



一种基于 El Gamal 加密算法的压缩数据收集方法

虞晓韩¹, 董克明², 李霞¹, 陈超¹

(1. 浙江工商大学, 浙江 杭州 310018; 2. 云南财经大学, 云南 昆明 650221)

摘要: 压缩感知技术在信号处理、图像处理、数据收集与分析等方面有很大优势, 是近年来的研究热点。研究了如何安全高效地运用压缩感知技术来收集无线传感器网络中的数据。传统的基于压缩感知技术的数据收集方法并不考虑数据收集的安全性, 而且网络内的所有节点都会参与每个测量值的收集。将 El Gamal 加密算法和基于稀疏随机矩阵的压缩感知技术相结合, 提出了一种基于 El Gamal 加密算法的稀疏压缩数据收集方法 (El Gamal based sparse compressive data gathering, ESCDG)。理论分析和数值实验表明, ESCDG 不仅能降低网络资源的消耗而且能抵御多项式算力的内部攻击和外部攻击。

关键词: 物联网; 数据收集; 压缩感知; 稀疏随机矩阵; 公钥加密

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019284

A compressive data gathering method based on El Gamal cryptography

YU Xiaohan¹, DONG Keming², LI Xia¹, CHEN Chao¹

1. Zhejiang Gongshang University, Hangzhou 310018, China

2. Yunnan University of Finance & Economics, Kunming 650221, China

Abstract: As a hot researching area, compressive sensing has advantages in signal process, image processing and data gathering and analysis. How to effectively and safely apply compressive sensing technology to collect data in wireless sensor networks is studied. Conventional compressive sensing technology based data collection method does not consider the security issue. Furthermore, every node in the network must participate in the process of collecting measurements. The El Gamal cryptography and sparse matrices based compressive sensing technology were combined, and the El Gamal based sparse compressive data gathering (ESCDG) method was proposed. Theoretical analysis and numerical experiments show that ESCDG not only decreases the network resource consumption but also be able to defense the internal and external attacks.

Key words: IoT, data gathering, compressive sensing, sparse random matrix, public cryptography

收稿日期: 2019-05-28; 修回日期: 2019-12-10

基金项目: 国家自然科学基金资助项目 (No.61701441, No.61703368, No.61801427); 浙江省自然科学基金资助项目 (No.LQ17F010001); 浙江科技计划重点项目 (No.2018C01084)

Foundation Items: The National Natural Science Foundation of China (No.61701441, No.61703368, No.61801427), Zhejiang Provincial Natural Science Foundation (No.LQ17F010001), Key Projects of Zhejiang Science and Technology Plan (No.2018C01084)



1 引言

物联网 (internet of things, IoT) 是目前信息技术领域非常热门的研究内容, 它可以让不同网络中不同类型的设备相互通信。预计到 2020 年全球工业物联网产值将达到 1 510 亿美元, 有望引领信息产业革命的第三次浪潮^[1-3]。世界上很多国家都把物联网技术的研究和发展提高到了国家战略高度, 我国也已经把物联网产业作为战略性新兴产业并发布了《中国制造 2025》等政策文件, 提出大力发展智能制造, 加速推动信息技术与制造技术的深度融合, 并进一步突出 2025 年、2035 年和 21 世纪中叶“三步走”目标^[3-5]。作为物联网的支撑技术之一, 无线传感器网络 (wireless sensor networks, WSN) 主要负责收集并传输网络数据等工作^[6]。WSN 由大量廉价且能量有限的传感器组成, 这些微小的传感器不断地采集周围环境的数据, 并将这些数据发送给数据中心/基站做进一步的分析和处理。对于很多物联网的应用 (例如森林火灾监测网络、战场环境感知与监测网络、工业物联网等), 如何安全、及时、高效地采集并传输数据给数据中心/基站是一个很重要的问题。

工业物联网一般是在复杂系统的关键单元上安装各种类型的传感器, 传感器通过无线通信的方式实现互联互通, 即 WSN。基于 WSN, 传感器可以将数据实时地发送给数据中心, 从而实现对设备运行情况的实时监测与控制^[7]。目前, 数据采集与监视控制系统 (supervisory control and data acquisition, SCADA) 被广泛地用在石油化工、燃气、电力和冶金等领域^[8-9]。SCADA 通过对传感器上的数据进行采集处理, 从而获得系统的运行状态, 实现对复杂系统的控制。随着工业系统的集成度不断提高, 网络内需要被监测的设备不断增多, 形成的工业物联网结构规模日益复杂。因此, 工业物联网面临的挑战也越来越多, 例如需要统一的标准、数据的存储与处理问题、功耗

问题和安全隐私问题等^[10-11]。作为工业物联网中至关重要的一步, 研究如何安全高效地采集与传输工业数据显得尤为迫切。

当前我国工业数据采集与传输技术的研究仍处于起步阶段, 面临着诸多问题。首先, 基础设施建设不足。传感器部署范围窄、数量少, 采集的工业数据量少、精度低而且效率低下。其次, 工业设备的联网率低。据中国工控网的数据显示, 截至 2017 年年底, 我国工业设备联网率仅为 10%, 而德国等发达国家的联网率已经达到至少 25%。再次, 在数据采集领域的协议标准繁多且互不兼容^[12]。大量的设备和系统开放程度不高, 没有接口对外开放, 也无文档说明, 导致大量工业数据无法和互联网互通, 无法满足全流程监控等生产需求。最后, 工业数据安全性低, 大量的设备没有防火墙, 并且在传输过程中也不对数据进行加密等处理。工业数据涉及用户的隐私与商业机密, 在采集与传输过程中存在一定的安全隐患, 容易遭到网络攻击导致数据被窃取等安全事件发生。

本文重点考虑 WSN 中数据采集与传输的能耗、时延与安全隐私问题。传统的无线传感器网络一般是通过利用数据的相关性, 将数据在网络内压缩后再发送给基站^[13-14]。考虑到无线传感器网络中的数据在时间和空间上都有很强的相关性这一特点, 参考文献 [15] 将压缩感知技术 (compressive sensing, CS), 应用于数据收集过程, 提出了压缩数据收集方法 (compressive data gathering, CDG)。和传统的数据收集方法不同, 压缩数据收集方法是收集原始数据的线性组合而非原始数据本身, 基站一旦收集足够多的线性组合就可以恢复出原始数据。压缩数据收集方法不仅可以平衡网络负载还能降低网络能耗。然而, 参考文献 [16] 证明了单纯地应用压缩感知技术并不一定能提高网络性能。因此, 他们提出了一种基于混合压缩感知技术的数据收集方法。参考文献 [17] 同时考虑了压缩数据收集方法的能耗和数

据收集时间,并将问题建模成一个联合最优化问题。首先证明了这个联合最优化问题是 NP 难问题。然后,在此基础上提出了一种 2-近似解法。该解法将原问题分解成前向树构造子问题和链接时序安排子问题,通过分别求解这两个子问题来求解原问题。参考文献[18]从网络容量和时延两方面研究了压缩数据收集方法,并提出了一种基于规则网络拓扑结构的压缩数据收集方法,该方法要求网络的拓扑结构必须是规则的。但是,实际应用中的网络拓扑结构往往是随机的。网络拓扑的随机性这一特点也会造成网络中数据的稀疏度不容易确定。考虑到这个因素,参考文献[19]提出了一种基于树变换的分群压缩数据收集方法 T-CDDA。该方法综合考虑了原始数据稀疏变换的基和基于此变换的分群数据收集路由。参考文献[20]考虑了数据收集的安全问题,并且提出了一种基于云和压缩感知的数据收集方法。该方法同时考虑了传感器、云和用户 3 个方面的安全问题。参考文献[13]也考虑了数据收集的安全问题,并且将同态加密算法应用于传统的压缩数据收集方法,该方法在数据收集效率上还有待提高。

本文将压缩感知技术和 El Gamal 公钥加密技术相结合,提出了一种基于 El Gamal 加密算法的稀疏压缩数据收集方法(El Gamal based sparse compressive data gathering, ESCDG)。传统的压缩数据收集方法是基于稠密的感知矩阵的。由于感知矩阵中的元素都是非零,网络内所有节点都参与测量值的收集。ESCDG 是一种基于稀疏随机矩阵的压缩数据收集方法,由于感知矩阵中存在一定比例的零元素,只有对测量值有贡献的节点才会参与数据收集。由于活跃的节点数量减少了,收集测量值所花费的能量和时间也会减少。理论分析和数值实验都表明,无论是能耗还是数据收集时间,ESCDG 都比传统的压缩数据收集方法表现好。同时,为了提高无线传感器网络收集数据的安全性,还应用了 El Gamal 公钥加密算法,这

使得 ESCDG 能够同时抵抗多项式算力的内部攻击和外部攻击。ESCDG 可以应用于工业现场设备的数据采集和工厂外智能产品/装备的数据采集。

2 预备知识

本节简要介绍压缩数据收集方法和 El Gamal 公钥加密方案。加粗的字体表示向量或者矩阵,小写字母 a_{ij} 和 x_i 分别表示矩阵 A 中第 i 行第 j 列元素和向量 x 中第 i 个元素。

2.1 压缩数据收集方法

如果向量 x 可以表示为 $x=Ps$ 且 s 中非零元素的个数最多为 k ,则称 x 是 k -稀疏的。考虑线性变换 $y=Ax$ 。其中, $A \in R^{m \times n}$ 是感知矩阵, $x \in R^n$ 是 k -稀疏的待恢复信号, $y \in R^m$ 是测量值向量。

给定 y 和 A ,存在某个常数 c_1 ,使得当 $m > c_1 k \ln \frac{n}{k}$

时,高斯随机矩阵能以很高的概率恢复原始信号 x ^[21]。参考文献[15]首先提出了 CDG 方法,该方法可以降低能耗和平衡网络负载。但是,传统的 CDG 方法一般采用稠密的感知矩阵。本文介绍一类适用于 CDG 方法的稀疏随机感知矩阵,基于稀疏随机感知矩阵的 CDG 方法可以进一步提高数据收集的效率。

定理 1 (Li 等人^[22]) 假设原始信号 x 是 k -稀疏的,并且 x 中的每个元素都非负。令 $a_{ij}=b_{ij}c_{ij}$,其中, b_{ij} 是从单位 α -稳定极大偏态分布中独立抽样而来的, c_{ij} 服从伯努利分布 $Ber(s)$ 。已知 A 、 y 和 $\varepsilon > 0$ 。如果 $m = \frac{k \ln(n/\varepsilon)}{1-e^{-sk}}$,那么 x 能够至少以概率 $1-\varepsilon$ 被恢复,且恢复的误差任意小。

由定理 1 可知,如果令 $s=\Theta(k-1)$,那么 $m=k \ln n$ 个测量值就足够恢复原始信号了。注意定理 1 假设 x 中的每个元素都是非负的。这个假设适用于现实世界中的大部分信号,因为现实世界中的信号都是有限的,而有限的信号都可以转换为非负的信号^[22]。本文默认稀疏随机矩阵 A 是给定的,



并用 s 表示矩阵中每个元素非零的概率。下面基于图 1 来比较基于稀疏随机矩阵的压缩数据收集方法的优势。图 1(a)所示为一个包含 6 个节点的线性网络。图 1(b)和图 1(c)分别表示了如何用传统的压缩数据收集方法和稀疏压缩数据收集方法收集 $y_1 = \sum_{i=1}^6 x_i a_{1i}$ 。假设 a_{12} 、 a_{15} 和 a_{16} 为零, 那么 $y_1 = x_1 a_{11} + x_3 a_{13} + x_4 a_{14}$ 。在图 1(c)中, 只有 N_1 、 N_3 和 N_4 需要参与数据收集。 N_6 不需要参与数据收集, 它是一个数据转发节点。很明显, 在这个例子中, 基于稀疏随机矩阵的压缩数据收集方法的性能比传统的压缩数据收集方法好。

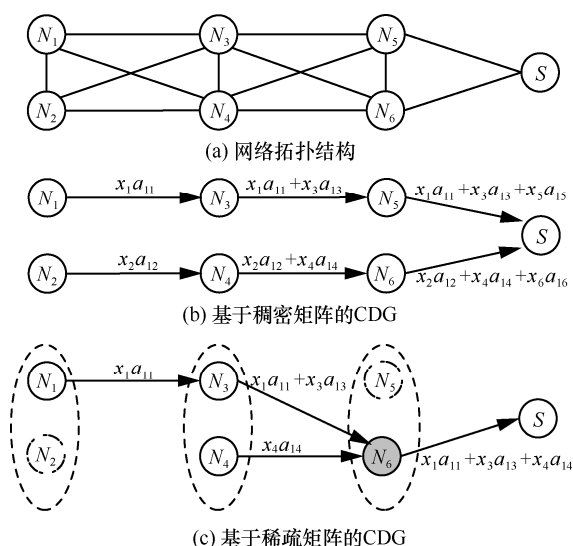


图 1 基于稠密矩阵和稀疏矩阵的压缩数据收集方法比较

2.2 El Gamal 公钥加密算法

本节介绍 El Gamal 公钥加密算法^[23]。一般地, 一个 El Gamal 加密系统由以下几个部分组成。

(1) 密钥产生系统 GEN

GEN 首先产生一个阶为 q (q 为大素数) 的循环群 G 。然后随机选择一个生成元 $g \in G$ 和随机整数 $x < q$ 。令 $h = g^x \bmod q$ 。公钥为 $PK = (q, g, h)$, 私钥为 $RK = (q, x)$ 。

(2) 加密算法 ENC

输入公钥 PK 、消息 $m \in G$ 和一个随机选择的整数 $y < q$, 加密算法 ENC 输出密文 $\langle c_1, c_2 \rangle$, 其中

$$c_1 = g^y \bmod q, \quad c_2 = h^y \bmod q.$$

(3) 解密算法 DEC

输入私钥 RK 和密文 $\langle c_1, c_2 \rangle$, 解密算法 DEC 输出密文 $m = c_2 / c_1^x \bmod q$ 。

El Gamal 加密系统是一种安全性很高的加密系统, 参考文献[23]已经证明它能够抵抗多项式攻击敌手 (计算能力为多项式时间) 的选择明文攻击和选择密文攻击。为了完整描述 El Gamal 加密系统, 还需要将明文二进制串映射到 G 中。这里介绍一种简单的映射方法: 首先, 选择一个大的强素数 r , 即 $q = (r-1)/2$ 也是一个素数。众所周知, 二次剩余模 r 的集合构成一个模 r 的乘法群 G , 其中 G 是一个阶为 q 的循环群; 然后, 把明文二进制串解释为一个整数并将这个整数加 1 得到 $w (1 \leq w \leq q)$, 这里加 1 是为了排除全零的可能性; 最后, 通过函数 $m = w^2 \bmod r$ 将 w 映射到 m 。

下面介绍 El Gamal 加密系统的一个特性: 同态性。该特性可将明文空间的运算一一映射到密文空间。

定义 1 假设 M 和 C 分别为明文空间和密文空间。公钥加密系统 (GEN, ENC, DEC) 是同态的, 如果它满足以下性质:

- 基于 M 和所有由 GEN 产生的公钥, ENC 输出的密文都属于 C ;
- 对满足 $m_1 = \text{DEC}(c_1)$ 和 $m_2 = \text{DEC}(c_2)$ 的任意 $m_1, m_2 \in M$ 和 $c_1, c_2 \in C$, 有: $\text{DEC}(c_1 \cdot c_2) = m_1 + m_2$, 其中, “ $\cdot$ ” 和 “ $+$ ” 分别是群 C 和 M 中的运算。

根据定义可知 El Gamal 加密是同态加密, 其明文空间和密文空间分别是 G 和 $G \times G$ 。压缩数据收集方法的目标是收集足够多的测量值, 每个测量值是各节点原始数据的线性组合。在收集某个测量值的过程中, 每个节点只对该测量值对应的线性组合中的某一项有贡献, 它只加密自己有贡献的那一项。因此, 每个节点在接收到上一个节点传来的数据后不需要将之进行解密, 直接将接

收到的数据和自己的密文进行组合即可。

3 基于 El Gamal 加密算法的稀疏压缩数据收集方法

3.1 网络模型

本文将时间分割成等长的时隙, 并假设发送一个数据分组的时间为一个时隙 (slot)。假设节点之间的通信采用半双工模式, 即节点同一时刻只能接收或者发送数据。为了保证通信的质量, 无线传感器网络的布置往往非常密集^[24-25]。因此, 本文研究由大量节点组成的平面网络。具体地说, 节点 $N_1, N_2, \dots, N_n$ 被随机均匀地分布在一个单位面积的正方形区域内, 基站位于整个网络的右上角 (如图 2 (a) 所示)。整个网络被分成 L^2 个小区并用 C_{ij} ($1 \leq i, j \leq L$) 表示, 每个小区是一个边长为 $l = 1/L$ 的小正方形。假设每个节点的通信距离为 $\sqrt{5}l$, 每个节点都可以和相邻的小区通信。为了保证整个网络的连通性, 每个小区的边长必须满足 $l \geq \Theta \sqrt{\frac{\ln n}{n}}$ ^[26]。本文假设 $l = \Theta \sqrt{\frac{\ln n}{n}}$ 。图 2(b) 显示了整个网络的数据流向。定义路径 P_i ($1 \leq i \leq L$) 为横向相连的小区 C_{ij} ($1 \leq j \leq L$) 的集合。

假设攻击者拥有足够多的资源 (包括多项式时间的计算能力、无限的存储空间和通信带宽等)

来实施攻击, 这些攻击者可以通过两种方式来攻击网络: 外部攻击和内部攻击。外部攻击者能够窃听整个网络的数据传输, 分析和比较整个网络的数据流动, 从而得到网络中数据流动的方向等信息。内部攻击者能够攻破网络中一个或者多个节点, 从而获取到这些被攻破节点中存储的所有数据。如果这些被攻破的节点存有私钥信息, 那么内部攻击者就很容易通过私钥恢复原始数据。

3.2 数据收集算法 ESCDG

将传统的压缩数据收集方法称为稠密压缩数据收集 (dense compressive data gathering, DCDG)。假设网络中的节点已经获取了原始数据 $x = (x_1, x_2, \dots, x_n)^T$ 。基站想要收集测量值向 $y = (y_1, y_2, \dots, y_m)^T$ 。定义收集 y_i ($1 \leq i \leq m$) 的过程为第 i 轮测量值/数据收集过程。在收集某个测量值的过程中, 只有系数非零的节点需要参与, 不参与的节点将进入睡眠模式。为了保证所有测量值都能顺利接收, 让每个小区至少有一个节点参与测量值的收集。假设每个节点都知道自己的 ID 和它所在的小区 ID, 节点 N_j 可以根据自己的 ID 和随机数产生器来生成系数向量 a_j (a_j 是随机矩阵中 A 的第 j 列)。注意节点 N_j 只知道本节点的系数向量 a_j , 它并不知道其他节点的系数向量。因此, 即使某些节点被攻破了攻击者也无法获取那些没

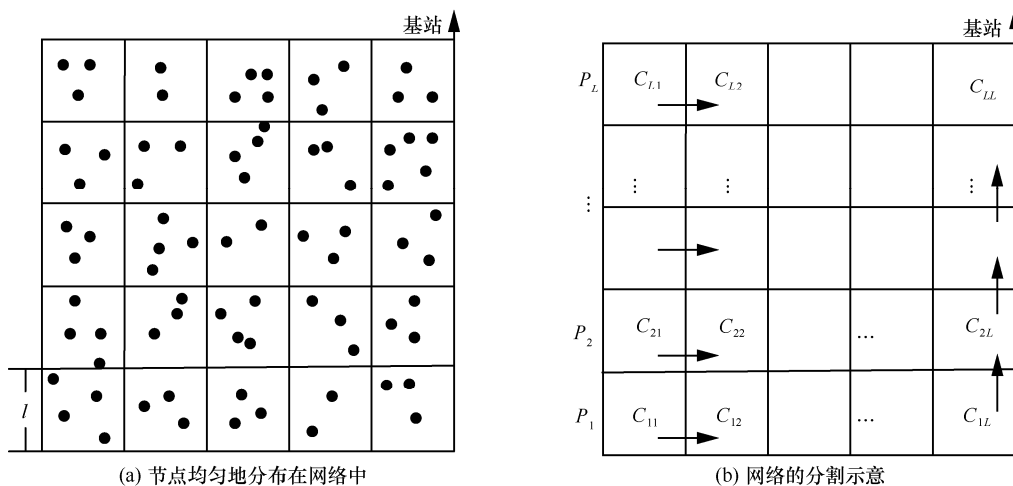


图2 单位面积的正方形网络



有被攻破的节点所产生的系数向量,从而保证感知矩阵 A 不会被攻击者窃取。为了区分每一轮测量值收集过程中节点的不同状态,本文将节点状态分为 3 类: O、I 和 R。状态 O 表示该节点不需要参与测量值收集。状态 I 表示该节点参与测量值的收集,但是并不接收上一个小区发来的数据。状态 R 表示该节点不仅需要参与测量值的收集,它还负责接收上一个小区发来的数据。不失一般性,考虑第 i 轮测量值收集过程。如果系数 $a_{ij}=0$ (或者 $a_{ij} \neq 0$), 就让节点 N_j ($j=1, 2, \dots, n$) 进入状态 O (或者 I)。网络中的小区可能出现如下情况。

- 小区内所有的节点都进入了状态 O, 这意味着该小区内所有节点都不需要参与本轮测量值的收集。但为了保证网络的连通性, 随机选择一个节点进入状态 R, 该节点就作为一个中继节点, 它只负责转发数据。
- 小区内至少有一个节点进入了状态 I。在此情况下, 从所有进入状态 I 的节点中随机选择一个节点进入状态 R, 该节点负责汇聚并转发数据。

综上所述, 只有状态为 I 或者 R 的节点处于活跃状态。

为了更好地理解, 下面将以图 1 所示为例进行说明。假设 c_j^i 为节点 N_j 在第 i 轮测量值收集过程中产生的密文, 即 $c_j^i = \text{ENC}(x_j a_{ij})$ 。把 N_1 和 N_2 、 N_3 和 N_4 、 N_5 和 N_6 分别布置在同一个小区内。在第 1 轮测量值收集过程中, 节点 N_2 和 N_5 的状态为 O, N_4 的状态为 I, N_1 、 N_3 和 N_6 的状态为 R。 N_1 首先发送密文 c_1^1 给 N_3 。然后, N_3 将自己的密文和收到的数据相乘得到 $c_1^1 c_3^1$ 。接着 N_3 和 N_4 分别将密文 $c_1^1 c_3^1$ 和 c_4^1 发送给 N_6 。 N_6 将接收到的数据相乘得到 $c_1^1 c_3^1 c_4^1$ 并转发给基站。基站在收到密文 $c_1^1 c_3^1 c_4^1$ 后, 用私钥和解密算法得到明文 $\text{DEC}(c_1^1 c_3^1 c_4^1) = x_1 a_{11} + x_3 a_{13} + x_4 a_{14} = y_1$ 。按照类似步骤, 基站可以收集到所有的测量值。最后,

基于 y 和 A , 基站应用压缩感知恢复算法得到原始数据 x 。

下面介绍 ESCDG 是如何收集图 2 中各节点数据的。其核心思想是节点一旦有“机会”便发送数据。定义 V_{ij}^r 和 T_{ij}^r 分别为在第 r 轮测量值收集过程中, 小区 C_{ij} 内活跃节点的数量和开始发送数据的时间。注意活跃节点即状态为 I 和 R 的节点。初始化 $T_{11}^1 = 0$ 。假设基站已经将数据收集的请求和公钥发送给所有节点。节点 N_j ($1 \leq j \leq n$) 基于自己的 ID 生成 a_j , 并与原始数据 x_j 相乘得到 $x_j a_j$ 。然后, 加密 $x_j a_j$ 得到密文 c_j^r ($1 \leq r \leq m$)。在之后的论述中, 本文所指的数据都是加密后的密文数据, 数据相乘指的是密文空间上的乘法运算。整个数据收集分为两个过程: 第一个过程为每条路径上的小区将数据发送给路径上最后一个小小区; 第二个过程是所有路径上最后一个小小区将数据发送给基站。在第一个过程中, 数据流动是横向链条式的, 即路径 P_i ($1 \leq i \leq L$) 上所有小区将数据发送给小区 C_{il} 。该过程具体包含以下几个步骤。

步骤 1 C_{11} 内状态为 I 的节点将第 1 轮数据发送给状态为 R 的节点。该节点将所有接收到的数据和自己的数据相乘并将结果发送给 C_{12} 内状态为 R 的节点。节点发送完数据后进入睡眠模式。每个小区在发送数据时都是状态为 I 的节点先将数据发送给状态为 R 的节点。然后, 该节点再将数据发送给下一个小区内状态为 R 的节点。在第 1 轮测量值收集过程中, 路径 P_1 上所有的小区都重复类似过程。因此, $T_{1j}^1 = \sum_{i=1}^{j-1} V_{1i}^1$ 。对于最后一个小小区, 随机选择一个节点, 并使它在每一轮测量值收集过程中的状态都为 R。每条路径上最后一个小小区内的其他节点将数据发送给本小区内状态为 R 的节点。

步骤 2 当 C_{15} 结束第 1 轮的数据发送后, C_{21} 马上开始发送第 1 轮数据。在第 1 轮数据收集过程中, C_{ij} 开始发送数据的时间都是在 $C_{(i-1)(j+4)}$ 结

束后, 即:

$$\begin{aligned} T_{ij}^1 &= T_{(i-1)(j+4)}^1 + V_{(i-1)(j+4)}^1, \\ 2 \leq i \leq L, 1 \leq j \leq L-4 \end{aligned} \quad (1)$$

在每轮测量值收集过程中, 所有路径上最后 4 个小区开始发送数据的时间都是在前一个小区发送结束后, 即:

$$T_{ij}^r = T_{i(j-1)}^r + V_{i(j-1)}^r, L-3 \leq j \leq L \quad (2)$$

步骤 3 当 C_{43} 完成第 1 轮数据发送后, C_{11} 马上开始第 2 轮。之后的每一轮, C_{11} 都在 C_{43} 完成前一轮数据发送后开始, 即 $T_{11}^r = T_{43}^{r-1} + V_{43}^{r-1}$, $r \geq 2$ 。

步骤 4 接收到 C_{11} 的第 2 轮数据后, C_{12} 先判断 C_{26} 是否完成了第 1 轮。如果 C_{26} 已经完成第 1 轮, C_{12} 马上开始发送第 2 轮数据; 如果 C_{26} 还未完成第 1 轮, C_{12} 等 C_{26} 完成后再发送第 2 轮数据。对于之后的每一轮, 路径 P_1 上除 C_{11} 外的所有小区都重复类似操作, 即:

$$T_{1j}^r = \max \{ T_{1(j-1)}^r + V_{1(j-1)}^r, T_{(i-1)(j+1)}^r + V_{(i-1)(j+1)}^r \}, \quad (3)$$

$$i \geq 2, 2 \leq j \leq L-4$$

步骤 5 当 C_{15} 完成第 1 轮数据发送后 C_{12} 开始发送第 2 轮数据。对于之后的每一轮。除了 P_1 外的所有路径上的第一个小区都在前一条路径第 5 个小区完成数据发送后再开始发送数据, 即:

$$T_{i1}^r = T_{(i-1)5}^r + V_{(i-1)5}^r, i, r \geq 2 \quad (4)$$

在每一轮测量值收集过程中, 如果 $i > j$, 那么 P_i 开始发送数据的时间比 P_j 晚。

步骤 6 当收到 C_{21} 发来的第 2 轮数据后, C_{22} 先等 C_{16} 完成第 2 轮。确认 C_{16} 已经完成第 2 轮后, C_{22} 再开始发送第 2 轮数据。之后的每一轮都重复类似操作, 即:

$$T_{ij}^r = \max \{ T_{i(j-1)}^r + V_{i(j-1)}^r, T_{(i-1)(j+4)}^r + V_{(i-1)(j+4)}^r \}, \quad (5)$$

$$i \geq 2, 2 \leq j \leq L-4$$

过程 1 结束后, C_{iL} ($1 \leq i \leq L$) 内某个节点 (即一直处于状态 R 的节点) 存储了 P_i 上所有小

区的数据。第二个过程就是这些节点组成一条路径将数据发送给基站。注意这些节点在每一轮数据收集过程中都要发送数据。每轮数据收集都由 C_{1L} 开始, 并且开始时间是在前一轮开始后的第 5 个时隙。路径上其他小区在接收到前一个小区发来的数据后即可开始发送数据。

4 理论分析

4.1 能量消耗

假设 E_c 和 E_p 分别为 DCDG 和 ESCDG 的能量消耗。因为节点大部分能耗来自于数据通信^[27], 本节只考虑节点通信所消耗的能量。换句话说, 能量消耗和数据发送次数成正比。在 DCDG 中, 每个节点在每一轮测量值收集过程中都要发送数据, 因此 $E_c = mn$ 。接下来考虑 ESCDG 的能耗。先考虑某轮测量值收集的能耗。假设小区内总的节点数量为 U , 活跃节点的数量为 V 。

引理 1 令 $L = \sqrt{\frac{n}{\ln n}}$ 。当 $n \rightarrow \infty$ 时, 小区内节点的数量以概率 1 趋向于 $U \in [1, e \ln n]$ 。

证明 已知任意小区为空的概率是 $(1-1/L^2)^n$ 。根据布尔不等式, 网络中至少有一个小区为空的概率最多是 $p_0 = L^2(1-1/L^2)^n$ 。由 $1-x \leq e^{-x}$ 可以得到:

$$p_0 \leq \frac{ne^{-\ln n}}{\ln n} = \ln^{-1} n \quad (6)$$

所以, 当 $n \rightarrow \infty$ 时, 至少有一个小区为空的概率趋向于 0。另一方面, U 服从二项分布 $B(n, 1/L^2)$ 。由切诺夫界可以得到, 对于所有 $\theta > 0$:

$$\Pr(U > e \ln n) \leq \frac{E[\exp(\theta U)]}{\exp(e \theta \ln n)} \quad (7)$$

令 $\theta=1$, 得到 $\Pr(U > e \ln n) \leq n^{-1}$ 。证毕。

当 $U=u$ 时, 定义 $p_i(u)$ 为小区内产生非零系数的节点数量是 i ($i=0, 1, \dots, u$) 的概率。可以得到:



$$V = \begin{cases} 1, & \text{以概率 } p_0 \\ i, & \text{以概率 } p_i(u) \end{cases} \quad (8)$$

其中, $p_i(u) = \binom{u}{i} s^i (1-s)^{u-i}$ (s 是感知矩阵中元素为非零的概率)。小区内活跃节点数量的期望为:

$$\begin{aligned} E[V] &= E[E[V|U]] = E[(1-s)^U + Us] = \sum_{j=0}^n jsq_j + \\ &\sum_{j=0}^n (1-s)^j q_j = \frac{sn}{L^2} + \sum_{j=0}^n \binom{n}{j} \left(\frac{1-s}{L^2}\right)^j \left(1 - \frac{1}{L^2}\right)^{n-j} = \quad (9) \\ &slbn + \left(1 - \frac{slbn}{n}\right)^n \end{aligned}$$

其中, $q_j = \binom{n}{j} \left(\frac{1}{L^2}\right)^j \left(1 - \frac{1}{L^2}\right)^{n-j}$ 。当网络中节点很多时, ESCDG 的能耗只有 DCDG 的 s 倍。

4.2 数据收集时间

本节证明当信号和感知矩阵都足够稀疏时, ESCDG 的数据收集时间从阶数角度来说是最优的。

引理 2 当网络规模很大时, 任何数据收集方法的数据收集时间以很大概率大于 $\Theta = \sqrt{\frac{n}{lbn}}$ 。

证明 考虑离基站最远的小区 (即 C_{11}) 发送完数据所需要的时间。已知 C_{11} 的节点数量 U 服从二项分布 $B(n, l^2)$, 其中 $l = \sqrt{\frac{lbn}{n}}$ 。由切比雪夫不等式可知, 存在一个常数 $\alpha \in (0, 0.5)$ 使得式 (10) 成立:

$$\Pr(|U - nl^2| \geq l\sqrt{n}(lbn)^\alpha) \leq \frac{nl^2(1-l^2)}{nl^2(lbn)^{2\alpha}} \leq (lbn)^{-2\alpha} \quad (10)$$

当 $n \rightarrow \infty$ 时, $(lbn)^{-2\alpha} \rightarrow 0$ 。变量 U 的偏差和中心的比值为:

$$\frac{l\sqrt{n}(lbn)^\alpha}{nl^2} = \frac{lbn}{(lbn)^{0.5+\alpha}} = (lbn)^{\alpha-0.5} \quad (11)$$

趋向于 0。因此, 当 n 很大时, U 的取值大概

率是 $\Theta(nl^2)$ 。因为 C_{11} 内节点不能同时发送数据, 且每个节点都要发送数据。所以, C_{11} 至少需要 $\Theta(nl^2)$ 个时隙来完成发送。又因为通信距离是 $\sqrt{5}l$, C_{11} 的数据至少需要 $\frac{\sqrt{2}}{\sqrt{5}l}$ 跳才能到达基站。

所以, C_{11} 的数据到达基站至少需要 $\Theta(nl^2 + l^{-1}) = \Theta\left(\sqrt{\frac{n}{lbn}}\right)$ 个时隙。证毕。

下面计算 ESCDG 的数据收集时间。假设某个数据方法 H 也把数据收集过程分为两个部分。其中, 过程二和 ESCDG 一样, 过程一给每个小区分配固定的时间 $h = (s+c) \ln n + \left(1 - \frac{s \ln n}{n}\right)^n$,

$c > 0$ 是常数。现在计算 H 的数据收集时间。在过程一中, 每条路径发送完一轮数据需要 Lh 个时隙, 而且 P_{i+1} 每轮开始的时间都比 P_i 晚 $5h$ 。因此, 收集每轮测量值的时间为 $6Lh$ 。又因为 C_{11} 每一轮开始的时间都比前一轮晚 $18h$, 所以, 过程一所需要的时间是 $(6Lh + 18mh)$ 。过程二将 L 个节点组成链条式网络, 每个节点发送 m 轮数据, 每轮开始时间都比前一轮晚 5 个时隙。因此, 过程二需要 $(L + 5m)$ 个时隙。综上所述, 方法 H 的数据收集时间是 $6Lh + 18mh + L + 5m = \Theta(Lh + mh)$ 。

定理 2 假设 $k = O\sqrt{\frac{n}{lbn^3}}$, $s = O(lb^{-1}n)$ 。当网络规模很大时, ESCDG 的数据收集时间大概率不会超过 $O\sqrt{\frac{n}{lbn}}$, 并且此时间从阶数的角度来说是最优的。

证明 因为当 $k = O\sqrt{\frac{n}{lbn^3}}$, $s = O(lb^{-1}n)$ 时, 方法 H 的数据收集时间是 $O\sqrt{\frac{n}{lbn}}$ 。所以, 只需要证明在 ESCDG 中, 每个小区在每轮测量值收集过程中发送数据的时间最多为 h 即可。已知

$$E[V] = slbn + \left(1 - \frac{slbn}{n}\right)^n, \text{ 可以得到:}$$

$$\begin{aligned}
E[V^2] &= E[E[V^2 | U]] = E[(1-s)U + (1-2E[V]) \cdot \\
(1-S)^U] &= s(1-s)lbn + (1-2E[V]) \cdot \\
\left(1 - \frac{slbn}{n}\right)^n &\leq s(1-s)lbn
\end{aligned} \quad (12)$$

由切比雪夫不等式可知：

$$\begin{aligned}
\Pr(|V - E[V]| \geq clbn) &\leq \frac{\text{Var}[V]}{(clbn)^2} \leq \\
\frac{s}{c^2 lbn} &= O(1 / lb^2 n)
\end{aligned} \quad (13)$$

当 $n \rightarrow \infty$ 时，每个小区活跃节点数 $V < h$ 的概率趋向于 1。证明完毕。

4.3 安全性分析

本节介绍 ESCDG 在外部攻击和内部攻击下的表现。常见的外部攻击有数据分组计数攻击 (packet counting attack) 和计时攻击 (timing attack) [28]。数据分组计数攻击是通过窃听非标准 (异常) 数量的数据分组来获得网络信息。在 ESCDG 中，每个节点在每一轮测量值收集过程中发送的数据分组数量基本一致。因此，ESCDG 可以抵御此类攻击。对于计时攻击，攻击者需要记录数据分组在网络中传输的时延以及到达和离开网络的时刻。在 ESCDG 中，活跃节点会对传输的数据分组进行编辑和修改 (把接收到的密文和自己的密文相乘)。换句话说，相同的数据分组在网络中存在的时间很短，计时攻击很难记录某个数据分组在网络中传输的时延以及离开网络的时刻。因此，ESCDG 能够抵御计时攻击。活跃节点会对数据分组进行编辑和修改这个特性使得 ESCDG 能够抵御任何需要追溯消息源的外部攻击。

对于内部攻击，攻击者往往先是攻破网络中的一部分节点，然后对网络内的数据流进行分析 (例如分析数据分组尺寸或者之间的相关性)，从而发动内部攻击。在 ESCDG 中，网络内节点接收到的数据是前面所有节点密文相乘后的结果，被攻破的节点对其他节点的信息一无所知，攻击

者无法通过分析数据分组之间间的相关性而获得信息。因此，ESCDG 也能抵御内部攻击。

5 数值实验

本节将从能量消耗和数据收集时间两个方面对 ESCDG 进行数值实验验证。考虑到比较的公平性，本节假设 DCDG 所使用的加密方法也是 El Gamal 公钥加密方案。数值实验采用实际测量的能耗数据 [29]：加密一次明文消耗能量 675.5 mJ，一次密文聚合计算消耗能量 5.04 μ J，节点发送一次加密后的数据分组消耗能量 460.8 μ J。假设整个感知区域的边长为 1 km，每个小区的边长为 $l = \sqrt{\frac{\log n}{n}}$ km，其中 n 是网络内节点的数量。因为每个节点的通信距离为 $\sqrt{5}l$ ，所以相邻小区内的节点都可以相互通信。节点发送一个数据分组需要花费 1 个时隙的时间。为了平滑随机性带来的影响，实验中的每个结果都是 100 次实验的平均值。下面说明数据传输时网络的拓扑结构变化，假设 100 个节点均匀地散布在网络中，感知矩阵中每个元素非零的概率设置为 $s=0.5$ 。图 3 显示了在收集某个测量值的过程中，网络内各节点所处的状态以及网络拓扑结构 (“+” “x” 和 “o” 分别表示状态为 R、I 和 O 的节点)。每个小区内状态为 I 或者 R 的节点将数据发送给下一个小区中状态为 R 的节点。由图 3 可知，网络中大约只有一半的节点需要发送数据。图 4 显示了整个网络的能耗随着收集的测量值数量的变换情况。在图 4 中，假设节点数量为 $n=1\ 000$ ，数据的稀疏度 $k=10$ ，感知矩阵中元素非零概率 $s=0.5$ 。由图 4 可知，ESCDG 的网络能耗低于 DCDG。随着收集的测量值数量增多，ESCDG 和 DCDG 之间网络能耗的差值越来越大，这个说明了 ESCDG 方法非常适合于大型网络。

图 5 显示了 ESCDG 和 DCDG 的能耗比值随着节点数量的增加而变化的情况。假设原始信号



的稀疏度 $k = 10$ ，并且考虑了 $s = 0.3$ 、 $s = 0.5$ 、 $s = 0.7$ 共 3 种情况下的能耗比值。从图 5 可知，随着网络节点数量的增加，ESCDG 相对于 DCDG 的能耗比值逐渐减少，并且这一比值逐渐趋向于 s 。这是因为在 ESCDG 的每一轮数据收集过程中，有

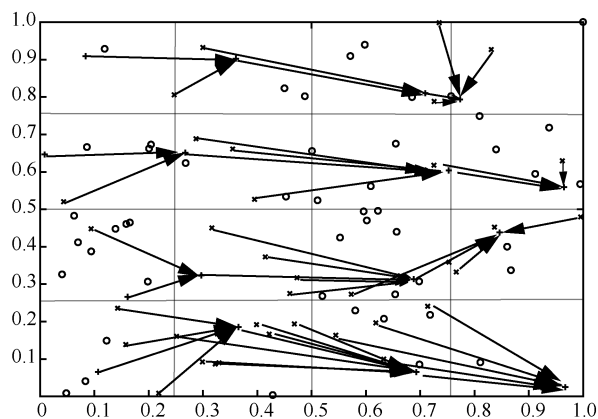


图 3 网络内各节点所处状态示意

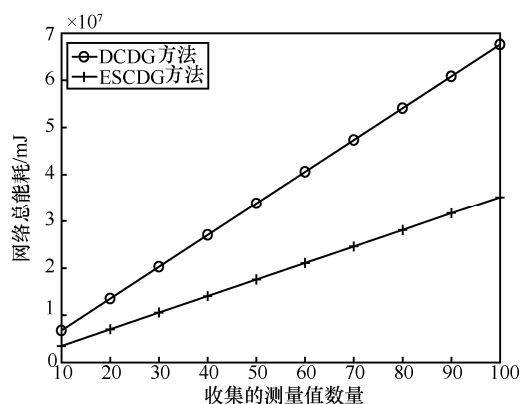


图 4 网络能耗随收集的测量值数量的变化情况

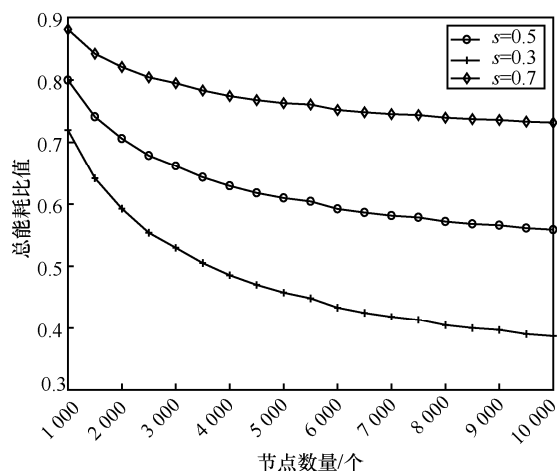


图 5 ESCDG 和 DCDG 的能耗比值随网络节点数量变化的情况

两类节点是活跃的：一类是对测量值有贡献的节点；另一类是虽然对测量值没有贡献，但为了保持网络连通而活跃的节点，随着节点数量的增加，网络变得越来越稠密，第一类节点的数量比例一直大约是 s 。然而，第二类节点的数量却越来越少，直到趋向于 0。

图 6 比较了在不同的稀疏度 s 情况下，ESCDG 和 DCDG 的能耗比值变化情况。在图 6 中，假设原始信号的稀疏度 $k = 10$ ，并且考虑了 $n = 1\,000$ 、 $n = 5\,000$ 、 $n = 10\,000$ 共 3 种情况下的能耗比值。可以得到，在节点数量不变的前提下，随着 s 的逐渐增大，ESCDG 和 DCDG 的能耗比值也逐渐增大，这个观察结果和图 5 的结论是一致的。

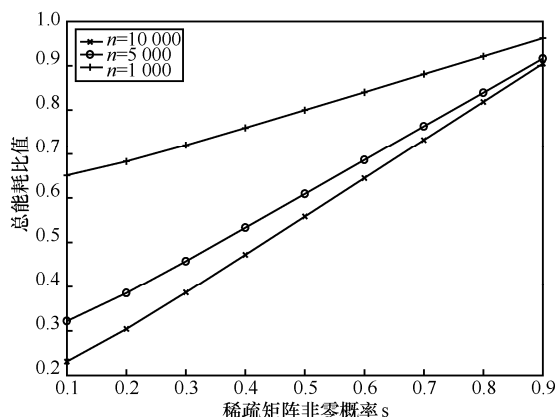


图 6 ESCDG 和 DCDG 的能耗比值随矩阵稀疏度的变化情况

下面比较一下 ESCDG 和 DCDG 在数据收集时间方面的性能。图 7 比较了 ESCDG 和 DCDG 的数据收集延迟（单位为时隙）随着节点数量的变化情况。根据上文理论分析结果，将图 7 中的参数设为 $s = \text{lb}^{-1}n$ ， $k = \sqrt{\frac{n}{\text{lb}^3 n}}$ 。由图 7 可以观察到，随着节点数量的增加，ESCDG 和 DCDG 的数据收集时间都是逐渐增加的。这个观察结果和引理 2 是一致的。因为任何方法的数据收集时间的阶数都大于 $\Theta(\sqrt{\frac{n}{\text{lb} n}})$ 。还可以观察到，对于相同的网络节点数量，ESCDG 的数据收集时间比 DCDG 的数据收集时间少。而且，随着节点数量的增加，ESCDG 和 DCDG 的数据收集时间之差越来越大。这说明对

于节点很多的无线传感器网络, ESCDG 对于网络性能的优化远远好于 DCDG。

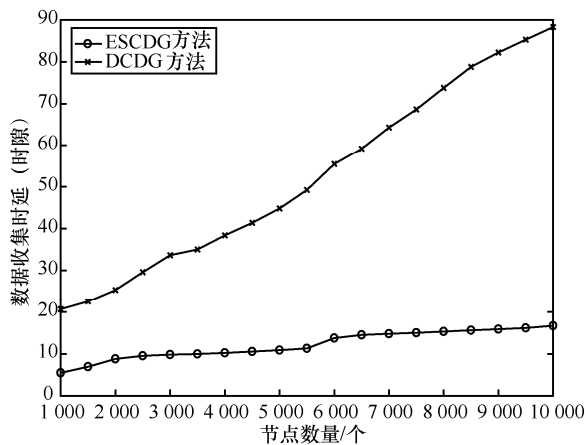


图7 ESCDG和DCDG的数据收集时间随网络节点数量的变化情况

6 结束语

本文介绍了一种基于稀疏随机矩阵和 El Gamal 公钥加密方案的压缩数据收集方法 ESCDG。该方法具有低能耗、低数据收集时延和高安全性的特点。在传统的无线传感器网络数据收集方法中, 离基站较近的节点需要转发较远的节点的数据, 这会造成整个网络的负载不均衡, 离基站较近的节点能耗速度快于离基站较远的节点。基于压缩感知技术的数据收集方法 CDG 很好地解决了这一问题。在 CDG 中, 网络内每个节点收发数据分组的次数大体是相同的, 这是由于每个节点发送的数据并不是原始数据, 而是原始数据的不同线性组合, 即测量值。传统的 CDG 方法是基于稠密的随机矩阵的, 每个节点都会参与每轮的测量值收集, 这在很大程度上限制了 CDG 的性能。本文提出的 ESCDG 通过利用感知矩阵的稀疏性, 进一步提高了 CDG 的性能。同时, 考虑到无线传感器网络一般是布置在公开环境中的, 为了提高数据收集的安全性, 本文还将 El Gamal 公钥加密算法应用到数据收集过程中, 这使得 ESCDG 能够同时抵抗常见的内部攻击和外部攻击。

参考文献:

- [1] GUBBI J, BUYYA R, MARUSIC S, et al. Internet of things (IoT): a vision, architectural elements, and future directions[J]. Future Generation Computer Systems, 2013, 29(7): 1645-1660.
- [2] 朱洪波, 杨龙祥, 于全. 物联网的技术思想与应用策略研究[J]. 通信学报, 2010, 31(11): 1-9.
ZHU H B, YANG L X, YU Q. Investigation of technical thought and application strategy for the Internet of Things[J]. Chinese Journal of Communications, 2010, 31(11): 1-9.
- [3] 张建雄, 吴晓丽, 杨震, 等. 基于工业物联网的工业数据采集技术研究与应用[J]. 电信科学, 2018, 34(10): 124-129.
ZHANG J X, WU X L, YANG Z, et al. Research and application of industrial data acquisition based on industrial internet of things[J]. Telecommunications Science, 2018, 34(10): 124-129.
- [4] 宁焕生, 徐群玉. 全球物联网发展及中国物联网建设若干思考[J]. 电子学报, 2010, 38(11): 2590-2599.
NING H S, XU Q Y. Research on global Internet of Things: developments and it's construction in China[J]. Chinese Journal of Electronics, 2010, 38(11): 2590-2599.
- [5] 国务院. 关于深化“互联网+先进制造业”发展工业互联网的指导意见[R]. 2017.
The State Council. Guidance on deepening the development of industrial internet in the “internet + advanced manufacturing industry”[R]. 2017.
- [6] PERERA C, ZASLAVSKY A, CHRISTEN P, et al. Context aware computing for the internet of things: a survey[J]. IEEE Communications Surveys & Tutorials, 2014(16): 414-454.
- [7] 李锐君, 周伟, 侯维岩. 面向先进制造环境的工业物联网关键技术研究[J]. 仪表技术, 2017(3): 4-7.
LI R J, ZHOU W, HOU W Y. Research on the key technologies of the Industrial Internet of Things for the advanced manufacturing environment[J]. Instrumental Technology, 2017(3): 4-7.
- [8] 湛锋, 秦冠军, 张长开, 等. 若干 SCADA 监控系统的设计比较[J]. 工业控制计算机, 2018, 31(2): 1-5.
ZHAN F, QIN G J, ZHANG C K, et al. Design comparison of several industrial monitoring SCADA systems[J]. Industrial Control Computer, 2018, 31(2): 1-5.
- [9] 陈海霞, 任庆海, 冯俊杰. SCADA 系统在污水泵站中的应用[J]. 电子测试, 2018(1): 93-95.
CHEN H X, REN Q H, FENG J J. Application of SCADA system in sewage pumping station[J]. Electronic Test, 2018(1): 93-95.
- [10] 朱佳佳, 赵妍. 物联网的应用与挑战[J]. 电信网技术, 2017(5): 26-30.
ZHU J J, ZHAO Y. Applications and challenges in Internet of Things[J]. Telecommunications Network Technology, 2017(5): 26-30.
- [11] 王清波. 物联网安全挑战与关键技术分析[J]. 无线互联科技, 2019, 16(1): 30-31.
WANG Q B. Analysis on the security challenges and key technologies of Internet of Things[J]. Wireless Internet Technology, 2019, 16(1): 30-31.

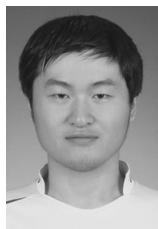


- [12] 常洁, 王艺, 李洁, 等. 工业通信网络现有架构的梳理总结和未来运营商的发展策略[J]. 电信科学, 2017, 33(11): 121-133.
- CHANG J, WANG Y, LI J, et al. Summary of existing framework in industrial communication networks and future development strategies for communication operators[J]. Telecommunications Science, 2017, 33(11): 121-133.
- [13] XIE K, WANG X, LIU X, et al. Interference-aware cooperative communication in multi-radio multi-channel wireless networks[J]. IEEE Transactions on Computers, 2016, 65(5): 1528-1542.
- [14] WU D, ARKHIPOV D I, KIM M, et al. ADDSEN: adaptive data processing and dissemination for drone swarms in urban sensing[J]. IEEE Transactions on Computers, 2017(66): 183-198.
- [15] LUO C, WW F, SUN J, et al. Compressive data gathering for large-scale wireless sensor networks[C]//The 15th Annual International Conference on Mobile Computer Networking, Sep 20-25, 2009, Beijing, China. New York: ACM Press, 2009: 145-156.
- [16] LUO J, XIANG L, ROSENBERG C. Does compressed sensing improve the throughput of wireless sensor networks [C]//The 44th IEEE International Conference on Communications, May 23-27, 2010, Cape Town, South Africa. Piscataway: IEEE Press, 2010: 1-6.
- [17] EBRAHIMI D, ASSI C. On the interaction between scheduling and compressive data gathering in wireless sensor networks[J]. IEEE Transactions on Wireless Communications, 2016(15): 2845-2858.
- [18] ZHENG H, XIAO S, WANG X, et al. Capacity and delay analysis for data gathering with compressive sensing in wireless sensor networks[J]. IEEE Transactions on Wireless Communications, 2013(12): 917-927.
- [19] ZHAO C, ZHANG W, YANG Y, et al. Treelet-based clustered compressive data aggregation for wireless sensor networks[J]. IEEE Transactions on Vehicle Technology, 2015(64): 4257-4267.
- [20] HSIEH S H, HUNG T H, LU C S, et al. A secure compressive sensing-based data gathering system via cloud assistance[J]. IEEE Access, 2018(6): 31840-31853.
- [21] CANDES E, WAKIN M. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine, 2008(25): 21-30.
- [22] LI P, ZHANG C H, ZHANG T. Sparse recovery with very sparse compressed counting[J]. arXiv preprint arXiv: 1401.0201, 2013.
- [23] KATZ J, LINDELL Y. Introduction to modern cryptography[M]. New York: CRC Press, 2007.
- [24] YU X H, BAEK S J. Energy-efficient collection of sparse data in wireless sensor networks using sparse random matrices[J]. ACM Transactions on Sensor Networks, 2017, 13(3): 22.
- [25] YU X H, BAEK S J. Joint routing and scheduling for data collection with compressive sensing to achieve order-optimal latency[J]. International Journal of Distributed Sensor Networks, 2017, 13(10): 1550147717737968.
- [26] GUPTA P, KUMAR P R. Critical power for asymptotic connectivity in wireless networks[M]//Stochastic Analysis, Control, Optimization and Applications. Birkhäuser: Massachusetts Avenue, 1999: 547-566.
- [27] RAGHUNATHAN V, SCHURGERS C, PARK S, et al. Energy-aware wireless microsensor networks[J]. IEEE Signal Processing Magazine, 2002, 19(2): 40-50.
- [28] LIN X, LU R, HUA FEI Z, et al. ASRPAKE: an anonymous secure routing protocol with authenticated key exchange for wireless ad hoc networks[C]//2007 IEEE International Conference on Communications, Jun 24-28, 2007, Glasgow, Scotland. Piscataway: IEEE Press, 2007: 1247-1253.
- [29] LIN Y H, CHANG S Y, SUN H M. CDAMA: concealed data aggregation scheme for multiple applications in wireless sensor networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 25(7): 1471-1483.

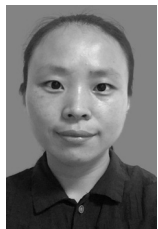
[作者简介]



虞晓韩(1987-), 男, 博士, 浙江工商大学讲师, 主要研究方向为无线传感器网络、压缩感知、数据收集和人工智能等。



董克明(1986-), 男, 博士, 云南财经大学讲师, 主要研究方向为无线传感器网络、数据挖掘和数据隐藏等。



李霞(1988-), 女, 博士, 浙江工商大学讲师, 主要研究方向为无线传感器网络、光传感器和可调带宽滤波器等。



陈超(1986-), 男, 博士, 浙江工商大学副教授, 主要研究方向为无线传感器网络、网络编码和机器学习等。