Maximum Precision-Lifetime Curve for Joint Sensor Selection and Data Routing in Sensor Networks

Kerem Bulbul* and Ozgur Ercetin

Sabanci University, Faculty of Engineering and Natural Sciences, Orhanli, Istanbul, 34956 Turkey

Abstract— **In many classes of monitoring applications employing battery-limited sensor networks, periodic sampling of an area with a given precision level is required. For such applications, we provide mathematical programming formulations for deriving the optimal trade-off curve between network lifetime and data precision, and design a practical heuristic for near-optimal operation. The properties of our models and the effectiveness of our heuristic are demonstrated by computational experiments.**

Keywords: sensor telemetry; sensor selection; routing; mathematical programming; link metric; heuristic.

I. INTRODUCTION

Sensor telemetry is the use of sensors and two-way wireless communications to gather information from products, people and places with sufficient accuracy. Recently, sensor telemetry is being developed for such diverse application areas as ecological habitat monitoring, military surveillance and industrial and commercial networked sensing [1].

In this paper, we focus on energy-efficient, *periodic sampling* based sensor telemetry, and we assume that the monitoring application periodically takes measurements from sample points that are equally apart in a given region. The quality of service (QoS) of the monitoring application depends on the number of sample points from which the measurements are taken, i.e., the precision level. In real-world applications, such as agricultural or structural health monitoring, this relationship between the application QoS and precision level is usually nonlinear. A low precision level definitely results in a low QoS because for a large portion of the region the ambient conditions remain unknown. However, a high precision level may result in over-sampling (and high energy consumption), because the ambient conditions do not usually change within short distances, e.g., the soil temperature on a farm. Thus, for efficient operation, it is essential to choose precision levels that do not result in under- or over-sampling.

In general, the sensor telemetry application can operate satisfactorily for a range of precision levels. Thus, the network is considered to be operational as long as data is collected with sufficient precision. Even if a number of nodes exhaust their batteries, the network may be able to continue to collect and deliver useful data to the monitoring station, alas with a lower precision. Therefore, in telemetry applications it is important to quantify the amount of data that can be collected starting with a given maximum initial precision level. In this paper, we develop *precision-lifetime curves* plotted for different initial maximum precision levels for this purpose. A precision-lifetime curve depicts the maximum possible amount of data that can be collected by the network at each supported

precision level. By examining these curves, the network manager can select an initial precision level that would provide an operation of the monitoring application with an acceptable QoS for as long as possible.

The paper is organised as follows: In Section II, we briefly discuss previous related work. In Section III, we introduce the system model used in the formulation of the lifetime optimisation problem. In Section IV, the derivation of the maximum precision-lifetime curve is formulated as a mixed-integer linear program. In Section V, we develop a fast heuristic for energy-efficient sensor selection and routing that yields a near-optimal precision-lifetime curve and that may be implemented in practice. In Section VI, we report the results of our computational experiments that demonstrate the effectiveness of the methods introduced, and we conclude in Section VII.

II. RELATED WORK

Our work is related to the connected sensor cover problem [2]. In the connected sensor cover problem, a set S of N nodes is deployed in a sensing field of area A, and each sensor node s_i has a sensing region denoted by A_i. A query Q is made for a region $A_Q \subseteq A$. Then, a set of sensor nodes $S' \subseteq S$ is selected so that $A_Q \subseteq \cup_i A_i$, and any pair of nodes in the connected sensor cover can communicate with each other either directly or indirectly over a multi-hop communication path. The minimum connected sensor cover problem finds the set of sensors with the minimum number of nodes, such that the two conditions above hold. The problem discussed in this paper differs from the minimum connected sensor cover problem in several aspects: First, we consider a sensor cover with a grid structure, where each node has the same sensing radius. Second, the best sensor cover is defined as the one that maximises the network lifetime. Furthermore, not all selected sensors sense, i.e., some of the selected nodes work as relay nodes. Finally, we require all sensor nodes to be connected to the monitoring station, but not necessarily to each other.

Our work is also related to the efforts on topology control and energy-efficient routing in wireless ad hoc and sensor networks in [5], [6], and [7]. Previously in the literature, topology control is investigated with the objective of providing end-to-end traffic, i.e., a connected and power/energy efficient path. Our work brings together application-specific requirements (in the form of measurement precision), topology control and routing. The closest model to ours is described in [5]; however, in [5] the objective is to detect the occurrence of an event in an energy-efficient manner rather than to continuously collect the ambient conditions of an area.

*Corresponding author: +90-216-483-9589 (tel), +90-216-483-9550 (fax).

In our work, we do not only select the best set of sensors to cover a region, but also determine the lifetime maximising route between these sensors and the monitoring station. Lifetime maximising routing is previously considered in [3] and [4], where the network is assumed to be *not* operational when the first node in the network exhausts its battery. Unlike these approaches, we consider the network to be operational as long as data with acceptable precision is collected from the network. Other network lifetime definitions are also proposed in the literature. For instance, [8] develops a maximum node-life and a maximum flow-life curve for a network with predetermined required rates of flows between pairs of nodes. Also, [9] investigates the rate allocation problem at the nodes of a network in order to maximize the total bit volume delivered to a monitoring station from the entire network over time while assuring fairness among all nodes. The models in these last two papers are similar in spirit to the models we develop in Section IV.

III. System Model

The sensor network is used to collect information from the sensors distributed in a field in order to support various (monitoring) applications. Each node in the sensor network has the same sensing capability. The nodes can be deployed in this field randomly or according to a predefined topology. If the nodes are deployed randomly, we assume that the nodes can determine their respective locations by using methods such as those described in [1, Chapter 3].

Let $U(K)$ be the utility gained from a query with precision K. The precision of the query can be defined in various ways. In general, the precision corresponds to the degree of how good the data collected represents the ambient conditions in the field. Specifically, we consider a monitoring application in which it is desired to take measurements in a field at points that are distributed as uniformly as possible, and the precision K is defined as the number measurement points along a cross section in either $x-$ or $y-$direction. In other words, the desired number of measurement points is K^2 if the precision of the query is K. We note that uniformly distributed measurements obviate the need for additional information on the spatial and temporal correlation among sensed data. Furthermore, we assume that in-network data fusion is not performed, since data fusion requires a priori statistical data to be kept for the area of interest.

The objective is to maximise the aggregate utility gained from the network over the network lifetime. Here, the network lifetime is defined as the latest time instant a query with an arbitrarily low (but pre-defined) precision level can be satisfied. The response to each query consumes a part of the limited node energy due to sensing, transmission and reception operations. Clearly, as the precision level for a query increases, the energy consumption in the network does also increase.

Assume that there are N sensor nodes in a field of dimensions $I \times I$, and (x_i, y_i) denotes the coordinates of sensor node i, $i = 1, \ldots, N$. Without loss of generality, the monitoring station, denoted by the index $i = 0$, is located at the coordinates $(0, 0)$. In a sample operation of the network,
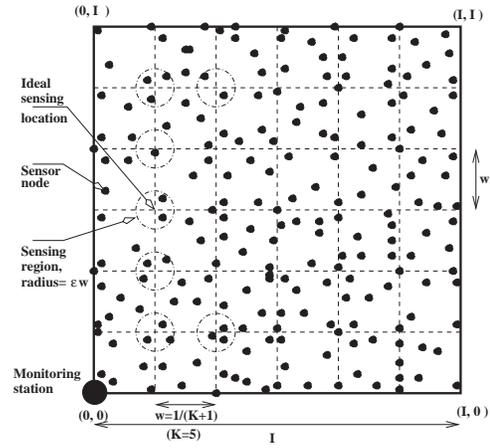
Fig. 1. Graphical representation of the parameters in the maximum precision-lifetime problem.

the telemetry application requests from the monitoring station to take measurements every $w = I/(K + 1)$ units along both $x-$ and $y-$ directions of the field corresponding to a query of precision K. Assuming that the area monitored is a square with one side length I, this forms a square grid lattice as demonstrated in Figure 1. Each point on this grid represents an *ideal measurement point* from which data is to be gathered. The location of the kth ideal measurement point is denoted by (a_k, b_k), $k = 1, \ldots, K^2$. However, if the nodes are randomly deployed, we may not always be able to find sensor nodes at these ideal positions. Therefore, upon receipt of the query, the monitoring station selects the set of sensor nodes among a number of candidate nodes. For each ideal measurement point (a_k, b_k), the candidate nodes lie within a circle of radius $\epsilon * w$ centered at (a_k, b_k) where ϵ is the user defined tolerance level. (See Figure 1.) In addition, by alternating between the sensor nodes in close proximity of the ideal measurement points, we can balance the energy consumption in the network and extend the lifetime. Alternatively, we could always choose the closest sensor nodes to the ideal measurement points; however, this policy would quickly deplete the energy of such nodes, and thus result in a shorter network lifetime.

The form of the utility function $U(K)$ depends on the application and the user requirements. For example, if we are just interested in whether a response from an arbitrary node in the network can be received or not, then $U(K) = c$, i.e., the utility gained from a query is the same for all precision levels. In more practical cases, $U(K)$ may increase with increasing precision, since more and higher quality information is obtained about the area. For instance, one may consider a linear utility function of the form $U(K) = a \cdot K + b$. In other more realistic cases, $U(K)$ may have a more exotic form, where the utility is low for low precision levels, with a rapid increase in utility for a range of precision levels in the middle, and so that the marginal utility starts decreasing beyond a threshold precision level. In any case, the definition of $U(K)$ is a subjective matter, and it is nearly impossible to figure out an exact form of the utility function.

In our work, instead of guessing a utility function and optimising the network operations according to this fictitious

utility function, we consider a lexicographic optimisation, i.e., we assume that $U(K) \gg U(K-1) \gg U(K-2) \ldots$. Therefore, in our lifetime optimisation problem, beginning with an initial user-specified *desired* precision level, we maximise the duration for which this precision level can be supported. When a query at this desired precision level fails due to the limited node energies, the next lower precision level that can be supported is selected, and the duration for which this precision level can be supported is maximised, and so on. The optimisation procedure terminates when the supportable precision level drops below a minimum *acceptable* precision level.

IV. LIFETIME MAXIMISATION PROBLEM

As discussed in the previous section, initially we would like to maximise the time t_1 to network partition given the user-defined initial precision K_0, i.e., the time when the sensor network fails to respond to a query for the first time because one or several nodes do not have sufficient energy to sense and/or to forward the packets to the monitoring station. At time t_1, we can declare that the network is dead, or we can decrease the precision of the queries so that $K_1 < K_0$ and continue to collect data from the network until the next point in time t_2 when one or more additional nodes drain up their energy, and the network fails to respond to the queries at precision K_1. Here, we adopt the second approach, i.e., we develop models and algorithms to construct a "maximum precision-lifetime curve". A maximum precision-lifetime curve is defined by an *ordered* set of pairs $MPLC = \{(t_m, K_{m-1}) | K_{m-1}$ is the precision for $t_{m-1} \leq t \leq t_m, \ m = 1, \ldots, M\}$ so that $t_0 = 0$, $t_m > t_{m-1}, \ m = 1, \ldots, M$, K_0 is the user defined initial precision level, and $K_m < K_{m-1}, \ m = 1, \ldots, M-1$. The length of the interval $t_{m-1} \leq t \leq t_m$ is denoted by $q_m = t_m - t_{m-1}$, and the precision level decreases at each *drop point* $(t_m, K_{m-1}), \ m = 1, \ldots, M$, where M is the number of drop points on the precision-lifetime curve. Our objective in this paper is to determine both the number of drop points M, and the associated times $t_1, \ldots, t_M$ on the maximum precision-lifetime curve as defined by our lexicographic network lifetime maximisation problem defined in the previous section.

The maximum precision-lifetime curve allows the user to decide when the network is dead based on the monitoring application's needs. For instance, if the precision level does not decrease quickly from one drop point to the next one on the maximum precision-lifetime curve, then the sensor network may still be able to collect useful information. Clearly, such an approach has advantages over declaring the network dead at the time of first network partition as in [3]. In general, we prefer collecting data at higher precisions sooner than later. Alternatively, it is possible to collect data at lower precisions, i.e., with fewer number of measurement points, from the start and operate the network for a longer period of time. However, such sparse data may not be useful. In other words, we assume that the initial precision K_0 defined by the user reflects the monitoring application's needs, and our objective is to operate the network at or close to this precision for as long as possible.

In order to generate the maximum precision-lifetime curve, we follow a two-step approach. First, by solving a linear program TTP-LP(M) (Time-To-Partition-LP) we maximise the time t_M until the next network partition given the previous $(M-1)$ drop points on the precision-lifetime curve and the current precision K_{M-1}. The next drop point is determined as (t_M, K_{M-1}) and added to $MPLC$. At time t_M, some nodes in the network exhaust their energy so that the network is no longer able to fulfill the queries at precision K_{M-1}. Therefore, at time t_M we set up and solve an integer program MP-IP($M+1$) (Maximum-Precision-IP) that determines the maximum possible precision K_M for the remaining network subject to the constraints imposed by the drop points that have already been identified. Note that at time zero, we have $MPLC = \{\}$, $t_0 = 0$, and K_0 is specified by the user.

For formulating TTP-LP(M), assume that $(M-1)$ drop points have already been determined on the maximum precision-lifetime curve, i.e., we know the total amount of sensor data for each measurement point that needs to be routed to the monitoring station for $t_{m-1} \leq t \leq t_m, \ m = 1, \ldots, M-1$. We are currently at time t_{M-1} and need to select $(K_{M-1})^2$ sensor nodes out of a total of N nodes in the network so that these sensor nodes are in sensing regions. The kth sensor node belongs to a set of candidate nodes S_k^M, and we allow the specific sensor node in the set S_k^M to change during $t_{M-1} \leq t \leq t_M$. A precise definition of the sets $S_k^m, \ k = 1, \ldots, (K_{m-1})^2, \ m = 1, \ldots, M$ will be given later. The objective at time t_{M-1} is to select one sensor node from each set S_k^m, and to route the packets from these sensor nodes such that the time to next network partition, i.e., $q_M = t_M - t_{M-1}$, is maximised, and ρ packets per measurement point per unit time are forwarded to the monitoring station. The main sources of energy consumption are the sensing, transmission and receiving operations. Let $E_i, \ i = 1, \ldots, N$, be the initial energy level of node i at time $t_0 = 0$, and let ξ and e_r be the constant amount of energy required for a single sensing and receiving operation by any node, respectively. The transmission energy depends on the distance between two nodes, and $e_{ij}, \ i = 1, \ldots N, \ j = 0, \ldots, N$, denotes the amount of energy consumed by node i when sending a unit size packet to node j. The messages generated by the $(K_{m-1})^2$ sensor nodes are routed to the monitoring station over multi-hop paths where the variable $f_{ij}^{km} \geq 0, \ k = 1, \ldots, K_{m-1}, \ m = 1, \ldots, M, \ i = 1, \ldots, N, \ j = 0, \ldots, N$, denotes the number of packets generated by a sensor node $k \in S_k^m$ and transmitted from node i to node j during the time period $t_{m-1} \leq t \leq t_m$. Finally, let $L_i, \ i = 1, \ldots N$, be the residual energy of node i at time t_M.

Ideally, during a time interval $t_{m-1} \leq t \leq t_m$, we would like to select $(K_{m-1})^2$ sensor nodes in the network such that they form a uniform grid in which the ideal sensor locations are separated by a distance $w_{K_{m-1}} = \frac{1}{K_{m-1}+1}$ in both $x-$ and $y-$directions. (See Figure 1.) However, for reasons discussed in detail in Section III, we only require that the kth sensor node is located within a circle of radius $\epsilon w_{K_{m-1}}$ around the kth ideal sensor location. In order to impose these constraints on the locations of the sensor nodes, we construct the sets $S_k^m = \{i \mid d_k(x_i, y_i) \leq \epsilon w_{K_{m-1}}\}, \ k = 1, \ldots, (K_{m-1})^2, \ m = 1, \ldots, M$, where $d_k(x_i, y_i)$ is the Euclidean distance of node i from the kth ideal sensor location. We make sure that the
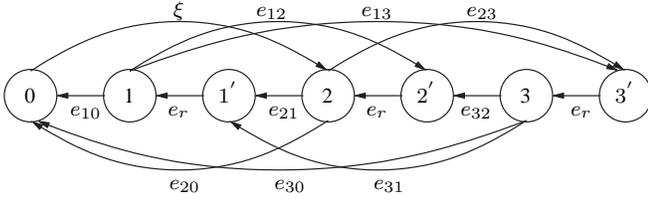
Fig. 2. Network structure and transmission energies for $N = 3$, $M = 1$, $K_0 = 1$, $S_1^1 = \{2\}$.

kth sensor node belongs to the set S_k^m during the time interval $t_{m-1} \leq t \leq t_m$.

Network flow problems are among the types of linear programs that can be solved most effectively [10]. Below, we formulate the problem of maximising the time q_M until the next network partition as a linear program TTP-LP(M) in which the feasible region has the structure of a *multi-commodity circulation problem* with side constraints for energy consumption, i.e., a network flow problem in which all nodes, including the monitoring station, have a net flow of zero for each commodity. In our formulation, we have $\sum_{m=1}^{M}(K_{m-1})^2$ commodities where a commodity C_k^m represents data collected by the nodes in the set S_k^m during the time interval $t_{m-1} \leq t \leq t_m$. In addition, we define a *virtual node* i' for each *real node* $i \neq 0$ that can only receive packets from a real node $j \neq i$, $1 \leq j \leq N$. The set of all real nodes (excluding the monitoring station) and the set of all virtual nodes are denoted by V and V', respectively. Also, define $\mathcal{V} = \{0\} \cup V \cup V'$. Virtual nodes can only send data to their associated real nodes, and they allow us to consider the receiving energy as the transmission energy from a virtual node to a real node. In order to ensure a net flow of zero for each commodity at the monitoring station, the sensing operations are represented by flows that originate from the monitoring station. However, in this case the transmission energy (equal to the sensing energy as defined below) is charged to the real node receiving the flow. In addition, real nodes can send data to the monitoring station in addition to any virtual node other than their own associated virtual node. Thus, we redefine the transmission energy matrix e as

follows: $e'_{uv} = \begin{cases} \xi & u = 0, \exists k, m \text{ s.t. } v \in S_k^m \\ e_r & u = i', v = i, i' \in V', i \in V \\ e_{ij} & i \in V, j \in \{0\} \cup V', j \neq i' \\ \infty & \text{otherwise} \end{cases}$.

An example for this network structure is given in Figure 2 for N=3, $M = 1$, $K_0 = 1$, $S_1^1 = \{2\}$. For each arc, the energy consumption for a unit flow through the arc is indicated in the figure.

Then, the linear program TTP-LP(M) below maximises the time q_M to the next network partition given that $(M-1)$ drop points already exist on the maximum precision-lifetime curve.

$$\max \; q_M \tag{1}$$

$$\sum_{i \in S_k^m} f_{0i}^{km} = \rho \cdot q_m \quad k = 1, \dots, (K_{m-1})^2,$$
$$m = 1, \dots, M - 1 \tag{2}$$

$$\sum_{i \in S_k^M} f_{0i}^{kM} = \rho \cdot q_M \quad k = 1, \dots, (K_{M-1})^2 \tag{3}$$

$$\sum_{j \in \mathcal{V}} f_{ij}^{km} - \sum_{j \in \mathcal{V}} f_{ji}^{km} = 0 \quad \forall i \in \mathcal{V}, \; k = 1, \dots, (K_{m-1})^2,$$
$$m = 1, \dots, M \tag{4}$$

$$\sum_{m=1}^{M} \sum_{k=1}^{(K_{m-1})^2} e'_{0i} f_{0i}^{km}$$
$$+ \sum_{m=1}^{M} \sum_{k=1}^{(K_{m-1})^2} \sum_{j \in \{0\} \cup V'} e'_{ij} f_{ij}^{km}$$
$$+ \sum_{m=1}^{M} \sum_{k=1}^{(K_{m-1})^2} e'_{i'i} f_{i'i}^{km} + L_i = E_i \quad \forall i \in V \tag{5}$$

$$q_M \geq 0 \tag{6}$$
$$f_{ij}^{km} \geq 0 \quad \forall i, j, k, m \tag{7}$$
$$L_i \geq 0 \quad \forall i \tag{8}$$

The objective function (1) maximises the time to next network partition. In general, there are many alternate flows that attain the same precision-lifetime curve. Hence, in order to find the maximum precision-lifetime curve, we allow TTP-LP(M) to optimise the flows that were previously obtained by TTP-LP(m), $m < M$, so that the current objective q_M is maximised as long as the previously determined total flow volumes into the monitoring station are satisfied for all commodities C_k^m, $k = 1, \dots, (K_{m-1})^2$, $m = 1, \dots, M - 1$. This is ensured by the constraints (2). Therefore, the set of flows $\{f_{ij}^{k1}\}$ determined at time $t_0 = 0$ by solving the model TTP-LP(1) may be different than the set of flows $\{f_{ij}^{k1}\}$ determined at time t_{M-1} by the model TTP-LP(M); however, the total flow delivered to the monitoring station from the set of nodes S_k^1, $k = 1, \dots, (K_0)^2$ is the same in both models. The constraints (3) prescribe that ρ packets per unit time are sent from each set S_k^M, $k = 1, \dots, (K_{M-1})^2$, to the monitoring station for q_M units of time. The net flow for any commodity at any node, including the monitoring station, is zero as reflected by the flow conservation constraints (4). The constraints (5) relate the initial energy levels E_i at time $t_0 = 0$ to the final energy levels L_i at time $t_M = t_{M-1} + q_M$.

The next drop point (t_M, K_{M-1}) is determined by solving TTP-LP(M) optimally, and at time $t = t_M$ the network is no longer able to deliver sensor data to the monitoring station at the current precision because one or several nodes exhaust their total energy. So, at time t_M we first have to determine the maximum precision $K_M \leq K_{M-1} - 1$ that can be supported by the remaining network before we can maximise q_{M+1} at this precision. Alternatively, we can consider this problem one of selecting the sets S_k^{M+1} so that K_M is maximised. Therefore, we define the sets $S_{k_\mu}^{M+1} = \{i \,|\, d_{k_\mu}(x_i, y_i) \leq \epsilon w_\mu\}$, $k_\mu = 1, \dots, (\mu)^2$, $\mu = 1, \dots, K_{M-1} - 1$, and the binary variables z_μ, $\mu = 1, \dots, K_{M-1} - 1$. If $z_\mu = 1$, then the sets $S_{k_\mu}^{M+1}$, $k_\mu = 1, \dots, (\mu)^2$, are selected for sensing during $t_M \leq t \leq t_{M+1}$ where t_{M+1} is still unknown. Then, the mixed integer model MP-IP($M + 1$) below maximises the

precision for the remaining network at time t_M.

$$\max \sum_{\mu=1}^{K_{M-1}-1} \mu z_\mu \tag{9}$$

$$\sum_{\mu=1}^{K_{M-1}-1} z_\mu = 1 \tag{10}$$

$$\sum_{i \in S_k^m} f_{0i}^{km} = \rho * q_m \quad k = 1, \ldots, (K_{m-1})^2,$$

$$m = 1, \ldots, M \tag{11}$$

$$\sum_{i \in S_{k_\mu}^{M+1}} f_{0i}^{k_\mu M+1} = \rho * z_\mu \quad k_\mu = 1, \ldots, (\mu)^2,$$

$$\mu = 1, \ldots, K_{M-1} - 1 \tag{12}$$

Constraints similar to

$$(4)\text{-}(8) \text{ in TTP-LP}(M). \tag{13}$$

In the model above, the constraint (10) ensures that exactly one set of uniformly distributed measurement points is selected, and the objective (9) maximizes the cardinality of this set. The constraints (11) correspond to the constraints (2) in the model TTP-LP(M) and prescribe that the previously determined total flow volumes into the monitoring station are satisfied for all commodities C_k^m, $k = 1, \ldots, (K_{m-1})^2$, $m = 1, \ldots, M$. If $z_\mu = 1$, then the sets $S_{k_\mu}^{M+1}$, $k_\mu = 1, \ldots, (\mu)^2$ must be able to sense and forward the corresponding sensor data to the monitoring station for at least one additional unit of time because $\rho * z_\mu = \rho$ in the constraints (12). In other words, the precision level until the next drop point on the maximum precision-lifetime curve is selected such that sensor data can be routed to the monitoring station during the time period $t_M \leq t \leq t_M + 1$. In addition, the model MP-IP($M+1$) has flow conservation and energy consumption constraints that are similar to those in TTP-LP(M) and are omitted here.

V. A Low Complexity Heuristic

The optimal precision-lifetime curves provide the network manager tactical information regarding the selection of the initial precision level, and the lifetime of a particular network satisfying the required quality of service. Unfortunately, deriving the optimal precision-lifetime curve may require solving several linear and integer programs successively. Therefore, effective heuristic approaches that can approximately realise the optimal precision-lifetime curve are required. Below, we discuss one such low computational complexity routing algorithm for a given initial precision level.

Our precision-lifetime algorithm consists of two interrelated parts: sensor selection and routing. The candidate sensor nodes that can respond to a particular query with a required precision K_{M-1} are given by the sets S_k^M, $k = 1, \ldots, (K_{M-1})^2$. Thus, one node from each set S_k^M, $k = 1, \ldots, (K_{M-1})^2$, is selected to respond to a query; effectively acting as the source of the kth commodity in the network. In our algorithm, given the current link metrics discussed later in this section, we first run reverse Dijkstra's algorithm [10] in order to find the shortest paths from all nodes in the network to the sink node. Then,

from each set S_k^M (excluding dead nodes) we select a sensor node that has the shortest path to the sink among the nodes in the set S_k^M. For each commodity, we send one unit flow per query over the calculated shortest paths. Note that some nodes may need to route several commodities for a given query if they belong to more than one shortest path. The link metrics are updated and the shortest paths are re-calculated every t_{upd} time units. Eventually, at some time t_M, the current precision level cannot be supported any more, since some of the nodes deplete their energies. Then, we set $K_M = K_{M-1} - 1$, and repeat the procedure at this precision level.

We define and update the link metrics in the shortest path problems in a similar way to that in the fast greedy heuristics discussed in [11]. The authors in [11] develop greedy heuristics for the maximum data extraction problem in energy-limited sensor networks. First, they present a multi-commodity maximum flow formulation for their problem. Then, they propose iterative algorithms based on the insights obtained from the dual linear program similar in spirit to the algorithms developed in [12].

In the discussion below, let l_{ij} be the cost of routing a unit flow over the link between the nodes i and j. It is clear that the link cost depends on the residual energy levels of nodes i and j, as well as the transmission energy between these nodes and the receiving and sensing energies. We use a link metric defined as follows:

$$l_{ij} = e_{ij} b^i + \xi b^i + e_r b^j, \tag{14}$$

where b^i is a parameter which is a function of the residual energy of node i. Note that the second term in (14) is only present if node i is a sensor node that responds to a query. Initially, b^i is set to $b_0^i = 1/E_i$, and b^i is updated each time the shortest paths are to be re-calculated if node i is currently receiving or sending data. Let b_k^i be the value of b^i after the kth update. Then,

$$b_k^i = b_{k-1}^i \exp(\delta \beta^i), \tag{15}$$

where δ is a control parameter and $\beta^i = 1 - E_i^{\text{rem}}/E_i$ where E_i^{rem} is the residual energy of node i at the time of the update. Thus, b^i is increased with each update, and the rate of increase of b^i increases over time because β^i increases. Note that this metric ensures that when all nodes have plenty of energy, the path with the minimum total consumed energy is preferred, whereas later it becomes more important to avoid nodes with low residual energy. In addition, each time the precision level is decreased we re-initialise $b^i = 1/E_i^{\text{rem}} \; \forall i$ and $E_i = E_i^{\text{rem}} \; \forall i$ because a change of precision presents a significant change in the network topology.

VI. Computational Experiments

In our numerical study, we have several objectives. First, we would like to demonstrate how a precision-lifetime curve can be employed for tactical decision making while choosing an initial precision level given the QoS requirements of the monitoring application. Second, we need to show that our heuristics exhibit a near-optimal and robust performance for a variety of network topologies. Due to the extensive computational
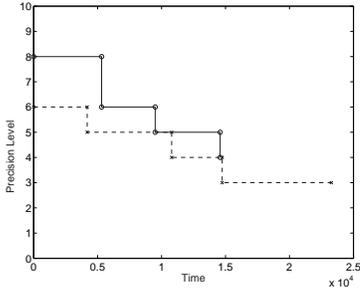
Fig. 3. Maximum precision-lifetime curves with different initial precision levels.

time required when solving the precision-lifetime optimisation problem in Section IV for 2-dimensional networks with a large number of nodes and a high initial precision level, we only report results in 1-dimensional networks for large K^0. (See Table II.) Clearly, a 1-dimensional network is a special case of a 2-dimensional network in which the number of measurement points is equal to the precision level, and the circular sensing regions around the ideal measurement points are reduced to intervals.

In all of our computational experiments, the initial energies of the nodes are uniformly distributed between 1.5 and 3.3. Each time unit, $\rho = 1000$ packets per measurement point per unit time are to be forwarded to the sink, and the node energies are decremented per unit flow sensed, transmitted or received according to the Crossbow specifications [13]. The parameters δ and t_{upd} in our heuristic are set to 0.1 and 15 time units, respectively. The mathematical programs in Section IV are modeled in ILOG OPL Studio 3.7.1 and solved by ILOG CPLEX 9.1. The heuristic in Section V is coded in Visual Basic.

Figure 3 depicts the trade-off between choosing different precision levels for a network with 80 nodes in which the nodes are distributed randomly on a line of length 10 meters (m) emanating from the sink. The tolerance level ϵ for determining the sets S_k^m is assumed to be 0.1. The network can maintain the precision levels $8, 6, 5,$ and 4 for a total network lifetime of $14,579$ time units for an initial precision level $K_0 = 8$ while the precision levels $6, 5, 4$ and 3 can be supported for a total network lifetime of $23,274$ time units if the initial precision level is decreased to $K_0 = 6$. In other words, if the monitoring application can tolerate precision levels down to 3 sample points per query, then choosing a slightly lower initial precision level provides a better network lifetime. Also note that the total data gathered at precision levels 5 and 4 are greater when we start collecting data at $K_0 = 6$. Thus, depending on the requirements of the monitoring application, it may be more advantageous in the long run to choose an initial precision level that is slightly lower than the highest value desired.

In order to demonstrate the effectiveness of the heuristic stated in Section V with different parameter settings, we generate two sets of problem instances. In the first set, we consider a 2-dimensional network in which the nodes are deployed randomly in a square area of dimensions $1m \times 1m$ and $\epsilon = 0.3$. We randomly generate 10 problem instances

for each $N = 250, 300, 350$ with $K_0 = 3$, and assume that the application can tolerate precision levels down to 1 measurement per query. The defining characteristic of these instances is that there is no power control in the network because the transmission energy is constant between any two pair of nodes in the network according to the Crossbow specifications [13].

For each instance, we generate both the optimal and heuristic precision-lifetime curves. In order to evaluate a precision-lifetime curve $PLC = \{(t_1, K_0), (t_2, K_1), \ldots, (t_M, K_{M-1})\}$, we use the function $f(PLC) = \sum_{m=1}^{M} 2^{K_{m-1}} (t_m - t_{m-1})$ where $t_0 = 0$. This evaluation function assigns a much greater weight to higher precision levels which is in line with our lexicographic lifetime optimisation problem. In Table I, we report statistics on the ratio $f^{\text{Heur}}/f^{\text{Opt}}$ computed over 10 instances for each $N = 250, 300$ and 350 where f^{Heur} and f^{Opt} denote the performance measures obtained for the curves constructed by the heuristic and the optimal algorithm, respectively. Clearly, the heuristic demonstrates an excellent performance, and although we do not report detailed results here, its computational time requirement is at least an order of magnitude less than the optimal approach.

TABLE I

PERFORMANCE OF THE HEURISTIC COMPARED TO THE MAXIMUM PRECISION-LIFETIME CURVE IN 2-DIMENSIONAL NETWORKS.

N	250	300	350
avg.	97.71%	97.37%	96.91%
min.	93.56%	94.12%	94.62%
max	100.00%	100.00%	98.79%

In the second set of problem instances, we consider a 1-dimensional network in which the nodes are deployed randomly on a line of length 10m and $\epsilon = 0.1$. Note that the transmission energies are not constant throughout the network in this case, and ϵ is chosen much smaller. We randomly generate 10 problem instances for each combination of $N = 50, 75, 100$ and $K_0 = 8, 7, 6$, and assume that the application can tolerate precision levels down to 4 measurements per query. The same performance measure as above is used, and the results are given in Table II. We note that the performance of the heuristic is more sensitive to increasing node density than the initial precision level. We also observe that the results are worse than those in Table I. However, recall that in the optimal algorithm we allow all flows from time zero on to be optimised for a set of drop points on the maximum precision-lifetime curve. (See Section IV.) Clearly, this is an ideal approach that is very hard to match in a practical implementation. Therefore, the results in Tables I-II are very promising.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we defined a new performance metric for the maximum data collection problem in battery-limited sensor networks. In this study, we focused on a data collection model which requires the periodic sampling of a region and formulated the maximum precision-lifetime problem as a series of linear and mixed integer programs. The resulting maximum precision-lifetime curve may be used by the network managers

TABLE II

PERFORMANCE OF THE HEURISTIC COMPARED TO THE MAXIMUM PRECISION-LIFETIME CURVE IN 1-DIMENSIONAL NETWORKS.

N	**50**			**75**			**100**		
K	6	7	8	6	7	8	6	7	8
avg.	90.81%	91.11%	91.03%	82.09%	81.42%	82.43%	79.22%	85.00%	85.34%
min.	76.56%	82.60%	75.48%	64.57%	64.57%	76.17%	66.10%	66.10%	76.15%
max.	100.00%	100.00%	100.00%	100.00%	92.93%	97.30%	90.27%	94.84%	96.40%

in calibrating the needs of the monitoring application according to the network capabilities. In addition, we developed a fast practical heuristic whose performance closely matches the optimal precision-lifetime curve without resorting to the solution of mathematical programs. As a future work, we will consider the incorporation of more elaborate definitions of data precision where we consider the spatial and temporal correlation of the data.

REFERENCES

[1] Krishnamachari, B., *Networking Wireless Sensors*, Cambridge University Press, NY, 2005.
[2] Gupta H., Zhou Z., Das S. R. and Gu Q., "Connected Sensor Cover: Self-Organization of Sensor Networks for Efficient Query Execution," *IEEE/ACM Transactions on Networking,* Vol. 14, No. 1, January 2006.
[3] Chang J, Tassiulas L. 2004. Maximum Lifetime Routing in Wireless Sensor Networks. *IEEE/ACM Transactions on Networking* August 2004; 12(4):609-619.
[4] Zussman G, Segall A. 2003. Energy Efficient Routing in Ad Hoc Disaster Recovery Networks. *Proc. IEEE Infocom 2003* April 2003.
[5] Simic SN, Sastry S. 2003. Distributed Environmental Monitoring Using Random Sensor Networks. *Proc. of IPSN'03* 2003.
[6] Li N, Hou JC. 2004. Topology Control in Heterogeneous Wireless Networks: Problems and Solutions. *Proc. IEEE Infocom 2004* March 2004.
[7] Narayanaswamy S, Kawadia V, Sreenivas RS, Kumar PR. 2002. Power Control in Ad-Hoc Networks: Theory, Architecture, Algorithm and Implementation of the COMPOW Protocol. *Proc. of European Wireless 2002* February 2002.
[8] Brown TX, Gabow HN, Zhang Q. 2001. Maximum Flow-Life Curve for a Wireless Ad Hoc Network. *Proc. of Symposium on Mobile Ad Hoc Networking and Computing* October 2001.
[9] Hou YT, Shi Y, Sherali HD. 2004. Rate allocation in wireless sensor networks with network lifetime requirement. *Proc. of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2004)* May 2004; 67-77.
[10] Ahuja RK, Magnanti TL, Orlin JB. *Network flows : theory, algorithms, and applications*. Prentice Hall: Englewood Cliffs, N.J., 1993.
[11] Sadagopan N, Krishnamachari B. Maximizing Data Extraction in Energy-Limited Sensor Networks. *Proc. IEEE Infocom 2004* March 2004.
[12] Garg N, Konemann J. Faster and Simpler Algorithms for Multicommodity Flow and Other Fractional Packing Problems. *FOCS '98* November 1998.
[13] http://www.xbow.com, July 2006.