Decentralized Chernoff Test in Sensor Networks

Anshuka Rangi and Massimo Franceschetti
University of California, San Diego
{arangi, massimo}@ucsd.edu

Stefano Marano
University of Salerno, Italy
marano@unisa.it

***Abstract*—We propose a decentralized, sequential and adaptive hypothesis test in sensor networks, which extends Chernoff's test to a decentralized setting. We show that the proposed test achieves the same asymptotic optimality of the original one, minimizing the expected cost required to reach a decision plus the expected cost of making a wrong decision, when the observation cost per unit time tends to zero. We also show that the proposed test is parsimonious in terms of communications. Namely, in the regime of vanishing observation cost per unit time, the expected number of channel uses required by each sensor to complete the test converges to four.**

***Index Terms*—Hypothesis Testing, Chernoff Test, KL-Divergence, Sensor Network**

I. Introduction

Inference systems based on sensor networks are attractive for a variety of reasons, including the increasingly low cost of the sensors, the inherent redundancy provided by the distributed structure of the network, the embedded computational capabilities of the sensors, and the availability of high-speed wireless communication channels [1]. When the network is used for detection, a set of hypotheses is tested based on the observations collected at the remote sensors, and an action is taken based on the results of these tests. Applications that fall in this framework include intrusion and target detection, and object classification and recognition [2]–[6].

One possible strategy to perform a statistical test using a sensor network is to send all observations from the sensors to a central processor, where the inference task is performed. Alternatively, in a decentralized setting, some preliminary processing can be performed at the sensors, and only a limited amount of pre-processed data is communicated to the central processor. This reduces the communication overhead, but may lead to sub-optimal performance. A natural question is what kind of local processing to perform at the sensor nodes, and what fusion scheme to adopt at the central processor, in order to reduce the communication burden while keeping a high level of detection performance. In this work, we address this question by proposing a decentralized statistical test in sensor networks that is optimal in terms of detection performance, while being parsimonious in terms of communication.

Hypothesis testing techniques are broadly classified as sequential or non-sequential tests, and adaptive or non-adaptive tests. Our focus is on a sequential and adaptive test. In a sequential test the number of observations needed to make a decision is not fixed in advance, but depends on the specific realization of the observed data. The test proceeds to collect and process data until a decision with a prescribed level of reliability can be made, and an important performance figure — in addition to the probability of correct decision — is the average number of observations required to end the test. In an adaptive test, the sensors' probing actions are chosen on the basis of the collected data in a causal manner. Hence, the sensors learn from the past, and adapt their future probing actions in a closed loop fashion.

Sequential tests were first introduced by Wald in [7]. One of such tests, the Sequential Probability Ratio Test (SPRT), was established to be asymptotically optimal for binary hypothesis testing in [8]. The asymptotic optimality of SPRT was extended to multi-hypothesis testing in [9], [10]. In the case of sequential and adaptive tests, Chernoff provided a test that is asymptotically optimal for binary hypotheses in his landmark paper [11]. Namely, as the observation cost per unit time vanishes, the test minimizes the sum of the expected cost required to reach a decision and the expected cost of making an incorrect decision. The asymptotic optimality of his test was extended to multi-hypothesis testing in [12]. The work in [13] discusses a specific application. The sequentiality and adaptivity gains for different tests were further studied in [14]–[16]. All of these results were established in a centralized setting.

Various works discuss extensions to a decentralized setting [17], [18]. Different techniques for combining the information at the central processor are considered in [1], [19]–[21]. In this context, asymptotically optimal sequential and non-adaptive tests have been developed [22], [23]. All of these results do not consider adaptive test, which are the main focus of our work.

We propose a Decentralized version of Chernoff's Test (DCT) for sensor networks that retains the asymptotic optimality of Chernoff's original solution. We provide an upper bound on the test performance in terms of expected risk. We also provide a matching converse, showing that any sequential test must achieve at least the same value of expected risk. Our solution is efficient in terms of communication overhead, compared to the trivial one where sensors blindly send all of their observations to the fusion center. We show that, as the observation cost per unit time vanishes, the expected number of times each sensor node uses the communication channel tends to four. Finally, discussing future work, we mention possible extensions of the test to a fully distributed scenario that does not require any fusion center operation.

The rest of the paper is organized as follows: Section II formulates the problem; Section III reviews the standard

Chernoff test and Section IV introduces its decentralized version; Section V informally describes the main idea behind the decentralized test and Section VI presents rigorous theoretical results. Section VII concludes the work.

II. Problem Formulation

We consider a sensor network composed of L sensors and one fusion center. The sensors and the fusion center can communicate with each other, while no direct mode of communication between the sensors is allowed. We consider M hypotheses and assume that only one of these hypotheses is true. At each time instant, each sensor can take a probing action, selected from a fixed set of actions $S = \{u_i\}_{i \in M}$, independently of all other sensors. For simplicity, we consider the cardinality of set S to be M, however, all results hold for the more general case as well. For $i, k \in [M]$, given that hypothesis h_i is true and probing action u_k is taken at sensor ℓ, we let $p_{i,\ell}^{u_k}$ denote the probability distribution of the observation received at the sensor ℓ following u_k. Given the true hypothesis, the observations received at any sensor are independent of the observations received at other sensors. Conversely, there may be a time-correlation between the observations at a given sensor, induced by the fact that the probing actions are observation-dependent.

Our performance measure – the risk – is analogous to the one considered in [11]. Given hypothesis h_i is true, the risk $\mathbb{R}_i^\delta$ of a sequential test δ is defined as

$$\mathbb{R}_i^\delta = c\,\mathbb{E}_i^\delta[N] + w_i\,\mathbb{P}_i^\delta(\hat{H} \neq h_i), \tag{1}$$

where $\mathbb{E}_i^\delta[N]$ is the conditional expected time required to reach a decision, c is the observation cost per unit time, $\hat{H}$ is the decision made, $\mathbb{P}_i^\delta(\hat{H} \neq h_i)$ is the conditional probability of wrong detection, and w_i is the cost of wrong detection. It follows that the risk corresponds to the sum of the expected cost required to reach a decision and the expected cost of making a wrong decision. Our objective is to design a test that minimizes the risk for all $i \in [M]$, as $c \to 0$.

We assume that observations corresponding to probing actions are instantly available at the sensors, the communication link between the sensor and the fusion center is noise free, and the information sent along this link is instantly available at the receiving end. The KL-divergence between the hypotheses is assumed to be finite for the entire action set S, namely, for all $\ell \in [L]$ and $i, j, k_1 \in [M]$, we have $D(p_{i,\ell}^{u_{k_1}} || p_{j,\ell}^{u_{k_1}}) < \infty$. Also, for all $\ell \in [L]$ and $i, j \in [M]$, there exists an action u_{k_1}, where $k_1 \in [M]$, such that $D(p_{i,\ell}^{u_{k_1}} || p_{j,\ell}^{u_{k_1}}) > 0$. This assumption entails little loss of generality, rules out trivialities, and is commonly adopted in the literature, see e.g. [11], [13]. Also, for all $\ell \in [L]$ and $i, j, k_1 \in [M]$, we assume $\mathbb{E}[\log(p_{i,\ell}^{u_{k_1}}(Y))/\log(p_{j,\ell}^{u_{k_1}}(Y))]^2 < \infty$.

III. Standard Chernoff Test

We start considering sensor ℓ alone, with no interactions with the fusion center or with other elements of the network. Chernoff test for this isolated sensor is as follows:

1) At step $k-1$, a temporary decision is made, based on the posterior probability of the hypotheses, given the past observations and actions. Specifically, the temporary decision is in favor of $h_{i^*_{k-1}}$ if

$$i^*_{k-1} = \arg\max_{i \in [M]} \mathbb{P}(H^* = h_i | y_\ell^{k-1}, u_\ell^{k-1}), \tag{2}$$

where H^* is the true hypothesis, $y_\ell^{k-1} = \{y_{1,\ell}, y_{2,\ell} \dots y_{k-1,\ell}\}$, $y_{i,\ell}$ is the observation at step i and sensor ℓ, $u_\ell^{k-1} = \{u_{1,\ell}, u_{2,\ell} \dots u_{k-1,\ell}\}$, and $u_{i,\ell}$ is the action at step i and sensor ℓ.

2) At step k, the action $u_{k,\ell}$ is randomly chosen among the elements of S, according to Probability Mass Function (PMF) $Q^\ell_{i^*_{k-1}}$, where:

$$Q^\ell_{i^*_{k-1}} = \arg\max_{q \in \mathcal{Q}} \min_{j \in [M] \setminus \{i^*_{k-1}\}} \sum_u q(u) D(p^u_{i^*_{k-1},\ell} || p^u_{j,\ell}),$$

in which $\mathcal{Q}$ denotes the set of all the possible PMFs over the alphabet $[M]$.

3) For all $i \in [M]$, update the posterior probability $\mathbb{P}(H^* = h_i | y_\ell^k, u_\ell^k)$.

4) The test stops at step N if the worst case log-likelihood ratio crosses a prescribed fixed threshold γ, i.e.,

$$\log \frac{p_{i^*_{N,\ell}}(y_\ell^N, u_\ell^N)}{\max_{j \neq i^*_N} p_{j,\ell}(y_\ell^N, u_\ell^N)} \geq \gamma, \tag{3}$$

where $p_{i^*_{N,\ell}}(y_\ell^N, u_\ell^N)$ is the posterior probability $\mathbb{P}(H^* = h_{i^*_N} | y_\ell^N, u_\ell^N)$ at sensor ℓ. If the test stops at step N, then the final decision is $h_{i^*_N}$. Otherwise, $k \leftarrow (k-1)$, and the procedures continues from 1).

IV. Decentralized Chernoff Test

As the observation cost per unit time tends to zero, the probability of wrong detection for the standard Chernoff test tends to zero [11]. It follows that minimizing the risk in (1) also corresponds to minimizing the expected number of samples required to reach a decision. When one sample is collected at each time step, minimizing the expected number of samples is obviously the same as minimizing the expected time for making a decision. However, this is not necessarily true in a decentralized setting.

To further illustrate this point, consider first minimizing the total expected number of samples collected by the L sensors to reach a decision, and assume that the amount of communication between sensors and fusion center is unconstrained. A straightforward design, which we call Fusion center based Chernoff Test (FCT), is as follows. The action set S is modified to S' with cardinality ML, where action $a_{i,\ell} \in S'$ corresponds to the selection of $u_i \in S$ and sensor $\ell \in [M]$. Then, a Chernoff test is performed on S' at the fusion center where the selection of $a_{i,\ell}$ corresponds to activating sensor ℓ, and enabling the activated sensor to use the probing action u_i in order to collect the corresponding observation, which is then delivered to the fusion center. It is not hard to see that, as the probability of wrong detection tends to zero, the FCT minimizes the total expected number of collected samples. The proof of this claim is similar to the one of the optimality of Chernoff test in [11] and is thus omitted.

The FCT also minimizes the total number of probing actions performed by the sensors. However, there is only one active sensor, out of L, per unit time, and all observations are communicated to the fusion center. Clearly, this is highly inefficient in terms of both communication overhead and decision time, and motivates introducing a different kind of test.

Our proposed DCT operates in two phases. In the initialization phase, each sensor ℓ sends a vector v_ℓ to the fusion center, where the elements of v_ℓ are, for all $i \in [M]$

$$v_{i,\ell} = \max_{q \in \mathcal{Q}} \min_{j \neq i} \sum_{u} q(u) D(p^u_{i,\ell} || p^u_{j,\ell}). \tag{4}$$

The quantity $v_{i,\ell}$ is a measure of the capability of sensor ℓ to detect hypothesis h_i (see [11] for a discussion), and plays a critical role in designing the test. After receiving v_ℓ from all sensors, the fusion center sends back to sensor ℓ a response vector ρ_ℓ, whose L entries are the scalars

$$\rho_{i,\ell} = v_{i,\ell}/I(i), \tag{5}$$

where $I(i) = \sum_{\ell=1}^{L} v_{i,\ell}$ is a measure of cumulative capability of the network to detect hypothesis h_i.

At this point, the test phase begins. All sensors perform a Chernoff test, independently of each other, consisting of steps 1-4 described in Section III, with an important difference: any time at sensor ℓ we have

$$\log \frac{p_{i_n^*,\ell}(y^n_\ell, u^n_\ell)}{\max_{j \neq i_n^*} p_{j,\ell}(y^n_\ell, u^n_\ell)} \geq \rho_{i_n^*,\ell} |\log c|, \tag{6}$$

then a local decision in favor of $h_{i_n^*}$ is communicated to the fusion center. This is not a stopping criterion for the test at sensor ℓ, but only a triggering condition for the communication between sensor ℓ and the fusion center. Thus, sensor ℓ continues to run the test until the fusion center sends a halting message.

The final decision $\hat{H}$ is made at the fusion center in favor of hypothesis h_i when the local decisions from all the sensors are in favor of h_i. After the final decision is made, the fusion center sends a halting message to all the sensors.

Apart from the initialization phase, the proposed DCT only requires the communication of an index $\in [M]$ during the test phase. Thus, the communication resources required are considerably less compared to the FCT, where continuous random variables are sent over the network at each step. In addition, our results show that, while maintaining the same asymptotic optimality of Chernoff's test as $c \to 0$, the oscillations in the local decisions at the sensors vanish, and each sensor tends to use the communication channel on average only four times: two in the initialization phase, one to communicate the local decision, and one to receive the halting message.

V. Informal Discussion

The key idea behind the proposed DCT is to determine the individual capabilities of the sensors for detecting the hypotheses. These capabilities — that depend on the true hypothesis H^* — are captured by the vector v_ℓ, whose i^{th} element is a measure of sensor capability to detect the hypothesis h_i. The fusion center gathers this information, and utilizes it to control the threshold at each sensor through the response vector $\rho_{i,\ell}$. At the fusion center, $I(i)$ is the measure of the cumulative detection capability of the network for hypothesis h_i, and $\rho_{i,\ell}$ denotes the fraction of this capability contributed by sensor ℓ for hypothesis h_i. To minimize the expected time to reach a decision, it is desirable to determine the threshold for each sensor ℓ such that all the sensors require roughly the same time to reach the triggering condition. This is analogous to dividing the task of hypothesis testing among the sensors based on their speed of performing the task, such that all the sensors finish their share of the task at roughly the same time.

VI. Theoretical Results

In the following theorems, N indicates the time required to make a decision, and C indicates the communication overhead, namely the number of times a sensor communicates with the fusion center. The superscripts $\mathcal{C}$ and δ refer to the DCT and to a generic decentralized sequential test, respectively.

Part (i) of Theorem 1 states that the probability of making a wrong decision can be made as small as desired by an appropriate choice of c. Part (ii) provides a bound on the expected time to reach the final decision, and part (iii) bounds the risk as an immediate consequence of parts (i) and (ii).

Theorem 1: (Direct). The following statements hold:
(i) For all $c \in (0, 1)$ and for all $i \in [M]$, given that hypothesis h_i is true, the probability that the DCT makes an incorrect decision is bounded as

$$\mathbb{P}^{\mathcal{C}}_i(\hat{H} \neq h_i) \leq \min\{(M-1)c, 1\}.$$

(ii) For all $i \in [M]$, given that hypothesis h_i is true, the expected decision time is

$$\mathbb{E}^{\mathcal{C}}_i[N] \leq (1 + o(1)) \frac{|\log c|}{I(i)}, \quad \text{as } c \to 0. \tag{7}$$

(iii) Combining (i) and (ii), the risk defined in (1) verifies

$$\mathbb{R}^{\mathcal{C}}_i \leq (1 + o(1)) \frac{c\,|\log c|}{I(i)}, \quad \text{as } c \to 0. \tag{8}$$

The following theorem provides a matching converse result.

Theorem 2: (Converse). For any sequential test δ, if for all $i \in [M]$ the probability of missed detection satisfies

$$\mathbb{P}^{\delta}_i(\hat{H} \neq h_i) = O(c\,|\log c|), \quad \text{as } c \to 0, \tag{9}$$

then we have

$$\mathbb{E}^{\delta}_i[N] \quad \geq \quad (1 + o(1)) \frac{|\log c|}{I(i)}, \tag{10}$$

$$\mathbb{R}^{\delta}_i \quad \geq \quad (1 + o(1)) \frac{c\,|\log c|}{I(i)}, \quad \text{as } c \to 0. \tag{11}$$

The following result is a consequence of Theorems 1 and 2. It shows the asymptotic optimality of the DCT, and presents the expected communication overhead, as $c \to 0$.

Theorem 3: For the DCT , for all $i \in [M]$ we have

$$\mathbb{E}_i^{\mathcal{C}}[N] = (1+o(1))\frac{|\log c|}{I(i)}, \tag{12}$$

$$\mathbb{R}_i^{\mathcal{C}} = (1+o(1))\frac{c\,|\log c|}{I(i)}, \quad \text{as } c \to 0, \tag{13}$$

$$\lim_{c\to 0} \mathbb{E}_i^{\mathcal{C}}[C] = 4. \tag{14}$$

To illustrate the proofs, we need the following additional notation. We let $y_\ell^k = \{y_{1,\ell}, y_{2,\ell} \ldots y_{k,\ell}\}$, where $y_{i,\ell}$ is the i^{th} observation sample at sensor ℓ; $u_\ell^k = \{u_{1,\ell}, u_{2,\ell} \ldots u_{k,\ell}\}$, where $u_{i,\ell}$ is the i^{th} action at sensor ℓ. We also let $A_{n,j}$ be the set of sample paths where the decision by the fusion center is made in favor of h_j at the n^{th} step, and we indicate a single sample path as $\{(u_1^n, y_1^n) \ldots (u_L^n, y_L^n)\}$. We indicate by $A_{n,j,\ell}$ the set of sample paths in $A_{n,j}$ corresponding to the ℓ^{th} sensor. Finally, we define

$$N_{i,\ell} = \inf\left\{ n : \sum_{k=1}^{n} \log \frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{\max_{j\neq i} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})} \geq \rho_{i,\ell}\,|\log c| \right\}.$$

Proof of Theorem 1: The proof consists of two parts. First, we write $\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i)$ as the probability of a countable union of disjoint sets of sample paths. An upper bound on this probability then follows from an upper bound on the probability of these disjoint sets, in conjunction with the union bound. Second, we upper bound $\mathbb{E}_i^{\mathcal{C}}[N]$ by the sum of the expected time required to reach the triggering condition (6) for hypothesis h_i, and the expected delay between the time of triggering and the final decision is taken in favor of hypothesis h_i at the fusion center. We then show that these expectations are the same at all sensors, so that (7) follows.

Consider the probability $\mathbb{P^C}_i(\hat{H} = h_j)$. This is same as the probability of the countable union of disjoint sets $A_{n,j}$. Thus, for $j \neq i$, we can write

$$\begin{aligned}
&\mathbb{P}_i^{\mathcal{C}}(A_{n,j}) \\
&= \int_{A_{n,j}} \prod_{\ell=1}^{L}\prod_{k=1}^{n} p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \ldots dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(a)}{=} \prod_{\ell=1}^{L}\int_{A_{n,j,\ell}} \prod_{k=1}^{n} p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) ..... dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(b)}{\leq} \prod_{\ell=1}^{L}\int_{A_{n,j,\ell}} c^{\rho_{j,\ell}} \prod_{k=1}^{n} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \ldots dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(c)}{=} c\prod_{\ell=1}^{L}\int_{A_{n,j,\ell}} \prod_{k=1}^{n} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \ldots dy_{n,\ell}(u_{n,\ell}) \\
&= c\prod_{\ell=1}^{L} \mathbb{P}_j^{\mathcal{C}}(\hat{H} = h_j \text{ at sample } n \text{ at } \ell^{th} \text{ sensor}) \\
&= c\,\mathbb{P}_j^{\mathcal{C}}(\hat{H} = h_j \text{ at sample } n),
\end{aligned} \tag{15}$$

where (a) follows from the definition of $A_{n,j,\ell}$; (b) follows from the definition of $N_{i,\ell}$; (c) follows from $\sum_{\ell=1}^{L} \rho_{j,\ell} = 1$. Now, we can bound $\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i)$ as follows

$$\begin{aligned}
&\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i) = \sum_{j\neq i} \mathbb{P}_i^{\mathcal{C}}(\hat{H} = h_j) = \sum_{j\neq i}\sum_{n=1}^{\infty} \mathbb{P}_i^{\mathcal{C}}(A_{n,j}) \\
&\leq \sum_{j\neq i}\sum_{n=1}^{\infty} c\,\mathbb{P}_j^{\mathcal{C}}(\hat{H} = h_j \text{ at sample } n) \\
&= \sum_{j\neq i} c\,\mathbb{P}_j^{\mathcal{C}}(\hat{H} = h_j) \leq c\,(M-1),
\end{aligned} \tag{16}$$

where the first inequality of the chain follows by (15). This proves part (i) of the theorem.

Let us now define

$$\tau(N_{i,\ell}) = \sup\left\{ n \ : \sum_{k=N_{i,\ell}+1}^{N_{i,\ell}+n} \log \frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{\max_{j\neq i} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})} \leq 0 \right\}.$$

The triggering condition (6) at the ℓ^{th} sensor is satisfied for all $n > N_{i,\ell} + \tau(N_{i,\ell})$, yielding

$$N \leq \max_{1\leq \ell \leq L}(N_{i,\ell} + \tau(N_{i,\ell}) + 1) \leq \max_{1\leq \ell\leq L} N_{i,\ell} + \sum_{\ell=1}^{L} \tau(N_{i,\ell}) + 1.$$

Taking the expectation of both sides, we get

$$\mathbb{E}_i^{\mathcal{C}}[N] \leq \mathbb{E}_i\left[\max_{1\leq\ell\leq L} N_{i,\ell}\right] + \sum_{\ell=1}^{L} \mathbb{E}[\tau(N_{i,\ell})] + 1. \tag{17}$$

We now bound the terms on the right-hand side of (17). As each sensor performs a Chernoff test individually, using [11, Lemma 2] we have, as $c \to 0$

$$\mathbb{E}_i[N_{i,\ell}] = (1+o(1))|\log c|/I(i), \tag{18}$$

which is independent of ℓ. Additionally, from [24, eq. (19)] we have, as $c \to 0$

$$\mathrm{Var}(N_{i,\ell}) = O(|\log c|).$$

Hence, as $c \to 0$, and for all $\ell \in [L]$, we have

$$\begin{aligned}
&\left(\mathbb{E}_i\left|N_{i,\ell} - (1+o(1))\frac{|\log c|}{I(i)}\right|\right)^2 \\
&\qquad \leq \mathbb{E}_i\left(N_{i,\ell} - (1+o(1))\frac{|\log c|}{I(i)}\right)^2 \\
&\qquad = \mathrm{Var}(N_{i,\ell}) = O(|\log c|).
\end{aligned} \tag{19}$$

The above yields,

$$\begin{aligned}
&\mathbb{E}_i\left[\max_{1\leq\ell\leq L} N_{i,\ell}\right] \\
&= \frac{|\log c|}{I(i)}(1+o(1)) + \mathbb{E}_i\left[\max_{1\leq\ell\leq L} N_{i,\ell} - (1+o(1))\frac{|\log c|}{I(i)}\right] \\
&\leq \frac{|\log c|}{I(i)}(1+o(1)) + \sum_{\ell=1}^{L}\mathbb{E}_i\left|N_{i,\ell} - (1+o(1))\frac{|\log c|}{I(i)}\right| \\
&= \frac{|\log c|}{I(i)}(1+o(1)) + O(\sqrt{|\log c|}),
\end{aligned} \tag{20}$$

where the inequality follows by $\max_\ell N_{i,\ell} \leq \sum_\ell |N_{i,\ell}|$, and the last equality follows by (19). The term $\mathbb{E}[\tau(N_{i,\ell})]$ on the right-hand side of (17) is finite since $\mathbb{E}[\log(p_{i,\ell}^{u_{i,\ell}}(y_{k,\ell})/\max_{j\neq i} p_{j,\ell}^{u_{i,\ell}}(y_{k,\ell}))]$ is the KL divergence between the two probability measures, which is positive and finite (see [25]). Thus, combining equation (17), (20) and the finiteness of $\mathbb{E}_i[\tau(N_{i,\ell})]$, as $c \to 0$ we get (7). ■

The proofs of Theorem 2 and 3 are omitted for space reasons, but they are available to the reader in [25].

VII. Conclusions and Future Work

We proposed a DCT which is parsimonious in terms of communications, and is asymptotically optimal in terms of detection performance, when the observation cost per unit time vanishes.

One major advantage of sensor networks is their robustness to node failures and external attacks. From this viewpoint, although we have presented our results assuming the presence of a fusion center, alternative solutions where all the information processing is completely distributed and there is no central unit, are certainly desirable. Our design could also be implemented in a fully distributed architecture. The key quantity $I(i)$ computed in the initialization phase can be obtained by *gossip* protocols using *consensus* techniques [26]–[28]. Similarly, once all the sensors reach the triggering condition (6), the final decision can also be easily computed in a distributed way.

Unlike the classic Chernoff test for isolated sensors, the proposed DCT has two critical times: the time required to reach the triggering condition in (6), and the delay between the time of triggering and the final decision at the fusion center. As in other sequential tests, unexpected long runs can occur in our setting as well, when these two times significantly deviate from their average. In the first case, one or more "outlier" sensors can take unusually long time to reach the local decision, and the remaining sensors would need to keep sending their decisions to the fusion center until the outliers have also reached their local decisions. This situation can lead to unusually large communication overhead, and can be triggered by even a single sensor. In the second case, when c is not sufficiently close to 0, sensors can reach incorrect local decisions which are likely to be different. This would also result in additional communication overhead until the time all sensors have agreed upon a single hypothesis. We plan to study these effects in our future work.

References

[1] P. K. Varshney, *Distributed detection and data fusion*. Springer Science & Business Media, 2012.

[2] Y. Hu, M. Dong, K. Ota, A. Liu, and M. Guo, "Mobile target detection in wireless sensor networks with adjustable sensing frequency," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1160–1171, 2016.

[3] Y. Liu, M. Dong, K. Ota, and A. Liu, "Activetrust: secure and trustable routing in wireless sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 2013–2027, 2016.

[4] J. Aslam, Z. Butler, F. Constantin, V. Crespi, G. Cybenko, and D. Rus, "Tracking a moving object with a binary sensor network," in *Proceedings of the 1st international conference on Embedded networked sensor systems*. ACM, 2003, pp. 150–161.

[5] P. Kulkarni, D. Ganesan, P. Shenoy, and Q. Lu, "Senseye: a multi-tier camera sensor network," in *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005, pp. 229–238.

[6] S. Ren, Q. Li, H. Wang, X. Chen, and X. Zhang, "Design and analysis of sensing scheduling algorithms under partial coverage for object detection in sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 3, pp. 334–350, 2007.

[7] A. Wald, *Sequential analysis*. Courier Corporation, 1973.

[8] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *The Annals of Mathematical Statistics*, pp. 326–339, 1948.

[9] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests. ii. accurate asymptotic expansions for the expected sample size," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1366–1383, 2000.

[10] V. Draglia, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests. i. asymptotic optimality," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2448–2461, 1999.

[11] H. Chernoff, "Sequential design of experiments," *The Annals of Mathematical Statistics*, vol. 30, no. 3, pp. 755–770, 1959.

[12] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2451–2464, 2013.

[13] M. Franceschetti, S. Marano, and V. Matta, "Chernoff test for strong-or-weak radar models," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 289–302, 2016.

[14] M. Naghshvar and T. Javidi, "Sequentiality and adaptivity gains in active hypothesis testing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 768–782, 2013.

[15] M. Naghshvar, T. Javidi *et al.*, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, 2013.

[16] N. A. Goodman, P. R. Venkata, and M. A. Neifeld, "Adaptive waveform design and sequential hypothesis testing for target recognition with active sensors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 105–113, 2007.

[17] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors ii. advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, 1997.

[18] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors part i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[19] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, no. 1, pp. 98–101, 1986.

[20] R. Jiang and B. Chen, "Fusion of censored decisions in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2668–2673, 2005.

[21] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 2, pp. 475–488, 2005.

[22] Y. Mei, "Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2072–2089, 2008.

[23] Y. Wang and Y. Mei, "Asymptotic optimality theory for decentralized sequential multihypothesis testing problems," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 7068–7083, 2011.

[24] D. Siegmund, "The variance of one-sided stopping rules," *The Annals of Mathematical Statistics*, vol. 40, no. 3, pp. 1074–1077, 1969.

[25] A. Rangi, M. Franceschetti, and S. Marano, "https://www.dropbox.com/s/6ktst33kqu9zzs3/decentralized chernofftestforsensornetworks.pdf."

[26] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, no. SI, pp. 2508–2530, 2006.

[27] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3375–3380, 2008.

[28] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypotheses," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 814–825, 2010.