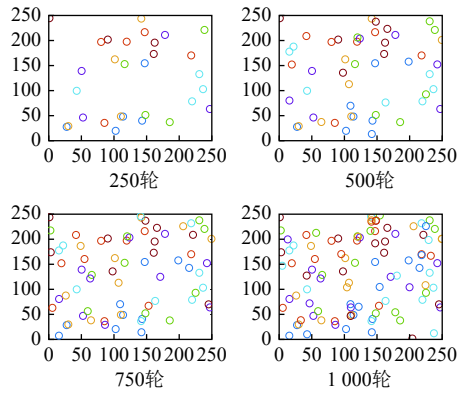


图 13 死亡节点的分布

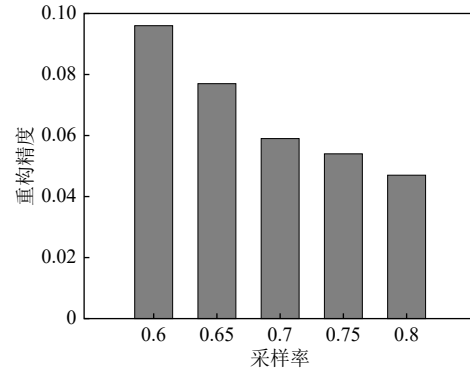


图 14 不同采样率的重构精度

5 结束语

为了平均网络的能耗我们首先采取随机采样的方法进行数据的收集, 没有采集任务的节点将处于休眠状态, 充分考虑链路的实际传输的状态来构建测量矩阵。

使用蚁群算法将到目的点的距离计算、节点间的链路状态情形以及下一跳的等级这 3 个指标来找出最优的中继节点, 并根据标记的信息素构建从簇头节点到汇聚节点的路由链路, 之后根据链路质量来进行包的多路径传输, 最后将计算任务卸载到边缘节点上进行数据的重构。仿真实验结果和实际场景的测试都表明, 本文提出的方法在平衡网络能耗和数据的重构的误差在丢包率 0.25 以内均表现出很好的效果。

References:

- [1] Boubiche S, Boubiche DE, Bilami A, Toral-Cruz H. Big data challenges and data aggregation strategies in wireless sensor networks. *IEEE Access*, 2018, 6: 20558–20571. [doi: [10.1109/ACCESS.2018.2821445](https://doi.org/10.1109/ACCESS.2018.2821445)]
- [2] Zhang DG, Zheng K, Zhang T, Wang X. A novel multicast routing method with minimum transmission for WSN of cloud computing service. *Soft Computing*, 2015, 19(7): 1817–1827. [doi: [10.1007/s00500-014-1366-x](https://doi.org/10.1007/s00500-014-1366-x)]
- [3] Zhang DG, Chen L, Zhang J, Chen J, Zhang T, Tang YM, Qiu JN. A multi-path routing protocol based on link lifetime and energy consumption prediction for mobile edge computing. *IEEE Access*, 2020, 8: 69058–69071. [doi: [10.1109/ACCESS.2020.2986078](https://doi.org/10.1109/ACCESS.2020.2986078)]
- [4] Zhang DG, Zhang T, Liu XH. Novel self-adaptive routing service algorithm for application in VANET. *Applied Intelligence*, 2019, 49(5): 1866–1879. [doi: [10.1007/s10489-018-1368-y](https://doi.org/10.1007/s10489-018-1368-y)]
- [5] Cui YY, Zhang DG, Zhang T, Chen L, Piao M, Zhu HL. Novel method of mobile edge computation offloading based on evolutionary game strategy for IoT devices. *AEU-Int'l Journal of Electronics and Communications*, 2020, 118: 153134. [doi: [10.1016/j.aeue.2020.153134](https://doi.org/10.1016/j.aeue.2020.153134)]
- [6] Zhang DG, Wu H, Zhao PZ, Liu XH, Cui YY, Chen L, Zhang T. New approach of multi-path reliable transmission for marginal wireless sensor network. *Wireless Networks*, 2020, 26(2): 1503–1517. [doi: [10.1007/s11276-019-02216-y](https://doi.org/10.1007/s11276-019-02216-y)]
- [7] Zhang DG, Zhang T, Dong Y, Liu XH, Cui YY, Zhao DX. Novel optimized link state routing protocol based on quantum genetic strategy for mobile learning. *Journal of Network and Computer Applications*, 2018, 122: 37–49. [doi: [10.1016/j.jnca.2018.07.018](https://doi.org/10.1016/j.jnca.2018.07.018)]
- [8] Zhang DG, Wang X, Song XD. New medical image fusion approach with coding based on SCD in wireless sensor network. *Journal of Electrical Engineering and Technology*, 2015, 10(6): 2384–2392. [doi: [10.5370/JEET.2015.10.6.2384](https://doi.org/10.5370/JEET.2015.10.6.2384)]
- [9] Zhang DG, Li G, Zheng K, Ming XC, Pan ZH. An energy-balanced routing method based on forward-aware factor for wireless sensor networks. *IEEE Trans. on Industrial Informatics*, 2014, 10(1): 766–773. [doi: [10.1109/TII.2013.2250910](https://doi.org/10.1109/TII.2013.2250910)]
- [10] Zhang DG, Wang X, Song XD, Zhao DX. A novel approach to mapped correlation of ID for RFID anti-collision. *IEEE Trans. on Services Computing*, 2014, 7(4): 741–748. [doi: [10.1109/TSC.2014.2370642](https://doi.org/10.1109/TSC.2014.2370642)]
- [11] Zhang DG, Ge H, Zhang T, Cui YY, Liu XH, Mao GQ. New multi-hop clustering algorithm for vehicular Ad Hoc networks. *IEEE Trans. on Intelligent Transportation Systems*, 2019, 20(4): 1517–1530. [doi: [10.1109/TITS.2018.2853165](https://doi.org/10.1109/TITS.2018.2853165)]
- [12] Zhang T, Zhang DG, Qiu JN, Zhang XD, Zhao PZ, Gong CL. A kind of novel method of power allocation with limited cross-tier

- interference for CRN. *IEEE Access* 2019, 7: 82571–82583. [doi: [10.1109/ACCESS.2019.2921310](https://doi.org/10.1109/ACCESS.2019.2921310)]
- [13] Yang JN, Ding M, Mao GQ, Lin ZH, Zhang DG, Luan TH. Optimal base station antenna downtilt in downlink cellular networks. *IEEE Trans. on Wireless Communications*, 2019, 18(3): 1779–1791. [doi: [10.1109/TWC.2019.2897296](https://doi.org/10.1109/TWC.2019.2897296)]
- [14] Zhang DG, Zhang T, Zhang J, Dong Y, Zhang XD. A kind of effective data aggregating method based on compressive sensing for wireless sensor network. *EURASIP Journal on Wireless Communications and Networking*, 2018, 2018: 159. [doi: [10.1186/s13638-018-1176-4](https://doi.org/10.1186/s13638-018-1176-4)]
- [15] Baron D, Duarte M, Sarvotham S, Wakin MB. An information-theoretic approach to distributed compressed sensing. In: *Proc. of the 43rd Allerton Conf. on Communication, Control, and Computing*. Monticello, 2005. 101–106.
- [16] Liu S, Zhang DG, Liu XH, Zhang T, Gao JX, Gong CL, Cui YY. Dynamic analysis for the average shortest path length of mobile Ad Hoc networks under random failure scenarios. *IEEE Access*, 2019, 7: 21343–21358. [doi: [10.1109/ACCESS.2019.2896699](https://doi.org/10.1109/ACCESS.2019.2896699)]
- [17] Jiao LC, Yang SY, Liu F, Hou B. Development and prospect of compressive sensing. *Acta Electronica Sinica*, 2011, 39(7): 1651–1662 (in Chinese with English abstract).
- [18] Zhang DG, Gao JX, Liu XH, Zhang T, Zhao DX. Novel approach of distributed & adaptive trust metrics for MANET. *Wireless Networks*, 2019, 25(6): 3587–3603. [doi: [10.1007/s11276-019-01955-2](https://doi.org/10.1007/s11276-019-01955-2)]
- [19] Sun ZY, Li CF, Yan B. CS-MDGA: A packet loss matching data gathering algorithm in sensor networks based on compressive sensing. *Acta Electronica Sinica*, 2020, 48(4): 723–733 (in Chinese with English abstract). [doi: [10.3969/j.issn.0372-2112.2020.04.014](https://doi.org/10.3969/j.issn.0372-2112.2020.04.014)]
- [20] Zhang DG, Gong CL, Jiang KW, Zhang XD, Zhang T. A kind of new method of Intelligent Trust Engineering Metrics (ITEM) for application of mobile Ad Hoc network. *Engineering Computations*, 2019, 37(5): 1617–1643. [doi: [10.1108/EC-12-2018-0579](https://doi.org/10.1108/EC-12-2018-0579)]
- [21] Liu XH, Zhang DG, Yan HR, Cui YY, Chen L. A new algorithm of the best path selection based on machine learning. *IEEE Access*, 2019, 7: 126913–126928. [doi: [10.1109/ACCESS.2019.2939423](https://doi.org/10.1109/ACCESS.2019.2939423)]
- [22] Zhang DG, Zhou S, Tang YM. A low duty cycle efficient MAC protocol based on self-adaption and predictive strategy. *Mobile Networks and Applications*, 2018, 23(4): 828–839. [doi: [10.1007/s11036-017-0878-x](https://doi.org/10.1007/s11036-017-0878-x)]
- [23] Zhang DG, Niu HL, Liu S. Novel PEECR-based clustering routing approach. *Soft Computing*, 2017, 21(24): 7313–7323. [doi: [10.1007/s00500-016-2270-3](https://doi.org/10.1007/s00500-016-2270-3)]
- [24] Sekar K, Suganya Devi K, Srinivasan P. Energy efficient data gathering using spatio-temporal compressive sensing for WSNs. *Wireless Personal Communications*, 2021, 117(2): 1279–1295. [doi: [10.1007/s11277-020-07922-x](https://doi.org/10.1007/s11277-020-07922-x)]
- [25] Wu XG, Xiong Y, Yang PL, Wan SH, Huang WC. Sparsest random scheduling for compressive data gathering in wireless sensor networks. *IEEE Trans. on Wireless Communications*, 2014, 13(10): 5867–5877. [doi: [10.1109/TWC.2014.2332344](https://doi.org/10.1109/TWC.2014.2332344)]
- [26] Li HB, Xiong QY, Shi WR, Zeng C, Fan M. Gradient based multipath reliable transmission strategy with fault tolerance for wireless sensor networks. *International Journal of Distributed Sensor Networks*, 2014, 10(4): 267421. (in Chinese with English abstract). [doi: [10.1155/2014/267421](https://doi.org/10.1155/2014/267421)]
- [27] Qiao JH, Zhang XY. Compressed sensing based data gathering in wireless sensor networks: A survey. *Journal of Computer Applications*, 2017, 37(11): 3261–3269
- [28] Li P, Wang JX, Ding CS. Energy-efficient data gathering scheme based on compressive sensing in wireless sensor networks. *Acta Automatica Sinica*, 2016, 42(11): 1648–1656 [doi: [10.16383/j.aas.2016.c160258](https://doi.org/10.16383/j.aas.2016.c160258)]
- [29] Wang Q, Lin DY, Yang PF, Zhang ZQ. An energy-efficient compressive sensing-based clustering routing protocol for WSNs. *IEEE Sensors Journal*, 2019, 19(10): 3950–3960. [doi: [10.1109/JSEN.2019.2893912](https://doi.org/10.1109/JSEN.2019.2893912)]
- [30] Benaddy M, El Habil B, El Ouali M, El Meslouhi O, Krit S. A mutlipath routing algorithm for wireless sensor networks under distance and energy consumption constraints for reliable data transmission. In: *Proc. of the 2017 Int'l Conf. on Engineering & MIS (ICEMIS)*. Monastir: IEEE, 2017. 1–4. [doi: [10.1109/ICEMIS.2017.8273076](https://doi.org/10.1109/ICEMIS.2017.8273076)]
- [31] Charbiwala Z, Chakraborty S, Zahedi S, He T, Bisdikian C, Kim Y, Srivastava MB. Compressive oversampling for robust data transmission in sensor networks. In: *Proc. of the 2010 IEEE INFOCOM*. San Diego: IEEE, 2010. 1–9. [doi: [10.1109/INFCOM.2010.5461926](https://doi.org/10.1109/INFCOM.2010.5461926)]
- [32] Han Z, Zhang X, Zhang DL, Zhang G, Ding SY. A data gathering algorithm based on compressive sensing in lossy wireless sensor networks. In: *Proc. of the 2nd Int'l Conf. on Frontiers of Sensors Technologies (ICFST)*. Shenzhen: IEEE, 2017. 146–153. [doi: [10.1109/ICFST.2017.8210492](https://doi.org/10.1109/ICFST.2017.8210492)]
- [33] Ghaderi MR, Tabataba Vakili V, Sheikhan M. Compressive sensing-based energy consumption model for data gathering techniques in wireless sensor networks. *Telecommunication Systems*, 2021, 77(1): 83–108. [doi: [10.1007/s11235-020-00748-9](https://doi.org/10.1007/s11235-020-00748-9)]
- [34] Ge GW. Data collection based on compressed sensing for wireless sensor networks [MS. Thesis]. Nanjing: Nanjing University of Posts and Telecommunications, 2017 (in Chinese with English abstract).

- [35] Liu Y, Liu GL, Kang WJ. Information collection scheme in underwater sensor networks with combined compressed sensing and LEACH. Chinese Journal of Sensors and Actuators, 2013, 26(3): 388–395 (in Chinese with English abstract). [doi: [10.3969/j.issn.1004-1699.2013.03.019](https://doi.org/10.3969/j.issn.1004-1699.2013.03.019)]
- [36] Ma Z, Zhang DG, Liu S, Song JJ, Hou YX. A novel compressive sensing method based on SVD sparse random measurement matrix in wireless sensor network. Engineering Computations, 2016, 33(8): 2448–2462. [doi: [10.1108/EC-09-2015-0269](https://doi.org/10.1108/EC-09-2015-0269)]
- [37] Zhang T, Zhang DG, Yan HR, Qiu JN, Gao JX. A new method of data missing estimation with FNN-based tensor heterogeneous ensemble learning for internet of vehicle. Neurocomputing, 2021, 420: 98–110. [doi: [10.1016/j.neucom.2020.09.042](https://doi.org/10.1016/j.neucom.2020.09.042)]

附中文参考文献:

- [17] 焦李成, 杨淑媛, 刘芳, 侯彪. 压缩感知回顾与展望. 电子学报, 2011, 39(7): 1651–1662.
- [19] 孙泽宇, 李传锋, 阎奔. 传感网中基于压缩感知的丢包匹配数据收集算法. 电子学报, 2020, 48(4): 723–733. [doi: [10.3969/j.issn.0372-2112.2020.04.014](https://doi.org/10.3969/j.issn.0372-2112.2020.04.014)]
- [27] 乔建华, 张雪英. 基于压缩感知的无线传感器网络数据收集研究综述. 计算机应用, 2017, 37(11): 3261–3269.
- [28] 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](https://doi.org/10.16383/j.aas.2016.c160258)]
- [34] 葛光伟. 基于压缩感知的无线传感网数据收集技术研究[硕士学位论文]. 南京: 南京邮电大学, 2017.
- [35] 刘亚, 刘功亮, 康文静. 压缩感知和LEACH结合的水下传感器网络信息采集方案. 传感技术学报, 2013, 26(3): 388–395. [doi: [10.3969/j.issn.1004-1699.2013.03.019](https://doi.org/10.3969/j.issn.1004-1699.2013.03.019)]



张德干(1969—), 男, 博士, 教授, 博士生导师, 主要研究领域为边缘计算, 物联网, 无线传感器网络, 移动算, 信息安全, 云计算.



张婷(1972—), 女, 博士, 教授, 主要研究领域为物联网, 车联网, 边缘计算.



董文森(1998—), 男, 硕士生, 主要研究领域为物联网, 无线传感器网络, 移动计算, 边缘计算.



赵彭真(1995—), 男, 硕士生, 主要研究领域为物联网, 无线传感器网络, 移动计算, 边缘计算等.



张捷(2000—), 男, 学士, 主要研究领域为车联网, 移动计算, 云计算.



朱浩丽(1996—), 女, 硕士生, 主要研究领域为物联网, 移动计算, 边缘计算等.