A NEW BROADCAST BASED DISTRIBUTED AVERAGING ALGORITHM OVER WIRELESS SENSOR NETWORKS

Franck Iutzeler† *Philippe Ciblat*† *Walid Hachem*‡ *Jérémie Jakubowicz*†

† Institut Télécom ; Télécom ParisTech ; CNRS LTCI
‡ CNRS LTCI ; Télécom ParisTech

ABSTRACT

The distributed estimation of the average value of the sensors initial measures is one of the most popular issue in the Wireless Sensor Networks (WSN) area. In WSNs, broadcasting data seems natural to exchange information quickly because of the broadcast nature of the Wireless channel. Nevertheless, although broadcast-based algorithms converge faster than pairwise algorithms, the obtained consensus is not necessarily the true average. By the means of additional side-information exchange, we propose a broadcast-based algorithm converging rapidly to the true average. The convergence of this new algorithm is established and its convergence speed is exhibited. We remark that the proposed algorithm outperforms the existing ones.

***Index Terms*—** distributed estimation, averaging, sensor network, broadcast, consensus

1. INTRODUCTION

Distributed algorithms over Wireless Sensors Networks (WSN) have been widely studied since the pioneer work in [1]; in particular, a lot of results have been obtained for the problem of averaging [2, 3]. However, only a few averaging algorithms take benefit of the broadcast nature of the wireless communication channels [4, 5]. In [4], at each clock tick, one (randomly chosen) sensor broadcasts its information to all its neighbors, then each neighbor averages its own value with the received one. With such an algorithm, the network's global sum is not preserved. This implies that the corresponding update matrix is not doubly-stochastic, and so preventing the algorithm to converge to the true average. Recently, to overcome this drawback, [5] has proposed a new broadcast-based algorithm relying on the transmission of two variables (instead of one) at each clock tick. Nevertheless any convergence speed analysis is provided.

In the literature, some algorithms have efficiently overcome the non doubly-stochasticity of the update matrix by introducing the principle of the *weighted* gossip [6, 7]. In such a scheme, the sensor exchange two variables: the first one represents the sum of the received information while the second one represents the importance level of the received information. In [6], such a *weighted gossip* principle is applied to a wired synchronous network without feedback. The absence of feedback leads to non doubly-stochastic update matrix. In [7], this principle is applied to wireless asynchronous network without feedback: actually, the (randomly chosen) sensor sends its variables to one (and only one) neighbor which does not send back its own variables.

In this paper, we thus propose to build an algorithm (called *Broadcast based Weighted Gossip -BWGossip-*) relying on the *weighted gossip* principle and taking benefit of the broadcast nature of the channel. Our proposed algorithm gathers the respective benefit of the broadcast approach (fast convergence speed) and of the weighted gossip approach (the true consensus). The main contribution of the paper is twofold: the algorithm design and its theoretical performance analysis.

The paper is organized as follows: in Section 2, we introduce our broadcast-based weighted gossip algorithm. In Section 3, we prove that the proposed algorithm converges to the true average. In Section 4, we prove that the square error is upper-bounded by an exponentially decreasing function with high probability. In Section 5, we provide an heuristic improvement to our algorithm by modifying the sensor clocks in a distributive manner without any additional cost. Our results are numerically illustrated in Section 6. Finally Section 7 is devoted to concluding remarks.

This work was partially funded by the French Defense Agency (DGA) and the Télécom/Eurécom Carnot Institute.

2. PROPOSED ALGORITHM

2.1. Signal Model

Let us consider a N-sensors network modeled by an unweighted undirected graph $\mathcal{G} = (V, E)$ where V is the set of vertices/sensors ($|V| = N$) and E is the set of edges/perfect links between the sensors. We assume $\mathcal{G}$ is connected. Each sensor i may exchange data with its neighborhood $\mathcal{N}_i = \{j \in V | (i, j) \in E\}$. Let $d_i = |\mathcal{N}_i|$ denote the degree of the sensor i. We also define $\mathbf{A}$ the so-called adjacency matrix of the graph, $\mathbf{D} = \mathrm{diag}(d_1, \cdots, d_N)$ the degree matrix and the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ [8].

Each sensor i has its own independent Poisson clock of parameter λ_i. At first, we will consider that all λ_i are identical and equal to λ, which is equivalent to a global clock of parameter $N\lambda$ and uniform selection of the awaking sensor. We will note t the instant of the t-th tick of the global clock. At $t = 0$, the sensor i only knows its individual measure $x_i(0)$. Let $x_{ave} = 1/N \sum_{i=1}^{N} x_i(0)$ be the average value of the initial measures. At time t and sensor i, the estimate average value is denoted by $x_i(t)$. The purpose of an averaging algorithm is that $x_i(t)$ goes to x_{ave} when t goes to infinity for each sensor i.

2.2. Broadcast based Weighted Gossip algorithm

Like [6, 7], the sensor i will update two local values $s_i(t)$ and $w_i(t)$ (at time t) whereas, in standard gossip algorithm, the sensor i updates directly $x_i(t)$. More precisely, $s_i(t)$ and $w_i(t)$ represent the *sum* of the received information and its *weight* related to how much information is passed through respectively. In the sequel, we denote $\mathbf{s}(t) = [s_1(t), \cdots, s_N(t)]^T$, $\mathbf{w}(t) = [w_1(t), \cdots, w_N(t)]^T$, and $\mathbf{x}(t) = [x_1(t), \cdots, x_N(t)]^T$.

The proposed algorithm is initialized as follows

- $\mathbf{s}(0) = \mathbf{x}(0)$
- $\mathbf{w}(0) = \mathbf{1}$

with $\mathbf{1}$ the column vector composed by N ones.

At time t, the vector of average estimates is obtained by $\mathbf{x}(t) = \mathbf{s}(t)/\mathbf{w}(t)$ where the division is done element-wise, and where $\mathbf{s}(t)$ and $\mathbf{w}(t)$ are updated as follows :
assuming that, at time t, the sensor i wakes up

- ► Sensor i broadcasts $\left(\frac{s_i(t)}{|\mathcal{N}_i|+1}; \frac{w_i(t)}{|\mathcal{N}_i|+1}\right)$
- ► At sensors in the neighborhood $\mathcal{N}_i$, we have: $\left\{\begin{array}{l} s_j(t+1) = s_j(t) + \frac{s_i(t)}{|\mathcal{N}_i|+1} \\ w_j(t+1) = w_j(t) + \frac{w_i(t)}{|\mathcal{N}_i|+1} \end{array}\right. , \forall j \in \mathcal{N}_i.$
- ► At sensor i, we have : $\left\{\begin{array}{l} s_i(t+1) = \frac{s_i(t)}{|\mathcal{N}_i|+1} \\ w_i(t+1) = \frac{w_i(t)}{|\mathcal{N}_i|+1} \end{array}\right.$
- ► All other sensors stay idle.

Using the matrix formalism, the proposed algorithm can be rewritten as follows

$$\left\{\begin{array}{l} \mathbf{s}^T(t) = \mathbf{s}^T(t-1)\mathbf{K}(t) = \mathbf{x}^T(0)\mathbf{P}(t) \\ \mathbf{w}^T(t) = \mathbf{w}^T(t-1)\mathbf{K}(t) = \mathbf{1}^T\mathbf{P}(t) \end{array}\right. \quad (1)$$

where $\mathbf{P}(t) = \mathbf{K}(1)\mathbf{K}(2)\ldots\mathbf{K}(t)$, $\mathbf{K}(t)$ is equal to $\mathbf{K}_i$ if the sensor i is active at time t, and

$$\mathbf{K}_i = \mathbf{I} - e_i e_i^T (\mathbf{I}+\mathbf{D})^{-1}\mathbf{L} \quad (2)$$

with e_i the i-th canonical vector. Notice that, albeit the matrix formalism is identical to [6, 7], the algorithms are different since the matrices $\mathbf{K}_i$ are different.

One can easily check that $\mathbf{K}(t)$ is row-stochastic (*i.e.*, $\mathbf{K}(t)\mathbf{1} = \mathbf{1}$) which leads to the following *mass-conservation* property

$$\left\{\begin{array}{l} \sum_{i=1}^N s_i(t) = \sum_{i=1}^N x_i(0) = N x_{ave} \\ \sum_{i=1}^N w_i(t) = N. \end{array}\right. \quad (3)$$

3. CONVERGENCE

One can straightforwardly check that the set of matrices $\{\mathbf{K}(t)\}_{t>0}$ satisfy the following properties.

P1) These matrices are (row) stochastic non-negative matrices with positive diagonals.

P2) The sequence if these matrices is i.i.d. [1].

We also have

P3) $\mathbb{E}[\mathbf{K}]$ is a primitive matrix.

To prove the previous property, we firstly lower-bound $\mathbb{E}[\mathbf{K}]$ as follows

$$\begin{aligned} \mathbb{E}[\mathbf{K}] &= \frac{1}{N}\sum_{i=1}^N \mathbf{I} - e_i e_i^T + e_i e_i^T \left[(\mathbf{I}+\mathbf{D})^{-1}(\mathbf{A}+\mathbf{I})\right] \\ &\geq \frac{N-1}{N}\mathbf{I} + \frac{1}{(d_{max}+1)N}(\mathbf{A}+\mathbf{I}) \geq 0 \end{aligned}$$

where $\geq$ stands for the element-wise inequality and d_{max} denotes the maximum degree of all the vertices. Since $\mathbf{A}$ is the adjacency matrix of a connected graph, $\exists m > 0, (\mathbf{I}+\mathbf{A})^m > 0$. Hence, for the same m, $\mathbb{E}[\mathbf{K}]^m \geq 1/(d_{max}N + N)^m(\mathbf{I}+\mathbf{A})^m > 0$, which implies that $\mathbb{E}[\mathbf{K}]$ is a primitive matrix.

In [7] (Theorem 4.1), it is proven that any weighted gossip algorithm such that $\mathbf{P1}$, $\mathbf{P2}$, and $\mathbf{P3}$ hold for $\mathbf{K}(t)$ converges to the true average. Therefore our proposed algorithm converges to x_{ave} as t goes to infinity.

[1] because at each global time t, a sensor (hence a matrix) is chosen uniformly as they have independent Poisson clocks with the same parameter λ.

4. CONVERGENCE SPEED

In this section, we will put the main contributions of the paper corresponding to the analysis of the Square Error (SE) of the proposed algorithm. We will prove that the SE is upper-bounded by an exponentially decreasing function with high probability. The convergence rate of this function is also exhibited.

First of all, one can easily remark that

$$\begin{aligned} |x_i(t) - x_{ave}|^2 &= \frac{|s_i(t) - x_{ave}w_i(t)|^2}{w_i(t)^2} \\ &= \frac{\left|\sum_{j=1}^N x_j(0)\left(\mathbf{P}_{ji}(t) - \frac{1}{N}\sum_{l=1}^N \mathbf{P}_{li}(t)\right)\right|^2}{w_i(t)^2}. \end{aligned}$$

By lower bounding $w_i(t)$ with its minimum and using Cauchy-Schwartz inequality, we obtain that

$$\begin{aligned} \mathrm{SE}(t) = \|\mathbf{x}(t) - x_{ave}\mathbf{1}\|_2^2 &= \sum_{i=1}^N |x_i(t) - x_{ave}|^2 \\ &\leq \Psi_1(t)\Psi_2(t) \end{aligned} \quad (4)$$

$$\text{where} \quad \Psi_1(t) = \frac{\|\mathbf{x}(0)\|_2^2}{|\min_k w_k(t)|^2}$$

$$\Psi_2(t) = \sum_{i=1}^N\sum_{j=1}^N \left|\left(\mathbf{P}^T(t)(\mathbf{I}-\mathbf{J})\right)_{ij}\right|^2$$

with $\mathbf{J} = (1/N)\mathbf{1}\mathbf{1}^T$.

In the sequel, we will prove, on the one hand, that $\Psi_1(t)$ is bounded with high probability and, on the other hand, that $\mathbb{E}[\Psi_2(t)]$ goes exponentially to zero when the number of iterations goes to infinity.

We prove the following theorem meaning that it is unlikely $\Psi_1(t)$ becomes very large, so a sensor talks too much compared to the other ones.

Theorem 1.

$$\Psi_1(t) = \mathcal{O}_P(1)$$

where $X_n = \mathcal{O}_P(Y_n)$ *stands for* $\forall\delta > 0$, $\exists C_\delta$ *such that* $\forall n$, $\mathbb{P}\{|X_n| \geq C_\delta |Y_n|\} < \delta$.

Proof. As in [6], in order to lower bound $\min_i w_i(t)$, we consider a time t_0 and a node n_0 whose weight is greater than 1 (there is obviously one because of the mass conservation exhibited in Eq. (3)). We know from [9] that the expectation of the diffusion time t_d (that is the time for any node to disseminate its information to the whole network) while broadcasting is $\mathbb{E}[t_d] \leq \Delta N + N(\Delta-1)\ln((N-1)/(\Delta-1)) = t_{max}$ with Δ the diameter of the graph. Hence, by Markov's inequality we know that the diffusion time is bounded in probability which means that at time $t = t_0 + t_d$, all the sensors will be informed with a small portion of the weight of n_0 which is greater than $\alpha = (d_{max}+1)^{-t_d} > 0$ because at each iteration the weights can be at most divided by $d_{max}+1$. Finally, let us remark that at $t=0$, all the sensors have weight 1 hence the established relation is true for all t. So, for all $t>0$, all weights will be greater than $\alpha > 0$ with high probability so $\Psi_1(t)$ is bounded with high probability. □

Our objective now is to find the behavior of $\Psi_2(t)$ when t goes to infinity. Actually, we will prove that $\Psi_2(t)$ is upper-bounded by an exponentially decreasing function with high probability. To do that, let us focus on the analysis of $\Phi_2(t) = \mathbb{E}[\Psi_2(t)]$.

Let us introduce

$$\Xi(t) = (\mathbf{I}-\mathbf{J})\,\mathbf{P}(t) \otimes (\mathbf{I}-\mathbf{J})\,\mathbf{P}(t) \tag{5}$$

where $\otimes$ stands for the Kronecker product. Since $\Psi_2(t)$ can be rewritten as $\|\mathbf{P}^T(t)(\mathbf{I}-\mathbf{J})\|_F^2$ with $\|.\|_F$ denoting the Froebenius norm, $\Psi_2(t)$ is the sum of the $((\mathbf{P}^T(t)\,(\mathbf{I}-\mathbf{J}))_{ij})^2$. These elementary terms are coefficients of the matrix $\Xi(t)$. Consequently, if $\mathbb{E}[\Xi(t)]$ vanishes exponentially to zero, $\Phi_2(t)$ also does at least at the same speed. Therefore, we will focus on $\mathbb{E}[\Xi(t)]$.

Using basic properties of the Kronecker product and operating the mathematical expectation given the natural filtration of the past events $\mathcal{F}_{t-1}$ enables us to obtain that

$$\begin{aligned} \Xi(t) &= \Xi(t-1).\,(\mathbf{K}(t)\otimes\mathbf{K}(t)) \\ \text{and}\quad \mathbb{E}\left[\Xi(t)|\mathcal{F}_{t-1}\right] &= \Xi(t-1).\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]. \end{aligned}$$

Then, remarking that $\Xi(t)\tilde{\mathbf{1}} = 0$ with $\tilde{\mathbf{1}} = \mathbf{1}\otimes\mathbf{1}$ leads to

$$\begin{aligned} \mathbb{E}\left[\Xi(t)|\mathcal{F}_{t-1}\right] &= \Xi(t-1).\left(\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right] - \tilde{\mathbf{1}}\mathbf{v}^T\right) \\ \text{and then}\quad \mathbb{E}\left[\Xi(t)\right] &= \Xi(0).\left(\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right] - \tilde{\mathbf{1}}\mathbf{v}^T\right)^t \end{aligned} \tag{6}$$

for any vector $\mathbf{v}$ and with $\Xi(0) = (\mathbf{I}-\mathbf{J})\otimes(\mathbf{I}-\mathbf{J})$. This enables us to prove the following result.

Lemma 1. *If there is a vector* $\mathbf{v}$ *such that* $\rho\left(\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right] - \tilde{\mathbf{1}}\mathbf{v}^T\right) < 1$*, then* $\mathbb{E}\left[\Xi(t)\right]$ *converges to zero as* t *goes to infinity.*

Proof. For all matrix norms, we can apply the submultiplicative inequality on Eq. (6) and follow the proof of Theorem 5.6.12 in [10] to obtain the result. □

By remarking $(\mathbf{I}-\mathbf{J})\,\mathbf{P}(t) = (\mathbf{I}-\mathbf{J})\,\mathbf{P}(t)\,(\mathbf{I}-\mathbf{J})$, Eq. (5) leads to the following result

$$\mathbb{E}\left[\Xi(t)\right] = \left(\left[(\mathbf{I}-\mathbf{J})\otimes(\mathbf{I}-\mathbf{J}\right].\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]\right)^t. \tag{7}$$

Lemma 2. $\mathbb{E}\left[\Xi(t)\right]$ *converges to zero as* t *goes to infinity if and only if* $\rho\left(((\mathbf{I}-\mathbf{J})\otimes(\mathbf{I}-\mathbf{J}))\,\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]\right) < 1$.

Proof. Given Eq. (7), $\mathbb{E}\left[\Xi(t)\right]$ can be written as $\mathbf{M}^k$ where $\mathbf{M}$ is an $N\times N$ real matrix. Then, using directly Theorem 5.6.12 in [10] leads to the result. □

The above lemmas enable us to see that the convergence of $\mathbb{E}\left[\Xi(t)\right]$ is closely related to the spectrum of $\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]$.

Lemma 3. *If* $\mathbf{K}$ *is as in Eq. (2), then it exists a vector* $\mathbf{v}$ *such that* $\rho\left(\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right] - \tilde{\mathbf{1}}\mathbf{v}^T\right) < 1$.

Proof. By construction, $\mathbb{E}[\mathbf{K}\otimes\mathbf{K}]$ is a non-negative matrix. It is also a primitive matrix. Indeed, $(\mathbb{E}[\mathbf{K}\otimes\mathbf{K}])^N \geq (\prod_{i=1}^N \mathbf{K}_i)\otimes(\prod_{i=1}^N \mathbf{K}_i) \geq 0$. Let us remark that $\prod_{i=1}^N \mathbf{K}_i \geq (1/(d_{max}+1))^N[\mathbf{I}+\mathbf{A}] \geq 0$. As $\mathbf{A}$ is the adjacency matrix of a connected graph, we know that it is irreducible so $\exists m' \in \mathbb{N}, m' < N-1 : (\mathbf{I}+\mathbf{A})^{m'} > 0$. So, by taking $m = Nm'$, $(\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right])^m > 0$ which means that $\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]$ is primitive.
As $\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]$ is a (row)-stochastic non-negative matrix, its spectral radius is 1 (see Lemma 8.1.21 in [10]). Moreover, it is easy to see that 1 is an eigenvalue associated with the eigenvector $\tilde{\mathbf{1}}$ and by the Peron-Froebenius theorem, we know that this eigenvalue has multiplicity 1. So, as this matrix is primitive, 1 is the unique eigenvalue of maximal modulus and its eigenspace is spanned by $\tilde{\mathbf{1}}$.
By using the Jordan normal form and the simple multiplicity of 1, we know that i) it exists a vector $\mathbf{v}_1$ equal to the left eigenvector corresponding to the eigenvalue 1, and ii) that the eigenvalues of $\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right] - \tilde{\mathbf{1}}\mathbf{v}_1^T$ are exactly the eigenvalues of $\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]$ except for the eigenvalue 1 which is now 0. As a consequence, the modulus of the eigenvalues of $\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right] - \tilde{\mathbf{1}}\mathbf{v}_1^T$ is strictly lower than 1. □

Putting Lemmas 1, 2 and 3 together, we get :

$$\rho\left(((\mathbf{I}-\mathbf{J})\otimes(\mathbf{I}-\mathbf{J}))\,\mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]\right) < 1. \tag{8}$$

We are now able to find an upper bound for $\Phi_2(t)$ decreasing exponentially to zero.

Theorem 2. *There is a constant* $C > 0$ *such that* $\forall\epsilon > 0$

$$\forall t > 0, \quad \Phi_2(t) \leq C\left(\Gamma+\epsilon\right)^t$$

with $\Gamma = \rho\left(((\mathbf{I}-\mathbf{J})\otimes(\mathbf{I}-\mathbf{J}))\,.\mathbb{E}[\mathbf{K}\otimes\mathbf{K}]\right)$.

Proof. From Eq. (7), and by using Lemma 5.6.13 in [10] and the matrix norm submultiplicativity, we obtain that there exists a constant $C' > 0$ such that $\forall t > 0, \forall(i,j) \in \{1,\cdots,N\}^2$,

$$\left(\mathbb{E}\left[\Xi(t)\right]\right)_{ij} \leq C'\left(\rho\left(((\mathbf{I}-\mathbf{J})\otimes(\mathbf{I}-\mathbf{J}))\,.\mathbb{E}[\mathbf{K}\otimes\mathbf{K}]\right)\right)^t.$$

As $\Phi_2(t)$ corresponds to a sum of N^2 elements of $\mathbb{E}\left[\Xi(t)\right]$, we have

$$\Phi_2(t) \leq N^2 C'\left(\rho\left(((\mathbf{I}-\mathbf{J})\otimes(\mathbf{I}-\mathbf{J}))\,.\mathbb{E}[\mathbf{K}\otimes\mathbf{K}]\right)+\epsilon\right)^t.$$

which concludes the proof. □

By using Markov's inequality on Theorem 2, we directly obtain the following theorem.

Theorem 3. *For any* $\epsilon > 0$*, we have*

$$\Psi_2(t) = \mathcal{O}_P\left((\Gamma+\epsilon)^t\right).$$

In Theorem 3, one can choose ϵ as small as possible[2]. Thus, as $\Gamma < 1$ (see Eq. (8)), $\Psi_2(t)$ vanishes exponentially with high probability. Combining Eq. (4), Eq. (8), Theorem 1, Theorem 3 and an Union's bound leads to the main result of this paper.

Theorem 4. *There exists* $0 < \Gamma < 1$ *such that* $\forall\epsilon > 0$

$$\mathrm{SE}(t) = \mathcal{O}_P\left((\Gamma+\epsilon)^t\right).$$

Roughly speaking (*i.e.*, by neglecting ϵ), one can write that $\mathrm{SE}(t) \preceq \exp\{-|\log(\Gamma)|t\}$ where $a \preceq b$ stands for "a is less or equal to a term proportional to b with high probability". The term $|\log(\Gamma)|$ corresponds to the convergence slope. Concerning our BWGossip algorithm, we have thus exhibit a lower-bound of its convergence slope.

[2] but one cannot choose $\epsilon = 0$ because even if $|||\mathbf{M}^t|||$ behave like $\rho(\mathbf{M})^t$ asymptotically for any norm, it is not necessary true for its coefficients (see p.299 in [10]).

5. AN INTUITIVE IMPROVEMENT: CLOCK CONTROL

So far, all the Poisson coefficients of the clocks were equals. This mean that all sensors were waking up uniformly and independently from their past actions. Intuitively, it would be more logical that a sensor *talking* a lot became less active during a long period.
Thanks to our BWGossip algorithm, each sensor knows whether it talks frequently or not (without additional cost) through its own weight value. Indeed, the more a sensor *talks*, the smaller its weight is. Therefore, our idea is to control the Poisson coefficient of each sensor with respect to their weight. We thus propose to consider the following rule for each Poisson coefficient

$$\lambda_i(t) = \alpha + (1-\alpha)w_i(t)$$

where $\alpha \in (0,1)$ is a tuning coefficient. Notice that the global clock remains unchanged since $\forall t > 0,\ \sum_{i=1}^{N} \lambda_i(t) = N$. The network does not so communicate more, but the talking sensors are just better chosen. The complexity of the algorithm is the same because the sensor whose weight changes has just to relaunch its Poisson clock. Even if the convergence and the convergence speed of the BWGossip with clock improvement have not been formally established, our simulations (see Fig. 1) show that it also converges exponentially to the average with higher speed if α is well chosen.

6. SIMULATIONS

In Figure 1, we plot the normalized mean square error for various averaging algorithms versus the number of clock ticks when 100 sensors are selected in a Random Geographic Graph [11] with a radius $r = \sqrt{4\log(N)/N}$. As already remarked, the Broadcast Gossip [4] does not converge to the average but decreases rapidly for the first iterations. The algorithm introduced by [5] has quite poor performance compared to the Random Gossip [2]. The BWGossip is clearly the fastest one, especially, when clock control management operates with appropriate α. In Figure 2, we plot the theoretical upper-bound of the convergence slope $|\log(\Gamma)|$ derived in Theorem 2 and the convergence slope obtained by linear regression on the logarithm of the empirical mean squared error (in Fig. 1(a), the BWGossip MSE (in log scale) is almost linear for t large enough suggesting the exponential decreasing of the MSE) versus the number of sensors N. We observe a very good agreement.

7. CONCLUSION

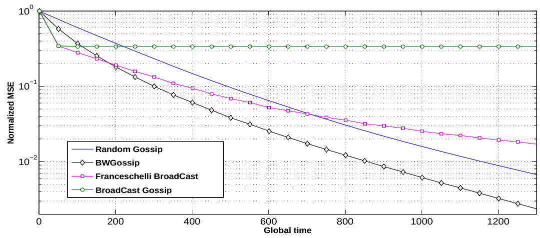
We provided a new averaging algorithm over Wireless Sensor Networks combining the speed of the broadcast-based algorithm and the convergence of the pairwise-based algorithm. We especially gave a good approximation of the convergence speed.

8. REFERENCES

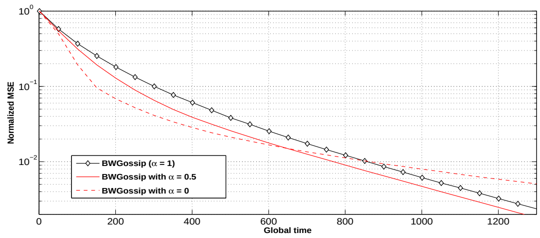
[1] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, M. I. T., Dept. of Electrical Engineering and Computer Science, 1984.

[2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[3] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic Gossip: Efficient Averaging for Sensor Networks," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1205–1216, 2008.



(a) Comparison with other averaging algorithms



(b) Effects of clock control

Fig. 1. Performance of the *BWGossip* algorithm.

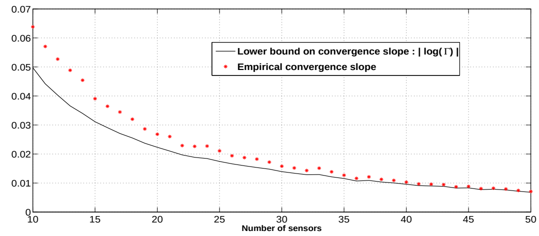


Fig. 2. Convergence slope of the *BWGossip* algorithm.

[4] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast Gossip Algorithms for Consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, 2009.

[5] M. Franceschelli, A. Giua, and C. Seatzu, "Distributed Averaging in Sensor Networks Based on Broadcast Gossip Algorithms," *IEEE Sensors J.*, vol. 11, no. 3, pp. 808–817, 2011.

[6] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. 44th Annual IEEE Symp. Foundations of Computer Science*, 2003, pp. 482–491.

[7] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted Gossip: Distributed Averaging using non-doubly stochastic matrices," in *Proc. (ISIT) Symp. IEEE Int Information Theory*, 2010, pp. 1753–1757.

[8] N. Biggs, *Algebraic graph theory*, Cambridge University Press, 1993.

[9] F. Iutzeler, J. Jakubowicz, W. Hachem, and P. Ciblat, "Distributed estimation of the maximum value over a Wireless Sensor Network," accepted to the 45-th Asilomar Conference on Signals, Systems, and Computer, available at `http://perso.telecom-paristech.fr/~ciblat/`.

[10] R.A. Horn and C.R. Johnson, *Matrix analysis*, Cambridge University Press, 2005.

[11] M. Penrose, *Random geometric graphs*, Oxford University Press, 2003.