Distributed Chernoff Test:
Optimal Decision Systems over Networks
Anshuka Rangi, Massimo Franceschetti and Stefano Marano

Abstract—We study active/adaptive decision making over sensor networks where the most appropriate sensors’ probing actions are chosen by continuously learning from the environment. Two network settings are considered: with and without central coordination. In the first case, the network interact with each other through a central entity, which plays the role of a fusion center. In the second case, all network processing takes place at the node level, in a fully distributed fashion. In both scenarios, we propose sequential and adaptive hypothesis tests, motivated by the classic Chernoff’s solution. We compare the performance of the proposed tests to the optimal sequential test over the network. In the presence of a fusion center, the proposed test achieves the same asymptotic optimality of the Chernoff test, minimizing the expected cost required to reach a decision plus the expected cost of making a wrong decision, when the observation cost per unit time tends to zero. The proposed test is also asymptotically optimal in the higher moments of the time required to reach a decision. Additionally, the test is parsimonious in terms of communications, and the expected number of channel uses per network node tends to a small constant. In the distributed setup, we provide sufficient conditions for which the proposed test achieves the same asymptotic optimality as Chernoff’s test. Under these conditions, the test is also asymptotically optimal in the higher moments of the time required to reach a decision. Additionally, this test is parsimonious in terms of communications in comparison to state-of-the-art schemes proposed in the literature.

Index Terms—Distributed detection, Active/Adaptive hypothesis testing, Chernoff test, Sequential testing, Internet of Things, Sensor Networks.

I. INTRODUCTION

With the boom in the Internet of Things, sensor-network based solutions have become increasingly popular for inference systems [1–3]. This is mainly due to the decreasing cost of the sensors, their increasing computational capabilities, the availability of high-speed communication channels, and the redundancy provided by the distributed structure of the network [4]. Inference systems have two key functionalities: decision making (or hypothesis testing) and estimation. We focus on designing optimal tests in decision-making scenarios. In this case, upon collecting observations the nodes of the network take a temporary decision in favor of a hypothesis, and this decision is used to perform future actions. Applications that fall in this framework include intrusion and target detection, and object classification and recognition [5–9].

A. Rangi and M. Franceschetti are with Department of Electrical and Computer Engineering of University of California, San Diego, USA. S. Marano is with Department of Information and Electrical Engineering and Applied Mathematics of University of Salerno, Italy. E-mails: {arangi, massimo}@ucsd.edu, marano@unisa.it

Previous studies are broadly classified into two categories: fusion center based and distributed. In the first case, all the nodes of the network are connected to a central unit — also referred to as fusion center — and two operative modalities are considered. In the first modality, the network nodes simply deliver their observations to the fusion center, where the inference task is performed. In the second modality, the nodes exploit their computational capability to perform preliminary processing on the observations, and only a limited amount of information is delivered to the fusion center for making the final decision. This reduces the communication overhead but may also result in a loss of performance.

In the distributed setup, network nodes are connected to each other via communication links, typically forming a sparse network, and there is no central processing unit. Thus, to perform an inference task, the network nodes need to perform computations locally, share their processed data with neighboring nodes, and collectively reach a decision.

A natural question in both the network settings is what kind of local processing to perform at the nodes, and what fusion scheme to adopt at the fusion center or at the network nodes, in order to reduce the communication burden while keeping a high level of performance. In this work, we address this question by proposing statistical tests for both settings, and compare their performance to the optimal test for the network.

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

The contributions of this work is twofold. We propose a Decentralized Chernoff Test (DCT) for the fusion center based setup, and a Consensus based Chernoff Test (CCT) for the distributed setup. We provide bounds on the test performance in terms of their risk, defined as the expected cost required to reach a decision plus the expected cost of making a wrong decision.
We also provide converse results showing the best possible performance of any sequential test over the network. We show that DCT is asymptotically optimal in terms of both the risk and the higher moments of the expected decision time, as the observation cost per unit time vanishes. Additionally, DCT is parsimonious in terms of communication: we show that, when the observation cost per unit time vanishes, the expected number of channel uses per node tends to four. We also provide sufficient conditions for which CCT retains the asymptotic optimality of Chernoff’s original solution, both in terms of risk and higher moments of decision time.

The rest of the paper is organized as follows: Section II discusses related work; Section III formulates the problem; Section IV reviews the standard Chernoff test; Section V introduces the Decentralized Chernoff Test (DCT); Section VI introduces the Consensus based Chernoff Test (CCT); Section VII presents theoretical results on DCT and CCT; Section VIII presents simulation results; Section IX concludes the work.

II. RELATED WORK

Sequential tests were first introduced by Wald in [10]. One of such tests, the Sequential Probability Ratio Test (SPRT) was established to be optimum for binary hypothesis testing in [11]. Among the feasible error exponents for sequential tests, SPRT achieves the best exponent. The asymptotic optimality of SPRT was proven for multi-hypothesis testing in [12], [13]. Sequential tests can be combined with adaptive schemes to enhance the performance, as these schemes adapt their choice of actions based on the past observations. In the case of sequential and adaptive tests, Chernoff provided the optimal test for binary composite hypotheses [14]. The asymptotic optimality of the Chernoff test was proved for multi-hypothesis testing in [15], see also [16] and references therein for an application. Later, the sequentiality and adaptivity gains for different classes of tests were studied, and it was established that sequential-adaptive tests outperform other classes of tests [17]. The gains can vary from application to application [18]-[20]. All these results are established in a setup where a single agent performs the task of hypothesis testing individually, without any collaboration with other agents.

Various works discuss the extension to an ensemble of networked sensors. Hypothesis testing was discussed for various configurations of networks in [21], [22]. Techniques for combining the information at the fusion center from various sensors are considered in [4], [23]-[25]. Minimization of the risk, which is dependent on both the decision time and the reliability of the decision, requires joint optimization over both the node level computations and the fusion center operations. The key challenges of this optimization problem are highlighted in [26]. An asymptotically optimal sequential (non-adaptive) test has been developed for hypothesis testing in the fusion center based setup [27], [28]. The measure of optimality adopted in [27], [28] is the same as the one used in our setup.

Previous works do not consider the performance of sequential and adaptive tests. The DCT proposed in this work goes in the direction of filling this gap, for networks with a star topology, namely networks in which each node is directly connected to a central coordination entity.

On the other hand, the CCT proposed in this work considers networks having a general graph structure and no central entity. In this case, various gossip protocols have been proposed for distributed estimation of the mean, sum, minimum and maximum of the node observations [29]-[34]. These protocols can be broadly classified into two categories: consensus protocols and running consensus protocols. In consensus protocols, the estimation task is performed after the collection of all the measurements (or observations) at the network nodes [29], [30], [32], [33]. Necessary and sufficient conditions for convergence are well studied, see e.g., [35]. In running consensus protocols, the collection of the measurements from the environment and the inference task are performed simultaneously at the network nodes [31], [34].

Motivated by these studies, protocols for distributed detection have been proposed, using different consensus protocols over belief vectors, where an element of the belief vector is the probability that a certain hypothesis is true given all the past information at a given node. The works on distributed detection focus on different strategies to transmit and combine the belief vectors over the network, and study the learning rate of these strategies [36]-[40]. A strategy based on distributed dual averaging was proposed in [39], using an optimization algorithm from [41]. The work in [36] proposes use of linear consensus strategies to combine the belief vectors. In this work, the learning rate is characterized in terms of the total variational error among the belief vectors across the network. Other works considered Bayesian strategies for updating and combining the belief vectors at the nodes, see [40]. The bounds on the asymptotic learning rate are presented in terms of KL-divergences and the centrality of the network nodes as times go to infinity [36], [40]. Under the assumption that the log-likelihood ratio is bounded, finite-time analysis on KL divergence cost has been studied in [38]. Under the same assumption, similar results have been obtained for networks modeled as time-varying graphs [37], [42]. Despite the huge literature on distributed detection, these works do not study sequential hypothesis testing in a distributed setup, which requires designing an appropriate stopping rule over the network and evaluating the corresponding expected decision time. Recently, a sequential (non-adaptive) hypothesis test which is asymptotically optimal in terms of risk has been proposed [43]. In the present work, we propose a sequential as well as adaptive hypothesis test in the distributed network setup. Unlike the previous literature, including [43], the proposed test does not perform consensus over the belief vector, and is parsimonious in terms of communication. Finally, the stopping criterion proposed in [43] is not applicable to CCT.

III. PROBLEM FORMULATION

We consider both a fusion center based setup and a fully distributed setup. In the fusion center based setup, we consider a network composed of L nodes and one fusion center. The network nodes and the fusion center are connected to
each other via communication links, while no direct mode of communication between the sensors is allowed.

In the distributed setup, we consider a network with a fully-flat architecture without any fusion center. All the communication and the information processing tasks take place at the node level, and are fully distributed because nodes exploit only locally available information. The network is composed of \( L \) nodes (or sensors) and is modeled as a graph \( G(L, E) \), where the set of nodes is denoted by \( L = \{1, 2, \ldots, L\} \), and the elements of \( E \) are the edges, namely unordered pairs of nodes \( \{\ell, j\} \), in which \( \{\ell, j\} \) represents the communication link between nodes \( \ell \) and \( j \), \( \ell \neq j \). The inter-node communication is allowed only over the edges in \( E \). The diameter \( d^G \) of the network is the maximum shortest hop-distance between any pair of nodes of \( G(L, E) \). We also denote by \( h^G \) the shortest height of all possible spanning trees of \( G(L, E) \). It is assumed that the network is connected, namely, there exists a path between any two nodes \( \ell \) and \( j \). This ensures that \( d^G \) and \( h^G \) are both finite.

The state of nature to be detected is one of \( M \) exhaustive and mutually exclusive hypotheses \( \{h_i\}_{i \in [M]} \), where the short-cut notation \( [M] = \{1, 2, \ldots, M\} \) is used. At each time instant, each node takes a probing action, selected from a fixed set of actions \( S = \{u_i\}_{i \in [M]} \). We assume that sensors select their actions independently of each other, and that the cardinality of \( S \) is equal to \( M \). Under this assumption, action \( u_i \) can be interpreted as the "best" action when the state of nature is \( h_i \), \( i \in [M] \). All the results of this paper can easily be extended to the more general case.

Suppose that the state of nature is \( h_i \), and consider node \( \ell \in L \). Let \( u_k, k \in [M] \), be the probing action taken by node \( \ell \) at a given time. Then, the probability distribution of the observation received at the node as a consequence of its probing action is denoted by \( p^{u_k}_{\ell,i} \). Given the true hypothesis \( h^* \), the observations received by any node are independent of the observations received by other network nodes. On the other hand, for a given node, observations collected at different time instants are not independent, because the probing actions are observation-dependent. The node learns from the past and tries to select the best action for the future.

The performance measure used in this work – the risk – is analogous to the one considered in [14]. Under true hypothesis \( H^* = h_i \), the risk \( R^i_\delta \) of a sequential test \( \delta \) is defined as
\[
R^i_\delta = c \mathbb{E}_\delta[N] + \omega_i \mathbb{P}_\delta(H \neq h_i),
\]
where \( N \) is the time required to reach a global decision in the network, \( c \) is the observation cost per unit time, \( H \) is the final decision, \( \mathbb{E} \) and \( \mathbb{P} \) are the expectation and the probability operators computed under \( H^* = h_i \), and \( \omega_i \) is the cost of a wrong decision. Note that the risk is the sum of the expected cost required to reach a decision and the expected cost of making a wrong decision.

We propose DCT and CCT for a fusion center based setup and a distributed setup, and evaluate their performance in terms of risk for all \( i \in [M] \), as \( c \to 0 \). Additionally, for both tests, we provide bounds on the higher moments of the time \( N \) required to reach a decision. In the fusion center based setup, the proposed test is asymptotically optimal in terms of risk as well as the higher moments of decision time, as \( c \to 0 \).

Assumptions and Notations: We assume that, following a sensor’s probing action, the observation corresponding to the probing action is instantly available at the sensor. In addition, we assume that the communication links between the sensors are noise free, and the information sent along these links is instantly available at the receiving end. The KL-divergence between the hypotheses is assumed to be finite for the entire time period under consideration. For all \( \ell \in [L] \) and \( i, j, \kappa \in [M] \), we have \( D(p^{u_k}_{\ell,i}||p^{u_k}_{\ell,j}) < \infty \). For all \( \ell \in [L] \) and \( i, j \in [M] \), there exists an action \( u_k \), where \( \kappa \in [M] \), such that \( D(p^{u_k}_{\ell,i}||p^{u_k}_{\ell,j}) > 0 \). This assumption entails little loss of generality, rules out trivialities, and is commonly adopted in the literature, see e.g., [14]. For all \( \ell \in [L] \) and \( i, j, \kappa \in [M] \), we assume \( \mathbb{E}[\log(p^{u_k}_{\ell,i}(Y))/\log(p^{u_k}_{\ell,j}(Y))]^2 < \infty \). We use the following notation: if \( v_1 = [v_{1,1}, \ldots, v_{k,1}] \) and \( v_2 = [v_{1,2}, \ldots, v_{k,2}] \) are two vectors of same dimension \( k \), then \( v_1 \leq v_2 \) means that \( v_{1,i} \leq v_{2,i} \) for all \( i \in [k] \). In addition, \( |v_{1,i}| \) is the vector of absolute values of the entries of \( v_1 \).

IV. STANDARD CHERNOFF TEST

We start by considering sensor \( \ell \) alone, with no interactions with other sensors of the network. The Chernoff test for this isolated sensor works as follows [14]:

1) At step \( k - 1 \), a temporary decision is made, based on the maximum posterior probability of the hypotheses, given the past observations and actions. Stated with a formula, the temporary decision is in favor of \( h^*_{i,k-1} \) if
\[
i^*_{k-1} = \arg \max_{i \in [M]} \mathbb{P}(H^* = h_i | y_{i,k-1}, u_{k-1}),
\]
where \( H^* \) is the true hypothesis, \( y_{i,k-1} = \{y_{i,1}, \ldots, y_{i,k-1}\} \), where \( y_{i,\ell} \) is the realization of the observation collected at time \( \ell \) (step) \( i \), \( u_{k-1} = \{u_{1,1}, \ldots, u_{k-1}\} \), and \( u_{i,\ell} \) is the realization of the action made at step \( i \).

2) At step \( k \), the action \( u_{i,k} \) is randomly chosen among the elements of action set \( S \), according to the Probability Mass Function (PMF) \( Q_{i,k-1}^\ell \), where
\[
Q_{i,k-1}^\ell = \arg \max_{q \in Q} \min_{j \in M_{i,k-1}} \sum_{u \in [M]} q(u) D(p_{i,k-1,\ell}^{u}||p_{j,k}^{u}),
\]
in which \( Q \) denotes the set of all the possible PMFs over the alphabet \( [M] \) of \( S \), and \( M_{i,k-1} = [M] \setminus \{i^*_{k-1}\} \).

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

4) The test stops at step \( N \) if the worst case log-likelihood ratio crosses a prescribed fixed threshold \( \gamma \), i.e.,
\[
\log \frac{p_{i,k}^{y_{i,k},u_k}(y_{i,k}, u_k)}{\max_{j \neq i^*_{k-1}} p_{j,k}^{y_{i,k},u_k}(y_{i,k}, u_k)} \geq \gamma,
\]
where \( p_{i,k}^{y_{i,k},u_k}(y_{i,k}, u_k) \) is the posterior probability \( \mathbb{P}(H^* = h_{i^*_{k-1}} | y_{i,k}, u_k) \) at sensor \( \ell \). If the test stops at step \( N \), then
the final decision is \( h_{i,\ell}^* \). Otherwise, \( k \leftarrow (k + 1) \), and the procedure continues from 1).

V. DECENTRALIZED CHERNOFF TEST

As the observation cost per unit time tends to zero, the probability of making a wrong decision for the standard Chernoff test tends to zero \cite{14}. It follows that minimizing the risk in \( (1) \) also corresponds to minimizing the expected number of samples required to reach a decision. When one sample is collected at each time step, minimizing the expected number of samples is obviously the same as minimizing the expected time for making a decision. However, this is not necessarily true in a network setting.

To further illustrate this point, consider first minimizing the total expected number of samples collected by the \( L \) nodes to reach a decision in a fusion center-based setup, and assume that the amount of communication between the nodes and the fusion center is unconstrained. A straightforward design, which we call Fusion center based Chernoff Test (FCT), is as follows. The action set \( S \) is modified to \( S' \) with cardinality \( ML \), where action \( a_{i,\ell} \in S' \) corresponds to the selection of \( u_i \in S \) and node \( \ell \in [M] \). Then, the Chernoff test is performed on \( S' \) at the fusion center where the selection of \( a_{i,\ell} \) corresponds to activating node \( \ell \), and enabling the activated node to use the probing action \( u_i \) to collect the corresponding observation, which is then delivered to the fusion center. It is not hard to see that, as the probability of wrong detection tends to zero, the FCT minimizes the total expected number of collected samples. The proof of this claim is similar to the proof of the optimality of the Chernoff test in \cite{14} and is thus omitted. The FCT also minimizes the total number of probing actions performed by the nodes in the network. However, there is only one active node, out of \( L \), per unit time, and all observations are communicated to the fusion center. Clearly, this is highly inefficient in terms of both communication overhead and decision time, and motivates introducing a different kind of test. Thus, we propose the DCT for minimizing the risk \( (1) \).

Our Decentralized Chernoff Test operates in two phases: initialization phase and test phase. In the initialization phase, assuming the probability distributions of the observations received at the nodes are unknown to the fusion center, each node \( \ell \) sends a vector \( v_\ell \) to the fusion center, where the elements of \( v_\ell \) are, for all \( i \in [M] \)

\[
    v_{i,\ell} = \max_{q \in Q} \min_{j \neq \ell} \sum_{u \in [M]} q(u) D(p_{i,\ell}^u || p_{j,\ell}^u). \tag{5}
\]

The quantity \( v_{i,\ell} \) is a measure of the capability of node \( \ell \) to detect hypothesis \( h_i \) (see \cite{14} for a discussion), and plays a critical role in designing the test. After receiving \( v_\ell \) from all the nodes, the fusion center sends back to each node \( \ell \) a response vector \( \rho_{\ell} \), whose \( M \) entries are the scalars

\[
    \rho_{i,\ell} = v_{i,\ell} / I(i), \tag{6}
\]

where

\[
    I(i) = \sum_{\ell=1}^L v_{i,\ell}, \tag{7}
\]

is a measure of cumulative capability of the network to detect hypothesis \( h_i \). At this point, the test phase begins. All the nodes perform the Chernoff test, independently of each other, consisting of steps 1-4 described in Section [V] with an important difference: any time at node \( \ell \) we have

\[
    \log \frac{p_{i,\ell}^*(y_{\ell}^n, u_{\ell}^n)}{\max_{j \neq \ell} p_{j,\ell}^*(y_{\ell}^n, u_{\ell}^n)} \geq \rho_{i,\ell} |\log c|, \tag{8}
\]

then a local decision in favor of \( h_{i,\ell}^* \) is communicated to the fusion center. The threshold in \( (8) \) is independent on both the hypothesis and the node \( \ell \), whereas it was a constant in the classic Chernoff test. Additionally, unlike the classic Chernoff test, the condition in \( (8) \) is not a stopping criterion for the test at node \( \ell \), but only a triggering condition for the communication between node \( \ell \) and the fusion center. Thus, node \( \ell \) continues to run the test until the fusion center sends a halting message.

The final decision \( \hat{H} \) is made at the fusion center in favor of hypothesis \( h_i \) when the local decisions from all the network nodes are in favor of \( h_i \). After the final decision is made, the fusion center sends a halting message to all the nodes. Unlike the classic Chernoff test, the nodes are guided by the fusion center to stop the test.

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

A. Informal discussion of DCT

The key idea behind the proposed DCT is to first determine the individual capabilities of the nodes for detecting the hypotheses. These capabilities — that depend on the true hypothesis \( h^* \) — are captured by the vector \( v_\ell \), whose \( i^{th} \) element is a measure of node’s \( \ell \) capability to detect the hypothesis \( h_i \). The fusion center gathers this information, and utilizes it to control the threshold at each node through the response vector \( \rho_{i,\ell} \). At the fusion center, \( I(i) \) is the measure of the cumulative detection capability of the network for hypothesis \( h_i \), and \( \rho_{i,\ell} \) denotes the fraction of this capability contributed by node \( \ell \) for hypothesis \( h_i \). To minimize the expected time to reach a decision, it is desirable to determine the threshold for each node \( \ell \) such that all the nodes require roughly the same time to reach the triggering condition in \( (8) \). This is analogous to dividing the task of hypothesis testing among the nodes based on their speed of performing the task, such that all the nodes finish their share of the task at roughly the same time.
VI. Consensus Based Chernoff Test

In this section we propose a Chernoff test designed for a distributed setup, which is referred to as the Consensus-based Chernoff test (CCT). The proposed CCT consists of three phases: initialization phase, test phase and decision phase. In the initialization phase, the network nodes reach a consensus regarding their cumulative capability to detect hypothesis $h_i$, $i \in [M]$. In the test phase, the nodes perform the Chernoff test locally and independently from each other. In the decision phase, the nodes reach a consensus regarding the final decision. The first two phases of CCT can be performed in parallel, while the last phase begins after the completion of the first two.

In the initialization phase, the goal of each node is to acquire knowledge about the cumulative capability of the network to detect the hypothesis. For node $\ell$, the measure of the capability to detect $h_i$ is given by $v_{i,\ell}$ defined in (3). Thus, the cumulative capability of the network to detect hypothesis $h_i$ is $I(i)$ defined in (7). Let $I = [I(1), \ldots, I(M)]$ be the corresponding vector. Unlike the fusion center based setup, there is no central entity to facilitate the computation of the vector $I$. Instead, the nodes can use a consensus protocol to estimate $I$, see, e.g., [29], [30], [32], [33]. The nodes estimate the arithmetic mean $I/L$ of the cumulative capability, which provides the desired estimate of $I$, assuming that $L$ is known to all the nodes. The distributed linear consensus protocol is of the form

$$\hat{I}_{\ell}^{n+1} = w_{\ell,\ell} \cdot \hat{I}_{\ell}^{n} + \sum_{j \in N_{\ell}} w_{\ell,j} \cdot \hat{I}_{j}^{n},$$

(9)

where $\hat{I}_{\ell}^{n} = [\hat{I}_{\ell}^{n}(1), \ldots, \hat{I}_{\ell}^{n}(M)]$ is the vector of estimated cumulative capabilities for the $M$ hypotheses at node $\ell$ and at time instance $n$, $w_{\ell,j}$ is the weight assigned by node $\ell$ to the estimate of node $j$, and $N_{\ell} = \{j|\{\ell,j\} \in E\}$ is the set of immediate neighbors of node $\ell$ in $G(L, E)$. At $n = 0$, the estimated cumulative capabilities are initialized as $\hat{I}_{\ell}^{0} = [v_{1,\ell}, \ldots, v_{M,\ell}]$, see (5). Note that the initialization $\hat{I}_{\ell}^{0}$ can be computed locally at the nodes. Since node $\ell$ does not communicate with nodes $\{j|j \notin N_{\ell} \cup \{\ell\}\}$, it follows that $w_{\ell,j} = 0$. Thus, for the network graph $G(L, E)$, the linear consensus protocol (9) can be written as

$$\hat{I}_{\ell}^{n+1} = W \cdot \hat{I}_{\ell}^{n},$$

(10)

where $W = [\hat{I}_{\ell}^{n}, \ldots, \hat{I}_{L}^{n}]^T$ is an $L \times M$ matrix of the estimate of the cumulative capability of the network at time instance $n$ at all nodes, and $W$ is an $L \times M$ matrix with elements $w_{\ell,j}$ where $\ell, j \in [L]$. The matrix $W$ is constrained to belong to the set

$$ \{W \in \mathbb{R}^{L \times L} | 0 < w_{\ell,j} < 1 \text{ if } j \in N_{\ell} \cup \{\ell\}, \text{ else } w_{\ell,j} = 0\}.$$

(11)

The following theorem presents necessary and sufficient conditions for the consensus protocol (10) to converge to the mean of the cumulative capability of the network $I/L$.

**Theorem 1.** ([35] Theorem 1) The consensus protocol (10) converges to the mean $I/L$ if and only if the following conditions hold

$$W \cdot I_{L \times 1} = I_{L \times 1},$$

(12)

and

$$R\left(W^{-1} \cdot I_{L \times L} \right) < 1,$$

(14)

where $R(\cdot)$ denotes the spectral radius of a matrix, and $I_{A \times B}$ is a $A \times B$ matrix of all ones. Additionally, the rate of convergence is proportional to the spectral radius at the left-hand side of (14).

Based on the above theorem, the computation of $W$ can be formulated as a convex optimization problem subject to (11), (12) and (13), and can be determined using standard techniques [35].

Our stopping rule for the initialization phase of CCT builds upon some ideas from [45] and is illustrated in Algorithm 1. In Algorithm 1, $y_{\ell}$ indicates the number of time instances since node $\ell$ is in a uniformly local $c/L^2$-consensus status. This status is defined by the following condition: for all $\ell \in [L]$ and $j \in N_{\ell}$,

$$|\hat{I}_{\ell}^{n} - \hat{I}_{j}^{n}| \leq c \cdot \frac{1}{L^2} \cdot 1_{1 \times M}.$$

(15)

The stopping rule for the initialization phase is defined as follows: if $y_{\ell} > L + 1$, the network has reached uniformly local $c/L^2$-consensus (or global $c/L$-consensus). Thus, as soon as $y_{\ell} > L + 1$, node $\ell$ sends a termination bit $m_{\ell}^{(1)} = 1$ (the superscript $^{(1)}$ indicates that this is the termination of the initialization phase) to inform its neighbors $N_{\ell}$ that the consensus to $I/L$ has been reached. When node $j$ receives $m_{\ell}^{(1)} = 1$, it halts the consensus, scales the estimate $\hat{I}_{\ell}^{n}$ by $L$ to get an estimate of $I$, and forwards $m_{\ell}^{(1)}$ to its neighbors $N_{j}$. At the termination of initialization phase, after scaling by $L$, for all $\ell, j \in [L]$, we get

$$|\hat{I}_{\ell}^{n} - \hat{I}_{j}^{n}| \leq c \cdot 1_{1 \times M}.$$

(16)

The second phase of CCT, namely the test phase, is presented in Algorithm 2. In this phase, all nodes perform the Chernoff test independently of each other, and compute $\log p_{i_{\ell}^n}(y^n, u^n)/\max_{j\neq i} p_{j_{\ell}^n}(y^n, u^n)$. Consider the generic sensor $\ell$. If at time $n$

$$\log \frac{p_{i_{\ell}^n}(y^n, u^n)}{\max_{j\neq i} p_{j_{\ell}^n}(y^n, u^n)} \geq \tilde{p}_{i_{\ell}^n}(\log c),$$

(17)

where $\tilde{p}_{i_{\ell}^n}(\cdot) = v_{i_{\ell}^n}(\cdot)/\hat{I}_{i_{\ell}^n}(\cdot)$, then sensor $\ell$ updates its local decision $H_{\ell}^{(3)}$ in favor of hypothesis $h_{i_{\ell}^n}$; otherwise, the local decision is set to NULL. Note that the consensus about $I/L$ in the first phase is independent of the computation of the log-likelihood in the second phase. Hence, the first two phases can be performed in parallel over the network. However, according to (17), the local decision at sensor $\ell$ requires a reliable
Algorithm 1 Initialization Phase of CCT

Initialize $n = 0$, and For all $\ell \in [L]$, $\hat{P}_\ell$, $y_\ell = 0$ and $z_\ell = 0$
while True do
    For all $\ell \in [L]$, broadcast local information $\hat{f}_\ell^{(n)}$ and $z_\ell$.
    Update the local cumulative capability using (9).
    $z_\ell = \min \{y_\ell, \min_{j \in N_\ell \cup \{\ell\}} z_j\} + 1$
    if $z_\ell > L + 1$ then
        Sensor $\ell$ broadcasts $m_\ell^{(1)} = 1$ and stop updating.
    Break; While;
end if
if $\max_{j \in N_\ell} |\hat{f}_j^{(n)} - \hat{f}_\ell^{(n)}| \leq c \cdot 1_{1 \times M}/L^2$ then
    $y_\ell = y_\ell + 1$
else
    $y_\ell = 0$
end if
$n = n + 1$
end while

Algorithm 2 Test Phase of CCT

For all $i \in [M]$ and $\ell \in [L]$, initialize $p_{i,\ell}(y^0, u^0) = 1/M$;
$n = 1$; $\hat{H}_\ell = \text{NULL}$
Input: Termination bit of initialization phase and decision phase, i.e., $m_\ell^{(1)}$ and $m_\ell^{(3)}$
while Final decision is not made i.e. $m_\ell^{(3)} \neq 1$ do
    For all $\ell \in [L]$, perform the Chernoff test.
    For all $\ell \in [L]$, collect $y_n, \ell$ and update $p_{i,\ell}(y_n, u_n)$.
    if (17) is true and $m_\ell^{(1)} = 1$ then
        $\hat{H}_\ell = \arg \max_{i \in [M]} p_{i,\ell}(y_n, u_n)$.
    end if
    $n = n + 1$
end while

Algorithm 3 Decision Phase of CCT

For all $\ell \in [L]$, initialize $d_\ell^{(n)} = x_n^{\ell} = 0, m_\ell^{(3)} = 0$
Input: Termination bit $m_\ell^{(1)}$
while $m_\ell^{(3)} == 1$ do
    if $m_\ell^{(3)} = 1$ is received from neighbor $j$ then
        Set the final decision, i.e., $H_\ell^{n} = H_j^{n-1}$
        Broadcast $m_\ell^{(3)}$ and $\hat{H}_\ell^{n}$.
    Break;
end if
For all $\ell \in [L]$, update $x_\ell^{n}$ according to (19).
For all $\ell \in [L]$, update $d_\ell^{n}$ according to (18).
if $d_\ell^{n} > L + 1$ then
    $m_\ell^{(3)} = 1$
    For all $\ell \in [L]$, broadcast $m_\ell^{(3)}$ and $\hat{H}_\ell^{n}$.
else
    For all $\ell \in [L]$, broadcast $d_\ell^{n}$ and $\hat{H}_\ell^{n}$.
end if

Decision Phase of CCT — to acquire this information. If the consensus algorithm requires communication of a real valued vector $\hat{P}_\ell$ only in its first phase to find the cumulative capabilities of the network. Since the termination time of Phase 1 is bounded, as we show later in (71), so is the number of communications of real valued vectors. Thus, CCT is parsimonious in terms of communication, in comparison to other schemes proposed in the literature [36–40].

A. Informal discussion of CCT

The key idea behind CCT is to determine the individual capabilities of the nodes for detecting the hypotheses. These capabilities — that depend on the true hypothesis $h^*$ — are captured by $v_i, \ell$. All the nodes in the network determine their cumulative capabilities to detect any hypothesis. Since there is no central entity to facilitate the communication of this information, they use a consensus algorithm — first phase of CCT — to acquire this information. If the consensus algorithm stops at time $N$, then $\hat{P}_\ell^{N}$ denotes the estimated fraction of the capability contributed by node $\ell$ for hypothesis $h_i$. To minimize the expected time to reach a decision, it is desirable
to determine the threshold for each node $\ell$ such that all the nodes require roughly the same time, following the termination of Phase 1, to reach the triggering condition (17). Given the estimate of cumulative capabilities $I^N$, this is analogous to dividing the task of hypothesis testing among the nodes based on their speed of performing the task, such that all the nodes finish their share of the task roughly at the same time. Phase 3 is a localized stopping criterion for the Chernoff test, and ensures that the nodes stop the test as they reach the same decisions. The quantities $x^\delta_{u}$ and $d^\delta_{u}$ capture this information mathematically, and percolate it over the network.

VII. Theoretical Results

We now present our main theoretical results, whose proofs are deferred to the Appendices.

A. Lower Bounds for a Sequential and an Adaptive Test

In this section, we present lower bounds on two different performance measures, namely risk and decision time, for any sequential and adaptive test. The superscript $\delta$ is appended to quantities that refer to a generic test and $N$ indicates the time required to take a decision.

**Theorem 2.** (Converse) Consider any sequential and adaptive test $\delta, \delta_i$. For all $i \in [M]$, suppose that the probability of missed detection is

$$P^\delta_i(\hat{H} \neq h_i) = O(c \log c), \quad c \to 0,$$

then we have, for all integers $r \geq 1$,

$$E^\delta_i[N^r] \geq \left(1 + o(1)\right) \frac{\log c}{I(i)}^r, \quad as \ c \to 0. \quad (20)$$

Using (27) with $r = 1$, we also have

$$E^\delta_i \geq \left(1 + o(1)\right) \frac{c \log c}{I(i)}, \quad as \ c \to 0. \quad (21)$$

The lower bounds provided by Theorem 2 hold for both the distributed setup and the fusion center based setup.

B. Decentralized Chernoff Test

We now provide upper bounds on the performance of DCT. In the following theorems, $C$ indicates the communication overhead, namely the number of times a node communicates with the fusion center. The superscript $D$ refers to the DCT.

Part (i) of Theorem 2 states that the probability of making a wrong decision can be made as small as desired by an appropriate choice of the observation cost $c$. Part (ii) provides a bound on the expected time to reach the final decision, and part (iii) bounds the risk as an immediate consequence of parts (i) and (ii).

**Theorem 3.** (Direct). The following statements hold:

(i) For all $c \in (0, 1)$ and all $i \in [M]$, the probability that the DCT takes an incorrect decision is

$$P^\delta_{D}(\hat{H} \neq h_i) \leq \min\{(M - 1)c, 1\}.$$

(ii) For all $\ell \in [L]$ and $i, j, \kappa \in [M]$, if $E\left[\log p_{i,\ell}^{u,\ell}(Y)/p_{j,\ell}^{u,\ell}(Y)\right]^2 < \infty$, then the expected decision time is

$$E^D_i[N] \leq \left(1 + o(1)\right) \frac{\log c}{I(i)}, \quad as \ c \to 0. \quad (23)$$

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

$$E^D_i \leq \left(1 + o(1)\right) \frac{c \log c}{I(i)}, \quad as \ c \to 0. \quad (24)$$

(iv) For all $\ell \in [L]$ and $i, j, \kappa \in [M]$ and all integers $r \geq 2$, if $E\left[\log p_{i,\ell}^{u,\ell}(Y)/p_{j,\ell}^{u,\ell}(Y)\right]^r < \infty$, then the $r^{th}$ moment of the decision time $N$ is

$$E^D_i[N^r] \leq \left(1 + o(1)\right) \frac{c \log c}{I(i)}^r, \quad as \ c \to 0. \quad (25)$$

In the above theorem, the bound on the expected decision time in (ii) requires the second moment of the log-likelihood ratio to be finite. In contrast, for all $r \geq 2$, the bound on the $r^{th}$ moment of the decision time requires only the $r^{th}$ moment (rather than the $(r + 1)^{th}$ moment) of the log-likelihood ratio to be finite.

The next result is a consequence of Theorems 2 and 3. It shows the asymptotic optimality of the DCT, and presents the expected communication overhead, as $c \to 0$.

**Theorem 4.** The following statements hold:

(i) For all $\ell \in [L]$ and $i, j, \kappa \in [M]$, if $E\left[\log p_{i,\ell}^{u,\ell}(Y)/p_{j,\ell}^{u,\ell}(Y)\right]^2 < \infty$, then the expected decision time is

$$E^D_i[N] = \left(1 + o(1)\right) \frac{\log c}{I(i)}, \quad as \ c \to 0. \quad (26)$$

Additionally,

$$E^D_i = \left(1 + o(1)\right) \frac{c \log c}{I(i)}, \quad as \ c \to 0. \quad (27)$$

(ii) For all $\ell \in [L]$ and $i, j, \kappa \in [M]$ and all integers $r \geq 2$, if $E\left[\log p_{i,\ell}^{u,\ell}(Y)/p_{j,\ell}^{u,\ell}(Y)\right]^r < \infty$, then the $r^{th}$ moment of the decision time is

$$E^D_i[N^r] = \left(1 + o(1)\right) \frac{c \log c}{I(i)}^r, \quad as \ c \to 0. \quad (28)$$

(iii) The expected communication overhead is

$$\lim_{c \to 0} E^D_i[C] = 4. \quad (29)$$

Combining Theorem 3 and Theorem 4, it follows that DCT is asymptotically optimal as the observation cost tends to zero. Since the probability of error tends to zero as $c \to 0$, DCT minimizes the expected decision time required to reach the correct decision over the network.

It is worth noting that the performance of DCT depends only on the cumulative capability $I(i)$ of the network to detect hypothesis $h_i$, and is independent of how the capabilities $\nu_{i,\ell}$ are distributed over the network. If two networks have the same cumulative capabilities, then the minimum expected decision time will be the same for both of them. These results hold irrespective of the number of nodes in the network.
C. Consensus-based Chernoff Test

We now provide upper bounds on the performance of CCT. Let us recall the definition of the ergodic coefficient \( \eta(W) \) of matrix \( W \) appearing in (10), which is

\[
\eta(W) = \min_{i \neq j} \sum_{k=1}^{L} \min\{w_{i,k}, w_{j,k}\}.
\]

We use the following lemma to provide our performance guarantees.

**Lemma 5.** [46, Proposition 1] If the network is connected then \( 0 < \eta(W_{i,j}) < 1 \).

In the following theorems, the superscript \( C \) is appended to quantities that refer to the CCT. Part (i) of Theorem 6 states that the probability of making a wrong decision can be made as small as desired by an appropriate choice of \( c \). Part (ii) provides a bound on the expected time to reach the final decision, and part (iii) bounds the risk as an immediate consequence of parts (i) and (ii). Finally, part (iv) presents the bound on the higher moments of the decision time \( N \) of CCT.

**Theorem 6.** (Direct). The following statements hold:

(i) For all \( c \in (0,1) \) and all \( i \in [M] \), the probability that CCT takes an incorrect decision is

\[
\mathbb{P}_{i}^{c}(H \neq h_{i}) \leq \min \left\{ (M-1)e^{-\frac{c}{1-\eta(W_{i,j})}}, 1 \right\},
\]

(ii) For all \( \ell \in [L], \ i, j, \kappa \in [M] \), if \( \mathbb{E} \left[ \log \frac{u_{i,j}^{\ell}(Y)}{p_{j,k}^{\ell}(Y)} \right]^{2} < \infty \), then the expected decision time is

\[
\mathbb{E}_{i}^{c}[N] \leq (1 + o(1)) \max \left\{ \frac{h^{\mathbb{G}} \cdot \log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(W_{i,j})^{h^{\mathbb{G}}})}, \frac{\log c}{I(i) - c} \right\},
\]

as \( c \to 0 \).

(iii) Combining (i) and (ii), the risk defined in (11) is

\[
\mathbb{R}_{i}^{c} \leq (1 + o(1)) \max \left\{ \frac{h^{\mathbb{G}} \cdot \log(1/\max_{j \in [L]} I(j))}{\log(1 - \eta(W_{i,j})^{h^{\mathbb{G}}})}, \frac{1}{I(i) - c} \right\} \cdot c|\log c|, \quad \text{as } c \to 0.
\]

(iv) For all \( \ell \in [L], \ i, j, \kappa \in [M] \) and all integers \( r \geq 2 \), if \( \mathbb{E} \left[ \log \frac{u_{i,j}^{\ell}(Y)}{p_{j,k}^{\ell}(Y)} \right]^{r} < \infty \), then the \( r \)-th moment of the expected decision time is

\[
\mathbb{E}_{i}^{c}[N^{r}] \leq (1 + o(1)) \max \left\{ \frac{h^{\mathbb{G}} \cdot \log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(W_{i,j})^{h^{\mathbb{G}}})}, \frac{\log c}{I(i) - c} \right\}^{r},
\]

as \( c \to 0 \).

The following result is a consequence of Theorems 2 and 6 and establishes the asymptotic optimality of CCT, under the condition that the maximum in (30), (31) and (32) is \( |\log c|/(I(i) - c) \).

**Theorem 7.** If the maximum in (30), (31) and (32) is \( |\log c|/(I(i) - c) \), then the following statements hold:

(i) For all \( \ell \in [L], \ i, j, \kappa \in [M] \), if \( \mathbb{E} \left[ \log \frac{u_{i,j}^{\ell}(Y)}{p_{j,k}^{\ell}(Y)} \right]^{2} < \infty \), then the expected decision time is

\[
\mathbb{E}_{i}^{c}[N] = (1 + o(1)) \frac{|\log c|}{I(i) - c}, \quad \text{as } c \to 0.
\]

Additionally, the risk is

\[
\mathbb{R}_{i}^{c} = (1 + o(1)) \frac{c|\log c|}{I(i) - c}, \quad \text{as } c \to 0.
\]

(ii) For all \( i \in [M] \) and all integers \( r \geq 2 \), if \( \mathbb{E} \left[ \log \frac{u_{i,j}^{\ell}(Y)}{p_{j,k}^{\ell}(Y)} \right]^{r} < \infty \), then the \( r \)-th moment of the expected decision time is

\[
\mathbb{E}_{i}^{c}[N^{r}] = (1 + o(1)) \left( \frac{|\log c|}{I(i) - c} \right)^{r}, \quad \text{as } c \to 0.
\]

Let \( \mathbb{G}^{2} \) be the number of spanning trees of \( G(\mathbb{L}, \mathbb{E}) \) of height \( h^{\mathbb{G}} \), and let \( W \) denote the smallest nonzero entry of \( W \), namely \( W = \min\{w_{i,j} : w_{i,j} > 0\} \). The following lemma provides three sufficient conditions for which the maximum in (30), (31) and (32) is \( |\log c|/(I(i) - c) \).

**Lemma 8.** If, for all \( i \in [M] \) and \( c \to 0 \), at least one of the following inequalities holds:

\[
I(i) \left| \log \left( \max_{j \in [M]} I(j) \right) \right| \leq \frac{\left| \log \left( 1 - \eta(W^{h^{\mathbb{G}}}) \right) \right|}{h^{\mathbb{G}}},
\]

\[
I(i) \left| \log \left( \max_{j \in [M]} I(j) \right) \right| \leq \frac{\left| \log \left( 1 - a^{\mathbb{G}} \cdot W^{h^{\mathbb{G}} / a^{\mathbb{G}}} \right) \right|}{h^{\mathbb{G}}},
\]

then the maximum in (30) is \( |\log c|/(I(i) - c) \).

We now briefly discuss the physical significance of the sufficient conditions presented in Lemma 8. The decision time of CCT (see (30), (31) and (32)) depends on two terms: \( A_{1} \) and \( A_{2} \), where

\[
A_{1} = \frac{h^{\mathbb{G}} \cdot \log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(W^{h^{\mathbb{G}}}))},
\]

\[
A_{2} = \frac{|\log c|}{I(i) - c}.
\]

Here, \( A_{1} \) corresponds to the expected time of the initialization phase of CCT. Since this phase performs consensus over the network, the expected time \( A_{1} \) is dependent on the network parameters \( h^{\mathbb{G}} \) and matrix \( W \). Likewise, \( A_{2} \) corresponds to the test phase of CCT where the Chernoff test is performed independently at all the nodes. Hence, similar to the test phase of DCT, the expected time \( A_{2} \) of this phase is independent of the
network parameters. Since the decision phase of CCT begins only after the termination of both the initialization phase and the test phase, \( \max\{A_1, A_2\} \) controls the expected decision time of CCT (see Theorem 6). The sufficient conditions in Lemma 8 ensure that the consensus in the initialization phase is reached before the triggering condition (17) in the test phase. In other words, consensus along the network should be faster than the time required by the Chernoff test to accumulate sufficient information to take a decision. Given that the initialization phase does not form a bottleneck for the test phase, CCT is asymptotically optimal as \( c \to 0 \).

As a final remark, note that the fusion center based setup can be considered a special case of the distributed setup. In the fusion center based setup, the cumulative capability vector \( I \) can be estimated (with no error) in two time steps at all the nodes, i.e., for \( n = 2 \) and \( \ell, j \in [L] \), the equivalent of (15) is

\[
|\hat{I}_\ell^c - \hat{I}_j^c| \leq 0 \cdot 1_{1 \times M},
\]

and is independent of \( c \). In the regime of vanishing cost \( c \to 0 \), \( A_2 \) dominates \( A_1 \), therefore \( \max\{A_1, A_2\} = A_2 \), which implies the asymptotic optimality of DCT.

VIII. NUMERICAL RESULTS

In this section, we evaluate the performance of both DCT and CCT by simulations, and compare the results to the theoretical bounds presented in the previous section. The performance of these tests is evaluated for different sizes of networks. In our experiments, the number of hypotheses is \( M = 3 \). The probability distribution \( p_{i,j}^{w_{i,j}} \) is Bernoulli with parameter \( p \), which is selected uniformly at random from \( (0, 1/3), (1/3, 2/3) \) and \( (2/3, 1) \) for \( i = 1, 2 \) and \( 3 \) respectively.

Figure 2 shows the risk of DCT in a fusion center based setup, as obtained by simulations. Figure 3 shows the corresponding value of the risk, as predicted by Theorem 3. The risk decreases as the observation cost \( c \) decreases. This is because the threshold in the triggering condition (8) increases, which ensures that the nodes have a greater confidence about their local decision. On the other hand, the risk decreases by increasing the number of sensors \( L \), because the cumulative capability of the network to detect the hypothesis, defined in (7), increases with \( L \), and the task of hypothesis testing is divided among a larger number of sensors. Hence, the final decision can be reached more quickly, and this decreases the risk. The trends are in agreement with the theoretical results obtained for DCT.

Our simulations also confirm the prediction that, on the average, only four communication rounds are required, per single sensor, see (29) in Theorem 4. The results of these simulations are not reported here for the sake of brevity. We only mention that, on rare occasions, for individual realizations it may happen that the number of communication rounds is substantially larger than four—a manifestation of the long-run phenomenon [47] p. 110. In practice, this can be remedied by resorting to a truncated version of the sequential test, for which the maximum number of probing actions is fixed, see [48], [49] and references therein, and see [50] for a simple implementation of truncation. A precise analysis of DCT using truncated tests is out of the scope of the present paper.

The performance of CCT is evaluated for two network topologies. In the first topology, given the number of network nodes \( L, \lfloor L/2 \rfloor \) sensors are connected to form a ring, and the remaining sensors are randomly connected to the sensors in the ring. An example of network with \( L = 10 \) is shown in Figure 1. In this case, the spanning height of the tree is in \( L \). In the second topology, given the number of network nodes \( L \), the nodes are connected to form a binary tree. In this case, the spanning height of the tree is \( O(\log_2 L) \).

Figure 4 shows the performance of CCT for the ring with random attachments, obtained by computer simulations. Figure 5 shows the value of risk according to Theorem 6. Like in the case of DCT, the risk of CCT decreases as the observation cost \( c \) decreases. Instead, the behavior as function of \( L \) is different. Unlike DCT, the risk of CCT increases by increasing the number of network nodes \( L \). This effect can be explained by observing that in CCT there is a trade-off between the time required by the initialization phase and the time required by the test phase. For the considered network \( G \) and consensus matrix \( W \), as the number of nodes \( L \) increases, the consensus scheme in the initialization phase will require more time in comparison to the test phase. Additionally, the time required by the test phase decreases with \( L \), for the same reasons as in the DCT case. Figures 6 and 7 show that the consensus between the sensors in the first phase of CCT becomes the dominating factor in the decision time. This is in agreement with the theoretical bounds provided in Theorem 6 because in this topology the maximum in the right-hand side of (30) is given by the first term, which corresponds to the time for the initialization phase, as detailed in the comments following Lemma 8.

Figures 6 and 7 show the performance of CCT for the tree topology, via simulations and using the theoretical predictions of Theorem 6 respectively. The risk of CCT decreases as \( c \) decreases. Unlike the ring topology with random attachments, the risk decreases by increasing \( L \) until \( L = 15 \), and then
increases. In this setup, for the initial values of $L$, the time required by the test phase is larger than the time for the initialization phase. Theoretically, the maximum at the right-hand side of (30) is given by the second term, which corresponds to the time for the test phase. On the contrary, for $L = 20$, the time of the initialization phase becomes dominant, which explains why the risk increases with $L$. Finally, comparing Figures 6 and 7, we see that the theoretical values of the risk are close to the results of numerical simulations.

**IX. Summary**

Networked sensor systems are ubiquitous solutions for inference problems in view of their improved robustness, scalability, versatility, performance, with respect to the classic centralized architectures. Initial implementations were based on inexpensive small sensors, with extremely limited hardware/software capabilities. Progressively, these devices acquired more and more functionalities, and are nowadays capable of active sensing, namely they can adapt the probing signal on the basis of previous measurements, in order to optimize their sensing capability. Thus, individual sensors have become intelligent devices which continuously learn from the past and decide their future actions, in a closed-loop adaptative scheme.

We considered two network configurations of these intelligent sensors: fusion center based setup and distributed setup. In the first configuration, the centralized entity (fusion center) coordinates the actions of the remote nodes, and takes the final decision. The second configuration does not have a central coordination, and all the processing takes place at the nodes: they actively collect measurements, exchange information with immediate neighbors, and collectively take a decision.
For the first configuration (fusion center based setting) we proposed a sequential active/adaptive decision system — referred to as DCT — which operates in two phases. First, there is a round of communication between the central entity and the remote nodes, needed to define the relative capability of each node to detect the hypotheses. This capability is then used to apportion the whole decision task among the nodes. Then, each node begins to continuously sense the environment of each node to detect the hypotheses. This capability is then referred to as DCT — which operates in two phases. First, we provided a sequential active/adaptive decision system —

For the second configuration (distributed setting), we exploited ideas from the DCT implementation, combined with gossip protocols that use consensus techniques, to design a fully distributed active/adaptive sequential decision system, which is referred to as CCT. Our approach is markedly different from those usually exploited in the literature, where real-valued belief vectors are continuously exchanged over the network to reach a consensus. Our implementation is, by design, more parsimonious in terms of communication.

CCT works in three phases. In the first phase, a consensus about the relative capability of the nodes to detect the state of nature is achieved by means of gossip protocols with local information exchange. In the second phase, nodes implement the Chernoff test and, once all the network nodes reach the triggering condition \( \ell \), the final decision is reached in a distributed way in the third phase of operation, by diffusing messages across the system that percolate the information of whether the other sensors have terminated their share of task.

We proved asymptotic optimality of CCT, under certain conditions. These depend on relative balancing between the time for the first and for the second phase of operation of CCT: optimality is retained when the time to reach a consensus (first phase) does not overcome the time for the second phase. The trade-off between expected times for the first two phases of operation is ruled by the consensus matrix \( W \) and by the network topology \( G \), as formalized in Theorem 6 and Lemma 8.

Appendix A

Proof of Theorem 2

Proof. Let \( H^* = h_i \) be the true hypothesis. The proof of Theorem 2 consists of two parts. First, for all hypotheses \( h_m \neq h_i \) and \( 0 < \epsilon < 1 \), we show that for the probability of error to be close to zero, the log-likelihood ratio should be in favor of \( h_i \), and should be greater than \(-(1 - \epsilon) \log c\) with high probability as \( c \to 0 \). Namely, the inequality

\[
S^N(h_i, h_m) = \sum_{\ell=1}^{L} \sum_{k=1}^{N} \log \frac{p_{h_i,\ell}(y_{k,\ell})}{p_{h_m,\ell}(y_{k,\ell})} \geq -(1 - \epsilon) \log c \quad (40)
\]

must hold with high probability, as \( c \to 0 \). Second, we show that for all \( 0 < \epsilon < 1 \) and \( n = -(1-\epsilon) \log c / I(\epsilon) \), it is unlikely that such inequality is not satisfied for some hypothesis \( h_m \neq h_i \).

We start by defining two sets of hypotheses \( \mathcal{H}_0 = \{ h_i \} \) and \( \mathcal{H}_1' = \{ h_m \}_{m \neq i} \). By (20), both type I and type II errors of the hypothesis test \( \mathcal{H}_0 \) vs. \( \mathcal{H}_1' \) are \( O(\epsilon \log c) \). Thus, by [14, Lemma 4], for all hypotheses \( h_m \neq h_i \) and \( 0 < \epsilon < 1 \), we have

\[
\mathbb{P}_i \left( S^N(h_i, h_m) \leq -(1 - \epsilon) \log c \right) = O(-c^\epsilon \log c). \quad (41)
\]

Fig. 6. Performance of CCT for the tree topology: risk vs. cost \( c \) for different number of sensors \( L \).

Fig. 7. Performance of CCT according to Theorem 6 for the tree topology: risk vs. cost \( c \) for different number of sensors \( L \).
Therefore, as \( c \to 0 \), the probability in (41) tends to 0, which concludes the first part of the proof.

Now, we show that for all \( \epsilon > 0 \), we have

\[
\lim_{n' \to \infty} \mathbb{P} \left( \max_{1 \leq n \leq n'} \min_{m \neq i} S^n(h_i, h_m) \geq n'(I(i) + \epsilon) \right) = 0.
\]

(42)

We have

\[
S^n(h_i, h_m) = \sum_{\ell = 1}^L \sum_{k = 1}^n \left( \log \frac{p_{i,\ell}^{u_k,\ell}(y_k,\ell)}{p_{m,\ell}^{u_k,\ell}(y_k,\ell)} - D(p_{i,\ell}^{u_k,\ell} || p_{m,\ell}^{u_k,\ell}) \right) + \sum_{\ell = 1}^L \sum_{k = 1}^n D(p_{i,\ell}^{u_k,\ell} || p_{m,\ell}^{u_k,\ell}) = M_1^n + M_2^n,
\]

where

\[
M_1^n = \sum_{\ell = 1}^L \sum_{k = 1}^n \left( \log \frac{p_{i,\ell}^{u_k,\ell}(y_k,\ell)}{p_{m,\ell}^{u_k,\ell}(y_k,\ell)} \right),
\]

and

\[
M_2^n = \sum_{\ell = 1}^L \sum_{k = 1}^n D(p_{i,\ell}^{u_k,\ell} || p_{m,\ell}^{u_k,\ell}).
\]

Then, for all \( 1 \leq n \leq n' \), we have

\[
\min_{m \neq i} M^n_{m} \leq \sum_{\ell = 1}^L \sum_{k = 1}^n \left( \log \frac{p_{i,\ell}^{u_k,\ell}(y_k,\ell)}{p_{m,\ell}^{u_k,\ell}(y_k,\ell)} \right) \leq \sum_{\ell = 1}^L \sum_{k = 1}^n v_{i,\ell}
\]

(45)

\[
\leq nI(i)
\]

(b)

\[
\leq n'I(i),
\]

(c)

where (a) follows from the definition of \( v_{i,\ell} \) in (5), (b) follows from the definition of \( I(i) \) in (7), and (c) follows from \( n \leq n' \). Now, if the event in (42) occurs for a fixed \( n_1 \), i.e.

\[
\min_{m \neq i} (M_1^{n_1} + M_2^{n_1}) \geq n'(I(i) + \epsilon),
\]

then there exists a hypothesis \( h_m \) such that \( M_1^{n_1} \geq n'\epsilon \). Thus, there exists a constant \( K' > 0 \) such that the probability in (42) becomes

\[
\mathbb{P} \left( \max_{1 \leq n \leq n'} \min_{m \neq i} S^n(h_i, h_m) \geq n'(I(i) + \epsilon) \right) \leq \sum_{m \neq i} \mathbb{P} \left( \max_{1 \leq n \leq n'} M_1^n \geq n'\epsilon \right)
\]

(46)

\[
\leq \frac{\epsilon}{n'} (M - 1) K',
\]

where (a) follows from the fact \( M_1^n \) is a martingale with mean zero and using the Doob-Kolmogorov extension of Chebyshev’s inequality [5]. Thus, (42) follows. As discussed in [14] Theorem 2, we have

\[
\mathbb{P}(N' \leq n_0 \leq N) \leq \mathbb{P}(\forall m \neq i : N \leq n_0 \text{ and } S^n(h_i, h_m) \geq n_0(I(i) + \epsilon)) + \mathbb{P}(\exists m \neq i : S^n(h_i, h_m) \leq n_0(I(i) + \epsilon)) + \mathbb{P}(\exists m \neq i : S^n(h_i, h_m) \leq n_0(I(i) + \epsilon)).
\]

The first and the second terms at the right-hand side of (47) approach zero by (42) and (41) respectively. Now, using (47), we also have

\[
\mathbb{P}(N' \leq n_0) = \mathbb{P}(N \leq n_0) \to 0,
\]

(48)

as \( c \to 0 \). The statement of the theorem now follows. \( \square \)

APPENDIX B
PROOFS FOR DCT AND CCT

A. Proof of Theorem 3

Proof. To prove Theorem 3, we need some additional notation. Let \( A_{n,j} \) be the set of sample paths where the decision made by the fusion center is in favor of \( h_j \) at the \( n^{th} \) step, and we indicate a single sample path as \( \{ (u^n_j, y^n_j) : (u^n_j, y^n_j) \} \). We indicate by \( A_{n,j} \) the set of sample paths in \( A_{n,j} \) corresponding to the \( \ell^{th} \) node. Finally, we define

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

The proof consists of two parts. First, we write \( \mathbb{P}_i^\mathcal{I}(H \neq h_i) \) as the probability of a countable union of disjoint sets of sample paths. An upper bound on this probability then follows from an upper bound on the probability of these disjoint sets, in conjunction with the union bound. Second, we upper bound \( \mathbb{E}_i^\mathcal{I}[N] \) by the sum of the expected time required to reach the triggering condition (8) at node \( \ell \) for \( H^* = h_i \), and the expected delay between the time of triggering and the time when the final decision is taken in favor of hypothesis \( h_i \) at the fusion center. We then show that these expectations are the same at all the nodes, so that (26) follows.

Consider the probability \( \mathbb{P}_i^\mathcal{I}(H = h_j) \). This is the same as the probability of the countable union of disjoint sets \( A_{n,j} \).
Thus, for all \( j \neq i \), we can write
\[
P_i^j(A_{n,j}) = \int_{A_{n,j}} \prod_{\ell=1}^{L} \prod_{k=1}^{n} p_{i,\ell}^{u_k,\ell}(y_{\ell,k}) \, dy_1,\ell(u_{1,\ell}) \ldots dy_n,\ell(u_{n,\ell})
\]
\[
= \prod_{\ell=1}^{L} \int_{A_{n,\ell}} \prod_{k=1}^{n} p_{i,\ell}^{u_k,\ell}(y_{\ell,k}) \, dy_1,\ell(u_{1,\ell}) \ldots dy_n,\ell(u_{n,\ell})
\]
\[
= \prod_{\ell=1}^{L} \int_{A_{n,\ell}} \prod_{k=1}^{n} p_{i,\ell}^{u_k,\ell}(y_{\ell,k}) \, dy_1,\ell(u_{1,\ell}) \ldots dy_n,\ell(u_{n,\ell})
\]
which is independent of \( \ell \). Using (52), we also have that for all \( \epsilon > 0 \) and \( n \geq (1 + \epsilon)|\log(c)/I(i)| \),
\[
P_i\left( \max_{1 \leq \ell \leq L} N_{i,\ell} \geq n \right) \leq \sum_{\ell=1}^{L} P_i(N_{i,\ell} \geq n),
\]
\[
\leq L K_i e^{-b_i n},
\]
where \( K_i = \min_j K_{i,j} \) and \( b_i = \min_j b_{i,j} \). For all \( r \geq 1 \), we have the bound on the \( r \)th moment of \( \max_{1 \leq \ell \leq L} N_{i,\ell} \), i.e.,
\[
E_i \left[ \left( \max_{1 \leq \ell \leq L} N_{i,\ell} \right)^r \right] \leq \left( 1 + o(1) \right) \frac{\log(c)}{I(i)}^r.
\]

Now, we bound the higher moments of \( \tau(N_{i,\ell}) \). Let \( N^* \) be the time instance such that for all \( n \geq N^* \), the local decision \( \hat{H} \) at node \( \ell \) is \( h^* \), i.e., \( \hat{H} = h^* \). Using [14, Lemma 1], there exists \( K > 0 \) and \( b > 0 \) such that
\[
P_i(N^* \geq n) \leq k \exp(-bn),
\]
which implies \( P_i(N^* < \infty) = 1 \). Then, node \( \ell \) follows time \( N^* \) selects the actions in an i.i.d. fashion according to the probability mass function given by (3).

Let \( G_{n,\ell} \) be the joint cumulative distribution function of the variables \( (y_{n,\ell}, u_{n,\ell}) \) at round \( n \) and node \( \ell \) for the Chernoff test. Also, let \( F_c \) be the joint cumulative distribution function of \( (y_{n,\ell}, u_{n,\ell}) \) under the true hypothesis \( h^* \) where the actions are selected according to \( Q_i^h \), (see (3)) at each round at sensor \( \ell \). Then, for all \( n \geq N^* \), \( G_{n,\ell} < F_c \). Since \( P_i(N^* < \infty) = 1 \), it follows that the distribution \( G_{n,\ell} \) converges to \( F_c \).

Given that for all \( n, (y_n, u_n) \sim F_c \) are i.i.d. random variables, we have that
\[
E_i \left[ \log(p_{i,\ell}^{u_k,\ell}(y_{\ell,k})/\max_{j \neq i} p_{j,\ell}^{u_k,\ell}(y_{\ell,k})) \right] = v_{i,\ell} > 0.
\]
Additionally, using (57), finiteness of the \( r \)th moment of log-likelihood ratio for \( r \geq 2 \), and by Corollary 10.1 and Lemma 11 in Appendix C, we have that
\[
E_i[\tau(N_{i,\ell})]^r < \infty,
\]
where the expectation is with respect to \( F_c \).

We now note that (58) also holds when the expectation is with respect to \( G_{n,\ell} \). To show this claim, first observe that \( E_i[\tau(N_{i,\ell})]^r \) is upper bounded by the two terms at the right hand side of (56) in Corollary 10.1. The first term is bounded, since the KL divergence between any two probability measures is finite. The second term can be split into two summations, one for \( 1 \leq n \leq N^* \), and the other for \( n \geq N^* + 1 \). The first summation is finite since \( N^* < \infty \) a.s., and the probability is at most one. By using Lemma 11 in Appendix C and \( G_{n,\ell} < F_c \), the second summation is also finite. It follows that (58) holds for the Chernoff test.

Since \( E_i[\log(p_{i,\ell}^{u_k,\ell}(y_{\ell,k})/\max_{j \neq i} p_{j,\ell}^{u_k,\ell}(y_{\ell,k}))]^2 \) is finite, using (58), the term \( E_i[\tau(N_{i,\ell})]^r \) on the right-hand side of (51) is finite and independent of \( c \). Now, combining equation (51), (55) and the finiteness of \( E_i[\tau(N_{i,\ell})] \), as \( c \to 0 \) we get (26). Thus, part (iii) of the theorem is proved.
Now,
\[ \mathbb{E}_i^D [N^r] \leq \mathbb{E}_i \left[ \max_{1 \leq i \leq L} \sum_{\ell \in [L]} \tau(N_{i,\ell}) + 1 \right]. \] (59)
The moments of \( \sum_{\ell \in [L]} \tau(N_{i,\ell}) \) are finite and independent of \( c \). Hence, the dominant term, dependent on \( c \), in the expansion of the right-hand side of (59) is given only by \( \max_{1 \leq i \leq L} N_{i,\ell} \). It follows that using (58) and (55), as \( c \to 0 \), we have
\[ \mathbb{E}_i^D [N^r] \leq \left( 1 + o(1) \right) \frac{\log c}{I(i)^r}, \] (60)
which proves part (iv) of the theorem.

B. Proof of Theorem 2
Proof. Combining Theorems 2 and 3 (55) and (54) follow immediately. We then turn to the proof of (29).

For all \( \ell \in [L] \), given that hypothesis \( h_1 \) is true, we have that as \( c \to 0 \) the probability of incorrect detection tends to zero. It follows that \( \bar{H} = h_1 \) and
\[ \mathbb{E}_i^D [N] = (1 + o(1)) \frac{\log c}{I(i)^2}, \] (61)
where the last equality follows from (55). Thus, as \( c \to 0 \) all the nodes reach the same local decision on average at the same time, and the average number of messages that each node sends to the fusion center to communicate this local decision is one. It follows that as \( c \to 0 \) the expected communication overhead is four: two in the initialization phase, one to communicate the local decision, and one to receive the halting message.

C. Proof of Theorem 3
Proof. Let \( B_{n,j} \) be the set of sample paths where the final decision \( \bar{H} \) is initiated in favor of \( h_j \) at the \( n \)th step, and we indicate a single sample path as \( \{(y_{1}^{u}, y_{2}^{v}, \ldots, y_{L}^{v})\} \). We indicate by \( B_{n,j,\ell} \) the set of sample paths in \( B_{n,j} \) corresponding to the \( \ell \)th node. \( N^c \) denotes the time taken to terminate the initialization phase of CCT. Now, we define the two times associated with the test phase of CCT:
\[ T_{i,\ell} = \inf \left\{ n : \sum_{k=1}^{n} \log \frac{p_{i,\ell}(y_{k,\ell})}{\max_{j \neq i} p_{j,\ell}(y_{k,\ell})} \geq \rho_{i,\ell}(N^c) \log c \right\}, \]
and
\[ \tau(T_{i,\ell}) = \sup \left\{ n : \sum_{k=T_{i,\ell}+1}^{n} \log \frac{p_{i,\ell}(y_{k,\ell})}{\max_{j \neq i} p_{j,\ell}(y_{k,\ell})} \leq 0 \right\}. \]
The proof consists of two parts. First, we write \( \mathbb{P}^C_i(\bar{H} \neq h_i) \) as the probability of a countable union of disjoint sets of sample paths. An upper bound on this probability then follows from an upper bound on the probability of these disjoint sets, in conjunction with the union bound. Second, \( \mathbb{E}_i^D [N] \) is dependent on the time required to reach and detect the consensus during the initialization phase, the time required to reach the triggering condition (17) in the test phase, and the time required to reach and detect that the nodes have reached a common decision about a hypothesis in the decision phase. Since initialization and test phases are performed in parallel followed by the decision phase, the stopping time \( N \) can be bounded as
\[ N \leq \max \left\{ N^c, \max_{1 \leq \ell \leq L} (T_{i,\ell} + \tau(T_{i,\ell})) \right\} + N^s, \] (62)
where \( N^s \) is the time taken during the decision phase of CCT. Consider the probability \( \mathbb{P}^C_i(\bar{H} = h_j) \). This is the same as the probability of the countable union of disjoint sets \( B_{n,j} \). Thus, for \( j \neq i \), we can write
\[ \mathbb{P}^C_i(B_{n,j}) = \int_{B_{n,j}} \prod_{\ell=1}^{L} \prod_{k=1}^{n} p_{i,\ell}(y_{k,\ell}) dy_{1,\ell}(u_{1,\ell}) \ldots dy_{n,\ell}(u_{n,\ell}) \]
\[ = \left( a \right) \int_{B_{n,j}} \prod_{\ell=1}^{L} \prod_{k=1}^{n} p_{i,\ell}(y_{k,\ell}) dy_{1,\ell}(u_{1,\ell}) \ldots dy_{n,\ell}(u_{n,\ell}) \]
\[ \leq \left( b \right) \prod_{\ell=1}^{L} \int_{B_{n,j}} e^{c(i)} \prod_{k=1}^{n} p_{j,\ell}(y_{k,\ell}) dy_{1,\ell}(u_{1,\ell}) \ldots dy_{n,\ell}(u_{n,\ell}) \]
\[ \leq e^{c(i)/(I(i)-c)} \int_{B_{n,j}} e^{c(i)} \prod_{\ell=1}^{L} e^{c(i)}(\bar{H} = h_j) \text{ at sample } n \text{ at } \ell \text{th sensor} \]
\[ = e^{c(i)/(I(i)-c)} \mathbb{P}_j^C(\bar{H} = h_j) \text{ at sample } n, \] (63)
where (a) follows from the definition of \( B_{n,j,\ell} \); (b) follows from the definition of \( T_{i,\ell} \); (c) follows from \( \sum_{\ell=1}^{L} p_{j,\ell}^{(n)} \leq I(i)/(I(i)-c) \). Now, we can bound \( \mathbb{P}^C_i(\bar{H} \neq h_i) \) as follows
\[ \mathbb{P}^C_i(\bar{H} \neq h_i) = \sum_{j \neq i} \mathbb{P}^C_i(\bar{H} = h_j) = \sum_{j \neq i} \sum_{n=1}^{\infty} \mathbb{P}_j^C(B_{n,j}) \]
\[ \leq \sum_{j \neq i} \sum_{n=1}^{\infty} e^{c(i)/(I(i)-c)} \mathbb{P}_j^C(\bar{H} = h_j) \text{ at sample } n \]
\[ = e^{c(i)/(I(i)-c)} (M - 1), \] (64)
where the inequality in the chain follows by (63). This proves part (i) of the theorem.

Let us bound the time required to terminate the initialization phase, i.e., \( N^c \). Since matrix \( W \) in (10) is row stochastic (see (13)) and the graph \( G(N,E) \) is connected, the ergodic coefficient \( \eta(W) \in (0,1) \) (see Lemma 5). It follows from (46) that for all \( k, n \in \mathbb{N} \) and \( \ell, j \in [L] \), we have
\[ e_{k,j}^{k+n} \leq (1 - \eta(W^n)) e_{k,j}^k, \]
where \( e_{k,j}^k = |\tilde{I}_{\ell} - I_{\ell}^j| \). Now, if the initialization phase reaches uniformly local \( c/L^2 \)-consensus at time instance \( k_0 \), then for all \( \ell, j \in [L] \), we have (see (46))
\[ e_{k,j}^{k_0} \leq c \cdot 1_{1 \times M}. \]
Thus, there exists \( k' \in \mathbb{N} \) such that \( h^G \cdot k' \leq k_0 \leq h^G \cdot (k' + 1) \).

Using \( (65) \), for all \( \ell, j \in [L] \), we have
\[
e^{k_0}_{\ell,j} \leq e^{G}_{\ell,j} \cdot k' \leq \left( 1 - \eta(h^G) \right)^{k'} e^{0}_{\ell,j} \leq \left( 1 - \eta(h^G) \right)^{k'} I,
\]
where (a) follows from \( \tilde{h}^G \cdot k' = h^G \cdot \tilde{h}^G \cdot (k' - 1) \) and Lemma [5] and (b) follows from the fact that for all \( \ell, j \in [L] \), \( e^{0}_{\ell,j} \leq I \). Since for all \( \ell, j \in [L] \), \( e^{0}_{\ell,j} \leq c \cdot 1 \times M \), using \( (66) \) we have
\[
1 - \eta(h^G) \leq c,
\]
\[
k' \leq \frac{\log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(h^G))}.
\]

Since \( k_0 \leq h^G(k' + 1) \), we have
\[
k_0 \leq h^G \left( \frac{\log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(h^G))} + 1 \right).
\]

Now, let \( k_d \) be the time to detect the uniformly local \( c/L^2 \)-consensus. From \( (45) \), we have
\[
k_d \leq h^G \left( \frac{-\log(d^G)}{\log(1 - \eta(h^G))} + 1 \right) + L + 1.
\]

Now, the time \( N^c \) for initialization phase is bounded as follows
\[
N^c \leq k_0 + k_d
\]
\[
\leq h^G \left( \frac{\log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(h^G))} + 1 \right) + h^G \left( \frac{-\log(d^G)}{\log(1 - \eta(h^G))} + 1 \right) + L + 1.
\]

The expected time of the test phase of CCT is bounded above by \( \mathbb{E}_i \left[ \max_{1 \leq \ell \leq L} (T_{i,\ell} + \tau(T_{i,\ell})) \right] \). Using \( (55) \) and \( (58) \) of Theorem [3] as \( c \to 0 \), we have
\[
\mathbb{E}_i \left[ \max_{1 \leq \ell \leq L} (T_{i,\ell} + \tau(T_{i,\ell})) \right] \leq \frac{\left| \log c \right|}{I(i) - c} (1 + o(1)).
\]

Now, we compute the time for the decision phase of CCT. The network will reach the final decision for all \( n > \max_{1 \leq \ell \leq L} \tau(T_{i,\ell}) + k_r \), where \( k_r \) is the time taken by the termination message \( m_i^{(3)} \) to reach every node after its initiation at any node. Thus, the time \( N^* \) of the decision phase is bounded above as
\[
N^* \leq \max_{1 \leq \ell \leq L} \tau(T_{i,\ell}) + k_r.
\]

Therefore, we have
\[
\mathbb{E}_i[N^*] \leq \sum_{\ell=1}^{L} \mathbb{E}_i[\tau(T_{i,\ell})] + \mathbb{E}_i[k_r].
\]

Using \( (58) \), the term \( \mathbb{E}_i[\tau(T_{i,\ell})] \) at the right-hand side of \( (73) \) is finite and independent of \( c \). Additionally, \( k_r < d^G + 1 \). Thus, \( \mathbb{E}_i[N^*] \) is finite and independent of \( c \).

Combining equations \( (71), (72) \), and the finiteness of \( \mathbb{E}_i[N^*] \), we get that \( (36) \) holds as \( c \to 0 \), proving part (ii) of the theorem.

Now we derive the bounds for the higher moments of the decision time \( N \). We have
\[
N \leq \max\{N^c, \max_{1 \leq \ell \leq L} (T_{i,\ell} + \tau(T_{i,\ell}))\} + N^*,
\]
\[
\leq \max\{N^c, \max_{1 \leq \ell \leq L} (T_{i,\ell})\} + 2 \max_{1 \leq \ell \leq L} \tau(T_{i,\ell}) + k_r,
\]
\[
\leq \max\{N^c, \max_{1 \leq \ell \leq L} (T_{i,\ell})\} + 2 \sum_{\ell \in [L]} \tau(T_{i,\ell}) + k_r.
\]

Now, we present the bound on the \( r \)-th moment of each term in the right-hand side of \( (74) \). Using \( (71) \), \( N^c \) is bounded above by a constant. As \( c \to 0 \), we have
\[
(N^c)^r \leq \left( 1 + o(1) \right) \frac{h^G \cdot \log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(h^G))}.
\]

Using \( (55) \), we have
\[
\mathbb{E}_i \left[ \max_{1 \leq \ell \leq L} T_{i,\ell}^r \right] = \left( 1 + o(1) \right) \frac{\log(c/\max_{j \in [L]} I(j))}{I(i) - c}.
\]

Using \( (58) \), the higher moments of the second term in the right-hand side of \( (74) \) are finite and independent of \( c \) by definition of \( \tau(T_{i,\ell}) \). Additionally, \( k_r \leq L + 1 \) < \( \infty \). Now,
\[
\mathbb{E}_i^c[N^r] \leq \mathbb{E}_i \left[ \max\{N^c, \max_{1 \leq \ell \leq L} (T_{i,\ell})\} + 2 \sum_{\ell \in [L]} \tau(T_{i,\ell}) + k_r \right]^r.
\]

The moments of \( \sum_{\ell \in [L]} \tau(T_{i,\ell}) + k_r \) are finite and independent of \( c \). The dominant terms, dependent on \( c \), in the expansion of the right-hand side of \( (77) \) depend only on \( \max\{N^c, \max_{1 \leq \ell \leq L} (T_{i,\ell})\} \). Therefore, as \( c \to 0 \), we have
\[
\mathbb{E}_i^c[N^r] \leq \left( 1 + o(1) \right) \frac{h^G \cdot \log(c/\max_{j \in [L]} I(j))}{\log(1 - \eta(h^G))} \cdot \frac{\left| \log c \right|}{I(i) - c}.
\]

which proves part (iv) of the theorem.

\( \square \)

D. Proof of Lemma [8]

Proof. As \( c \to 0 \), we have \( I(i) - c \approx I(i) \). Using this approximation, if \( (36) \) holds, then the max operation in \( (30) \) results in \( |\log c|/(I(i) - c) \). The submultiplicative property of the ergodic coefficient implies that
\[
\eta(h^G) \leq \eta(h^G).
\]

Using this fact and \( (36) \), it follows that \( (37) \) is a sufficient condition as well. Now, \( (38) \) follows from \( (37) \) and \( h^G \geq 1 \).

\( \square \)

E. Decision phase of CCT

Lemma 9. If \( d^N_{i,\ell} > L + 1 \), then there exists a time \( k \leq N \) at which the local decision of all the nodes are the same, i.e., \( \min_{j \in [L]} x^k_j \geq 1 \). This decision is the same as the local decision \( H^N_{i,\ell} \) of node \( i \) at time \( N \).

Proof. At time \( N \) and node \( \ell \), if \( d^N_{i,\ell} > L + 1 \), then for all \( k \in N_{i,\ell} \), \( d^N_{i,\ell} > L \) and \( x^k_{i,\ell} \geq L \). If the shortest distance
between the node \( \ell \) and \( j \) is \( s_{\ell,j} \), then \( d_j^{N-s_{\ell,j}} > L - s_{\ell,j} + 1 \). Thus, for all \( j \in [L] \), \( d_j^{N-d^\circ} > d_j^{N-s_{\ell,j}} + s_{\ell,j} - d^\circ > 1 \) as \( s_{\ell,j} \leq d^\circ \leq L \). This implies, for all \( j \in [L] \), \( x_j^{N-d^\circ-1} \geq 1 \). Thus, the first statement of the claim follows.

Now, we prove the second statement by contradiction. Let the decision at time \( N - d^\circ - 1 \), for all \( j \in [L] \), be \( H_j^{N-d^\circ-1} = h' \) which is different from \( H_j^N \). At sensor \( \ell \), let the decision change from \( h' \) to \( H_{\ell}^N \) at time \( n \). Then,

\[
N - d^\circ - 1 < n \leq N.
\]

Therefore,

\[
x^N_{\ell} = 1,
\]

which implies

\[
d_{\ell}^{N+1} \leq 2.
\]

Now,

\[
d_{\ell}^N = d_{\ell}^{N+1} + N - n - 1
\]
\[
\leq 2 + N - n - 1
\]
\[
< 2 + d^\circ
\]
\[
\leq 2 + L.
\] (79)

However, \( d_{\ell}^N \geq L + 2 \) by the statement of the Lemma. Hence, by contradiction, we conclude that the second statement of our claim holds.

**Appendix C**

**Proof of Miscellaneous Results**

In this section, we present results used to bound time \( \tau(N_i, \ell) \) in Theorem 3 and 6 (see (58)). Let \( X_1, \ldots, X_n \) be i.i.d. random variables and let the time

\[
T = \sup \left\{ n : \sum_{k=1}^n X_k < 0 \right\}.
\] (80)

This is the last time at which

\[
S_n > 0,
\] (81)

where \( S_n = \sum_{k=1}^n X_k \).

**Lemma 10.** If \( E[X_1] < \infty \) and \( E[X_1] < \epsilon_0 < 0 \), then

\[
E[T^r] \leq \left( \frac{2}{\epsilon_0} \right)^r E[S^{r \epsilon_0}] + \sum_{k=1}^\infty k^{r-1} P(S_k + k\epsilon_0/2 > 0),
\] (82)

where \( S^r = \max_{j>1} S_k \).

**Proof.** Proof of the lemma follows from some basic definitions and bounds from probability theory. We have

\[
E[T^r] = \sum_{k=1}^\infty k^{r-1} P(T \geq k)
\]
\[
= \sum_{k=1}^\infty \int_{r}^{\infty} \epsilon \leq \max S_j > 0 \] \( j \geq k \)
\[
= \sum_{k=1}^\infty k^{r-1} P \left( \max_{j \geq k} (S_j - S_k) + S_k > 0 \right)
\]
\[
= \sum_{k=1}^\infty k^{r-1} P \left( S^* + S_k > 0 \right).
\] (83)

Now,

\[
\sum_{k=1}^\infty k^{r-1} P(S^* + S_k > 0)
\]
\[
= \sum_{k=1}^\infty \int_{r}^{\infty} \epsilon \leq m \) \( dm
\]
\[
= \int_{r}^{\infty} \sum_{k=1}^\infty k^{r-1} P(S_k > 0) \) \( \leq m \) \( dm
\]
\[
+ \int_{r}^{\infty} \sum_{k=1}^\infty k^{r-1} P(S_k + \epsilon_0 k/2 > 0) \) \( m \) \( dm
\]
\[
\leq \left( \frac{2}{\epsilon_0} \right)^r E[S^{r \epsilon_0}] + \sum_{k=1}^\infty k^{r-1} P(S_k + k\epsilon_0/2 > 0).
\] (84)

**Corollary 10.1.** Let

\[
T = \sup \left\{ n : \sum_{k=1}^n X_k < 0 \right\}.
\] (85)

If for any \( \epsilon_0 > 0 \), we have \( E[X_1] > \epsilon_0 \) and \( E[X_1] < \infty \), then

\[
E[T^r] \leq \left( \frac{2}{\epsilon_0} \right)^r E \left[ \min_k S_k \right]^r
\]
\[
+ \sum_{k=1}^\infty k^{r-1} P(S_k - k\epsilon_0/2 < 0).
\] (86)

**Proof.** The proof follows from replacing \( S_k \) by \( -S_k \) in Lemma 10.

**Lemma 11.** Let \( X_1, \ldots, X_n \) be a sequence of independent and identically distributed random variables with finite \( r \)th moment, i.e., for all \( r \geq 2 \), \( E[|X_1|^r] < \infty \). We have

\[
\sum_{n=1}^\infty n^{r-1} P \left( \sum_{k=1}^n X_k > n \right) < \infty
\] (87)

**Proof.** Without loss of generality, we assume \( E(X_1) = 0 \). Otherwise, \( X_1 \) can be replaced by \( X_1 - E(X_1) \), and the result will follow. Event \( A = \{ \sum_{k=1}^n X_k > n \} \) is written as a subset of the union of three events i.e. \( A \subset A^{(1)} \cup A^{(2)} \cup A^{(3)} \).
We bound the probability of these three events, and show that for all $i \in [3]$, we have
\[ \sum_{n=1}^{\infty} n^{-1} \mathbb{P}(A^{(i)}_n) < \infty. \] (88)

Thus, (87) follows from (88).

Let $2^i \leq n < 2^{i+1}$. The events $A^{(1)}_n$, $A^{(2)}_n$ and $A^{(3)}_n$ are defined as follows:
\[ A^{(1)}_n = \{ \text{There exists } k \leq n \text{ such that } |X_k| > 2^{i-2} \}, \]
\[ A^{(2)}_n = \{ \text{There exists at least two integers } k_1, k_2 \leq n \text{ such that } |X_{k_1}| > n^{4/5} \text{ and } |X_{k_2}| > n^{4/5} \}, \]
\[ A^{(3)}_n = \left\{ \left| \sum_{k \in \mathbb{N}^r} X_k \right| > 2^{i-2} \right\}, \]

where $N^r = [n] \setminus \{k_1, k_2\}$, and $k_1$ and $k_2$ are the indices such that $|X_{k_1}| > n^{4/5}$ and $|X_{k_2}| > n^{4/5}$. If the event $A^{(1)}_n \cup A^{(2)}_n \cup A^{(3)}_n$ does not occur, then we have
\[ \left| \sum_{k \in [n]} X_k \right| < 2^{i-2} + 2^{i-2} < n. \]

Hence, $A \subset A^{(1)}_n \cup A^{(2)}_n \cup A^{(3)}_n$, and
\[ \mathbb{P}(A) \leq \mathbb{P}(A^{(1)}_n) + \mathbb{P}(A^{(2)}_n) + \mathbb{P}(A^{(3)}_n). \]

Therefore,
\[ \sum_{n=1}^{\infty} n^{-1} \mathbb{P}(A) \leq \sum_{n=1}^{\infty} n^{-1} \mathbb{P}(A^{(1)}_n) + \sum_{n=1}^{\infty} n^{-1} \mathbb{P}(A^{(2)}_n) + \sum_{n=1}^{\infty} n^{-1} \mathbb{P}(A^{(3)}_n). \] (89)

Now, we bound the probability of all three events at the right-hand side of the above equation.

Let $a_i = \mathbb{P}(X_k \geq 2^i)$. We have
\[ \sum_{i=0}^{\infty} 2^{i-1} a_i \leq \sum_{i=0}^{\infty} 2^i (a_i - a_{i+1}) \]
\[ \leq \mathbb{E}[X_k^+] \quad \text{(c)} \]
where (a) follows from the fact that $\int_{y_1}^{y_2} ydy \leq \int_{y_1}^{y_2} ydy$, (b) follows trivially from the definition of $\mathbb{E}[X_k^+]$, and (c) follows from the assumption in the statement of the lemma. Thus, using (90), we have
\[ \sum_{i=0}^{\infty} 2^{i-1} a_i < \infty. \] (91)

Now, we bound the probability of event $A^{(1)}_n$
\[ \sum_{n=1}^{\infty} n^{-1} \mathbb{P}(|X_k| \geq 2^i) \leq \sum_{i=0}^{\infty} \sum_{2^i \leq n < 2^{i+1}} 2^{(i+1)(r-1)} a_i \]
\[ = \sum_{i=0}^{\infty} \sum_{2^i \leq n < 2^{i+1}} 2^{i-1} 2^{r-1} a_i \]
\[ = \sum_{i=0}^{\infty} \sum_{2^i \leq n < 2^{i+1}} 2^{(r-1)+r-1} a_i \]
\[ < \infty, \]
where the first inequality follows from $\int_{y_1}^{y_2} ydy \leq \int_{y_1}^{y_2} ydy$, and the last inequality follows from (91).

Since the $r$th moment is finite, for all $k \in \mathbb{N}$ there exists a finite $K$ such that
\[ \mathbb{P}(|X_k| > u) \leq K/u^r. \] (93)

Now, we bound the probability of event $A^{(2)}_n$
\[ \mathbb{P}(A^{(2)}_n) \leq \sum_{1 \leq k_1 < k_2 \leq n} \mathbb{P}(|X_{k_1}| > n^{4/5} \text{ and } |X_{k_2}| > n^{4/5}) \]
\[ \leq n^2 \cdot \mathbb{P}(|X_1| > n^{4/5}) \mathbb{P}(|X_2| > n^{4/5}) \]
\[ \leq K^2 \cdot n^2 \cdot n^{-4r/5} \cdot n^{-4r/5} \]
where (a) follows from the definition of the event and union bound, (b) follows from the independence of random variables and number of possible combinations of $k_1$ and $k_2$, and (c) follows from (93). Therefore, we have
\[ \sum_{n=1}^{\infty} n^{-1} \mathbb{P}(A^{(2)}_n) \leq \sum_{n=1}^{\infty} K^2 \cdot n^{-1} n^{r-1} \cdot n^2 \cdot n^{-4r/5} \cdot n^{-4r/5} \]
\[ = \sum_{n=1}^{\infty} K^2 \cdot n^{-3r/5+1} \]
\[ \leq \frac{b}{\infty}, \]
where (a) follows from (94), and (b) follows as $r \geq 2$.

Now, we bound the probability of event $A^{(3)}_n$. Let
\[ X_k^+ = \begin{cases} X_k & \text{if } |X_k| \leq n^{4/5} \\ 0 & \text{otherwise}. \end{cases} \] (96)

Now, let $\mathbb{E}[X_k^+] = \epsilon_n$ and $Y_k = X_k^+ - \epsilon_n$. Thus, $\mathbb{E}[Y_k] = 0$. There exists a finite constant $K'$ such that
\[ \mathbb{E}\left[ \sum_{k=1}^{n} Y_k^{2r} \right] \]
\[ \leq \mathbb{E}\left[ \sum_{k=1}^{n} (Y_k^{2r}) \right] + n \cdot \mathbb{E}\left[ (Y_k^{2r}) \right] \cdot \mathbb{E}\left[ Y_k \right] + \ldots \]
\[ \leq \mathbb{E}\left[ \sum_{k=1}^{n} n^{4r/5} (Y_k)^r \right] + n \cdot \mathbb{E}\left[ n^{4(r-1)/5} (Y_k)^r \right] \mathbb{E}[Y_k] + \ldots \]
\[ \leq K' \cdot n^{4r/5} n^{-r+1} \cdot 0.5r/n, \]
where (a) follows from the binomial expansion of $(\sum_{k=1}^{n} Y_k)^{2r}$ and independence of the random variables, (b) follows from (95), and (c) follows from the fact that the largest binomial coefficient in the expansion of $(\sum_{k=1}^{n} Y_k)^{2r}$ is $O(n^{r-1}+0.5r/n)$, and the $r$th moment of $Y_k$ is finite. Thus, using (97), we have
\[ \mathbb{P}\left( \sum_{k=1}^{n} Y_k \right) > n/16 \leq K' \cdot n^{4r/5} n^{-r+1} \cdot 0.5r/n n^{2r}. \] (98)

Now, using (96), as $n \to \infty$, $\epsilon_n \to 0$. Thus, there exists a $N_e$ such that for all $n > N_e$, $\epsilon_n < 1/16$. Therefore, for all
\( n > N_c, \ X_k^+ \leq Y_k + 1/16 \). Hence, the probability of event \( A_n^{(3)} \), for all \( n > N_c \), is
\[
\mathbb{P}(A_n^{(3)}) = \mathbb{P}(\sum_{k=1}^{n} X_k^+) > 2^{n-2} \\
\leq \mathbb{P}(\sum_{k=1}^{n} Y_k > n/16) \\
\leq K' \cdot n^{4r/5} n^{-r-1+0.5r/n} n^{-2r},
\]
where the last inequality follows from (98). Thus, we have
\[
\sum_{n=1}^{\infty} n^{-r-1} \mathbb{P}(A_n^{(3)})
\leq \sum_{n=1}^{N_c} n^{-r-1} \mathbb{P}(A_n^{(3)}) + \sum_{n=N_c}^{\infty} n^{-r-1} K' \cdot n^{4r/5} n^{-r-1+0.5r/n} n^{-2r}
\leq K'' + \sum_{n=N_c}^{\infty} n^{-r-1} K' \cdot n^{4r/5} n^{-r-1+0.5r} n^{-2r}
\leq K'' + \sum_{n=N_c}^{\infty} K' \cdot n^{-6r/5}
\leq \infty,
\]
where \( a \) follows from (99), \( b \) uses the fact that the finite sum of finite numbers is finite and is denoted by the constant \( K'' \), and \( n \geq 1 \), \( c \) follows from the simplification of the previous inequality, and \( d \) follows from the fact that \( r \geq 2 \). Finally, using (92), (95) and (100), (87) follows.

\[
\text{ACKNOWLEDGMENT}
\]

This work was partially supported by NSF awards number CNS-1446891 and CCF-1717942.

\section*{REFERENCES}


