A Censoring Strategy for Decentralized Estimation in Energy-Constrained Adaptive Diffusion Networks

Rocío Arroyo-Valles, Sina Maleki, Geert Leus
Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology
Email: {m.d.r.arroyovalles, s.maleki, g.j.t.leus}@tudelft.nl

***Abstract*—This paper presents a censoring strategy for distributed estimation over adaptive networks in scenarios where energy resources are limited. Sensors apply selective communication policies in order to save energy for being able to transmit more important information later. Simulation results show an enhancement in network lifetime, by reducing communication processes among nodes with a slightly degraded result, compared with energy unconstrained schemes.**

***Index Terms*—Wireless sensor networks, energy efficiency, decision rules, adaptive networks, distributed estimation.**

I. Introduction

Adaptive Networks (ANs) consist of a collection of spatially distributed nodes that are connected through a particular topology. These networks allow solving estimation problems in a distributed manner, by exploiting cooperation among neighboring nodes that can only access local information. Thus, each node in the AN should rely on local information captured and sent by its neighbors. Several rules to combine and assign weights to the neighboring local information have been proposed in the literature [1], [2], [3], but many of them ignore the different noise profiles that nodes may have across the network because they operate under different Signal-to-Noise (SNR) ratio conditions. The work in [4] proposes a rule to assign weights based on the estimation of the noise variance in real time.

When ANs are implemented on Wireless Sensor Networks (WSNs), interactions between nodes, and in general all tasks carried out by a sensor node, have a cost in terms of energy consumption. But resources are not infinite. In fact, energy is the most critical constraint in WSNs. For this reason, research efforts in WSNs are aimed at reducing the energy consumption due to communication processes.The question is which information should be exchanged among neighboring nodes and how often it should be exchanged in order to obtain a good estimate while keeping interaction among nodes to a minimum. Selective transmission (censoring) strategies allow nodes to make autonomous decisions about transmitting or not transmitting the available data to neighboring nodes, according to the relevance of the information and the energy cost [5]. This way, the network lifetime will be prolonged, thereby also guaranteeing a good overall performance. Therefore, an important and key aspect is how sensors determine the relevance of the information they should transmit.

In this work, the problem of distributed estimation over adaptive networks is considered when energy constraints are taken into account. Starting from the Adapt-then-Combine (ATC) diffusion algorithm proposed in [6] and applying the combining weight rule included in [4], nodes apply selective communication policies when exchanging their local estimate information with their neighbors in order to save energy for being able to transmit more relevant information later. Up to now, the energy concern in this scenario has only been taken into account in probabilistic diffusion networks, which were initially proposed to cover the dynamic topology case given that links and nodes may be subject to failures [7]. In the probabilistic diffusion approach, a node only communicates with a randomly selected subset of its direct neighbors. This way, the traffic of the network is reduced. In [8], a link probability control strategy for probabilistic diffusion networks with communication resource constraints is proposed. It is shown that communication processes are reduced and the system performance is improved compared to algorithms that control links with static probabilities. Further, a data selection method formulated as a censoring problem for parameter estimation in WSNs was explored in [9], where sensors have to send their measurements to a fusion center. However, all these works neither include explicitly the cost of the different actions taken by them nor take into account how important the information coming from each node is. In this paper, as in a censoring scheme, the decision whether to share (or not) the local information is made comparing a function of the data with a threshold, which is computed as a solution of a Markov Decision Process (MDP). Furthermore, unlike the probabilistic diffusion approaches, the instantaneous neighborhood of our approach is determined by the local decisions made by nodes and not selected by each node randomly with a certain probability. This way, at every time instant, only the most useful information is exchanged. Results will show an enhancement in network lifetime at the cost of a slightly degraded performance.

The rest of the paper is structured as follows. Section II describes the adaptive diffusion strategy and Section III presents the sensor model and the optimal selective transmitter. Section IV describes the general procedure to apply selective communication strategies to adaptive networks while Section V depicts the simulation experiments and results. Conclusions and future lines are drawn in Section VI.

The work was partially supported by projects PRI-PIBIN-2011-1266 and TEC2011-22480, the Spanish Ministry of Education, under the National Program of Human Resources Mobility from the I+D+i 2008-2011 National plan, and the NWO-STW under the VICI program (project 10382).

978-1-4673-5577-3/13/$31.00 ©2013 IEEE

II. Adaptive diffusion strategies

Consider a network of N nodes with a predefined network topology. Two nodes are neighbors if they can share information, and denote by $\mathcal{N}_k$ the neighborhood of node k including k itself. In adaptive diffusion, the neighborhood $\mathcal{N}_k$ remains fixed when nodes are static. At each time $i \geq 0$, each node k has access to a scalar measurement $z_k(i) \in \mathbb{C}$ and a regression vector $\mathbf{u}_k(i) \in \mathbb{C}^{M\times 1}$ of length M, both arising from realizations of zero-mean random processes $\tilde{z}_k$ and $\tilde{\mathbf{u}}_k$, respectively, which are related via $z_k(i) = \mathbf{u}_k^T(i)\mathbf{w}^o + v_k(i)$, where $\mathbf{w}^o \in C^{M\times 1}$ is the unknown vector we wish to estimate and $v_k(i) \in \mathbb{C}$ accounts for noise, which is assumed to be zero-mean with variance $\sigma_{v,k}^2$, and it is independent of the other variables. The objective of the distributed estimation problem is to generate an estimate of $\mathbf{w}^o$ in a distributed manner by solving the following optimization problem

$$\min_{\mathbf{w}} \sum_{k=1}^{N} \mathbb{E}\left|z_k(i) - \mathbf{u}_k^T(i)\mathbf{w}\right|^2. \tag{1}$$

Several diffusion adaptation schemes for solving (1) have been developed (see [1] and references therein), and one of them is the Adapt-then-Combine (ATC) diffusion algorithm [6]. It consists of two steps. The first one involves local adaptation, where each node k updates its local estimate $\boldsymbol{\varphi}_k(i)$ using the combined estimate from the previous iteration, $\mathbf{w}_k(i-1)$, and its own data, $\{z_k(i); \mathbf{u}_k(i)\}$:

$$\boldsymbol{\varphi}_k(i) = \mathbf{w}_k(i-1) + \eta_k(i)\mathbf{u}_k^*(i)\left[z_k(i) - \mathbf{u}_k^T(i)\mathbf{w}_k(i-1)\right], \tag{2}$$

where the normalized least-mean squares (NLMS) algorithm has been considered, and $\eta_k(i) = \frac{\hat{\eta}_k}{\delta + \|\mathbf{u}(i)\|^2}$, with $\hat{\eta}_k$ a positive step size, δ is a regularization factor (a small positive constant), * denotes complex conjugate and $\|\cdot\|$ the Euclidean distance.

The second step is a combination stage where the intermediate estimates $\{\boldsymbol{\varphi}_l(i)\}$ from the neighborhood of node k ($l \in \mathcal{N}_k$) are combined through the coefficients $\{a_{l,k}\}$ to obtain the updated weight estimate $\mathbf{w}_k(i)$:

$$\mathbf{w}_k(i) = \sum_{l\in\mathcal{N}_k} a_{l,k}\boldsymbol{\varphi}_l(i), \tag{3}$$

where the combining weights $\{a_{l,k}\}$, collected into the $N \times N$ matrix $\mathbf{A}$ as $[\mathbf{A}]_{l,k} = a_{l,k}$ should satisfy: $\mathbf{A1} = \mathbf{1}$ and $a_{l,k} = 0$ if $l \notin \mathcal{N}_k$, where $\mathbf{1}$ stands for a column vector with all elements equal to one.

To select the combining weights, several approaches have been proposed [1]. However, most of the rules ignore the noise profiles across the network and since some nodes may be noisier than others, it is not enough to rely on a sensor connectivity to assign weights to neighboring estimates. In [4], after solving an optimization problem to find the optimal combining weights, an algorithm is proposed which includes the noise variance, here labeled as the Adaptive Combining Weights (ACW) algorithm. More specifically, node k combines the intermediate estimates $\{\boldsymbol{\varphi}_l(i)\}$ from its neighbors in an inversely proportional manner to the noise variance. Defining $\sigma_{l,k}^2(i)$ as a scaled estimate of the noise variance $\sigma_{v,l}^2$ for node k at time i we can compute a set of time-dependent weights $\{a_{l,k}(i)\}$ as

$$a_{l,k}(i) = \begin{cases} \frac{\sigma_{l,k}^{-2}(i)}{\sum_{j\in\mathcal{N}_k} \sigma_{j,k}^{-2}(i)} & \text{if } l \in \mathcal{N}_k \\ 0, \ \text{otherwise} \end{cases} \tag{4}$$

where $\sigma_{l,k}^2(i)$ is updated following

$$\sigma_{l,k}^2(i) = (1-\nu_k)\sigma_{l,k}^2(i-1) + \nu_k \left\|\boldsymbol{\varphi}_l(i) - \mathbf{w}_k(i-1)\right\|^2 \tag{5}$$

and ν_k is a positive step size. Thus, nodes with smaller variances will be given larger weights.

But since adaptive diffusion strategies are implemented over WSNs, which are energy-constrained, communications should be kept to a minimum in order not to waste energy and to last longer, without decreasing significantly the performance. It is not enough to compute optimally the combining weights because the number of packet exchanges is not reduced. Hence, a way of reducing communications among nodes is applying selective communication (censoring) strategies. Next, we describe the sensor model and the selective transmission policy before explaining how it will be applied to adaptive diffusion networks.

III. Selective communication policy

A. Sensor model

1) State vector: Following the analysis carried out in [10], the node state will be characterized by two variables: the energy at a given node at time i, $e(i)$, and the importance (relevance) of the message to be sent at time i, $y(i)$. The node state vector is defined as $\mathbf{s}(i) = [e(i), y(i)]^T$.

2) Actions: At time i, the sensor must make a decision $d(i)$ about sending or not sending the current information message. We take $d(i) = 1$ if the message is transmitted, while $d(i) = 0$ if discarded. Thus, the decision rule is a function of the state vector; i.e., $d(i) = d(\mathbf{s}(i)) = d(e(i), y(i))$.

3) State dynamics: Sensors deplete their batteries according to the taken actions. The available energy after time i can be expressed recursively as

$$e(i+1) = e(i) - d(i)c_1(i) - (1-d(i))c_0(i), \tag{6}$$

where $c_1(i)$ represents the energy consumption due to a transmission, and $c_0(i)$ is the energy consumed when the message is discarded. The latter may include the cost of sensing data, the cost of data reception or the cost of idle states. Parameter $c_1(i)$ accounts for all the previous costs together with the cost of transmitting the message. We assume that the energy consumption may have some random components, so that $c_1(i)$ and $c_0(i)$ can be viewed as stochastic processes.

4) Rewards: With $u(\cdot)$ standing for the Heaviside step function (with the convention $u(0) = 1$), the *reward* at time i for a node that decides to transmit a message will be

$$r(i) = y(i)u(e(i) - c_1(i)). \tag{7}$$

It means that whenever a node transmits a message, the reward it obtains is the message importance.

Defining the total reward up to time i as $t(i) = \sum_{p=0}^{i} d(p)r(p)$, decisions will be made in order to maximize the total expected reward (i.e., maximize (on average) the importance sum of all transmitted messages), defined as $\mathbb{E}\{t_\infty\} = \mathbb{E}\{\lim_{i\to\infty} t(i)\}$.

B. Optimal selective transmission

Provided that messages are properly quantified, nodes could eventually discard low graded messages with the expectation of transmitting more important upcoming messages. Therefore, from the perspective of the network efficiency, a selective transmitting policy at each node may serve to optimize energy resources to transmit only the most relevant information.

Considering the fact that the sensor model described above has the structure of a Markov Decision Process (MDP) [11], optimal sequences of decision rules in the form $d(i) = d(e(i), y(i))$ and maximizing $\mathbb{E}\{t_\infty\}$ are derived in [5]. It turns out that, under some stationarity conditions and for large $e(i)$, the optimal decision rules only depend on $y(i)$ and can be expressed in the form

$$d(i) = u\left(\frac{y(i)}{\Delta(y(i))} - \tau\right), \tag{8}$$

where $\Delta(y(i)) = \mathbb{E}\{c_1(i) - c_0(i)|y(i)\}$ and τ is the solution of

$$\mathbb{E}\{c_0(i)\}\tau = \mathbb{E}\{(y(i) - \Delta(y(i))\tau)^+\}, \tag{9}$$

with $(h)^+ = hu(h)$, for any h. The stationarity conditions mentioned above state that $c_1(i)$ and $c_0(i)$ are stationary stochastic processes.

1) Estimating asymptotic thresholds: In general, the value of τ satisfying (9) cannot be analytically computed, but it can be estimated iteratively using [10]:

$$\tau(i) = (1 - 1/i)\,\tau(i-1) + (y(i) - \Delta(y(i))\tau(i-1))^+/i\epsilon_0, \tag{10}$$

where $\epsilon_0 = \mathbb{E}\{c_0(i)\}$ can be estimated by sample averaging. We assume that $\Delta(y(i))$ does not depend on $y(i)$ (because the energy consumption is independent of the importance values). Therefore, we consider $\Delta(y(i)) = \Delta$ which can be estimated by sample averaging, too.

Considering $\mu(i) = \Delta\tau(i)$ and a deterministic energy model (where E_R and E_T are the energy spent on receiving or sensing, and transmitting states, respectively, so that $c_1(i) = E_T + E_R$ and $c_0(i) = E_R$), the optimal forwarding threshold is computed as

$$\mu(i) = \left(1 - \frac{1}{i}\right)\mu(i-1) + \frac{\rho}{i}\cdot(y(i) - \mu(i-1))^+, \tag{11}$$

where $\rho = \frac{E_T}{E_R}$. Note that we do not consider the energy consumption due to idle states because sensors always receive data at every time instant i.

IV. Applying selective transmission policies to adaptive diffusion networks

The selective transmission policy can be integrated into the ATC algorithm in the combination step in order to reduce the number of communication processes while it is assured that the most relevant information is exchanged. Every node is able to make a decision whether to transmit its local estimate to its neighbors (which can then be used during the combination step to obtain a better estimate of the parameter of interest), or to censor it because it is similar to the one transmitted previously, irrelevant or noisier and less reliable than the others due to a low SNR. Fig. 1 shows schematically the cooperation strategies followed by the selective ATC algorithm.

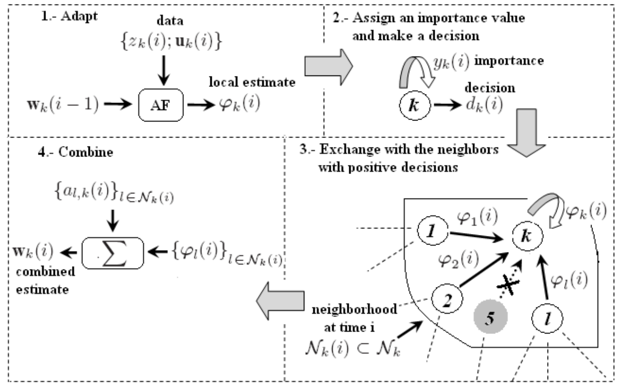

Fig. 1. Illustration of the adaptive ATC strategy combined with selective transmitters; AF stands for adaptive filter.

Note that when censoring is applied, the neighborhood of node k is not static anymore but dynamic (it may change from one time instant to the following) due to the decisions made by its neighbors. Therefore, the instantaneous neighborhood $\mathcal{N}_k(i)$ is a subset of the neighborhood established according to the network topology, i.e., $\mathcal{N}_k(i) \subset \mathcal{N}_k$. Also, nodes can receive local estimates from its neighbors even if they do not send. It means that the graph is directed so if node n does not send to node k ($n \notin \mathcal{N}_k(i)$), k may still send updates to n.

In order to apply selective policies, it is crucial that nodes assign an importance value to the local estimates (information to be transmitted). Moreover, the importance value is the key component of the reward in (7). It should reflect the contribution of a node's local estimate to improve its neighboring estimates in order to obtain a better global estimate of $\mathbf{w}^o$. This way, the extra estimation error due to the fewer communication processes among sensors is minimized.

Let us denote $\boldsymbol{\varphi}_k(i)$ as the local estimate from node k at time i, $\mathbf{w}_k(i-1)$ as the combined estimate at node k at time $i-1$ and $a_{kk}(i-1)$ as the weight that node k assigns to its own estimate at time $i-1$, which is inversely related to the estimated noise variance. Every node then assigns a local importance given by

$$y_k(i) = \|\boldsymbol{\varphi}_k(i) - \mathbf{w}_k(i-1)\|\, a_{kk}(i-1). \tag{12}$$

Nodes quantify how different the current local estimate and the previous combined estimate are and adjust it with their combination weight. Thus, if both estimates are similar or the estimated noise is high, nodes will assign a low importance based on the intuition that censoring estimates that are highly deviated from the true value increases the estimation error more than censoring similar estimates.

V. Simulations

In this section, we assess the performance of the proposed algorithm. Several approaches have been compared, including the non-cooperative one (sensors do not exchange information with their neighbors, i.e., $\mathbf{w}_k(i) = \boldsymbol{\varphi}_k(i)$), the ACW algorithm [4], the ACW algorithm with an initial percentage of sensors in sleep mode (ACW-Sleeping), the ACW algorithm where nodes reduce the updating frequency and behave as the non-cooperative one during the remaining time (ACW-Slow), the

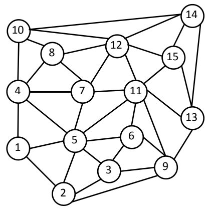

Fig. 2. Network topology

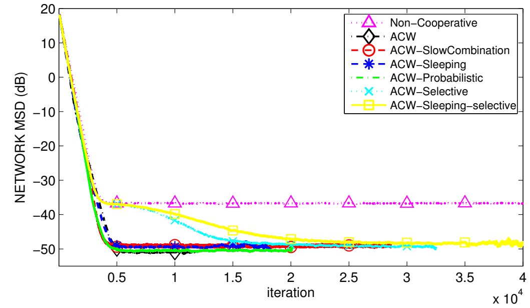

Fig. 3. Learning curve of the network MSD for the different algorithms.

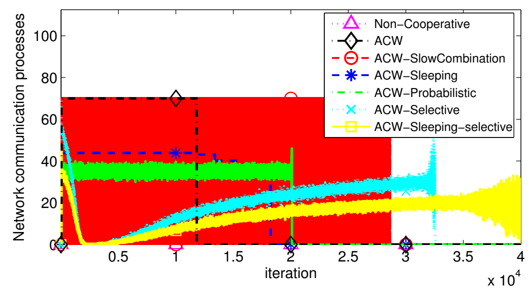

Fig. 4. Evolution of the number of communication processes.

ACW algorithm combined with selective transmission (ACW-Selective), which is explained in Section IV, and the ACW-Selective algorithm combined with sleeping nodes (ACW-Sleeping-Selective). Furthermore, results are also compared with the ACW probabilistic diffusion algorithm, which decides with a certain probability which links from its neighborhood are active and disabled at every time instant. In this case, the instant neighborhood may also change.

The metric that is used to compare the performance of the algorithms is the network mean square deviation (MSD). Defining the weight error vector as $\tilde{\mathbf{w}}_k(i) = \mathbf{w}^o - \mathbf{w}_k(i)$, and $\text{MSD}_k(i) = \|\tilde{\mathbf{w}}_k(i)\|^2$ [1], the network MSD error is defined as $\text{MSD}(i) = \frac{1}{N}\sum_{k=1}^{N}\text{MSD}_k(i)$, with steady-state value obtained when $i \to \infty$.

The regressors $\mathbf{u}_k(i)$ follow a multidimensional zero-mean Gaussian distribution with covariance matrix equal to the identity matrix. The observation noise follows a Gaussian distribution with zero mean and variance selected so that the SNR at each node is taken from a uniform distribution between $[-20, -30]$ dB. The regressors and the observation noise are independent. Besides, the observations and regression vectors are the same for all the algorithms, as well as the initialization parameter values.

The network consists of 15 nodes with a topology shown in Fig. 2, which is also used in [8]. The vector to estimate, $\mathbf{w}^o$, follows a uniform distribution between -1 and 1, with $M = 50$. The maximum number of iterations is $T_{max} = 40000$, the initial energy of the sensors is 200000, the transmitting energy consumption is $E_T = 1$, $E_R = 2$ is the energy for reception and sensing and finally $\hat{\eta}_k = 0.1$ and $\nu_k = 0.2$ for all nodes [4]. The percentage of nodes that apply a sleeping policy is 25% (for the sleeping approaches) and the probability that a link between two neighbors is active during a time instant is 50% (for the probabilistic approach). The combination interval for the ACW-Slow algorithm is 3, i.e., sensors exchange information with their neighbors every 3 iterations. The network is dead when the first node runs out of battery. All results are averaged over 50 simulation runs. Note that we show average results; however, each simulation has a different duration. Hence, we average according to the number of alive networks at every time instant.

Fig. 3 shows the network MSD. We can observe that the ACW algorithm achieves the lowest network MSD; however, the batteries are becoming empty very fast due to the huge number of communication processes with the neighboring nodes to carry out the combination step. On the other hand, the network lasts much longer using a selective transmission (see also Table I for the specific values) while the estimation error does not increase excessively. Furthermore, the ACW-Sleeping-Selective algorithm also achieves a reasonable performance, mainly in the stationary case, because the penalty it pays is only a slight degradation in performance. However, these algorithms have a slower convergence after the first time instants due to the censoring of transmissions.

Method	Lifetime (iterations)
Non-Cooperative	40000
ACW	11813
ACW-Slow	28619
ACW-Sleeping - 25 %	13795
ACW-Probabilistic	20029
ACW-Selective	31738
ACW-Sleeping-Selective	35236

TABLE I
AVERAGE NETWORK LIFETIME FOR ALL THE ALGORITHMS.

Besides, we can see that the ACW-Selective and the ACW-Sleeping-Selective algorithms both follow the non-cooperative one during some time instants because most of the sensors decide not to share the local estimates with their neighbors, and therefore, every sensor uses only its own information for computing the estimate. Fig. 4, which shows the number of communication processes in the network, illustrates this behavior as the number of exchanges approximates 0. But it is constant for the ACW algorithm and oscillates between 0 and the previous constant amount for the ACW-Slow.

Fig. 5 depicts the evolution of the remaining energy along time. For the sake of clarity, we keep the average energy constant once the network lifetime expires. It is worth mentioning that the methods which apply a mechanism for reducing the number of communications have a smaller slope than the ACW. Among them, the algorithms that apply selective

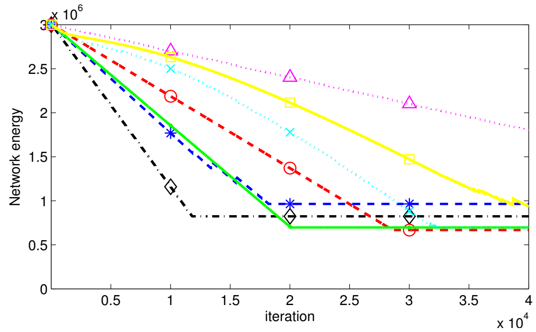

Fig. 5. Evolution of the remaining energy in the network for the different algorithms. The legend is the same as in Fig. 3.

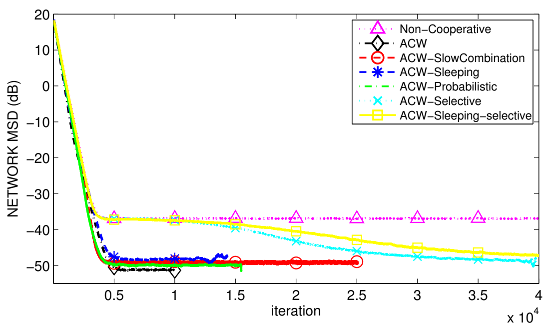

Fig. 6. Learning curve of the network MSD for $E_T = 4$.

transmission are the best ones in terms of balancing energy. The discontinuities that appear in the ACW-Sleeping algorithm are due to averaging. The sensors that are put into sleep mode differ from one simulation to another. Due to this, there are simulations that last longer, because the number of communication processes is smaller (see also the discontinuities in Fig. 4), while the MSD error increases slightly.

The network MSD evolution for a higher transmission energy, $E_T = 4$, is depicted in Fig. 6. The greater E_T, with respect to E_R, the larger the gain in terms of network lifetime for those algorithms that apply selective transmission schemes, but the slower the convergence. As the energy consumed for transmitting a local estimate is higher, the sensors should make the right decision about which estimate to transmit in order to maximize the reward. This way, the number of censored transmissions increases.

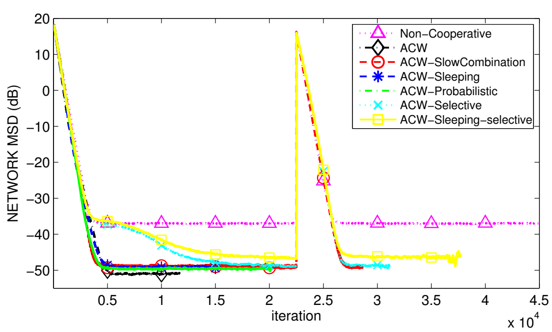

Fig. 7. Learning curve of the network MSD when an abrupt change is introduced in the middle of the iterations.

Next, let us consider the network topology of Fig. 2 when $T_{max} = 45000$. An abrupt change is introduced in the middle of the maximum number of iterations, so we can analyze the ability to reconverge. The network MSD evolution is shown in Fig. 7. It can be seen that some algorithms, such as the ACW, ACW-Probabilistic or the ACW-Sleeping, do not have enough energy to detect the change of the parameter vector to estimate, and the ACW-Slow depletes them earlier than those that are selective. The reconvergence of the selective algorithms is much faster after the abrupt change, because nodes have learnt from the past.

VI. Conclusions and Future Work

A selective communication policy for parameter estimation in energy-constrained adaptive networks was explored in this paper. Nodes should make decisions whether to transmit their local estimates to neighboring nodes or to censor them in order to save energy, while assuring that the most relevant information is exchanged. The energy consumption should be taken into account in adaptive diffusion scenarios in order not to miss abrupt changes of the parameter vector to estimate. Simulation results showed that the proposed selective algorithm achieved error values that are comparable with those that did not censor information, while increasing the network lifetime.

The presented results are based on heuristics. Future work includes the search for the optimal number of nodes to put into sleep without degrading the network performance. Also, a faster convergence of the selective schemes is of importance.

References

[1] A.H. Sayed, *Diffusion Adaptation over Networks*. In *E-Reference Signal Processing*, R. Chellapa and S. Theodoridis, editors, Elsevier, 2013. Also available as arXiv:1205.4220v1, May 2012.

[2] J. Fernandez-Bes, J.A. Azpicueta-Ruiz, M.T.M. Silva, and J. Arenas-Garcia, "A novel scheme for diffusion networks with least-squares adaptive combiners," in *IEEE Int'l Workshop on Machine Learning for Signal Processing (MLSP'12)*, sept. 2012, pp. 1 –6.

[3] N. Takahashi, I. Yamada, and A.H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Trans. on Signal Processing*, vol. 58, no. 9, pp. 4795 –4810, sept. 2010.

[4] Sheng-Yuan Tu, and A.H. Sayed, "Optimal Combination Rules for Adaptation and Learning over Networks," in *4th IEEE Int'l Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'11)*, dec. 2011, pp. 317 –320.

[5] R. Arroyo-Valles, A. G. Marques, and J. Cid-Sueiro, "Optimal Selective Transmission under Energy Constraints in Sensor Networks," *IEEE Trans. on Mobile Computing*, vol. 8, no. 11, pp. 1524 – 1538, 2009.

[6] F.S. Cattivelli, and A.H. Sayed, "Diffusion LMS Strategies for Distributed Estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1035 –1048, march 2010.

[7] C.G. Lopes, and A.H. Sayed, "Diffusion adaptive networks with changing topologies," in *IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP'08)*, april 2008, pp. 3285 –3288.

[8] N. Takahashi and I. Yamada, "Link probability control for probabilistic diffusion least-mean squares over resource-constrained networks," in *IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP'10)*, march 2010, pp. 3518 –3521.

[9] E.J. Msechu, and G.B. Giannakis, "Distributed measurement censoring for estimation with wireless sensor networks," in *IEEE 12th Int'l Workshop on Signal Processing Advances in Wireless Comm. (SPAWC'11)*, june 2011, pp. 176 –180.

[10] R. Arroyo-Valles, A. G. Marques, and J. Cid-Sueiro, "Optimal Selective Forwarding for Energy Saving in Wireless Sensor Networks," *IEEE Trans. on Wireless Comm.*, vol. 10, no. 1, pp. 164–175, 2011.

[11] D.J. White, *Markov Decision Processes*. Wiley-Interscience, 1993.