Improved Performance Properties of the CISPRT Algorithm for Distributed Sequential Detection

Kun Liu and Yajun Mei

Abstract

In distributed sequential detection problems, local sensors observe raw local observations over time, and are allowed to communicate local information with their immediate neighborhood at each time step so that the sensors can work together to make a quick but accurate decision when testing binary hypotheses on the true raw sensor distributions. One interesting algorithm is the Consensus-Innovation Sequential Probability Ratio Test (CISPRT) algorithm proposed by Sahu and Kar (IEEE Trans. Signal Process., 2016), but unfortunately its performance properties are not well studied. In this article, we present improved, non-asymptotic properties of the CISPRT algorithm for Gaussian data in term of network connectivity no matter how large the number of sensors is. Moreover, we also provide sharp upper bound on the information loss of the CISPRT algorithm as compared to the centralized optimal SPRT algorithm in the asymptotic regime when Type I and II error probabilities go to 0. Numerical simulations suggest that our results are useful under the practical setting when the number of sensors is moderately large.

Index Terms

CISPRT, distributed learning, network connectivity, sequential detection, oracle properties.

I. INTRODUCTION

Distributed online learning becomes increasingly important in many real-world applications such as cognitive radio networks (Liang, Lai and Halloran [1] [2]), social recommender systems (Tekin, Zhang and van der Schaar [3], Yan et al. [4]), natural language processing (Gimpel, Das and Smith [5]). Under a general setting, there are $N$ sensors or agents taking raw observations over time in a system, and each local sensor can only communicate

Manuscript received September 13, 2017.
K. Liu and Y. Mei are with The H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA (email: \{kliu80, ymei3\}@gatech.edu). This research is partially supported by NSF grant CMMI-1362876.
its local information with the immediate neighbors at each time. Such local information communication can be conducted adaptively or sequentially over time so that sensors can work together to reach consensus quickly. The advantages of distributed settings are to protect intrinsic privacy of sensitive data (Yan et al. [4]), increase computational capacity (Rabbat and Nowak [6], Galstyan et al. [7], Zhang et al. [8]), and mitigate collection and storage burden of modern large datasets (Rabbat and Nowak [9], Lorenzo and Barbarossa [10]).

There are many important distributed online learning problems in engineering and statistics, and one of them is the distributed sequential detection, see Blum, Kassam and Poor [11], Viswanathan and Varshney [12], Zhou et al. [13], where the distributed sensors work together to quickly and correctly decide which is the true underlying probability measures or models for raw sensor observations. Had the local sensors been able to send local information to a central location, often called the fusion center, for further analysis, extensive research has been done along two distinct directions. The first one is when the fusion center has access to all raw sensor observations, which is the centralized sequential detection problem. This is well studied in the classical subfield of sequential analysis in statistics (Basseville and Nikiforov [14], Siegmund [15], Tartakovsky, Nikiforov and Basseville [16], Wald [17]). In particular, the optimal centralized procedure is the well-known Sequential Probability Ratio Test (SPRT), see Wald [17]. The other direction is the decentralized sequential detection, where the local sensors send quantized sensor messages to the fusion center to make a global decision, see Fellouris [18], Hashemi and Rhodes [19], Mei [20], Veeravalli, Basar and Poor [21].

Unfortunately, in the distributed sequential detection problem, there is no fusion center, and the local sensors need to work together to make a decision. Research is rather limited in the literature: very few efficient algorithms have been proposed partly because it involves complicated communication strategies between local sensors and their neighborhood sensors. One exception is the Consensus+Innovation SPRT (CISPRT) algorithm developed in Sahu and Kar [22] that is based on the weighted average of local log-likelihood ratio tests, see Kar and Moura [23] [24], Kar, Aldosari and Moura [25] for the motivation and more background. The CISPRT algorithm is novel and interesting, as each local sensor utilizes local information not only from itself and its immediate neighbor sensors, but also from remote connected sensors that are 2-hop or more hops away from itself. Also see Li and Wang [26] for an interesting generalization of the CISPRT algorithm under the fixed $q$-round message passing protocol.

Intuitively, the performance of a distributed algorithm including the CISPRT will depend on the neighborhood structure of local sensors, or the network connectivity. For any pre-specified neighborhood structure of local sensors, Sahu and Kar [22] provided the lower and upper bounds on the performance properties of the CISPRT for Gaussian data, and also derived an upper bound on its information loss as compared to the optimal centralized SPRT. However, extensive numerical studies demonstrated that while the lower bounds are reasonable, the upper bounds are too
loose. Indeed, in the special case when each local sensor is connected to all other sensors, each local sensor can be treated as the fusion center, and the CISPRT becomes the centralized SPRT. However, the upper bounds in Sahu and Kar [22] is much larger than the classical well-known results for the centralized SPRT, especially with a large number of local sensors. This led us to raise an open problem whether one can derive sharper upper bounds that are meaningful regardless of the number of sensors under any reasonable practical setting.

The main objective of this article is to provide a positive answer to this open problem by deriving refined upper bounds on the performance properties of the CISPRT algorithm for Gaussian data in term of the network connectivity regardless of the number of sensors. Note that it is mathematically challenging to investigate the properties of any sequential detection procedures in the scenario of large number $N$ of sensors: Sahu and Kar [22] provided a novel approach to analyze the oracle properties of the CISPRT for Gaussian data for any number $N$ of sensors, but unfortunately their results are too crude. Here we propose to further polish their approach and improve their theoretical properties of the CISPRT algorithm for Gaussian data regardless of how large the number $N$ of sensors is.

Our main contributions are several-fold. First, the classical techniques for performance analysis in sequential hypothesis testing are renewal theory and overshoot analysis, but unfortunately they are not suitable for high-dimensional data or for a large number $N$ of local sensors, since the overshoot constant often exponentially increase as the function of the number $N$ of local sensors. Our approaches provide alternative tools through a careful tail probability analysis of Gaussian distributions that can explicitly characterize the effects of the dimensionality or the number $N$ of sensors on the performance of the sequential tests for Gaussian data. Second, we derive oracle upper bounds on the performance properties of the CISPRT algorithm for Gaussian data under any pre-specified neighborhood structure of local sensors regardless of the number $N$ of sensors. Our upper bounds are non-asymptotic and much sharper than those in Sahu and Kar [22], as our approaches work on a more detailed analysis on the infinite sum of tail probabilities. Finally, in the asymptotic regime when Type I and II error probabilities go to 0, we provide a tight upper bound on the information loss of the CISPRT as compared to the centralized optimal SPRT algorithm. Our results show that the more the number of sensor or the sparser the network neighborhood connectivity is, the larger the information loss is.

The remainder of this article is organized as follows. In Section II, we discuss the formulation of the distributed sequential detection problem as well as the CISPRT algorithm and its properties derived in Sahu and Kar [22]. In Section III, we present our main theoretical results on the refined oracle properties of the CISPRT. Simulation studies results are presented in Section IV, and the detailed proof of our main theorem, Theorem III.1, is provided in Section V. Finally, some conclusion remarks are included in Section VI.
II. Preliminaries and Background

A. Distributed Sequential Detection Problems

Consider a network system of $N$ sensors that take observations over time. At each time step $t = 1, 2, \ldots$, the $i$-th sensor observes an observation $y_i(t)$ for $i = 1, \ldots, N$. There are two hypotheses on the local sensor observations $y_i(t)$’s. Under the null hypothesis $H_0$, the $y_i(t)$’s are $N(-\mu, \sigma^2)$ and under the alternative hypothesis $H_1$, the $y_i(t)$’s are $N(\mu, \sigma^2)$. Here the sensor observations $y_i(t)$’s are assumed to be independent and identically distributed (iid) over time and across sensors, conditional on each hypothesis.

Under the distributed sequential detection setting, the objective is for each local sensor to work with its neighborhood sensors to make a quick and accurate decision on which of these two hypotheses is true. In particular, each local sensor can only communicate its local information with its (one-hop) neighborhood sensors. Here we assume that the neighborhood structure of sensors are pre-specified, and can be represented as an undirected graph $G = (V, E)$: the $i$-th vertex in $V$ represents the $i$-th sensor, and there is an edge between the $i$-th vertex(sensor) and the $j$-th vertex(sensor), i.e., $(i, j) \in E$, if and only if the corresponding sensors are neighbors and can communicate local information with each other. Here we assume that the graph $G = (V, E)$ is simple, i.e., without self loops and multiple edges. For the $i$-th sensor, its neighborhood is given by $\Omega_i = \{j \in V | (i, j) \in E\}$, and its degree is given by the cardinality $d_i = |\Omega_i|$. See Mesbahi and Egerstedt [27] for more graph theoretic methods in network systems.

For a distributed sequential procedure $D_d$, it consists of $(T_i, \delta_i)_{i=1}^N$, where $T_i$ is the number of time steps the $i$-th local sensor needed to make a local decision $\delta_i \in \{0, 1\}$. Here $T_i$ is a local stopping time in the sense that $\{T_i = t\}$ depends on the information from the $i$-th local sensor as well as its neighborhood up to time $t$. The local decision $\delta_i = 0$ or 1 means that the $i$-th local senors accepts $H_0$ or $H_1$, respectively.

The performance of a distributed sequential procedure $D_d = (T_i, \delta_i)_{i=1}^N$ is evaluated by its local expected sample sizes, $E_i[T_i]$ and $E_0[T_i]$’s, and its local error probabilities, $P_0[\delta_i = 1]$ and $P_1[\delta_i = 0]$. Ideally one would like all these four local quantities are simultaneously as small as possible for all local sensors, which is impossible. As mentioned in Sahu and Kar [22], one useful formulation is to find a distributed sequential procedure $D_d = (T_i, \delta_i)_{i=1}^N$ that (asymptotically) minimizes

$$\max_{i=1,2,\ldots,N} E_1[T_i]$$

subject to the local false alarm constraints:

$$P_0[\delta_i = 1] \leq \alpha \text{ and } P_1[\delta_i = 0] \leq \beta$$
for all $i = 1, 2, \cdots, N$, where $0 < \alpha, \beta \leq 1/2$ are the pre-specified false alarm bounds. Instead of $E_1[T_i]$ in (1), one might also be interested in minimizing $E_0[T_i]$, or more generally, the Bayesian-type criterion $\pi_0 E_0[T_i] + (1 - \pi_0) E_1[T_i]$. Fortunately, in the context of sequential tests, Wald’s (optimal centralized) SPRT can minimize each and every of these criteria, and thus we consider the criterion of $E_1[T_i]$ here.

Note that the global objective function in (1) can also be replaced by
\[
\min_{i=1,2,\cdots,N} E_1[T_i]
\]
or other functions if one wants. Here we do not discuss the appropriateness of different formulations or the corresponding optimality theories, and our focus is to investigate the performance of a specific distributed sequential procedure. For that reason, our results below deal with the local expected sample sizes $E_1[T_i]$’s themselves, since it is straightforward to extend these local results to the global level such as that in (1).

B. SPRT and CISPRT

Let us first consider the centralized setup when the graph of the network neighborhood structure is complete in the sense that at each time step each local sensor has access to all raw observations over the graph. This is equivalent to the scenario with the fusion center, as each and every local sensor can be regarded as the fusion center. In such scenario, Wald’s SPRT is the optimal centralized sequential test under the formulation of (1) and (2). To define the SPRT, denote the local log-likelihood ratio of the $i$-th sensor at time step $t$ by
\[
\eta_i(t) = \log \frac{f_1(y_i(t))}{f_0(y_i(t))} = \frac{2\mu}{\sigma^2} y_i(t),
\]
and denote the centralized likelihood ratio statistic up to time $t$ by
\[
S_c(t) = \sum_{s=1}^{t} \sum_{i=1}^{N} \eta_i(s) = S_c(t-1) + \sum_{i=1}^{N} \eta_i(t)
\]
for all $t \geq 1$. The centralized SPRT is then defined by the stopping time
\[
T_c = \inf \{ t : S_c(t) \not\in [\gamma_c^l, \gamma_c^h] \},
\]
where $\gamma_c^l$ and $\gamma_c^h$ are two pre-specified constants so as to satisfy the false alarm constraints in (2). In particular, a common used though slightly conservative choice is
\[
\gamma_c^l = \log \frac{1 - \beta}{\alpha} \quad \text{and} \quad \gamma_c^h = \log \frac{1 - \alpha}{\beta}.
\]
Moreover, since \( y_i(t) \sim N(-\mu, \sigma^2) \) or \( N(\mu, \sigma^2) \) under \( H_0 \) or \( H_1 \), the Kullback-Leibler divergence at each local sensor is

\[
m = E_1(\eta_i(t)) = \frac{2\mu^2}{\sigma^2}
\]

and thus the centralized Kullback-Leibler divergence of the joint observation \( Y(t) = (Y_1(t), \ldots, Y_N(t)) \) is \( Nm \).

Furthermore, as shown in Wald [17], subject to the false alarm constraint in (2), for any sequential test \( T \), distributed or centralized,

\[
E_1(T) \geq \mathcal{M}(\alpha, \beta),
\]

where the universal lower bound is given by

\[
\mathcal{M}(\alpha, \beta) = \frac{1}{Nm} \left[ (1 - \beta) \log \frac{1 - \beta}{\alpha} + \beta \log \frac{\beta}{1 - \alpha} \right].
\]

Also the centralized SPRT \( T_c \) in (5) attains this lower bound asymptotically for fixed \( N \) and \( m \) as \( \alpha, \beta \rightarrow 0 \).

Now let us switch to the distributed setup for a general neighborhood structure where each local sensor can only communicate with its neighborhood sensors. In Sahu and Kar [22], the authors proposed an interesting CISPRT algorithm where each local sensor makes a local decision based on the weighted average of the local likelihood ratio statistics from itself and its neighborhood sensors. Specifically, at time step \( t \), each \( i \)-th local sensor computes its local test statistics recursively:

\[
S_i(t) = w_{ii}S_i(t-1) + \sum_{j \in \Omega_i} w_{ij}S_j(t-1) + w_{ii}\eta_i(t) + \sum_{j \in \Omega_i} w_{ij}\eta_j(t),
\]

for \( t = 1, 2, \ldots \), where the initial value \( S_i(0) = 0 \) and \( \Omega_i \) is the (one-hop) neighborhood of the \( i \)-th sensor. Here the \( w_{ij} \)'s are pre-specified weights satisfying

\[
w_{ij} \geq 0, \ w_{ii} + \sum_{j \in \Omega_i} w_{ij} = 1, \ \forall i, j
\]

and the discussion on the choices of the weights \( w_{ij} \)'s will be postponed a little bit.

Under the matrix notation, let us collect the weights \( w_{ij} \)'s in an \( N \times N \) matrix \( W \), where \( w_{ij} = 0 \) if \( (i,j) \notin E \). Denote by \( S(t) \) and \( \eta(t) \) as the \( N \times 1 \) vectors \( (S_1(t), \ldots, S_N(t))^T \) and \( (\eta_1(t), \ldots, \eta_N(t))^T \). The local test statistics can be updated recursively as

\[
S(t) = W \left(S(t-1) + \eta(t)\right).
\]
for $t \geq 1$.

For the CISPRT, each $i$-th sensor makes a local decision at time

$$T_i = \inf \{ t \geq 1 : S_i(t) \notin [\gamma_i^l, \gamma_i^h] \} ,$$

for some pre-specific thresholds $\gamma_i^l < 0 < \gamma_i^h$. When stopping, the $i$-th sensor makes a local decision

$$\delta_i = \begin{cases} 
0, & \text{if } S_i(T_i) \leq \gamma_i^l; \\
1, & \text{if } S_i(T_i) \geq \gamma_i^h.
\end{cases} \quad (13)$$

From the pure mathematical viewpoint, the stopping times $T_i$’s in (12) depend on the properties of $S_i(t)$’s which are a component of the $N$-dimensional random walks $S(t)$’s. One may be able to apply the classical renewal theory to analyze the “asymptotic” properties of the stopping times $T_i$’s in (12), but unfortunately such an approach will involve “constant” terms for overshoot analysis that are exponentially increasing as the dimension $N$ increases. In particular, when the number $N$ of sensors is large, such constant terms can be huge, and thus the corresponding asymptotic analysis can be meaningless under any reasonable practical setting of distributed detection. Here we provide an alternative approach that yields the same first-order asymptotic result as in the classical renewal theory for one or low dimension $N$, but has a potential to derive useful oracle properties of stopping times under the setting of high dimension $N$ for Gaussian data.

C. Spectral Graph Theory and Weight Matrix $W$

In this subsection, let us present some basic materials for spectral graph theory that are related to the main assumption for our network structure and the design of the weight matrix $W$ in (10). Also see Mesbahi and Egerstedt [27] for a more complete introduction of graph theory.

Recall our network neighbor structure is characterized by the undirected simple graph $G = (V, E)$. In spectral graph theory, the degree matrix $D$ is a $N \times N$ diagonal matrix with the $i$-th diagonal being $d_i$, the degree of the $i$-th vertex. The adjacency matrix $A$ is a $(0, 1)$-matrix with zeros on its diagonal and $A_{ij} = 1$ if and only if the $i$-th vertex and the $j$-th vertex are connected for all $1 \leq i \neq j \leq N$. The Laplacian matrix is then defined as

$$L = D - A . \quad (14)$$
Alternatively, for the Laplacian matrix $L$, its elements are given by

$$L_{i,j} = \begin{cases} 
  d_i & \text{if } i = j; \\
  -1 & \text{if } i\text{-th and } j\text{-th vertex connected for } i \neq j; \\
  0 & \text{otherwise.}
\end{cases}$$

The Laplacian matrix $L$ is a positive semidefinite matrix, and thus has $N$ non-negative eigenvalues:

$$0 = \lambda_1(L) \leq \lambda_2(L) \leq \cdots \leq \lambda_N(L).$$

Moreover, the number of times 0 appears as an eigenvalue of the Laplacian matrix $L$ is the number of connected components in the graph. Equivalently, a graph is connected if and only if $\lambda_2(L) > 0$, see Chung [28].

Our main assumption on the network neighborhood structure is as follows.

**Assumption II.1.** The graph $G = (V, E)$ is connected, or equivalently, the second smallest eigenvalue of the Laplacian matrix $L$ is positive, i.e., $\lambda_2(L) > 0$.

Next, let us discuss the choices of the weight matrix $W$ in (11). From the technical or algorithm viewpoint, the weight matrix $W$ can be arbitrary as long as $W$ is a stochastic matrix in the sense of satisfying (10). However, in the context of distributed sequential detection, the implicit assumption is that $w_{ij} > 0$ if and only if the $i$-th and $j$-th sensors are neighbors. There are still many reasonable choices for the weight matrix $W$, and one useful one is to define

$$W = I_{N \times N} - \delta L,$$  

(16)

where $I_{N \times N}$ is the $N \times N$ identity matrix and $\delta$ is a small positive constant so that all elements of $W$ are positive and thus (10) holds. Under this choice of the weight matrix $W$, for a given $i$-th sensor/vertex, it assigns a small but equal weight of $w_{ij} = \delta$ to all of its $d_i$ neighbor sensors, and assigns most weight $w_{ii} = 1 - \delta d_i$ to itself.

An interesting fact of the weight matrix $W$ in (16) is that it is symmetric ($w_{ij} = w_{ji}$) and irreducible, and the latter is due to the fact that the graph is connected under Assumption II.1. In addition, it is straightforward to show that as a stochastic matrix satisfying (10), the matrix $W$ has the largest eigenvalue 1, and the second largest eigenvalue, denote by $r$, is strictly less than 1. Moreover, the second largest eigenvalue $r$ can be characterized by

$$r = ||W - J||,$$  

(17)

where $J = \frac{1}{N} 11^T$ is an $N \times N$ matrix with all entries being the constant $1/N$, and vector $1$ is a all one $N$
dimensional vector.

It turns out that the second largest eigenvalue of \( W \), or \( r \) in (17), will play an essential role in our theoretical analysis. Among all weight matrices \( W \) of the form in (16), it is shown in Xiao and Boyd [29] that the second largest eigenvalue \( r \) of \( W \) is minimized at

\[
\delta_{opt} = 2/(\lambda_N(L) + \lambda_2(L))
\]

with the minimum value

\[
r_{opt} = \frac{\lambda_N(L) - \lambda_2(L)}{\lambda_N(L) + \lambda_2(L)},
\]

where \( \lambda_N(L) \) and \( \lambda_2(L) \) are the largest and second smallest eigenvalues in (15) for the Laplacian matrix \( L \).

III. IMPROVED PROPERTIES OF CISPRT

In this section, we derive our main theoretical properties of the CISPRT procedure in (12) and (13) under Assumption II.1 when the network is connected. Note that there are two thresholds, \( \gamma_l \) and \( \gamma_h \), in the CISPRT procedure in (12). At the high level, the upper bound \( \gamma_h \) is closely related to Type I error probability and the expected sample size under \( H_1 \), whereas the lower bound \( \gamma_l \) is closely related to Type II error probability and the expected sample size under \( H_0 \). For simplicity, below we will focus on the effects of the upper bound \( \gamma_h \) on the Type I error probability and the expected sample size under \( H_1 \) of the CISPRT, as the effects of the lower bound \( \gamma_l \) are similar.

Let us first summarize the main theoretical results of Sahu and Kar [22]. For a given weight matrix \( W \), denote by

\[
\rho = 1 - \exp \left( -\frac{N m}{4(Nr^2 + 1)} \right),
\]

where \( m \) is the Kullback-Leibler divergence in (7) and \( r \) is the second largest eigenvalue of \( W \) and can be rewritten as in (17). Sahu and Kar [22] showed that the Type I error of the CISPRT is given by

\[
P_0(\delta_i = 1) \leq 2\rho^{-1} \exp \left( -\frac{7}{8} \frac{N}{Nr^2 + 1} \gamma_h \right),
\]

and the expected sample size of the CISPRT satisfies

\[
\frac{1 - 2\epsilon}{m} \gamma_h - \frac{\text{constant}}{m} \leq E(T_i) \leq \frac{5}{4} \frac{1}{m} \gamma_h + \rho^{-1} + 1
\]
if $\alpha = \beta = \epsilon < 1/2$. Note that the original upper bound in Sahu and Kar [22] does not have the constant 1 in (22), but we found out that their proof contains a minor mistake to count the number of integers in the interval $0 \leq t \leq a$ as $a$, not $a + 1$. Thus we add 1 here so that the results in (22) are mathematically rigorous. Moreover, in the asymptotic setting when $\alpha = \beta = \epsilon \to 0$, Sahu and Kar [22] compares the expected stopping times of the CISPRT with the universal lower bound:

$$1 \leq \limsup_{\epsilon \to 0} \frac{E_1(T_i)}{M(\epsilon)} \leq \frac{10}{7} (N r^2 + 1)$$  \hspace{1cm} (23)

for all $i = 1, \cdots, N$. Here $M(\epsilon)$ is the universal lower bound in (8) when $\alpha = \beta = \epsilon$, i.e., $M(\epsilon) = M(\alpha = \epsilon, \beta = \epsilon)$.

Unfortunately, the upper bound results in (21)-(23) are too loose. In particular, extensive numerical simulations in Sahu and Kar [22] demonstrate that the Monte Carlo simulation of $E_1(T_i)$ is very close to $\frac{1}{m} \gamma_i^h$ for small $\epsilon$, i.e., while the lower bound in (22) is reasonable, the upper bound in (22) is too loose. This also raised another open theoretical question whether the factor $10/7$ in the right-hand side of (23) is loose or not. In this paper, we improve the upper bounds in (21)-(23) in the sense that the main factors $7/8, 5/4$ and $10/7$ will be reduced to 1, respectively. The price we paid is to add extra terms that can be thought of as second-order terms.

Our main results can be summarized in the following two theorems under two different setups: one is non-asymptotic and the other is asymptotic.

**Theorem III.1.** For the CISPRT, at any given $i$-th local sensor, the Type I error probability satisfies

$$P_0(\delta_i = 1) \leq \rho^{-1} \exp \left\{ - \frac{N}{N r^2 + 1} \gamma_i^h + \log \left( 1 + \frac{N}{4(N r^2 + 1)} \gamma_i^h \right) + 1 \right\},$$  \hspace{1cm} (24)

and its expected sample size under $H_1$ satisfies

$$E_1(T_i) \leq \frac{1}{m} \gamma_i^h + \frac{1}{2} \rho^{-1/2} \left( \sqrt{\frac{1}{m} \gamma_i^h + 1} \right) + 1.$$  \hspace{1cm} (25)

Here $\rho$ is a constant in (20), and $m$ is the Kullback-Leibler divergence in (7).

**Theorem III.2.** In the asymptotic setting when $\alpha = \beta = \epsilon \to 0$, if the CISPRT satisfies the local false alarm constraints in (2), then we have

$$1 \leq \limsup_{\epsilon \to 0} \frac{E_1(T_i)}{M(\epsilon)} \leq N r^2 + 1.$$  \hspace{1cm} (26)

for all $i = 1, 2, \cdots, N$. Moreover, the upper bound (26) is attained when the graph is complete and the weight matrix $W$ is defined in (16) with $\delta = \delta_{opt}$ in (18).
Note that Theorem III.1 deals with the non-asymptotic properties of the CISPRT, and holds regardless of how large the number \(N\) of sensors is. Its proof is very technical and will be postponed to Section V. Meanwhile, Theorem III.2 is to bound the asymptotic information loss of the CISPRT as compared to the centralized optimal SPRT algorithm in the asymptotic regime when Type I and II error probabilities go to 0. By Theorem III.2, the factor \(10/7\) in (23) can only be reduced to the smallest possible value 1, and cannot be reduced further. In that sense, our result in (26) is sharp under the general setting of network neighborhood structure. Below we will prove Theorem III.2.

**Proof of Theorem III.2:** The lower bound in (26) follows directly from the fact that the universal lower bound \(\mathcal{M}(\epsilon) = \mathcal{M}(\alpha = \epsilon, \beta = \epsilon)\) holds for any stopping time \(T\) satisfying the false alarm constraints, no matter whether \(T\) is decentralized or centralized. In addition, by (8), as \(\alpha = \beta = \epsilon \to 0\),

\[
\mathcal{M}(\epsilon) \sim \frac{1}{Nm} \log \epsilon^{-1}.
\]  

(27)

Here and below we denote by \(x(\epsilon) \sim y(\epsilon)\) if and only if \(\lim_{\epsilon \to 0} (x(\epsilon)/y(\epsilon)) = 1\).

To prove the upper bound in (26), in order for the CISPRT to satisfy the Type I error constraints in (2) with \(\alpha = \epsilon \to 0\), by (24), a conservative choice of \(\gamma_i^h\) will be

\[
\gamma_i^h \sim \frac{Nr^2}{N} + 1 \log \epsilon^{-1},
\]  

(28)

as \(\epsilon \to 0\), for any given \(N, \rho,\) and \(r\). Plugging this into (25), and focusing on the first-order term, we have

\[
E_1(T_i) \leq (1 + o(1)) \frac{Nr^2}{Nm} + \frac{1}{Nm} \log \epsilon^{-1}
\]  

(29)

The upper bound in (26) then follows immediately from (27) and (29).

To prove that the upper bound in (26) is attainable for complete graph, let us first investigate the eigenvalues in (15) for the corresponding Laplacian matrix \(L\). For the complete graph, we have \(L = NI_{N \times N} - 11^T\), where the vector \(1\) is a all one \(N\) dimensional vector. That is, for the Laplacian matrix \(L\) of the complete graph, the diagonals have the value of \(N - 1\), and all other entries have the value of \(-1\). An elementary algebra shows that the eigenvalues of \(L\) are \(\lambda_1 = 0\) and \(\lambda_2 = \lambda_3 = \cdots = \lambda_N = N\).

On the one hand, plugging these eigenvalues for the complete graph into (18) and (16), we have \(\delta_{opt} = 1/N\) and the optimal weight matrix becomes \(W_{opt} = I_{N \times N} - (1/N)(NI_{N \times N} - 11^T) = \frac{1}{N}11^T\). By the definition of (9), for the CISPRT with the optimal weight matrix under the complete graph scenario, each local sensor is to put equal weights to all raw sensor observations, and each and every local sensor essentially runs the optimal
centralized SPRT in (4). Hence, for the complete graph, the CISPRT with the optimal weight matrix is equivalent to the optimal centralized SPRT, and thus we have \( \lim \sup_{\epsilon \to 0} \frac{E_1(T_i)}{M(\epsilon)} = 1 \).

On the other hand, by (19), with the optimal weight matrix for the complete graph, we have \( r = r_{opt} = 0 \), and thus the upper bound becomes \( N^{r_{opt}^2} + 1 = 1 \). Combining the above results together shows that the upper bound is attainable and sharp, completing the proof of Theorem III.2.

IV. Simulation Studies

In this section, we report our simulation study results to illustrate the usefulness of our improved performance properties of CISPRT algorithm.

We use random graph to generate the neighborhood structure of sensor as follow. Assume the \( N \) sensors correspond to \( N \) random points in a unit square \([0,1] \times [0,1]\). Two sensors are connected if and only if the distance of the corresponding two points is less than the connectivity parameter \( g \). In our simulation studies, we consider two choices of \( N = 30 \) and \( 300 \), and two choices of the connectivity parameter \( g = 0.3 \) and \( 0.9 \). In other words, we consider a total of \( 2 \times 2 = 4 \) different network structures. For each of these four given networks, the raw sensor observations \( y_i(t) \)'s are assumed to be \( N(\mu_1, \sigma^2) \) under the alternative hypothesis \( H_1 \) and \( N(-\mu_1, \sigma^2) \) under the null hypothesis \( H_0 \). Here we set the mean \( \mu_1 = 1/\sqrt{2} \) and \( \sigma = 1 \), so that the local Kullback-Leibler divergence in (7) becomes \( m = 2 \mu^2/\sigma^2 = 1 \) at each local sensor.

Our focus is on the performance of the CISPRT in (13) in each of these four given networks, as the Type I and Type II error probabilities constraints \( \alpha = \beta = \epsilon \) vary from \( 10^{-8} \) to \( 10^{-4} \) with step size \( 10^{-6} \). For simplicity, we consider the symmetric scenario when the lower and upper bounds \( \gamma^l_i \) and \( \gamma^h_i \) of the CISPRT in (13) are given by \( \gamma^l_i \equiv -\gamma \) and \( \gamma^h_i \equiv \gamma \) for all \( i = 1, \cdots, N \) for some \( \gamma > 0 \). In our simulation studies, we choose \( \gamma \) based on our improved upper bound on Type I error probability by equating (24) to \( \epsilon \) and then finding the corresponding numerical value of \( \gamma \) for each \( \epsilon \). For each given threshold \( \gamma \), we simulate the expected sample size of the CISPRT in (13) under \( H_1 \) based on \( M = 2000 \) Monte Carlo runs.

For the purpose of easy understanding, Figure 1 compare our improved upper bound (red solid line) in (25) with three other estimates of \( E_1(T_i) \) of the CISPRT: (i) Sahu and Kar’s upper bound (blue dashed line) in (22); (ii) the Monte Carlo simulated expected sample size, \( E_1(T_i) \) (purple dotted line); and (iii) the lower bound (green dotdash line) in (22). From the plots, our upper bounds on the expected sample size are much closer to the Monte Carlo simulated expected sample size than Sahu and Kar’s original upper bound, no matter the number of sensors and the connectivity parameter.
It is interesting to see the simulated $E_1(T_i)$ falls between our improved upper bound and the lower bound. However, note that the actual lower bound in (22) includes an uncomputable constant, and thus it is not as close to the Monte Carlo simulated $E_1(T_i)$ as Figure 1 illustrates. Also note that the first order term of our improved upper bound is same as that of the lower bound as $\epsilon$ goes to 0, and this confirms that our improved upper bound is sharp up to first-order.

V. PROOF OF THEOREM III.1

This section is devoted to prove Theorem III.1 under the non-asymptotic setting.

Let us fix the $i$-th local sensor, and investigate the properties of the stopping time $T_i$ in (12) of the CISPRT at this specific local sensor that are related to the upper bound $\gamma_i^h$. Since the detailed proof is technical and involves many subscripts, we decide to abuse the notation and denote the stopping time $T_i$ and the upper bound $\gamma_i^h$ in (12) simply by $T$ and $\gamma$.

Let us first provide the high-level idea to prove Theorem III.1. Note that the Type I error probability can be written as

$$\mathbb{P}_0(\delta_i = 1) = \mathbb{P}_0(S_i(T) \geq \gamma),$$  

(30)

where $S_i(T)$ is the value of the local test statistic in (9) at the stopping time $T$. Note that in the classical sequential analysis for the centralized setting, it is standard to use the change of measures arguments, and rewrite the Type I error probability as its equivalent form of $E_1(e^{-S_c(T)}I(S(T) \geq \gamma))$, where $S_c(T)$ is the centralized log-likelihood ratio in (4). The analysis on this error probability analysis and the expected sample size $E_1(T)$ is then based on the renewal theory and overshoot analysis for random walks over time $t$. Unfortunately, such approach breaks down for distributed setting when the centralized test statistic $S_c(t)$ and the local test statistic $S(t)$ can be completely different. Moreover, with large number $N$ sensors, the overshoot analysis often involves constants that are exponentially increasing as $N$ increases and thus the corresponding analysis can be meaningless under the practical setting.

Sahu and Kar [22] proposed an alternative method to bound the Type I error probability and expected sample size directly. Specifically, note that

$$\mathbb{P}_0(\delta_i = 1) = \mathbb{P}_0(S_i(T) \geq \gamma) = \sum_{t=1}^{\infty} \mathbb{P}_0(T = t, S_i(t) \geq \gamma) \leq \sum_{t=1}^{\infty} \mathbb{P}_0(S_i(t) \geq \gamma) \leq \sum_{t=1}^{\infty} Q\left(\frac{\gamma - \mu_0^*(t)}{\sqrt{V_0^*(t)}}\right).$$  

(31)

Here $Q(u) = \mathbb{P}(N(0, 1) > u)$, and the local test statistics $S_i(t)$ are Gaussian distributed under $H_0$, say, $N(\mu_0^*(t), V_0^*(t))$, ...
Fig. 1: A comparison of four different estimates of $E_1(T_i)$ of the CISPRT under four different setting of random graph depending on the number $N$ of sensors and the connectivity parameter $g$. In each plot, four curves represent four different estimates as $\alpha = \beta = \epsilon$ varies, and these four methods rank from largest to smallest are as follows: (i) The blue dashed line is Sahu and Kar’s upper bound in (22); (ii) The red solid line is our improved upper bound in (25); (iii) The purple dotted line is the Monte Carlo simulated estimate of $E_1(T_i)$; and (iv) The green dotdash line is the lower bound in (22). The plots confirms that our upper bound in (25) provides the correct first-order approximation of $E_1(T_i)$.

at any fixed $t$, since the raw sensor observations are Gaussian. Meanwhile, the expected sample size, $E_1(T)$, is
bounded by

\[ E_1(T) = \sum_{t=0}^{\infty} P_1(T > t) \leq \sum_{t=0}^{\infty} P_1(S_i(t) < \gamma) \leq \sum_{t=0}^{\infty} Q \left( \frac{\gamma - \mu_1^*(t)}{\sqrt{V_1^*(t)}} \right), \tag{32} \]

where the local test statistics \( S_i(t) \) are Gaussian \( N(\mu_1^*(t), V_1^*(t)) \) under \( H_1 \).

Due to the similarity between (31) and (32), below we will focus on the Type I error probability analysis in (31).

By (7) and (9), Sahu and Kar (2016) showed that

\[ \mu_0^*(t) = -mt \quad \text{and} \quad V_0^*(t) \leq \frac{2m(Nr^2 + 1)t}{N}. \tag{33} \]

Sahu and Kar [22] then combined these above results together to bound the infinite sum in (31) by splitting the interval \( t \in [1, \infty) \) into four subintervals:

\[ [1, \frac{\gamma}{2m}], \quad [\frac{\gamma}{2m}, \frac{\gamma}{m}], \quad [\frac{\gamma}{m}, \frac{2\gamma}{m}], \quad [\frac{2\gamma}{m}, \infty). \tag{34} \]

Bounding the sum in each of these four subintervals leads Sahu and Kar (2016) to derive the result in (21).

The direct approach in (31) - (34) is non-asymptotic, but unfortunately it is too crude in general. Indeed, if we apply them directly to investigate the Type I error probability or expected sample sizes of the centralized SPRT, the results will be much looser as compared with those from the classical renewal theory: while the first-order terms have the same order, the coefficients from the direct approach in (31) - (34) are much larger.

After a careful analysis, we find out that the main reason is caused by the middle two subintervals in (34), and the direct approach in (31) - (34) can be refined to provide better bounds if we further split each of the middle two subintervals into \( k \) sub-subintervals, for some suitable choice of \( k \) that will be optimally determined later. In fact, when we applied this new refined approach to investigate the Type I error probability or expected sample sizes of the centralized SPRT, then the corresponding results are first-order asymptotically equivalent to those from the classical renewal theory. This suggests that the refined direct approach yields an accurate upper bound for complete graph regardless of the number \( N \) of sensors, and thus may also lead to good bounds for other network structures.

Now we are ready to provide the detailed, rigorous proof of (24). First, we further split each of the middle two subintervals into \( k \) sub-subintervals as follows. Let \( \ell = \frac{\gamma}{2m} \) or \( \frac{\gamma}{m} \), and we propose to further split the subinterval \([\ell, 2\ell]\) as \( k \) sub-subintervals:

\[ \left[ \frac{k+j-1}{k} \ell, \frac{k+j}{k} \ell \right] \quad \text{for} \quad j = 1, \ldots, k. \tag{35} \]

Relation (24) in Theorem III.1 can be proved by bounding the infinite sum in (31) through these subintervals.
Second, we will use heavily the following well-known fact for $N(0, 1)$ distribution:

$$Q(x) = P(N(0, 1) > x) \leq \frac{1}{2} \exp \left(-\frac{x^2}{2}\right) \text{ for all } x > 0.$$  \hfill (36)

Also $Q(x)$ is decreasing as a function of $x$, and thus replacing $V_0^*(t)$ by its upper bound in (33) yields an upper bound of (31).

Next, by (31), (33) and (36), we have

$$P_0(\delta_i = 1) \leq \sum_{t=1}^{\infty} Q \left( \frac{\gamma + mt}{\sqrt{2mt(Nr^2 + 1)}} \right) \leq \frac{1}{2} \sum_{t=1}^{\infty} \exp \left\{ -\frac{N(\gamma + mt)^2}{4mt(Nr^2 + 1)} \right\},$$

where $A_1, A_2, A_3$ and $A_4$ denote the corresponding sum over $t$ in the subintervals in (34), whereas $A_{2j}$ and $A_{3j}$ are over the sub-subintervals in (35) for $\ell = \frac{\gamma}{2m}$ or $\frac{\gamma}{m}$.

Sahu and Kar [22] proved their results by bounding $A_1, A_2, A_3$ and $A_4$, and here we refine their results by bounding $A_{2j}$'s and $A_{3j}$'s. The main mathematical tool is the simple fact that when $a \leq t \leq b$, for $c = N/(4(Nr^2 + 1))$,

$$\sum_{t=a}^{b} \exp \left\{ -c \left( \frac{\gamma + mt}{mt} \right) \right\} \leq \sum_{t=a}^{b} \exp \left\{ -c \left( \frac{\gamma^2}{mb} + 2\gamma + mt \right) \right\}$$

$$= \exp \left\{ -c \left( \frac{\gamma^2}{mb} + 2\gamma \right) \right\} \frac{\exp(-cm(a)) - \exp(-cm(b + 1))}{1 - \exp(-cm)},$$

$$\leq \rho^{-1} \exp \left\{ -c \left( \frac{\gamma^2}{mb} + 2\gamma \right) \right\} \exp(-cm),$$

where the constant $\rho = 1 - \exp(-cm)$ is defined in (20).

In order to help casual readers better understanding our main ideas, let us first provide the bounds of Sahu and Kar [22] on $A_1$ and $A_2$. Applying (38) to the case of $a = 1$ and $b = \gamma/(2m)$, we have

$$A_1 \leq \rho^{-1} \exp(-4c\gamma) \exp(-cm) \leq \rho^{-1} \exp \left( -\frac{N\gamma}{N r^2 + 1} \right).$$  \hfill (38)

where the last relations follows from the fact that $\exp(-cm) < 1$ and the definition of $c = N/(4(Nr^2 + 1))$.

Similarly, applying (38) to the case of $a = \frac{\gamma}{2m}$ and $b = \frac{\gamma}{m}$, we have

$$A_2 \leq \rho^{-1} \exp(-3c\gamma) \exp(-\frac{1}{2}c\gamma) = \rho^{-1} \exp(-\frac{7}{2}c\gamma)$$

$$= \rho^{-1} \exp \left( -\frac{7}{8} \frac{N\gamma}{N r^2 + 1} \right).$$  \hfill (39)
It is easy to see that $A_3$ satisfies (39), whereas $A_4$ satisfies (38). A combination of (37) with the bounds in (38) and (39) yields (21), which is the upper bound of $P_0(\delta_i=1)$ derived in Sahu and Kar [22].

To improve the upper bound in (21) of Sahu and Kar [22], our key observation is that the bound in (39) for $A_2$ and $A_3$ can be further reduced. For that purpose, let us consider the $A_{2j}$ over the $j$-th sub-subinterval in (35), and apply (38) to the case of $a = \frac{k+j-1}{k}$ and $b = \frac{k+j}{2k}$. Then for $j = 1, 2, \cdots, k$, we have

$$A_{2j} \leq \rho^{-1} \exp \left( -2c_2 \frac{2k+j}{k+j} \right) \exp \left( -c_2 \frac{k+j-1}{2k} \right)$$

$$= \rho^{-1} \exp(-2c_2) \exp \left( -c_2 \left( \frac{2k}{k+j} + \frac{k+j-1}{2k} \right) \right)$$

$$\leq \rho^{-1} \exp(-2c_2) \exp \left( -c_2 (2 - \frac{1}{2k}) \right)$$

$$= \rho^{-1} \exp \left\{ -\frac{8k-1}{8k} \frac{N\gamma}{N + 1} \right\}, \quad (40)$$

where the second to last relation follows from the simple fact that $u + 1/u \geq 2$ for $u = 2k/(k+j)$, and the last equation is from the definition of $c = N/(4(Nr^2 + 1))$.

It is useful to discuss the implication of (40) and compare the corresponding upper bound on $A_2$ with those in (39). By (40), we have

$$A_2 \leq k \rho^{-1} \exp \left\{ -\frac{8k-1}{8k} \frac{N\gamma}{N + 1} \right\} \quad (41)$$

It is interesting to see that (39) is a special case of (41) with $k = 1$. By increasing the value of $k$ or the number of sub-subintervals, we can reduce the factor $7/8$ in the exponent term of (39) to a much smaller value close to 1, and the price we pay is the multiplication factor $k$.

Similarly, the same technique of (40) is applied to $A_{3j}$ or $A_3$, and we have

$$A_{3j} \leq \rho^{-1} \exp \left\{ -\frac{4k+3}{4k+4} \frac{N\gamma}{N + 1} \right\} \quad (42)$$

for $j = 1, \cdots, k$. On the one hand, when $k = 1$, the upper bounds in (40) and (42) are the same with the factor $7/8$ in the exponential term. On the other hand, for a general $k > 1$, the upper bound in (42) is larger, which can also be used to bound all $2(k+1)$ terms in $A_1, A_{2j}$ and $A_4$. By (37), the Type I error probability of the CISPRT is bounded by

$$P_0(\delta_i=1) \leq \frac{1}{2} (A_1 + \sum_{j=1}^{k} A_{2j} + \sum_{j=1}^{k} A_{3j} + A_4)$$

$$\leq (k+1) \rho^{-1} \exp \left\{ -\frac{4k+3}{4k+4} \frac{N\gamma}{N + 1} \right\}$$

$$= \rho^{-1} \exp \left\{ -\frac{N\gamma}{N + 1} \right\} \times \exp \left\{ \log(u) + \frac{1}{u} \frac{N\gamma}{4(N + 1)} \right\}, \quad (43)$$
where $u = k + 1$. A simple calculus analysis shows that for any given $D > 0$, the function $\log(u) + D/u$ is minimized at $u_{\text{opt}} = D$ with the minimum value of $\log(D) + 1$. However, a subtly here is that we should restrict $u$ to be integers. The good news is that (43) holds for any integer $u = k + 1$ and thus we can choose a specific integer $u^* = \lceil D \rceil$, the smallest integer $\geq D$. For this specific choice of $u^*$, we have

$$\log(u^*) + \frac{D}{u^*} \leq \log(D + 1) + \frac{D}{D} = \log(D + 1) + 1,$$

(44)

which is asymptotically equivalent to the minimum bound $\log(D) + 1$ for large $D$. Combining the above results together yields (24), which completes the proof of Type I error properties of the CISPRT algorithm.

The proof of (25) for the expected sample size is similar, except with different subintervals. By (7) and (9), we can show that

$$\mu_1^\ast(t) = mt \quad \text{and} \quad V_1^\ast(t) \leq \frac{2m(Nr^2 + 1)t}{N}.
$$

(45)

By (32), for the CISPRT,

$$E_1(T) \leq \sum_{i=0}^{\infty} Q\left(\frac{tm - \gamma}{\sqrt{2tm(Nr^2 + 1)}/N}\right)
= \frac{B_1 + B_2 + B_3 + B_4}{N},$$

(46)

where $B_1, B_2, B_3, B_4$ correspond to the summation over $t$ in each of the following four subintervals:

$$[0, \frac{\gamma}{m}], \ [\frac{\gamma}{m}, \frac{3\gamma}{2m}], \ [\frac{3\gamma}{2m}, \frac{2\gamma}{m}], \ [\frac{2\gamma}{m}, \infty).$$

(47)

It turns out that the bounds derived in Sahu and Kar [22] for $B_1, B_3, B_4$ are tight, and the bound on $B_2$ is too loose. To be more specific, Sahu and Kar [22] used the similar technique for Type I error to show that $B_3 \leq \frac{1}{2}\rho^{-1}$ and $B_4 \leq \frac{1}{2}\rho^{-1}$, and also bound $B_1$ and $B_2$ by

$$B_1 = \sum_{0 \leq t \leq \gamma/m} Q\left(\frac{tm - \gamma}{\sqrt{2tm(Nr^2 + 1)}/N}\right) \leq \sum_{0 \leq t \leq \gamma/m} 1 \leq \frac{\gamma}{m} + 1,$$

$$B_2 = \sum_{\frac{\gamma}{m} < t \leq \frac{3\gamma}{2m}} Q\left(\frac{tm - \gamma}{\sqrt{2tm(Nr^2 + 1)}/N}\right) \leq \sum_{\frac{\gamma}{m} < t \leq \frac{3\gamma}{2m}} 1 = \frac{\gamma}{4m},$$

$$B_3 \leq \frac{1}{2}\rho^{-1}, \quad \text{and} \quad B_4 \leq \frac{1}{2}\rho^{-1}$$

(48)

Here $B_1$ and $B_2$ are based on the two simple facts: (1) $Q(u) = P(\mathcal{N}(0,1) > u) \leq 1$ for all $-\infty < u < \infty$ and (2) $Q(u) \leq 1/2$ when $u > 0$. 


To improve the upper bound of $B_2$, we propose to further split the subinterval $\left[ \frac{\gamma}{m}, \frac{3\gamma}{2m} \right]$ into $k$ sub-intervals:

$$\left[ \frac{j+2 \gamma}{j+1 m}, \frac{j+1 \gamma}{j m} \right]$$

for $j = 2, \cdots, k$. (49)

Denote by $B_2^{(1)}$ and $B_2^{(j)}$ the summation as in $B_2$ in (48) except when $t$ is over the first and $j$-th sub-interval in (49), respectively, for $j = 2, \cdots, k$. For the first subinterval in (49), by the simple fact that $Q(u) \leq \frac{1}{2}$ when $u \geq 0$, we have

$$B_2^{(1)} = \frac{1}{2} \left( \frac{1}{k+1} \frac{\gamma}{m} \right). \quad (50)$$

For the $j$-th subinterval in (49) with $j = 2, \cdots, k$, we propose to explore relation (36), which provides a much improved bound for $Q(u)$ than the constant $1/2$ when $u > 0$. That is, by (36), for $j = 2, \cdots, k$, we have

$$B_2^{(j)} \leq \frac{1}{2} \sum_{j=2}^{j+1} \frac{\gamma}{m} \exp \left\{ -\frac{N(\gamma - mt)^2}{4mt(Nr^2 + 1)} \right\} \quad (51)$$

Our remaining arguments are similar to those in (38), with a minor twist to reflect the change of mean from $\mu_0^*(t)$ to $\mu_1^*(t)$. To be more specific, as in (38), it is not difficult to see that

$$\sum_{l=a}^{b} \exp \left\{ -c \left( \gamma \right) \frac{mt}{\gamma} \right\}$$

$$\leq \rho^{-1} \exp \left\{ -c \left( \frac{\gamma^2}{mb} - 2\gamma \right) \right\} \exp(-cma) = \rho^{-1},$$

when $a = \frac{j+2 \gamma}{j+1 m}$ and $b = \frac{\gamma}{2}$. Combining this with (51) yields that

$$B_2^{(j)} \leq \frac{1}{2} \rho^{-1}, \quad (53)$$

where the right-hand side upper bound does not depend on $\gamma$. By (50) and (53), we have

$$B_2 \leq \frac{1}{2} \frac{1}{k+1} \frac{\gamma}{m} + \sum_{j=2}^{k} \frac{1}{2} \rho^{-1} = \frac{1}{k+1} \frac{\gamma}{2m} + (k-1) \frac{1}{2} \rho^{-1}. \quad (54)$$

Hence, by (46), (48) and (54), the expected sample size of $T$ under $H_1$ satisfies

$$E_1(T) \leq B_1 + B_2 + B_3 + B_4$$

$$\leq \frac{\gamma}{m} + 1 + \left[ \frac{1}{k+1} \frac{\gamma}{2m} + \frac{k-1}{2} \rho^{-1} \right] + \frac{1}{2} \rho^{-1} + \frac{1}{2} \rho^{-1}.$$
\[
\frac{\gamma}{m} + \frac{\gamma}{2(k + 1)m} + \frac{k + 1}{2}\rho^{-1} + 1, \quad (55)
\]
for any integer \(k \geq 1\). When \(k = 1\), this is just the upper bound of \(E_1(T)\) in (22) derived by Sahu and Kar [22].

Here we will choose \(k\) suitably to drive a better upper bound of \(E_1(T)\). In particular, in (55), we can choose the integer \(k = k^* = \lceil \sqrt{\frac{\gamma m}{\rho}} - 1 \rceil\), and by the similar arguments in (44), it is straightforward that (25) follows at once from (55), which completes the proof of the theorem.

VI. CONCLUSIONS

In this paper, we investigate the distributed sequential detection problem over a pre-specified network structure. Our focus is on improving the non-asymptotic upper bound of the CISPRT algorithm proposed by Sahu and Kar [22]. We derive a tight upper bound through a novel approach to bound the infinite sum of tail probabilities of Gaussian distributions. Moreover, we provide sharp upper bound on the information loss of the CISPRT algorithm as compared to the centralized optimal SPRT in the asymptotic regime as Type I and II error probabilities go to 0.

Several future directions can be pursued for distributed sequential detection. First, instead of binary simple hypothesis testing, it will be interesting to develop efficient algorithms when there are nuisance parameters in the alternative hypothesis. Second, instead of all local sensors having different distributions simultaneously under the alternative hypothesis \(H_1\), one may want to develop efficient algorithm where only a few unknown subset of local sensors have distributions from \(H_1\). This might be closely related to sparsity detection or false discovery rate in the modern statistics literature. Finally, in our paper the neighborhood or network structure is pre-specified, and it will be useful to investigate the time-varying network structure.

REFERENCES


