Cost-aware stochastic compressive data gathering for wireless sensor networks

Huang, Jiajia; Soong, Boon-Hee

2018

Huang, J., & Soong, B.-H. (2019). Cost-aware stochastic compressive data gathering for wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 68(2), 1525-1533. doi:10.1109/TVT.2018.2887091

<https://hdl.handle.net/10356/143046>

<https://doi.org/10.1109/TVT.2018.2887091>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/TVT.2018.2887091>

Downloaded on 26 Apr 2023 21:11:05 SGT

Cost-Aware Stochastic Compressive Data Gathering for Wireless Sensor Networks

Jiajia Huang, *Member, IEEE*, Boon-Hee Soong, *Senior Member, IEEE*

Abstract—Data gathering is a crucial function of wireless sensor networks (WSNs). In resource-limited WSNs, it is critical to improve cost-efficiency and prolong network lifetime. Sensor networks utilizing deterministic routing paths are particularly vulnerable to attacks. Additionally, repeated use of the same path will introduce load unbalance. In this paper, we propose a cost-aware stochastic compressive data gathering for WSNs. In contrast to traditional compressive sensing (CS) based algorithms that implicitly assume uniform transmission cost, our proposed scheme will consider the cost diversity into the CS data gathering framework. The Markov chain based model will be used to characterize the stochastic data gathering process. An optimization problem is formulated to minimize the total expected cost subjected to the constraints on global degree of randomness and the recovery error. Our proposed algorithm requires less total expected cost to achieve certain level of recovery accuracy. It also prolongs network lifetime due to its load balancing features. Extensive simulations on both synthetic and real data show that the proposed algorithm significantly outperforms benchmark algorithms.

Index Terms—Wireless sensor networks, compressive sensing, stochastic routing, absorbing Markov chain, sparse random projections.

I. INTRODUCTION

The Internet of Things (IoT) enables the objects to sense, communicate, and make decisions with its underlying technologies [1]. As a major component of IoT, wireless sensor networks (WSNs) are being increasingly integrated into many aspects of our lives in areas such as medical health care, intelligent transport systems, and smart cities, just to name a few. The most crucial function of WSNs is data gathering, whereby sensor nodes collect environmental data and transmit to the sink node. Sensor devices do this with tremendous speed, creating an immense amount of data that presents a serious challenge to energy-constrained sensor nodes.

The primary problem faced by WSNs lies in prolonging network lifetimes. There are several challenges to this. Firstly, the sensor nodes are energy-constrained. They are almost impossible to recharge after deployment. Thus, energy-efficiency is an important criterion when designing data gathering algorithms for WSNs. Secondly, as the sensor nodes possess limited computational capability, they cannot afford complex and expensive computation. Moreover, reliable delivery of

critical information is vital for successful operation. However, the unreliable nature of communication links in a hostile environments poses a major challenge for the reliable delivery of information to the sink nodes [2], [3].

The conventional method of data gathering in WSNs is through deterministic routing. All the data packets are transmitted and forwarded via the shortest path to the sink nodes. Deterministic tree-based algorithms work well in small static networks. However, as a network grows in size, several problems surface. Firstly, the deterministic tree-based data gathering inherently introduces load unbalance. Likewise, the repeated use of the same shortest path will drain out the nodes along the shortest path quickly, which shortens the network lifetime. In addition, deterministic tree-based algorithm cannot cope with changing environment and node failure. Finally, the predictability of deterministic routing algorithms makes it vulnerable to malicious attacks and eavesdropping.

In contrast to deterministic routing, stochastic routing randomly selects the next hop with a certain probability. By allowing randomness in routing, stochastic routing proves beneficial to both dynamic and static networks. In static networks, the randomness of stochastic routing helps to reduce traffic congestion, resulting in energy-efficient load balancing. The random selection of next hop in stochastic routing makes it robust to node failures and frequent topology changes in dynamic wireless sensor networks. Furthermore, for both static and dynamic networks, the randomness in stochastic routing offers inherent security against malicious attacks, because the routing path is unpredictable.

Data correlation can be exploited to reduce data transmissions and achieve energy-efficiency [4]. Cooperative in-network compression algorithms such as [5] require information exchange for data compression, which introduces significant communication and computation overhead. Distributed source coding (DSC), such as Slepian-Wolf coding [6], [7] reduces the computation complexity in the sensor nodes. However, it requires prior information about the data correlation structure and performs well only when the correlation structure is static. As a promising alternative, the compressive sensing (CS) technique can be applied to data gathering WSNs to reduce data transmission [8]–[11]. Compressive sensing have been shown to recover the data vector with high probability with less measurements than the data vector dimension [12]. CS technique imposes light computation load for the sensor, the resource-demanding recovery process is performed at the powerful sink node. In addition, CS technique requires no prior information about the data correlation structure. Hence, it can cope with changing correlation patterns. CS can be applied in

Copyright (c) 2018 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. J. Huang was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. She is now with the Institute for Infocomm Research, A*STAR. E-mail: huang_jiajia@i2r.a-star.edu.sg.

B.-H. Soong is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. E-mail: ebhsoong@ntu.edu.sg.

tree-based data gathering networks to balance the traffic load across the network, such as in [8], [11]. Furthermore, CS with a sparse measurement matrix can be used to further reduce the number of transmissions in the network [10], [13]–[15].

An implicit assumption for CS is that the transmission cost for each node is the same. Plenty of previous works focus on reducing the number of transmissions while achieving acceptable recovery accuracy. However, in practical sensor networks, the transmission cost may vary due to factors such as the link quality, distance, etc. The cost diversity should be considered in the CS framework to achieve cost-efficiency. Several papers have considered cost diversity under different scenarios [16]–[18]. Xiang *et al.* [16] proposed a hybrid CS based algorithm for multi-hop WSNs with heterogeneous link transmission costs. They considered compressive sensing with a deterministic tree-based routing topology. While in our proposed scheme, randomized routing policy is considered. Additionally, the authors in [16] assumed a dense measurement matrix. Our proposed algorithm assumes a sparse random measurement matrix, which further reduces transmission energy. Xu *et al.* [17] proposed a cost-aware compressive sensing for networked sensing systems. They aimed to achieve a good trade-off between the total sampling cost and the recovery accuracy. A regularized column sum (RCS) was used as a predictive metric for recovery accuracy. Meanwhile, Chen *et al.* [18] proposed a cost-aware activity scheduling for sleeping wireless sensor networks. Their intent was to minimize the sampling cost with a constraint on the regularized mutual coherence. Nevertheless, both [17], [18] considered a single-hop network. In a word, we propose a novel scheme which considers cost diversity together with sparse measurement matrices to achieve energy efficiency and prolong network lifetime for multi-hop WSNs.

In this paper, we combine stochastic routing and compressive sensing to develop a cost-aware stochastic compressive data gathering scheme for multi-hop WSNs. The objective is to collect enough random projections for reliable data recovery through stochastic compression with minimal expected total energy cost. To achieve this, we propose an analytical framework based on Markov chain to model the stochastic compressive data gathering process. Then, we formulate an optimization problem with the objective to minimize expected energy cost, subjected to the constraints on recovery error and global randomness of data gathering process. We seek to construct an optimized transition probability matrix so that the equivalent measurement matrix constructed from the absorbing Markov chain can guarantee the recovery of data with certain error bound. Stochastic routing is used to deliver randomly selected compressed sensor readings to the sink node through several dynamic routes. As mentioned previously, this approach brings benefits such as load balancing and security. However, the resulting equivalent CS measurement matrix will be a sparse non-uniform random matrix, which is different from a traditional uniform sparse random CS measurement matrix whose recovery performance is guaranteed using ℓ_1 -minimization recovery. Hence, a recovery algorithm based on sketching decoder is exploited. In summary, the contributions of this paper can be described as the following.

- We develop a cost-aware stochastic compressive data gathering scheme for multi-hop WSNs.
- We exploit a Markov chain based model to characterize the stochastic compressive data gathering process and formulate an optimization problem to minimize total expected cost.
- We derive the formula to describe the relationship between the transition probability matrix of the Markov chain and the equivalent measurement matrix of the stochastic compressive data gathering process.
- By solving the optimization problem, we construct an optimized equivalent sparse measurement matrix with optimized compression probability for compressive data gathering and recovery.
- Extensive simulation results are provided to show that our proposed algorithm outperforms benchmark algorithms.

The remainder of this paper is organized as follows. Section II sets up the network model and describes the process of stochastic compressive sensing. In Section III, the cost-aware stochastic compressive data gathering problem is formulated as a constrained optimization problem. Section IV presents the procedure to solve for optimal compression probability. In Section V, the low complexity sketching decoder to recover the CS data is presented. In Section VI, we evaluate the performance of the proposed algorithm through simulation. Section VII concludes the paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Network model

Consider a multi-hop wireless sensor network consisting of $n - 1$ nodes and a sink. The WSN can be modelled as a directed acyclic graph (DAG) [19], $G = (V, E)$, where V is the set of nodes in the network, E denotes the wireless links. Each node is assumed to know its geographical information and neighbors. Neighbors of node i include the predecessors set M_i and the successors set N_i . Successors set of node i can be defined as $N_i = \{j | d_{jn} \leq d_{in}, d_{ij} \leq r_0, j \in V\}$, where n is the sink node, d_{ij} is the Euclidean distance between node i and node j , r_0 is the radius of communication range. Similarly, the predecessors set of node i can be defined as $M_i = \{j | d_{jn} \geq d_{in}, d_{ij} \leq r_0, j \in V\}$. The outgoing traffic of node i will be transmitted to its successors with a certain probability. Moreover, we assume that, there is at least one path from each node to the sink node. We define a cost matrix C with entries c_{ij} denoting the cost on link (i, j) .

The sensor nodes are deployed in a sensing field to sense the environment and report their readings to the sink node through multi-hop. The sensor reading for node i is denoted as u_i . The data vector collected by all the sensor nodes is denoted as $\mathbf{u} = [u_1, \dots, u_n]^T$. Previous research works [13], [20] have shown that various natural data, such as temperature data vector for a smooth sensing field, are compressible. Therefore, we assume that the data vector $\mathbf{u} \in \mathbb{R}^n$ is compressible in the transform domain Ψ . More specifically, $\mathbf{u}$ can be represented as $\mathbf{u} = \Psi\boldsymbol{\theta}$, where Ψ is an $n \times n$ orthonormal transform matrix (e.g. Fourier or wavelet transform basis [21]), $\boldsymbol{\theta} = \Psi^T \mathbf{u}$ is an $n \times 1$ transform coefficient vector. The magnitude of $\boldsymbol{\theta}$

decays according to power law [22]. That is, if we order θ in magnitude, such that $|\theta|_{(1)} \geq |\theta|_{(2)} \geq \dots \geq |\theta|_{(n)}$. Then, the i th largest transform coefficient satisfies,

$$|\theta|_{(i)} \leq D_0 i^{-\frac{1}{q}}, \quad i = 1, 2, \dots, n, \quad (1)$$

where D_0 is a constant, q controls the rate of decay of the transform coefficients.

The approximation error obtained by keeping the largest k coefficients and setting the remaining coefficients to zero, is denoted as,

$$\|\mathbf{u} - \hat{\mathbf{u}}_k\|_2 = \|\theta - \hat{\theta}_k\|_2 \leq \alpha_q D_0 k^{-\frac{1}{q} + \frac{1}{2}}, \quad (2)$$

where α_q is a constant only depends on q .

B. Recovery of Compressible Data

For a compressible data vector $\mathbf{u}$, by obtaining L random linear projections $\mathbf{y} = \Phi \mathbf{u}$, where Φ is an $L \times n$ measurement matrix, traditional CS theory guarantees that data vector $\mathbf{u}$ can be recovered from $\mathbf{y}$ with high probability [12]. To be more specific, data recovery can be achieved by solving the ℓ_1 -norm minimization problem as shown below in (3).

$$\min_{\theta \in \mathbb{R}^n} \|\theta\|_{\ell_1} \quad \text{subject to} \quad \mathbf{y} = \Phi \Psi \theta, \quad (3)$$

where $\|\theta\|_{\ell_1} = \sum_i |\theta_i|$. Let $\hat{\theta}$ be the optimal solution for the above optimization problem. The reconstructed data vector is given by $\hat{\mathbf{u}} = \Psi \hat{\theta}$. ℓ_1 -minimization can be solved in polynomial time by linear programming [23], [24].

The choice of measurement matrix Φ will affect the recovery accuracy of the data vector. The ℓ_1 -minimization algorithm performs an accurate recovery if the measurement matrix satisfies the restricted isometry property (RIP) [25], [26]. Previous research works [12], [23] have shown that dense random measurement matrices such as Gaussian random matrices and Bernoulli random matrices can satisfy RIP condition. Hence, they can guarantee the recovery accuracy using ℓ_1 -minimization algorithm. All the elements are nonzero for dense random matrices.

Conversely, the equivalent measurement matrix of the proposed stochastic compressive data gathering algorithm (which will be introduced later) is a non-uniform sparse random matrix, which doesn't satisfy the RIP condition. Therefore, a low complexity sketching decoder [13] with optimal error bound is used in our proposed algorithm for recovery.

C. Stochastic Compressive Data Gathering

The stochastic data gathering process can be modelled using absorbing Markov Chain. The collection of one random projection starts with an initiate node. The initiate node randomly selects a successor from its neighbors with a certain probability. It subsequently transmits its compressed data to the selected node. The selected node compresses its own reading into the packet and transmits towards the sink node. It can be observed that the random selection of the successor node depends only on the assigned transition probability and is not related to previous packet transmission. Moreover, all

Fig. 1: Stochastic compressive data gathering.

the compressed data packets are routed towards a single destination. Consequently, the stochastic routing of one random projection can be modelled as a discrete absorbing Markov Chain [27], with a set of states $S = \{s_1, s_2, \dots, s_n\}$ and a transition probability matrix $\mathbf{P}$. The entry p_{ij} of $\mathbf{P}$ gives the transition probability from state s_i to s_j between time slot t and $(t+1)$. Node i is represented by the state s_i . Without loss of generality, we assume that the sink node n corresponds to the absorbing state s_n while other nodes correspond to transient states $s_1, s_2, \dots, s_{n-1}$. With this model, the transition probability p_{ij} specifies the probability that a packet is transmitted from node i to node j .

As shown in Fig. 1, one random walk from the initiate node towards the sink node results in one random projection. For example, the random projection y_l received at the sink node is produced by the l -th random walk, corresponds to the product of the l -th row of the measurement matrix and the data vector, $y_l = \Phi_l \mathbf{u}$, where Φ_l denotes the l -th row of the measurement matrix. In the following, we will describe how to obtain the l -th random projection y_l . We assume that the l -th random walk consists of the following nodes $\{\beta_1, \dots, \beta_{d_l}\}$, where d_l is the total length of the l th random walk, β_1 is the initiate node and β_{d_l} is the sink node.

To begin with, the initiate node β_1 generates a packet with the reading $y_1^1 = u_{\beta_1}$. Subsequently, one of its neighbors is randomly chosen as β_2 with certain probability. And the generated packet is transmitted to β_2 .

Upon receiving the packet, β_2 generates a packet aggregating its own sensor reading with the received data by either adding or subtracting its own reading. $y_2^2 = y_1^1 + u_{\beta_2}$ or $y_2^2 = y_1^1 - u_{\beta_2}$. Then, y_2^2 is transmitted to a randomly selected neighbors.

Accordingly, for the i -th node of path l , β_i , it generates the packet with aggregated data $y_i^i = y_{i-1}^{i-1} \pm u_{\beta_i}$.

The packet will be aggregated until it reaches the sink node, which will produce the l -th linear random projection $y_l = y_{d_l}^{d_l}$. L random walks result in L random projections. Once the sink node obtains all the L random projections $\mathbf{y} = [y_1, \dots, y_L]^T$, it can recover the data vector with the recovery algorithm, which will be introduced later in Section V.

The long term behavior of the random walk determines the compression probability of each node, which also affects the recovery quality. We define the compression probability

π_j as the probability that the sensor reading of node j is aggregated to form the random projection. In other words, compression probability is the probability that the j th entries of the measurement matrix is nonzero.

D. Equivalent Measurement Matrix and Compression Probability

As discussed earlier, given an initiate node, the long term behavior of the corresponding absorbing Markov chain determines the compression probability of each node. Therefore, given the L random walks to the sink node, we can construct the following equivalent measurement matrix,

$$\phi_{lj} = \begin{cases} +1 & \text{with prob. } \frac{\pi_j}{2}, \\ 0 & \text{with prob. } 1 - \pi_j, \\ -1 & \text{with prob. } \frac{\pi_j}{2}, \end{cases} \quad (4)$$

where $l = 1, \dots, L$ corresponds to the number of random walks. The l -th random projection is produced by $\sum_{j=1}^n \phi_{lj} u_j$. The probability that $\phi_{lj} \neq 0$ is defined as the compression probability π_j , which is related to the occurrence frequency of node j in the absorbing Markov Chain. Here, we assume that node 1 is the initiate node. Starting from node 1, L independent random walks to the sink node will produce L random projections. Then the sink node recovers the original data from the random projections.

Since the compression probability relates to the long term behavior of the absorbing Markov chain, in order to derive the compression probability, we need to write the transition probability matrix in the following canonical form [27],

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{r} \\ \mathbf{0}^T & 1 \end{bmatrix},$$

where $\mathbf{Q}$ is an $(n-1) \times (n-1)$ matrix representing the transition probabilities of the transient nodes, $\mathbf{r}$ is an $(n-1) \times 1$ matrix representing the transition probabilities of the links terminating at the sink node and $\mathbf{0}$ is an $(n-1) \times 1$ column vector with all zero entries.

Given the transition probability matrix $\mathbf{P}$, the *fundamental matrix* of $\mathbf{P}$ can be computed as $\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1}$. The entry f_{ij} of $\mathbf{F}$ denotes the expected number of times the process is in transient state s_j if it is initiated in state s_i . In other words, f_{ij} denotes the average occurrence of node j if the initiate node is i . Assume node 1 is the initiate node. If started from node 1, the expected number of occurrence of node j is f_{1j} . Then the probability of compression π_j can be given by the following formula,

$$\pi_j = f_{1j}. \quad (5)$$

The above equation (5) establishes the relationship between the transition probability of the Markov chain and the equivalent measurement matrix of the compressive data gathering. π_j controls the sparsity of the equivalent measurement matrix, larger π_j indicates a denser measurement matrix and possibly more energy consumption.

E. Expected Energy Cost

The energy cost is a key performance measurement for a data gathering WSN. In deterministic routing, the route to deliver the data to the sink node with the least energy cost can be found by shortest path algorithm. In the context of stochastic routing, where all the available paths between the source and the sink node are utilized, the expected energy cost is used as a performance metric.

The expected cost to transmit a packet from any initiate node i to the sink node is denoted as

$$w_i = \begin{cases} \sum_{j \in N_i} p_{ij}(c_{ij} + w_j) & \text{if } i \neq n \\ 0 & \text{if } i = n \end{cases} \quad (6)$$

In matrix form, the expected cost vector in (6) can be written as

$$\mathbf{w} = \text{diag}(\mathbf{P}\mathbf{C}^T) + \mathbf{P}\mathbf{w}, \quad (7)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_{n-1}, 0]^T$ is the expected cost vector.

F. The Randomness of the Data Gathering Process

A deterministic shortest path to the sink node will result in the minimum cost. However, the deterministic routing path is vulnerable to malicious attacks and results in load unbalance. Introducing stochastic routing can improve the robustness of the network and reduce congestion. On the other hand, the randomness of the data gathering process affects the expected energy cost. The expected energy cost increases as the randomness increases. The Shannon entropy [28] of a node is used to measure the local randomness of the next hop choice for a node. Here, the Shannon entropy of the transition probabilities can be given as

$$h_i = - \sum_{j \in N_i} p_{ij} \log p_{ij}, \quad \forall i \neq n. \quad (8)$$

Moreover, the global entropy H is defined as the sum of weighted individual values as

$$H = \sum_{i=1}^{n-1} f_{1i} h_i = - \sum_{i=1}^{n-1} f_{1i} \sum_{j \in N_i} p_{ij} \log p_{ij}, \quad (9)$$

where each individual entropy is weighted by the expected number of visits to this node. The global entropy measures the global degree of randomness of the data gathering process.

III. PROBLEM FORMULATION

By Modelling the stochastic compressive routing as an absorbing Markov chain, the expected energy cost vector is given by (6).

With sparse compressive sensing technique, only $L \ll n$ random walks to the sink node is sufficient for data recovery. In this paper, we aim to derive the optimal transition probability matrix that minimizes the total expected cost to collect L compressed measurement at the sink node subjected to the constraint on the bound on CS recovery error and the global degree of randomness of the data gathering process.

Node 1 is selected as the initiate node, where w_1 is the expected cost to transmit a packet from node 1 to the

Fig. 2: The framework of cost-aware stochastic compressive data gathering.

sink node. It takes L random walks to produce L random projections. Consequently, the total expected cost to collect L random projections at the sink node is given by Lw_1 . Then, the problem can be formulated as the following.

$$\underset{\mathbf{P}}{\text{minimize}} \quad Lw_1 \quad (10)$$

$$\text{subject to} \quad -\sum_{i=1}^{n-1} f_i \sum_{j \in N_i} p_{ij} \log p_{ij} = H_0, \quad (11)$$

$$L = O(k^2 \log n), \quad (12)$$

$$\mathbf{w} = (\mathbf{I} - \mathbf{P})^{-1} \text{diag}(\mathbf{P}\mathbf{C}^T), \quad (13)$$

$$0 \leq p_{ij} \leq 1, \forall (i, j) \in E, \quad (14)$$

$$\sum_{j \in N_i} p_{ij} = 1, \quad \forall i \neq n, \quad (15)$$

where constraint (11) is the global randomness constraint, constraint (12) specifies the number of random walks to guarantee data recovery, constraint (13) establishes the relationship between the expected cost of one random walk and the transition probability, constraint (15) states that the summation of the probability to neighbors is 1. We solve the above optimization problem for the optimal transition probability, following which the compression probability can be obtained by equation (5).

The framework of our cost-aware stochastic compressive data gathering can be described by the flow chart shown in Fig. 2.

IV. OPTIMAL COMPRESSION PROBABILITY

In this section, we derive the formula to compute the optimal probability to minimize the total expected data gathering cost by solving the optimization problem. With the optimal transition probability, we can compute the optimal compression probability with equation (5). We solve the problem with the Lagrange multiplier method. To begin with, the Lagrangian for the optimization problem can be obtained as the following.

$$\begin{aligned} \mathcal{L} = & Lw_1 + \sum_{i=1}^{n-1} \lambda_i [w_i - \sum_{j \in N_i} p_{ij}(c_{ij} + w_j)] + \lambda_n(w_n) \\ & + \sum_{i=1}^{n-1} \mu_i (\sum_{j \in N_i} p_{ij} - 1) + \eta [\sum_{i=1}^{n-1} f_i \sum_{j \in N_i} p_{ij} \log p_{ij} + H_0] \\ & + \rho(L - k^2 \log n), \end{aligned}$$

where λ_i, η and ρ are the Lagrangian multipliers.

Setting $\frac{\partial \mathcal{L}}{\partial p_{kl}} = 0$, we obtain

$$\begin{aligned} & -\lambda_k(c_{kl} + w_l) + \mu_k + \eta f_k (\log p_{kl} + 1) \\ & + \eta \sum_{i=1}^{n-1} \left(\frac{\partial f_i}{\partial p_{kl}} \right) \sum_{j \in N_i} p_{ij} \log p_{ij} = 0. \end{aligned}$$

Substituting $h_i = -\sum_{j \in N_i} p_{ij} \log p_{ij}$, the above equation can be simplified into

$$p_{kl} = \exp \left[\frac{\lambda_k(c_{kl} + w_l)}{\eta f_k} - \frac{\mu_k + f_k}{\eta f_k} + \frac{\eta \sum_{i=1}^{n-1} \left(\frac{\partial f_i}{\partial p_{kl}} \right) h_i}{\eta f_k} \right]. \quad (16)$$

Using the fact that the summation of the probability to neighbors is 1, $\sum_{l' \in N_k} p_{kl'} = 1$, with some manipulation, we obtain

$$\begin{aligned} & \exp \left[\frac{-\mu_k - \eta f_k}{\eta f_k} \right] \\ & = \frac{1}{\sum_{l' \in N_k} \exp \left[\frac{\lambda_k(c_{kl'} + w_{l'}) + \eta \sum_{i=1}^{n-1} \left(\frac{\partial f_i}{\partial p_{kl'}} \right) h_i}{\eta f_k} \right]}. \end{aligned} \quad (17)$$

Substituting equation (17) into (16), we can have the update equation for probability as the following,

$$p_{kl} = \frac{\exp \left[\frac{\lambda_k(c_{kl} + w_l) + \frac{1}{f_k} \sum_{i=1}^{n-1} \left(\frac{\partial f_i}{\partial p_{kl}} \right) h_i}{\eta f_k} \right]}{\sum_{l' \in N_k} \exp \left[\frac{\lambda_k(c_{kl'} + w_{l'}) + \eta \sum_{i=1}^{n-1} \left(\frac{\partial f_i}{\partial p_{kl'}} \right) h_i}{\eta f_k} \right]}. \quad (18)$$

Then, with $\frac{\partial \mathcal{L}}{\partial w_k} = 0$, we obtain

$$\lambda_k - \sum_{i \in M_k} \lambda_i p_{ik} = 0, \text{ for } k \neq 1, \quad (19)$$

where M_k is the set of predecessors of node k . Since the initiate node has no predecessors, we obtain $\lambda_1 = -1$. Defining $\tau_k = -\lambda_k$, then we have

$$\tau_k = \begin{cases} \sum_{i \in M_k} \tau_i p_{ik} & \text{if } k \neq 1 \\ 1 & \text{if } k = 1 \end{cases} \quad (20)$$

Defining $\zeta_l = \frac{1}{f_k} \sum_{i=1}^{n-1} \left(\frac{\partial f_i}{\partial p_{kl}} \right) h_i$, the optimal transition probability from node k to node l is given by

$$p_{kl} = \frac{\exp \left[\frac{-\tau_k(c_{kl} + w_l) + \zeta_l}{\eta f_k} \right]}{\sum_{l' \in N_k} \exp \left[\frac{-\tau_k(c_{kl'} + w_{l'}) + \zeta_{l'}}{\eta f_k} \right]}. \quad (21)$$

Theoretically, the optimal $\mathbf{P}$ can be found by iteratively solving equation (21). However, it is impossible to get an analytical expression of multiplier η . Therefore, we use Algorithm 1 to solve for optimal $\mathbf{P}$ with η as an input. Parameter $\eta > 0$ is an indicator of global randomness, which is a monotonically increasing function of H , where larger η indicates larger global randomness. Given η , the optimal transition probability can be found with the iterative method.

After obtaining the optimal transition probability, the optimal compression probability for node j , $\pi_j = f_{1j}$, which corresponds to the occurrence frequency of node j when the initiate node is node 1. Algorithm 1 summarizes the procedure of solving optimal compression probability.

Algorithm 1: Solve for optimal compression probability

Input : Graph $G = (V, E)$, initiate node = 1, sink node = n , cost matrix $\mathbf{C}$, randomness parameter η

Output: Optimal compression probability π^*

- 1 Initialize the transition probability as pure random, $p_{ij} = \frac{1}{\deg(i)}$, where $\deg(i)$ is the outdegree of node i ;
- 2 **repeat**
- 3 Compute h_i using equation (8) ;
- 4 Solve $\mathbf{w} = \text{diag}(\mathbf{P}\mathbf{C}^T) + \mathbf{P}\mathbf{w}$;
- 5 Compute global entropy H with equation (9) ;
- 6 Compute ζ_j for each node j ;
- 7 **for** $j \in N_i$ **do**
- 8 | Update p_{ij} with the equation (21) ;
- 9 **end**
- 10 **until** P converges;
- 11 Compute the corresponding fundamental matrix of $\mathbf{P}$, $\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1}$;
- 12 $\pi_j = f_{1j}$

V. DATA RECOVERY USING SKETCHING DECODER

According to [10], [13], with the measurement matrix Φ , transform matrix Ψ , and the linear random projections $\mathbf{y}$, we first recover the transform coefficients vector, recorded as $\hat{\theta} = [\hat{\theta}_1, \dots, \hat{\theta}_n]^T$. Then the data vector can be recovered as $\hat{\mathbf{u}} = (\hat{\theta}\Psi^{-1})^T$. The detailed low complexity sketching decoder is described as in Algorithm 2.

Previous studies have shown that, with $L = O(k^2 \log n)$ random projections, the recovery algorithm can achieve error bound that is comparable to k best approximation [10], [13]. More specifically, the recovery error bound of sparse compressive data gathering is described by the Theorem 1 as the following.

Theorem 1. Suppose data vector $\mathbf{u} \in \mathbb{R}^{n \times 1}$ satisfies $\frac{\|\mathbf{u}\|_\infty}{\|\mathbf{u}\|_2} \leq M$. A sparse random measurement matrix Φ with probability distribution π , and $L = B(\delta, \epsilon)k^2 \log n$. Let $\mathbf{y} = \frac{1}{\sqrt{L}}\Phi\mathbf{u}$, consider an orthonormal transform $\Psi \in \mathbb{R}^{n \times n}$ and the corresponding transform coefficients $\theta = \Psi^T\mathbf{u}$. If the k largest transform coefficients in magnitude give an approximation with error $\|\mathbf{u} - \hat{\mathbf{u}}_{\text{opt}}\|_2^2 \leq \delta\|\mathbf{u}\|_2^2$, then given only $\mathbf{y}$, Φ and Ψ , the sparse stochastic data gathering gives an estimate $\hat{\mathbf{u}}$ with error

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_2^2 \leq (1 + \epsilon)\delta\|\mathbf{u}\|_2^2, \quad (22)$$

with probability at least $1 - n^{-\gamma}$. $\gamma > 0$ is a constant.

Theorem 1 states that, with the given L number of linear projections, the recovery error of proposed stochastic compress-

Algorithm 2: Sketching decoder for non-uniform sparse random projection recovery.

Input : Transform basis matrix Ψ , $L_1 = C_1 \frac{1+\xi+\chi H^2}{\epsilon^2}$ and $L_2 = C_2(1 + \gamma) \log n$ and $L = L_1 L_2$

Output: Recovered data vector $\hat{\mathbf{u}}$

- 1 Construct measurement matrix Φ with optimized compression probability ;
- 2 Divide the measurement matrix Φ into L_2 sub-matrices, $\{\Phi_1, \Phi_2, \dots, \Phi_{L_2}\}$, each of size $L_1 \times n$;
- 3 Divide transform matrix Ψ into n column vector $\{\psi_1, \dots, \psi_n\}$, where $\psi_j \in \mathbb{R}^n$ is the j -th column of the original transform matrix. ;
- 4 Divide the random projection of the data vector into L_2 sub-matrices, $\{\mathbf{a}_1 = \frac{1}{\sqrt{L_1}}\Phi_1\mathbf{u}, \dots, \mathbf{a}_{L_2} = \frac{1}{\sqrt{L_1}}\Phi_{L_2}\mathbf{u}\}$, each of size $L_1 \times 1$;
- 5 **for** $j = 1 : n$ **do**
- 6 **for** $l = 1 : L_2$ **do**
- 7 | Project the basis vector to obtain $\mathbf{b}_l^j = \Phi_l \psi_j$;
- 8 | Compute the inner product of $\mathbf{a}_l$ and $\mathbf{b}_l^j$, obtaining $z_l^j = (\mathbf{a}_l)^T \mathbf{b}_l^j$;
- 9 **end**
- 10 Compute estimate of the j -th element of the transformation coefficients, $\hat{\theta}_j = \text{median}\{z_1^j, \dots, z_{L_2}^j\}$;
- 11 **end**
- 12 Keep the largest k coefficients in $\hat{\theta}$, and set the remaining to zero ;
- 13 The data vector can be recovered as $\hat{\mathbf{u}} = (\hat{\theta}\Psi^{-1})^T$;

sive data gathering can achieve an error which is only $(1 + \epsilon)$ times of the optimal error bound for a k -sparse data vector.

VI. PERFORMANCE EVALUATION

Fig. 3: Synthetic data vector and the corresponding transform coefficients using Fourier basis.

In this section, we evaluate the performance of the proposed scheme on both synthetic data and real data. For synthetic data,

(a) The trade-off between number of transmissions and the recovery error. Synthetic data, network size $n = 200$.

(b) The trade-off between number of transmissions and the recovery error. Real data, DCT transform, network size $n = 54$.

Fig. 4: The trade-off between the number of transmissions and the recovery error.

(a) The trade-off between overall expected energy cost and the recovery error. Synthetic data, network size $n = 200$.

(b) The trade-off between overall expected energy cost and the recovery error. Real data, DCT transform, network size $n = 54$.

Fig. 5: Trade-off between overall expected energy cost and the recovery error.

the sparsity of the data vector can be controlled. The synthetic data is generated as follows: we randomly generate an acyclic directed graph with $n = 200$ nodes, each connected link (i, j) is associated with a normalized cost c_{ij} ranges from 0 to 1. We assume that the data vector collected by the n nodes is k -sparse on the frequency domain. That is, the Fourier transform coefficients have only k significant values in magnitude. Fig. 3 gives an example of a synthetic data vector for a network of 200 nodes. It can be observed that the data vector is 4 sparse in the frequency domain, which can be recovered using CS method. While the real data is the temperature data imported from Intel Lab [29], and the data vector is sparsified using Discrete cosine transform (DCT).

The performance of the proposed algorithm is evaluated from three aspects: the recovery error, the total expected cost and the network lifetime. The recovery error is defined as $e = \frac{\|\mathbf{u} - \hat{\mathbf{u}}\|_2^2}{\|\mathbf{u}\|_2^2}$, where $\mathbf{u}$ is the original data vector, and $\hat{\mathbf{u}}$ is the recovered data vector. We assume that each node has a certain

amount of initiate energy. The network lifetime is defined as the time duration until the first node runs out of energy, which is measured in time slots.

For comparison, random walk based routing [30] with CS technique and biased random walk [31] with k -nearest neighbor (KNN) method [32] are also implemented.

First of all, for CS based algorithms, Fig. 4 plots the number of random projections to achieve a certain level of accuracy. It can be observed that, for both our proposed and benchmark algorithm, the recovery error reduces as the number of random projections increases. In addition, less degree of global randomness results in larger number of random projections. For example, in Fig. 4a, when there are 200 nodes in the network, 140 random projections are required to achieve approximately 0.4 recovery error for $H = 4.29$, while only 60 random projections are required to achieve the same level of accuracy for global randomness of $H = 7.31$. This may due to the fact that the equivalent measurement matrix of an absorbing Markov Chain with a smaller degree of randomness is more

(a) Network lifetime vs the recovery error. Synthetic data, network size $n = 200$.

(b) Network lifetime vs the recovery error. Real data, DCT transform, network size $n = 54$.

Fig. 6: Network lifetime vs the recovery error.

sparse, leading to more random projections for data recovery.

Fig. 5 investigates the total expected cost for the proposed algorithm. It can be observed that, for a given degree of randomness, to achieve lower recovery error, the overall energy cost is increased. On the other hand, to achieve the same recovery error, our proposed cost-aware algorithm requires less energy cost than the benchmark algorithms. For example, in Fig. 5a, to achieve recovery error less than 30%, the proposed algorithm achieves about $\frac{1}{2}$ total energy cost of the random walk based-CS algorithm and $\frac{1}{10}$ total expected energy cost of the biased random walk-KNN algorithm. The proposed algorithm minimizes the total expected cost for a certain degree of randomness.

Fig. 6 plots the network lifetime as a function of recovery error. It can be observed that, for a given algorithm, in order to achieve better recovery accuracy, network lifetime will decrease as the recovery error reduces. For the same level of accuracy, the proposed algorithm achieves longer network lifetime compare to benchmark algorithms. As illustrated in Fig. 6a, to achieve the recovery error of 0.4, the network lifetime of the proposed algorithm with $H = 7.31$ is about 1.4 times longer than random walk based-CS algorithm and 2.5 times longer than biased random walk-KNN algorithm. Evidently, extensive simulations on both synthetic data and real data indicate that our proposed algorithm minimizes the total expected cost and prolongs network lifetime.

VII. CONCLUDING REMARKS

In this paper, we proposed a cost-aware stochastic compressive data gathering scheme for multi-hop WSNs. We proposed a Markov chain based method to analyze the relationship between the total expected cost and recovery accuracy to collect and recover the data vector. An optimization problem to minimize the total expected cost for data gathering subjected to the recovery accuracy and global randomness constraints is formulated. The results from our simulations have shown that the proposed algorithm prolonged network lifetime whilst achieving a good trade-off between total expected cost and the recovery accuracy. Moving forward, the proposed scheme

can also be extended to consider more complicated scenarios with multiple mobile sinks, which were not represented in our findings.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their constructive inputs that help to improve the quality of this paper. This research was funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (Sin-BerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

REFERENCES

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.
- [2] U. S. Wijetunge, R. Subramanian, A. Pollok, and S. Perreau, "Maximizing packet delivery ratio in stochastic routing: An analytical perspective," in *2014 IEEE International Conference on Communications (ICC)*, 2014, Conference Proceedings, pp. 415–420.
- [3] M. Zhao, P. H. J. Chong, and H. C. Chan, "An energy-efficient and cluster-parent based rpl with power-level refinement for low-power and lossy networks," *Computer Communications*, 2016.
- [4] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [5] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, "Network correlated data gathering with explicit communication: Np-completeness and algorithms," *IEEE/ACM Transactions On Networking*, vol. 14, no. 1, pp. 41–54, 2006.
- [6] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [7] J. Zheng, P. Wang, and C. Li, "Distributed data aggregation using slepian-wolf coding in cluster-based wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2564–2574, 2010.
- [8] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proceedings of the 15th annual international conference on Mobile computing and networking*. ACM, 2009, pp. 145–156.

- [9] X. Liu, L. Jun, and C. Rosenberg, "Compressed data aggregation: Energy-efficient and high-fidelity data collection," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1722–1735, 2013.
- [10] X. Y. Liu, Y. Zhu, L. Kong, C. Liu, Y. Gu, A. V. Vasilakos, and M. Y. Wu, "Cdc: Compressive data collection for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2188–2197, 2015.
- [11] J. Huang and B.-H. Soong, "Hybrid compressive sensing for delay-efficient sustainable data gathering," in *IEEE 85th Vehicular Technology Conference (VTC), Sydney, Australia*, 2017.
- [12] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [13] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Proceedings of the 6th international conference on Information processing in sensor networks*. ACM, 2007, pp. 331–339.
- [14] Z. Haifeng, Y. Feng, T. Xiaohua, G. Xiaoying, W. Xinbing, and X. Shilin, "Data gathering with compressive sensing in wireless sensor networks: A random walk based approach," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 1, pp. 35–44, 2015.
- [15] D. Ebrahimi and C. Assi, "On the interaction between scheduling and compressive data gathering in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.
- [16] X. Liu, L. Jun, and A. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks," in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*, 2011, pp. 46–54.
- [17] L. Xu, X. Hao, N. Lane, X. Liu, T. Moscibroda, and S. Nath, *Cost-aware compressive sensing for networked sensing systems*. Proceedings of the 14th International Conference on Information Processing in Sensor Networks, 2015.
- [18] W. Chen and I. Wassell, "Cost-aware activity scheduling for compressive sleeping wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. PP, no. 99, pp. 1–1, 2016.
- [19] N. Christofides, *Graph theory: An algorithmic approach*. Academic Press, Inc., New York, 1975.
- [20] M. Vetterli and J. Kovacevic, *Wavelets and subband coding*. Prentice-Hall, 1995.
- [21] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 92–101, 2008.
- [22] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best," *Journal of the American mathematical society*, vol. 22, no. 1, pp. 211–231, 2009.
- [23] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.
- [24] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [25] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*. IEEE, 2008, pp. 798–805.
- [26] A. Gilbert and P. Indyk, "Sparse recovery using sparse matrices," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 937–947, 2010.
- [27] C. M. Grinstead and J. L. Snell, *Introduction to probability*. American Mathematical Soc., 2012.
- [28] M. Saerens, Y. Achbany, F. Fouss, and L. Yen, "Randomized shortest-path problems: Two related models," *Neural computation*, vol. 21, no. 8, pp. 2363–2404, 2009.
- [29] "Intellab data," 2004, accessed: 2017-04-18. [Online]. Available: <http://db.csail.mit.edu/labdata/labdata.html>
- [30] I. Mabrouki, X. Lagrange, and G. Floc, "Random walk based routing protocol for wireless sensor networks," in *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, p. 71.
- [31] R. Beraldi, R. Baldoni, and R. Prakash, "A biased random walk routing protocol for wireless sensor networks: The lukewarm potato protocol," *IEEE Transactions on Mobile Computing*, vol. 9, no. 11, pp. 1649–1661, 2010.
- [32] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.