Asymptotically optimum tests for decentralized sequential testing in continuous time

Georgios Fellouris
Department of Statistics
Columbia University
New York, USA
Email: gaf2106@columbia.edu

George V. Moustakides
Department of Electrical and Computer Engineering
University of Patras,
26500 Patras, Greece
Email: moustaki@upatras.gr

***Abstract*— We propose an asymptotically optimum test for the problem of decentralized sequential hypothesis testing in continuous time, in the case where the sensors have full local memory and no feedback from the fusion center. According to our scheme, the sensors perform locally repeated SPRTs and communicate, asynchronously, their one-bit decisions to the fusion center. The fusion center in turn uses the received information to perform a centralized SPRT in order to make the final decision. The expected time for a decision of the proposed scheme differs from the optimum continuous-time centralized SPRT only by a constant. This fact suggests order-2 asymptotic optimality of our test as compared to existing schemes that are optimal of order-1. Moreover, simulation experiments reveal that the performance of our scheme is significantly better than that of the discrete-time centralized SPRT.**

Keywords: SPRT, Decentralized, Sequential Testing.

SPECIAL SESSION: Distributed Inference and Decision-Making in Multisensor Systems
ORGANIZERS: A. Tartakovsky & V. Veeravalli.

I. Introduction

Consider the geometry depicted in Fig. 1 where K sensors observe K statistically independent *continuous-time* processes $\{\xi_{t,i}\}_{t\geq 0},\ i = 1,\dots,K$. There are two simple hypotheses about these processes, H_0 and H_1. According to H_0, the law of the process $\{\xi_{t,i}\}_{t\geq 0}$ is described by the probability measure $\mathbb{P}_{0,i}$, whereas according to H_1 by $\mathbb{P}_{1,i}$, $i = 1,\dots,K$. The goal is to choose between the two hypotheses as soon as possible. The decision is made at a *fusion center* which receives, sequentially, *discrete-time* information $\{z_{n,i}\}_{n\geq 0}, n \in \mathbb{Z}$, conveyed from the sensors with the help of standard (wireless) digital communication systems. In order to limit the need in communication bandwidth between sensors and fusion center the communication rate must be kept low. This requirement practically excludes transmission of samples obtained from the continuous-time signals by conventional deterministic canonical sampling and demands for sampling strategies that are more efficient.

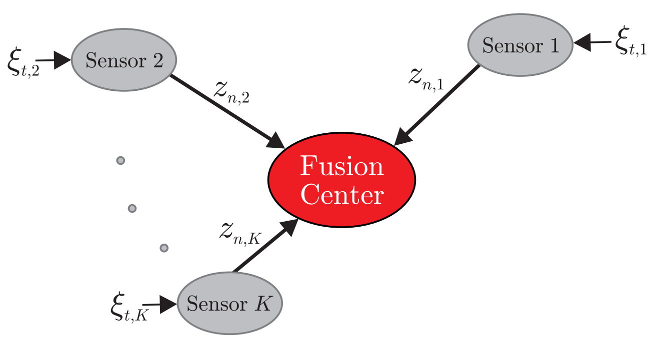

Fig. 1. Geometry of the decentralized hypothesis testing problem.

We note that a decentralized decision strategy is comprised of two parts 1) the *sampling strategy* at the sensors and 2) the fusion center *decision policy*. Sampling strategies define the type of information to be transmitted from the sensors to the fusion center , whereas decision policies how this information should be utilized by the fusion center to produce its final decision. We also distinguish the decentralized schemes from the *centralized* structures in which the fusion center has complete access to the continuous-time processes $\{\xi_{t,i}\}$. It is clear that the application of a centralized optimum test gives rise to the *ultimate point of reference* in performance.

The decentralized detection problem was first introduced by Tsitsiklis [7]. Later, in Veerevali et.al [8] we find a detailed presentation of the different sampling models that can be defined. In this work, Veeravalli et.al produce -under a Bayesian setting- optimum schemes for the models with no and partial local memory i.e. when each sensor has access to the past decisions of all sensors, but not access to its own previous observations or only to its current observation. Mei [3], also under a Bayesian setting, developed a sampling/detection scheme for the model with full local memory and no feedback, i.e. when sensors remember all their past acquired samples, but they have no access to the previous decisions of the other sensors. The proposed test was shown to be asymptotically optimum, in the sense that the ratio of its performance and the performance of the centralized test tends to 1, as the appropriate error probability tends to 0.

The current decentralized literature on this problem mainly refers to discrete-time signals and to *synchronous* communication between sensors and fusion center, silently assuming the existence of a *global clock*. Transferring this methodology to the continuous-time case requires the processes $\{\xi_{t,i}\}$ to be sampled concurrently, using canonical sampling. The acquired samples $\xi_{nT,i}$ (T being the constant sampling period) need to be further processed with the help of an additional sampler (more accurately quantizer) in order to produce the signals $\{z_{n,i}\}$ to be transmitted. A discussion of asynchronous sampling in the problem of distributed sequential hypothesis testing can be found in Samarasooriya et.al [4], where the number of local decisions of the sensors is modeled by a Poisson process.

Moreover, the work that has been done so far in the problem of distributed hypothesis testing follows mainly the Bayesian approach. A very interesting exception can be found in Hussain [1], where the frequentist approach is followed and the suggested test in this work is the analogue in discrete time of the test that we present here. However in [1] no theoretical justification of any form is provided regarding the proposed scheme.

In this work, we examine sensors with *full local memory* and we assume that there is no feedback from the fusion center. We take the frequentist approach, i.e our goal is to minimize the expected time for a decision under each hypothesis for given Type I and Type II error probabilities.

The rest of the paper is organized as follows: in Section II we introduce the notation that we will use and we define mathematically the problem. Moreover, we discuss the optimal centralized test and define its random-sampling modification. In Section III, we develop our test and in Section IV we prove its order-2 asymptotic optimality in the case where each observed continuous-time signal at each sensor is a Brownian motion with constant drift (not necessarily the same in all sensors). In Section V, we develop an exact simulation algorithm for the simulation of the suggested scheme and we compare it with the optimum continuous-time test, its discrete version and Mei's [3] approach. We conclude in Section VI.

II. Problem formulation and random sampling

A. Problem formulation

Let $(\Omega_i, \mathcal{F}_{\infty,i})$ be a probability space, on which we define the stochastic process $\{\xi_{t,i}\}_{t\geq 0}$, whose statistics are given by $\mathbb{P}_{0,i}$ under the null hypothesis H_0 and by $\mathbb{P}_{1,i}$ under the alternative hypothesis H_1, where $\mathbb{P}_{0,i}$ and $\mathbb{P}_{1,i}$ are probability measures on $(\Omega_i, \mathcal{F}_{\infty,i}), i = 1, \ldots, K$.

If we consider now the probability space $(\Omega, \mathcal{F}_\infty)$, where $\Omega = \Omega_1 \times \ldots \times \Omega_K$ and $\mathcal{F}_\infty = \mathcal{F}_{\infty,1} \times \ldots \times \mathcal{F}_{\infty,K}$, the K-dimensional process $\{\xi_t = (\xi_{t,1}, \ldots, \xi_{t,K})\}_{t\geq 0}$ is defined on this space and we have the following hypothesis testing problem about its law, which we we will denote by $\mathbb{P}$:

$$H_0 : \mathbb{P} = \mathbb{P}_0 \ , \ H_1 : \mathbb{P} = \mathbb{P}_1 \tag{1}$$

where $\mathbb{P}_0 = \mathbb{P}_{0,1} \times \ldots \times \mathbb{P}_{0,K}$ and $\mathbb{P}_1 = \mathbb{P}_{1,1} \times \ldots \times \mathbb{P}_{1,K}$.

Let now $\{\mathcal{F}_{t,i}\}$ be the filtration generated by the observed continuous-time signal $\{\xi_{t,i}\}$ and $\{u_{t,i}\}$ the local log-likelihood ratio at Sensor-i. Then:

$$u_{t,i} = \log \frac{d\mathbb{P}_{0,i}}{d\mathbb{P}_{\infty,i}} \left(\mathcal{F}_{t,i}\right). \tag{2}$$

Let also $\{\tilde{\mathcal{F}}_t\}$ be the filtration generated by the samples $\{z_{n,i}\}$ that are sent from the sensors to the fusion center, i.e. $\{\tilde{\mathcal{F}}_t\}$ is all the received information at the fusion center up to time t. We also distinguish the filtration $\{\mathcal{F}_t\}$ where $\mathcal{F}_t = \sigma\{u_{s,i};\ s \leq t; i = 1, \ldots, K\}$ is the received information at the fusion center for the continuous-time centralized test, i.e. the continuous-time signals from each sensor. Note that because $\{u_{t,i}\}$ is a sufficient statistic for Sensor-i in the centralized test, transmitting these signals instead of the observed $\{\xi_{t,i}\}$ produces no performance loss for the optimum decision structures.

Therefore, we can think of a decentralized decision strategy as a triplet $(\{\tilde{\mathcal{F}}_t\}, \tilde{\mathcal{T}}, \tilde{d})$,where $\{\tilde{\mathcal{F}}_t\}$ is the filtration at the fusion center, $\tilde{\mathcal{T}}$ a stopping time with respect to this filtration and $\tilde{d}$ a $\tilde{\mathcal{F}}_{\tilde{\mathcal{T}}}$- measurable random variable which takes values on the set $\{0, 1\}$. $\{\tilde{\mathcal{F}}_t\}$ reflects the sampling strategy, the structure of the information that we transmit from the sensors to the fusion center, whereas $(\tilde{\mathcal{T}}, \tilde{d})$ is the decision policy at the fusion center, which should be compatible with the chosen sampling strategy.

Thus, the decentralized hypothesis testing problem -under a non-Bayesian setting- is the minimization of the expected time for a decision under each hypothesis, for given Type I and Type II error probabilities, *jointly* over all possible sampling strategies at the sensors and decision policies at the fusion center.

B. The optimal centralized test

Let us now recall the optimal test for the solution of the centralized sequential hypothesis testing problem, which is Wald's Sequential Probability Ratio Test (SPRT) and is defined as follows:

$$u_t = u_{t,1} + u_{t,2} + \cdots + u_{t,K} \tag{3}$$

$$\mathcal{S} = \inf_{t\geq 0}\{t : u_t \notin (-A, B)\} \tag{4}$$

$$d_{\mathcal{S}} = \mathbf{1}_{\{u_{\mathcal{S}}=B\}}, \tag{5}$$

where $A, B > 0$ are two constant thresholds; $\mathcal{S}$ is the SPRT stopping time, i.e. the first time the test statistics leaves the open interval $(-A, B)$; and $d_{\mathcal{S}}$ is an $\mathcal{F}_{\mathcal{S}}$-measurable random variable, according to which H_0 is accepted if the lower threshold is crossed, whereas H_1 is accepted if the upper threshold is crossed. Optimality of the SPRT in continuous-time was established by Shiryaev [5] for Brownian Motion (BM) with constant drifts under each hypothesis. In particular, following Wald's setup [9], $\mathcal{S}$ solves the following constrained optimization problem

$$\inf_{\mathcal{T}} \mathbb{E}[\mathcal{T}]; \quad \text{subject to } \mathbb{P}_0[d_{\mathcal{T}} = 1] \leq \alpha \text{ and } \mathbb{P}_1[d_{\mathcal{T}} = 0] \leq \beta. \tag{6}$$

Here $\mathbb{E}[\cdot]$ denotes expectation with respect to the probability measure induced by either of the two hypotheses and $\alpha, \beta > 0$ are such that $\alpha + \beta < 1$. In other words, the SPRT minimizes the expected time for a decision under *both* hypotheses H_0 and H_1 among all sequential tests $(\mathcal{T}, d)$ with Type I and II error probabilities no larger than α and β, respectively. Time $\mathcal{T}$ is an integrable $\{\mathcal{F}_t\}$- stopping time and $d_{\mathcal{T}}$ and $\mathcal{F}_{\mathcal{T}}$-measurable random variable with values on $\{0, 1\}$.

Optimality of SPRT according to the above sense is guaranteed as long as the SPRT boundaries A, B are chosen so that the error probabilities are satisfied with equalities, which implies:

$$A = \log\left(\frac{1-\alpha}{\beta}\right),\ B = \log\left(\frac{1-\beta}{\alpha}\right). \quad (7)$$

or equivalently

$$\alpha = \frac{e^A - 1}{e^{A+B} - 1},\ \beta = \frac{e^B - 1}{e^{A+B} - 1}. \quad (8)$$

C. SPRT with adapted random sampling

Let $\{t_n^i\}$ be a strictly increasing sequence of sampling instances with $\lim_{n\to\infty} t_n^i = \infty\,(\mathbb{P}_0, \mathbb{P}_1\text{-a.s.})$, where each $t_{n,i}$ is a stopping time (s.t.) adapted to $\{\mathcal{F}_{t,i}\}$. Consider now the sampled version $\{u_{t_n^i,i}\}$ of the local log-likelihood ratio and suppose that these values are available at the fusion center at the sampling times $\{t_n^i\}$. Replacing the continuous-time log-likelihood ratios $\{u_{t,i}\}$ with their sampled versions $\{u_{t_n^i,i}\}$, gives rise to the following test statistic

$$\tilde{u}_t = u_{t^1_{n_t},1} + \cdots + u_{t^K_{n_t},K} \quad (9)$$

where $t^i_{n_t} = \max_{t_n^i \geq 0}\{t_n^i : t_n^i \leq t\}$ is the last sampling instant before (and including) t at Sensor-i. In other words, at every time t we add the latest available local log-likelihood ratios. The test now at the fusion center continuous as in the case of the continuous-time SPRT and it is defined as follows

$$\tilde{\mathcal{S}} = \inf_{t\geq 0}\{t : \tilde{u}_t \notin (-\tilde{A}, \tilde{B})\}, \quad (10)$$

$$\tilde{d}_{\tilde{\mathcal{S}}} = \mathbf{1}_{\{\tilde{u}_{\tilde{\mathcal{S}}} \geq \tilde{B}\}}, \quad (11)$$

where again $\tilde{A}, \tilde{B}$ are selected to satisfy the constraints in the error probabilities with equality.

We observe that the Fusion decision policy $(\tilde{\mathcal{S}}, \tilde{d}_{\tilde{\mathcal{S}}})$ in this approach is based on the sequential information expressed with the help of the filtration $\{\tilde{\mathcal{F}}_t\}$, where $\tilde{\mathcal{F}}_t = \sigma\{u_{t_n^i,i};\ t_n^i \leq t; i = 1,\ldots,K\}$ denotes all the *asynchronously* received information at the fusion center up to time t.

Note that, transmitting the samples $\{u_{t_n^i,i}\}$ is equally difficult as in the case of deterministic canonical sampling with $t_n^i = nT$, since these quantities are real numbers. *Given* the latter sampling strategy, we recall that the resulting test at the Fusion center is the discrete-time SPRT which is also optimum (for discrete time), as long as the continuous-time signals $\{\xi_{t,i}\}$ are Levy processes.

In the next section we are going to introduce a suitable selection of sampling instances $\{t_n^i\}$ that allows for the communication of the samples $\{u_{t_n^i,i}\}$ simply by transmitting *1-bit* information.

III. Proposed sampling/detection strategies

Crucial point in being able to implement the test in (10) is the availability, at the fusion center, of the samples $\{u_{t_n^i,i}\}$ of the local log-likelihood ratios. We observe that we can write

$$u_{t_n^i,i} = [u_{t_n^i,i} - u_{t_{n-1}^i,i}] + [u_{t_{n-1}^i,i} - u_{t_{n-2}^i,i}] + \cdots + [u_{t_1^i,i} - u_{t_0^i,i}] \quad (12)$$

where we define $t_0^i = 0$ and assume that $u_{0,i} = 0$. It is therefore sufficient for Sensor-i to transmit the differences $[u_{t_n^i,i} - u_{t_{n-1}^i,i}]$ between consecutive sampling times. The key idea is to select the sequence of s.t. $\{t_n^i\}$ so that these differences constitute 1-bit information. In fact, as we shall see next, this is not very complicated.

For Sensor-i select before hand two boundaries $-A_i < 0 < B_i$ which are also known to the fusion center. Suppose that t_{n-1}^i is already set, then define t_n^i as

$$t_n^i = \inf_{t > t_{n-1}^i}\{t : u_{t,i} - u_{t_{n-1}^i,i} \notin (-A_i, B_i)\}.$$

If $\{u_{t,i}\}$ has continuous paths we observe that at time t_n^i the difference $u_{t_n^i,i} - u_{t_{n-1}^i,i}$ will hit either $-A_i$ or B_i and this information can be transmitted, at time t_n^i, to the fusion center using 1 bit. Indeed, if $z_{t_n^i,i}$ is the information to be transmitted we define

$$z_{t_n^i,i} = \begin{cases} 1 & \text{if } u_{t_n^i,i} - u_{t_{n-1}^i,i} \geq B_i \\ 0 & \text{if } u_{t_n^i,i} - u_{t_{n-1}^i,i} \leq -A_i. \end{cases}$$

Once t_n^i has been set we repeat the same process for $t_{n+1}^i, \ldots,$ etc. The procedure we just described is simply a *repeated* Sequential Probability Ratio Test (SPRT) where every time the test statistics $u_{t,i} - u_{t_{n-1}^i,i}$ hits one of the two boundaries $-A_i, B_i$ we restart the SPRT by updating the term $u_{t_{n-1}^i,i}$. The interesting point is that by properly selecting the two boundaries $-A_i, B_i$ we have complete control over the *average sampling period* since the latter is simply the "detection delay" of the corresponding SPRT.

At the fusion center, whenever we receive a new information bit from any sensor, we update the corresponding local log-likelihood ratio from (12) and we apply the test according to (10). In fact, it is possible to directly update $\tilde{u}_t$ as follows: suppose that at time t the fusion center receives the bit $z_{t,i}$ from Sensor-i, then

$$\tilde{u}_t = \tilde{u}_{t-} + (1 - z_{t,i})(-A_i) + z_{t,i}B_i, 0 \quad (13)$$

where $\tilde{u}_{t-}$ denotes the test statistics right before the bit arrival. It is not difficult to verify that (13) yields the same update as computing $\tilde{u}_t$ following (9).

IV. The Brownian Motion case

Let us now focus on the special case where $\{\xi_{t,i}\}$ is a standard BM with drift equal to 0 under H_0 and $\mu_i \neq 0$ under H_1. We can then verify that $u_{t,i} = -0.5\mu_i^2 t + \mu_i \xi_{t,i}$. First we treat the centralized SPRT test. For this case we have the following convenient formulas:

$$\mathbb{E}_0[\mathcal{S}] = \frac{2}{\mu_1^2 + \cdots + \mu_K^2}\left[\frac{A(e^B - 1) - B(e^A - 1)}{e^{A+B} - 1}\right] \quad (14)$$

$$\mathbb{E}_1[\mathcal{S}] = \frac{2}{\mu_1^2 + \cdots + \mu_K^2}\left[\frac{B(e^A - 1) - A(e^B - 1)}{e^{A+B} - 1}\right]. \quad (15)$$

Regarding now the proposed test, the next theorem demonstrates that our scheme differs from the optimum, only by a bounded quantity.

Theorem 1: Let $\xi_{t,i}, i = 1, \ldots, K$, be as above and fix the local boundaries $-A_i < 0 < B_i, i = 1, \ldots, K$. At the Fusion center consider the proposed test $\tilde{\mathcal{S}}$ with boundaries $-\tilde{A} < 0 < \tilde{B}$. Let also $\mathcal{S}$ denote the optimal continuous-time centralized SPRT test with thresholds $-A < 0 < B$ and assume that both pairs (A, B) and $(\tilde{A}, \tilde{B})$ are chosen so that the corresponding decision strategies $(\mathcal{S}, d_{\mathcal{S}})$ and $(\tilde{\mathcal{S}}, d_{\tilde{\mathcal{S}}})$ satisfy the error probability constraints with equality, then we have

$$0 \leq \mathbb{E}_0[\tilde{\mathcal{S}}] - \mathbb{E}_0[\mathcal{S}] \leq \tilde{C},$$

$$0 \leq \mathbb{E}_1[\tilde{\mathcal{S}}] - \mathbb{E}_1[\mathcal{S}] \leq \tilde{C},$$

uniformly over β and α, where $\tilde{C}$ is some positive constant.

Proof: First of all, we observe that by the definition of the suggested repeated SPRT sampling we have

$$-A_i < u_{t,i} - u_{t^i_{n_t},i} < B_i, \forall i = 1, \ldots, K,$$

therefore

$$\sum(-A_i) < u_t - \tilde{u}_t < \sum B_i \Rightarrow |u_t - \tilde{u}_t| < C < \infty,$$

where $C = \max(\sum A_i, \sum B_i)$. This observation has the following important implications

- The stopping time of our test $\tilde{\mathcal{S}}$ can be bounded from above and from below by two SPRT stopping times. Indeed:

$$\begin{aligned}
\tilde{\mathcal{S}} &= \inf\{t > 0 : \tilde{u}_t \leq -\tilde{A} \\
&\qquad \text{or } \tilde{u}_t \geq +\tilde{B}\} \\
&= \inf\{t > 0 : (\tilde{u}_t - u_t) + u_t \leq -\tilde{A} \\
&\qquad \text{or } (\tilde{u}_t - u_t) + u_t \geq \tilde{B}\} \\
&= \inf\{t > 0 : u_t \leq -(\tilde{u}_t - u_t) - \tilde{A} \\
&\qquad \text{or } u_t \geq \tilde{B} - (\tilde{u}_t - u_t)\}.
\end{aligned}$$

  Now using our initial observation, we have: $\mathcal{S}_l \leq \tilde{\mathcal{S}} \leq \mathcal{S}_u$, where

$$\mathcal{S}_l = \inf_{t \geq 0}\{t : u_t \notin (-\tilde{A} + C, \tilde{B} - C)\} \quad (16)$$

$$\mathcal{S}_u = \inf_{t \geq 0}\{t : u_t \notin (-\tilde{A} - C, \tilde{B} + C)\}. \quad (17)$$

- the thresholds of the stopping times $\tilde{\mathcal{S}}, \mathcal{S}$ that correspond to the same error probabilities have a bounded distance, i.e. $|\tilde{B} - B| \leq C'$ and $|\tilde{A} - A| \leq C'$. Indeed using the definition of $\mathcal{S}_l, \mathcal{S}_u$ we observe

$$\mathbb{E}_0[\mathbf{1}_{\{u_{\mathcal{S}_l} \geq \tilde{B} - C\}}] \geq \mathbb{E}_0[\mathbf{1}_{\{\tilde{u}_{\tilde{\mathcal{S}}} \geq \tilde{B}\}}] \geq \mathbb{E}_0[\mathbf{1}_{\{u_{\mathcal{S}_u} \geq \tilde{B} + C\}}].$$

  Recalling now that $\mathbb{E}_0[\mathbf{1}_{\{\tilde{u}_{\tilde{\mathcal{S}}} \geq \tilde{B}\}}] = \alpha = \mathbb{E}_0[\mathbf{1}_{\{u_{\mathcal{S}} \geq B\}}]$ and using (8) we have

$$e^{-(\tilde{B} - C) + o(1)} \geq e^{-B + o(1)} \geq e^{-(\tilde{B} + C) + o(1)}$$

  which leads to the desired conclusion that $|B - \tilde{B}| \leq C + o(1) \leq C'$. In exactly similar way we can prove the other inequality.

From the optimality of the continuous-time SPRT $\mathcal{S}$, we have: $\mathbb{E}_j[\mathcal{S}] \leq \mathbb{E}_j[\tilde{\mathcal{S}}], j = 0, 1$. Combining this with the previous observations, we have:

$$0 \leq \mathbb{E}_j[\tilde{\mathcal{S}}] - \mathbb{E}_j[\mathcal{S}] \leq \mathbb{E}_j[\mathcal{S}_u] - \mathbb{E}_j[\mathcal{S}], j = 0, 1. \quad (18)$$

But, both $\mathcal{S}_u, \mathcal{S}$ are SPRT stopping times, therefore the corresponding expectations $\mathbb{E}_j[\mathcal{S}_u], \mathbb{E}_j[\mathcal{S}], j = 0, 1$ will be given by the formulae (14) and (15), for the appropriate thresholds. Moreover, we observe that $\alpha \to 0$ is equivalent to $B \to \infty$, and since $|B - \tilde{B}| = O(1)$ it will also be $\tilde{B} \to \infty$. Similarly, $\beta \to 0$ is equivalent to $A \to \infty$ and $\tilde{A} \to \infty$. Thus, it suffices to show that the limit as $B, \tilde{B} \to \infty$ of

$$-\Bigg[\frac{(\tilde{B} + C)(e^{\tilde{A}+C} - 1) + (-\tilde{A} - C)e^{\tilde{A}+C}(e^{\tilde{B}+C} - 1)}{e^{\tilde{A}+\tilde{B}+2C} - 1} - \frac{B(e^A - 1) - Ae^A(e^B - 1)}{e^{A+B} - 1}\Bigg]$$

and the limit as $A, \tilde{A} \to \infty$ of

$$\Bigg[\frac{-(\tilde{A} + C)(e^{\tilde{B}+C} - 1) + (\tilde{B} + C)e^{\tilde{B}+C}(e^{\tilde{A}+C} - 1)}{e^{\tilde{A}+\tilde{B}+2C} - 1} - \frac{-A(e^B - 1) + Be^B(e^A - 1)}{e^{A+B} - 1}\Bigg]$$

are both bounded, which can be easily verified. ■

Remark: The above theorem suggests -as in Mei [3]- that feedback from the fusion center to the sensors is not necessary for designing asymptotically optimum decentralized tests, when the sensors have full local memory.

V. Simulations

In this section we perform simulation experiments in order to illustrate the asymptotic optimality of the suggested scheme and compare its performance with the optimal centralized SPRT as well as its discrete time version. Specifically, we implement an *exact* simulation algorithm, where we simulate the intersampling times in each sensor. Therefore, we avoid simulating Brownian Motions in the sensors, which would require a fine time-discretization scheme and would lead to a less accurate and more computationally expensive simulation algorithm. In order to implement such an algorithm, we need to compute the pdf of the intersampling times in each sensor and sample from it.

Because of the Markov property of the Brownian motion, it is clear that the intersampling times $\{t_{n,i} - t_{n-1,i}\}$ and the signals $\{z_{n,i}\}$ are sequences of iid random variables under each hypothesis, thus it suffices to examine the distribution of $(t_{1,i}, z_{1,i})$ under each hypothesis. This is what the next proposition does, but before we state it, we introduce the following notation:

- We denote by $h(\cdot; c, \theta)$ the pdf of the stopping time T_c, i.e. the first time a Brownian motion with constant drift θ (starting from 0) hits the point $c \neq 0$. $h(\cdot; c, \theta)$ is the inverse gaussian density (also known as Wald density) and is given by the following formula:

$$h(t;c,\theta)=\frac{|c|}{\sqrt{2\pi t^3}}\exp\left\{-\frac{(c-\theta t)^2}{2t}\right\},\, t>0.$$

- In the case that the drift $\theta=0$, we simplify the notation and we write: $h(\cdot;c)$ instead of $h(\cdot;c,0)$.
- We also use the following notation:

$$g(t;c,d)=\sum_{n=-\infty}^{\infty}h(t;2n(c+d)+c,0)\,,t>0,$$

where $c,d>0$.

Proposition 1: a). The density of $t_{1,i}$ on the events $\{z_{1,i}=1\}$ and $\{z_{1,i}=0\}$, under the null hypothesis, is given by the formulas

$$\mathbb{P}_{0,i}(t_{1,i}\in dt;z_{1,i}=1)=e^{\frac{-B_i}{2}-\frac{\mu_i^2}{8}t}g\left(t;\frac{B_i}{\mu_i},\frac{A_i}{\mu_i}\right)dt \quad (19)$$

$$\mathbb{P}_{0,i}(t_{1,i}\in dt;z_{1,i}=0)=e^{\frac{A_i}{2}-\frac{\mu_i^2}{8}t}g\left(t;\frac{A_i}{\mu_i},\frac{B_i}{\mu_i}\right)dt, \quad (20)$$

and under the alternative by

$$\mathbb{P}_{1,i}(t_{1,i}\in dt;z_{1,i}=1)=e^{\frac{B_i}{2}-\frac{\mu_i^2}{8}t}g\left(t;\frac{B_i}{\mu_i},\frac{A_i}{\mu_i}\right)dt$$

$$\mathbb{P}_{1,i}(t_{1,i}\in dt;z_{1,i}=0)=e^{\frac{-A_i}{2}-\frac{\mu_i^2}{8}t}g\left(t;\frac{A_i}{\mu_i},\frac{B_i}{\mu_i}\right)dt.$$

b). The sequence of signals $\{z_{n,i}\}_{n\in\mathbf{N}}$ in sensor i is a sequence of iid Bernoulli random variables under each hypothesis, with parameter $\pi_{0,i}=\frac{e^{A_i}-1}{e^{A_i+B_i}-1}$ under H_0 and $\pi_{1,i}=\frac{e^{A_i+B_i}-e^{B_i}}{e^{A_i+B_i}-1}$ under H_1. c). The expectation of the intersampling time $\tau_{1,i}$, under the null hypothesis, is

$$\mathbb{E}_0[\tau_{1,i}]=\frac{2}{\mu_i^2}\left[(1-\pi_{0,i})A_i-\pi_{0,i}B_i\right],$$

and under the alternative

$$\mathbb{E}_1[\tau_{1,i}]=\frac{2}{\mu_i^2}\left[\pi_{1,i}B_i-(1-\pi_{1,i})(-A_i)\right].$$

d). Let $\pi_{j,i}(t)$ be the conditional probability that the i^{th} sensor hits the upper threshold given the corresponding intersampling time, under hypothesis $H_j, j=0,1$. Then it will be:

$$\begin{aligned}\pi_{0,i}(t) &\equiv P_0(z_{1,i}=1|t_{1,i}=t)\\ &= \frac{\mathbb{P}_{0,i}(t_{1,i}\in dt;z_{1,i}=1)}{\mathbb{P}_{0,i}(t_{1,i}\in dt)},\end{aligned}$$

$$\begin{aligned}\pi_{1,i}(t) &\equiv P_1(z_{1,i}=1|t_{1,i}=t)\\ &= \frac{\mathbb{P}_{1,i}(t_{1,i}\in dt;z_{1,i}=1)}{\mathbb{P}_{0,i}(t_{1,i}\in dt)},\end{aligned}$$

e). The conditional densities of the intersampling time $t_{1,i}$ given $z_{1,i}=0$ and given $z_{1,i}=1$ will be: under the null hypothesis

$$\mathbb{P}_{0,i}(t_{1,i}\in dt|z_{1,i}=1)=\frac{\mathbb{P}_{0,i}(t_{1,i}\in dt;z_{1,i}=1)}{\pi_{0,i}},$$

$$\mathbb{P}_{0,i}(t_{1,i}\in dt|z_{1,i}=0)=\frac{\mathbb{P}_{0,i}(t_{1,i}\in dt;z_{1,i}=0)}{1-\pi_{0,i}},$$

and under the alternative hypothesis

$$\mathbb{P}_{1,i}(t_{1,i}\in dt|z_{1,i}=1)=\frac{\mathbb{P}_{1,i}(t_{1,i}\in dt;z_{1,i}=1)}{\pi_{1,i}},$$

$$\mathbb{P}_{1,i}(t_{1,i}\in dt|z_{1,i}=0)=\frac{\mathbb{P}_{1,i}(t_{1,i}\in dt;z_{1,i}=0)}{1-\pi_{1,i}}.$$

f). The supremum of the ratio of each of the pdf's in a), b) and e) over the exponential pdf with rate $\frac{\mu_i^2}{8}$ is bounded.

Proof: It is well known (see for example [2, Page 99]) that $g(\cdot;\frac{A_i}{\mu_i},\frac{B_i}{\mu_i})$ is the density of the first exit time of a standard Brownian motion from the interval $(-\frac{A_i}{\mu_i},\frac{B_i}{\mu_i})$ *on the event that the lower boundary is crossed*, whereas $g(t;\frac{B_i}{\mu_i},\frac{A_i}{\mu_i})$ is the corresponding density *on the event that the upper boundary is crossed.*

For a) we follow a standard application of Girsanov's theorem, similarly as in [2, Page 196]; b) and c) are well known results that can be obtained from applications of Optional Sampling Theorem to the log-likelihood ratio u_t; d) and e) are applications of Bayes Rule, whereas f) follows from the boundedness of the pdfs that we obtained in the previous questions. ■

Remarks:

- This proposition implies that, under each hypothesis, we can simulate the conditional intersampling density given the information on the boundary that was crossed through an acceptance-rejection scheme, using the exponential pdf with rate $\frac{\mu_i^2}{8}$ as the candidate density.
- We can choose the sampling thresholds in the sensors in such a way that the sampling frequency in all sensors is the same. In that case, we have to equate the expectations in part (2) of the above proposition with the desired sampling frequency and solve the resulting non-linear equation. In order to determine the thresholds in a unique way, we need another meaningful constraint at each sensor. In the absence of such a constraint, we can simply choose the thresholds to be symmetric, i.e $A_i=B_i$. In that case, the simulation of the scheme has some very appealing properties as the next proposition suggests.

Proposition 2: In the case of symmetric boundaries, $A_i=B_i$, we have the following properties

1) The pdf of the intersampling time $t_{1,i}$ is the *same* under the two hypotheses and will be denoted by $\tilde{h}(\cdot;B_i,\mu_i)$ (in analogy with the corresponding one-sided density). It will be

$$\tilde{h}(t;B_i,\mu_i)=\cosh\left(\frac{B_i}{2}\right)e^{-\frac{\mu_i^2}{8}t}g\left(t;\frac{B_i}{\mu_i},\frac{B_i}{\mu_i}\right) \quad (21)$$

2) The signal $z_{1,i}$ is a Bernoulli random variable with parameter $\pi_{0,i}=\frac{1}{1+e^{B_i}}$ (under H_0) and $\pi_{1,i}=\frac{e^{B_i}}{1+e^{B_i}}$ (under H_1).
3) The signal $z_{1,i}$ is independent from the intersampling time $t_{1,i}$ under each hypothesis, i.e. using the notation

of Proposition 1 we have under the null hypothesis

$$\pi_{0,i}(t) = \pi_{0,i} = \frac{1}{1+e^{B_i}}, \forall t > 0$$

and under the alternative

$$\pi_{1,i}(t) = \pi_{1,i} = \frac{e^{B_i}}{1+e^{B_i}}, \forall t > 0.$$

4) The ratio of the pdfs $\tilde{h}(\cdot; B_i, \mu_i)$ and $h(\cdot; \frac{B_i}{\mu_i}, \frac{\mu_i}{2})$ is bounded above by 2. In particular

$$\sup_{t>0} \frac{\tilde{h}(t; B_i, \mu_i)}{h(t; \frac{B_i}{2}, \frac{\mu_i}{2})} = 1 + e^{-B_i}.$$

Proof: The first three parts follow from substitution $A_i = B_i$ in Proposition 1. The fourth part follows from the observation that

$$\sum_{n=-\infty}^{\infty} \frac{h\left(t; (4n+1)\frac{B_i}{\mu_i}, 0\right)}{h(t; \frac{B_i}{\mu_i}, 0)} \leq 1.$$

This concludes the proof. ∎

Remark: Proposition 2 suggests that in the case of symmetric boundaries, we can sample the signals and the interarrival densities *independently*. Moreover, it suggests $h(\cdot; \frac{B_i}{2}, \frac{\mu_i}{2})$ as a good candidate density for the simulation of the intersampling density using an acceptance-rejection scheme. The exponential density remains of course a potential candidate, but the inverse Gaussian density appears to be much more efficient, in the sense that the percentage of rejected samples is guaranteed to be smaller than 50%.

A. Simulation Algorithm

We can now state the following algorithm for the simulation of our scheme, which consists of two steps in each simulation run. The first one is the computation of the stopping time $\tilde{S}$ under each of the two hypotheses. The second is the computation of the Type I and Type II error probabilities.

Thus, for each hypothesis $H_j, j = 0, 1$, we have:

1) Set $\tilde{u}^{(j)} = 0$.
2) Sample $z_{1,i} \sim \text{Bernoulli}(\pi_{j,i}), i = 1, \ldots, K$.
3) Sample $t_{1,i}$ from its conditional interarrival density given $z_{1,i}$, $i = 1, \ldots, K$.
4) Set $k = \arg\min_i \{t_{1,i}\}$ and $t_{\text{new}} = t_{1,k}$.
5) Set $\tilde{u}^{(j)} = \tilde{u}^{(j)} + B_k z_{1,k} + (-A_k)(1 - z_{1,k})$.
6) If $-\tilde{A} < \tilde{u}^{(j)} < \tilde{B}$, then:
   - Set: $t_{1,k} = t_{1,k} + t_{\text{new}}$.
   - If $k = \arg\min_i\{t_{1,i}\}$, go back to step (5), otherwise go back to step (2).
7) If $\tilde{u}^{(j)} \geq \tilde{B}$ or $\tilde{u}^{(j)} \leq -\tilde{A}$, then $\tilde{S} = t_{1,k}$.

Now in order to compute the Type I and Type II error probabilities of our scheme, $\tilde{\gamma}$ and $\tilde{\delta}$ respectively, we apply the classical trick of importance sampling (see Siegmund [6] for details). Thus, the algorithm continues as follows:

- $\tilde{\gamma} = e^{-\tilde{u}^{(1)}} \mathbf{1}_{\{\tilde{u}^{(1)} \geq \tilde{B}\}}$ and $\tilde{\delta} = e^{\tilde{u}^{(0)}} \mathbf{1}_{\{\tilde{u}^{(0)} \leq -\tilde{A}\}}$.

Repeating this procedure many times, we can obtain the Monte-Carlo estimates and standard errors of the quantities of interest, i.e. $\mathbb{E}_0[\tilde{S}], \mathbb{E}_0[\tilde{S}], \tilde{\gamma}, \tilde{\delta}$.

In the special case of symmetric boundaries, the above algorithm can be modified, so that step (3) takes the following form:

- Sample $t_{1,i}$ from the unconditional interarrival density of sensor i, $\tilde{h}(\cdot; -B_i, B_i, \mu_i)$ using an acceptance-rejection scheme, with $h(\cdot; B_i, \mu_i)$ as the candidate.

B. Experiments

We proceed with a numerical example, where we have $K = 2$ sensors with $\mu_1 = \mu_2 = 1$. We apply three tests. First, the continuous-time centralized SPRT which serves as a point of reference. Its performance is analytically given in (14) and (15), as a function of the thresholds A, B. Second, we simulate the discrete-time centralized SPRT, with signals sampled with a constant period $T = 2.71$ and sent to the fusion center, without quantization, to perform the discrete-time SPRT test. Third is our scheme with local boundaries $B_i = -A_i = 3$ producing an *average sampling period* of 2.71 (from part (4) of Proposition 1), which matches the period T of the centralized discrete-time test. This selection is necessary for a fair comparison. Finally, we also simulate Mei's scheme introduced in [3]. Fig. 2 depicts the average expected time for a

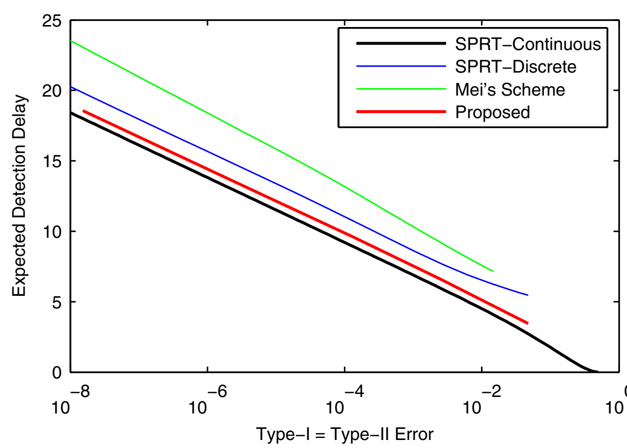

Fig. 2. Expected detection delay as a function of the error probability.

decision under H_1 as a function of the probability of Type I = Type II error. We observe that the proposed test performs even better than the discrete-time SPRT, exhibiting a performance which is very close to the optimum.

VI. CONCLUSIONS

In this work, we examined the problem of decentralized sequential hypothesis testing *in continuous time* in the case where the sensors have full local memory and no feedback from the fusion center. We suggested a very easily implementable scheme which entails *asynchronous* communication of *1-bit* decisions of the sensors to the fusion center. Moreover, we proved that in the case where the observed processes in the sensors are (drifted) Brownian motions, the proposed scheme exhibits a strong asymptotic optimality property, in particular it is asymptotically optimal of order-2, as compared to Mei's scheme [3], which is optimal of order-1. We illustrated this optimality with simulation experiments, which indicated that the performance of our scheme is even superior to that of the centralized discrete-time SPRT and very close to the optimal performance of the centralized continuous-time SPRT. Thus, we can achieve better performance in the decentralized problem than in the centralized one, as long as we implement *random sampling* instead of the conventional deterministic sampling.

REFERENCES

[1] A.M. Hussain, "Multisensor Distributed Sequential Detection," *IEEE Tran. Aer. Elect. Syst.*, vol. 30, no. 3, pp. 698-708.

[2] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus, Second Edition*, Springer-Verlag, New York, 1991.

[3] Y. Mei, "Asymptotic optimality theory for sequential hypothesis testing in sensor networks," submitted to the *IEEE Tran. Inf. Th.*, under review, 2008.

[4] V.N.S. Samarasooriya and P.K. Varshney, "Sequential Approach to asynchronous decision fusion," *Opt. Eng.*, vol. 35, no. 3, pp. 625-633, 1996.

[5] A.N. Shiryaev, *Optimal Stopping Rules*. Springer-Verlang, New York,1978.

[6] D. Siegmund, "Importance sampling in the Monte Carlo study of sequential tests," *Annals of Statistics*, vol. 4, no. 4, pp.673-684, 1976.

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

[8] V.V. Veeravalli, T. Basar and H.V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Tran. Inf. Th.*, vol. 39, no. 3, pp. 433-442, 1993.

[9] A. Wald, *Sequential Analysis*. Wiley, New York, 1945.