Decentralized Sequential Hypothesis Testing Using Asynchronous Communication

Georgios Fellouris and George V. Moustakides, *Senior Member, IEEE*

***Abstract*—An asymptotically optimum test for the problem of decentralized sequential hypothesis testing is presented. The induced communication between sensors and fusion center is asynchronous and limited to 1-bit data. When the sensors observe continuously stochastic processes with continuous paths, the proposed test is *order-2* asymptotically optimal, in the sense that its inflicted performance loss is bounded. When the sensors take discrete time observations, the proposed test achieves *order-1* asymptotic optimality, i.e., the ratio of its performance over the optimal performance tends to 1. Moreover, we show theoretically and corroborate with simulations that the performance of the suggested test in discrete time can be significantly improved when the sensors sample their underlying continuous time processes more frequently, a property which is not enjoyed by other centralized or decentralized tests in the literature.**

***Index Terms*—Decentralized detection, sequential hypothesis testing, Sequential Probability Ratio Test (SPRT).**

I. Introduction

SEQUENTIAL hypothesis testing, first introduced by Wald [1], is one of the most classical and well-studied problems of sequential analysis, with applications in areas such as industrial quality control, signal detection, design of clinical trials, etc [2], [3]. In the last two decades, there has been an intense interest in the decentralized formulation of the problem [4]–[13]. In this setup, the sequentially acquired information for decision making is distributed across a number of sensors and is transmitted to a global decision maker (fusion center), which is responsible for making the final decision. The main difference in the decentralized version of the problem is that the sensors are required to *quantize* their observations before transmitting them to the fusion center, in other words, the sensors must send to the fusion center messages that belong to a *finite alphabet* [4]. This requirement is imposed by the need for data compression, smaller communication bandwidth and robustness of the sensor network, which are crucial issues in application areas such as signal processing, mobile and wireless communication, multisensor data fusion, internet security, robot networks and others. For a review of decentralized nonsequential testing we refer to [5].

Manuscript received August 30, 2009; revised August 07, 2010; accepted August 18, 2010. Date of current version December 27, 2010. This work was supported (in part) by the AFOSR under Grant FA9550-08-1-0376.

G. Fellouris is with the Statistics Department, Columbia University, New York, NY 10027 USA (e-mail: gaf2106@columbia.edu).

G.V. Moustakides is with the Department of Electrical and Computer Engineering, University of Patras, 26500 Rion, Greece (e-mail: moustaki@upatras.gr).

Communicated by M. Lops, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2010.2090249

Depending on the *local memory* that the sensors possess and whether there exists *feedback* from the fusion center, Veeravalli *et al.* [6] proposed five different configurations for the sensor network. In the same work, the authors found the optimal decentralized sequential test, under a Bayesian setting, in the case of full feedback and local memory restricted to past decisions. Also under a Bayesian setting, the case of no feedback and no local memory was treated in [7] and the case of full local memory with no feedback in [8], [9]. However, in the last two cases no exactly optimal decentralized sequential test has been discovered. For a review of decentralized sequential testing we refer to [10].

Our work here differs from other articles in the literature of decentralized detection in a number of ways. First of all, we introduce the configuration of *partial* local memory, thus we assume that at each time instant each sensor has access to the value of a summary statistic—that summarizes its previous observations—and uses this value, together with its current observation, in order to send a quantized signal to the fusion center. Under this configuration, an asymptotically optimal scheme was suggested by Mei in [11] under a Bayesian setting.

Moreover, instead of restricting ourselves to a discrete time framework, we also consider a continuous time setup. The latter is of course an idealization, since the sensors in practice cannot record their observations continuously; however, it allows us to isolate the performance loss due to discrete sampling at the sensors and provides us with insight that leads to more efficient schemes in discrete time.

Finally, unlike most decentralized schemes that require synchronous communication of the sensors with the fusion center, we suggest that the sensors transmit their messages asynchronously and at random times (see [12] for a different asynchronous scheme). In particular, we suggest that the communication times of sensor i should be *stopping times* with respect to the observed filtration at sensor i. We call this type of communication *adapted*.

A special case of adapted communication is *Lebesgue* (or *level-triggered*) sampling, which induces naturally a *1-bit* communication between sensors and fusion center. Lebesgue sampling, combined with a Sequential Probability Ratio Test at the fusion center, gives rise to a detection structure which is known as Decentralized Sequential Probability Ratio Test (D-SPRT) and was introduced by Hussain [13] in a discrete time context. However, Hussain did not provide any theoretical support for this test.

Our main contributions are that we define D-SPRT in more detail and prove its asymptotic optimality both in discrete and continuous time. More specifically, we prove that if the sensors observe continuously the paths of independent Itô processes, D-SPRT is *order-2* asymptotically optimal, since its associated performance loss remains bounded. When the sensors take discrete time (independent and identically distributed) observations, we prove that D-SPRT achieves *order-1* asymptotic optimality, i.e., the ratio of its performance over the optimal centralized performance tends to 1 as long as the communication between sensors and fusion center is *not very frequent, but reasonably infrequent*. Moreover, we show that the performance of the discrete time D-SPRT can be significantly improved, approaching the performance of the continuous time scheme, if the sensors sample their underlying continuous time processes more frequently. Finally, we present simulation experiments which corroborate our theoretical findings and show that the continuous time D-SPRT is more efficient than the discrete time centralized SPRT, whereas the discrete time D-SPRT is more efficient than the decentralized scheme presented in [11].

This paper is organized as follows: Section I contains the Introduction. In Section II, we formulate the sequential hypothesis testing problem for the discrete and continuous time case under a centralized and decentralized setup. Moreover, we introduce the concept of adapted sampling and emphasize on Lebesgue sampling and D-SPRT. In Section III, we recall the main optimality results for the centralized formulation, since these tests serve as a point of reference for their decentralized counterparts. Section IV presents the asymptotic optimality properties of D-SPRT in the context of continuous time and continuous path observations, while in Section V, we develop the analogous results, at the expense of a more involved analysis, for the discrete time case. In this section we also examine the notion of oversampling, which reconciles the behavior of D-SPRT in discrete and continuous time and provides some important design guidelines. Finally, in Section VI, we conclude and discuss generalizations of our work.

II. Centralized Versus Decentralized Sequential Testing

Suppose we have a sensor network consisting of K sensors as depicted in Fig. 1. Each sensor i observes *sequentially* a realization of a stochastic process $\{\xi_t^i\}_{t\geq 0}$ with distribution P^i. We assume that the processes $\{\xi_t^1\},\dots,\{\xi_t^K\}$ are independent and we denote by $\{\mathscr{F}_t^i\}_{t\geq 0}$ the filtration generated by $\{\xi_t^i\}_{t\geq 0}$. We also denote with P the probability measure of $\{(\xi_t^1,\dots,\xi_t^K)\}_{t\geq 0}$ and by $\{\mathscr{F}_t^i\}_{t\geq 0}$ the filtration generated by this vector process. From the assumption of independence across sensors, we have $\mathsf{P}=\mathsf{P}^1\times\cdots\times\mathsf{P}^K$.

Consider now the following two hypotheses for the probability measure P

$$\mathsf{H}_0:\mathsf{P}=\mathsf{P}_0;\quad \mathsf{H}_1:\mathsf{P}=\mathsf{P}_1 \tag{1}$$

where $\mathsf{P}_j=\mathsf{P}_j^1\times\cdots\times\mathsf{P}_j^K, j=0,1$, and $\mathsf{P}_j^i, j=0,1; i=1,\dots,K$ are known probability measures. Thus, $\mathsf{H}_0,\mathsf{H}_1$ are two simple hypotheses. For simplicity, we assume that the measures $\mathsf{P}_0^i,\mathsf{P}_1^i$ are locally equivalent, therefore we can define the local log-likelihood ratio process at each sensor i and for each time instant $t\in[0,\infty)$ as

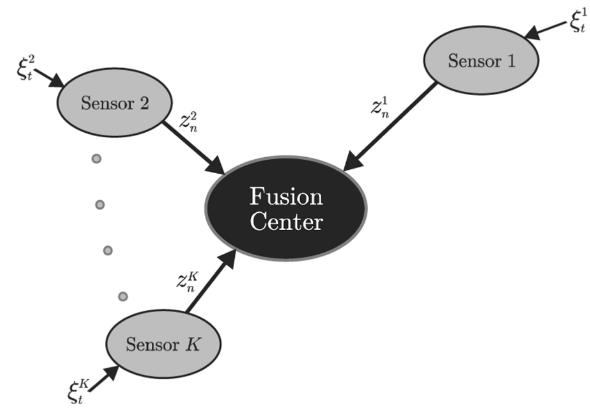


Fig. 1. Schematic representation of a decentralized sensor network.

$$u_t^i=\log\frac{d\mathsf{P}_1^i}{d\mathsf{P}_0^i}\left(\mathscr{F}_t^i\right);\quad u_0^i=0. \tag{2}$$

Moreover, due to the independence of observations across sensors, we can write the global log-likelihood ratio $\{u_t\}$ in the sensor network as the sum of its local counterparts, i.e.,

$$u_t=\log\frac{d\mathsf{P}_1}{d\mathsf{P}_0}(\mathscr{F}_t)=\sum_{i=1}^K u_t^i,\quad 0\leq t<\infty. \tag{3}$$

Since the sensors can communicate information to the fusion center only at a sequence of discrete times, we assume that the fusion center receives sequentially from each sensor i the data $\{z_n^i\}$ at a strictly increasing sequence of time instants $\{\tau_n^i\}_{n\in\mathbb{N}}$. Each τ_n^i is an $\{\mathscr{F}_t^i\}$-adapted stopping time with $\tau_0^i=0$ and $\mathsf{P}_j(\tau_n^i<\infty)=1,\forall n\in\mathbb{N}, j=0,1$ and $i=1,\dots,K$. We call this communication scheme *adapted sampling (or communication)* and we refer to the stopping times $\{\tau_n^i\}$ as the *sampling (or communication) times* in sensor i. Each z_n^i constitutes a summary of the acquired information $\{\mathscr{F}_t^i\}_{\tau_n^i}^i$ up to time τ_n^i and, as we mentioned in the Introduction, it takes values in a finite alphabet. Here, we are going to assume that this set is *binary*. We should also emphasize that we do not consider any feedback scheme from the fusion center towards the sensors.

Adapted sampling clearly implies *asynchronous* communication between the sensors and the fusion center at *random* time instants. Thus, the number of messages sent from sensor i to the fusion center up to (and including) any time instant t is in general random, different for each sensor and will be denoted by m_t^i.

We should mention that adapted sampling is a general framework that can incorporate various sampling mechanisms already used in the literature, in particular:

- When $\tau_n^i-\tau_{n-1}^i=h,\forall n\in\mathbb{N}$, adapted sampling reduces to *canonical deterministic sampling* with constant sampling period $h>0$, common to all sensors.
- When $\{\tau_n^i-\tau_{n-1}^i\}_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables, *independent* of the observation process $\{\xi_t^i\}$, adapted sampling becomes *independent random sampling*, as long as we properly enlarge the local filtration at each sensor. Notice that if the intersampling periods

$\{\tau_n^i - \tau_{n-1}^i\}_{n\in\mathbb{N}}$ are independent and exponentially distributed with the same mean, we recover the sampling scheme suggested in [12].

- When the sampling times depend on the observed sequence and are given by the following recursion:

$$\tau_n^i = \inf\left\{t > \tau_{n-1}^i : u_t^i - u_{\tau_{n-1}^i}^i \notin (-\underline{\Delta}_i, \bar{\Delta}_i)\right\} \quad (4)$$

where $\underline{\Delta}_i, \bar{\Delta}_i > 0$ are proper thresholds, then we call the resulting scheme *Lebesgue* (or *level-triggered*) sampling.

Although not evident at first, we should emphasize that the fusion center is the recipient not only of the data sequences $\{z_n^i\}$ but also of the sampling times $\{\tau_n^i\}$, that may carry information relevant to the hypothesis testing problem.

In parallel to the communication activity, the fusion center uses the received data in order to decide whether to continue or stop receiving additional observations from the sensors. In the latter case it proceeds to make a final decision between the two hypotheses.

Under a *decentralized setup*, the fusion center has access to the filtration $\{\mathscr{G}_t\}_{t\geq 0}$, where $\mathscr{G}_t = \sigma\{(z_n^i, \tau_n^i), \tau_n^i \leq t; i = 1,\dots,K\}$ is the σ-algebra generated by all pairs (z_n^i, τ_n^i) received up to time t. Thus, the fusion center must use an $\{\mathscr{G}_t\}$-adapted stopping time T to decide about stopping or continuing sampling and, after stopping, an $\{\mathscr{G}_t\}$-measurable decision function $d_T \in \{0,1\}$ to select one of the two hypotheses.

Under the *centralized setup*, at any time t the fusion center gains access to the *entire* information acquired by the sensors up to this time, which is described by the σ-algebra $\{\mathscr{F}_t\} = \sigma\{\xi_s^i, 0 < s \leq t; i = 1,\dots,K\}$. Thus, the fusion center can use an $\{\mathscr{F}_t\}$-adapted stopping time T and an $\{\mathscr{F}_T\}$-measurable decision function $d_T \in \{0,1\}$ to stop sampling and provide a decision between the two hypotheses.

Our intention is to define the pair (T, d_T) optimally under both the centralized and the decentralized formulation. Following Wald [1], for any $\alpha, \beta > 0$ we define the class of sequential tests for which the type-I and type-II error probabilities are below the two levels α, β, respectively, that is

$$\mathscr{C}_{\alpha,\beta} = \{(T, d_T) : \mathsf{P}_0(d_T = 1) \leq \alpha \text{ and } \mathsf{P}_1(d_T = 0) \leq \beta\}. \quad (5)$$

We can now define the following constrained optimization problem.

Problem 1: Given $\alpha, \beta > 0$ such that $\alpha + \beta < 1$, find a sequential test $(\mathscr{T}, d_T) \in \mathscr{C}_{\alpha,\beta}$ so that

$$\mathsf{E}_j[\mathscr{T}] = \inf_{(T,d_T)\in\mathscr{C}_{\alpha,\beta}} \mathsf{E}_j[T], \quad j = 0,1, \quad (6)$$

where we denote with $\mathsf{E}_j[\cdot]$ the expectation under hypothesis $\mathsf{H}_j, j = 0,1$. If we seek the test among the $\{\mathscr{F}_t\}$-adapted schemes, we refer to the optimum centralized version, whereas if we limit ourselves to $\{\mathscr{G}_t\}$-adapted sequential tests, then we obtain the optimum decentralized procedure. Note that we attempt to find a *single* test that *simultaneously* minimizes two different criteria (the expected decision delay under the two hypotheses). It was Wald's remarkable insight that led him first to conjecture [1], and then prove [14], that a test with such an extraordinary optimality property indeed exists.

Let us also introduce a second problem, proposed by Liptser and Shiryaev [15], which constitutes a slight variant of Problem 1.

Problem 2: Given $\alpha, \beta > 0$ such that $\alpha + \beta < 1$, find a sequential test $(\mathscr{T}, d_{\mathscr{T}}) \in \mathscr{C}_{\alpha,\beta}$, so that

$$\begin{aligned} -\mathsf{E}_0[u_{\mathscr{T}}] &= \inf_{(T,d_T)\in\mathscr{C}_{\alpha,\beta}} (-\mathsf{E}_0[u_T]) \\ \mathsf{E}_1[u_{\mathscr{T}}] &= \inf_{(T,d_T)\in\mathscr{C}_{\alpha,\beta}} \mathsf{E}_1[u_T]. \end{aligned} \quad (7)$$

Recalling that $\{u_t\}$ is the running log-likelihood ratio of the two probability measures, it is clear that the two expectations $\mathsf{E}_1[u_t]$ and $-\mathsf{E}_0[u_t]$ give rise to nonnegative and increasing functions of time. These two time functions constitute, in Information Theory, a popular divergence measure known as the Kullback-Leibler (K-L) divergence. This interesting information theoretic criterion reduces to the usual average detection delay when the signals are i.i.d. (in discrete time) or Brownian motions with constant drift (in continuous time).

It is clear that the performance of any decentralized scheme is inferior than that of the optimum centralized test. This is true for two major reasons. First, because a decentralized test has access to less information ($\{z_n^i\}$ being a summary of $\{\xi_t^i\}$), but also because of loss in *time resolution* ($\{\tau_n^i\}$ being a sampled version of the actual time t). The main goal of this work is to find decentralized schemes where this performance loss can be quantified and propose methods for controlling it.

Regarding the decentralized version of Problems 1 and 2, we must emphasize that, the way they are stated, it is assumed that the sampling/quantization policy, namely the mechanism by which the pairs $\{(z_n^i, \tau_n^i)\}$ are generated from the observation sequence $\{\xi_t^i\}$, is already specified. Of course, one might extend both problems by including an additional minimization over the sampling/quantization policy, thus optimizing all parts of the decentralized test. Finding however optimum, per se, decentralized tests that solve the extended version of the two problems turns out to be an extremely challenging task. For example, even if we fix the sampling policy and require all sensors to communicate every time they take an observation, the resulting extended optimization problem is in general intractable from a dynamic programming point of view [10]. For this reason, we focus on suboptimum procedures.

To assess the quality of any decentralized test, since the optimum decentralized test is not available, we can compare it against the *centralized* optimum scheme, which is known in several important cases. We are in particular interested in *asymptotically optimum* tests. If $\mathscr{T}$ denotes the stopping time corresponding to the optimum centralized test that solves Problem 1 or 2 and T the stopping time of a decentralized (or even centralized) competitor, then we distinguish the following degrees of asymptotic optimality[1]:

We will say that a test is *asymptotically optimal of order-1*, if for $j=0,1$ and as $\alpha,\beta\to 0$, we have

$$\frac{\mathsf{E}_j[T]}{\mathsf{E}_j[\mathscr{T}]}=1+o(1),\ \text{or}\ \frac{\mathsf{E}_j[u_T]}{\mathsf{E}_j[u_{\mathscr{T}}]}=1+o(1) \tag{8}$$

for Problems 1 and 2, respectively.

We will say that a test is *asymptotically optimal of order-2*, if for $j=0,1$ and as $\alpha,\beta\to 0$, we have

$$\mathsf{E}_j[T]-\mathsf{E}_j[\mathscr{T}]=O(1),\ \text{or}\ \mathsf{E}_j[u_T]-\mathsf{E}_j[u_{\mathscr{T}}]=O(1) \tag{9}$$

for Problems 1 and 2, respectively.

It is clear that order-2 asymptotic optimality is stronger than order-1, since expected delays and K-L divergences increase without bound as $\alpha,\beta\to 0$. These definitions provide an intuitive classification of asymptotically optimal behavior, since the average run length (ARL) curve of an order-1 asymptotically scheme may diverge from the optimal ARL curve, whereas in the case of an order-2 asymptotically optimal scheme the two curves will have a bounded distance.

Before establishing any form of asymptotic optimality, we need to recall the major results of the optimum centralized theory.

III. Optimum Centralized Sequential Testing

The optimization problems defined in (6) and (7) are associated with the celebrated Sequential Probability Ratio Test (SPRT) proposed by Wald [1], which is defined as follows

$$\mathscr{T}=\inf\{t>0: u_t\notin(-A,B)\}$$
$$d_{\mathscr{T}}=\begin{cases}1 & \text{if } u_{\mathscr{T}}\geq B\\ 0 & \text{if } u_{\mathscr{T}}\leq -A\end{cases} \tag{10}$$

where $A,B>0$ are two thresholds and $\mathscr{T}$ is the first time the global log-likelihood ratio process $\{u_t\}$ leaves the open interval $(-A,B)$. The decision function $d_{\mathscr{T}}$ is an $\mathscr{F}_{\mathscr{T}}$-measurable random variable, according to which H_0 (H_1) is accepted if the lower (upper) threshold is first crossed. In *continuous time*, Shiryaev [16] considered the following hypothesis testing problem:

$$\mathsf{H}_0:\xi_t^i=w_t^i;\quad \mathsf{H}_1:\xi_t^i=\mu^i t+w_t^i \tag{11}$$

where $\{w_t=(w_t^1,\ldots,w_t^K)\}_{t\geq 0}$ is a K-dimensional Wiener process and $\mu=(\mu^1,\ldots,\mu^K)\in\mathbb{R}^K$ are constant drifts with $\mu^i\neq 0,\forall i$. The local log-likelihood ratio is equal to $u_t^i=-0.5(\mu^i)^2 t+\mu^i\xi_t^i$ and by summing the local components we can compute u_t and apply the SPRT. The SPRT was shown in [16] to be optimum in the sense of Problem 1 and Problem 2 as long as the thresholds A,B are chosen so that the error probability constraints in (5) are satisfied with equalities. Moreover, we have the following exact formulas for the optimum performance:

$$\mathsf{E}_0[\mathscr{T}]=\frac{2}{||\mu||^2}\mathcal{H}(\alpha,\beta);\quad \mathsf{E}_1[\mathscr{T}]=\frac{2}{||\mu||^2}\mathscr{H}(\beta,\alpha) \tag{12}$$

where $\mathcal{H}(x,y)=x\log(\frac{x}{1-y})+(1-x)\log(\frac{1-x}{y})$, whereas the optimal thresholds that guarantee that the two error probability constraints are satisfied with equality are given by

$$A=\log\left(\frac{1-\alpha}{\beta}\right),\quad B=\log\left(\frac{1-\beta}{\alpha}\right). \tag{13}$$

Liptser and Shiryaev [15] considered the following significantly richer class of hypothesis testing problems:

$$\mathsf{H}_0:\xi_t^i=w_t^i;\quad \mathsf{H}_1:\xi_t^i=\int_0^t\mu_s^i\,ds+w_t^i, \tag{14}$$

where as before $\{w_t=(w_t^1,\ldots,w_t^K)\}_{t\geq 0}$ is a K-dimensional Wiener process and $\{\mu_t=(\mu_t^1,\ldots,\mu_t^K)\}_{t\geq 0}$ is a K-dimensional $\{\mathscr{F}_t\}$-adapted process satisfying[2]

$$\mathsf{P}_j\left(\int_0^\infty||\mu_s||^2\,ds=\infty\right)=1$$
$$\mathsf{P}_j\left(\int_0^t||\mu_s||^2\,ds<\infty\right)=1$$
$$\mathsf{E}_j\left[\exp\left(\frac{1}{2}\int_0^t||\mu_s||^2\,ds\right)\right]<\infty \tag{15}$$

for all $t\geq 0, j=0,1$. The local log-likelihood ratio u_t^i takes the form

$$u_t^i=\int_0^t\mu_s^i d\xi_s^i-\int_0^t 0.5(\mu_s^i)^2\,ds \tag{16}$$

which again allows for the computation of u_t and the application of SPRT. Moreover, the K-L divergence can be written in the following form:

$$-\mathsf{E}_0[u_t]=\mathsf{E}_0\left[\int_0^t 0.5||\mu_s||^2\,ds\right]$$
$$\mathsf{E}_1[u_t]=\mathsf{E}_1\left[\int_0^t 0.5||\mu_s||^2\,ds\right] \tag{17}$$

which reveals the nonnegative and time increasing nature of this alternative criterion.

The SPRT is optimum in the sense of Problem 2 in the case of Itô processes [15], but also for general continuous-path processes [18], yielding the following optimal performance:

$$-\mathsf{E}_0[u_{\mathscr{T}}]=\mathcal{H}(\alpha,\beta),\quad \mathsf{E}_1[u_{\mathscr{T}}]=\mathcal{H}(\beta,\alpha) \tag{18}$$

with the thresholds A,B given by (13) in order to satisfy the two constraints in (5) with equality.

In *discrete time*, it is known that when the vector sequence $\{\xi_t\}$ with $\xi_t=(\xi_t^1,\ldots,\xi_t^K)$ is i.i.d. with independent components under both hypotheses, SPRT is optimum in the sense of

[1]We recall the difference between the notations $\Theta(\cdot),O(\cdot)$ and $o(\cdot)$. If ω is a parameter that tends to 0 or ∞ and $\mathcal{A}(\omega),\mathcal{B}(\omega)$ functions of ω then $\mathcal{A}(\omega)=\Theta(\mathcal{B}(\omega))$ means that $|\mathcal{A}(\omega)|/|\mathcal{B}(\omega)|$ is uniformly bounded away from 0 and ∞; $A(\omega)=O(\mathcal{B}(\omega))$ that the same ratio is bounded away from ∞ and $A(\omega)=o(\mathcal{B}(\omega))$ that $\mathcal{A}(\omega)|/|\mathcal{B}(\omega)|\to 0$ as ω tends to 0 or ∞.

[2]The last condition in (15) is known as the Novikov condition and assures that $\{e^{u_t}\}$ is a P_0-martingale and $\{e^{-u_t}\}$ a P_1-martingale. Alternative, more relaxed conditions that guarantee the martingale property can be found in [17, p. 199].

Problem 1 and Problem 2. In particular, suppose that under the two hypotheses we have

$$\mathsf{H}_0 : \xi_t \sim F_0(\xi^1, \ldots, \xi^K) = \prod_{i=1}^{K} F_0^i(\xi^i)$$
$$\mathsf{H}_1 : \xi_t \sim F_1(\xi^1, \ldots, \xi^K) = \prod_{i=1}^{K} F_1^i(\xi^i) \qquad (19)$$

where $F_j^i(x)$ denotes the common cumulative distribution function (cdf) of the data acquired by sensor i when hypothesis H_j is true and "$\sim$" means "distributed according to". For this case, the local log-likelihood ratio at time t takes the form

$$u_t^i = \sum_{k=1}^{t} \log \frac{dF_1^i\left(\xi_k^i\right)}{dF_0^i\left(\xi_k^i\right)} \qquad (20)$$

and by summing over i we can compute the global log-likelihood ratio u_t and apply the SPRT. The proof of optimality of SPRT was first given by Wald and Wolfowitz in [14]. In fact, this proof constitutes the first optimality result of Sequential Analysis.

We can now make the following remarks.

- The SPRT has also been proven to be optimal in the case where the $\{\xi_t^i\}$ are independent homogeneous Poisson processes [19]. This problem however is not particularly interesting under the decentralized setup, since an arrival at a sensor can be signaled to the fusion center using simply one bit of information.
- In discrete time, SPRT is known to be optimum only in the i.i.d. case. Unfortunately, no analog to the Itô class result for Problem 2 has been developed so far.

From the optimum centralized theory, we conclude that in order to apply the SPRT we need the global log-likelihood ratio $\{u_t\}$ or more precisely its local components $\{u_t^i\}$ coming from the sensors. Our goal in the next sections will be to propose efficient *approximations* for these processes that will replace them in the definition of SPRT and give rise to a decentralized SPRT-like test. The efficiency of this test will then be compared against the optimum SPRT in order to establish its asymptotic optimality.

IV. Decentralized Sequential Testing in Continuous Time

Since we are in the continuous time case, t is real, taking values in $[0, \infty)$. Let us assume, but without for the moment explaining how, that the fusion center is capable of reproducing *exactly* the local log-likelihood ratio u_t^i at the sampling instants $t = \tau_n^i$ by using only the received information $\{z_n^i\}$ from sensor i. It then makes sense to approximate u_t^i between sampling times with its most recently reproduced value. In order to write this more formally, we recall that m_t^i denotes the number of messages transmitted by sensor i up to time t. Thus, at time t, $\tau_{m_t^i}^i$ is the most recent communication time for sensor i and $u_{\tau_{m_t^i}^i}^i$ the most recently reproduced log-likelihood ratio value. Our suggestion is to approximate u_t^i with $\tilde{u}_t^i = u_{\tau_{m_t^i}^i}^i$.

We emphasize that we have exact equality between $\tilde{u}_t^i$ and u_t^i at $t = \tau_n^i$, because we assume that the fusion center is capable of reproducing exactly the corresponding log-likelihood ratio at the sampling times $\{\tau_n^i\}$. Then, the fusion center can produce an approximation $\tilde{u}_t$ for the global log-likelihood ratio u_t by summing the available local approximations

$$\tilde{u}_t = \sum_{i=1}^{K} \tilde{u}_t^i = \sum_{i=1}^{K} u_{\tau_{m_t^i}^i}^i, \quad 0 \le t < \infty \qquad (21)$$

Unlike the local approximation $\tilde{u}_t^i$, which is exact at the times $\{\tau_n^i\}$, the global approximation $\tilde{u}_t$ can be exactly equal to u_t at a sampling instant only if all sensors transmit synchronously, otherwise u_t and $\tilde{u}_t$ will be different.

Replacing now $\{u_t\}$ with $\{\tilde{u}_t\}$ in the definition of SPRT in (10), we obtain an SPRT-like test of the form

$$\tilde{\mathscr{T}} = \inf\{t \ge 0 : \tilde{u}_t \notin (-\tilde{A}, \tilde{B})\}$$
$$d_{\tilde{\mathscr{T}}} = \begin{cases} 1 & \text{if } \tilde{u}_{\tilde{\mathscr{T}}} \ge \tilde{B} \\ 0 & \text{if } \tilde{u}_{\tilde{\mathscr{T}}} \le -\tilde{A} \end{cases} \qquad (22)$$

where again the thresholds $\tilde{A}, \tilde{B} > 0$ are selected to satisfy the error probability constraints in (5) with equality. The test we just described constitutes the fusion center policy we propose under the decentralized setup. Let us now explain how the fusion center can make an exact reproduction of the local log-likelihood ratios.

A. Lebesgue Sampling as a Quantization Strategy

Of course, the simplest way the fusion center can reproduce the log-likelihood ratio is by receiving the corresponding value directly from the sensor. However, this would require a communication protocol that is not limited to 1-bit information. The interesting point is that, after careful consideration, the 1-bit communication constraint can be satisfied in the case of Lebesgue sampling.

Recalling that $\{\tau_n^i\}$ denotes the sequence of communication times for sensor i, we have that the local log-likelihood ratio at time τ_n^i can be written as

$$u_{\tau_n^i}^i = \sum_{k=1}^{n} \left[u_{\tau_k^i}^i - u_{\tau_{k-1}^i}^i\right] \qquad (23)$$

suggesting that the fusion center only needs the increments $u_{\tau_k^i}^i - u_{\tau_{k-1}^i}^i$ in order to recover the exact value $u_{\tau_n^i}^i$ at the sampling instant τ_n^i. But when $\{u_t^i\}$ has *continuous paths* and we adopt the Lebesgue sampling scheme (4), each of these increments can take only two values. Indeed, due to path continuity, the process $u_t^i - u_{\tau_{n-1}^i}^i$ will hit at time τ_n^i one of the two thresholds, $-\underline{\Delta}_i$ or $\bar{\Delta}_i$. By assuming that the values $\underline{\Delta}_i, \bar{\Delta}_i$ are selected before hand and are made available to the fusion center, it then becomes easy to communicate the exact value of the increment $u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i$ by simply transmitting the following 1-bit information

$$z_n^i = \begin{cases} 1, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i = \bar{\Delta}_i \\ 0, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i = -\underline{\Delta}_i. \end{cases} \qquad (24)$$

Using the sequence $\{z_n^i\}$ and (23), the fusion center can reproduce u_t^i exactly at the sampling times $\{\tau_n^i\}$ and then form $\tilde{u}_t$, which is required in the SPRT-like test defined in (22). Actually, with this particular communication protocol it is possible to update directly the test statistic $\tilde{u}_t$, without passing through the local statistics $\tilde{u}_t^i$. Indeed, every time the fusion center receives the 1-bit information z_n^i from sensor i, it must simply add to the existing $\tilde{u}_t$ either $-\underline{\Delta}_i$ or $\bar{\Delta}_i$ depending on z_n^i being 0 or 1, respectively. This observation suggests that the process $\{\tilde{u}_t\}$ is *piecewise constant* exhibiting jumps every time the fusion center receives information from one or more sensors.

Lebesgue sampling in conjunction with the stopping and decision mechanism defined in (22) gives rise to the Decentralized Sequential Probability Ratio Test. This is in fact the continuous time version of the scheme suggested in [13] and constitutes the decentralized sequential test that will be in the center of our attention. We emphasize that the D-SPRT is a valid decentralized scheme, since the communication activity it requires is limited to 1-bit data. Before examining the optimality characteristics of the D-SPRT, let us identify certain important properties of this detection structure.

- Lebesgue sampling at each sensor can be seen as a local *repeated* SPRT with thresholds $\underline{\Delta}_i, \bar{\Delta}_i$. Using (13), one can also prove that for every $n \in \mathbb{N}$

$$-\underline{\Delta}_i = \log \frac{\mathsf{P}_1\left(z_n^i = 0\right)}{\mathsf{P}_0\left(z_n^i = 0\right)}; \quad \bar{\Delta}_i = \log \frac{\mathsf{P}_1\left(z_n^i = 1\right)}{\mathsf{P}_0\left(z_n^i = 1\right)}. \qquad (25)$$

  Consequently, for the update of the estimate $\tilde{u}_t$, the fusion center uses the log-likelihood ratio of the received bits z_n^i.
- The local thresholds $\underline{\Delta}_i, \bar{\Delta}_i$ control the average intersampling period, which is an increasing function of these two parameters. Recalling that we have two different hypotheses, we understand that the average intersampling period will depend on the true hypothesis. If we require the two average periods to have specific prescribed values, then we can uniquely identify the local thresholds for the Brownian motion or the Itô process case, using (12) (or (18) if we want to specify the K-L divergence) and (13). In other data models, the two thresholds can be specified using simulations.
- From the definition of the Lebesgue sampling scheme it is easy to see that $|u_t^i - \tilde{u}_t^i| \leq \underline{\Delta}_i + \bar{\Delta}_i, t \geq 0$, suggesting that

$$|u_t - \tilde{u}_t| \leq C = \sum_{i=1}^{K} (\underline{\Delta}_i + \bar{\Delta}_i), \quad t \geq 0. \qquad (26)$$

  Thus, at any time t, the "approximate" log-likelihood ratio $\tilde{u}_t$ differs from the "true" log-likelihood ratio u_t at most by the constant C.
- As we argued above, $\{\tilde{u}_t\}$ is piecewise constant. Assuming it is right continuous with left limits, the difference $\tilde{u}_t - \tilde{u}_{t-}$ expresses the possible jump in the process at time t. The largest in absolute value jump occurs when all sensors communicate at the same time and transmit data of the same sign. It is easy to verify that the maximal jump can also be bounded by the parameter C that was defined in (26), i.e.,

$$|\tilde{u}_t - \tilde{u}_{t-}| \leq C. \qquad (27)$$

- We recall that, in addition to the data sequence $\{z_n^i\}$, each sensor transmits indirectly to the fusion center the sequence $\{\tau_n^i\}$ of communication times. As we argued before, the pairs (z_n^i, τ_n^i) constitute the complete set of information received by the fusion center generating the filtration $\{\mathscr{G}_t\}$. It is also evident that the statistics of (z_n^i, τ_n^i) differ under each hypothesis suggesting that both components of the pair may carry information about the true hypothesis. We realize however that D-SPRT makes use only of the data $\{z_n^i\}$ ignoring completely $\{\tau_n^i\}$. Dropping this amount of information may inflict a performance loss, however it turns out to be practically advantageous. Indeed, any efficient use of the pair (z_n^i, τ_n^i) would require the knowledge (or computation) of the corresponding joint pdf under the two hypotheses. Unfortunately, this is possible only for the Brownian motion model [17] and, even in this case, it is in the form of a complicated series expansion.

B. Asymptotic Optimality of the D-SPRT

Let us now establish a strong asymptotic optimality property for D-SPRT in continuous time. This is the goal of our next theorem.

Theorem 1: Suppose that $\tilde{\mathscr{T}}, d_{\tilde{\mathscr{T}}}$ is the D-SPRT test defined in (22), with thresholds $\tilde{A}, \tilde{B}$ selected to satisfy the error probability constraints in (5) with equality. Then

$$\tilde{A} \leq |\log \beta| + C; \quad \tilde{B} \leq |\log \alpha| + C. \qquad (28)$$

Furthermore, D-SPRT is asymptotically optimum of order-2 in the case of Problem 1 and Problem 2 with Brownian motion signals with constant drifts and in the case of Problem 2 with Itô processes.

Proof: To prove (28), we apply a change of measures and use (26), this yields

$$\begin{aligned}\beta = \mathsf{P}_1(\tilde{u}_{\tilde{\mathscr{T}}} \leq -\tilde{A}) &= \mathsf{E}_0\left[e^{u_{\tilde{\mathscr{T}}}} \mathbb{1}_{\{\tilde{u}_{\tilde{\mathscr{T}}} \leq -\tilde{A}\}}\right] \\ &= \mathsf{E}_0\left[e^{\tilde{u}_{\tilde{\mathscr{T}}} + (u_{\tilde{\mathscr{T}}} - \tilde{u}_{\tilde{\mathscr{T}}})} \mathbb{1}_{\{\tilde{u}_{\tilde{\mathscr{T}}} \leq -\tilde{A}\}}\right] \leq e^{-\tilde{A}+C} \qquad (29)\end{aligned}$$

which proves the first inequality in (28). Similarly, we can show the second inequality.

Regarding the order-2 asymptotic optimality, we are going to prove only the case of Itô processes and Problem 2, since this reduces to Problem 1 in the case of Brownian motions with constant drifts. According to the second relation in (9), under hypothesis H_0 we need to prove that

$$(-\mathsf{E}_0[u_{\tilde{\mathscr{T}}}]) - (-\mathsf{E}_0[u_{\mathscr{T}}]) = O(1). \qquad (30)$$

Note that the left-hand side in (30) is always nonnegative since the SPRT, by being optimum, delivers the smallest K-L divergence. Consequently, what is left to show is that the difference can be upper bounded by a constant.

Recall that $\{\tilde{u}_t\}$ is piecewise constant, therefore stopping can occur only with a jump. According to (27) the jumps of this process cannot exceed the bound C defined in (26). Since before stopping the process $\tilde{u}_t$ takes values in the interval $(-\tilde{A}, \tilde{B})$,

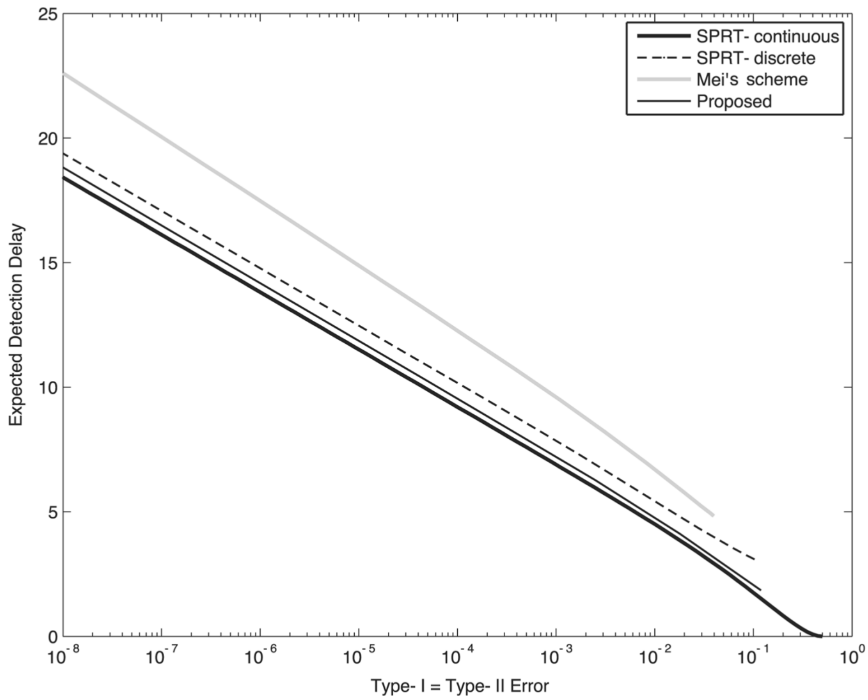


Fig. 2. Relative performance of centralized and decentralized schemes in continuous time with $K = 2$ sensors and testing between H_0 : Brownian motions with drift 0 and H_1 : Brownian motions with drift 1.

after stopping we have $\tilde{u}_{\tilde{\mathscr{T}}} \geq -\tilde{A} - C$. Using this observation, (26) and (28), we can write

$$\begin{aligned}\mathsf{E}_0[u_{\tilde{\mathscr{T}}}] &= \mathsf{E}_0[\tilde{u}_{\tilde{\mathscr{T}}} + (u_{\tilde{\mathscr{T}}} - \tilde{u}_{\tilde{\mathscr{T}}})] \\ &\geq (-\tilde{A} - C) - C \geq -|\log\beta| - 3C. \end{aligned} \quad (31)$$

From (18), we have that the performance of the SPRT satisfies $-\mathsf{E}_0[u_{\mathscr{T}}] = |\log\beta| + \alpha|\log\beta| + o(1)$ as $\alpha, \beta \to 0$. Normally, α and β are selected to have the same order of magnitude yielding $\alpha|\log\beta| = o(1)$, however for the validity of our theorem we can even tolerate cases where $\alpha|\log\beta| = O(1)$, that is, cases where α and β are of drastically different orders of magnitudes (e.g., $\beta = O(1/|\log\alpha|)$). Consequently, assuming that α and β converge to 0 so that $\alpha|\log\beta| + \beta|\log\alpha| = O(1)$, we can replace $|\log\beta|$ with the optimal SPRT performance $-\mathsf{E}_0[u_{\mathscr{T}}]$ in (31) and prove (30) under H_0. Adopting similar arguments for the upper threshold $\tilde{B}$, we can prove (9) under H_1. This concludes the proof. ■

C. Simulation Experiments

We now present a simulation experiment in the context of Problem 1 with continuous time observations defined as in (11). Specifically, each sensor observes a standard Brownian motion under H_0 and a Brownian motion with a constant drift under H_1. We consider the case of $K = 2$ sensors with the two constant drifts under H_1 to have the values $\mu^1 = \mu^2 = 1$.

We compare D-SPRT against the continuous time (centralized) SPRT, the discrete time (centralized) SPRT and Mei's [11] decentralized test. Unlike D-SPRT, the last two tests are based on synchronous communication between sensors and fusion center. Thus, in order to implement these schemes, we assume that each sensor observes its underlying process at the times $t = 0, h, 2h, \ldots$ and communicates with the fusion center at all these times, where h is a known positive constant. In the case of centralized discrete time SPRT, the sensors transmit the exact values that they observe, whereas in the case of Mei's scheme they send quantized signals using an alphabet of size 3.

For the comparison to be fair, we must equate the average intersampling periods of Lebesgue sampling with the constant communication period h of canonical deterministic sampling. Selecting the local thresholds to have values $\bar{\Delta}_i = \underline{\Delta}_i = 2$ gives $\mathsf{E}_0[\tau_1^i] = \mathsf{E}_1[\tau_1^i] = 3.0464$, which must become the value for the period of deterministic sampling, namely $h = 3.0464$. In Fig. 2 we can see that the distance between the D-SPRT and the optimal performance remains bounded, which agrees with the order-2 asymptotic optimality result of Theorem 1. Mei's scheme [11] on the other hand, known to be order-1 asymptotically optimum, exhibits performance that slowly diverges from the optimum.

The other important conclusion that we can draw from our graph is that D-SPRT exhibits a distinct performance improvement over the discrete time centralized SPRT, which is based on canonical deterministic sampling. We recall that this algorithm is optimum in discrete time, but under the continuous time setup it is asymptotically optimum of order-2. As we argued in the Introduction, Lebesgue sampling is preferable to canonical deterministic sampling from a practical point of view since it does not require synchronization. Motivated by our simulations we conjecture that, even under the centralized setup, this form of sampling achieves better performance than canonical deterministic sampling.

V. Decentralized Sequential Testing in Discrete Time

We consider the same formulation as in Section IV, only now time t is discrete with $t \in \mathbb{N}$. At each sensor i, $\{\xi_t^i\}$ is a sequence of independent observations under each hypothesis H_j with common cumulative distribution functions $F_j^i(x), j = 0, 1$. Denoting with $\ell_t^i = \log(dF_1^i(\xi_t^i)/dF_0^i(\xi_t^i))$ the log-likelihood ratio of the sample ξ_t^i, we assume that $\mathsf{P}_j^i(\ell_t^i \neq 0) > 0$, in other words that the two densities are not equal with probability 1. We then have that the global log-likelihood ratio u_t is given by

$$u_t = \sum_{i=1}^{K} u_t^i = \sum_{i=1}^{K}\sum_{k=1}^{t} \ell_k^i = u_{t-1} + \sum_{i=1}^{K} \ell_t^i. \quad (32)$$

The corresponding SPRT is optimum in the sense of Problem 1 and 2 [14], provided that the two thresholds A, B are selected to satisfy the probability constraints in (5) with equality. We recall that, in discrete time there is no other data model for which we know the solution for either Problem 1 or Problem 2 (i.e., there is no equivalent to the Itô processes case).

Since the centralized SPRT will again become the point of reference for any decentralized test, it is necessary to quantify its performance. Unfortunately, in discrete time there are no exact expressions as in continuous time and we therefore need to resort to asymptotic formulas and bounds. For the performance of SPRT we have the following lower bounds [2, p. 21]:

$$\begin{aligned} -\mathsf{E}_0[u_{\mathscr{T}}] &\geq \mathcal{H}(\alpha, \beta) = |\log \beta| + o(1) \\ \mathsf{E}_1[u_{\mathscr{T}}] &\geq \mathcal{H}(\beta, \alpha) = |\log \alpha| + o(1) \end{aligned} \quad (33)$$

which replace the exact equalities of the continuous time and continuous path case depicted in (18).

Our intention is to apply the same D-SPRT scheme we introduced in the continuous time case, namely Lebesgue sampling combined with an SPRT-like test, where we approximate properly the global log-likelihood ratio u_t. Unfortunately, this transfer from continuous to discrete time is not as straightforward as one might expect. The main reason is that with Lebesgue sampling we are no longer able to reproduce exactly the local log-likelihood ratios at the corresponding sampling times, because of the *overshoot effect* occurring at the local SPRTs. This rather unfortunate difference is responsible for a substantial complication in the corresponding discrete time analysis.

For simplicity, we will limit ourselves to the case where the two error levels α, β decrease to 0 at the same rate, meaning that the ratio α/β is uniformly bounded away from 0 and ∞ (or according to our definition $\beta = \Theta(\alpha)$).

A. Lebesgue Sampling and D-SPRT in Discrete Time

In each sensor i, the Lebesgue sampling scheme (4) produces a sequence $\{\tau_n^i\}$ of $\{\mathscr{F}_t^i\}$-adapted stopping times, only now, due to the overshoot effect, the local SPRT statistic $u_t^i - u_{\tau_{n-1}^i}^i$ does not necessarily hit the two thresholds, $-\underline{\Delta}_i$ and $\bar{\Delta}_i$, at time τ_n^i. Consequently, the information z_n^i sent over the channel can express only the *side* by which the statistic $u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i$ exits the interval $(-\underline{\Delta}_i, \bar{\Delta}_i)$. More precisely

$$z_n^i = \begin{cases} 1, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \geq \bar{\Delta}_i \\ 0, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \leq -\underline{\Delta}_i \end{cases} \quad (34)$$

which is the equivalent of (24).

The question that now arises is how the fusion center should utilize the sequence of messages $\{z_n^i\}$. We recall that in the continuous time and continuous path case the fusion center, in view of (25), uses the log-likelihood ratio of the received bits z_n^i to update the estimate $\tilde{u}_t$. Consequently, it is natural to use the same idea here (as was also originally suggested in [13]) and define the following two quantities for each sensor:

$$-\underline{\Lambda}_i = \log \frac{\mathsf{P}_1\left(z_n^i = 0\right)}{\mathsf{P}_0\left(z_n^i = 0\right)}; \quad \bar{\Lambda}_i = \log \frac{\mathsf{P}_1\left(z_n^i = 1\right)}{\mathsf{P}_0\left(z_n^i = 1\right)}. \quad (35)$$

Both values $\underline{\Lambda}_i, \bar{\Lambda}_i$ can be precomputed (e.g., by simulations) and made known to the fusion center.

As we argued above, we are interested in the sequence of overshoots $\{\eta_n^i\}$, where

$$\begin{aligned} \eta_n^i = &\left(u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i + \underline{\Delta}_i\right) \mathbb{1}_{\{u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \leq -\underline{\Delta}_i\}} \\ &+ \left(u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i - \bar{\Delta}_i\right) \mathbb{1}_{\{u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \geq \bar{\Delta}_i\}}. \end{aligned} \quad (36)$$

The maximal expected overshoot is a parameter that will play a very important role in our analysis. We define it as follows:

$$\theta = \max_j \max_i \mathsf{E}_j\left[\left|\eta_n^i\right|\right] \quad (37)$$

and we know [20] that it is finite if $\mathsf{E}_j[(\ell_t^i)^2] < \infty, j = 0, 1, i = 1, \ldots, K$.

In the continuous time and continuous path case, since there is no overshoot, the thresholds $\underline{\Delta}_i, \bar{\Delta}_i$ coincide with the quantities $\underline{\Lambda}_i, \bar{\Lambda}_i$. In discrete time, this is no longer true. The next lemma quantifies their relative size.

Lemma 1: Let $\underline{\Delta}_i, \bar{\Delta}_i > 0$ denote the thresholds for the local SPRT and $\underline{\Lambda}_i, \bar{\Lambda}_i$ be defined as in (35, then

$$\underline{\Lambda}_i \geq \underline{\Delta}_i; \quad \bar{\Lambda}_i \geq \bar{\Delta}_i \quad (38)$$

$$\underline{\Lambda}_i = \underline{\Delta}_i + O(\theta); \quad \bar{\Lambda}_i = \bar{\Delta}_i + O(\theta). \quad (39)$$

Proof: The proof is presented in the Appendix. ■

The fusion center, every time it receives an information bit z_n^i, it updates its existing statistic $\tilde{u}_t$ by either adding $-\underline{\Lambda}_i$ when $z_n^i = 0$ or $\bar{\Lambda}_i$ when $z_n^i = 1$. Recalling that m_t^i denotes the number of bits transmitted by sensor i up to time t, we can write for the D-SPRT statistic that $\tilde{u}_t = \sum_{i=1}^{K} \tilde{u}_t^i$ where

$$\tilde{u}_t^i = \sum_{n=1}^{m_t^i} \lambda_n^i; \text{ with } \lambda_n^i = -\underline{\Lambda}_i \mathbb{1}_{\{z_n^i = 0\}} + \bar{\Lambda}_i \mathbb{1}_{\{z_n^i = 1\}}. \quad (40)$$

The K-L information numbers of the sequence $\{\lambda_n^i\}$ also play an important role in our analysis. We have the following estimates depicted in the next lemma.

Lemma 2: For the K-L information numbers of the sequence $\{\lambda_n^i\}$ we can write

$$I_0^i = -\mathsf{E}_0[\lambda_n^i] \geq \frac{\underline{\Delta}_i\left(e^{\bar{\Delta}_i} - 1\right) + \bar{\Delta}_i\left(e^{-\underline{\Delta}_i} - 1\right)}{e^{\bar{\Delta}_i} - e^{-\underline{\Delta}_i}} > 0$$

$$I_1^i = \mathsf{E}_1[\lambda_n^i] \geq \frac{\underline{\Delta}_i\left(e^{-\underline{\Delta}_i}-1\right)+\bar{\Delta}_i\left(e^{\bar{\Delta}_i}-1\right)}{e^{\bar{\Delta}_i}-e^{-\underline{\Delta}_i}} > 0. \tag{41}$$

Additionally, if $\underline{\Delta}_i, \bar{\Delta}_i$ are either of the size of a constant bounded away from 0, or tend to ∞ in such a way that $\underline{\Delta}_i/\bar{\Delta}_i$ is bounded away from 0 and ∞ (i.e., $\underline{\Delta}_i = \Theta(\bar{\Delta}_i)$), the previous expressions simplify to

$$I_0^i \geq \Theta(\underline{\Delta}_i), \quad I_1^i \geq \Theta(\bar{\Delta}_i). \tag{42}$$

Proof: The proof is presented in the Appendix. ■

The analysis of the classical SPRT algorithm relies on Wald's (second) identity. In order to be able to analyze D-SPRT, it turns out that we need an equivalent result. The next lemma introduces a version of Wald's second identity that is suitable for our needs.

Lemma 3: Let $\{\tau_n^i\}$ denote the sequence of sampling times in sensor i generated by the Lebesgue sampling scheme (4). Consider a sequence $\{\zeta_n^i\}$ of i.i.d. random variables where each ζ_n^i is a function of the samples $\xi^i_{\tau^i_{n-1}+1}, \ldots, \xi^i_{\tau^i_n}$ acquired by the sensor during the nth intersampling period and assume $\mathsf{E}_j[|\zeta_n^i|] < \infty$. If T denotes any $\{\mathscr{F}_t\}$-adapted stopping time with finite P_j-expectation and m_T^i is the number of sampling times τ_n^i occurred up to and including time T, then for $j = 0, 1$, we have

$$\mathsf{E}_j\left[\sum_{n=1}^{m_T^i+1}\zeta_n^i\right] = \mathsf{E}_j\left[\zeta_1^i\right]\left(\mathsf{E}_j\left[m_T^i\right]+1\right). \tag{43}$$

As an immediate consequence we can obtain the following two estimates.

i) For $\zeta_n^i \geq 0$, we have

$$\mathsf{E}_j\left[\sum_{n=1}^{m_T^i}\zeta_n^i\right] \leq \mathsf{E}_j\left[\zeta_1^i\right]\left(\mathsf{E}_j\left[m_T^i\right]+1\right). \tag{44}$$

ii) If $\{\zeta_n^i\}$ is a sequence with $|\zeta_n^i| \leq M < \infty$ for all n, then

$$\left|\mathsf{E}_j\left[\sum_{n=1}^{m_T^i}\zeta_n^i\right] - \mathsf{E}_j\left[\zeta_1^i\right]\mathsf{E}_j\left[m_T^i\right]\right| \leq 2M. \tag{45}$$

Proof: The proof is presented in the Appendix. ■

One might wonder why it is necessary to set the upper limit in (43) to $m_T^i + 1$ instead of the classical m_T^i we encounter in Wald's original identity. Unfortunately, if the upper limit is replaced by m_T^i, then in the proof (specifically in (62)) the random variable ζ_n^i will be combined with $\mathbb{1}_{\{m_T^i \geq n\}}$ instead of $\mathbb{1}_{\{m_T^i \geq n-1\}}$. As it turns out, these two quantities are not necessarily independent as is the case between ζ_n^i and $\mathbb{1}_{\{m_T^i \geq n-1\}}$ and therefore Wald's identity cannot be assured.

Unlike in continuous time, due to the overshoot effect, there is now an accumulation of errors which results in the difference $u_t - \tilde{u}_t$ being unbounded and no longer limited by a constant. However, by properly selecting the local thresholds, we will see that we can force this difference to grow at a much slower pace than each of its components $u_t, \tilde{u}_t$. In turn, this possibility will allow us to prove interesting asymptotic optimality properties for the discrete time D-SPRT. Since the difference of the two statistics plays a crucial role in our analysis, with the next lemma we obtain an estimate of its size.

Lemma 4: For any $\{\mathscr{F}_t\}$-adapted P_j-integrable stopping time T we have for $j = 0, 1$

$$\mathsf{E}_j[|u_T - \tilde{u}_T|] \leq O(\theta)\left(\frac{|\mathsf{E}_j[\tilde{u}_T]| + 2C'}{\min_i I_j^i} + K\right) + C \tag{46}$$

where $C' = \sum_{i=1}^K(\underline{\Lambda}_i + \bar{\Lambda}_i), C = \sum_{i=1}^K(\underline{\Delta}_i + \bar{\Delta}_i)$ and θ is the maximal expected overshoot defined in (37).

Proof: The proof makes use of Lemma 3 and is presented in the Appendix. ■

B. Asymptotic Optimality

We have concluded the presentation of the background material that is necessary for establishing our main result. Before going to the next theorem that introduces a key estimate for the performance of D-SPRT, we introduce an additional quantity that expresses the order of magnitude of the local thresholds. We will assume that there exists a quantity Δ such that for all i, we have $\underline{\Delta}_i = \Theta(\Delta)$ and $\bar{\Delta}_i = \Theta(\Delta)$. This is necessary because in order to establish the desired asymptotic optimality property, at some point we require the local thresholds to tend to infinity. With this assumption all local thresholds increase at the same rate. After this clarification, we can now state the next key theorem.

Theorem 2: Let $\mathscr{T}, \tilde{\mathscr{T}}$ denote the stopping times for the centralized SPRT and D-SPRT, respectively. We then have the following estimates for the thresholds of D-SPRT:

$$\tilde{A} \leq |\log\beta|; \quad \tilde{B} \leq |\log\alpha|. \tag{47}$$

Additionally, for $j = 0, 1$, we can write

$$\begin{aligned} |\mathsf{E}_0[u_{\tilde{\mathscr{T}}}] - \mathsf{E}_0[u_{\mathscr{T}}]| &\leq \frac{O(\theta)}{\Theta(\Delta)}|\log\beta| + \Theta(\Delta) \\ |\mathsf{E}_1[u_{\tilde{\mathscr{T}}}] - \mathsf{E}_1[u_{\mathscr{T}}]| &\leq \frac{O(\theta)}{\Theta(\Delta)}|\log\alpha| + \Theta(\Delta). \end{aligned} \tag{48}$$

Proof: The proof is presented in the Appendix. ■

We note that (47) is the analog of (28) in discrete time. In fact, it constitutes a better approximation than (28) but at the expense of a more involved proof. Inequality (48) refers to the difference of the K-L divergences between the SPRT stopping time $\mathscr{T}$ and the D-SPRT stopping time $\tilde{\mathscr{T}}$. Since we are in the i.i.d. case, we know that the K-L divergence is proportional to the expected delay and the proportionality factor is simply the K-L information number. The inequalities in (48) reflect the trade-off that underlies the choice of the local thresholds Δ. Indeed, overly small local thresholds will induce frequent communication with the fusion center, thus resulting in rapid error accumulation due to the overshoot effect. This is captured by the first term in the right-hand side of (48). If, on the other hand, we use overly large local thresholds, then this will generate long decision delays due to infrequent communication with the fusion center and coarse time resolution. This part is captured by the second term in (48).

Clearly, there is a compromising value for the local threshold size Δ that can optimize the performance of the test.

If we consider the ratio

$$0 \leq \frac{\mathsf{E}_j[\tilde{\mathscr{T}}] - \mathsf{E}_j[\mathscr{T}]}{\mathsf{E}_j[\mathscr{T}]} = \frac{|\mathsf{E}_j[u_{\tilde{\mathscr{T}}}] - \mathsf{E}_j[u_{\mathscr{T}}]|}{|\mathsf{E}_j[u_{\mathscr{T}}]|} \leq \frac{O(\theta)}{\Theta(\Delta)} + \frac{\Theta(\Delta)}{|\log \alpha|} \tag{49}$$

and let Δ become a function of α such that $\lim_{\alpha \to 0} \frac{\theta}{\Delta} = \lim_{\alpha \to 0} \frac{\Delta}{|\log \alpha|} = 0$, then the right-hand side of (49) tends to 0 establishing order-1 asymptotic optimality. After some simple reasoning, we can deduce that the best choice is $\Delta = \Theta(\sqrt{\theta |\log \alpha|})$, which equates the two terms in (49), yielding

$$0 \leq \frac{\mathsf{E}_j[\tilde{\mathscr{T}}] - \mathsf{E}_j[\mathscr{T}]}{\mathsf{E}_j[\mathscr{T}]} \leq \Theta\left(\sqrt{\frac{\theta}{|\log \alpha|}}\right). \tag{50}$$

The optimal value we obtained for Δ is the optimum local threshold size, expressed in an "order of magnitude" form. Observe also that the convergence rate to 0 of the right-hand side in the previous expression is of the same order as the one obtained in [11], for every constant value of θ.

C. *Oversampling*

From (48), we can see that if we also select θ to tend to 0 with α as $\theta = O(1/|\log \alpha|)$ and we use the optimum value for Δ (which turns out to be bounded), then the inflicted performance loss of the discrete-time D-SPRT is asymptotically bounded as $\alpha, \beta \to 0$ and we recover our continuous time result. Of course, for this observation to be of practical interest, we need to be able to control the maximum expected overshoot θ. This is in general possible if the observations at the sensors come from sampling continuous time process with continuous paths.

Suppose for example that the underlying sensor process at each sensor i is a standard Brownian motion under H_0 and a Brownian motion with drift μ^i under H_1. Moreover, suppose that these underlying continuous time processes are sampled using canonical deterministic sampling with a sampling period h. Thus, each sensor i observes the increments $\{\xi_t^i\}$, which are independent and identically distributed with $\xi_t^i \sim \mathcal{N}(0,h)$ under H_0 and $\xi_t^i \sim \mathcal{N}(\mu^i h, h)$ under H_1. The corresponding log-likelihood ratio is $\ell_t^i = -0.5(\mu^i)^2 h + \mu^i \xi_t^i$.

In this case, we can show that θ is determined by the sampling period h, which can be controlled by the designer of the scheme. In particular, we will show that $\theta = O(h^{1/4})$. In order to prove this claim, it suffices to show that $\mathsf{E}_j[|\eta_n^i|]$ is of the order $O(h^{1/4})$ for each $i = 1, \ldots, K, j = 0, 1, n \in \mathbb{N}$ and since our processes are stationary it suffices to consider only $n = 1$.

Indeed, recalling that $\tau_0^i = u_{\tau_0^i}^i = 0$, we have

$$\begin{aligned} \tau_1^i &= \inf\left\{t > 0 : u_t^i \notin (-\underline{\Delta}_i, \bar{\Delta}_i)\right\} \\ \eta_1^i &= \left(u_{\tau_1^i}^i + \underline{\Delta}_i\right) \mathbb{1}_{\{u_{\tau_1^i}^i \leq -\underline{\Delta}_i\}} + \left(u_{\tau_1^i}^i - \bar{\Delta}_i\right) \mathbb{1}_{\{u_{\tau_1^i}^i \geq \bar{\Delta}_i\}}. \end{aligned} \tag{51}$$

Note now that we can write $\tau_1^i = \min\{\underline{\tau}_1^i, \bar{\tau}_1^i\}$, where

$$\begin{aligned} \underline{\tau}_1^i &= \inf\left\{t > 0 : u_t^i \leq -\underline{\Delta}_i\right\} \\ \bar{\tau}_1^i &= \inf\left\{t > 0 : u_t^i \geq \bar{\Delta}_i\right\}. \end{aligned} \tag{52}$$

Using these definitions, the overshoot takes the form

$$\eta_1^i = \left(u_{\underline{\tau}_1^i}^i + \underline{\Delta}_i\right) \mathbb{1}_{\{u_{\tau_1^i}^i \leq -\underline{\Delta}_i\}} + \left(u_{\bar{\tau}_1^i}^i - \bar{\Delta}_i\right) \mathbb{1}_{\{u_{\tau_1^i}^i \geq \bar{\Delta}_i\}} \tag{53}$$

from which we can easily deduce that

$$\mathsf{E}_j\left[|\eta_1^i|\right] \leq \mathsf{E}_0\left[-\left(u_{\underline{\tau}_1^i}^i + \underline{\Delta}_i\right)\right] + \mathsf{E}_1\left[u_{\bar{\tau}_1^i}^i - \bar{\Delta}_i\right]. \tag{54}$$

From [20, Theorem 3] and Lyapunov's inequality, we have for any $r \geq 1$ that

$$\begin{aligned} \sup_{\underline{\Delta}_i > 0} \mathsf{E}_0\left[-\left(u_{\underline{\tau}_1^i}^i + \underline{\Delta}_i\right)\right] &\leq \left[\frac{r+2}{r+1} \frac{\mathsf{E}_0\left[|\ell_1^i|^{r+1}\right]}{\left|\mathsf{E}_0\left[\ell_1^i\right]\right|}\right]^{1/r} \\ \sup_{\bar{\Delta}_i > 0} \mathsf{E}_1\left[u_{\bar{\tau}_1^i}^i - \bar{\Delta}_i\right] &\leq \left[\frac{r+2}{r+1} \frac{\mathsf{E}_1\left[|\ell_1^i|^{r+1}\right]}{\left|\mathsf{E}_1\left[\ell_1^i\right]\right|}\right]^{1/r}. \end{aligned} \tag{55}$$

If we set $r = 2$, then the upper bound turns out to be $\Theta(h^{1/4})$, which proves our claim.

In order to illustrate the relevance of this result for the design of D-SPRT in practice, we perform simulation experiments in this framework with $K = 2, \mu^i = 1$ and we have selected two values for the sampling period, namely $h = 1$ and 0.1 and two values for the local thresholds, specifically $\bar{\Delta}_i = \underline{\Delta}_i = \Delta = 1$ and 2. We compare the discrete time D-SPRT with the (optimal) discrete time centralized SPRT and also with the decentralized test suggested by Mei in [11], which is also asymptotically optimal of order-1.

Fig. 3 depicts the K-L divergence of the competing schemes. We recall that in this case the K-L divergence is proportional to the expected detection delay. We decided to present the former measure instead of the latter, because the K-L divergence is independent of the size of the samples, while the decision delay varies drastically with this quantity (smaller samples tend to need more time to reach the same threshold).

We observe that D-SPRT exhibits a notable performance improvement when we go from the value $h = 1$ to $h = 0.1$. This is in complete accordance with our previous analysis, since $h = 0.1$ generates likelihood ratios and overshoots of smaller size than $h = 1$. The optimum SPRT on the other hand and Mei's scheme are relatively insensitive to this change of size in the samples. For D-SPRT, it is basically the overshoot accumulation reflected in the first term of the right-hand side in (48) that is reduced when we use smaller h, incurring an overall performance improvement. It is also worth emphasizing that for the D-SPRT the communication frequency (expressed in continuous time) between the sensors and the fusion center *stays relatively unchanged* under both values of h, while in the two other schemes it increases by a factor of ten.

Finally, in Fig. 3 we can also observe that the D-SPRT performance is not monotone with respect to the value of the local threshold $\underline{\Delta}_i = \bar{\Delta}_i = \Delta$. Indeed, case $\Delta = 2$ is better than $\Delta = 1$ for smaller values of α. Additionally, the error probability values where $\Delta = 2$ prevails are increasing with the size of the samples. This performance can be explained by our analysis. We recall that the optimum local threshold is $\Theta(\sqrt{\theta|\log\alpha|})$ suggesting that the error probability where any specific Δ is optimum is roughly $\alpha = \Theta(\exp(-\Delta^2/\theta))$. Consequently, a larger threshold yields better performance at a smaller error probability and this value is an increasing function of θ, the maximal expected overshoot defined in (37).

VI. Conclusions and Generalizations

We have presented and rigorously analyzed a decentralized scheme for sequential hypothesis testing. The detection structure relies on local SPRTs, which are implemented repeatedly at each sensor and used for random sampling of the observed data stream. This sampling scheme naturally induces an asynchronous, 1-bit communication protocol between sensors and fusion center, a practically desirable characteristic. By performing a detailed analysis we were able to prove interesting asymptotic optimality properties for the proposed test. In particular, we established its asymptotic optimality of order-2 in continuous time and of order-1 in discrete time. Moreover, we emphasized theoretically, as well as through simulations, the ability of the suggested scheme to improve significantly its performance when the sensors oversample their underlying Brownian motions, a property which is not enjoyed by other decentralized or centralized schemes. Overall, our decentralized detection method exhibits performance very close to that of the optimum centralized test and outperforms other decentralized tests in the literature.

We have assumed throughout this paper that the sensor processes are independent under each hypothesis. However, this is not always necessary for the implementation and proof of asymptotic optimality of the suggested scheme. Indeed, when the sensors observe continuously the paths of *correlated* drifted Brownian motions under each hypothesis, the suggested scheme can be implemented with a slight modification and its order-2 asymptotic optimality property remains valid [21]. In general, though, when the sensors observe correlated processes, finding an asymptotically optimum decentralized sequential scheme is a much more challenging task and remains an open problem.

Appendix

Proof of Lemma 1: To prove the lemma, note that

$$\frac{\mathsf{P}_1(z_n^i=0)}{\mathsf{P}_0(z_n^i=0)} = \mathsf{E}_0\left[e^{u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}}\,\middle|\, u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}} \le -\underline{\Delta}_i\right]. \tag{56}$$

Since

$$\mathsf{E}_0\left[e^{u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}}\,\middle|\, u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}} \le -\underline{\Delta}_i\right] \le e^{-\underline{\Delta}_i} \tag{57}$$

this proves (38). For (39), using Jensen's inequality in (56), we can write

$$\mathsf{E}_0\left[e^{u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}}\,\middle|\, u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}} \le -\underline{\Delta}_i\right] \ge e^{-\underline{\Delta}_i}e^{-\mathscr{D}} \tag{58}$$

where

$$\begin{aligned}\mathscr{D} &= \mathsf{E}_0\left[-(u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}+\underline{\Delta}_i)\,\middle|\, u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}} \le -\underline{\Delta}_i\right]\\ &= \frac{\mathsf{E}_0\left[-\left(u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}+\underline{\Delta}_i\right)\mathbb{1}_{\{u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}\le-\underline{\Delta}_i]\}}\right]}{\mathsf{P}_0(u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}} \le -\underline{\Delta}_i)}\\ &= \frac{\mathsf{E}_0\left[-\left(u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}+\underline{\Delta}_i\right)\mathbb{1}_{\{u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}}\le-\underline{\Delta}_i]\}}\right]}{1-\mathsf{P}_0\left(u^i_{\tau^i_n}-u^i_{\tau^i_{n-1}} \ge \bar{\Delta}_i\right)}\\ &\le \frac{\theta}{1-e^{-\bar{\Delta}_i}}\end{aligned} \tag{59}$$

where in the last inequality, we used the fact that the numerator is an overshoot and therefore bounded by θ and in the denominator we used Wald's approximation (which provides an upper bound) for the error probability of the local SPRT exiting from the wrong side. Replacing the bound for $\mathscr{D}$ in (58), taking the logarithm and recalling (38), we conclude

$$0 \le \underline{\Lambda}_i - \underline{\Delta}_i \le \frac{\theta}{1-e^{-\bar{\Delta}_i}}. \tag{60}$$

Assuming that $\bar{\Delta}_i$ is bounded away from 0, the previous right-hand side becomes $O(\theta)$ and proves the lemma. ■

Proof of Lemma 2: Let us prove the first inequality in (41). Note that

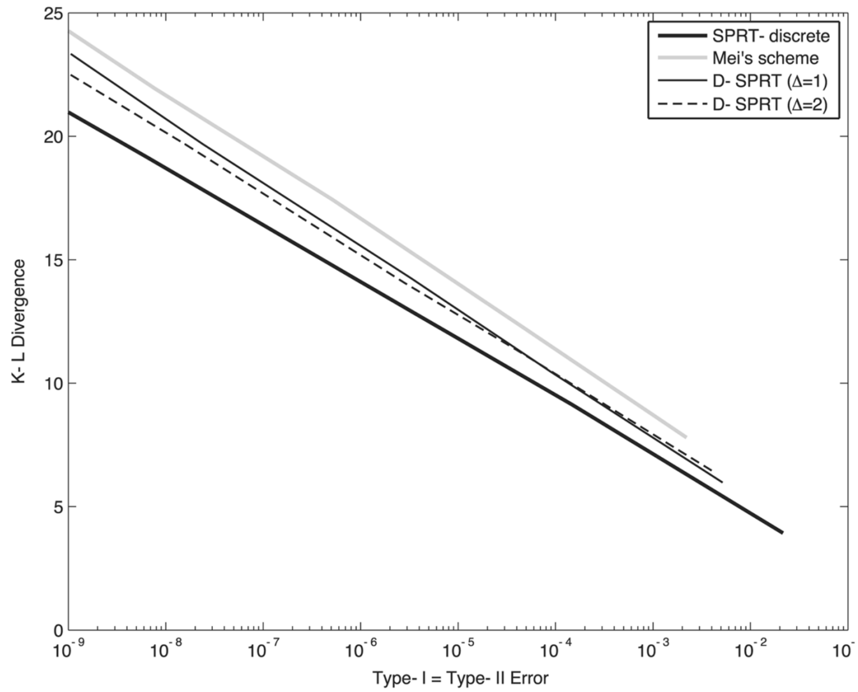
$$I_0^i = -\mathsf{E}_0\left[\lambda_n^i\right] = \frac{\underline{\Lambda}_i(e^{\bar{\Lambda}_i}-1)+\bar{\Lambda}_i(e^{-\underline{\Lambda}_i}-1)}{e^{\bar{\Lambda}_i}-e^{-\underline{\Lambda}_i}} > 0. \tag{61}$$

By direct differentiation we can verify that the function $\mathcal{K}(x,y) = \{x(e^y-1)+y(e^{-x}-1)\}/(e^y-e^{-x})$ is monotonically increasing in both its arguments, when $x,y \ge 0$. Consequently, from (38), namely that $\underline{\Lambda}_i, \bar{\Lambda}_i$ exceed $\underline{\Delta}_i, \bar{\Delta}_i$ respectively, we immediately deduce the final inequality. Proving (42) is straightforward. ■

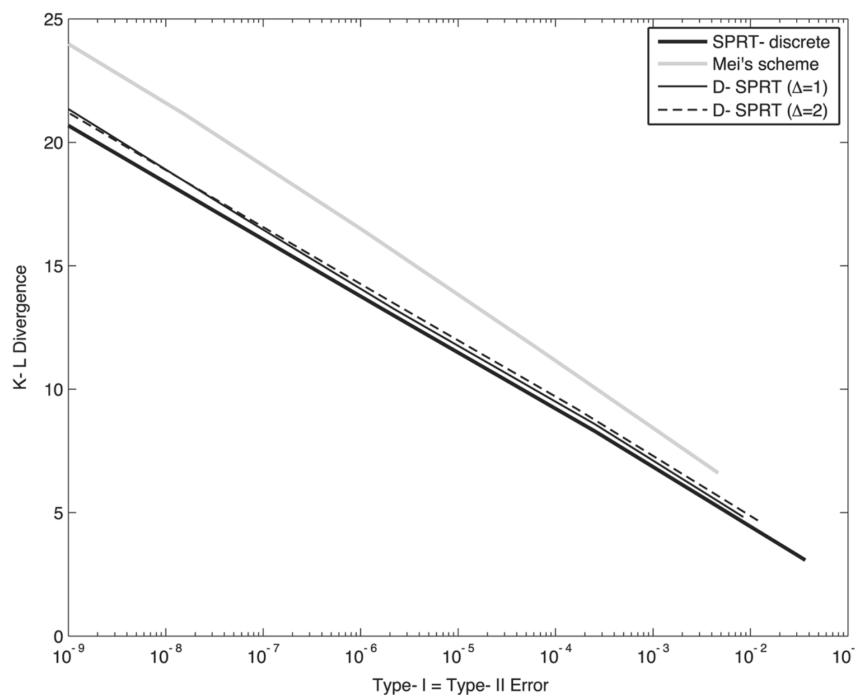
Proof of Lemma 3: For simplicity, we drop the subscript j that refers to the true hypothesis. We observe that

$$\mathsf{E}\left[\sum_{n=1}^{m_T^i+1}\zeta_n^i\right] = \mathsf{E}\left[\sum_{n=1}^{\infty}\zeta_n\mathbb{1}_{\{m_T^i\ge n-1\}}\right]. \tag{62}$$

Note that $\{m_T^i \ge n-1\} = \{T \ge \tau^i_{n-1}\}$. By recalling that τ^i_{n-1} is an $\{\mathscr{F}^i_t\}$-adapted stopping time, this suggests that it is also $\{\mathscr{F}_t\}$-adapted. Because of the latter observation we can assess that the event $\{T \ge \tau^i_{n-1}\}$ is $\mathscr{F}_{\tau^i_{n-1}-1}$-measurable (since $\{T \ge \tau\}$ is $\mathscr{F}_{\tau-1}$-measurable, this being true even if τ is an $\{\mathscr{F}_t\}$-adapted stopping time). Consequently, ζ_n^i is independent of $\mathbb{1}_{\{m_T^i\ge n-1\}}$. Using this observation and interchanging summation and expectation in (62), we have



(a)



(b)

Fig. 3. Relative performance of centralized and decentralized tests in discrete time with $K = 2$ sensors and testing between H_0 : Normal $\mathcal{N}(0, h)$ and H_1 : Normal $\mathcal{N}(h, h)$ random variables with (a) $h = 1.0$ and (b) $h = 0.1$.

$$\begin{aligned}\mathsf{E}\left[\sum_{n=1}^{\infty}\zeta_n\mathbb{1}_{\{m_T^i\geq n-1\}}\right] &= \sum_{n=1}^{\infty}\mathsf{E}\left[\zeta_n\mathbb{1}_{\{m_T^i\geq n-1\}}\right]\\ &= \mathsf{E}[\zeta_1]\sum_{n=1}^{\infty}\mathsf{P}\left(m_T^i\geq n-1\right)\\ &= \mathsf{E}[\zeta_1]\,\mathsf{E}[m_T] \end{aligned}\tag{63}$$

which is what we wanted to prove.

Notice that interchanging summation and expectation is straightforward when the ζ_n's are positive random variables. In the general case we can write $\zeta_n^i = \max\{\zeta_n^i, 0\} - \max\{-\zeta_n^i, 0\}$. The result then follows if we assume $\mathsf{E}[\max\{\zeta_n^i, 0\}] < \infty$ and $\mathsf{E}[\max\{-\zeta_n^i, 0\}] < \infty$ or equivalently $\mathsf{E}[|\zeta_n^i|] < \infty$, which is exactly what we have assumed.

Proof of Lemma 4: To prove (46), note that $|u_t - \tilde{u}_t| \leq \sum_{i=1}^{K}|u_t^i - \tilde{u}_t^i|$. Using (40) we observe that we can write

$$\left|u_t^i - \tilde{u}_t^i\right| \le \left|u_t^i - u_{\tau_{m_t^i}}^i\right| + \sum_{n=1}^{m_t^i} \left|\left[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i\right] - \lambda_n^i\right|. \tag{64}$$

From the definition of the Lebesgue sampling we have $|u_t^i - u_{\tau_{m_t^i}}^i| \le \underline{\Delta}_i + \bar{\Delta}_i$. Now note that if $u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i$ exits from the lower end, then $|[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] - \lambda_n^i| = |[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] + \underline{\Lambda}_i| \le |[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] + \underline{\Delta}_i| + (\underline{\Lambda}_i - \underline{\Delta}_i)$, with the last inequality coming from (38). Similarly, if $u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i$ exits from the upper end, then $|[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] - \lambda_n^i| \le |[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] - \bar{\Delta}_i| + (\bar{\Lambda}_i - \bar{\Delta}_i)$. In both cases, we see that $|[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] - \lambda_n^i| \le |\eta_n^i| + R_i$, with η_n^i the overshoot defined in (36). Consequently, we can further upper bound (64) using the overshoot. Replacing t with T then taking expectation and using (44), we obtain

$$\begin{aligned}
&\mathsf{E}\left[\left|u_T^i - \tilde{u}_T^i\right|\right] \\
&\quad \le \underline{\Delta}_i + \bar{\Delta}_i + \mathsf{E}\left[\left|\eta_n^i\right|\right]\left(\mathsf{E}\left[m_T^i\right] + 1\right) \\
&\quad \le \underline{\Delta}_i + \bar{\Delta}_i + \max_i \mathsf{E}\left[\left|\eta_n^i\right|\right]\left(\mathsf{E}\left[m_T^i\right] + 1\right).
\end{aligned} \tag{65}$$

Summing over i and recalling the definition of θ in (44), we obtain

$$\mathsf{E}[|u_T - \tilde{u}_T|] \le O\theta\left(\sum_{i=1}^{K} \mathsf{E}\left[m_T^i\right] + K\right) + C. \tag{66}$$

Using now (40), we can write

$$-\mathsf{E}_0\left[u_T^i\right] = -\mathsf{E}_0\left[\sum_{n=1}^{m_T^i} \lambda_n^i\right] \ge -\mathsf{E}_0\left[\lambda_n^i\right]\mathsf{E}_0\left[m_T^i\right] - 2(\underline{\Lambda}_i + \bar{\Lambda}_i) \tag{67}$$

where for the last inequality, we used (45) and the fact that $|\lambda_n^i| \le \underline{\Lambda}_i + \bar{\Lambda}_i$. Since by definition $I_0^i = -\mathsf{E}_0[\lambda_n^i]$ is the K-L information number for the random sequence $\{\lambda_n^i\}$, we strengthen the inequality by minimizing over i. Summing the result over i yields

$$-\mathsf{E}_0[u_T] \ge \left(\min_i I_o^i\right) \sum_{i=1}^{K} \mathsf{E}_0\left[m_T^i\right] - 2C'. \tag{68}$$

Solving for the sum and replacing in (66) yields the desired inequality under H_0. Similar proof applies under H_1. ■

Proof of Theorem 2: The most important part of the proof is demonstrating the validity of the estimates in (47). For this task, we need to introduce some additional notation; thus, we denote by z_n the nth binary message that arrives at the fusion center *irrespectively of the sensor which sent it* and by k_n the identity of the sensor which transmitted the nth message. The flow of information at the fusion center is then described by the filtration $\{\mathcal{C}_n\}$, where $\mathcal{C}_n = \sigma((z_1, k_1)\ldots,(z_n, k_n))$. In order to avoid technical complications we assume that at any given time the fusion center receives at most one message from the sensors. The extension to the general case is merely technical and we refer to [21] for details.

The fusion center likelihood under H_j after the arrival of the first n messages is

$$\begin{aligned}
&\mathsf{P}_j((z_1, k_1), \ldots, (z_n, k_n)) \\
&\quad = \mathsf{P}_j(k_1, \ldots, k_n) \prod_{l=1}^{n} \mathsf{P}_j(z_l \,|\, z_1, \ldots, z_{l-1}, k_1, \ldots, k_n) \\
&\quad = \mathsf{P}_j(k_1, \ldots, k_n) \prod_{l=1}^{n} \mathsf{P}_j(z_l \,|\, k_l) \\
&\quad = \mathsf{P}_j(k_1, \ldots, k_n) \prod_{l=1}^{n} \mathsf{P}_j^{k_l}(z_l).
\end{aligned} \tag{69}$$

The first equality uses simply the definition of conditional probability. The second equality is based on the fact that z_l (the value of the lth transmitted message at the fusion center) is independent of all other messages $\{(z_j, k_j), j \neq l\}$ *conditionally on* k_l (the identity of the sensor from which the lth message was transmitted). For the third equality, we simply used our notation that the probability measure of a sample coming from sensor i is denoted as P_j^i, where j refers to the true hypothesis.

The likelihood ratio after the arrival of the nth message is

$$\frac{\mathsf{P}_1((z_1, k_1), \ldots, (z_n, k_n))}{\mathsf{P}_0((z_1, k_1)\ldots,(z_n, k_n))} = \phi_n\, e^{\tilde{v}_n} \tag{70}$$

where—recalling the definition of the log-likelihood ratios $\bar{\Lambda}_i, \underline{\Lambda}_i$—we define

$$\begin{aligned}
\phi_n &= \frac{\mathsf{P}_1(k_1, \ldots, k_n)}{\mathsf{P}_0(k_1, \ldots, k_n)} \\
\tilde{v}_n &= \sum_{j=1}^{n} [\bar{\Lambda}_{k_j}\, z_j + \underline{\Lambda}_{k_j}\,(1 - z_j)].
\end{aligned} \tag{71}$$

The process $\tilde{v}_n$ is of course closely related to the process $\tilde{u}_t$, we defined in (21). Note that $\tilde{u}_t$ is expressed in terms of the global time t whereas $\tilde{v}_n$ in terms of the *number of messages* n received by the fusion center. To explicitly specify their dependence, let $\{\tau_n\}$ be the increasing sequence of communication times between *any* sensor and the fusion center, where τ_n is the time instant (in global time) that the fusion center receives its nth message. Then the two processes are related through the equality $\tilde{v}_n = \tilde{u}_{\tau_n}$.

The fusion center policy defined in (22) can also be expressed in terms of number of messages at the fusion center as

$$\begin{aligned}
\tilde{\mathscr{N}} &= \inf\{n \in \mathbb{N} : \tilde{v}_n \notin (-\tilde{A}, \tilde{B})\} \\
d_{\tilde{\mathscr{N}}} &= \begin{cases} 1, & \text{if } \tilde{v}_{\tilde{\mathscr{N}}} \ge \tilde{B} \\ 0, & \text{if } \tilde{v}_{\tilde{\mathscr{N}}} \le -\tilde{A} \end{cases}
\end{aligned} \tag{72}$$

and we clearly have $\tilde{\mathscr{T}} = \tau_{\tilde{\mathscr{N}}}$ and $d_{\tilde{\mathscr{T}}} = d_{\tilde{\mathscr{N}}}$. Now, $\tilde{\mathscr{N}}$ is a $\{\mathcal{C}_n\}$-stopping time which represents the number of messages that are collected by the fusion center until a decision is reached by D-SPRT, whereas $d_{\tilde{\mathscr{N}}}$ is $\mathcal{C}_{\tilde{\mathscr{N}}}$-measurable random variable which represents the D-SPRT decision rule. Since $\{d_{\tilde{\mathscr{N}}} = 0\} = \{\tilde{v}_{\tilde{\mathscr{N}}} \le -\tilde{A}\} \in \mathcal{C}_{\tilde{\mathscr{N}}}$, with a change of measure we have

TABLE I
NOTATION FOR LEBESGUE SAMPLING IN DISCRETE TIME

ξ_t^i	observed value at sensor i at time t
ℓ_t^i	log-likelihood ratio of ξ_t^i
u_t^i	log-likelihood ratio at sensor i up to time t
u_t	global log-likelihood ratio up to time t
$\overline{\Delta}_i(-\underline{\Delta}_i)$	upper (lower) local threshold at sensor i
τ_n^i	time of nth transmission from sensor i
z_n^i	transmitted message from sensor i at time τ_n^i
λ_n^i	log-likelihood ratio of z_n^i
η_n^i	overshoot of $u_t^i - u_{\tau_{n-1}}^i$ above $\overline{\Delta}_i$ or below $-\underline{\Delta}_i$
$\overline{\Lambda}_i(-\underline{\Lambda}_i)$	log-likelihood ratio of the event $\{z_n^i = 1\}$ ($\{z_n^i = 0\}$)
m_t^i	number of transmitted messages from sensor i up to time t
$\tau_{m_t^i}^i$	most recent communication time for sensor i until time t
$\tilde{u}_t^i$	approximation of u_t^i at the fusion center
$\tilde{u}_t$	approximation of u_t at the fusion center
τ_n	time of nth message received by the fusion center
$\tilde{v}_n$	global log-likelihood ratio of the first n messages

$$\beta = \mathsf{P}_1(d_{\tilde{\mathscr{N}}} = 0) = \mathsf{E}_1\left[\mathbb{1}_{\{\tilde{v}_{\tilde{\mathscr{N}}} \le -\tilde{A}\}}\right]$$
$$= \mathsf{E}_0\left[e^{\tilde{v}_{\tilde{N}}}\,\phi_{\tilde{\mathscr{N}}}\,\mathbb{1}_{\{\tilde{v}_{\tilde{\mathscr{N}}} \le -\tilde{A}\}}\right] \le e^{-\tilde{A}}\,\mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}] \qquad (73)$$

and taking logarithms in both sides, we obtain $\tilde{A} \le |\log\beta| + \log \mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}]$, thus it suffices to show that $\mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}] = 1$. But $\{\phi_n\}$ is a likelihood ratio, thus it is a $(\mathsf{P}_0, \{\mathcal{C}_n\})$-martingale with P_0-expectation equal to 1. Therefore, it suffices to show that we can apply optional sampling theorem. This is indeed possible due to the special form of the $\{\mathcal{C}_n\}$-stopping time $\tilde{\mathscr{N}}$. Since $\tilde{\mathscr{N}}$ is P_0-a.s. finite, it suffices to show that $\mathsf{E}_0[|\phi_{\tilde{\mathscr{N}}}|] < \infty$ and $\lim_{n\to\infty} \mathsf{E}_0[\phi_n \mathbb{1}_{\{n<\tilde{\mathscr{N}}\}}] = 0$. Since ϕ_n is a $\mathcal{C}_n$-measurable random variable and $\{n < \tilde{\mathscr{N}}\} \in \mathcal{C}_n$, from a change of measure, we obtain

$$\mathsf{E}_0\left[\phi_n \mathbb{1}_{\{n<\tilde{\mathscr{N}}\}}\right] = \mathsf{E}_1\left[e^{-\tilde{v}_n}\phi_n^{-1}\phi_n \mathbb{1}_{\{n<\tilde{\mathscr{N}}\}}\right]$$
$$= \mathsf{E}_1\left[e^{-\tilde{v}_n}\mathbb{1}_{\{n<\tilde{N}\}}\right]$$
$$\le e^{\max\{\tilde{A},\tilde{B}\}}\mathsf{P}_1(n < \tilde{\mathscr{N}}) \to 0 \qquad (74)$$

as $n \to \infty$. Notice that the inequality is due to the fact that $-\tilde{A} < \tilde{v}_n < \tilde{B}$ for $n < \tilde{N}$, whereas for the limit we have used the fact that $\tilde{N}$ is P_1-a.s. finite.

Similarly, we have

$$\mathsf{E}_0[|\phi_{\tilde{\mathscr{N}}}|] = \mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}] = \mathsf{E}_1\left[e^{-\tilde{v}_{\tilde{\mathscr{N}}}}\phi_{\tilde{\mathscr{N}}}^{-1}\phi_{\tilde{\mathscr{N}}}\right]$$
$$= \mathsf{E}_1[e^{-\tilde{v}_{\tilde{\mathscr{N}}}}] \le e^{\max\{\tilde{A},\tilde{B}\}+C'} < \infty. \qquad (75)$$

Therefore, we can apply the optional sampling theorem and obtain $\mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}] = 1$, which proves the first inequality in (47). The second inequality can be shown in an analogous way.

To prove the second part of the theorem, namely (48), we return to the global time and we consider the inequality under H_0. Note that

$$\mathsf{E}_0[u_{\tilde{\mathscr{T}}}] \ge \mathsf{E}_0[\tilde{u}_{\tilde{\mathscr{T}}}] - \mathsf{E}_0[|u_{\tilde{\mathscr{T}}} - \tilde{u}_{\tilde{\mathscr{T}}}|]. \qquad (76)$$

Using (46) from Lemma 4, the inequality becomes

$$\mathsf{E}_0[u_{\tilde{\mathscr{T}}}] \ge (1+\Phi)\mathsf{E}_0[\tilde{u}_{\tilde{\mathscr{T}}}] - C - 2\Phi C' - K\max_i \mathsf{E}_0[|\eta_n^i|] \qquad (77)$$

where $\Phi = O(\theta)/(\min_i I_0^i)$. As in the continuous time case, we have $\tilde{u}_{\tilde{\mathscr{T}}} \ge -\tilde{A} - C'$ and using (47) we can write $\tilde{u}_{\tilde{\mathscr{T}}} \ge -|\log\beta| - C'$ which also implies $\mathsf{E}_0[\tilde{u}_{\tilde{\mathscr{T}}}] \ge -|\log\beta| - C'$. Replacing the latter in (77) results in

$$\mathsf{E}_0[u_{\tilde{\mathscr{T}}}] + |\log\beta| \ge -\Phi|\log\beta|$$
$$-(1+3\Phi)C' - C - K\max_i \mathsf{E}_0[|\eta_n^i|]. \qquad (78)$$

If we replace, in the left-hand side of the previous inequality, $|\log\beta|$ with the optimum performance $-\mathsf{E}_0[u_{\mathscr{T}}]$, because of (33), we strengthen the inequality obtaining

$$(-\mathsf{E}_0[u_{\tilde{\mathscr{T}}}]) - (-\mathsf{E}_0[u_{\mathscr{T}}]) \le \Phi|\log\beta|$$
$$+(1+3\Phi)C' + C + K\max_i \mathsf{E}_0[|\eta_n^i|] + o(1). \qquad (79)$$

Note now that $C = \Theta(\Delta)$ and for the overshoot we have $\max_i \mathsf{E}_0[|\eta_n^i|] \le \theta$. In our analysis we consider Δ to be, either of the order of a constant or to tend to infinity and θ to be either of the order of a constant or to tend to 0. Because of this assumption and Lemma 1, we have $\underline{\Delta}_i, \bar{\Lambda}_i$ that are $\Theta(\Delta)$ meaning that $C' = \Theta(\Delta)$. Because of Lemma 2, we conclude that $\min_i I_0^i \ge \Theta(\Delta)$, consequently $\Phi \le O(\theta)/\Theta(\Delta)$. Substituting these order of magnitudes in (79) yields

$$(-\mathsf{E}_0[u_{\tilde{\mathscr{T}}}]) - (-\mathsf{E}_0[u_{\mathscr{T}}])$$
$$= \frac{O(\theta)}{\Theta(\Delta)}|\log\beta| + \Theta(\Delta) + O(\theta) + o(1). \qquad (80)$$

Finally, due to the relative size of Δ and θ, we can also conclude that $\Theta(\Delta) + O(\theta) + o(1) = \Theta(\Delta)$ which proves the desired version of the inequality. Similar steps can be applied to prove the theorem under hypothesis H_1. ■

REFERENCES

[1] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.

[2] D. Siegmund, *Sequential Analysis, Tests and Confidence Intervals*. New York: Springer, 1985.

[3] B. K. Ghosh and P. K. Sen, *Handbook of Sequential Analysis*. New York: Marcel Dekker, 1991.

[4] R. R. Tenney and N. R. Sandell, Jr., "Detection with distributed sensors," *IEEE Trans. Aerosp. Elect. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.

[5] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*. Greenwich, CT: JAI Press, 1990.

[6] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 433–442, Mar. 1993.

[7] J. N. Tsitsiklis, "On threshold rules in decentralized detection," in *Proc. 15th IEEE Conf. Decision Control*, Athens, Greece, 1986, pp. 232–236.

[8] H. R. Hashemi and I. B. Rhodes, "Decentralized sequential detection," *IEEE Trans. Inf. Theory*, vol. 35, no. 3, pp. 509–520, May 1989.

[9] V. V. Veeravalli, "Comments on 'Decentralized sequential detection'," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1428–1429, Jul. 1992.

[10] V. V. Veeravalli, "Sequential decision fusion: Theory and applications," *J. Franklin Inst.*, vol. 336, pp. 301–322, 1999.

[11] Y. Mei, "Asymptotic optimality theory for sequential hypothesis testing in sensor networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2072–2089, May 2008.
[12] V. N. S. Samarasooriya and P. K. Varshney, "Sequential approach to asynchronous decision fusion," *Opt. Eng.*, vol. 35, no. 3, pp. 625–633, 1996.
[13] A. M. Hussain, "Multisensor distributed sequential detection," *IEEE Trans. Aerosp. Elect. Syst.*, vol. 30, no. 3, pp. 698–708, Jul. 1994.
[14] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Statist.*, vol. 19, pp. 326–339, 1948.
[15] R. L. Liptser and A. N. Shiryaev, *Statistics of Random Processes, II Applications*, 2nd ed. New York: Springer, 2001.
[16] A. N. Shiryaev, *Optimal Stopping Rules*. New York: Springer, 1978.
[17] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed. New York: Springer, 1991.
[18] A. Irle, "Extended optimality of sequential probability ratio tests," *Ann. Statist.*, vol. 12, pp. 380–386, 1984.
[19] G. Peskir and A. N. Shiryaev, "Sequential Testing Problems for Poisson Processes," *Ann. Statist.*, vol. 28, no. 3, pp. 837–859, 2000.
[20] G. Lorden, "On excess over the boundary," *Ann. Math. Stat.*, vol. 41, no. 2, pp. 520–527, 1970.
[21] G. Fellouris, "Decentralized Sequential Decision Making With Asynchronous Communication," Ph.D. dissertation, Dept. Statistics, Columbia Univ., New York, 2010.

Georgios Fellouris received the diploma in applied mathematics in 2004 and the Masters in Mathematical Modeling in the Economy and the Technology in 2005 from the School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Greece. In 2010, he received the Ph.D. in statistics from the Department of Statistics, Columbia University, New York.

He currently holds a post-doctoral position at the Department of Mathematics, University of Southern California, Los Angeles. His research interests lie in the areas of statistical inference for stochastic processes, sequential analysis, decentralized detection and estimation.

George V. Moustakides (M'82–SM'97) was born in Drama, Greece, in 1955. He received the diploma in electrical and mechanical engineering from the National Technical University of Athens, Greece, in 1979, the M.Sc. degree in systems engineering from the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, in 1980, and the Ph.D degree in electrical engineering and computer science from Princeton University, Princeton NJ, in 1983.

From 1983 to 1986, he was with INRIA, France, as a Junior Researcher and during 1987 to 1990 he held a research position at the Computer Technology Institute of Patras, Greece. In 1991 he joined the Computer Engineering and Informatics department, University of Patras, Greece, as an Associate Professor, and in 1996 he became Professor in the same department. From 2002 to 2006, he was a faculty member with the department of Computer and Communication Engineering, University of Thessaly, Volos, Greece and in 2007 he joined the department of Electrical and Computer Engineering, University of Patras, Patras, Greece. From 2001 to 2004, he was also a Senior Researcher (Directeur de Recherche) with INRIA, France. His interests include wireless communications, statistical signal processing, sequential detection and signal processing for hearing aids.