In-Network Iterative Distributed Estimation for Power-constrained Wireless Sensor Networks

Santosh Shah, Baltasar Beferull-Lozano
Group of Information and Communication Systems
Instituto de Robótica y Tecnologías de la Información y las Comunicaciones
Universidad de Valencia, 46980, Paterna (Valencia), Spain
{Santosh.Shah, Baltasar.Beferull}@uv.es

Abstract—In this paper, we consider the problem of power-efficient distributed estimation of a localized event in the large-scale Wireless Sensor Networks (WSNs). In order to increase the power efficiency in these networks, we develop a joint optimization problem that involves both selecting a subset of active sensors and the routing structure so that the quality of estimation at a given querying node is the best possible subject to a total imposed communication cost. We first formulate our problem as an optimization problem and show that it is NP-Hard. Then, we design two algorithms: a fixed-tree relaxation-based and a novel and very efficient iterative distributed to optimize jointly the sensor selection and the routing structure. We also provide a lower bound for our optimization problem and show that our iterative distributed algorithm provides a performance that is close to this bound. Although there is no guarantee that the gap between this lower bound and the optimal solution of the main problem is always small, our numerical experiments support that this gap is actually very small in many cases. An important result from our work is that the fact that because of the interplay between communication cost and gain in estimation when fusing measurements from different sensors, the traditional Shortest Path Tree (SPT) routing structure, widely used in practice, is no longer optimal, that is, our routing structures provide a better trade-off between the overall communication cost and estimation accuracy. Comparing to more conventional sensor selection and fixed routing algorithms, our proposed joint sensor selection and routing algorithms yield a significant amount of energy saving.

I. INTRODUCTION

Given a large-scale WSN and a querying (sink) node requesting a certain task to be performed, selecting a proper subset of sensors from which information is taken and optimizing the routing structure can lead to very important power savings [1]. In this paper, we focus on the deterministic parameter produced by a localized source target, and consider the scenario where the network performs distributed estimation of this parameter. In a distributed estimation setting of the limited power resources in WSNs, there is a need to choose a subset of active sensors whose data can be used for estimation and a routing tree to route the information to a sink node, which performs the final estimation. This motivates us to perform joint optimization on the sensor selection and the routing so that the quality of estimation has to be maximized for a given power budget, we also show that joint optimization performs better than independent optimization [2].

In this paper, we show that the SPT based only on Communication Cost (SPT-CC) is not the optimal routing structure. Fig. 1 illustrates a simple example where the SPT-CC is not the optimal routing structure when an Estimate-and-Forward (EF) [3] is used, that is, when each sensor performs a parameter estimation and fuses all other measurements it receives from its child sensors together with its own measurement and then sends only one flow of data to its parent sensor in the chosen routing structure, instead of simply forwarding the data, which would generate more flows. As shown in Fig. 1, choosing in this particular case all other sensors, the same gains in estimation are obtained however, taking into account that $f_c(.)$ is increasing with distance, the total communication cost corresponding to SPT-CC is larger, assuming that we generate observations at each sensor.

There has been already a substantial amount of research in sensor selection and data transmission for WSNs with an application view to estimation. The problem of choosing a subset of sensor measurements from a set of possible sensor measurements has been analyzed in [4]–[6], where measurements are transmitted directly to the sink, therefore, no routing. In [4], authors propose a centralized solution based on performing a relaxation of an integer optimization problem using an efficient interior point method. A distributed version of this interior point method is introduced in [5]. In [6], a sampling framework based on linear minimum variance unbiased estimation is proposed to select a subset of sensors. In [7], [8], a tradeoff between the number of active sensors and the energy used (in data transmission directly to the sink) by each active sensor is presented to minimize the Mean...
Square Error (MSE) estimation [9]. A problem of bandwidth constrained distributed estimation of a deterministic parameter has been proposed in [10], [11], where each sensor compresses its observation into a few bits and transmit directly to the fusion center for parameter estimation. In [2], an Innovation Diffusion (ID) using an EF approach is proposed, where the objective is to select a subset of minimum number of active sensors in order to conserve energy and prolong system lifetime. The main drawback of this algorithm is that it does not trade-off estimation accuracy and communication cost jointly. Their multi-hop routing is determined by the next best activating sensor that minimizes their objective function without taking into account communication cost. However, to the best of our knowledge none of these works consider jointly the sensor selection and routing problem.

The main contributions of this paper are: (a) we define and analyze the problem of joint sensor selection and routing for distributed estimation, in terms of both algorithm design and complexity; (b) we show that this problem is NP-Hard; (c) we present a fixed-tree algorithm that is based on a relaxation of our optimization problem and that decouples the choice of routing structure with the sensor selection; (d) then we provide an iterative distributed algorithm using an EF strategy which performs this joint optimization iteratively; (e) we also provide a lower bound for the optimal solution of our optimization problem and show experimentally that our iterative distributed algorithm generates a solution that is close to this lower bound. The rest of the paper is structured as follows: In section II, we present our problem formulation and analyze its NP-hardness. Section III describes our joint sensor selection and routing algorithms and their complexity analysis. In section IV, we present our simulation results, showing the performance of our algorithms. Finally, in section V, we present our conclusions.

II. PROBLEM FORMULATION

![Fig. 2. Symmetric network connectivity graph G, where S is the sink node. We consider the problem of estimating some deterministic parameter generated by a spatially localized source. There is a sink node, whose aim is to obtain the best possible estimation under some power constraints, as it is the case in battery-powered WSNs. We consider stationary sensors with one-to-one communication links assuming the existence of some MAC protocol that resolves the network collisions. We also assume a WSN modeled as an undirected network connectivity graph \( G = (V, E) \), where \( V \) is a set of sensors and \( E \) is a set of one-to-one orthogonal communication links. In general terms, our problem can be naturally formulated as an optimization problem [12] as follows:

\[
\begin{align*}
\text{minimize} & \quad \text{Distortion} \\
\text{subject to} & \quad \text{Comm} \_\text{cost} \leq P_{\text{max}}
\end{align*}
\]

where \( P_{\text{max}} \) is the maximum power budget for the WSN in terms of communication cost, our objective is the average estimation error (Distortion) obtained at the sink node, \( T \subseteq G \) is in general a non-spanning tree (as represented in Fig. 2 by the continuous links) and \( \text{Comm} \_\text{cost} \) is the total communication cost associated to the non-spanning tree \( T \). Thus, our optimization problem involves optimizing jointly both: selecting a subset of sensors and selecting the routing structure. The cost of delivering a measurement to the sink node grows, usually as the number of sensors grows, although it depends strongly on the chosen routing structure. On the other hand, the distortion associated to the estimation can be usually reduced as the number of chosen sensors grows (notice that it depends on the SNRs). Thus, there is a trade-off between distortion and communication cost, which motivates us to perform pareto optimization over these two metrics.

A. Network Model

We assume that two sensors are neighbors if one sensor is in the transmission range of the other sensor and each sensor has a unique identity \( k \in \{1, 2, \ldots, N\} \), where \( N \) is the number of sensors. Then, there is a certain distance range \( d_{th} \) such that if the distance \( d_{ij} \) between two sensors \( i \) and \( j \) is less than or equal to \( d_{th} \), then, there is a communication link between them, which ensures that each sensor can be reached to other sensors in a graph by multi-hop links. This results in a network connectivity symmetric graph \( G \), as illustrated in Fig. 2. We denote the \((N+1)\)-th sensor as the sink node, where the final estimation is to be obtained. In this paper, we assume that each sensor is equipped with an omnidirectional antenna and that the receiver has a noise power level \( N_0 \). We consider that each sensor can adjust its communication cost (transmission power) such that a desired SNR \( S_{th} \) at the receiver is achieved. In this sense, the communication cost from an active sensor \( i \) to its active neighbor \( j \in \mathcal{N}(i) \) is assumed to be:

\[
J_c(d_{ij}) \propto d_{ij}^\alpha N_0 S_{th} B
\]

where \( \alpha (2 < \alpha < 6) \) is the path-loss exponent and \( B \) is the fixed block-length\(^1\) used in each transmission. This simple communication cost model has been experimentally supported in [2], [13], as an approximately valid model for analyzing routing structures in WSN related applications.

B. Signal Model

We use a linear function in our signal model because of its simplicity, and because it leads to practical estimation approaches that give closed form estimators. As we will see later, even with a linear model our optimization problem is already NP-hard, that is, a linear model still maintaining essentially the complexity of our problem. In fact, assuming a

\(^1\)The case of adaptive block-length assignment is left as future work.
linear model, the optimal estimator, as well as its performance, can be readily obtained. Let us consider the model:
\[ y_k = h_k \theta + z_k, \quad k = 1, 2, \ldots, N \]  
(2)
where \( y_k \in \mathbb{R} \) is a scalar observation of sensor \( k \), \( \theta \in \mathbb{R} \) is an unknown deterministic parameter to be estimated, whose observation is distorted by a scalar \( h_k \in \mathbb{R} \) and corrupted by additive Gaussian noise \( z_k \), which is taken to be independent and identically distributed (i.i.d.) with pdf \( \mathcal{N}(0, \sigma^2) \), where \( \sigma^2 \) is assumed to be known. We assume that the scalar \( h_k \) follows the signal strength (e.g. energy) decay model [14]: \( y_k = \theta / d_{k,t}^\theta + z_k \) where \( y_k \) can be the energy sensor reading at sensor \( k \), \( h_k = 1 / d_{k,t}^\theta \) is a decreasing function of the distance \( d_{k,t} \) from the sensor \( k \) to the source target \( t \) of the parameter that we are interested to estimate, and \( \beta \) is the signal decay exponent which is assumed to be known (or estimated via training sequences [14], [15]).

C. Parameter Estimation
The well known Best Linear Unbiased Estimator (BLUE) [9] is the optimal estimator for linear problem (2) giving the smallest possible MSE, thus, it coincides in this case with the Minimum Mean Square Error (MMSE) estimator. The optimal estimator of \( \theta \) is readily given by:
\[ \hat{\theta} = \frac{\sum_{k=1}^{N} h_k y_k}{\sum_{k=1}^{N} h_k^2} \]  
(3)
and the associated MSE is given by:
\[ MSE_{\hat{\theta}} = \left( \frac{\sum_{k=1}^{N} h_k^2}{\sigma^2} \right)^{-1} \]  
(4)

D. Joint Sensor Selection and Routing Optimization Problem
Let us assume a variable \( b_k \in \{0, 1\}, k = 1, 2, \ldots, N \), denoting the status of each sensor, namely, \( b_k = 1 \) denotes that the \( k \)-th sensor is active (i.e. chosen) and \( b_k = 0 \) denotes that the \( k \)-th sensor is inactive (i.e. not chosen). We want to minimize the overall distortion (MSE) in estimation subject to a total communication cost, and where we can choose both a subset of sensors and a routing structure. This can be formulated analytically as follows:

\[
\begin{align*}
\text{minimize} & \quad MSE_{\hat{\theta}} = \left( \frac{\sum_{k=1}^{N} b_k h_k^2}{\sigma^2} \right)^{-1} \\
\text{subject to} & \quad \sum_{k=1}^{N} b_k f_c(d_{k,p_k}) = P_{\text{max}} - \Delta \\
& \quad b_k \leq b_{p_k}; \quad \Delta \geq 0 \\
& \quad b_k \in \{0, 1\}
\end{align*}
\]  
(5)
where \{\( b_k, \Delta \)\}_{k=1}^{N} are chosen so that the resulting routing structure \( T \) is a subtree of the \( G \), \( p_k \) is the identity of the parent of the \( k \)-th sensor, and \( \Delta \) is the power gap, that is, maximum power allowed \( P_{\text{max}} \) minus the total power incurred. Here, the constraint \( b_k \leq b_{p_k} \) ensures that no sensor is selected if its parent on the tree is not selected, and therefore, it ensures that the selected sensors form a valid routing subtree \( T \subseteq G \) rooted at the sink node. We let \( p^* \) denote the optimal objective value of this problem.

E. NP-hardness

Theorem 1. The joint optimization problem (5) of sensor selection and multi-hop routing structure for distributed estimation under a total power constraint, is NP-Hard.

Proof: The proof is based on performing a polynomial time reduction [16] from the Directed Hamiltonian Path (DHP) to our problem, that is, mapping every instance from the DHP problem to our problem (see Appendix for detailed proof).

III. JOINT SENSOR SELECTION AND ROUTING ALGORITHMS

A. Fixed-Tree Relaxation-Based Algorithm
In this section, we first consider a simple low-complexity algorithm by decoupling the estimation process and the routing structure controlling the communication cost. Firstly, we generate an SPT-CC rooted at the sink node that provides the identity of the parent of each sensor, we store as the set \{\( k, p_k \)\}. Then we re-write the optimization problem for only the sensor selection so that the routing structure used for these sensors will be a subtree of the SPT-CC that is also rooted at the sink node. We call this fixed-tree relaxation-based algorithm. Considering the idea of an EF, a relaxed version of problem (5) is given by:

\[
\begin{align*}
\text{minimize} & \quad MSE_{\hat{\theta}} = \left( \frac{\sum_{k=1}^{N} \hat{b}_k h_k^2}{\sigma^2} \right)^{-1} \\
\text{subject to} & \quad \sum_{k=1}^{N} \hat{b}_k f_c(d_{k,p_k}) = P_{\text{max}} - \Delta \\
& \quad \hat{b}_k \leq b_{p_k}; \quad \Delta \geq 0 \\
& \quad 0 \leq \hat{b}_k \leq 1
\end{align*}
\]  
(6)
where \( \hat{b}_k \) is the relaxed version of the variable \( b_k \). Let us denote \{\( b_k^* \)\}_{k=1}^{N} a solution of this relaxed problem.

Notice that problem (6) is a convex problem, but it is not equivalent to the original problem (5) since \{\( b_k^* \)\}_{k=1}^{N} will not be binary in general. We use the solution of problem (6) to perform a suboptimal subset selection \( \hat{S} \) by sorting the optimal values \{\( b_k^* \)\}_{k=1}^{N} in descending order and selecting the subset of \( K \) largest \( b_k^* \)'s satisfying the power constraint as long as \( \Delta \geq 0 \). Then, denoting \{\( \hat{b}_k \)\}_{k=1}^{N} a set of binary values such that \( \hat{b}_k = 1 \) if \( k \in \hat{S} \) and \( \hat{b}_k = 0 \) if \( k \notin \hat{S} \), we have that:

\[ L_{f \times d} = \frac{\sum_{k=1}^{N} \hat{b}_k h_k^2}{\sigma^2} \geq p^* \]  
(7)
When making the sorting, because of the constraint \( \hat{b}_k \leq b_{p_k} \), it is a tree and because of the relaxation the routing solution will be a subset of SPT-CC.

In our network setting, sink node is not allowed to take the measurement of the event, that is, \h_{N+1} = 0, then the lower
bound is obtain by solving this problem:

\[
\begin{align*}
\text{minimize} & \quad MSE_\delta = \left(\sum_{k=1}^{N+1} b_k^* h_k^2 / \sigma^2 \right)^{-1} \\
\text{subject to} & \quad \sum_{k=1}^{N+1} b_k^* f_c(d_k, p_k) = P_{\text{max}} - \Delta \\
& \quad b_k^* = 1; b_k \leq b_{\text{p}} \\
& \quad \Delta \geq 0; 0 \leq b_k^* \leq 1
\end{align*}
\]  

(8)

We use the solution \( b_k^* \) of the problem (8) to obtain an optimal routing structure that has to be routed at the sink node. Consider the minimum spanning tree of the directed graph \( G' = (V, E') \), where it is assumed that the directed edge between sensor \( i \) and \( j \) has a cost \( c_{ij} = b_j^* f_c(d_{ij}) \) and similarly \( c_{ji} = b_j^* f_c(d_{ij}) \) for the other direction from sensor \( j \) to \( i \). Later, we update the routing structure based on the condition that if \( c_{ij} < c_{ji} \) then sensor \( j = p_i \) will be the parent of sensor \( i \) otherwise, \( i = p_j \) will be the parent of sensor \( j \). The constraint \( b_{N+1} = 1 \) in (8) enforces this tree to be routed at the sink node. Finally, we store the new set \( \{k, p_k\} \) based on the condition at each sensor pair, then calculate the total communication cost of this routing structure and the lower bound \( L \) that is given by the value of the objective in (8), where \( b_{N+1} = 0 \), thus, the lower bound \( L \) is given by:

\[
L = \left(\sum_{k=1}^{N+1} b_k^* h_k^2 / \sigma^2 \right)^{-1} \leq p^*  
\]  

(9)

By observing the gap \( \delta_{f,x} = L_{f,x} - L \), we can have an assessment on how good this suboptimal approximation is.

We can verify this lower bound \( L \) using Newton’s method by solving our relaxed problem (6) approximately using the log barrier method ([12], Sec. 11.2) taking care of the new constraint (10) which depends on \( \nu \).

\[
\begin{align*}
\text{minimize} & \quad \phi(b^*) = \left(\sum_{k=1}^{N+1} b_k^* h_k^2 / \sigma^2 \right)^{-1} - \\
& \quad - \frac{1}{\nu} \sum_{k=1}^{N} \left(\log(b_k^*) + \log(1 - b_k^*) + \log(b_{p_k}^* - b_k^*) + \log(\Delta) \right) \\
\text{subject to} & \quad \sum_{k=1}^{N} b_k^* f_c(d_k, p_k) = P_{\text{max}} - \Delta \\
\end{align*}
\]  

(10)

where \( \nu > 0 \) is a parameter that sets the accuracy of the approximation. The function \( \phi \) is convex and smooth, therefore, problem (10) can be efficiently solved by the Newton method. Let \( p^* \) denote the solution of problem (10), which depends on \( \nu \).

A standard result in interior-point methods (this can be seen in [12], Sec. 11.2.2) is that \( p^* \) is no more than \( 2N/\nu \) suboptimal for the problem (6), that is:

\[
\left(\sum_{k=1}^{N+1} b_k^* h_k^2 / \sigma^2 \right)^{-1} \leq \left(\sum_{k=1}^{N+1} b_k^* h_k^2 / \sigma^2 \right)^{-1} + 2N / \nu = L + 2N / \nu
\]

which implies:

\[
L' = \left(\sum_{k=1}^{N+1} b_k^* h_k^2 / \sigma^2 \right)^{-1} - 2N / \nu \leq p^* 
\]  

(11)

and as \( \nu \to \infty \), we get the lower bound \( L \). We can use this bound to choose \( \nu \) so that increase in gap contributed by term \( 2N/\nu \) is small.

We now describe Newton’s method for solving our problem (10), (see [12], Sec. 10.2 for more details). We take \( \text{diag}(b^*) f_c = (P_{\text{max}} / N) \tilde{f}_c \) as an initial (feasible) point, and the definition of Newton’s search step \( \Delta b_{\text{nn}}^* \) is modified at each step to take the equality constraints into account, which can be expressed as:

\[
\Delta b_{\text{nn}}^* = -\left(\nabla^2 \phi\right)^{-1} \nabla \phi + \left(\frac{\tilde{f}_c}{f_c} \nabla^2 \phi\right)^{-1} \nabla \phi \left(\nabla^2 \phi\right)^{-1} f_c 
\]

(12)

where \( \nabla \phi \) and \( \nabla^2 \phi \) are the gradient and Hessian of the function \( \phi \), respectively. We then use a backtracking line search to choose a step size \( \delta \in (0, 1] \), and update \( b^* \) by \( b^* := b^* + \delta \Delta b_{\text{nn}}^* \). We stop when the Newton decrement \( -\nabla \phi(b^*)^T \Delta b_{\text{nn}}^* \) is small, \( \delta \in (0, 1] \), where \( \delta \) is a small tolerance. The total number of steps required is typically ten or fewer in our problem.

For completeness, we give expressions for the derivatives of \( \phi \). Its gradient \( \nabla \phi \) is given by:

\[
(\nabla \phi)_k = -\left(\frac{b_k^* h_k^2}{\sigma^2}\right)^{-1} - \frac{1}{\nu} \left(\frac{1}{b_k^*} - \frac{1}{1 - b_k^*}\right)
\]

(13)

The Hessian \( \nabla^2 \phi \) is given by:

\[
\nabla^2 \phi = 2 \left(\frac{\sum_{k=1}^{N} b_k^* h_k^2}{\sigma^2}\right)^{-1} + \frac{1}{\nu} \text{diag} \left(\frac{1}{(b_k^*)^2} + \frac{1}{(1 - b_k^*)^2} \right) + \cdots + \frac{1}{(b_{N}^*)^2} + \frac{1}{(1 - b_{N}^*)^2}  
\]

(14)

B. Iterative Distributed Algorithm

The main disadvantages of the fixed-tree relaxation-based algorithm are the following: (a) this algorithm assumes that the SPT-CC is the optimal routing structure from which sensors are selected, thus ignoring the interplay between communication cost and estimation; (b) the sensor selection is optimized in a centralized manner, thus, making it less convenient to be scalable in WSNs. As we show in this paper, the SPT-CC is not in general the optimal transmission structure because of the interplay between communication cost and estimation gain obtained when fusing measurements from different sensors. In other words, each routing decision affects both the communication cost and estimation gain, as already illustrated in the simple example of Fig. 1.

In this section, we present an scalable iterative distributed algorithm that performs jointly the sensor selection and routing with the possibility to trade-off the metrics of distortion and communication cost. In this case, there is not a pre-selected structured routing tree and our algorithm should iterate to select both sensors and routes jointly. Because of the limited look-ahead ability, our algorithm will provide an overall sub-optimal solution to our original problem (5), however, since it takes into account both metrics, as we will show below, this algorithm provides better results than the fixed-tree relaxation-based and the algorithm ID in [2].
The idea of the algorithm is as follows. The whole process starts from the sensor that first detects the phenomenon (on an average) with the largest SNR, equivalently, the largest \( h_j \) (sensor \( a \) in Fig. 3(a) and it is close to target). Notice that the most power efficient way to send this measurement to the sink node is using the corresponding shortest path on the SPT-CC. This motivates us to select this single path as an initialization of our algorithm, forming a backbone (black thick path in Fig. 3(a) from sensor \( a \) to sink \( S \)). Thus, we initially select all the intermediate sensors in this path and their associated routes from the SPT-CC (as illustrated in Fig. 3(a), intermediate sensors \( b \) and \( c \)). Next, each of the currently selected sensors calculates locally an objective function \( h_j^{-2} + \gamma_j f_c(d_{ij}) \) for all its 1-hop neighbors in the original connectivity graph (illustrated by dash links in Fig. 3(a)) and stores the identity of the neighbor that minimizes its associated objective function (such as neighbors 1, 3, and 8 as illustrated in Fig. 3(b)). Finally, we activate the best neighbor that minimize this objective function among all 1-hop neighbors (in this case, sensor 1, see Fig. 3(c)), where communication between sensors can be done by message passing. The local objective function \( h_j^{-2} + \gamma_j f_c(d_{ij}) \) is a combination of the estimation gain resulting from the corresponding inactive sensor and the communication cost incurred when selecting that sensor. While performing the selection of a new sensor, we also check whether an alternative route through the new selected sensor is more power efficient or not. We denote this operation as backtracking (see Fig. 4 and Algorithm 1), which improves the quality of estimation.

![Fig. 3. Illustration of the choice of neighbor for sensors \( \{a, b, c\} \in i \) and \( j \in N(i) \) represents that sensor \( i \) has \( j \) neighbors. Here, the dark sensors and thick black routes in the left and middle figures denote the initially chosen backbone path, that is chosen previously from the SPT-CC.](image)

![Fig. 4. Successful backtracking operation performed by the algorithm, where \( p_i \) is the parent of sensor \( i \).](image)

### Algorithm 1: Iterative Distributed Algorithm

**Require:** \( \hat{S}_\text{iter}, T, P, \) and \( P_{\text{max}} \)

- **T:** multi-hop path from target \( t \) to sink \( S \) (backbone)
- \( \hat{S}_\text{iter} \): sensors in \( T \)
- \( P \): total communication cost of \( T \), initially, \( \gamma_j = 1 \)

**while** \( P < P_{\text{max}} \) **do**

- Let \( A_{i,j} \) : sensor \( i \) has \( j \) neighbors \( (j \in N(i)) \)
  
  \[ \{i,j\} = \arg\min_{\{i\in\hat{S}_\text{iter}, j\in A_{i,j}\}} (h_j^{-2} + \gamma_j f_c(d_{ij})) \]
  
  \( T = T \cup (i,j) \), \( \hat{S}_\text{iter} = \hat{S}_\text{iter} \cup \{j\} \)

  \( P_{\text{tot}K} = P; P_{\text{tot}K} = P_{\text{tot}K} + f_c(d_{ij}); P = P_{\text{tot}K} \)

  **if** \( (f_c(d_{jp_i}) < f_c(d_{ip_i})) \) **then**
  
  \( P_{\text{tot}K} = P_{\text{tot}K} - f_c(d_{ip_i}) + f_c(d_{jp_i}) \)

  \( p_j = p_i, \ p_i = j; P = P_{\text{tot}K} \)
  
  **end if**

  **- Backtracking:**

  **if** \( h_j < h_{j_{\text{pre}}} \) **then** Decrease \( \gamma_{j_{\text{next}}} \)
  
  **else** Increase \( \gamma_{j_{\text{next}}} \)
  
  **end if**

  **else**

  \( q = (P_{\text{tot}K} - P_{\text{max}}); T = (T \cap (i,j)) \); \( \hat{S}_\text{iter} = (\hat{S}_\text{iter} \cap \{j\}) \]

  \( P = P_{\text{tot}K} - r; r = r + 1; \gamma_{j_{\text{next}}} = \gamma_j \times r/q \)
  
  **end if**

**end while**

For a given sensor these two metrics (estimation gain and communication cost) are weighted by a factor \( \gamma_j \) to perform the suitable trade-off locally at each sensor. We start with the value of \( \gamma_j = 1 \) and then, we update \( \gamma_j \) that determines the relative importance of the estimation gain and the communica-

...
when the choice of a sensor causes the new total measured power ($P_{\text{tot},k}$) to exceed the power $P_{\text{max}}$, then $\gamma_j$ is updated differently: first, we do not choose that sensor and do not add the associated communication cost, then we calculate the current power gap ($P_{\text{tot},k} - P_{\text{max}}$) and iterate our algorithm to bring this gap as close to zero as possible. Next, by searching for the sensor that needs less communication cost, thus, we place a higher weight on the communication cost. For this, we simply increase $\gamma_{next}$ by a factor proportional to the number of iterations and inversely proportional to the gap (so that if we have gap very small, we can provide even higher weight on communication cost), till we find the best additional sensors to close this gap. The formal description of the process is provided in Algorithm 1. Our iterative distributed algorithm selects a subset of sensors $\tilde{S}_{\text{iter}}$, which will also produce a solution $L_{\text{iter}} \geq P^*$, thus we can define an associated gap, which will be given by $\delta_{\text{iter}} = L_{\text{iter}} - L$.

C. Computational Complexity Analysis

Next, we provide a complexity analysis for the algorithms we described previously.

1) Fixed-Tree Relaxation-Based Algorithm: The main computational complexity in the fixed-tree relaxation-based algorithm comprises the following parts: SPT-CC from Bellman-Ford algorithm that runs in $O(MN)$ (where $M$ is the number of edges), solving the relaxed optimization problem requires $O(N^3)$ operations (using interior-point methods [12]), since this problem is convex (objective is to be maximized with convex constraints) at the sink node. These methods typically require a few tens of iterations and each iteration can be carried out with a complexity of $O(N^3)$ operations.

2) Iterative Distributed Algorithm: Our iterative distributed algorithm performs the following operations: (a) calculating SPT-CC to generate a backbone from $t$ to sink $S$ takes $O(MN)$ operations, (b) there is an outer loop that runs for all $N$ sensors and within this loop, there is a subset selection operation controlled by the given power budget that chooses $K$ sensors, which takes in total $O(KN^2 \log N)$ operations.

Thus, both algorithm have similar order of complexity, however the fixed-tree relaxation algorithm is not fully distributed and in addition, as we show in the next section, its performance degrades because it ignores the coupling between routing and sensor selection for estimation.

IV. SIMULATION RESULTS

In this section, we show the performance comparison of the proposed algorithms through numerical simulations. For the sake of consistency and ease in comparing the results, we consider only a fixed target location. We consider a WSN with $N = 200$ sensors deployed randomly in a square region. For the communication cost between sensors $i$ and $j$, we assume $\alpha = 4$ in our model $f_c(d_{ij}) \propto d_{ij}^\alpha N_i S_i B$. The measurement gain in our example is assumed $h_i = 1/d_i^\beta$ with $\beta = 1$. We tested our algorithms using 100 different network topologies with the sink node located at the center of the region. For each network topology, we have run algorithms for a range of maximum power constraints values $P_{\text{max}}$.

Fig. 5. Power gaps: $\Delta_{\text{fed}} = P_{\text{max}} - P_{\text{tot},K}$ in fixed-tree relaxation; $\Delta_{\text{iter}} = P_{\text{max}} - P_{\text{tot},K}$ in iterative distributed.

In order to show the nonoptimality of the fixed-tree relaxation-based algorithm, we calculate the average power gaps $\Delta = P_{\text{max}} - P_{\text{tot},K}$ for both algorithms. We calculate the average gap using the estimated mean of the gaps obtained for all 100 topologies. Fig. 5 shows a calculation of the substantial gap that results from using the fixed-tree relaxation-based algorithm. It can be seen also that in the case of our iterative distributed algorithm the resulting gap is small.

Fig. 6. The subset of activated sensors and associated routing structure routed at the sink node for a power budget $P_{\text{max}} = 10$: (a) fixed-tree relaxation algorithm; (b) iterative distributed algorithm.

Fig. 7. Comparison of the proposed and related algorithm with the lower bound $L$ given by (9).

The subset of activated sensors and associated routing structure routed at the sink node for a power budget $P_{\text{max}} = 10$ are shown in Fig. 6(a) and Fig. 6(b). It can be seen by some arrows in Fig. 6(a) that the transmission links that are in fixed-
tree relaxation based algorithm consume more power, which are not present in iterative distributed algorithm, as a result our iterative distributed algorithm provides better estimation in distortion by activating some other sensors with the same power budget.

In order to show the performance of any algorithm obtained when selecting a subset of sensors and a certain routing structure, we evaluate the lower bound $L$ (given by (9)). The bound $L_{f_{xed}}$ is obtained using the simple sorting selection rule (given by (7)), and a better bound $L_{iter}$ (produced by Algorithm 1). Fig. 7 shows $L_{ID}$, $L_{f_{xed}}$, $L_{iter}$, and $L$, where $L_{ID}$ is the solution obtained from the Innovation Diffusion (ID) based algorithm [2]. Notice that the solution $L_{ID}$ is based on an EF approach but optimizing independently both metrics without any trade-off. Fig. 7 shows the performance in MSE for different values of the maximum power budgets $P_{max}$ with the sink located at the center. Our iterative distributed algorithm outperforms the fixed-tree relaxation-based and the algorithm ID as compared to the lower bound $L$, depending on the available power budget. We also demonstrate the other choice of $\gamma_j$: the choice with the $\gamma_j >> 1$, that is, very high weight on communication cost, therefore, sensors that require low communication cost to transmit information to their parents are given more priority to activate, even-though some of them are far from the target. On the other hand choice with the $\gamma_j << 1$, that is, very low weight on communication cost, which always forces to activate the sensors that are close to target avoiding communication cost metric. In both the cases, we obtain very poor distortion in estimation as can be seen in the Fig. 7. The best case is with the adaptive $\gamma_j$ as explained in the Algorithm 1.

![Fig. 8. Gaps; $\delta_{ID}$, $\delta_{f_{xed}}$, and $\delta_{iter}$.](image)

The gaps with the lower bound for the algorithm ID and for the proposed algorithms are presented in Fig. 8. These plots show that our iterative distributed algorithm performs better than both the fixed-tree relaxation-based and the algorithm ID for each power budget.

As shown by our results above, the routing tree based on the SPT-CC routing structure is not optimal in general. In many situations, we see a notable gap between the maximum power allowed and the total incurred power when using fixed-tree relaxation-based algorithm.

In Fig. 9, we illustrate an example how distortion behaves by adding more and more sensors in the fixed-size field of estimation and fixed power budget $P_{max} = 10$. We assume random deployment of the sensors and notice that as new sensors are deployed the previous ones are kept.

![Fig. 9. The MSE estimation performance due to a subset of sensors activated from different network size scenarios (100 sensors to 400 sensors) for a given fixed power budget $P_{max} = 10$.](image)

V. CONCLUSION

This paper considers the estimation problem in a scenario where the WSN covers a large geographical area, and is monitoring a localized event. In the recent literature, some solutions have been proposed, but most of them try only to reduce the problem to only the cardinality of the selected sensor set, ignoring the optimization of the network topology (routing). However, network topology is indeed an important factor in the problem since in general, the cost of receiving one measurement that is far from the fusion center consumes more energy than one that is closer to it.

APPENDIX

Theorem 1. The joint optimization problem (5) of sensor selection and multi-hop routing structure for distributed estimation under a total power constraint, is NP-Hard.

Proof: We prove the complexity of our problem using a reduction technique [16]. Basically, a problem $L$ can be reduced to another problem $L'$ ($L \leq_p L'$) if any instance of $L$ can be converted to an instance of $L'$ (by a mapping that is computable in polynomial time), so that the solution to the instance of $L'$ allows to solve the instance of $L$. Thus, if $L$ reduces to another problem $L'$, then $L'$ is as hard to solve than $L$.

In general, the joint sensor selection and routing problem belongs to a class of combinatorial optimization problems [17] and is typically NP-Hard. Moreover, minimizing $(\sum_{k=1}^n b_k h_k^2)^{-1}$ is equivalent to:

\[
\begin{align*}
\min \log \left( \sum_{k=1}^n b_k h_k^2 \right)^{-1} &= \min \left( -\log \sum_{k=1}^n b_k h_k^2 \right) \\
\iff \max \log \sum_{k=1}^n b_k h_k^2 &= \max \sum_{k=1}^n b_k h_k^2 \\
\iff \min \left( -\sum_{k=1}^n b_k h_k^2 \right)
\end{align*}
\]

Again, for simplicity, we assume that our problem is to find a single-path in our graph of $n$ sensors (as shown in Fig. 10) and we perform the reduction from another well-known NP-Hard problem to this problem, which is a simplification of our problem (5). Notice that a communication cost can be expressed in terms of the set of flows $f_{ij} \in \{0, 1\}$ connecting
sensors $i$ and $j$, as it is usual in network flow problems. Considering these facts, one can use a scalarization of the two objectives (instead of putting one as a constraint) to form an equivalent overall weighted objective, which can be written as:

$$\text{minimize} \quad \sum_{k=1}^{n} b_k h_k^2 + \omega \sum_{(i,j) \in E} f_{ij} f_c(d_{ij})$$  \hspace{1cm} (15)$$

where $\omega > 0$ is a global scalarization parameter that controls the importance of both the metrics in (15). Notice that, for each power budget constraint $\sum_{(i,j) \in E} f_{ij} f_c(d_{ij})$ the value of $\omega$ changes as a function of $P_{\text{max}}$. Our goal is to find an optimal subset of sensors that includes closest sensor $n$ to the target, which is equivalent to a path between the sensor $n$ and the sink node $S$ in our graph. The first term in the objective function for a subset of $K$ sensors can be re-written as (see Fig. 10):

$$\sum_{k=p}^{n} b_k h_k^2 = b_p h_p^2 + b_{p+1} h_{p+1}^2 + \cdots + b_{n-1} h_{n-1}^2 + b_n h_n^2$$

$$\sum_{k=p}^{n} b_k h_k^2 = f_{qp} h_p^2 + f_{(q-1)p} h_q^2 + \cdots + f_{n(p-1)} h_{n-1}^2 + b_n h_n^2$$

then, the objective (15) for a subset of $K$ sensors becomes:

$$\text{minimize} \quad \sum_{(i,j) \in E'} f_{ij} (\omega f_c(d_{ij}) - h_i^2) - h_n^2$$  \hspace{1cm} (16)$$

where the number of edges in a subset $E' \subset E$ is $K$.

The objective (16) is independent of $b_k$ (no sensor selection), and it is equivalent to a Direct Hamiltonian Path (DHP) problem from the sensor $n$ to $S$ in a directed graph $G = (V, E)$ where the edges have a cost equal to $f_{ij}(d_{ij}) = \omega f_c(d_{ij}) - h_i^2$. Notice that, when we weight less the communication cost term (i.e. a low $\omega$), we can create negative cycles in the equivalent graph and the standard DHP problem becomes unbounded (infinite flow on negative weights). Interestingly, this problem is much harder than the usual DHP problem when there are no negative cycles, where for instance Karp’s algorithm [18] can be applied (polynomial time). In fact, it can be shown that the general DHP with negative weights is NP-Hard [19].

**Proof of NP-Hard:** Let us consider activating sensor (say $i$ in Fig. 10) then we prove the NP-Hardness of our problem by showing that the problem of generating a single path from $i$ to $S$, namely the $(i,S)$ DHP is NP-Hard. In order to do so, we show that the DHP problem, which is known to be NP-complete [16], reduces to $(i,S)$ DHP. We take an instance of the DHP problem, that is an undirected symmetric graph $G = (V, E)$ as shown in Fig. 10 and add to it an extra vertex $i$, which is connected to every vertex $v \in V$, adding the corresponding edge, that is, we add an extra edge $(i,v)$. This will generate another undirected symmetric graph $G_i = (V_i, E_i)$. If there is a Hamiltonian path in $G_i$, there will be an edge for $i$ to the one of the vertex of $G$, which will ensure that the new Hamiltonian path starts at $i$ and ends at $S$ covering each vertex (in the selected subset) of $G$. Therefore, it ensures that there is a Hamiltonian path $(i,S)$ in $G_i$, Conversely, if there is a Hamiltonian path $(i,S)$ in $G_i$, the vertex excluding $i$ will also form a Hamiltonian path in $G$. Since this transformation (adding extra vertex and edge) can be done in polynomial time, we have that $DHP \leq_P (i,S) DHP$. Moreover, since $(i,S) DHP \in NP$, we can conclude that $(i,S) DHP$ is NP-complete and its optimization version is NP-Hard.

**REFERENCES**


