

# Adaptive Distributed Estimation over Wireless Sensor Networks with Packet Losses

Alberto Speranzon

Unilever R&D Port Sunlight  
Quarry Road East  
CH63 3JW Bebington  
United Kingdom

alberto.speranzon@unilever.com

Carlo Fischione

UC Berkeley  
Hearst Ave.  
94720 Berkeley  
CA

fischion@eecs.berkeley.edu

Björn Johansson and Karl Henrik Johansson

School of Electrical and Engineering  
Royal Institute of Technology  
Osqudas väg 10  
100-44 Stockholm, Sweden

{bjorn.johansson|kallej}@ee.kth.se

**Abstract**—A distributed adaptive algorithm to estimate a time-varying signal, measured by a wireless sensor network, is designed and analyzed. The presence of measurement noises and of packet losses is considered. Each node of the network locally computes adaptive weights that guarantee to minimize the estimation error variance. Decentralized conditions on the weights, which ensure the stability of the estimates throughout the overall network, are also considered. A theoretical performance analysis of the scheme is carried out both in the presence of perfect and lossy links. Numerical simulations illustrate performance for various network topologies and packet loss probabilities.

**Index Terms**—Distributed filtering; Wireless sensor networks; Networked Embedded System.

## I. INTRODUCTION

The robustness, flexibility and reduced cost of Wireless Sensor Networks (WSNs) motivate the development of new classes of estimation algorithms, which need to be designed considering the limited computational and communication capabilities of such systems. Collaboration is suitable to overcome intrinsic limitations in processing measurements from only a single sensor, which often provide data with correlated data and quantization noise [1].

The main contribution of this paper is a theoretical framework to the design of a collaborative distributed filter for tracking a time-varying signal with a WSN. The approach ensures that the filter is stable and exhibits good behavior with respect to packet losses. We state and solve an optimization problem that provides optimal filter coefficients that guarantees the stability of the estimation error even in presence of lossy communications. The filter design is based on new global constraints, which are different from those adopted in the earlier work [2], [3].

Studies on the properties of distributed algorithm are found in the area of parallel computing [4]. More recently, they have found renewed interest in several other research areas, including communication networks, multi-robot coordination, and signal processing. Here we limit the discussion to

The work by A. Speranzon has been partially supported by the European Commission through the Marie Curie Transfer of Knowledge project BRIDGET (MKTD-CD\_2005 029961). C. Fischione was with KTH when this work was done. The work by C. Fischione, B. Johansson, K. H. Johansson is in the framework of the RUNES Integrated Project, contract number FP6-IST-004536, and the HYCON Network of Excellence, contract number FP6-IST-511368. It was partially funded also by the Swedish Foundation for Strategic Research and by the Swedish Research Council.

some recent and relevant contributions. In [5], the problem of distributed estimation of an average by a wireless sensor network is presented. It is assumed that nodes take a set of initial samples, and then iteratively exchange the averages of the samples collected by other nodes. The authors show that each node reaches asymptotically the global average. In [6], a more general approach is investigated. The authors study a distributed average computation of a time-varying signal measured by each sensor, when the signal is affected by a zero-mean noise. The same linear filter was also considered in [7], where the weights are computed to speed up the time of the average computation. A related consensus filter for distributed sensor fusion was proposed in [8], where the focus is on designing a distributed low pass filter. In [2], we extended the algorithms in [5]–[8] by designing the filter weights so that the variance of the estimation errors is minimized. The covariance matrix estimation problem of this filter was considered in [3].

The remainder of the paper is organized as follows: In Section II, the problem of adaptive distributed estimation algorithm over WSNs is posed. In Section III, conditions that guarantee the stability of the estimation errors are investigated in the presence of packet losses. Such conditions are included in the optimization problem studied in Section IV, where the filter coefficients are derived. The covariance of the estimates is discussed in Section VI. In Section VII, numerical results illustrate the performance of the distributed filter. Finally, conclusions and future perspectives are described in Section VIII.

**Notation:** Given a stochastic variable  $x$  we denote by  $\mathbb{E}x$  its expected value. When necessary, we write  $\mathbb{E}_y x$  to denote that the expected value is taken with respect to the probability density function of  $y$ . With  $\|\cdot\|$  we denote the  $\ell^2$ -norm of a vector or the spectral norm of a matrix. Given a matrix  $A$ , we denote with  $\ell_m(A)$  and  $\ell_M(A)$  the minimum and maximum eigenvalue, respectively, and we refer to its largest singular value as  $\gamma(A)$ . Given the matrix  $B$ , we denote with  $A \circ B$  the Hadamard (element-wise) product between  $A$  and  $B$ . With  $I$  and  $\mathbb{1}$  we denote the identity matrix and the vector  $(1, \dots, 1)^T$ , respectively.

## II. PROBLEM FORMULATION

Consider a WSN with  $N > 1$  sensors. At every time instant, each sensor in the network takes a noisy measure

of a scalar signal  $d(t)$ :

$$u(t) = d(t)\mathbf{1} + v(t),$$

where  $u = (u_1, \dots, u_N)^T$  is the vector collecting the sensor measurements and  $v = (v_1, \dots, v_N)^T$  is vector of the additive noises. We assume that  $v(t) \sim \mathcal{N}(0, \sigma^2 I)$  and that  $t \in \mathbb{N}_0 = \{0, 1, \dots\}$ .

We model the network as a weighted graph  $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{1, \dots, N\}$  is the vertex set and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  is the edge set. A weighting function  $\mathcal{W} : \mathcal{E} \times \mathbb{N}_0 \rightarrow \mathbb{R} : (i, j)(t) \mapsto g_{ij}(t)$ , which assigns a weight to each edge of the graph. We denote the set of neighbors of node  $i \in \mathcal{V}$  as  $\mathcal{N}_i(t) = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\} \cup (i, i)$ . The nodes are supposed to transmit over wireless channels using same frequencies, thus communication channels are symmetric [9]. This means that if there exists a link  $(i, j)$ , then there is also the link  $(j, i)$ . Notice, however, that the weights are such that  $g_{ij}(t) \neq g_{ji}(t)$ .

We assume each node  $i$  computes an estimate  $x_i(t)$  of  $d(t)$  by taking a linear combination of its own and of its neighbors' estimates and measurements:

$$x_i(t) = \sum_{j \in \mathcal{N}_i(t)} k_{ij}(t) \phi_{ij}(t) x_j(t-1) + \sum_{j \in \mathcal{N}_i(t)} h_{ij}(t) \phi_{ij}(t) u_i(t), \quad (1)$$

where  $k_{ij}(t) \phi_{ij}(t)$  and  $h_{ij}(t) \phi_{ij}(t)$  are weights assigned to the arc  $(j, i)$ . In particular,  $k_{ij}(t), h_{ij}(t) \in \mathbb{R}$  are deterministic values, whereas  $\phi_{ij}(t)$ , with  $i \neq j$ , are binary random variables modelling packet losses of the link from node  $j$  to node  $i$ . We assume that the estimate  $x_i(t-1)$  and measurements  $u_i(t)$  are available at node  $i$  at time  $t$ , so that  $\phi_{ii}(t) = 1$  for all  $t \geq 0$ . For  $i \neq j$ , we assume that the random variables  $\phi_{ij}(t)$  are identically distributed with probability mass function  $\Pr(\phi_{ij}(t) = 1) = p$ , and  $\Pr(\phi_{ij}(t) = 0) = q = 1 - p$ , where  $p \in [0, 1]$  denotes the successful packet reception probability. We also make the natural assumption that the measurement noise  $v(t)$  and the random variables  $\phi_{ij}(t)$  are independent. We write (1) in matrix form as

$$\begin{aligned} x(t) &= (K(t) \circ \Phi(t)) x(t-1) + (H(t) \circ \Phi(t)) u(t) \\ &= K_{\phi(t)}(t) x(t-1) + H_{\phi(t)}(t) u(t), \end{aligned}$$

where  $x = (x_1, \dots, x_N)^T$  and

$$[K(t)]_{ij} = \begin{cases} k_{ij}(t), & \text{if } j \in \mathcal{N}_i(t); \\ 0, & \text{otherwise.} \end{cases}$$

The matrix  $H(t)$  is defined similarly and the elements of  $\Phi(t)$  are equal to  $\phi_{ij}$ . In next section we derive conditions on  $K(t)$  and  $H(t)$  that guarantee the estimation error to converge.

### III. CONVERGENCE OF THE ESTIMATION ERROR

Define the estimation error  $e(t) = x(t) - d(t)\mathbf{1}$ . Introduce  $\delta(t) = d(t) - d(t-1)$ , so that the error dynamics can be described as

$$\begin{aligned} e(t) &= K_{\phi(t)}(t) e(t-1) + d(t) (K_{\phi(t)}(t) + H_{\phi(t)} - I) \mathbf{1} \\ &\quad - \delta(t) K_{\phi(t)}(t) \mathbf{1} + H_{\phi(t)}(t) v(t). \end{aligned} \quad (2)$$

We will assume that  $\delta(t)$  is slowly varying, so  $|\delta(t)| < \Delta$  for some small  $\Delta > 0$ . The following results holds:

*Proposition 3.1:* Consider the system (2) and assume that:

- (i)  $((K(t) + H(t) - I) \circ \bar{\Phi}(t)) \mathbf{1} = 0$  for any realization  $\bar{\Phi}(t)$  of the stochastic process  $\Phi(t)$ ,
- (ii)  $|\delta(t)| < \Delta$  for all  $t \in \mathbb{N}_0$ ,
- (iii)  $\gamma(K(t)) \leq \gamma_{\max} < 1$  for all  $t \in \mathbb{N}_0$ .

Then,

$$\lim_{t \rightarrow +\infty} \|\mathbb{E}_{\phi} \mathbb{E}_v e(t)\| \leq \frac{\Delta \sqrt{N} \gamma_{\max}}{1 - \gamma_{\max}}.$$

*Proof:* Taking the expected value of (2), with respect to the pdf's of  $v(t)$  and  $\phi_{ij}(t)$  and using assumption (i), we have that

$$\mathbb{E}_{\phi} \mathbb{E}_v e(t) = (K(t) \circ Q) \mathbb{E}_{\phi} \mathbb{E}_v e(t-1) - \delta(t) (K(t) \circ Q) \mathbf{1}.$$

Let us, for simplicity, define  $s(t) = \mathbb{E}_{\phi} \mathbb{E}_v e(t)$ . The dynamics of  $s$  is thus given by a deterministic time-varying linear system. Let us consider the function  $V(s(t)) = \|s(t)\|$ . Then we have that

$$V(s(t)) \leq (\|K(t) \circ Q\| - 1) V(s(t-1)) + \|K(t) \circ P\| \Delta \sqrt{N}.$$

Notice that

$$\begin{aligned} \|K(t) \circ Q\| &= \|K(t) \circ p \mathbf{1} \mathbf{1}^T + K(t) \circ q I\| \\ &= \|p K(t) + K(t) \circ q I\| \leq p \|K\| + \|K(t) \circ q I\| \\ &\leq p \|K(t)\| + q \|K(t)\| = \|K(t)\|, \end{aligned}$$

where we have used that for any unitarily invariant norm and for any matrix  $A$  and  $B$ , it holds that  $\|A \circ B\|^2 \leq \|A^* A\| \|B^* B\|$  [10], that  $\|K^T K\|^{1/2} = \|K\|$ , and that  $\|q^2 I\|^{1/2} = q$ . Hence, assumption (iii) gives that  $\|K(t) \circ Q\| \leq \gamma_{\max} < 1$ . Therefore, it follows

$$V(s(t)) \leq \gamma_{\max}^t V(s(t-1)) + \gamma_{\max} \frac{1 - \gamma_{\max}^t}{1 - \gamma_{\max}} \Delta \sqrt{N},$$

from where, taking the limit  $t \rightarrow +\infty$ , we get the result.  $\blacksquare$  Proposition 3.1 provides conditions under which the estimation error converges to a neighborhood of the origin. The estimate will thus have a bias. From the proof of the Proposition 3.1, it is clear that the bias depends on  $\|K(t)\|$ ,  $\sqrt{N}$  and  $\Delta$ . If the signal  $d(t)$  is slowly varying,  $\Delta \ll 1$ , and  $\|K(t)\|$  small, then the bias will be small. An unbiased estimate can be obtained setting  $K(t)\mathbf{1} = 0$  along with the conditions in Proposition (3.1). However, under this supplementary condition, the minimum variance estimator has  $K(t) = 0$ , and each node just takes the average of the communicated measurements.

In the next session, we will pose an optimization problem that allows us to distributely determine optimal weights such that the error variance is minimized.

### IV. MINIMUM VARIANCE DISTRIBUTED ESTIMATOR

In order to design a minimum variance distributed estimator, we need to consider how the error variance changes over time. The error variance, with respect to the measurement noise, is  $P(t) = \mathbb{E}_v (e(t) - \mathbb{E}_v e(t))(e(t) - \mathbb{E}_v e(t))^T$ , and its update equation is

$$P(t) = K_{\phi(t)}(t) P(t-1) K_{\phi(t)}^T(t) + \sigma^2 H_{\phi(t)}(t) H_{\phi(t)}^T(t).$$

As cost function, we consider the trace of  $P(t)$ , and thus the problem is to determine  $K_{\phi(t)}(t)$  and  $H_{\phi(t)}(t)$  that minimize  $\text{tr } P(t)$  for a given  $P(t-1)$ . Notice that the matrix  $K_{\phi(t)}(t)$  and  $H_{\phi(t)}(t)$  need to fulfill constraints imposed by Proposition 3.1 to guarantee the convergence of the error. Furthermore, observe that minimizing  $\text{tr } P(t)$  corresponds to minimizing the error variance at each node. We can thus distribute the computation among the nodes in the network.

The error variance at node  $i$  can be written as

$$\mathbb{E}_v e_i^2(t) = k_{i_{\phi(t)}}(t)^T P(t-1) k_{i_{\phi(t)}}(t) + \sigma^2 h_{i_{\phi(t)}}(t)^T h_{i_{\phi(t)}}(t), \quad (3)$$

where  $k_{i_{\phi(t)}}(t)^T$  and  $h_{i_{\phi(t)}}(t)^T$  are the  $i$ -th rows of the matrices  $K_{\phi(t)}(t)$  and  $H_{\phi(t)}(t)$ , respectively.

Let us introduce  $M_i(t) = \text{card}(\mathcal{N}_i(t))$  and let us collect in a vector  $\epsilon_i(t) \in \mathbb{R}^{M_i(t)}$  the estimation errors available at node  $i$  at time  $t$ . Assume that the nodes have a unique identifier, so that it sorts its neighbors according to the identifier, and sort the elements of  $\epsilon_i(t)$  accordingly:

$$\epsilon_i(t) = (e_{i_1}(t), \dots, e_{i_{M_i(t)}}(t))^T, \quad i_1 < \dots < i_{M_i(t)} \leq i_{M_i \max},$$

where  $M_{i \max} = \max_t M_i(t)$ . Consider

$$\mathbb{E}_v (e_i^2 | \Phi(t) = \varphi(t)) = \kappa_{i_{\varphi(t)}}^T(t) \Gamma_i(t-1) \kappa_{i_{\varphi(t)}}(t) + \sigma^2 \eta_{i_{\varphi(t)}}^T(t) \eta_{i_{\varphi(t)}}(t). \quad (4)$$

where  $\varphi(t)$  is a realization of the packet loss process  $[\Phi(t)]_i$ . The vectors  $\kappa_{i_{\varphi(t)}}(t)$  and  $\eta_{i_{\varphi(t)}}(t)$ , corresponding to the outcome of the packet loss process, are the non-zero elements of the vectors  $k_{i_{\varphi(t)}}(t)$  and  $h_{i_{\varphi(t)}}(t)$ , respectively, ordered according to the node order described above. In the following, to keep a lighter notation, we consider  $\kappa_i(t)$  and  $\eta_i(t)$  instead of  $\kappa_{i_{\varphi(t)}}(t)$  and  $\eta_{i_{\varphi(t)}}(t)$ , respectively.

The matrix  $\Gamma_i(t-1) \in \mathbb{R}^{M_i(t)} \times \mathbb{R}^{M_i(t)}$  in (4) is the error covariance matrix associated to  $\epsilon_i(t-1)$ , namely

$$\Gamma_i(t-1) = \mathbb{E}_v (\epsilon_i(t-1) - \mathbb{E}_v (\epsilon_i(t-1))) \times (\epsilon_i(t-1) - \mathbb{E}_v (\epsilon_i(t-1)))^T.$$

Note that  $\Gamma_i(t-1)$  is a sub-matrix of  $P(t-1)$ , obtained by puncturing from  $P(t-1)$  the rows and columns corresponding to neighbors of node  $i$  that do not communicate at time  $t$ . We did not so far explain how  $\Gamma_i(t)$  is computed by each node when data is lost. This issue is discussed in detail in the next section.

To obtain the optimal weights  $\kappa_i(t)$  and  $\eta_i(t)$  that minimize the estimation error variance in (4), each node needs to solve the following optimization problem

$$\begin{aligned} \min_{\kappa_i(t), \eta_i(t)} & \kappa_i^T(t) \Gamma_i(t-1) \kappa_i(t) + \sigma^2 \eta_i^T(t) \eta_i(t) \\ \text{s.t.} & (\kappa_i(t) + \eta_i(t))^T \mathbf{1} = 1 \\ & \gamma(K(t)) < 1. \end{aligned} \quad (5)$$

The constraints comes from Proposition 3.1. The first one is the distributed equivalent to  $(K_{\phi(t)}(t) + H_{\phi(t)}(t) - I) \mathbf{1} = \mathbf{1}$ . The second constraint is global, since  $K(t)$  depends on all  $k_i(t)$ ,  $i = 1, \dots, N$ . For  $i = 1, \dots, N$ , let us define the set  $\Theta_i(t) = \{j \neq i : \mathcal{N}_j(t) \cap \mathcal{N}_i(t) \neq \emptyset\}$ , which is the collection of nodes located at two hops distance from node

$i$  plus neighbor nodes of  $i$ , at time  $t$ . The following result holds (the proof is along the lines of Proposition 3.1 in [11]).

*Proposition 4.1:* Suppose there exist  $\psi_i(t) > 0$ ,  $i = 1, \dots, N$ , such that

$$\psi_i(t) + \sqrt{\psi_i(t)} \sum_{j \in \Theta_i(t)} \sqrt{\psi_j(t)} \leq \gamma_{\max}. \quad (6)$$

If  $\|\kappa_i(t)\|^2 \leq \psi_i(t)$ ,  $i = 1, \dots, N$ , then  $\gamma(K(t)) \leq \gamma_{\max} < 1$ .

*Remark 4.1:* The bound in Proposition 4.1 is rather conservative, because no a-priori knowledge of the network topology is used, and the proof rely on the conservative Geršgorin discs. Many other results bounding the eigenvalues of a matrix by its elements are known, e.g., [12, pag. 378–389]. However, we have found none that provides bounds requiring only local information for a node  $i$ . Note also that Perron-Frobenius theory cannot be applied, because we have no assumption on the sign of the elements of  $K(t)$ .  $\diamond$

In the optimization problem (5) we can now substitute the global constraint with the constraint  $\|\kappa_i\|^2 \leq \psi_i(t)$  where  $\psi_i(t)$  satisfies the set of nonlinear inequalities (6). Suitable values of  $\psi_i(t)$  are computed in a distributed way, solving (6) at the equality and using the component solution method [4], [11]. The optimization problem (5) with the new constraint is similar to a Quadratically Constrained Quadratic Problem [13]. Assume the matrix  $\Gamma(t-1)$  is positive definite. Since also  $\sigma^2 > 0$ , the cost function in (5) is convex. The problem also admits a strict interior point solution, corresponding to  $\kappa_i(t) = 0$  and  $\eta_i(t) \mathbf{1} = 1$ . Thus Slater's condition is satisfied and strong duality holds [13, pag. 226]. The problem, however, does not have a closed form solution and thus we need to rely on numerical algorithms to derive the optimal  $\kappa_i(t)$  and  $\eta_i(t)$ . We have the following proposition whose proof is in [11].

*Proposition 4.2:* For a given covariance matrix  $\Gamma_i(t-1)$ , the values of  $\kappa_i(t)$  and  $\eta_i(t)$  that minimizes (5) are given by

$$\kappa_i(t) = \frac{\sigma^2 (\Gamma_i(t-1) + \lambda_i(t) I)^{-1} \mathbf{1}}{\sigma^2 \mathbf{1}^T (\Gamma_i(t-1) + \lambda_i(t) I)^{-1} \mathbf{1} + M_i(t)}, \quad (7)$$

$$\eta_i(t) = \frac{\mathbf{1}}{\sigma^2 \mathbf{1}^T (\Gamma_i(t-1) + \lambda_i(t) I)^{-1} \mathbf{1} + M_i(t)}, \quad (8)$$

with the optimal Lagrange multiplier  $\lambda_i(t) \in [0, \sigma^2 / \sqrt{M_i(t) \psi_i(t)} - \ell_m(\Gamma_i(t-1))]$ .

Previous proposition provides an interval within which the optimal  $\lambda_i$  can be found. Simple search algorithms can be considered to numerically solve  $(\kappa_i^*)^T \kappa_i^* - \psi_i$  for  $\lambda_i$ , such as, for example, the bisection one [13].

## V. COVARIANCE ESTIMATION

Since the estimator is a discrete linear time-varying system and the stochastic process  $x(t)$  and thus  $\epsilon(t)$  is not stationary, the estimation of the error covariance  $\Gamma(t)$  needed in Proposition 4.2 at next time step is not an easy task. However, if we consider the signals in the quasi-stationary sense, estimation based on samples guarantees to give good results. It turns out that, if a quasi-stationary signal is the input of a BIBO time-varying linear system, then its output is also quasi-stationary [14, pag. 34] [15]. In particular in our

case the input signal is the measurement signal  $u(t)$  which is (component-wise) stationary and ergodic and thus also quasi-stationary. This implies that also  $x(t)$  is quasi-stationary (component-wise) since it is the output of a uniformly exponentially stable time-varying linear system. We can estimate the error covariance using the sample covariance. In order to do that, let us introduce the vector  $\hat{\epsilon}_i(t)$  collecting the estimates of the estimation error of node  $i$ , at time  $t$ . Such a vector is defined using the same rules for  $\epsilon_i(t)$ , where the nodes are sorted according to the identifier:

$$\hat{\epsilon}_i(t) = (\hat{\epsilon}_{i_1}(t), \dots, \hat{\epsilon}_{i_{M_i(t)}}(t))^T, \quad i_1 < \dots < i_{M_i(t)} \leq i_{M_i \max}.$$

Using previous definition, the sample mean  $\hat{m}_{\epsilon_i}(t)$  and covariance estimate of  $\hat{\Gamma}_i(t)$  can be computed as

$$\hat{m}_{\epsilon_i}(t) = \frac{1}{t} \sum_{i=0}^t \hat{\epsilon}_i(t) \quad (9)$$

$$\hat{\Gamma}_i(t) = \frac{1}{t} \sum_{i=0}^{t-1} (\hat{\epsilon}_i(t) - \hat{m}_{\epsilon_i}(t))(\hat{\epsilon}_i(t) - \hat{m}_{\epsilon_i}(t))^T \quad (10)$$

where  $\hat{\epsilon}_i(t) = x^i(t) - \hat{d}(t)$ , and  $\hat{d}(t)$  is an unbiased estimator of the signal we would like to track.

*Remark 5.1:* There are two major issues concerning the computations of (9) and (10). First, an unbiased estimate of  $d(t)$  must be available. Indeed, biased estimate would lead to a non-consistent estimation of the covariance matrix. Second, packet losses introduce zeros in the vector  $\hat{\epsilon}_i(t)$ , whereas the reception of packets from nodes not connected at the previous time instance requires an initialization of the corresponding entries of the covariance matrix (10).

We approach the issue mentioned in the previous remark in the sequel.

#### A. Unbiased Covariance Estimation

The problem reduces to estimate  $\hat{d}(t)$  from available data. A node  $i$  in the network has the estimates  $x_{i_j}(t)$  and measurements  $u_{i_j}(t)$  with  $i_j \in \mathcal{N}_i$ . Let  $x^i(t)$  and  $u^i(t)$  be the collection of all these values. From the design of the estimator described above this data set is modelled as

$$x^i(t) = d(t)\mathbf{1} + \xi(t) + w(t), \quad u^i(t) = d(t)\mathbf{1} + v(t),$$

where  $\xi(t) \in \mathbb{R}^{M_i(t)}$  is the bias of the estimates and  $w(t)$  is zero-mean Gaussian noise modelling the variance of the estimator. In summary, node  $i$  has available  $2M_i(t)$  data values in which half of the data are corrupted by a small biased term  $\xi(t)$  and a low variance noise  $w(t)$  and the other half is corrupted by zero-mean Gaussian noise  $v(t)$  with high variance.

It is clear that using only  $u^i(t)$  to estimate  $\hat{d}(t)$  we can obtain an unbiased estimate of  $d(t)$ , however its covariance is rather large since  $M_i(t)$  is typically small. Using only measurements we then over-estimate the error covariance and this results in poor performance. On the other hand using  $x^i(t)$  only determines an under-estimate of the covariance, which rapidly makes the weights  $\eta_i(t)$  vanish. In this case the signal measurements are discarded and thus tracking becomes impossible.

In order to use both  $x^i(t)$  and  $u^i(t)$  a minimum least square problem can be posed. However, numerical technique have to be used to get the optimal solution. Since this is computationally expensive, we consider an alternative approach based on a regularized problem and the use of the Generalized Cross-Validation method [16]. See [3] for details. Another approach could be based on the use of stochastic least square. Unfortunately, such an approach requires knowledge of the covariance matrix of  $x^i(t)$ , which is not easily available. We do not consider this approach in the present paper.

#### B. Packet losses

The computation of the average (9) and covariance matrix (10) has to take into account that packets are lost. Indeed, if node  $i$  does not receive estimates  $x_j(t)$  from some node  $j \in \mathcal{N}_i(t)$ , then  $\hat{\epsilon}_j(t)$  should be put to zero in  $\hat{\epsilon}_i(t)$ , and the covariance matrix  $\hat{\Gamma}_i(t)$  is filled with zeros along the  $j$ th row and  $j$ th column. This corresponds to say that when packets are lost, statistical information associated to the estimates from the node  $j$  are lost.

If node  $i$  receives estimates  $x_j(t)$  from some new node  $j \in \mathcal{N}_i(t)$ , then the corresponding  $\hat{\epsilon}_j(t)$  should simply be inserted in the corresponding position in  $\epsilon_i(t)$ . We put  $\max_k \hat{\Gamma}_i(t-1)_{kk}$  in the position  $jj$ , i.e., we initialize the variance of the error for the new neighbor of node  $i$  to be the maximum variance for previous set of neighbors. This choice is motivated by the fact that nodes are collaborating in estimating  $d(t)$ , and their estimation error is decreasing and alike, as time goes on. Another possibility could be to initialize  $\hat{\Gamma}_i(t)_{jj}$  to  $\sigma^2$ . However, such choice is a rather rough over approximation of the real covariance matrix, and numerical results we have performed showed that it strongly deteriorates performance.

Once the estimation of the covariance matrix has been updated according to the packet loss process described above, and using (9) and (10), we can extract from  $\hat{\Gamma}_i(t)$  the matrix  $\tilde{\Gamma}_i(t)$  corresponding to the index of the neighbors communicating with  $i$ . This guarantees the invertibility of the estimate of the covariance matrix. Such a matrix will then be used to compute the filter coefficients (8).

## VI. PERFORMANCE ANALYSIS

It is interesting to understand the filter performance in terms of its variance. Given the optimal weights  $\kappa_i(t)$  and  $\eta_i(t)$  for node  $i$ , we have the following results:

*Proposition 6.1:* For any  $\lambda_i \geq 0$  and for any packet loss realization  $\varphi(t) \in \Phi(t)$ , the optimal value of  $\kappa_i(t)$  and  $\eta_i(t)$  are such that the error variance at node  $i$  satisfies

$$\mathbb{E}(e_i^2 | \Phi(t)) = \varphi(t) \leq \frac{\sigma^2}{M_i(t)}.$$

*Proof:* The variance at each time instance is given by equation (4). Using the optimal values in the error variance

we have

$$\begin{aligned} \mathbb{E}(e_i^2|\Phi(t)) &= \varphi(t) = \frac{\sigma^2}{M_i(t) + \sigma^2 \mathbf{1}^T(\Gamma_i(t-1) + \lambda_i(t)I)^{-1} \mathbf{1}} \\ &\quad - \frac{\sigma^4 \lambda_i(t) \mathbf{1}^T(\Gamma_i(t-1) + \lambda_i I)^{-2} \mathbf{1}}{(M_i(t) + \sigma^2 \mathbf{1}^T(\Gamma_i(t-1) + \lambda_i(t)I)^{-1} \mathbf{1})^2} \\ &\leq \frac{\sigma^2}{M_i(t) + \sigma^2 \mathbf{1}^T(\Gamma_i(t-1) + \lambda_i(t)I)^{-1} \mathbf{1}}. \end{aligned}$$

Since  $\Gamma_i(t-1) \succ 0$  and  $\lambda_i \geq 0$ , then  $\mathbf{1}^T(\Gamma_i(t-1) + \lambda_i I)^{-1} \mathbf{1} > 0$ . Thus the proposition follows. ■

Notice that previous proposition guarantees that the estimation error at each time  $t$ , and in each node, is always upper-bounded by the variance of the estimator that just takes the averages of the  $M_i(t)$  entries of  $u_i(t)$ . If we assume that  $\ell_M(\Gamma(t-1) + \lambda_i(t)I) \leq \sigma^2$  then we can improve the bound in Proposition 6.1. This assumption is hard to prove formally because the matrix  $\Gamma(t-1)$  is estimated in a rather involved way, as we described in Section V, and also because  $\lambda_i$  is computed numerically. However, numerical simulations, which we do not report due to space limitations, show that such an assumption holds for  $q \in [0, 0.3]$ .

Let consider in detail the system (3). The optimal values of  $k_{i_\phi(t)}(t)$  and  $h_{i_\phi(t)}(t)$

$$\begin{aligned} k_{i_\phi(t)}(t) &= \frac{\sigma^2(\Gamma_{i_\phi(t)}(t-1) + \lambda_{i_\phi(t)}I)^{-1} \mathbf{1}_{\phi(t)}}{\sigma^2 \mathbf{1}_{\phi(t)}^T(\Gamma_{i_\phi(t)}(t-1) + \lambda_{i_\phi(t)}I)^{-1} \mathbf{1}_{\phi(t)} + \mathbf{1}_{\phi(t)}^T \mathbf{1}_{\phi(t)}} \\ h_{i_\phi(t)}(t) &= \frac{\mathbf{1}_{\phi(t)}}{\sigma^2 \mathbf{1}_{\phi(t)}^T(\Gamma_{i_\phi(t)}(t-1) + \lambda_{i_\phi(t)}I)^{-1} \mathbf{1}_{\phi(t)} + \mathbf{1}_{\phi(t)}^T \mathbf{1}_{\phi(t)}} \end{aligned} \quad (11)$$

where  $\mathbf{1}_{\phi(t)} = \mathbf{1} \circ (\phi(t)_{ii_1}, \dots, \phi(t)_{ii_{M_i \max}})^T$  and the covariance matrix  $\Gamma_{i_\phi(t)}(t-1) \in \mathbb{R}^{M_i \max} \times \mathbb{R}^{M_i \max}$  is a covariance matrix that can be estimated for any packet loss process. We define this matrix so that its non-zero elements are the entries of  $\Gamma_i(t-1)$ , which are corresponding to packets received. Notice that under the hypothesis that  $\ell_M(\Gamma_i(t-1)) \leq \sigma^2$ , it holds that  $\ell_M(\Gamma_{i_\phi(t)}(t-1)) \leq \sigma^2$ , since this last matrix is obtain from the first introducing pairs of zero row-columns. We also define the inverse  $(\Gamma_{i_\phi(t)}(t-1) + \lambda_{i_\phi(t)}I)^{-1}$  such that the non-zero entries are as those of the matrix  $(\Gamma_i(t-1) + \lambda_i I)^{-1}$ , corresponding to any received packet, and zero otherwise. The  $\lambda_{i_\phi(t)}$  are, for any  $\phi(t)$ , non-negative scalars. We then have the following result.

*Proposition 6.2:* Let assume that  $\ell_M(\Gamma_{i_\phi(t)}(t-1) + \lambda_i(t)I) \leq \sigma^2$ , for any  $\phi(t) \in \Phi(t)$ , and that packets are dropped with probability  $q$ . Then

$$\mathbb{E}_\phi \mathbb{E}_v e_i^2 \leq \frac{1-q^{M_i \max}}{1-q} \frac{\sigma^2}{2M_i \max}. \quad (12)$$

*Proof:* The variance at node  $i$  is given by (3), and the optimal values of  $k_{i_\phi}$  and  $h_{i_\phi}$  are as in (11). Following the same steps as in the proof of Proposition 6.1, we have that

$$\mathbb{E}_\phi \mathbb{E}_v e_i^2 \leq \mathbb{E}_\phi \frac{\sigma^2}{\mathbf{1}_\phi^T \mathbf{1}_\phi + \sigma^2 \mathbf{1}_\phi^T (\Gamma_{i_\phi} + \lambda_{i_\phi} I)^{-1} \mathbf{1}_\phi}. \quad (13)$$

Previous inequality holds since the argument of the statistical expectation is always positive and we are taking the

expectation over a positive distribution, thus the sign of the argument is maintained [12, pag. 392]. From the assumption that  $\ell_M(\Gamma_{i_\phi(t)}(t-1) + \lambda_i I) \leq \sigma^2$ , it follows

$$\mathbf{1}_\phi^T (\Gamma_{i_\phi} + \lambda_i I)^{-1} \mathbf{1}_\phi \geq \frac{\mathbf{1}_\phi^T \mathbf{1}_\phi}{\sigma^2}.$$

Using previous inequality in (13), we have

$$\mathbb{E}_\phi \mathbb{E}_v e_i^2 \leq \mathbb{E}_\phi \frac{\sigma^2}{2 \mathbf{1}_\phi^T \mathbf{1}_\phi}.$$

The random variable  $\mathbf{1}_\phi^T \mathbf{1}_\phi = \sum_{i_j=1}^{M_i \max} \phi_{ii_j}$  is binomial with parameter  $p$ , thus the right hand side of previous inequality is the first negative moment of a binomial random variable, which can be computed as in [17, pag. 432, eq. 3.4], obtaining the sought result. ■

Notice that for  $q = 0$ , namely no packet errors and with the assumption that the error covariance estimate is bounded, previous Proposition allow us to refine the bound obtained in Proposition 6.1, where no hypothesis on the error covariance where done.

## VII. NUMERICAL RESULTS

In this section we report some numerical results with the purpose to compare our strategy with other solutions from the literature. In particular, we considered a solution, denoted  $E_{av}$ , where the new estimate is the average between previous estimates and the current input. We also consider a solution where the weights are computed based on the Laplacian matrix associated to the communication graph, which we denoted with  $E_{lap}$ . The estimator proposed in this paper is denoted with  $E_{new}$ . Simulations were carried out for two networks, with  $N = 20$  and  $N = 35$  nodes randomly distributed on an environment of size  $N/2 \times N/2$ . Pair of nodes communicate only if their distance is less than  $2\sqrt{N}$ . The signal to track is shown in the first plot of Fig. 2.

Simulations were carried out for four cases of packet losses: perfect connectivity,  $p = 0.9$ ,  $p = 0.7$ , and  $p = 0.5$ . The noise variance was chosen to be  $\sigma^2 = 1.5$ , which corresponds to a high peak signal to noise ratio.

Time series of the estimates for our solution  $E_{new}$ , the Average Estimator  $E_{av}$  and the Laplacian  $E_{lap}$  are reported in Fig. 2 for the cases of  $N = 20$  and  $p = 0.3$ . The good behavior of  $E_{new}$  can be clearly noticed. It can be show that it remains basically the same also for other scenarios of number of nodes and packet loss.

In order to compare the different solutions, we consider the average of the mean square estimation error over all the nodes of the network, which we denote MSE. As performance index we consider

$$\mu_i = \frac{\text{MSE}(E_i) - \text{MSE}(E_{new})}{\text{MSE}(E_i)}, \quad i \in \{av, lap\}.$$

A summary of the results of the simulations is shown in Fig. 1, where each curve is referred to the pair estimator-network size. The performance improvement of our filter is evident in any situation of packet losses. Specifically, when  $N = 20$ , averaging over different  $q$ , there is an improvement of  $\mu_{av}^{N=20} = 52\%$  with respect to the Average estimator and

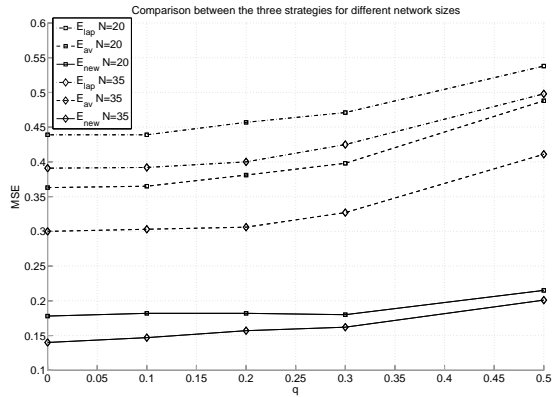


Fig. 1. Performance comparison among the different filtering strategies. The curves refer to the three estimators, Average, Laplacian and the one proposed in this paper for networks with  $N = 20$  and  $N = 35$ .

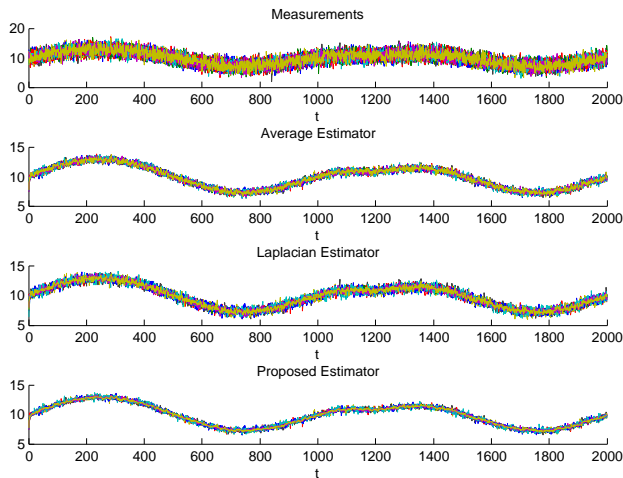


Fig. 2. Signals representing the estimate and the measurements of  $d(t)$  for each of the 20 nodes in the network when  $p = 0.3$ .

of about  $\mu_{\text{lap}}^{N=20} = 59\%$  with respect to the Laplacian. In the case of  $N = 35$ , we obtain a  $\mu_{\text{av}}^{N=35} = 51\%$  with respect to Average estimator and of about  $\mu_{\text{lap}}^{N=35} = 62\%$  with respect to the Laplacian. In Fig. 3 we show the value of  $\gamma(K(t))$  for the simulation with  $N = 20$  nodes and for different values of  $q$ . As it can be clearly seen the maximum singular value is always below 1, thus confirming the validity of the local constraint in (6) adopted to the design of the filter coefficients.

## VIII. CONCLUSIONS

In this paper, we have presented a decentralized cooperative estimation algorithm for tracking a time-varying signal using a wireless sensor network. A mathematical framework is proposed to design a filter, which is supposed to run locally in each node of the network. As a relevant contribution, performance analysis has been carried out including time varying communication networks with Bernoulli models of the wireless channel. Theoretical analysis shows that the filter is stable, and the variance of the estimation error is bounded

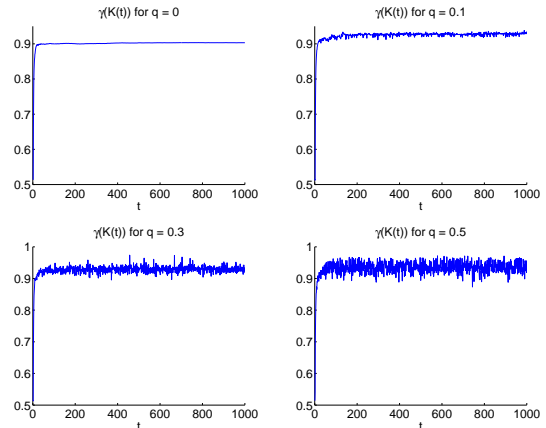


Fig. 3. Value of  $\gamma(K(t))$  as function of  $t$  for different values of  $q$  for the case  $N = 20$ .

even in the presence of large packet losses probabilities. Numerical results illustrate the validity of our approach.

## ACKNOWLEDGMENT

The authors thank Prof. M. Johansson, and Prof. B. Wahlberg for interesting and fruitful discussions.

## REFERENCES

- [1] A. Savvides, W. L. Garber, R. L. Moses, and M. B. Srivastava, "An analysis of error inducing parameters in multihop sensor node localization," *IEEE Transaction on Mobile Computing*, Vol. 4, No. 6, 2005.
- [2] A. Speranzon, C. Fischione, and K. H. Johansson, "Distributed and collaborative estimation over wireless sensor networks," in *In Proc. of IEEE CDC*, 2006.
- [3] —, "A distributed estimation algorithm for tracking over wireless sensor networks," in *IEEE ICC*, 2007.
- [4] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [5] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," *In Proc. of IEEE IPSN*, 2005.
- [6] L. Xiao, S. Boyd, and S. J. Kim, "Distributed average consensus with least-mean-square deviation," *Submitted to Journal of Parallel and Distributed Computing*, 2006.
- [7] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *System Control Letter*, 2004.
- [8] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," *In Proc. of IEEE CDC*, 2005.
- [9] G. L. Stüber, *Principles of Mobile Communication*. Kluwer Academic Publishers, 1996.
- [10] R. A. Horn and R. Mathias, "An analog of the Cauchy-Schwarz inequality for Hadamard products and unitarily invariant norms," *SIAM J. on Mat. Anal. and App.*, vol. 11, no. 4, pp. 481–498, 1990.
- [11] A. Speranzon, C. Fischione, and K. H. Johansson, "A distributed minimum variance estimator for wireless sensor networks," School of Electrical Engineering, KTH, Sweden, Tech. Rep., 2007.
- [12] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [14] L. Ljung, *System Identification: Theory for the User (2nd Edition)*. Prentice Hall PTR, 1998.
- [15] L. Ljung and B. Wahlberg, "Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra," *Adv. in App. Prob.*, vol. 24, no. 2, 1992.
- [16] G. H. Golub and U. Von Matt, "Tikhonov regularization for large scale problems," Stanford SCCM, Stanford, Tech. Rep. 97-03, 1997.
- [17] M. T. Chao and W. E. Strawderman, "Negative moments of positive random variables," *Journal of the American Statistical Association*, vol. 67, no. 388, 1972.