An Efficient Clustering and Path Planning Strategy for Data Collection in Sensor Networks Based on Space-Filling Curves

Yuan Yan and Yasamin Mostofi

Abstract— **In this paper, we consider a scenario where a mobile robot is tasked with periodically collecting data from a fixed wireless sensor network. Our goal is to minimize the total energy cost of the operation, including the communication cost from the sensors to the robot and the motion cost of the robot. We propose a strategy that properly combines the ideas of clustering and using a mobile robot for data collection. Our approach is based on using space-filling curves, which results in a computationally-efficient algorithm. It can furthermore handle realistic communication environments by utilizing probabilistic channel predictors that go beyond disk models. We mathematically characterize an upper bound for the performance of our proposed algorithm, which shows how the energy saving is related to the total number of generated bits in the network, and the communication and motion parameters. Finally, we verify the effectiveness of our proposed framework in a simulation environment, where a considerable reduction in energy consumption is achieved as compared to the case of no clustering.**

I. INTRODUCTION

In recent years, considerable progress has been made in the area of mobile sensor networks and networked robotic systems [1]–[7]. One application is to use mobile robots (mobile sinks) to harvest data from a wireless sensor network [7]–[11]. In such a scenario, the sensors can be scheduled to transmit information bits to a mobile robot when it gets closer to them, resulting in a smaller communication cost. Along this line, various approaches have been proposed for addressing different aspects of this problem, such as path planning and speed control of the mobile robots [7]–[9], and routing design of the network [11]. However, using mobile robots not only causes long latency for data collection, due to the limited speed of the robots, but also consumes motion energy.

In the wireless sensor network literature, clustering is another popular approach for energy saving [12], [13]. In such a framework, the sensors are grouped into a number of clusters with an elected cluster head to manage the data collection in its corresponding cluster. However, each cluster head needs to incur high energy to forward the gathered information bits to a remote station. It therefore makes sense to utilize both clustering and mobility control for data collection in sensor networks.

In [14]–[18], the ideas of clustering and using mobile robots for data collection are combined. More specifically, in [14], both communication cost of the sensors and motion cost of the robot are considered. However, the trajectory of the robot is assumed to be fixed. In [15]–[17], piecewise linear trajectories are designed to reduce the communication cost of the sensors. However, the motion cost of the robot is not considered. Moreover, the aforementioned works do not consider realistic fading communication environments and do not provide theoretical analysis or performance guarantees.

In this paper, we consider a scenario where a mobile robot is tasked with periodically collecting up-to-date data from a wireless sensor network. Our goal is to minimize the total energy cost of the operation, including the communication cost from the sensors to the robot and the motion cost of the robot. We propose a strategy that properly combines the ideas of clustering and using a mobile robot for data collection. In such a strategy, the robot first groups the sensors into a number of clusters. A stop position is then chosen in each cluster for the robot to collect the data from the sensors in the corresponding cluster. The robot then periodically visits all the stop positions and gathers the data from the network. In this paper, we consider a single-hop transmission strategy from the sensors to the robot. Moreover, we assume communication over realistic fading links and utilize our previously-proposed probabilistic channel assessment framework for path planning [19], [20].

The main challenge of our considered scenario is the need to jointly design the clustering strategy of the sensor network, as well as the stop positions and path of the robot, which can be computationally prohibitive. We then propose an approach based on utilizing space-filling curves [21], [22] to design our algorithm. A space-filling curve is a one-dimensional curve that passes through all the points of a 2D square. By utilizing space-filling curves and their locality property, we show how the coupled clustering, stop position selection and path planning problems can be solved efficiently. Moreover, we characterize an upper bound for our proposed algorithm which shows the dependence on the total number of generated bits in the network, as well as the communication and motion parameters. Finally, our results indicate a considerable energy saving as compared to the case of no clustering.

The rest of the paper is organized as follows. Section II describes the motion and communication models, and introduces the basic properties and applications of the space-filling curves. Section III presents our proposed approach and mathematically characterizes its upper bound. Section IV shows the performance of the proposed approach in a

Yuan Yan and Yasamin Mostofi are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, USA (email:{yuanyan,ymostofi}@ece.ucsb.edu).

This work is supported in part by NSF NeTS award # 1321171.

simulation environment. We conclude in Section V.

II. Problem Setup

We consider a scenario where a mobile robot (mobile sink) is tasked with periodically collecting the data from a fixed wireless sensor network. The robot does not know the exact positions of the sensors, but only the probability distribution of their positions. We assume that there are a total of m stationary sensors uniformly distributed in a square workspace $\mathcal{W}$ with the side length of D. Each sensor collects data from the environment with a data rate of $\widetilde{\rho}$ bits/second. A mobile robot is then tasked with gathering the up-to-date information bits from all the sensors in a given period T. Our goal is to minimize the total energy cost of the whole operation in each period, including the communication energy cost of the sensors and the motion cost of the robot. Clearly, the motion cost is minimized if the robot stays at the center of the workspace while the sensors transmit their gathered bits to it. However, such a strategy is not energy efficient for communication. On the other hand, the communication cost can be minimized if the robot visits each sensor to download the data. This strategy, however, not only causes a high motion cost, but also possibly violates the given time period constraint.

In this paper, we thus propose a strategy that properly combines the ideas of clustering and robotic path planning for data collection in a wireless sensor network. In such a strategy, all the sensors are divided into a number of clusters. A stop position is chosen in each cluster for the robot to collect the data (via single-hop wireless transmissions) from the sensors in the corresponding cluster. The robot then periodically visits all the stop positions and gathers the data from the network. Fig. 1 shows our considered scenario. In general, the motion cost increases as the number of clusters increases since the robot needs to travel a longer distance in each period (the motion cost is proportional to the length of travel, as we shall discuss in Section II-C). On the other hand, the communication cost decreases as the number of clusters increases, since the transmission distances of the sensors decrease. Hence, in this problem, the robot needs to jointly optimize the clustering of the network, its stop positions and the order in which it visits the stop positions, in order to minimize the total energy cost.

Note that even if all the stop positions are given beforehand, the path planning problem of the robot becomes the classic Traveling Salesman Problem (TSP), which is NP-hard. Therefore, in this paper, we focus on designing computationally-efficient algorithms to solve this problem. We further show how the robot can probabilistically assess the communication cost by using realistic channel models that go beyond the widely-utilized disk models.

In the rest of this section, we first introduce the concept of space-filling curves which will be used in the subsequent sections for the algorithm design and performance analysis. Then, we briefly present the communication and motion models that are used in this paper.

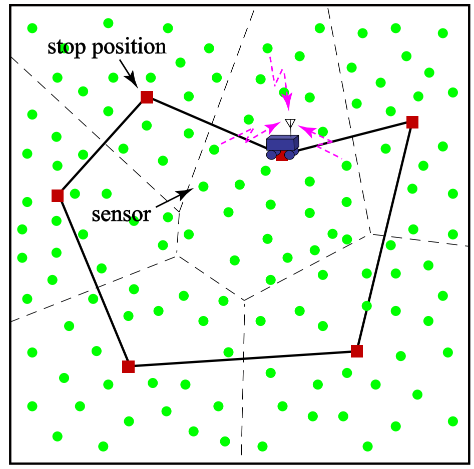

Fig. 1. An example of using a mobile robot to gather the data from a sensor network. The sensors are divided into 6 clusters. In each period, the robot visits all the stop positions to collect the data from the sensors in the corresponding clusters.

A. *Space-Filling Curves [21], [22]*

A space-filling curve is a one-dimensional curve that passes through every point of a two-dimensional square. Some of the most celebrated ones are the Hilbert curve, Peano curve and Sierpiński curve. Readers are referred to [21] for more details on various space-filling curves and how to recursively construct them.

Because of their recursive and self-similar construction, one of the most important properties of space-filling curves is the *locality* property, which means that any two points that are close in the one-dimensional space are mapped to two points that are close in the 2D space. More specifically, let $\text{SF}(\cdot) : \mathcal{C} \rightarrow \mathcal{W}$ denote the continuous mapping of a space-filling curve from the unit circle to the $[0, D]^2$ square workspace, where $\mathcal{C} = \{\phi \mid 0 \leq \phi \leq 1\}$ represents the unit circle. Then, we have [21], [23]

$$\|q_1 - q_2\| \leq C_{\text{SF}} \times D \times \mu(\phi_1, \phi_2), \tag{1}$$

where $\phi_1, \phi_2 \in \mathcal{C}$, $q_1 = \text{SF}(\phi_1) \in \mathcal{W}$, $q_2 = \text{SF}(\phi_2) \in \mathcal{W}$, C_{SF} is a constant depending on the type of the space-filling curve, $\mu(\phi_1, \phi_2) = \min\{|\phi_1 - \phi_2|, 1 - |\phi_1 - \phi_2|\}^{1/2}$, and $\|\cdot\|$ and $|\cdot|$ denote the Euclidean norm and absolute value of the argument respectively. See Fig. 2 for an illustration of the mapping. For Sierpiński curves, for instance, $C_{\text{SF}} = 2$ [21], [23].

Due to this *locality* property, space-filling curves are widely used in computational science [21]. The application that is most related to this paper is to solve the Traveling Salesman Problem (TSP) as follows [23], [24]. First, map all the points in the square into the unit circle using $\text{SF}^{-1}(\cdot)$. Then, order the mapped points on the unit circle in a clockwise or counter-clockwise direction. Finally, build the tour by connecting the points in the square based on this order. Fig. 2 shows an example of solving the TSP problem by using this approach. When the points are i.i.d. and the number of points becomes large, the heuristic tour is roughly 35% away from the optimum [24]. However, the computational complexity of this approach is extremely low. In [22], we showed how to utilize the space-filling curves to solve a communication-aware dynamic coverage problem. In this paper, we utilize them for clustering and data collection in a sensor network.

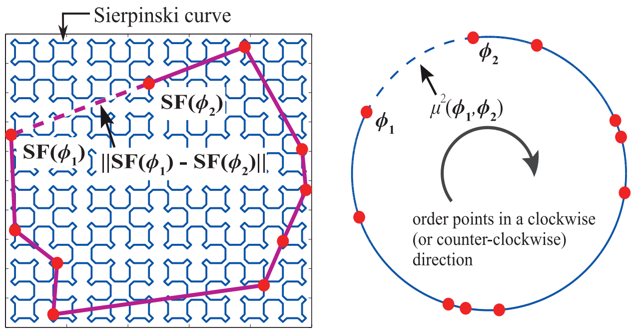

Fig. 2. The blue curve in the left figure shows the construction of the Sierpiński curve after four recursions in the 2D workspace. The right figure then shows the corresponding 1D curve (circle). The red dots in the left and right figures show the positions of the points in the square and their corresponding mapping in the unit circle respectively. The figure then shows an illustration of the mapping SF($\cdot$) and $\mu(\cdot,\cdot)$. The figure also shows how this mapping can be used to solve a TSP problem, where the line segments that connect the points in the square form the tour obtained based on the ordering of the points in the unit circle [22].

B. Spectral Efficiency and Communication Energy Model

In the communication literature, Bit Error Rate (BER) is the metric that is used to characterize the communication reception quality. BER characterizes how many bits are flipped (received in error) and is a function of Channel to Noise Ratio (CNR), transmission rate and transmit power. More specifically, the communication quality in the transmission to the robot is acceptable if the BER is below a given threshold. Assume that the commonly-used MQAM modulation is used for communication [25]. We then have the following approximated expression for BER [25]: $p_b \approx 0.2\exp\big(-1.5\widetilde{P}_{\text{C}}\gamma(p,q)/(2^R-1)\big)$, where p_b is the BER, $\widetilde{P}_{\text{C}}$ denotes the communication transmit power, R represents the spectral efficiency (transmission rate divided by the given bandwidth), and $\gamma(p,q)$ is the received CNR in the transmission from the transmitter at position p to the receiver at position q. Then, given a target BER of $p_{b,\text{th}}$, the minimum required communication power for transmission with a spectral efficiency R can be characterized as $\widetilde{P}_{\text{C}}(p,q) = (2^R-1)/(K\gamma(p,q))$, where $K = -1.5/\ln(5p_{b,\text{th}})$. Then, the total communication energy cost of each sensor in a period T can be characterized as $\widetilde{E}_{\text{C}}(p,q) = \widetilde{P}_{\text{C}}(p,q)\rho T/R$, where $\rho = \widetilde{\rho}/B$, B is the given communication bandwidth, and $\rho T/R$ is the required transmission time from a sensor in a period T. We take $\rho \leq R$, i.e. the sensing rate is assumed less than or equal to the given communication rate. This is a natural assumption since the task will not be feasible otherwise.

In practice, the true value of the received CNR ($\gamma(p,q)$) may not be available for all the p, q pairs. Then, the robot can use our previously-proposed probabilistic channel assessment framework to estimate the CNR, based on a small number of a priori sample measurements in the same environment [19], [20]. Since we are assessing the channel quality probabilistically, the anticipated communication energy cost $\widetilde{E}_{\text{C}}(p,q)$ becomes a random variable as well. We then take the average of $\widetilde{E}_{\text{C}}(p,q)$ (over the predicted channel) as our communication cost:

$$E_{\text{C}}(p,q) = P_{\text{C}}(p,q)\frac{\rho T}{R} = \frac{2^R-1}{K}\mathbb{E}\left\{\frac{1}{\gamma(p,q)}\right\}\frac{\rho T}{R}, \quad (2)$$

where $P_{\text{C}}(p,q) = \big((2^R-1)/K\big)\mathbb{E}\{1/\gamma(p,q)\}$. Generally, $\gamma(p,q)$ can be modeled as a lognormal random variable [19], [20]. In this paper, we consider the case where the channel is spatially uncorrelated, which results in the following:

$$\mathbb{E}\left\{\frac{1}{\gamma(p,q)}\right\} = \exp\left(\left(\frac{\ln 10}{10}\right)^2\frac{\sigma^2_{\text{dB}}}{2}\right)\underbrace{\widetilde{\alpha}\|p-q\|^{n_{\text{PL}}}}_{\text{path loss}},$$
$$= \alpha\|p-q\|^{n_{\text{PL}}},$$

where n_{PL} is the path loss exponent (typically around $2 \sim 6$ [25]), σ^2_{dB} is the variance of the channel in the dB domain, $\widetilde{\alpha}$ is the path loss parameter and $\alpha = \exp\big((\ln 10/10)^2(\sigma^2_{\text{dB}}/2)\big)\,\widetilde{\alpha}$.[1] Hence, the average communication energy cost decreases as the path loss exponent decreases and/or the variance decreases. Moreover, for this uncorrelated channel model, it can be seen that it is always optimal for a sensor to transmit its bits to the closest stop position of the robot.

C. Motion Energy Model

Experimental studies suggest that the motion power cost of a mobile robot can be approximated by a polynomial function of its velocity [26]. In this paper, we use the following linear model to characterize the motion power cost of the robot [26]: $P_{\text{M}} = \begin{cases} \kappa_1 v + \kappa_2 & \text{if } 0 < v \leq v_{\max}, \\ 0 & \text{if } v = 0, \end{cases}$ where P_{M} is the motion power, v denotes the velocity of the robot, κ_1 and κ_2 are positive constants, and $v_{\max}$ is the maximum velocity of the robot. This model is a very good fit to the Pioneer 3DX robot, when the velocity is smaller than 0.9 m/s [26]. Then, the motion energy cost for traveling along a trajectory with length d can be found as follows: $E_{\text{M}} = \kappa_1 d + \kappa_2 t_{\text{mo}}$, where E_{M} is the motion energy cost and $t_{\text{mo}} \geq d/v_{\max}$ is the total motion time spent along this trajectory. Note that E_{M} is minimized when $t_{\text{mo}} = d/v_{\max}$, i.e. when the robot travels with its maximum speed. In this case, the motion energy cost becomes $\kappa_{\text{M}} d$, where $\kappa_{\text{M}} = \kappa_1 + \kappa_2/v_{\max}$.

III. Joint Clustering and Path Planning for Data Collection

As mentioned previously, the main challenge of our considered problem is to jointly decide the number of clusters, the stop positions and the optimal tour to visit them. In particular, the stop positions affect not only the communication cost of the sensors (since each sensor transmits to its closest stop position), but also the motion cost of the robot. Even if the stop positions are given, finding the optimal tour is still NP-hard. However, this difficulty can be considerably

[1] The underlying channel parameters can be estimated as shown in [19], [20].

reduced if we map the 2D environment into the 1D space-filling curve domain. Based on the locality property, each sensor only needs to transmit to the stop position that is the closest in the space-filling curve domain. Then, the tour of the robot can be sub-optimally but very efficiently found. An upper bound on the length of the tour can also be characterized by using inequality (1). In this section, we show how to achieve these goals.

A. Our Proposed Approach Based on Space-Filling Curves

Let N denote the number of clusters. We first consider the case where there is no clustering, i.e. there is only one cluster (the whole workspace). This case will then serve as a benchmark for comparison. It can be easily seen that the optimal stop position is the center of the workspace in this case. As a result, the optimal total average energy can be found as

$$E^{\star}_{\text{tot}}(N=1) = \int_{-D/2}^{D/2}\int_{-D/2}^{D/2} \frac{\rho T \beta m}{D^2}\|q\|^{n_{\text{PL}}} dq$$
$$= \rho T \beta m D^{n_{\text{PL}}} \int_{-0.5}^{0.5}\int_{-0.5}^{0.5} \|q\|^{n_{\text{PL}}} dq = \rho T \beta m D^{n_{\text{PL}}} C_1, \quad (3)$$

where $\beta = (2^R - 1)/(KR)\alpha$, $C_1 = \int_{-0.5}^{0.5}\int_{-0.5}^{0.5}\|q\|^{n_{\text{PL}}}dq$, and $E^{\star}_{\text{tot}}(N=1)$ denotes the optimal total energy cost in the 2D workspace for the case of no clustering. Note that in this case, there is no motion cost for the robot.

Next, consider the case of N clusters, for $N \in \{2, 3, \cdots\}$ (optimization of N will follow shortly). We map the 2D workspace $\mathcal{W}$ to $\mathcal{C}$ by using $\text{SF}^{-1}(\cdot)$. Since the sensors are uniformly distributed in $\mathcal{W}$, they are also uniformly distributed in $\mathcal{C}$ [23]. We then divide $\mathcal{C}$ into N equal-sized arcs of the length $1/N$. Based on the locality property of the space-filling curves, we expect that the sensors that are mapped to the same arc are also close to each other in $\mathcal{W}$. Hence, we group the sensors that are mapped to the same arc as a cluster. We then choose the stop positions as the centers of the corresponding arcs.

By using (1) and (2), the total average communication energy cost in $\mathcal{W}$ can then be bounded from above as follows:

$$E_{\text{comm}}(N) \leq N m \rho T \int_0^{1/N} \beta \left(C_{\text{SF}} D \sqrt{\left|\phi - \frac{1}{2N}\right|} \right)^{n_{\text{PL}}} d\phi$$
$$= \frac{m\rho T \beta C_{\text{SF}}^{n_{\text{PL}}} D^{n_{\text{PL}}}}{n_{\text{PL}}/2 + 1} \left(\frac{1}{2N}\right)^{\frac{n_{\text{PL}}}{2}},$$

where $E_{\text{comm}}(N)$ denotes the total average communication energy cost in $\mathcal{W}$ by mapping our clustering and stop positions from 1D back to the 2D space.

Since all the stop positions are the centers of the corresponding arcs in 1D, the total motion energy cost in $\mathcal{W}$ can be bounded from above as follows: $E_{\text{mo}}(N) \leq N \times \kappa_{\text{M}} C_{\text{SF}} D \sqrt{1/N} = \kappa_{\text{M}} C_{\text{SF}} D \sqrt{N}$, where $E_{\text{mo}}(N)$ represents the motion energy cost in $\mathcal{W}$ by mapping our stop positions and path planning strategy from 1D back to the 2D space. Hence, we have the following upper bound for the total cost of the whole network:

$$E_{\text{tot}}(N) = E_{\text{comm}}(N) + \varpi E_{\text{mo}}(N) \leq E_{\text{tot,upper}}(N)$$
$$= \frac{m\rho T \beta C_{\text{SF}}^{n_{\text{PL}}} D^{n_{\text{PL}}}}{n_{\text{PL}}/2 + 1} \left(\frac{1}{2N}\right)^{\frac{n_{\text{PL}}}{2}} + \sqrt{N} C_{\text{SF}} \varpi \kappa_{\text{M}} D,$$
$$\text{for } N = \{2, 3, \cdots\}, \quad (4)$$

where $E_{\text{tot}}(N)$ is the total energy cost in the 2D workspace by mapping our clustering, stop positions and path planning strategy from 1D back to the 2D space, $E_{\text{tot,upper}}(N)$ denotes the corresponding upper bound based on the characterization in the space-filling curve domain, and $\varpi > 0$ is a weight balancing the importance of the communication cost of the sensor network and the motion cost of the robot. For the case of $N = 1$, we take $E_{\text{tot,upper}}(1) = E^{\star}_{\text{tot}}(1)$.

In this paper, we assume that the sensors in one cluster transmit their data to the robot on different frequency bands. Thus, the communication time for the robot to collect the information bits in each cluster is $\rho T/R$. As a result, we require $T \geq \rho T/R$, i.e. $\rho \leq R$, for the case of no clustering to guarantee the feasibility of the task, as we have assumed in Section II-B. For the case of $N \in \{2, 3, \cdots\}$, the total motion time budget for the robot is $\widetilde{T} = (1 - N\rho/R)T$. Then, to guarantee that the robot can collect all the data in the given period T, the total length of the path should be smaller than or equal to $\widetilde{T} v_{\max}$: $C_{\text{SF}} D \sqrt{N} \leq \widetilde{T} v_{\max}$. Hence, we have the following sufficient conditions:

$$\widetilde{T} = (1 - N\rho/R)T > 0,$$
$$\text{and} \quad (C_{\text{SF}} D)^2 N \leq (T v_{\max})^2 (1 - N\rho/R)^2.$$

After some straightforward calculations, we have the following upper bound for N:

$$N \leq N_{\max} = \max\left\{1, \left\lfloor \frac{R}{\rho} + \frac{1}{2}\left(\frac{R C_{\text{SF}} D}{\rho v_{\max} T}\right)^2 - \frac{1}{2}\left(\frac{R C_{\text{SF}} D}{\rho v_{\max} T}\right)^2 \sqrt{1 + \frac{4\rho}{R}\left(\frac{v_{\max} T}{C_{\text{SF}} D}\right)^2} \right\rfloor \right\}. \quad (5)$$

Note that if the second term in the maximization operator of (5) is smaller than or equal to 1, the robot can only choose $N = 1$ (no clustering) to guarantee the feasibility.

From (4), it can be seen that the total energy cost in the 2D workspace $E_{\text{tot}}(N)$ is guaranteed to be small if we minimize its upper bound $E_{\text{tot,upper}}(N)$. Next, we show some properties of $E_{\text{tot,upper}}(N)$ and find the optimal $N^{\star}$ to minimize it.

First, note that the communication and motion energy costs in $E_{\text{tot,upper}}(N)$ are monotonically decreasing and increasing with respect to N respectively. Hence, if the communication energy cost was zero, then we would have $N^{\star} = 1$ (i.e. no clustering). On the other hand, if the motion cost was zero (for instance, $\varpi = 0$), then we would have $N^{\star} = N_{\max}$, since it minimizes the upper bound of the total communication energy cost in the space-filling curve domain. For the case where both communication and motion energy costs are non-zero, we start with minimizing $E_{\text{tot,upper}}(N)$

for a continuous variable N_c, for $2 \leq N_\text{c} \leq N_{\max}$. The first-order derivative of $E_\text{tot,upper}(N_\text{c})$ with respect to N_c is as follows:

$$\begin{aligned}\frac{\partial E_\text{tot,upper}(N_c)}{\partial N_c} &= -\frac{m\rho T\beta C_\text{SF}^{n_\text{PL}} D^{n_\text{PL}} n_\text{PL}}{(n_\text{PL}/2+1)(2N_c)^{\frac{n_\text{PL}}{2}+1}} + \frac{C_\text{SF}\varpi\kappa_\text{M} D}{2\sqrt{N_c}} \\ &= \left(\frac{1}{N_c}\right)^{\frac{n_\text{PL}}{2}+1}\left(\frac{C_\text{SF}\varpi\kappa_\text{M} D}{2} N_c^{\frac{n_\text{PL}+1}{2}} \right. \\ &\qquad \left. - \frac{m\rho T\beta C_\text{SF}^{n_\text{PL}} D^{n_\text{PL}} n_\text{PL}}{(n_\text{PL}/2+1)2^{\frac{n_\text{PL}}{2}+1}}\right). \end{aligned} \tag{6}$$

By equating (6) to zero, we have

$$N_\text{c}^\star = \left(\frac{m\rho T\beta C_\text{SF}^{n_\text{PL}-1} D^{n_\text{PL}-1} n_\text{PL}}{(n_\text{PL}/2+1)2^{\frac{n_\text{PL}}{2}}\varpi\kappa_\text{M}}\right)^{\frac{2}{n_\text{PL}+1}}. \tag{7}$$

Moreover, it can be seen that (6) is monotonically decreasing when $2 \leq N_c \leq \min\{\max\{2, N_\text{c}^\star\}, N_{\max}\}$, and is monotonically increasing when $\min\{\max\{2, N_\text{c}^\star\}, N_{\max}\} \leq N_c \leq N_{\max}$.[2] Thus, the global minimum of $E_\text{tot,upper}(N_c)$, for $2 \leq N_\text{c} \leq N_{\max}$, is achieved at $\begin{cases} 2 & \text{if } N_\text{c}^\star < 2 \\ N_\text{c}^\star & \text{if } 2 \leq N_\text{c}^\star \leq N_{\max} \\ N_{\max} & \text{if } N_{\max} < N_\text{c}^\star \end{cases}$. As a result, the optimal $N^\star$ to minimize $E_\text{tot,upper}(N)$, for $N \in \{1, \cdots, N_{\max}\}$, is the following:

$$N^\star = \begin{cases} \arg\min_{N\in\{1,2\}}\{E_\text{tot,upper}(N)\} & \text{if } N_\text{c}^\star < 2, \\ \arg\min_{N\in\{1,\lfloor N_\text{c}^\star\rfloor,\lceil N_\text{c}^\star\rceil\}}\{E_\text{tot,upper}(N)\} & \\ & \text{if } 2 \leq N_\text{c}^\star \leq N_{\max}, \\ \arg\min_{N\in\{1,N_{\max}\}}\{E_\text{tot,upper}(N)\} & \text{if } N_{\max} < N_\text{c}^\star. \end{cases} \tag{8}$$

Based on (4) and (8), we then propose Algorithm 1 to solve our considered clustering and path planning problem. In the space-filling curve domain, we first find $N^\star$ (the optimal N) by minimizing $E_\text{tot,upper}(N)$, for $N \in \{1, \cdots, N_{\max}\}$, by using (8). If $N^\star = 1$, then we return the strategy of no clustering. Otherwise, we divide $\mathcal{C}$ into $N^\star$ equal-sized arcs where the sensors that are mapped to the same arc form a cluster. The stop positions are chosen as the centers of the corresponding arcs. We then find the tour of the robot by ordering the stop positions in $\mathcal{C}$. Finally, we map the clusters, the stop positions and the tour back to the 2D workspace.

B. Performance Analysis of Our Proposed Approach

In this section, we show an upper bound for the performance of our proposed approach for the case of $2 \leq N_\text{c}^\star \leq N_{\max}$.

Theorem 1: If $2 \leq N_\text{c}^\star \leq N_{\max}$ (these conditions can be satisfied when, for instance, T is large and ρ is small as compared to R), we have the following upper bound for

[2]We can prove that $E_\text{tot,upper}(N_c)$ is a quasi-convex function, for $2 \leq N_\text{c} \leq N_{\max}$.

Algorithm 1 Our proposed approach based on space-filling curves

1: map all the sensors in the 2D workspace $\mathcal{W}$ to $\mathcal{C}$
2: minimize $E_\text{tot,upper}(N)$, for $N \in \{1, \cdots, N_{\max}\}$, by using (8) to find $N^\star$
3: if $N^\star = 1$
4: return no clustering strategy
5: else
6: equally divide $\mathcal{C}$ into $N^\star$ arcs with the sensors in the same arc forming a cluster
7: choose the center of each arc as the stop position of the corresponding cluster
8: find the tour by ordering the stop positions
9: map the clusters, the stop positions and the tour back to the 2D workspace $\mathcal{W}$
10: end

$E_\text{tot,upper}(N^\star)$:

$$E_\text{tot,upper}(N^\star) \leq \left(\sqrt{2}+\frac{1}{n_\text{PL}}\right)\left(\frac{n_\text{PL}}{n_\text{PL}/2+1}\right)^{\frac{1}{n_\text{PL}+1}} \times \left(\frac{1}{2}\right)^{\frac{n_\text{PL}}{2(n_\text{PL}+1)}} (m\rho T\beta)^{\frac{1}{n_\text{PL}+1}} (\varpi\kappa_\text{M})^{\frac{n_\text{PL}}{n_\text{PL}+1}} (C_\text{SF} D)^{\frac{2n_\text{PL}}{n_\text{PL}+1}}. \tag{9}$$

Moreover, compared to the case of no clustering, we have

$$\begin{aligned}\underbrace{\frac{E_\text{tot}(N^\star)}{E_\text{tot}^\star(1)}}_{\text{energy saving in 2D}} &\leq \underbrace{\frac{E_\text{tot,upper}(N^\star)}{E_\text{tot,upper}(1)}}_{\text{energy saving in 1D}} \\ &\leq C_2\left(\frac{\varpi\kappa_\text{M}}{m\rho T\beta}\right)^{\frac{n_\text{PL}}{n_\text{PL}+1}} D^{-\frac{n_\text{PL}^2-n_\text{PL}}{n_\text{PL}+1}} \\ &= C_2\left(\frac{\varpi\kappa_\text{M}}{m\rho T\beta D^{n_\text{PL}-1}}\right)^{\frac{n_\text{PL}}{n_\text{PL}+1}}, \end{aligned} \tag{10}$$

where

$$\begin{aligned} C_2 &= \frac{1}{C_1}\left(\sqrt{2}+\frac{1}{n_\text{PL}}\right)\left(\frac{n_\text{PL}}{n_\text{PL}/2+1}\right)^{\frac{1}{n_\text{PL}+1}} \\ &\qquad \times \left(\frac{1}{2}\right)^{\frac{n_\text{PL}}{2(n_\text{PL}+1)}} C_\text{SF}^{\frac{2n_\text{PL}}{n_\text{PL}+1}} \\ &< \frac{(\sqrt{2}+1/2)C_\text{SF}^2}{C_1}, \quad \text{for } n_\text{PL} \geq 2. \end{aligned}$$

Proof: For the case that $2 \leq N_\text{c}^\star \leq N_{\max}$, the optimal energy cost has the following upper bound:

$$\begin{aligned} &E_\text{tot,upper}(N^\star) \leq E_\text{tot,upper}(\lceil N_\text{c}^\star\rceil) \\ &= \frac{m\rho T\beta C_\text{SF}^{n_\text{PL}} D^{n_\text{PL}}}{n_\text{PL}/2+1}\left(\frac{1}{2\lceil N_\text{c}^\star\rceil}\right)^{\frac{n_\text{PL}}{2}} + \sqrt{\lceil N_\text{c}^\star\rceil} C_\text{SF}\varpi\kappa_\text{M} D \\ &< \frac{m\rho T\beta C_\text{SF}^{n_\text{PL}} D^{n_\text{PL}}}{n_\text{PL}/2+1}\left(\frac{1}{2N_\text{c}^\star}\right)^{\frac{n_\text{PL}}{2}} + \sqrt{2N_\text{c}^\star} C_\text{SF}\varpi\kappa_\text{M} D \\ &= \left(\sqrt{2}+\frac{1}{n_\text{PL}}\right)\left(\frac{n_\text{PL}}{n_\text{PL}/2+1}\right)^{\frac{1}{n_\text{PL}+1}}\left(\frac{1}{2}\right)^{\frac{n_\text{PL}}{2(n_\text{PL}+1)}} \\ &\qquad \times (m\rho T\beta)^{\frac{1}{n_\text{PL}+1}} (\varpi\kappa_\text{M})^{\frac{n_\text{PL}}{n_\text{PL}+1}} (C_\text{SF} D)^{\frac{2n_\text{PL}}{n_\text{PL}+1}}, \end{aligned}$$

where the second inequality follows from the fact that the first term on the left hand side of the equation is monotonically decreasing with respect to $N_c^\star$ and the second term is monotonically increasing with respect to $N_c^\star$ for $N_c^\star \geq 2$. This confirms (9) in the theorem.

Moreover, equation (10) can be easily obtained by substituting the upper bound of (9) into $E_{\text{tot,upper}}(N^\star)/E^\star_{\text{tot}}(1)$. ∎

It can be seen from Theorem 1 that the upper bound of our proposed algorithm depends on the ratio $\varpi\kappa_{\text{M}}/(m\rho T\beta)$. Hence, the upper bound decreases if the motion cost (κ_{M}) decreases and/or the number of information bits generated in each period ($m\rho T$) increases. Also, the upper bound decreases as σ^2_{dB} increases, since β is monotonically increasing with respect to σ^2_{dB}. Finally, it is easy to see that $(\varpi\kappa_{\text{M}}/(m\rho T\beta))^{\frac{n_{\text{PL}}}{n_{\text{PL}}+1}}$ is monotonically decreasing with respect to n_{PL}, if $\varpi\kappa_{\text{M}}/(m\rho T\beta) < 1$, and $D^{-\frac{n^2_{\text{PL}}-n_{\text{PL}}}{n_{\text{PL}}+1}}$ is monotonically decreasing with respect to n_{PL}, for $n_{\text{PL}} \geq \sqrt{2}-1$ (this can be easily satisfied as in reality, n_{PL} is typically around $2 \sim 6$, as discussed in Section II-B). Hence, the upper bound decreases as n_{PL} increases, if $\varpi\kappa_{\text{M}}/(m\rho T\beta)$ is small. These observations are intuitive, since as the communication demand (the total number of generated information bits in each period) becomes higher and/or the channel quality gets worse, more stop positions are preferred in order to reduce the communication cost, as can be seen in (7). As a result, we expect that in general, the robot will save more energy, as compared to the case of no clustering.

IV. Simulation Results

Consider the case where the workspace is a 1000 m $\times$ 1000 m square region. There are total 400 sensors uniformly distributed in the workspace. The channel in the workspace has the following realistic channel parameters [27]: $n_{\text{PL}} = 4.57$ and $\sigma^2_{\text{dB}} = 16$. Moreover, we choose $R = 8$ bits/Hz/s, $p_{b,\text{th}} = 10^{-6}$, $B = 10$ MHz and the receiver noise power is chosen to be the realistic value of -104 dBm. We also use the real motion parameters of the Pioneer 3DX robot as follows [28]: $\kappa_1 = 7.4$, $\kappa_2 = 0.29$ and $v_{\max} = 1$ m/s. Furthermore, we use Sierpiński curves in our proposed algorithm. Finally, we choose $\varpi = 1$.

Fig. 3 shows the clusters of the sensors, the stop positions and the tour of the robot. In this example, we choose a sensing rate of $\widetilde{\rho} = 60$ Kbps for each sensor and $T = 3600$ s. It can be seen that in general, the sensors which are close to each other are grouped together as a cluster. However, this is not always the case as solving based on space-filling curves results in a considerably efficient (computation-wise) algorithm at a possible loss of optimality. Still, the predicted total energy cost of the whole network is only 0.68% of the energy cost of the case of no clustering, indicating a significant reduction in energy usage. Fig. 4 further shows the benefit of our proposed approach for different sensing rates. It can be seen that our approach can save the total energy consumption considerably, as compared to the case of no clustering. It is worth noting that the ratio does not monotonically increase as the sensing rate increases. This is because our proposed strategy optimizes the energy in the space-filling curve domain, which is an upper bound for the true performance. Finally, Fig. 5 shows the optimal number of clusters found by using our proposed approach. It can be seen that the number of clusters increases as the sensing rate increases. Similar results can also be obtained for different channel and motion parameters. We skip the details for brevity.

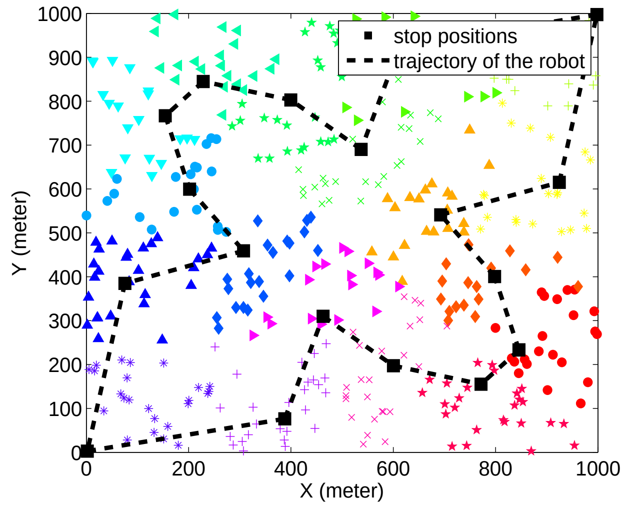

Fig. 3. The figure shows the clusters of the sensors, the stop positions and the tour of the robot by using our proposed algorithm of Section III-A. The markers with the same type and color represent a cluster of sensors. In this example, the predicted total energy cost of the whole network is only 0.68% of the energy cost of the case of no clustering.

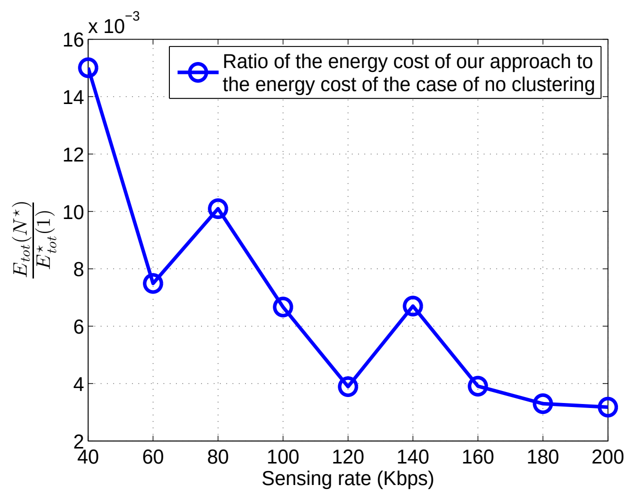

Fig. 4. The figure shows the ratio of the energy cost of our approach to the energy cost of the case of no clustering for different sensing rates. It can be seen that our approach can save the total energy consumption considerably.

V. Conclusions

In this paper, we considered a scenario where a mobile robot is tasked with periodically collecting data from a wireless sensor network. We proposed an energy-aware strategy that properly combines the ideas of clustering and using a mobile robot for data collection. Our approach is based on using space-filling curves, which results in

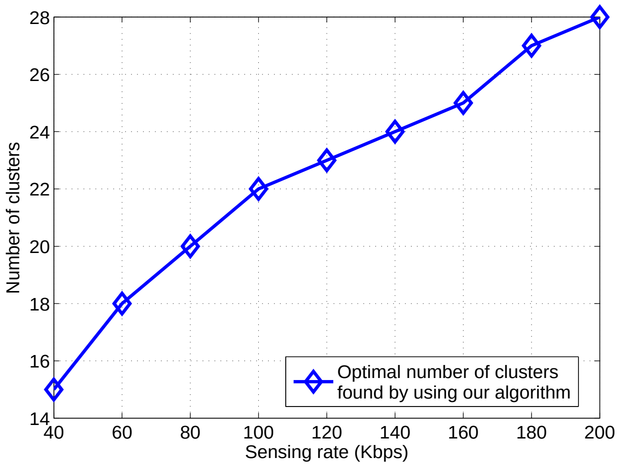

Fig. 5. The figure shows the optimal number of clusters. It can be seen that the number of clusters increases as the sensing rate increases.

a computationally-efficient algorithm. It can furthermore handle realistic communication environments by utilizing probabilistic channel predictors that go beyond disk models. We mathematically characterized an upper bound for the performance of our proposed algorithm which shows how the energy saving is related to the total number of generated bits in the network, and the communication and motion parameters. Finally, we verified the effectiveness of our proposed framework in a simulation environment, where a considerable reduction in energy consumption is achieved as compared to the case of no clustering.

References

[1] F. Bullo, E. Frazzoli, M. Pavone, K. Savla, and S. L. Smith. Dynamic vehicle routing for robotic systems. *Proc. of the IEEE*, 2011.

[2] M. M. Zavlanos, M. B. Egerstedt, Y. C. Hu, and G. J. Pappas. Graph-theoretic connectivity control of mobile robot networks. *Proceedings of the IEEE*, 99(9):1525 – 1540, July 2011.

[3] Cory Dixon and Eric W. Frew. Optimizing cascaded chains of unmanned aircraft acting as communication relays. *IEEE Journal on Selected Areas in Communications*, 30(5):883–898, 2012.

[4] J. Fink, A. Ribeiro, and V. Kumar. Motion planning for robust wireless networking. In *IEEE Int'l Conf. on Robotics and Automation*, 2012.

[5] J. Cortes, S. Martinez, and F. Bullo. Coordination and geometric optimization via distributed dynamical systems. *SIAM Journal on Control and Optimization*, 44(5):1543–1574, 2005.

[6] S.S. Ponda, L.B. Johnson, A.N. Kopeikin, H.-L. Choi, and J.P. How. Distributed planning strategies to ensure network connectivity for dynamic heterogeneous teams. *IEEE Journal on Selected Areas in Communications*, 30(5):861–869, 2012.

[7] D. Bhadauria, O. Tekdas, and V. Isler. Robotic data mules for collecting data over sparse sensor fields. *Journal of Field Robotics*, 28(3):388–404, 2011.

[8] R. Sugihara and R. Gupta. Improving the data delivery latency in sensor networks with controlled mobility. In *Distributed Computing in Sensor Systems*, pages 386–399. 2008.

[9] R. Sugihara and R. Gupta. Optimal speed control of mobile node for data collection in sensor networks. *IEEE Transactions on Mobile Computing*, 9(1):127–139, 2010.

[10] Y. Yun and Y. Xia. Maximizing the lifetime of wireless sensor networks with mobile sink in delay-tolerant applications. *IEEE Transactions on Mobile Computing*, 9(9):1308–1318, 2010.

[11] J. Luo and J.-P. Hubaux. Joint sink mobility and routing to maximize the lifetime of wireless sensor networks: the case of constrained mobility. *IEEE/ACM Trans. on Networking*, 18(3):871–884, 2010.

[12] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. on Wireless Comm.*, 1(4):660–670, 2002.

[13] O. Younis and S. Fahmy. Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on Mobile Computing*, 3(4):366–379, 2004.

[14] A. Somasundara, A. Kansal, D. Jea, D. Estrin, and M. Srivastava. Controllably mobile infrastructure for low energy embedded networks. *IEEE Transactions on Mobile Computing*, 5(8):958–973, 2006.

[15] M. Ma and Y. Yang. Sencar: an energy-efficient data gathering mechanism for large-scale multihop sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 18(10):1476–1488, 2007.

[16] G. Xing, T. Wang, W. Jia, and M. Li. Rendezvous design algorithms for wireless sensor networks with a mobile base station. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pages 231–240, 2008.

[17] O. Jerew, K. Blackmore, and W. Liang. Mobile base station and clustering to maximize network lifetime in wireless sensor networks. *Journal of Electrical and Computer Engineering*, 2012.

[18] Y. Yang, M. Fonoage, and M. Cardei. Improving network lifetime with mobile wireless sensor networks. *Computer Communications*, 33(4):409–419, 2010.

[19] Y. Mostofi, M. Malmirchegini, and A. Ghaffarkhah. Estimation of communication signal strength in robotic networks. In *Proc. of the 50th IEEE Int'l Conf. on Robotics and Automation*, May 2010.

[20] M. Malmirchegini and Y. Mostofi. On the spatial predictability of communication channels. *IEEE Trans. on Wireless Comm.*, 2012.

[21] M. Bader. *Space-Filling Curves: An Introduction with Applications in Scientific Computing*, volume 9. Springer, 2012.

[22] Y. Yan and Y. Mostofi. Efficient communication-aware dynamic coverage using space-filling curves. In *American Control Conference*, Portland, OR, 2013.

[23] J. Bartholdi III and L. Platzman. An $O(N \log N)$ planar travelling salesman heuristic based on spacefilling curves. *Operations Research Letters*, 1(4):121–125, 1982.

[24] J. Bartholdi III and L. Platzman. Heuristics based on spacefilling curves for combinatorial problems in euclidean space. *Management Science*, 34(3):291–305, 1988.

[25] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[26] Y. Mei, Y.-H. Lu, Y. C. Hu, and C. S. G. Lee. Deployment of mobile robots with energy and timing constraints. *IEEE Transactions on Robotics*, 22(3):507–522, June 2006.

[27] W. M. Smith. *Urban Propagation Modeling for Wireless Systems*. PhD thesis, 2004.

[28] Y. Mei, Y.-H. Lu, Y. C. Hu, and C. S. G. Lee. Energy-efficient motion planning for mobile robots. In *Proc. of IEEE Int'l Conf. on Robotics and Automation*, pages 4344–4349, New Orleans, LA, April 2004.