QUICKEST CHANGE DETECTION IN ANONYMOUS HETEROGENEOUS SENSOR NETWORKS

Zhongchang Sun[†] *Shaofeng Zou*[†] *Qunwei Li*[*]

[†]University at Buffalo, the State University of New York
[*]Ant Financial, China
Email:zhongcha@buffalo.edu szou3@buffalo.edu qunwei.qw@antfin.com

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 in this paper, where at each time step, the fusion center receives n unordered samples. The fusion center does not know which sensor each sample comes from, and thus does not know its exact distribution. In this paper, a simple optimality proof is derived for the Mixture Likelihood Ratio Test (MLRT), which was constructed and proved to be optimal for the non-sequential anonymous setting in [1]. For the QCD problem, a mixture CuSum algorithm is constructed in this paper, and is further shown to be optimal under Lorden's criterion [2].

***Index Terms*—** Anonymous, hypothesis testing, heterogeneous, mixture CuSum, sequential change detection.

1. INTRODUCTION

Suppose a network consists of n sensors and a fusion center. At some unknown time, an event occurs in the network, and causes a change in the data generating distribution of the sensors. The goal is to detect the change as quickly as possible subject to false alarm constraints. We consider a general setting with heterogeneous sensors, where the sensors can be clustered into K groups, and different groups follow different data generating distributions. It is assumed that the number of sensors and the data generating distributions in each group are known. In this paper, we investigate the scenario where the sensors are anonymous. Specifically, the fusion center does not know which sensor each sample comes from (see e.g., [3,4] for anonymous data collection approaches). The anonymous and heterogeneous setting finds a wide range of applications in sensor networks in social settings [5], where human participants are involved, and thus privacy and anonymity are required to protect the participants.

The quickest change detection (QCD) problem in sensor networks has been widely studied in the literature [6–14]. In these papers, one CuSum algorithm can be implemented at each sensor, and be further combined to design algorithms with certain optimality guarantee. In this paper, we are interested in the anonymous setting, where at each time step, the fusion center receives n arbitrarily permuted observations, and the permutations at different time steps may be different. Therefore, the fusion center does not know which samples over time come from one particular sensor. Existing approaches are not applicable since the fusion center is not able to compute one CuSum statistic for each sensor.

In this paper, we first revisit the non-sequential setting with anonymous heterogeneous sensors in [1], where one sample is collected from each sensor. In [1], a Mixture Likelihood Ratio Test (MLRT) was developed for this composite hypothesis testing problem, where the group assignment is the unknown parameter. The MLRT was shown to be optimal under the Neyman-Pearson setting [1]. In this paper, we provide a simple proof for the optimality of the MLRT. The basic idea is to construct a composite binary hypothesis testing problem with uniform priors on all possible group assignments of the samples, and to show that the optimal test for the case with Bayesian priors is also optimal under the Neyman-Pearson setting.

We further study the QCD problem in anonymous heterogeneous sensor networks, and design a mixture CuSum algorithm. The basic idea is design a CuSum type algorithm using the mixture likelihood ratio. We show that the mixture CuSum algorithm is optimal under Lorden's criterion [2]. We also provide numerical results to demonstrate that our mixture algorithm outperforms two other algorithms using the Bayesian approach and the generalized likelihood ratio approach for the unknown group assignments.

2. PROBLEM STATEMENT

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 $p_{\theta,k}$, $\theta \in \{0,1\}$. The centralized setting is considered, where there is a fusion center. The sensors are anonymous, i.e., the fusion center doesn't know which group 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.

2.1. Binary Composite Hypothesis Testing

We first revisit the binary hypothesis testing problem in [1], where one sample is collected from each sensor. 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 labelings by $\mathcal{S}_{n,\lambda}$, where $\lambda = \{n_1, \ldots, n_K\}$.

The problem is a composite hypothesis testing problem, where σ is the unknown parameter for both $\theta = 0$ and 1:

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

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

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

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

The Neyman-Pearson setting is studied, where the goal is to solve the following problem for any $\alpha \in [0, 1]$:

$$\inf_{\phi:P_F(\phi)\le\alpha} P_M(\phi). \tag{4}$$

2.2. Quickest Change Detection

In the QCD setting, unordered 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 observed samples at time t by $X^n[t]$. Before the change, i.e., $t < \nu$, $X^n[t] \sim \mathbb{P}_{0,\sigma_t} \triangleq \prod_{i=1}^n p_{0,\sigma_t(i)}$, for some unknown $\sigma_t \in \mathcal{S}_{n,\lambda}$. After the change, i.e., $t \ge \nu$, $X^n[t] \sim \mathbb{P}_{1,\sigma_t} \triangleq \prod_{i=1}^n p_{1,\sigma_t(i)}$, 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 $t_1 \ne t_2$. We further assume that $X^n[t_1]$ is independent from $X^n[t_2]$ for any $t_1 \ne t_2$.

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, and we define the worst-case average detection delay (WADD) under Lorden's criterion [2] and worst-case average run length (WARL) for any stopping time τ as follows:

$$\text{WADD}(\tau) \triangleq \sup_{\nu\ge1} \sup_{\Omega} \operatorname{ess\,sup} \mathbb{E}_\Omega^\nu \left[(\tau-\nu)^+ | \mathbf{X}^n[1,\nu-1]\right],$$

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

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 i are labeled according to σ_i, and $\mathbf{X}^n[1, \nu-1] = \{X^n[1], \ldots, X^n[\nu-1]\}$.

Denote by $\mathcal{F}_t$ the σ-algebra generated by the observations of all the nodes up to time t, for $t \in \mathbb{N}$. The goal is to design a stopping rule that minimizes the WADD subject to a constraint on the WARL:

$$\inf_{\tau:\text{WARL}\ge\gamma} \text{WADD}(\tau). \tag{6}$$

3. MIXTURE LIKELIHOOD RATIO TEST AND A SIMPLE PROOF

For the binary composite hypothesis testing problem in Section 2.1, Chen and Huang constructed an MLRT, and showed that the MLRT is optimal [1]. In this section, we will first briefly review the optimality proof in [1], and then we will present a simple version of the proof.

3.1. Mixture Likelihood Ratio Test

Recall the mixture likelihood ratio $\ell(x^n)$ in [1]:

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

Then the MLRT was defined in [1] 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. *[1, Theorem 3.1] Consider the binary composite hypothesis testing problem under the Neyman-Pearson setting in Section 2.1. The MLRT ϕ^* is optimal.*

The key idea of the proof in [1] is to reduce the original composite hypothesis testing problem in Section 2.1 into a simple one through the ordering map $\Pi(x^n)$, and then apply 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 $x_{i_1} \ge x_{i_2} \ge \ldots \ge x_{i_n}$. In the proof, due to the introduction of the ordering map, a careful examination of the measurability needs to be conducted. The proof in [1] 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 ϕ. 6) Then, the MLRT test is optimal among all tests.

3.2. A Simpler Proof for Lemma 1

In this section, 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 uniform priors on all $\sigma \in \mathcal{S}_{n,\lambda}$ under both hypotheses, and define the average type-I and type-II error probabilities for any test ϕ:

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

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

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

$$\mathcal{H}_0 : \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} P_{0,\sigma}, \quad (11)$$

$$\mathcal{H}_1 : \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma \in \mathcal{S}_{n,\lambda}} P_{1,\sigma}, \quad (12)$$

for which the optimal test (the same as the MLRT) is the likelihood ratio test between (11) and (12) [15].

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 $\sigma' = \sigma \circ \pi$. 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 $\theta = 0, 1$,

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

It then follows that

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

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

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

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

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

4. MIXTURE-CUSUM ALGORITHM FOR THE QCD PROBLEM

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

$$T^*(b) = \inf\{t : \max_{1 \le k \le t} \sum_{i=k}^{t} \log \ell(x^n[i]) \ge b\}. \quad (16)$$

We show that the mixture CuSum algorithm is exactly optimal under Lorden's criterion [2].

Theorem 1. *Consider the QCD problem in Section 2.2, the mixture CuSum algorithm in* (16) *is exactly optimal.*

Proof. We follow a similar idea to the one in Section 3.2. 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:

$$\begin{aligned} \widetilde{\text{WADD}}(\tau) &= \sup_{\nu \ge 1} \operatorname{ess\,sup} \widetilde{\mathbb{E}}^{\nu}[(\tau - \nu)^+ | \widetilde{\mathbf{X}}^n[1, \nu - 1]], \\ \widetilde{\text{ARL}}(\tau) &= \widetilde{\mathbb{E}}^{\infty}[\tau], \quad (17) \end{aligned}$$

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 $\widetilde{\mathbf{X}}^n[1, \nu - 1]$ are i.i.d. from $\widetilde{\mathbb{P}}_0$. It was shown that the CuSum algorithm is optimal under Lorden's criterion [16]. Therefore, T^* in (16) is 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 in Section 3.2, we can show that for any stopping time τ,

$$\widetilde{\text{WADD}}(\tau) \le \text{WADD}(\tau) \text{ and } \widetilde{\text{ARL}}(\tau) \ge \text{WARL}(\tau). \quad (18)$$

We will then show that T^* achieves the equality in (18), which will complete the proof. Due to the fact that T^* 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

$$\begin{aligned} &\operatorname{ess\,sup} \mathbb{E}_{\Omega}^{\nu}[(T^* - \nu)^+ | \mathbf{X}^n[1, \nu - 1]] \\ &\qquad = \operatorname{ess\,sup} \mathbb{E}_{\Omega'}^{\nu}[(T^* - \nu)^+ | \mathbf{X}^n[1, \nu - 1]], \\ &\mathbb{E}_{\Omega}^{\infty}[T^*] = \mathbb{E}_{\Omega'}^{\infty}[T^*] \end{aligned} \tag{19}$$

We note that to establish (18) and the optimality of T^*, the proof is more involved than the binary hypothesis testing case in Section 3.2 because of the esssup and the conditional expectation. The details can be found in Appendix. □

5. NUMERICAL RESULTS

In this section, we provide some numerical results. We compare our optimal mixture CuSum test with two other algorithms based on the Bayesian approach and the generalized likelihood ratio approach of tackling the unknown group assignments. For the Bayesian approach of tackling the group assignments, 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 have exact n_k sensors in each group k. We then compute the following likelihood ratio:

$$l_b(x^n[t]) = \frac{\prod_{i=1}^{n} \left(\sum_{k=1}^{K} \frac{n_k}{n} P_{1,k}(x_i[t]) \right)}{\prod_{i=1}^{n} \left(\sum_{k=1}^{K} \frac{n_k}{n} P_{0,k}(x_i[t]) \right)}. \tag{20}$$

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

$$l_g(x^n[t]) = \frac{\sup_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(x^n[t])}{\sup_{\sigma \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(x^n[t])}. \tag{21}$$

We then design CuSum-type tests using (20) and (21), which are referred to as the Bayesian and Generalized CuSums.

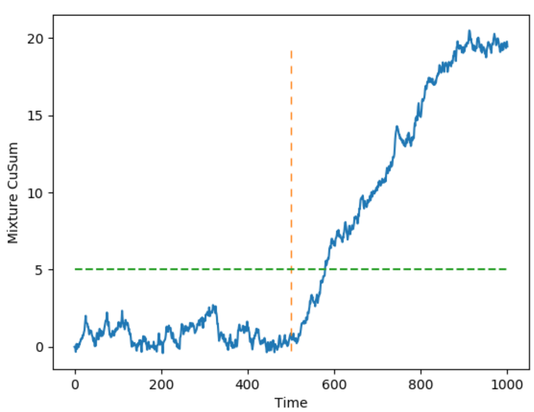

Fig. 1. Evolution Path of The Mixture CuSum

We 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. 1, we set the change point to be 500. We plot one sample evolution path of the mixture CuSum algorithm. It can be seen that before change point, the test statistic fluctuates around zero, and after change point, it starts to increase with a positive drift.

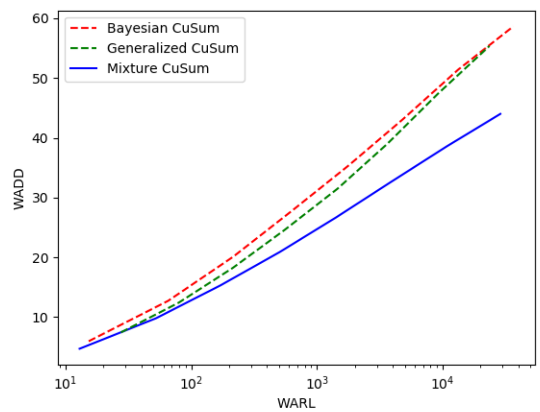

Fig. 2. Comparison of The Three Algorithms.

In Fig. 2, we plot the WADD as a function of the WARL. We repeat the experiment for 10000 times and take the average. It can be seen from Fig. 2 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 $\mathbb{P}_{0,\sigma}$ for some $\sigma \in \mathcal{S}_{n,\lambda}$.

6. CONCLUSION

In this paper, we studied the hypothesis testing problem in anonymous heterogeneous sensor networks. We first revisited the non-sequential setting studied in [1], 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 [2]. We note that asymptotic optimality results can also be obtained under Pollak's criterion [17]. Our results demonstrated that exact knowledge of numbers of sensors in each group leads to a better performance.

Although being optimal, our mixture CuSum algorithm needs to compute the average of the likelihood over all possible group assignments, and thus is computationally expensive when the number of sensors is large. It is of future interest to design computationally efficient algorithms for large networks. Moreover, it is assumed that after the change all the sensors change their data generating distributions. It is also of interest to investigate the case where only an unknown subset of sensors change their data generating distributions.

7. REFERENCES

[1] 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*, 2019.

[2] G. Lorden et al., "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.

[3] 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)*. IEEE, 2007, pp. 1–8.

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

[5] 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.

[6] 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.

[7] 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.

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

[9] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *The Annals of Statistics*, pp. 670–692, 2013.

[10] 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.

[11] 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.

[12] 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.

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

[14] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley, "Quickest detection of dynamic events in networks," *accepted, IEEE Trans. Inform. Theory*, 2019.

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

[16] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.

[17] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, pp. 206–227, 1985.

Appendix

A. PROOF OF (18)

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,

$$\widehat{X}^n[t] \sim \begin{cases} \widetilde{\mathbb{P}}_0, & \text{if } t < \nu \\ \mathbb{P}_{1,\sigma_t}, & \text{if } t \geq \nu. \end{cases} \tag{22}$$

For any stopping time τ, define the worst-case average detection delay for the model in (22) as follows:

$$\begin{aligned}&\widehat{\mathrm{WADD}}(\tau)\\ &= \sup_{\nu\geq 1}\sup_{\sigma_\nu,\ldots,\sigma_\infty}\operatorname{ess\,sup}\widehat{\mathbb{E}}^{\nu}_{\sigma_\nu,\ldots,\sigma_\infty}[(\tau-\nu)^+|\widehat{\mathbf{X}}^n[1,\nu-1]],\end{aligned} \tag{23}$$

where $\widehat{\mathbb{E}}^{\nu}_{\sigma_\nu,\ldots,\sigma_\infty}$ denotes the expectation when the data is distributed according to (22). To prove that $\mathrm{WADD}(\tau) \geq \widetilde{\mathrm{WADD}}(\tau)$, we will first show that $\mathrm{WADD}(\tau) = \widehat{\mathrm{WADD}}(\tau)$, and then show that $\widehat{\mathrm{WADD}}(\tau) \geq \widetilde{\mathrm{WADD}}(\tau)$.

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

$$\begin{aligned}&\mathrm{WADD}(\tau)\\ &= \sup_{\nu\geq 1}\sup_{\Omega}\operatorname{ess\,sup}\mathbb{E}^{\nu}_{\Omega}[(\tau-\nu)^+|\mathbf{X}^n[1,\nu-1]]\\ &= \sup_{\nu\geq 1}\sup_{\omega\in\mathcal{N}}\sup_{\mu\in\mathcal{M}}\operatorname{ess\,sup}\mathbb{E}^{\nu}_{\omega}[(\tau-\nu)^+|\mathbf{X}^n[1,\nu-1]],\end{aligned} \tag{24}$$

where $\mathbb{E}^{\nu}_{\omega}$ 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 $\widehat{\mathrm{WADD}}$ 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 $\mathbb{E}^{\nu}_{\omega}[(\tau-\nu)^+|\mathbf{X}^n[1,\nu-1]]$. Then, WADD and $\widehat{\mathrm{WADD}}$ can be written as

$$\begin{aligned}\mathrm{WADD}(\tau) &= \sup_{\nu\geq 1}\sup_{\omega\in\mathcal{N}}\sup_{\mu\in\mathcal{M}}\operatorname{ess\,sup} f_\omega(\mathbf{X}^n[1,\nu-1]),\\ \widehat{\mathrm{WADD}}(\tau) &= \sup_{\nu\geq 1}\sup_{\omega\in\mathcal{N}}\operatorname{ess\,sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]).\end{aligned} \tag{25}$$

It then suffices to show that

$$\sup_{\mu\in\mathcal{M}}\operatorname{ess\,sup} f_\omega(\mathbf{X}^n[1,\nu-1]) = \operatorname{ess\,sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]). \tag{26}$$

For any $\omega \in \mathcal{N}$ and $\mu \in \mathcal{M}$, let

$$\begin{aligned} b_{\omega,\mu} &= \operatorname{ess\,sup} f_\omega(\mathbf{X}^n[1,\nu-1]) \\ &= \inf\{b : \mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b) = 0\}, \end{aligned} \tag{27}$$

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

$$\begin{aligned} b^*_\omega &= \inf\Big\{b : \int_{\mathbf{x}^n[1,\nu-1]} \mathbb{1}_{\{f_\omega(\mathbf{x}^n[1,\nu-1])>b\}} \\ &\quad \times \mathrm{d}\prod_{t=1}^{\nu-1} \widetilde{\mathbb{P}}_0(x^n(t)) = 0\Big\} \\ &= \inf\Big\{b : \int_{\mathbf{x}^n[1,\nu-1]} \mathbb{1}_{\{f_\omega(\mathbf{x}^n[1,\nu-1])>b\}} \\ &\quad \times \mathrm{d}\prod_{t=1}^{\nu-1} \frac{1}{|\mathcal{S}_{n,\lambda}|} \sum_{\sigma_t \in \mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma_t}(x^n(t)) = 0\Big\} \\ &= \inf\Big\{b : \int_{\mathbf{x}^n[1,\nu-1]} \mathbb{1}_{\{f_\omega(\mathbf{x}^n[1,\nu-1])>b\}} \\ &\quad \times \mathrm{d}\frac{1}{|\mathcal{M}|} \sum_{\mu\in\mathcal{M}} \mathbb{P}_\mu(\mathbf{x}^n[1,\nu-1]) = 0\Big\} \\ &= \inf\Big\{b : \frac{1}{|\mathcal{M}|} \sum_{\mu\in\mathcal{M}} \mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b) = 0\Big\}. \end{aligned} \tag{28}$$

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

$$\mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b^*_\omega) = 0. \tag{29}$$

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

$$\sup_{\mu\in\mathcal{M}} b_{\omega,\mu} \le b^*_\omega. \tag{30}$$

Conversely, let $\sup_{\mu\in\mathcal{M}} b_{\omega,\mu} = b'$. For any $\mu \in \mathcal{M}$, we have that

$$\mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b') = 0. \tag{31}$$

Then,

$$\frac{1}{|\mathcal{M}|} \sum_{\mu\in\mathcal{M}} \mathbb{P}_\mu(f_\omega(\mathbf{X}^n[1,\nu-1]) > b') = 0. \tag{32}$$

This further implies that

$$b^*_\omega \le b' = \sup_{\mu\in\mathcal{M}} b_{\omega,\mu}. \tag{33}$$

Combining (30) and (33), we have that

$$\sup_{\mu\in\mathcal{M}} b_{\omega,\mu} = b^*_\omega, \tag{34}$$

and thus

$$\sup_{\mu\in\mathcal{M}} \operatorname{ess\,sup} f_\omega(\mathbf{X}^n[1,\nu-1]) = \operatorname{ess\,sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]). \tag{35}$$

This implies that

$$\mathrm{WADD}(\tau) = \widehat{\mathrm{WADD}}(\tau). \tag{36}$$

Step 2. The next step is to show that $\widehat{\mathrm{WADD}}(\tau) \ge \widetilde{\mathrm{WADD}}(\tau)$. We will first show that

$$\begin{aligned} &\sup_{\omega\in\mathcal{N}} \operatorname{ess\,sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \\ &\ge \operatorname{ess\,sup} \sup_{\omega\in\mathcal{N}} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]). \end{aligned} \tag{37}$$

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 $\hat{b} = \sup_{\omega\in\mathcal{N}} \operatorname{ess\,sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1])$. For any $\omega \in \mathcal{N}$, we have that

$$\widetilde{\mathbb{P}}^\nu\Big(f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \ge \hat{b}\Big) = 0. \tag{38}$$

Since $\mathcal{N}$ is countable, it then follows that

$$\begin{aligned} &\widetilde{\mathbb{P}}^\nu\Big(\sup_{\omega\in\mathcal{N}} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \ge \hat{b}\Big) \\ &\le \widetilde{\mathbb{P}}^\nu\Big(\cup_{\omega\in\mathcal{N}} \big\{f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) > \hat{b}\big\}\Big) \\ &\le \sum_{\omega\in\mathcal{N}} \widetilde{\mathbb{P}}^\nu\Big(f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) > \hat{b}\Big) = 0. \end{aligned} \tag{39}$$

Therefore,

$$\begin{aligned} \hat{b} &= \sup_{\omega\in\mathcal{N}} \operatorname{ess\,sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \\ &\ge \operatorname{ess\,sup} \sup_{\omega\in\mathcal{N}} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]). \end{aligned} \tag{40}$$

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

$$\begin{aligned} &\sup_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda}^{\otimes(T-\nu+1)}}} \sum_{t=\nu+1}^{T} (t-\nu)\mathbb{P}_{\sigma_\nu,...,\sigma_T}(\tau = t|\widehat{\mathbf{X}}^n[1,\nu-1]) \\ &\ge \sum_{t=\nu+1}^{T} (t-\nu) \frac{1}{|\mathcal{S}_{n,\lambda}|^{(T-\nu+1)}} \\ &\quad \times \sum_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda}^{\otimes(T-\nu+1)}}} \mathbb{P}_{\sigma_\nu,...,\sigma_T}(\tau = t|\widehat{\mathbf{X}}^n[1,\nu-1]) \\ &= \sum_{t=\nu+1}^{T} (t-\nu)\widetilde{\mathbb{P}}^\nu(\tau = t|\widetilde{\mathbf{X}}^n[1,\nu-1]). \end{aligned} \tag{41}$$

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

$$\widehat{\mathbb{E}}_\omega^\nu[(\tau-\nu)|\widehat{\mathbf{X}}^n[1,\nu-1]] \geq \widetilde{\mathbb{E}}^\nu[(\tau-\nu)|\widetilde{\mathbf{X}}^n[1,\nu-1]], \tag{42}$$

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 (40) and (42), we have that

$$\begin{aligned}
&\widehat{\text{WADD}}(\tau) \\
&= \sup_{\omega\in\mathcal{N}} \operatorname{ess\,sup} f_\omega(\widehat{\mathbf{X}}^n[1,\nu-1]) \\
&\geq \operatorname{ess\,sup} \widetilde{\mathbb{E}}^\nu[(\tau-\nu)^+|\widetilde{\mathbf{X}}^n[1,\nu-1]] \\
&= \widetilde{\text{WADD}}(\tau).
\end{aligned} \tag{43}$$

Combining (36) and (43), 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.

B. T^* ACHIEVES EQUALITY IN (18)

We will show that the mixture CuSum algorithm achieves the equality in (18), i.e.,

$$\widehat{\text{WADD}}(T^*) = \widetilde{\text{WADD}}(T^*). \tag{44}$$

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 T^* is symmetric, it follows that for any $i \geq \nu$, and any $\sigma_i, \sigma_i' \in \mathcal{S}_{n,\lambda}$,

$$\begin{aligned}
&\operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_{\sigma_\nu,...,\sigma_i,...,\sigma_\infty}[(T^*-\nu)^+|\widehat{\mathbf{X}}^n[1,\nu-1]] \\
&= \operatorname{ess\,sup} \widehat{\mathbb{E}}^\nu_{\sigma_\nu,...,\sigma_i',...,\sigma_\infty}[(T^*-\nu)^+|\widehat{\mathbf{X}}^n[1,\nu-1]].
\end{aligned} \tag{45}$$

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

$$\begin{aligned}
&\sup_{\omega\in\mathcal{N}} \operatorname{ess\,sup} \widehat{\mathbb{E}}_\omega^\nu[(T^*-\nu)^+|\widehat{\mathbf{X}}^n[1,\nu-1]] \\
&= \operatorname{ess\,sup} \widehat{\mathbb{E}}_\omega^\nu[(T^*-\nu)^+|\widehat{\mathbf{X}}^n[1,\nu-1]].
\end{aligned} \tag{46}$$

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

$$\begin{aligned}
&\sup_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda}{}^{\otimes(T-\nu+1)}}} \sum_{t=\nu+1}^{T} (t-\nu)\mathbb{P}_{\sigma_\nu,...,\sigma_T}(T^*=t|\widehat{\mathbf{X}}^n[1,\nu-1]) \\
&= \sum_{t=\nu+1}^{T} (t-\nu)\frac{1}{|\ \mathcal{S}_{n,\lambda}\ |^{(T-\nu+1)}} \\
&\quad \times \sum_{\substack{\{\sigma_\nu,...,\sigma_T\} \\ \in \mathcal{S}_{n,\lambda}{}^{\otimes(T-\nu+1)}}} \mathbb{P}_{\sigma_\nu,...,\sigma_T}(T^*=t|\widehat{\mathbf{X}}^n[1,\nu-1]) \\
&= \sum_{t=\nu+1}^{T} (t-\nu)\widetilde{\mathbb{P}}^\nu(T^*=t|\widetilde{\mathbf{X}}^n[1,\nu-1]).
\end{aligned} \tag{47}$$

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

$$\widehat{\mathbb{E}}_\omega^\nu[(T^*-\nu)^+|\widehat{\mathbf{X}}^n[1,\nu-1]] = \widetilde{\mathbb{E}}^\nu[(T^*-\nu)^+|\widetilde{\mathbf{X}}^n[1,\nu-1]]. \tag{48}$$

From (46) and (48), it follows that

$$\begin{aligned}
&\widehat{\text{WADD}}(T^*) \\
&= \sup_{\nu\geq1}\sup_{\omega\in\mathcal{N}} \operatorname{ess\,sup} \widehat{\mathbb{E}}_\omega^\nu[(T^*-\nu)^+|\widehat{\mathbf{X}}^n[1,\nu-1]] \\
&= \sup_{\nu\geq1} \operatorname{ess\,sup} \widehat{\mathbb{E}}_\omega^\nu[(T^*-\nu)^+|\widetilde{\mathbf{X}}^n[1,\nu-1]] \\
&= \sup_{\nu\geq1} \operatorname{ess\,sup} \widetilde{\mathbb{E}}^\nu[(T^*-\nu)^+|\widetilde{\mathbf{X}}^n[1,\nu-1]] \\
&= \widetilde{\text{WADD}}(T^*).
\end{aligned} \tag{49}$$

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