



# Measurement Aggregation and Routing Techniques for Energy-Efficient Estimation in Wireless Sensor Networks

Iordanis Koutsopoulos, Maria Halkidi

## ► To cite this version:

Iordanis Koutsopoulos, Maria Halkidi. Measurement Aggregation and Routing Techniques for Energy-Efficient Estimation in Wireless Sensor Networks. WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, May 2010, Avignon, France. pp.43-52. inria-00498339

**HAL Id: inria-00498339**

**<https://inria.hal.science/inria-00498339>**

Submitted on 7 Jul 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Measurement Aggregation and Routing Techniques for Energy-Efficient Estimation in Wireless Sensor Networks

Iordanis Koutsopoulos      Maria Halkidi

**Abstract**—Wireless sensor networks are fundamentally different from other wireless networks due to energy constraints and spatial correlation among sensor measurements. Mechanisms that efficiently compress and transport sensor data in the network are needed. We consider the problem of maximizing lifetime of wireless sensor networks that are entitled with the task of estimating an unknown parameter or process and thus need to adhere to estimation error specifications. We investigate optimal endogenous sensor measurement rate control, in-network data aggregation and routing for achieving the goal above. Sensors take measurements and aggregate incoming data from neighbors in a single outgoing flow by applying appropriate aggregation weights. By doing so, they control the variance of outgoing flow. Each sensor controls its measurement rate and aggregation weights, and aggregated measurement data are routed to the FC for Maximum Likelihood (ML) estimation.

The challenge is to find an optimal compromise between eliminating data redundancy and maintaining data representation accuracy so as to adhere to estimation quality constraints and reduce the volume of transported data, thus improving network lifetime. Sensor spatial correlation, measurement accuracies, link qualities and energy reserves affect sensor measurement rates, data aggregation and routes to the FC. On the other hand, measurement rates, aggregation, and sensor characteristics impact the estimation error. We show that the problem can be decomposed into separate optimization problems where each sensor autonomously takes its measurement rate, aggregation and routing decisions. We design an iterative primal-dual algorithm that relies on low overhead feedback from the FC to the nearest sensors, and on sensor neighbor Lagrange multiplier exchanges. Our work strikes the optimal fundamental tradeoff between network lifetime, in-network data aggregation and estimation quality and yields a solution based on distributed sensor coordination.

## I. INTRODUCTION

Wireless sensor networks are densely deployed for monitoring purposes such as estimation of a quantity, tracking of a process, or detection of an event. The fundamental characteristic of sensor networks that distinguishes them from other classes of wireless networks is the *spatial correlation* among measurements of co-located sensors. Two or more neighboring sensors will most likely perceive the event or

phenomenon in a similar manner and their measurements will be correlated. If the phenomenon is a point event or process, it will trigger only sensors around the phenomenon location, and these will provide correlated measurements, while sensors further away will provide much different measurements. The amount of correlation among sensor measurements depends on relative proximity of sensor nodes and their distance to the event source. The same situation arises when tracking a point process with time-varying location, e.g due to mobility. If the quantity to be estimated is spatially homogeneous, sensors in the same locality provide correlated measurements as well.

Performance guarantees in sensor networks can be a small estimation error, or small probabilities of false alarm and missed detection. Such operational objectives are mapped onto optimization objectives that are different from conventional ones such as end-to-end throughput maximization or delay minimization for which wireless networks are traditionally deployed. Network control actions should take into account these objectives so that the network performs the tasks it is entitled to with best performance while consuming energy in a prudent fashion and increasing network operational lifetime.

This paper studies the interplay between network lifetime and estimation quality by considering:

- optimal distributed endogenous sensor measurement rate control,
- in-network data aggregation,
- data routing to a fusion center (FC) that performs the estimation.

The inherent tradeoff is that a high measurement rate improves estimation quality, yet it is inefficient in terms of network lifetime since gathering data to the FC consumes more energy. Data aggregation can further aid in increasing network lifetime, if nodes appropriately fuse their own measurement data flow with those received by other sensors so as to produce one outgoing flow. The challenge in aggregation is to compromise data accuracy (captured by the variance of the outgoing flow) as little as possible by performing appropriate data processing, while considerably reducing the measurement load to be transferred in the network.

Furthermore, sensor measurement rate control and data aggregation need to take into account spatial correlation

Iordanis Koutsopoulos is with the Department of Computer and Communications Engineering, University of Thessaly, Volos, GR 38221, Greece. Email: jordan@uth.gr.

Maria Halkidi is with the Department of Digital Systems, University of Piraeus, GR 18534, Greece. Email: mhalk@unipi.gr.

of neighboring sensors and be in synergy with routing, so that low redundancy and high accuracy data are routed to the FC. Sensor characteristics, such as energy reserves and measurement qualities, as well as wireless link quality affect the measurement rate, aggregation and routing decisions. The effort should be focused on the requirement that the rate with which sensor energy is replenished is balanced across the network. On the other hand, measurement rates and data aggregation affect estimation quality, since they create different regimes about the joint probability distribution of measurements at the FC, and hence they affect the estimation error.

#### A. Related work

A seminal work in lifetime maximization for wireless sensor networks is the one in [1], where routing of sensor data with certain generation rates to multiple FCs is addressed. In this work, there is no consideration of estimation constraints or spatial correlation. Spatial correlation in wireless sensor networks has been studied in various contexts [2]. The work in [3] considers exploiting spatial correlation for designing efficient medium access control protocols with low contention overhead under a distortion constraint. The authors create a tessellation of the area through vector quantization and select a representative sensor to transmit from each region. The authors in [4] study sensor activation for maximizing a general utility function that captures network objectives for rechargeable sensors. In that paper, spatial correlation is taken into account through correlated discharge and recharge processes; that is, neighboring sensors that perceive a local phenomenon in a similar manner have their batteries emptied almost at similar times.

In the context of information theory, the premise is that spatially correlated sensor sources are jointly encoded or compressed to eliminate redundant information and transmit minimal necessary amount of bits. The problem becomes deciding on source coding rates, that is, numbers of bits that each source should transmit. In [5], spatial correlation is captured by conditional entropy, and a framework for joint compression of sensor data is proposed that minimizes total cost of data aggregation to sinks. In [6], a similar model is used for the min-cost joint source coding and routing under no contention. A similar problem is considered in [7], where the contributions are a metric that captures both mechanisms and a near-optimal clustering scheme. In [8], a distributed optimization method is proposed for joint source coding, routing and access control such that total energy is minimized.

Various aspects of estimation objectives have been considered in the literature [9]-[11]. In [9], the authors study energy minimization through finding the number of quantization levels for given estimation error. The work in [11] studies the impact of transmit power on estimation error. Power allocation for minimum estimation error subject to power constraint, or power minimization subject to an estimation

error constraint, possess the water-filling structure. At the optimal solution, only a subset of sensors transmit and the rest are off. Routing and sensor placement to maximize network lifetime was studied in [12] where the distortion depends on sensor positions relative to some cluster-heads. An effort to formulate the lifetime distortion tradeoff with rate - distortion theory was reported in [13]. Finally, [14] studies cooperative routing and defines a link metric to optimally aggregate data for detecting a random field.

Data aggregation in its simplest version is the operation by which multiple input packets at a node are aggregated into one output packet [15]. The problem of where to perform aggregation so as to minimize the number of transmissions per unit data is shown to be equivalent to the minimum Steiner tree problem, thus it is NP-Hard. The work in [16] proposes heuristics for the minimization of a cost function that depends both on the number of transmissions and the number of bits per transmission. In [17], [18], some clustering based heuristics are presented for the maximum lifetime data aggregation problem. The authors in [19] present a gradient-based distributed approach for routing and in-network aggregation. Aggregation is imposed by the fact that the fixed input rate at each node exceeds the sum of capacity of output links. The objective here is to minimize end-to-end distortion. A different view on aggregation is offered in [20]. The problem faced by each node is to decide when to perform aggregation of incoming arriving packets under the uncertainty of randomness of the arrival process and that of transmission channel availability. A stochastic control formulation addresses the tradeoff of waiting more and thus aggregate larger amount of incoming traffic versus respecting certain delay constraints. Finally, in [21], the authors consider the joint problem of data aggregation and node activation scheduling in order to reduce end-to-end delivery delay.

#### B. Our contribution

We study lifetime maximization in multi-hop wireless sensor networks under given estimation error specifications. The key novel idea is that, besides routing to the FC, we allow each sensor to control the *endogenous* number of measurements per unit time (that is, the measurement rate), the aggregation of input flows to a single outgoing flow, and the rate of this outgoing flow. The endogenous measurement rate is essentially the rate with which a sensor *samples* the process under estimation. Measurement rates and local aggregation decisions are influenced by sensor spatial correlations such that data redundancy is reduced. Since data aggregation removes some data redundancy to the expense of potentially increasing the uncertainty of the aggregated flow, it should be applied judiciously, for instance in sensors with large spatial correlation with their neighbors or if error specifications allow it. Input data flows should be appropriately combined and aggregated to a single output flow, such that this uncertainty is the minimum

possible and the aggregated flow does not compromise the estimation error requirement. Furthermore data aggregation and measurement rate control impact subsequent routing decisions and thus network lifetime, and ultimately affect estimation error at the FC.

Our contributions are as follows: (i) we capture the fundamental tradeoff between estimation accuracy and network lifetime, and we bring into the picture all msjor mechanisms that shape this tradeoff, namely sensor measurement rate control, data aggregation and routing, (ii) we formulate the optimization problem of maximum lifetime subject to a constraint on estimation error, (iii) we devise an iterative primal-dual distributed algorithm, where each sensor separately takes its measurement rate, aggregation and routing decisions with the aid of feedback from the neighbor sensors or the FC in the form of Lagrange multipliers. Sensors with a direct link to the FC receive direct feedback from the FC and their neighbors, and perform a different adaptation than the nodes with no direct link to the FC.

Our measurement rate control approach can be viewed as *compression*, but at the *measurement packet level*, rather than at bit level. By addressing the problem at packet level, we capture the precise dependence between estimation uncertainty due to sensor observation packets and energy efficiency. En route to the solution, we introduce a metric that captures estimation error for correlated measurements. This is inspired by the total squared cross-correlation (TSC) in code design in CDMA [22]. Our results msnifest that optimal network lifetime can be achieved through autonomous control of sensor measurement rate control, aggregation combining selection and routing based on spatial correlations for given estimation error specification. The rest of the paper is organized as follows. In section II we present the model and assumptions. In section III we formulate the optimization problem and solve it with the distributed algorithm. Section IV presents numerical results and section V concludes our study.

## II. SYSTEM MODEL

### A. Sensor measurement model

We consider a set  $\mathcal{N}$  of  $m$  sensors, represented by a directed graph  $G(\mathcal{N}, \mathcal{A})$ , where  $\mathcal{A}$  is the set of links. Define  $\mathcal{S}_i^{in}$  and  $\mathcal{S}_i^{out}$  as the set of nodes that can reach  $i$  and the set of nodes that can be reached by  $i$  with a certain transmit power. Each sensor has initial energy reserve  $E_i$ .

We consider a clock-driven system. Sensors observe a slowly time-varying, unknown, spatially homogeneous phenomenon process. Every  $T_s$  time units, typically a few seconds, sensors submit measurements to the FC. We call each such interval of duration  $T_s$  an *epoch*. The process under observation is a sequence  $\{\theta_t\}_{t=1,2,\dots}$ , where  $\theta_t$  is the unknown parameter value of the process at epoch  $t$ , assumed to remain fixed for the entire epoch duration. Within a given epoch  $t$ , a sensor  $i$  takes endogenous measurements at time

instants  $\tau$ ,

$$x_i(\tau) = \theta_t + n_i(\tau) \quad (1)$$

where  $x_i(\tau)$  is the measurement at time  $\tau$  and  $\theta_\tau = \theta_t$ . The noise process  $n_i(\cdot)$  captures uncertainty of sensor  $i$  observation due to different perception of the process and due to residual measurement errors. For each  $i$ ,  $n_i(t)$  is Gaussian, zero mean, wide-sense stationary and uncorrelated in time: for any  $t, t'$ , *temporal correlation*  $R_i(t, t') = 0$  if  $t \neq t'$  and  $R_i(t, t) = \sigma_i^2$  otherwise. Quantity  $\sigma_i^2 = \mathbb{E}[n_i^2(t)]$  is the variance of  $n_i(t)$  for any  $t$  and captures measurement inaccuracy. Each  $n_i(t)$  is independent from  $\{\theta_t\}_{t=1,\dots}$ . This work is not concerned with estimation of point phenomena processes, which require a different modeling approach in which the observation and the measurement noise should also depend on the distance from the source of the phenomenon.

For each pair of sensors  $i$  and  $j$ , noise processes are *spatially correlated* due to sensor device proximity. Spatial correlation is time-invariant, namely for any  $t, t'$ , the spatiotemporal correlation  $R_{ij}(s, s'; t, t')$  between sensors  $i$  and  $j$  at locations  $s$  and  $s'$  is  $R_{ij}(s, s')$  and depends only on their locations. By leaving out temporal correlation, we wish to focus on the impact of spatial correlation. Define the symmetric  $m \times m$  *spatial correlation matrix*  $\mathbf{C}$ , whose  $[i, j]$ -th element,  $\rho_{ij} = \mathbb{E}[n_i(t)n_j(t')]$  is the spatial correlation between noise processes  $n_i(t)$  and  $n_j(t)$  of sensors  $i$  and  $j$  at all times  $(t, t')$ . Pairwise correlations  $\rho_{ij}$  are assumed to be non-negative, and they are non-zero only for  $j \in \mathcal{S}_i^{in} \cup \mathcal{S}_i^{out}$ , and matrix  $\mathbf{C}$  is positive definite.

Each epoch consists of two subintervals of fixed duration. First, a subinterval in which sensors take measurements and perform aggregation of incoming data. Control information is also exchanged during this interval. Second, a transmission interval where sensors transmit their aggregated data to their neighbors or to the FC.

During the first subinterval, the FC broadcasts necessary control information to sensors within its range as detailed in section III. During the same subinterval, each sensor collects its measurements. Different measurements of a sensor are uncorrelated in time. However, any measurement of a sensor is *spatially correlated* with any other measurement of another nearby sensor. Also, messaging from neighbor sensors takes place during this phase. Once a sensor receives incoming data from its neighbors, it aggregates its input data flows with its own measurements and generates a single outgoing flow, by applying appropriate aggregation coefficients as detailed later. During the transmission subinterval, sensors forward aggregated measurement data toward the FC in multi-hop fashion. We assume no quantization or compression takes place at the bit level. We also assume that the length of the epoch is large enough such that all aggregated measurements reach the FC. At the end of the epoch, the FC makes the estimation.



### B. Sensor aggregation and transmission model

Let  $N_i(t)$  be the number of measurements generated by sensor  $i$  at epoch  $t$ . Let the number of epochs  $T$  grow large, and define  $r_i = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T N_i(t)$  to be the *average measurement rate* (in measurements/sec or in bits/sec) of sensor  $i$ . As discussed later, the vector of measurement numbers  $\mathbf{N} = (N_1, \dots, N_m)$  per epoch, or the measurement rate vector  $\mathbf{r} = (r_1, \dots, r_m)$  will be a control variable vector.

We adopt a *fluid* model for information flow, such that information forwarded from sensor  $i$  to  $j$  is treated as a real-valued flow  $f_{ij} \geq 0$ , denoting average amount of bits per unit of time. Thus,  $f_{ij} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T N_{ij}(t)$ , where  $N_{ij}(t)$  is the number of measurements that  $i$  sends to  $j$  at epoch  $t$ . Each sensor  $i$  receives data flows  $\{f_{ji} : j \in \mathcal{S}_i^{in}\}$  from its neighbors and aggregates them with its own measurement flow of rate  $r_i$ .

In this work, we consider the class of *linear combining aggregation policies* that operate as follows. At each epoch  $t$ , a sensor  $i$  receives  $N_{ji}(t)$  measurements from neighbor  $j \in \mathcal{S}_i^{in}$  and has generated  $N_i(t)$  measurements of its own. It then needs to decide on the *output* number of measurements,  $\tilde{N}_i(t)$  that it will create by aggregating input measurements and its own measurements. Equivalently, a sensor  $i$  decides on the amount of outgoing flow  $\tilde{f}_i = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \tilde{N}_i(t)$ .

Once it decides on  $\tilde{f}_i$ , sensor  $i$  generates  $\tilde{N}_i(t)$  weighted sums of these measurements. Each weighted sum, indexed by  $\ell = 1, \dots, \tilde{N}_i(t)$  is of the form  $X_i + \sum_{j \in \mathcal{S}_i^{in}} w_{ji} X_{ji}$  where  $X_i$  is a random variable denoting a randomly selected measurement out of the  $N_i(t)$  ones generated by  $i$ ,  $X_{ji}$  is a random variable denoting a randomly selected measurement out of the  $N_{ji}(t)$  ones respectively, and  $w_{ji}$  are appropriate aggregation weights. The  $\tilde{N}_i(t)$  aggregate measurements form a single outgoing flow of amount

$$\tilde{f}_i \leq r_i + \sum_{j \in \mathcal{S}_i^{in}} f_{ji}, \quad (2)$$

where the inequality constraint is set so that the concept of aggregation in reducing the amount of input data flow is meaningful. The model of measurement aggregation is depicted in Fig. 1. The outgoing flow  $\tilde{f}_i$  is then split into flows to neighbors in  $\mathcal{S}_i^{out}$ . The collection of aggregation coefficient vectors  $\mathbf{w}_i = (w_{ji} : j \in \mathcal{S}_i^{in})$  for sensor  $i = 1, \dots, m$ , the aggregation rate vector  $(\tilde{f}_1, \dots, \tilde{f}_m)$  and the flow routing decisions  $\mathbf{f}_i = (f_{ij} : j \in \mathcal{S}_i^{out})$  for each sensor  $i$  are our continuous-valued control variables to be discussed later. We also assume that processes  $\{N_i(t)\}$ ,  $\{N_j(t)\}$  for  $i \neq j$  are uncorrelated in time.

A sensor consumes energy only during transmission. The energy consumed by sensor  $i$  to transmit an information unit (e.g. a measurement packet) to sensor  $j \in \mathcal{S}_i^{out}$  is  $e_{ij}$ . This captures wireless link gain between  $i$  and  $j$ , namely path loss and fading and well as the requirement for a sufficient signal to noise ratio (SNR) at a receiver. Define as  $\mathcal{S}_{FC}$

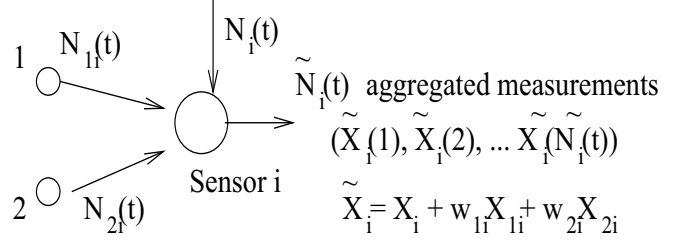


Fig. 1. Measurement aggregation at sensor  $i$  with two in-neighbors, at epoch  $t$ . Sensor  $i$  receives  $N_{1i}(t)$  measurements from neighbor 1 and  $N_{2i}(t)$  measurements from neighbor 2. It also generates  $N_i(t)$  measurements of its own. Then, sensor  $i$  generates  $\tilde{N}_i(t)$  aggregated measurements. Call  $\tilde{X}_i$  the random variable denoting one of the  $\tilde{N}_i(t)$  aggregated measurements. Each of the  $\tilde{N}_i(t)$  aggregated measurements,  $\tilde{X}_i$ , is formed as follows: pick one of  $N_{1i}(t)$  measurements, call it  $X_{1i}$ , then pick one of the  $N_{2i}(t)$  measurements,  $X_{2i}$  and pick one of the  $N_i(t)$  measurements,  $X_i$ . Then,  $\tilde{X}_i = X_i + w_{1i} X_{1i} + w_{2i} X_{2i}$ . The aggregation policy consists in deciding how many aggregated measurements  $\tilde{N}_i(t)$  will be generated, as well as the aggregation weights  $w_{1i}, w_{2i}$ .

the set of sensors  $i$  for which  $\text{FC} \in \mathcal{S}_i^{out}$ . These are the sensors with direct link to the FC. Let  $e_i$  denote the energy consumed per information unit from sensor  $i \in \mathcal{S}_{FC}$  to the FC.

### C. Estimation at the FC

Within each epoch, the FC obtains measurements from sensors in  $\mathcal{S}_{FC}$  and computes an estimate  $\hat{\theta}_t$  of  $\theta_t$  in the Maximum Likelihood (ML) sense. ML is a valid estimation in the absence of prior knowledge about the phenomenon  $\theta_t$ . We drop epoch index  $t$  in the sequel. Consider first the simplest scenario that each sensor sends one measurement to the FC and no intermediate aggregation takes place. An ensemble of  $m$  measurements  $\{x_i\}_{i=1}^m$  is available at the FC. The joint probability density function (p.d.f.) of sensor measurement vector  $\mathbf{x} = (x_1, \dots, x_m)^T$  is

$$p_{\theta}(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} (\det \mathbf{C})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \theta \mathbf{1})^T \mathbf{C}^{-1} (\mathbf{x} - \theta \mathbf{1})\right\} \quad (3)$$

namely it is a Gaussian p.d.f with mean  $\theta \mathbf{1}$  and correlation matrix  $\mathbf{C}$ , where  $\mathbf{1}$  is the  $m \times 1$  vector of 1's. The ML estimate of  $\theta$ , is  $\hat{\theta}_{ML} = \arg \max_{\theta} \log p_{\theta}(\mathbf{x}) = (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}) / (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})$ .

The criterion for estimation quality is the mean squared error (MSE),  $\mathbb{E}[(\theta - \hat{\theta}_{ML})^2]$ . To make estimation quality independent of  $\theta$ , we consider the class of unbiased estimators, i.e. those estimators for which  $\mathbb{E}(\hat{\theta}_{ML}) = \theta$ . Then, the MSE equals  $\text{var}(\hat{\theta}_{ML})$ . We thus seek to minimize  $\text{var}(\hat{\theta}_{ML})$  (where the expectation is with respect to randomness of observations), through a minimum variance unbiased (MVU) estimator. In our case, the ML estimate is unbiased, and  $\text{var}(\hat{\theta}_{ML}) = [\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}]^{-1}$ . For spatially uncorrelated sensors, we have  $\mathbf{C}^{-1} = \text{diag}(1/\sigma_1^2, \dots, 1/\sigma_m^2)$ , and  $\hat{\theta}_{ML} = \left(\sum_{i=1}^m \frac{1}{\sigma_i^2}\right)^{-1} \cdot \left(\sum_{i=1}^m \frac{x_i}{\sigma_i^2}\right)$  and  $\text{var}(\hat{\theta}_{ML}) =$

$\left(\sum_{i=1}^m \frac{1}{\sigma_i^2}\right)^{-1}$ . If all sensors are identical ( $\sigma_i^2 = \sigma^2$  for all  $i$ ), then  $\hat{\theta}_{ML} = \frac{1}{m} \sum_{i=1}^m x_i$  and  $\text{var}(\hat{\theta}_{ML}) = \sigma^2/m$ .

### III. MAX-LIFETIME MEASUREMENT RATE CONTROL, AGGREGATION AND ROUTING SUBJECT TO ESTIMATION ERROR CONSTRAINTS

#### A. Estimation error vs. numbers of endogenous measurements of sensors

We first obtain the joint p.d.f. of sensor measurements and then proceed to the expression of  $\text{var}(\hat{\theta}_{ML})$  for different numbers of measurements per sensor. For now, we focus on the impact of number of measurements of sensors on the estimation error without considering data aggregation.

1) *Spatially uncorrelated sensors*: Assume that at a given epoch, sensor  $i$  takes  $N_i$  measurements,  $i = 1, \dots, m$  denoted by vector  $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_{N_i}^{(i)})$  and send them to the FC. Let  $n_i, i = 1, \dots, m$  be indices with  $1 \leq n_i \leq N_i$ . Measurements of a given sensor are uncorrelated to each other. Since sensors are also spatially uncorrelated, all sensor measurements are uncorrelated with each other, and under the Gaussian assumption, they are independent. The joint p.d.f. of measurement vector  $\mathbf{x}$  is

$$p_{\theta}(\mathbf{x}) = \prod_{i=1}^m \left( \prod_{n_i=1}^{N_i} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_i^2} (x_{n_i}^{(i)} - \theta)^2\right\} \right). \quad (4)$$

The ML estimate of  $\theta$ ,  $\hat{\theta}_{ML}$ , is

$$\hat{\theta}_{ML} = \left( \sum_{i=1}^m \sum_{n_i=1}^{N_i} \frac{x_{n_i}^{(i)}}{\sigma_i^2} \right) \cdot \left( \sum_{i=1}^m \frac{N_i}{\sigma_i^2} \right)^{-1}, \quad (5)$$

and  $\text{var}(\hat{\theta}_{ML}) = \left( \sum_{i=1}^m \frac{N_i}{\sigma_i^2} \right)^{-1}$ . Thus, estimation error depends on the number of measurements  $N_i$  of each sensor  $i$ .

2) *Spatially correlated sensors*: The situation here is more complex due to correlations among measurements of different sensors. We start from the case of two sensors, 1 and 2, with  $N_1 = 1$  and  $N_2 = 2$  that are sent to the FC. For the joint p.d.f. of vector  $\mathbf{x} = (x_1^{(1)}, x_1^{(2)}, x_2^{(2)})$ , we argue as follows:

$$p_{\theta}(x_1^{(1)}, x_1^{(2)}, x_2^{(2)}) = p_{\theta}(x_1^{(2)} | x_1^{(1)}) p_{\theta}(x_2^{(2)} | x_1^{(1)}) p_{\theta}(x_1^{(1)}),$$

since measurements of sensor 2,  $x_1^{(2)}, x_2^{(2)}$  are conditionally independent, given the measurement  $x_1^{(1)}$  of 1. Finally,

$$p_{\theta}(x_1^{(1)}, x_1^{(2)}, x_2^{(2)}) = \frac{p_{\theta}(x_1^{(1)}, x_1^{(2)}) p_{\theta}(x_1^{(1)}, x_2^{(2)})}{p_{\theta}(x_1^{(1)})}, \quad (6)$$

and the joint p.d.f.  $p_{\theta}(x_1^{(1)}, x_1^{(2)}, x_2^{(2)})$  can be expressed in terms of distributions  $p_{\theta}(x_1^{(1)}, x_1^{(2)})$  and  $p_{\theta}(x_1^{(1)}, x_2^{(2)})$  which are jointly Gaussian with correlation matrix equal to the  $2 \times 2$  spatial correlation matrix of sensors 1 and 2, while  $p_{\theta}(x_1^{(1)})$  is Gaussian with variance  $\sigma_1^2$ .

Similarly, the joint p.d.f. of measurement vectors  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$  where sensors take  $N_1, N_2 > 1$  measurements, is

$$p_{\theta}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \frac{\prod_{n_1=1}^{N_1} \prod_{n_2=1}^{N_2} p_{\theta}(x_{n_1}^{(1)}, x_{n_2}^{(2)})}{\left( \prod_{n_1=1}^{N_1} p_{\theta}(x_{n_1}^{(1)}) \right)^{N_2-1} \left( \prod_{n_2=1}^{N_2} p_{\theta}(x_{n_2}^{(2)}) \right)^{N_1-1}}. \quad (7)$$

and the variance of estimation error is

$$\text{var}(\hat{\theta}_{ML}) = \left( N_1 N_2 \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} - \frac{(N_2 - 1) N_1}{\sigma_1^2} - \frac{(N_1 - 1) N_2}{\sigma_2^2} \right)^{-1}, \quad (8)$$

where  $\mathbf{1}$  is the  $2 \times 1$  vector of 1's.

Consider now the general case of  $m$  sensors, where sensor  $i$  takes  $N_i$  measurements. For sensor subset  $\mathcal{S} \subseteq \mathcal{M}$ , define  $\mathbf{C}_{\mathcal{S}}$  to be the sub-matrix of  $\mathbf{C}$  that consists only of rows and columns corresponding to sensors in  $\mathcal{S}$ . Let  $|\mathcal{S}|$  be the cardinality of  $\mathcal{S}$ . After some tedious algebra, we obtain:

$$\text{var}(\hat{\theta}_{ML}) = \left[ \sum_{\mathcal{S} \subseteq \mathcal{M}} \gamma_{\mathcal{S}} \left( \prod_{i \in \mathcal{S}} N_i \prod_{j \in \mathcal{M} \setminus \mathcal{S}} (N_j - 1) \right) \mathbf{1}^T \mathbf{C}_{\mathcal{S}}^{-1} \mathbf{1} \right]^{-1}, \quad (9)$$

where  $\gamma_{\mathcal{S}} = (-1)^{|\mathcal{S}|}$ , if the number of sensors  $m$  is even, and  $\gamma_{\mathcal{S}} = (-1)^{|\mathcal{S}|+1}$  if  $m$  is odd, and  $\mathbf{1}$  is a vector of ones of appropriate dimension.

Let  $\text{var}(\hat{\theta}_{ML}) = [h(\mathbf{N})]^{-1}$ . Each term in the sum in  $h(\mathbf{N})$  denotes mutual coupling among measurements of sensor subset  $\mathcal{S}$ . To compute  $\text{var}(\hat{\theta}_{ML})$ , one needs to compute  $2^m - 1$  terms, each of which involves finding  $\mathbf{C}_{\mathcal{S}}^{-1}$ . Depending on the deployment application, the number of sensors vary from less than ten to some hundreds. For few sensors,  $\text{var}(\hat{\theta}_{ML})$  can be precisely evaluated. For larger number of sensors, some methods are needed to reduce computational load. One method could be to split the set of sensors  $\mathcal{M}$  into subsets by clustering, compute precisely an estimate and error variance for each sensor subset, and fuse all estimates.

Here, we introduce a solution inspired from CDMA code design. Each user code is a vector, and the pairwise code cross-correlation is their inner product. The total squared cross-correlation (TSC) metric, defined as the sum of squares of pairwise code cross-correlations quantifies mutual interference among CDMA codes and is used in designing low-interference, high capacity systems [22]. From (8), note that for two sensors  $i$  and  $j$ , it is  $\text{var}(\hat{\theta}_{ML}) = \left( N_i N_j a_{ij} + \frac{N_i}{\sigma_i^2} + \frac{N_j}{\sigma_j^2} \right)^{-1}$ , with  $a_{ij} = \mathbf{1}^T \mathbf{C}_{ij}^{-1} \mathbf{1} - \frac{1}{\sigma_i^2} - \frac{1}{\sigma_j^2}$ , and  $\mathbf{C}_{ij}$  the  $2 \times 2$  correlation matrix that corresponds to sensors  $i$  and  $j$ . Thus,  $a_{ij}$  may be viewed as the coupling between measurements of sensors  $i$  and  $j$ ; if  $\rho_{ij} = 0$ , then  $a_{ij} = 0$ . We construct the following metric for estimation error variance, which we call *Total Pairwise Correlation Approximation (TPCA)* of variance of estimation error:

$$\text{TPCA}(\hat{\theta}_{ML}) = \left( \sum_{i=1}^m \sum_{j \in \mathcal{S}_i, j \neq i}^m N_i N_j a_{ij} + \sum_{i=1}^m \frac{N_i}{\sigma_i^2} \right)^{-1} = \tilde{h}(\mathbf{N})^{-1}. \quad (10)$$

The expression above stands for the average estimation error at a given epoch. To obtain an expression for average estimation error over all epochs,  $\bar{h}(\mathbf{r}) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \tilde{h}(\mathbf{N}^{(t)})$ , we need to express it through  $\mathbf{r}$ . It can be shown that

$$\bar{h}(\mathbf{r}) = T_s^2 \sum_i \sum_{j \in \mathcal{S}_i, j \neq i} a_{ij} r_i r_j + T_s \sum_i \frac{r_i}{\sigma_i^2}. \quad (11)$$

### B. Problem statement and formulation

The fundamental tradeoff between estimation quality and lifetime is that more measurements yield small estimation error, but they require more energy to gather at the FC. Since sensors have different observation quality, different energy reserves and different consumed energy per measurement, it is necessary to provide sensors with the possibility to submit *different* average number of measurements per unit time, i.e. different measurement rates. Essentially sensors sample the process they are meant to estimate with different rates. When sensor spatial correlations come into play, endogenous sensor measurement rates should be such that sensors in closer proximity collectively transmit fewer measurements than if sensors were uncorrelated if this reduces redundancy. Due to the coupling among measurements of different sensors because of spatial correlation, different measurement rate vectors give rise to different joint probability distribution of measurements, and thus different estimation errors. By controlling measurement rates, we create transmission regimes that provide the FC with adequate measurements for the specified estimation error while adhering to low energy consumption.

Data aggregation effectively reduces the number of messages transported in the network. Each sensor combines its measurements with incoming ones from neighboring nodes with appropriate weights and generates one outgoing flow. First, each sensor needs to decide the rate of generated aggregate traffic flow. Larger aggregate flow rates aid more in supporting a given estimation error variance, yet they incur more traffic load and thus higher energy consumption in the network. By controlling aggregation weights, each sensor controls the variance of the aggregated flow and therefore influences future aggregations and ultimately the estimation error. The challenge of aggregation is to eliminate data redundancy while producing low uncertainty (variance) data, since these will be eventually forwarded to the FC through a cascade of aggregations in the network. Routing at each sensor aims at selecting the portions of aggregate measurement traffic to transmit to each neighbor. The joint problem is to determine the measurement rate of each sensor, the aggregation policy and the routes through which aggregate measurements are transported from sensors to the FC, such that network lifetime is maximized.

1) *Aggregation method:* To show the benefit of aggregation, consider the following simple example. Consider two sensors that take one measurement each and send it to the FC. The measurement of each sensor has variance  $\sigma^2$  and

sensor measurements are spatially correlated with correlation  $\rho$ . Without aggregation, sensors send two measurements and the variance of estimation error is  $(\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^{-1} = \frac{1}{2}(\sigma^2 + \rho)$ . Suppose now that one sensor, say sensor 2 aggregates its measurement  $X_2$  with the measurement of sensor 1,  $X_1$  with weight  $w$ , generates  $X_1 + wX_2$  and sends it to the FC. The value of  $w$  that minimizes the variance of the aggregated measurement,  $\mathbb{E}[(X_1 + wX_2)^2]$  is  $w^* = -\rho/\sigma^2$  and the variance of estimator error is  $\sigma^2 - \frac{\rho^2}{\sigma^2}$ . This is clearly smaller than the one without aggregation if  $\rho > \sigma^2/2$ . This simple example manifests that aggregation is beneficial if sensors are adequately correlated.

First, we discuss the impact of aggregation. Fix attention to sensor  $j$  that receives flows from its in-neighbors  $k \in \mathcal{S}_j^{in}$  and aggregates them to one flow with a weight vector  $\mathbf{w}_j = (w_{kj} : k \in \mathcal{S}_j^{in})$ . The key observations that we build upon in the sequel are: (i) *aggregation changes the variance of outgoing flow*, (ii) *aggregation changes the cross-correlation between output data of sensor  $j$  and its neighbors*. The latter observation stems from the fact that if sensor  $j$  performs aggregation, its output traffic is a weighted mix of input flows of neighbors in  $\mathcal{S}_j^{in}$  and its own traffic. Therefore the cross-correlation between the output (aggregate) traffic of  $j$  and the endogenous traffic of a neighbor  $i$  depends on the cross-correlation of endogenous measurement traffic of  $j$  and that of  $i$  and the cross-correlation between the traffic of sensors in  $\mathcal{S}_j^{in}$  and that of  $i$ .

Cross-correlations also clearly depend on aggregation weights. Suppose  $j$  performs aggregation as follows,  $\tilde{X}_j = X_j + \sum_{k \in \mathcal{S}_j^{in}} w_{kj} X_{kj}$ , and it sends the aggregate traffic to sensor  $i$ . Let  $\tilde{\rho}_{ji}$  be the new cross-correlation, after the mixing above at  $j$ . It is

$$\tilde{\rho}_{ji} = \mathbb{E}[X_i \tilde{X}_j] = \rho_{ij} + \sum_{k \in \mathcal{S}_j^{in}} w_{kj} \tilde{\rho}_{ik}. \quad (12)$$

In order for  $i$  to know the new cross-correlation,  $j$  should pass to its neighbors in  $\mathcal{S}_j^{out}$  its in-neighbor IDs and the aggregation weights  $\mathbf{w}_j$ . Node  $i$  will parse the list of its neighbors and if  $k$  belongs to that list,  $i$  will use the current value  $\tilde{\rho}_{ik}$  in the expression above to compute the new cross-correlation.

Let vector  $\mathbf{w}^{(2)} = (w_{ji}^2 : j \in \mathcal{S}_i^{in})$  be the vector of squares of aggregation weights. For sensor  $i$ , define the vector of variances of incoming flows,  $\tilde{\sigma}_{i,in}^2 = (\tilde{\sigma}_j^2 : j \in \mathcal{S}_i^{in})$ , the vector of cross-correlation of aggregated traffic of in-neighbors of  $i$  with the endogenous traffic of  $i$ ,  $\tilde{\rho}_i = (\tilde{\rho}_{ji} : j \in \mathcal{S}_i^{in})$ , and the matrix of cross-correlations between aggregated traffic of in-neighbors of  $i$ ,  $\tilde{\mathbf{R}}_i = (\tilde{\rho}_{kl} : k, l \in \mathcal{S}_i^{in})$ . Then, since the aggregated measurements are given by  $\tilde{X}_i = X_i + \sum_{j \in \mathcal{S}_i^{in}} w_{ji} X_{ji}$ , the variance of the aggregated flow coming out of sensor  $i$  is

$$\tilde{\sigma}_i^2 = \sigma_i^2 + 2\mathbf{w}_i^T \tilde{\rho}_i + \mathbf{w}_i^{(2)T} \tilde{\sigma}_{i,in}^2 + 2\mathbf{w}_i^T \tilde{\mathbf{R}}_i \mathbf{w}_i. \quad (13)$$

To compute the variance above, node  $j$  should pass to sensors in  $\mathcal{S}_j^{out}$  a control packet with the variance  $\tilde{\sigma}_j^2$  of

its outgoing flow. Sensor  $i$  can also find  $\tilde{\rho}_{ji}$  from messages passed by  $j \in \mathcal{S}_i^{in}$  as outlined above. Sensor  $i$  can deduce matrix  $\tilde{\mathbf{R}}_i$  from cross-correlations broadcasted by sensors  $j \in \mathcal{S}_i^{in}$ .

2) *Network lifetime*: As mentioned above, with the adopted fluid model, information forwarded between sensors  $i$  and  $j \in \mathcal{S}_i^{out}$  is a real-valued flow  $f_{ij} \geq 0$  denoting average amount of bits per unit of time. Let  $\mathbf{f} = (f_{ij} : (i, j) \in \mathcal{A})$  be the vector of link flows. The lifetime of sensor  $i$  is  $L_i(\mathbf{f}) = E_i / (\sum_{j \in \mathcal{S}_i^{out}} e_{ij} f_{ij})$ , where the denominator denotes energy consumption rate of sensor  $i$ . Network lifetime is defined as the time until the battery of the first sensor empties, namely it is  $\min_{i \in \mathcal{N}} L_i(\mathbf{f})$ .

3) *Estimation error constraint and nodes close to the FC*: We are given a constraint  $\varepsilon$  on average estimation error. Expression  $[\bar{h}(\mathbf{r})]^{-1}$  in (11) gives the average estimation error, assuming that the generated measurement rate vector  $\mathbf{r}$  reaches the FC after routing takes place. In our setup, sensor measurement flows are *aggregated* and mixed with other flows throughout the network and hence the expression above needs to be modified. Observe that the estimation at the FC is based on aggregate flows that the FC receives from sensors in  $\mathcal{S}_{FC}$ .

Call the FC node  $x$  and let  $N_{ix}(t)$  be the number of measurements forwarded to the FC by sensor  $i \in \mathcal{S}_{FC}$  at epoch  $t$ . Let  $f_{ix} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T N_{ix}(t)$ . The average estimation error at the FC is

$$\bar{h}(\mathbf{f}) = T_s^2 \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_{FC}, j \neq i} \tilde{\alpha}_{ij} f_{ix} f_{jx} + T_s \sum_{i \in \mathcal{S}_{FC}} \frac{f_{ix}}{\tilde{\sigma}_{ix}^2}. \quad (14)$$

where  $\tilde{\sigma}_{ix}^2$  is the variance of the aggregate flows that  $i \in \mathcal{S}_{FC}$  sends to the FC receiver, and

$$\tilde{\alpha}_{ij} = \frac{\tilde{\sigma}_{ix}^2 + \tilde{\sigma}_{jx}^2 - 2\tilde{\rho}_{ij}}{\tilde{\sigma}_{ix}^2 \cdot \tilde{\sigma}_{jx}^2 - \tilde{\rho}_{ij}^2} - \frac{1}{\tilde{\sigma}_{ix}^2} - \frac{1}{\tilde{\sigma}_{jx}^2}. \quad (15)$$

The problem of maximizing lifetime subject to estimation error constraints can be formulated as:

$$\max_{\mathbf{r}, \mathbf{f}, \mathbf{W}} \min_{i \in \mathcal{N}} L_i(\mathbf{f})$$

with  $\mathbf{W} = (\mathbf{w}_i : i \in \mathcal{N})$  being the ensemble of aggregation policies of sensors, subject to:

$$\sum_{j \in \mathcal{S}_i^{out}} f_{ij} \leq r_i + \sum_{j \in \mathcal{S}_i^{in}} f_{ji}, \text{ for all } i \in \mathcal{N} \quad (16)$$

$$\frac{T_s^2}{2} \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_{FC}, j \in \mathcal{N}_i} \tilde{\alpha}_{ij} f_{ix} f_{jx} + T_s \sum_{i \in \mathcal{S}_{FC}} \frac{f_{ix}}{\tilde{\sigma}_{ix}^2} = \frac{1}{\varepsilon}, \quad (17)$$

$$\tilde{\sigma}_i^2 = \sigma_i^2 + 2\mathbf{w}_i^T \tilde{\boldsymbol{\rho}}_i + \mathbf{w}_i^{(2)T} \tilde{\boldsymbol{\sigma}}_{i,in}^2 + 2\mathbf{w}_i^T \tilde{\mathbf{R}}_i \mathbf{w}_i, \text{ for all } i \in \mathcal{N} \quad (18)$$

with  $\mathbf{r} \geq \mathbf{0}$ ,  $\mathbf{f} \geq \mathbf{0}$  and  $\mathcal{N}_i = \mathcal{S}_i^{in} \cup \mathcal{S}_i^{out}$ . Constraint (16) refers to the fact that the total amount of aggregate flow (to be routed to out-neighbors) should not exceed the total amount of incoming traffic for each sensor  $i$ . Constraint (17) describes the estimation error constraint at the FC which

relates to the amount of incoming traffic at the FC. Note that we multiply the first term by  $1/2$  to transform it into a form that does not need additional coordination among sensors. Finally, constraint (18) relates the variance of aggregate flow to those of the incoming flows.

Define a new variable  $z = \min_{i \in \mathcal{N}} L_i(\mathbf{f})$ , and set  $\hat{\mathbf{r}} = z\mathbf{r}$  and  $\hat{\mathbf{f}} = z\mathbf{f}$  to get the equivalent formulation,

$$\max_{\hat{\mathbf{r}}, \hat{\mathbf{f}}, \mathbf{W}, z} z$$

subject to:

$$\sum_{j \in \mathcal{S}_i^{out}} \hat{f}_{ij} \leq \hat{r}_i + \sum_{j \in \mathcal{S}_i^{in}} \hat{f}_{ji}, \text{ for all } i \in \mathcal{N} \quad (19)$$

$$\frac{T_s^2}{2} \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_{FC}, j \in \mathcal{N}_i^{in}} \tilde{\alpha}_{ij} \hat{f}_{ix} \hat{f}_{jx} + T_s \sum_{i \in \mathcal{S}_{FC}} \frac{z \hat{f}_{ix}}{\tilde{\sigma}_{ix}^2} = \frac{z^2}{\varepsilon}, \quad (20)$$

$$\sum_{j \in \mathcal{S}_i^{out}} e_{ij} \hat{f}_{ij} \leq E_i, \text{ for all } i \in \mathcal{N} \quad (21)$$

and constraint (18). Constraint (21) is due to the new variable  $z$ . It is  $\hat{\mathbf{r}} \geq \mathbf{0}$ ,  $\hat{\mathbf{f}} \geq \mathbf{0}$ ,  $z > 0$ , where the last constraint is due to the new variable  $z$ .

### C. Primal-dual algorithm

We form the Lagrangian

$$\begin{aligned} L(z, \hat{\mathbf{r}}, \hat{\mathbf{f}}, \mathbf{W}, \tilde{\boldsymbol{\sigma}}_i^2, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\xi}) = & -z + \sum_{i \in \mathcal{N}} \lambda_i \left( \sum_{j \in \mathcal{S}_i^{out}} \hat{f}_{ij} - \hat{r}_i - \sum_{j \in \mathcal{S}_i^{in}} \hat{f}_{ji} \right) \\ & + \mu \left( \frac{T_s^2}{2} \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_{FC}, j \in \mathcal{N}_i} \tilde{\alpha}_{ij} \hat{f}_{ix} \hat{f}_{jx} + T_s \sum_{i \in \mathcal{S}_{FC}} \frac{z \cdot \hat{f}_{ix}}{\tilde{\sigma}_{ix}^2} - \frac{z^2}{\varepsilon} \right) \\ & + \sum_{i \in \mathcal{N}} \xi_i (\tilde{\sigma}_i^2 - \sigma_i^2 - 2\mathbf{w}_i^T \tilde{\boldsymbol{\rho}}_i - \mathbf{w}_i^{(2)T} \tilde{\boldsymbol{\sigma}}_{i,in}^2 - 2\mathbf{w}_i^T \tilde{\mathbf{R}}_i \mathbf{w}_i) \end{aligned}$$

where  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\xi}$ ,  $\boldsymbol{\mu}$  are the Lagrange multipliers corresponding to the constraints of outgoing flow, output variance and estimation error respectively. Note that the following hold:

$$\sum_{i \in \mathcal{N}} \lambda_i \sum_{j \in \mathcal{S}_i^{out}} \hat{f}_{ij} = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{S}_i^{out}} \lambda_j \hat{f}_{ij}$$

and

$$\sum_i \xi_i \sum_{j \in \mathcal{S}_i^{in}} w_{ji}^2 \sigma_{ji}^2 = \sum_i \sum_{j \in \mathcal{S}_i^{out}} \xi_j w_{ij}^2 \tilde{\sigma}_i^2$$

Due to the latter property, we can write:

$$\sum_i \xi_i (\tilde{\sigma}_i^2 - \mathbf{w}_i^{(2)T} \tilde{\boldsymbol{\sigma}}_{i,in}^2) = \sum_i \xi_i \left( \sum_{j \in \mathcal{S}_i^{out}} \frac{\tilde{\sigma}_i^2}{|\mathcal{S}_i^{out}|} - \xi_j w_{ij}^2 \right) \tilde{\sigma}_i^2 \quad (22)$$

The Lagrangian can be divided into a part corresponding to sensors in  $\mathcal{S}_{FC}$  and a part for sensors not in  $\mathcal{S}_{FC}$ .



1) *Sensors not in direct range of the FC*: Consider first the part concerning sensors  $i \notin \mathcal{S}_{FC}$ , call it  $L_1(\cdot)$ . Lagrangian  $L_1(\cdot)$  can be decomposed to a sum of terms, each of which stands for one sensor  $i$  and is of the form,

$$\sum_{j \in \mathcal{S}_i^{out}} (\lambda_i - \lambda_j) \hat{f}_{ij} - \lambda_i \hat{r}_i + \sum_{j \in \mathcal{S}_i^{out}} \left( \frac{\xi_i}{|\mathcal{S}_i^{out}|} - \xi_j w_{ij}^2 \right) \tilde{\sigma}_i^2 + \xi_i (-\sigma_i^2 - 2\mathbf{w}_i^T \tilde{\rho}_i - \mathbf{w}_i^{(2)T} \tilde{\sigma}_{i,in}^2 - 2\mathbf{w}_i^T \tilde{\mathbf{R}}_i \mathbf{w}_i).$$

The primal problem for each sensor  $i \notin \mathcal{S}_{FC}$

$$\min_{\hat{\mathbf{r}}, \hat{\mathbf{f}}, \tilde{\sigma}_i^2, \lambda, \xi} L_1(z, \hat{\mathbf{r}}, \hat{\mathbf{f}}, \mathbf{W}, \tilde{\sigma}_i^2, \lambda, \xi) \quad \text{s.t.} \quad \sum_{j \in \mathcal{S}_i} e_{ij} \hat{f}_{ij} \leq E_i, \quad \forall i, \quad (23)$$

can be divided into: i) a *endogenous measurement rate control problem*, where each sensor adapts  $\hat{r}_i$ , (ii) a *routing problem* that controls  $\hat{\mathbf{f}}$ , and iii) an *aggregation weight adaptation problem* with which sensor  $i$  adjusts vector  $\mathbf{w}_i$ . As a result of the latter, each sensor computes the variance of aggregate data,  $\tilde{\sigma}_i$  and passes it to its neighbors. In the endogenous measurement control problem, sensor  $i$  does one step of the gradient update,

$$\hat{r}_i^{(t)} = \hat{r}_i^{(t-1)} + s_t \lambda_i^{(t)} \quad (24)$$

where  $s_t > 0$  is the step size, which should be a decreasing function of  $t$ . The routing problem consists of a linear programming (LP) problem that each sensor  $i$  solves to determine next-hop traffic flow variables  $\hat{f}_{ij}$ ,

$$\min_{\hat{f}_{ij}} \sum_{j \in \mathcal{S}_i^{out}} (\lambda_i - \lambda_j) \hat{f}_{ij} \quad \text{s.t.} \quad \sum_{j \in \mathcal{S}_i} e_{ij} \hat{f}_{ij} \leq E_i. \quad (25)$$

Note that for (25), Lagrange multipliers from sensors  $j \in \mathcal{S}_i^{out}$  need to be communicated to  $i$ . For the third problem, each sensor  $i$  adjusts its output variance  $\sigma_i^2$  by running one iteration of the gradient algorithm,

$$\tilde{\sigma}_i^{2(t)} = \tilde{\sigma}_i^{2(t-1)} - s_t \left( \sum_{j \in \mathcal{S}_i^{out}} \left( \frac{\xi_i^{(t)}}{|\mathcal{S}_i^{out}|} - \xi_j^{(t)} w_{ij}^2 \right) \right). \quad (26)$$

Again, multipliers  $\xi_j$  from sensors  $j \in \mathcal{S}_i^{out}$  need to be passed to sensor  $i$ . Moreover, each sensor adapts the aggregation weights by taking a step of the gradient update,

$$\mathbf{w}_i^{(t)} = \mathbf{w}_i^{(t-1)} + 2s_t [\tilde{\rho}_i + \mathbf{w}_i^{(t-1)} \circ \tilde{\sigma}_{i,in}^2 + (\tilde{\mathbf{R}}_i^{(t)} + \tilde{\mathbf{R}}_i^{T(t)}) \mathbf{w}_i^{(t-1)}]. \quad (27)$$

and passes these weights to its out-neighbors which need them to compute their output variance according to (26). In the above,  $\mathbf{a} \circ \mathbf{b}$  is the vector whose elements are the pairwise products of corresponding elements of  $\mathbf{a}$  and  $\mathbf{b}$ .

2) *Sensors in direct range of the FC*: Consider the part of the Lagrangian  $L_2(\cdot)$  for sensors  $i \in \mathcal{S}_{FC}$ . Now,  $L_2(\cdot)$  cannot decompose into a sum of terms, one for each sensor, and it is not possible to differentiate between routing and aggregation as in the case of sensors  $i \notin \mathcal{S}_{FC}$ . Furthermore, the energy constraints (21) are relaxed and incur multiplier vector  $\nu$ .

The Lagrangian is,

$$\begin{aligned} L_2(z, \hat{\mathbf{r}}, \hat{\mathbf{f}}, \mathbf{W}, \tilde{\sigma}_i^2, \lambda, \mu, \xi, \nu) = & -z + \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_i^{out}} (\lambda_i - \lambda_j) \hat{f}_{ij} - \sum_{i \in \mathcal{S}_{FC}} \hat{r}_i \\ & + \mu \left( \frac{T_s^2}{2} \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_{FC}, j \in \mathcal{N}_i} \tilde{\alpha}_{ij} \hat{f}_{ix} \hat{f}_{jx} + T_s \sum_{i \in \mathcal{S}_{FC}} \frac{z \cdot \hat{f}_{ix}}{\tilde{\sigma}_{ix}^2} - \frac{z^2}{\varepsilon} \right) \\ & + \sum_{i \in \mathcal{S}_{FC}} (-\sigma_i^2 - 2\mathbf{w}_i^T \tilde{\rho}_i - \mathbf{w}_i^{(2)T} \tilde{\sigma}_{i,in}^2 - 2\mathbf{w}_i^T \tilde{\mathbf{R}}_i \mathbf{w}_i) \\ & + \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_i^{out}} \left( \frac{\xi_i}{|\mathcal{S}_i^{out}|} - \xi_j w_{ij}^2 \right) \tilde{\sigma}_i^2 \\ & + \sum_{i \in \mathcal{S}_{FC}} \nu_i \left( \sum_{j \in \mathcal{S}_i^{out}} e_{ij} \hat{f}_{ij} - E_i \right). \end{aligned}$$

Note that each sensor  $i \in \mathcal{S}_{FC}$  needs to determine its flow to the FC,  $\hat{f}_{ix}$  (which will be taken into account in the estimation procedure), as well as the flows to its out-neighbors,  $j \in \mathcal{S}_i^{out}, j \neq x$ . The energy constraint of sensor  $i$  is written,

$$e_{ix} \hat{f}_{ix} + \sum_{j \in \mathcal{S}_i^{out}, j \neq x} e_{ij} \hat{f}_{ij} \leq E_i \quad \forall i \in \mathcal{S}_{FC}. \quad (28)$$

To adapt its out-flows, each sensor makes the following gradient descent steps,

$$\hat{f}_{ij}^{(t)} = [\hat{f}_{ij}^{(t-1)} - s_t [(\lambda_i^{(t)} - \lambda_j^{(t)}) + \nu^{(t)} e_{ij}]]^+, \quad j \neq x \quad (29)$$

where  $y^+ = y$  if  $y > 0$  and 0 otherwise, and

$$\begin{aligned} \hat{f}_{ix}^{(t)} = & [\hat{f}_{ix}^{(t-1)} - s_t (\lambda_i^{(t)} + \frac{T_s \mu^{(t)} z^{(t)}}{\tilde{\sigma}_i^2} \\ & + \frac{\mu^{(t)} T_s^2}{2} \sum_{j \in \mathcal{S}_{FC}, j \in \mathcal{N}_i} f_{jx}^{(t)} \tilde{\alpha}_{ij} + \nu^{(t)} e_{ij})]^+. \end{aligned}$$

Also, sensors perform the output variance adaptation,

$$\begin{aligned} \tilde{\sigma}_{ix}^{2(t)} = & \tilde{\sigma}_{ix}^{2(t-1)} - s_t \left[ \sum_{j \in \mathcal{S}_i^{out}} \left( \frac{\xi_i^{(t)}}{|\mathcal{S}_i^{out}|} - \xi_j^{(t)} w_{ij}^2 \right) + \mu^{(t)} \right. \\ & \left. \left( \frac{T_s^2}{2} \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_{FC}, j \neq i} \frac{\partial \tilde{\alpha}_{ij}}{\partial \tilde{\sigma}_{ix}^2} \hat{f}_{ix}^{(t)} \hat{f}_{jx}^{(t)} - T_s \sum_{i \in \mathcal{S}_{FC}} \frac{z \hat{f}_{ix}}{\tilde{\sigma}_{ix}^2} \right) \right] \quad (30) \end{aligned}$$

Sensors in  $\mathcal{S}_{FC}$  update their aggregation weights as in (27). The steps of the primal-dual algorithm, is summarized below.

- **STEP 0: Initialization.** Each sensor  $i$  initializes multipliers  $\lambda_i^{(0)}$ ,  $\xi_i^{(0)}$  and  $w_i^{(0)}$ . Each sensor  $i \in \mathcal{S}_{FC}$  initializes  $\nu_i^{(0)}$ . The FC initializes  $\mu^{(0)}$  and  $z^{(0)}$ .
- **STEP 1:** Each sensor  $i \in \mathcal{N}$  updates  $\lambda_i^{(t)}$ :

$$\lambda_i^{(t)} = [\lambda_i^{(t-1)} + s_t \left( \sum_{j \in \mathcal{S}_i^{out}} \hat{f}_{ij}^{(t-1)} - \hat{r}_i^{(t-1)} - \sum_{j \in \mathcal{S}_i^{in}} \hat{f}_{ji}^{(t-1)} \right)]^+$$

and broadcasts it to its neighborhood.

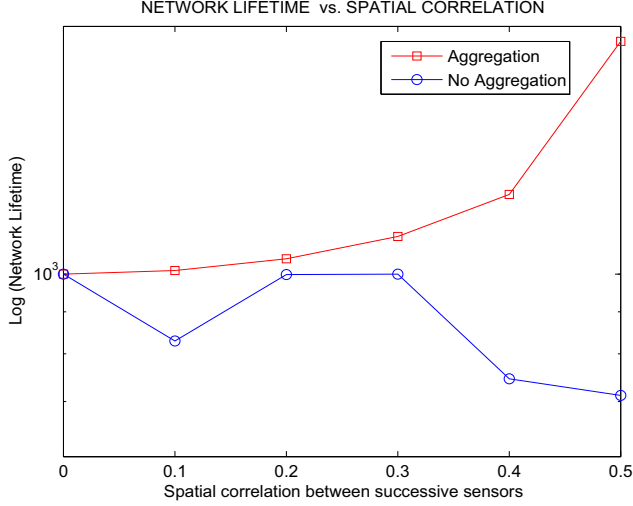


Fig. 2. Normalized lifetime versus cross-correlation for the case of aggregation and no aggregation for  $\varepsilon = 10^{-6}$ .

- **STEP 2:** Each sensor  $i \in \mathcal{N}$  adjusts  $\xi_i$ :

$$\xi_i^{(t)} = \xi_i^{(t-1)} + s_t (\tilde{\sigma}_i^{2(t-1)} - \sigma_i^2 - 2\mathbf{w}_i^{T(t-1)} \tilde{\mathbf{p}}_i - \mathbf{w}_i^{(2)(t-1)} \tilde{\sigma}_{i,in}^{2(t-1)} - 2\mathbf{w}_i^{T(t-1)} \tilde{\mathbf{R}}_i^{(t)} \mathbf{w}_i^{(t-1)}) \quad (31)$$

and broadcasts it to its neighborhood.

- **STEP 3:** The FC updates multiplier  $\mu$  based on,

$$\begin{aligned} \mu^{(t)} = \mu^{(t-1)} + s_t \left( \frac{T_s^2}{2} \sum_{i \in \mathcal{S}_{FC}} \sum_{j \in \mathcal{S}_{FC}, j \in \mathcal{N}_i} \tilde{\alpha}_{ij} \hat{f}_{ix}^{(t-1)} \hat{f}_{jx}^{(t-1)} \right. \\ \left. + T_s \sum_{i \in \mathcal{S}_{FC}} \frac{z^{(t-1)} \cdot \hat{f}_{ix}^{(t-1)}}{\tilde{\sigma}_{ix}^{2(t-1)}} - \frac{z^{(t-1)^2}}{\varepsilon} \right) \quad (32) \end{aligned}$$

- **STEP 4:** Each sensor  $i \in \mathcal{S}_{FC}$  computes:

$$\nu_i^{(t)} = [\nu_i^{(t-1)} + s_t (e_{ix} \hat{f}_{ix}^{(t-1)} + \sum_{j \in \mathcal{S}_{out}, j \neq x} e_{ij} \hat{f}_{ij}^{(t-1)} - E_i)]^+$$

- **STEP 5:** The FC updates  $z^{(t)}$  by solving  $\frac{\partial L(\cdot)}{\partial z} = 0$ :

$$z^{(t)} = \frac{\varepsilon}{2} \left( T_s \sum_{j \in \mathcal{S}_{FC}} \frac{f_{ix}^{(t-1)}}{\tilde{\sigma}_i^{2(t-1)}} - \frac{1}{\mu^{(t)}} \right)$$

- **STEP 6:** The FC broadcasts  $z^{(t)}$  and  $\mu^{(t)}$  to all sensors.
- **STEP 7:** Sensors update their measurement rate according to  $\hat{r}_i^{(t)} = \hat{r}_i^{(t-1)} + s_t \lambda_i^{(t)}$ .
- **STEP 8a:** Each sensor  $i \notin \mathcal{S}_{FC}$  solves the routing problem (25) and finds  $\hat{f}_{ij}^{(t)}$ .
- **STEP 8b:** Each sensor  $i \in \mathcal{S}_{FC}$  updates  $\hat{f}_{ij}^{(t)}$  and  $\hat{f}_{ix}^{(t)}$  according to (29) and (30).
- **STEP 9:** Each sensor  $i \notin \mathcal{S}_{FC}$  updates the variance of aggregated flow,  $\tilde{\sigma}_i^2$  according to (26) while sensors  $i \in \mathcal{S}_{FC}$  do the same according to (30).
- **STEP 10:** Each sensor  $i \in \mathcal{N}$  updates its aggregation weights  $\mathbf{w}_i$  according to (27).

- **STEP 11:** Each sensor passes its output variance  $\tilde{\sigma}_i^{2(t)}$  and its aggregation coefficients  $\mathbf{w}_i^{(t)}$  to all its neighbors.
- **STEP 12:** Each sensor  $i$  sets its generation rate  $r_i^{(t)} = \hat{r}_i^{(t)} / z^{(t)}$ , i.e it generates  $N_i^{(t)} = T_s r_i^{(t)}$  measurements during epoch  $t$ . It determines flows as  $f_{ij}^{(t)} = \hat{f}_{ij}^{(t)} / z^{(t)}$ .
- **STEP 13:** Each sensor updates cross-correlations  $\tilde{\rho}_i^{(t)}$ ,  $\tilde{\mathbf{R}}_i^{(t)}$  based on received information in Step 10.
- **STEP 14:** Aggregated measurements from sensors in  $\mathcal{S}_{FC}$  reach FC at  $t$ .
- **STEP 15:**  $t \leftarrow t + 1$ . Go to Step 1. Continue until convergence.

The algorithm is decentralized, with minimal feedback from the FC. At each epoch, dual variables  $\lambda^{(t)}$ ,  $\mu^{(t)}$ ,  $\xi^{(t)}$  and  $\nu^{(t)}$  are updated. These are performed with gradient ascent steps. Next, a one-shot minimization is performed by the FC to find  $z^{(t)}$ . In Step 7 each sensor  $i$  adjusts endogenous measurement rate  $\hat{r}_i^{(t)}$ . Each sensor not in  $\mathcal{S}_{FC}$  independently takes routing decisions by solving the LP problem (25). Sensors in  $\mathcal{S}_{FC}$  adjust their routing. Subsequently, adaptation of aggregation weights  $\mathbf{w}_i^{(t)}$  is performed. If the sequence of steps satisfies  $\lim_{t \rightarrow \infty} s_t = 0$ ,  $\sum_t s_t = \infty$ , the algorithm converges to at least a local optimum of the original problem.

#### IV. NUMERICAL EVALUATION

To demonstrate the tradeoff between lifetime and estimation error and the way this is shaped by our approach, we consider a toy structure of  $m = 10$  sensors deployed in a straight line. The topology of sensors in a straight line arises often in peripheral monitoring. Sensor  $i > 1$  receives data from sensor  $i - 1$ , it aggregates them with its own measurements and forwards it to sensor  $i + 1$ . Sensor  $i = 1$  forwards its data to 2 and sensor  $m$  is the only sensor with direct link to the FC. All distances  $d_{i,i+1}$  and distance  $d_{m,x}$  of sensor  $m$  to the FC are  $d$ .

For spatial correlation, we assume  $\rho(d) = \exp[-\frac{d}{3}]$ . Thus, cross-correlation between consecutive sensors is the same, say  $\rho$ . Also  $\sigma_i^2 = 1$  for all  $i$ . Let  $e \sim d^2$  be the energy consumed to reliably transmit a packet to a neighbor and  $E$  be the initial energy of each sensor. We take  $E/e = 10^9$  for the purposes of the experiments. Non-neighboring sensors are uncorrelated. For each sensor, we consider only the energy consumed for transmission. We chose this example so as to abstract out the routing and stress the benefits of spatial correlation-aware sensor network aggregation and endogenous measurement rate control.

In Figure 2, we depict the logarithm of normalized lifetime,  $z/(E/e)$  for the network above for the cases when aggregation may or may not be part of the formulation. The estimation error specification was taken to be  $\varepsilon = 10^{-6}$ . If aggregation is present, an aggregation weight needs to be computed by each sensor  $i$ . Sensor  $i$  combines traffic from  $i - 1$  to its own with this weight. If aggregation is not

included, only the measurement rate control matters, and lifetime is determined by the load carried by sensor  $m$ .

A first observation is that network lifetime always increases with spatial correlation, since sensors take coordinated decisions with their neighbors and perform appropriate measurement control or aggregation. Secondly, aggregation is always beneficial in terms of energy consumption; the estimation error specification is met with less transported traffic in the network, and hence network lifetime is increased. The difference is more evident for larger values of spatial correlation: for instance, for  $\rho = 0.5$ , we have normalized lifetime values of 1818.2 and 712.1 time units respectively for the cases of no aggregation and aggregation. Note that for  $\rho = 0$ , namely for spatially uncorrelated sensors, both cases lead to the same normalized lifetime of value 1000 time units. This is explained by the fact that if sensors are uncorrelated, aggregation cannot improve the variance of the outgoing traffic, and hence it renders itself void. An interesting phenomenon occurs for values of correlation between 0.2 and 0.3 for the case of no aggregation, where lifetime increases.

## V. CONCLUSION

We addressed the problem of maximum lifetime in a sensor network subject to a constraint on estimation error. Our main contribution is the joint decentralized orchestration of endogenous sensor measurement rates, local measurement aggregation, and routing decisions, and the design of a distributed primal-dual algorithm that relies on lightweight local sensor coordination and feedback from the FC. Depending on spatial correlation, sensors control their measurement rates and aggregation policies only when necessary to avoid redundant data generation.

A similar approach can be applied to study distributed inference and detection problems in sensor networks, albeit with different performance objectives, such as small probability of false alarm or missed detection. An enhanced model would be needed in that case, that captures the spatial and temporal uncertainty of event occurrence and the different perceived measurements by sensors depending on their distance from the event. In our work, we adhered to a distributed optimization-driven approach and the focus was only on the algorithmic aspect of the problem. The translation of the proposed algorithm to a practical approach would require among other issues the specification of exchanged control messages among sensors. We plan to address these issues in a future work.

## VI. ACKNOWLEDGMENTS

The authors acknowledge support of the European Commission through projects C4C (FP7-ICT-223844) and OP-NEX (FP7-ICT-224218).

## REFERENCES

- [1] J.-H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks", *IEEE/ACM Trans. Networking*, vol.12, no.4, pp.609-619, Aug. 2004.
- [2] M.C. Vuran, O.B. Akan and I.F. Akyildiz, "Spatio-Temporal correlation: Theory and applications for wireless sensor networks," *Computer Net. Journal (Elsevier)*, vol.45, no.3, pp. 245-261, June 2004.
- [3] M.C. Vuran and I.F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks", *IEEE/ACM Trans. Networking*, vol.14, no.2, pp. 316-329, April 2006.
- [4] N. Jaggi, K. Kar and A. Krishnamurthy, "Near-optimal activation policies in rechargeable sensor networks under spatial correlations", *ACM Trans. Sensor Net.*, vol.4, no.3, May 2008.
- [5] S.J. Baek, G. de Veciana and X. Su, "Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation", *IEEE J. Select. Areas Commun.*, vol.22, no.6, pp.1130-1140, Aug. 2004.
- [6] R. Cristescu, B. Beferull-Lozano, M. Vetterli and R. Wattenhofer, "Network correlated data gathering with explicit communication: NP-Completeness and algorithms", *IEEE/ACM Trans. Networking*, vol.14, no.1, pp.41-54, Feb. 2006.
- [7] S. Patten, B. Krishnamachari and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks", *Proc. ACM Symp. Inf. Proc. in Sensor Networks (IPSN)*, 2004.
- [8] K. Yuen, B. Liang and B. Li, "A distributed framework for correlated data gathering in sensor networks", *IEEE Trans. Veh. Tech.*, vol.57, no.1, pp.578-593, Jan. 2008.
- [9] A. Krasnoperov, J.-J. Xiao and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noise", *EURASIP J. Wireless Com. Networking*, vol.5, no.4, pp.473-482, Sept. 2005.
- [10] J. Matamoros and C. Anton-Haro, "Opportunistic random access for distributed parameter estimation in wireless sensor networks", *Europ. Signal Proc. Conf. (EUSIPCO)*, Poznan, Poland 2007.
- [11] S. Cui, J. Xiao, A. Goldsmith, Z. Luo and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing", *IEEE Trans. Signal Proc.*, vol.55, no.9, pp.4683-4695, Sept. 2007.
- [12] V. Sharma, E. Frazzoli and P.G. Voulgaris, "Improving lifetime data gathering and distortion for mobile sensing networks", *Proc. IEEE Sensor and Ad Hoc Comm. Net. (SECON)*, 2004.
- [13] A. Kansal, A. Ramamoorthy, M.B. Srivastava and G.J. Pottie, "On sensor network lifetime and data distortion", *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, 2005.
- [14] Y. Sung, S. Misra, L. Tong and A. Ephremides, "Cooperative routing for distributed detection in large sensor networks", *IEEE J. Select. Areas Commun.*, vol.25, no.2, pp.471-483, Feb. 2007.
- [15] B. Krishnamachari, D. Estrin, S. Wicker, "The impact of data aggregation in wireless sensor networks", *Proc. Int. Conf. on Distr. Comp. Systems*, 2002.
- [16] R. Cristescu, B. Beferull-Lozano and M. Vetterli, "Networked Slepian-Wolf: Theory, algorithms and scaling laws", *IEEE Trans. Inf. Theory*, vol.51, no.2, pp.4057-4073, Dec. 2005.
- [17] K. Kalpakis, K. Dasgupta and P.Namjoshi, "Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks", *Computer Networks*, vol.42, no.6, pp.697-716, Aug. 2003.
- [18] H. Ozgur Han and I. Korpoglu, "Power efficient data gathering and aggregation in wireless sensor networks", *SIGMOD Record*, vol.32, no.4, pp.66-71, Dec. 2003.
- [19] R. Khalili and J. Kurose, "A distributed minimum-distortion routing algorithm with in-network data processing", in *Proc. IEEE INFOCOM Mini-Conf.*, 2008.
- [20] Z. Ye, A.A. Abouzeid and J. Ai, "Optimal policies for distributed data aggregation in wireless sensor networks", in *Proc. IEEE INFOCOM*, 2007.
- [21] C. Joo, J.-G.Choi and N.B.Shroff, "Delay performance of scheduling with data aggregation in wireless sensor networks", in *Proc. IEEE INFOCOM*, 2010.
- [22] J.L. Massey and T. Mittelholzer, "Welch's bound and sequence sets for code-division multiple access systems", in *Sequences II, Methods in Communication, Security and Computer Science*, A. Capocelli, A. De Santis and U. Vaccaro, Eds. New York: Springer-Verlag, 1993.