

面向有损链路的传感网压缩感知数据收集算法*

韩哲¹, 张霞¹, 李鸥¹, 张策¹, 张大龙²

¹(解放军信息工程大学 信息与系统工程学院, 河南 郑州 450000)

²(郑州大学 信息工程学院, 河南 郑州 450000)

通讯作者: 张霞, E-mail: zhangxiaatzz@sina.com



摘要: 基于压缩感知的数据收集算法在能量受限、数据冗余的无线传感网中有巨大的应用潜力, 现有研究大多假定无线链路理想。通过实验说明, 有损链路丢包会严重影响压缩感知数据收集算法的数据重构质量。提出了一种基于重传与时间序列相关性预测(CS data gathering based on retransmission and time series correlation prediction, 简称CS-RTSC)的数据收集算法, 将有损链路上的丢包建模为随机丢包和块状丢包, 设计了基于滑动窗统计的丢包类型预判算法, 在检测到链路丢包时判断丢包类型, 对随机丢包采用重传恢复, 对块状丢包设计了基于时间序列相关性预测算法恢复。仿真结果表明: 该算法能够有效降低有损链路丢包对CS数据收集的影响; 在网络丢包率达到30%时, CS数据重构的相对误差仅比理想链路下的CS相对重构误差高0.1%。

关键词: 无线传感网; 压缩感知; 有损链路; 丢包类型预判; 时间序列相关性

中图法分类号: TP393

中文引用格式: 韩哲, 张霞, 李鸥, 张策, 张大龙. 面向有损链路的传感网压缩感知数据收集算法. 软件学报, 2017, 28(12): 3257-3273. <http://www.jos.org.cn/1000-9825/5246.htm>

英文引用格式: Han Z, Zhang X, Li O, Zhang C, Zhang DL. Data gathering algorithm based on compressive sensing under lossy WSN. Ruan Jian Xue Bao/Journal of Software, 2017, 28(12): 3257-3273 (in Chinese). <http://www.jos.org.cn/1000-9825/5246.htm>

Data Gathering Algorithm Based on Compressive Sensing Under Lossy WSN

HAN Zhe¹, ZHANG Xia¹, LI Ou¹, ZHANG Ce¹, ZHANG Da-Long²

¹(School of Information and System Engineering, The PLA Information Engineering University, Zhengzhou 450000, China)

²(School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China)

Abstract: Data gathering algorithm based on compressive sensing (CS) has enormous application potential in wireless sensor network (WSN) in which there is limited energy and a lot of redundant data. However, most existing studies assume that network is based on ideal link. This paper illustrates a situation by experiment that existing CS reconstruction quality will be seriously affected by lossy link, and proposes a CS data gathering algorithm based on retransmission and time series correlation prediction (CS-RTSC). The type of packet loss is modeled as element random loss (ERL) and block random loss (BRL). The loss type prediction algorithm based on sliding window statistics is designed to determine the type of packet loss when link packet loss occurs. Retransmission recovery is applied for ERL, and time series correlation prediction algorithm is designed to recover the loss for BRL. The simulation result indicates that the proposed algorithm can effectively reduce the impact of lossy link in CS data gathering. When the packet loss ratio is up to 30%, the relative error of CS reconstruction signal is only 0.1% higher than that of the CS reconstruction signal in the ideal link.

Key words: wireless sensor network; compressive sensing; lossy link; prediction of packet loss type; time series correlation

* 基金项目: 国家科技重大专项(2014zx03006003)

Foundation item: National Science and Technology Major Projects of China (2014zx03006003)

收稿时间: 2016-08-25; 修改时间: 2016-10-21; 采用时间: 2016-11-17; jos 在线出版时间: 2017-03-24

CNKI 网络优先出版: 2017-03-24 17:09:26, <http://kns.cnki.net/kcms/detail/11.2560.TP.20170324.1709.008.html>

无线传感网(wireless sensor network,简称 WSN)通过大量的小型、低成本、低能耗传感节点协作地采集和处理数据,提供全天候、细颗粒度、多元监测信息,在战场侦察、生物栖息地环境监测、精细农业等领域具有巨大应用潜力.由于传感节点通常密集布署,网络所收集的原始数据具有较强的时间、空间相关性,导致传输存在大量冗余的原始数据浪费了传感节点的能量资源.

压缩感知(compressive sensing,简称 CS)技术由于其优异的压缩性能和编码端简单、解码端复杂的特点,被应用到 WSN 的数据收集中^[1].基于 CS 的无线传感网数据收集算法不仅可以压缩网内数据,且能均衡网络能耗,延长网络寿命.文献[2]最早实现在树形拓扑下将 CS 与 WSN 数据收集相结合,提出了压缩感知数据收集算法(compressive data gathering,简称 CDG).为进一步降低网内传输数据量,研究者们通过改变网络路由与观测矩阵^[3-6]、利用网内数据的时空相关性^[7-9]等方式提高数据的压缩率,降低能耗.然而,上述研究均假定 WSN 的链路是理想的,即链路上无误码,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,而目前,针对有损链路下的压缩感知数据收集算法的研究还很少.文献[10]指出:当全网节点数据存在少量的高斯独立随机丢失时,CS 重构算法可利用数据间的相关性将丢失信息重构,而实际中,造成 WSN 链路丢包的原因复杂,并不能简单地用高斯独立随机丢包模型描述.文献[11]分析了 Inter Indoor、GreenOrbs 和 Ocean Sense 系统采集的传感数据,指出了传感数据间存在较强的时间相关性,并提出了 CS 与时空相关性相结合的联合优化算法恢复丢失节点数据.但该算法仅适用于 Sink 端数据库的丢失数据恢复,而未考虑数据采集过程中的丢包恢复问题.文献[12]提出了 SRS-DG 算法,每轮随机选择部分节点参与数据收集,在 Sink 端,根据接收到数据的节点号构造极稀疏观测矩阵,重构全网原始数据.该方法虽针对解决不可靠链路下的 CS 数据收集问题,但由于采用极稀疏假设,对于全网节点数据空间相关性弱的环境不适用;另外,该算法本质上仍是采用传统的数据收集方式收集节点数据,将产生 Sink 附近节点能耗提前耗尽的能耗不均衡问题.

基于上述研究背景,本文首先通过实验说明了树状路由下 CS 收集的数据重构精度对丢包高度敏感;然后在分析链路丢包成因的基础上,将其分成随机丢包和块状丢包两种,提出了 CS-RTSC 算法;在节点发现链路丢包时,首先利用基于滑动窗的丢包预判方法判断丢包类型,对随机丢包直接采用重传机制恢复,对块状丢包则设计了时间序列相关性预测的方式补包,以较低的通信代价在误差限度内恢复丢包.仿真结果表明:该算法相比现有典型算法能够有效降低有损链路丢包对 CS 数据收集的影响,在网络丢包率达到 30%时,CS 数据重构的相对误差仅比理想链路上的相对重构误差高 0.1%.

1 CS 数据收集算法在有损链路下面临的问题

1.1 问题描述

假设传感网随机部署 N 个节点,其采集数据记为 $d=(d_1, d_2, \dots, d_N)^T$, d 在稀疏基 $\Psi_{N \times N}$ 下稀疏,观测矩阵为 $\Phi=(\phi_{ij})_{M \times N}$,其中, $M \ll N$,观测向量为 $Y=(y_i)_{M \times 1} = \Phi \Psi^T \cdot d$,则 Sink 节点可以通过求解公式(1)、公式(2)所示的优化问题,在一定精度约束下重构原数据^[1]:

$$Y = \Phi S = \Phi \Psi^T \cdot d = \Theta \cdot d \quad (1)$$

$$\hat{d} = \arg \min \|d\|_1 \quad \text{s.t. } Y = \Phi \cdot S \quad (2)$$

在 WSN 中,基于 CS 的每轮数据收集分为 M 次独立观测分别进行,可描述为公式(3):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{M1} & \phi_{M2} & \cdots & \phi_{MN} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \quad (3)$$

图 1 为在树状路由下的 CS 数据收集过程,假设有损链路下 CS 数据收集的第 i 次观测中($1 \leq i \leq M$),节点 S_8

发送至 S_9 的数据 $\sum_{j=1}^8 \phi_{i,j} d_j$ 丢失,则 S_9 的子节点 S_1, S_2, \dots, S_8 的采样数据均丢失,第 i 次观测的观测值 y_i 变为

$$y_i = [\phi_{i,1} \dots \phi_{i,8} \phi_{i,9} \dots \phi_{i,N}] \begin{bmatrix} 0 \\ \vdots \\ 0 \\ d_9 \\ \vdots \\ d_N \end{bmatrix} \quad (4)$$

可见,CS 数据收集过程中,一次丢包将造成多个节点的采集数据丢失,这种丢包的关联效应是由 CS 压缩采样过程中叠加传输多跳链路的各节点采集数据而造成的,发生丢包的节点 S_k 越靠近 Sink,丢包的关联效应影响越大;特别的,若 Sink 的一跳邻居节点数据丢失,则关联效应会致使全网节点的采集数据均丢失.另外,由公式(3)可知,每轮 CS 数据收集分为 M 次观测分别进行,而每次观测中网内链路均为有损链路,若以 y'_i 表示在有损链路上第 i 次观测得到的存在误差的观测值,其中 $i=1,2,\dots,M$,则每轮得到的观测向量为 $\mathbf{Y}'_{M \times 1} = (y'_1, y'_2, \dots, y'_M)^T$,其中, M 个观测值都存在不同程度的误差.因此, Sink 根据 $\mathbf{Y}'_{M \times 1}$ 经最优化重构后得到的数据精度将极大地降低.

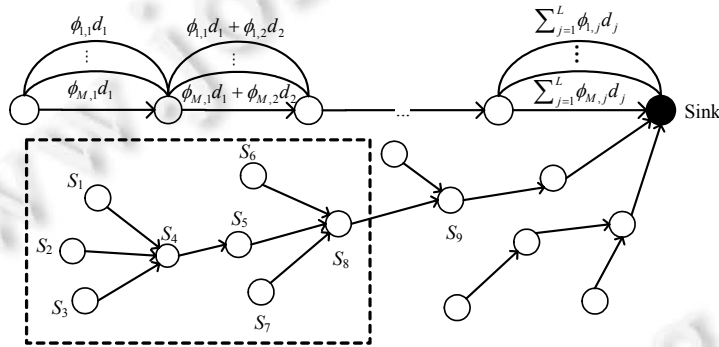


Fig.1 Data gathering based on CS under tree routing
图 1 基于树状路由的 CS 数据收集过程

下面通过实验,定量评估有损链路对 CS 数据收集性能的影响.实验中,采用经典的 CDG 压缩感知数据收集算法,性能度量指标为相对重构误差 η ,其定义见公式(5).

$$\eta = \frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\sqrt{\sum_{n=0}^{N-1} (\hat{x}_n - x_n)^2}}{\sqrt{\sum_{n=0}^{N-1} x_n^2}} \quad (5)$$

其中, \mathbf{x} 表示原始信号向量, $\hat{\mathbf{x}}$ 表示重构信号向量.算法的相对重构误差值越小,表明其重构精度越高.

图 2(a)对比了在理想无损链路和平均丢包率为 5%的有损链路的 CS 数据收集算法的性能.随观测次数 M 的增加,理想链路下的 η 急剧下降,当 $M > 15$ 之后, η 趋于稳定且低于 0.02;而对于平均丢包率为 5%的链路,虽然随 M 增加相对误差有一定降低,但其值仍大于 0.8.由此可知,CS 数据收集算法在实际有损链路下的数据重构质量远低于理想链路下评估结果.在有损链路下,由于丢包的关联效应,简单地通过增加观测次数 M 并不能有效提高重构精度.图 2(b)是固定观测次数 M 时,相对误差 η 随链路丢包率的变化.考虑到在理想链路下 M 超过 20 时 η 已低于 0.01,本次实验中取 M 为 24.由图可知:随链路丢包率的增加,CS 数据的相对重构误差 η 迅速增大,当链路丢包率为 2%时, $\eta > 0.3$.这说明即使链路仅存在轻度的丢包,也会严重降低 CS 收集算法的性能.换言之,基于 CS 的数据收集算法对丢包高度敏感,因此亟待研究 CS 数据收集算法在有损无线链路下的改进方法.

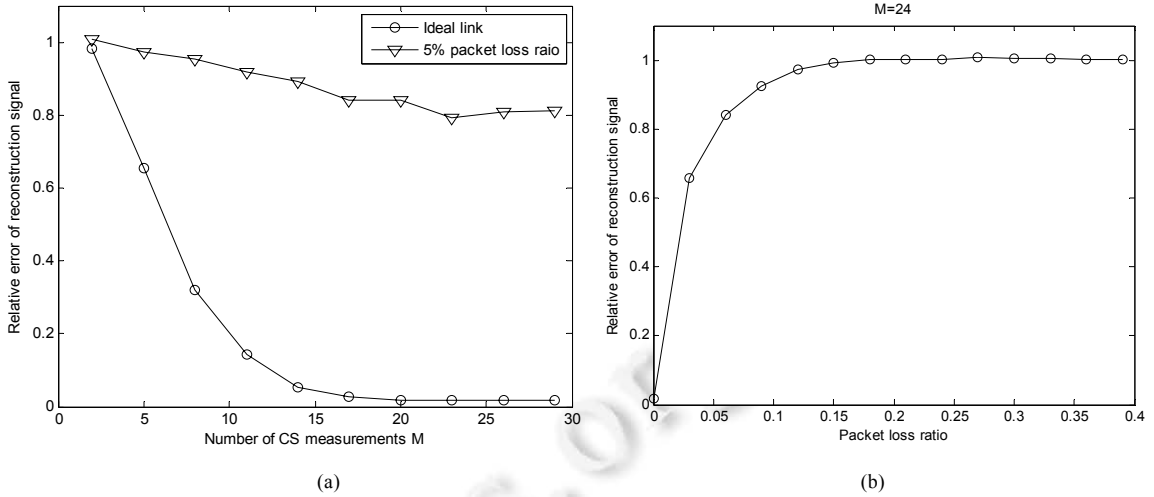


Fig.2 Impact of packet loss on the performance of data gathering based on CS in lossy link

图 2 有损链路丢包对 CS 数据收集的性能影响

1.2 有损链路丢包建模与分析

解决链路丢包的直接思路是在发生丢包时重传.对于对丢包高度敏感的 CS 数据收集算法,采用有限次重传能否改善提高数据重构精度呢?为了回答该问题,本文对有损链路丢包进行建模,并通过实验分析重传对不同丢包模型下的 CS 数据收集算法性能的影响.借鉴文献[13-15]并结合 WSN 的实际情况,将链路丢包分为随机丢包 (element random loss,简称 ERL)、块状丢包(block random loss,简称 BRL)和混合丢包(combination loss,简称 CL)这 3 类,如图 3 所示.

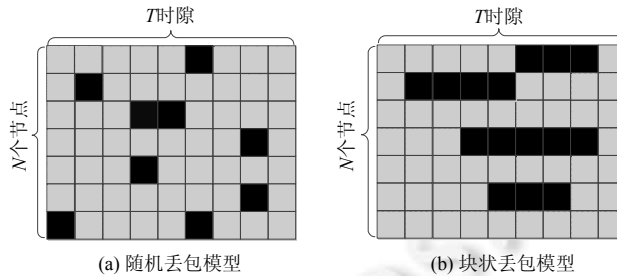


Fig.3 Diagram of packet loss model

图 3 丢包模型示意图

• ERL 模型

随机丢包通常由外界突发噪声、发包冲突等造成,其发生随机独立,若无线信道的误比特率为 P_b ,数据包长为 B 字节,则节点的丢包概率为

$$P_{ERL} = 1 - (1 - P_b)^{8B} \tag{6}$$

• BRL 模型

块状丢包通常由网络拥塞、链路相关突发等造成,丢包之间具有相关性.即:若链路发生丢包后,将出现相关连续丢包现象,丢包概率随时间衰减.块状丢包的丢包概率见公式(7)、公式(8).

$$P_b(i) = \exp\left(-\frac{i^2}{2c^2}\right) \tag{7}$$

$$P_{BRL} = \begin{cases} P_b(i), & P_b(i) > P_{ERL} \\ P_{ERL}, & P_b(i) < P_{ERL} \end{cases} \tag{8}$$

$P_B(i)$ 表示第 i 个数据包丢失的概率; $c \in [0, 0.2]$ 为衰减常数,其值越大,对应块状丢包的丢包概率衰减速率越快; P_{BRL} 为块状丢包的丢包概率.

• CL 模型

混合丢包是同时存在随机丢包和块状丢包的混合模型,因为无线链路传输环境的复杂性,该模型最接近实际情况;但在进行丢包分析时,通常将其拆分为随机丢包和块状丢包分别处理.

图 4 给出了不同丢包模型下,重传机制对 CS 数据收集算法性能的影响.图 4(a)是在随机丢包下的实验结果,当链路丢包率相同时,随最大重传次数的增加,CS 数据的相对重构误差降低,在链路丢包率为 20%时,最大重传次数为 1 的 CS 数据恢复相对误差为 0.74,而最大重传次数为 5 的相对误差下降为 0.017.图 4(b)是在块状丢包下的实验结果,当链路丢包率相同时,增加最大重传次数虽然也可降低数据的恢复相对误差,但效果不理想,当链路丢包率为 20%时,最大重传次数为 5 的数据恢复相对误差为 0.78,而最大重传次数为 20 的数据恢复相对误差为 0.51,CS 的重构精度仍然很低.通过该组实验说明:简单的重传能够以较小的代价克服随机丢包对 CS 数据收集算法带来的性能影响;但对于块状丢包以较大的能耗代价仍不能有效克服影响,仍需寻找其他方法.

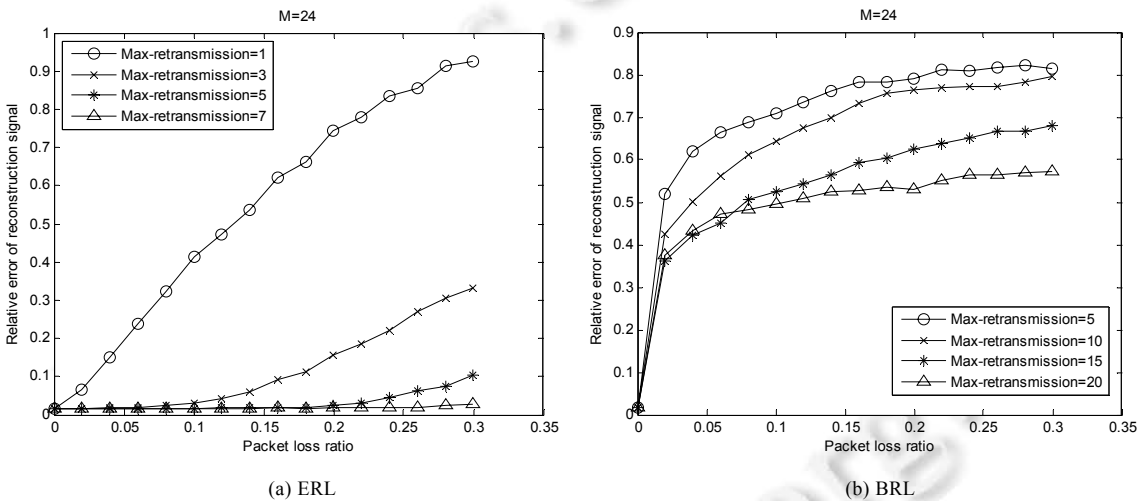


Fig.4 Impact of retransmissions on the reconstruction accuracy of CS data gathering algorithm in the lossy link

图 4 有损链路下 CS 数据收集算法重构精度受重传次数的影响

2 基于重传与时间序列相关性预测的 CS 数据收集算法

通过第 1 节的分析可知,重传能以较小的代价克服随机丢包的影响^[16],但不能有效解决块状丢包的影响.考虑到传感网的采集数据具有较强的时间相关性^[17],本文利用具有时间相关性的历史数据补包的方法来改善块状丢包时的压缩感知数据收集算法的性能.在网络运行过程中,发生丢包的节点如何判定链路当前丢包的类型,仍需研究相应的预判机制.本节首先介绍丢包类型预判方法,然后设计基于时间序列相关的预测算法,最后给出有损链路下的压缩感知数据收集算法——CS-RTSC.

2.1 基于滑动窗的丢包类型预判

用变量 X_i 表示接收节点 S 的第 i 个数据包接收情况,令 $X_i=1$ 表示第 i 个数据包被成功接收, $X_i=0$ 表示第 i 个数据包未被成功接收.由于链路质量受噪声干扰水平、网络拥塞程度等因素影响,链路丢包类型是时变的,若基于整个序列 $\{X_i\}$ 的统计情况判断丢包类型,其结果并不能反映当前链路的最新状态,因此,采用滑动窗判断节点当前的丢包状态,窗口由最新 L 次数据包接收结果组成,如图 5 所示.

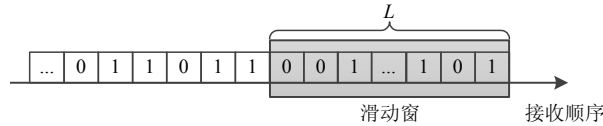


Fig.5 Statistical sequence of data reception

图 5 数据接收情况统计序列

由于随机丢包和块状丢包具有不同的丢包特点,不同链路状态下,滑动窗口内的数据包接收结果存在不同的分布.为方便描述窗口内的数据包接收分布情况,引入随机变量 z ,见定义 1.

定义 1. 定义随机变量 z 为滑动窗内的丢包个数与窗口长度的比值,即:

$$z = \frac{\sum_{i=1}^L (1 - X_i)}{L} \tag{9}$$

其中, L 是滑动窗长度.一方面 L 的大小会影响判断的时效性,另一方面将影响随机变量 z 的分布.本文选择的滑动窗长度 $L=20$,在该长度下,不仅满足时效性的要求,且随机变量 z 的分布也具有较好的规律性.利用仿真工具分别生成 ERL 丢包和 BRL 丢包序列各 10 000 个,记为 $\{X_i\}$, $\{X'_i\}$;针对两类序列样本,分别以 5% 的显著水平对随机样本 z 进行 Jarque-Bera 检验,发现其分布均服从正态规律,把概率密度分布函数分别记为 $f_1(z)$ 和 $f_2(z)$,见公式 (10)、公式(11).

$$f_1(z) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(z - \mu_1)^2}{2\sigma_1^2}\right) \tag{10}$$

$$f_2(z) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(z - \mu_2)^2}{2\sigma_2^2}\right) \tag{11}$$

其中, μ_1, σ_1 和 μ_2, σ_2 分别是 ERL 模型和 BRL 模型下随机变量 z 的均值和方差.当无线信道的误比特率 P_b 和数据包长度 B 确定后,分布参数 $\mu_1, \sigma_1, \mu_2, \sigma_2$ 的值也确定.例如:假设当前无线信道的误比特率为 $P_b=10^{-3}$,网内数据包长为 50 字节,则根据随机丢包和块状丢包的丢包模型,通过对其各自丢包的分布统计,如图 6 所示,利用参数估计即可得到随机丢包和块状丢包在当前链路条件下各自的随机变量 z 的概率密度分布函数,其中,随机丢包下 z 的分布参数为 $\mu_1=0.33, \sigma_1=0.21$,块状丢包下 z 的分布参数为 $\mu_2=0.58, \sigma_2=0.20$.因此在网络运行前, $f_1(z)$ 和 $f_2(z)$ 可作为判断丢包类型的先验信息提前输送至网内的各节点.

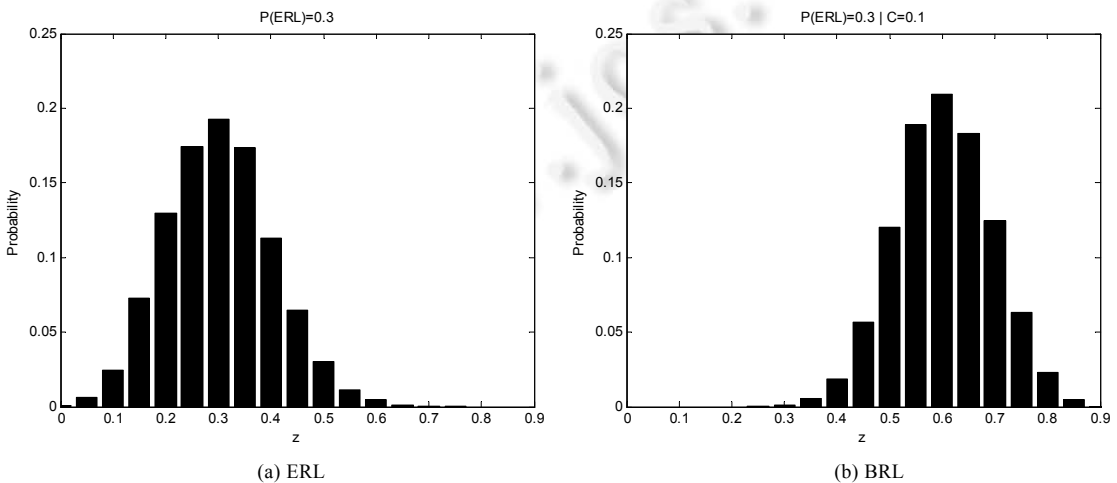


Fig.6 Statistical distribution of random variable in a sliding window

图 6 节点滑动窗内随机变量的统计分布

由于滑动窗内观测量 z 为随机变量,在不同的丢包类型下具有不同的概率密度分布函数,因此,当前链路的丢包类型判断问题可转换为假设检验问题.即存在两个假设,分别记为:当前丢包状态为随机丢包 H_0 和当前丢包状态为块状丢包 H_1 .在这两种假设下,随机变量 z 的概率密度分布函数分别为

$$f(z | H_0) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(z - \mu_1)^2}{2\sigma_1^2}\right) \quad (12)$$

$$f(z | H_1) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(z - \mu_2)^2}{2\sigma_2^2}\right) \quad (13)$$

若用 D_i 表示判决 H_i 为真($i=0,1$),则该假设检验判决结果存在两种错误:虚警概率 $P_f=P(D_1|H_0)$,即将 ERL 判决为 BRL;漏警概率 $P_m=P(D_0|H_1)$,即将 BRL 判决为 ERL.针对本文的应用场景,漏警 P_m 的代价要大于虚警 P_f .这是由于在判决过程中,若将随机丢包判为块状丢包,则对应的代价仅为丢包恢复精度上的小部分损耗;而若将块状丢包判为随机丢包,则可能造成存在关联效应的丢包.因此,引入代价函数 C_{ij} 表示当假设 H_i 为真时,判决 H_j 为真的代价,其取值分别为: $C_{10}=\alpha_1, C_{01}=\alpha_2$,其中, $\alpha_1 > \alpha_2 > 0, C_{00}=C_{11}=0$.选择最小风险 Bayes 判决作为假设检验的似然比判决准则.

根据最小风险 Bayes 判决准则,若似然比:

$$l(z) = \frac{f(z | H_1)}{f(z | H_0)} > \frac{C_{10} - C_{00}}{C_{01} - C_{11}} \cdot \frac{P(H_0)}{P(H_1)} \quad (14)$$

则判 H_1 为真;否则,判 H_0 为真.假设链路中出现块状丢包与随机丢包概率的比值为 R ,即 $P(H_1)/P(H_0)=R$,将公式(12)、公式(13)代入公式(14)可得:

$$z > \frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2 + \sigma_1\sigma_2\sqrt{(\mu_1 - \mu_2)^2 + 2(\sigma_2^2 - \sigma_1^2)\ln\frac{\alpha_1\sigma_2}{\alpha_2\sigma_1R}}}{\sigma_2^2 - \sigma_1^2} \quad (15)$$

其中,不等式右侧为常量,记为判决门限 ξ .因此,网络运行过程中,在滑动窗内计算随机变量 z 值,通过判断 z 值与门限 ξ 大小,即可判断当前链路的丢包状态:若 $z > \xi$,则判定当前丢包状态为块状丢包;反之为随机丢包.当前链路的丢包状态确定后,即可根据当前状态对下一次丢包的丢包类型做出预判,具体过程见算法 1.

算法 1. 丢包预判算法.

输入:接收序列 $\{X_i\}$,信道误比特率 P_b 及包长 B ,代价函数 C_{ij} ,比值 R ;

输出:丢包类型预判结果.

1. **WHEN** packet loss occurs in time slot T
2. Obtain sliding window $\{X_{T-L+1}, X_{T-L+2}, \dots, X_T\}$ from received sequence
3. Compute the number u of $X_i=0$ in sliding window $\{X_{T-L+1}, X_{T-L+2}, \dots, X_T\}$
4. $z=u/L$
5. Obtain $\mu_1, \mu_2, \sigma_1, \sigma_2$ based on P_b and B
6. Compute threshold η
7. **IF** $z < \xi$ **THEN**
8. output ERL
9. update received sequence $\{X_i\}$
10. **ELSE IF** $z > \xi$ **THEN**
11. output BRL
12. update received sequence $\{X_i\}$ with $\{X_{T+1}=1\}$
13. **END IF**

当在 CS 数据收集的 T 时隙时,节点 S 发现自身的接收数据包丢失,则从自身内存中取最近 L 次数据包接收情况组成滑动窗 $\{X_{T-L+1}, X_{T-L+2}, \dots, X_T\}$,基于该滑动窗,结合公式(9)计算当前链路状态下的随机变量 z 值.由于随

机丢包和块状丢包下随机变量 z 的概率密度分布函数 $f_1(z)$ 和 $f_2(z)$ 在网络运行已经确定,因此判决门限 ξ 也为常量.通过比较 z 值和判决门限 ξ 的大小,则可判断当前链路的丢包状态.预判出当前链路的丢包类型,并对丢包进行处理后需更新序列 $\{X_i\}$.对于随机丢包,若重传成功,则在序列 $\{X_i\}$ 中加入“1”作为最新的接收情况;若失败,则在序列 $\{X_i\}$ 中加入“0”.对于块状丢包,采用时间序列相关性预测恢复丢包后,若以“0”更新序列 $\{X_i\}$ 的最新接收情况,滑动窗内 z 值持续增大,则下一次丢包类型预判结果仍为块状丢包,进而会出现预判结果的“自锁”现象;而若以“1”更新序列 $\{X_i\}$ 的最新接收情况,观测量 z 值将缓慢下降,实际上随着时间的推移,块状丢包间的相关性会逐渐减弱,观测量 z 值也将要下降,因此以“1”更新序列 $\{X_i\}$ 符合实际情况,所以在预测出块状丢包并处理后,在序列 $\{X_i\}$ 中加入“1”作为最新的接收情况.

2.2 时间序列相关性预测算法

传感网由于其自身的特点,通常被应用在环境数据收集、监测等领域,而该领域的传感数据通常具有较强的时间相关性.文献[11]提出了利用数据间的时间、空间相关性结合 CS 重构算法恢复传感网 Sink 端数据库内丢包数据的方法,获得了较高的恢复精度.因此,对于 CS 数据收集过程中出现的块状丢包,本文探讨利用数据的时间相关性恢复块状丢包的数据.

2.2.1 构建时间相关性序列

基于树状路由的 CS 数据收集过程中,若链路上发生丢包,则丢失数据为丢包接收节点一支的所有子节点加权叠加数据包和,如图 7 所示.该丢失数据并非节点直接采集到的原始数据 d_j ,而是多个节点采集的原始数据经加权线性叠加后的叠加和 $\sum \phi_j d_j$.由于节点采集的原始数据 d_j 在其时间序列上具有较强的时间相关性,而线性变化并不会改变数据本身之间的相关性,因此若参与叠加的所有节点加权系数 ϕ_j 相同,则节点接收到的加权叠加包 $\sum \phi_j d_j$ 在其时间序列上也具有时间相关性.

由于压缩感知每轮数据收集分为 M 次观测分别进行,而每次观测的观测系数 ϕ_{ij} 均不同,但对于不同轮的同一观测,观测系数 ϕ_{ij} 是相同的,则可选择节点不同轮同一观测的接收历史数据构造时间相关性序列.如图 8 所示:若节点在其第 $T+1$ 轮的第三次观测中接收数据丢失(如图 8 中黑色部分),则构建时间相关性序列的数据应为图 8 中灰色条状区域的数据,其长度为 k .记节点在第 i 轮第 i 次观测中接收到的加权叠加数据包为 h_i^i ,则节点的 k 阶时间相关性序列记为 $H_k = (h_{T-k+1}^i, h_{T-k+2}^i, \dots, h_T^i)^T$,其中, $i \in [1, 2, \dots, M]$.

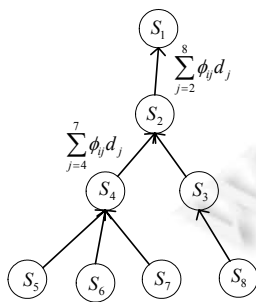


Fig.7 Process of CS-based data gathering under tree routing

图 7 基于树状路由的 CS 数据收集

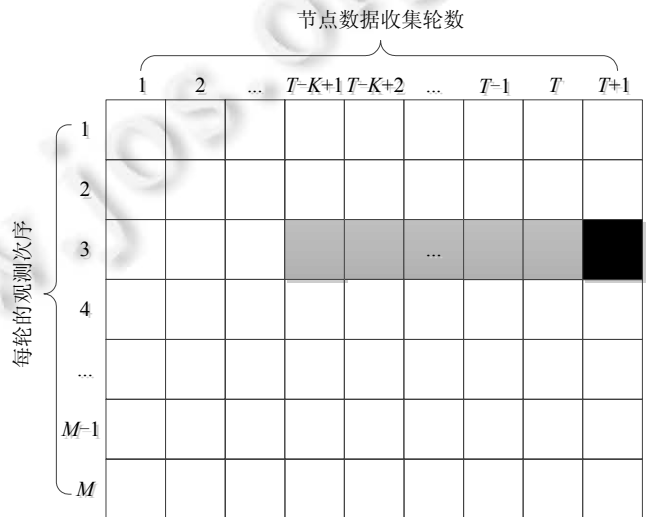


Fig.8 Data storage distribution of receiving node

图 8 接收节点的数据存储分布

2.2.2 时间相关性预测

为提高时间相关性序列数据预测的精度,本文选择加权移动平均模型^[18]作为序列预测模型,加权平均模型见公式(16).

$$\hat{h}_{t+1} = \frac{w_1 h_{t-k+1} + w_2 h_{t-k+2} + \dots + w_k h_t}{w_1 + w_2 + \dots + w_k} \tag{16}$$

其中, \hat{h}_{t+1} 为序列的加权移动预测值; w_i 为序列中数据 h_{t-k+i} 的权值,它反映了数据 h_{t-k+i} 对预测值 \hat{h}_{t+1} 的影响程度. 对于具有时间相关性的数据序列,数据之间的相关性强弱与其数据间的时间间隔长短有关,时间间隔越短,数据间的相关性越强;反之则越弱.因此,在利用节点的 k 阶时间相关性序列 H_k 预测第 $T+1$ 轮第 i 次观测的接收数据 \hat{h}_{T+1}^i 时,序列中各数据的权值选择受各数据与预测值 \hat{h}_{T+1}^i 间的时间间隔影响,时间间隔越短,则对应数据的权值越大;反之则越小.根据序列中各数据距预测时隙的时间间隔长短,设置见表 1 所示的权值.

Table 1 Weighted coefficient of time series
表 1 时间序列加权系数

数据收集轮	$T-k+1$	$T-k+2$...	$T-1$	T
数据值	h_{T-k+1}^i	h_{T-k+2}^i		h_{T-1}^i	h_T^i
权值 W	1	2		$k-1$	k

若引入加权向量 $A=(a_1, a_2, \dots, a_k)$, 其中, $a_n = n / \sum_{j=1}^k j, a_n (n \in [1, k])$ 表示序列中数据 h_{T-k+n}^i 的归一化权值,则节点在第 $T+1$ 轮第 i 次观测的接收预测数据 \hat{h}_{T+1}^i 为

$$\hat{h}_{T+1}^i = A \cdot H_k = \sum_{n=1}^k \frac{n}{1+2+\dots+k} h_{T-k+n}^i \tag{17}$$

假设节点 S 在 $T+1$ 轮的第 i 次观测中发生块状丢包,则节点根据自身内存中存储的历史接收数据,结合图 8 所示方法构建 k 阶时间相关序列 H_k ;之后,根据预测公式(17)可得丢包的预测数据 \hat{h}_{T+1}^i .其中,预测阶数 k 值需预先确定,且预测阶数 k 的不同取值将影响节点丢包的预测精度,本文根据预测误差反馈动态调整 k 值.定义最大允许预测误差为 e_{max} ,设置初始值 $k_0=3$,当链路丢包采用时间序列相关性补包后,等待未来最近的一次正常接收的真实数据 h_{real} ;当正确接收到真实传输数据 h_{real} 后,利用当前 k 值,结合公式(17)再次预测真实数据 h_{real} 对应轮的数据预测值 \hat{h}_{real} ,计算预测误差 $e = |h_{real} - \hat{h}_{real}|$.若预测误差 $e \leq e_{max}$,则不需调整 k 值;若预测误差 $e > e_{max}$,则分别取调整步长为 $+\Delta k$ 和 $-\Delta k$ 调整 k 值,并分别再次计算当前预测误差 e_1, e_2 ,若 $|e_1| < |e_2|$,则选取步长 $+\Delta k$ 更新 k 值;否则,选取步长 $-\Delta k$ 更新 k 值.具体过程见算法 2.

算法 2. 节点 S_j 在第 $T+1$ 轮的第 i 次观测中接收数据丢失的时间序列相关性预测算法.

输入:加权向量 A ,调整步长 Δk ,最大允许误差 e_{max} ;

输出:发送至下一跳节点的数据包 $(\hat{h}_{T+1}^i + \phi_j d_j)$,下一次丢包预测阶数 k .

1. **IF** node S_j found received date h_{T+1}^i loss **THEN**
2. Obtain vector H_k from the memory of node S
3. $\hat{h}_{T+1}^i = A \cdot H_k$
4. node S send $(\hat{h}_{T+1}^i + \phi_j d_j)$ to the next hop node
5. **END IF**
- /*Updating the order k of temporal correlation prediction*/
6. **WHILE** node S receives real date x_{real} from the last hop node in the next slots, **DO**
7. $\hat{h}_{real} = A \cdot H_k$
8. $e = |h_{real} - \hat{h}_{real}|$

```

9.   IF  $e \leq e_{max}$  THEN
10.     $k=k$ 
11.  ELSE IF  $e > e_{max}$  THEN
12.     $k_1=k+\Delta k, k_2=k-\Delta k$ 
13.     $\hat{h}_{real1} = A \cdot H_{k_1}, \hat{h}_{real2} = A \cdot H_{k_2}$ 
14.     $e_1 = |h_{real} - \hat{h}_{real1}|, e_2 = |h_{real} - \hat{h}_{real2}|$ 
15.    IF  $e_1 > e_2$  THEN
16.       $k=k_2$ 
17.    ELSE IF  $e_1 \leq e_2$  THEN
18.       $k=k_1$ 
19.    END IF
20.  END IF
21. END WHILE
    
```

2.3 CS-RTSC数据收集算法

针对有损链路下的丢包及现存压缩感知数据收集算法在丢包时不能有效重构的问题,本文提出了 CS-RTSC 算法,采用重传与时间序列相关性相结合的方式,保证了压缩感知数据收集算法在有损链路下的正常实施,算法流程如图 9 所示.

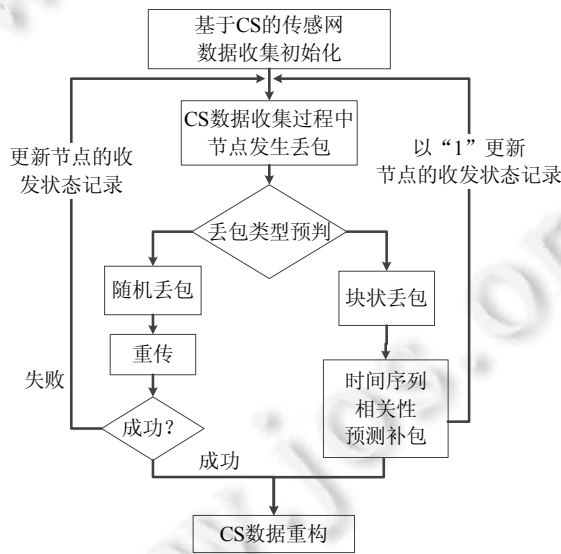


Fig.9 Flow chart of CS-RTSC algorithm

图 9 CS-RTSC 算法流程图

CS-RTSC 算法的数据收集过程分为 3 个阶段分别进行:第 1 阶段为传感网初始化阶段,该阶段完成的是全网节点的组网、网内各链路接收状态的先验信息积累和 CS 数据采样的节点观测向量配置;第 2 阶段为 CS 数据收集阶段,该阶段实现在有损链路下对网内各节点采集数据的有效 CS 采样和收集;第 3 阶段为采样数据的 CS 重构阶段,该阶段针对采样数据进行重构,以获得网内各节点的原始采集数据.算法的具体过程如下.

(1) 传感网络初始化.

首先,全网节点由 Sink 端发起,构建最小生成树(minimum spanning tree,简称 MST)路由,完成节点组网;其

次,Sink 向全网广播心跳数据包,网内各节点在接收到 Sink 端心跳包后,开启自身定时器,在时间段 T_1 内,全网节点以传统的发送、转发的数据收集方式沿路由向 Sink 端传递节点自身心跳包,此过程中,各节点实时统计记录自身接收链路上数据包的接收情况,并将结果存储在其内存中,作为滑动窗丢包预判机制的先验信息,即,初始化的接收序列 $\{X_i\}$,直至时间 T_1 结束;最后,Sink 先向全网广播随机种子 ξ ,节点 i 收到随机种子 ξ 后与自身的节点 ID 号结合 (ξ, ID_i) ,生成属于自身的一系列观测矩阵 $(\phi_{i1}, \phi_{i2}, \dots, \phi_{iM})^T$ 存储到自身内存中。

(2) CS 数据收集.

在 CS 数据收集过程中,各节点按照路由将采集数据 d_i 与对应的观测系数 ϕ_{ij} 相乘,依次叠加中转传送至 Sink 端.此过程中,当节点发现自身的接收数据包丢失,根据当前无线信道的误比特率 P_b 及数据包长度 B 确定随机变量 z 在两类丢包模型下的概率密度分布函数 $f_1(z)$ 和 $f_2(z)$,进而根据公式(15)计算丢包类型的判决门限 ξ ,在滑动窗内,计算窗内随机变量 z 的值,比较 z 值与门限 ξ 的大小.当 $z < \eta$ 时,选择重传机制恢复丢包,设定最大重传次数为 \max_num ,重传过程中,实时更新滑动窗.若达到最大重传次数后,节点的接收数据仍未被成功恢复,则基于最新更新的滑动窗计算观测量 z ,比较 z 值与门限 ξ 的大小,根据二者的大小情况再次预判当前丢包类型,并根据预判结果采用相应机制恢复丢包.当 $z > \eta$ 时,选择时间序列相关性预测恢复丢包,构建丢包节点的 k 阶时间相关性序列 H_k ,根据公式(17)计算丢包的预测值 \hat{h}_{T+1}^i ,将预测值作为丢包节点接收到的数据包向下一跳节点传输,完成本次 CS 观测采样,同时完成对预测阶数 k 的更新.

(3) 数据重构.

Sink 节点在收到一轮数据收集的 M 个观测值后,构建观测向量 $\mathbf{Y}=(y_1, y_2, \dots, y_M)^T$,Sink 根据随机种子 ξ 和全网节点 ID 号重建观测矩阵 $(\phi_{ij})_{M \times N}$,结合稀疏基 $\Psi_{N \times N}$,利用 CS 重构算法重构出稀疏信号 \mathbf{S} ,通过计算 $\mathbf{d}=\Psi \mathbf{S}$ 即可重构出原始信号向量 \mathbf{d} .

根据上述过程可知:全网各节点对丢包类型的预判、丢包的重发和时间相关性补包预测的处理均为简单的顺序结构语句,其执行的复杂度为 $O(1)$.因此,对于性能受限的单节点,算法可保证在线实时运行.对于整个网络,CS-RTSC 算法执行一轮数据收集,全网至少需收发 $N \times M$ 个数据包,其中, N 为全网节点个数, M 为观测次数.由于 CS 数据收集过程中,收发一数据包在算法执行上为一次乘法和加法线性运算,并不涉及复杂的循环等操作,且根据 CS 理论,观测次数 $M \geq O(K \log N)$, K 为数据稀疏度,可视为一个常量,因此对于全网而言,CS-RTSC 算法的复杂度为 $O(M \log N)$.相较于传统的 CS 稠密投影数据收集算法,该算法并不增加额外的计算开销.

3 仿真与分析

假设传感网络的部署规模为中小型网络,网络的拓扑结构不轻易发生变化,网内各节点周期地从环境中采集传感数据,且传感数据的数据量较小.为验证 CS-RTSC 算法的有效性,本文选取来自 GreenOrbs 系统 225 个传感节点采集的 100 期数据,每期数据采集间隔为 10min.仿真环境如下:传感节点被随机均匀布设在 30×30 的监测区域内,节点以 MST 构建全网路由,汇聚节点 Sink 位于网络中央,坐标为(15,15).传感节点周期地采集监控部署区域内的环境数据,并以多跳路由的 CS 数据收集方式将数据传送至 Sink 端,在 Sink 端重构各节点的原始采集数据.压缩采样过程中的稀疏基采用 DCT 基;观测矩阵采用文献[12]中给出的观测矩阵,见公式(18).

$$\phi_{ij} = \sqrt{s} \begin{cases} +1, & \text{with prob. } \frac{1}{2s} \\ 0, & \text{with prob. } 1 - \frac{1}{s} \\ -1, & \text{with prob. } \frac{1}{2s} \end{cases} \quad (18)$$

其中,参数 s 可控制观测矩阵的稀疏程度,本文取 $\frac{1}{s} = 1$,即,采用稠密观测矩阵;Sink 端重构算法采用 OMP 算法.

无线传感网络的能耗采用文献[19]中所述的能耗模型,见公式(19)、公式(20).

$$E_T(d, l) = (\alpha_1 + \alpha_2 \times d^m) \times l \quad (19)$$

$$E_R(l) = \alpha_1 \times l \quad (20)$$

其中, $E_T(d, l)$ 表示发送 l 比特数据的能耗, d 表示发送距离, α_1 表示发送和接收的电路能耗, α_2 表示距离衰减系数, n 为路径损耗因子 ($2 < n < 5$, 在自由空间一般取 $n=2$); $E_R(l)$ 表示接收 l 比特数据的能耗。

本文选择 CDG 算法^[3]、CS-Hybrid 算法^[5]及 SRS-DG 算法^[12]作为对比算法, 其中, CDG、CS-Hybrid 算法均为经典的树状路由下 CS 数据收集算法, 二者不同的是 CDG 算法的全网节点均参与 CS 数据收集, 而 CS-Hybrid 算法仅有满足转发数据包数大于观测次数 M 的节点才以 CS 方式收集数据; SRS-DG 算法以极稀疏的方式收集和重构全网数据, 其数据收集过程仍采用传统的传感网数据收集方式。性能评估指标为相对重构误差和网络寿命, 其中,

- 相对重构误差的定义见公式(5), 相对重构误差越小, 则算法的重构精度越高。由于本文选择的仿真数据来自 GreenOrbs 系统的真实温度数据, 数据间具有较强的空间相关性, 因此设定阈值 $\omega=0.03$ 。当相对重构误差 $\eta \leq \omega$ 时, 认为重构成功; 否则认为重构失败;
- 网络寿命定义为网络从开始运行到第 1 个节点死亡的时间, 网络寿命越长, 则算法的能耗性能越好。

设置仿真参数如下: CS 观测次数 M 为 20, 丢包的最大重传次数 $\max_num=5$, 丢包则采用由 70% 随机丢包和 30% 块状丢包组成的混合丢包, 时间序列相关性预测补包阶数 $k=5$ 。

3.1 CS-RTSC 算法性能

如图 10(a) 所示为 CS-RTSC 算法、CDG 算法、CS-Hybrid 算法及 SRS-DG 算法在重构精度上的性能对比图。由图可知: 随着链路平均丢包率的增加, CDG 算法与 CS-Hybrid 算法的数据相对重构误差迅速增加, 而 CS-RTSC 算法和 SRS-DG 算法的相对重构误差则增加相对缓慢; 当链路丢包率达到 40% 时, 有重传机制保障的 CS-Hybrid 算法和 CDG 算法的重构相对误差达到 0.6, 远超过了成功重构的阈值 ω , 而 CS-RTSC 算法的相对误差仅为 0.018, 小于阈值 ω , 成功重构出全网采集数据。造成二者相对重构误差差距大的原因为: 在有损链路下, CS-RTSC 算法采用重传与时间相关性预测相结合的方式保证了 CS 数据收集过程中“欠采样”数据的完整性; 而 CDG 算法和 CS-Hybrid 算法采用有限次的重传, 虽能部分恢复丢失的“欠采样”数据, 但不能保证其完整性; 因此在有损链路下, CS-RTSC 算法能够有效改善丢包对 CS 数据收集的影响。

图 10(b) 为 CS-RTSC 算法与 SRS-DG 算法重构性能在细节上的性能对比, 两个算法在链路丢包率低于 40% 时, 相对重构误差均低于阈值 ω , 即能够成功重构出全网数据; 虽然在链路平均丢包率低于 30% 时, 二者重构性能相差不大, 而当链路平均丢包率大于 30%, SRS-DG 算法的相对重构误差会迅速增加。这是由于链路丢包率影响 SRS-DG 算法在 Sink 端构建极稀疏观测矩阵的行数, 即 CS 的观测次数 M , 当平均丢包率增大时, CS 的观测次数随其减小, 进而造成 CS 观测次数不足, 降低了重构精度; 而对于本文提出的 CS-RTSC 算法, 链路平均丢包率仅造成 CS 数据收集过程中的丢包恢复数量的增加, 但并不影响 CS 的观测次数, 因此对相对重构误差的影响也较小。

为讨论数据源的相关性对算法重构精度的影响, 本文采用文献[7]提出的事件源衰减的空间相关性模型生成不同空间相关性的数据源, 其中, 通过调整事件源衰减系数 n 可改变数据源的空间相关性, 衰减系数越大, 数据源的空间相关性越弱。由于 CS-RTSC 算法与 CDG 算法、CS-Hybrid 算法采用相同的观测矩阵和重构算法, 因此其重构精度受数据源相关性的影响程度相同, 在此选择 CS-RTSC 算法代表此 3 种算法的性能。图 11 给出了 CS-RTSC 算法与 SRS-DG 算法在不同的事件源衰减系数下, 数据的重构相关误差的对比图, 其中, 设置网络的平均丢包率为 20%。由图可知: 在不同的数据空间相关性下, CS-RTSC 算法的重构精度优于 SRS-DG 算法, 且数据的空间相关性越弱, CS-RTSC 算法的重构精度优势越明显。这是由于 SRS-DG 算法采用了极稀疏的观测矩阵, 仅将小部分的采集数据直接作为压缩感知的观测值, 重构全网节点的原始数据, 因此网络的数据源必须具有较强的空间相关性才能使 SRS-DG 算法有效; 而 CS-RTSC 算法采用稠密观测矩阵, 全网节点的采集数据均参与压缩采样过程, 因此该算法对数据源相关性的要求也相对较弱。在实际应用中, 由于不同传感网系统采集到数据的空间相关性强弱不同, 同一个传感网系统采集数据的空间相关性强弱也随时间的改变而变化, 所以相较于 SRS-DG 算法, 本文提出的 CS-RTSC 算法具有更好的适用性。

在网络寿命方面, 图 12 给出在相同的重构精度下, 当链路平均丢包率为 20% 时, 不同算法的数据收集轮数 r

与节点死亡数的关系.其中,设置节点的初始能量为 1J,数据包长度 B 为 50 字节,数据压缩率 $\rho=M/N=0.11$,节点平均发送距离 $d=4$,收发的电路能耗 α_1 为 50nJ/bit,距离衰减系数 α_2 为 10pJ/bit/m².如图 12 所示:对于 SRS-DG 算法,死亡节点出现相对分散,虽然其采用极稀疏方法使得网络的整体发包量减少,但网络能耗极不均衡,Sink 的一跳邻居节点在数据收集的前 400 轮几乎已经死亡殆尽,而使整个网络提前失效;而对于 CDG 算法、Hybird-CS 算法和 CS-RTSC 算法,由于数据收集过程中采用压缩感知,所以节点死亡较为集中;另外,由于压缩感知每轮数据收集需全网节点参与,所以当网络中第 1 个节点死亡即认为网络寿命终止.由仿真结果可得,Hybird-CS 算法、CDG 算法、CS-RTSC 算法分别在第 774 轮、第 870 轮、第 1 010 轮出现第 1 个死亡节点.因此,在链路平均丢包率为 20%时,CS-RTSC 算法相对于 Hybird-CS 算法和 CDG 算法分别增加了 30%和 16%的网络寿命,有效保证了能耗均衡,延长了网络寿命.

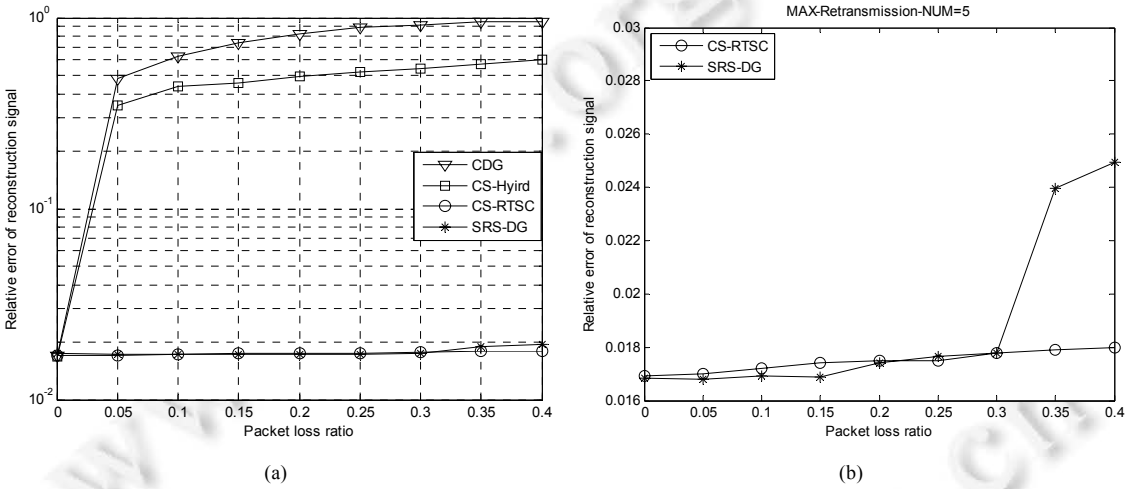


Fig.10 Comparison of algorithms reconstruction performance

图 10 算法数据重构性能对比

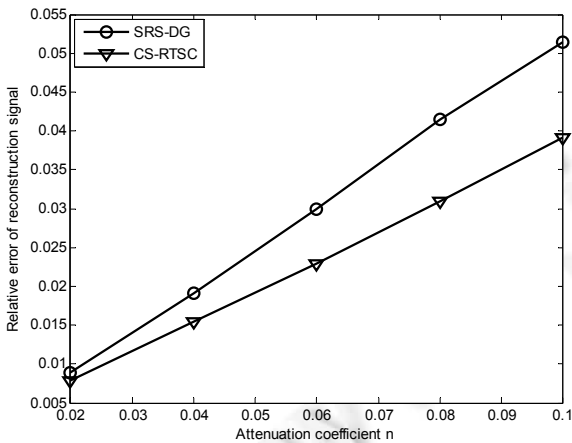


Fig.11 Impact of data spatial correlation on reconstruction performance

图 11 数据空间相关性对算法重构性能影响

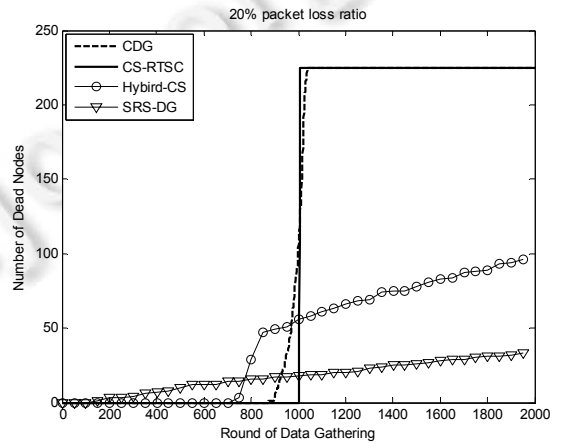


Fig.12 Comparison of network lifetime

图 12 网络寿命对比

3.2 影响算法性能因素分析

对于 CS-RTSC 算法,影响其性能的因素主要有:算法的数据压缩率 $\rho=M/N$ 、时间序列相关性补包的预测阶数 k 、重传允许的最大重传次数 \max_num 和混合丢包的组成比例等.其中,用有损链路下块状丢包出现概率与随机丢包出现概率的比值 $R=p_{BRL}/p_{ERL}$ 来描述混合丢包的组成比例.本节将分别通过仿真实验分析上述因素对性能的影响.

图 13 为在 $\max_num=5$ 、时间序列相关性预测阶数为 3、 $R=0.5$ 的条件下,CS-RTSC 算法相对重构误差随数据压缩率的变化关系.随数据压缩率 ρ 的增加,算法的相对重构误差逐渐减小,压缩率 ρ 在不小于 0.075 时,3 种链路条件下均可实现成功重构;当 $\rho>0.11$ 后,CS-RTSC 算法的重构精度趋于稳定.由于数据压缩率一方面影响 CS 数据重构的精度,另一方面也影响网络的能耗,随数据压缩率的增大,CS 数据重构精度提高,而网络能耗却增加,因此在选择数据压缩率 ρ 时,需权衡网络能耗和数据的重构精度.另外,根据上述仿真结果,当数据的压缩率 ρ 相同时,CS-RTSC 算法的数据相对重构误差随链路的平均丢包率变化不大.如:当数据压缩率为 0.089 时,理想链路、平均丢包率为 20%和 40%链路的相对重构误差分别为 0.0168,0.0175 和 0.0195,三者之间彼此相差仅为 0.002 左右;而当数据压缩率为 0.11 时,三者之间彼此相差不足 0.001.因此,CS-RTSC 算法的重构精度随链路丢包率变化较小.

图 14 为时间序列相关性补包的预测阶数 k 与 CS-RTSC 算法重构性能的关系,其中,数据压缩率 $\rho=0.11$, $\max_num=5$ 和 $R=0.5$.通过改变 GreenOrbs 系统数据的采集间隔来改变数据序列的时间相关性,采集间隔越大,数据序列的时间相关性越弱;反之则越强.仿真结果表明:随预测阶数 k 增大,CS-RTSC 算法相对重构误差逐渐增大.这是由于在节点采集数据的时间序列相关性中,相邻数据之间的相关性最强,所以仅采用丢包上一时隙的数据预测丢包数据时,对应的预测误差最小.另外,数据序列的时间相关性不同,重构误差随预测阶数 k 增大的速率也不相同,数据序列的时间相关性越弱,重构精度越差,预测阶数 k 对算法重构精度的影响越大.因此 CS-RTSC 算法在环境数据发生较大波动时,重构性能将变差.而在实际数据采集过程中,由于 WSN 采集数据的时间相关性会随时间、空间的变化而变化,且发生块状丢包后可能存在连续预测的情况,因此,为保证算法的重构性能,本文采用基于重构精度反馈的自适应预测阶数 k .

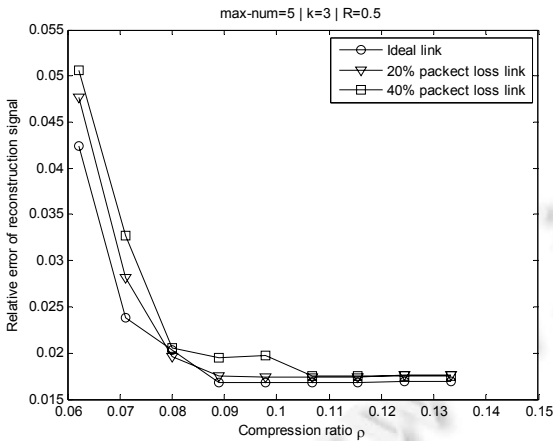


Fig.13 Relationship between data compression ratio and relative reconstruction error of CS-RTSC

图 13 数据压缩率与 CS-RTSC 相对重构误差的关系

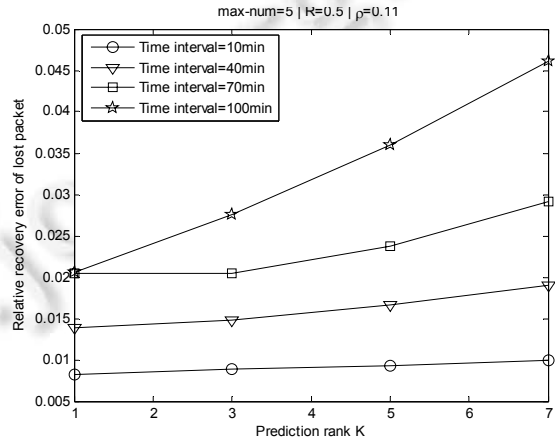


Fig.14 Impact of prediction rank on recovery accuracy of packet loss

图 14 预测阶数对丢包恢复精度的影响

在数据压缩率 $\rho=0.11$ 、时间序列相关性预测阶数 $k=3$ 和 $R=0.5$ 的条件下,仿真最大重传次数 \max_num 对 CS-RTSC 算法性能的影响如图 15 所示.由图可知:随最大重传次数的增加,算法重构的相对误差逐渐减小,并且减小速率逐渐变慢.如在丢包率为 20%时,最大重传次数为 1,3,5,7 的相对重构误差分别为 0.310,0.056,0.019,0.016.

由于算法成功重构的阈值 $\omega=0.03$,则为保证算法的重构成功率,通常选择的最大重传次数应不小于 5.

对于 CS-RTSC 算法,链路上出现的不同丢包类型对算法重构精度的影响程度也不相同,因此,本文在最大重传次数为 5、预测阶数为 3、数据压缩率 $\rho=0.11$ 的条件下,仿真块状丢包出现概率与随机丢包出现概率的比值 $R=p_{BRL}/p_{ERL}$ 与 CS-RTSC 算法相对重构误差的关系如图 16 所示.其中,比值 R 越大,混合丢包中块状丢包所占的比例就越大.由图可知:当链路的平均丢包率相同时,混合丢包中块状丢包的比例越大,算法的重构精度就越差.由于块状丢包采用时间序列预测处理,预测结果存在预测误差,而随机丢包采用重传机制可恢复丢失的原始数据,所以块状丢包比例越大,CS 压缩的原始数据就存在越大的误差,进而经重构后存在的误差也越大.另外,如图 16 所示:在平均丢包率为 30%时,比值 R 由 0 增至 2,算法的相对重构误差从 0.017 1 增加至 0.017 5,其误差的增幅相对较小.此现象说明本文设计的时间序列相关性预测丢包算法所能达到的重构精度与原始数据包达到的重构精度相差较小,也即本文对块状丢包的处理算法具有有效性.

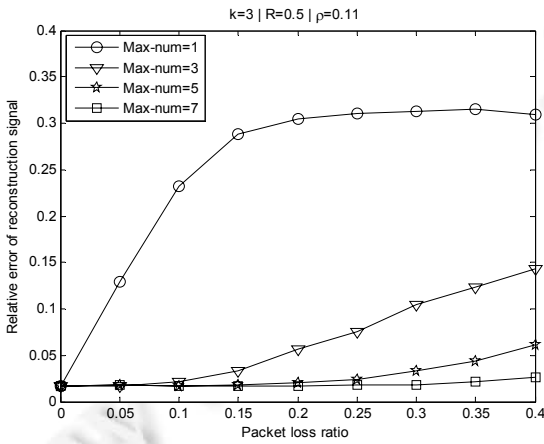


Fig. 15 Impact of maximum retransmission number on the performance of CS-RTSC

图 15 最大重传次数对 CS-RTSC 性能的影响

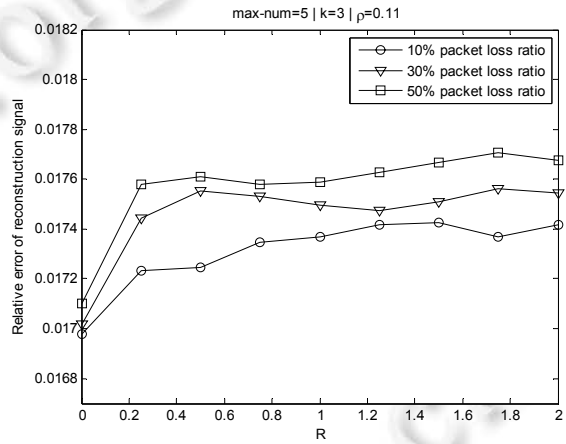


Fig. 16 Impact of CL composition on the performance of CS-RTSC

图 16 混合丢包的组成比例对 CS-RTSC 性能影响

4 总结

本文面向在中小型无线传感网络下,有损链路丢包造成现存 CS 数据收集算法不能有效重构的问题,提出了 CS-RTSC 算法.在 CS 数据收集过程中,若链路上出现丢包,提出采用滑动窗实时统计分析节点当前的丢包类型,对于随机类型丢包采用重传机制恢复,对于块状丢包设计了时间相关性序列预测恢复,从而降低了有损链路丢包的影响,保证了 CS 数据收集在有损链路下的正常实施.仿真结果表明:该算法可有效保证 CS 数据收集在有损链路下的正常实施,且具有较好的重构精度;相对于仅采用重传恢复丢包的 CS 数据收集算法,有效延长了网络寿命.另外,本文还分析了影响 CS-RTSC 算法性能的因素,分别讨论了算法的数据压缩率 $\rho=M/N$ 、时间序列相关性补包的预测阶数 k 、重传允许的最大重传次数 \max_num 和混合丢包的组成比例等因素对算法性能的影响.

通过对 CS-RTSC 算法影响因素的分析,该算法在网络环境数据较为平稳且网内传输时延低、流量负载轻的条件下,具有更好的性能.这是由于若环境数据剧烈变化,一方面,压缩感知重构算法的性能将会受限;另一方面,网内数据的时间相关性变弱,将使时间序列相关性预测丢包的精度下降,从而影响算法的重构性能.对于网内的传输时延,若时延过大,将会使节点对数据包接收情况造成误判,从而影响了当前链路状态的判断;对于网络流量负载,若网内各节点的流量负载过大,则容易造成网络拥塞,进而增加了节点丢包的块状丢包比例.根据分析可知,块状丢包的比例增高将会降低算法的重构精度.因此,CS-RTSC 算法在轻流量负载的网络中具有更好的性能.

References:

- [1] Shi GM, Liu DH, Gao DH, Liu Z, Lin J, Wang LJ. Advances in theory and application of compressed sensing. Chinese Journal of Electronics, 2009,37(5):1070–1081 (in Chinese with English abstract). [doi: 10.3321/j.issn:0372-2112.2009.05.028.]
- [2] Luo C, Wu F, Sun J, Chen CW. Compressive data gathering for large-scale wireless sensor networks. In: Proc. of the 15th Annual Int'l Conf. on Mobile Computing and Networking. New York: ACM Press, 2009. 145–156. [doi: 10.1145/1614320.1614337]
- [3] Luo C, Wu F, Sun J, Chen CW. Efficient measurement generation and pervasive sparsity for compressive data gathering. IEEE Trans. on Wireless Communications, 2010,9(12):3728–3738. [doi: 10.1109/TWC.2010.092810.100063]
- [4] Ebrahimi D, Assi C. Compressive data gathering using random projection for energy efficient wireless sensor networks. Ad Hoc Networks, 2014,16(1570-8705):105–119. [doi: 10.1016/j.adhoc.2013.12.004]
- [5] Nguyen MT, Teague KA. Tree-Based energy-efficient data gathering in wireless sensor networks deploying compressive sensing. In: Proc. of the 23rd Wireless and Optical Communication Conf. (WOCC). New York: IEEE Press, 2014. 1–6.
- [6] Wang W, Garofalakis M, Ramchandran K. Distributed sparse random projections for refinable approximation. In: Proc. of the 6th Int'l Conf. on Information Processing in Sensor Networks. New York: ACM Press, 2007. 331–339. [doi: 10.1109/IPSNS.2007.4379693]
- [7] Zhang C, Zhang X, Li O, Wang C, Zhang DL. Data gathering using dynamic clustering based on WSN compressive sensing algorithm. Journal of Computer Research and Development, 2016,53(9):2000–2008 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2016.20150459]
- [8] Mahmudimanesh M, Khelil A, Suri N. Balanced spatio-temporal compressive sensing for multi-hop wireless sensor networks. In: Proc. of the 9th Int'l Conf. on Mobile Ad-Hoc and Sensor Systems (MASS). New York: IEEE Press, 2012. 389–397. [doi: 10.1109/MASS.2012.6502539]
- [9] Wang J, Ji SQ, Cheng Y. Spatial-Temporal correlation-based low-latency compressed sensing in WSNs. Int'l Journal of Grid and Distributed Computing, 2015,8(3):53–64. [doi: 10.14257/ijgcd.2015.8.3.06]
- [10] Li P, Hastie TJ, Church KW. Very sparse random projections. In: Proc. of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2006. 287–296. [doi: 10.1145/1150402.1150436]
- [11] Kong LH, Xia MY, Liu XY, Wu MY, Liu X. Data loss and reconstruction in sensor networks. In: Proc. of the 2013 IEEE INFOCOM. New York: IEEE Press, 2013. 1654–1662. [doi: 10.1109/INFOCOM.2013.6566962]
- [12] Wu XG, Yang PL, Jung T, Xiong Y, Zheng X. Compressive sensing meets unreliable link: Sparsest random scheduling for compressive data gathering in lossy WSNs. In: Proc. of the 15th ACM Int'l Symp. on Mobile Ad Hoc Networking and Computing. New York: ACM Press, 2014. 13–22. [doi: 10.1145/2632951.2632969]
- [13] Srinivasan K, Kazandjieva MA, Agarwal S, Levis P. The β -factor: Measuring wireless link burstiness. In: Proc. of the 6th ACM Conf. on Embedded Network Sensor Systems. New York: ACM Press, 2008. 29–42. [doi: 10.1145/1460412.1460416]
- [14] Munir S, Lin S, Hoque E, Nirjon SM, Stankovic JA, Whitehouse K. Addressing burstiness for reliable communication and latency bound generation in wireless sensor networks. In: Proc. of the 9th ACM/IEEE Int'l Conf. on Information Processing in Sensor Networks. New York: ACM/IEEE Press, 2010. 303–314. [doi: 10.1145/1791212.1791248]
- [15] Guo ZQ, Wang Q, Wan YD, Wu L. Modeling of burstiness for IEEE 802.15.4 radio link and optimizing for retransmission mechanism in industrial environments. Ruan Jian Xue Bao/Journal of Software, 2012,23:77–86 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/12009.htm>
- [16] Wan CY, Campbell AT, Krishnamurthy L. Pump-Slowly, fetch-quickly (PSFQ): A reliable transport protocol for sensor networks. IEEE Journal on Selected Areas in Communications, 2005,23(4):862–872. [doi: 10.1109/JSAC.2005.843554]
- [17] Liu YH, He Y, Li M, Wang JL, Liu KB, Mo LF, Dong W, Yang Z, Xi M, Zhao JZ, Li XY. Does wireless sensor network scale? A measurement study on GreenOrbs. IEEE Trans. on Parallel and Distributed Systems, 2013,24(10):1983–1993. [doi: 10.1109/TPDS.2012.216]
- [18] Kapgate D. Weighted moving average forecast model based prediction service broker algorithm for cloud computing. Int'l Journal of Computer Science and Mobile Computing, 2014,3(2):71–79.
- [19] Wang C, Zhang X, Li O. Sparse random projection algorithm based on minimum energy tree in wireless sensor network. Journal of Communications, 2015,10(9):740–746.

附中文参考文献:

- [1] 石光明,刘丹华,高大化,刘哲,林杰,王良君.压缩感知理论及其研究进展.电子学报,2009,37(5):1070-1081. [doi: 10.3321/j.issn:0372-2112.2009.05.028]
- [7] 张策,张霞,李鸥.基于压缩感知的无线传感网动态分簇数据收集算法.计算机研究与发展,2016,53(9):2000-2008. [doi: 10.7544/j.issn1000-1239.2016.20150459]
- [15] 郭志强,王沁,万亚东,吴龙.工业环境中 IEEE 802.15.4 链路突发性建模及重传机制优化.软件学报,2012,23:77-86. <http://www.jos.org.cn/1000-9825/12009.htm>



韩哲(1991—),男,河南洛阳人,硕士生,主要研究领域为无线传感网络,压缩感知,嵌入式系统.



张策(1991—),男,博士,主要研究领域为无线传感网络,压缩感知.



张霞(1979—),女,博士,副教授,主要研究领域为无线传感网络协议,协议逆向分析,可见光通信.



张大龙(1976—),男,博士,讲师,主要研究领域为无线通信和移动通信协议,大规模数据处理.



李鸥(1961—),男,博士,教授,博士生导师,主要研究领域为无线传感网络,认知无线电,无线自组织网络.

www.jos.org.cn