

Exact and Approximate Balanced Data Gathering in Energy-Constrained Sensor Networks¹

Patrik Floréen^a, Petteri Kaski^b, Jukka Kohonen^a,
Pekka Orponen^b

^a *Helsinki Institute for Information Technology HIIT, Basic Research Unit,
Department of Computer Science, P.O. Box 68,
FI-00014 University of Helsinki, Finland.*

^b *Laboratory for Theoretical Computer Science, P.O. Box 5400,
FI-02015 Helsinki University of Technology, Finland.*

Abstract

We consider the problem of gathering data from a wireless multi-hop network of energy-constrained sensor nodes to a common base station. Specifically, we aim to balance the total amount of data received from the sensor network during its lifetime against a requirement of sufficient coverage for all the sensor locations surveyed. Our main contribution lies in formulating this balanced data gathering task, studying the effects of balancing, and proposing an approximation algorithm for the problem. Based on an LP network flow formulation, we present experimental results on both optimal and approximate data routing designs, in open transmission ranges and with impenetrable obstacles between the nodes.

1 Introduction

Wireless networks consisting of a large number of miniature electromechanical devices with sensing, computing and communication capabilities are rapidly

Email addresses: patrik.floreen@cs.helsinki.fi (Patrik Floréen),
petteri.kaski@hut.fi (Petteri Kaski), jukka.kohonen@cs.helsinki.fi (Jukka Kohonen), pekka.orponen@hut.fi (Pekka Orponen).

¹ A preliminary version of this paper has been presented at the First International Workshop on Algorithmic Aspects of Wireless Sensor Networks (ALGOSENSORS 2004), Turku, Finland, July 2004. Research supported by the Academy of Finland, Grants 202203 (PF and JK), 202205 (PK) and 204156 (PO).

becoming a reality, due to the accumulation of advances in digital electronics, wireless communications and microelectromechanical technology [11,12,17,29]. Prospective applications of such devices cover a wide range of domains [2,7,9,10,23].

One generic type of application for sensor networks is the continuous monitoring of an extended geographic area at relatively low data rates [2,5,32]. The information provided from all points of the sensor field is then gathered via multi-hop communications to a base station for further processing. We are here envisaging a scenario where environmental data are frequently and asynchronously collected over an area, and all information is to be gathered for later postprocessing of best possible quality, including detection of possibly faulty data. This means that data aggregation [24,25] cannot be employed.

Significant design constraints are imposed by multi-hop routing and the limited capabilities and battery power available at the sensor nodes. A number of recent papers have addressed e.g. optimal sensor placement [6,13,14,22,28] and energy-efficient routing designs and protocols [8,16,20,24,26,27,31] with the objective of maximizing lifetime [3,4,24,28] or data volume [21].

We envisage the sensor placement to be fixed beforehand, either by an application expert according to the needs of the particular application at hand, or randomly, for example by scattering the sensors from an airplane. For the sake of achieving a comprehensive view of the whole area to be monitored, not only should the total amount of data received at the base station be maximized, but the different sensors should be able to get through to the base station some minimum amount of data.

The idea of incorporating a certain balancing requirement on the data gathering has also recently been proposed in [26,27] and in [13], as well as in the preliminary conference version of this paper [18]. In [26,27] the authors put forth a more general model of information extraction that takes into account the nonlinear relation between transmission power and information rate. Our problem formulation can be seen as a linearized, computationally feasible version of this approach. Another difference between [26,27] and our work pertains to the expression of the balancing, or fairness, requirement. Article [13] considers the problem of maximizing the lifetime of a sensor network, and explicates this task in terms of an integer program that counts the number of “rounds” the network is operational, assuming that each sensor sends one data packet in each round. This formulation entails a strict fairness condition among the sensors, requiring them all to send exactly equal amounts of data. We allow an adjustable trade-off between maximizing the total amount of data received at the base station and the minimum amount of data received from each sensor. Moreover our program formulation does not require integer variables.

It should be noted that our linear-programming based solution relies on in-

formation about the energy costs of transmitting and receiving a unit of data between each pair of nodes, and about the energy supply at each node. This information suffices; knowledge of node locations as such is not required. Our model readily adjusts to obstacles and other deviations from simple radio-link models as long as the transmission and reception costs can be determined by the nodes themselves, either by simply probing at different power levels, or using more sophisticated means such as a Received Signal Strength Indicator (RSSI) [30].

Our linear program formulation also requires all the information to be available at a single location. This assumption is realistic only if the operation time of the network is long and the amount of control traffic small. Otherwise, routing decisions must be made based on local information. Our results thus provide an upper bound on what is actually achievable using distributed protocols with local information.

This paper is organized as follows. In Sect. 2 we formulate the balanced data gathering task as an optimization problem that admits an exact solution via linear programming. In Sect. 3 we extend our previous work in [18] by designing a faster flow-based algorithm that, for any approximation ratio $\alpha > 1$, finds an approximate solution in time $O(N^3 \log N)$, where N is the number of nodes in the network. In Sect. 4 we present experimental results for a number of different network topologies. For large instances, the experiments show that the approximation algorithm outperforms linear programming in terms of execution time. The paper is concluded in Sect. 5.

2 Optimization of balanced data gathering

A sensor network consists of three types of nodes. The sensor nodes (sources) generate data, which is to be gathered via multi-hop transmission to a base station (sink). The network may also contain relay nodes that only forward, but do not generate data. Each node has a limited supply of energy, which constrains its ability to receive and transmit data.

Formally, let V be the set of all nodes in the network, consisting of disjoint subsets S (the source nodes), R (the relay nodes) and $\{t\}$ (the sink). We denote by N the total number of nodes, and by n the number of source nodes. Each node $i \in V$ has an initial energy supply of E_i ; as a special case, we may set $E_t = \infty$. We assume that the actual transmissions are infrequent enough for collisions and signal interference not to occur. Furthermore, assuming that the sensors generate data asynchronously and in small packets, the process can be modeled as a flow.

<p>maximize $F_\lambda = (1 - \lambda) \frac{1}{n} \sum_{i \in S} q_i + \lambda \min_{i \in S} q_i$</p> <p>subject to</p> $\sum_{j \in V} f_{tj} = 0, \tag{1}$ $\sum_{j \in V} f_{ij} = q_i + \sum_{j \in V} f_{ji}, \quad i \in S, \tag{2}$ $\sum_{j \in V} f_{ij} = \sum_{j \in V} f_{ji}, \quad i \in R, \tag{3}$ $f_{ij} \geq 0, \quad i, j \in V, \tag{4}$ $f_{ii} = 0, \quad i \in V, \tag{5}$ $q_i \geq 0, \quad i \in S, \tag{6}$ $\sum_{j \in V} \tau_{ij} f_{ij} + \sum_{j \in V} \rho f_{ji} \leq E_i, \quad i \in V. \tag{7}$
--

Fig. 1. Model for balanced data gathering.

The flow-based model for data gathering is depicted in Fig. 1. The variable q_i , where $i \in S$, models the quantity of data generated at source node i and ultimately received at the sink. The flow variable f_{ij} , where $i, j \in V$, models the data transmitted from node i and received at node j .

Formally, a *flow* f is a function that associates a nonnegative value f_{ij} to every edge ij , where $i, j \in V$, such that the constraints from (1) to (6) are satisfied. The quantity variables q_i can be expressed in terms of the flow as $q_i(f) = \sum_j f_{ij} - \sum_j f_{ji}$, $i \in S$.

A flow is *feasible* if it also satisfies the energy constraint (7). The energy cost of transmitting a unit of data from node i to node j is given by a parameter τ_{ij} (transmission cost) and the cost of receiving a unit of data is given by a parameter ρ (reception cost).

Our model places no restrictions on the values of the parameters τ_{ij} and ρ . As an example, in the commonly used simple radio-link models [30], τ_{ij} would be taken to be $\tau_{\text{elec}} + \tau_{\text{amp}} \cdot d_{ij}^\nu$, where τ_{elec} corresponds to the energy consumed by the transmitter electronics and $\tau_{\text{amp}} \cdot d_{ij}^\nu$ corresponds to the energy consumed by the transmit amplifier to achieve an acceptable signal-to-noise ratio at the receiving node; d_{ij} is the physical distance between nodes i and j and the path loss exponent ν , $2 \leq \nu \lesssim 4$, models the decay of the radio signal in the ambient medium. The parameter ρ corresponds to the energy consumed by the receiver electronics.

One goal of the data flow design for the network is to maximize the total quantity of data, or equivalently, the average $\text{avg}_{i \in S} q_i = \frac{1}{n} \sum_{i \in S} q_i$. However, taking this as the singular objective may lead to the “starvation” of some of

the sensor nodes: typically, the average data quantity objective is maximized by data flows that only forward data generated close to the sink, and do not expend any energy on relaying data generated at distant parts of the network.

To counterbalance this tendency, we take as another objective to maximize the minimum quantity $\min_{i \in S} q_i$. The two objectives are combined as a utility function

$$F_\lambda := (1 - \lambda) \operatorname{avg}_{i \in S} q_i + \lambda \min_{i \in S} q_i, \quad (8)$$

where the balancing parameter λ , $0 \leq \lambda \leq 1$, controls the trade-off between the two conflicting objectives. If strict balancing between sensors is desired, $\lambda = 1$ can be selected.

Even though the utility function F_λ is not linear in the variables q_i , the problem can be converted into a linear program (LP) by replacing the term $\min_{i \in S} q_i$ with an additional variable μ , and adding constraints $q_i \geq \mu$ for all $i \in S$. The linear program can then be solved using standard techniques, giving an optimal data gathering flow solution for a given λ . Note that a linear programming approach is taken also in [13,24,28,31].

We remark that the model here differs slightly from our earlier model [18] in that each source has unlimited capacity of generating data. If desired, an upper bound on q_i can easily be introduced by splitting each source node i into a pair consisting of a source node i and a relay node i' , where $\tau_{ii'} = 1$ and E_i is the desired upper bound.

In practice, an optimal feasible flow f can be used to route approximately q_i unit-size data packets from each source node $i \in S$ to the sink node, assuming that all the q_i and f_{ij} values are large. At each node i , simply forward the first $\lfloor f_{i1} \rfloor$ packets to node 1, the next $\lfloor f_{i2} \rfloor$ packets to node 2, the next $\lfloor f_{i3} \rfloor$ packets to node 3, and so on. A somewhat more elegant solution is to randomize the routing strategy, so that each incoming packet at node i is forwarded to node j with probability $f_{ij} / \sum_{j'} f_{ij'}$.

3 An approximation algorithm

In this section we develop a polynomial-time approximation algorithm for the balanced data gathering problem in Fig. 1. The algorithm is based on the approximation framework for fractional packing problems in [19]. By a fractional packing problem we mean a linear program of the form $\max\{c^T x : Ax \leq b, x \geq 0\}$, where the matrix A and the vector b are nonnegative.

To arrive at a packing problem equivalent to the balanced data gathering

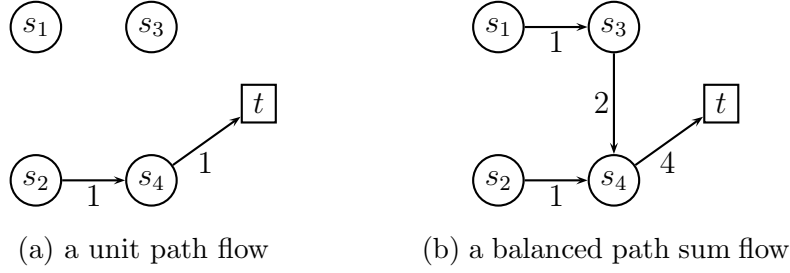


Fig. 2. Examples of elementary flows.

problem, we require some further definitions and lemmata. A flow f is *balanced* if $q_i(f) = q_j(f)$ for all $i, j \in S$. The *support* of a flow is the set of all edges ij carrying positive flow. A flow is *acyclic* if its support is acyclic. A *unit path flow* (from a node $i \in S$) is a flow such that $q_i = 1$ and the support of the flow is a path from i to t . A *balanced path sum flow* is a sum of unit path flows, one from each source node $i \in S$. An *elementary flow* is a unit path flow or a balanced path sum flow (see Fig. 2). We write \mathcal{E} for the set of all elementary flows. Similarly, we write \mathcal{E}_p and \mathcal{E}_b for the sets of all unit path flows and balanced path sum flows, respectively. To avoid degenerate cases, we assume $n > 1$ so that $\mathcal{E}_p \cap \mathcal{E}_b = \emptyset$.

Lemma 1 *For every (feasible) flow f , there exists an acyclic (feasible) flow f' such that $q_i(f) = q_i(f')$ for all $i \in S$.*

Proof. If f is acyclic, we are done. Otherwise, let C be a cycle in the support of f . Let $c = \min\{f_{ij} : ij \in C\}$. Decreasing the flow on every edge of C by c results in a flow f' with at least one less edge in its support and $q_i(f) = q_i(f')$ for all $i \in S$. Repeating the transformation if necessary eventually gives an acyclic flow. Since flow is only decreased, the transformation cannot produce an infeasible flow from a feasible flow. \square

Lemma 2 *Every acyclic flow f can be expressed as a nonnegative linear combination $f = \sum_{e \in \mathcal{E}} x_e e$ such that $\min_{i \in S} q_i(f) = \sum_{e \in \mathcal{E}_b} x_e$.*

Proof. If f is the zero flow, we are done. Otherwise, we decompose the flow step by step into elementary flows so that each step k removes at least one edge from the support of the flow. Thus, there are at most $N(N-1)/2$ steps because the support is acyclic. Initially, put $f^{(0)} = f$ and $x_e = 0$ for all $e \in \mathcal{E}$. The iteration step is divided into two cases.

If $\min_{i \in S} q_i(f^{(k)}) > 0$, we proceed as follows. As $q_i(f^{(k)}) > 0$ there exists a unit path flow from i whose support is contained in the support of f . Let $e \in \mathcal{E}_b$ be a balanced path sum flow obtained as the sum of such unit path flows, one for every $i \in S$. Set x_e to the maximum value such that $x_e e_{ij} \leq f_{ij}^{(k)}$ for all $i, j \in V$, and let $f^{(k+1)} = f^{(k)} - x_e e$ be the flow for the next iteration. Because

e carries one unit of flow from every source node $i \in S$ to the sink t , we have $q_i(f^{(k+1)}) = q_i(f^{(k)}) - x_e$ for all $i \in S$.

Otherwise; that is, when $\min_{i \in S} q_i(f^{(k)}) = 0$ and $f^{(k)}$ is still nonzero, there exists a node $i \in S$ with $q_i(f^{(k)}) > 0$ because the support of $f^{(k)}$ is acyclic. In this case, let $e \in \mathcal{E}_p$ be a unit path flow from i such that the support of e is contained in the support of $f^{(k)}$. Set x_e to the maximum value such that $x_e e_{ij} \leq f_{ij}$ for all $i, j \in V$, and let $f^{(k+1)} = f^{(k)} - x_e e$ be the flow for the next iteration.

Let u be the least integer such that $\min_{i \in S} q_i(f^{(u)}) = 0$. By the structure of the iteration, $0 = \min_{i \in S} q_i(f^{(u)}) = \min_{i \in S} q_i(f^{(0)}) - \sum_{e \in \mathcal{E}_b} x_e e = \min_{i \in S} q_i(f) - \sum_{e \in \mathcal{E}_b} x_e$. \square

Define the constraint matrix A so that the entry a_{ie} is the cost for node $i \in V$ for receiving and transmitting the elementary flow $e \in \mathcal{E}$; that is, $a_{ie} = \sum_{j \in V} \tau_{ij} e_{ij} + \sum_{j \in V} \rho_{eji}$. Furthermore, put $b_i = E_i$ for all $i \in V$.

Lemma 3 *A flow $f = \sum_{e \in \mathcal{E}} x_e e$ is feasible if and only if $Ax \leq b$.*

Proof. Substitute the expression for f into (7). \square

To complete the packing problem, we must express the functional

$$F_\lambda(f) = (1 - \lambda) \frac{1}{n} \sum_{i \in S} q_i(f) + \lambda \min_{i \in S} q_i(f) \quad (9)$$

for a flow $f = \sum_{e \in \mathcal{E}} x_e e$ in terms of the elementary flows. Applying

$$\sum_{i \in S} q_i(e)/n = \begin{cases} 1/n & \text{if } e \in \mathcal{E}_p; \text{ and} \\ 1 & \text{if } e \in \mathcal{E}_b \end{cases}$$

for all $e \in \mathcal{E}$, we obtain

$$F_\lambda(f) = (1 - \lambda) \frac{1}{n} \sum_{e \in \mathcal{E}_p} x_e + (1 - \lambda) \sum_{e \in \mathcal{E}_b} x_e + \lambda \min_{i \in S} \left(\sum_{e \in \mathcal{E}} x_e q_i(e) \right).$$

For all $i \in S$ and $e \in \mathcal{E}_b$, it holds that $q_i(e) = 1$. Thus

$$\sum_{e \in \mathcal{E}} x_e q_i(e) \geq \sum_{e \in \mathcal{E}_b} x_e,$$

implying

$$F_\lambda(f) \geq (1 - \lambda) \frac{1}{n} \sum_{e \in \mathcal{E}_p} x_e + \sum_{e \in \mathcal{E}_b} x_e. \quad (10)$$

In particular, equality holds in (10) if a decomposition given by Lemma 2 is used for f . These observations lead to the following packing objective function. Associate with every elementary flow $e \in \mathcal{E}$ the coefficient

$$c_e = \begin{cases} (1 - \lambda)/n & \text{if } e \in \mathcal{E}_p; \text{ and} \\ 1 & \text{if } e \in \mathcal{E}_b. \end{cases}$$

Thus, (10) is equivalent to $F_\lambda(f) \geq c^T x$, and equality holds if a decomposition given by Lemma 2 is used for f .

Theorem 4 *Let x be an optimal solution to the packing linear program*

$$\max\{c^T x : Ax \leq b, x \geq 0\}.$$

Then, $f = \sum_{e \in \mathcal{E}} x_e e$ is an optimal solution to the balanced data gathering problem in Fig. 1.

Proof. By Lemma 3, f is feasible. To reach a contradiction, suppose that f is not optimal. Then, there exists a feasible flow f' with $F_\lambda(f') > F_\lambda(f)$. By Lemma 1 we can assume that f' is acyclic. Let $f' = \sum_{e \in \mathcal{E}} x'_e e$ be a decomposition given by Lemma 2. By Lemma 3, x' is a feasible solution to the packing linear program. Furthermore, since equality holds in (10) for f' , we have $c^T x' = F_\lambda(f') > F_\lambda(f) \geq c^T x$, a contradiction. \square

Initialization Given the approximation parameter $\alpha > 1$, put $\epsilon = 1 - \alpha^{-1/2}$ and $\delta = (1 + \epsilon)((1 + \epsilon)N)^{-1/\epsilon}$. Initialize $x_e = 0$ for all $e \in \mathcal{E}$, and let $y_i = \delta/b_i$ for all $i \in V$. Start the iteration.

Iteration Repeat the following four steps in sequence until $b^T y \geq 1$.

- (1) Let $g = \arg \min_{e \in \mathcal{E}} \sum_{i \in V} a_{ie} y_i / c_e$.
- (2) Let $k = \arg \min_{i \in V} b_i / a_{ig}$.
- (3) Update $x_g \leftarrow x_g + b_g / a_{kg}$.
- (4) Update $y_i \leftarrow y_i \cdot (1 + \epsilon \cdot a_{ig} b_k / (a_{kg} b_i))$ for all $i \in V$.

Termination Multiply all components of x by $1/\log_{1+\epsilon}((1 + \epsilon)/\delta)$ and report x as an approximate optimal solution. Stop.

Fig. 3. The packing approximation algorithm adapted from [19].

Figure 3 shows the approximate solution algorithm for packing LPs from [19] that has been adapted to the present context. The following theorem is an immediate consequence of [19, Theorem 3.1].

Theorem 5 *For any given $\alpha > 1$, the algorithm in Fig. 3 computes an α -approximation to the balanced data gathering problem in time $N \lceil \epsilon^{-1} \log_{1+\epsilon} N \rceil m(N)$, where $\epsilon = 1 - \alpha^{-1/2}$ and $m(N)$ is the time required by steps (1) through (4) of the iteration.*

As it turns out, we can achieve $m(N) = O(N^2)$ by applying the single-source shortest paths algorithm of Dijkstra [15] on an auxiliary directed graph with nonnegative edge weights. For a nonnegative vector y with components indexed by V , let $H(y)$ be the complete directed graph with vertex set V such that for all $i, j \in V$, the weight of edge ij is $w_{ij} = \tau_{ij}y_i + \rho y_j$. As an immediate consequence, we have:

Lemma 6 *Let $e \in \mathcal{E}$ have support T . Then, $\sum_{ij \in T} e_{ij} w_{ij} = \sum_{i \in V} a_{ie} y_i$.*

In particular, a unit path flow $e \in \mathcal{E}_p$ that minimizes $\ell(e) = \sum_{i \in V} a_{ie} y_i$ corresponds to a shortest path from a source node to the sink node t in $H(y)$. Reversing the direction of the edges in $H(y)$, one shortest path computation originating from t allows us to obtain ℓ -minimizing unit path flows from all source nodes. The sum of such minimum unit path flows over the source nodes is an ℓ -minimizing balanced path sum flow. Because the coefficients c_e are constant over the unit path flows and the balanced path sum flows, respectively, these observations enable us to determine in time $m(N) = O(N^2)$ an elementary flow $e \in \mathcal{E}$ that minimizes $\sum_{i \in V} a_{ie} y_i / c_e$.

We conclude this section by noting that although network flow problems of various kinds have been extensively studied (see [1]), the present problem differs from the traditional setting. Here a flow through a node incurs a local energy cost, which must not exceed the energy available at the node. These constraints are inherently local, in contrast to the global cost functional encountered in a traditional minimum-cost flow problem. Furthermore, the combined utility function F_λ differs from the standard maximum-flow type setting. Compared with the augmenting path decomposition, we decompose a flow into two types of elementary flows having different contributions to the utility function.

In analogy to the max-flow min-cut characterization for a maximum flow, it would be of interest to develop a fully combinatorial characterization of an optimum solution in the present context. Such a characterization, if one exists, is likely to lead to a fully combinatorial exact algorithm for the balanced data gathering problem.

4 Experimental results

4.1 Exact solutions and the effect of balancing

We first study the effect of balanced data gathering in various kinds of networks. In each network, the exact optimum flow is found by solving the linear program formed from the model shown in Fig. 1, using the `linprog` routine

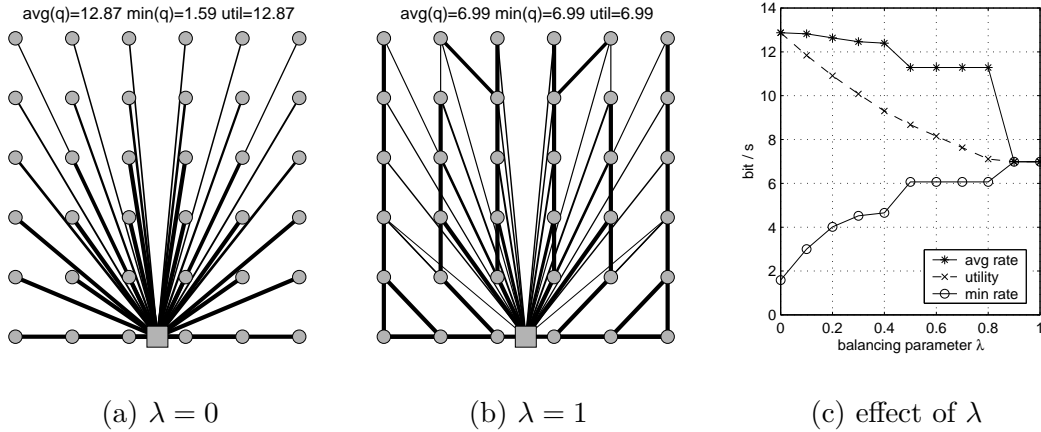


Fig. 4. Regular grid sensor network in an unobstructed area.

of MATLAB's Optimization Toolbox.

In each simulation experiment, the network consists of 36 sensor nodes, in a square area of dimensions $1 \text{ km} \times 1 \text{ km}$, and the base station located at the center of the south side. All sensors have an energy constraint of 20 J. Transmission and reception costs are computed as $\tau_{ij} = 100 \frac{\text{nJ}}{\text{bit}} + 0.01 \frac{\text{nJ}}{\text{bit m}^2} \cdot d_{ij}^2$, and $\rho = 100 \frac{\text{nJ}}{\text{bit}}$. These values are comparable to those in [4,20,21].

In Fig. 4 the area is unobstructed and the sensor nodes are placed in a regular 6×6 grid. If no balancing is required ($\lambda = 0$), each node selfishly transmits as much of its own data as its energy constraint allows, leading to very small data quantities from nodes far away from the base station. As higher balancing parameters are used, the distant nodes get a larger share of the network's transport resources, and accordingly, the area is more evenly covered by observations. This comes at a cost of reducing the average data quantity. It should be noted that with intermediate values of λ , the solutions are not simply linear combinations of the two extreme cases. At $\lambda = 0.5$ the minimum quantity is already increased almost by a factor of four compared to the $\lambda = 0$ case, while the average quantity is reduced by only 12%.

As an example of nonuniformly distributed sensors, in Fig. 5 most of the area is covered by a lake where no sensor nodes can be placed; the sensors are placed randomly around it. However, radio transmission is unaffected by the lake. If balancing is required, much of the flow is routed around the lake, where it can be sent over shorter links than those crossing the lake.

Since our model allows arbitrary transmission costs τ_{ij} , it is not restricted to idealized transmission conditions like the popular unit disk model. We can study situations where there are large impenetrable obstacles within the area, as illustrated in Fig. 6. This is done by assigning an infinite transmission cost to any link that intersects an obstacle. The sensors are again randomly placed.

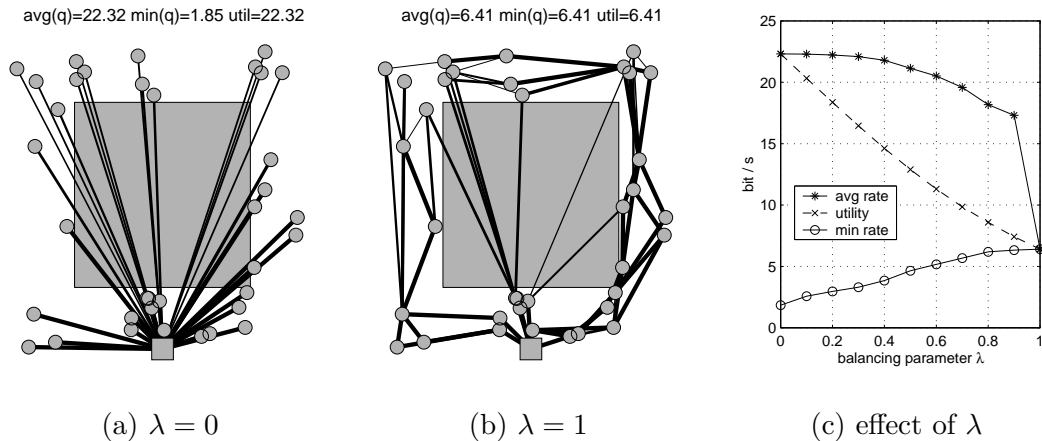


Fig. 5. Random sensor network around a square-shaped lake.

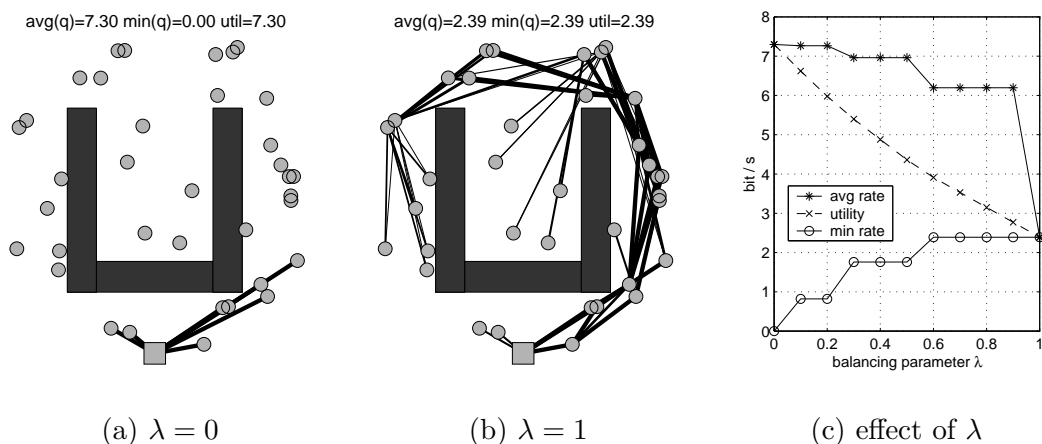


Fig. 6. Random sensor network around a U-shaped impenetrable wall.

In the $\lambda = 0$ case, the few sensors that have line-of-sight to the base station expend all of their energy for their own data, with the result that sensors behind the wall are unable to transmit anything (minimum quantity equals 0). When λ is increased, sensors begin to co-operate, and all sensors can reach the base station through multihop transmission.

Finally, in Fig. 7 we study a different kind of obstruction. Scattered throughout the area there are 100 small square obstacles that block transmission. By this we intend to model difficult environment, such as a dense forest. Due to the large number of these small obstacles, each node can directly see some of the other nodes (on the average 12 out of the 36), and only a few sensors have line-of-sight to the base station. Balanced data gathering requires routing the flow around the obstacles in a nontrivial way, as can be seen in the figure.

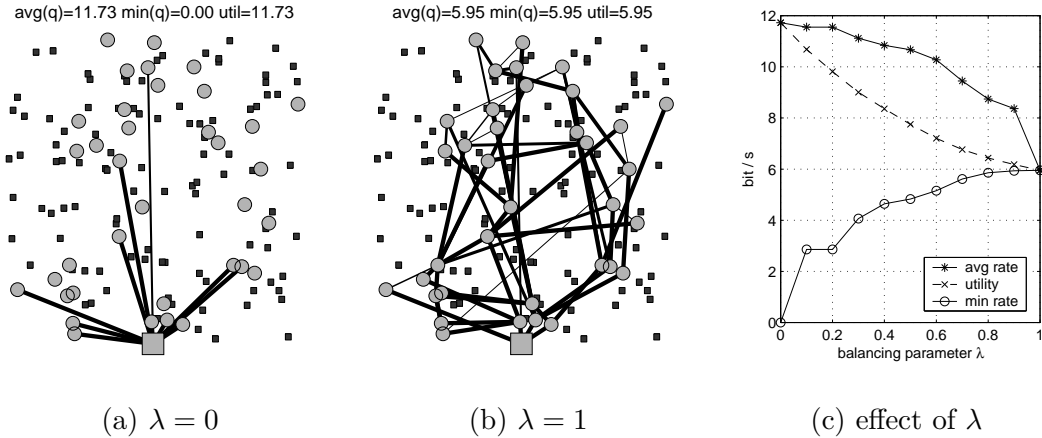


Fig. 7. Random sensor network in an area containing small, randomly scattered obstructions.

4.2 Approximate solutions

The next set of experiments concerns the applicability of the approximation algorithm from Sect. 3 to balanced data gathering. The algorithm was implemented in C, and applied to network instances similar to those in the previous section.

Keeping in mind that the algorithm admits an arbitrary approximation ratio parameter α , it is interesting to study how varying its value affects the solution quality and the running time of the algorithm. Another question of interest is whether the approximation algorithm can successfully handle large network instances, for which solving the LP exactly is impractical.

For the first experiment, we take random networks of various sizes (36, 64, 81, 100, 144 and 196 sensors) surrounding an impenetrable wall, as in Fig. 6 in the previous section. The balancing parameter is fixed at $\lambda = 0.5$. The instances are solved both exactly with an LP solver, and with the approximation algorithm for values $\alpha \in \{1.1, 1.2, 1.5\}$. The results are shown in Fig. 8.

It can be seen that the actually achieved approximation ratio (that is, the ratio of utility F_λ between the exact solution and the approximate solution) is systematically smaller than the bound α guaranteed by Theorem 5.

According to Theorem 5, for a given α , the approximation algorithm runs in time $O(N^3 \log N)$. The experimental results are in line with this. Note that if very high precision is desired, solving the LP exactly may be faster than running the approximation algorithm with a very small value of α ; the break-even point naturally depends on the implementation. However, for e.g. $\alpha = 1.5$, the approximation algorithm provides quite reasonable approximations and

runs (in our experiments) significantly faster than the linear program solver.

We also conducted the same experiment using the other network topologies from Sect. 4.1: the grid network in an unobstructed area, the random network around a lake, and the random network in a densely obstructed area. In all cases, the results were similar to those shown above. Approximation ratios achieved in practice were systematically better than the bound α ; for example, $\alpha = 1.5$ usually gives ratios between 1.2 and 1.25.

In the last experiment, we study how the performance of the approximation algorithm depends on the obstruction density in a randomly obstructed network. For this experiment, we first place 300 small obstructions at random, and 100 sensors at random in the remaining area. We then remove the obstructions gradually, 50 at a time, leading finally to an unobstructed network. The balancing parameter is fixed at $\lambda = 0.5$ and the approximation ratio at $\alpha = 1.5$. The results are illustrated in Fig. 9. The approximation algorithm finds good approximations in all cases. The achieved approximation ratio varies between 1.12 and 1.27.

As expected, the utility attained in the network decreases as more links are obstructed. However, even in a densely obstructed situation, such as the one shown in Fig. 9(a), the network is able to attain about one fourth of the utility of the unobstructed network. It should be noted that with 300 obstacles, transporting the flow is quite difficult, as each node can see only a few other nodes (9 on the average), and some nodes can see only one other node.

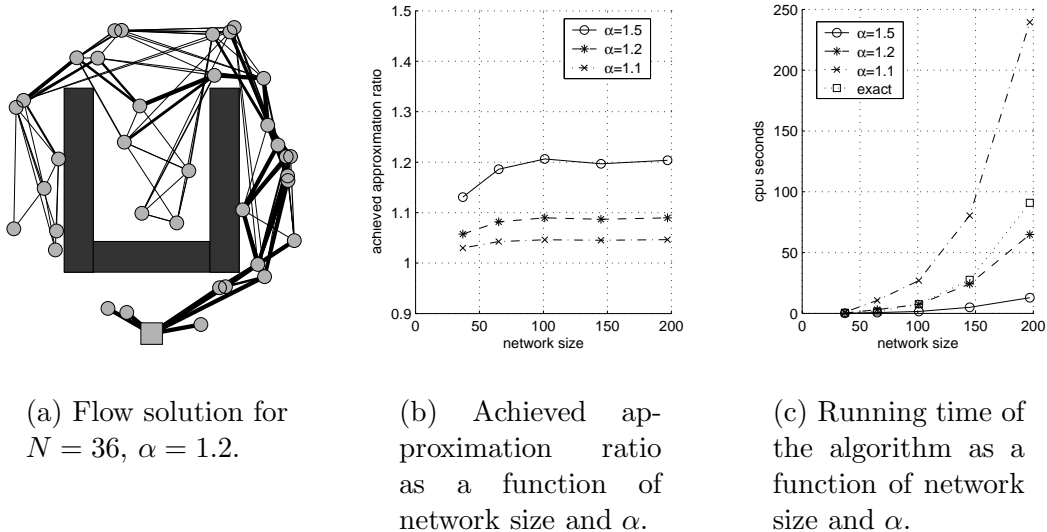
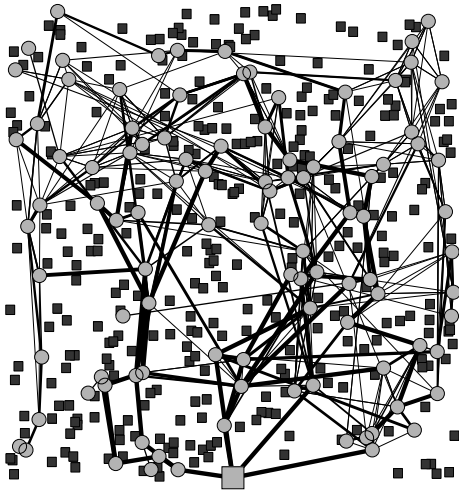
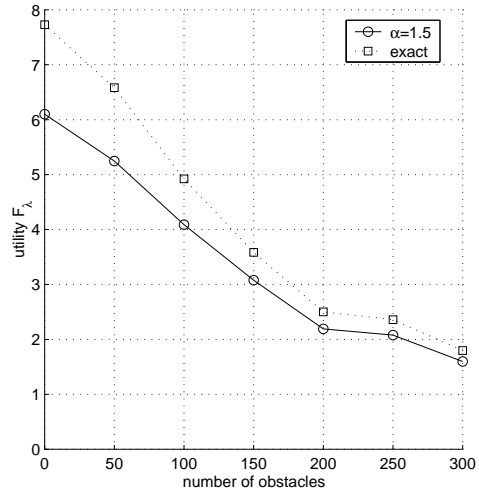


Fig. 8. Approximate solutions for flow maximization in the case $\lambda = 0.5$. Random sensor networks of various sizes, around a U-shaped impenetrable wall.



(a) Approximate flow solution for 300 obstacles.



(b) Utility F_λ as a function of number of obstacles.

Fig. 9. Effect of obstacle density. Random sensor network in an area scattered with random obstructions.

5 Conclusions

We have considered the problem of energy-efficient data gathering in sensor networks, with special emphasis on the goal of balancing the average volume of data collected against sufficient coverage of the monitored area. We have formulated a linear programming model of the task of finding optimal routes for the data produced at the sensor nodes, given a balancing requirement in terms of a balancing parameter $\lambda \in [0, 1]$.

Experiments with the model show that for reasonable values of the balancing parameter, a significant increase in coverage is achieved, without any great decrease in the average amount of data gathered per node.

We have also developed a polynomial-time approximation algorithm for the task, using an approximation framework for fractional packing problems. Our experiments confirm that the approximation algorithm is computationally feasible.

In our experiments with obstacles, sensor networks were seen to be fairly robust against even a fairly high number of obstructions. This was achieved through the use of global optimization at a central location, where information about all link costs in the network was available. It remains to be studied how closely this global optimum can be approximated by distributed algorithms that have access to local information only. The effect of possible node faults

during the operation of the network is also a topic for further research.

References

- [1] R. K. Ahuja, T. L. Magnanti, J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Upper Saddle River NJ, 1993.
- [2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, “Wireless sensor networks: a survey.” *Computer Networks* 38 (2002), 393–422.
- [3] M. Bhardwaj, A. Chandrakasan, “Bounding the lifetime of sensor networks via optimal role assignment.” *Proc. 21st Ann. Joint Conf. of the IEEE Computer and Communications Societies (InfoCom’02, New York NY, June 2002)*, Vol. 3, 1587–1596. IEEE, New York NY, 2002.
- [4] M. Bhardwaj, T. Garnett, A. Chandrakasan, “Upper bounds on the lifetime of sensor networks.” *Proc. IEEE Intl. Conf. on Communications (ICC’01, Helsinki, Finland, June 2001)*, Vol. 3, 785–790. IEEE, New York NY, 2001.
- [5] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, J. Zhao, “Habitat monitoring: application driver for wireless communication technology.” *Proc. 2001 ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean (San José, Costa Rica, April 2001)*, 20–41. ACM Press, New York NY, 2001.
- [6] K. Chakrabarty, S. S. Iyengar, H. Qi, E. Cho, “Grid coverage for surveillance and target location in distributed sensor networks.” *IEEE Transactions on Computers* 51 (2002), 1448–1453.
- [7] C.-Y. Chong, S. P. Kumar, “Sensor networks: evolution, opportunities, and challenges.” In [29], pp. 1247–1256.
- [8] M. Chu, H. Haussecker, F. Zhao, “Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks.” *Intl. J. on High Performance Computing Applications* 16 (2002), 293–313.
- [9] *Communications of the ACM* 43 (2000), May. Special Section on Embedding the Internet.
- [10] *Communications of the ACM* 45 (2002), December. Special Section on Issues and Challenges in Ubiquitous Computing.
- [11] *Communications of the ACM* 47 (2004), June. Special Section on Wireless Sensor Networks.
- [12] *Computer* 37 (2004), August. Cover Feature: Sensor Networks.
- [13] K. Dasgupta, M. Kukreja, K. Kalpakis, “Topology-aware placement and role assignment for energy-efficient information gathering in sensor networks.” *Proc. 8th IEEE Symp. on Computers and Communication (ISCC’03, Kemer-Antalya, Turkey, July 2003)*, 341–348. IEEE, New York NY, 2003.

- [14] S. S. Dhillon, K. Chakrabarty, S. S. Iyengar, “Sensor placement for grid coverage under imprecise detections.” *Proc. 5th ISIF Intl. Conf. on Information Fusion (FUSION’02, Annapolis MD, July 2002)*, 1581–1587. International Society for Information Fusion, Chatillon, France, 2002.
- [15] E. W. Dijkstra, “A note on two problems in connexion with graphs.” *Numerische Mathematik 1* (1959), 269–271.
- [16] D. Estrin, R. Govindan, J. Heidemann, S. Kumar, “Next century challenges: scalable coordination in sensor networks.” *ACM/IEEE Intl. Conf. on Mobile Computing and Networks (MobiCom’99, Seattle WA, August 1999)*, 263–270. ACM Press, New York NY, 1999.
- [17] D. Estrin et al., *Embedded, Everywhere: A Research Agenda for Networked Systems of Embedded Computers*. The National Academies Press, Washington DC, 2001.
- [18] E. Falck, P. Floréen, P. Kaski, J. Kohonen, P. Orponen, “Balanced data gathering in energy-constrained wireless sensor networks.” *Algorithmic Aspects of Wireless Sensor Networks: Proc. First Intl. Workshop (ALGOSENSORS 2004, Turku, Finland, July 2004)*, 59–70. Springer-Verlag, LNCS 3121, Berlin, 2004.
- [19] N. Garg, J. Könemann, “Faster and simpler algorithms for multicommodity flow and other fractional packing problems.” *Proc. 39th Ann. Symp. on Foundations of Computer Science (Palo Alto, CA, Nov. 8–11, 1998)*, 300–309. IEEE Computer Society, Los Alamitos, CA, 1998.
- [20] W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks.” *Proc. 33rd Hawaii Intl. Conf. on System Sciences (HICSS’00, Maui HI, January 2000)*, 8020–8029. IEEE Computer Society, Los Alamitos CA, 2000.
- [21] B. Hong, V. K. Prasanna, “Constrained Flow Optimization with Applications to Data Gathering in Sensor Networks.” *Algorithmic Aspects of Wireless Sensor Networks: Proc. First Intl. Workshop (ALGOSENSORS 2004, Turku, Finland, July 2004)*, 187–200. Springer-Verlag, LNCS 3121, Berlin, 2004.
- [22] A. Howard, M. J. Matarić, S. Sukhatme, “Mobile sensor network deployment using potential fields: a distributed, scalable solution to the area coverage problem.” In H. Asama, T. Arai, T. Fukuda, T. Hasegawa (Eds.), *Distributed Autonomous Robotic Systems 5*. Springer-Verlag, Berlin, 2002.
- [23] *IEEE Spectrum 41* (2004), July. Special Report: Sensor Nation.
- [24] K. Kalpakis, K. Dasgupta, P. Namjoshi, “Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks.” *Computer Networks 42* (2003), 697–716.
- [25] B. Krishnamachari, D. Estrin, S. Wicker, “The impact of data aggregation in wireless sensor networks.” *Intl. Workshop on Distributed Event-Based Systems (DEBS’02, Vienna, Austria, July 2002)*, 575–578. IEEE, New York NY, 2002.

- [26] B. Krishnamachari, F. Ordóñez, “Analysis of energy-efficient, fair routing in wireless sensor networks through non-linear optimization.” *Proc. IEEE 58th Vehicular Technology Conference (VTC 2003-Fall, Orlando FL, October 2003)*, Vol. 5, 2844–2848. IEEE Computer Society, Los Alamitos CA, 2003.
- [27] F. Ordóñez and B. Krishnamachari, “Optimal information extraction in energy-limited wireless sensor networks.” *IEEE J. Selected Areas in Communications* 22 (6) (2004), Special Issue on Fundamental Performance Limits of Wireless Sensor Networks, 1121–1129.
- [28] J. Pan, Y. T. Hou, L. Cai, Y. Shi, S. X. Shen, “Topology control for wireless sensor networks.” *Proc. 9th Ann. Intl. Conf. on Mobile Computing and Networking (MobiCom’03)*, 286–299. ACM Press, New York NY, 2003.
- [29] *Proceedings of the IEEE* 91 (2003) 8 (August). Special Issue on Sensor Networks and Applications.
- [30] T. S. Rappaport, *Wireless Communications: Principles & Practice*. Prentice Hall, Upper Saddle River NJ, 1996.
- [31] N. Sadagopan, B. Krishnamachari, “Maximizing data extraction in energy-limited sensor networks.” *Proc. 23rd Conf. of the IEEE Computer and Communications Society (InfoCom’04, Hong Kong, March 2004)*.
- [32] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, D. Estrin, “Habitat monitoring with sensor networks.” In [11], pp. 34–40.