Decisions under Binary Messaging over Adaptive Networks

Stefano Marano
DIEM, University of Salerno,
via Giovanni Paolo II 132, I-804084 Fisciano (SA), Italy
E-mail: marano@unisa.it

Ali H. Sayed
Ecole Polytechnique Federale de Lausanne EPFL,
School of Engineering, CH-1015 Lausanne, Switzerland
E-mail: ali.sayed@epfl.ch

***Abstract*—We consider an adaptive network made of interconnected agents engaged in a binary decision task. It is assumed that the agents cannot deliver full-precision messages to their neighbors, but only binary messages. For this scenario, a modified version of the ATC diffusion rule for the agent state evolution is proposed with improved decision performance under adaptive learning scenarios. An approximate analytical characterization of the agents' state is derived, giving insight into the network behavior at steady-state and enabling numerical computation of the decision performance. Computer experiments show that the analytical characterization is accurate for a wide range of the parameters of interest.**

I. INTRODUCTION

Inferential sensor networks designed to operate in nonstationary environments have received considerable attention in recent years, due to their wide range of applications [1], [2]. It has been recognized that diffusion protocols with constant step-size ensure good properties of continuous adaptation and learning over such distributed networks, enabling the system to continuously track drifts in the underlying state of nature [3]–[5]. One operative modality of the diffusion scheme is the adapt-then-combine (ATC) rule, in which each agent of the network first updates (adapts) its status exploiting the new data available at the current time, and then makes the status available to the neighbor agents for exploiting the node diversity (combination) [2]. The combination stage requires the exchange of messages among the agents and the common assumption is that these messages are delivered with full precision. In some practical cases, however, the delivering of full-precision messages is not possible because the connection links have limited capabilities.

With this motivation, we propose a modified version of the ATC rule, tailored to the assumption that the inter-agent messages can be only binary. In the binary decision problem that we address, these one-bit messages are the agent decisions based on the most recent observations, which are not biased by the past states of nature, so that the network tracks efficiently the possible modifications of the operational conditions. The modified scheme of the agents' state evolution poses new challenges arising from the interplay between the binary information coming from the neighbors, and the continuous nature of the agent state. This leads to a nontrivial distribution for the agent state in the long run, for which no exact closed-form analytical characterization is available. However, we provide reasonably simple analytical approximations, from which the decision performance can be easily obtained numerically.

The network design is addressed in the next section, and the system analysis is conducted in Sec. III. One example of application is given in Sec. IV, followed by final remarks in Sec. V. An extended version of this work can be found in [6].

The work of A. H. Sayed was supported in part by NSF grant CCF-1524250.

II. ADAPTIVE NETWORK DESIGN

We consider a connected network made of S agents (or nodes) labeled by $k = 1, \dots, S$. The symbol $\mathcal{N}_k$ denotes the set of nodes connected to agent k, including k itself, which are referred to as the neighbors of agent k. The network is engaged in a binary decision problem among two states of nature, $\mathcal{H}_0$ and $\mathcal{H}_1$. The state of nature can vary over time, and the network must track these changes.

Agent k at (discrete) time $n \geq 1$ collects the random scalar measurement $\mathbf{r}_k(n)$ and computes the log-likelihood ratio $\mathbf{x}_k(n) \triangleq \log \frac{f_{\mathbf{r},1}(\mathbf{r}_k(n))}{f_{\mathbf{r},0}(\mathbf{r}_k(n))}$, where $f_{\mathbf{r},h}(\mathbf{r}_k(n))$ denotes the probability density function (PDF) of $\mathbf{r}_k(n)$ under hypothesis $\mathcal{H}_h$, $h = 0, 1$. Measurements are assumed to be absolutely continuous random variables admitting a density, and the log-likelihood is assumed to exist. Let $\mathbf{y}_k(n)$ be the state variable at node k and time n. We consider the adaptive system described by the following equations:

$$\mathbf{v}_k(n) = (1-\mu)\,\mathbf{y}_k(n-1) + \mu\,\mathbf{x}_k(n), \tag{1a}$$

$$\mathbf{y}_k(n) = a_k \mathbf{v}_k(n) + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{N} a_{k\ell}\,\widetilde{\mathbf{x}}_\ell(n), \qquad n \geq 1, \tag{1b}$$

in which $\widetilde{\mathbf{x}}_\ell(n)$ is a one-bit quantization of $\mathbf{x}_\ell(n)$, namely,

$$\widetilde{\mathbf{x}}_\ell(n) \triangleq \mathcal{I}(\mathbf{x}_\ell(n) \geq \gamma_{\rm loc})\,\mathbb{E}_1\mathbf{x} + \mathcal{I}(\mathbf{x}_\ell(n) < \gamma_{\rm loc})\,\mathbb{E}_0\mathbf{x}, \tag{2}$$

where $\gamma_{\rm loc}$ is a local quantization threshold, $\mathcal{I}$ denotes the indicator function, and $\mathbb{E}_h$ is the expectation operator under hypothesis $\mathcal{H}_h$. The quantized variable $\widetilde{\mathbf{x}}_\ell(n)$ can be regarded as the local decision made by agent ℓ at time n, exploiting only the current observation $\mathbf{r}_\ell(n)$. If $\gamma_{\rm loc} = 0$, this decision obeys the maximum-likelihood optimality criterion [7].

The parameter μ appearing in (1a) — known as the "step-size" in the adaptive network literature — scales the fresh measurements relative to the accumulated state of the agent. Typically, $\mu \ll 1$, and $\mu = 0.1$ is a common figure in our experiments. The combination coefficients $\{a_{k\ell}\}$ appearing in (1b) are nonnegative scalars that satisfy $\sum_{\ell=1}^{N} a_{k\ell} = 1$, for all $k = 1\ldots, S$. Clearly, $a_{k\ell} = 0$ if agents k and ℓ are not neighbors. The self-combination coefficient $a_{kk} > 0$ is denoted by a_k for simplicity. When there is no danger of confusion, we also denote $\mathbf{r}_k(n)$ and $\mathbf{x}_k(n)$ simply by $\mathbf{r}$ and $\mathbf{x}$, respectively.

Note in (2) that, by definition, $\mathbb{E}_1\mathbf{x}$ is the KL divergence [8] from $f_{\mathbf{r},1}(\mathbf{r})$ to $f_{\mathbf{r},0}(\mathbf{r})$, while $\mathbb{E}_0\mathbf{x}$ is the negative of the KL divergence from $f_{\mathbf{r},0}(\mathbf{r})$ to $f_{\mathbf{r},1}(\mathbf{r})$. Divergences are nonnegative and we assume that they are strictly positive and finite.

From (1b), note also that when the state of nature changes, node ℓ informs its neighbors about this modification by delivering to them the local decision $\widetilde{\mathbf{x}}_\ell(n)$, which is not affected by the previous measurements. An alternative scheme in which the combination rule (1b) is replaced by $\mathbf{y}_k(n) = a_k\mathbf{v}_k(n) + \sum_{\ell\neq k} a_{k\ell}\widetilde{\mathbf{v}}_\ell(n)$, where $\widetilde{\mathbf{v}}_\ell(n)$ is a quantized version of $\mathbf{v}_\ell(n)$, would respond more slowly to drifts in the underlying state of nature, as discussed in [6]. Thus, the rationale behind (1a)-(1b) is to enhance the *adaptive* capabilities of the network.

The forthcoming analysis is focused on the *decision* performance of the network at steady-state ($n \to \infty$). Namely, we assume an infinitely long period of stationarity for the underlying state of nature, and consider the single-threshold decision rule of agent k: $\lim_{n\to\infty}\mathbf{y}_k(n) \geq \gamma \Rightarrow$ decide $\mathcal{H}_1$, $\lim_{n\to\infty}\mathbf{y}_k(n) < \gamma \Rightarrow$ decide $\mathcal{H}_0$, where γ is a given *decision* threshold [not to be confused with the quantization threshold $\gamma_{\rm loc}$ in (2)]. The goal is to compute the false-alarm and detection probability of agent k, defined as $P_f \triangleq \mathbb{P}_0(\lim_{n\to\infty}\mathbf{y}_k(n) \geq \gamma)$ and $P_d \triangleq \mathbb{P}_1(\lim_{n\to\infty}\mathbf{y}_k(n) \geq \gamma)$, where $\mathbb{P}_h$ is the probability operator under hypothesis $\mathcal{H}_h$.

When the state of nature is constant, the observations $\{\mathbf{r}_k(n)\}$, and consequently the log-likelihoods $\{\mathbf{x}_k(n)\}$, are assumed i.i.d. (independent and identically distributed) across all sensors $k = 1,\ldots,S$, and time indices $n \geq 1$.

III. System Analysis

A. Steady-State Distribution of the Agents

Iterating (1a)-(1b), simple algebraic manipulations yield an explicit formula for $\mathbf{y}_k(n)$, as follows:

$$\mathbf{y}_k(n) = \underbrace{\eta_k^n\,\mathbf{y}_k(0)}_{\text{transient}} + \underbrace{\mu\,a_k\sum_{i=1}^{n}\eta_k^{i-1}\,\mathbf{x}_k(n-i+1)}_{\triangleq \mathbf{u}_k(n)} + \underbrace{\sum_{i=1}^{n}\sum_{\ell=1}^{S}\eta_k^{i-1}\,c_{k\ell}\,\widetilde{\mathbf{x}}_\ell(n-i+1)}_{\triangleq \mathbf{z}_k(n)}, \tag{3}$$

where $\eta_k \triangleq (1-\mu)a_k$, and $c_{k\ell} = a_{k\ell}$ for $k \neq \ell$, while $c_{kk} = 0$. Now since $0 < \eta_k < 1$, the transient contribution in (3) disappears in the long run, and the decision properties of the system will depend on the limiting distribution of the *continuous* component $\mathbf{u}_k(\infty) \triangleq \lim_{n\to\infty}\mathbf{u}_k(n)$ and the *discrete* component $\mathbf{z}_k(\infty) \triangleq \lim_{n\to\infty}\mathbf{z}_k(n)$.

Consider first $\mathbf{u}_k(\infty)$, and let $F_{\mathbf{u},h}(u)$, $u \in \Re$, be its cumulative distribution function (CDF) under $\mathcal{H}_h$. Let $\Phi_{\mathbf{x},h}(t) \triangleq \log\mathbb{E}_h e^{j\,t\,\mathbf{x}}$, $t \in \Re$, be the log-characteristic functions of $\mathbf{x}$ under hypothesis $\mathcal{H}_h$, where $j = \sqrt{-1}$. We make the assumptions that $\mathbf{x}$ admits a density, that $\Phi_{\mathbf{x},h}(t)$ is known and that it can be expanded in power series with radius of convergence $0 < \tau_{\mathbf{x},h} \leq \infty$. Namely, for $|t| < \tau_{\mathbf{x},h}$, we have the identity $\Phi_{\mathbf{x},h}(t) = \sum_{n=1}^{\infty}\varphi_{n,h}\,t^n$.

Theorem 1: *If $\tau_{\mathbf{x},h} = \infty$, the CDF $F_{\mathbf{u},h}(u)$ of $\mathbf{u}_k(\infty)$ admits the following representation: for $u \in \Re$,*

$$F_{\mathbf{u},h}(u) = \frac{1}{2} - \frac{1}{2\pi j}\int_{-\infty}^{\infty}\exp\left\{\sum_{n=1}^{\infty}\varphi_{n,h}\frac{(\mu\,a_k t)^n}{1-\eta_k^n} - jut\right\}\frac{dt}{t}.$$

Sketch of the proof: A detailed proof is provided in [6]. Here we illustrate the main ideas. Let $\Phi_{\mathbf{u},h}(t)$ be the log-characteristic function of $\mathbf{u}_k(\infty)$. A known inversion formula to obtain the CDF $F_{\mathbf{u},h}(u)$ from the corresponding log-characteristic function $\Phi_{\mathbf{u},h}(t)$ is [9]: $F_{\mathbf{u},h}(u) = \frac{1}{2} - \frac{1}{2\pi j}\int_{-\infty}^{\infty} e^{\Phi_{\mathbf{u},h}(t)-jut}\,dt/t$, and therefore the proof reduces to showing that $\Phi_{\mathbf{u},h}(t) = \sum_{n=1}^{\infty}\varphi_{n,h}\frac{(\mu\,a_k t)^n}{1-\eta_k^n}$. The key ingredient to show this is the following structural property of the random variable $\mathbf{u}_k(\infty)$:

$$\mathbf{x}' + \eta_k\,\frac{\mathbf{u}_k(\infty)}{\mu\,a_k} \stackrel{d}{=} \frac{\mathbf{u}_k(\infty)}{\mu\,a_k}, \tag{4}$$

where $\stackrel{d}{=}$ denotes equality in distribution, and $\mathbf{x}'$ is an independent copy of the random variable $\mathbf{x}$. From (4) we get

$$\Phi_{\mathbf{x},h}(t) + \Phi_{\mathbf{u},h}\left(\frac{\eta_k\,t}{\mu\,a_k}\right) = \Phi_{\mathbf{u},h}\left(\frac{t}{\mu\,a_k}\right), \tag{5}$$

and it can be verified by direct substitution that the functional equation (5) is satisfied by $\Phi_{\mathbf{u},h}(t) = \sum_{n=1}^{\infty}\varphi_{n,h}\frac{(\mu\,a_k t)^n}{1-\eta_k^n}$. □

Theorem 2: *For $0 < \tau_{\mathbf{x},h} \leq \infty$, the following approximation holds:*

$$F_{\mathbf{u},h}(u) \approx \frac{1}{2} - \frac{2}{\pi}\sum_{n=0}^{\bar{n}_u}\frac{1}{(2n+1)}\,\mathrm{Im}\left\{\exp\left[-j\frac{u\,\delta_u}{\mu a_k}\frac{2n+1}{2} + \Phi_{\mathbf{x},h}\left(\frac{2n+1}{2}\delta_u\right) + \sum_{m=1}^{\bar{m}_u}\frac{\eta_k^m\,\varphi_{m,h}}{1-\eta_k^m}\left(\frac{2n+1}{2}\delta_u\right)^m\right]\right\}, \tag{6}$$

where $0 < \delta_u < \tau_{\mathbf{x},h}/[\eta_k(\bar{n}_u + \frac{1}{2})]$ is sufficiently small, and the integers $\bar{n}_u$ and $\bar{m}_u$ are sufficiently large.

Proof: The proof follows from [10], [11], by exploiting (4) and (5). The details are given in [6]. □

In (6), apart from truncating the two series by choosing sufficiently large integers $\bar{n}_u$ and $\bar{m}_u$, the approximation can be controlled by a suitable choice of δ_u, as discussed in [6].

The above theorem provides an approximation for the CDF $F_{\mathbf{u},h}(u)$ of the continuous component $\mathbf{u}_k(\infty)$ in (3). Next, we focus on the CDF $F_{\mathbf{z},h}(z)$, $z \in \Re$, of the discrete component $\mathbf{z}_k(\infty)$. First, let us introduce a new random variable $\mathbf{z}_k^\star$ with the same distribution as $\mathbf{z}_k(\infty)$:

$$\begin{aligned}\mathbf{z}_k(\infty) &= \lim_{n\to\infty}\sum_{\ell=1}^{S} c_{k\ell}\sum_{i=1}^{n}\eta_k^{i-1}\,\widetilde{\mathbf{x}}_\ell(n-i+1)\\ &\overset{d}{=}\sum_{\ell=1}^{S} c_{k\ell}\sum_{i=1}^{\infty}\eta_k^{i-1}\,\widetilde{\mathbf{x}}_\ell(i)\triangleq \mathbf{z}_k^\star. \end{aligned}\tag{7}$$

Second, let us introduce a normalized version of $\widetilde{\mathbf{x}}_\ell(n)$, taking values in the set $\{-1,+1\}$:

$$\mathbf{b}_\ell(i)\triangleq\frac{2\,\widetilde{\mathbf{x}}_\ell(i)-(\mathbb{E}_1\mathbf{x}+\mathbb{E}_0\mathbf{x})}{\mathbb{E}_1\mathbf{x}-\mathbb{E}_0\mathbf{x}}.\tag{8}$$

Using (8), we rewrite $\mathbf{z}_k^\star$ as follows:

$$\mathbf{z}_k^\star=\sum_{\ell=1}^{S}\frac{c_{k\ell}}{1-\eta_k}\,\frac{(\mathbb{E}_1\mathbf{x}+\mathbb{E}_0\mathbf{x})+(\mathbb{E}_1\mathbf{x}-\mathbb{E}_0\mathbf{x})\,\mathbf{z}_{k\ell}}{2},\tag{9}$$

where

$$\mathbf{z}_{k\ell}=(1-\eta_k)\sum_{i=1}^{\infty}\eta_k^{i-1}\mathbf{b}_\ell(i).\tag{10}$$

The series in (10) is known as Bernoulli convolution [12], and we proceed now to obtain a suitable approximation for its distribution, which in turn gives an approximation for the distribution of $\mathbf{z}_k^\star$.

Let us assume that $\mathcal{H}_1$ is in force and suppose that $p_d \triangleq \mathbb{P}_1(\mathbf{x}\geq\gamma_{\rm loc})$ is sufficiently large, where p_d is the agent detection probability obtained by exploiting only the current observation. By introducing an integer ω_k, we split the series on the right-hand side of (10) into two parts, as follows

$$\mathbf{z}_{k\ell}=(1-\eta_k)\sum_{i=1}^{\omega_k}\eta_k^{i-1}\mathbf{b}_\ell(i)+(1-\eta_k)\sum_{i=\omega_k+1}^{\infty}\eta_k^{i-1}\mathbf{b}_\ell(i).\tag{11}$$

The second summand on the right-hand side of (11) can be bounded as:

$$\underbrace{-\eta_k^{\omega_k}\leq}_{\substack{=-\eta_k^{\omega_k}\\ \text{iff }\{\mathbf{b}_\ell(i)=-1,\,\forall i>\omega_k\}}}\ (1-\eta_k)\sum_{i=\omega_k+1}^{\infty}\eta_k^{i-1}\mathbf{b}_\ell(i)\ \underbrace{\leq\eta_k^{\omega_k}}_{\substack{=\eta_k^{\omega_k}\\ \text{iff }\{\mathbf{b}_\ell(i)=+1,\,\forall i>\omega_k\}}}\tag{12}$$

and, because of the assumption that p_d is large, namely "+1" is the most likely outcome of $\mathbf{b}_\ell(i)$, the expression in (12) can be approximated by its upper bound $\eta_k^{\omega_k}$. From (11), this yields

$$\mathbf{z}_{k\ell}\approx(1-\eta_k)\sum_{i=1}^{\omega_k}\eta_k^{i-1}\mathbf{b}_\ell(i)+\eta_k^{\omega_k}\triangleq\widehat{\mathbf{z}}_{k\ell},\tag{13}$$

with $\widehat{\mathbf{z}}_{k\ell}\geq\mathbf{z}_{k\ell}$. In view of (12), the approximation in (13) entails a bounded error, namely, $0\leq\widehat{\mathbf{z}}_{k\ell}-\mathbf{z}_{k\ell}\leq 2\eta_k^{\omega_k}$. Inserting $\widehat{\mathbf{z}}_{k\ell}$ into (9) in place of $\mathbf{z}_{k\ell}$ gives an approximation for $\mathbf{z}_k^\star$, which we denote by $\widehat{\mathbf{z}}_k^\star$. Now, to control the approximation, let us fix a sufficiently small $\epsilon_k>0$. Simple algebra shows that

$$\omega_k\geq\frac{\log\frac{\mathbb{E}_1\mathbf{x}-\mathbb{E}_0\mathbf{x}}{\epsilon_k(1-\eta_k)}}{\log\frac{1}{\eta_k}}\qquad\Rightarrow\qquad 0\leq\widehat{\mathbf{z}}_k^\star-\mathbf{z}_k^\star\leq\epsilon_k,\tag{14}$$

and we therefore select ω_k as the smallest integer that verifies the first inequality in (14) for a sufficiently small ϵ_k.

Approximation (13) consists of replacing $\mathbf{z}_{k\ell}$ by $\widehat{\mathbf{z}}_{k\ell}$, which is a discrete random variable taking on 2^{ω_k} values. These values and the corresponding probabilities can be easily computed, and this gives the probability mass function (PMF) of $\widehat{\mathbf{z}}_{k\ell}$. Indeed, from (13) we see that each realization of $\widehat{\mathbf{z}}_{k\ell}$ corresponds to a different pattern of ±1 in the string $\mathbf{b}_\ell(1),\dots,\mathbf{b}_\ell(\omega_k)$, and the probability of this realization depends on the number of "+1" in the string, as follows: $p_d^{\#\text{of}+1}(1-p_d)^{\#\text{of}-1}$. Because $c_{k\ell}\neq0$ for $\ell\in\mathcal{N}_k\backslash\{k\}$, replacing $\mathbf{z}_{k\ell}$ by $\widehat{\mathbf{z}}_{k\ell}$ in (9), we see that the resulting random variable $\widehat{\mathbf{z}}_k^\star$ is the sum of $|\mathcal{N}_k|-1$ variables, each with alphabet of cardinality 2^{ω_k}, implying that $\widehat{\mathbf{z}}_k^\star$ takes on at most $2^{\omega_k(|\mathcal{N}_k|-1)}$ values. The PMF of $\widehat{\mathbf{z}}_k^\star$ can be obtained by convolving the $|\mathcal{N}_k|-1$ PMFs of $\widehat{\mathbf{z}}_{k\ell}$, $\ell\in\mathcal{N}_k\setminus\{k\}$, which are random variables independent of each other. This requires $|\mathcal{N}_k|-2$ convolutions. Implementing these convolutions, we get the set of realizations of $\widehat{\mathbf{z}}_k^\star$, say $\{\widehat{z}_{i,h}\}$, and the corresponding probabilities, say $\{\widehat{\nu}_{i,h}\}$, with $i=1,\dots,L\leq2^{\omega_k(|\mathcal{N}_k|-1)}$.

We have thus derived the CDF $\widehat{F}_{\mathbf{z},h}(z)$ of $\widehat{\mathbf{z}}_k^\star$, which is a staircase function with steps of size $\{\widehat{\nu}_{i,h}\}$ at the L points $\{\widehat{z}_{i,h}\}$. Since $\widehat{\mathbf{z}}_k^\star$ is an approximation of $\mathbf{z}_k^\star$, the CDF $\widehat{F}_{\mathbf{z},h}(z)$ is used as an approximation of the CDF of $\mathbf{z}_k^\star$. This latter is the same of the CDF $F_{\mathbf{z},h}(z)$ of $\mathbf{z}_k(\infty)$, as shown in (7). Note that, because of the inequalities $0\leq\widehat{\mathbf{z}}_k^\star-\mathbf{z}_k^\star\leq\epsilon_k$ in (14), we have the following relationship between $F_{\mathbf{z},h}(z)$ and $\widehat{F}_{\mathbf{z},h}(z)$:

$$F_{\mathbf{z},h}(z-\epsilon_k)\leq\widehat{F}_{\mathbf{z},h}(z)\leq F_{\mathbf{z},h}(z),\qquad z\in\Re.\tag{15}$$

We have characterized the CDFs $F_{\mathbf{u},h}(u)$ and $F_{\mathbf{z},h}(z)$ of the continuous and of the discrete components $\mathbf{u}_k(\infty)$ and $\mathbf{z}_k(\infty)$, respectively. Consider now the state $\mathbf{y}_k(\infty)$ of agent k at steady-state. Neglecting the transient component in (3), we have $\mathbf{y}_k(\infty)=\mathbf{u}_k(\infty)+\mathbf{z}_k(\infty)$, where the two random variables at the right-hand side are independent. Then, the CDF $F_{\mathbf{y},h}(y)$ of the variable $\mathbf{y}_k(\infty)$ is given by the convolution formula [13], where $f_{\mathbf{u},h}(u)$ is the density of $\mathbf{u}(\infty)$:

$$F_{\mathbf{y},h}(y)=\int_{-\infty}^{\infty}F_{\mathbf{z},h}(\xi)\,f_{\mathbf{u},h}(y-\xi)\,d\xi.\tag{16}$$

Using (15) in (16), we have:

$$F_{\mathbf{y},h}(y-\epsilon_k)=\int_{-\infty}^{\infty}F_{\mathbf{z},h}(\xi)\,f_{\mathbf{u},h}(y-\epsilon_k-\xi)\,d\xi\tag{17a}$$

$$=\int_{-\infty}^{\infty}F_{\mathbf{z},h}(\xi-\epsilon_k)\,f_{\mathbf{u},h}(y-\xi)\,d\xi\tag{17b}$$

$$\leq\int_{-\infty}^{\infty}\widehat{F}_{\mathbf{z},h}(\xi)\,f_{\mathbf{u},h}(y-\xi)\,d\xi\tag{17c}$$

$$\leq\int_{-\infty}^{\infty}F_{\mathbf{z},h}(\xi)\,f_{\mathbf{u},h}(y-\xi)\,d\xi=F_{\mathbf{y},h}(y).\tag{17d}$$

If ϵ_k is small enough such that $f_{\mathbf{u},h}(u)$ is almost constant over intervals of length ϵ_k, the integral in (17a) is almost equal to the integral in (17d), and therefore $F_{\mathbf{y},h}(y)$ is almost equal to the integral in (17c). This latter reduces to a finite summation, because $\widehat{F}_{\mathbf{z},h}(z)$ is a staircase function with finite number of steps, and we arrive at the approximation

$$F_{\mathbf{y},h}(y) \approx \sum_{i=1}^{L} \widehat{\nu}_{i,h}\, F_{\mathbf{u},h}(y - \widehat{z}_{i,h}). \tag{18}$$

B. Computational Considerations

The number L of summands in (18) may be as large as $2^{\omega_k(|\mathcal{N}_k|-1)}$, which can be problematic for numerical implementations. The reduction of the computational burden to obtain $F_{\mathbf{y},h}(y)$ is now in order. Recall that we are working under $\mathcal{H}_1$, with the assumption that $1 - p_d \ll 1$.

The cardinality 2^{ω_k} of the random variable $\widehat{\mathbf{z}}_{k\ell}$ in (13) can be substantially reduced as follows. Since $1 - p_d$ is small, the number of occurrences of "-1" in the string $\mathbf{b}_\ell(1), \ldots, \mathbf{b}_\ell(\omega_k)$ is expected to be small. Accordingly, suppose that at most two occurrences of the least likely digit "-1" occur. Then, the resulting variable takes values in an alphabet of cardinality $1 + \omega_k + \omega_k(\omega_k - 1)/2$, whose elements correspond to the patterns of ± 1 in the string $\mathbf{b}_\ell(1), \ldots, \mathbf{b}_\ell(\omega_k)$ with at most two "-1". The probability of occurrence of these patterns is given by $p_d^{\#\text{of}+1}(1-p_d)^{\#\text{of}-1}$. In order to obtain a valid random variable, the aggregate probability of patterns with more than two occurrences of "-1" that we have disregarded, must be apportioned amongst the elements of the alphabet. This apportionment is described in detail in [6] and the resulting random variable is shown in Table I, where the subindex k to ω_k and η_k is omitted. The random variable defined in Table I is used as an approximation of $\widehat{\mathbf{z}}_{k\ell}$.

value	pattern	probability
$1 - 2(1-\eta) - 2\eta(1-\eta)$	$- - + + \cdots + + +$	$(1-p_d)^2$
$1 - 2(1-\eta) - 2\eta^2(1-\eta)$	$- + - + \cdots + + +$	$(1-p_d)^2 p_d$
$\vdots$	$\vdots$	$\vdots$
$1 - 2(1-\eta) - 2\eta^{\omega-1}(1-\eta)$	$- + + + \cdots + + -$	$(1-p_d)^2 p_d^{\omega-2}$
$1 - 2(1-\eta)$	$- + + + \cdots + + +$	$(1-p_d) p_d^{\omega-1}$
$1 - 2\eta(1-\eta) - 2\eta^2(1-\eta)$	$+ - - + \cdots + + +$	$(1-p_d)^2 p_d$
$1 - 2\eta(1-\eta) - 2\eta^3(1-\eta)$	$+ - + - \cdots + + +$	$(1-p_d)^2 p_d^2$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$
$1 - 2\eta^{\omega-3}(1-\eta) - 2\eta^{\omega-2}(1-\eta)$	$+ + + + \cdots - - +$	$(1-p_d)^2 p_d^{\omega-1}$
$1 - 2\eta^{\omega-3}(1-\eta) - 2\eta^{\omega-1}(1-\eta)$	$+ + + + \cdots - + -$	$(1-p_d)^2 p_d^{\omega-2}$
$1 - 2\eta^{\omega-3}(1-\eta)$	$+ + + + \cdots - + +$	$(1-p_d) p_d^{\omega-1}$
$1 - 2\eta^{\omega-2}(1-\eta) - 2\eta^{\omega-1}(1-\eta)$	$+ + + + \cdots + - -$	$(1-p_d)^2 p_d^{\omega-2}$
$1 - 2\eta^{\omega-2}(1-\eta)$	$+ + + + \cdots + - +$	$(1-p_d) p_d^{\omega-1}$
$1 - 2\eta^{\omega-1}(1-\eta)$	$+ + + + \cdots + + -$	$(1-p_d) p_d^{\omega-1}$
1	$+ + + + \cdots + + +$	p_d^{ω}

TABLE I

APPROXIMATION OF THE RANDOM VARIABLE $\widehat{\mathbf{z}}_{k\ell}$.

Additional simplifications for lightening the computational burden to obtain $F_{\mathbf{y},h}(y)$ are possible. First, the cardinality of the random variable in Table I can be further reduced by aggregating in a single value realizations that are closer than $2\eta_k^{\omega_k}$ (if any). Second, the same type of aggregation can be operated after implementing each of the $|\mathcal{N}_k| - 2$ convolutions that correspond to the summation in (9). In both cases, the probability assigned to the aggregated symbol is the sum of the probabilities of the realizations merged together. These aggregations follow the same spirit of approximation (13), which entails an error $\leq 2\eta_k^{\omega_k}$.

So far, we have assumed that $\mathcal{H}_1$ is in force, but it can be shown that the same arguments used to obtain the CDF of $\mathbf{y}_k(\infty)$ apply under hypothesis $\mathcal{H}_0$. In this case the assumption is $p_f \triangleq \mathbb{P}_0(\mathbf{x} \geq \gamma_{\text{loc}}) \ll 1$, and the final result can be obtained by replacing p_d with $1 - p_f$ in the previous derivations, and elaborating on $-\mathbf{z}_{k\ell}$ rather than $\mathbf{z}_{k\ell}$. The details are omitted.

IV. EXAMPLE

Consider a network made of $S = 10$ agents whose topology is encoded in the connection matrix (19), where the presence of 1 in position (m, n) means that agents m and n are neighbors:

$$\begin{pmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1
\end{pmatrix}. \tag{19}$$

Let the combination coefficients $a_{k\ell}$ be as follows:

$$a_{k\ell} = \begin{cases} a, & \ell = k, \\ \frac{1-a}{|\mathcal{N}_k|-1}, & \ell \in \mathcal{N}_k \setminus \{k\}, \\ 0, & \text{otherwise}, \end{cases} \tag{20}$$

with $0 < a < 1$, and assume $\gamma_{\text{loc}} = 0$ see (2).

Suppose that the i.i.d. observations $\{\mathbf{r}_k(n)\}$ are exponentially distributed, under both hypotheses, as follows: $\mathbf{r} \sim \mathcal{E}(\lambda_h)$ under $\mathcal{H}_h$, $h = 0, 1$. The symbol $\mathbf{r} \sim \mathcal{E}(\lambda)$ means that the probability density function of $\mathbf{r}$ is $\lambda \exp(-\lambda r)$, for $r \geq 0$. We assume $0 < \lambda_1 < \lambda_0$, and define $\lambda_e \triangleq \lambda_0/\lambda_1 > 1$. The log-likelihood is $\mathbf{x}_k(n) = \lambda_1(\lambda_e - 1)\mathbf{r}_k(n) - \log \lambda_e$, and by straightforward algebra the forthcoming results follow.

Under $\mathcal{H}_0$, the log-characteristic function of $\mathbf{x}$ is

$$\Phi_{\mathbf{x},0}(t) = \log \frac{\lambda_e^{1-j\,t}}{\lambda_e - j\,t\,(\lambda_e - 1)}. \tag{21}$$

For $|t| < \tau_{\mathbf{x},0}$, this function can be expanded in series as $\Phi_{\mathbf{x},0}(t) = \sum_{n=1}^{\infty} \varphi_{n,0} t^n$, with

$$\varphi_{n,0} = \begin{cases} \left(1 - \lambda_e^{-1} - \log \lambda_e\right) j, & n = 1, \\ \frac{1}{n}\left[\left(1 - \lambda_e^{-1}\right) j\right]^n, & n > 1, \end{cases} \tag{22}$$

and the radius of convergence of the series is $\tau_{\mathbf{x},0} = \frac{\lambda_e}{\lambda_e - 1}$.

The log-characteristic function of $\mathbf{x}$ under $\mathcal{H}_1$ can be obtained by a known property of the log-likelihood [14, Eq. (90), p. 44], which yields $\Phi_{\mathbf{x},1}(t) = \Phi_{\mathbf{x},0}(t - j)$. This gives

$$\Phi_{\mathbf{x},1}(t) = \log \frac{\lambda_e^{-j\,t}}{1 - j\,t\,(\lambda_e - 1)}, \tag{23}$$

which admits the series expansion $\Phi_{\mathbf{x},1}(t) = \sum_{i=1}^{\infty} \varphi_{n,1} t^n$, $|t| < \tau_{\mathbf{x},1}$, where

$$\varphi_{n,1} = \begin{cases} (\lambda_e - 1 - \log \lambda_e)\, j, & n = 1, \\ \frac{1}{n}\left[(\lambda_e - 1)\, j\right]^n, & n > 1, \end{cases} \tag{24}$$

with radius of convergence $\tau_{\mathbf{x},1} = \frac{1}{\lambda_e - 1}$.

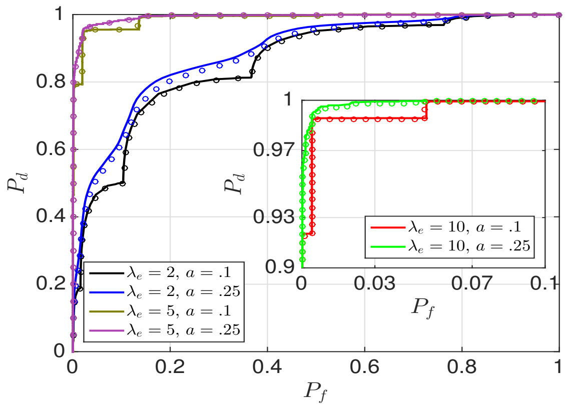

Fig. 1. Decision performance of agent $k = 3$, assuming $\mu = 0.1$ and different combinations of λ_e and a. The solid curves refer to the approximate theoretical expressions and small circles denote simulation points.

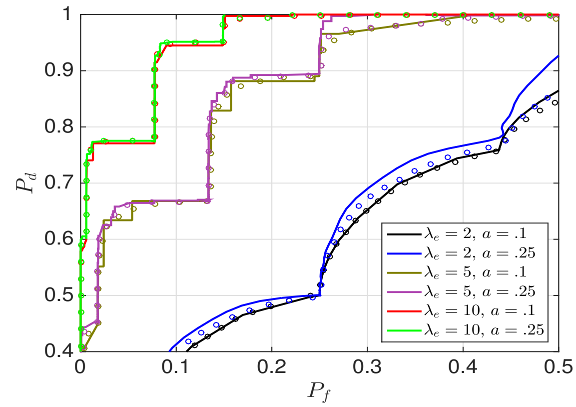

Fig. 2. Decision performance of agent $k = 9$, assuming $\mu = 0.1$ and different combinations of λ_e and a. The solid curves refer to the approximate theoretical expressions and small circles denote simulation points.

Using (21)-(24), Theorem 2 gives an approximation for $F_{\mathbf{u},h}(u)$, $h = 0, 1$, which, inserted in (18), gives an approximation for $F_{\mathbf{y},h}(y)$. In computing the sequences $\{\widehat{z}_{i,h}\}$ and $\{\widehat{\nu}_{i,h}\}$ appearing in (18) the numerical simplifications described in Sec. III-B can be applied. Using $F_{\mathbf{y},h}(y)$, $h = 0, 1$, the binary decision performance of agent k is immediately derived in the form of false alarm probability $P_f = \mathbb{P}_0(\mathbf{y}_k(\infty) \geq \gamma)$ and detection probability $P_d = \mathbb{P}_1(\mathbf{y}_k(\infty) \geq \gamma)$. The decision performance of agent k is compared to the results of computer simulations using 10^5 independent Monte Carlo runs and $n = 100$ iterations. This comparison is shown in Figs. 1 and 2 for the highly connected agent $k = 3$ and the weakly connected agent $k = 9$.

V. Final Comments

This work considers an adaptive network made of connected agents engaged in a binary decision problem, where the individual agents can process data locally with full precision, but the inter-agent messages can be only binary. The system design ensures good adaptation properties, and the theoretical analysis allows us to derive the approximate decision performance in a semi-analytical form that requires the sum of two truncated series and the numerical implementation of certain convolutions. Both procedures are reasonably easy and do not entail severe computational burden.

The analysis shows a nontrivial agent performance which arises from the interaction between the continuous component of the agent state, and the discrete component related to the messages delivered from its neighbors.

As seen in the example shown in Figs. 1 and 2, the decision performance of the network improves when the self-combination coefficient a grows. This gain in performance is paid in terms of lower adaptation capability (which improves for smaller values of a) revealing an inherent system trade-off. We also find that the performance curves are smoother for larger values of a. This can be explained by recalling that the larger a is, the less credit the agent gives to the messages coming from its neighbors, and therefore the weaker the effect of the discrete component becomes.

Computer experiments show that the accuracy of the approximation developed tends to improve when the number of the agent neighbors grows and when the self-combination coefficient a_k decreases. The approximation seems satisfying for a wide range of system parameters μ and combination coefficients $\{a_{k\ell}\}$. However, in the limiting regime $\mu \to 0$ (vanishing step-size) and $a_k \to 1$ (weak agent interactions) the approximation fails. For this scenario a different approach is provided in [6], where it is shown that a version of the central limit theorem applies, leading to a Gaussian approximation for the agents' steady-state distributions.

References

[1] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[2] ——, "Adaptation, learning, and optimization over networks," in *Foundations and Trends in Machine Learning*. Boston-Delft: NOW Publishers, 2014, vol. 7, no. 4–5, pp. 311–801.

[3] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. on Signal Processing*, vol. 59, no. 5, pp. 1917–1932, 2011.

[4] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Diffusion-based adaptive distributed detection: Steady-state performance in the slow adaptation regime," *IEEE Trans. on Information Theory*, vol. 62, no. 8, pp. 4710–4732, Aug. 2016.

[5] ——, "Distributed detection over adaptive networks: Refined asymptotics and the role of connectivity," *IEEE Trans. Signal and Inf. Process. over Networks*, vol. 2, no. 4, pp. 442–460, Dec. 2016.

[6] S. Marano and A. H. Sayed, "Detection under one-bit messaging over adaptive networks," *submitted*. Also available as arXiv:1803.06725 [cs.MA], March 2018.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. New Jersey: Prentice Hall, 1998.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: Wiley-Interscience, 2006.

[9] J. Gurland, "Inversion formulae for the distribution of ratios," *The Annals of Statistics*, vol. 19, pp. 228–237, 1948.

[10] N. G. Shephard, "Inversion formulae for the distribution of ratios," *J. Statist. Comput. Simul.*, vol. 39, pp. 37–46, 1991.

[11] R. B. Davies, "Numerical inversion of a characteristic function," *Biometrika*, vol. 60, no. 2, pp. 415–417, 1973.

[12] P. Erdös, "On a family of symmetric Bernoulli convolutions," *Amer. J. Math*, vol. 61, pp. 974–975, 1939.

[13] W. Feller, *An Introduction to Probability and Its Applications, Volume 2*. New York: John Wiley & Sons, 1971.

[14] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part I*. New York: John Wiley & Sons, 1968 (reprinted, 2001).