ADAPTIVE DISTRIBUTED COMPRESSED SENSING FOR DYNAMIC HIGH-DIMENSIONAL HYPOTHESIS TESTING

Nicolò Michelusi and Urbashi Mitra

Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, USA

ABSTRACT

In this paper, a framework for dynamic high-dimensional hypothesis testing in wireless sensor networks is presented. The sensor nodes (SNs) collect and transmit to a fusion center (FC), in a distributed fashion, compressed measurements of a time-correlated hypothesis vector. The FC, based on the measurements collected, tracks the hypothesis vector, and feeds back minimal information about the uncertainty in the current estimate, which enables adaptation of the SNs' data collection and transmission strategy. The policy of the SNs is optimized with the overall objective of minimizing the detection error probability, under sensing and transmission cost constraints incurred by each SN. A Bernoulli approximation on the detection error is employed, which enables a significant reduction in the optimization complexity and the design of scalable estimators based on sparse approximation recovery algorithms. Simulation results demonstrate that, for a target 5% detection error, the adaptive scheme attains 90% and 50% cost savings with respect to a memoryless scheme which does not exploit the time-correlation and a non-adaptive one, respectively.

***Index Terms*—** Hypothesis testing, stochastic optimization, distributed systems, sensor networks

1. INTRODUCTION

Wireless sensor networks (WSNs) make it possible to monitor the environment by means of tiny sensors (SNs) distributed over the field to perform data acquisition, processing and communication tasks [1]. In this paper, we consider a WSN where the fusion center (FC) tracks a time-correlated binary hypothesis vector, by collecting low-dimensional (compressed) noisy measurements from nearby SNs. This situation arises, for instance, in cognitive radio networks [2], where secondary users need to track the busy/idle state of channels to optimize their access. In particular, we devise a scheme which exploits the time-correlation in the process: by leveraging the estimate in the previous slot and collecting new measurements in the present slot, the FC needs only to estimate a *sparse* residual uncertainty vector, thus enabling the use of sparse approximation and recovery techniques. However, due to the distributed operation of the SNs, the amount of measurements collected at the FC is uncertain and dynamic, thus inducing fluctuations in the detection performance. Similar to our previous works [3, 4], we propose a scheme where the FC feeds back minimal information about the uncertainty in the current estimate, thus enabling adaptation of the SNs' data collection and transmission strategy. The data collection and reporting strategy is optimized via dynamic programming (DP) [5], so as to minimize the detection error probability at the FC, under cost constraints incurred by each SN. In order to tackle the high dimensionality of the problem, we propose a Bernoulli approximation for the detection error, which enables a significant reduction in the state space and in the optimization complexity and the design of scalable estimators based on sparse approximation recovery algorithms. The optimal policy prescribes that, when the estimation quality is good (low uncertainty), the SNs remain idle, in order to preserve energy. On the other hand, when the estimation quality is poor (high uncertainty), the SNs react by collecting more measurements, at higher cost.

Feedback schemes in distributed estimation and detection settings have been proposed in [6–9]. In [6], a two message feedback architecture for binary hypothesis testing is considered, in which the second message of each SN is based on full or partial knowledge of the first message of the other SNs. In [8, 9], the FC estimates a finite state Markov chain and a random field, respectively, and, in each slot, feeds back the posterior state distribution, which is used by the SNs to adapt their quantizers to minimize the mean squared estimation error. In these works, fixed rate quantizers are employed, so that the cost of data acquisition and transmission is not explicitly accounted for. In contrast, we employ a cross-layer perspective, *i.e.*, the data reporting strategy is optimized to trade-off these costs and the detection error probability.

Compressive sensing (CS) [10, 11] enables the recovery of sparse signals, using only a small number of measurements. In [12], a CS framework to track the support of a time-correlated Bernoulli-Gaussian signal is presented. Therein, a centralized scheme is employed. In contrast, in our work, we employ a distributed setting where each SN collects a compressed measurement of the hypothesis vector; we devise a feedback scheme which enables adaptation of the SNs. Distributed CS is studied in [13, 14], for a static signal model. In [13], a number of SNs measure signals that are each indi-

This research has been funded in part by the following grants: ONR N00014-09-1-0700, CCF-0917343, CCF-1117896, CNS-1213128, AFOSR FA9550-12-1-0215, and DOT CA-26-7084-00. N. Michelusi is in part supported by AEIT (Italian association of electrical engineering) through the scholarship "Isabella Sassi Bonadonna 2013".

N. Michelusi and U. Mitra are with the Dept. of Electrical Engineering, University of Southern California. email: michelus@usc.edu.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

vidually sparse in some basis and also correlated from SN to SN. Intra- and inter-signal correlation structures are exploited to enable signal acquisition via CS. In [15], a Kronecker CS framework is developed, which enables the acquisition of multidimensional signals, *e.g.*, spatially and temporally sparse/compressible. In [14], a distributed basis pursuit algorithm is designed to minimize the communication between nodes in the network. Recovery of static binary sparse signals via CS has been investigated in [16, 17], and its application to spectrum sensing has been studied in [2].

This paper is organized as follows. In Sec. 2, we present the system model and the optimization problem. In Sec. 3, we present the Bernoulli approximation, based on which a low-complexity feedback scheme is developed in Sec. 4. In Sec. 5, we present numerical results. Sec. 6 concludes the paper with some final remarks.

2. SYSTEM MODEL

Let $\mathbf{b}_k$ be an N-dimensional *binary hypothesis vector*, taking value in $\{0,1\}^N$, with the following temporal dynamics:

$$\mathbf{b}_{k+1} = \mathbf{b}_k \oplus \mathbf{w}_k, \tag{1}$$

where k is the time-slot index, $\oplus$ denotes the component-wise binary XOR operation, $\mathbf{w}_k$ is an N-dimensional *evolution vector*, taking value in $\{0,1\}^N$, with i.i.d. Bernoulli components with probability $\mathbb{P}(\mathbf{w}_{k,i}=1)=p_W$, i.i.d. over time. Therefore, from (1), each component of $\mathbf{b}_k$ evolves as a two state Markov chain, with transition probabilities $P_B(b_1|b_0) \triangleq \mathbb{P}(\mathbf{b}_{k+1,i}=b_1|\mathbf{b}_{k,i}=b_0),\ b_0,b_1\in\{0,1\}$, where $P_B(b|b) = 1-p_W,\ \forall b\in\{0,1\}$. At steady state, $\mathbb{P}(\mathbf{b}_{k,i}=1)=0.5$.[1]

A set of N_S SNs collect noisy measurements of $\mathbf{b}_k$ as

$$y_{n,k} = \mathbf{a}_{n,k}^T\mathbf{b}_k + z_{n,k},\ \forall n=1,2,\dots,N_S, \tag{2}$$

where $y_{n,k}\in\mathbb{R}$ is the *compressed measurement* collected at node n in slot k, $z_{n,k}\sim\mathcal{N}(0,\sigma_R^2)$ is Gaussian noise, i.i.d. over time and across SNs, $\mathbf{a}_{n,k}^T$ is the measurement vector, and the superscript "T" denotes the matrix transpose. For simplicity, we assume that $\mathbf{a}_{n,k}\sim\mathcal{N}(\mathbf{0},\sigma_A^2\mathbf{I}_N)$, i.i.d. over time and across SNs, where $\mathbf{I}_p$ is the $p\times p$ unit matrix. However, the following analysis can be extended to the more general case where $\mathbf{a}_{n,k}$ is non-Gaussian, non-i.i.d. Note that, for energy efficiency purposes (data collection and transmission costs), we assume that only a scalar measurement is acquired at each SN in each slot ($y_{n,k}$ at SN n), rather than a vector of measurements. However, this framework can be extended to multi-dimensional measurements as well.

Examples: The vector $\mathbf{b}_k$ may represent, for instance, the busy/idle state of channels in cognitive radio networks [2], and the measurement vector $\mathbf{a}_{n,k}$ is thus the result of the channel attenuation between primary users and cognitive radios, as well as of filtering operations occurring at each terminal. Alternatively, $\mathbf{b}_k$ could be the global state of a wireless network, so that $y_{n,k}$ may represent a compressed view of the network state at node n, aggregated from nearby nodes, *e.g.*, using consensus strategies (some entries of $\mathbf{a}_{n,k}$ may be equal to zero in this case, depending on the network topology; this analysis is left for future research). The estimate of $\mathbf{b}_k$ may then be used at the FC to implement centralized control schemes that adapt to the network state [18].

[1]The case $P_B(1|1)\neq P_B(0|0)$ is of interest and will be considered as a future work.

Each SN, in each slot k, collects the measurement and reports it to the FC via a single-hop wireless link, with common probability α_k, and remains idle otherwise, to preserve energy. The set of SNs that report their measurement to the FC is denoted as $\mathcal{S}_k$, whose cardinality $S_k=|\mathcal{S}_k|$ is a binomial random variable with N_S trials and mean $N_S\alpha_k$, $S_k\sim\mathcal{B}(N_S,\alpha_k)$. Let $\mathbf{r}_k$ be the (possibly empty, if $S_k=0$) measurement vector collected at the FC in slot k, given by

$$\mathbf{r}_k = \mathbf{A}_k^T\mathbf{b}_k + \mathbf{z}_k, \tag{3}$$

where $\mathbf{A}_k=[\mathbf{a}_{n,k}]_{n\in\mathcal{S}_k}$ is the measurement matrix, known to the FC, and $\mathbf{z}_k=[z_{n,k}]_{n\in\mathcal{S}_k}$ is the column noise vector. Note that the size of $\mathbf{r}_k$ is random, due to the probabilistic transmission decision of each SN. Given the sequence of measurement vectors $\mathbf{r}_0^k=(\mathbf{r}_0,\mathbf{r}_1,\dots,\mathbf{r}_k)$ and of measurement matrices $\mathbf{A}_0^k=(\mathbf{A}_0,\mathbf{A}_1,\dots,\mathbf{A}_k)$, the FC estimates $\mathbf{b}_k$ using the *maximum-a-posteriori* (MAP) estimator

$$\hat{\mathbf{b}}_k = \boldsymbol{\beta}_k(\mathbf{r}_0^k,\mathbf{A}_0^k) \triangleq \arg\max_{\mathbf{b}_k\in\{0,1\}^N}\mathbb{P}(\mathbf{b}_k|\mathbf{r}_0^k,\mathbf{A}_0^k). \tag{4}$$

We denote the detection error probability, given $\mathbf{r}_0^k$, $\mathbf{A}_0^k$, as

$$P_{E,k}(\mathbf{r}_0^k,\mathbf{A}_0^k) \triangleq \frac{1}{N}\mathbb{E}\left[\left\|\boldsymbol{\beta}_k(\mathbf{r}_0^k,\mathbf{A}_0^k)-\mathbf{b}_k\right\|_F^2\middle|\mathbf{r}_0^k,\mathbf{A}_0^k\right]. \tag{5}$$

The FC, at the beginning of each slot, broadcasts the value of the common transmission probability $\alpha_k(\mathbf{r}_0^{k-1},\mathbf{A}_0^{k-1})\in[0,1]$ employed by the SNs in slot k to make their transmission decision. Its value is based, possibly, on the history $\mathbf{r}_0^{k-1}$, $\mathbf{A}_0^{k-1}$. We assume that each SN incurs the cost of one unit to perform data collection and transmission to the FC, whereas it incurs no cost in staying idle, so that the expected cost is $\alpha_k(\mathbf{r}_0^{k-1},\mathbf{A}_0^{k-1})$. We define the time-average detection error and average cost of each SN over a finite time-horizon of length K, for a given MAP estimator $\boldsymbol{\beta}=(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_{K-1})$ and feedback policy $\alpha=(\alpha_0,\alpha_1,\dots,\alpha_{K-1})$, as

$$\begin{aligned}\bar{P}_E^K(\boldsymbol{\beta},\alpha) &\triangleq \tfrac{1}{K}\mathbb{E}_{\boldsymbol{\beta},\alpha}\left[\textstyle\sum_{k=0}^{K-1}P_{E,k}(\mathbf{r}_0^k,\mathbf{A}_0^k)\right],\\ \bar{C}_{SN}^K(\boldsymbol{\beta},\alpha) &\triangleq \tfrac{1}{K}\mathbb{E}_{\boldsymbol{\beta},\alpha}\left[\textstyle\sum_{k=0}^{K-1}\alpha_k(\mathbf{r}_0^{k-1},\mathbf{A}_0^{k-1})\right],\end{aligned} \tag{6}$$

The objective is to determine the optimal $(\boldsymbol{\beta}^*,\alpha^*)$ that trades off detection error and cost incurred by each SN, defined as

$$(\boldsymbol{\beta}^*,\alpha^*) = \arg\min P_E^K(\boldsymbol{\beta},\alpha)+\lambda\bar{C}_{SN}^K(\boldsymbol{\beta},\alpha), \tag{7}$$

for some $\lambda>0$. Let $\pi_k(\mathbf{b}_k)=\mathbb{P}(\mathbf{b}_k|\mathbf{r}_0^{k-1},\mathbf{A}_0^{k-1})$, $\mathbf{b}_k\in\{0,1\}^N$, be the prior belief of $\mathbf{b}_k$ at time k, before collecting the measurements in slot k, and $\hat{\pi}_k(\mathbf{b}_k)=\mathbb{P}(\mathbf{b}_k|\mathbf{r}_0^k,\mathbf{A}_0^k)$, be the posterior belief of $\mathbf{b}_k$ at time k, after collecting the measurements in slot k. We have the recursive expressions

$$\begin{aligned}\hat{\pi}_k(\mathbf{b}_k) &= \tfrac{\mathbb{P}(\mathbf{r}_k|\mathbf{b}_k,\mathbf{A}_k)\pi_k(\mathbf{b}_k)}{\sum_{\mathbf{b}_k'}\mathbb{P}(\mathbf{r}_k|\mathbf{b}_k',\mathbf{A}_k)\pi_k(\mathbf{b}_k')},\\ \pi_{k+1}(\mathbf{b}_{k+1}) &= \textstyle\sum_{\mathbf{b}_k}\mathbb{P}(\mathbf{b}_{k+1}|\mathbf{b}_k)\hat{\pi}_k(\mathbf{b}_k),\end{aligned} \tag{8}$$

where $\{\mathbf{r}_k|\mathbf{b}_k, \mathbf{A}_k\} \sim \mathcal{N}(\mathbf{A}_k^T\mathbf{b}_k, \sigma_R^2\mathbf{I}_{S_k})$, and $\mathbb{P}(\mathbf{b}_{k+1}|\mathbf{b}_k) = \prod_{i=1}^N P_B(\mathbf{b}_{k+1,i}|\mathbf{b}_{k,i})$. Then, it can be shown that π_k is a sufficient statistic for decision making in slot k, so that $(\boldsymbol{\beta}^*, \alpha^*)$ can be restricted to be a function of π_k only. Then, from (4),

$$\hat{\mathbf{b}}_k = \boldsymbol{\beta}^*(\hat{\pi}_k) = \arg\max_{\mathbf{b}_k\in\{0,1\}^N} \hat{\pi}_k(\mathbf{b}_k) = \mathbf{b}_k \oplus \mathbf{e}_k, \quad (9)$$

where $\mathbf{e}_k$ is the detection error, and we have removed the dependence of $\boldsymbol{\beta}^*$ on k, since it only depends on $\hat{\pi}_k$.

The optimal feedback control $\alpha_k^*(\pi_k)$ can be determined recursively using DP [5]. However, this optimization, as well as the MAP estimator (9), have high complexity. In particular, (9) requires a search over a high dimensional space, of cardinality 2^N. The belief π_k needs to be computed via (8), which involves a marginalization over a high dimensional space. Finally, DP involves an expectation over $\mathbf{b}_k \in \{0,1\}^N$, the local transmission decisions of the SNs, and $\mathbf{r}_k$. These are all combinatorial problems, whose complexity grows exponentially with N_S and N. In the next section, we propose a Bernoulli approximation for the detection error $\mathbf{e}_k$ in (9), which enables a significant complexity reduction.

3. BERNOULLI APPROXIMATION

Using (9) and given the posterior belief of $\mathbf{b}_k$ in slot k, $\hat{\pi}_k$, the distribution of $\mathbf{e}_k$ is given by

$$\mathbb{P}(\mathbf{e}_k|\hat{\pi}_k) = \hat{\pi}_k(\hat{\mathbf{b}}_k \oplus \mathbf{e}_k),\ \forall \mathbf{e}_k \in \{0,1\}^N. \quad (10)$$

Herein, due to the intractability of using the true probability mass function $\mathbb{P}(\mathbf{e}_k|\hat{\pi}_k)$, we approximate the detection error $\mathbf{e}_k$ as a Bernoulli vector with i.i.d. entries with probability

$$P_{E,k} = \sum_{\mathbf{e}_k\in\{0,1\}^N} \frac{1}{N}\|\mathbf{e}_k\|_0\, \hat{\pi}_k(\hat{\mathbf{b}}_k \oplus \mathbf{e}_k), \quad (11)$$

denoted as the *average detection error*, where $\|\mathbf{x}\|_0$ is the ℓ_0 norm of $\mathbf{x}$. Letting $\hat{\mathbf{w}}_k = \mathbf{e}_k \oplus \mathbf{w}_k$ in (9) and (1), we obtain

$$\mathbf{b}_{k+1} = \hat{\mathbf{b}}_k \oplus \mathbf{e}_k \oplus \mathbf{w}_k = \hat{\mathbf{b}}_k \oplus \hat{\mathbf{w}}_k. \quad (12)$$

Then, using the Bernoulli approximation on $\mathbf{e}_k$, it follows that $\hat{\mathbf{w}}_k$ has Bernoulli i.i.d. entries with probability

$$P_{\hat{W},k} \triangleq \mathbb{P}(\hat{\mathbf{w}}_{k,i} = 1|P_{E,k}) = p_W + P_{E,k}(1-2p_W), \quad (13)$$

so that, given the estimate $\hat{\mathbf{b}}_k$ and $P_{\hat{W},k}$, π_{k+1} can be derived from (12) accordingly. We term $P_{\hat{W},k}$ as the *uncertainty state*, since it captures the amount of uncertainty in the unobserved $\mathbf{b}_{k+1}$, given the estimate $\hat{\mathbf{b}}_k$ of $\mathbf{b}_k$ in slot k, and we use it in place of the true belief state π_k. Given the vector of measurements collected at the FC in slot $k+1$, $\mathbf{r}_{k+1} = \mathbf{A}_{k+1}^T\mathbf{b}_{k+1} + \mathbf{z}_{k+1}$, from Bayes' rule we obtain

$$\hat{\pi}_{k+1}(\mathbf{b}_{k+1}) = \frac{\mathbb{P}(\mathbf{r}_{k+1}|\mathbf{b}_{k+1}, \mathbf{A}_{k+1})\mathbb{P}(\mathbf{b}_{k+1}|\hat{\mathbf{b}}_k, P_{\hat{W},k})}{\sum_{\mathbf{b}'_{k+1}} \mathbb{P}(\mathbf{r}_{k+1}|\mathbf{b}'_{k+1}, \mathbf{A}_{k+1})\mathbb{P}(\mathbf{b}'_{k+1}|\hat{\mathbf{b}}_k, P_{\hat{W},k})}.$$

Since $\{\mathbf{r}_{k+1}|\mathbf{b}_{k+1}, \mathbf{A}_{k+1}\} \sim \mathcal{N}(\mathbf{A}_{k+1}^T\mathbf{b}_{k+1}, \sigma_R^2\mathbf{I}_{S_{k+1}})$, using (12) we thus obtain

$$\hat{\pi}_{k+1}(\mathbf{b}_{k+1}) \propto \exp\left\{-\frac{1}{2\sigma_R^2}\left\|\mathbf{r}_{k+1} - \mathbf{A}_{k+1}^T\mathbf{b}_{k+1}\right\|_F^2\right\} \times \mathbb{P}(\hat{\mathbf{w}}_k = \mathbf{b}_{k+1} \oplus \hat{\mathbf{b}}_k|P_{\hat{W},k}). \quad (14)$$

Equivalently, we can express the above likelihood function in terms of $\hat{\mathbf{w}}_k$. In particular, since $\hat{\mathbf{w}}_k$ has Bernoulli i.i.d. entries with probability $P_{\hat{W},k}$ and $\mathbf{b}_{k+1,i} = \hat{\mathbf{w}}_{k,i}(1-\hat{\mathbf{b}}_{k,i}) + (1-\hat{\mathbf{w}}_{k,i})\hat{\mathbf{b}}_{k,i}$, we obtain

$$\hat{\pi}_{k+1}(\mathbf{b}_{k+1}) \propto \exp\left\{-\frac{1}{2\sigma_R^2}\left\|\hat{\mathbf{r}}_{k+1} - \hat{\mathbf{A}}_{k+1}^T\hat{\mathbf{w}}_k\right\|_F^2\right\} \times (1-P_{\hat{W},k})^{N-\|\hat{\mathbf{w}}_k\|_0} P_{\hat{W},k}^{\|\hat{\mathbf{w}}_k\|_0}, \quad (15)$$

where we have defined the *residual measurement vector* $\hat{\mathbf{r}}_{k+1}$ and the *residual measurement matrix* $\hat{\mathbf{A}}_{k+1}$, given by

$$\hat{\mathbf{r}}_{k+1} \triangleq \mathbf{r}_{k+1} - \mathbf{A}_{k+1}^T\hat{\mathbf{b}}_k,\ \hat{\mathbf{A}}_{k+1} \triangleq (\mathbf{I}_N - 2\mathrm{diag}(\hat{\mathbf{b}}_k))\mathbf{A}_{k+1}.$$

Then, using (12) and taking the logarithm of (15), we obtain

$$\boldsymbol{\beta}_{k+1}^*(\hat{\pi}_{k+1}) = \arg\max_{\mathbf{b}_{k+1}\in\{0,1\}^N} \ln \hat{\pi}_{k+1}(\mathbf{b}_{k+1}) = \hat{\mathbf{b}}_k \oplus \hat{\mathbf{w}}_k^*, \quad (16)$$

where $\hat{\mathbf{w}}_k^*$ is the MAP estimator of $\hat{\mathbf{w}}_k$, defined as

$$\hat{\mathbf{w}}_k^* = \arg\min_{\hat{\mathbf{w}}_k\in\{0,1\}^N}\left\{\left\|\hat{\mathbf{r}}_{k+1} - \hat{\mathbf{A}}_{k+1}^T\hat{\mathbf{w}}_k\right\|_F^2 + \mu_k\|\hat{\mathbf{w}}_k\|_0\right\}, \quad (17)$$

where we have defined $\mu_k \triangleq 2\sigma_R^2 \ln\left(\frac{1-P_{\hat{W},k}}{P_{\hat{W},k}}\right)$. In particular, $\hat{\mathbf{w}}_k^*$ can be determined using sparse approximation algorithms. To this end, we perform a convex relaxation of (17), by extending the optimization over the convex space $\hat{\mathbf{w}}_k \in [0,1]^N$, rather than the discrete set $\hat{\mathbf{w}}_k \in \{0,1\}^N$, and by relaxing the ℓ_0 regularization term $\mu_k\|\hat{\mathbf{w}}_k\|_0$ with the ℓ_1 term $\mu_k\|\hat{\mathbf{w}}_k\|_1 = \mu_k\sum_i|\hat{\mathbf{w}}_{k,i}|$, so that (17) becomes a quadratic programming problem, whose solution is denoted as $\tilde{\mathbf{w}}_k^* \in [0,1]^N$. After obtaining $\tilde{\mathbf{w}}_k^*$, $\hat{\mathbf{w}}_k^* \in \{0,1\}^N$ can be finally approximated using a minimum distance criterion.

4. APPROXIMATE FEEDBACK CONTROL α

We now determine an approximate feedback control scheme $\alpha_k(P_{\hat{W},k-1})$, which leverages the Bernoulli approximation proposed in Sec. 3. Note that $\alpha_k(P_{\hat{W},k-1})$ is now a function of the uncertainty state $P_{\hat{W},k-1}$, rather than the prior belief π_k, due to the approximation employed. Let $\rho(S_k, P_{\hat{W},k-1})$ be the average detection error probability, when $\hat{\mathbf{w}}_{k-1}$ in (12) has i.i.d. components with probability $P_{\hat{W},k-1}$ and S_k measurements are received at the FC. In particular, $\rho(S_k, P_{\hat{W},k-1})$ is obtained by marginalizing with respect to both the observation vector $\mathbf{r}_k$ and the measurement matrix $\mathbf{A}_k$, whose entries are generated i.i.d. from $\mathcal{N}(\mathbf{0}, \sigma_A^2)$, for dimensionality reduction purposes. This approximation inherently assumes that S_k and $P_{\hat{W},k-1}$ are the most important factors which determine the detection performance of the MAP estimator, whereas the actual realizations of $\mathbf{r}_k$ and $\mathbf{A}_k$ are not as influential. We can thus approximate the average detection error probability and cost as

$$\begin{aligned} \hat{P}_E^K(\alpha) &\simeq \tfrac{1}{K}\mathbb{E}_\alpha\left[\textstyle\sum_{k=0}^{K-1}\rho(S_k, P_{\hat{W},k-1})\right], \\ \hat{C}_{SN}^K(\alpha) &\simeq \tfrac{1}{K}\mathbb{E}_\alpha\left[\textstyle\sum_{k=0}^{K-1}\alpha_k(P_{\hat{W},k-1})\right], \end{aligned} \quad (18)$$

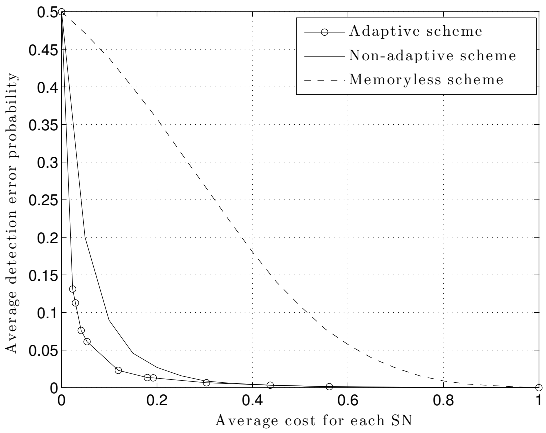

Fig. 1. Average detection error vs average transmission cost for each SN.

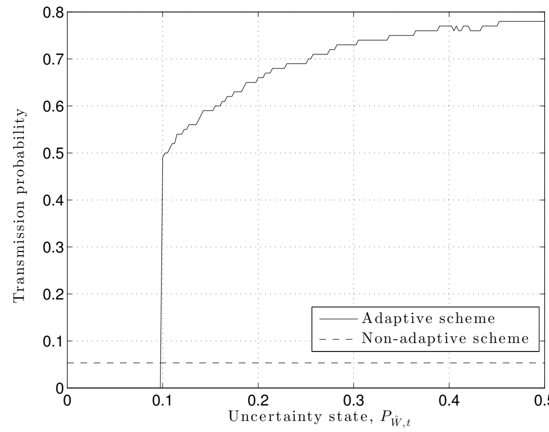

Fig. 2. Transmission probability as a function of the uncertainty state, for the adaptive and non-adaptive schemes.

where, using (13), the uncertainty state is updated as

$$P_{\hat{W},k}=\phi(S_k;P_{\hat{W},k-1})\triangleq p_W+\rho(S_k,P_{\hat{W},k-1})(1-2p_W), \quad (19)$$

and $S_k \sim \mathcal{B}(N_S, \alpha_k(P_{\hat{W},k-1}))$. The optimal feedback control policy can then be derived using the DP algorithm, by solving recursively, for $k = K-1, K-2, \dots, 0$,

$$\hat{C}_{\alpha^*}^{k,K}(P_{\hat{W},k-1}) = \max_{\alpha_k\in[0,1]} \sum_{s=0}^{N_S} \mathbb{P}(S_k = s|\alpha_k)\rho(s, P_{\hat{W},k-1})$$
$$+\lambda\alpha_k + \sum_{s=0}^{N_S} \mathbb{P}(S_k = s|\alpha_k)\hat{C}_{\alpha^*}^{k+1,K}(\phi(s; P_{\hat{W},k-1})), \quad (20)$$

with initial condition $\hat{C}_{\alpha^*}^{K,K}(P_{\hat{W},K}) = 0$. The optimizer above is the optimal policy $\alpha_k^*(P_{\hat{W},k-1})$.

Note that, leveraging the proposed approximation, we have marginalized the dynamics of the system with respect to the specific realization of the measurement process $\mathbf{r}_0^{K-1}$ and measurement matrices $\mathbf{A}_0^{K-1}$, making the problem tractable. The state is thus captured by the uncertainty state $P_{\hat{W},k-1} \in [0, 1]$, rather than the belief π_k, and its evolution is only a function of the (random) number of measurements collected at the FC, S_k, via the state update equation (19). A significant reduction in complexity is thus achieved.

5. NUMERICAL RESULTS

In this section, we present numerical results. The parameters are: $N_S = 10$, $N = 10$, $p_W = 0.01$, $\sigma_A^2 = 1$ and $\sigma_R^2 = 0.1$. We consider the following schemes:

- *Adaptive scheme* (AS): solution of the DP algorithm (20); the transmission probability of each SN in each slot is adapted to the uncertainty state $P_{\hat{W},k-1}$;
- *Non-adaptive scheme* (NAS): each SN transmits with fixed probability in each slot, independently of $P_{\hat{W},k-1}$;
- *Memoryless scheme* (MS): the FC does not exploit the temporal correlation in the process $\mathbf{b}_k$; the SNs transmit with a fixed probability in each slot.

The MAP estimator (17) is solved via exhaustive search (optimal decoder), so that the ultimate performance bounds are characterized. However, sparse approximation recovery algorithms can be employed, as explained in Sec. 3, providing no additional insights. The error curve $\rho(S_k, P_{\hat{W},k-1})$ is obtained via Monte-Carlo simulation. In Fig. 1, we plot the average detection error probability as a function of the cost incurred by each SN, obtained via simulation over a time-horizon of length $K_{sim} = 1000$. We notice that AS and NAS attain a significant performance improvement with respect to MS. AS further improves NAS in the detection error versus cost trade-off. For instance, for a target 5% detection error, AS attains 90% and 50% cost savings with respect to MS and NAS, respectively. In fact, as can be seen in Fig. 2, AS allows the SNs to remain idle and save energy when the uncertainty state is small, *i.e.*, the detection accuracy is high at the FC, whereas the SNs report more aggressively when the detection accuracy at the FC is poor, in order to improve it. Note that, at initialization (slot $k = 0$), we have $P_{\hat{W},0} = 0.5$, and therefore the FC is highly uncertain about $\mathbf{b}_0$. This initial uncertainty is corrected by activating more SNs (transmission probability ~ 0.8), so that a large number of measurements is collected at the FC, providing a good initial estimate of $\mathbf{b}_0$.

6. CONCLUSIONS

In this paper, we have presented a framework for dynamic high-dimensional hypothesis testing in WSNs, based on compressed measurements collected at the FC from nearby SNs. The FC feeds back minimal information about the uncertainty in the current estimate, which enables adaptation of the SNs' data collection and transmission strategy. The policy of the SNs is optimized, with the overall objective of minimizing the detection error probability, under sensing and transmission cost constraints incurred by each SN. In order to tackle the high dimensionality of the problem, we employ a Bernoulli approximation for the detection error, which enables a significant reduction in the state space and in the optimization complexity and the design of scalable estimators based on sparse approximation recovery algorithms. Simulation results demonstrate significant cost savings by performing adaptation (50%, for a target 5% detection error) and by exploiting the temporal correlation in the process (90% cost saving).

7. REFERENCES

[1] K. Romer and F. Mattern, “The design space of wireless sensor networks,” *IEEE Wireless Communications*, vol. 11, no. 6, pp. 54–61, 2004.

[2] J. Meng, W. Yin, H. Li, E. Hossain, and Z. Han, “Collaborative Spectrum Sensing from Sparse Observations in Cognitive Radio Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 327–337, 2011.

[3] N. Michelusi and U. Mitra, “Fusion Center Feedback for Quasi-Decentralized Estimation in Sensor Networks,” in *51th Annual Allerton Conference on Communication, Control, and Computing*, 2013.

[4] ——, “Distributed estimation in sensor networks with quality feedback: A general framework,” in *IEEE GlobalSIP Symposium on New Sensing and Statistical Inference Methods*, Austin, TX, USA, Dec. 2013.

[5] D. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific, Belmont, Massachusetts, 2005.

[6] W. P. Tay, “The Value of Feedback in Decentralized Detection,” *IEEE Transactions on Information Theory*, vol. 58, no. 12, pp. 7226–7239, 2012.

[7] O. Kreidl, J. Tsitsiklis, and S. Zoumpoulis, “On Decentralized Detection With Partial Information Sharing Among Sensors,” *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1759–1765, 2011.

[8] M. Huang and S. Dey, “Dynamic Quantizer Design for Hidden Markov State Estimation Via Multiple Sensors With Fusion Center Feedback,” *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2887–2896, 2006.

[9] A. Dogandzic and K. Qiu, “Decentralized Random-Field Estimation for Sensor Networks Using Quantized Spatially Correlated Data and Fusion-Center Feedback,” *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 6069–6085, 2008.

[10] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[11] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[12] J. Ziniel and P. Schniter, “Dynamic Compressive Sensing of Time-Varying Signals Via Approximate Message Passing,” *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5270–5284, 2013.

[13] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, “Distributed Compressive Sensing,” *CoRR*, vol. abs/0901.3403, 2009.

[14] J. Mota, J. Xavier, P. Aguiar, and M. Puschel, “Distributed Basis Pursuit,” *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1942–1956, 2012.

[15] M. Duarte and R. Baraniuk, “Kronecker Compressive Sensing,” *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, Feb 2012.

[16] M. Stojnic, “Recovery thresholds for ℓ_1 optimization in binary compressed sensing,” in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2010, pp. 1593–1597.

[17] U. Nakarmi and N. Rahnavard, “BCS: Compressive sensing for binary sparse signals,” in *Military Communications Conference (MILCOM)*, 2012, pp. 1–5.

[18] M. Levorato, S. Narang, U. Mitra, and A. Ortega, “Reduced dimension policy iteration for wireless network control via multiscale analysis,” in *IEEE Global Communications Conference (Globecom)*, 2012, pp. 3886–3892.