Quickest Change Detection in Anonymous Heterogeneous Sensor Networks

Zhongchang Sun   Shaofeng Zou   Ruizhi Zhang   Qunwei Li

Abstract—The problem of quickest change detection (QCD) in anonymous heterogeneous sensor networks is studied. There are n heterogeneous sensors and a fusion center. The sensors are clustered into K groups, and different groups follow different data-generating distributions. At some unknown time, an event occurs in the network and changes the data-generating distribution of the sensors. The goal is to detect the change as quickly as possible, subject to false alarm constraints. The anonymous setting is studied, where at each time step, the fusion center receives n unordered samples, and the fusion center does not know which sensor each sample comes from, and thus does not know its exact distribution. A simple optimality proof is first derived for the mixture likelihood ratio test, which was constructed and proved to be optimal for the non-sequential anonymous setting in [2]. For the QCD problem, a mixture CuSum algorithm is further constructed, and is further shown to be optimal under Lorden's criterion. For large networks, a computationally efficient test is proposed and a novel theoretical characterization of its false alarm rate is developed. Numerical results are provided to validate the theoretical results.

Index Terms—Hypothesis testing, mixture CuSum, sequential change detection, computationally efficient, optimal.

I. Introduction

IN quickest change detection (QCD) problem [3]–[9], a decision maker collects samples sequentially from a stochastic environment. At some unknown time, an event occurs and causes a change in the data-generating distribution. The goal of the decision maker is to detect the change as quickly as possible subject to a constraint on the false alarm. The QCD problem in sensor networks has been widely studied in the literature [10]–[23]. In these studies, it is usually assumed that the fusion center knows which sensor that each sample comes from, and thus the statistical property of the sample is known. However, in a wide range of modern practical applications, the nodes are anonymous and heterogeneous. In this case, only unordered and anonymous samples are available to the fusion center, and the fusion center doesn't know what data generating distribution that each sample follows.

In this paper, we investigate the QCD problem using anonymized samples. We consider a general scenario with heterogeneous sensors, where the sensors can be clustered into K groups with different data generating distributions, and the fusion center does not know which sensor each sample comes from. At some unknown time, an event occurs in the network, and changes the data-generating distribution of the nodes. The goal is to detect the change as quickly as possible subject to false alarm constraints using anonymized samples (see Fig. 1).
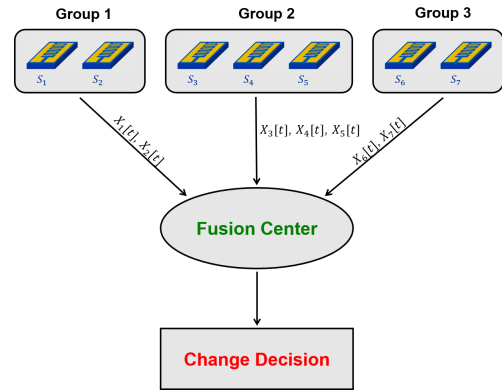

Fig. 1. QCD in anonymous heterogeneous sensor network. Sensors in different groups have different distributions. At each time t, the fusion center collects unordered samples $X_1[t], \cdots, X_7[t]$ from sensors $S_1, \cdots, S_7$.

Statistical inference for anonymous and heterogeneous sensor networks finds a wide range of practical applications. For example, in large-scale Internet of things (IoT) networks [24]–[27], devices are commonly small and low-cast sensing devices powered by battery, and are usually deployed in a massive scale. In such large-scale networks, the communication cost of identifying individual sensors increases drastically as the number of sensors grows [2], which is not affordable for battery powered small IoT sensing devices that are expected to survive for years without battery change. Moreover, sensors in IoT networks are usually heterogeneous for various industrial and consumer applications, e.g., pressure sensor, light sensor, temperature sensor, humidity sensor, seismic sensor and electrochemical sensor. The same type of sensors deployed in different regions also exhibits heterogeneity in their data generating distributions, e.g., electrochemical sensors that are near to or far away from the air pollution source and climate sensors on different sides of the same mountain. The second example is crowdsourcing, which is an evolving distributed problem-solving and business production model [28]–[30]. Crowdsourcing aims to collect data, ideas, micro-tasks from a large and relatively open group of people. With human participants, anonymity is necessarily needed to protect pri-

Zhongchang Sun and Shaofeng Zou are with the Department of Electrical Engineering, University at Buffalo, Buffalo, NY 14228 USA (e-mail: zhongcha@buffalo.edu, szou3@buffalo.edu). Ruizhi Zhang is with the Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68588 USA (e-mail: rzhang35@unl.edu). Qunwei Li is with the Ant Group, Hangzhou, China (e-mail: qunwei.qw@antfin.com)

vacy [31]–[38]. Based on their skill level and background, e.g., education, country, and language, participants can be divided into groups that are heterogeneous. QCD finds a wide range of applications in these networks, e.g., environmental change (air/water quality) monitoring, fake news detection in social networks, pandemic outbreak detection and seismic wave detection. In these applications, a change in the data-generating distributions occurs due to an abrupt event which is of interest to be detected quickly.

A. Related Works

The problem studied in this paper is closely related to the problem of QCD under the multiple-channel setup [10]–[15], [18]–[23], where samples are collected from multiple sensors sequentially, and the goal is to detect a change in the data-generating distribution of some unknown subset of the sensors or all the sensors. These works assume that the sensors are non-anonymous, i.e., it is known that which sensor that each sample comes from. In the non-anonymous setting, algorithms can be designed by combining the CuSum statistics each calculated for one sensor. These algorithms inherit the nice property of the CuSum algorithm [39] which can be updated in an online and recursive fashion, and thus is computationally efficient. It was shown in these studies that such a type of algorithms are asymptotically optimal for various scenarios. In this paper, we are interested in the anonymous setting, where at each time step the fusion center receives n arbitrarily permuted (unordered) observations, and the permutations at different time steps may be different due to anonymity. Then, the fusion center does not know which samples over time come from one particular sensor. Therefore, existing approaches based on the idea of combining local CuSum statistics are not applicable any more since the fusion center is not able to compute one CuSum statistic for each node.

In anonymous networks, the fusion center does not know the exact distribution of each sample due to the uncertainty caused by the anonymity [2]. The group label that assigns the samples to different types of sensors is an unknown parameter of the distribution. Therefore, the QCD problem in anonymous networks can be viewed as a composite QCD problem with unknown pre- and post-change distributions [40]–[46], The main difference lies in that the unknown parameter in our problem is changing with time, i.e., the group label may not be the same at different time steps, and thus the samples are not identically distributed in the pre- or post-change regime. As will be shown in our numerical results, a generalized likelihood ratio based test does not work well here. Furthermore, we do not assume that the distributions belong to any parameterized family of distributions, e.g., exponential family.

The problem of quickest detection of a moving anomaly was studied in [47], [48], where an unknown sensor is affected by an anomaly with an unknown trajectory that emerges in the network at some unknown time. In [47], the statistical behavior of the samples is modeled using a hidden Markov model [49], and the trajectory is modeled as a deterministic and unknown one in [48]. Our work is different from the one in [47] since we do not put any assumption on the prior of

group label (trajectory of the anomaly in [47]). The study in [48] is related to ours in that the samples can be equivalently viewed as being collected from anonymous sensors since the node affected by the anomaly is unknown.

The offline hypothesis testing problem in the anonymous setting was investigated in [2], where one sample is collected from each sensor. A mixture likelihood ratio test (MLRT) was developed, and was further shown to be optimal under the Neyman-Pearson setting. Here, we consider the QCD problem under the anonymous setting with sequential samples and time-varying group labels. We are interested in the tradeoff between the false alarm rate and average detection delay, which requires construction of sequential tests and more involved optimality analysis.

In Table I, we summarize the difference between our paper and other related works. We note that the fusion center may be able to recover the group identity if it performs, e.g., hypothesis testing, and the error probability depends on the distance between the distributions of different groups. With unordered samples, perfect anonymity can only be guaranteed if distributions among different groups are exactly the same. Designing optimal mechanisms to achieve perfect anonymity is not the focus of this paper, and might be of independent interest. In this paper, we focus on the design of optimal quickest detection algorithms for the scenario with unordered samples.

B. Main Contributions

We first revisit the non-sequential hypothesis testing problem with anonymous heterogeneous sensors. We provide a simple proof for the optimality of the MLRT [2]. The basic idea is to construct a binary composite hypothesis testing problem with uniform priors on all possible group labels, and to show that the optimal test for the case with a uniform Bayesian prior is also optimal under the minimax setting.

For the QCD problem in anonymous networks, we design a mixture CuSum algorithm, and prove that the mixture CuSum algorithm is exactly optimal under Lorden's criterion [50]. To show its exact optimality, we build a novel connection among several simple QCD problems and the QCD problem under the anonymous setting. The major challenge in our analysis is due to that we are optimizing the worst-case performance over all possible change-point, group labels and pre-change observations.

The computational complexity of the mixture CuSum algorithm at each time step increases almost exponentially in the number of nodes, and thus is not efficient when the network is large. We then propose a computationally efficient test based on the asymptotic behavior of the mixture CuSum test statistic when the network is large. The basic idea is to approximate the mixture CuSum statistic by a convex optimization problem with linear constraints, the computational complexity of which is independent of the number of sensors. We provide a comprehensive discussion of its performance. We also derive a lower bound on its worst-case average run length to false alarm, so that a threshold can be chosen analytically for false alarm control in practice.

Existing studies	Our paper
[10]–[15], [18]–[23]: QCD problem in sensor networks, sensors are non-anonymous	Sensors are anonymous, i.e., fusion center does not know which sensor each sample comes from
[40]–[46]:Composite QCD problem with unknown pre- and post-change distributions	Unknown parameters change with time, and generalized likelihood ratio based test is not applicable
[47], [48]: quickest detection of moving anomaly	Samples are unordered at each time step
[2]: offline binary hypothesis testing problem in anonymous networks	Samples are sequentially collected

TABLE I
COMPARISON OF RELATED WORKS

We provide numerical results to demonstrate the performance of our proposed algorithms. We compare our algorithms with two other heuristic algorithms based on the Bayesian approach and the generalized likelihood ratio approach, and show that our mixture CuSum has the best performance, and our computationally efficient test also performs better than the other two tests. We also compare their computational complexity, and show that our the computationally efficient algorithm is much more efficient for large networks.

C. Paper Organization

In Section II, we present the problem formulation. In Section III, we revisit the binary hypothesis testing problem in the anonymous setting, and derive a simple optimality proof for the MLRT. In Section IV, we develop the mixture CuSum algorithm, and prove its exact optimality under Lorden's criterion. We further develop a computationally efficient test and characterize its performance theoretically. In Section V, we provide numerical results to validate our theoretical assertions. In Section VI, we present some concluding remarks.

II. PROBLEM FORMULATION

Consider a network consisting of n sensors. The sensors are heterogeneous and can be divided into K groups. Each group k has n_k sensors, $1 \leq k \leq K$. The distributions of the observations in group k are $p_{\theta,k}$, $\theta \in \{0,1\}$. Let $\boldsymbol{P}_\theta = [p_{\theta,1} \cdots p_{\theta,K}]^T$. We assume that $\boldsymbol{\alpha}^T \boldsymbol{P}_0 \neq \boldsymbol{\alpha}^T \boldsymbol{P}_1$. The centralized setting is considered, where there is a fusion center. The sensors are anonymous, i.e., the fusion center does not know which group of sensors that each observation comes from. The fusion center only knows the distributions $p_{\theta,k}$, $\theta \in \{0,1\}$ and the number of sensors n_k in each group k.

A. Binary Composite Hypothesis Testing

We first revisit the binary hypothesis testing problem in [2]. The goal is to distinguish between the two hypotheses: $\mathcal{H}_0 : \theta = 0$ and $\mathcal{H}_1 : \theta = 1$.

Denote by $X^n = \{X_1, \ldots, X_n\}$ the n collected samples. Denote by $\sigma(i) \in \{1, \ldots, K\}$ the label of the group that X_i comes from, i.e., $X_i \sim p_{\theta,\sigma(i)}$. Due to the anonymity, $\sigma(i), i = 1, \ldots, n$, are *unknown* to the fusion center. There are $\binom{n}{n_1, \ldots, n_K}$ possible $\sigma : \{1, \ldots, n\} \to \{1, \ldots, K\}$ satisfying $|\{i : \sigma(i) = k\}| = n_k, \forall k = 1, \ldots, K$. We denote the collection of all such labels by $\mathcal{S}_{n,\lambda}$, where $\lambda = \{n_1, \ldots, n_K\}$.

Given σ, the n collected samples are assumed to be independent. The problem is a composite hypothesis testing problem, where σ is the unknown parameter for both $\theta = 0$ and 1:

$$\mathcal{H}_\theta : X^n \sim \mathbb{P}_{\theta,\sigma} \triangleq \prod_{i=1}^n p_{\theta,\sigma(i)}, \text{ for some } \sigma \in \mathcal{S}_{n,\lambda}. \quad (1)$$

The worst-case type-I and type-II error probabilities for a decision rule ϕ are defined as

$$P_F(\phi) \triangleq \max_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{E}_{0,\sigma}[\phi(X^n)], \quad (2)$$

$$P_M(\phi) \triangleq \max_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{E}_{1,\sigma}[1 - \phi(X^n)], \quad (3)$$

where $\mathbb{E}_{\theta,\sigma}$ denotes the expectation under $\mathbb{P}_{\theta,\sigma}$, for $\theta \in \{0,1\}$ and $\sigma \in \mathcal{S}_{n,\lambda}$. The Neyman-Pearson setting is studied, where the goal is to solve the following problem for any $\zeta \in [0,1]$:

$$\inf_{\phi:P_F(\phi) \leq \zeta} P_M(\phi). \quad (4)$$

B. Quickest Change Detection

In the QCD setting, anonymized samples are observed sequentially. At some unknown time ν, an event occurs in the network, and changes the data-generating distributions of the sensors. Specifically, denote the i-th sample at time t by $X_i^n[t]$ and all the observed samples at time t by $X^n[t]$. Before the change, i.e., $t < \nu$, $X^n[t] \sim \mathbb{P}_{0,\sigma_t}$, for some *unknown* $\sigma_t \in \mathcal{S}_{n,\lambda}$. After the change, i.e., $t \geq \nu$, $X^n[t] \sim \mathbb{P}_{1,\sigma_t}$, for some *unknown* $\sigma_t \in \mathcal{S}_{n,\lambda}$. We note that σ_t may change with time, i.e., σ_{t_1} may not be the same as σ_{t_2}, for $t_1 \neq t_2$. We assume that for any $t \geq 0$, given σ_t, the samples in $X^n[t]$ are independent. We further assume that $X^n[t_1]$ is independent from $X^n[t_2]$ for any $t_1 \neq t_2$.

The objective is to detect the change at time ν as quickly as possible subject to false alarm constraints. In this paper, we consider a deterministic unknown change point ν. We define the worst-case average detection delay (WADD) under Lorden's criterion [50] and worst-case average run length (WARL) for any stopping time τ as follows:

$$\text{WADD}(\tau) \triangleq \sup_{\nu \geq 1} \sup_{\Omega} \text{ess sup } \mathbb{E}_\Omega^\nu \left[(\tau - \nu)^+ | \mathbf{X}^n[1, \nu - 1] \right],$$

$$\text{WARL}(\tau) \triangleq \inf_{\Omega} \mathbb{E}_\Omega^\infty[\tau], \quad (5)$$

where $\Omega = \{\sigma_1, \sigma_2, ..., \sigma_\infty\}$, $\mathbb{E}_\Omega^\nu$ denotes the expectation when the change is at ν, and the observations at time t are labeled according to σ_t, and $\mathbf{X}^n[1, \nu - 1] = \{X^n[1], \ldots, X^n[\nu - 1]\}$.

The goal is to design a stopping rule that minimizes the WADD subject to a constraint on the WARL:

$$\inf_{\tau:\text{WARL}(\tau) \geq \gamma} \text{WADD}(\tau). \quad (6)$$

C. Notations

In this section, we list the notations used in this paper.

- n denotes the number of sensors, K denotes the number of groups and n_k denotes the number of sensor in group k.
- $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_K]^T$, where $\alpha_k = \lim_{n\to\infty} \frac{n_k}{n}$ denotes the asymptotic fraction of sensors of group k.
- $\sigma(i) \in \{1, \ldots, K\}$ denotes the label of the group that X_i comes from, i.e., $X_i \sim p_{\theta,\sigma(i)}$, and $\mathcal{S}_{n,\lambda}$ denotes the collection of all $\sigma(i)$, where $\lambda = \{n_1, \ldots, n_K\}$.
- $H(\boldsymbol{\alpha})$ denotes the entropy of $\boldsymbol{\alpha}$.
- Π_{X^n} denotes the empirical distribution of samples X^n, and $T(\Pi_{X^n})$ denotes the type class of Π_{X^n}.
- $\mathcal{P}_n$ denotes the set of types with denominator n.
- $D(P||Q)$ denotes the Kullback-Leibler (KL) divergence between P and Q.
- $f(x) \sim g(x)$ as $x \to x_0$ if $f(x) = g(x)(1 + o(1))$ as $x \to x_0$.

III. MLRT AND A SIMPLE OPTIMALITY PROOF

For the binary composite hypothesis testing problem in Section II-A, Chen and Huang constructed a mixture likelihood ratio test (MLRT), and showed that the MLRT is optimal under the Neyman-Pearson setting in (4) [2]. In this section, we first briefly review the optimality proof in [2], and then we present a simple version of the proof.

Define the mixture likelihood ratio $\ell(X^n)$ as follows:

$$\ell(X^n) = \frac{\sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n)}{\sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n)}. \tag{7}$$

Then the MLRT was defined in [2] as

$$\phi^*(X^n) = \begin{cases} 1, & \text{if } \ell(X^n) > \eta \\ \beta, & \text{if } \ell(X^n) = \eta \\ 0, & \text{if } \ell(X^n) < \eta, \end{cases} \tag{8}$$

where $\beta \in [0,1]$, η is the threshold, and they are chosen to meet the false alarm constraint.

Lemma 1. *[2, Thm. 3.1] The MLRT ϕ^* is optimal for* (4).

The key idea of the proof in [2] is to reduce the original composite hypothesis testing problem in Section II-A into a simple one through the ordering map $\Pi(X^n)$, and then apply the Neyman-Pearson lemma. The ordering map $\Pi(X^n)$ of X^n is defined as $\Pi(X^n) = (X_{i_1}, X_{i_2}, \ldots, X_{i_n})$, such that $X_{i_1} \geq X_{i_2} \geq \ldots \geq X_{i_n}$. In the proof, due to the introduction of the ordering map, a careful examination of the measurability needs to be conducted. The proof in [2] can be summarized by the following steps. 1) In the auxiliary space induced by the ordering mapping, the induced probability measure is independent of σ, and thus the corresponding problem in the auxiliary space is a simple hypothesis testing problem. 2) In the auxiliary space, applying the Neyman-Pearson lemma, the optimal test is obtained. 3) Any symmetric test in the original sample space is equivalent to a test in the auxiliary space in terms of type-I and type-II error probabilities, where a test ϕ is symmetric if $\phi(x^n) = \phi(\pi(x^n))$ for any x^n and any permutation π. 4) The optimal test in the auxiliary space is

the MLRT and is symmetric, which means that among all symmetric tests, the MLRT is optimal. 5) For any test ψ, one can always symmetrize it and construct a symmetric test ϕ, which is as good as ψ. 6) Then, the MLRT is optimal among all tests.

In the following, we present a simple proof for the optimality of the MLRT. Our proof does not need to use the ordering map, and is much simpler.

Proof. We consider a Bayesian setting with a uniform prior on σ under both hypotheses, and define the average type-I and type-II error probabilities for any test ϕ:

$$\widetilde{P}_F(\phi) \triangleq \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{E}_{0,\sigma}[\phi(X^n)], \tag{9}$$

$$\widetilde{P}_M(\phi) \triangleq \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{E}_{1,\sigma}[1 - \phi(X^n)]. \tag{10}$$

Then under the Bayesian setting, this problem reduces to the following simple binary hypothesis testing problem:

$$\mathcal{H}_0 : \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}, \tag{11}$$

$$\mathcal{H}_1 : \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}, \tag{12}$$

for which the optimal test (the same as the MLRT) is the likelihood ratio test between (11) and (12) [51].

It can be verified that for any permutation $\pi(X^n) = (X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$, $\phi^*(X^n) = \phi^*(\pi(X^n))$. For any π, let $\sigma' = \sigma \circ \pi$, where "$\circ$" denotes the composition of two functions, i.e., $f \circ g(x) = f(g(x))$. Then $\mathbb{E}_{\theta,\sigma}[\phi^*(\pi(X^n))] = \mathbb{E}_{\theta,\sigma\circ\pi}[\phi^*(X^n)] = \mathbb{E}_{\theta,\sigma'}[\phi^*(X^n)]$. For any $\sigma' \in \mathcal{S}_{n,\lambda}$, a π can be found so that $\sigma \circ \pi = \sigma'$. Thus, for any $\sigma, \sigma' \in \mathcal{S}_{n,\lambda}$ and $\theta = 0, 1$,

$$\mathbb{E}_{\theta,\sigma'}[\phi^*(X^n)] = \mathbb{E}_{\theta,\sigma}[\phi^*(X^n)]. \tag{13}$$

It then follows that

$$\begin{aligned} P_F(\phi^*) &= \max_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{E}_{0,\sigma}[\phi^*(X^n)] = \mathbb{E}_{0,\sigma}[\phi^*(X^n)] \\ &= \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{E}_{0,\sigma}[\phi^*(X^n)] = \widetilde{P}_F(\phi^*). \end{aligned} \tag{14}$$

Similarly, it can be shown that $P_M(\phi^*) = \widetilde{P}_M(\phi^*)$.

From (9) and (10), it follows that for any test ϕ,

$$\begin{aligned} \widetilde{P}_F(\phi) &\leq P_F(\phi), \\ \widetilde{P}_M(\phi) &\leq P_M(\phi). \end{aligned} \tag{15}$$

Since ϕ^* is optimal for the problem of minimizing $\widetilde{P}_M(\phi)$ subject to $\widetilde{P}_F(\phi) \leq \epsilon$, then ϕ^* is also optimal for problem of minimizing $P_M(\phi)$ subject to $P_F(\phi) \leq \epsilon$. $\square$

IV. MIXTURE CUSUM ALGORITHM AND A COMPUTATIONALLY EFFICIENT TEST

A. Mixture CuSum Algorithm

Motivated by the fact that the MLRT is optimal for the binary composite hypothesis testing problem, we construct the following mixture CuSum algorithm:

$$\tau^*(b) = \inf \left\{ t : \max_{1 \leq j \leq t} \sum_{i=j}^{t} \log \ell(X^n[i]) \geq b \right\}. \quad (16)$$

Let $W[t] = \max_{1 \leq j \leq t} \sum_{i=j}^{t} \log \ell(X^n[i])$. The test statistic $W[t]$ has the following recursion:

$$W[t+1] = (W[t])^+ + \log \ell(X^n[t+1]), W_0 = 0. \quad (17)$$

The following theorem shows that the mixture CuSum algorithm is exactly optimal under Lorden's criterion [50] in (6).

Theorem 1. *Consider the QCD problem in Section II-B, the mixture CuSum algorithm in* (16) *is exactly optimal under Lorden's criterion in* (6).

Proof Sketch. Consider a simple QCD problem with samples independent and identically distributed (i.i.d.) according to the pre-change distribution $\widetilde{\mathbb{P}}_0 = \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}$ and the post-change distribution $\widetilde{\mathbb{P}}_1 = \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}$, respectively. For this pair of pre- and post-change distributions, define the $\widetilde{\text{WADD}}$ and $\widetilde{\text{ARL}}$ for any stopping rule τ as follows:

$$\widetilde{\text{WADD}}(\tau) = \sup_{\nu \geq 1} \text{ess sup} \, \widetilde{\mathbb{E}}^\nu[(\tau - \nu)^+ | \widetilde{\mathbf{X}}^n[1, \nu-1]],$$
$$\widetilde{\text{ARL}}(\tau) = \widetilde{\mathbb{E}}^\infty[\tau], \quad (18)$$

where $\widetilde{\mathbb{E}}^\nu$ denotes the expectation when the change is at ν, the pre- and post-change distributions are $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$, and $\widetilde{\mathbf{X}}^n[t], 1 \leq t \leq \nu - 1$, are i.i.d. from $\widetilde{\mathbb{P}}_0$. For this new problem, the goal is to solve

$$\inf_{\tau : \widetilde{\text{ARL}}(\tau) \geq \gamma} \widetilde{\text{WADD}}(\tau) \quad (19)$$

for some prescribed $\gamma > 0$.

It was shown that the CuSum algorithm is exactly optimal for the problem in (19) under Lorden's criterion in [52]. Therefore, τ^* in (16) is exactly optimal for the QCD problem defined by pre- and post-change distributions $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$.

Following similar ideas as ones in Section III, we can show that for any stopping time τ,

$$\widetilde{\text{WADD}}(\tau) \leq \text{WADD}(\tau) \text{ and } \widetilde{\text{ARL}}(\tau) \geq \text{WARL}(\tau). \quad (20)$$

We will then show that τ^* achieves the equality in (20), which will complete the proof. Due to the fact that τ^* is symmetric, i.e., it is invariant to any permutation of $X^n[j]$, $\forall j = 1, 2, \ldots$. For any Ω and Ω', it follows that

$$\text{ess sup} \, \mathbb{E}_\Omega^\nu[(\tau^* - \nu)^+ | \mathbf{X}^n[1, \nu-1]]$$
$$= \text{ess sup} \, \mathbb{E}_{\Omega'}^\nu[(\tau^* - \nu)^+ | \mathbf{X}^n[1, \nu-1]],$$
$$\mathbb{E}_\Omega^\infty[\tau^*] = \mathbb{E}_{\Omega'}^\infty[\tau^*]. \quad (21)$$

To establish (20) and the optimality of τ^*, the proof is more involved than the binary hypothesis testing case in Sec. III due to the esssup and the conditional expectation. $\square$

The missing details of the proof can be found in Appendix A and Appendix B. The asymptotic optimality under Pollak's formulation [53] can also be derived similarly, and is ignored in this paper however due to space limitation.

The mixture likelihood ratio $\ell(X^n[i])$ needs to compute the average of the likelihood over all possible $\sigma \in \mathcal{S}_{n,\lambda}$. Note that the size of $\mathcal{S}_{n,\lambda}$ is $\binom{n}{n_1, \cdots, n_K}$. From the exponential bounds on the size of a type class [54], we have that $\frac{2^{nH\left(\left[\frac{n_1}{n} \cdots \frac{n_K}{n}\right]\right)}}{(n+1)^{|\mathcal{X}|}} \leq \binom{n}{n_1, \cdots, n_K} \leq 2^{nH\left(\left[\frac{n_1}{n} \cdots \frac{n_K}{n}\right]\right)}$, where $H\left(\left[\frac{n_1}{n} \cdots \frac{n_K}{n}\right]\right)$ denotes the entropy of $\left[\frac{n_1}{n} \cdots \frac{n_K}{n}\right]$. As $n \to \infty$, we have that $\lim_{n \to \infty} H\left(\left[\frac{n_1}{n} \cdots \frac{n_K}{n}\right]\right) = H(\boldsymbol{\alpha})$. Therefore, the computational complexity of mixture CuSum increases almost exponentially with n, which limits its practical applications in large networks. This motivates the need for computationally efficient tests for large networks. There are a wide range of applications in which the number of nodes is very large, e.g., IoT networks with thousands of sensors, smart grids with a large number of PMUs, crowdsourcing, and wireless sensor networks.

B. A Computationally Efficient Algorithm

In this section, we focus on discrete distributions, that is, the cardinality of $\mathcal{X}$ is finite, where $\mathcal{X}$ denotes the alphabet of the distributions $p_{\theta,k}, \forall \theta \in \{0, 1\}, k \in \{1, 2, \cdots, K\}$. We note that our mixture CuSum algorithm and its exact optimality result apply to general distributions, which are not necessarily discrete. Denote by $\mathcal{P}_{\mathcal{X}}$ the set of all distributions supported on $\mathcal{X}$. We propose a computationally efficient algorithm and then derive a lower bound on its WARL so that a threshold can be chosen analytically for false alarm control.

We first introduce some useful results that motivate the design of our algorithm. Let Π_{X^n} denote the empirical distribution of samples X^n, and let $T(\Pi_{X^n})$ denote the type class of Π_{X^n}. Then, it can be shown that [1]

$$\frac{\sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n)}{\sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n)} = \frac{\mathbb{P}_{1,\sigma}\left(T(\Pi_{X^n})\right)}{\mathbb{P}_{0,\sigma}\left(T(\Pi_{X^n})\right)}. \quad (22)$$

The right hand side of equation in (22) is a function of the empirical distribution Π_{X^n}. Let $\mathcal{P}_n$ denote the set of types with denominator n. For $n \geq 1$, let $Q_n \in \mathcal{P}_n$ be a sequence of distributions and $\lim_{n \to \infty} Q_n = Q$. The computation of the mixture likelihood ratio in (22) can be approximated by an optimization problem when n is large using the fact [2] that

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}_{\theta,\sigma}\left(T(Q_n)\right)$$
$$= -\inf_{\substack{\boldsymbol{U} = (U_1, \ldots, U_K) \in (\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T \boldsymbol{U} = Q}} \sum_{k=1}^{K} \alpha_k D(U_k || p_{\theta,k}). \quad (23)$$

The right hand side of (23) is a convex optimization problem with linear constraints, which can be solved efficiently using standard optimization tools [55], [56]. Its computational complexity is independent of the number of sensors. Therefore, for large n, the mixture over σ in (22) can be approximated by

---

[1] See Lemma 4.1 in [2] for the proof.
[2] See Lemma 5.2 in [2] for the proof.

solving a convex optimization problem whose computational complexity is independent of the network size n.

Let $\boldsymbol{P} = [P_1 \cdots P_K]^T$, where $P_k \in \mathcal{P}_{\mathcal{X}}$, for $1 \leq k \leq K$. For any $Q \in \mathcal{P}_{\mathcal{X}}$, define the following function of Q:

$$f_{\boldsymbol{P}}(\boldsymbol{\alpha}, Q) = \inf_{\substack{(U_1,\ldots,U_K) \in (\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T \boldsymbol{U} = Q}} \sum_{k=1}^{K} \alpha_k D(U_k || P_k). \quad (24)$$

Intuitively, an algorithm for the problem in Section II can be constructed by approximating the log of the mixture likelihood ratio at time t in the mixture CuSum algorithm using $n\big(f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[t]})\big)$. However, the lower bound on the WARL for this algorithm is difficult to derive due to the "inf" in the test statistic. We construct a novel test that can be updated recursively, and for which a lower bound on WARL can be theoretically derived. Moreover, as will be numerically demonstrated, this test has a WADD-WARL trade-off that is close to the optimal mixture CuSum, while also being computationally efficient.

Let $\hat{\nu}_t$ denote the change point estimate at time t. Denote by $\hat{t} \triangleq t - \hat{\nu}_t + 1$. We then design our detection statistic to approximate $W[t]$ in (16):

$$\widehat{W}[t] = \hat{t} n \big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t, t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t, t]}) \big]. \quad (25)$$

Instead of using a maximum likelihood approach to estimate $\hat{\nu}_t$ as in (16), which is not computationally efficient here, since $\hat{\nu}_t$ also appears in $\mathbf{X}^n[\hat{\nu}_t, t]$, we design a recursive way of updating $\hat{\nu}_t$. Let $\hat{\nu}_0 = 0$. If $\widehat{W}[t] \leq 0$, $\hat{\nu}_{t+1} = t + 1$, and if $\widehat{W}[t] > 0$, $\hat{\nu}_{t+1} = \hat{\nu}_t$. Then, $\Pi_{\mathbf{X}^n[\hat{\nu}_t, t]}$ can also be updated recursively: if $\widehat{W}[t] \leq 0$, $\Pi_{\mathbf{X}^n[\hat{\nu}_{t+1}, t+1]} = \Pi_{\mathbf{X}^n[t+1]}$, and if $\widehat{W}[t] > 0$, $\Pi_{\mathbf{X}^n[\hat{\nu}_{t+1}, t+1]} = \frac{\hat{t} \Pi_{\mathbf{X}^n[\hat{\nu}_t, t]} + \Pi_{\mathbf{X}^n[t+1]}}{\hat{t} + 1}$.

We next provide a heuristic explanation of how $\widehat{W}[t]$ evolves in the pre- and post-change regimes. According to the Glivenko–Cantelli theorem [57], before the change point ν, as $n \to \infty$, $\Pi_{\mathbf{X}^n[\hat{\nu}_t, t]}$ converges to $\boldsymbol{\alpha}^T \boldsymbol{P}_0$ almost surely. It can be easily seen that $f_{\boldsymbol{P}}(\boldsymbol{\alpha}, Q) \geq 0$ for any $\boldsymbol{\alpha}, \boldsymbol{P}$ and Q. The equality holds if and only if $\boldsymbol{\alpha}^T \boldsymbol{P} = Q$. This implies that $f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_0) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_0) < 0$. Therefore, before the change point ν, for large n, $\widehat{W}[t]$ has a negative drift. Similarly, after the change point ν, for large n, $\widehat{W}[t]$ has a positive drift of $f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1)$, and evolves towards ∞. This motivates us to construct the following computationally efficient test:

$$\tau_e = \inf \left\{ t \geq 1 : \widehat{W}[t] \geq b \right\}. \quad (26)$$

The computation cost of τ_e mainly lies in the update of the empirical distribution and the optimization step. The computational complexity of updating the empirical distribution increases linearly with n, and the computational complexity of the optimization step is independent of n. Therefore, the computationally efficient test is more efficient than the optimal mixture CuSum algorithm when n is large. Table II summarizes the computational complexity of the mixture CuSum algorithm and the computationally efficient algorithm.

In the following theorem, we present a lower bound on the WARL for our computationally efficient test in (26).

	Mixture CuSum	Efficient algorithm
Complexity	$O(2^{nH(\boldsymbol{\alpha})})$	$O(n)$

TABLE II
COMPUTATIONAL COMPLEXITY: MIXTURE CUSUM V.S. COMPUTATIONALLY EFFICIENT ALGORITHM.

Theorem 2. *Define* $\Gamma \triangleq \big\{ \mu \in \mathcal{P}_{\mathcal{X}} : f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \mu) > f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \mu) \big\}$. *Let*

$$h = \inf_{\substack{(U_1,\ldots,U_K) \in (\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T \boldsymbol{U} \in \Gamma}} \sum_{k=1}^{K} n_k D(U_k || P_{0,k}). \quad (27)$$

Then $h > 0$ *and for any* Ω,

$$\mathbb{E}_{\Omega}^{\infty} \big[\tau_e(b) \big] \geq \frac{e^b}{\big(\frac{b}{h} + 1 \big) \big(\prod_k |\mathcal{P}_{\frac{b}{h} n_k}| \big)}. \quad (28)$$

In the following, we provide a proof sketch, and the full proof can be found in Appendix C.

Proof Sketch. Let $Y = \inf\{t \geq 1 : \widehat{W}[t] \leq 0\}$ be the first regeneration time. For any Ω and $m \geq 1$, from Sanov's theorem [54], we can show that

$$\mathbb{P}_{\Omega}^{\infty}(Y > m) \leq \left(\prod_k |\mathcal{P}_{mn_k}| \right) e^{-mh}. \quad (29)$$

Define regeneration times $Y_0 = 0$ and for $r \geq 0$, $Y_{r+1} = \inf \big\{ t > Y_r : \widehat{W}[t] \leq 0 \big\}$. Let $R = \inf\{r : Y_r \leq \infty$ and $\widehat{W}[t] \geq b$ for some $Y_r < t \leq Y_{r+1}\}$ denote the index of the first cycle in which $\widehat{W}[t]$ crosses b. Note that according to the recursive update rule of $\hat{\nu}_t$ and $\widehat{W}[t]$, the test statistics in cycle $r+1$ are independent of the samples in cycles $1, \cdots, r$. For any Ω, we have that

$$\mathbb{E}_{\Omega}^{\infty}[\tau_e(b)] \geq \mathbb{E}_{\Omega}^{\infty}[R] = \sum_{r=0}^{\infty} \mathbb{P}_{\Omega}^{\infty}(R \geq r). \quad (30)$$

For any Ω and $m \geq 1$, we have that

$$\mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < Y) \leq \mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < m) + \mathbb{P}_{\Omega}^{\infty}(Y > m). \quad (31)$$

Consider the first term in (31) $\mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < m)$, by applying Sanov's theorem [54], we have that for any Ω,

$$\mathbb{P}_{\Omega}^{\infty}\big(\tau_e(b) < m\big) \leq m \left(\prod_k |\mathcal{P}_{mn_k}| \right) e^{-b}. \quad (32)$$

Let $m = \frac{b}{h}$. Combing (29) and (32), it follows that

$$\mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < Y) \leq \left(\frac{b}{h} + 1 \right) \left(\prod_k |\mathcal{P}_{\frac{b}{h} n_k}| \right) e^{-b}. \quad (33)$$

From (33) and the independence among the cycles [58], we have that

$$\mathbb{P}_{\Omega}^{\infty}(R \geq r) \geq \left(1 - \left(\frac{b}{h} + 1 \right) \left(\prod_k |\mathcal{P}_{\frac{b}{h} n_k}| \right) e^{-b} \right)^r, \quad (34)$$

Therefore, from (30) and (34), for any Ω,

$$\mathbb{E}_{\Omega}^{\infty}[\tau_e(b)] \geq \frac{e^b}{\big(\frac{b}{h} + 1 \big) \big(\prod_k |\mathcal{P}_{\frac{b}{h} n_k}| \big)}. \quad (35)$$

This completes the proof. $\qquad\square$

To guarantee that $\inf_{\Omega} \mathbb{E}_{\Omega}^{\infty}\big[\tau_e(b)\big] \geq \gamma$, it suffices to choose b such that $\frac{e^b}{\left(\frac{b}{h}+1\right)\left(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\right)} = \gamma$ and $b \sim \log \gamma$.

Note that an upper bound on the WADD for τ_e is difficult to obtain. To understand the detection delay of the proposed computationally efficient test, we then study the case when the change occurs at $\nu = 1$. We have the following result.

Proposition 1. *Consider the case with $\nu = 1$. Then, as $t \to \infty$,*

$$n\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]})\big] \to nf_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1),$$

almost surely.

Proof. According to the Glivenko–Cantelli theorem [57], as $t \to \infty$, under the post-change distribution, the empirical distribution $\Pi_{\mathbf{X}^n[1,t]}$ convergences to $\boldsymbol{\alpha}^T \boldsymbol{P}_1$ almost surely. Due to the fact that $f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1) = 0$, we have that $\frac{1}{t}(t-1+1)n\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]})\big]$ converges to $nf_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1)$ almost surely. $\qquad\square$

Intuitively, Proposition 1 implies that if the change is at $\nu = 1$ and regeneration does not happen, then the detection delay of the computationally efficient algorithm increases linearly with the threshold b at the rate of $1/(nf_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1))$.

We then present the following universal lower bound on the WADD, and show that the slope is also $1/(nf_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1))$ when n is large.

Proposition 2. *For large γ, we have that*

$$\inf_{\tau: WARL \geq \gamma} WADD(\tau) \sim \frac{\log \gamma}{D\left(\widetilde{\mathbb{P}}_1 || \widetilde{\mathbb{P}}_0\right)}(1 + o(1)). \tag{36}$$

Moreover, as $n \to \infty$,

$$\lim_{n \to \infty} \frac{1}{n} D\left(\widetilde{\mathbb{P}}_1 || \widetilde{\mathbb{P}}_0\right) = f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1). \tag{37}$$

Proof Sketch. It was shown in Section IV-A that the mixture CuSum τ^* is exactly optimal for the QCD problem in Section II-B. Then, as $\gamma \to \infty$, we have that $\inf_{\tau: WARL \geq \gamma} WADD(\tau) = WADD(\tau^*)$. Further note that for the mixture CuSum τ^*, τ^* achieves the equality in (20). Then, we have that $\widetilde{WADD}(\tau^*) = WADD(\tau^*)$. Since τ^* is optimal for the simple QCD problem in (19), from Theorem 4 in [41], as $\gamma \to \infty$, it follows that

$$WADD(\tau^*) = \widetilde{WADD}(\tau^*) \sim \frac{\log \gamma}{D\left(\widetilde{\mathbb{P}}_1 || \widetilde{\mathbb{P}}_0\right)}(1 + o(1)). \tag{38}$$

The proof of (37) can be found in Appendix D. $\qquad\square$

By combining Propositions 1 and 2, it can be seen that the tradeoff between the WADD and WARL for our computationally efficient test is close to the optimal one when n is large. This demonstrates the advantage of our test that for large networks, it has a similar statistical efficiency comparing to the optimal test, and has a significantly reduced computational complexity.

V. SIMULATION RESULTS

A. Mixture CuSum Algorithm

We first show an example evolution path of the mixture CuSum algorithm. Set $n = 2$ and $K = 2$, i.e., one sensor in each group. For group 1, the pre- and post-change distributions are $\mathcal{N}(0,1)$ and $\mathcal{N}(0.5,1)$, respectively. For group two, the pre- and post-change distribution are $\mathcal{N}(2,1)$ and $\mathcal{N}(1.5,1)$, respectively. In Fig. 2, we set the change point to be 500 and $b = 5$. We plot one sample evolution path of the mixture CuSum algorithm. It can be seen that before the change point, the test statistic fluctuates around zero, and after the change point, it starts to increase with a positive drift.

We then compare our optimal mixture CuSum test with two other heuristic algorithms based on the Bayesian approach and the generalized likelihood ratio approach to tackling the unknown group assignments. For the Bayesian approach, we pretend that each sample comes from group k with probability n_k/n, for $k = 1, \ldots, K$, independently, so that on average the k-th group has n_k sensors, although we actually have exact n_k sensors in each group k. We then compute the following likelihood ratio:

$$l_b(x^n[t]) = \frac{\prod_{i=1}^n \left(\sum_{k=1}^K \frac{n_k}{n} p_{1,k}(x_i[t])\right)}{\prod_{i=1}^n \left(\sum_{k=1}^K \frac{n_k}{n} p_{0,k}(x_i[t])\right)}. \tag{39}$$

The generalized likelihood ratio for the sample $x^n[t]$ is

$$l_g(x^n[t]) = \frac{\sup_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(x^n[t])}{\sup_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(x^n[t])}. \tag{40}$$

We then design CuSum-type tests using (39) and (40), which are referred to as Bayesian and Generalized CuSums. The test statistics of these three algorithms are all symmetric, and therefore for different Ω, the average detection delay and average run length are the same.

In Fig. 3, we plot the WADD as a function of the WARL. It can be seen that our mixture CuSum algorithm outperforms the other two algorithms. Moreover, the relationship between the WADD and log of the WARL is linear. The slope of these three curves should be the reciprocal of the expectation of the corresponding likelihood ratio under $\mathbb{P}_{1,\sigma}$ for some $\sigma \in \mathcal{S}_{n,\lambda}$. Due to the fact that the distributions are continuous, our computationally efficient test is not applicable here.

B. Computationally Efficient Algorithm

For the computationally efficient algorithm, we first consider a simple example with $n = 2$, $K = 2$, $n_1 = 1$ and $n_2 = 1$. The pre- and post-change distributions for group 1 are binomial distribution $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.3)$, respectively, and for group 2 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.7)$, respectively. We plot a sample evolution path of the efficient algorithm in Fig. 4. Similar to the mixture CuSum, before the change point, the test statistic fluctuates around zero, and after the change point, it starts to increase with a positive drift.

We then compare the performance of our efficient algorithm with the optimal mixture CuSum algorithm, the Bayesian CuSum algorithm and the Generalized CuSum algorithm, and
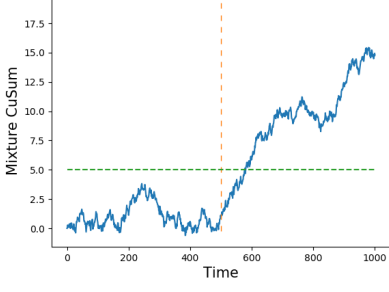
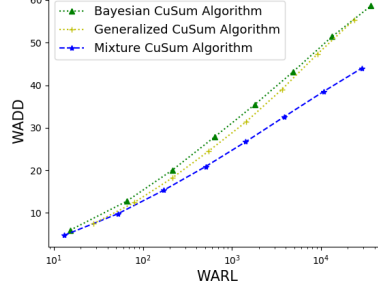Fig. 2. Evolution path of the Mixture CuSum.

Fig. 3. Comparison of the Bayesian CuSum algorithm, the Generalized CuSum algorithm and the Mixture CuSum algorithm: $n = 2, K = 2$
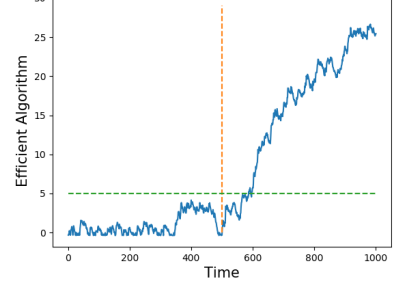

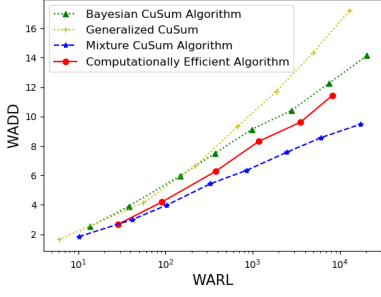Fig. 4. Evolution path of the computationally efficient algorithm: $n = 2, K = 2$.

Fig. 5. Comparison of the Bayesian CuSum algorithm, the Generalized CuSum algorithm, the Mixture CuSum algorithm and the computationally efficient algorithm: $n = 2, K = 2$.
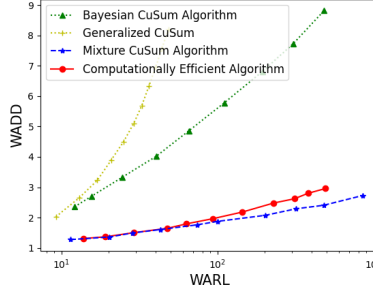

Fig. 6. Comparison of the Bayesian CuSum algorithm, the Generalized CuSum algorithm, the Mixture CuSum algorithm and the computationally efficient algorithm: $n = 8, K = 2$.
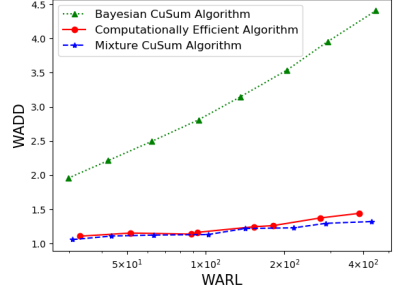

Fig. 7. Comparison of the Bayesian CuSum algorithm, the Mixture CuSum algorithm and the computationally efficient algorithm: $n = 20, K = 2$.
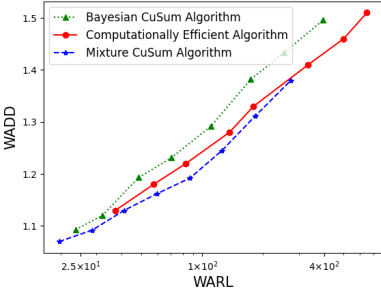

Fig. 8. Comparison of the Bayesian CuSum algorithm, the Mixture CuSum algorithm and the computationally efficient algorithm: $n = 10, K = 4$.
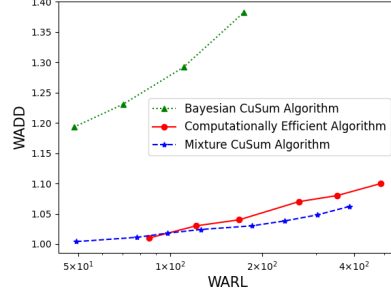

Fig. 9. Comparison of the Bayesian CuSum algorithm, the Mixture CuSum algorithm and the computationally efficient algorithm: $n = 18, K = 4$.
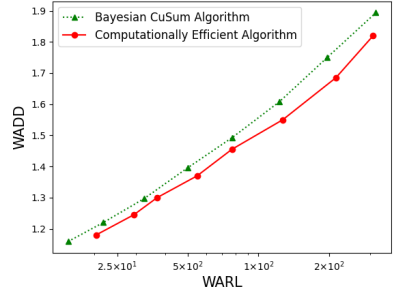

Fig. 10. Comparison of the Bayesian CuSum algorithm and the computationally efficient algorithm: $n = 100, K = 4$.

repeat the experiment for $n = 8$, $n_1 = 4$, $n_2 = 4$ and $n = 20$, $n_1 = 10, n_2 = 10$ with the same distributions.

For the three cases with $n = 2$, $n = 8$ and $n = 20$, we plot the WADD as a function of the WARL in Figs. 5, Fig. 6 and Fig. 7. It can be seen that mixture CuSum outperforms the other three tests, and our computationally efficient test has a better performance than the intuitive Bayesian CuSum and Generalized CuSum. For the case with $n = 20$, $n_1 = 10, n_2 = 10$, the performance of the Generalized CuSum algorithm is much worse than the other three algorithms, therefore is not included in Fig. 7. More importantly, comparing Fig. 5, Fig. 6 and Fig. 7, we can see that as n increases, the slope of the WADD-WARL tradeoff curve of the efficient algorithm gets

closer to the one of the optimal mixture CuSum algorithm. This conforms to the design of our computationally efficient test which aims to approximate the optimal mixture CuSum when n is large, and our theoretical discussion in Propositions 1 and 2.

We then consider the case with $K = 4$. The pre- and post-change distributions for group 1 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.3)$, respectively, for group 2 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.7)$, respectively, for group 3 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.25)$, respectively, for group 4 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.75)$, respectively. In Fig. 8, we plot the WADD as a function of the WARL with $n = 10, n_1 = n_2 = n_3 = 2$ and $n_4 = 4$. In Fig. 9, we plot the WADD as a function of the WARL with

$n = 18, n_1 = n_2 = n_3 = 3$ and $n_4 = 9$. From Fig. 8 and Fig. 9, it can be seen that the mixture CuSum has the best performance, and our computationally efficient algorithm outperforms other heuristic algorithms, and is close to the optimal mixture CuSum algorithm. We also compare the performance of Bayesian CuSum algorithm and the efficient algorithm when $n = 100$ and $K = 4$. We note that for the case $n = 100$, the computational complexity is too high for the mixture CuSum algorithm to be simulated. We set $n_1 = n_2 = n_3 = 20$ and $n_4 = 40$, and the pre- and post-change distributions for group 1 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.4)$, respectively, for group 2 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.45)$, respectively, for group 3 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.35)$, respectively, for group 4 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.6)$, respectively. From Fig. 10, it can be seen that the performance of our efficient algorithm is better than the Bayesian CuSum algorithm.

In Fig. 11, we show the computational efficiency of our proposed algorithms. Specifically, we compare the running time of computing one step update of our computationally efficient algorithm and the optimal mixture CuSum algorithm (on Intel Core i5-8265U CPU). From Fig. 5, one can see that as n increases, the running time of the mixture CuSum increases exponentially, while the running time of our computationally efficient test stays almost the same.
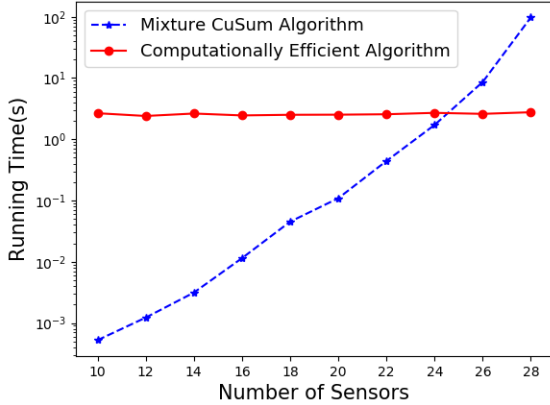

Fig. 11. Comparison of the computational complexity between the Mixture CuSum algorithm and the computationally efficient algorithm.

VI. CONCLUSION

In this paper, we studied the statistical inference problem in anonymous heterogeneous sensor networks. We first revisited the non-sequential setting studied in [2], and provided a simple optimality proof for the MLRT. We then extended our approach to the problem of QCD with anonymous heterogeneous sensors, and constructed a mixture CuSum algorithm. We showed that the mixture CuSum algorithm is optimal under Lorden's criterion [50]. We note that asymptotic optimality results can also be obtained under Pollak's criterion [53]. Although being optimal, our mixture CuSum algorithm is computationally expensive when the number of sensors is large. We then proposed a computationally efficient algorithm

with a novel recursive update rule of the change point estimate and the test statistic. We further developed its WARL lower bound for practical false alarm control. Our numerical results showed that the mixture CuSum algorithm has the best performance and the computationally efficient algorithm also outperforms other heuristic algorithms. Moreover, when the number of sensor is large, the computationally efficient algorithm is much more efficient than the optimal mixture CuSum algorithm. Our results provide useful tools and insights to investigate various kinds of statistical inference problems in anonymous networks.

One possible extension is to the case where the samples in different time steps are not independent [41]. It is also of interest to investigate when samples are quantized and sensors can only receive binary codewords [59]. In this case, such quantizing measurement should be incorporated into the design of mixture CuSum algorithm. Moreover, in this paper, it is assumed that after the change all the sensors change their data-generating distributions simultaneously. Therefore, another possible future direction is to consider the case where only an unknown subset of sensors are affected by the change. Moreover, the change may also be dynamic and propagate following some unknown pattern. In many practical applications, the data-generating distributions may not be available beforehand, and data-driven approaches in anonymous heterogeneous networks need to be developed.

APPENDIX A
PROOF OF (20)

We construct a new sequence of random variables $\{\widehat{X}^n[t]\}_{t=1}^{\infty}$. Before the change point, $\widehat{X}^n[t]$ are i.i.d. according to the mixture distribution $\widetilde{\mathbb{P}}_0 = \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}$. After the change point, $\widehat{X}^n[t]$ follows the distribution $\mathbb{P}_{1,\sigma_t}$ for some $\sigma_t \in \mathcal{S}_{n,\lambda}$. Specifically,

$$\widehat{X}^n[t] \sim \begin{cases} \widetilde{\mathbb{P}}_0, & \text{if } t < \nu \\ \mathbb{P}_{1,\sigma_t}, & \text{if } t \geq \nu. \end{cases} \qquad (41)$$

For any stopping time τ, define the worst-case average detection delay for the model in (41) as follows:

$$\widehat{\text{WADD}}(\tau)$$
$$= \sup_{\nu \geq 1} \sup_{\sigma_\nu,...,\sigma_\infty} \text{ess sup} \widehat{\mathbb{E}}^\nu_{\sigma_\nu,...,\sigma_\infty}[(\tau - \nu)^+ | \widehat{\mathbf{X}}^n[1, \nu-1]],$$

where $\widehat{\mathbb{E}}^\nu_{\sigma_\nu,...,\sigma_\infty}$ denotes the expectation when the data is distributed according to (41). To prove that $\text{WADD}(\tau) \geq \widetilde{\text{WADD}}(\tau)$, we will first show that $\text{WADD}(\tau) = \widehat{\text{WADD}}(\tau)$, and then show that $\widehat{\text{WADD}}(\tau) \geq \widetilde{\text{WADD}}(\tau)$.

Step 1. Denote by $\mathcal{M}$ the collection of all $\{\sigma_1, ..., \sigma_{\nu-1}\}$, and μ is an element in $\mathcal{M}$. Denote by $\mathcal{N}$ the collection of all $\{\sigma_\nu, ..., \sigma_\infty\}$, and ω is an element in $\mathcal{N}$. Thus, $\Omega = \{\mu, \omega\}$. Then, the WADD can be written as

$$\text{WADD}(\tau)$$
$$= \sup_{\nu \geq 1} \sup_{\Omega} \text{ess sup} \mathbb{E}^\nu_\Omega[(\tau - \nu)^+ | \mathbf{X}^n[1, \nu-1]]$$
$$= \sup_{\nu \geq 1} \sup_{\omega \in \mathcal{N}} \sup_{\mu \in \mathcal{M}} \text{ess sup} \mathbb{E}^\nu_\omega[(\tau - \nu)^+ | \mathbf{X}^n[1, \nu-1]], \quad (42)$$

where $\mathbb{E}_\omega^\nu$ denotes the expectation when change point is ν, and after the change point, the data follows distribution $\prod_{t=\nu}^\infty \mathbb{P}_{1,\sigma_t}$. We note that $\widehat{X}^n[t]$ and $X^n[t]$, for $t \geq \nu$, have the same distribution $\mathbb{P}_{1,\sigma_t}$. Therefore, the difference between WADD and $\widehat{\text{WADD}}$ lies in that they take esssup with respect to different distributions, i.e., the distributions of $\mathbf{X}^n[1,\nu-1]$ and $\widehat{\mathbf{X}}^n[1,\nu-1]$ are different. Let $f_\omega(\mathbf{X}^n[1,\nu-1])$ denote $\mathbb{E}_\omega^\nu[(\tau-\nu)^+|\mathbf{X}^n[1,\nu-1]]$. Then, WADD and $\widehat{\text{WADD}}$ can be written as

$$\text{WADD}(\tau) = \sup_{\nu \geq 1} \sup_{\omega \in \mathcal{N}} \sup_{\mu \in \mathcal{M}} \text{ess sup} f_\omega(\mathbf{X}^n[1,\nu-1]),$$

$$\widehat{\text{WADD}}(\tau) = \sup_{\nu \geq 1} \sup_{\omega \in \mathcal{N}} \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]). \quad (43)$$

It then suffices to show that $\sup_{\mu \in \mathcal{M}} \text{ess sup} f_\omega(\mathbf{X}^n[1,\nu-1]) = \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$.

For any $\omega \in \mathcal{N}$ and $\mu \in \mathcal{M}$, let $b_{\omega,\mu} = \text{ess sup} f_\omega(\mathbf{X}^n[1,\nu-1]) = \inf\{b : \mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b) = 0\}$, where $\mathbb{P}_\mu$ denotes the probability measure when the data is generated according to $\mathbb{P}_{0,\sigma_1},...,\mathbb{P}_{0,\sigma_{\nu-1}}$ before change point ν.

Let $b_\omega^* = \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$. It can be shown that

$$b_\omega^* = \inf\left\{b : \int_{\mathbf{x}^n[1,\nu-1]} \mathbb{1}_{\{f_\omega(\mathbf{x}^n[1,\nu-1])>b\}}\right.$$
$$\left. \times d\prod_{t=1}^{\nu-1} \widetilde{\mathbb{P}}_0(x^n(t)) = 0\right\}$$

$$= \inf\left\{b : \int_{\mathbf{x}^n[1,\nu-1]} \mathbb{1}_{\{f_\omega(\mathbf{x}^n[1,\nu-1])>b\}}\right.$$
$$\left. \times d\prod_{t=1}^{\nu-1} \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma_t \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma_t}(x^n(t)) = 0\right\}$$

$$= \inf\left\{b : \int_{\mathbf{x}^n[1,\nu-1]} \mathbb{1}_{\{f_\omega(\mathbf{x}^n[1,\nu-1])>b\}}\right.$$
$$\left. \times d\frac{1}{|\mathcal{M}|} \sum_{\mu \in \mathcal{M}} \mathbb{P}_\mu(\mathbf{x}^n[1,\nu-1]) = 0\right\}$$

$$= \inf\left\{b : \frac{1}{|\mathcal{M}|} \sum_{\mu \in \mathcal{M}} \mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b) = 0\right\}.$$

It then follows that for any $\mu \in \mathcal{M}$,

$$\mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b_\omega^*) = 0. \quad (44)$$

Therefore, for any $\mu \in \mathcal{M}$, we have that $b_{\omega,\mu} \leq b_\omega^*$. Then

$$\sup_{\mu \in \mathcal{M}} b_{\omega,\mu} \leq b_\omega^*. \quad (45)$$

Conversely, let $\sup_{\mu \in \mathcal{M}} b_{\omega,\mu} = b'$. For any $\mu \in \mathcal{M}$, we have $\mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b') = 0$. Then, $\frac{1}{|\mathcal{M}|}\sum_{\mu \in \mathcal{M}} \mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b') = 0$. This further implies that

$$b_\omega^* \leq b' = \sup_{\mu \in \mathcal{M}} b_{\omega,\mu}. \quad (46)$$

Combining (45) and (46), we have that $\sup_{\mu \in \mathcal{M}} b_{\omega,\mu} = b_\omega^*$, and thus $\sup_{\mu \in \mathcal{M}} \text{ess sup} f_\omega(\mathbf{X}^n[1,\nu-1]) = \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$. This implies that

$$\text{WADD}(\tau) = \widehat{\text{WADD}}(\tau). \quad (47)$$

Step 2. The next step is to show that $\widehat{\text{WADD}}(\tau) \geq \widetilde{\text{WADD}}(\tau)$. We will first show that $\sup_{\omega \in \mathcal{N}} \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \geq \text{ess sup} \sup_{\omega \in \mathcal{N}} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$. Denote by $\widetilde{\mathbb{P}}^\nu$ the probability measure when the change is at ν, the pre- and post-change distributions are $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$, respectively. Let $\hat{b} = \sup_{\omega \in \mathcal{N}} \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$. For any $\omega \in \mathcal{N}$, we have that $\widetilde{\mathbb{P}}^\nu(f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \geq \hat{b}) = 0$. Since $\mathcal{N}$ is countable, it then follows that

$$\widetilde{\mathbb{P}}^\nu\left(\sup_{\omega \in \mathcal{N}} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \geq \hat{b}\right)$$
$$\leq \widetilde{\mathbb{P}}^\nu\left(\cup_{\omega \in \mathcal{N}} \{f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) > \hat{b}\}\right)$$
$$\leq \sum_{\omega \in \mathcal{N}} \widetilde{\mathbb{P}}^\nu\left(f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) > \hat{b}\right) = 0. \quad (48)$$

Therefore,

$$\hat{b} = \sup_{\omega \in \mathcal{N}} \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$$
$$\geq \text{ess sup} \sup_{\omega \in \mathcal{N}} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]). \quad (49)$$

Before the change point ν, $\widehat{X}^n[t]$ and $\widetilde{X}^n[t]$ follow the same distribution. For any $T \geq \nu+1$, we have that

$$\sup_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda}^{\otimes(T-\nu+1)}}} \sum_{t=\nu+1}^T (t-\nu)\mathbb{P}_{\sigma_\nu,...,\sigma_T}(\tau=t|\widehat{\mathbf{X}}^n[1,\nu-1])$$

$$\geq \sum_{t=\nu+1}^T (t-\nu)\frac{1}{|\mathcal{S}_{n,\lambda}|^{(T-\nu+1)}}$$
$$\times \sum_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda}^{\otimes(T-\nu+1)}}} \mathbb{P}_{\sigma_\nu,...,\sigma_T}(\tau=t|\widehat{\mathbf{X}}^n[1,\nu-1])$$

$$= \sum_{t=\nu+1}^T (t-\nu)\widetilde{\mathbb{P}}^\nu(\tau=t|\widetilde{\mathbf{X}}^n[1,\nu-1]). \quad (50)$$

As $T \to \infty$, we have that

$$\widehat{\mathbb{E}}_\omega^\nu[(\tau-\nu)|\widehat{\mathbf{X}}^n[1,\nu-1]] \geq \widetilde{\mathbb{E}}^\nu[(\tau-\nu)|\widetilde{\mathbf{X}}^n[1,\nu-1]], \quad (51)$$

where $\mathbb{P}_{\sigma_\nu,...,\sigma_T}$ denotes the probability measure when the observations from time ν to time T are generated according to $\mathbb{P}_{\sigma_\nu},...,\mathbb{P}_{\sigma_T}$.

From (49) and (51), we have that

$$\widehat{\text{WADD}}(\tau)$$
$$= \sup_{\omega \in \mathcal{N}} \text{ess sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$$
$$\geq \text{ess sup} \widetilde{\mathbb{E}}^\nu[(\tau-\nu)^+|\widetilde{\mathbf{X}}^n[1,\nu-1]]$$
$$= \widetilde{\text{WADD}}(\tau). \quad (52)$$

Combining (47) and (52), it follows that $\text{WADD}(\tau) = \widehat{\text{WADD}}(\tau) \geq \widetilde{\text{WADD}}(\tau)$. Similarly, it can be shown that $\text{WARL}(\tau) \leq \widetilde{\text{ARL}}(\tau)$. This concludes the proof.

APPENDIX B
τ^* ACHIEVES EQUALITY IN (20)

We will show that the mixture CuSum algorithm achieves the equality in (20), i.e.,

$$\widehat{\text{WADD}}(\tau^*) = \widetilde{\text{WADD}}(\tau^*). \tag{53}$$

For any $\{\sigma_\nu, ..., \sigma_i, ..., \sigma_\infty\}$, consider another element in $\mathcal{N}$, $\{\sigma_\nu, ..., \sigma_i', ...\sigma_\infty\}$. Due to the fact that τ^* is symmetric, it follows that for any $i \geq \nu$, and any $\sigma_i, \sigma_i' \in \mathcal{S}_{n,\lambda}$,

$$\operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_{\sigma_\nu,...,\sigma_i,...,\sigma_\infty}[(\tau^* - \nu)^+ | \widehat{\mathbf{X}}^n[1, \nu - 1]]$$
$$= \operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_{\sigma_\nu,...,\sigma_i',...,\sigma_\infty}[(\tau^* - \nu)^+ | \widehat{\mathbf{X}}^n[1, \nu - 1]]. \tag{54}$$

Therefore, $\widehat{\text{WADD}}(\tau^*)$ doesn't depend on ω, which further implies that

$$\sup_{\omega \in \mathcal{N}} \operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_\omega[(\tau^* - \nu)^+ | \widehat{\mathbf{X}}^n[1, \nu - 1]]$$
$$= \operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_\omega[(\tau^* - \nu)^+ | \widehat{\mathbf{X}}^n[1, \nu - 1]]. \tag{55}$$

For any $T \geq \nu + 1$, we have that

$$\sup_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda} \otimes (T-\nu+1)}} \sum_{t=\nu+1}^T (t - \nu) \mathbb{P}_{\sigma_\nu,...,\sigma_T}(\tau^* = t | \widehat{\mathbf{X}}^n[1, \nu - 1])$$

$$= \sum_{t=\nu+1}^T (t - \nu) \frac{1}{|\mathcal{S}_{n,\lambda}|^{(T-\nu+1)}}$$
$$\times \sum_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda} \otimes (T-\nu+1)}} \mathbb{P}_{\sigma_\nu,...,\sigma_T}(\tau^* = t | \widehat{\mathbf{X}}^n[1, \nu - 1])$$

$$= \sum_{t=\nu+1}^T (t - \nu) \widetilde{\mathbb{P}}^\nu(\tau^* = t | \widetilde{\mathbf{X}}^n[1, \nu - 1]). \tag{56}$$

As $T \to \infty$, we have that

$$\widehat{\mathbb{E}}^\nu_\omega[(\tau^* - \nu)^+ | \widehat{\mathbf{X}}^n[1, \nu - 1]] = \widetilde{\mathbb{E}}^\nu[(\tau^* - \nu)^+ | \widetilde{\mathbf{X}}^n[1, \nu - 1]]. \tag{57}$$

From (55) and (57), it follows that

$$\widehat{\text{WADD}}(\tau^*)$$
$$= \sup_{\nu \geq 1} \sup_{\omega \in \mathcal{N}} \operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_\omega[(\tau^* - \nu)^+ | \widehat{\mathbf{X}}^n[1, \nu - 1]]$$
$$= \sup_{\nu \geq 1} \operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_\omega[(\tau^* - \nu)^+ | \widetilde{\mathbf{X}}^n[1, \nu - 1]]$$
$$= \sup_{\nu \geq 1} \operatorname{ess\,sup} \widetilde{\mathbb{E}}^\nu[(\tau^* - \nu)^+ | \widetilde{\mathbf{X}}^n[1, \nu - 1]]$$
$$= \widetilde{\text{WADD}}(\tau^*). \tag{58}$$

Similarly, it can be shown that $\widehat{\text{ARL}}(\tau^*) = \text{WARL}(\tau^*)$.

APPENDIX C
PROOF OF THEOREM 2

Let $Y = \inf\{t \geq 1 : \widehat{W}[t] \leq 0\}$ be the first regeneration time. For any Ω and $m \geq 1$, we have that

$$\mathbb{P}^\infty_\Omega(Y > m) = \mathbb{P}^\infty_\Omega(\widehat{W}[t] > 0, \forall t \in [1, m])$$
$$\leq \mathbb{P}^\infty_\Omega\left(nm[f_{\mathbf{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]}) - f_{\mathbf{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]})] > 0\right).$$

Let $\Gamma \triangleq \{\mu \in \mathcal{P}_\mathcal{X} | f_{\mathbf{P}_0}(\boldsymbol{\alpha}, \mu) > f_{\mathbf{P}_1}(\boldsymbol{\alpha}, \mu)\}$. We have that

$$\mathbb{P}^\infty_\Omega\left(nm[f_{\mathbf{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]}) - f_{\mathbf{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]})] > 0\right)$$
$$= \mathbb{P}^\infty_\Omega\{\Pi_{\mathbf{X}^n[1,m]} \in \Gamma\}$$
$$= \sum_{\substack{(U_1,...,U_K) \in \mathcal{P}_{mn_1} \times ... \times \mathcal{P}_{mn_K} \\ \boldsymbol{\alpha}^T \mathbf{U} \in \Gamma}} \prod_{k=1}^K p_{0,k}^{\otimes mn_k}(T_{mn_k}(U_k))$$
$$\leq \sum_{\substack{(U_1,...,U_K) \in \mathcal{P}_{mn_1} \times ... \times \mathcal{P}_{mn_K} \\ \boldsymbol{\alpha}^T \mathbf{U} \in \Gamma}} e^{-\sum_{k=1}^K mn_k D(U_k || p_{0,k})}$$
$$\leq \left(\prod_k |\mathcal{P}_{mn_k}|\right)$$
$$\cdot \exp\left(-\inf_{\substack{(U_1,...,U_K) \in \mathcal{P}_{mn_1} \times ... \times \mathcal{P}_{mn_K} \\ \boldsymbol{\alpha}^T \mathbf{U} \in \Gamma}} \sum_{k=1}^K mn_k D(U_k || p_{0,k})\right)$$
$$\leq \left(\prod_k |\mathcal{P}_{mn_k}|\right) e^{-hm}, \tag{59}$$

where the last step is due to the fact that $\mathcal{P}_{mn_k} \subseteq \mathcal{P}_\mathcal{X}$, $\forall 1 \leq k \leq K$. Note that $f_{\mathbf{P}}(\boldsymbol{\alpha}, Q) \geq 0$ for any Q and the equality holds if and only if $\boldsymbol{\alpha}^T \mathbf{P} = Q$ almost everywhere. We then have that $\boldsymbol{\alpha}^T \mathbf{P}_0 \notin \Gamma$ and $h > 0$. Therefore, for any Ω and $m \geq 1$,

$$\mathbb{P}^\infty_\Omega(Y > m) \leq \left(\prod_k |\mathcal{P}_{mn_k}|\right) e^{-mh}. \tag{60}$$

Define regeneration times $Y_0 = 0$ and for $r \geq 0$, $Y_{r+1} = \inf\{t > Y_r : \widehat{W}[t] \leq 0\}$. Let $R = \inf\{r : Y_r \leq \infty \text{ and } \widehat{W}[t] \geq b \text{ for some } Y_r < t \leq Y_{r+1}\}$ denote the index of the first cycle in which $\widehat{W}[t]$ crosses b. Note that according to the recursive update rule of $\hat\nu_t$ and $\widehat{W}[t]$, the test statistics in cycle $r + 1$ are independent of the samples in cycles $1, \cdots, r$. For any Ω, we have that

$$\mathbb{E}^\infty_\Omega[\tau_e(b)] \geq \mathbb{E}^\infty_\Omega[R] = \sum_{r=0}^\infty \mathbb{P}^\infty_\Omega(R \geq r). \tag{61}$$

For any Ω and $m \geq 1$, we have that

$$\mathbb{P}^\infty_\Omega(\tau_e(b) < Y)$$
$$= \mathbb{P}^\infty_\Omega(\tau_e(b) < Y, Y \leq m) + \mathbb{P}^\infty_\Omega(\tau_e(b) < Y, Y > m)$$
$$\leq \mathbb{P}^\infty_\Omega(\tau_e(b) < m) + \mathbb{P}^\infty_\Omega(Y > m). \tag{62}$$

Consider the first term in (62) $\mathbb{P}^\infty_\Omega(\tau_e(b) < m)$:

$$\mathbb{P}^\infty_\Omega(\tau_e(b) < m) = \mathbb{P}^\infty_\Omega\left(\max_{1 \leq t < m} \widehat{W}[t] \geq b\right)$$
$$\leq \sum_{1 \leq t < m} \mathbb{P}^\infty_\Omega\left(\widehat{W}[t] \geq b\right)$$
$$= \sum_{1 \leq t < m} \mathbb{P}^\infty_\Omega\left(n\hat{t}[f_{\mathbf{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat\nu_t, t]}) - f_{\mathbf{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat\nu_t, t]})] \geq b\right).$$

Define $\Gamma_{b,t} \triangleq \left\{\mu \in \mathcal{P}_\mathcal{X} | n\hat{t}[f_{\mathbf{P}_0}(\boldsymbol{\alpha}, \mu) - f_{\mathbf{P}_1}(\boldsymbol{\alpha}, \mu)] \geq b\right\}$. For all $\mu \in \Gamma_{b,t}$, we have that $n\hat{t} f_{\mathbf{P}_0}(\boldsymbol{\alpha}, \mu) \geq b + n\hat{t} f_{\mathbf{P}_1}(\boldsymbol{\alpha}, \mu) \geq b$, where the last inequality is due to the facts that $\hat{t} \geq 0$ and

$f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \mu) \geq 0$. For any Ω and $1 \leq t < m$, following the same idea as the one in (59), we have that

$$\mathbb{P}_{\Omega}^{\infty}\Big(n\hat{t}\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t, t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t, t]})\big] > b\Big)$$

$$= \mathbb{P}_{\Omega}^{\infty}\Big\{\Pi_{\mathbf{X}^n[\hat{\nu}_t, t]} \in \Gamma_{b, t}\Big\}$$

$$\leq \left(\prod_k |\mathcal{P}_{\hat{t}n_k}|\right)$$

$$\cdot \exp\left(-\inf_{\substack{(U_1, \ldots, U_K) \in (\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T \boldsymbol{U} \in \Gamma_{b, t}}} \sum_{k=1}^{K} n_k \hat{t} D(U_k \| p_{0, k})\right)$$

$$\leq \left(\prod_k |\mathcal{P}_{mn_k}|\right) e^{-b}. \tag{63}$$

We then have that for any Ω,

$$\mathbb{P}_{\Omega}^{\infty}\big(\tau_e(b) < m\big) \leq m\left(\prod_k |\mathcal{P}_{mn_k}|\right) e^{-b}. \tag{64}$$

Let $m = \frac{b}{h}$. Combing (60) and (63), we have that

$$\mathbb{P}_{\Omega}^{\infty}\big(\tau_e(b) < Y\big) \leq \left(\frac{b}{h} + 1\right)\left(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\right) e^{-b}. \tag{65}$$

It then follows that

$$\mathbb{P}_{\Omega}^{\infty}(R \geq r) = \mathbb{P}_{\Omega}^{\infty}\Big(\widehat{W}[t] < b, \forall t \leq Y_r\Big)$$

$$= \mathbb{P}_{\Omega}^{\infty}\Big(\widehat{W}[t] < b, \forall Y_{m-1} \leq t \leq Y_m, \forall 1 \leq m \leq r\Big)$$

$$= \prod_{m=1}^{r} \mathbb{P}_{\Omega}^{\infty}\Big(\widehat{W}[t] < b, \forall Y_{m-1} \leq t \leq Y_m\Big)$$

$$\geq \left(1 - \left(\frac{b}{h} + 1\right)\left(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\right) e^{-b}\right)^r, \tag{66}$$

where the second equality is due to (65) and the independence among the cycles [58]. Therefore, for any Ω,

$$\mathbb{E}_{\Omega}^{\infty}[\tau_e(b)] \geq \sum_{r=0}^{\infty}\left(1 - \left(\frac{b}{h} + 1\right)\left(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\right) e^{-b}\right)^r$$

$$= \frac{e^b}{\left(\frac{b}{h} + 1\right)\left(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\right)}. \tag{67}$$

This completes the proof.

APPENDIX D
PROOF OF (37)

From (22), we have that for any $\sigma \in \mathcal{S}_{n, \lambda}$,

$$\log \frac{\widetilde{\mathbb{P}}_1(X^n)}{\widetilde{\mathbb{P}}_0(X^n)} = \log \frac{\mathbb{P}_{1, \sigma}\big(T(\Pi_{X^n})\big)}{\mathbb{P}_{0, \sigma}\big(T(\Pi_{X^n})\big)}. \tag{68}$$

Let $\mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_\theta, \epsilon) = \{\mu \in \mathcal{P}_{\mathcal{X}} : \sup_{x \in \mathcal{X}} |\mu(x) - \boldsymbol{\alpha}^T \boldsymbol{P}_\theta(x)| \leq \epsilon\}$ denote the ball centered at $\boldsymbol{\alpha}^T \boldsymbol{P}_\theta$ with radius $\epsilon > 0$. According to the Glivenko–Cantelli theorem [57], we then have that for any $\sigma \in \mathcal{S}_{n, \lambda}$ and $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}_{\theta, \sigma}\Big\{\sup_{x \in \mathcal{X}} |\Pi_{X^n}(x) - \boldsymbol{\alpha}^T \boldsymbol{P}_\theta(x)| > \epsilon\Big\} = 0. \tag{69}$$

It then follows that for any $\sigma \in \mathcal{S}_{n, \lambda}$ and $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}_{\theta, \sigma}\Big\{\Pi_{X^n} \notin \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_\theta, \epsilon)\Big\}$$

$$= \lim_{n \to \infty} \mathbb{P}_{\theta, \sigma}\Big\{\sup_{x \in \mathcal{X}} |\Pi_{X^n}(x) - \boldsymbol{\alpha}^T \boldsymbol{P}_\theta(x)| > \epsilon\Big\} = 0. \tag{70}$$

It was shown in Lemma 5.3 in [2] that $f_{\boldsymbol{P}_\theta}(\boldsymbol{\alpha}, P)$ is a continuous function of P for any $\theta \in \{0, 1\}$. Therefore, $f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, P) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, P)$ is a continuous function of P. Then we have that for any $\epsilon > 0$, there exists an $\eta(\epsilon) > 0$ such that $\forall P \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)$,

$$f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1) - \eta(\epsilon) < f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, P) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, P)$$

$$< f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1) + \eta(\epsilon), \tag{71}$$

where $\eta(\epsilon) \to 0$ as $\epsilon \to 0$. We then have that

$$\lim_{n \to \infty} \frac{1}{n} D\Big(\widetilde{\mathbb{P}}_1 \| \widetilde{\mathbb{P}}_0\Big)$$

$$= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}_{\widetilde{\mathbb{P}}_1}\Big[\log \mathbb{P}_{1, \sigma}\big(T(\Pi_{X^n})\big) - \log \mathbb{P}_{0, \sigma}\big(T(\Pi_{X^n})\big)\Big]$$

$$\overset{(a)}{\leq} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}_{\widetilde{\mathbb{P}}_1}\left[\log\left(\prod_k |\mathcal{P}_{n_k}|\right) - \log\left(\prod_{k=1}^{K} \frac{1}{(n_k + 1)^{|\mathcal{X}|}}\right)\right.$$

$$- \inf_{\substack{(U_1, \ldots, U_K) \in \mathcal{P}_{n_1} \times \ldots \times \mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T \boldsymbol{U} = \Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k \| p_{1, k})$$

$$\left. + \inf_{\substack{(U_1, \ldots, U_K) \in \mathcal{P}_{n_1} \times \ldots \times \mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T \boldsymbol{U} = \Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k \| p_{0, k})\right]$$

$$= \lim_{n \to \infty} \frac{1}{n} \widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\big) \mathbb{E}_{\widetilde{\mathbb{P}}_1}\left[\log\left(\prod_k |\mathcal{P}_{n_k}|\right)\right.$$

$$- \inf_{\substack{(U_1, \ldots, U_K) \in \mathcal{P}_{n_1} \times \ldots \times \mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T \boldsymbol{U} = \Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k \| p_{1, k})$$

$$+ \inf_{\substack{(U_1, \ldots, U_K) \in \mathcal{P}_{n_1} \times \ldots \times \mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T \boldsymbol{U} = \Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k \| p_{0, k})$$

$$\left. - \log\left(\prod_{k=1}^{K} \frac{1}{(n_k + 1)^{|\mathcal{X}|}}\right)\Big| \Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\right]$$

$$+ \lim_{n \to \infty} \frac{1}{n} \widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \notin \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\big) \mathbb{E}_{\widetilde{\mathbb{P}}_1}\left[\log\left(\prod_k |\mathcal{P}_{n_k}|\right)\right.$$

$$- \inf_{\substack{(U_1, \ldots, U_K) \in \mathcal{P}_{n_1} \times \ldots \times \mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T \boldsymbol{U} = \Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k \| p_{1, k})$$

$$+ \inf_{\substack{(U_1, \ldots, U_K) \in \mathcal{P}_{n_1} \times \ldots \times \mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T \boldsymbol{U} = \Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k \| p_{0, k})$$

$$\left. - \log\left(\prod_{k=1}^{K} \frac{1}{(n_k + 1)^{|\mathcal{X}|}}\right)\Big| \Pi_{X^n} \notin \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\right]$$

$$\overset{(b)}{=} \lim_{n \to \infty} \frac{1}{n} \widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\big) \mathbb{E}_{\widetilde{\mathbb{P}}_1}\Bigg[$$

$$- \inf_{\substack{(U_1,...,U_K)\in\mathcal{P}_{n_1}\times...\times\mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k||p_{1,k})$$

$$+ \inf_{\substack{(U_1,...,U_K)\in\mathcal{P}_{n_1}\times...\times\mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k||p_{0,k})$$

$$\left| \Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T\boldsymbol{P}_1,\epsilon) \right]$$

$$+ \lim_{n\to\infty} \frac{1}{n} \widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T\boldsymbol{P}_1,\epsilon)\big) \bigg(\log\Big(\prod_k |\mathcal{P}_{n_k}|\Big)$$

$$- \log\Big(\prod_{k=1}^{K} \frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big)\bigg)$$

$$\overset{(c)}{\le} f_{\boldsymbol{P}_0}(\boldsymbol{\alpha},\boldsymbol{\alpha}^T\boldsymbol{P}_1) + \eta(\epsilon), \tag{72}$$

where the inequality (a) is due to the bound of the probability of type classes [54]: $\frac{1}{(n_k+1)^{|\mathcal{X}|}} 2^{-n_k D(U_k||p_{\theta,k})} \le p_{\theta;k}^{\otimes n_k}\big(T_{n_k}(U_k)\big) \le 2^{-n_k D(U_k||p_{\theta,k})}$, the equality (b) is due to the fact that $\lim_{n\to\infty} \widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \notin \mathcal{B}(\boldsymbol{\alpha}^T\boldsymbol{P}_1,\epsilon)\big) = 0$ and the inequality (c) is due to (71) and the fact that $\lim_{n\to\infty} \frac{1}{n} \widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T\boldsymbol{P}_1,\epsilon)\big) \Big(\log\Big(\prod_k |\mathcal{P}_{n_k}|\Big) - \log\Big(\prod_{k=1}^{K} \frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big)\Big) = 0$.

For the lower bound, following the same idea as in (72), we have that

$$\lim_{n\to\infty} \frac{1}{n} D\big(\widetilde{\mathbb{P}}_1||\widetilde{\mathbb{P}}_0\big)$$

$$\ge \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\widetilde{\mathbb{P}}_1} \Bigg[\log\Big(\prod_{k=1}^{K} \frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big) - \log\Big(\prod_k |\mathcal{P}_{n_k}|\Big)$$

$$- \inf_{\substack{(U_1,...,U_K)\in\mathcal{P}_{n_1}\times...\times\mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k||p_{1,k})$$

$$+ \inf_{\substack{(U_1,...,U_K)\in\mathcal{P}_{n_1}\times...\times\mathcal{P}_{n_K} \\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^{K} n_k D(U_k||p_{0,k}) \Bigg]$$

$$\ge f_{\boldsymbol{P}_0}(\boldsymbol{\alpha},\boldsymbol{\alpha}^T\boldsymbol{P}_1) - \eta(\epsilon). \tag{73}$$

By (72) and (73), we have that

$$\lim_{n\to\infty} \frac{1}{n} D\big(\widetilde{\mathbb{P}}_1||\widetilde{\mathbb{P}}_0\big) = f_{\boldsymbol{P}_0}(\boldsymbol{\alpha},\boldsymbol{\alpha}^T\boldsymbol{P}_1). \tag{74}$$

REFERENCES

[1] Z. Sun, S. Zou, and Q. Li, "Quickest change detection in anonymous heterogeneous sensor networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 5925–5929, 2020.

[2] W. N. Chen and I. H. Wang, "Anonymous heterogeneous distributed detection: Optimal decision rules, error exponents, and the price of anonymity," *IEEE Trans. Inform. Theory*, vol. 65, no. 11, pp. 7390–7406, 2019.

[3] V. V. Veeravalli and T. Banerjee, "Quickest change detection," *Academic press library in signal processing: Array and statistical signal processing*, vol. 3, pp. 209–256, 2013.

[4] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[5] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge University Press, 2009.

[6] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, 1993.

[7] D. O. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. Springer Series in Statistics, Springer, 1985.

[8] A. G. Tartakovsky, *Sequential Change Detection and Hypothesis Testing: General Non-iid Stochastic Models and Asymptotically Optimal Rules*. CRC Press, 2019.

[9] L. Xie, S. Zou, Y. Xie, and V. V. Veeravalli, "Sequential (quickest) change detection: Classical results and new directions," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 2, pp. 494–514, 2021.

[10] A. G. Tartakovsky and V. V. Veeravalli, "Change-point detection in multichannel and distributed systems," *Applied Sequential Methodologies: Real-World Examples with Data Analysis*, vol. 173, pp. 339–370, 2004.

[11] A. G. Tartakovsky, B. L. Rozovskii, R. B. 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 Proc.*, vol. 54, no. 9, pp. 3372–3382, 2006.

[12] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.

[13] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *Ann. Statist.*, pp. 670–692, 2013.

[14] G. Fellouris and G. Sokolov, "Second-order asymptotic optimality in multisensor sequential change detection," *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 3662–3675, 2016.

[15] V. Raghavan and V. V. Veeravalli, "Quickest change detection of a Markov process across a sensor array," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1961–1981, 2010.

[16] T. He, S. Ben-David, and L. Tong, "Nonparametric change detection and estimation in large-scale sensor networks," *IEEE Trans. Signal Proc.*, vol. 54, no. 4, pp. 1204–1217, 2006.

[17] D. Ciuonzo and P. S. Rossi, "Dechade: Detecting slight changes with hard decisions in wireless sensor networks," *International Journal of General Systems*, vol. 47, no. 5, pp. 535–548, 2018.

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

[19] M. Ludkovski, "Bayesian quickest detection in sensor arrays," *Sequential Analysis*, vol. 31, no. 4, pp. 481–504, 2012.

[20] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley, "Quickest detection of dynamic events in networks," *IEEE Trans. Inform. Theory*, vol. 66, no. 4, pp. 2280–2295, 2020.

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

[22] 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, 2008.

[23] S. Zou, V. V. Veeravalli, J. Li, D. Towsley, and A. Swami, "Distributed quickest detection of significant events in networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 8454–8458, 2019.

[24] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[25] B. B. P. Rao, P. Saluia, N. Sharma, A. Mittal, and S. V. Sharma, "Cloud computing for internet of things amp; sensing based applications," in *2012 Sixth International Conference on Sensing Technology (ICST)*, pp. 374–380, 2012.

[26] S. Li, L. D. Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and internet of things," *IEEE Trans. Industr. Inform.*, vol. 9, no. 4, pp. 2177–2186, 2013.

[27] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 4, pp. 2347–2376, 2015.

[28] M.-C. Yuen, I. King, and K.-S. Leung, "A survey of crowdsourcing systems," in *IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing*, pp. 766–773, 2011.

[29] E. Estellés-Arolas and F. G.-L. de Guevara, "Towards an integrated crowdsourcing definition," *Journal of Information Science*, vol. 38, no. 2, pp. 189–200, 2012.

[30] D. Brabham, *Crowdsourcing*. The MIT Press Essential Knowledge series, MIT Press, 2013.

[31] B. Zhou, J. Pei, and W. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," *ACM SIGKDD Explorations Newsletter*, vol. 10, no. 2, pp. 12–22, 2008.

[32] R. Cummings, S. Krehbiel, Y. Mei, R. Tuo, and W. Zhang, "Differentially private change-point detection," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, pp. 10825–10834, 2018.

[33] C. Dwork, "Differential privacy: A survey of results," in *Proc. International Conference on Theory and Applications of Models of Computation*, pp. 1–19, Springer, 2008.

[34] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.

[35] J. Horey, M. M. Groat, S. Forrest, and F. Esponda, "Anonymous data collection in sensor networks," in *Proc. Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services (MobiQuitous)*, pp. 1–8, 2007.

[36] M. M. Groat, W. Hey, and S. Forrest, "KIPDA: k-indistinguishable privacy-preserving data aggregation in wireless sensor networks," in *Proc. IEEE INFOCOM*, pp. 2024–2032, 2011.

[37] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "Statistical framework for source anonymity in sensor networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–6, 2010.

[38] A. Wadaa, S. Olariu, L. Wilson, M. Eltoweissy, and K. Jones, "On providing anonymity in wireless sensor networks," in *Proc. International Conference on Parallel and Distributed Systems*, pp. 411–418, 2004.

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

[40] D. Siegmund and E. S. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *Ann. Statist.*, pp. 255–271, 1995.

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

[42] T. Banerjee and V. V. Veeravalli, "Data-efficient minimax quickest change detection with composite post-change distribution," *IEEE Trans. Inform. Theory*, vol. 61, no. 9, pp. 5172–5184, 2015.

[43] S. Zou, G. Fellouris, and V. V. Veeravalli, "Quickest change detection under transient dynamics: Theory and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 65, no. 3, pp. 1397–1412, 2018.

[44] Y. Mei, "Sequential change-point detection when unknown parameters are present in the pre-change distribution," *Ann. Statist.*, vol. 34, no. 1, pp. 92–122, 2006.

[45] B. Brodsky and B. Darkhovsky, "Asymptotically optimal sequential change-point detection under composite hypotheses," in *Proc. IEEE Conference on Decision and Control*, pp. 7347–7351, 2005.

[46] S. Pergamenchtchikov and A. G. Tartakovsky, "Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis," *Journal of Multivariate Analysis*, vol. 174, p. 104541, 2019.

[47] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Sequential algorithms for moving anomaly detection in networks," *Sequential Analysis*, vol. 39, no. 1, pp. 6–31, 2020.

[48] G. Rovatsos, G. V. Moustakides, and V. V. Veeravalli, "Quickest detection of moving anomalies in sensor networks," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 2, pp. 762–773, 2021.

[49] C.-D. Fuh and A. G. Tartakovsky, "Asymptotic Bayesian theory of quickest change detection for hidden Markov models," *IEEE Trans. Inform. Theory*, vol. 65, no. 1, pp. 511–529, 2018.

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

[51] P. Moulin and V. V. Veeravalli, *Statistical Inference for Engineers and Data Scientists*. Cambridge University Press, 2018.

[52] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, pp. 1379–1387, Dec. 1986.

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

[54] T. M. Cover, *Elements of Information Theory*. John Wiley & Sons, 2006.

[55] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1." http://cvxr.com/cvx, Mar. 2014.

[56] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control* (V. D. Blondel, S. P. Boyd, and H. Kimura, eds.), pp. 95–110, Springer-Verlag, 2008.

[57] H. G. Tucker, "A generalization of the Glivenko-Cantelli theorem," *Ann. Math. Statist.*, vol. 30, pp. 828–830, 09 1959.

[58] S. Asmussen, *Applied Probability and Queues*, vol. 51. Springer Science & Business Media, 2008.

[59] X. Cheng, D. Ciuonzo, and P. S. Rossi, "Multibit decentralized detection through fusing smart and dumb sensors based on rao test," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 2, pp. 1391–1405, 2020.

Zhongchang Sun (S'20) is a PhD student at the Department of Electrical Engineering, University at Buffalo, the State University of New York. He received the B.S. degree from Beijing Institute of Technology, Beijing, China in 2019. His research interests are on hypothesis testing, quickest change detection and distributionally robust optimization.

Shaofeng Zou (S'14-M'16) is an Assistant Professor, at the Department of Electrical Engineering, University at Buffalo, the State University of New York. He received the Ph.D. degree in Electrical and Computer Engineering from Syracuse University in 2016. He received the B.E. degree (with honors) from Shanghai Jiao Tong University, Shanghai, China, in 2011. He was a postdoctoral research associate at the Coordinated Science Lab, University of Illinois at Urbana-Champaign during 2016-2018. Dr. Zou's research interests include reinforcement learning, machine learning, statistical signal processing and information theory. He received the National Science Foundation CRII award in 2019.
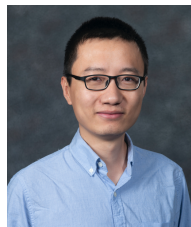
Ruizhi Zhang is an Assistant Professor in the Department of Statistics at University of Nebraska-Lincoln. He received his B.S. degree in Mathematics from Hua Loo-Keng Talent Program in Mathematics at University of Science and Technology of China (USTC) in 2014, graduated with honors. He received his Ph.D. degree in Statistics in the School of Industrial and Systems Engineering at Georgia Institute of Technology. His research interests include change-point detection, sequential analysis, robust statistics, high-dimensional statistical inference, functional data analysis.
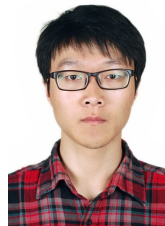
Qunwei Li (S'16) received the B.S. and M.S. degrees in electrical engineering with honors from Xidian University, Xi'an, China, in 2011 and 2014. He received the Ph.D. degree in electrical engineering from Syracuse University, Syracuse, NY, USA, in 2018. He was a postdoctoral researcher in the Center for Applied Scientific Computing (CASC) at Lawrence Livermore National Laboratory (LLNL), CA, USA, from 2018 to 2019. His research interests include human decision making, adversarial deep learning, optimization algorithms, and recommender systems. Dr. Li received the Syracuse University Graduate Fellowship Award in 2014 and the All University Doctoral Prize 2018 by Syracuse University for superior achievement in completed dissertations.