Quickest Change Detection of a Markov Process Across a Sensor Array

Vasanthan Raghavan, *Member, IEEE*, and Venugopal V. Veeravalli, *Fellow, IEEE*

***Abstract*—Recent attention in quickest change detection in the multisensor setting has been on the case where the densities of the observations change at the same instant at all the sensors due to the disruption. In this work, a more general scenario is considered where the change propagates across the sensors, and its propagation can be modeled as a Markov process. A centralized, Bayesian version of this problem is considered, with a fusion center that has perfect information about the observations and *a priori* knowledge of the statistics of the change process. The problem of minimizing the average detection delay subject to false alarm constraints is formulated in a dynamic programming framework. Insights into the structure of the optimal stopping rule are presented. In the limiting case of rare disruptions, it is shown that the structure of the optimal test reduces to thresholding the *a posteriori* probability of the hypothesis that no change has happened. Under a certain condition on the Kullback-Leibler (K-L) divergence between the post- and the pre-change densities, it is established that the threshold test is asymptotically optimal (in the vanishing false alarm probability regime). It is shown via numerical studies that this *low-complexity* threshold test results in a substantial improvement in performance over *naive* tests such as a single-sensor test or a test that incorrectly assumes that the change propagates instantaneously.**

***Index Terms*—Change-point problems, distributed decision-making, optimal fusion, quickest change detection, sensor networks, sequential detection.**

I. Introduction

An important application area for distributed decision-making systems is in environment surveillance and monitoring. Specific applications include: i) intrusion detection in computer networks and security systems [1], [2]; ii) monitoring cracks and damages to vital bridges and highway networks [3]; iii) monitoring catastrophic faults to critical infrastructures such as water and gas pipelines, electricity connections, supply chains, etc., [4]; iv) biological problems characterized by an event-driven potential including monitoring human subjects for epileptic fits, seizures, dramatic changes in physiological behavior, etc., [5], [6]; v) dynamic spectrum access and allocation problems [7]; vi) chemical or biological warfare agent detection systems to protect against terrorist attacks; vii) detection of the onset of an epidemic; and viii) failure detection in manufacturing systems and large machines. In all of these applications, the sensors monitoring the environment take observations that undergo a change in statistical properties in response to a disruption (change) in the environment. The goal is to detect the point of disruption (change-point) as quickly as possible, subject to false alarm constraints.

In the standard formulation of the change detection problem, studied over the last fifty years, there is a sequence of observations whose density changes at some unknown point in time and the goal is to detect the change-point as soon as possible. Two classical approaches to quickest change detection are: i) the *minimax* approach [8], [9], where the goal is to minimize the worst-case delay subject to a lower bound on the mean time between false alarms; and ii) the *Bayesian* approach [10]–[12], where the change-point is assumed to be a random variable with a density that is known *a priori* and the goal is to minimize the average (expected) detection delay subject to a bound on the probability of false alarm. Significant advances in both the minimax and the Bayesian theories of change detection have been made, and the reader is referred to [8]–[21] for a representative sample of the body of work in this area. The reader is also referred to [8], [15], [17], [21]–[26] for performance analyses of the standard change detection approaches in the minimax context, and [27], [28] in the Bayesian context.

Extensions of the above framework to the multisensor case where the information available for decision-making is *distributed* has also been explored [28]–[32]. In this setting, the observations are taken at a set of L distributed sensors, as shown in Fig. 1. The sensors may send either quantized/unquantized versions of their observations or local decisions to a *fusion center*, subject to communication delay, power, and bandwidth constraints, where a final decision is made, based on all the sensor messages. In particular, in much of this work [28]–[31], it is assumed that the statistical properties of *all* the sensors' observations change at the same time.

However, in many scenarios such as detecting pollutants and biological warfare agents, the change process is governed by the movement of the agent through the medium. Thus, it is more suitable to consider the case where the statistics of each sensor's observations may change at different points in time. This problem is studied in [32] where the authors consider a

Manuscript received December 19, 2008; revised July 25, 2009. Current version published March 17, 2010. This work was supported in part by the NSF by Grant CCF-0830169, and by the U.S. Army Research Office MURI Grant W911NF-06-1-0094, through a subcontract from Brown University at the University of Illinois. This paper was presented in part at the 11th International Conference on Information Fusion, Cologne, Germany, June–July 2008 and the IEEE International Symposium on Information Theory, Seoul, South Korea, June–July 2009.

V. Raghavan was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3052, Australia (e-mail: vasanthan_raghavan@ieee.org).

V. V. Veeravalli is with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: vvv@illinois.edu).

Communicated by F. Hlawatsch, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2010.2040869

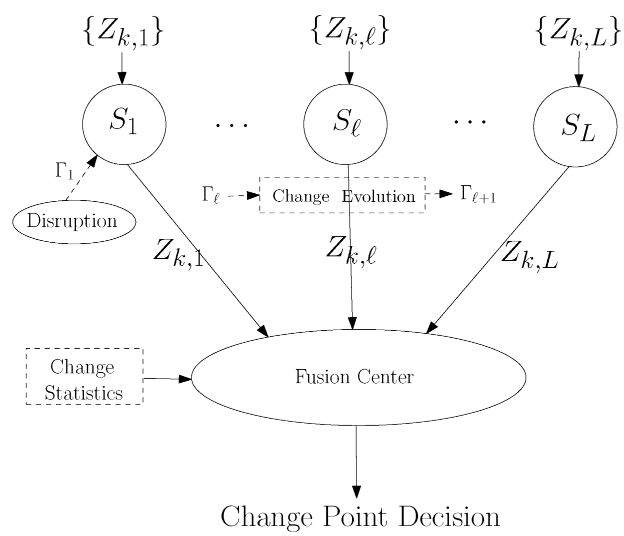

Fig. 1. Change-point detection across a linear array of sensors.

one-shot scheme with each sensor running a cumulative sum (CUSUM) algorithm. The sensors communicate with the fusion center only when they are ready to signal an alarm. It is established in [32] that a scheme where the fusion center employs a minimal strategy of declaring change upon receiving the first asynchronous signal from the sensors is asymptotically optimal in an extended minimax sense.

In this work, we consider a Bayesian version of this problem and assume that the point of disruption (that needs to be detected) is a random variable with a geometric distribution. More general disruption models can be considered, but the case of a geometric prior has an intuitive and appealing interpretation due to the *memorylessness* property of the geometric random variable. In addition, the practically relevant rare disruption regime can be obtained by letting the geometric parameter go to zero. We assume that the L sensors are placed in an array or a line and they observe the change as it propagates through them. The progression of change in only one strictly determined direction can be thought as a first approximation to more realistic situations. The inter-sensor delay is modeled with a Markov model and, in particular, the focus is on the case where the inter-sensor delay is also geometric. This model can be viewed as a first order approximation to more general propagation models, with the zeroth-order model being the case where the statistical properties of the sensors' observations change at the same time.

We study the centralized case, where the fusion center has complete information about the observations at all the L sensors, the change process statistics, and the pre- and the post-change densities. This is applicable in scenarios where: i) the fusion center is geographically collocated with the sensors so that ample bandwidth is available for reliable communication between the sensors and the fusion center; and ii) the impact of the disruption-causing agent on the statistical dynamics of the change process and the statistical nature of the change so induced can be modeled accurately. Note that under the centralized model, the special case where the change happens at the same time at all sensors corresponds to the standard (single sensor) quickest change detection problem [12] with an L-vector observation.

Our work differs from Hadjiliadis *et al.* in two ways. First, the nature of the one-shot scheme implies that the complexity of decision-making is at the sensor level in [32], whereas it is at the fusion center here. In terms of the natural trade-off between reliability of decision-making (due to a L-fold sensor diversity) and device-level complexity of the sensor network, we can view these two works as corresponding to the two extreme cases. Secondly, prior information about the change process (including the direction of change propagation) incorporated in the Bayesian setting here should allow us to improve performance over the minimax problem of [32].

Summary of Main Contributions: The goal of the fusion center is to come up with a strategy (or a stopping rule) to declare change, subject to false alarm constraints. Towards this goal, we first show that the problem studied here fits the standard dynamic programming (DP) framework [33] with the sufficient statistics given by the *a posteriori* probabilities of the state of the system conditioned on the observation process. We then establish a recursion for the sufficient statistics, which generalizes the recursion for the case when all the sensors observe the change at the same instant, which is equivalent to the single sensor case studied in [33, p. 133]. We further go on to establish the structure of the optimal stopping rule for change detection. This rule takes the form of the smallest time of cross-over (intersection) of a linear functional (or hyperplane) in the space of sufficient statistics with a nonlinear concave function. While further analytical characterization of the optimal stopping rule is difficult in general, in the extreme scenario of a rare disruption regime, we show that the structure of this rule reduces to a simple threshold test on the *a posteriori* probability that no change has happened. This low-complexity test is denoted as ν_A (corresponding to an appropriate choice of threshold A) for simplicity.

While ν_A is obtained as a limiting form of the optimal test, this does not necessarily imply that it is a "good" test. We therefore proceed to establish that ν_A is asymptotically optimal (as the false alarm probability PFA vanishes) under a certain condition on the Kullback-Leibler (K-L) divergence between the post- and the pre-change densities. Meeting this condition becomes easier as the change propagates more quickly across the sensor array, and in the extreme case where the sensor observations change at the same time, this condition reduces to the mild one that the K-L divergence be positive.

The difference between the noninstantaneous and instantaneous change propagation settings is more apparent in the nonasymptotic, but small PFA regime. Asymptotic optimality of a particular test in the instantaneous change propagation setting translates to an L-fold increase in the slope of average detection delay (ADD) versus $\log(\mathsf{PFA})$ in the regime where the false alarm probability is small, but not vanishing (e.g., $\mathsf{PFA} \approx 10^{-4}$ or 10^{-5}). However, if the change propagates slowly across the sensor array, numerical studies indicate that not all of the L sensors' observations *may* contribute to the performance of ν_A in this regime. Nevertheless, as $\mathsf{PFA} \to 0$, all the L sensors are expected (in general) to contribute to the slope.

Thus, while it is not clear if ν_A is asymptotically optimal in general, or even if all the sensors' observations contribute to its performance in the nonasymptotic regime, numerical studies also show that ν_A can result in substantial performance improvement over naive tests such as: (i) the *single sensor test*, where only the first sensor's observation is used in decision-making; or (ii) the *mismatched test*, where all the sensors' observations are used in decision-making, albeit with a wrong model that change propagates instantaneously. This improvement in performance is especially apparent in regimes of practical importance, where the disruption is rare, and the propagation is reasonably quick, but noninstantaneous across the sensors. The performance improvement possible with ν_A, in addition to its low-complexity, make it an attractive choice for many practical applications with a basis in multisensor change process detection.

Organization: This paper is organized as follows. The change process detection problem is formally set up in Section II. In Section III, this problem is posed in a dynamic programming framework and the sufficient statistics of the dynamic program (DP) are identified. The structure of the optimal stopping rule in the general case and the rare disruption regime are illustrated in Section IV. The limiting form of the optimal test is denoted as ν_A for simplicity. The main statements regarding the asymptotic optimality of ν_A are stated in Section V. These results are established in Section VI and the Appendices. A discussion of the main results and numerical studies to illustrate our results are provided in Section VII. Concluding remarks are made in Section VIII.

II. Problem Formulation

Consider a distributed system with an array of L sensors, as in Fig. 1, that observes an L-dimensional discrete-time stochastic process $\boldsymbol{Z}_k = [Z_{k,1}, \ldots, Z_{k,L}]$, where $Z_{k,\ell}$ is the observation at the ℓth sensor at the kth time instant. A disruption in the sensing environment occurs at the random time instant Γ_1, and hence, the density[1] of the observations at each sensor undergoes a change from the null density f_0 to the alternate density f_1.

Change Process Model: We consider a *change process* where the change-point evolves across the sensor array. In particular, the change-point as seen by the ℓth sensor is denoted as Γ_ℓ. We assume that the evolution of the change process is Markovian across the sensors. That is,

$$\begin{aligned} P(\{\Gamma_{\ell_1+\ell_2+\ell_3} &= m_1+m_2+m_3\} \\ &\times |\{\Gamma_{\ell_1+\ell_2} = m_1+m_2\}, \{\Gamma_{\ell_1} = m_1\}) \\ &= P(\{\Gamma_{\ell_1+\ell_2+\ell_3} = m_1+m_2+m_3\} \\ &\times |\{\Gamma_{\ell_1+\ell_2} = m_1+m_2\}) \end{aligned}$$

for all ℓ_i and $m_i \geq 0, i = 1,2,3$. Further simplification of the analysis is possible under a *joint-geometric* model on $\{\Gamma_\ell\}$. Under this model, the change-point (Γ_1) evolves as a geometric random variable with parameter ρ, and inter-sensor change propagation is modeled as a geometric random variable with parameter $\{\rho_{\ell-1,\ell}, \ell = 2, \ldots, L\}$. That is

$$P(\{\Gamma_1 = m\}) = \rho(1-\rho)^m, \qquad m \geq 0$$

and

$$\begin{aligned} P(\{\Gamma_\ell &= m_1+m_2\}|\{\Gamma_{\ell-1} = m_2\}) \\ &= \rho_{\ell-1,\ell}(1-\rho_{\ell-1,\ell})^{m_1}, \qquad m_1 \geq 0 \end{aligned}$$

independent of $m_2 \geq 0$ for all ℓ such that $2 \leq \ell \leq L$. We will find it convenient[2] to set $\rho_{0,1} = \rho$ and $\rho_{L,L+1} = 0$ so that $\rho_{\ell-1,\ell}$ is defined for all $\ell = 1, \ldots, L+1$.

While a joint-geometric model is consistent with the Markovian assumption as only the inter-sensory (one-step) propagation parameters are modeled, the change-points at the individual sensors themselves are *not* geometric. The joint-geometric model can be viewed as a first order approximation of more realistic propagation scenarios. In particular, note that $\rho \to 1$ corresponds to the case where instantaneous disruption (that is, the event $\{\Gamma_1 = 0\}$) has a high probability of occurrence. On the other hand, $\rho \to 0$ uniformizes the change-point in the sense that the disruption is equally likely to happen at any point in time. This case where the disruption is "rare" is of significant interest in practical systems [15], [18], [28]–[31]. This is also the case where we will be able to make insightful statements about the structure of the optimal stopping rule. Similarly, we can also distinguish between two extreme scenarios at sensor ℓ depending on whether $\rho_{\ell-1,\ell} \to 0$ or $\rho_{\ell-1,\ell} \to 1$. The case where $\rho_{\ell-1,\ell} \to 1$ corresponds to instantaneous change propagation at sensor ℓ and $\{\Gamma_\ell = \Gamma_{\ell-1}\}$ with high probability. The case where $\rho_{\ell-1,\ell} \to 0$ corresponds to uniformly likely propagation delay. The widely-used assumption of instantaneous change propagation across sensors is equivalent to assuming $\rho_{\ell-1,\ell} = 1$ for all $\ell = 2, \ldots, L$.

Observation Model: To simplify the study, we assume that the observations (at every sensor) are independent, conditioned[3] on the change hypothesis corresponding to that sensor, and are identically distributed pre- and post-change, respectively. That is,

$$Z_{k,\ell} \sim \begin{cases} \text{i.i.d.} & f_0 \quad \text{if } k < \Gamma_\ell, \\ \text{i.i.d.} & f_1 \quad \text{if } k \geq \Gamma_\ell. \end{cases}$$

We will describe the above assumption as that corresponding to an "i.i.d. observation process." Let $D(f_1, f_0)$ denote the Kullback-Leibler divergence between f_1 and f_0. That is,

$$D(f_1, f_0) = \int \log\left(\frac{f_1(x)}{f_0(x)}\right) f_1(x)dx.$$

We also assume that the measure described by f_0 is *absolutely continuous* with respect to that described by f_1. That is, if $f_1(x) = 0$ for some x, then $f_0(x) = 0$. This condition ensures that $E_{\bullet|f_1}[\frac{f_0(\bullet)}{f_1(\bullet)}] = 1$.

Performance Metrics: We consider a *centralized, Bayesian* setup where a fusion center has complete knowledge of the observations from all the sensors, $I_k \triangleq \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_k\}$, in addition to knowledge of statistics of the change process (equivalently,

[1]We assume that the pre-change (f_0) and the post-change (f_1) densities exist.

[2]This is also consistent with an equivalent $(L+2)$-sensor system where sensor indices run through $\{\ell = 0, \ldots, L+1\}$.

[3]More general observation (correlation) models are important in practical settings. This will be the subject of future work.

$\{\rho_{\ell-1,\ell}\}$) and statistics[4] of the observation process (equivalently, f_0 and f_1). The fusion center decides whether a change has happened or not based on the information, I_k, available to it at time instant k (equivalently, it provides a stopping rule or stopping time τ).

The two conflicting performance measures for quickest change detection are the probability of false alarm, $\mathsf{PFA} \triangleq P(\{\tau < \Gamma_1\})$, and the average detection delay, $\mathsf{ADD} \triangleq E[(\tau - \Gamma_1)^+]$, where $x^+ = \max(x, 0)$. This conflict is captured by the Bayes risk, defined as

$$\begin{aligned} R(c) &\triangleq \mathsf{PFA} + c\mathsf{ADD} \\ &= E[\mathbb{1}(\{\tau < \Gamma_1\}) + c(\tau - \Gamma_1)^+] \end{aligned}$$

for an appropriate choice of per-unit delay cost c, where $\mathbb{1}(\{\cdot\})$ is the indicator function of the event $\{\cdot\}$. We will be particularly interested in the regime where $c \to 0$. That is, a regime where minimizing PFA is more important than minimizing ADD, or equivalently, the asymptotics where $\mathsf{PFA} \to 0$.

The goal of the fusion center is to determine

$$\tau_{\mathsf{opt}} = \arg \inf_{\tau \in \boldsymbol{\Delta}_\alpha} \mathsf{ADD}(\tau)$$

from the class of change-point detection procedures $\boldsymbol{\Delta}_\alpha = \{\tau : \mathsf{PFA}(\tau) \leq \alpha\}$ for which the probability of false alarm does not exceed α. In other words, the fusion center needs to come up with a strategy (a stopping rule τ) to minimize the Bayes risk.

III. Dynamic Programming Framework

It is straightforward to check that [12, pp. 151-152] the Bayes risk can be written as

$$R(c) = P(\{\Gamma_1 > \tau\}) + cE[\sum_{k=0}^{\tau-1} P(\{\Gamma_1 \leq k\})].$$

Towards solving for the optimal stopping time, we restrict attention to a finite-horizon, say the interval $[0, T]$, and proceed via a dynamic programming (DP) argument.

The state of the system at time k is the vector $\boldsymbol{S}_k = [S_{k,1}, \ldots, S_{k,L}]$ with $S_{k,\ell}$ denoting the state at sensor ℓ. The state $S_{k,\ell}$ can take the value 1 (post-change), 0 (pre-change), or t (terminal). The system goes to the terminal state t, once a change-point decision τ has been declared. The state evolves as follows:

$$S_{k,\ell} = f\left(S_{k-1,\ell}, \Gamma_\ell, 1_{\{\tau \leq k\}}\right)$$

where the transition function is given as

$$f(s, \gamma, a) = \begin{cases} 0 & \text{if } \gamma > k, \quad s \neq t, \quad a = 0 \\ 1 & \text{if } \gamma \leq k, \quad s \neq t, \quad a = 0 \\ t & \text{if } s = t \quad \text{or} \quad a = 1 \end{cases}$$

with $\boldsymbol{S}_0 = \mathbf{0}$. Since $\boldsymbol{S}_{k-1}$ captures the information contained in $\{\Gamma_\ell \leq j\}$ for $0 \leq j \leq k-1$ and all ℓ, given $\boldsymbol{S}_{k-1}$, $\{\Gamma_\ell \leq k\}$ is independent of $\{\Gamma_\ell \leq j, j \leq k-1\}$ for all ℓ. Thus, the state evolution satisfies the Markov condition needed for dynamic programming.

The state is not observable directly, but only through the observations. The observation equation can be written as

$$Z_{k,\ell} = V_{k,\ell}^{(S_{k,\ell})} \mathbb{1}(\{S_{k,\ell} \neq t\}) + \xi \mathbb{1}(\{S_{k,\ell} = t\}), \quad \ell \geq 1$$

where $V_{k,\ell}^{(0)}$ and $V_{k,\ell}^{(1)}$ are the kth samples from independently generated infinite arrays of i.i.d. data according to f_0 and f_1, respectively. When the system is in the terminal state, the observations do not matter (since a change decision has already been made) and are hence denoted by a dummy random variable, ξ. It is clear that the observation uncertainty $(V_{k,\ell}^{(0)}, V_{k,\ell}^{(1)})$ satisfies the necessary Markov conditions for dynamic programming since they are i.i.d. in time.

Finally, the expected cost (Bayes risk) can be expressed as the expectation of an additive cost over time by defining

$$g_k(\boldsymbol{S}_k) = c\mathbb{1}(\{S_{k,1} = 1\})$$

and a terminal cost $\mathbb{1}(\{S_{k,1} = 0\})$. Thus the problem fits the standard dynamic programming framework with termination [33], with the sufficient statistic (belief state) being given by

$$P(\{\boldsymbol{S}_k = \boldsymbol{s}_k\} | I_k)$$

where $I_k = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_k\}$ for k such that $\boldsymbol{S}_k \neq t$, i.e., $S_{k,\ell} \in \{0, 1\}$ for each ℓ. Note that this sufficient statistic is described by 2^L conditional probabilities, corresponding to the 2^L values that $\boldsymbol{s}_k$ can take. We will next see that this sufficient statistic can be further reduced[5] to only L independent probability parameters in the general case.

The fusion center determines τ, and, hence, the minimum expected cost-to-go at time k for the above DP problem can be seen to be a function of I_k. For a finite horizon T, the cost-to-go function is denoted as $\widetilde{J}_k^T(I_k)$ and is of the form (see [33, p. 133], [29], for examples of similar nature)

$$\begin{aligned} \widetilde{J}_T^T(I_T) &= P(\{\Gamma_1 > T\} | I_T) \\ \widetilde{J}_k^T(I_k) &= \min\Big\{ P(\{\Gamma_1 > k\} | I_k), cP(\{\Gamma_1 \leq k\} | I_k) \\ &\qquad + E\left[\widetilde{J}_{k+1}^T(I_{k+1}) | I_k \right] \Big\}, \qquad 0 \leq k < T \end{aligned}$$

where I_0 is the empty set. The first term in the above minimization corresponds to the cost associated with stopping at time k, while the second term corresponds to the cost associated with proceeding to time $k+1$ without stopping. The minimum expected cost for the finite-horizon optimization problem is $\widetilde{J}_0^T(I_0)$.

Recursion for the Sufficient Statistics: We define an $(L+1)$-tuple of conditional probabilities, $\{p_{k,\ell}, \ell = 1, \ldots, L+1\}$

$$p_{k,\ell} \triangleq P(\{\Gamma_{\ell-1} \leq k, \Gamma_\ell > k\} | I_k).$$

We now show that $\boldsymbol{p}_k \triangleq [p_{k,1}, \ldots, p_{k,L+1}]$ can be obtained from $\boldsymbol{p}_{k-1}$ via a recursive approach. For this, we note that the

[4]We assume that the fusion center has knowledge of f_0 and f_1 so that it can use this information to declare that a change has happened. Relaxing this assumption is important in the context of practical applications and is the subject of current work.

[5]This should not be entirely surprising as our assumption of a line (or array) geometry imposes a "natural" ordering on the sensors' change-points. They can be arranged in nondecreasing order: $\Gamma_\ell \geq \Gamma_{\ell-1}$ for all ℓ.

underlying probability space Ω in the setup can be partitioned as

$$\Omega = \bigcup_{\ell=1}^{L+1} T_{k,\ell}$$

where

$$T_{k,\ell} \triangleq \{\Gamma_{\ell-1} \leq k, \Gamma_\ell > k\}.$$

The event where no sensor has observed the change is denoted as $T_{k,1}$. On the other hand, $T_{k,\ell}$ (for $\ell \geq 2$) corresponds to the event where the maximal index of the sensor that has observed the change before time instant k is $\ell - 1$.

Observe that $p_{k,\ell}$ is the probability of $T_{k,\ell}$ conditioned on I_k. To show that $p_{k,\ell}$ can be written in terms of $\boldsymbol{p}_{k-1}$, the observations $\boldsymbol{Z}_k$ and the prior probabilities, we partition $T_{k,\ell}$ further as

$$T_{k,\ell} = \bigcup_{j=1}^{\ell} U_{k,\ell,j}$$

$$U_{k,\ell,j} \triangleq \{\Gamma_{j-1} \leq k-1, \Gamma_j = k, \ldots, \Gamma_{\ell-1} = k, \Gamma_\ell \geq k+1\}, \quad 1 \leq j \leq \ell.$$

Note that $U_{k,\ell,j} \cap T_{k-1,j} = U_{k,\ell,j}$. Using the new partition $\{U_{k,\ell,j}, j = 1, \ldots, \ell\}$ and applying Bayes' rule repeatedly, it can be checked that $p_{k,\ell}$ can be written as

$$p_{k,\ell} = \frac{\sum_{m=1}^{\ell} f(\boldsymbol{Z}_k | I_{k-1}, U_{k,\ell,m}) P(U_{k,\ell,m} | I_{k-1})}{\sum_{j=1}^{L+1} \sum_{m=1}^{j} f(\boldsymbol{Z}_k | I_{k-1}, U_{k,j,m}) P(U_{k,j,m} | I_{k-1})} \triangleq \frac{\mathcal{N}_\ell}{\sum_{j=1}^{L+1} \mathcal{N}_j} \tag{1}$$

where $f(\cdot|\cdot)$ denotes the conditional probability density function of $\boldsymbol{Z}_k$ and $\mathcal{N}_\ell$ denotes the numerator term.

From the i.i.d. assumption on the statistics of the observations, the first term within the summation for $\mathcal{N}_\ell$ can be written as

$$f(\boldsymbol{Z}_k | I_{k-1}, U_{k,\ell,m}) = \prod_{j=1}^{\ell-1} f_1(Z_{k,j}) \prod_{j=\ell}^{L} f_0(Z_{k,j}) = \prod_{j=1}^{\ell-1} L_{k,j} \prod_{j=1}^{L} f_0(Z_{k,j})$$

where $L_{k,j} \triangleq \frac{f_1(Z_{k,j})}{f_0(Z_{k,j})}$ is the likelihood ratio of the two hypotheses given that $Z_{k,j}$ is observed at the jth sensor at the kth instant. For the second term, observe from the definitions that

$$P(U_{k,\ell,m} | I_{k-1}) = P(T_{k-1,m} | I_{k-1}) \frac{P(U_{k,\ell,m})}{P(T_{k-1,m})}.$$

Thus, we have

$$\mathcal{N}_\ell = \left(\sum_{m=1}^{\ell} \frac{P(U_{k,\ell,m})}{P(T_{k-1,m})} p_{k-1,m} \right) \prod_{m=1}^{\ell-1} L_{k,m} \prod_{m=1}^{L} f_0(Z_{k,m}) \triangleq \left(\sum_{m=1}^{\ell} w_{k,\ell,m} p_{k-1,m} \right) \boldsymbol{\Phi}_{\mathsf{obs}}(k,\ell) \tag{2}$$

where the first part is a weighted sum of $p_{k-1,m}$ with weights decided by the prior probabilities, and the second part of the evolution equation, $\boldsymbol{\Phi}_{\mathsf{obs}}(k,\ell)$, can be viewed as that part that depends only on the observation $\boldsymbol{Z}_k$.

Many observations are in order at this stage:

- The above expansion for $\mathcal{N}_\ell$ can be explained intuitively: If the maximal sensor index observing the change by time k is $\ell - 1$, then the maximal sensor index observing the change by time $k-1$ should be from the set $\{0, \ldots, \ell-1\}$.
- Using the joint-geometric model for $\{\Gamma_\ell\}$, it can be shown that $w_{k,\ell,m}$ is of the form

$$w_{k,\ell,m} = \frac{P(U_{k,\ell,m})}{P(T_{k-1,m})} = (1 - \rho_{\ell-1,\ell}) \cdot \prod_{j=m-1}^{\ell-2} \rho_{j,j+1} \triangleq (1 - \rho_{\ell-1,\ell}) \cdot w_m^\ell$$

$$\mathcal{N}_\ell = \prod_{m=1}^{\ell-1} L_{k,m} \prod_{m=1}^{L} f_0(Z_{k,m}) \cdot (1 - \rho_{\ell-1,\ell}) \times \left(\sum_{m=1}^{\ell} p_{k-1,m} \cdot w_m^\ell \right) \tag{3}$$

  with the understanding that the product term in the definition of w_m^ℓ is vacuous (and is to be replaced by 1) if $m = \ell$. It is important to note that the joint-geometric assumption renders the weights $(w_{k,\ell,m})$ associated with $p_{k-1,m}$ independent of k. This will be useful later in establishing convergence properties for the DP.
- It is important to note that given a fixed value of ℓ, $p_{k,\ell}$ is dependent on the entire vector $\boldsymbol{p}_{k-1}$ and not on $p_{k-1,\ell}$ alone. Thus, the recursion for $\mathcal{N}_\ell$ implies that $\boldsymbol{p}_k$ forms the sufficient statistic and the function $\widetilde{J}_k^T(I_k)$ can be written as a function of only $\boldsymbol{p}_k$, say $J_k^T(\boldsymbol{p}_k)$. The finite-horizon DP equations can then be rewritten as

$$J_T^T(\boldsymbol{p}_T) = p_{T,1}$$
$$J_k^T(\boldsymbol{p}_k) = \min\{p_{k,1}, \quad c(1 - p_{k,1}) + A_k^T(\boldsymbol{p}_k)\}$$

  with

$$A_k^T(\boldsymbol{p}_k) \triangleq E\left[J_{k+1}^T(\boldsymbol{p}_{k+1}) \,|\, I_k\right] = \int \left[J_{k+1}^T(\boldsymbol{p}_{k+1}) f(\boldsymbol{Z}_{k+1} \,|\, I_k)\right]\big|_{\boldsymbol{Z}_{k+1}=\boldsymbol{z}} \, d\boldsymbol{z}.$$

  Note that the previously established recursion for $\boldsymbol{p}_{k+1}$ implies that $\boldsymbol{p}_{k+1} = g(\boldsymbol{p}_k, \boldsymbol{Z}_{k+1})$ for an appropriate choice of $g(\cdot,\cdot)$ (the precise form of $g(\cdot,\cdot)$ is clear from (1) and (2)) which ensures that the right-hand side is indeed a function of $\boldsymbol{p}_k$.
- It is easy to check that the general framework reduces to the special case when all the change-points coincide with Γ_1. In this case, as in [29], we define $p_k \triangleq P(\{\Gamma_1 \leq k\} | I_k)$. Note that only $T_{k,1}$ and $T_{k,L+1}$ are nonempty sets with

$$T_{k,1} = \{\Gamma_1 \geq k+1\}, \quad T_{k,L+1} = \{\Gamma_1 \leq k\}$$
$$p_{k,L+1} = p_k, \quad p_{k,1} = 1 - p_k, \quad p_{k,\ell} = 0, \quad \ell = 2, \ldots, L.$$

Furthermore, the recursion for p_k reduces to

$$p_k = \frac{\mathcal{N}}{\prod_{j=1}^{L} f_0(Z_{k,j})(1-p_{k-1})(1-\rho)+\mathcal{N}}$$
$$\mathcal{N} = \prod_{j=1}^{L} f_1(Z_{k,j})((1-p_{k-1})\rho + p_{k-1})$$

which coincides with [29, eq. (13)-(15)]. This case can also be obtained from (3) by setting $\rho_{\ell-1,\ell}=1$ for all ℓ with $2 \leq \ell \leq L$.

IV. Structure of the Optimal Stopping Rule ($\tau_{\sf opt}$)

The goal of this section is to study the structure of the optimal stopping rule, $\tau_{\sf opt}$. For this, we follow the same outline as in [33] and study the infinite-horizon version of the DP problem by letting $T \to \infty$.

Theorem 1: Let $\boldsymbol{p} = [p_1, \ldots, p_{L+1}]$ be an element of the standard L-dimensional simplex $\mathcal{P}$, defined as, $\mathcal{P} \triangleq \{\boldsymbol{p} : \sum_{j=1}^{L+1} p_j = 1\}$. The infinite-horizon cost-to-go for the DP is of the form

$$J(\boldsymbol{p}) = \min\{p_1, c(1-p_1) + A_J(\boldsymbol{p})\}$$

where the function $A_J(\boldsymbol{p})$: i) is concave in $\boldsymbol{p}$ over $\mathcal{P}$; ii) is bounded as $0 \leq A_J(\boldsymbol{p}) \leq 1$; and iii) satisfies $A_J(\boldsymbol{p}) = 0$ over the hyperplane $\{\boldsymbol{p} : p_1 = 0\}$.

Proof: See Appendix A. ■

At this stage, it is a straightforward consequence that the optimal stopping rule is of the form

$$\tau_{\sf opt} = \inf_k \{p_{k,1}(1+c) - c < A_J(\boldsymbol{p}_k)\}.$$

That is, a change is declared when the hyperplane on the left side is exceeded by $A_J(\boldsymbol{p}_k)$ and no change is declared, otherwise. We will next see that this test characterization reduces to a degenerate one as $\rho \to 0$.

To establish this degeneracy, we define the following one-to-one and invertible transformation:

$$q_{k,\ell} = \frac{p_{k,\ell}}{\rho p_{k,1}} \Longleftrightarrow p_{k,\ell} = \frac{q_{k,\ell}}{\sum_{j=1}^{L+1} q_{k,j}}, \quad \ell = 1, \ldots, L+1$$

which is equivalent to

$$p_{k,1} = \frac{1}{1 + \rho \sum_{j=2}^{L+1} q_{k,j}}$$

and

$$p_{k,\ell} = \frac{\rho \quad q_{k,\ell}}{1 + \rho \sum_{j=2}^{L+1} q_{k,j}}, \qquad \ell = 2, \ldots, L+1.$$

We can write $q_{0,\ell}$ in terms of the priors as

$$q_{0,1} = \frac{p_{0,1}}{\rho p_{0,1}} = \frac{1}{\rho}$$
$$q_{0,\ell} = \frac{p_{0,\ell}}{\rho p_{0,1}}$$
$$= \frac{P(\{\Gamma_1 = \cdots = \Gamma_{\ell-1} = 0, \Gamma_\ell > 0\})}{\rho P(\{\Gamma_1 > 0\})}$$
$$= \frac{\prod_{j=0}^{\ell-2} \rho_{j,j+1}(1-\rho_{\ell-1,\ell})}{\rho(1-\rho)}, \qquad \ell = 2, \ldots, L+1.$$

Note that while $p_{k,\ell}$ are conditional probabilities of certain events, and hence, lie in the interval $[0,1]$, the range of $q_{k,\ell}$ is in general $[0,\infty)$.

It can be checked that the evolution equation can be rewritten in terms of $q_{k,\ell}$ as

$$q_{k,\ell} = \frac{1-\rho_{\ell-1,\ell}}{1-\rho} \cdot \prod_{j=1}^{\ell-1} L_{k,j} \cdot \left(\sum_{j=1}^{\ell} q_{k-1,j} w_j^{\ell} \right). \quad (4)$$

It is interesting to note from (4) that the update for $q_{k,\ell}$ is a weighted sum of $q_{k-1,j}, j = 1, \ldots, \ell$ with progressively decreasing weight as j increases. Similarly, we can define $J_k^T(\cdot)$ and $A_k^T(\cdot)$ in terms of $\boldsymbol{q}_k$. Using the transformation $\{q_{k,\ell}\}$, $\tau_{\sf opt}$ is seen to have the form

$$\boxed{\tau_{\sf opt} = \inf_k \left\{ \sum_{\ell=2}^{L+1} q_{k,\ell} > \frac{1-A_J(\boldsymbol{q}_k)}{\rho(c+A_J(\boldsymbol{q}_k))} \right\}.}$$

When all Γ_ℓ coincide, we have

$$q_{k,L+1} = \frac{p_k}{\rho(1-p_k)} \triangleq q_k$$
$$q_{k,1} = \frac{1}{\rho}, \quad q_{k,\ell} = 0, \qquad \ell = 2, \ldots, L.$$

Further, it is straightforward to check that the evolution in (4) reduces to

$$q_{k,L+1} = \frac{\prod_{j=1}^{L} L_{k,j}}{1-\rho} \cdot (1 + q_{k-1,L+1}). \quad (5)$$

Thus, the space of sufficient statistics and the optimal test reduce to a one-dimensional variable ($p_k = P(\{\Gamma_1 \leq k\}|I_k)$ or equivalently, q_k) and a threshold test on p_k (or equivalently, on q_k), respectively. In the general case, unless something more is known about the structure of $A_J(\cdot)$ (which is possible if there is some structure on $\{\rho_{\ell-1,\ell}\}$), we cannot say more about $\tau_{\sf opt}$. Nevertheless, the following theorem establishes its structure in the practical setting of a rare disruption regime ($\rho \to 0$). The limiting test thresholds (from below) the *a posteriori* probability that no-change has happened, and is denoted as ν_A.

Theorem 2: The test structure corresponding to $\tau_{\sf opt}$ converges in probability to a simple threshold operation in the asymptotic limit as $\rho \to 0$. This limiting test is of the form

$$\boxed{\nu_A = \begin{cases} \text{Stop} & \text{if } \log\left(\sum_{\ell=2}^{L+1} q_{k,\ell}\right) \geq A \\ \text{Continue} & \text{if } \log\left(\sum_{\ell=2}^{L+1} q_{k,\ell}\right) < A \end{cases}}$$

for an appropriate choice of threshold A.

Proof: See Appendix B. ■

The test ν_A is of low-complexity because of the following properties: i) a simple recursion formula (4) for the sufficient statistics; ii) a threshold operation for stopping; and iii) the threshold value that can be precomputed given the PFA constraint (see Prop. 3).

The fact that $\tau_{\sf opt} \overset{\rho \downarrow 0}{\to} \nu_A$ for an appropriate choice of A *does not* imply that ν_A is asymptotically (as PFA $\to 0$ or as $\rho \to 0$)

optimal. However, the low-complexity of this test, in addition to Theorem 2, and the fact that the structure of $A_J(\boldsymbol{q}_k)$ (and, hence, τ_{opt}) are not known suggest that it is a good candidate test for change detection across a sensor array. In fact, we will see this to be the case when we establish sufficient conditions under which ν_A is asymptotically optimal.

V. Main Results on ν_A

Towards this end, our main interest is in understanding the performance (ADD versus PFA) of ν_A for any general choice of threshold A.

Special Cases of Change Parameters: To build intuition, we start by considering some special scenarios of change propagation modeling. The first scenario corresponds to the case where one (or more) of the $\rho_{\ell-1,\ell}$ is 1. The following proposition addresses this setting.

Proposition 1: Consider an L-sensor system described in Section II, parameterized by $\{\rho_{\ell-1,\ell}\}$, where $\rho_{\ell',\ell'+1} = 1$ for some ℓ' and $\max_{j\neq \ell'} \rho_{j,j+1} < 1$. This system is equivalent to an $(L-1)$-sensor system, parameterized by $\{\delta_{\ell,\ell+1}\}$, where

$$\begin{aligned} \delta_{j,j+1} &= \rho_{j,j+1}, \qquad j \leq \ell' - 1 \\ \delta_{j,j+1} &= \rho_{j+1,j+2}, \qquad j \geq \ell' \end{aligned}$$

with the $(\ell'+1)$th sensor observing (a combination of) $Z_{k,\ell'+1}$ and $Z_{k,\ell'+2}$ with a geometric delay parameter of $\delta_{\ell',\ell'+1} = \rho_{\ell'+1,\ell'+2}$.

Proof: The proof is straightforward by studying the evolution of $\{q_{k,\ell}\}$ for the original L-sensor system. From (4), it can be seen that $q_{k,\ell'+1} = 0$ (identically) for all k and the reduced $(L-1)$-dimensional system discards this redundant information, while the observation corresponding to the $(\ell'+1)$th sensor is carried over to the $(\ell'+2)$th original sensor. ■

The second scenario corresponds to the case where one (or more) of the $\rho_{\ell-1,\ell}$ is 0.

Proposition 2: Consider an L-sensor system, parameterized by $\{\rho_{\ell-1,\ell}\}$, with ℓ' indicating the smallest index such that $\rho_{\ell',\ell'+1} = 0$. This system is equivalent to an ℓ'-sensor system with the same parameters as that of the original system. It is as if sensors $(\ell'+1)$ and beyond do not exist (or contribute) in the context of change detection.

Proof: The proof is again straightforward by considering the evolution of $\{q_{k,\ell}\}$ in (4) and noting that $q_{k,j}, j \geq \ell'+2$ are identically 0 for all k. ■

It is useful to interpret Propositions. 1 and 2 via an "information flow" paradigm. If change propagation is instantaneous across a sensor (corresponding to the first case), it is as if the fusion center is *oblivious* to the presence of that sensor conditioned upon the previous sensors' observations. In this setting, the detection delay corresponding to that sensor is zero, as would be expected from the fact that the geometric parameter is 1. In the second case, information flow to the fusion center (concerning change) is *cut-off or blocked* past the first sensor with a geometric parameter of 0. That is, the observations made by sensors $\{\ell'+1,\ldots,L\}$ (if any) do not contribute information to the fusion center in helping it decide whether the disruption has happened or not. Apart from these extreme cases of oblivious/blocking sensors, we can assume without loss in generality that

$$0 < \min_{\ell} \rho_{\ell-1,\ell} \leq \max_{\ell} \rho_{\ell-1,\ell} < 1.$$

Continuity arguments suggest that if some $\rho_{\ell-1,\ell}$ is small (but nonzero), it should be natural to expect that the ℓth sensor and beyond *may not* "effectively" contribute any information to the fusion center. We will interpret this observation after establishing performance bounds for ν_A.

Probability of False Alarm: We first show that letting $A \to \infty$ in ν_A corresponds to considering the regime where PFA $\to 0$.

Proposition 3: The probability of false alarm with ν_A can be upper bounded as

$$\mathsf{PFA} \leq \frac{1}{1+\rho\cdot\exp(A)}.$$

That is, if $\alpha \leq 1$ and the threshold A is set as $A = \log(\frac{1}{\rho\alpha})$, then PFA $\leq \alpha$.

Proof: The proof is elementary and follows the same argument as in [28] and [34]. Note that $p_{k,1}$ and ν_A can also be written as

$$\begin{aligned} p_{k,1} &= P(\{\Gamma_1 > k\}|I_k) \\ \nu_A &= \inf_k \left\{ p_{k,1} \leq \frac{1}{1+\rho\cdot\exp(A)} \right\}. \end{aligned}$$

Thus, we have

$$\mathsf{PFA} = P(\{\nu_A < \Gamma_1\}) = E[p_{\nu_A,1}] \leq \frac{1}{1+\rho\cdot\exp(A)}. \quad ■$$

Universal Lower Bound on ADD*:* We now establish a lower bound on ADD for the class of stopping times $\boldsymbol{\Delta}_\alpha$. That is, any stopping time τ should have an ADD larger than the lower bound if PFA is to be smaller than α.

Proposition 4: Consider the class of stopping times $\boldsymbol{\Delta}_\alpha = \{\tau : \mathsf{PFA}(\tau) \leq \alpha\}$. Under the assumption that $\min_{\ell=2,\ldots,L} \rho_{\ell-1,\ell} > 0$, we have

$$\inf_{\tau\in\boldsymbol{\Delta}_\alpha} \mathsf{ADD}(\tau) \geq \frac{|\log(\alpha)|\cdot(1+\mathrm{o}(1))}{LD(f_1,f_0)+|\log(1-\rho)|} \qquad \text{as } \alpha \to 0$$

where the $\mathrm{o}(1)$ term converges to zero as $\alpha \to 0$.

Proof: The proof follows on similar lines as [28, Lemma 1 and Theorem 1], but with some modifications to accommodate the change process setup. See Appendix C. ■

Upper Bound on ADD *of* ν_A*:* We will now establish an upper bound on ADD of ν_A.

Theorem 3: Let $\{\rho_{\ell-1,\ell}\}$ be such that $0 < \min_\ell \rho_{\ell-1,\ell} \leq \max_\ell \rho_{\ell-1,\ell} < 1$. Further, assume that $D(f_1,f_0)$ be such that there exists some j satisfying $\ell \leq j \leq L$ and

$$D(f_1,f_0) > \frac{1}{j-\ell+1}\log\left(\frac{\sum_{p=0}^{\ell-1}(1-\rho_{p,p+1})}{1-\rho_{j,j+1}}\right) \qquad (6)$$

for all $2 \leq \ell \leq L$. Then, the performance of ν_A with $A = \log(\frac{1}{\rho\alpha})$ is given by

$$\boxed{E[\nu_A] \leq \frac{|\log(\rho\alpha)|\cdot(1+\mathrm{o}(1))}{LD(f_1,f_0)+|\log(1-\rho)|} \quad \text{as} \quad \alpha \to 0.} \quad ■$$

Corollary 1: Combining Proposition 4 and Theorem 3, it can be seen that ν_A is asymptotically optimal (as $\alpha \to 0$) for any fixed $\rho > 0$. In other words

$$\inf_{\tau \in \mathbf{\Delta}_\alpha} \mathsf{ADD}(\tau) \sim E[\nu_A]$$

where we have used the notation $X_\alpha \sim Y_\alpha$ as $\alpha \to 0$ to mean $\lim_{\alpha \to 0} \frac{X_\alpha}{Y_\alpha} = 1$. ■

The proof of Theorem 3 in the general case of an arbitrary number (L) of sensors with an arbitrary choice of $\{\rho_{\ell-1,\ell}\}$ results in cumbersome analysis. Hence, it is worthwhile to consider the special case of two sensors that can be captured by just two change parameters: ρ and $\rho_{1,2}$. The main idea that is necessary in tackling the general case is easily exposed in the $L = 2$ setting in Section VI. The general case is carefully studied in Appendix D.

VI. Average Detection Delay: Special Case $(L = 2)$

The main statement in the $L = 2$ case is the following result.

Proposition 5: $(L = 2)$ The stopping time ν_A is such that $\nu_A \to \infty$ as $A \to \infty$. Further, if $D(f_1, f_0)$ satisfies

$$D(f_1, f_0) > \log(2 - \rho - \rho_{1,2})$$

we also have

$$\lim_{A \to \infty} \frac{E[\nu_A]}{A} \leq \frac{1}{2D(f_1, f_0) + |\log(1-\rho)|}. \quad ■$$

We will work our way to the proof of the above statement by establishing some initial results.

Proposition 6: If $0 < \{\rho, \rho_{1,2}\} < 1$, we can recast $\{q_{k,\ell}\}$ as follows:

$$q_{k,1} = \frac{1}{\rho}$$

$$q_{k,2} = \underbrace{\left(\frac{1-\rho_{1,2}}{1-\rho}\right)^k \cdot \left(1 + \frac{1-\rho_{1,2}}{1-\rho}\right)}_{\alpha_{k,2}} \cdot \underbrace{\prod_{m=1}^{k} L_{m,1}}_{C_{k,1}} \cdot \underbrace{\prod_{m=0}^{k-2} (1 + \zeta_{m,2})}_{J_{k,2}}$$

$$\zeta_{m,2} = \frac{1-\rho}{(1-\rho_{1,2}) \cdot (1 + q_{m,2}) \cdot L_{m+1,1}}$$

$$q_{k,3} = \underbrace{\frac{\rho_{1,2}}{(1-\rho)^k} \cdot \left(1 + \frac{1-\rho_{1,2}}{1-\rho} + \frac{1}{1-\rho}\right)}_{\alpha_{k,3}} \cdot \underbrace{\prod_{m=1}^{k} L_{m,1} L_{m,2}}_{C_{k,1} C_{k,2}} \cdot \underbrace{\prod_{m=0}^{k-2} (1 + \zeta_{m,3})}_{J_{k,3}}$$

$$\zeta_{m,3} = \frac{\rho_{1,2} \cdot (1 - \rho + (1-\rho_{1,2}) \cdot L_{m+1,1} \cdot (1 + q_{m,2}))}{L_{m+1,1} L_{m+1,2} \cdot (\rho_{1,2} + \rho_{1,2} q_{m,2} + q_{m,3})}.$$

Proof: We start with the recursions

$$q_{k,2} = \frac{(1-\rho_{1,2})}{1-\rho} \cdot L_{k,1} \cdot (1 + q_{k-1,2})$$

$$q_{k,3} = \frac{L_{k,1} L_{k,2}}{1-\rho} \cdot (\rho_{1,2} + \rho_{1,2} q_{k-1,2} + q_{k-1,3}).$$

The expression for $q_{k,2}$ is obtained by isolating the term $(1 + q_{k-j,2})$ at every stage as j increases from 2 to k. The expression for $q_{k,3}$ is obtained by isolating the term $(\rho_{1,2} + \rho_{1,2} q_{k-j,2} + q_{k-j,3})$ at every stage as j increases. ■

The test ν_A can now be rewritten as

$$\begin{aligned}\nu_A &= \inf_k \{\log(q_{k,2} + q_{k,3}) > A\} \\ &= \inf_k \{\log(\alpha_{k,2} \cdot C_{k,1} \cdot J_{k,2} + \alpha_{k,3} \cdot C_{k,1} C_{k,2} \cdot J_{k,3}) > A\} \\ &= \inf_k \left\{\log(\alpha_{k,2} \cdot C_{k,1} \cdot J_{k,2}) + \log\left(1 + C_{k,2} \cdot \frac{\alpha_{k,3}}{\alpha_{k,2}} \cdot \frac{J_{k,3}}{J_{k,2}}\right) > A\right\}.\end{aligned}$$

We need the following preliminaries in the course of our analysis.

Lemma 1: Since $q_{m,2} \geq 0$, note that $J_{k,2}$ can be trivially upper bounded as

$$J_{k,2} \leq \prod_{m=1}^{k-1} \left(1 + \frac{1-\rho}{(1-\rho_{1,2}) \cdot L_{m,1}}\right). \quad ■$$

Lemma2: If $\{x, x_1, x_2, \ldots\}$ are i.i.d. with $x \geq 0$ and $E[\log(x)] > 0$, then

$$\frac{1}{k} \log\left(1 + \prod_{m=1}^{k} x_m\right) - \frac{\sum_{m=1}^{k} \log(x_m)}{k} \overset{k \to \infty}{\to} 0$$

a.s. and in mean.

If $\{x, x_1, x_2, \ldots\}$ are i.i.d. with $x \geq 0$ and $E[\log(x)] \leq 0$, then

$$\frac{1}{k} \log\left(1 + \prod_{m=1}^{k} x_m\right) \overset{k \to \infty}{\to} 0 \quad \text{a.s. and in mean.}$$

Note that both these conclusions are true even if $\{x_m\}$ are not i.i.d. (or even independent) as long as the condition on the sign of $E[\log(x)]$ can be replaced with an almost sure (and in mean) statement on the sign of $\lim_n \frac{1}{n} \sum_{m=1}^{n} \log(x_m)$ (or an appropriate variant thereof). ■

The following statement, commonly referred to as the Blackwell's elementary renewal theorem [35, pp. 204-205], is needed in our proofs.

Lemma 3: Let x_m be i.i.d. positive random variables and define T_m as follows:

$$T_m = T_{m-1} + x_m, \quad m \geq 1 \quad \text{and} \quad T_0 = 0.$$

The number of renewals in $[0, t]$ is $N_t = \inf_k \{T_k > t\}$. Then, we have

$$\frac{N_t}{t} \to \frac{1}{\mu} \quad \text{a.s. as} \quad t \to \infty$$

and

$$\frac{E[N_t]}{t} \to \frac{1}{\mu} \quad \text{as } t \to \infty$$

where $\mu \triangleq E[x_m] \in (0, \infty]$. ■

Proof of Proposition 5 ($L = 2$): We will postpone the proof of the first statement to Appendix D when we consider the general case in Proposition 9. For the second statement, we first use the bound for $J_{k,2}$ from Lemma 1 and the fact that $\zeta_{m,\ell} \geq 0$, and thus we have

$$\begin{aligned}
&\log\left(1 + C_{k,2} \cdot \frac{\alpha_{k,3}}{\alpha_{k,2}} \cdot \frac{J_{k,3}}{J_{k,2}}\right) \\
&\geq \log\left(1 + C_{k,2} \cdot \frac{\alpha_{k,3}}{\alpha_{k,2}} \cdot \frac{1}{\prod_{m=1}^{k-1}\left(1 + \frac{1-\rho}{(1-\rho_{1,2})L_{m,1}}\right)}\right) \\
&\geq \log\left(1 + \prod_{m=1}^{k} \frac{\rho_{1,2}^{1/k} \cdot L_{m,2}}{(1-\rho_{1,2}) \cdot \left(1 + \frac{1-\rho}{(1-\rho_{1,2})L_{m,1}}\right)}\right).
\end{aligned}$$

Now, observe that

$$\begin{aligned}
&E\left[\log\left(\frac{L_{m,2}}{(1-\rho_{1,2}) \cdot \left(1 + \frac{1-\rho}{(1-\rho_{1,2})L_{m,1}}\right)}\right)\right] \\
&= D(f_1, f_0) + \log\left(\frac{1}{1-\rho_{1,2}}\right) \\
&\quad - E\left[\log\left(1 + \frac{1-\rho}{(1-\rho_{1,2})L_{m,1}}\right)\right] \\
&\geq D(f_1, f_0) + \log\left(\frac{1}{1-\rho_{1,2}}\right) \\
&\quad - \log\left(1 + E\left[\frac{1-\rho}{(1-\rho_{1,2})L_{m,1}}\right]\right) \\
&= D(f_1, f_0) - \log(2 - \rho - \rho_{1,2}) > 0
\end{aligned}$$

where the first equality follows since $\rho_{1,2} > 0$ (change has to eventually happen at the second sensor to ensure that $E[\log(L_{m,2})] = D(f_1, f_0)$), the second step follows from Jensen's inequality and the third equality from the fact that $E_{f_1}[\frac{1}{L_{m,1}}] = 1$. Using this fact in conjunction with Lemma 2 and noting that $\rho_{1,2} > 0$, as $k \to \infty$, we have

$$\begin{aligned}
&\log(\alpha_{k,2} \cdot C_{k,1} \cdot J_{k,2}) + \log\left(1 + C_{k,2} \cdot \frac{\alpha_{k,3}}{\alpha_{k,2}} \cdot \frac{J_{k,3}}{J_{k,2}}\right) \\
&\geq \log(C_{k,1} C_{k,2} \cdot \alpha_{k,3} \cdot J_{k,3}) \\
&\geq \underbrace{\sum_{m=1}^{k} \log\left(\frac{\rho_{1,2}^{1/k} \cdot L_{m,1} \cdot L_{m,2}}{1-\rho}\right)}_{L_k}.
\end{aligned}$$

The above relationship implies that $\nu_A \leq \nu_{L,A}$ where

$$\nu_{L,A} \triangleq \inf_k \{L_k > A\}.$$

Applying Lemma 3 (since the entries in the definition of $\nu_{L,A}$ are independent) and the first statement of the theorem that $\nu_A \to \infty$ as $A \to \infty$, we have

$$\frac{E[\nu_A]}{A} \leq \frac{E[\nu_{L,A}]}{A} \overset{A \to \infty}{\to} \frac{1}{2D(f_1, f_0) + |\log(1-\rho)|}.$$

■

The general case where $L > 2$ is discussed in Appendix D.

VII. Discussion and Numerical Results

Discussion: A loose sufficient condition for all the L sensors to contribute to the slope of ADD of ν_A is that

$$\begin{aligned}
D(f_1, f_0) > &\max_{\ell=1,\ldots,L-1} \min_{j \geq \ell+1} \frac{1}{j-\ell} \\
&\cdot \log\left(\frac{\sum_{p=0}^{\ell}(1-\rho_{p,p+1})}{1-\rho_{j,j+1}}\right) \\
&\triangleq \gamma_u.
\end{aligned}$$

Another sufficient condition is that

$$D(f_1, f_0) > \max_{\ell=1,\ldots,L-1} \frac{1}{L-\ell} \cdot \log\left(1 - \rho + \sum_{j=1}^{\ell}(1-\rho_{j,j+1})\right).$$

That is, if ρ is such that

$$\rho \geq \sum_{\ell=2}^{L}(1-\rho_{\ell-1,\ell})$$

then $\gamma_u \leq 0$ and the condition of Theorem 3 reduces to a mild one that the K-L divergence between f_1 and f_0 be positive. A special setting where the above condition is true (irrespective of the rarity of the disruption-point) is the regime where change propagates across the sensor array "quickly." The case instantaneous propagation is an extreme example of this regime and Theorem 3 recaptures this extreme case.

In more general regimes where change propagates across the sensor array "slowly," either the disruption-point should become less rare (independent of the choice of f_1 and f_0) or that the densities f_1 and f_0 be sufficiently discernible (independent of the rarity of the disruption-point) so that all the L sensors can contribute to the asymptotic slope. When these conditions fail to hold, it is not clear whether the theorems are applicable, or even if all the L sensors contribute to the slope of $E[\nu_A]$. Nevertheless, it is reasonable to conjecture that as long as $\min_\ell \rho_{\ell-1,\ell} > 0$, then all the L sensors contribute to the asymptotic slope.

However, the difference between the asymptotic and the nonasymptotic regimes needs a careful revisit. Following the initial remark (Proposition 2) on the extreme case of blocking sensors (where some $\rho_{\ell-1,\ell} = 0$), in the more realistic case where some $\rho_{\ell-1,\ell}$ may be small (but nonzero), it is possible that if $D(f_1, f_0)$ is smaller than some threshold value (determined by the change propagation parameters), not all of the L sensors may "effectively" contribute to the slope of ADD, at

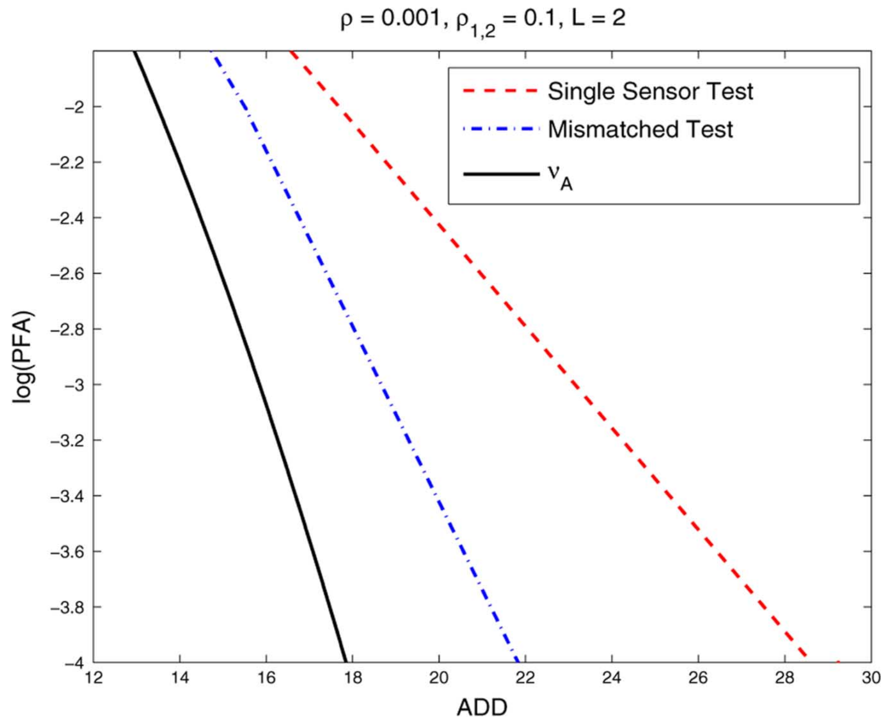

Fig. 2. Probability of false alarm versus Average detection delay for a $L = 2$ setting with $\rho = 0.001$ and $\rho_{1,2} = 0.1$.

least for reasonably small, but nonasymptotic values of PFA. For example, see the ensuing discussion where numerical results illustrate this behavior at PFA values of 10^{-4} to 10^{-5} for some choice of change propagation parameters, *even* when the condition in Theorem 3 is met. When the condition in Theorem 3 is not met, such a behavior is expected to be more typical.

The final comment is on the approach pursued in this paper. While the approach pursued in Section VI and Appendix D results in interesting conclusions, it is not clear if this approach is *fundamental* in the sense that this is the only approach possible for characterizing ADD versus PFA. Furthermore, this approach assumes the existence of $\{\gamma_{\ell,j}\}$ (see Appendix D). Even if these quantities exist and are hence, theoretically computable, such a computation is complicated by the fact that $\{\zeta_{m,\ell}, m = 1, \ldots, k\}$ are correlated. Thus, verification of the exact condition in Proposition 10 (equivalently, computing $\ell^\star$) has to be achieved either via Monte Carlo methods or by bounding $\Delta_{\ell,j}$, as done here. Furthermore, correlation of $\{\zeta_{m,\ell}\}$, and hence, y_m [see (11)]) implies that statistics of ν_A have to be obtained using nonlinear renewal theoretic techniques for general (correlated) random variables [36]. This is the subject of current work.

Numerical Study I—Performance Improvement With ν_A: Given that the structure of τ_{opt} is not known in closed-form, we now present numerical studies to show that ν_A results in substantial improvement in performance over both a single sensor test (which uses the observations only from the first sensor and ignores the other sensor observations) and a test that uses the observations from all the sensors but under a mismatched model (where the change-point for all the sensors is assumed to be the same), even under realistic modeling assumptions.

The first example corresponds to a two sensor system where the occurrence of change is modeled as a geometric random variable with parameter $\rho = 0.001$. Change propagates from the first sensor to the second with the geometric parameter $\rho_{1,2} = 0.1$. The pre- and post-change densities are $\mathcal{CN}(0,1)$ and $\mathcal{CN}(1,1)$, respectively so that $D(f_1, f_0) = 0.50$. While the threshold for ν_A is set as in Proposition 3, the thresholds for the single sensor and mismatched tests are set as in [28]. The recursion for the sufficient statistic of the mismatched test follows the description in [29]. Fig. 2 depicts the performance of the three tests obtained via Monte Carlo methods and shows that ν_A can result in an improvement of at least 4 units of delay at even marginally large PFA values on the order of 10^{-3}.

The second example corresponds to a five sensor system where $\rho = 0.005$. Change propagates across the array according to the following model: $\rho_{1,2} = 0.1, \rho_{2,3} = 0.2, \rho_{3,4} = 0.5$ and $\rho_{4,5} = 0.7$. The pre- and the post-change densities are $\mathcal{CN}(0,1)$ and $\mathcal{CN}(0.75,1)$ so that $D(f_1, f_0) \approx 0.2813$. With $D(f_1, f_0)$ and the change parameters as above, Theorem 3 assures us that at least $L = 2$ sensors contribute to the ADD versus PFA slope asymptotically. On the other hand, Fig. 3 shows that more than two sensors indeed contribute to the slope. Thus, it can be seen that Theorem 3 provides only a sufficient condition on performance bounds. It is also worth noting the transition in slope (unlike the case in [29]) for both the mismatched test and ν_A as PFA decreases from moderately large values to zero, whereas the slope of the single sensor test (as expected) remains constant.

Numerical Study II—Performance Gap Between the Tests: We now present a second case-study with the main goal being the understanding of the relative performance of ν_A with respect to the single sensor and the mismatched tests. We again consider a $L = 2$ sensor system and we vary the change process parameters, ρ and $\rho_{1,2}$, in this study. The pre- and the post-change densities are $\mathcal{CN}(0,1)$ and $\mathcal{CN}(1.2,1)$ so that $D(f_1, f_0) = 0.72$.

Figs. 4 and 5(b) show the performance of the three tests with varying ρ parameters for a fixed choice of $\rho_{1,2}$. We observe that the gap in performance between the single sensor test and ν_A

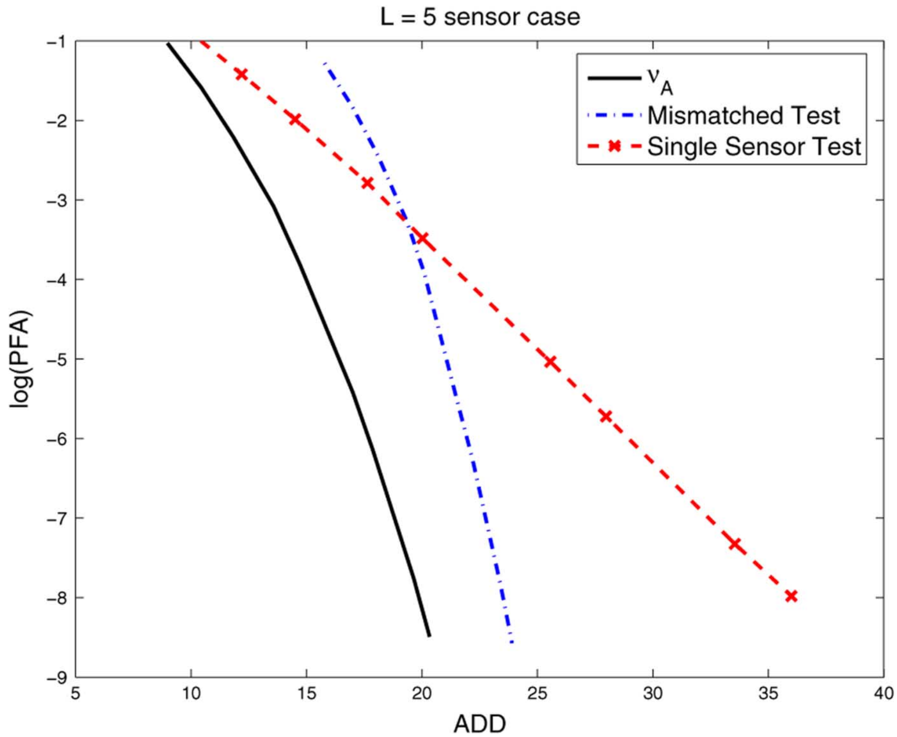

Fig. 3. Probability of false alarm versus Average detection delay for a typical $L = 5$ setting.

increases as ρ decreases, whereas the gap between ν_A and the mismatched test stays fairly constant. Similarly, Fig. 5 shows the performance of the three tests with varying $\rho_{1,2}$ parameters for a fixed choice of ρ. We observe from these plots that the gap between the mismatched test and ν_A increases as $\rho_{1,2}$ decreases, whereas the gap between the single sensor test and ν_A increases as $\rho_{1,2}$ increases.

The choice of $D(f_1, f_0) = 0.72$ is such that the sufficient condition in Theorem 3 are satisfied, independent of the change parameters. Hence, we expect the slope of the ADD versus PFA plot to be of the form $\frac{1}{2D(f_1,f_0)+|\log(1-\rho)|}$ *asymptotically* as PFA $\to 0$. Nevertheless, Fig. 5(c) and (d) show that, when both ρ and $\rho_{1,2}$ are small, the slope of ν_A is only as good as (or slightly better than) the single sensor test, which is known to have a slope of the form $\frac{1}{D(f_1,f_0)+|\log(1-\rho)|}$. Thus, we see that even though our theory guarantees that both the sensors' observations contribute in the eventual performance of ν_A *asymptotically*, we may not see this behavior for reasonable choices of PFA such as 10^{-4}. The case of observation models not meeting the conditions of Theorem 3 is expected to show this trend for even lower PFA values.

To summarize these observations, if ADD_{ν_A}, $\mathsf{ADD}_{\mathsf{MM}}$ and $\mathsf{ADD}_{\mathsf{SS}}$ denote the average detection delays for ν_A, mismatched and single sensor tests (respectively) for some fixed choice of PFA, then

$$\mathsf{ADD}_{\mathsf{MM}} - \mathsf{ADD}_{\nu_A} \propto \frac{1}{\rho_{1,2}} \text{ and independent of } \rho$$
$$\mathsf{ADD}_{\mathsf{SS}} - \mathsf{ADD}_{\nu_A} \propto \frac{\rho_{1,2}}{\rho}.$$

It is interesting to note from the above equations that $\rho_{1,2}$ impacts the gap between the two tests in a contrasting way. The test ν_A is expected to result in significant performance improvement in the regime where ρ is small, but $\rho_{1,2}$ is neither too small nor too large. In fact, this regime where ν_A is expected to result in significant performance improvement is the precise regime that is of importance in practical contexts. This is so because we can expect the occurrence of disruption (e.g., cracks in bridges, intrusions in networks, onset of epidemics, etc.) to be a rare phenomenon. Once the disruption occurs, we expect change to propagate across the sensor array fairly quickly due to the geographical (network proximity in the case of computer networks) proximity of the other sensors, but not so quick that the extreme case of instantaneous propagation is applicable. Classifying the regime of $\{\rho_{\ell-1,\ell}\}$ and $D(f_1, f_0)$ where significant performance improvement is possible with ν_A is ongoing work. It is also of interest to come up with better test structures in the regime where ν_A does not lead to a significant performance improvement.

VIII. Concluding Remarks

We considered the centralized, Bayesian version of the change process detection problem in this work and posed it in the classical dynamic programming framework. This formulation of the change detection problem allows us to establish the sufficient statistics for the DP under study and a recursion for the sufficient statistics. While we obtain the broad structure of the optimal stopping rule (τ_{opt}), any further insights into it are rendered infeasible by the complicated nature of the infinite-horizon cost-to-go function. Nevertheless, τ_{opt} reduces to a threshold rule (denoted in this work as ν_A) in the rare disruption regime.

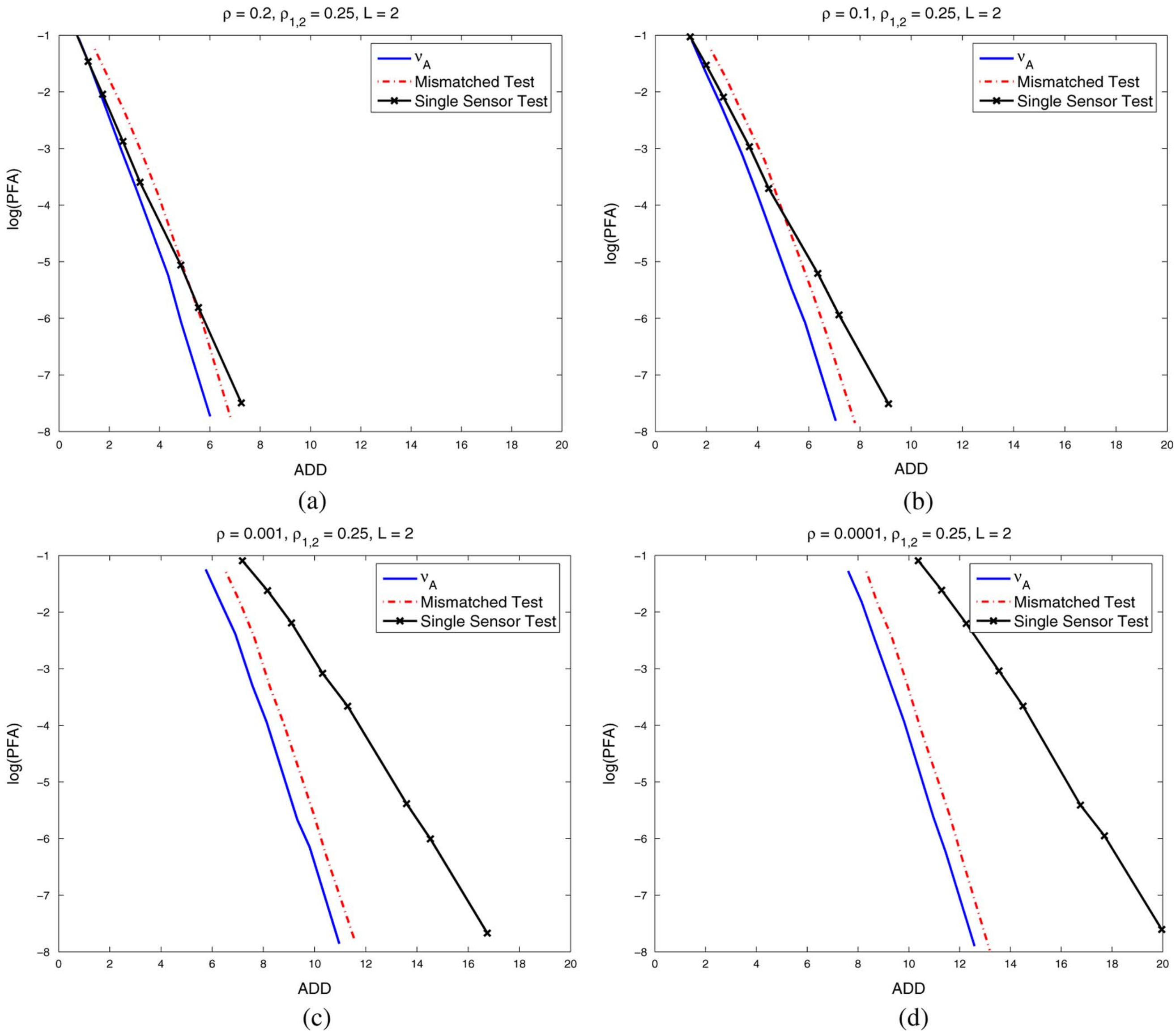

Fig. 4. Probability of false alarm versus Average detection delay for a $L = 2$ setting with different model parameters.

The test ν_A possesses the following properties and thus serves as an attractive test for practical applications that can be modeled with a change process: i) it is of low-complexity; ii) under certain mild sufficient conditions (more specifically, if the K-L divergence $D(f_1, f_0)$ is more than a number determined by the parameters of the change process), it is asymptotically optimal in the small PFA regime; and iii) numerical studies suggest that it can lead to substantially improved performance over naive tests. Nevertheless, the asymptotic expansion of ADD in terms of $\log(\mathsf{PFA})$ is not enough to determine how small the false alarm probability should be in order for this expansion and asymptotic optimality of ν_A to hold. Studies indicate that PFA should be chosen significantly smaller than those needed for good approximations in the simpler quickest detection problems solved earlier by the same approach.

Apart from the recent work of [32], the change process detection problem has not been studied in detail. Thus, there exists potential for extending this work in multiple new directions. While we established the asymptotic optimality of ν_A when $D(f_1, f_0) \geq \gamma_u$, it is unclear as to what happens when $D(f_1, f_0) < \gamma_u$. In other words, is $\ell^\star = L + 1$ when $D(f_1, f_0) < \gamma_u$ given that $\gamma_u > 0$? It is most likely that ν_A is asymptotically optimal even in this regime as long as $\min_\ell \rho_{\ell-1,\ell} > 0$, but establishing this result may require new analysis tools. However, if ν_A is not asymptotically optimal in this regime, it is of interest to design better low-complexity stopping rules, e.g., threshold tests on weighted sums of the *a posteriori* probabilities based on further study of the structure of τ_{opt}.

More careful asymptotic analysis of ν_A and the performance gap between ν_A and other tests would involve tools from nonlinear renewal theory [25], [28], [36] and is the subject of current attention. Such an asymptotic study could in turn drive the design of better test structures. Our numerical results also illustrate and motivate the need for nonasymptotic characterization (e.g., piece-wise linear approximations of the ADD versus $\log(\mathsf{PFA})$ curve) of the proposed tests.

Extensions of this work to more general observation models are important in the context of practical applications. For example, non-iid [28] and Hidden-Markov models [23] have

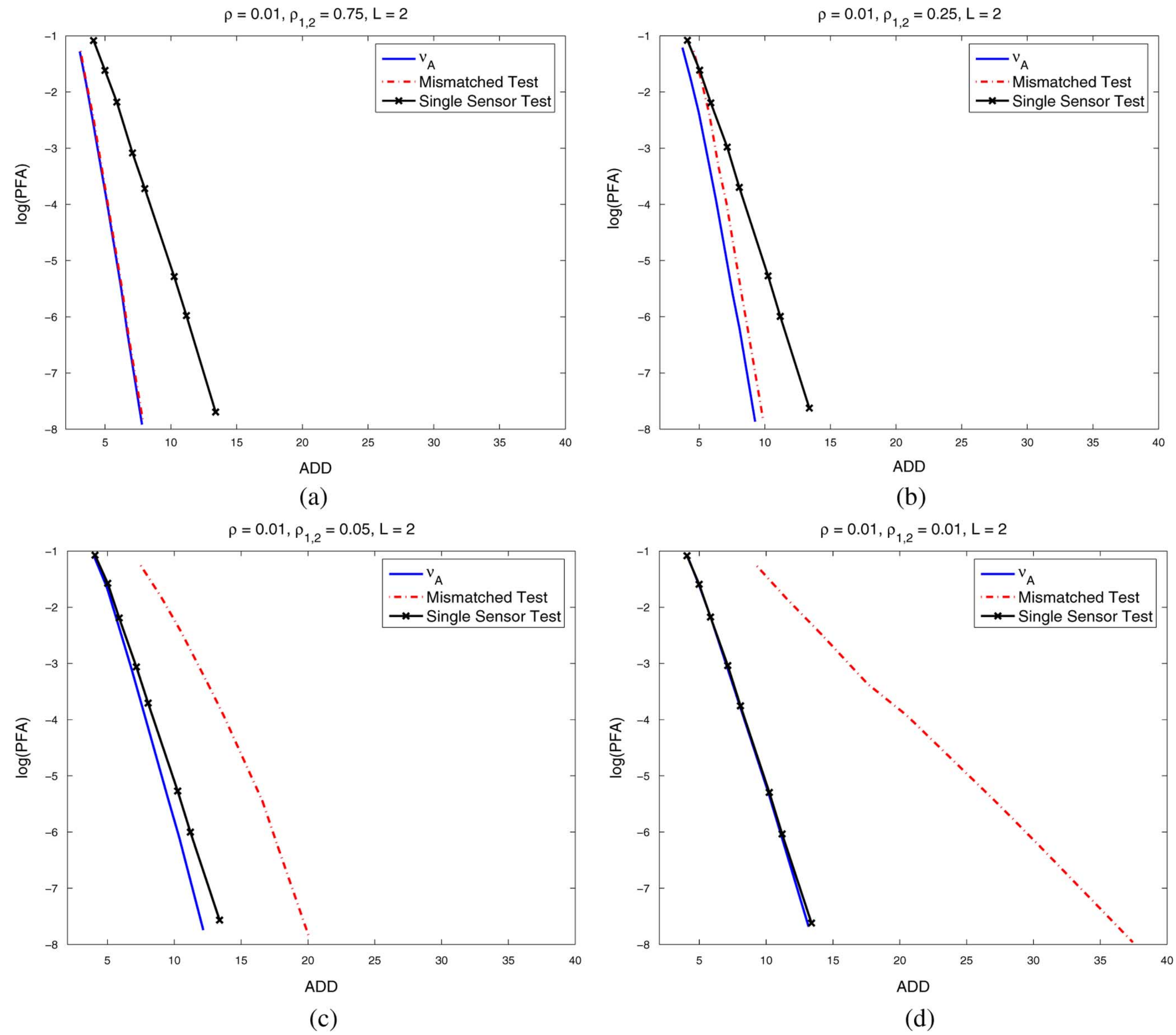

Fig. 5. Probability of false alarm versus Average detection delay for a $L = 2$ setting with different model parameters.

found increased interest in biological problems determined by an event-driven potential [5], [6]. Practical applications will in turn drive the need for understanding quickest change detection with certain specific observation models.

APPENDIX

A. Proof of Theorem 1

Before considering the infinite-horizon DP, we will study the finite-horizon version and establish some properties along the directions of [29], [37], [33]. A straightforward induction argument shows that if T is fixed

$$0 \leq J_k^T(\boldsymbol{p}) \leq 1 \quad \text{for all } 0 \leq k \leq T$$
$$0 \leq A_k^T(\boldsymbol{p}) \leq 1 \quad \text{for all } 0 \leq k \leq T.$$

Similarly, it is easy to observe that for any k, $A_k^T(\boldsymbol{p})$ and $J_k^T(\boldsymbol{p})$ equal zero if $p_1 = 0$. A routine induction argument (illustrated at the end of the main part of this proof) establishes the concavity of $A_k^T(\cdot)$ and $J_k^T(\cdot)$.

We now consider the infinite-horizon DP and show that it is well-defined. (That is, we remove the restriction that the stopping time is finite and let $T \to \infty$.) Towards this end, we need to establish that $\lim_T J_k^T(\cdot)$ exists, which is done as follows: By an induction argument, we note that for any $\boldsymbol{p}$ and T fixed, we have

$$J_k^T(\boldsymbol{p}) \leq J_{k+1}^T(\boldsymbol{p}), \quad 0 \leq k \leq T-1.$$

It is important to note that this conclusion critically depends on the joint-geometric assumption of the change process [in particular, the *memorylessness* property that results in the independence of $w_{k,\ell,m}$ on k in (3)] and the i.i.d. nature of the observation process conditioned on the change-point.

Using a similar induction approach, observe that for any $\boldsymbol{p}$ and k fixed, $J_k^{T+1}(\boldsymbol{p}) \leq J_k^T(\boldsymbol{p})$. Heuristically, this can also be seen to be true because the set of stopping times increases with

T. Since $J_k^T(\boldsymbol{p}) \geq 0$ for all k and T, for any fixed k, we can let $T \to \infty$ and we have

$$\lim_T J_k^T(\boldsymbol{p}) = \inf_{T:T>k} J_k^T(\boldsymbol{p}) \triangleq J_k^\infty(\boldsymbol{p}).$$

Furthermore, the *memorylessness* property and the i.i.d. observation process results in the invariance of $J_k^\infty(\boldsymbol{p})$ on k. This is shown by a simple time-shift argument. Denote this common limit as $J(\boldsymbol{p})$.

A simple dominated convergence argument [35] then shows that $\lim_T A_k^T(\boldsymbol{p})$ is well-defined and independent of k. If we denote this limit as $A_J(\boldsymbol{p})$, we have

$$\begin{aligned} A_J(\boldsymbol{p}) &= \int J(\boldsymbol{p})|_{\boldsymbol{Z}=\boldsymbol{z}} f(Z|I_\bullet)|_{\boldsymbol{Z}=\boldsymbol{z}} d\boldsymbol{z} \\ &= \int J(\boldsymbol{p})|_{\boldsymbol{Z}=\boldsymbol{z}} \left\{ \sum_{j=1}^{L+1} \left((1-\rho_{j-1,j}) \cdot \sum_{m=1}^{j} w_m^j p_m \right) \right. \\ &\qquad\qquad \left. \times \boldsymbol{\Phi}_{\mathsf{obs}}(\bullet, j) \right\} |_{\boldsymbol{Z}=\boldsymbol{z}} d\boldsymbol{z} \end{aligned}$$

where we have denoted the dependence of $J(\boldsymbol{p})$ on $\boldsymbol{z}$ by the notation $J(\boldsymbol{p})|_{\boldsymbol{Z}=\boldsymbol{z}}$ and the fact that $\boldsymbol{\Phi}_{\mathsf{obs}}(k,j)|_{\boldsymbol{Z}=\boldsymbol{z}}$ is independent of k is denoted as $\boldsymbol{\Phi}_{\mathsf{obs}}(\bullet, j)$. Hence, the infinite-horizon cost-to-go can be written as

$$J(\boldsymbol{p}) = \min\{p_1,\ c(1-p_1) + A_J(\boldsymbol{p})\}.$$

The structure of $A_J(\boldsymbol{p})$ follows from the finite-horizon characterization by letting $T \to \infty$.

Establishing Concavity of $A_k^T(\cdot)$ and $J_k^T(\cdot)$: We now show that $A_k^T(\boldsymbol{p}_k)$ and $J_k^T(\boldsymbol{p}_k)$ are concave in $\boldsymbol{p}_k$. First, note that $J_T^T(\boldsymbol{p}_T) = p_{T,1}$ is concave in $\boldsymbol{p}_T$ because it is affine. Using the recursion for $\boldsymbol{p}_T$, it is straightforward to check that

$$A_{T-1}^T(\boldsymbol{p}_{T-1}) = E[J_T^T(\boldsymbol{p}_T)|I_{T-1}] = p_{T-1,1}\cdot(1-\rho).$$

Using this in the definition of $J_{T-1}^T(\boldsymbol{p}_{T-1})$, we have

$$J_{T-1}^T(\boldsymbol{p}_{T-1}) = \begin{cases} p_{T-1,1}, & 0 \leq p_{T-1,1} \leq \frac{c}{c+\rho} \\ c + p_{T-1,1}(1-\rho-c), & \frac{c}{c+\rho} \leq p_{T-1,1} \leq 1. \end{cases}$$

Since both $A_{T-1}^T(\boldsymbol{p}_{T-1})$ and $J_{T-1}^T(\boldsymbol{p}_{T-1})$ are affine and piecewise-affine (It is important to note that the slope of the second affine part, which is $1-\rho-c$, is smaller than the first $(=1)$.) in $\boldsymbol{p}_{T-1,1}$ respectively, they are concave.

We now assume that $J_{k+1}^T(\boldsymbol{p}_{k+1})$ is concave in $\boldsymbol{p}_{k+1}$ and show that $A_k^T(\boldsymbol{p}_k)$ is also concave in $\boldsymbol{p}_k$. For this, consider $\lambda A_k^T(\boldsymbol{p}_k^1) + (1-\lambda)A_k^T(\boldsymbol{p}_k^2)$ with $\boldsymbol{p}_k^1$ and $\boldsymbol{p}_k^2$ being two elements in the standard L-dimensional simplex. We have

$$\begin{aligned} &\lambda A_k^T\left(\boldsymbol{p}_k^1\right) + (1-\lambda)A_k^T\left(\boldsymbol{p}_k^2\right) \\ &\quad= \int \left[\lambda J_{k+1}^T(\boldsymbol{p}_{k+1}^1)\mu_1 + (1-\lambda)J_{k+1}^T\left(\boldsymbol{p}_{k+1}^2\right)\mu_2\right]|_{\boldsymbol{Z}_{k+1}=\boldsymbol{z}} d\boldsymbol{z} \\ &\quad= \int \left[\mu J_{k+1}^T\left(\boldsymbol{p}_{k+1}^1\right) + (1-\mu)J_{k+1}^T\left(\boldsymbol{p}_{k+1}^2\right)\right] \\ &\qquad \times (\lambda\mu_1 + (1-\lambda)\mu_2)|_{\boldsymbol{Z}_{k+1}=\boldsymbol{z}} d\boldsymbol{z} \end{aligned}$$

where

$$\begin{aligned} \mu_i &= f(\boldsymbol{Z}_{k+1}|I_k)|_{\boldsymbol{p}_k=\boldsymbol{p}_k^i} \\ &= \sum_{j=1}^{L+1}\left[\left(\sum_{m=1}^{j} w_{k+1,j,m} p_{k,m}^i\right)\boldsymbol{\Phi}_{\mathsf{obs}}(k+1,j)\right], \qquad i=1,2 \\ \mu &= \frac{\lambda\mu_1}{\lambda\mu_1 + (1-\lambda)\mu_2}. \end{aligned}$$

Using the concavity of $J_{k+1}^T(\cdot)$, we can upper bound the above as follows:

$$\begin{aligned} &\lambda A_k^T\left(\boldsymbol{p}_k^1\right) + (1-\lambda)A_k^T\left(\boldsymbol{p}_k^2\right) \\ &\leq \int \left[J_{k+1}^T\left(\mu\boldsymbol{p}_{k+1}^1 + (1-\mu)\boldsymbol{p}_{k+1}^2\right)(\lambda\mu_1 + (1-\lambda)\mu_2)\right]|_{\boldsymbol{Z}_{k+1}=\boldsymbol{z}} d\boldsymbol{z}. \end{aligned}$$

If we define

$$\boldsymbol{p}_k^3 \triangleq \lambda\boldsymbol{p}_k^1 + (1-\lambda)\boldsymbol{p}_k^2$$

it is straightforward to check that

$$\boldsymbol{p}_{k+1}^3 = \mu\boldsymbol{p}_{k+1}^1 + (1-\mu)\boldsymbol{p}_{k+1}^2.$$

Using these facts, we have

$$\lambda A_k^T(\boldsymbol{p}_k^1) + (1-\lambda)A_k^T(\boldsymbol{p}_k^2) \leq A_k^T(\lambda\boldsymbol{p}_k^1 + (1-\lambda)\boldsymbol{p}_k^2)$$

thus establishing the concavity of $A_k^T(\cdot)$. The concavity of $J_k^T(\cdot)$ follows since the minimum and sum of concave functions is concave. An inductive argument completes the proof. ■

B. Proof of Theorem 2

We will show that

$$\tau_{\mathsf{opt}} \overset{\rho\downarrow 0}{\to} \begin{cases} \text{Stop} & \text{if } \sum_{j=2}^{L+1} q_{k,j} \geq \frac{1}{c} \\ \text{Continue} & \text{if } \sum_{j=2}^{L+1} q_{k,j} \leq \frac{1-h(\rho)}{c} \end{cases}$$

(in probability) for an appropriately chosen function $h(\rho)$ that satisfies $\lim_{\rho\to 0} h(\rho) = 0$. We start with the finite-horizon DP and define $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Psi}_k$ as follows:

$$\begin{aligned} \boldsymbol{\Phi}_k &\triangleq \frac{1}{1+\rho\sum_{j=2}^{L+1} q_{k,j}} - J_k^T(\boldsymbol{q}_k), \qquad 0 \leq k \leq T \\ \boldsymbol{\Psi}_k &\triangleq A_k^T(\boldsymbol{q}_k) - \frac{1-\rho}{1+\rho\sum_{j=2}^{L+1} q_{k,j}}, \qquad 0 \leq k \leq T-1. \end{aligned}$$

The main idea behind the proof is to show that $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Psi}_k$ are bounded by a function of ρ (that goes to 0 as $\rho \to 0$), *uniformly* for all k. Thus, the structure of the test in the limit as $\rho \to 0$ can be obtained.

Towards this goal, note from Appendix A that $\boldsymbol{\Phi}_T = \boldsymbol{\Psi}_{T-1} = 0$. Also, note that $J_{T-1}^T(\boldsymbol{q}_{T-1})$ can be written as

$$J_{T-1}^T(\boldsymbol{q}_{T-1}) = \begin{cases} \frac{1-\rho+\rho c\sum_{j=2}^{L+1} q_{T-1,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-1,j}}, & 0 \leq \sum_{j=2}^{L+1} q_{T-1,j} \leq \frac{1}{c} \\ \frac{1}{1+\rho\sum_{j=2}^{L+1} q_{T-1,j}} & \sum_{j=2}^{L+1} q_{T-1,j} \geq \frac{1}{c} \end{cases}$$

which can be equivalently written as

$$\mathbf{\Phi}_{T-1} = \rho \cdot \frac{1 - c\sum_{j=2}^{L+1} q_{T-1,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-1,j}} \cdot \mathbb{1}\left(\left\{\sum_{j=2}^{L+1} q_{T-1,j} \leq \frac{1}{c}\right\}\right).$$

Note that $0 \leq \mathbf{\Phi}_{T-1} \leq \rho$ and we have

$$0 \leq E[\mathbf{\Phi}_{T-1}|I_{T-2}] \triangleq -\mathbf{\Psi}_{T-2} = \rho g_2(\rho)$$

where

$$g_2(\rho) \triangleq E\left[\underbrace{\frac{1-c\sum_{j=2}^{L+1} q_{T-1,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-1,j}} \mathbb{1}\left(\left\{\sum_{j=2}^{L+1} q_{T-1,j} \leq \frac{1}{c}\right\}\right)}_{X_\rho}\middle| I_{T-2}\right].$$

Now observe that X_ρ can be rewritten as

$$X_\rho = \frac{1 - c\sum_{j=2}^{L+1} q_{T-1,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-1,j}} \cdot \mathbb{1}\left(\left\{p_{T-1,1} \geq \frac{c}{c+\rho}\right\}\right).$$

Furthermore, $X_\rho \leq 1$ for all ρ and the set within the indicator function (above) converges to the empty set as $\rho \downarrow 0$. Thus, a straightforward consequence of the bounded convergence theorem for conditional expectation [35] is that

$$\lim_{\rho\downarrow 0} g_2(\rho) = 0$$
$$\frac{\mathbf{\Psi}_{T-2}}{\rho} \stackrel{\rho\downarrow 0}{\to} 0$$

independent of the choice of T.

Plugging the above relation in the expression for $J_{T-2}^T(\boldsymbol{q}_{T-2})$, we have

$$\begin{aligned} J_{T-2}^T(\boldsymbol{q}_{T-2}) &= \min\left\{\frac{1}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}}, \frac{1-\rho+\rho c\sum_{j=2}^{L+1} q_{T-2,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}} + \mathbf{\Psi}_{T-2}\right\} \\ &= \min\left\{\frac{1}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}}, \frac{1-\rho\left(1-\frac{\mathbf{\Psi}_{T-2}}{\rho}\right)+\rho c\sum_{j=2}^{L+1} q_{T-2,j}\left(1+\frac{\mathbf{\Psi}_{T-2}}{c}\right)}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}}\right\} \\ &= \frac{1}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}} - \mathbf{\Phi}_{T-2} \end{aligned}$$

$$\begin{aligned} \mathbf{\Phi}_{T-2} &= \frac{\rho - \mathbf{\Psi}_{T-2} - \rho(c+\mathbf{\Psi}_{T-2})\sum_{j=2}^{L+1} q_{T-2,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}} \cdot \mathbb{1}\left(\left\{\sum_{j=2}^{L+1} q_{T-2,j} \leq \frac{1}{c}\cdot\frac{1-\frac{\mathbf{\Psi}_{T-2}}{\rho}}{1+\frac{\mathbf{\Psi}_{T-2}}{c}}\right\}\right) \\ &= \rho\cdot\left[\frac{1-c\sum_{j=2}^{L+1} q_{T-2,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}} + g_2(\rho)\right] \cdot \mathbb{1}\left(\left\{p_{T-2,1} \geq \frac{c-\rho g_2(\rho)}{c+\rho}\right\}\right) \end{aligned}$$

with $0 \leq \mathbf{\Phi}_{T-2} \leq \rho(1+g_2(\rho))$. As before, it is straightforward to check that the set within the indicator function converges to the empty set as $\rho \downarrow 0$ and we can write $\mathbf{\Psi}_{T-3}$ as

$$\begin{aligned} -\mathbf{\Psi}_{T-3} &= E[\mathbf{\Phi}_{T-2}|I_{T-3}] = \rho g_3(\rho) \\ g_3(\rho) &= E\left[\left(\frac{1-c\sum_{j=2}^{L+1} q_{T-2,j}}{1+\rho\sum_{j=2}^{L+1} q_{T-2,j}} + g_2(\rho)\right) \cdot \mathbb{1}\left(\left\{p_{T-2,1} \geq \frac{c-\rho g_2(\rho)}{c+\rho}\right\}\right)\middle| I_{T-3}\right] \end{aligned}$$

with

$$\lim_{\rho\downarrow 0} g_3(\rho) = 0 \text{ and } \frac{\mathbf{\Psi}_{T-3}}{\rho} \stackrel{\rho\downarrow 0}{\to} 0.$$

Following the same logic inductively, it can be checked that

$$\frac{\mathbf{\Psi}_{T-k}}{\rho} \stackrel{\rho\downarrow 0}{\to} 0, \quad 1 \leq k \leq T$$

independent of the choice of T. That is, we have

$$J_k^T(\boldsymbol{q}_k) = \min\left\{\frac{1}{1+\rho\sum_{j=2}^{L+1} q_{k,j}}, \frac{1-\rho+\rho c\sum_{j=2}^{L+1} q_{k,j}}{1+\rho\sum_{j=2}^{L+1} q_{k,j}} + \mathbf{\Psi}_k\right\}.$$

Thus, the test structure reduces to stopping when

$$\sum_{j=2}^{L+1} q_{k,j} \geq \frac{1}{c}\cdot\frac{1-\frac{\mathbf{\Psi}_k}{\rho}}{1+\frac{\mathbf{\Psi}_k}{c}}$$

and using the limiting form for $\mathbf{\Psi}_k$ as $\rho \to 0$, we have the threshold structure (as stated). The proof is complete by going from the finite-horizon DP to the infinite-horizon version as in the proof of Theorem 1. Note that while we expect the limiting test structure in the finite-horizon setting to be dependent on T, it is not seen to be the case in this work because $\rho = 0$ is a discontinuity point for the DP. ■

C. Proof of Proposition 4

We first intend to show that a version of [28, Lemma 1] holds in our case. More precisely, our goal is to show that for any $\epsilon \in (0,1)$, we have

$$\lim_{\alpha\to 0} \sup_{\tau\in\mathbf{\Delta}_\alpha} P_k(\{k \leq \tau < k+(1-\epsilon)L_\alpha\}) = 0$$

where $P_k(\{\cdot\})$ denotes the probability measure when $\Gamma_1 = k$ and

$$L_\alpha \triangleq \frac{\log\left(\frac{1}{\alpha}\right)}{LD(f_1, f_0) + |\log(1-\rho)|}.$$

Note that $L_\alpha \to \infty$ as $\alpha \to 0$. Following along the logic of the proof of [28, Lemma 1] here, it can be seen that

$$\begin{aligned} &P_k(\{k \le \tau < k + (1-\epsilon)L_\alpha\}) \\ &\quad \le \exp((1-\epsilon^2)qL_\alpha)P_\infty(\{k \le \tau < k + (1-\epsilon)L_\alpha\}) \\ &\qquad + P_k\left(\left\{\max_{0\le n<(1-\epsilon)L_\alpha} \mathcal{Z}_{k+n}^k \ge (1-\epsilon^2)qL_\alpha\right\}\right) \end{aligned} \tag{7}$$

where $q \triangleq LD(f_1, f_0)$, $P_\infty(\{\cdot\})$ denotes the probability measure when no change happens, and

$$\mathcal{Z}_{k+n}^k = \sum_{\ell=1}^{L} \sum_{i=\Gamma_\ell}^{k+n} \log\left(\frac{f_1(Z_{i,\ell})}{f_0(Z_{i,\ell})}\right)$$

with $\Gamma_1 = k$.

For the first term in (7), we have the following. With the appropriate definitions of q and L_α, and the tail probability distribution of a geometric random variable, it is again easy to check (as in the proof of Lemma 1) that for any $\tau \in \mathbf{\Delta}_\alpha$, we have

$$\exp((1-\epsilon^2)qL_\alpha)P_\infty(\{k \le \tau < k + (1-\epsilon)L_\alpha\}) \to 0 \quad \text{as } \alpha \to 0$$

for any $\epsilon \in (0,1)$ and all $k \ge 1$. For the second term in (7), we need a condition analogous to [28, equation (3.2)]:

$$P_k\left(\left\{\frac{1}{M}\max_{0\le n<M} \mathcal{Z}_{k+n}^k \ge (1+\epsilon)q\right\}\right) \overset{M\to\infty}{\to} 0 \quad \text{for all} \quad \epsilon > 0 \quad \text{and} \quad k \ge 1.$$

This is trivial since the following is true:

$$\frac{\mathcal{Z}_{k+n}^k}{n} \overset{\text{a.s.}}{\to} LD(f_1, f_0) \quad \text{as} \quad n \to \infty \tag{8}$$

for all $k \in [1, \infty)$.

The above condition follows from the following series of steps. First, note that the strong law of large numbers for i.i.d. random variables implies that

$$\frac{\mathcal{Z}_{k+n}^k}{n} + \frac{1}{n}\sum_{\ell=2}^{L} \underbrace{\sum_{i=\Gamma_1}^{\Gamma_\ell - 1} \log\left(\frac{f_1(Z_{i,\ell})}{f_0(Z_{i,\ell})}\right)}_{z_\ell} \overset{\text{a.s.}}{\to} LD(f_1, f_0) = q \quad \text{as } n \to \infty.$$

Then, it can be easily checked that

$$E[z_\ell] = D(f_1, f_0)\sum_{j=2}^{\ell} \frac{(1-\rho_{j-1,j})}{\rho_{j-1,j}}.$$

Since $\min_\ell \rho_{\ell-1,\ell} > 0$ from the statement of the proposition, we have $E[z_\ell] \in (0, \infty)$ for all $\ell = 2, \ldots, L$, and hence, (8) holds. Applying the condition in (8) with $M = (1-\epsilon)L_\alpha$ as $\alpha \to 0$, we have the equivalent of [28, Lemma 1].

The proposition follows by application of an equivalent version of [28, Theorem 1, eq. (3.14)] which follows exactly as in [28]. ∎

D. Average Detection Delay: General Case ($L \ge 3$)

We now consider the general case where $L \ge 3$. The main statement here is as follows.

Proposition 5 ($L \ge 3$): If $D(f_1, f_0)$ is such that (6) is satisfied, we have

$$\lim_{A\to\infty} \frac{E[\nu_A]}{A} \le \frac{1}{LD(f_1, f_0) + |\log(1-\rho)|}.$$ ∎

As in Section VI, we will work towards the proof of this statement. For this, the following generalizations of Proposition 6 and Lemma 1 are necessary.

Proposition 7: We have

$$q_{k,\ell} = \alpha_{k,\ell} \cdot \prod_{j=1}^{\ell-1} \underbrace{\prod_{m=1}^{k} L_{m,j}}_{C_{k,j}} \cdot \underbrace{\prod_{m=0}^{k-2} (1+\zeta_{m,\ell})}_{J_{k,\ell}}, \quad \ell = 2, \ldots, L+1$$

where

$$\alpha_{k,2} = \left(\frac{1-\rho_{1,2}}{1-\rho}\right)^k \cdot \left(1 + \frac{1-\rho_{1,2}}{1-\rho}\right)$$

$$\alpha_{k,\ell} = \left(\frac{1-\rho_{\ell-1,\ell}}{1-\rho}\right)^k \cdot \prod_{j=1}^{\ell-2} \rho_{j,j+1} \cdot \left(\sum_{j=0}^{\ell-1} \frac{1-\rho_{j,j+1}}{1-\rho}\right), \quad \ell \ge 3$$

$$\zeta_{m,\ell} = \frac{(1-\rho_{\ell-1,\ell})^{-1}}{\prod_{j=1}^{\ell-1} L_{m+1,j}} \cdot \frac{\sum_{j=1}^{\ell-1} q_{m,j} w_j^\ell \mathcal{C}_{m+1,j,\ell}}{\sum_{j=1}^{\ell} q_{m,j} w_j^\ell}$$

$$\mathcal{B}_{m,n,\ell} = \sum_{p=n-1}^{\ell-1} (1-\rho_{p,p+1}) \cdot \prod_{j=1}^{p} L_{m,j}, \quad n = 1, \ldots, \ell$$

$$\mathcal{C}_{m,n,\ell} = \mathcal{B}_{m,n,\ell} - (1-\rho_{\ell-1,\ell}) \cdot \prod_{j=1}^{\ell-1} L_{m,j}, \quad n = 1, \ldots, \ell.$$

Proof: The proof is provided in Appendix E for the sake of completeness. Also, see Appendix E for how this proposition can be reduced to the case of [29]. ∎

Lemma 4: The following upper bound for $\zeta_{m,\ell}$ is obvious when $\max_\ell \rho_{\ell-1,\ell} < 1$:

$$\begin{aligned} \zeta_{m,\ell} &\le \frac{\mathcal{B}_{m+1,1,\ell}}{(1-\rho_{\ell-1,\ell}) \cdot \prod_{j=1}^{\ell-1} L_{m+1,j}} \\ &= \frac{\sum_{p=0}^{\ell-2}(1-\rho_{p,p+1}) \prod_{j=1}^{p} L_{m+1,j}}{(1-\rho_{\ell-1,\ell}) \cdot \prod_{j=1}^{\ell-1} L_{m+1,j}}. \end{aligned}$$ ∎

From Proposition 7, ν_A can be conveniently rewritten as

$$\nu_A = \inf_k \left\{\log\left(\sum_{\ell=2}^{L+1} \alpha_{k,\ell} \cdot C_{k,1} \cdots C_{k,\ell-1} \cdot J_{k,\ell}\right) > A\right\}.$$

Unlike the setting in Section VI, the structure of ν_A (as of now) is not amenable to studying ADD (in further detail). This is because it has the form of log of sum of random variables (see [34] for similar difficulties in the multihypothesis testing problem). We alleviate this difficulty by rewriting the test statistic in terms of quantities whose asymptotics can be easily studied.

Proposition 8: We have the following expansion for the test statistic:

$$\begin{aligned}&\log\left(\sum_{\ell=2}^{L+1}\alpha_{k,\ell}\cdot C_{k,1}\cdots C_{k,\ell-1}\cdot J_{k,\ell}\right)\\&=\log(\alpha_{k,2}\cdot C_{k,1}\cdot J_{k,2})\\&\quad+\sum_{\ell=2}^{L}\log\left(1+\frac{\eta_{k,\ell}\cdot\alpha_{k,\ell+1}\cdot C_{k,\ell}\cdot J_{k,\ell+1}}{\alpha_{k,\ell}\cdot J_{k,\ell}}\right)\\&=\log\left(\left(\frac{1-\rho_{1,2}}{1-\rho}\right)^k\cdot\frac{2-\rho-\rho_{1,2}}{1-\rho}\cdot C_{k,1}\cdot J_{k,2}\right)\\&\quad+\sum_{\ell=2}^{L}\log\left(1+\eta_{k,\ell}\cdot\beta_{k,\ell}\cdot C_{k,\ell}\cdot\frac{J_{k,\ell+1}}{J_{k,\ell}}\right)\end{aligned}$$

where

$$\begin{aligned}\beta_{k,\ell}&=\frac{\alpha_{k,\ell+1}}{\alpha_{k,\ell}}=\rho_{\ell-1,\ell}\cdot\left(1+\frac{1-\rho_{\ell,\ell+1}}{\sum_{m=0}^{\ell-1}1-\rho_{m,m+1}}\right)\\&\quad\cdot\left(\frac{1-\rho_{\ell,\ell+1}}{1-\rho_{\ell-1,\ell}}\right)^k,\qquad\ell=2,\dots,L\\\eta_{k,\ell+1}&=\frac{\eta_{k,\ell}\cdot\beta_{k,\ell}\cdot C_{k,\ell}\cdot\frac{J_{k,\ell+1}}{J_{k,\ell}}}{1+\eta_{k,\ell}\cdot\beta_{k,\ell}\cdot C_{k,\ell}\cdot\frac{J_{k,\ell+1}}{J_{k,\ell}}},\qquad\ell=2,\dots,L-1\end{aligned}$$

with $\eta_{k,2}=1$.

Proof: The proof is straightforward by using the induction principle. ■

The following proposition establishes the general asymptotic trend of ν_A.

Proposition 9: The test ν_A is such that $\nu_A\to\infty$ a.s. as $A\to\infty$.

Proof: See Appendix E. ■

As we try to understand ν_A further, it is important to note that the behavior of the decision statistic of ν_A is determined (only) by the trends of

$$x_{k,\ell}\triangleq\beta_{k,\ell}\cdot C_{k,\ell}\cdot\frac{J_{k,\ell+1}}{J_{k,\ell}},\qquad\ell=2,\dots,L.$$

This is so because the asymptotics of $\{\eta_{k,\ell}\}$ are also primarily determined by the trends of $\{x_{k,\ell}\}$. We now develop the generalized version of the heuristic in Section VI for the upper bound of ADD. Consider the case where $L=4$. The second piece in the description of the test statistic (in Proposition 8) can be written as

$$\mathcal{L}\triangleq\log(1+\eta_{k,2}x_{k,2})+\log(1+\eta_{k,3}x_{k,3})+\log(1+\eta_{k,4}x_{k,4})$$

where the evolution of $\eta_{k,\ell}$ and $x_{k,\ell}$, $\ell=2,3,4$ is described in Proposition 8. In the regime where $k\to\infty$, note that if $x_{k,2}\to\infty$ (with high probability), then $\eta_{k,3}\to1$. On the other hand, if $x_{k,2}\to0$ (with high probability), then $\eta_{k,3}\to x_{k,2}$. Thus, we can identify (and partition) eight cases as follows:

$$\begin{aligned}&\text{Case 1}: x_{k,2}\to0,\ x_{k,2}x_{k,3}\to0,\ x_{k,2}x_{k,3}x_{k,4}\to0\\&\quad\Longrightarrow\frac{\eta_{k,3}}{x_{k,2}}\to1,\ \frac{\eta_{k,4}}{x_{k,2}x_{k,3}}\to1\\&\quad\Longrightarrow\mathcal{L}\to0\\&\text{Case 2}: x_{k,2}\to0,\ x_{k,2}x_{k,3}\to0,\ x_{k,2}x_{k,3}x_{k,4}\to\infty\\&\quad\Longrightarrow\frac{\eta_{k,3}}{x_{k,2}}\to1,\quad\frac{\eta_{k,4}}{x_{k,2}x_{k,3}}\to1\\&\quad\Longrightarrow\frac{\mathcal{L}}{\log(x_{k,2}x_{k,3}x_{k,4})}\to1\\&\text{Case 3}: x_{k,2}\to0,\ x_{k,2}x_{k,3}\to\infty,\ x_{k,4}\to0\\&\quad\Longrightarrow\eta_{k,3}\to x_{k,2},\quad\eta_{k,4}\to1\\&\quad\Longrightarrow\frac{\mathcal{L}}{\log(x_{k,2}x_{k,3})}\to1\\&\text{Case 4}: x_{k,2}\to0,\ x_{k,2}x_{k,3}\to\infty,\ x_{k,4}\to\infty\\&\quad\Longrightarrow\eta_{k,3}\to x_{k,2},\quad\eta_{k,4}\to1\\&\quad\Longrightarrow\frac{\mathcal{L}}{\log(x_{k,2}x_{k,3}x_{k,4})}\to1\\&\text{Case 5}: x_{k,2}\to\infty,\ x_{k,3}\to0,\ x_{k,3}x_{k,4}\to0\\&\quad\Longrightarrow\eta_{k,3}\to1,\quad\eta_{k,4}\to x_{k,3}\\&\quad\Longrightarrow\frac{\mathcal{L}}{\log(x_{k,2})}\to1\\&\text{Case 6}: x_{k,2}\to\infty,\ x_{k,3}\to0,\quad x_{k,3}x_{k,4}\to\infty\\&\quad\Longrightarrow\eta_{k,3}\to1,\quad\eta_{k,4}\to x_{k,3}\\&\quad\Longrightarrow\frac{\mathcal{L}}{\log(x_{k,2}x_{k,3}x_{k,4})}\to1\\&\text{Case 7}: x_{k,2}\to\infty,\ x_{k,3}\to\infty,\ x_{k,4}\to0\\&\quad\Longrightarrow\eta_{k,3}\to1,\quad\eta_{k,4}\to1\\&\quad\Longrightarrow\frac{\mathcal{L}}{\log(x_{k,2}x_{k,3})}\to1\\&\text{Case 8}: x_{k,2}\to\infty,\ x_{k,3}\to\infty,\ x_{k,4}\to\infty\\&\quad\Longrightarrow\eta_{k,3}\to1,\quad\eta_{k,4}\to1\\&\quad\Longrightarrow\frac{\mathcal{L}}{\log(x_{k,2}x_{k,3}x_{k,4})}\to1.\end{aligned}$$

In all the eight cases, we have a universal description for $\mathcal{L}$ (as $k\to\infty$) that holds with high probability:

$$\begin{aligned}\mathcal{L}&\overset{k\to\infty}{\approx}\sum_{m=2}^{\ell^\star-1}\log(x_{k,m})\\\ell^\star&=\arg\min_{2\le\ell\le4}\\&\qquad\times\left\{\lim_{k\to\infty}\left(\prod_{m=\ell}^{j}x_{k,m}\right)\to0\text{ for all }\quad j\ge\ell\right\}.\end{aligned}$$

If $\ell^\star=2$, then the above summation is replaced by 0, and if there exists no $\ell\in\{2,3,4\}$ such that the above condition holds, then $\ell^\star$ is set to 5.

The following proposition provides a precise mathematical formulation of the above heuristic.

Proposition 10: Let the following limit be well defined and be denoted as $\gamma_{\ell,j}$

$$\gamma_{\ell,j}\triangleq\lim_{k\to\infty}\frac{1}{k}\sum_{m=1}^{k}\log\left(\frac{1+\zeta_{m,j+1}}{1+\zeta_{m,\ell}}\right).$$

Define $\ell^\star$ as

$$\ell^\star \triangleq \arg \min_{\ell: 2 \leq \ell \leq L} \{\Delta_{\ell,j} \leq 0 \text{ for all } j = \ell, \dots, L\},$$
$$\Delta_{\ell,j} = \log\left(\frac{1-\rho_{j,j+1}}{1-\rho_{\ell-1,\ell}}\right) + (j-\ell+1)D(f_1, f_0) + \gamma_{\ell,j}. \tag{9}$$

If there exists no element in the set for the $\arg\min$ operation in (9), we set $\ell^\star = L+1$. Then, as $A \to \infty$ (and, hence, $k = \nu_A \to \infty$ a.s. from Proposition 9), we have

$$\frac{1}{k}\sum_{\ell=2}^{L} \log(1+\eta_{k,\ell}x_{k,\ell}) - \frac{1}{k}\sum_{\ell=2}^{\ell^\star-1}\log(x_{k,\ell}) \to 0 \quad \text{a.s.} \tag{10}$$

If $\ell^\star = 2$, then the second term in the above expression is set to 0.

Proof: See Appendix E. ∎

Following Propositions 9 and 10, as $A \to \infty$, ν_A can be restated as

$$\nu_A \to \inf_k \left\{ \sum_{m=1}^{k} \left(\log\left(\frac{1-\rho_{1,2}}{1-\rho}\right) + \log(L_{m,1}) + \log(1+\zeta_{m,2}) + \frac{1}{k}\sum_{\ell=2}^{\ell^\star-1}\log(x_{k,\ell}) \right) > A \right\}$$
$$= \inf_k \left\{ \sum_{m=1}^{k} y_m > A \right\}$$
$$y_m = \log\left(\frac{1-\rho_{\ell^\star-1,\ell^\star}}{1-\rho}\right) + \sum_{j=1}^{\ell^\star-1}\log(L_{m,j}) + \log(1+\zeta_{m,\ell^\star}) \tag{11}$$

with $\ell^\star$ defined in (9).

Observe that if the condition in Proposition 10 is satisfied, the first $\ell^\star - 1$ sensors contribute to the slope of ADD and the rest of the sensors $\ell^\star, \dots, L$ (if any) do not contribute to the slope. While this characterization seems to be hard to utilize subsequently, it is important to understand the conditions under which $\ell^\star = L+1$. In this direction, (6) provides a simple condition such that the observations from all the L sensors contribute to the slope.

Proof of Proposition 5 ($L \geq 3$): First, using Lemma 4 note that, we can bound $\Delta_{\ell,j}$ as

$$\Delta_{\ell,j} \geq (j-\ell+1)D(f_1,f_0) + \log(1-\rho_{j,j+1}) - E\left[\log\left(\sum_{p=0}^{\ell-1}\frac{(1-\rho_{p,p+1})}{\prod_{i=p+1}^{\ell-1}L_{\bullet,i}}\right)\right].$$

Using Jensen's inequality and noting that

$$E_{f_1}\left[\frac{1}{\prod_{i=p+1}^{\ell-1}L_{\bullet,i}}\right] = 1$$

and (6) is sufficient to ensure that for all $\ell = 2, \dots, L$, there exists some $j \geq \ell$ such that $\Delta_{\ell,j} > 0$. It is important to realize that the above condition is necessary as well as sufficient for $\ell^\star = L+1$. Thus, under the assumption that (6) holds, invoking Proposition 9 as $A \to \infty$ (that is, letting $k = \nu_A \to \infty$ a.s. and using Proposition 10), ν_A can be written as

$$\nu_A \overset{A\to\infty}{\to} \inf_k \left\{ \sum_{m=1}^{k} \left(\sum_{\ell=1}^{L}\log(L_{m,\ell}) + \log\left(\frac{1}{1-\rho}\right) + \log(1+\zeta_{m,L+1}) \right) > A \right\}.$$

Note that since $\zeta_{m,L+1} \geq 0$, we have

$$\sum_{m=1}^{k}\left(\sum_{\ell=1}^{L}\log(L_{m,\ell}) + \log\left(\frac{1}{1-\rho}\right) + \log(1+\zeta_{m,L+1})\right) \geq \underbrace{\sum_{m=1}^{k}\left(\sum_{\ell=1}^{L}\log(L_{m,\ell}) + \log\left(\frac{1}{1-\rho}\right)\right)}_{L_k}$$

and hence, $\nu_A \leq \nu_{L,A}$ where

$$\nu_{L,A} \triangleq \inf_k\{L_k > A\}.$$

Thus, we have

$$\frac{E[\nu_A]}{A} \leq \frac{E[\nu_{L,A}]}{A} \overset{A\to\infty}{\to} \frac{1}{LD(f_1,f_0) + \log\left(\frac{1}{1-\rho}\right)}$$

where the convergence is again due to Lemma 3. ∎

E. Completing Proofs of Statements in Appendix D

Proof of Proposition 7: We start from (4) and apply the recursion relationship for $\{q_{k-1,\ell}\}$. Noting that $w_m^j w_j^\ell = w_m^\ell$ for all j such that $m \leq j \leq \ell$, we can collect the contributions of different terms and write $\sum_{j=1}^{\ell} q_{k-1,j}w_j^\ell$ as

$$\sum_{j=1}^{\ell} q_{k-1,j}w_j^\ell = \frac{1}{1-\rho}\cdot\sum_{j=1}^{\ell} q_{k-2,j}w_j^\ell \mathcal{B}_{k-1,j,\ell}$$

where $\{\mathcal{B}_{k-1,j,\ell}\}$ is as defined in the statement of the proposition. Thus, we have

$$\sum_{j=1}^{\ell} q_{k-1,j}w_j^\ell = \{1+\zeta_{k-2,\ell}\} \cdot \frac{(1-\rho_{\ell-1,\ell})\prod_{j=1}^{\ell-1}L_{k-1,j}}{1-\rho}\cdot\left(\sum_{j=1}^{\ell}q_{k-2,j}w_j^\ell\right)$$
$$\zeta_{k-2,\ell} = \frac{\sum_{j=1}^{\ell-1}q_{k-2,j}w_j^\ell\mathcal{C}_{k-1,j,\ell}}{(1-\rho_{\ell-1,\ell})\prod_{j=1}^{\ell-1}L_{k-1,j}}\cdot\frac{1}{\sum_{j=1}^{\ell}q_{k-2,j}w_j^\ell}.$$

Iterating the above equation, we have the conclusion in the statement of the proposition.

It is useful to reduce Proposition 7 to the case of [29] when $\rho_{\ell-1,\ell} = 1$ for all $\ell = 2, \ldots, L$. For this, note that $\alpha_{k,\ell}$ (and, hence, $q_{k,\ell}$) are identically zero for all $2 \leq \ell \leq L$. Thus, we have

$$q_{k,L+1} = \alpha_{k,L+1} \cdot \prod_{j=1}^{L} \prod_{m=1}^{k} L_{m,j} \cdot \prod_{m=0}^{k-2} (1 + \zeta_{m,L+1}).$$

We then have the following reductions:

$$\alpha_{k,L+1} = \frac{1}{(1-\rho)^k} \cdot \left(1 + \frac{1}{1-\rho}\right)$$
$$\zeta_{m,L+1} = \frac{1}{\prod_{j=1}^{L} L_{m+1,j}} \cdot \frac{\mathcal{B}_{m+1,1,L+1}}{1 + q_{m,L+1}}$$
$$\mathcal{B}_{m+1,1,L+1} = 1 - \rho$$

and hence

$$\begin{aligned} q_{k,L+1} &= \frac{\prod_{j=1}^{L} L_{k,j}}{1-\rho} \\ &\quad \cdot \prod_{m=0}^{k-1} \left\{ \frac{1}{1 + q_{m-1,L+1}} + \frac{\prod_{j=1}^{L} L_{m,j}}{1-\rho} \right\} \\ &= \frac{\prod_{j=1}^{L} L_{k,j}}{1-\rho} \cdot \frac{1}{\prod_{m=-1}^{k-2} (1 + q_{m,L+1})} \\ &\quad \cdot \prod_{m=0}^{k-1} \left\{ 1 + \frac{\prod_{j=1}^{L} L_{m,j} (1 + q_{m-1,L+1})}{1-\rho} \right\} \end{aligned}$$

with the initial condition that $q_{-1,L+1} = 0$ and $L_{0,j} = 1$ for all j. It is straightforward to establish via induction that the only way in which the above recursion can hold is if $q_{k,L+1}$ satisfies

$$q_{k,L+1} = \frac{\prod_{j=1}^{L} L_{k,j}}{1-\rho} \cdot (1 + q_{k-1,L+1})$$

which, as expected, is the same recursion as (5). ■

Proof of Proposition 9: First, note that if we can find $\{U_k\}$ such that for all k

$$\log \left(\sum_{\ell=2}^{L+1} \alpha_{k,\ell} \cdot C_{k,1} \cdots C_{k,\ell-1} \cdot J_{k,\ell} \right) \leq U_k$$

then $\nu_A \geq \nu_{U,A}$ where

$$\nu_{U,A} \triangleq \inf_k \{U_k > A\}.$$

We use Lemma 4 to obtain the following bound and the associated $\{U_k\}$:

$$\begin{aligned} &\sum_{\ell=2}^{L+1} \alpha_{k,\ell} \cdot C_{k,1} \cdots C_{k,\ell-1} \cdot J_{k,\ell} \\ &\quad \leq \sum_{\ell=2}^{L+1} \frac{(1-\rho_{\ell-1,\ell}) \cdot \prod_{j=1}^{\ell-1} L_{k,j} \cdot D_\ell}{1-\rho} \\ &\quad \cdot \prod_{m=1}^{k-1} \frac{\sum_{p=0}^{\ell-1} (1-\rho_{p,p+1}) \prod_{j=1}^{p} L_{m,j}}{1-\rho} \\ &\quad \leq \frac{1}{1-\rho} \cdot \left(\sum_{\ell=2}^{L+1} D_\ell \cdot \prod_{j=1}^{\ell-1} L_{k,j} \right) \\ &\quad \cdot \prod_{m=1}^{k-1} \frac{\sum_{p=0}^{L} (1-\rho_{p,p+1}) \prod_{j=1}^{p} L_{m,j}}{1-\rho} \\ &\quad \leq \frac{D}{1-\rho} \cdot \left(\sum_{p=1}^{L} \frac{1-\rho_{p,p+1}}{1-\rho} \cdot \prod_{j=1}^{p} L_{k,j} \right) \\ &\quad \cdot \prod_{m=1}^{k-1} \frac{\sum_{p=0}^{L} (1-\rho_{p,p+1}) \prod_{j=1}^{p} L_{m,j}}{1-\rho} \\ &\quad \leq \frac{D}{1-\rho} \cdot \prod_{m=1}^{k} \frac{\sum_{p=0}^{L} (1-\rho_{p,p+1}) \prod_{j=1}^{p} L_{m,j}}{1-\rho} \end{aligned}$$

where

$$D_\ell = \prod_{j=1}^{\ell-2} \rho_{j,j+1} \cdot \left(\sum_{j=0}^{\ell-1} \frac{1-\rho_{j,j+1}}{1-\rho} \right)$$

$D = 1 + \max_{\ell=1,\ldots,L} \frac{\ell}{1-\rho_{\ell,\ell+1}}$. With the above bound, we have

$$\begin{aligned} \nu_A \geq \inf_k \Bigg\{ &\sum_{m=1}^{k} \log \left(\sum_{p=0}^{L} (1-\rho_{p,p+1}) \prod_{j=1}^{p} L_{m,j} \right) \\ &- k \log(1-\rho) > A + \log \left(\frac{1-\rho}{D} \right) \Bigg\}. \end{aligned}$$

The conclusion follows by using Lemma 3 and noting that

$$E \left[\log \left(\frac{\sum_{p=0}^{L} (1-\rho_{p,p+1}) \prod_{j=1}^{p} L_{m,j}}{1-\rho} \right) \right] \in (0, \infty). \quad ■$$

Proof of Proposition 10: This proof is a formal write-up of the heuristic presented before the statement of Proposition 10. Following the definition of $\eta_{k,j}$ and the fact that $0 \leq \eta_{k,j} \leq 1$, we have

$$\eta_{k,j} x_{k,j} \leq \prod_{m=\ell^\star}^{j} x_{k,m}, \quad j \geq \ell^\star.$$

Suppose there exists an $\ell^\star \leq L$ as defined in (9), invoking Lemma 2 with the fact that $\Delta_{\ell^\star,j} \leq 0$ for all $j \geq \ell^\star$, we have

$$\frac{1}{k} \sum_{\ell=\ell^\star}^{L} \log(1 + \eta_{k,\ell} x_{k,\ell}) \stackrel{k\to\infty}{\to} 0 \text{ a.s. and in mean.}$$

Thus, we have

$$\begin{aligned} &\frac{1}{k} \sum_{\ell=2}^{L} \log(1 + \eta_{k,\ell} x_{k,\ell}) \\ &\quad - \frac{1}{k} \sum_{\ell=2}^{\ell^\star-1} \log(1 + \eta_{k,\ell} x_{k,\ell}) \\ &\quad \stackrel{k\to\infty}{\to} 0 \quad \text{a.s. and in mean.} \end{aligned}$$

The main contribution to (10) is now established via induction. Since $\eta_{k,2} = 1$, we can expand the sum as (modulo the a.s. and in mean convergence parts)

$$\frac{1}{k}\sum_{\ell=2}^{\ell^\star-1}\log(1+\eta_{k,\ell}x_{k,\ell}) - \frac{1}{k}\log\left(1+\sum_{\ell=2}^{\ell^\star-1}\prod_{m=2}^{\ell}x_{k,m}\right) \overset{k\to\infty}{\to} 0.$$

If $\ell^\star = 2$, it is clear that the proposition is true. If $3 \leq \ell^\star \leq L+1$, since $2 < \ell^\star$, by the definition of $\ell^\star$, there exists (a smallest choice) $j_2 \geq 2$ such that

$$\prod_{m=2}^{j_2} x_{k,m} \overset{k\to\infty}{\to} \infty$$

with

$$\prod_{m=2}^{p} x_{k,m} \overset{k\to\infty}{\to} 0 \text{ or } \mathcal{O}(1) \text{ for all } 2 \leq p \leq j_2 - 1$$

provided the set $[2,\ldots,j_2-1]$ is not empty. There are two possibilities: $j_2 = \ell^\star - 1$ or $j_2 \leq \ell^\star - 2$. (Note that $j_2 \geq \ell^\star$ results in a contradiction since it will imply $\prod_{m=\ell^\star}^{j_2} x_{k,m} \to \infty$, but we know this is not true from the definition of $\ell^\star$). In the first case, we are done upon invoking Lemma 2. In the second case, iterating by replacing 2 with j_2+1 (as many times as necessary) and finally invoking Lemma 2 and noting the main contribution of the sum in (10), we arrive at the conclusion of the proposition. ■

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their detailed and careful review that helped in improving the presentation of this paper.

REFERENCES

[1] A. G. Tartakovsky, B. Rozovskii, R. Blazek, and H. Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3372–3382, Sept. 2006.

[2] J. S. Baras, A. Cardenas, and V. Ramezani, "Distributed change detection for worms, DDOS and other network attacks," in *Proc. Amer. Cont. Conf. (ACC)*, Boston,, MA, 2004, pp. 1008–1013.

[3] K. Mechitov, W. Kim, G. Agha, and T. Nagayama, "High-frequency distributed sensing for structure monitoring," *Trans. Soc. Instr. and Cont. Eng.*, vol. E-S-1, no. 1, pp. 109–114, 2006.

[4] National Science Foundation Workshop on Monitoring and Controlling the Nation's Critical Infrastructures, Final Rep., 2006 [Online]. Available: http://www.ece.wisc.edu/~nowak/ci

[5] L. A. Farwell and E. Donchin, "Talking off the top of your head: A mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophys.*, vol. 70, pp. 510–523, 1988.

[6] R. Ratnam, J. B. M. Goense, and M. E. Nelson, "Change-point detection in neuronal spike train activity," *Neurocomputing*, vol. 52–54, pp. 849–855, 2003.

[7] H. Li, C. Li, and H. Dai, "Quickest spectrum sensing in cognitive radio," in *Proc. 42nd IEEE Conf. Inf. Syst. Sci. (CISS)*, Princeton, NJ, Mar. 2008, pp. 203–208.

[8] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.

[9] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, vol. 13, no. 1, pp. 206–227, Mar. 1985.

[10] A. N. Shiryaev, "The detection of spontaneous effects," *Sov. Math. Dokl.*, vol. 2, pp. 740–743, 1961.

[11] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Prob. Appl.*, vol. 8, no. 1, pp. 22–46, Jan. 1963.

[12] A. N. Shiryaev, *Optimal Stopping Rules*. New York: Springer-Verlag, 1978.

[13] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag, 1985.

[14] A. G. Tartakovsky, *Sequential Methods in the Theory of Information Systems (in Russian)*. Moscow: Radio i Svyaz, 1991.

[15] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[16] T. L. Lai, "Sequential analysis: Some classical problems and new challenges (with discussion)," *Stat. Sinica*, vol. 11, pp. 303–408, 2001.

[17] A. G. Tartakovsky and V. V. Veeravalli, , N. Mukhopadhyay, S. Datta, and S. Chattopadhyay, Eds., "Change-point detection in multichannel and distributed systems with applications," in *Applications of Sequential Methodologies*. New York: Marcel Dekker, 2004, pp. 331–363.

[18] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge , U.K.: Cambridge University Press, 2008.

[19] T. L. Lai, "Sequential change-point detection in quality control and dynamical systems," *J. Roy. Statist. Soc. Ser. B (Meth)*, vol. 57, no. 4, pp. 613–658, 1995.

[20] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[21] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.

[22] M. Beibel, "Sequential detection of signals with known shape and unknown magnitude," *Stat. Sinica*, vol. 10, pp. 715–729, 2000.

[23] C. D. Fuh, "SPRT and CUSUM in hidden Markov models," *Ann. Statist.*, vol. 31, pp. 942–977, 2003.

[24] G. Peskir and A. N. Shiryaev, "Solving the Poisson disorder problem," in *Advances in Finance and Stochastics*. Berlin: Springer, 2002, pp. 295–312.

[25] A. G. Tartakovsky, *Extended Asymptotic Optimality of Certain Change-Point Detection Procedures*, 2003, submitted for publication.

[26] B. Yakir, "A note on optimal detection of a change in distribution," *Ann. Statist.*, vol. 25, pp. 2117–2126, 1997.

[27] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2917–2929, Jul. 1998.

[28] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *SIAM Theory Prob. Appl.*, vol. 49, no. 3, pp. 458–497, 2005.

[29] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[30] A. G. Tartakovsky and V. V. Veeravalli, "An efficient sequential procedure for detecting changes in multichannel and distributed systems," in *Proc. 5th IEEE Int. Conf. Inf. Fusion*, Annapolis, MD, Jul. 2002, vol. 1, pp. 41–48.

[31] A. G. Tartakovsky and V. V. Veeravalli, "Asymptotically optimal quickest change detection in distributed sensor systems," *Sequential Analysis*, vol. 27, no. 4, pp. 441–475, Nov. 2008.

[32] O. Hadjiliadis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3346–3359, Jul. 2009.

[33] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[34] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1994–2007, Nov. 1994.

[35] R. A. Durrett, *Probability: Theory and Examples*, 2nd ed. New York: Duxbury, 1995.

[36] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*. Philadelphia, PA: SIAM, 1982.

[37] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 433–442, Mar. 1993.

Vasanthan Raghavan (S'01–M'06) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 2001, the M.S. and Ph.D. degrees in electrical and computer engineering in 2004 and 2006, respectively, and the M.A. degree in mathematics in 2005, all from the University of Wisconsin-Madison.

He is currently a Research Fellow with The University of Melbourne, Parkville, Australia. He was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, from 2006 to 2009. His research interests span multiantenna communication techniques, quickest change detection, information theory, multihop networking, robust control, and random matrix theory.

Venugopal V. Veeravalli (S'86–M'92–SM'98–F'06) received the B.Tech. degree in 1985 from the Indian Institute of Technology, Bombay, (Silver Medal Honors), the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, all in electrical engineering.

He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently a Professor with the Department of Electrical and Computer Engineering, and a Research Professor in the Coordinated Science Laboratory. He served as a Program Director for communications research at the U.S. National Science Foundation, Arlington, VA, from 2003 to 2005. He has previously held academic positions with Harvard University, Rice University, and Cornell University. His research interests include distributed sensor systems and networks, wireless communications, detection and estimation theory, and information theory.

Dr. Veeravalli was on the Board of Governors of the IEEE Information Theory Society from 2004 to 2007. He was an Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2000 to 2003, and an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 1999 to 2000. Among the awards he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award, the National Science Foundation CAREER Award, and the Presidential Early Career Award for Scientists and Engineers (PECASE).