Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Information fusion strategies from distributed filters in packet-drop networks

A. Chiuso, L. Schenato

Abstract— **In this paper we study different distributed estimation schemes for stochastic discrete time linear systems where the communication between the sensors and the estimation center is subject to random packet loss. Sensors are provided with computational and memory resources so that they can potentially perform data processing of the measurements before sending their information. In particular, we explore three different strategies. The first, named measurement fusion (MF), optimally fuses the raw measurements received so far from all sensors. The second strategy, named optimal partial estimate fusion (OPEF), optimally fuses at the central node the last local state estimates received from each sensor. The last strategy, named open loop partial estimate fusion (OLPEF), simply sums local state estimates received from each sensor and replace the lost ones with their open loop prediction. We provide some analytical results about the performance of these three schemes in special regimes conditions, namely low and high process noise. We also show through numerical simulations that, although none of the three schemes achieves the ideal performance of a scheme with infinite bandwidth communication between sensors and the central node, the OPEF scheme provides almost ideal performance.**

I. Introduction

The rapid proliferation of large wireless interconnected systems capable of sensing and computation is posing several challenges due to the unavoidable lossy nature of the wireless channel. These challenges are particularly evident in control and estimation applications since packet loss and random delay degrade the overall system performance, thus motivating the development of novel tools and algorithms, as illustrated in the survey [8]. In this work we focus on the problem of estimating a stochastic discrete time linear system observed by a number of sensors which can communicate with a central node via a wireless lossy channel.

A. Previous Work

There is a vast literature regarding distributed estimation and sensor fusion with perfect communication links. It is well known that the optimal solution in the standard scenario where all sensors are co-located with the estimation center, is given by the centralized Kalman filter (CKF) [2]. In the seminal papers [13] and [9] it was shown that it is not necessary to send the raw measurements to the central node to recover the CKF estimate, but it is possible to reconstruct it from local Kalman filter (LKF) estimates generated by each sensor. In particular, the CKF can be obtained as the output of a linear filter which uses the LKF estimates as inputs. The idea behind the fusion of LKF estimates rather than the raw measurements was motivated by the need of distributing part of the computational burden of the central estimation center to the sensors. More recently Wolfe et al. [14] showed that the computational load of the central node can be reduced even further by running on each sensor a local filter which generates a partial estimate of the state so that the central node just need to sum them together to recover the CKF estimate. We refer to this strategy as partial estimate fusion (PEF). Moreover, this strategy does not even require uncorrelated measurement noise among sensors, differently from [13]. There are also dedicated distributed estimation algorithms such as the federated filters proposed by Carlson [5]. However, the framework adopted in all these works did not include packet loss nor delay, and the topology was supposed to be known to all sensors and the central node. Sensor fusion, whose goal is to devise efficient numerical algorithms to fuse measurements (and not local estimates) from heterogeneous sensors like radars and GPS with possibly different random delays or missing data, is also a deeply investigated area, in particular in the context of moving target tracking [4]. For example in [3] and [15] the authors showed how to perform optimal estimation with time-varying delay and out-of-order packets without requiring the storage of large memory buffers and the inversion of many matrices. In [12] the authors provided lower and upper bounds for optimal estimator subject to random measurement loss, and in [10] those results were extended to multiple distributed sensors subject to simultaneous packet loss and random delay. Finally, the recent paper [11] analyzes some tradeoffs between communication, computation and estimation performance in multi-hop tree networks.

B. Motivations

Differently from distributed estimation with perfect communication and sensor fusion, little attention has been given to fusion of local estimates from multiple sensors subject to random packet loss and random delay. In fact, it has been shown in [7] that sending the LKF estimates allows the central node to construct a better state estimate than sending the raw measurements, even in the presence of packet loss. This is because the local estimate includes the information about all previous measurements, therefore as soon as the central node receives the local estimate it can reconstruct the optimal estimate even if some previous packets were lost. Differently, by sending the raw measurements, if a measurement is lost then the information that it conveys is lost forever. This observation, which is valid in general only when a single sensor is considered, suggests that sending

L. Schenato is with the Department of Information Engineering, University of Padova, Italy schenato@dei.unipd.it
A. Chiuso is with the Department of Management and Engineering, University of Padova, Vicenza, Italy, chiuso@dei.unipd.it
This work has been partially supported by European Union project FeedNetBack, by CARIPARO Foundation by the PRIN Project "New Methods and Algorithms for Identification and Adaptive Control of Technological Systems".

a local estimate of the state $\hat{x}^i_{t|t}$ is the right thing to do also in the context of lossy communication. Indeed we will show that, when there is no process noise (Q=0), sending partial estimates, as suggested in [14], allows to recover the CKF filter as if all measurements up to the latest received packet from each node were received at the central node. However, this is not the case when there is process noise. Moreover, it is not clear how to modify the LKF fusion or the PEF schemes proposed by [13] and [14] when packets are lost, since these strategies rely on the assumption that all packet will be received. A naive adaptation of these schemes to include missing packets, would be the use an open loop estimate based on the last received packet, suggested by the fact that $\mathbb{E}[x_t \,|\, y_1, \ldots, y_{t-\tau}] = A^\tau \mathbb{E}[x_{t-\tau} \,|\, y_1, \ldots, y_{t-\tau}] = A^\tau \hat{x}_{t-\tau|t-\tau}$, where τ is the delay of the last packet received by the central node. However, as observed in [1], this strategy can lead to much worse performance than simple measurement fusion (MF), i.e. the strategy based on the transmission of the raw measurements.

C. Contribution

Motivated, by these considerations, in this paper we explore in more detail the problem of distributed estimation where the communication between the sensors and the estimation center is subject to packet loss. Also we assume that sensors are "smart", i.e. they can preprocess the measured data, e.g. computing local state estimates. We first show that with multiple sensors it is not possible to find a distributed estimation algorithm transmitting a packet of bounded size which provides the same performance of a CKF based on all measurements from each sensor till the last received packet.

This ideal filter is referred as infinite bandwidth filter (IBF). Based on this negative result, we propose three suboptimal strategies, the first is based on standard measurement fusion (MF), the second on the optimal (best linear) fusion of partial state estimate (OPEF), and the last on the simple sum of partial state estimates by substituting the ideal current partial state estimate with its open loop estimate if some packets are lost (OLPEF). We prove that the last two strategies can achieve the optimal performance when there is no process noise ($Q = 0$), while in the opposite regime when there is no measurement noise ($R = 0$), none of the proposed filters achieve the optimal performance, although the measurement fusion scheme seems to be very close to the optimal in numerical simulations. We also observed through numerical simulations that the approach based on optimal fusion of partial estimates (OPEF), although not optimal, provides a performance with is very close to the infinite bandwidth filter (IBF) in any noise regime and even for high packet loss rates.

II. Problem formulation

A. Modeling

We consider a discrete time linear stochastic systems observed by N sensors:

$$\begin{array}{rcl} x_{t+1} & = & Ax_t + w_t \\ y^i_t & = & C_i x_t + v^i_t, \quad i = 1, \ldots, N \end{array} \tag{1}$$

where $x \in \mathbb{R}^n$, $y_i \in \mathbb{R}^{m_i}$, w_t and v^i_t are uncorrelated, zero-mean, white Gaussian noises with covariances $\mathbb{E}[w_t w_t^T] = Q$, and $\mathbb{E}[v^i_t (v^j_t)^T] = R_{ij}$, i.e. we also allow for correlated measurement noise. More compactly, if we define the compound measurement column noise vector $v_t = (v^1_t, \ldots, v^N_t) \in \mathbb{R}^m, m = \sum_i m_i$, we have $\mathbb{E}[v_t v_s^T] = R\delta(t-s)$, where the (i,j)-th block of the matrix $R \in \mathbb{R}^{m \times m}$ is $[R]_{ij} = R_{ij} \in \mathbb{R}^{m_i \times m_j}$. The initial condition x_0 is again a zero-mean Gaussian random variable uncorrelated with the noises and covariance $\mathbb{E}[x_0 x_0^T] = P_0$. We also assume that $R > 0$, the pair $(A, Q^{1/2})$ is reachable and (A, C) is observable, where $C^T = [C_1^T \; C_2^T \; \ldots \; C_N^T]$, which are sufficient conditions for the existence of a stable estimator.

The sensors are not directly connected with each other and can send messages to a common central node through a lossy communication channel, i.e. there is a non zero probability that the message is not delivered correctly. We model the packet dropping events through a binary random variable $\gamma^i_t \in \{0, 1\}$ such that:

$$\gamma^i_t = \begin{cases} 0 & \text{if packet sent at time } t \text{ by node } i \text{ is lost} \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

Each sensor is provided with computational and memory resources to (possibly) preprocess information before sending it to the central node. More formally, at each time instant t each sensor i sends the preprocessed information $z^i_t \in \mathbb{R}^\ell$:

$$z^i_t = f^i_t(y^i_1, y^i_2, \ldots, y^i_t) = f^i_t(y^i_{1:t}) \tag{3}$$

where ℓ is bounded and $f^i_t()$ are causal functions of the local measurements. Natural choices are $z^i_t = y^i_t$, i.e. the latest measurement, or the output of a (time varying) linear filter:

$$\begin{array}{l} \xi^i_t = F^i_t \xi^i_{t-1} + G^i_t y^i_t \\ z^i_t = H^i_t \xi^i_t + D^i_t y^i_t \end{array}$$

as for example a local Kalman filter.

The objective is to find the best mean square estimate given the information available at time t at the central node. More formally, let us define the information set $\mathcal{I}_t = \bigcup_{i=1}^N \mathcal{I}^i_t$ available at the central node, where $\mathcal{I}^i_t = \{z^i_k \,|\, \gamma^i_k = 1, k = 1, \ldots, t\}$, then the best mean square estimate and its corresponding error covariance at the central node are given by $\hat{x}_{t|t} = \mathbb{E}[x_t \,|\, \mathcal{I}_t]$ and $P_{t|t} = var(x_t - \hat{x}_{t|t} \,|\, \mathcal{I}_t) = \mathbb{E}[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T \,|\, \mathcal{I}_t]$. It is evident that also the error covariance $P_{t|t}$ is random variable since it depends on the specific packet drop history represented by the random variables γ^i_t. Also, the error covariance is a function of specific preprocessing strategy defined by the functions $f^i_t()$. If we do not constrain the dimension of the messages transmitted by each node to be bounded, then an optimal strategy is to send all measurements up to that instant, i.e. $z^i_t = y^i_{1:t}$. Using this strategy the corresponding information sets available at the central node are $\overline{\mathcal{I}}^i_t = \{\varnothing\}$ if $\gamma^i_k = 0, \forall k = 1, \ldots, t$ or $\overline{\mathcal{I}}^i_t = \{y^i_{1:t-\tau^i_t}\}$, where $\tau^i_t = t - \text{argmax}\{k | \gamma^i_k = 1, = 1, \ldots, t)$ is the delay of the last packet received from node i at time t. In this idealized situation, the minimum mean square estimate (MMSE) is given by $\hat{x}^*_{t|t} = \mathbb{E}[x_t \,|\, \bigcup_i \overline{\mathcal{I}}^i_t] = \mathbb{E}[x_t \,|\, y^1_{1:t-\tau^1_t}, \ldots, y^N_{1:t-\tau^N_t}]$; we shall also call this estimator *infinite bandwidth filter* (IBF). Its error covariance $P^*_{t|t} = Var(x_t - \hat{x}^*_{t|t} \,|\, \mathcal{I}_t)$ is clearly a lower bound for any linear estimator independently of the preprocessing $f_t^i()$ performed by each node for any possible packet loss sequence, i.e.

$$P^*_{t|t} \leq P_{t|t},\ \forall f_t^i(), \forall \gamma_t^i.$$

Our objective is to find preprocessing schemes $f_t^i(y^i_{1:t})$ with bounded size output z_t^i which can achieve the lower bound on error covariance $P^*_{t|t}$. The next theorem shows that it is not possible:

Theorem 1: Let us consider the state estimate $\hat{x}_{t|t}$ and $\hat{x}^*_{t|t}$ defined as above. Then there do not exist (possibly nonlinear) functions $z_t^i = f_t^i(y^i_{1:t}) \in \mathbb{R}^\ell$ with bounded size $\ell < \infty$ such that $P^*_{t|t} = P_{t|t}$ for any possible packet loss sequence, i.e.

$$\nexists f_t^i() \,|\, P_{t|t} = P^*_{t|t},\ \forall \gamma_t^i$$

Proof: We will prove the theorem by providing an explicit example. Let us consider the following scalar dynamical systems with two sensors:

$$\begin{aligned} x_{t+1} &= x_t + w_t \\ y_t^1 &= x_t + v_t^1; \quad y_t^2 = x_t + v_t^2 \end{aligned}$$

where x_0, w_t, v_t^1, v_t^2 are uncorrelated zero-mean white random variables with covariance $\sigma_x = \sigma_w = \sigma_{v^1} = \sigma_{v^2} = 1$, respectively. We consider two different packet arrival scenarios:

$$\begin{aligned} a &: \{\gamma_2^1 = 1;\, \gamma_1^1 = \gamma_1^2 = \gamma_2^2 = 0\}, \\ b &: \{\gamma_1^1 = \gamma_1^2 = 0;\, \gamma_2^1 = \gamma_2^2 = 1\} \end{aligned}$$

i.e. at time $t = 2$ in scenario (a) only the second packet from the first sensor arrived successfully to the central node, while in scenario (b) both packets corresponding to time $t = 2$ were received but the packets corresponding to time $t = 1$ were lost. We start by showing that there do not exist *linear* functions of the measurement $z_t^i = f_t^i(y^i_{1:t}) = \sum_{k=1}^t \alpha^i_{t,k} y^i_k$ of size one, i.e. $z_t^i \in \mathbb{R}$, that can retrieve the optimal mean square estimate $\hat{x}^*_{t|t}$ for the two scenarios just illustrated. In fact, let us consider scenario (a) which leads to $\hat{x}^{*,a}_{2|2} = \mathbb{E}[x_2 \,|\, y_1^1, y_2^1] = \bar{\alpha}_1^{1,a} y_1^1 + \bar{\alpha}_2^{1,a} y_2^1$, where $\bar{\alpha}_1^{1,a} \neq \bar{\alpha}_2^{1,a} \neq 0$, and we made explicit with the superscript $()^a$ that the actual optimal mean square estimate depends on the particular packet loss sequence history. Therefore, in order to have $\hat{x}^a_{2|2} = \mathbb{E}[x_2 \,|\, z_2^1] = \beta_1^a z_2^1$ equal to $\hat{x}^{*,a}_{2|2}$, we must have $\beta_1^a[\alpha^1_{2,1}\, \alpha^1_{2,2}] = [\bar{\alpha}_1^{1,a}\, \bar{\alpha}_2^{1,a}]$. Differently, in scenario (b), the optimal mean square estimate $\hat{x}^{*,b}_{2|2} = \mathbb{E}[x_2 \,|\, y_1^1, y_2^1, y_1^2, y_2^2] = \bar{\alpha}_1^{1,b} y_1^1 + \bar{\alpha}_2^{1,b} y_2^1 + \bar{\alpha}_1^{2,b} y_1^2 + \bar{\alpha}_2^{2,b} y_2^2$, where $[\bar{\alpha}_1^{1,b}\, \bar{\alpha}_2^{1,b}] \neq \gamma\, [\bar{\alpha}_1^{1,a}\, \bar{\alpha}_2^{1,a}], \forall \gamma \in \mathbb{R}$, i.e. the two vectors of coefficients are not parallel. This implies that also $[\bar{\alpha}_1^{1,b}\, \bar{\alpha}_2^{1,b}] \neq \gamma\, [\alpha^1_{2,1}\, \alpha^1_{2,1}], \forall \gamma \in \mathbb{R}$. Therefore, since the estimate $\hat{x}^b_{2|2} = \mathbb{E}[x_2 \,|\, z_2^1, z_2^2] = \beta_1 z_2^1 + \beta_2 z_2^2$ and $z_2^1 = \gamma(\alpha^1_{2,1} y_1^1 + \alpha^1_{2,2} y_2^1)$ for some γ, it follows that $\hat{x}^b_{2|2} \neq x^{*,b}_{2|2}$. This concludes the proof that there do not exist linear functions of dimension one that allow to retrieve the optimal estimate for all possible packet loss sequences.

This results continue to hold even if we consider more general *nonlinear* functions $z_t^i = f_t^i(y^i_{1:t})$. In fact, as shown in the specific example above, in order to reconstruct the optimal estimate, $\bar{z}^a = f_2^1(y_1^1, y_2^1) = \bar{\alpha}_1^{1,a} y_1^1 + \bar{\alpha}_2^{1,a} y_2^1$ in the first scenario and $\bar{z}^b = f_2^1(y_1^1, y_2^1) = \bar{\alpha}_1^{1,b} y_1^1 + \bar{\alpha}_2^{1,b} y_2^1$ in the second, must hold. Since the two pair of coefficients are not parallel, the central node can also reconstruct y_1^1, y_2^1 from $\bar{z}^a, \bar{z}^b$. This is equivalent to saying that the function $z_2^1 = f_2^1(y_1^1, y_2^1)$ maps two real numbers into a single real number, and that the central node can reconstruct the two real numbers from the single real number z_2^1, which is clearly impossible [1].

The proof for arbitrary but finite packet size ℓ, i.e $z_t^i \in \mathbb{R}^\ell$ can be obtained similarly by properly constructing $\ell + 1$ different packet loss scenarios for which the gains of the optimal linear combination of the measurements are linearly independent, which means that there do not exists linear functions $f_t^i()$ which always recover the optimal mean square estimate $x^*_{t|t}$. Also similarly to the proof above, this can be extended to general nonlinear functions $f_t^i()$. ∎

The previous theorem states that there is no hope to find a preprocessing with bounded message size which can achieve the error covariance $P^*_{t|t}$ of the infinite bandwidth filter (IBF) since it is not possible to know in advance what the packet loss event will be. We will therefore propose two suboptimal estimation strategies which provide the optimal solution in the special case of perfect communication link, i.e. when there is no packet loss. The first, referred as measurement fusion (MF), consists in sending the raw measurements:

$$\begin{aligned} z_t^i &= y_t^i \\ \hat{x}^{MF}_{t|t} &= \mathbb{E}[x_t \,|\, \mathcal{I}_t^i,\, i = 1, \ldots, N] \\ \mathcal{I}_t^i &= \{y_k^i \,|\, \gamma_k^i = 1,\, k = 1, \ldots, t\} \end{aligned} \tag{4}$$

The second, referred as optimal estimate fusion (OEF), consists in sending filtered estimates from each sensor and then optimally combining the most recent ones from each sensor at the central node:

$$\begin{aligned} z_t^i &= \Gamma_t^i z^i_{t-1} + G_t^i y_t^i \\ \hat{x}^{OEF}_{t|t} &= \mathbb{E}[x_t \,|\, z^i_{t-\tau_t^i}, i = 1, \ldots, N] = \textstyle\sum_{i=1}^N \Phi_t^i z^i_{t-\tau_t^i} \end{aligned} \tag{5}$$

for suitable choices of the matrices Γ_t^i and G_t^i which will be discussed in the Section IV.

III. Measurement fusion

In this section we briefly summarize how to iteratively compute the estimate based on the measurement fusion strategy. Let us first define the following variables:

$$\bar{C}_t = \begin{bmatrix} \gamma_t^1 C_1 \\ \vdots \\ \gamma_t^N C_N \end{bmatrix}, \quad \bar{y}_t = \begin{bmatrix} \gamma_t^1 y_t^1 \\ \vdots \\ \gamma_t^N y_t^N \end{bmatrix},$$

$$\bar{R}_t = \begin{bmatrix} \gamma_t^1 R_{11} & \cdots & \gamma_t^1 \gamma_t^N R_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_t^N \gamma_t^1 R_{N1} & \cdots & \gamma_t^N R_{NN} \end{bmatrix}$$

which can be obtained from the centralized matrices C and R, and from the lumped column measurement vector $y_t = (y_t^1\, y_t^2\, \ldots\, y_t^N)^T$ by replacing the rows and columns

[1] Of course one could argue that in an infinite bandwidth setup there is essentially no limitation on the number ℓ in (3); however, when bandwidth limitations come into play, resolution requirements would of course impose an upper bound on ℓ. It would also be possible to consider "smart" coding schemes which, however, would have to depend also on the specific packet loss sequence.

corresponding to the lost packet with zeros. It was shown in [10] that the state estimate for the measurement fusion strategy is given by:

$$\hat{x}^{MF}_{t|t} = (I - \bar{C}_t L_t) A \hat{x}^{MF}_{t-1|t-1} + \bar{L}_t \bar{y}_t \quad (6)$$
$$P^{MF}_{t|t} = P_{t|t-1} - P_{t|t-1}\bar{C}_t^T(\bar{C}_t P_{t|t-1}\bar{C}_t^T + \bar{R}_t)^{\dagger}\bar{C}_t P_{t|t-1} \quad (7)$$
$$\bar{L}_t = P_{t|t-1}\bar{C}_t^T(\bar{C}_t P_{t|t-1}\bar{C}_t^T + \bar{R}_t)^{\dagger} \quad (8)$$
$$P_{t+1|t} = A P^{MF}_{t|t} A^T + Q \quad (9)$$

where the symbol † indicates the Moore-Penrose pseudoinverse. The previous equations correspond to a time-varying Kalman filter which depends on the packet loss sequence. Note that only measurements that have arrived are used to the computation of the estimate $\hat{x}^{MF}_{t|t}$, i.e. the dummy zero measurement in $\bar{y}_t$ are not used as if they were real measurements, but are discarded.

The measurement fusion strategy has the advantage to be computed recursively and exactly with the inversion of one matrix of (at most) the size of the lumped measurement vector y_t. On the other hand, if a packet is lost, then the information corresponding to the measurement in that packet is lost forever, while sending filtered version of the output as in the optimal estimate fusion (OEF) this information might be partially recovered. In fact, as we will see in Section V there are noise regimes, namely in the absence of process noise, in which the MF performs considerably worse than OEF.

IV. State estimate fusion

In this section we consider the second strategy mentioned above, named OEF. According to this strategy, the i-th node sends an "estimate" of the state computed via

$$z^i_t = \Gamma^i_t z^i_{t-1} + G^i_t y^i_t \quad (10)$$

and the central node has to compute the optimal fusion rule

$$\hat{x}^{OEF}_{t|t} = \mathbb{E}[x_t \,|\, z^i_{t-\tau^i_t}, i = 1, \dots, N] = \sum_{i=1}^{N} \Phi^i_t z^i_{t-\tau^i_t} \quad (11)$$

where $t - \tau^i_t$ is the last time in which the central node has received a packet from node i. The conditional expectation will be computed assuming a Gaussian measure[2].

Besides computing the coefficients Φ^i_t one has also to decide how each node processes its own measurements, i.e. how Γ^i_t and G^i_t are chosen.

Before discussing these choices, we first describe how the gains Φ^i_t can be computed. Let us define: $\Phi_t := \left[\Phi^1_t, ..., \Phi^N_t\right]$ and $z_{t,\tau} := \left[\begin{array}{ccc} \left(z^1_{t-\tau^1_t}\right)^{\top} & \dots & \left(z^N_{t-\tau^N_t}\right)^{\top} \end{array} \right]^{\top}$. Of course, the optimal fusion coefficients of Eqn. (11) can be computed as:

$$\Phi_t = \mathbb{E}\left[x_t z^T_{t,\tau}\right] \mathbb{E}\left[z_{t,\tau} z^T_{t,\tau}\right]^{-1} \quad (12)$$

We shall now outline a procedure which allows to compute the covariance matrices $\mathbb{E}\left[x_t z^T_{t,\tau}\right]$ and $\mathbb{E}\left[z_{t,\tau} z^T_{t,\tau}\right]$. To this purpose let us define the augmented state vector $s_t := \left(x_t, z^1_t, .., z^N_t\right)$. Combining equations (1) and (10) it is immediate to see that

$$s_t = \Psi_t s_{t-1} + B^w_t w_{t-1} + B^v_t v_t \quad (13)$$

where

$$\Psi_t := \left[\begin{array}{cccc} A & 0 & \dots & 0 \\ G^1_t C_1 A \Gamma^1_t & \Gamma^1_t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ G^N_t C_N A \Gamma^N_t & 0 & 0 & \Gamma^N_t \end{array} \right]$$

$$B^w_t := \left[\begin{array}{c} I \\ G^1_t C_1 \\ \vdots \\ G^M_t C_M \end{array} \right] \quad B^v_t := \left[\begin{array}{ccc} 0 & \dots & 0 \\ G^1_t & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & G^M_t \end{array} \right]$$

From this equation the covariance function $\Sigma_{h,k} := \mathbb{E}[s_h s^T_k]$ can be easily computed, starting from the initial condition

$$\Sigma_{0,0} := \left[\begin{array}{cccc} \mathbb{E}[x_0 x^T_0] & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{array} \right].$$

Observe now that all the elements of $\mathbb{E}\left[x_t z^T_{t,\tau}\right]$ and $\mathbb{E}\left[z_{t,\tau} z^T_{t,\tau}\right]$ are indeed elements of $\Sigma_{h,k}$ for suitable values of h and k.

It is also convenient to note that also the conditional variance of $\hat{x}^{OEF}_{t|t}$ given the sequence $\{\gamma^i_s\}_{s=1,..,t}$ can be computed using the standard formula for the error covariance

$$Var\{\hat{x}^{OEF}_{t|t} | \gamma^i_s, \ s \le t\} = Var\{x_t\} - \Phi_t \mathbb{E}\left[z_{t,\tau} z^T_{t,\tau}\right] \Phi^T_t \quad (14)$$

This equation will be useful in evaluating the performance of different choices of the local pre-processing strategies Γ^i_t and G^i_t. Of course it can also be used to monitor on-line the performance of the estimator $\hat{x}^{OEF}_{t|t}$.

Note that the error covariance of OEF, $P^{OEF}_{t|t}$ that uses only the latest packet received from each sensor node, is larger than the one that could be obtained from all received packets, $P_{t|t}$, at the price of a higher computational cost, i.e.:

$$P^*_{t|t} \le P_{t|t} \le P^{OEF}_{t|t} \ \forall \gamma^i_t.$$

The optimal choice of the "local" filter matrices Γ^i_t and G^i_t in Eqn. (10) is far from being a trivial task even if topology and statistics of the model are completely known. Therefore, in order to gain some further intuition, we explore and compare some sensible choices of the matrices Γ^i_t and G^i_t.

A. Optimal Partial Estimate Fusion

This strategy is suggested by the observation that, in the absence of packet losses, one could compute the gains in a centralized manner and distribute the computations to each sensor. To be more precise, assume all measurements were available to a common location, i.e. that there where no packet losses. We shall denote with $x^{CKF}_{t|t} := \mathbb{E}[x_t | y^i_{1:t}, i = 1, .., N]$ the centralized Kalman filter (CKF); its evolution is governed by the equations:

$$\begin{array}{rcl} \hat{x}^{CKF}_{t|t} & = & F_t \hat{x}^{CKF}_{t-1|t-1} + L_t y_t \\ F_t & = & (I - L_t C) A \end{array} \quad (15)$$

[2] Alternatively one could think of $\mathbb{E}[\cdot \mid \cdot]$ as being the best *linear* estimator.

where the gain $L_t = [L_t^1\, L_t^2 \cdots L_t^N]$ is the centralized Kalman filter gain computed as

$$\begin{aligned} P_{t+1} &= (A - K_tC)P_t(A - K_tC)^T + K_tRK_t^T + Q \\ L_t &= P_tC^T(CP_tC^T + R)^{-1} \\ K_t &= AL_t \end{aligned}$$

Note now that, defining z_t^i to be the solution of

$$z_t^i = F_t z_{t-1}^i + L_t^i y_t^i, \qquad (16)$$

the CKF estimate $\hat{x}_{t|t}^{CKF}$ is given by $\hat{x}_{t|t}^{CKF} = \sum_{i=1}^N z_t^i$. For these reason we shall call the z_t^i's "partial estimates". This strategy has been suggested in [14] for distributed estimation to the purpose of reducing the power consumption. Note that Eqn. (16) falls in the class Eqn. (10) with $\Gamma_t^i := F_t$ and $G_t^i := L_t^i$.

In the presence of packet losses, only $z_{t-\tau_t^i}^i$ are available to the central node and, with this information, the best (linear) estimate is given by

$$\hat{x}_{t|t}^{OPEF} = \mathbb{E}[x_t \mid z_{t-\tau_t^i}^i, i = 1, \ldots, N] = \sum_{i=1}^N \Phi_t^i z_{t-\tau_t^i}^i \qquad (17)$$

where the superscript ^{OPEF} stands for optimal partial estimate fusion and the coefficients Φ_t^i are computed as described in the previous section.

B. Optimal Kalman Filter fusion

Note that, in the previous strategy, the local filter at each node depends upon all the other sensors; this is only reasonable either if the network topology is fixed or if the central node can communicate to each sensor the new filter parameters if the network changes.

Alternatively each sensor could compute the best estimate given its own measurements, which is a local in nature, .i.e.

$$\begin{aligned} \hat{z}_t^{i,l} &= F_t^i \hat{z}_{t-1}^{i,l} + L_t^{i,l} y_t^i \\ F_t^i &= (I - L_t^{i,l} C_i)A \end{aligned}$$

where the gains[3] $L_t^{i,l}$ are the local Kalman filter gains computed as

$$\begin{aligned} P_{t+1}^i &= (A - K_t^{i,l}C_i)P_t^i(A - K_t^{i,l}C_i)^T + K_t^{i,l}R_{ii}\left(K_t^{i,l}\right)^T + Q \\ L_t^{i,l} &= P_t^iC_i^T(C_iP_t^iC_i^T + R_{ii})^{-1} \\ K_t^{i,l} &= AL_t^{i,l} \end{aligned}$$

We shall call the optimal estimate based on the $z_{t-\tau_t^i}^{i,l}$'s optimal Kalman estimate fusion (OKEF):

$$\hat{x}_{t|t}^{OKEF} = \mathbb{E}[x_t \mid z_{t-\tau_t^i}^{i,l}, i = 1, \ldots, N] = \sum_{i=1}^N \Phi_t^{i,l} z_{t-\tau_t^i}^{i,l} \qquad (18)$$

Unfortunately, as discussed in [13], even in the absence of packet losses the optimal estimate cannot be recovered as a linear function of the z_t^i's.

[3]The superscript ^{i,l} reminds that $z_t^{i,l}$ is the local estimate of the state at the $i-$th sensor, where the gain $L^{i,l}$ is computed using the *local* Kalman filter equations.

C. Open Loop Partial Estimate Fusion

The third strategy, referred to as open loop partial estimate fusion (OLPEF), aims at simplifying the optimal partial estimate fusion; in fact the preprocessing of the measurement is the same, i.e. z_t^i are computed as in the OPEF strategy, but it does not compute the optimal linear combination of the estimates at the central node.

$$\begin{aligned} z_t^i &= F_t z_{t-1}^i + L_t^i y_t^i \\ F_t^i &= (I - L_tC)A \\ \hat{x}_{t|t}^{OLPEF} &= \textstyle\sum_{i=1}^N A^{\tau_t^i} z_{t-\tau_t^i}^i \end{aligned} \qquad (19)$$

The rationale behind this strategy is that, since in the absence of packet losses $\hat{x}_{t|t} = \sum_i z_t^i$, when z_t^i is not available one could compute an estimate by propagating (in "open loop") the last partial estimate $z_{t-\tau_t^i}^i$ using the approximation $z_t^i \simeq A^{\tau_t^i} z_{t-\tau_t^i}^i$. Note that $P_{t|t}^* \leq P_{t|t} \leq P_{t|t}^{OPEF} \leq P_{t|t}^{OLPEF}$, $\forall \gamma_t^i$ where the last inequality follows from the fact that last messages are not fused optimally in the OLPEF strategy.

Remark 1: If sensor nodes either appear or disappear OLPEF would most probably fail. Differently it is to be expected that both OPEF and OKEF will be able to compensate for this changes providing sensible estimates since the weights Φ_t^i and $\Phi_t^{i,l}$ in Eqn. (17) and Eqn. (18) are chosen adaptively based on the received packets.

V. Special cases

A. Small process noise regime (Q=0)

An important regime is when the state evolution can be described by a deterministic linear map, i.e. when the process noise is very small. We shall study the limiting case $Q = 0$, i.e. no process noise. We shall also restrict our attention to the case in which the measurement noises are uncorrelated, i.e. $R = \text{block diag}\{R_1, .., R_N\}$.

Proposition 1: Let us consider the proposed estimation schemes, namely MF and OPEF, OLPEF, OKEF and IBF for $Q = 0$ and $R = \text{block diag}\{R_1, .., R_N\}$. Then

$$P_{t|t}^* = P_{t|t}^{OPEF} = P_{t|t}^{OKEF} = P_{t|t}^{OLPEF} < P_{t|t}^{MF}$$

Proof: See [6] ∎

B. Small measurement noise regime (R=0)

Another important regime to be considered is when the measurement noise R is much smaller as compared to the process noise Q. This is a regime for which only recent measurements convey relevant information. One might wonder whether one of the proposed fusion schemes, namely the MF and the OPEF, can always provide the best achievable estimate $\hat{x}_{t|t}^*$, or, at least, if one is always better then the other. The next proposition shows that the answer to both questions is negative.

Proposition 2: Let us consider the two proposed estimation schemes, namely MF and OPEF, for $R = 0$ and $Q > 0$. Then there exist scenarios for which $P_{t|t}^{MF} > P_{t|t}^{OPEF}$ and scenarios for which $P_{t|t}^{MF} < P_{t|t}^{OPEF}$

Proof: See [6] ∎

The proposition shows how in general none of two strategies MF and OPEF is superior to the other also in the limiting regime $R = 0$. As a consequence, it also shows that none of them always achieves the optimal filter performance $x_{t|t}^*$.

VI. Simulations

In order to illustrate and compare the methodologies described above, we consider the following simulation example. The measurement vector y_t, of dimension 7, is generated by (1) with

$$A = \begin{bmatrix} 0.99 & 1 \\ 0 & 0.99 \end{bmatrix} \quad C = \begin{bmatrix} 2 & .4 & 1 & 1 & 0.4 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T.$$

The noises v_t and w_t are uncorrelated, zero mean Gaussian white noises with covariances, respectively,

$$\begin{aligned} \mathbb{E}[v_t v_t^T] &= R = \text{diag}\{10,\, 20,\, 40,\, 0.5,\, 2,\, 1,\, 40\} \\ \mathbb{E}[w_t w_t^T] &= \mu_Q Q = \mu_Q \text{diag}\{0.001,\, 0.001\} \end{aligned}.$$

The parameter μ_Q will be varied to study the behavior under different regimes, i.e. different ratios between the model and the measurements noises.

Figure 1 reports the error variance of the first component of the state as a function of μ_Q, for the packet drop probability $\mathbb{P}[\gamma_t^i = 0] = 0.5$. Note that the conditional variance given the packet drop sequence $\{\gamma_t^i\}$ has been computed in closed form as discussed in Section IV for all methods except OLPEF. The unconditional variance is obtained simulating a sufficiently long sequence of packet drop sequence and averaging the conditional variance over that sequence. The same could also have been done for the OLPEF; however this is rather involved from the computational point of view and hence the variance for OLPEF has been computed purely by Monte Carlo simulations.

For small values of μ_Q the OLPEF behaves very similarly to OPEF. This is reasonable since, for small process noise it make sense to "trust" the model and hence propagate estimates in open loop. Note also that MF is the worst strategy for small μ_Q; this is also in line with the results in Section V predicting that OPEF is better than MF for $Q = 0$.

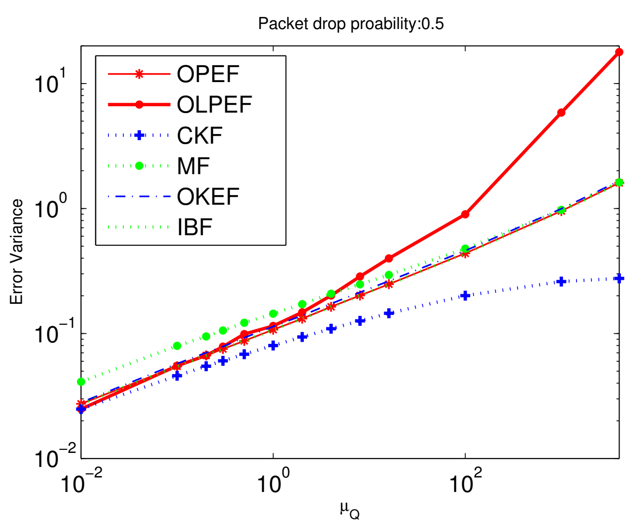

Fig. 1. Error Variance vs. μ_Q. The curve relative to OPEF coincide, to any practical purpose, with that corresponding to IBF.

VII. Conclusions

In this paper we explored the problem of distributed estimation subject to random packet loss between the sensors and the central location where the best state estimate is required. We showed, that differently from classical setup with perfect communication, random packet dropping destroys the possibility for the sensors to optimally design their preprocessing scheme since they cannot predict in advance the loss sequence. Nonetheless, we have observed through numerical simulations that optimally fusing partial estimates from each sensor provides a performance that is very close to the ideal performance. This opens up a number of future research directions. The first is to provide some upper bound on the performance of the OPEF strategy and show that it is not too far from the ideal performance of the IBF. Another relevant area of research is to provide numerically efficient algorithms to compute the OPEF. In fact, it requires the inversion of large size matrices which might be too computationally demanding, therefore approximation schemes for OPEF are ought. Finally, it is not clear how the OPEF scheme can be extended to rooted tree networks, i.e. sensors cannot send packets directly to the central node, but they have to route them though other sensors as it typically happens in Wireless Sensor Networks.

References

[1] A. Agnoli, A. Chiuso, P. D'Errico, A. Pegoraro, and L. Schenato, "Sensor fusion and estimation strategies for data traffic reduction in rooted wireless sensor networks," in *Proc. of International Symposium on Communications, Control and Signal Processing (ISCCSP 2008)*, La Valletta, Malta, 2008.

[2] B. Anderson and J. Moore, *Optimal Filtering*. Prentice Hall, 1979.

[3] Y. Bar-Shalom, H. Chen, and M. Mallick, "One-step solution for the general out-of-sequence measurement problem in tracking," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 40, no. 1, pp. 27–37, 2004.

[4] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.

[5] N. Carlson, "Federated square root filter for decentralized parallel processors," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 26, no. 3, pp. 517–525, 1990.

[6] A. Chiuso and L. Schenato, "Information fusion strategies from distributed filters in packet-drop networks," University of Padova, Tech. Rep., 2008, available at `http://paduaresearch.cab.unipd.it/1138/`.

[7] V. Gupta, D. Spanos, B. Hassibi, and R. M. Murray, "Optimal LQG control across a packet-dropping link," *Systems and Control Letters*, vol. 56, no. 6, pp. 439–446, 2007.

[8] J. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE, special issue*, vol. 95, no. 1, pp. 138–162, January 2007.

[9] B. Levy, D. Castañon, G. Verghese, and A. Willsky, "A scattering framework for decentralized estimation problems," *Automatica*, vol. 19, no. 4, pp. 373–384, 1983.

[10] L. Schenato, "Optimal sensor fusion for distributed sensors subject to random delay and packet loss," in *to appear in Proc. of IEEE Conf. on Decision and Control*, New Orleans, U.S.A., 2007.

[11] L. Shi, K. Johansson, and R. Murray, "Estimation over wireless sensor networks: Tradeoff between communication, computation and estimation qualities," in *Proc. of IFAC World Congress*, Seoul, Korea, 2008.

[12] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, September 2004.

[13] A. Willsky, M. Bello, D. Castanon, B. Levy, and G. Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," *IEEE Transactions on Automatic Control*, vol. 27, no. 4, pp. 799–813, 1982.

[14] J. Wolfe and J. Speyer, "A low-power filtering scheme for distributed sensor networks," in *Proceeding of IEEE Conference on Decision and Control (CDC'03)*, 2003, pp. 6325–6326.

[15] K.-S. Zhang, X. R. Li, and Y.-M. Zhu, "Optimal update with out-of-sequence observations for distributed filtering," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 1992–2004, 2005.