DIFFUSION ADAPTIVE NETWORKS WITH CHANGING TOPOLOGIES

Cassio G. Lopes and *Ali H. Sayed*

Department of Electrical Engineering
University of California
Los Angeles, CA, 90095.
Email: {cassio, sayed@ee.ucla.edu}

ABSTRACT

Adaptive networks (AN) have been recently proposed to address distributed estimation problems [1]–[4]. Here we extend prior work to changing topologies and data-normalized algorithms. The resulting framework may also treat signals with general distributions, rather than Gaussian, provided that certain data statistical moments are known. A byproduct of this formulation is a probabilistic diffusion adaptive network: a simpler yet robust variant of the standard diffusion algorithm [2].

***Index Terms*—** Adaptive filters, distributed estimation, adaptive networks, diffusion, cooperative systems, distributed adaptive filters

1. INTRODUCTION

In this work we extend the concept of diffusion adaptive networks [1, 2] to the dynamic topology case, where nodes and links may be subject to failure. There are several potential applications that may benefit from these adaptive structures [5]. The proposed framework allows for data-normalized updates and leads, among other possibilities, to diffusion normalized LMS (dNLMS) and diffusion affine projection (dAPA) algorithms. The analysis below applies to signals with arbitrary distributions provided that the data statistical moments are known. We derive models for the transient and steady-state behavior of the diffusion algorithm. The results establish the interesting observation that a probabilistic diffusion protocol enables an adaptive network to limit communication among nodes to a small fraction of its full version without presenting a significant decrease in performance, as illustrated by simulations and theory.

2. DATA-NORMALIZED DIFFUSION ADAPTIVE NETWORKS

We want to estimate an $M \times 1$ unknown vector w^o from measurements collected at N nodes in a network. Each node k has access to time realizations $\{d_k(i), u_{k,i}\}$, $k = 1, \ldots, N$, of zero-mean random data $\{\boldsymbol{d}_k, \boldsymbol{u}_k\}$, with $d_k(i)$ a scalar measurement and $u_{k,i}$ a regression row vector; both at time i. The measurements are assumed to obey the linear model [6]:

$$\boldsymbol{d}_k = \boldsymbol{u}_k w^o + \boldsymbol{v}_k \tag{1}$$

where $\boldsymbol{v}_k$ is background noise, assumed independent over time and space and with variance $\sigma^2_{v,k}$. In our notation, all vectors are column vectors with the exception of the regressors $\boldsymbol{u}_k$, which are chosen as row vectors for convenience of exposition. We also denote random quantities by boldface letters (such as $\{\boldsymbol{d}_k, \boldsymbol{u}_k\}$) and use normal font to refer to their realizations (such as $\{d_k(i), u_{k,i}\}$).

To estimate w^o, we resort to adaptive networks [2, 4]. An adaptive network results from equipping the nodes of the network with local learning rules or adaptive filters. The available communication topology is then exploited to implement a cooperation protocol among the nodes in order to efficiently exploit spatial and temporal information. Different learning rules allied with different cooperation protocols give rise to different adaptive networks. Figure 1 presents an adaptive network operation under a diffusion protocol, where each node k has a neighborhood $\mathcal{N}_{k,i}$ at time i, defined as the set of nodes linked to k (including k itself). Each node $\ell \in \mathcal{N}_{k,i}$ has an estimate $\psi_\ell^{(i-1)}$ of w^o. The diffusion adaptive scheme works as follows. First, at node k, an aggregate estimate $\phi_k^{(i-1)}$ is generated by linearly combining the neighbors' estimates in order to exploit spatial diversity and, subsequently, the local estimate is updated from $\psi_k^{(i-1)}$ to $\psi_k^{(i)}$, i.e.,

$$\begin{aligned} \phi_k^{(i-1)} &= \sum_{\ell \in \mathcal{N}_{k,i}} c_{k\ell}(i)\, \psi_\ell^{(i-1)} \\ \psi_k^{(i)} &= \phi_k^{(i-1)} + H_k u_{k,i}^* \big(d_k(i) - u_{k,i}\phi_k^{(i-1)}\big) \end{aligned} \tag{2}$$

where the coefficients $\{c_{kl}\}$ denote a set of *local* combiners satisfying $\sum_\ell c_{k\ell}(i) = 1$. Moreover, H_k is a matrix that is generally dependent on the local neighborhood regression data. For example, choosing $H_k = \mu_k I_M$ leads to the standard dLMS algorithm [2], while choosing

$$H_k = \frac{\mu_k}{\epsilon + \|u_{k,i}\|^2}\, I_M$$

leads to a distributed normalized LMS algorithm, or dNLMS for short. Other choices are possible, such as a distributed affine projection algorithm (dAPA).

The aggregation mapping leading to $\phi_k^{(i-1)}$ in (2) could be any general (nonlinear) function of the nearby estimates. One linear mapping frequently used is the *Metropolis* rule:

$$\begin{cases} c_{k\ell} = 1/\max(n_k, n_\ell) \quad \text{if } k \neq \ell \text{ are } \textit{linked} \\ c_{k\ell} = 0 \quad \text{if } k \text{ and } \ell \text{ not } \textit{linked} \\ c_{kk} = 1 - \sum_{\ell \in \mathcal{N}_k / k} c_{k\ell} \quad \text{for } k = \ell \end{cases} \tag{3}$$

Other possible mappings are the nearest neighbor and the Laplacian rules.

This material was based on work supported in part by the National Science Foundation under awards ECS-0601266 and ECS-0725441. The work of Mr. Lopes was also supported by a fellowship from CAPES, Brazil, under award 1168/01-0.

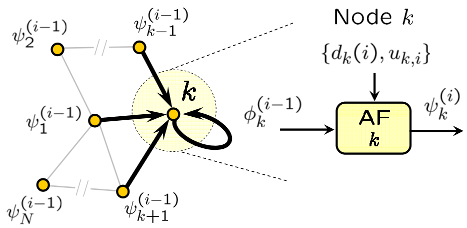

Fig. 1. A network with diffusion cooperation strategy.

3. NETWORK GLOBAL MODEL

Algorithm (2) combines the effect of several adaptive filter updates, and a dynamically changing network topology (since $\mathcal{N}_{k,i}$ and $c_{k,l}(i)$ can vary with time). In order to analyze the performance of such systems we resort to state-space representations. We introduce the global random quantities:

$$\boldsymbol{\psi}^i \triangleq \text{col}\{\boldsymbol{\psi}_1^{(i)}, \ldots, \boldsymbol{\psi}_N^{(i)}\}, \quad \boldsymbol{\phi}^{i-1} \triangleq \text{col}\{\boldsymbol{\phi}_1^{(i-1)}, \ldots, \boldsymbol{\phi}_N^{(i-1)}\}$$
$$\boldsymbol{U}_i \triangleq \text{diag}\{\boldsymbol{u}_{1,i}, \ldots, \boldsymbol{u}_{N,i}\}, \quad \boldsymbol{d}_i \triangleq \text{col}\{\boldsymbol{d}_1(i), \ldots, \boldsymbol{d}_N(i)\}$$
$$\boldsymbol{H}(\boldsymbol{U}_i) \triangleq \text{diag}\left\{\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_N\right\}$$

which collect the data across all N nodes at a particular time snapshot. A global state-space model for (2) is then given by [1, 3]

$$\boxed{\boldsymbol{\psi}^i \;=\; \boldsymbol{G}_i\boldsymbol{\psi}^{i-1} + \boldsymbol{H}\boldsymbol{U}_i^*\left(\boldsymbol{d}_i - \boldsymbol{U}_i\boldsymbol{G}_i\boldsymbol{\psi}^{i-1}\right)} \tag{4}$$

where $\boldsymbol{G}_i = \boldsymbol{C}_i \otimes I_M$. This model captures the dynamic network topology in terms of the combining matrix $\boldsymbol{C}_i = [\boldsymbol{c}_{k\ell}(i)]$. It can accommodate several different adaptive algorithms, time-varying networks, networks subject to link and node failures, and time delays. The matrix $\boldsymbol{C}_i$ captures information about the instantaneous network topology: a nonzero entry $\boldsymbol{c}_{k\ell}(i)$ means that node k is connected to node ℓ at time i. Moreover, $\boldsymbol{C}_i$ satisfies $\boldsymbol{C}_i q_N = q_N$, where $q = \text{col}\{1, 1, \ldots, 1\}$ is $N \times 1$.

4. RANDOM TOPOLOGY MODEL

We will consider undirected graphs ($c_{kl}(i) = c_{lk}(i)$) for simplicity, but the concepts and derivations extend to directed links as well. We model the topology dynamics by assuming that the links and nodes are random entities. We assume that at any given time i, the (now random) link weight $\boldsymbol{c}_{k\ell}(i)$, which connects node k to node ℓ, will assume either a nominal value $c_{k\ell} = c_{\ell k}$ with probability $p_{k\ell} = p_{\ell k}$, or it will be zero with probability q_{kl}:

$$[\boldsymbol{C}_i]_{k\ell} = \begin{cases} c_{k\ell} & \text{with } \; p_{k\ell} \\ 0 & \text{with } \; q_{k\ell} = 1 - p_{k\ell}. \end{cases} \tag{5}$$

The nodes are modelled in a similar fashion, with probability of occurrence equals to p_{kk} – see Fig. 2. For simplicity, we assume in this work that a nominal topology C_0, comprised of a fixed number of nodes N and n_l links, is subject to link failures[1]. Thus, the n_l links gives rise to 2^{n_l} different subnetworks C_ℓ with probability p_ℓ each, comprised of existing and faulty links. The probabilities $\{p_l\}$ are related to the $\{p_{kl}\}$. When a link is removed, a zero is introduced in the corresponding positions in $\boldsymbol{C}_i$. When a node is shut down, all the associated entries are zeroed out in $\boldsymbol{C}_i$.

[1]The case of variable number of nodes will be studied elsewhere.

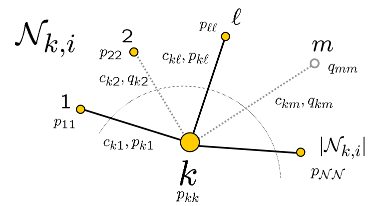

Fig. 2. Mean topology calculation.

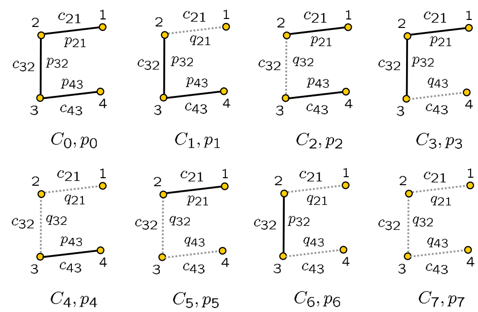

Fig. 3. The subnetworks associated with the nominal matrix C_0.

In this scenario, we define the mean topology matrices $G = E\boldsymbol{G}_i$ and $\mathcal{G} = E(\boldsymbol{G}_i \odot \boldsymbol{G}_i^{*T})$ where $\odot$ denotes the block Kronecker product, namely,

$$G = \sum_{\ell=1}^{2^{n_l}} p_\ell G_\ell \quad \text{and} \quad \mathcal{G} = \sum_{\ell=1}^{2^{n_l}} p_\ell\big(G_\ell \odot G_\ell^{*T}\big) \tag{6}$$

where $p_\ell = P\{\boldsymbol{C}_i = C_\ell\}$, $G_\ell = \big(C_\ell \otimes I_M\big)$ and $\odot$ denotes the block Kronecker product [1], [3]. Figure 3 depicts a simple example for $N = 4$ and $n_l = 3$. For instance, the subnetwork C_2 happens with probability $p_2 = p_{21}q_{32}p_{43}$.

5. MEAN TRANSIENT ANALYSIS

Let

$$w^{(o)} \triangleq q_N \otimes w^o \quad \text{and} \quad \widetilde{\boldsymbol{\psi}}^i \triangleq w^{(o)} - \boldsymbol{\psi}^i \tag{7}$$

From the data model (1) we have $\boldsymbol{d}_i = \boldsymbol{U}_i w^{(o)} + \boldsymbol{v}_i$, where $\boldsymbol{v}_i = \text{col}\{\boldsymbol{v}_1(i), \ldots, \boldsymbol{v}_N(i)\}$. Since $\boldsymbol{G}_i w^{(o)} = w^{(o)}$, by subtracting $w^{(o)}$ from the left side and $\boldsymbol{G}_i w^{(o)}$ from the right side of (4) we get

$$\boxed{\widetilde{\boldsymbol{\psi}}^i = (I_{NM} - \boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{U}_i)\,\boldsymbol{G}_i\widetilde{\boldsymbol{\psi}}^{i-1} - \boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{v}_i} \tag{8}$$

Assuming temporal and spatial independence of the regressors, and that the perturbations in the network topology are not correlated with the captured data and taking expectations of both sides leads to

$$\boxed{E\widetilde{\boldsymbol{\psi}}^i = \big[I_{NM} - E\big(\boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{U}_i\big)\big]\,E\boldsymbol{G}_i\;E\widetilde{\boldsymbol{\psi}}^{i-1}} \tag{9}$$

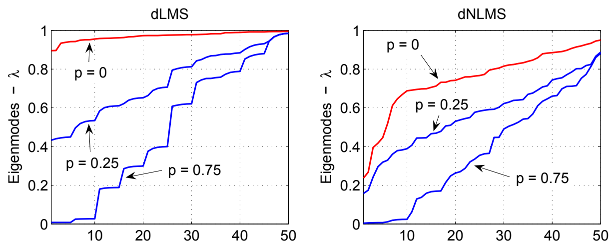

Fig. 4. Mean network modes for diffusion LMS and diffusion NLMS in a network with $N = 10, M = 5$, and white noise with diverse power across the network [1], [4].

Therefore, the mean evolution of the global weight error vector depends on the data moment $E\boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{U}_i$ and on the mean topology matrix $E\boldsymbol{G}_i$.

Figure 4 presents the eigenmodes of an adaptive network running the dLMS algorithm (left) and the dNLMS algorithm (right). Note how cooperation decreases the modes of the network and enhances stability ($p_{kl} = p$ is the probability that link $k\ell$ exists).

6. MEAN-SQUARE TRANSIENT ANALYSIS

Due to space limitations, we will not present the full derivations here. Define the global mean-square deviation (MSD) and excess mean-square error (EMSE) measures as

$$\text{MSD} = \frac{1}{N}E\|\widetilde{\boldsymbol{\psi}}^{i-1}\|^2 \quad \text{and} \quad \text{EMSE}\frac{1}{N}E\|\widetilde{\boldsymbol{\psi}}^{i-1}\|^2_{R_u} \tag{10}$$

where $R_u = E\boldsymbol{U}_i^*\boldsymbol{U}_i$. We resort to energy conservation arguments [3, 4, 6] to derive a model for the network mean-square evolution in terms of weighted norms of $\widetilde{\boldsymbol{\psi}}^{i-1}$, i.e., $\|\widetilde{\boldsymbol{\psi}}^{i-1}\|^2_\Sigma$. For instance, making $\Sigma = I_{NM}$ or $\Sigma = R_u$ retrieves (10). We start by defining the local output estimation error at node k as

$$\boldsymbol{e}_k(i) = \boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{\phi}_k^{(i-1)} \tag{11}$$

and collect the errors across the network into the global vector $\boldsymbol{e}_i = \{\boldsymbol{e}_1(i), \boldsymbol{e}_2(i), \ldots, \boldsymbol{e}_N(i)\}$, so that

$$\boldsymbol{e}_i \;=\; \boldsymbol{d}_i - \boldsymbol{U}_i\boldsymbol{G}_i\boldsymbol{\psi}^{i-1} \;=\; \boldsymbol{U}_i\boldsymbol{G}_i\widetilde{\boldsymbol{\psi}}^{i-1} + \boldsymbol{v}_i \tag{12}$$

substituting the weight error vector definitions in (4), together with (12) yields

$$\widetilde{\boldsymbol{\psi}}^{i} \;=\; \boldsymbol{G}_i\widetilde{\boldsymbol{\psi}}^{i-1} - \boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{e}_i \tag{13}$$

Now performing the weighted energy balance on both sides of (13) for some arbitrary $NM \times NM$ Hermitian matrix $\Sigma \geq 0$, and taking expectations gives

$$\boxed{\begin{aligned} E\|\widetilde{\boldsymbol{\psi}}^{i}\|^2_\Sigma \;&=\; E\|\widetilde{\boldsymbol{\psi}}^{i-1}\|^2_{\Sigma'} \;+\; E\boldsymbol{v}_i^*\boldsymbol{U}_i\boldsymbol{H}\Sigma\boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{v}_i && (14)\\ \Sigma' \;&=\; E\boldsymbol{G}_i^*\Sigma\boldsymbol{G}_i \;-\; E\boldsymbol{G}_i^*\Sigma\boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{U}_i\boldsymbol{G}_i \\ &\quad - E\boldsymbol{G}_i^*\boldsymbol{U}_i^*\boldsymbol{U}_i\boldsymbol{H}\Sigma\boldsymbol{G}_i + E\boldsymbol{G}_i^*\boldsymbol{U}_i^*\boldsymbol{U}_i\boldsymbol{H}\Sigma\boldsymbol{H}\boldsymbol{U}_i^*\boldsymbol{U}_i\boldsymbol{G}_i && (15)\end{aligned}}$$

where we resorted to commonly adopted independence assumptions for mathematical tractability. It is important to remark though, that these assumptions are made for analysis *only*, they do not compromise the spatial-temporal nature of the problem, neither its distributiveness [4]. The recursive variance relation (14) describes the evolution of weighted norms of $\widetilde{\boldsymbol{\psi}}^{i}$ in terms of data statistical moments. Note that the weighting matrix Σ is implicit in the recursion. Therefore, to find closed form expressions for the MSD and EMSE we need to resort to vectorization techniques and block Kronecker products [1], [3], [6]. The derivation is laborious and we will present a particular solution here for evolving topologies and LMS-type updates observing Gaussian regressors, for which we choose $H = D = \text{diag}\{\mu_1 I_M, \ldots, \mu_N I_M\}$. Subsequently we apply a Gaussian transformation to (14) via $\overline{\boldsymbol{\psi}}^{i} = T^*\widetilde{\boldsymbol{\psi}}^{i}$, $\overline{\boldsymbol{U}}_i = \boldsymbol{U}_i T$, and $\overline{\boldsymbol{G}}_i = T^*\boldsymbol{G}_i T$, where $R_u = T\Lambda T^*$, T is unitary, and $\Lambda = \text{diag}\{\Lambda_1, \ldots, \Lambda_N\}$, with $\Lambda_k > 0$ and diagonal. These steps lead to a recursion that describes the mean-square performance of the network in terms of the free vectorized parameter $\overline{\sigma} = \text{bvec}\{\overline{\Sigma}\}$: [2]

$$\boxed{\begin{aligned} E\|\overline{\boldsymbol{\psi}}^{i}\|^2_{\overline{\sigma}} \;&=\; E\|\overline{\boldsymbol{\psi}}^{i-1}\|^2_{\overline{F}\overline{\sigma}} \;+\; b^T\overline{\sigma} && (16)\\ \overline{F} \;&=\; \overline{\mathcal{G}}\cdot\big[I_{N^2M^2} \;-\; (I_{NM}\odot\Lambda D) \\ &\qquad -\; (\Lambda D\odot I_{NM}) \;+\; (D\odot D)\mathcal{A}\big]\overline{\sigma} && (17)\end{aligned}}$$

where $\overline{\mathcal{G}} = E\big(\overline{\boldsymbol{G}}_i \odot \overline{\boldsymbol{G}}_i^{*T}\big)$, $b = \text{bvec}\{R_v D^2\Lambda\}$, $R_v = \Lambda_v \odot I_M$ and $\Lambda_v = \text{diag}\{\sigma^2_{v,1}, \ldots, \sigma^2_{v,N}\}$. Choosing $\overline{\sigma} = \text{bvec}\{I_{NM}\}$ in (16) gives the MSD. Selecting $\overline{\sigma} = \text{bvec}\{\Lambda\}$ gives the EMSE.

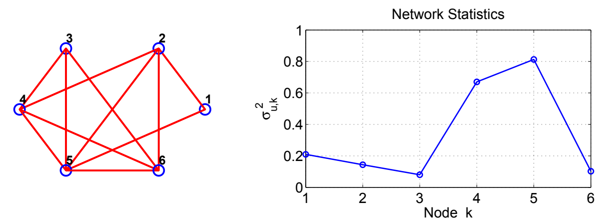

Fig. 5. Network topology (left) and statistics (right).

7. PROBABILISTIC DIFFUSION LMS

An interesting variant of the diffusion protocols studied in [1, 2] arises when we let the nodes communicate only with a subset of their direct neighbors. By doing so, we may save valuable energy and communication resources.

It was argued in [4] that the intermediate averaging iterations in consensus techniques are not necessary, provided that the step-size is small enough. We now observe that the amount of interaction among the nodes can actually be reduced to a good extent *without degrading* the performance in the mean-square sense.

Figure 5 presents the network settings for a probabilistic diffusion example with the corresponding eigenmodes captured in Fig. 6, for different values of the link probability p. Despite the fact that the faster modes are altered by decreasing p (see left plot), the maximum eigenmode, which ultimately determines the convergence rate, remains practically unaltered, as depicted in the right plot of Fig. 6. In all simulations the background noise was set to $\sigma^2_{v,k} = 10^{-3}$.

[2] The $M \times M$ blocks of $\overline{\Sigma}$ are first stacked onto one single block column $\overline{\Sigma}^c$ (an $N^2M \times M$ matrix); then moving along $\overline{\Sigma}^c$, each individual block is vectorized into an $M^2 \times 1$ column vector, so that $\overline{\sigma}$ is $N^2M^2 \times 1$.

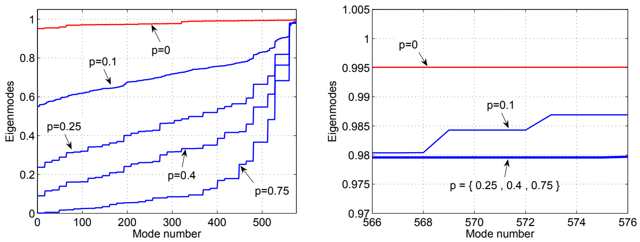

Fig. 6. Mean-square network eigenmodes (left) and zoom (right).

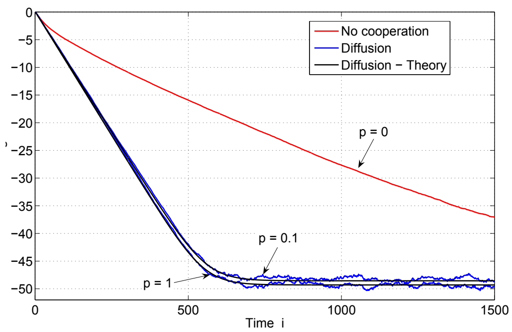

Fig. 7. Global MSD performance for $p = 0$ (no cooperation), $p = 0.1$ (probabilistic diffusion) and $p = 1$ (standard diffusion).

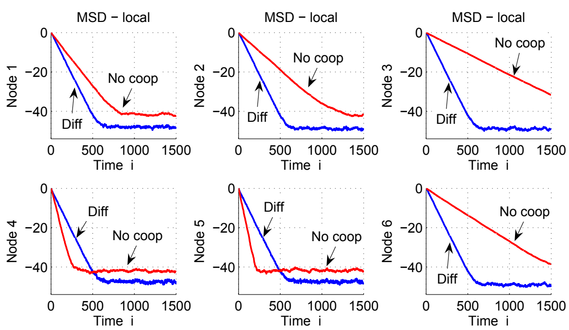

Fig. 8. MSD per node (local).

In Fig. 7 we present the global MSD evolution for $p = 1$ and $p = 0.1$. In other words, in the first case all available links are used; in the second, only about 10% of the links are employed. Still a significant improvement over the non-cooperative counterpart is noticed. Figure 8 presents the local MSD evolution for the nodes across the network. Note the equalization effect in the diffusion curves: despite the quite low cooperation resources, the nodes present a strikingly similar mean-square performance. Figure 9 presents the instantaneous network traffic: the network usage is quite below its full capacity, which is defined as the number of available links at time i.

We simulated another example with a larger network and with signals presenting diverse power $\sigma^2_{u,k}$ and correlation indices α_k. The same noise level $\sigma^2_v = 10^{-3}$ was employed [1]–[4]. The same effect takes place: now with considerable improvement even for a link probability as low as $p = 0.01$, as Fig. 10 shows.

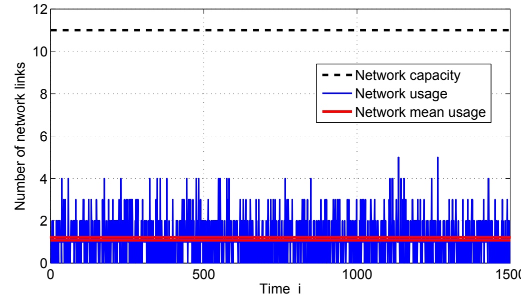

Fig. 9. Network traffic.

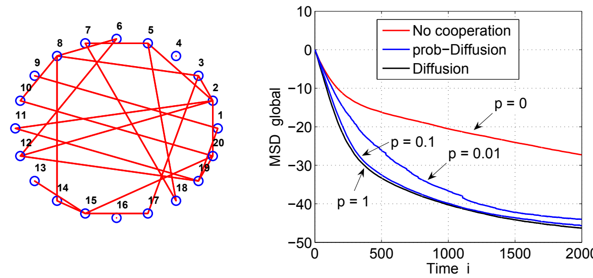

Fig. 10. Left: Network topology. Right: MSD performance.

8. CONCLUDING REMARKS

As simulations and theory have shown, diminishing the amount of message exchanges among nodes remarkably does not degrade the network performance for a wide range of link probability p. In other words, much less communication resources may be employed to achieve a pre-defined performance level in diffusion protocols.

We are currently extending the adaptive diffusion protocol [2], [7] to compensate for node failures.

9. REFERENCES

[1] C. G. Lopes and A. H. Sayed, "Steady-state performance of adaptive diffusion least-mean squares," *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, Madison, WI, August 2007.

[2] C. G. Lopes and A. H. Sayed, "Diffusion least-mean-squares over adaptive networks," *Proc. ICASSP*, Honolulu, Hawaii, vol. 3, pp. 917-920, April 2007.

[3] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: formulation and performance analysis," *IEEE Transactions on Signal Processing*, to appear, 2008.

[4] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064-4077, August 2007.

[5] D. Li, K. D. Wong, Y. H. Hu and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, Issue 2, March 2002, pp. 17-29.

[6] A. H. Sayed. *Fundamentals of Adaptive Filtering*, Wiley, NJ, 2003.

[7] J. Arenas-Garcia, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1078–1090, March 2006.