Detection over Diffusion Networks: Asymptotic Tools for Performance Prediction and Simulation

(Invited Paper)

Vincenzo Matta*, Paolo Braca†, Stefano Marano*, Ali H. Sayed‡
*DIEM, University of Salerno, Fisciano (SA), Italy, {vmatta, marano}@unisa.it
†NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy, paolo.braca@cmre.nato.int
‡Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA, sayed@ee.ucla.edu

***Abstract*—Exploiting recent progress [1]–[4] in the characterization of the detection performance of diffusion strategies over adaptive multi-agent networks: i) we present two theoretical approximations, one based on asymptotic normality and the other based on the theory of exact asymptotics; and ii) we develop an efficient simulation method by tailoring the importance sampling technique to diffusion adaptation. We show that these theoretical and experimental tools complement each other well, with their combination offering a substantial advance for a reliable *quantitative* detection-performance assessment. The analysis provides insight into the interplay between the network topology, the combination weights, and the inference performance, revealing the *universal* behavior of diffusion-based detectors over adaptive networks.**

***Index Terms*—Distributed detection, adaptive network, diffusion, large deviations, exact asymptotics, importance sampling.**

I. Introduction, Background and Main Results

The paradigm of distributed detection over adaptive networks can be exemplified as follows. Several spatially dispersed agents, linked together by a given network topology, continually sense streaming data from the environment. The statistical properties of the data depend upon an unknown binary state of nature. At *any* time instant, *all* the individual agents arrive at their own decision about the actual state of nature by implementing some distributed processing strategy that relies on using their own observations along with the information exchanged through cooperation with their neighbors. The goal is for the networked agents to achieve a detection performance that is superior to what they would attain if the agents acted individually without cooperation. In addition, when operating in real-world scenarios, it is essential to endow the agents with adaptive capabilities in order to enable them to track the drifts in the statistical conditions in the data, in the environmental conditions, and in the network topology, among other possibilities. Therefore, the agents must be able to track such drifts promptly (transient performance), while concurrently ensuring small detection error probabilities when the underlying statistics remain stable for a sufficiently long time (steady-state performance).

The literature about distributed detection is abundant. We refer the reader to [5]–[7] as entry points on the subject. In particular, we shall focus on *fully-decentralized* detection problems, where there is no fusion center and only local interactions among the agents are allowed. For such scenarios, solutions based on decentralized consensus strategies with decaying step-size have been proposed in [8]–[12], and the detection performance of these algorithms has been characterized in different asymptotic frameworks [8]–[12]. In comparison, a distinguishing feature of our work resides in the emphasis on *adaptive* solutions. To enable continuous adaptation, it has been shown that diffusion strategies with *constant* step-sizes (as opposed to the decaying step-size employed in consensus implementations [8]–[12]) are successful in combining the requirements of tracking and learning, and offer wider stability ranges and enhanced steady-state performance. While several results are available for diffusion strategies in connection to their *estimation* performance [13]–[16], only a handful of results have been obtained for the *detection* performance. For example, in [17] the problem of using diffusion algorithms for detection purposes has been considered, with reference to a Gaussian problem. More recently, the general problem of distributed detection over adaptive networks has been addressed in [1]–[4], where the fundamental scaling laws governing the steady-state error probabilities in the asymptotic regime of small step-sizes have been established.

The work of A. H. Sayed was supported in part by NSF grants CCF-1524250 and ECCS-1407712.

A. Diffusion Implementation

Assume that, at time n, the k-th network agent, for $k = 1, 2, \ldots, S$, computes its local statistic $\boldsymbol{x}_k(n)$ (the observation itself, or a suitable function thereof), whose expectation and variance will be denoted by $\mathbb{E}[\boldsymbol{x}]$ and σ_x^2, respectively. Data are assumed to be spatially and temporally independent and identically distributed (i.i.d.). We start by examining a useful diffusion implementation, namely, the ATC (Adapt-then-Combine) implementation, which is preferred since it possesses some inherent advantages in terms of a slightly improved mean-square-error performance relative to other forms [18]. In ATC diffusion, the state of agent k at time n is adjusted as follows [15], [16]:

$$\boldsymbol{y}_k(n) = \sum_{\ell=1}^{S} a_{k,\ell}\{\boldsymbol{y}_\ell(n-1) + \mu[\boldsymbol{x}_\ell(n) - \boldsymbol{y}_\ell(n-1)]\}, \qquad (1)$$

where $0 < \mu \ll 1$ is a small step-size parameter, and the combination weights $\{a_{k,\ell}\}$ are nonnegative and *convex*. It is convenient to collect the combination weights into a square matrix $A = [a_{k,\ell}]$, which turns out to be right-stochastic, namely, $a_{k,\ell} \geq 0, \quad A\mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ is a column-vector with all entries equal to 1. Let us denote the n-th power of A by $B_n = [b_{k,\ell}(n)] \triangleq A^n$. Throughout this article, we assume that A has second largest eigenvalue magnitude strictly less than one (a condition automatically satisfied when there is always a path with nonzero weights between any pair of nodes, and at least one node in the network has a self-loop, i.e., $a_{k,k} > 0$ for some agent k, see [13], [14]), which yields [19]:

$$b_{k,\ell}(n) \overset{n\to\infty}{\longrightarrow} p_\ell \quad \Leftrightarrow \quad B_n \overset{n\to\infty}{\longrightarrow} \mathbb{1}p, \qquad (2)$$

where the limiting combination (row) vector $p = [p_1, p_2, \ldots, p_S]$ (the Perron eigenvector) satisfies $pA = p$, $p_\ell > 0$, and $\sum_{\ell=1}^{S} p_\ell = 1$ — see, e.g., [13], [14].

Following a well-established approach in the adaptation literature [18] we now focus on i) the steady-state properties (as $n \to \infty$), and ii) the small step-size regime ($\mu \to 0$). We start by examining the steady-state behavior of $\boldsymbol{y}_k(n)$ for a given step-size μ. It is convenient to evaluate explicitly the recursion in (1), and introduce the following random variable:

$$\boldsymbol{y}_k^\star(n) = \sum_{i=1}^{n}\sum_{\ell=1}^{S}\mu(1-\mu)^{i-1}b_{k,\ell}(i)\boldsymbol{x}_\ell(i), \tag{3}$$

which, since observations are i.i.d. across time, *shares the same distribution* of the diffusion output $\boldsymbol{y}_k(n)$ with null initial state $\boldsymbol{y}_\ell(0)$ (a legitimate assumption since we are focusing on the *steady-state* behavior). In [1], [2] it has been shown that the diffusion output $\boldsymbol{y}_k(n)$ converges in distribution to a limiting random variable $\boldsymbol{y}_{k,\mu}^\star$ that can be represented as:

$$\boxed{\boldsymbol{y}_{k,\mu}^\star \triangleq \sum_{i=1}^{\infty}\sum_{\ell=1}^{S}\mu\,(1-\mu)^{i-1}b_{k,\ell}(i)\boldsymbol{x}_\ell(i)} \tag{4}$$

Moreover, it is possible to prove that $\mathbb{E}[\boldsymbol{y}_{k,\mu}^\star] = \mathbb{E}[\boldsymbol{x}]$, and that

$$\sigma_{k,\mu}^2 \triangleq \mathrm{VAR}[\boldsymbol{y}_{k,\mu}^\star] = \sigma_x^2\sum_{i=1}^{\infty}\sum_{\ell=1}^{S}\mu^2(1-\mu)^{2(i-1)}b_{k,\ell}^2(i). \tag{5}$$

B. The Detector

Each agent in the network performs a binary hypothesis test in an *adaptive* and *fully distributed* manner. In this setting, the agents collect an increasing amount of streaming observations, whose statistical properties depend upon an *unknown* binary state of nature, which is represented by a pair of hypotheses, say, $\mathcal{H}_0$ and $\mathcal{H}_1$. The local statistics $\boldsymbol{x}_k(n)$ are spatially and temporally i.i.d. conditioned on the hypothesis. Throughout the article we assume that $\mathbb{E}_0[\boldsymbol{x}] \neq \mathbb{E}_1[\boldsymbol{x}]$ (which means, when $\boldsymbol{x}$ is a log-likelihood ratio, that the detection problem is not singular [20]), and, without loss of generality, that $\mathbb{E}_0[\boldsymbol{x}] < \mathbb{E}_1[\boldsymbol{x}]$. At time n, the k-th agent is tasked to produce a decision about the state of nature, based upon its state value $\boldsymbol{y}_k(n)$. According to the theoretical considerations made in [1]–[4], the decision rule is of the form:

$$\boxed{\boldsymbol{y}_k(n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lesseqgtr}} \gamma} \tag{6}$$

for some threshold value γ. The performance of this test can be assessed in terms of its limiting Type-I and Type-II error probabilities, which are defined as, respectively: $\alpha_{k,\mu} \triangleq \mathbb{P}_0[\boldsymbol{y}_{k,\mu}^\star > \gamma]$, and $\beta_{k,\mu} \triangleq \mathbb{P}_1[\boldsymbol{y}_{k,\mu}^\star \le \gamma]$.

C. Main Result

The statistical characterization of the steady-state output $\boldsymbol{y}_{k,\mu}^\star$ is usually a formidable task. However, for the case of Gaussian observations, where the fact that linear combinations preserve Gaussianity implies that $\boldsymbol{y}_{k,\mu}^\star$ is Gaussian, for each value of $\mu \in (0,1)$ and for each value of $k \in \{1,2,\dots,S\}$, we have [17]:

$$\mathbb{P}[\boldsymbol{y}_{k,\mu}^\star > \gamma] = Q\left(\frac{\gamma - \mathbb{E}[\boldsymbol{x}]}{\sigma_{k,\mu}}\right) \quad \text{(for Gaussian observations)}, \tag{7}$$

where $Q(\cdot)$ is the standard $Q-$function for the Gaussian distribution. For fixed values of $\mathbb{E}[\boldsymbol{x}]$ and σ_x^2, and for $\gamma = 0$, the corresponding probability curves of the different network agents are depicted in Fig. 1, with reference to the network represented in the inset plot, and to the uniform averaging combination rule [13], [14]. Three main observations emerge:

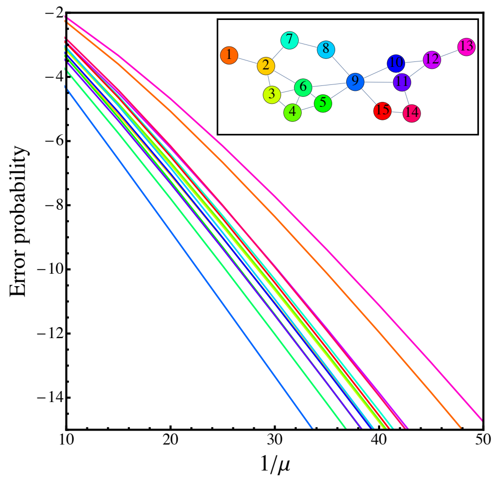

Fig. 1. Steady-state error probabilities for the fully Gaussian case according to (7).

O1) The error probabilities decay exponentially fast as functions of $1/\mu$, approximately with the same decaying rate.
O2) Different agents exhibit different performance, depending on their connectivity.
O3) In accordance with the exponential decay, small changes in μ can lead to substantial variation in the error proabilities. Therefore, *very* small error probabilities can be achieved.

It is useful to remark that, for a *non-adaptive* system (with diminishing step-size), it has been shown that the error probabilities decay exponentially fast *as functions of time*. Accordingly, such probabilities might reach "astronomically" small values as time elapses. In our *adaptive* setting (with constant step-size), the situation is different since the error probabilities stabilize as time elapses. The possibility of reducing the error probabilities is now related to decreasing the step-size. From these observations, at least two fundamental questions arise:
Q1) Figure 1 has been generated by using the exact analytical formulas that are available for the Gaussian case. Are there analytical formulas and tools for more general cases?
Q2) Are the observed features representative of a *universal* behavior of adaptive detection over diffusion networks?

The answer to both questions turns out to be in the affirmative. Specifically, Q2 is answered through Q1, since only *analytical* formulas can ascertain a *universal* behavior. With regards to Q1, it is virtually impossible to obtain exact characterization of the error curves for general detection problems. One must resort to an *asymptotic* (in the small step-size regime) analysis. In Sect. II we present two asymptotic tools. The first tool is based upon a Central-Limit-Theorem (CLT) Gaussian approximation, which is exact in the *small*-deviations regime where the threshold γ scales as $\sqrt{\mu}$, and the error probabilities become stable (rather than vanishing) as μ goes to zero. For detection applications, the latter regime is particularly useful in the framework of *locally* optimum detection [9], [21], where the distance (i.e., the signal-to-noise ratio) between the hypotheses is small. In contrast, the Gaussian asymptotics fail to predict the correct tail behavior when the error probabilities vanish as the step-size decreases. The second tool used in the analysis is based upon a *large*-deviations analysis, which is able to capture the correct tail behavior, a property that turns out to be especially relevant in the light of the previous observation O3. In this respect, the performance-prediction tools that will arise from the analysis will complement each other well. However, there is still an important gap. In fact, since both approximations are *asymptotic*, they must be complemented by proper tools to perform simulation. In the light of observation O3, Monte Carlo simulations become prohibitive. This difficulty can be overcome by resorting to importance sampling techniques. Nevertheless, in order to design properly the importance sampling simulator, it is crucial to have knowledge of the asymptotic tail behavior of the involved random variables. In Sect. III, we shall tackle and solve this problem. As a result, we will end up with an ensemble of tools (different theoretical approximations, and the importance sampling technique), whose joint application enables an accurate and reliable quantitative analysis of adaptive detection over diffusion networks.

II. Asymptotic Tools to Performance Prediction

A. Normal Approximations

In [1], [2] it is shown that

$$\frac{\boldsymbol{y}^\star_{k,\mu}-\mathbb{E}[\boldsymbol{x}]}{\sqrt{\mu\sigma^2_{\rm lim}}} \overset{\mu\to 0}{\rightsquigarrow} \mathcal{N}(0,1) \tag{8}$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution, the symbol $\rightsquigarrow$ denotes convergence in distribution, and where

$$\sigma^2_{\rm lim} \triangleq \lim_{\mu\to 0}\sigma^2_{k,\mu}/\mu = \sigma^2_x/2\sum_{\ell=1}^{S}p^2_\ell. \tag{9}$$

In view of Slutsky's Theorem, Eqs. (8) and (9) imply that [20]:

$$\boxed{\frac{\boldsymbol{y}^\star_{k,\mu}-\mathbb{E}[\boldsymbol{x}]}{\sigma_{k,\mu}} \overset{\mu\to 0}{\rightsquigarrow} \mathcal{N}(0,1)} \tag{10}$$

which amounts to using (7) as an approximation of the diffusion-output performance. While the formulations in (8) and (10) are asymptotically equivalent in distribution, we see that (10) can be interpreted as a refinement of the first-order approximation in (8), with the limiting variance being replaced by the actual variances (which depend upon the agent index k). Accordingly, it is expected that (10) offers a better performance in practice, since the actual variances might embody useful information about the connectivity features of the agents.

B. Large Deviations and Exact Asymptotics

We first introduce some useful quantities. The Logarithmic Moment Generating Function (LMGF) of the local statistics $\boldsymbol{x}_k(n)$ is $\psi(t) \triangleq \ln\mathbb{E}[e^{t\boldsymbol{x}_k(n)}]$. Throughout the article, we assume that $\psi(t)<+\infty$ for all $t\in\mathbb{R}$. Likewise, the LMGF of the steady-state variable $\boldsymbol{y}^\star_{k,\mu}$ is defined as $\phi_{k,\mu}(t)\triangleq\ln\mathbb{E}[e^{t\boldsymbol{y}^\star_{k,\mu}}]$. We will be primarily concerned with the limiting scaled LMGF:

$$\phi(t)\triangleq\lim_{\mu\to 0}\mu\,\phi_{k,\mu}(t/\mu)=\sum_{\ell=1}^{S}\int_0^{p_\ell t}\frac{\psi(\tau)}{\tau}d\tau, \tag{11}$$

where the existence and the value of the above limit have been ascertained in [3], [4]. Finally, we introduce the Fenchel-Legendre transform of $\phi(t)$: $\Phi(\gamma)\triangleq\sup_{t\in\mathbb{R}}[\gamma t-\phi(t)]$, and the domain where $\Phi(\gamma)$ is finite: $\mathcal{D}_\Phi=\{\gamma\in\mathbb{R}:\ \Phi(\gamma)<\infty\}$, whose interior will be denoted by $\mathcal{D}^o_\Phi$ [22].

It was established in [1]–[4] that, as the step-size μ vanishes, the steady-state random variable $\boldsymbol{y}^\star_{k,\mu}$ obeys a Large Deviations Principle (LDP) with rate function $\Phi(\gamma)$, namely, that, for $\gamma\in\mathcal{D}^o_\Phi$:

$$\boxed{\begin{array}{lcll}\lim_{\mu\to 0}\mu\ln\mathbb{P}[\boldsymbol{y}^\star_{k,\mu}>\gamma] &=& -\Phi(\gamma), & \gamma>\mathbb{E}[\boldsymbol{x}]\\ \lim_{\mu\to 0}\mu\ln\mathbb{P}[\boldsymbol{y}^\star_{k,\mu}\le\gamma] &=& -\Phi(\gamma), & \gamma<\mathbb{E}[\boldsymbol{x}]\end{array}} \tag{12}$$

A known limitation of large deviations resides in the fact that it neglects all sub-exponential corrections. For instance, assume network agents 1 and 2 exhibit asymptotic error probabilities $P_1=e^{-\Phi/\mu}$ and $P_2=10\,e^{-\Phi/\mu}=e^{-(1/\mu)[\Phi+o(1)]}$, where $o(1)$ stands for any correction such that $o(1)\to 0$ as $\mu\to 0$. These two probabilities have the same exponent Φ, but we shall always have $P_2=10\,P_1$, because the factor 10 acts as a sub-exponential term.

A refined analysis can be pursued by seeking an asymptotic approximation, $\mathscr{P}_{k,\mu}(\gamma)$, that ensures the much stronger conclusion: $\mathbb{P}[\boldsymbol{y}^\star_{k,\mu}>\gamma]=\mathscr{P}_{k,\mu}(\gamma)[1+o(1)]$. This framework is commonly referred to as *exact asymptotics* [22], [23].

The exact asymptotics for detection over diffusion networks were obtained in [3], [4]. The result can be formally stated as follows. Assume that $\boldsymbol{x}_k(n)$ is not lattice. Let $\gamma\in\mathcal{D}^o_\Phi$, with $\gamma>\mathbb{E}[\boldsymbol{x}]$, and let θ_γ be the unique solution to the stationary equation $\phi'(\theta_\gamma)=\gamma$. Then, $\theta_\gamma>0$ and, for $k=1,2,\ldots,S$:

$$\boxed{\mathscr{P}_{k,\mu}(\gamma)\triangleq\sqrt{\frac{\mu}{2\pi\theta^2_\gamma\,\phi''(\theta_\gamma)}}\,e^{-\frac{1}{\mu}[\Phi(\gamma)+\epsilon_{k,\mu}(\theta_\gamma)]}} \tag{13}$$

with

$$\epsilon_{k,\mu}(t)=[\phi(t)-\mu\phi_{k,\mu}(t/\mu)]+\frac{[\phi'(t)-\phi'_{k,\mu}(t/\mu)]^2}{2\,\phi''(t)}, \tag{14}$$

where the ratio $\epsilon_{k,\mu}(\theta_\gamma)/\mu$ remains bounded as $\mu\to 0$.

We see that $\mathscr{P}_{k,\mu}(\gamma)=e^{-(1/\mu)[\Phi(\gamma)+o(1)]}$, where the term $o(1)$ collects all the sub-exponential corrections. In particular, the (agent-dependent) correction $\epsilon_{k,\mu}(\theta_\gamma)$ takes into account the network topology and the combination weights, and is *peculiar* to the adaptive network.

III. Importance Sampling and Cramér's transform

Consider a (continuous, for ease of description) random variable $\boldsymbol{y}$, with probability density function (pdf) $f(y)$. Consider also another pdf $\tilde f(y)$ that does not vanish (except for zero-measure sets) when $f(y)>0$, and introduce the (likelihood-ratio) weighting function $w(y)=f(y)/\tilde f(y)$. It then holds that:

$$\mathbb{P}[\boldsymbol{y}>\gamma]=\int_\gamma^\infty w(y)\tilde f(y)dy=\mathbb{E}_{\tilde f}\left[w(\boldsymbol{y})\mathcal{I}_{\{\boldsymbol{y}>\gamma\}}\right], \tag{15}$$

where $\mathcal{I}_{\mathcal{E}}$ is the indicator of an event $\mathcal{E}$, and $\mathbb{E}_{\tilde f}[\cdot]$ denotes expectation computed over the *transformed* pdf $\tilde f(y)$. Thus, the quantity to be estimated can be regarded as the expectation, under the transformed pdf, of the indicator of the event $\{\boldsymbol{y}>\gamma\}$, *weighted* by the function $w(y)$. The rationale behind importance sampling is that, by an appropriate choice of the weighting function, it is possible to map an event that is rare under the original sampling pdf, $f(y)$, into an event that is *not* rare under the new sampling pdf $\tilde f(y)$. In this way, the number of Monte Carlo iterations needed to estimate the expectation is reduced, because (important) samples are generated around the body (not the tail) of the new distribution. An accurate estimate of the probability tails is enabled by the weighting function $w(y)$.

When working with random variables obeying the LDP, there is a classical way to select the transformed pdf $\tilde f(y)$. This is usually referred to as *exponential twisting* of $f(y)$, and amounts to selecting $w(y)=e^{-\eta y+\ln\mathbb{E}[e^{\eta\boldsymbol{y}}]}$ — see [22]. The aforementioned change of measure was originally proposed by Cramér [24] to compute large-deviations exponents, and is accordingly also known as Cramér's transform. The choice of a given exponential twisting (i.e., the choice of the parameter η) is critical in determining the accuracy of the estimates produced by the importance sampling algorithm. Interestingly, theoretical studies suggest to use for importance sampling the *same* exponential twisting needed to compute the error exponents — see, e.g., [25]. For the classical and simplest case of summation of i.i.d. random variables, the latter kind of exponential twisting is well-known — see, e.g., [22]. For our specific problem of adaptive distributed detection, the solution is more involved, and relies on knowledge of the large-deviations characterization of the diffusion output. This solution is accurately examined in [4], Appendix B, and amounts to selecting $\eta = \theta_\gamma/\mu$, where θ_γ is the solution to $\phi'(\theta_\gamma) = \gamma$. Denoting by

$$\phi_{k,\mu}(t;n) \triangleq \ln \mathbb{E}[e^{t\boldsymbol{y}_k^\star(n)}] = \sum_{i=1}^{n}\sum_{\ell=1}^{S} \psi\left(\mu(1-\mu)^{i-1} b_{k,\ell}(i) t\right), \quad (16)$$

the LMGF of the random variable $\boldsymbol{y}_k^\star(n)$ in (3), such choice yields:

$$\boxed{w(y) = e^{-\frac{\theta_\gamma}{\mu}\, y + \phi_{k,\mu}\left(\frac{\theta_\gamma}{\mu};n\right)}} \quad (17)$$

Unfortunately, in order to generate samples according to $\tilde{f}(y)$, the weighting function $w(y)$ is not sufficient, since one needs to know also the pdf $f(y)$ of the random variable $\boldsymbol{y}_k^\star(n)$. As already observed, knowledge of this distribution is seldom available. To overcome this issue, one may consider running first a Monte Carlo simulation by generating random instances of the local statistics $\boldsymbol{x}_\ell(i)$, and then evaluating $\boldsymbol{y}_k^\star(n)$ through (3). This means that we must find the appropriate change of measure applied to $\boldsymbol{x}_\ell(i)$ that would induce the desired exponential change of measure on the random variable $\boldsymbol{y}_k^\star(n)$. To this aim, let us denote by $\pi(x)$ the pdf of $\boldsymbol{x}_\ell(i)$. We now show that the desired goal can be achieved by drawing the (i,ℓ)-th sample $\boldsymbol{x}_\ell(i)$ from the pdf:

$$\boxed{\tilde{\pi}_{i,\ell}(x) = e^{\eta_{i,\ell} x - \psi(\eta_{i,\ell})}\pi(x)} \quad (18)$$

with $\eta_{i,\ell} \triangleq (1-\mu)^{i-1} b_{k,\ell}(i)\,\theta_\gamma$. All these samples are still generated independently, but now they are no longer identically distributed. Note also that the above transformation depends upon the index k of the agent under consideration, even if the subscript has been suppressed for ease of notation. Let us introduce the joint ensemble $\boldsymbol{X} = \{\boldsymbol{x}_\ell(i)\}$, for $i = 1, 2, \ldots, n$ and for $\ell = 1, 2, \ldots, S$. The expectation in (15) can then be rewritten as:

$$\mathbb{P}[\boldsymbol{y}_k^\star(n) > \gamma] = \mathbb{E}[\mathcal{I}_{\{\boldsymbol{y}_k^\star(n)>\gamma\}}] = \mathbb{E}_{\tilde{\pi}}\left[\frac{\pi(\boldsymbol{X})}{\tilde{\pi}(\boldsymbol{X})}\mathcal{I}_{\{\boldsymbol{y}_k^\star(n)>\gamma\}}\right]. \quad (19)$$

But since the $\boldsymbol{x}_\ell(i)$ are i.i.d., we have from (18):

$$\frac{\pi(\boldsymbol{X})}{\tilde{\pi}(\boldsymbol{X})} = e^{-\sum_{i=1}^n \sum_{\ell=1}^S \eta_{i,\ell}\boldsymbol{x}_\ell(i) + \sum_{i=1}^n \sum_{\ell=1}^S \psi(\eta_{i,\ell})} = w(\boldsymbol{y}_k^\star(n)), \quad (20)$$

where we further applied the definition of $w(y)$ and Eqs. (3) and (16). This result shows that (19) corresponds to the expectation in (15) with the pdf $\tilde{f}(y)$ chosen as in (17).

IV. Numerical Examples

We consider the canonical shift-in-mean detection problem:

$$\mathcal{H}_0 : \boldsymbol{d}_k(n) \sim \mathscr{L}(d), \qquad \mathcal{H}_1 : \boldsymbol{d}_k(n) \sim \mathscr{L}(d-\rho), \quad (21)$$

with shift parameter $\rho > 0$, and with noise distributed according to the Laplace pdf (with unitary scale parameter) $\mathscr{L}(d) = 1/2\, e^{-|d|}$. In the above formulation, $\boldsymbol{d}_k(n)$ denotes the measurement collected by agent k at time n. The corresponding local statistic $\boldsymbol{x}_k(n)$ is chosen as the log-likelihood ratio, $\ln \frac{\mathscr{L}(\boldsymbol{d}_k(n)-\rho)}{\mathscr{L}(\boldsymbol{d}_k(n))}$, which yields:

$$\boldsymbol{x}_k(n) = \begin{cases} -\rho, & \boldsymbol{d}_k(n) < 0, \\ +\rho, & \boldsymbol{d}_k(n) > \rho, \\ 2\boldsymbol{d}_k(n) - \rho, & \boldsymbol{d}_k(n) \in [0,\rho]. \end{cases} \quad (22)$$

Let $\pi_0(x)$ and $\pi_1(x)$ denote the (generalized) pdfs of $\boldsymbol{x}_k(n)$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. It suffices to focus on $\pi_0(x)$, since a shift-in-mean with respect to a symmetric pdf yields, for the pdfs of the log-likelihood ratio, $\pi_1(x) = \pi_0(-x)$. From (22) we get:

$$\pi_0(x) = \frac{1}{2}\delta(x+\rho) + \frac{e^{-\rho}}{2}\delta(x-\rho) + \frac{1}{4}e^{-\frac{x+\rho}{2}}\Pi\left(\frac{x}{2\rho}\right), \quad (23)$$

where $\Pi(x)$ is a unit-width rectangular window centered at 0, and $\delta(x)$ is the Dirac-delta function. The corresponding LMGF of $\boldsymbol{x}_k(n)$ is computable in closed form:

$$\psi_0(t) = \ln\left(\frac{e^{-t\rho}}{2} + \frac{e^{(t-1)\rho}}{2} + \frac{e^{-\rho/2}\rho}{2}\,\text{sinch}[\rho(t-1/2)]\right),$$

where $\text{sinch}(x) \triangleq \sinh(x)/x$, with $\text{sinch}(0) = 1$. We now show how to implement the importance sampling method described in Sect. III. Applying an exponential twisting with parameter η to $\pi_0(x)$ in (23), straightforward algebra yields:

$$\tilde{\pi}_0(x) = q_-\delta(x+\rho) + q_+\delta(x-\rho) + [1 - q_- - q_+]\, g(x), \quad (24)$$

where

$$q_- \triangleq \frac{e^{-\eta\rho}}{2e^{\psi_0(\eta)}}, \qquad q_+ \triangleq \frac{e^{(\eta-1)\rho}}{2e^{\psi_0(\eta)}}, \quad (25)$$

and

$$g(x) \triangleq \frac{1}{2\rho}\,\frac{e^{x(\eta-1/2)}}{\text{sinch}[\rho(\eta-1/2)]}\Pi\left(\frac{x}{2\rho}\right). \quad (26)$$

We see from (24) that $\tilde{\pi}_0(x)$ is a mixture of three components, such that $\boldsymbol{x}_k(n)$ is equal to $-\rho$ (resp., ρ) with probability q_- (resp., q_+), and is sampled from the pdf $g(x)$ otherwise.

With regards to the combination policy, we consider the uniform averaging rule [13], [14]. Denoting by $\mathcal{N}_k$ the k-th agent's neighborhood (including k itself), such a rule prescribes setting $a_{k,\ell} = 1/|\mathcal{N}_k|$ for $\ell \in \mathcal{N}_k$, and zero otherwise. This choice provides a right-stochastic A, whose Perron eigenvector has entries given by $p_\ell = |\mathcal{N}_\ell|/\sum_{m=1}^S |\mathcal{N}_m|$ [13], [14]. With regards to the detection threshold, we follow the procedure described in [1], [2], obtaining $\gamma = 0$, which implies $\alpha_{k,\mu} = \beta_{k,\mu}$, since $\pi_1(x) = \pi_0(-x)$. Finally, in our simulations, we refer to a sufficiently large time horizon (steady-state performance) and evaluate the error probabilities for different values of the step-size.

In Fig. 2 we consider the network represented in the inset plot. The performance of the agents is displayed as a function of $1/\mu$, and different agents are marked with different colors. Specifically, we show the two theoretical approximations [exact asymptotics in (13) and normal approximation in (10)], along with the performance obtained through the importance sampling recipe described in Sect. III. The estimated error probability curves confirm that: $i)$ the error probability decays exponentially, approximately with the same decaying rate for all agents; and $ii)$ the curves are ordered so as to reflect the network connection structure. The normal approximation turns out to be accurate, especially in the leftmost region. We see that the variances in (10) seem to contain useful information about the different detection performance at different agents. However, as indicated by the theory, the predictions obtained with the normal approximation cannot be assumed true as the step-size decreases. The empirical probability converges towards the exact asymptotics as the step-size decreases. The dependency between the network structure and the detection performance at different agents is correctly embodied in the exact asymptotics, as witnessed by the correct ordering of the curves. In contrast to what happens for the normal approximation, the predictions offered by the exact asymptotics,

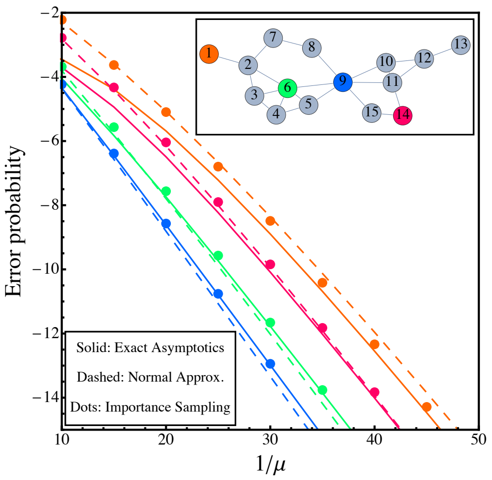

Fig. 2. Steady-state error probabilities $\alpha_{k,\mu} = \beta_{k,\mu}$ for the Laplace example with $\rho = 0.6$. The network topology is depicted in the inset, and the combination policy is the uniform averaging rule. The performance of agents 1, 6, 9 and 14 is displayed.

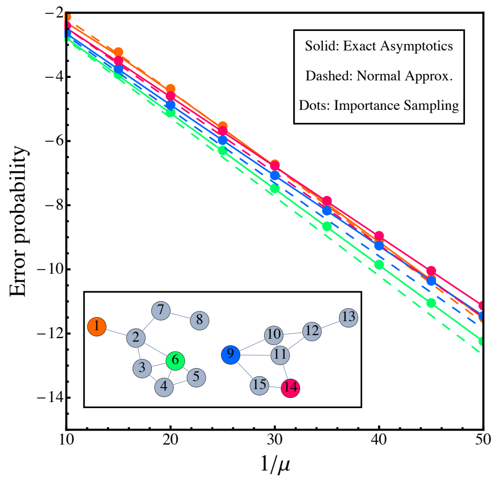

Fig. 3. Same setting of Fig. 2, with agent 9 disconnected from agents 5, 6 and 8.

for less connected agents, seem less accurate in the leftmost part of the plot. In a sense, the two theoretical approximations complement each other. Moreover, thanks to the simulation tool presented in this work, we have been able to perform a careful empirical examination of the detection performance, despite the very small values of the error probabilities. Remarkably, we see that using the importance sampling tool in conjunction with the available theoretical formulas allows an almost complete extrapolation of the system performance.

In Fig. 3 we repeat the analysis with reference to the two different networks shown in the inset plot, which are obtained from the network in the inset plot of Fig. 2 by disconnecting agent 9 from agents 5, 6 and 8. The general conclusions are similar to those reported from the analysis of Fig. 2. Joint inspection of Figs. 2 and 3 reveals further interesting properties. First, the benefits of cooperation emerge clearly, since the performance of the same individual agents when the information percolates through the entire *joint* network, improves with respect to the performance corresponding to the case of two disconnected subnets. It is particularly interesting to compare the performances of agents 1 and 14 in the latter scenario. We see that agent 1 is less connected than agent 14. This would explain the fact that, in the leftmost region of the plot, agent 1 has a worse performance. However, as μ decreases, the performance of agent 1 becomes better than that of agent 14. One wonders why. The reason is because it can be verified that the subnet of agent 1 has an error exponent (i.e., a slope of the probability curve) higher than that of the subnet of agent 14. According to our large-deviations analysis, this fact implies that the probability curve of agent 1 *must* cross that of agent 14.

In summary, we conclude that the presented tools (normal approximation + exact asymptotics + importance sampling) provide a substantial advance in the quantitative performance analysis of adaptive detection over diffusion networks.

References

[1] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Diffusion-based adaptive distributed detection: steady-state performance in the slow adaptation regime," Digital Object Identifier 10.1109/TIT.2016.2580665.

[2] P. Braca, S. Marano, V. Matta, and A. H. Sayed, "Large deviations analysis of adaptive distributed detection," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014, pp. 6112–6116.

[3] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Exact asymptotics of distributed detection over adaptive networks," in *Proc. IEEE ICASSP*, Brisbane, Australia, Apr. 2015, pp. 3377–3381.

[4] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Distributed detection over adaptive networks: Refined asymptotics and the role of connectivity," available as arXiv:1601.07011v1 [cs.MA], Jan. 2016.

[5] P. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, 1997.

[6] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors: Part I–Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[7] R. Blum, S. Kassam, and V. Poor, "Distributed detection with multiple sensors: Part II–Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

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

[9] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing statistical hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.

[10] ——, "Consensus-based Page's test in sensor networks," *Signal Processing*, vol. 91, no. 4, pp. 919–930, Apr. 2011.

[11] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4396, Sep. 2011.

[12] D. Bajovic, D. Jakovetic, J. Moura, J. Xavier, and B. Sinopoli, "Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5987–6002, Nov. 2012.

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

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

[15] J. Chen and A. H. Sayed, "On the learning behavior of adaptive networks — Part I: Transient analysis," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3487–3517, Jun. 2015.

[16] J. Chen and A. H. Sayed, "On the learning behavior of adaptive networks — Part II: Performance analysis," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3518–3548, Jun. 2015.

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

[18] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.

[19] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.

[20] H. Shao, *Mathematical Statistics*, 2nd ed., Springer, 2003.

[21] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1987.

[22] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.

[23] R. R. Bahadur, and R. R. Rao, "On deviations of the sample mean," *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 1015–1027, Dec. 1960.

[24] H. Cramér, "Sur un nouveau théorème-limite de la théorie des probabilités," in: *Actualités Scientifiques et Industrielles*, Hermann, Paris, no. 736, 1938.

[25] J. A. Bucklew, *Large Deviations Techniques in Decision, Simulation and Estimation*. Wiley, NY, 1990.