2018 IEEE Conference on Decision and Control (CDC)
Miami Beach, FL, USA, Dec. 17-19, 2018

Consensus-Based Chernoff Test in Sensor Networks

Anshuka Rangi, Massimo Franceschetti and Stefano Marano

Abstract— **We propose a sequential and adaptive hypothesis test that operates in a completely distributed setting, relying on a sensor network where no single data-fusion center is present. The test is inspired by Chernoff's optimal solution, originally derived in a centralized setting. We compare the performance of our test with the optimal sequential test in sensor networks and provide sufficient conditions for which the proposed test achieves asymptotic optimality, 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. Under these conditions, the proposed test is also shown to be asymptotically optimal with respect to the higher moments of the time required to reach a decision.**

I. Introduction

With the boom in the Internet of Things, sensor-network based solutions have become increasingly popular for inference systems. Their advantages include the increasingly low cost of the sensors, their embedded computational capabilities, the inherent redundancy provided by the structure of the network, and the availability of high-speed wireless communication channels [1]. In a typical setup, a set of hypotheses is tested based on the observations collected at the sensors, and the result of the test is used to choose future actions to be performed in the network. Applications that fall in this framework include intrusion and target detection, and object classification and recognition [2]–[7]. These systems can be broadly classified into three types: centralized, decentralized, and completely distributed. In a centralized setting, the sensors send all of their observations to a central processor, where the inference task is performed. In a decentralized setting, the computational capabilities of the sensors are exploited to perform some amount of preliminary processing, before sending a limited amount of information to the central processor. This reduces the communication overhead, possibly at the price of sub-optimal performance. In a completely distributed setting, sensors are connected to each other via communication links, typically forming a sparse network, and there is no central processing unit. Thus, the sensors need to perform computations locally, share their processed data with neighboring sensors, and collectively reach a decision. A natural question is what kind of local processing to perform and what fusion schemes to adopt at the sensor nodes in order to reduce the communication overhead while keeping a high level of detection performance. In this paper we address this question by proposing a completely distributed statistical test for sensor networks that generalizes Chernoff's optimal one derived in a centralized setting [8].

Hypothesis testing techniques can be broadly classified as sequential or non-sequential tests, and adaptive or non-adaptive tests. Our focus is on a sequential and adaptive test. In a sequential test, the number of observations needed to take a decision is not fixed in advance, but depends on 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 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.

Our contributions are as follows. We propose a sequential and adaptive test that generalizes Chernoff's classic test to a completely distributed setting. We provide an upper bound on the test performance in terms of expected risk, the expected cost required to reach a decision plus the expected cost of making a wrong decision. We also provide upper bounds on the higher order moments of the time required to reach a decision. Finally, we derive bounds on the best possible performance of any sequential test in sensor networks and provide sufficient conditions for our test to retain the same asymptotic optimality of Chernoff's original centralized solution. These results extend and build upon our previous ones for hypothesis testing in a decentralized setting and in the presence of a fusion center [9].

There is a huge literature on distributed estimation and detection over networks. We now briefly recall some key results regarding the fully distributed setting. For references on sequential and adaptive tests in centralized and decentralized settings see [9]. Various gossip protocols have been proposed for distributed computation of different functions, like mean, sum, minimum and maximum, of different network parameters [10]–[15]. These protocols can be divided into two categories: consensus protocols and running-consensus protocols. In consensus protocols, estimation of the desirable parameter occurs after the measurements are collected at the sensors [10], [11], [13], [14]. On the other hand, in running-consensus protocols, the sensing of the environment and the estimation of parameters are performed simultaneously [12], [15]. Necessary and sufficient conditions for the convergence are studied in both of these cases [16].

Similarly, protocols for distributed detection are based on the computation of a belief about a hypothesis and its propagation over the network occurs in a similar fashion

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. {arangi,massimo}@ucsd.edu, marano@unisa.it

978-1-5386-1395-5/18/$31.00 ©2018 IEEE

as in running-consensus protocols. Works in this area focus on different strategies to transmit and combine the belief vectors of the hypotheses over the network, and study the learning rate of these strategies [17]–[21]. The analysis of these techniques in terms of expected decision time, however, has been limited and our work provides an asymptotically optimal solution for this.

The rest of the paper is organized as follows: Section II formulates the problem; Section III reviews the Chernoff test and Section IV introduces its consensus based version; Section V describes the main idea behind the distributed test and Section VI presents rigorous theoretical results; Section VII present the simulations; Section VIII concludes the work.

II. Problem Formulation

We consider a distributed sensor 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 by L sensors and is modeled as a graph $\mathscr{G}(\mathscr{L},\mathscr{E})$, where the set of nodes $\mathscr{L}=\{1,2,\ldots,L\}$ represents the L sensors, and the elements of $\mathscr{E}$ are the edges, namely unordered pairs of nodes $\{(\ell,j)\}$, in which (ℓ,j) represents the communication link between sensors ℓ and j, $\ell\neq j$. The inter-sensor communication is allowed only over the edges $\mathscr{E}$ of $\mathscr{G}(\mathscr{L},\mathscr{E})$. The diameter $d^{\mathscr{G}}$ of the network is the maximum shortest hop-distance between any pair of nodes. We also denote by $h^{\mathscr{G}}$ the shortest height of all possible spanning trees of $\mathscr{G}(\mathscr{L},\mathscr{E})$. It is assumed that the network is connected, namely, there exists a path between any two sensors ℓ and j. Thus, $d^{\mathscr{G}}$ and $h^{\mathscr{G}}$ 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 $[M]=\{1,\ldots M\}$ is used. At each time instant, each sensor 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 the set S is equal to M. Under this latter assumption, action u_i can be interpreted as the "best" action when the state of nature is h_i. All the results of this paper can be extended to the more general case.

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

The performance measure used in this work – the risk – is analogous to the one considered in [8]. Under true hypothesis h_i, the risk $\mathbb{R}_i^{\delta}$ of a sequential test δ is defined as follows:

$$\mathbb{R}_i^{\delta}=c\,\mathbb{E}_i^{\delta}[N]+\omega_i\mathbb{P}_i^{\delta}(\hat{H}\neq h_i), \tag{1}$$

where N is the time required to reach a global decision in the network, c is the observation cost per unit time, $\hat{H}$ is the final decision, $\mathbb{E}_i$ and $\mathbb{P}_i$ are the expectation and the probability operators computed under $H^*=h_i$, and ω_i is the cost of a wrong decision. We propose a test for a distributed sensor network and evaluate its performance in terms of risk for all $i\in[M]$, as $c\to 0$. We provide bounds on the higher moments of the time N required to reach a decision, and also provide sufficient conditions under which the proposed test is asymptotically optimal in terms of risk as well as the higher moments of the decision time N, as $c\to 0$.

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 action set S, namely, for all $\ell\in[L]$ and $i,j,k_1\in[M]$, we have $D(p_{i,\ell}^{u_{k_1}}||p_{j,\ell}^{u_{k_1}})<\infty$. Also, for all $\ell\in[L]$ and $i,j\in[M]$, there exists an action u_{k_1}, where $k_1\in[M]$, such that $D(p_{i,\ell}^{u_{k_1}}||p_{j,\ell}^{u_{k_1}})>0$. This assumption entails little loss of generality, rules out trivialities, and is commonly adopted in the literature, see e.g. [8]. Also, for all $\ell\in[L]$ and $i,j,k_1\in[M]$, we assume $\mathbb{E}[\log(p_{i,\ell}^{u_{k_1}}(Y))/\log(p_{j,\ell}^{u_{k_1}}(Y))]^2<\infty$. We shall use of 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 $v_1\preceq v_2$ means $v_{i,1}\leq v_{i,2}$ for all $i\in[k]$. In addition, $|v_1|$ is the vector of absolute values of the entries of v_1.

III. Standard Chernoff Test

We start with considering sensor ℓ alone, with no interactions with other sensors of the network. The Chernoff test for this isolated sensor works as follows:

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

$$i_{k-1}^*=\arg\max_{i\in[M]}\mathbb{P}(H^*=h_i|y_\ell^{k-1},u_\ell^{k-1}), \tag{2}$$

where H^* is the true hypothesis, $y_\ell^{k-1}=\{y_{1,\ell},\ldots y_{k-1,\ell}\}$, where $y_{i,\ell}$ is the realization of the observation collected at time index (step) i, $u_\ell^{k-1}=\{u_{1,\ell},\ldots u_{k-1,\ell}\}$, and $u_{i,\ell}$ is the realization of the action made at step i.

2) At step k, the action $u_{k,\ell}$ is randomly chosen among the elements of action set S, according to the Probability Mass Function (PMF) $Q_{i_{k-1}^*}^{\ell}$, where:

$$Q_{i_{k-1}^*}^{\ell}=\arg\max_{q\in\mathscr{Q}}\min_{j\in M_{i_{k-1}^*}}\sum_u q(u)D(p_{i_{k-1}^*,\ell}^u||p_{j,\ell}^u),$$

in which $\mathscr{Q}$ 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 probabilities $\mathbb{P}(H^*=h_i|y_\ell^k,u_\ell^k)$.

4) The test stops at step N if the worst case log-likelihood ratio crosses a prescribed fixed threshold γ, i.e.,

$$\log \frac{p_{i^*_{N,\ell}}(y_\ell^N, u_\ell^N)}{\max_{j \neq i^*_N} p_{j,\ell}(y_\ell^N, u_\ell^N)} \geq \gamma, \tag{3}$$

where $p_{i^*_{N,\ell}}(y_\ell^N, u_\ell^N)$ is the posterior probability $\mathbb{P}(H^* = h_{i^*_N} | y_\ell^N, u_\ell^N)$ at sensor ℓ. If the test stops at step N, then the final decision is $h_{i^*_N}$. Otherwise, $k \leftarrow (k+1)$, and the procedures continues from 1).

IV. Consensus Based Chernoff Test

We propose a version of the above Chernoff test designed for fully-flat sensor networks, which is referred to as the consensus-based Chernoff test (CCT). The proposed CCT consists of three phases: consensus among the sensors regarding their cumulative capability to detect hypothesis h_i, $i \in [M]$, performing a Chernoff test locally at each sensor, and consensus regarding the decision among the sensors. The first two phases of CCT can be performed in parallel while the last phase begins after the completion of the first two phases.

In the first phase, the goal of each sensor is to acquire knowledge about the cumulative capability of the network to detect hypothesis h_i, $i \in [M]$. For sensor ℓ, the measure of the capability to detect h_i is given by

$$v_{i,\ell} = \max_{q \in \mathcal{Q}} \min_{j \neq i} \sum_u q(u) D(p^u_{i,\ell} || p^u_{j,\ell}). \tag{4}$$

Thus, the cumulative capability of the network to detect hypothesis h_i is

$$I(i) = \sum_{\ell=1}^{L} v_{i,\ell}, \qquad i \in [M], \tag{5}$$

and $I = [I(1), \ldots, I(M)]$ is the corresponding vector. Since there is no central entity to facilitate the computation of the quantities in (5), the sensors use local information and consensus techniques to acquire this knowledge. Consensus techniques allow to compute the arithmetic mean of remotely-collected observations by exploiting only information locally available to the sensors, see e.g., [10], [11], [13], [14]. Assuming that the number of sensors L is known to all the sensors, we use a linear consensus technique to estimate the arithmetic mean I/L of the cumulative capability, which multiplied by L provides the desired estimate of I. The distributed linear consensus technique is of the form

$$\hat{I}_\ell^{n+1} = w_{\ell,\ell} \cdot \hat{I}_\ell^n + \sum_{j \in \mathcal{N}_\ell} w_{\ell,j} \cdot \hat{I}_j^n, \tag{6}$$

where $\hat{I}_\ell^n = [\hat{I}_\ell^n(1), \ldots, \hat{I}_\ell^n(M)]$ is the vector of estimated cumulative capabilities for M hypotheses at sensor ℓ and time instance n, $w_{\ell,j}$ is the weight assigned by sensor ℓ to the estimate of sensor j, and $\mathcal{N}_\ell = \{j | \{\ell, j\} \in \mathcal{E}\}$ is the set of immediate neighbors of sensor ℓ in $\mathcal{G}(\mathcal{L}, \mathcal{E})$. At $n = 0$, the estimated cumulative capabilities are initialized as $\hat{I}_\ell^0 = [v_{1,\ell}, \ldots, v_{M,\ell}]$, and the initialization $\hat{I}_\ell^0$ can be computed locally at the sensors. Since the sensor ℓ does not communicate to sensors $\{j | j \notin \mathcal{N}_\ell \cup \{\ell\}\}$, it follows that $w_{\ell,j} = 0$. Thus, for the network graph $\mathcal{G}(\mathcal{L}, \mathcal{E})$, the linear consensus technique (6) can be written as

$$\hat{I}^{n+1} = W \cdot \hat{I}^n, \tag{7}$$

where $\hat{I}^n = [\hat{I}_1^n, \ldots, \hat{I}_L^n]^T$ is an $L \times M$ matrix of the estimate of the cumulative capability of the network at time instance n at all sensors, and W is an $L \times L$ matrix with elements $w_{\ell,j}$ where $\ell, j \in [L]$. The matrix W is constrained to belong to

$$\mathcal{T} = \{W \in \mathbb{R}^{L \times L} | w_{\ell,j} > 0 \text{ if } j \in \mathcal{N}_\ell \cup \{\ell\} \text{ else } w_{\ell,j} = 0\}. \tag{8}$$

The consensus iteration (7) will converge to the mean of the cumulative capability of the network I/L if and only if [16]

$$\mathbf{1}^T \cdot W = \mathbf{1}^T, \tag{9}$$

$$W \cdot \mathbf{1} = \mathbf{1}, \tag{10}$$

and

$$R\left(W - \frac{\mathbf{1}_{L\times 1} \cdot \mathbf{1}_{1\times L}}{L}\right) < 1, \tag{11}$$

where $R(.)$ denotes the spectral radius, and $\mathbf{1}_{A\times B}$ is a $A \times B$ matrix of all ones. Since the rate of convergence of the consensus scheme (7) is proportional to $R(W - \mathbf{1}_{L\times 1} \cdot \mathbf{1}_{1\times L}/L)$, the computation of W can be formulated as a convex optimization problem subject to (8), (9) and (10), and can be determined using standard techniques [16].

Another challenge in the Phase 1 of CCT is the stopping rule of the consensus algorithm. Assuming the number of sensors in the network is known to all the sensors, a localized stopping rule is proposed in [22]. Using this rule, the first phase of CCT is summarized in Algorithm 1. As soon as $z_\ell > L+1$, sensor ℓ sends a termination bit $m_t^1 = 1$ of Phase 1 to inform its neighbors $\mathcal{N}_\ell$ that the consensus to $\approx I/L$ has been reached. When sensor j receives $m_t^1 = 1$, it halts the consensus, scale the estimate $\hat{I}_\ell^n$ by L to get an estimate of I, and forwards m_t^1 to its neighbors $\mathcal{N}_j$. Using threshold c/L^2 at each sensor (see Algorithm 1), the stopping rule ensures that the network has reached an uniformly local c/L^2-consensus, i.e. for all $\ell \in [L]$ and $j \in \mathcal{N}_\ell$,

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

Therefore, for all $\ell, j \in [L]$,

$$|\hat{I}_\ell^n - \hat{I}_j^n| \preceq \frac{c}{L} \cdot \mathbf{1}_{1\times M}, \tag{12}$$

because the diameter $d^{\mathcal{G}}$ of the network is at most L. After scaling by L, at termination of Phase 1, for all $\ell, j \in [L]$,

$$|\hat{I}_\ell^n - \hat{I}_j^n| \preceq c \cdot \mathbf{1}_{1\times M}. \tag{13}$$

In the second phase of CCT, all sensors perform a Chernoff test independently of each other, and compute $\log p_{i^*_n,\ell}(y^n, u^n) / \max_{j \neq i^*_n} p_{j,\ell}(y^n, u^n)$. Following the termination of Phase 1, if at time n and sensor ℓ

$$\log \frac{p_{i^*_n,\ell}(y^n, u^n)}{\max_{j \neq i^*_n} p_{j,\ell}(y^n, u^n)} \geq \hat{\rho}^n_{i^*_n,\ell} |\log c|, \tag{14}$$

where $\hat{\rho}^n_{i^*_n,\ell} = v_{i^*_n,\ell}/\hat{I}^n_\ell(i^*_n)$, then the local decision $\hat{H}^n_\ell$ is updated in favor of hypothesis $h_{i^*_n}$, otherwise, it is set to null. Note that the consensus about I/L in the first phase is independent of the Chernoff test and the computation of log-likelihood in the second phase. Hence, these can be performed in parallel over the network. However, in (14), the local decision at the sensors requires reliable estimate of $\rho^n_{i^*_n,\ell} = v_{i^*_n,l}/I(i^*_n)$ which is only attained after the termination of the first phase. Hence, the local decision $\hat{H}^n_\ell$ at node ℓ is made only after receiving the termination bit $m^1_t = 1$.

Phase 3 follows after the update of the local decision at the sensors. Each sensor ℓ communicates its local decision $\hat{H}^n_\ell$, if any, to $\mathcal{N}_\ell$. It also communicates d^n_ℓ defined as

$$d^n_\ell = \min\{\min_{j\in\mathcal{N}_\ell\cup\{\ell\}} d^{n-1}_j, x^{n-1}_\ell\} + 1, \tag{15}$$

where for all $\ell \in [L]$, $d^0_\ell = 0$ and

$$x^n_\ell = \begin{cases} x^{n-1}_\ell + 1 & \text{if } \forall j \in \mathcal{N}_\ell, \hat{H}^n_\ell = \hat{H}^n_j \text{ and } \hat{H}^n_\ell = \hat{H}^{n-1}_\ell, \\ 1 & \text{if } \forall j \in \mathcal{N}_\ell, \hat{H}^n_\ell = \hat{H}^n_j \text{ and } \hat{H}^n_\ell \neq \hat{H}^{n-1}_\ell, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Say x^n_ℓ is some constant k, then the decision of $\mathcal{N}_\ell$ is the same as the local decision $\hat{H}^n_\ell$ of sensor ℓ for the past k time instances. The value of d^n_ℓ is responsible for the percolation of this information along the sensor network. Using (16), if a sensor j does not report its local decision, then $x^n_\ell = 0$ in the neighborhood of j. A sensor ℓ stops the test at time instance N if $d^N_\ell > L+1$. This ensures that there exists a time instance $k \le N$ at which the local decision of all the sensors is same i.e. $\min_{j\in[L]} x^k_j \ge 1$ (Appendix I of [23]). Additionally, if $d^N_\ell > L+1$, then sensor ℓ informs its neighbors $\mathcal{N}_\ell$ that the final decision is $\hat{H}^N_\ell$, and sends a termination bit $m^3_t = 1$ of Phase 3 to terminate the test. When a sensor j receives $m^3_t = 1$ and the final decision, it halts the test and forwards m^3_t along with the final decision to its neighbors $\mathcal{N}_j$. All the sensors will receive $m^3_t = 1$ after at most $d^{\mathcal{G}}$ time instances.

Algorithm 1 Phase 1 of CCT

```
Initialize n = 0, and For all ℓ ∈ [L], Î^0_ℓ, y_ℓ = 0 and z_ℓ = 0
while True do
    For all ℓ ∈ [L], broadcast local information Î^(n)_ℓ and z_ℓ.
    Update the local cumulative capability using (6).
    z_ℓ = min{y_ℓ, min_{j∈N_ℓ∪{ℓ}} z_j} + 1
    if z_ℓ > L + 1 then
        Sensor ℓ broadcasts m^1_t = 1 and stop updating.
        Break While;
    end if
    if max_{j∈N_ℓ} |Î^(n)_ℓ − Î^(n)_j| ⪯ c · 1_{1×M}/L^2 then
        y_ℓ = y_ℓ + 1
    else
        y_ℓ = 0
    end if
    n = n + 1
end while
```

In the literature, distributed hypothesis testing is performed while communicating the posterior probabilities for all hypotheses which are real valued vectors over the network [17]–[21]. On the other hand, CCT only communicates the local decision of the sensors. CCT requires communication of $\hat{I}^n_\ell$, a real valued vector, in its first phase to find the cumulative capabilities of the network. However, since the termination time of Phase 1 is bounded [23], the number of communications of real valued vectors is bounded. Thus, CCT is parsimonious in terms of the communication in comparison to the schemes in the literature [17]–[21].

V. Informal discussion

The key idea behind the CCT is to determine the individual capabilities of the sensors for detecting the hypotheses. These capabilities — that depend on the true hypothesis h^* — are captured by $v_{i,\ell}$. All the sensors 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 instance N, then $\hat{\rho}^N_{i,\ell}$ denotes the estimated fraction of the capability contributed by sensor ℓ for hypothesis h_i. To minimize the expected time to reach a decision, it is desirable to determine the threshold (see right hand side of (14)) for each sensor ℓ such that all the sensors require roughly the same time, following the termination of Phase 1, to reach the triggering condition (14). Given the estimate of cumulative capabilities $\hat{I}^N$, this is analogous to dividing the task of hypothesis testing among the sensors based on their speed of performing the task, such that all the sensors finish their share of the task roughly at the same time. Phase 3 is a localized stopping criterion of the Chernoff test, and ensures the sensors stop the test as they reach the same decisions. x^n_ℓ and d^n_ℓ capture this information mathematically, and percolate it over the network.

VI. Theoretical Results

In the following theorems, N indicates the time required to make a decision. The superscripts $\mathscr{C}$ and δ refer to the CCT and to a generic decentralized sequential test, respectively. The ergodic coefficient of the matrix W is

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

The proofs of theorems can be found in the appendices of [23]. 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 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). (iv) presents the bound on the higher moments of the decision time N of CCT. First, we present a lemma which is used in the proof of Theorem 2.

Lemma 1: If the network is connected, then $0 < \eta(W^{h^{\mathscr{G}}}) < 1$ [24, Proposition 1].

Theorem 2: (Direct). The following statements hold:
(i) For all $c \in (0,1)$ and for all $i \in [M]$, given that hypothesis h_i is true, the probability that the CCT makes an incorrect decision is bounded as

$$\mathbb{P}_i^{\mathscr{C}}(\hat{H} \neq h_i) \leq \min\{(M-1)c^{\frac{1}{1-c/I(i)}}, 1\}.$$

(ii) For all $i \in [M]$, given that hypothesis h_i is true, the expected decision time is

$$\begin{aligned}&\mathbb{E}_i^{\mathscr{C}}[N]\\ &\leq (1+o(1)) \max\left\{\frac{h^{\mathscr{G}} \cdot \log(c/\max_{j\in[L]} I(j))}{\log\left(1-\eta(W^{h^{\mathscr{G}}})\right)}, \frac{|\log c|}{I(i)-c}\right\},\end{aligned} \tag{17}$$

as $c \to 0$.
(iii) Combining (i) and (ii), the risk defined in (1) verifies

$$\begin{aligned}\mathbb{R}_i^{\mathscr{C}} \leq &(1+o(1)) \max\left\{\frac{h^{\mathscr{G}} \cdot |\log(1/\max_{j\in[L]} I(j))|}{|\log(1-\eta(W^{h^{\mathscr{G}}}))|}, \frac{1}{I(i)-c}\right\}\\ &\cdot c|\log c|,\end{aligned} \tag{18}$$

as $c \to 0$.
(iv) For all $\ell \in [L]$, $i, j, k_1 \in [M]$ and $r \geq 2$, if $\mathbb{E}\left[\log p_{i,\ell}^{u_{k_1,\ell}}(Y)/p_{j,\ell}^{u_{k_1,\ell}}(Y)\right]^r < \infty$, then

$$\begin{aligned}&\mathbb{E}_i^{\mathscr{C}}[N^r]\\ &\leq \left((1+o(1)) \max\left\{\frac{h^{\mathscr{G}} \cdot \log(c/\max_{j\in[L]} I(j))}{\log(1-\eta(W^{h^{\mathscr{G}}}))}, \frac{|\log c|}{I(i)-c}\right\}\right)^r.\end{aligned} \tag{19}$$

as $c \to 0$.

The following theorem provides a converse result.

Theorem 3: (Converse). For any sequential test δ, if for all $i \in [M]$ the probability of missed detection satisfies

$$\mathbb{P}_i^{\delta}(\hat{H} \neq h_i) = O(c\,|\log(c)|), \quad \text{as } c \to 0, \tag{20}$$

then we have

$$\mathbb{E}_i^{\delta}[N^r] \quad \geq \quad \left((1+o(1))\frac{|\log c|}{I(i)}\right)^r, \text{ as } c \to 0. \tag{21}$$

Using the above result of $r = 1$, we have

$$\mathbb{R}_i^{\delta} \quad \geq \quad (1+o(1))\frac{c\,|\log c|}{I(i)}, \quad \text{as } c \to 0. \tag{22}$$

The following result is a consequence of Theorems 2 and 3. It provides a sufficient condition for the asymptotic optimality of the CCT as $c \to 0$.

Theorem 4: Let δ^* be the optimal sequential test in the sensor network. For the CCT , if the maximum in (17) is $|\log c|/(I(i)-c)$, then for all $i \in [M]$ we have

$$\lim_{c\to 0} \frac{\mathbb{E}_i^{\mathscr{C}}[N]}{\mathbb{E}_i^{\delta^*}[N]} \quad = \quad 1, \tag{23}$$

$$\lim_{c\to 0} \frac{\mathbb{R}_i^{\mathscr{C}}}{\mathbb{R}_i^{\delta^*}} \quad = \quad 1. \tag{24}$$

The following corollary provides sufficient conditions for which the maximum in (17) results in $|\log c|/(I(i)-c)$.

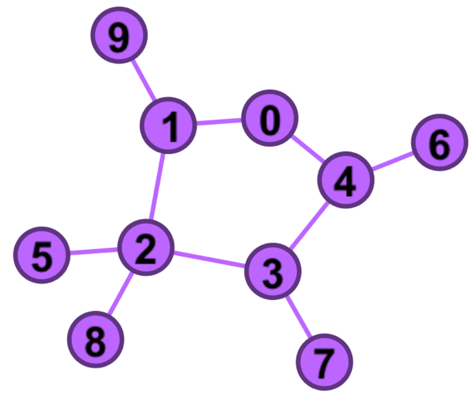

Fig. 1. Example of a sensor network for L=10

Corollary 4.1: The following are sufficient conditions for CCT to be asymptotically optimal. As $c \to 0$,
(i) For all $i \in [M]$, we have

$$I(i)\left|\log\left(\max_{j\in[M]} I(j)\right)\right| \leq \frac{\left|\log\left(1-\eta(W^{h^{\mathscr{G}}})\right)\right|}{h^{\mathscr{G}}}. \tag{25}$$

(ii) For all $i \in [M]$, we have

$$I(i)\left|\log\left(\max_{j\in[M]} I(j)\right)\right| \leq \frac{\left|\log\left(1-a^{\mathscr{G}} \cdot \underline{W}^{h^{\mathscr{G}}}\right)\right|}{h^{\mathscr{G}}}, \tag{26}$$

where $a^{\mathscr{G}}$ is the number of spamming trees of $\mathscr{G}(\mathscr{L},\mathscr{E})$ of height $h^{\mathscr{G}}$, and $\underline{W} = \min\{w_{\ell,j} : w_{\ell,j} > 0\}$.
(iii) For all $i \in [M]$, we have

$$I(i)\left|\log\left(\max_{j\in[M]} I(j)\right)\right| \leq \frac{\left|\log\left(1-a^{\mathscr{G}} \cdot \underline{W}^{d^{\mathscr{G}}}\right)\right|}{d^{\mathscr{G}}}. \tag{27}$$

We now briefly discuss the physical significance of the sufficient conditions presented in Corollary 4.1. The consensus in Phase 1 of CCT should be reached before the triggering condition (14) for Chernoff Test is satisfied. In other words, consensus along the network should be faster than the time required by the Chernoff test to accumulate sufficient information to make a decision.

VII. Simulation Results

We now evaluate the performance of CCT for different sensor networks. Given a number of sensors L, we connect $\lceil L/2 \rceil$ of them to form a ring topology, and the remaining ones are randomly connected to the sensors in the ring. An example network is depicted in Figure 1. We assume the number of hypotheses is $M = 3$. The probability distribution $p_{i,\ell}^{u_k}$ is a Bernoulli distribution with parameter p, which is selected uniformly at random from $(0,1/3)$,$(1/3,2/3)$ and $(2/3,1)$ for $h_i = 1,2$ and 3 respectively. In Figure 2, the expected decision time of CCT increases with the number of sensors. Using Part (ii) of Theorem 2, as L increases, there is a trade-off between the time required to reach consensus in

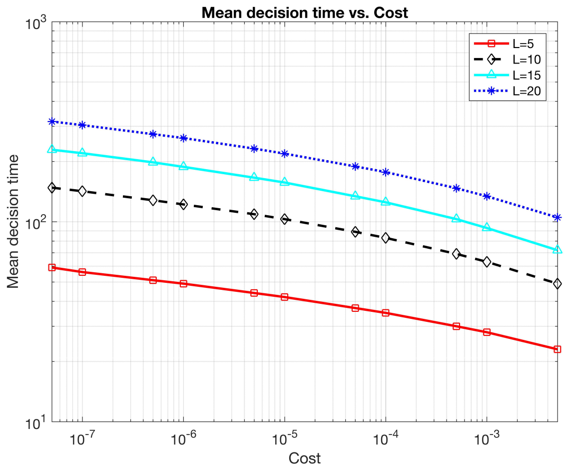

Fig. 2. Performance of CCT for different number of sensors

Phase 1 and the time required to reach the local decision (see (14)) in Phase 2. As L increases, the diameter of the network increases thus the consensus in Phase 1 will require more time. On the other hand, as L increases, the cumulative capabilities of the network will increase, thus, reducing the time required to reach the local decision (see (14)). According to Figure 2, the time required to reach consensus in Phase 1 of CCT becomes the dominating factor in the decision time for this network. This is in accordance with the theoretical bounds in Part (ii) of Theorem 2. The first term in the max operation in (17), corresponding to the time of Phase 1, becomes dominant, and increases with L.

Another key observation from Figure 2 is that the expected decision time reduces with the increase in observation cost per unit time c. As c increases, the stopping criterion of Phase 1 and the decision criterion in Phase 2 i.e. (14) are relaxed, and can be reached earlier. The observation is in accordance with the theoretical bounds provided in Theorem 2.

VIII. Final Remarks

The literature of distributed hypothesis testing has mainly focused on the communication of real-valued belief vectors, which is analogous to communicating posterior probability of the hypotheses, over the network. In contrast, our solution is parsimonious in terms of communication. In Phase 2 and Phase 3 of CCT, communication is limited to local decisions and (15), which can be encoded in $\log_2(M)+\log_2(L)+1$ bits. Although Phase 1 of CCT requires communication of real valued vectors, however, the number of these communications is bounded.

In the literature, the analysis of distributed hypothesis testing schemes is limited to the asymptotic learning rate as time tends to infinity. In contrast, in our work we have studied the probability of missed detection, expected decision time, and the higher moments of the decision time.

References

[1] P. K. Varshney, *Distributed detection and data fusion*. Springer Science & Business Media, 2012.

[2] Y. Hu, M. Dong, K. Ota, A. Liu, and M. Guo, "Mobile target detection in wireless sensor networks with adjustable sensing frequency," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1160–1171, 2016.

[3] A. Rangi and M. Franceschetti, "Multi-armed bandit algorithms for crowdsourcing systems with online estimation of workers' ability," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 1345–1352.

[4] Y. Liu, M. Dong, K. Ota, and A. Liu, "Activetrust: secure and trustable routing in wireless sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 2013–2027, 2016.

[5] J. Aslam, Z. Butler, F. Constantin, V. Crespi, G. Cybenko, and D. Rus, "Tracking a moving object with a binary sensor network," in *Proceedings of the 1st international conference on Embedded networked sensor systems*. ACM, 2003, pp. 150–161.

[6] P. Kulkarni, D. Ganesan, P. Shenoy, and Q. Lu, "Senseye: a multi-tier camera sensor network," in *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005, pp. 229–238.

[7] S. Ren, Q. Li, H. Wang, X. Chen, and X. Zhang, "Design and analysis of sensing scheduling algorithms under partial coverage for object detection in sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 3, pp. 334–350, 2007.

[8] H. Chernoff, "Sequential design of experiments," *The Annals of Mathematical Statistics*, vol. 30, no. 3, pp. 755–770, 1959.

[9] A. Rangi, M. Franceschetti, and S. Marano, "Decentralized Chernoff test in sensor networks," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 501–505.

[10] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, no. SI, pp. 2508–2530, 2006.

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

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

[13] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[14] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*. IEEE, 2003, pp. 482–491.

[15] P. Braca, S. Marano, and V. Matta, "Running consensus in wireless sensor networks," in *Information Fusion, 2008 11th International Conference on*. IEEE, 2008, pp. 1–6.

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

[17] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[18] A. Nedić, A. Olshevsky, and C. A. Uribe, "Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 5884–5889.

[19] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed detection: Finite-time analysis and impact of network topology," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3256–3268, 2016.

[20] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Transactions on Automatic control*, vol. 57, no. 3, pp. 592–606, 2012.

[21] A. Lalitha, A. Sarwate, and T. Javidi, "Social learning and distributed hypothesis testing," in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 551–555.

[22] P. Xie, K. You, and C. Wu, "How to stop consensus algorithms, locally?" *arXiv preprint arXiv:1703.05106*, 2017.

[23] A. Rangi, M. Franceschetti, and S. Marano, "Distributed chernoff test: Optimal decision systems over networks," *arXiv preprint arXiv:1809.04587*, 2018.

[24] J.-G. Dong and L. Qiu, "Flocking of the cucker-smale model on general digraphs," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 5234–5239, 2017.