# Distributed Sequential Detection for Gaussian Shift-in-Mean Hypothesis Testing

Anit Kumar Sahu, *Student Member, IEEE* and Soummya Kar, *Member, IEEE*

## Abstract

This paper studies the problem of sequential Gaussian shift-in-mean hypothesis testing in a distributed multi-agent network. A sequential probability ratio test (SPRT) type algorithm in a distributed framework of the *consensus+innovations* form is proposed, in which the agents update their decision statistics by simultaneously processing latest observations (innovations) sensed sequentially over time and information obtained from neighboring agents (consensus). For each pre-specified set of type I and type II error probabilities, local decision parameters are derived which ensure that the algorithm achieves the desired error performance and terminates in finite time almost surely (a.s.) at each network agent. Large deviation exponents for the tail probabilities of the agent stopping time distributions are obtained and it is shown that asymptotically (in the number of agents or in the high signal-to-noise-ratio regime) these exponents associated with the distributed algorithm approach that of the optimal centralized detector. The expected stopping time for the proposed algorithm at each network agent is evaluated and is benchmarked with respect to the optimal centralized algorithm. The efficiency of the proposed algorithm in the sense of the expected stopping times is characterized in terms of network connectivity. Finally, simulation studies are presented which illustrate and verify the analytical findings.

## Index Terms

Distributed Detection, Multi-agent Networks, Consensus, Sequential Probability Ratio Tests, Large Deviations

## 1. INTRODUCTION

*A. Background and Motivation*

The focus of this paper is on sequential simple hypothesis testing in multi-agent networks in which the goal is to detect the (binary) state of the environment based on observations at the agents. By sequential we mean, instead of considering fixed sample size hypothesis tests in which the objective is to minimize the probabilities of decision error (the false alarm and the miss) based on a given deterministic number of samples or observation data collected by the network agents, we are interested in the design of testing procedures that in the *quickest* time or using the *minimal* amount of sensed data samples at the agents can distinguish between the two hypotheses with guaranteed accuracy given in terms of pre-specified tolerances on false alarm and miss probabilities. The motivation behind studying sequential as opposed to fixed sample size testing is that in most practical agent networking scenarios, especially in applications that are time-sensitive and/or resource constrained, the priority is to achieve inference as quickly as possible by expending the minimal amount of resources (data samples, sensing energy and communication). Furthermore, we focus on distributed application environments which are devoid of fusion centers[1] and in which inter-agent collaboration or information exchange is limited to a pre-assigned, possibly sparse, communication structure.

Under rather generic assumptions on the agent observation models, it is well-known that in a (hypothetical) centralized scenario or one in which inter-agent communication is all-to-all corresponding to a complete communication graph, the sequential probability ratio test (SPRT) ([1]) turns out to be the optimal procedure for sequential testing of binary hypotheses; specifically, the SPRT minimizes the expected detection time (and hence the number of agent observation samples that need to be processed) while achieving requisite error performance in terms of specified probability of false alarm ($\alpha$) and probability of miss ($\beta$) tolerances. The SPRT and its variants have been applied in various contexts, see, for example, spectrum sensing in cognitive radio networks ([2]–[4]), target tracking [5], to name a few. However, the SPRT, in the current multi-agent context, would require computing a (centralized) decision statistic at all times, which, in turn, would either require all-to-all communication among the agents or access to the entire network data at all times at a fusion center. In contrast, restricted by a pre-assigned possibly sparse collaboration structure among the agents, in this paper we present and characterize a distributed sequential detection algorithm, the $\mathcal{CISPRT}$, based on the *consensus+innovations* approach

---

[1]By fusion center or center, we mean a hypothetical decision-making architecture in which a (central) entity has access to all agent observations at all times and/or is responsible for decision-making on behalf of the agents.

(see, for example [6], [7]). Specifically, focusing on a setting in which the agent observations over time are conditionally Gaussian and independent and identically distributed (i.i.d.), we study the $\mathcal{CISPRT}$ sequential detection procedure in which each network agent maintains a local (scalar) test statistic which is updated over time by simultaneously assimilating the test statistics of neighboring agents at the previous time instant (a consensus potential) and the most recent observations (innovations) obtained by the agent and its neighbors. Also, similar in spirit to the (centralized) SPRT, each agent chooses two (local) threshold parameters (design choices) and the test termination at an agent (and subsequent agent decision on the hypotheses) is determined by whether the local test statistic at the agent lies in the interval defined by the thresholds or not. This justifies the nomenclature that the $\mathcal{CISPRT}$ is a distributed SPRT type algorithm of the consensus+innovations form. The main contributions of this paper are as follows:

**Main Contribution 1: Finite Stopping Property.** We show that, given any value of probability of false alarm $\alpha$ and probability of miss $\beta$, the $\mathcal{CISPRT}$ algorithm can be designed such that each agent achieves the specified error performance metrics and the test procedure terminates in finite time almost surely (a.s.) at each agent. We derive closed form expressions for the local threshold parameters at the agents as functions of $\alpha$ and $\beta$ which ensures that the $\mathcal{CISPRT}$ achieves the above property.

**Main Contribution 2: Asymptotic Characterization.** By characterizing the stopping time distribution of the $\mathcal{CISPRT}$ at each network agent, we compute large deviations decay exponents of the stopping time tail probabilities at each agent, and show that the large deviations exponent of the $\mathcal{CISPRT}$ approaches that of the optimal centralized in the asymptotics of $N$, where $N$ denotes the number of agents in the network. In the asymptotics of vanishing error metrics (i.e., as $\alpha, \beta \to 0$), we quantify the ratio of the expected stopping time $T_{d,i}(\alpha,\beta)$ for reaching a decision at an agent $i$ through the $\mathcal{CISPRT}$ algorithm and the expected stopping time $T_c(\alpha,\beta)$ for reaching a decision by the optimal centralized (SPRT) procedure, i.e., the quantity $\frac{\mathbb{E}[T_{d,i}(\alpha,\beta)]}{\mathbb{E}[T_c(\alpha,\beta)]}$, which in turn is a metric of efficiency of the proposed algorithm as a function of the network connectivity. In particular, we show that the efficiency of the proposed $\mathcal{CISPRT}$ algorithm in terms of the ratio $\frac{\mathbb{E}[T_{d,i}(\alpha,\beta)]}{\mathbb{E}[T_c(\alpha,\beta)]}$ is upper bounded by a constant which is a function of the network connectivity and can be made close to one by choosing the network connectivity appropriately, thus establishing the benefits of inter-agent collaboration in the current context.

**Related Work.** Detection schemes in multi-agent networks which involve fusion centers, where all agents in the network transmit their local measurements, local decisions or local likelihood ratios to a fusion agent which subsequently makes the final decision (see, for example, [8]–[11]) have been well studied. Consensus-based approaches for fully distributed but single snapshot processing, i.e., in which the agents first collect their observations possibly over a long time horizon and then deploy a

consensus-type protocol [12]–[14] to obtain distributed information fusion and decision-making have also been explored, see, for instance, [15], [16]. Generalizations and variants of this framework have been developed, see for instance [17] which proposes truncated versions of optimal testing procedures to facilitate efficient distributed computation using consensus; scenarios involving distributed information processing where some of the agents might be faulty or there is imperfect model information (see, for example, [18], [19]) have also been studied. More relevant to the current context are distributed detection techniques that like the $\mathcal{CISPRT}$ procedure perform simultaneous assimilation of neighborhood decision-statistics and local agent observations in the same time step, see, in particular, the running consensus approach [20], [21], the diffusion approach [22]–[24] and the consensus+innovations approach [25]–[27]. These works address important questions in fixed (but possibly large) sample size distributed hypothesis testing, including asymptotic characterization of detection errors [21], [24], fundamental performance limits as characterized by large deviations decay of detection error probabilities in generic nonlinear observation models and random networks [25], [26], and detection with noisy communication links [27]. A continuous time version of the running consensus approach [21] was studied in [28] recently with implications on sequential distributed detection; specifically, asymptotic properties of the continuous time decision statistics were obtained and in the regime of large thresholds bounds on expected decision time and error probability rates were derived. However, there is a fundamental difference between the mostly fixed or large sample size procedures discussed above and the proposed $\mathcal{CISPRT}$ sequential detection procedure – technically speaking, the former focuses on analyzing the probability distributions of the detection errors as a function of the sample size and/or specified thresholds, whereas, in this paper, we design thresholds, stopping criteria and characterize the probability distributions of the (random) stopping times of sequential distributed procedures that aim to achieve quickest detection given specified tolerances on the detection errors. Addressing the latter requires novel technical machinery in the design and analysis of dynamic distributed inference procedures which we develop in this paper.

We also contrast our work with sequential detection approaches based in other types of multi-agent networking scenarios. In the context of decentralized sequential testing in multi-agent networks, funda-mental methodological advances have been reported, see, for instance, [29]–[36], which address very general models and setups. These works involve fusion center based processing where all agents in the network either transmit their local decisions, measurements or their quantized versions to a fusion center. In contrast, in this paper we restrict attention to Gaussian binary testing models only, but focus on a fully distributed paradigm in which there is no fusion center and inter-agent collaboration is limited to a pre-assigned, possibly sparse, agent-to-agent local interaction graph.

**Paper Organization :** We briefly summarize the organization of the rest of the paper. Section 1-B presents notation to be used throughout the paper. The sensing models and the abstract problem formulation are stated and discussed in Section 2-A. Section 2-B presents preliminaries on centralized sequential detection and motivates the distributed setup pursued in this paper. Section 3 presents the $\mathcal{CISPRT}$ algorithm. The main results of the paper are stated in Section 4 which includes the derivation of the thresholds for the $\mathcal{CISPRT}$ algorithm, the stopping time distribution for the $\mathcal{CISPRT}$ algorithm and the key technical ingredients concerning the asymptotic properties and large deviation analysis for the stopping time distributions of the centralized and distributed setups. It also includes the characterization of the expected stopping times of the $\mathcal{CISPRT}$ algorithm and its centralized counterpart in asymptotics of vanishing error metrics. Section 6 presents simulation studies. The proofs of the main results appear in Section 7, whereas, Section 8 concludes the paper.

*B. Notation*

We denote by $\mathbb{R}$ the set of reals, $\mathbb{R}_+$ the set of non-negative reals, and by $\mathbb{R}^k$ the $k$-dimensional Euclidean space. The set of $k \times k$ real matrices is denoted by $\mathbb{R}^{k \times k}$. The set of integers is denoted by $\mathbb{Z}$, whereas, $\mathbb{Z}_+$ denotes the subset of non-negative integers and $\overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{\infty\}$. We denote vectors and matrices by bold faced characters. We denote by $A_{ij}$ or $[A]_{ij}$ the $(i,j)$th entry of a matrix $\mathbf{A}$; $a_i$ or $[a]_i$ the $i$th entry of a vector $\mathbf{a}$. The symbols $\mathbf{I}$ and $\mathbf{0}$ are used to denote the $k \times k$ identity matrix and the $k \times p$ zero matrix respectively, the dimensions being clear from the context. We denote by $\mathbf{e}_i$ the $i$th column of $\mathbf{I}$. The symbol $\top$ denotes matrix transpose. The $k \times k$ matrix $\mathbf{J} = \frac{1}{k}\mathbf{1}\mathbf{1}^\top$ where $\mathbf{1}$ denotes the $k \times 1$ vector of ones. The operator $\|\cdot\|$ applied to a vector denotes the standard Euclidean $\mathcal{L}_2$ norm, while applied to matrices it denotes the induced $\mathcal{L}_2$ norm, which is equivalent to the spectral radius for symmetric matrices. All the logarithms in the paper are with respect to base $e$ and represented as $\log(\cdot)$. Expectation is denoted by $\mathbb{E}[\cdot]$ and $\mathbb{E}_\theta[\cdot]$ denotes expectation conditioned on hypothesis $H_\theta$ for $\theta \in \{0,1\}$. $\mathbb{P}(\cdot)$ denotes the probability of an event and $\mathbb{P}_\theta(.)$ denotes the probability of the event conditioned on hypothesis $H_\theta$ for $\theta \in \{0,1\}$. $\mathbb{Q}(.)$ denotes the $\mathbb{Q}$-function which calculates the right tail probability of a normal distribution and is given by $\mathbb{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$, $x \in \mathbb{R}$. We will use the following property of the $\mathbb{Q}(.)$ function, namely for any $x > 0$, $\mathbb{Q}(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}$. For deterministic $\mathbb{R}_+$-valued sequences $\{a_t\}$ and $\{b_t\}$, the notation $a_t = O(b_t)$ denotes the existence of a constant $c > 0$ such that $a_t \leq c b_t$ for all $t$ sufficiently large, whereas, $a_t = o(b_t)$ indicates $a_t/b_t \to 0$ as $t \to \infty$.

**Spectral Graph Theory.** For an undirected graph $G = (V, E)$, $V$ denotes the set of agents or vertices with cardinality $|V| = N$, and $E$ the set of edges with $|E| = M$. The unordered pair $(i, j) \in E$ if there exists an edge between agents $i$ and $j$. We only consider simple graphs, i.e. graphs devoid of self loops and multiple edges. A path between agents $i$ and $j$ of length $m$ is a sequence $(i = p_0, p_1, \ldots, p_m = j)$ of vertices, such that $(p_n, p_{n+1}) \in E$, $0 \le n \le m-1$. A graph is connected if there exists a path between all the possible agent pairings. The neighborhood of an agent $i$ is given by $\Omega_i = \{j \in V \mid (i, j) \in E\}$. The degree of agent $i$ is given by the cardinality $d_i = |\Omega_i|$. The structure of the graph may be equivalently represented by the symmetric $N \times N$ adjacency matrix $\mathbf{A} = [A_{ij}]$, where $A_{ij} = 1$ if $(i, j) \in E$, and 0 otherwise. The degree matrix is represented by the diagonal matrix $\mathbf{D} = diag(d_1 \ldots d_N)$. The graph Laplacian matrix is represented by

$$\mathbf{L} = \mathbf{D} - \mathbf{A}.$$

The Laplacian is a positive semidefinite matrix, hence its eigenvalues can be sorted and represented in the following manner

$$0 = \lambda_1(\mathbf{L}) \le \lambda_2(\mathbf{L}) \le \ldots \lambda_N(\mathbf{L}).$$

Furthermore, a graph is connected if and only if $\lambda_2(\mathbf{L}) > 0$ (see [37] for instance). We stress that under our notation, $\lambda_2$, also known as the Fiedler value (see [37]), plays an important role because it acts as an indicator of whether the graph is connected or not.

## 2. PROBLEM FORMULATION

*A. System Model*

The $N$ agents deployed in the network decide on either of the two hypothesis $H_0$ and $H_1$. Each agent $i$ at (discrete) time $t$ makes a scalar observation $y_i(t)$ of the form

$$Under \quad H_\theta : y_i(t) = \mu_\theta + n_i(t), \quad \theta = 0, 1.$$

For the rest of the paper we consider $\mu_1 = \mu$ and $\mu_0 = -\mu$, and assume that, the agent observation noise processes are independent and identically distributed (i.i.d.) Gaussian processes under both hypotheses formalized as follows:

**Assumption A1.** *For each agent $i$ the noise sequence $\{n_i(t)\}$ is i.i.d. Gaussian with mean zero and variance $\sigma^2$ under both $H_0$ and $H_1$. The noise sequences are also spatially uncorrelated, i.e., $\mathbb{E}_\theta[n_i(t)n_j(t)] = \mathbf{0}$ for all $i \ne j$ and $\theta \in \{0, 1\}$.*

Collect the $y_i(t)$'s, $i = 1, 2, \ldots N$ into the $N \times 1$ vector $\mathbf{y}(t) = (y_1(t), \ldots, y_N(t))^\top$ and the $n_i(t)$'s, $i = 1, 2, \ldots N$ into the $N \times 1$ vector $\mathbf{n}(t) = (n_1(t), \ldots, n_N(t))^\top$.

The log-likelihood ratio at the $i$-th sensor at time index $t$ is calculated as follows:-

$$\eta_i(t) = \frac{f_1(y_i(t))}{f_0(y_i(t))} = \frac{2\mu y_i(t)}{\sigma^2},$$

where $f_0(\cdot)$ and $f_1(\cdot)$ denote the probability distribution functions (p.d.f.s) of $y_i(t)$ under $H_0$ and $H_1$ respectively.

We note that,

$$\eta_i(t) \sim \begin{cases} \mathcal{N}(m, 2m), & H = H_1 \\ \mathcal{N}(-m, 2m), & H = H_0, \end{cases} \tag{1}$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian p.d.f. and $m = \frac{2\mu^2}{\sigma^2}$. The Kullback-Leibler divergence at each agent is given by

$$KL = m. \tag{2}$$

### B. Sequential Hypothesis Testing – Centralized or All-To-All Communication Scenario

We start by reviewing concepts and results from (centralized) sequential hypothesis testing theory, see [38] for example, to motivate our distributed hypothesis testing setup. Broadly speaking, the goal of sequential simple hypothesis testing is as follows: given pre-specified constraints on the error metrics, i.e., upper bounds $\alpha$ and $\beta$ on the probability of false alarm $\mathbb{P}_{FA}$ and probability of miss $\mathbb{P}_M$, the decision-maker keeps on collecting observations sequentially over time to decide on the hypotheses $H_1$ or $H_0$, i.e., which one is true; the decision-maker also has a *stopping criterion* or *stopping rule* based on which it decides at each time (sampling) instant whether to continue sampling or terminate the testing procedure. Finally, after termination, a (binary) decision is computed as to which hypothesis is in force based on all the obtained data. A sequential testing procedure is said to be *admissible* if the stopping criterion, i.e., the decision whether to continue observation collection or not, at each instant is determined solely on the basis of observations collected thus far. Naturally, from a resource optimization viewpoint, the decision-maker seeks to design the sequential procedure (or equivalently the stopping criterion) that minimizes the expected number of observation samples (or equivalently time) required to achieve a decision with probabilities of false alarm and miss upper bounded by $\alpha$ and $\beta$ respectively. To formalize in the current context, first consider a setup in which inter-agent communication is all-to-all, i.e., at each time instant each agent has access to all the sensed data of all other agents. In this complete network scenario, each

agent behaves like a (hypothetical) center and the information available at any agent $n$ at time $t$ is the sum-total of network observations till $t$, formalized by the $\sigma$-algebra [39]

$$\mathcal{G}_c(t) = \sigma\left\{y_i(s), \ \forall i = 1, 2, \ldots N \text{ and } \forall 1 \leq s \leq t\right\}.$$

An admissible test $D_c$ consists of a stopping criteria, where at each time $t$ the agents' (or the center in this case) decision to stop or continue taking observations is adapted to (or measurable with respect to) the $\sigma$-algebra $\mathcal{G}_c(t)$. Denote by $T_{D_c}$ the termination time of $D_c$, a random time taking values in $\mathbb{Z}_+ \cup \{\infty\}$. Formally, by the above notion of admissibility, the random time $T_{D_c}$ is necessarily a stopping time with respect to (w.r.t.) the filtration $\{\mathcal{G}_c(t)\}$, see [39], and, in this paper, we restrict attention to tests $D_c$ that terminate in finite time a.s., i.e., $T_{D_c}$ takes values in $\mathbb{Z}_+$ a.s. Denote by $\mathbb{E}_\theta[T_{D_c}]$ the expectation of $T_{D_c}$ under $H_\theta$, $\theta = 0, 1$, and $\widehat{H}_{D_c} \in \{0, 1\}$ the decision obtained after termination. (Note that, assuming $T_{D_c}$ is finite a.s., the random variable $\widehat{H}_{D_c}$ is measurable w.r.t. the stopped $\sigma$-algebra $\mathcal{G}_{T_{D_c}}$.) Let $\mathbb{P}_{FA}^{D_c}$ and $\mathbb{P}_M^{D_c}$ denote the associated probabilities of false alarm and miss respectively, i.e.,

$$\mathbb{P}_{FA}^{D_c} = \mathbb{P}_0\left(\widehat{H}_{D_c} = 1\right) \quad \text{and} \quad \mathbb{P}_M^{D_c} = \mathbb{P}_1\left(\widehat{H}_{D_c} = 0\right).$$

Now, denoting by $\mathcal{D}_c$ the class of all such (centralized) admissible tests, the goal in sequential hypothesis testing is to obtain a test in $\mathcal{D}_c$ that minimizes the expected stopping time subject to attaining specified error constraints. Formally, we aim to solve[2]

$$\min_{D_c \in \mathcal{D}_c} \mathbb{E}_1[T_{D_c}],$$
$$\text{s.t. } \mathbb{P}_{FA}^{D_c} \leq \alpha, \mathbb{P}_M^{D_c} \leq \beta, \tag{3}$$

for specified $\alpha$ and $\beta$. Before proceeding further, we make the following assumption:

**Assumption A2.** *The pre-specified error metrics, i.e., $\alpha$ and $\beta$, satisfy $\alpha, \beta \in (0, 1/2)$.*

Noting that the (centralized) Kullback-Leibler divergence, i.e., the divergence between the probability distributions induced on the joint observation space $\mathbf{y}(t)$ by the hypotheses $H_1$ and $H_0$, is $Nm$ where

---

[2]Note, in (3) the objective is to minimize the expected stopping time under hypothesis $H_1$. Alternatively, we might be interested in minimizing $\mathbb{E}_0[T_{D_c}]$ over all admissible tests; similarly, in a Bayesian setup with prior probabilities $p_0$ and $p_1$ on $H_0$ and $H_1$ respectively, the objective would consist of minimizing the overall expected stopping time $p_0\mathbb{E}_0[T_{D_c}] + p_1\mathbb{E}_1[T_{D_c}]$. However, it turns out that, in the current context, the Wald's SPRT [40] (to be discussed soon) can be designed to minimize each of the above criteria. Hence, without loss of generality, we adopt $\mathbb{E}_1[T_{D_c}]$ as our test design objective and use it as a metric to determine the relative performance of tests.

$m$ is defined in (1), we obtain (see [40]) for each $D_c \in \mathcal{D}_c$ that attains $\mathbb{P}_{FA}^{D_c} \leq \alpha$ and $\mathbb{P}_M^{D_c} \leq \beta$,

$$\mathbb{E}_1[T_{D_c}] \geq \mathcal{M}(\alpha, \beta),$$

where the universal lower bound $\mathcal{M}(\alpha, \beta)$ is given by

$$\mathcal{M}(\alpha, \beta) = \frac{(1 - \beta) \log(\frac{1-\beta}{\alpha}) + \beta \log(\frac{\beta}{1-\alpha})}{Nm}. \tag{4}$$

**Optimal (centralized) tests: Wald's SPRT.** We briefly review Wald's sequential probability ratio test (SPRT), see [40], that is known to achieve optimality in (3). To this end, denote by $S_c(t)$ (the centralized) test statistic

$$S_c(t) = \sum_{s=1}^{t} \frac{\mathbf{1}^\top}{N} \eta(s), \tag{5}$$

where $\eta(s)$ denotes the vector of log-likelihood ratios $\eta_i(s)$'s at the agents [3]. The SPRT consists of a pair of thresholds (design parameters) $\gamma_c^l$ and $\gamma_c^h$, such that, at each time $t$, the decision to continue or terminate is determined on the basis of whether $S_c(t) \in [\gamma_c^l, \gamma_c^h]$ or not. Formally, the stopping time of the SPRT is defined as follows:

$$T_c = \inf\{t \mid S_c(t) \notin [\gamma_c^l, \gamma_c^h]\}. \tag{6}$$

At $T_c$ the following decision rule is followed:

$$H = \begin{cases} H_0, & S_c(T_c) \leq \gamma_c^l \\ H_1, & S_c(T_c) \geq \gamma_c^h. \end{cases} \tag{7}$$

The optimality of the SPRT w.r.t. the formulation (3) is well-studied; in particular, in [40] it was shown that, for any specified $\alpha$ and $\beta$, there exist choices of thresholds $(\gamma_c^l, \gamma_c^h)$ such that the SPRT (6)-(7) achieves the minimum in (3) among all possible admissible tests $D_c$ in $\mathcal{D}_c$.

For given $\alpha$ and $\beta$, exact analytical expressions of the optimal thresholds are intractable in general. A commonly used choice of thresholds, see [1], is given by

$$\gamma_c^h = \log\left(\frac{1-\beta}{\alpha}\right)$$
$$\gamma_c^l = \log\left(\frac{\beta}{1-\alpha}\right), \tag{8}$$

---

[3]Both the sum and average (over $N$) can be taken as the test statistics for the centralized detector. We divide by $N$ for notational simplicity, so that the centralized decision statistic update becomes a special case of the $\mathcal{CISPRT}$ decision statistic update studied in Section 3.

which, although not strictly optimal in general, ensures that $\mathbb{P}^c_{FA} \leq \alpha$ and $\mathbb{P}^c_M \leq \beta$. (For SPRT procedures we denote by $\mathbb{P}^c_{FA}$ and $\mathbb{P}^c_M$ the associated probabilities of false alarm and miss respectively, which depend on the choice of thresholds used.) Nonetheless, the above choice (8) yields *close* to optimal behavior, and is, in fact, *asymptotically* optimal; formally, supposing that $\alpha = \beta = \epsilon$, the SPRT with thresholds given by (8) guarantees that (see [38])

$$\lim_{\epsilon \to 0} \frac{\mathbb{E}_1[T_c]}{\mathcal{M}(\epsilon, \epsilon)} = 1,$$

where $\mathcal{M}(\cdot)$ is defined in (4). In the sequel, given a testing procedure $D_c \in \mathcal{D}_c$ and assuming $\alpha = \beta = \epsilon$, we will study the quantity $\limsup_{\epsilon \to 0} \left( \mathbb{E}_1[T_{D_c}] / \mathcal{M}(\epsilon, \epsilon) \right)$ as a measure of its efficiency. Also, by abusing notation, when $\alpha = \beta = \epsilon$, we will denote $\mathcal{M}(\epsilon) \doteq \mathcal{M}(\epsilon, \epsilon)$.

## C. Subclass of Distributed Tests

The SPRT (6)-(7) requires computation of the statistic $S_c(t)$ (see (5)) at all times, which, in turn, requires access to all agent observations at all times. Hence, the SPRT may not be implementable beyond the fully centralized or all-to-all agent communication scenario as discussed in Section 2-B. Motivated by practicable agent networking applications, in this paper we are interested in distributed scenarios, in which inter-agent communication is restricted to a preassigned (possibly sparse) communication graph. In particular, given a graph $G = (V, E)$, possibly sparse, modeling inter-agent communication, we consider scenarios in which inter-agent cooperation is limited to a single round of message exchanges among neighboring agents per observation sampling epoch. To formalize the distributed setup and the corresponding subclass $\mathcal{D}_d$ of distributed tests, denote by $\mathcal{G}_{d,i}(t)$ the information available at an agent $i$ at time $t$. The information set includes the observations sampled by $i$ and the messages received from its neighbors till time $t$, and is formally given by the $\sigma$-algebra

$$\mathcal{G}_{d,i}(t) = \sigma\left\{ y_i(s), m_{i,j}(s), \quad \forall 1 \leq s \leq t, \forall j \in \Omega_i \right\}. \tag{9}$$

The quantity $m_{i,j}(s)$ denotes the message received by $i$ from its neighbor $j \in \Omega_i$ at time $s$, assumed to be a vector of constant (time-invariant) dimension; the exact message generation rule is determined by the (distributed) testing procedure $D_d$ in place and, necessarily, $m_{i,j}(s)$ is measurable w.r.t. the $\sigma$-algebra $\mathcal{G}_{d,j}(s)$. Based on the information content $\mathcal{G}_{d,i}(t)$ at time $t$, an agent decides on whether to continue taking observations or to stop in the case of which, it decides on one of the hypothesis $H_0$ or $H_1$. A distributed testing procedure $D_d$ then consists of message generation rules, and, local stopping and decision criteria

at the agents. Intuitively, and formally by (9) and the fact that $m_{i,j}(s)$ is measurable w.r.t $\mathcal{G}_{d,j}(s)$ for all $(i,j)$ and $s$, we have

$$\mathcal{G}_{d,i}(t) \subset \mathcal{G}_c(t) \quad \forall i, t,$$

i.e., the information available at an agent $i$ in the distributed setting is a subset of the information that would be available to a hypothetical center in a centralized setting as given in Section 2-B. Formally, this implies that the class of distributed tests $\mathcal{D}_d$ is a subset of the class of centralized or all-possible tests $\mathcal{D}_c$ as given in Section 2-B, i.e., $\mathcal{D}_d \subset \mathcal{D}_c$. (Intuitively, it means any distributed test can be implemented in a centralized setup or by assuming all-to-all communication.) In this paper, we are interested in characterizing the distributed test that conforms to the communication restrictions above and is optimal in the following sense:

$$\min_{D_d \in \mathcal{D}_d} \max_{i=1,2,\ldots,N} \mathbb{E}_1[T_{D_d,i}],$$

$$\text{s.t. } \mathbb{P}_{FA}^{D_d,i} \leq \alpha, \mathbb{P}_M^{D_d,i} \leq \beta, \forall i = 1, 2, \ldots, N. \tag{10}$$

In the above, $T_{D_d,i}$ denotes the termination (stopping) time at an agent $i$ and $\mathbb{P}_{FA}^{D_d,i}$, $\mathbb{P}_M^{D_d,i}$, the respective false alarm and miss probabilities at $i$. Note that, since $\mathcal{D}_d \subset \mathcal{D}_c$, for any distributed test $D_d$ we have $\mathbb{E}_1[T_{D_d,i}] \geq \mathbb{E}_1[T_c]$ for all $i$ at any specified $\alpha$ and $\beta$, i.e., a distributed procedure cannot outperform the optimal centralized procedure, the SPRT given by (6)-(7). Rather than solving (10), in this paper, we propose a distributed testing procedure of the consensus+innovations type (see Section 3), which is efficiently implementable and analyze its performance w.r.t. the optimal centralized testing procedure. In particular, we study its performance as a function of the inter-agent communication graph and show that as long as the network is *reasonably* well-connected, but possibly much sparser than the complete or all-to-all network, the suboptimality (in terms of the expected stopping times at the agents) of the proposed distributed procedure w.r.t. the optimal centralized SPRT procedure is upper bounded by a constant factor much smaller than $N$. Our results clearly demonstrate the benefits of collaboration (even over a sparse communication network) as, in contrast, in the non-collaboration case (i.e., each agent relies on its own observations only) each agent would require $N$ times the expected number of observations to achieve prescribed $\alpha$ and $\beta$ as compared to the optimal centralized scenario.[4]

---

[4]In the non-collaboration setup, the optimal procedure at an agent is to perform an SPRT using its local observation sequence only; w.r.t. the centralized, this implies that the effective SNR at an agent reduces by a factor $1/N$ and hence (see [40]) the agent would require $N$ times more observations (in expectation) to achieve the same level of false alarm and miss.

## 3. A DISTRIBUTED SEQUENTIAL DETECTOR

To mitigate the high communication and synchronization overheads in centralized processing, we propose a distributed sequential detection scheme where network communication is restricted to a more *localized* agent-to-agent interaction scenario. More specifically, in contrast to the fully centralized setup described in Section 2-B, we now consider sequential detection in a distributed information setup in which inter-agent information exchange or cooperation is restricted to a preassigned (arbitrary, possibly sparse) communication graph, whereby an agent exchanges its (scalar) test statistic and a scalar function of its latest sensed information with its (one-hop) neighbors. In order to achieve reasonable detection performance with such localized interaction, we propose a distributed sequential detector of the *consensus+innovations* form. Before discussing the details of our algorithm, we state an assumption on the inter-agent communication graph.

**Assumption A3.** *The inter-agent communication graph is connected, i.e. $\lambda_2(\mathbf{L}) > 0$, where $\mathbf{L}$ denotes the associated graph Laplacian matrix.*

**Decision Statistic Update**. In the proposed distributed algorithm, each agent $i$ maintains a test statistic $P_{d,i}(t)$, which is updated recursively in a distributed fashion as follows :

$$
P_{d,i}(t+1) = \frac{t}{t+1} \left( w_{ii} P_{d,i}(t) + \sum_{j \in \Omega_i} w_{ij} P_{d,j}(t) \right)
$$
$$
+ \frac{1}{t+1} \left( w_{ii} \eta_i(t+1) + \sum_{j \in \Omega_i} w_{ij} \eta_j(t+1) \right),
\tag{11}
$$

where $\Omega_i$ denotes the communication neighborhood of agent $i$ and the $w_{ij}$'s denote appropriately chosen combination weights (to be specified later).

We collect the weights $w_{ij}$ in an $N \times N$ matrix $\mathbf{W}$, where we assign $w_{ij} = 0$, if $(i,j) \notin E$. Denoting by $\mathbf{P}_d(t)$ and $\eta(t)$ as the vectors $[P_{d,1}(t), P_{d,2}(t), \ldots, P_{d,N}(t)]^\top$ and $[\eta_1(t), \eta_2(t), \ldots, \eta_N(t)]^\top$ respectively, (11) can be compactly written as follows:-

$$
\mathbf{P_d}(t+1) = \mathbf{W} \left( \frac{t}{t+1} \mathbf{P_d}(t) + \frac{1}{t+1} \eta(t+1) \right).
\tag{12}
$$

Now we state some design assumptions on the weight matrix $\mathbf{W}$.

**Assumption A4.** *We design the weights $w_{ij}$'s in (11) such that the matrix $\mathbf{W}$ is non-negative, symmetric, irreducible and stochastic, i.e., each row of $\mathbf{W}$ sums to one.*

We remark that, if Assumption A4 is satisfied, then the second largest eigenvalue in magnitude of $\mathbf{W}$, denoted by $r$, turns out to be strictly less than one, see for example [14]. Note that, by the stochasticity of $\mathbf{W}$, the quantity $r$ satisfies

$$r = ||\mathbf{W} - \mathbf{J}||. \tag{13}$$

For connected graphs, a simple way to design $\mathbf{W}$ is to assign equal combination weights, in which case we have,

$$\mathbf{W} = \mathbf{I} - \delta\mathbf{L}, \tag{14}$$

where $\delta$ is a suitably chosen constant. As shown in [41], [42], Assumption A4 can be enforced by taking $\delta$ to be in $(0, 2/\lambda_N(\mathbf{L}))$. The smallest value of $r$ is obtained by setting $\delta$ to be equal to $2/(\lambda_2(\mathbf{L}) + \lambda_N(\mathbf{L}))$, in which case we have,

$$r = ||\mathbf{W} - \mathbf{J}|| = \frac{(\lambda_N(\mathbf{L}) - \lambda_2(\mathbf{L}))}{(\lambda_2(\mathbf{L}) + \lambda_N(\mathbf{L}))}. \tag{15}$$

**Remark 3.1.** *It is to be noted that Assumption A4 can be enforced by appropriately designing the combination weights since the inter-agent communication graph is connected (see Assumption A3). Several weight design techniques satisfying Assumption A4 exist in the literature (see, for example, [41]). The quantity $r$ quantifies the rate of information flow in the network, and in general, the smaller the $r$ the faster is the convergence of information dissemination algorithms (such as the consensus or gossip protocol on the graph, see for example [14], [42], [43]). The optimal design of symmetric weight matrices $\mathbf{W}$ for a given network topology that minimizes the value $r$ can be cast as a semi-definite optimization problem [41].*

**Stopping Criterion for the Decision Update**. We now provide a stopping criterion for the proposed distributed scheme. To this end, let $S_{d,i}(t)$ denote the quantity $tP_{d,i}(t)$, and let $\gamma_{d,i}^h$ and $\gamma_{d,i}^l$ be thresholds at an agent $i$ (to be determined later) such that agent $i$ stops and makes a decision only when,

$$S_{d,i}(t) \notin [\gamma_{d,i}^l, \gamma_{d,i}^h] \tag{16}$$

for the first time. The stopping time for reaching a decision at an agent $i$ is then defined as,

$$T_{d,i} = \inf\{t \ \ |S_{d,i}(t) \notin [\gamma_{d,i}^l, \gamma_{d,i}^h]\}, \tag{17}$$

and the following decision rule is adopted at $T_{d,i}$ :

$$H = \begin{cases} H_0 & S_{d,i}(T_{d,i}) \leq \gamma_{d,i}^l \\ H_1 & S_{d,i}(T_{d,i}) \geq \gamma_{d,i}^h. \end{cases} \tag{18}$$

We refer to this distributed scheme (11), (17) and (18) as the *consensus+innovations* SPRT ($\mathcal{CISPRT}$) hence forth.

**Remark 3.2.** *It is to be noted that the decision statistic update rule is distributed and recursive, in that, to realize (11) each agent needs to communicate its current statistic and a scalar function of its latest sensed observation to its neighbors only; furthermore, the local update rule (11) is a combination of a consensus term reflecting the weighted combination of neighbors' statistics and a local innovation term reflecting the new sensed information of itself and its neighbors. Note that the stopping times $T_{d,i}$'s are random and generally take different values for different agents. It is to be noted that the $T_{d,i}$'s are in fact stopping times with respect to the respective agent information filtrations $\mathcal{G}_{d,i}(t)$'s as defined in (9). For subsequent analysis we refer to the stopping time of an agent as the stopping time for reaching a decision at an agent.*

We end this section by providing some elementary properties of the distributed test statistics.

**Proposition 3.3.** *Let the Assumptions A1, A3 and A4 hold. For each $t$ and $i$, the statistic $S_{d,i}(t)$, defined in (16)-(18), is Gaussian under both $H_0$ and $H_1$. In particular, we have*

$$\mathbb{E}_0[S_{d,i}(t)] = -mt \quad and \quad \mathbb{E}_1[S_{d,i}(t)] = mt,$$

*and*

$$\mathbb{E}_0\left[(S_{d,i}(t) + mt)^2\right] = \mathbb{E}_1\left[(S_{d,i}(t) - mt)^2\right] \leq \frac{2mt}{N} + \frac{2mr^2(1 - r^{2t})}{1 - r^2}.$$

*Proof:* Recall from (1), $\eta_i(t)$ is distributed as $\mathcal{N}(m, 2m)$, $\forall i = 1, 2, \cdots, N$, when conditioned on hypothesis $H_1$ and where $m$ is the Kullback-Leibler divergence as defined in (2). Hence,

$$\mathbb{E}_1[S_{d,i}(t)] = \sum_{j=1}^{t} \mathbf{e}_i^\top \mathbf{W}^{t+1-j} \mathbb{E}_1[\eta(j)]$$

$$= m \sum_{j=1}^{t} \mathbf{e}_i^\top \mathbf{W}^{t+1-j} \mathbf{1}$$

$$\Rightarrow \mathbb{E}_1[S_{d,i}(t)] = mt.$$

We note that $\mathbf{S}_\eta = Cov(\eta(t)) = 2m\mathbf{I}$. By standard algebraic manipulations we have,

$$Var(S_{d,i}(t)) = \mathbb{E}_1\left[(S_{d,i}(t) - mt)^2\right] = \sum_{j=1}^{t} \mathbf{e}_i^\top \mathbf{W}^{t+1-j} \mathbf{S}_\eta \mathbf{W}^{t+1-j} \mathbf{e}_i$$

$$= \sum_{j=1}^{t} \mathbf{e_i}^\top (\mathbf{W^{t-j}} - \mathbf{J})\mathbf{S}_\eta(\mathbf{W^{t-j}} - \mathbf{J})\mathbf{e_i} + \sum_{j=1}^{t} \mathbf{e_i}^\top \mathbf{J S}_\eta \mathbf{J e_i}$$

$$= 2m \sum_{j=1}^{t} \mathbf{e_i}^\top (\mathbf{W}^{2(t-j)} - \mathbf{J})\mathbf{e_i} + 2m \sum_{j=1}^{t} \mathbf{e_i}^\top \mathbf{J e_i}$$

$$= 2m\| \sum_{j=1}^{t} \mathbf{e_i}^\top (\mathbf{W}^{2(t-j)} - \mathbf{J})\mathbf{e_i}\| + \frac{2mt}{N}$$

$$\leq 2m \sum_{j=1}^{t} \|\mathbf{e_i}^\top (\mathbf{W}^{2(t-j)} - \mathbf{J})\mathbf{e_i}\| + \frac{2mt}{N}$$

$$\leq 2m \sum_{j=1}^{t} \|\mathbf{e_i}^T \mathbf{e_i}\|\|\mathbf{W}^{2(t-j)} - \mathbf{J}\| + \frac{2mt}{N}$$

$$= 2m \sum_{j=0}^{t-1} r^{2j} + + \frac{2mt}{N}$$

$$\leq \frac{2mt}{N} + \frac{2m(1 - r^{2t})}{1 - r^2}.$$

The assertion for hypothesis $H_0$ follows in a similar way. ∎

## 4. MAIN RESULTS

We formally state the main results in this section, the proofs being provided in Section 7.

### A. Thresholds for the $\mathcal{CISPRT}$

In this section we derive thresholds for the $\mathcal{CISPRT}$, see (16)-(18), in order to ensure that the procedure terminates in finite time a.s. at each agent and the agents achieve specified error probability requirements. We emphasize that in the proposed approach, a particular agent has access to its one hop neighborhood's test statistics and latest sensed information only. Moreover the latest sensed information is accessed through a scalar function of the latest observation of the agents in an agent's neighborhood. Recall, by (11) and (16) the (distributed) test statistic at agent $i$ is given by

$$S_{d,i}(t) = \sum_{j=1}^{t} \mathbf{e}_i^\top \mathbf{W}^{t+1-j} \eta_j.$$

For the proposed $\mathcal{CISPRT}$, we intend to derive thresholds which guarantee the error performance in terms of the error probability requirements $\alpha$ and $\beta$, i.e., such that $\mathbb{P}_{FA}^{d,i} \leq \alpha$ and $\mathbb{P}_{M}^{d,i} \leq \beta$, $\forall i = 1, 2, \ldots, N$, where $\mathbb{P}_{FA}^{d,i}$ and $\mathbb{P}_{M}^{d,i}$ represent the probability of false alarm and the probability of miss for the $i$th agent defined as

$$\mathbb{P}_{FA}^{d,i} = \mathbb{P}_0(S_{d,i}(T_{d,i}) \geq \gamma_{d,i}^h)$$
$$\mathbb{P}_{M}^{d,i} = \mathbb{P}_1(S_{d,i}(T_{d,i}) \leq \gamma_{d,i}^l),$$

with $T_{d,i}$ as defined in (17).

**Theorem 4.1.** *Let the Assumptions A1-A4 hold.*

*1) Then, for each $\alpha$ and $\beta$ there exist $\gamma_{d,i}^h$ and $\gamma_{d,i}^l$, $\forall i = 1, 2, \ldots, N$, such that $\mathbb{P}_{FA}^{d,i} \leq \alpha$ and $\mathbb{P}_{M}^{d,i} \leq \beta$ and the test concludes in finite time a.s. i.e.*

$$\mathbb{P}_1(T_{d,i} < \infty) = 1, \forall i = 1, 2, \ldots, N,$$

*where $T_{d,i}$ is the stopping time for reaching a decision at agent $i$.*

*2) In particular, for given $\alpha$ and $\beta$, any choice of thresholds $\gamma_{d,i}^h$ and $\gamma_{d,i}^l$ satisfying*

$$\gamma_{d,i}^h \geq \frac{8(k+1)}{7N} \left( \log \left( \frac{2}{\alpha} \right) - \log(1 - e^{\frac{-Nm}{4(k+1)}}) \right) = \gamma_d^{h,0} \tag{19}$$

$$\gamma_{d,i}^l \leq \frac{8(k+1)}{7N} \left( \log \left( \frac{\beta}{2} \right) + \log(1 - e^{\frac{-Nm}{4(k+1)}}) \right) = \gamma_d^{l,0}, \tag{20}$$

*where $m$ is defined in (2) and $k$ is defined by*

$$Nr^2 = k,$$

*with $r$ as in (13), achieves a.s. finite stopping at an agent $i$ while ensuring that $\mathbb{P}_{FA}^{d,i} \leq \alpha$ and $\mathbb{P}_{M}^{d,i} \leq \beta$.*

The first assertion ensures that for any set of pre-specified error metrics $\alpha$ and $\beta$ (satisfying Assumption A2), the $\mathcal{CISPRT}$ can be designed to achieve the error requirements while ensuring finite stopping a.s. It is to be noted that the ranges associated with the thresholds in (19)-(20) provide sufficient threshold design conditions for achieving pre-specified performance, but may not be necessary. The thresholds chosen according to (19)-(20) are not guaranteed to be optimal in the sense of the expected stopping time of the $\mathcal{CISPRT}$ algorithm and there might exist better thresholds (in the sense of expected stopping time) that achieve the pre-specified error requirements.

**Remark 4.2.** *We remark the following: 1) We have shown that the $\mathcal{CISPRT}$ algorithm can be designed so as to achieve the pre-specified error metrics at every agent $i$. This, in turn, implies that the probability of not reaching decision consensus among the agents can be upper bounded by $N\beta$ when conditioned on $H_1$ and $N\alpha$ when conditioned on $H_0$. It is to be noted that with $\alpha \to 0$ and $\beta \to 0$, the probability of not reaching decision consensus conditioned on either of the hypothesis goes to $0$ as well; 2) The factor $k$ in the closed form expressions of the thresholds in* (19) *and* (20) *relates the value of the thresholds to the rate of flow of information $r$ and, hence, in turn, can be related to the degree of connectivity of the inter-agent communication graph under consideration, see* (13)-(14) *and the accompanying discussion. From Assumption A4, we have that $r < 1$. As $r$ goes smaller, which intuitively means increased rate of flow of information in the inter-agent network, the value of thresholds needed to achieve the pre-specified error metrics become smaller i.e. the interval $[\gamma_{d,i}^l, \gamma_{d,i}^h]$ shrinks for all $i = 1, 2, \ldots, N$.*

*B. Probability Distribution of $T_{d,i}$ and $T_c$*

We first characterize the stopping time distributions for the centralized SPRT detector (see Section 2-B) and those of the distributed $\mathcal{CISPRT}$. Subsequently, we compare the centralized and distributed stopping times by studying their respective large deviation tail probability decay rates.

**Theorem 4.3.** *([44], [45]) Let the Assumptions A1 and A2 hold and given the SPRT for the centralized setup in* (5)-(7)*, we have*

$$\mathbb{P}_1(T_c > t) \geq \exp\left(\frac{N\mu\gamma_c^l}{\sigma^2}\right) K_t^\infty\left(\gamma_c^h\right) - \exp\left(\frac{N\mu\gamma_c^h}{\sigma^2}\right) K_t^\infty\left(\gamma_c^l\right), \tag{21}$$

*where*

$$K_t^S(a) = \frac{\sigma^2\pi}{N(\gamma_c^h - \gamma_c^l)^2} \sum_{s=1}^S \frac{l(-1)^{l+1}}{\frac{Nm}{4} + \frac{\sigma^2 s^2 \pi^2}{2N(\gamma_c^h - \gamma_c^l)^2}} \exp\left(-\left(\frac{Nm}{4} + \frac{\sigma^2 s^2 \pi^2}{2N(\gamma_c^h - \gamma_c^l)^2}\right)t\right) \sin\left(\frac{s\pi a}{\gamma_c^h - \gamma_c^l}\right),$$

*whereas, $T_c$ is defined in* (6) *and $\gamma_c^h$ and $\gamma_c^l$ are the associated SPRT thresholds chosen to achieve specified error requirements $\alpha$ and $\beta$.*

The above characterization of the stopping distribution of Wald's SPRT was obtained in [44], [45]. In particular, this was derived by studying the first passage time distribution of an associated continuous time Wiener process with a constant drift; intuitively, the continuous time approximation of the discrete time SPRT consists of replacing the discrete time likelihood increments by a Wiener process accompanied by a constant drift that reflects the mean of the hypothesis in place. This way, the sequence obtained by sampling the continuous time process at integer time instants is equivalent in distribution to the

(discrete time) Wald's SPRT considered in this paper. The term on the R.H.S. of (21) is exactly equal to the probability that the first passage time of the continuous time Wiener process with left and right boundaries $\gamma_c^l$ and $\gamma_c^h$ respectively is greater than $t$, whereas, is, in general, a lower bound for the discrete time SPRT (as given in Theorem 4.3) as increments in the latter happen at discrete (integer) time instants only.

We now provide a characterization of the stopping time distributions of the $\mathcal{CISPRT}$ algorithm.

**Lemma 4.4.** *Let the assumptions A1-A4 hold. Consider the $\mathcal{CISPRT}$ algorithm given in* (12), (17) *and* (18) *and suppose that, for specified $\alpha$ and $\beta$, the thresholds $\gamma_{d,i}^h$ and $\gamma_{d,i}^l$, $i = 1, \cdots, N$, are chosen to satisfy the condtions derived in* (19) *and* (20). *We then have,*

$$\mathbb{P}_1(T_{d,i} > t) \leq \mathbb{Q}\left(\frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right), \quad \forall i = 1, 2, \ldots, N, \tag{22}$$

*where $T_{d,i}$ is the stopping time of the $i$-th agent to reach a decision as defined in* (17).

### C. Comparison of stopping times of the distributed and centralized detectors

In this section we compare the stopping times $T_c$ and $T_{d,i}$ by studying their respective large deviation tail probability decay rates. We utilize the bounds derived in Theorem 4.3 and Lemma 4.4 to this end.

**Corollary 4.5.** *Let the hypotheses of Lemma 4.3 hold. Then we have the following large deviation characterization for the tail probabilities of $T_c$:*

$$\liminf_{t \to \infty} \frac{1}{t} \log(\mathbb{P}_1(T_c > t)) \geq -\frac{Nm}{4} - \frac{\sigma^2\pi^2}{2N(\gamma_c^h - \gamma^l)^2}.$$

It is to be noted that the exponent is a function of the thresholds $\gamma_c^h$ and $\gamma_c^l$ and with the decrease in the error constraints $\alpha$ and $\beta$, $\frac{Nm}{4} + \frac{\sigma^2\pi^2}{2N(\gamma_c^h - \gamma^l)^2} \approx \frac{Nm}{4}$.

**Theorem 4.6.** *Let the hypotheses of Lemma 4.4 hold. Then we have the following large deviation characterization for the tail probabilities of the $T_{d,i}$'s:*

$$\limsup_{t \to \infty} \frac{1}{t} \log(\mathbb{P}_1(T_{d,i} > t)) \leq -\frac{Nm}{4} \quad , \forall i = 1, 2, \ldots, N.$$

Importantly, the upper bound for the large deviation exponent of the $\mathcal{CISPRT}$ in Theorem 4.6 is independent of the inter-agent communication topology as long as the connectivity conditions Assumptions A3-A4 hold. Finally, in the asymptotic regime, i.e., as $N$ goes to $\infty$, since $\frac{\sigma^2\pi^2}{2N(\gamma_c^h - \gamma^l)^2} = o(Nm)$, we have that the performance of the distributed $\mathcal{CISPRT}$ approaches that of the centralized SPRT, in the sense of stopping time tail exponents, as $N$ tends to $\infty$.

*D. Comparison of the expected stopping times of the centralized and distributed detectors*

In this section we compare the expected stopping times of the centralized SPRT detector and the proposed $\mathcal{CISPRT}$ detector. Recall that $\mathbb{E}_j[T_{d,i}]$ and $\mathbb{E}_j[T_c]$ represent the expected stopping times for reaching a decision for the $\mathcal{CISPRT}$ (at an agent $i$) and its centralized counterpart respectively, where $j \in \{0,1\}$ denotes the hypothesis on which the expectations are conditioned on. Without loss of generality we compare the expectations conditioned on Hypothesis $H_1$, similar conclusions (with obvious modifications) hold when the expectations are conditioned on $H_0$ (see also Section 2-B).

Also, for the sake of mathematical brevity and clarity, we approximate $\alpha = \beta = \epsilon$ in this subsection. Recall Section 2-B and note that, at any instant of time $t$, the information $\sigma$-algebra $\mathcal{G}_{d,i}(t)$ at any agent $i$ is a subset of $\mathcal{G}_c(t)$, the information $\sigma$-algebra of a (hypothetical) center, which has access to the data of all agents at all times. This implies that any distributed procedure (in particular the $\mathcal{CISPRT}$) can be implemented in the centralized setting, and, since $M(\epsilon)$ (see (4)) constitutes a lower bound on the expected stopping time of any sequential test achieving error probabilities $\alpha = \beta = \epsilon$, we have that

$$\frac{\mathbb{E}_1[T_{d,i}]}{\mathcal{M}(\epsilon)} \geq 1, \quad \forall \ i = 1, 2, \ldots, N, \tag{23}$$

for all $\epsilon \in (0, 1/2)$. In order to provide an upper bound on the ratio $\mathbb{E}_1[T_{d,i}]/\mathcal{M}(\epsilon)$ and, hence, compare the performance of the proposed $\mathcal{CISPRT}$ detector with the optimal centralized detector, we first obtain a characterization of $\mathbb{E}_1[T_{d,i}]$ in terms of the algorithm thresholds as follows.

**Theorem 4.7.** *Let the assumptions A1-A4 hold and let $\alpha = \beta = \epsilon$. Suppose that the thresholds of the $\mathcal{CISPRT}$ be chosen as $\gamma_{d,i}^h = \gamma_d^{h,0}$ and $\gamma_{d,i}^l = \gamma_d^{l,0}$ for all $i = 1, \cdots, N$, where $\gamma_d^{h,0}$ and $\gamma_d^{l,0}$ are defined in (19)-(20). Then, the stopping time $T_{d,i}$ of the $\mathcal{CISPRT}$ at an agent $i$ satisfies*

$$\frac{(1-2\epsilon)\gamma_{d,i}^h}{m} - \frac{c}{m} \leq \mathbb{E}_1[T_{d,i}] \leq \frac{5\gamma_{d,i}^h}{4m} + \frac{1}{1 - e^{\frac{-Nm}{4(k+1)}}},$$

*where $k = Nr^2$, $r$ is as defined in (13), and $c > 0$ is a constant that may be chosen to be independent of the thresholds and the $\epsilon$.*

It is to be noted that, when $\alpha = \beta = \epsilon$, then $\gamma_{d,i}^h = -\gamma_{d,i}^l$ from (19) and (20). The upper bound derived in the above assertion might be loose, owing to the approximations related to the non-elementary $\mathbb{Q}$-function. We use the derived upper bound for comparing the performance of the $\mathcal{CISPRT}$ algorithm with that of its centralized counterpart. The constant $c > 0$ in the lower bound is independent of the thresholds $\gamma_{d,i}^l$ and $\gamma_{d,i}^h$ (and hence, also independent of the error tolerance $\epsilon$) and is a function of the network topology and the Gaussian model statistics only. Explicit expressions and bounds on $c$ may be obtained by refining

the various estimates in the proofs of Lemma 7.2 and Theorem 4.7, see Section 4. However, for the current purposes, it is important to note that $c = o(\gamma_d^{h,0})$, i.e., as $\epsilon$ goes to zero or equivalently in the limit of large thresholds $c/\gamma_d^{h,0} \to 0$. Hence, as $\epsilon \to 0$, the more readily computable quantity $\frac{(1-2\epsilon)\gamma_{d,i}^h}{m}$ may be viewed as a reasonably good approximation to the lower bound in Theorem 4.7.

**Theorem 4.8.** *Let the hypotheses of Theorem 4.7 hold. Then, we have the following characterization of the ratio of the expected stopping times of the $\mathcal{CISPRT}$ and the centralized detector in asymptotics of the $\epsilon$,*

$$1 \leq \limsup_{\epsilon \to 0} \frac{\mathbb{E}_1[T_{d,i}]}{\mathcal{M}(\epsilon)} \leq \frac{10(k+1)}{7}, \quad \forall i = 1, 2, \ldots, N, \tag{24}$$

*where $k = Nr^2$ and $r$ is as defined in (13).*

Theorem 4.8 shows that the $\mathcal{CISPRT}$ algorithm can be designed in such a way that with pre-specified error metrics $\alpha$ and $\beta$ going to $0$, the ratio of the expected stopping time for the $\mathcal{CISPRT}$ algorithm and its centralized counterpart are bounded above by $\frac{10(k+1)}{7}$ where the quantity $k$ depends on $r$ which essentially quantifies the dependence of the $\mathcal{CISPRT}$ algorithm on the network connectivity.

**Remark 4.9.** *It is to be noted that the derived upper bound for the ratio of the expected stopping times of the $\mathcal{CISPRT}$ algorithm and its centralized counterpart may not be a tight upper bound. The looseness in the upper bound is due to the fact that the set of thresholds chosen are oriented to be sufficient conditions and not necessary. As pointed out in Remark 4.2 there might exist possibly better choice of thresholds for which the pre-specified error metrics are satisfied. Hence, given a set of pre-specified error metrics and a network topology the upper bound of the derived assertion above can be minimized by choosing the optimal weights for $\mathbf{W}$ as shown in [41]. It can be seen that the ratio of expected stopping times of the isolated SPRT based detector case, i.e., the non-collaboration case, and the centralized SPRT based detector is $N$ (see Section 2-B). So, for the $\mathcal{CISPRT}$ case in order to make savings as far as the stopping time is concerned with respect to the isolated SPRT based detector, $\frac{10(k+1)}{7} \leq N$ should be satisfied. Hence, we have that $r \leq \sqrt{\frac{7N-10}{10N}}$ is a sufficient condition for the same.*

## 5. DEPENDENCE OF THE $\mathcal{CISPRT}$ ON NETWORK CONNECTIVITY: ILLUSTRATION

In this section, we illustrate the dependence of the $\mathcal{CISPRT}$ algorithm on the network connectivity, by considering a class of graphs. Recall from section 3 that the quantity $r$ quantifies the *rate of information flow* in the network, and in general, the smaller the $r$ the faster is the convergence of information dissemination algorithms (such as the consensus or gossip protocol ([14], [42], [43]) on the graph and

the optimal design of symmetric weight matrices $\mathbf{W}$ for a given network topology that minimizes the value $r$ can be cast as a semi-definite optimization problem [41].

To quantify the dependance of the $\mathcal{CISPRT}$ algorithm on the graph topology, we note that the limit derived in (24) is a function of $\mathbf{W}$ and can be re-written as follows :

$$\limsup_{\epsilon \to 0} \frac{\mathbb{E}_1[T_{d,i}]}{\mathcal{M}(\epsilon)} \leq \frac{10(Nr^2 + 1)}{7} \doteq \mathcal{R}(\mathbf{W}), \tag{25}$$

i.e., the derived upper bound $\mathcal{R}(\mathbf{W})$ is a function of the chosen weight matrix $W$. Based on (25), naturally, a weight design guideline would be to design $\mathbf{W}$ (under the network topological constraints) so as to minimize $\mathcal{R}(\mathbf{W})$, which, by (25) and as discussed earlier corresponds to minimizing $r = \|\mathbf{W} - \mathbf{J}\|$. This leads to the following upper bound on the achievable performance of the $\mathcal{CISPRT}$:

$$\limsup_{\epsilon \to 0} \frac{\mathbb{E}_1[T_{d,i}]}{\mathcal{M}(\epsilon)} \leq \min_{\mathbf{W}} \mathcal{R}(\mathbf{W}).$$

By restricting attention to constant link weights, i.e., $\mathbf{W}$'s of the form $(\mathbf{I} - \delta\mathbf{L})$ and noting that

$$\min_{\delta} \|\mathbf{I} - \delta\mathbf{L} - \mathbf{J}\| = \frac{(\lambda_N(\mathbf{L}) - \lambda_2(\mathbf{L}))}{(\lambda_2(\mathbf{L}) + \lambda_N(\mathbf{L}))},$$

(see (15)), we further obtain

$$\limsup_{\epsilon \to 0} \frac{\mathbb{E}_1[T_{d,i}]}{\mathcal{M}(\epsilon)} \leq \min_{\mathbf{W}} \mathcal{R}(\mathbf{W}) \leq \min_{\delta} \mathcal{R}(\mathbf{I} - \delta\mathbf{L}) = \frac{10}{7} + \frac{10N(\lambda_N(\mathbf{L}) - \lambda_2(\mathbf{L}))^2}{7(\lambda_2(\mathbf{L}) + \lambda_N(\mathbf{L}))^2}. \tag{26}$$

The final bound obtained in (26) might not be tight, being an upper bound (there may exist $\mathbf{W}$ matrices not of the form $\mathbf{I} - \delta\mathbf{L}$ with smaller $r$) to a possibly loose upper bound derived in (24), but, nonetheless, directly relates the performance of the $\mathcal{CISPRT}$ to the spectra of the graph Laplacian and hence the graph topology. From (26) we may further conclude that networks with smaller value of the ratio $\lambda_2(\mathbf{L})/\lambda_N(\mathbf{L})$ tend to achieve better performance. This leads to an interesting graph design question: given resource constraints, specifically, say a restriction on the number of edges of the graph, how to design inter-agent communication networks that tend to minimize the eigen-ratio $\lambda_2(\mathbf{L})/\lambda_N(\mathbf{L})$ so as to achieve improved $\mathcal{CISPRT}$ performance. To an extent, such graph design questions have been studied in prior work, see [15], which, for instance, shows that expander graphs tend to achieve smaller $\lambda_2(\mathbf{L})/\lambda_N(\mathbf{L})$ ratios given a constraint on the total number of network edges.

## 6. SIMULATIONS

We generate planar random geometric networks of 30, 300 and 1000 agents. The $x$ coordinates and the $y$ coordinates of the agents are sampled from an uniform distribution on the open interval $(0, 1)$.
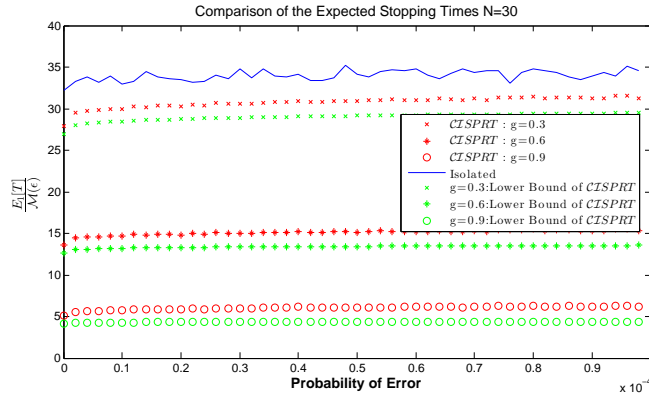
Fig. 1: Comparison of Stopping Time Distributions for N=30

We link two vertices by an edge if the distance between them is less than or equal to $g$. We go on re-iterating this procedure until we get a connected graph. We construct the geometric network for each of $N = 30, 300$ and $1000$ cases with three different values of $g$ i.e. $g = 0.3, 0.6$ and $0.9$. The values of $r$ obtained in each case is specified in Table I. We consider two cases, the $\mathcal{CISPRT}$ case and the

| r | g=0.3 | g=0.6 | g=0.9 |
|---|---|---|---|
| N=30 | 0.8241 | 0.5580 | 0.2891 |
| N=300 | 0.7989 | 0.6014 | 0.2166 |
| N=1000 | 0.7689 | 0.5940 | 0.2297 |

TABLE I: Values of $r$

non-collaborative case. We consider $\alpha = \beta = \epsilon$ and ranging from $10^{-8}$ to $10^{-4}$ in steps of $10^{-6}$. For each such $\epsilon$, we conduct 2000 simulation runs to empirically estimate the stopping time distribution $\mathbb{P}_1(T > t)$ of a randomly chosen agent (with uniform selection probability) for each of the cases. From these empirical probability distributions of the stopping times, we estimate the corresponding expected stopping times. Figure 4 shows the instantaneous behavior of the test statistics in the case of $N = 300$ with $\epsilon = 10^{-10}$. In Figures 1, 2 and 3 it is demonstrated that the ratio of the expected stopping time of the $\mathcal{CISPRT}$ algorithm and the universal lower bound $\mathcal{M}(\epsilon)$ is less than that of the ratio of the expected stopping times of the isolated (non-collaborative) case and $\mathcal{M}(\epsilon)$. The ratio of the theoretical lower bound of the expected stopping time of the $\mathcal{CISPRT}$ derived in Theorem 4.7 and $\mathcal{M}(\epsilon)$ was also studied. More precisely, we compared the experimental ratio of the expected stopping times of the $\mathcal{CISPRT}$ and $\mathcal{M}(\epsilon)$
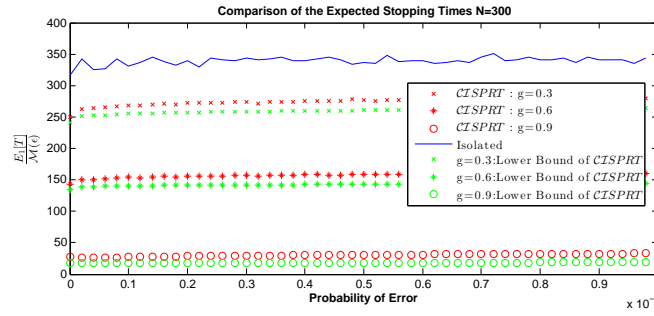
Fig. 2: Comparison of Stopping Time Distributions for N=300
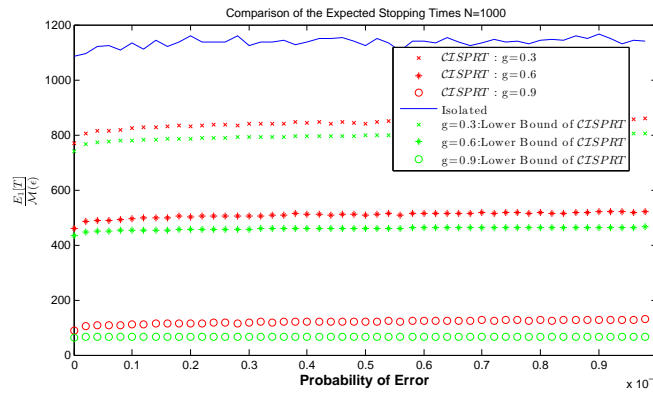


Fig. 3: Comparison of Stopping Time Distributions for N=1000
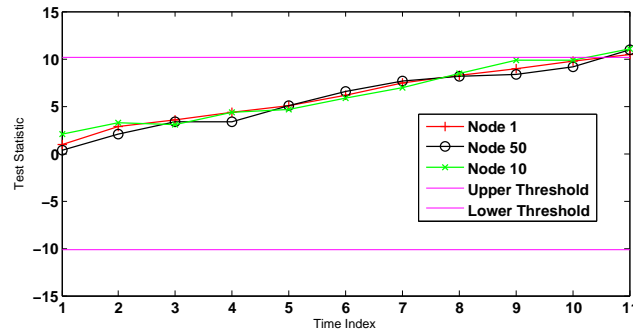


Fig. 4: Instantaneous behavior of $S_{d,i}(t)$

with the ratio of the quantity $\frac{(1-2\epsilon)\gamma_{d,i}^h}{m}$ (the small $\epsilon$ approximation of the theoretical lower bound given in Theorem 4.7, see also the discussion provided in Section 4 after the statement of Theorem 4.7) and $\mathcal{M}(\epsilon)$. It can be seen that the experimental ratio of the expected stopping times of the $\mathcal{CISPRT}$ and $\mathcal{M}(\epsilon)$ is very close to the ratio of the (approximate) theoretical lower bound of expected stopping time of the $\mathcal{CISPRT}$ and $\mathcal{M}(\epsilon)$, which shows that the lower bound derived in Theorem 4.7 is reasonable. Figure 4 is an example of a single run of the algorithm which shows the instantaneous behavior of the distributed test statistic $S_{d,i}(t)$ for $N = 300$, where we have plotted three randomly chosen agents i.e. $i = 1$, $i = 10$ and $i = 50$.

## 7. PROOFS OF MAIN RESULTS

*Proof of Theorem 4.1:* Let $\hat{A} = e^{\gamma_{d,i}^l}$ and $\hat{B} = e^{\gamma_{d,i}^h}$ where $\gamma_{d,i}^h$ and $\gamma_{d,i}^l \in \mathbb{R}$ are thresholds (to be designed) for the $\mathcal{CISPRT}$. In the following derivation, for a given random variable $z$ and an event $A$, we use the notation $\mathbb{E}[z; A]$ to denote the expectation $\mathbb{E}[z\mathbb{I}_A]$. Let $T$ denote the random time which can take values in $\overline{\mathbb{Z}}_+$ given by

$$T = \inf \left\{ t | S_{d,i}(t) \notin \left[ \gamma_{d,i}^l, \gamma_{d,i}^h \right] \right\}. \tag{27}$$

First, we show that for any $\gamma_{d,i}^h$ and $\gamma_{d,i}^l \in \mathbb{R}$,

$$\mathbb{P}_0 \left( T < \infty \right) = \mathbb{P}_1 \left( T < \infty \right) = 1, \tag{28}$$

i.e., the random time $T$ defined in (27) is a.s. finite under both the hypotheses. Indeed, we have,

$$\mathbb{P}_1 \left( T > t \right) \leq \mathbb{Q} \left( \frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}} \right)$$

$$\Rightarrow \lim_{t \to \infty} \mathbb{P}_1 \left( T > t \right) = 0$$

$$\Rightarrow \mathbb{P}_1 \left( T < \infty \right) = 1.$$

The proof for $H_0$ follows in a similar way.

Now, since (28) holds, the quantity $S_{d,i}(T)$ is well-defined a.s. under $H_0$. Now, noting that, under $H_0$, for any $t$, the quantity $S_{d,i}(t)$ is Gaussian with mean $-mt$ and variance upper bounded by $\frac{2mt}{N} +$

$\frac{2mr^2(1-r^{2t})}{1-r^2}$ (see Proposition 3.3), we have,

$$\mathbb{P}_{FA}^{d,i} = \mathbb{P}_0(S_{d,i}(T) \geq \log \hat{B}) = \sum_{t=1}^{\infty} \mathbb{P}_0(T = t, S_{d,i}(t) \geq \log \hat{B})$$

$$\leq \sum_{t=1}^{\infty} \mathbb{P}_0(S_{d,i}(t) \geq \log \hat{B})$$

$$\leq \sum_{t=1}^{\infty} \mathbb{Q}\left(\frac{\log \hat{B} + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right). \tag{29}$$

To obtain a condition for $\gamma_{d,i}^h$ in the $\mathcal{CISPRT}$ such that $\mathbb{P}_{FA}^{d,i} \leq \alpha$, let's define $k > 0$ such that $k = Nr^2$. Now, note that $k$ thus defined satisfies

$$\frac{2mr^2(1-r^{2t})}{1-r^2} \leq \frac{2mkt}{N}, \quad \forall t. \tag{30}$$

Then we have, by (29)-(30),

$$\mathbb{P}_{FA}^{d,i} \leq \sum_{t=1}^{\infty} \mathbb{Q}\left(\frac{\log \hat{B} + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right) \leq \sum_{t=1}^{\infty} \mathbb{Q}\left(\frac{\log \hat{B} + mt}{\sqrt{\frac{2mt(k+1)}{N}}}\right) \leq \frac{1}{2} \sum_{t=1}^{\infty} e^{\frac{-(\gamma_{d,i}^h)^2 - m^2t^2 - 2\gamma_{d,i}^h mt}{\frac{4mt(k+1)}{N}}}$$

$$= \frac{e^{-\frac{N\gamma_{d,i}^h}{2(k+1)}}}{2}\left(\sum_{t=1}^{\lfloor \frac{\gamma_{d,i}^h}{2m} \rfloor} e^{\frac{-N(\gamma_{d,i}^h)^2 - Nm^2t^2}{4mt(k+1)}} + \sum_{t=\lfloor \frac{\gamma_{d,i}^h}{2m} \rfloor + 1}^{\lfloor \frac{\gamma_{d,i}^h}{m} \rfloor} e^{\frac{-N(\gamma_{d,i}^h)^2 - Nm^2t^2}{4mt(k+1)}}\right.$$

$$\left. + \sum_{t=\lfloor \frac{\gamma_{d,i}^h}{m} \rfloor + 1}^{\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor} e^{\frac{-N(\gamma_{d,i}^h)^2 - Nm^2t^2}{4mt(k+1)}} + \sum_{t=\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor + 1}^{\infty} e^{\frac{-N(\gamma_{d,i}^h)^2 - Nm^2t^2}{4mt(k+1)}}\right)$$

$$\leq \frac{e^{-\frac{N\gamma_{d,i}^h}{2(k+1)}}}{2}\left(e^{-\frac{N\gamma_{d,i}^h}{2(k+1)}} \underbrace{\sum_{t=1}^{\lfloor \frac{\gamma_{d,i}^h}{2m} \rfloor} e^{\frac{-Nmt}{4(k+1)}}}_{(1)} + e^{-\frac{N\gamma_{d,i}^h}{4(k+1)}} \underbrace{\sum_{t=\lfloor \frac{\gamma_{d,i}^h}{2m} \rfloor + 1}^{\lfloor \frac{\gamma_{d,i}^h}{m} \rfloor} e^{\frac{-Nmt}{4(k+1)}}}_{(2)}\right.$$

$$\left. + e^{-\frac{N\gamma_{d,i}^h}{8(k+1)}} \underbrace{\sum_{t=\lfloor \frac{\gamma_{d,i}^h}{m} \rfloor + 1}^{\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor} e^{\frac{-Nmt}{4(k+1)}}}_{(3)} + \underbrace{\sum_{t=\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor + 1}^{\infty} e^{\frac{-Nmt}{4(k+1)}}}_{(4)}\right)$$

$$\leq \frac{e^{-\frac{N\gamma_{d,i}^h}{2(k+1)}}}{2(1 - e^{-\frac{Nm}{4(k+1)}})}\left(e^{-\frac{N\gamma_{d,i}^h}{2(k+1)}} + e^{-\frac{N\gamma_{d,i}^h}{4(k+1)}} e^{-\frac{N\gamma_{d,i}^h}{8(k+1)}} + e^{-\frac{N\gamma_{d,i}^h}{8(k+1)}} e^{-\frac{N\gamma_{d,i}^h}{4(k+1)}} + e^{-\frac{N\gamma_{d,i}^h}{2(k+1)}}\right)$$

$$\leq \frac{2e^{-\frac{7N\gamma_{d,i}^h}{8(k+1)}}}{1 - e^{-\frac{Nm}{4(k+1)}}}. \tag{31}$$

In the above set of equations we use the fact that $\mathbb{Q}(x)$ is a non-increasing function, the inequality $\mathbb{Q}(x) \leq \frac{1}{2}e^{\frac{-x^2}{2}}$, and we upper bound $(1) - (4)$ by their infinite geometric sums.

We now note that, a sufficient condition for $\mathbb{P}_{FA}^{d,i} \leq \alpha$ to hold is the following:

$$\frac{2e^{-\frac{7N\gamma_{d,i}^h}{8(k+1)}}}{1 - e^{-\frac{Nm}{4(k+1)}}} \leq \alpha. \tag{32}$$

Solving (32), we have that, any $\gamma_{d,i}^h$ that satisfies

$$\gamma_{d,i}^h \geq \gamma_d^{h,0} = \frac{8(k+1)}{7N}\left(\log\left(\frac{2}{\alpha}\right) - \log(1 - e^{-\frac{Nm}{4(k+1)}})\right), \tag{33}$$

achieves $\mathbb{P}_{FA}^{d,i} \leq \alpha$ in the $\mathcal{CISPRT}$.

Proceeding as in (29) and (31) we have that, any $\gamma_{d,i}^l$ that satisfies

$$\gamma_{d,i}^l \leq \gamma_d^{l,0} \doteq \frac{8(k+1)}{7N}\left(\log\left(\frac{\beta}{2}\right) + \log(1 - e^{-\frac{Nm}{4(k+1)}})\right), \tag{34}$$

achieves $\mathbb{P}_M^{d,i} \leq \beta$ in the $\mathcal{CISPRT}$.

Clearly, by the above, any pair $(\gamma_{d,i}^h, \gamma_{d,i}^l)$ satisfying $\gamma_{d,i}^h \in [\gamma_d^{h,0}, \infty)$ and $\gamma_{d,i}^l \in (-\infty, \gamma_d^{l,0}]$ (see (33) and (34)) ensures that $\mathbb{P}_{FA}^{d,i} \leq \alpha$ and $\mathbb{P}_M^{d,i} \leq \beta$. The a.s. finiteness of the corresponding stopping time $T_{d,i}$ (see (17)) under both $H_0$ and $H_1$ follows readily by arguments as in (28). ∎

**Remark 7.1.** *It is to be noted that the derived thresholds are sufficient conditions only. The approximations (see $(1) - (4)$ in (31)) made in the steps of deriving the expressions of the thresholds were done so as to get a tractable expression of the range. By solving the following set of equations*

$$\frac{1}{2}\sum_{t=1}^{\infty} e^{\frac{-N(\gamma_{d,i}^l)^2 - Nm^2t^2 + 2N\gamma_{d,i}^l mt}{4mt(k+1)}} \leq \beta$$

$$\frac{1}{2}\sum_{t=1}^{\infty} e^{\frac{-N(\gamma_{d,i}^h)^2 - Nm^2t^2 - 2N\gamma_{d,i}^h mt}{4mt(k+1)}} \leq \alpha$$

*numerically, tighter thresholds can be obtained.*

*Proof of Lemma 4.4:* Let us define the event $A_s^i$ as $\{\gamma_{d,i}^l \leq S_{d,i}(s) \leq \gamma_{d,i}^h\}$. Now, note that

$$\mathbb{P}_1(T_{d,i} > t) = \mathbb{P}_1(\cap_{s=1}^t A_s^i),$$

and

$$\mathbb{P}_1(\cap_{s=1}^t A_s^i) \leq \mathbb{P}_1(A_t^i).$$

By Proposition 3.3, under $H_1$, for any $t$, the quantity $S_{d,i}(t)$ is Gaussian with mean $mt$ and variance upper bounded by $\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}$. Hence we have, for all $i = 1, 2, \ldots, N$.

$$\mathbb{P}_1(T_{d,i} > t) \leq \mathbb{Q}\left( \frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}} \right). \tag{35}$$

∎

*Proof of Corollary 4.5:* For simplicity of notation, let $a = \frac{Nm}{4}$ and $b = \frac{\sigma^2\pi^2}{2N(\gamma_c^h - \gamma_c^l)^2}$. From (21), we have,

$$\frac{1}{t}\log(\mathbb{P}_1(T_c > t)) \geq \frac{1}{t}\log\left( \exp\left( \frac{N\mu\gamma_c^l}{\sigma^2} \right) K_t^\infty\left( \gamma_c^h \right) - \exp\left( \frac{N\mu\gamma_c^h}{\sigma^2} \right) K_t^\infty\left( \gamma_c^l \right) \right)$$

$$= \frac{1}{t}\log\left( \exp\left( -(a+b)t \right) \right)$$

$$+ \frac{1}{t}\log\left( b\sum_{s=1}^{\infty} \frac{s(-1)^{s+1}}{a + s^2 b} \exp\left( -b(s^2-1)t \right) \right.$$

$$\times \left( \exp\left( \frac{N\mu\gamma_c^l}{\sigma^2} \right) \sin\left( \frac{s\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) - \exp\left( \frac{N\mu\gamma_c^h}{\sigma^2} \right) \sin\left( \frac{s\pi\gamma_c^l}{\gamma_c^h - \gamma_c^l} \right) \right) \right). \tag{36}$$

For all $t, S \geq 1$, let

$$U(t, S) = \frac{1}{t}\log\left( b\sum_{s=1}^{S} \frac{s(-1)^{s+1}}{a + s^2 b} \exp\left( -b(s^2-1)t \right) \right.$$

$$\times \left( \exp\left( \frac{N\mu\gamma_c^l}{\sigma^2} \right) \sin\left( \frac{s\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) - \exp\left( \frac{N\mu\gamma_c^h}{\sigma^2} \right) \sin\left( \frac{s\pi\gamma_c^l}{\gamma_c^h - \gamma_c^l} \right) \right) \right)$$

and let $g = \exp\left( \frac{N\mu\gamma_c^h}{\sigma^2} \right) + \exp\left( \frac{N\mu\gamma_c^l}{\sigma^2} \right)$.

Note that for all $t \geq 1$, the limit

$$\lim_{S \to \infty} U(t, S) \tag{37}$$

exists and is finite (by Theorem 4.3), and similarly for all $S \geq 1$,

$$\lim_{t \to \infty} U(t, S) = \lim_{t \to \infty} \frac{1}{t}\log\left( b\sum_{s=1}^{S} \frac{s(-1)^{s+1}}{a + s^2 b} \exp\left( -b(s^2-1)t \right) \right.$$

$$\times \sin\left( \frac{s\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) \left( \exp\left( \frac{N\mu\gamma_c^l}{\sigma^2} \right) + (-1)^{s+1} \exp\left( \frac{N\mu\gamma_c^h}{\sigma^2} \right) \right) \right)$$

$$= \lim_{t \to \infty} \frac{1}{t}\log\left( \frac{bg}{a+b} \sin\left( \frac{\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) \right)$$

$$= 0, \tag{38}$$

where we use the fact that only the largest exponent in a finite summation of exponential terms contributes to its log-normalised limit as $t \to \infty$ and

$$\sin\left( \frac{s\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) = (-1)^s \sin\left( \frac{s\pi\gamma_c^l}{\gamma_c^h - \gamma_c^l} \right).$$

Finally, using the fact that there exists a constant $c_5 > 0$ (independent of $t$ and $S$) such that for all $t, S \geq 1$,

$$U(t, S) \leq \frac{1}{t} \log \left( bg \sum_{s=1}^{S} \frac{s}{a + s^2 b} \exp\left(-b(s^2 - 1)t\right) \right) \leq c_5, \tag{39}$$

we may conclude that the convergence in (37)-(38) are uniform in $S$ and $t$ respectively. This in turn implies that the order of the limits may be interchanged and we have that

$$\lim_{t \to \infty} \lim_{S \to \infty} U(t, S) = \lim_{S \to \infty} \lim_{t \to \infty} U(t, S) = 0. \tag{40}$$

Hence, we have from (36) and (40),

$$\liminf_{t \to \infty} \frac{1}{t} \log(\mathbb{P}_1(T_c > t)) \geq -(a + b)$$

$$+ \lim_{t \to \infty} \lim_{S \to \infty} \frac{1}{t} \log \left( b \sum_{s=1}^{S} \frac{s(-1)^{s+1}}{a + s^2 b} \exp\left(-b(s^2 - 1)t\right) \right.$$

$$\times \sin \left( \frac{s\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) \left( \exp\left( \frac{N\mu\gamma_c^l}{\sigma^2} \right) - (-1)^s \exp\left( \frac{N\mu\gamma_c^h}{\sigma^2} \right) \right) \right)$$

$$= -(a + b) + \lim_{S \to \infty} \lim_{t \to \infty} \frac{1}{t} \log \left( b \sum_{s=1}^{S} \frac{s(-1)^{s+1}}{a + s^2 b} \exp\left(-b(s^2 - 1)t\right) \right.$$

$$\times \sin \left( \frac{s\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) \left( \exp\left( \frac{N\mu\gamma_c^l}{\sigma^2} \right) + (-1)^{s+1} \exp\left( \frac{N\mu\gamma_c^h}{\sigma^2} \right) \right) \right)$$

$$= -(a + b) + \lim_{S \to \infty} \lim_{t \to \infty} \frac{1}{t} \log \left( \frac{bg}{a + b} \sin \left( \frac{\pi\gamma_c^h}{\gamma_c^h - \gamma_c^l} \right) \right)$$

$$= -(a + b) = -\left( \frac{Nm}{4} + \frac{\sigma^2 \pi^2}{2N(\gamma_c^h - \gamma_c^l)^2} \right).$$

∎

*Proof of Theorem 4.6:* We use the following upper bound for $\mathbb{Q}$ function in the proof below

$$\mathbb{Q}(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \tag{41}$$

From (22),(41) and (39), we have,

$$\limsup_{t\to\infty} \frac{1}{t} \log(\mathbb{P}_1(T_{d,i} > t))$$

$$\leq \limsup_{t\to\infty} \frac{1}{t} \log\left( \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\frac{2mt}{N} + 2m\frac{r^2(1-r^{2t})}{(1-r^2)}}}{(-\gamma_{d,i}^h + mt)} e^{\frac{-N(-\gamma_{d,i}^h + mt)^2}{4mt + 4mN\frac{r^2(1-r^{2t})}{(1-r^2)}}} \right)$$

$$\leq \limsup_{t\to\infty} \frac{1}{t} \left( \log\left( \frac{\sqrt{\frac{2mt}{N} + 2m\frac{r^2(1-r^{2t})}{(1-r^2)}}}{\sqrt{2\pi}(mt - \gamma_{d,i}^h)} \right) - \frac{N(\gamma_{d,i}^h)^2}{4mt + 4m\frac{r^2(1-r^{2t})}{(1-r^2)}} \right.$$

$$\left. - \frac{Nmt}{4 + 4\frac{r^2(1-r^{2t})}{(t(1-r^2))}} + \frac{Nm\gamma_{d,i}^h t}{2mt + 2mN\frac{r^2(1-r^{2t})}{(1-r^2)}} \right)$$

$$\Rightarrow \limsup_{t\to\infty} \frac{1}{t} \log(\mathbb{P}_1(T_{d,i} > t)) \leq -\frac{Nm}{4}.$$

∎

The proof of Theorem 4.7 requires an intermediate result that estimates the divergence between the agent statistics over time.

**Lemma 7.2.** *Let the Assumptions A1, A3 and A4 hold. Then, there exists a constant $c_1$, depending on the network topology and the Gaussian model statistics only, such that*

$$\mathbb{E}_1 \left[ \sup_{t\geq 0} \|S_{d,i}(t) - S_{d,j}(t)\| \right] \leq c_1$$

*for all agent pairs $(i, j)$.*

*Proof:* Denoting by $\mathbf{S}_d(t) = t\mathbf{P}_d(t)$ the vector of the agent test statistics $S_{d,i}(t)$'s, we have by (12),

$$\mathbf{S}_d(t+1) = W\left(\mathbf{S}_d(t) + \eta(t+1)\right). \tag{42}$$

Let $\overline{S}_d(t)$ denote the average of the $S_{d,i}(t)$'s, i.e.,

$$\overline{S}_d(t) = (1/N) \cdot (S_{d,1}(t) + \cdots + S_{d,N}(t)),$$

Noting that $J\mathbf{S}_d(t) = \overline{S}_d(t)\mathbf{1}$ and $WJ = JW = J$, we have from (42)

$$\mathbf{v}_{t+1} = (W - J)\mathbf{v}_t + \mathbf{u}_{t+1}, \tag{43}$$

where $\mathbf{v}_t$ and $\mathbf{u}_t$, for all $t \geq 0$, are given by

$$\mathbf{v}_t = \mathbf{S}_d(t) - \overline{S}_d(t)\mathbf{1}$$

and

$$\mathbf{u}_{t+1} = (W - J)\,\eta(t+1).$$

It is important to note that the sequence $\{\mathbf{u}_t\}$ is i.i.d. Gaussian and, in particular, there exists a constant $c_2$ such that $\mathbb{E}_1[\|\mathbf{u}_t\|^2] \leq c_2$ for all $t$.

Now, by (43) we obtain

$$\|\mathbf{v}_{t+1}\| \leq r\|\mathbf{v}_t\| + \|\mathbf{u}_{t+1}\|,$$

where recall $r = \|W - J\| < 1$. Since the sequence $\{\mathbf{u}_t\}$ is i.i.d. and $\mathcal{L}_2$-bounded, an application of the Robbins-Siegmund's lemma (see [46]) yields

$$\mathbb{E}_1\left[\sup_{t \geq 0} \|\mathbf{v}_t\|\right] \leq c_3 < \infty,$$

where $c_3$ is a constant that may be chosen as a function of $r$, $c_2$ and $\mathbb{E}_1[\|\mathbf{v}_0\|]$. Now, noting that, for any pair $(i, j)$,

$$\mathbb{E}_1\left[\sup_{t \geq 0} \|S_{d,i}(t) - S_{d,j}(t)\|\right] \leq \mathbb{E}_1\left[\sup_{t \geq 0} \|S_{d,i}(t) - \overline{S}_d(t)\|\right] + \mathbb{E}_1\left[\sup_{t \geq 0} \|S_{d,j}(t) - \overline{S}_d(t)\|\right] \leq 2c_3,$$

the desired assertion follows. ∎

*Proof of Theorem 4.7:* We prove the upper bound in Theorem 4.7 first. Since $\mathbb{P}_1(T_{d,i} < \infty) = 1$, for the upper bound we have,

$$\mathbb{E}_1[T_{d,i}] = \sum_{t=0}^{\infty} \mathbb{P}_1(T_{d,i} > t)$$

$$\overset{(a)}{\leq} \sum_{0}^{\infty} \mathbb{Q}\left(\frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right)$$

$$= \underbrace{\sum_{0}^{\lfloor \frac{\gamma_{d,i}^h}{m} \rfloor} \mathbb{Q}\left(\frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right)}_{(1)} + \underbrace{\sum_{\lfloor \frac{\gamma_{d,i}^h}{m} \rfloor + 1}^{\lfloor \frac{3\gamma_{d,i}^h}{2m} \rfloor} \mathbb{Q}\left(\frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right)}_{(2)}$$

$$+ \underbrace{\sum_{\lfloor \frac{3\gamma_{d,i}^h}{2m} \rfloor + 1}^{\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor} \mathbb{Q}\left(\frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right)}_{(3)} + \underbrace{\sum_{\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor + 1}^{\infty} \mathbb{Q}\left(\frac{-\gamma_{d,i}^h + mt}{\sqrt{\frac{2mt}{N} + \frac{2mr^2(1-r^{2t})}{1-r^2}}}\right)}_{(4)}$$

$$\overset{(b)}{\leq} \frac{\gamma_{d,i}^h}{m} + \frac{\gamma_{d,i}^h}{4m} + \frac{1}{2}e^{\frac{N\gamma_{d,i}^h}{2(k+1)}} \sum_{\lfloor \frac{3\gamma_{d,i}^h}{2m} \rfloor + 1}^{\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor} e^{\frac{-(N\gamma_{d,i}^h)^2 - Nm^2t^2}{4m(k+1)t}} + \frac{1}{2(1 - e^{\frac{-Nm}{4(k+1)}})}$$

$$\leq \frac{5\gamma_{d,i}^h}{4m} + \frac{1}{2(1 - e^{\frac{-Nm}{4(k+1)}})} + \frac{1}{2}e^{\frac{3N\gamma_{d,i}^h}{8(k+1)}} \sum_{\lfloor \frac{3\gamma_{d,i}^h}{2m} \rfloor + 1}^{\lfloor \frac{2\gamma_{d,i}^h}{m} \rfloor} e^{\frac{-Nmt}{4(k+1)}}$$

$$\leq \frac{5\gamma_{d,i}^h}{4m} + \frac{1}{1 - e^{\frac{-Nm}{4(k+1)}}},$$

where $(a)$ is due to the upper bound derived in Lemma 4.4 and $(b)$ is due to the following : 1) $\forall t \in [0, \lfloor \frac{\gamma_{d,i}^h}{m} \rfloor]$ in (1), $-\gamma_{d,i}^h + mt$ is negative and hence every term in the summation can be upper bounded by 1; 2) $\forall t \in [\lfloor \frac{\gamma_{d,i}^h}{m} \rfloor + 1, \lfloor \frac{3\gamma_{d,i}^h}{2m} \rfloor]$ in (2), $-\gamma_{d,i}^h + mt$ is positive and hence every term in the summation can be upper bounded by $\frac{1}{2}$; and 3) for the terms (3) and (4), the inequality $\mathbb{Q}(x) \leq \frac{1}{2}e^{-x^2/2}$ is used and the sums are upper bounded by summing the resulting geometric series.

In order to obtain the lower bound, we first note that conditioned on hypothesis $H_1$, at the stopping time $T_{d,i}$, an agent exceeds the threshold $\gamma_{d,i}^h$ with probability at least $1 - \epsilon$ and is lower than the threshold $\gamma_{d,i}^l$ with probability at most $\epsilon$. Moreover, with $\alpha = \beta = \epsilon$, $\gamma_{d,i}^h = -\gamma_{d,i}^l$.

Now, denote by $E_i^h$ the event $E_i^h = \{S_{d,i}(T_{d,i}) \geq \gamma_{d,i}^h\}$ and by $E_i^l$ the event $E_i^l = \{S_{d,i}(T_{d,i}) \leq \gamma_{d,i}^l\}$. Since $\mathbb{P}_1(T_{d,i} < \infty) = 1$, we have that

$$\mathbb{E}_1\left[S_{d,i}(t)\right] = \mathbb{E}_1\left[S_{d,i}(t).\mathbb{I}_{E_i^h}\right] + \mathbb{E}_1\left[S_{d,i}(t).\mathbb{I}_{E_i^l}\right], \tag{44}$$

where $\mathbb{I}_{\{.\}}$ denotes the indicator function. We now lower bound the quantities on the R.H.S. of (44). Note that $\gamma_{d,i}^h \geq 0$ and $S_{d,i}(t) \geq \gamma_{d,i}^h$ on $E_i^h$. Hence

$$\mathbb{E}_1\left[S_{d,i}(t).\mathbb{I}_{E_i^h}\right] \geq \gamma_{d,i}^h \mathbb{P}_1\left(E_i^h\right) \geq (1 - \epsilon)\gamma_{d,i}^h. \tag{45}$$

Now recall the construction in the proof of Lemma 7.2 and note that by (42) we have

$$S_{d,i}(t) = S_{d,i}(t - 1) - \sum_{j \in \Omega_i} w_{ij}\left(S_{d,i}(t - 1) - S_{d,j}(t - 1)\right) + \eta_i(t).$$

Hence, we have that

$$S_{d,i}(T_{d,i}).\mathbb{I}_{E_i^l} \tag{46}$$

$$\geq S_{d,i}(T_{d,i} - 1).\mathbb{I}_{E_i^l} - \sum_{j \in \Omega_i} w_{ij}\|S_{d,i}(T_{d,i} - 1) - S_{d,j}(T_{d,i} - 1)\| - \|\eta_i(T_{d,i})\|$$

$$\geq S_{d,i}(T_{d,i} - 1).\mathbb{I}_{E_i^l} - \sum_{j \in \Omega_i} w_{ij} \sup_{t \geq 0} \|S_{d,i}(t) - S_{d,j}(t)\| - \|\eta_i(T_{d,i})\|.$$

Now, observe that on the event $E_i^l$, $S_{d,i}(T_{d,i} - 1) > \gamma_{d,i}^l$ a.s. Since $\gamma_{d,i}^l < 0$ and $\mathbb{P}_1(E_i^l) \leq \epsilon$ (by hypothesis), we have that

$$\gamma_{d,i}^l \epsilon \leq \gamma_{d,i}^l \mathbb{P}_1(E_i^l) \tag{47}$$

$$= \mathbb{E}_1 \left[ \gamma_{d,i}^l . \mathbb{I}_{E_i^l} \right] \leq \mathbb{E}_1 \left[ S_{d,i}(T_{d,i} - 1) . \mathbb{I}_{E_i^l} \right].$$

Note that, by Lemma 7.2, we have

$$\mathbb{E}_1 \left[ \sum_{j \in \Omega_i} w_{ij} \sup_{t \geq 0} \| S_{d,i}(t) - S_{d,j}(t) \| \right]$$

$$\leq \sum_{j \in \Omega_i} w_{ij} \mathbb{E}_1 \left[ \sup_{t \geq 0} \| S_{d,i}(t) - S_{d,j}(t) \| \right]$$

$$\leq |\Omega_i| c_1.$$

Finally, by arguments similar to [47], [48] for characterizing expected overshoots in stopped random sums (see, in particular, Theorem 1 in [48]) it follows that there exists a constant $c_4$ (depending on the Gaussian model statistics and the network topology only) such that

$$\mathbb{E}_1 \left[ \| \eta_i(T_{d,i}) \| \right] \leq c_4. \tag{48}$$

In particular, note that, the constant $c_4$ in (48) may be chosen to be independent of the thresholds and, hence, the error tolerance parameter $\epsilon$. Substituting (47)-(48) in (46) we obtain

$$\mathbb{E}_1 \left[ S_{d,i}(T_{d,i}) . \mathbb{I}_{E_i^l} \right] \geq \gamma_{d,i}^l \epsilon - |\Omega_i| c_1 - c_4.$$

This together with (44)-(45) yield

$$\mathbb{E}_1 \left[ S_{d,i}(T_{d,i}) \right] \geq (1 - \epsilon) \gamma_{d,i}^h + \gamma_{d,i}^l \epsilon - |\Omega_i| c_1 - c_4 \tag{49}$$

$$= (1 - 2\epsilon) \gamma_{d,i}^h - c,$$

where the last equality follows by noting that $\gamma_{d,i}^h = -\gamma_{d,i}^l$ and taking the constant $c$ to $c = |\Omega_i| c_1 + c_4$.

We note that the event $\{T_{d,i} = t\}$ is independent of $\eta_i, i > t$. We also have from Theorem 4.1 that

$\mathbb{P}_1(T_{d,i} < \infty) = 1$. Hence, we have,

$$\mathbb{E}_1[S_{d,i}(T_{d,i})] = \mathbb{E}_1[\sum_{j=1}^{T_{d,i}} \mathbf{e_i}^\top \mathbf{W}^{t+1-j} \eta(j)]$$

$$= \mathbb{E}_1 \left[ \sum_{j=1}^{\infty} \mathbb{I}_{\{T_{d,i} \geq j\}} \mathbf{e_i}^\top \mathbf{W}^{T_{d,i}+1-j} \eta(j) \right]$$

$$= \sum_{j=1}^{\infty} \mathbb{E}_1 \left[ \mathbb{I}_{\{T_{d,i} \geq j\}} \mathbf{e_i}^\top \mathbf{W}^{T_{d,i}+1-j} \right] \mathbb{E}_1 [\eta(j)]$$

$$= m \sum_{j=1}^{\infty} \mathbb{E}_1 \left[ \mathbb{I}_{\{T_{d,i} \geq j\}} \mathbf{e_i}^\top \mathbf{W}^{T_{d,i}+1-j} \right] \mathbf{1}$$

$$= m \sum_{j=1}^{\infty} \mathbb{E}_1 \left[ \mathbb{I}_{\{T_{d,i} \geq j\}} \mathbf{e_i}^\top \mathbf{W}^{T_{d,i}+1-j} \mathbf{1} \right]$$

$$= m \mathbb{E}_1 [T_{d,i}]. \tag{50}$$

Combining (50) and (49) we have,

$$\frac{(1-2\epsilon)\gamma_{d,i}^h}{m} - \frac{c}{m} \leq \mathbb{E}_1[T_{d,i}]$$

and the desired assertion follows.

∎

*Proof of Theorem 4.8:* From (23), we first note that

$$\frac{\mathbb{E}_1[T_{d,i}]}{\mathbb{E}_1[T_c]} \geq 1, \forall i = 1, 2, \ldots, N. \tag{51}$$

From the upper bound for the stopping time distribution derived for the $\mathcal{CISPRT}$ in (35), we have the following upper bound for $\mathbb{E}_1[T_{d,i}]$

$$\mathbb{E}_1[T_{d,i}] \leq \frac{5\gamma_{d,i}^h}{4m} + \frac{1}{1 - e^{\frac{-Nm}{4(k+1)}}}. \tag{52}$$

We choose the threshold $\gamma_{d,i}^h$ to be

$$\gamma_{d,i}^h = \gamma_d^{h,0} = \frac{8(k+1)}{7N}(\log(\frac{2}{\epsilon}) - \log(1 - e^{\frac{-Nm}{4(k+1)}})). \tag{53}$$

Using (52) and (53), we have

$$\limsup_{\epsilon \to 0} \frac{\mathbb{E}_1[T_{d,i}]}{\mathbb{E}_1[T_c]} \leq \lim_{\epsilon \to 0} \frac{\frac{10}{7}(k+1)\log(\frac{2}{\epsilon}) + O(1)}{(1-2\epsilon)\log(\frac{1-\epsilon}{\epsilon})}.$$

Noting that,

$$\limsup_{\epsilon \to 0} \frac{O(1)}{(1-2\epsilon)\log(\frac{1-\epsilon}{\epsilon})} = 0,$$

we obtain

$$\limsup_{\epsilon \to 0} \frac{\mathbb{E}_1[T_{d,i}]}{\mathbb{E}_1[T_c]} \leq \frac{10(k+1)}{7}. \tag{54}$$

Combining (54) and (51), the result follows. ∎

## 8. CONCLUSION

In this paper we have considered sequential detection of Gaussian binary hypothesis observed by a sparsely interconnected network of agents. The $\mathcal{CISPRT}$ algorithm we proposed combines two terms : a *consensus* term that updates at each sensor its test statistic with the test statistics provided by agents in its one-hop neighborhood and an *innovation* term that updates the current agent test statistic with the new local sensed information. We have shown that the $\mathcal{CISPRT}$ can be designed to achieve a.s. finite stopping at each network agent with guaranteed error performance. We have provided explicit characterization of the large deviation decay exponents of tail probabilities of the $\mathcal{CISPRT}$ stopping and its expected stopping time as a function of the network connectivity. The performance of the $\mathcal{CISPRT}$ was further benchmarked w.r.t. the optimal centralized sequential detector, the SPRT. The techniques developed in this paper are of independent interest and we envision their applicability to other distributed sequential procedures. An interesting future direction would be to consider networks with random time-varying topology. We also intend to develop extensions of the $\mathcal{CISPRT}$ for setups with correlated and non-linear non-Gaussian observation models.

## 9. ACKNOWLEDGMENT

## REFERENCES

[1] A. Wald, "Sequential tests of statistical hypotheses," *Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 1945.

[2] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Sequential detection of cyclostationary signal for cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4480–4485, 2009.

[3] A. Jayaprakasam and V. Sharma, "Sequential detection based cooperative spectrum sensing algorithms in cognitive radio," in *2009 First UK-India International Workshop on Cognitive Wireless Systems (UKIWCWS).* IEEE, 2009, pp. 1–6.

[4] S. Chaudhari, V. Koivunen, and H. V. Poor, "Autocorrelation-based decentralized sequential detection of ofdm signals in cognitive radios," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2690–2700, 2009.

[5] S. D. Blostein and H. S. Richardson, "A sequential detection approach to target tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 1, pp. 197–212, 1994.

[6] S. Kar and J. M. Moura, "Distributed linear parameter estimation in sensor networks: Convergence properties," in *42nd Asilomar Conference on Signals, Systems and Computers 2008*. IEEE, 2008, pp. 1347–1351.

[7] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575 – 3605, June 2012.

[8] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, 2003.

[9] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, no. 2, pp. 297–344, 1993.

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

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

[12] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, Jun. 2003.

[13] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, January 2007.

[14] A. G. Dimakis, S. Kar, J. M. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.

[15] S. Kar, S. Aldosari, and J. M. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, p. 2609, 2008.

[16] S. Kar and J. M. Moura, "Consensus based detection in sensor networks: Topology optimization under practical constraints," *Proc. 1st Intl. Wrkshp. Inform. Theory Sensor Networks*, 2007.

[17] J. Zhang and R. S. Blum, "Asymptotically optimal truncated multivariate gaussian hypothesis testing with application to consensus algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 431–442, 2014.

[18] Q. Zhou, S. Kar, L. Huie, H. V. Poor, and S. Cui, "Robust distributed least-squares estimation in sensor networks with node failures," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*. IEEE, 2011, pp. 1–6.

[19] Q. Zhou, S. Kar, L. Huie, and S. Cui, "Distributed estimation in sensor networks with imperfect model information: an adaptive learning-based approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012*. IEEE, 2012, pp. 3109–3112.

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

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

[22] F. Cattivelli and A. H. Sayed, "Diffusion lms-based distributed detection over adaptive networks," in *Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, 2009*. IEEE, 2009, pp. 171–175.

[23] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks based on diffusion estimation schemes," in *IEEE 10th Workshop on Signal Processing Advances in Wireless Communications, 2009. SPAWC'09*. IEEE, 2009, pp. 61–65.

[24] ——, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1917–1932, 2011.

[25] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. Moura, "Distributed detection via gaussian running consensus: Large deviations asymptotic analysis," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4381–4396, 2011.

[26] D. Jakovetic, J. M. Moura, and J. Xavier, "Distributed detection over noisy networks: Large deviations analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4306–4320, 2012.

[27] S. Kar, R. Tandon, H. V. Poor, and S. Cui, "Distributed detection in noisy sensor networks," in *IEEE International Symposium on Information Theory Proceedings (ISIT), 2011*. IEEE, 2011, pp. 2856–2860.

[28] V. Srivastava and N. E. Leonard, "Collective decision-making in ideal networks: The speed-accuracy tradeoff," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 121–132, 2014.

[29] H. Hashemi and I. B. Rhodes, "Decentralized sequential detection," *IEEE Transactions on Information Theory*, vol. 35, no. 3, pp. 509–520, 1989.

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

[31] A. Nayyar and D. Teneketzis, "Sequential problems in decentralized detection with communication," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5410–5435, 2011.

[32] V. V. Veeravalli, T. Başar, and H. V. Poor, "Decentralized sequential detection with sensors performing sequential tests," *Mathematics of Control, Signals and Systems*, vol. 7, no. 4, pp. 292–305, 1994.

[33] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 534–548, 2011.

[34] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press Cambridge, 2009, vol. 40.

[35] A. G. Tartakovsky and V. V. Veeravalli, "Change-point detection in multichannel and distributed systems," *Applied Sequential Methodologies: Real-World Examples with Data Analysis*, vol. 173, pp. 339–370, 2004.

[36] A. Tartakovsky and X. Rong Li, "Sequential testing of multiple hypotheses in distributed systems," in *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, vol. 2. IEEE, 2000, pp. THC4–17.

[37] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.

[38] H. Chernoff, *Sequential analysis and optimal design*. Siam, 1972, vol. 8.

[39] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*. Springer-Verlag Berlin, 1987, vol. 288.

[40] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326–339, 1948.

[41] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[42] S. Kar and J. M. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3315–3326, 2008.

[43] ——, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, 2009.

[44] D. A. Darling and A. Siegert, "The first passage problem for a continuous markov process," *The Annals of Mathematical Statistics*, pp. 624–639, 1953.

[45] P. Hieber and M. Scherer, "A note on first-passage times of continuously time-changed brownian motion," *Statistics & Probability Letters*, vol. 82, no. 1, pp. 165–172, 2012.

[46] P. Baldi, L. Mazliak, and P. Priouret, *Martingales and Markov chains: solved exercises and elements of theory*. CRC Press, 2002.

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

[48] G. Lorden, "On excess over the boundary," *The Annals of Mathematical Statistics*, pp. 520–527, 1970.