On Network Quickest Change Detection with Uncertain Models: An Experimental Study

James Z. Hare*, Yuchen Liang†, Lance M. Kaplan* and Venugopal V. Veeravalli‡
* DEVCOM Army Research Laboratory, USA. {james.z.hare.civ, lance.m.kaplan.civ}@army.mil
† Ohio State University, USA. liang.1439@osu.edu
‡ University of Illinois Urbana-Champaign, USA. vvv@illinois.edu

***Abstract*—We study the problem of Quickest Change Detection (QCD) in a complex networked system consisting of a set of heterogeneous agents that sequentially feed information to a central fusion center. At any unknown deterministic time, a persistent anomaly occurs, causing the distribution of observations from an unknown distinguishable subset of agents to simultaneously change from a nominal (pre-change) distribution to an anomalous (post-change) distribution, and the goal of the fusion center is to detect the change as quickly as possible subject to a false alarm constraint. Traditionally, various fusion rules have been proposed that assume that the distributions at each agent are either completely known or unknown and are locally solved using the Cumulative Sum (CuSum) and Generalized Likelihood Ratio (GLR) statistics, respectively. When an agent has access to training data, the *Uncertain Likelihood Ratio* (ULR) test generalizes distributional assumptions using *uncertain* distributions. However, the ULR has not been implemented for network change detection. This paper empirically studies incorporating the ULR statistics into the existing fusion rules for QCD and compares the average detection delay. Our results show that the ULR test can improve the average detection delay over the GLR tests using certain fusion techniques, while approaching the detection delay of the CuSum tests as the training data increases. Our results provide insights into future theoretical analysis to improve network QCD with imprecise knowledge of the distributions.**

***Index Terms*—Quickest Change Detection, Uncertain Models, Information Fusion**

I. Introduction

Complex networked systems are present in many applications such as the Internet of Battlefield Things (IoBT) [1], smart grids, highway monitoring systems, multiple unmanned aerial systems, remote diagnostics and troubleshooting, sensor networks, to name a few [2], [3]. They typically consist of a network of heterogeneous agents that sequentially transmit multiple data streams of information across a network to perform desired tasks, such as event detection. In decentralized networks, information is fed to a fusion center that typically cannot combine observations directly, requiring fusion algorithms to combine heterogeneous information. Additionally, event's may be present across the entire network or a subset of the network, and may only affect an unknown subset of agents, requiring combining all of the information in order to collectively infer the event. Furthermore, the agent's prior knowledge of the event characteristics may be *uncertain* and the fusion architecture must account for uncertainty to avoid making confidently wrong decisions.

A critical problem in complex networked systems is to detect the event that a system's operating condition has changed from a nominal state to an anomalous state as quickly as possible. The problem of network Quickest Change Detection (QCD) provides a framework to sequentially detect if a change has occurred [4]. This framework assumes that at any unknown deterministic time, an anomaly occurs that causes the distribution of observations of the agents to change from a nominal (pre-change) distribution to an anomalous (post-change) distribution, where the goal is to identify the change as quickly as possible. The anomaly is typically assumed to be persistent and that the subset of agents affected by the change is known [5], [6]. However, this has been relaxed and applied to multichannel detection [7]–[10], anonymous networks [11]–[13], and moving anomalies [14]–[16].

The objective of the network QCD problem presented in this paper is from a minimax standpoint, which aims to design a test statistic that minimizes the worst-case detection delay or conditional average detection delay subject to a constraint based on the average run length to false alarm [17]. When the fusion center knows the pre- and post-change distributions exactly, the optimal network test is the Cumulative Sum (CuSum) test, which maximizes the summation of the local agents' cumulative log-likelihood ratios computed using their sequential observations [7]. Alternatively, the Shriyaev-Roberts (SR) test is also asymptotically optimal, which sums the local agents' cumulative log-likelihood ratios over the possible change point locations.

However, complex networked systems operate in a highly dynamic and uncertain environment where complete knowledge of the agents' distributions is not always available. When the distributions of each agent are completely unknown, the Generalized Likelihood Ratio (GLR) test has been proposed, which replaces the known likelihoods in the CuSum and SR procedures with estimated likelihoods [18], [19], where the parameters of the distributions are estimated using Maximum Likelihood Estimates (MLE) [20], [21]. When the pre-change distributions of each agent are known exactly and the set of affected agents is known, the GLR tests are asymptotically optimal for the network QCD problem, see [18], [19]. However, the set of affected agents is typically unavailable in the GLR setting since the post-change distributions are unknown, requiring adaptations to the fusion rule to achieve robust detection delays.

Borrowing ideas from the network CuSum literature, Mei [22] found that when the number of affected agents is small compared to the total number of agents, computing the maximum over the local statistics eat each agent is shown to improve the detection delay. As the number of affected agents grows, the summation of local test statistics improves the detection delay compared to the maximum rule. In either case, both tests are first-order asymptotically optimal, and the difference in detection delay with finite samples is due to the second-order term of the expected detection delay. Alternatively, mixture procedures have been proposed to account for unknown affected agents using the GLR test by placing a weight on the local GLR statistics based on the agents' estimated distinguishability [21], [23]. However, these fusion rules were only studied for completely known distributions or unknown post-change distributions at each agent.

In some situations, the agents may have imprecise knowledge of the distributions that can be exploited. The Mixture rule has been proposed from a Bayesian perspective as a single agent test and is asymptotically optimal in the minimax formulation [24]. This test assumes that the pre-change distribution is known and estimates the expected likelihood of the post-change distribution of each agent by averaging the likelihood over an assumed prior distribution of the possible parameters. However, to the best of our knowledge, fusion rules based on locacl Mixture test statistics has not been studied from the network QCD problem, even though the single agent results should apply under certain conditions.

Recently, the *Uncertain Likelihood Ratio* (ULR) test for a single agent has been proposed to generalize the Mixture rule by allowing both the pre- and post-change distributions to be *uncertain* [25], [26]. The ULR test incorporates training data into the test statistic to provide imprecise knowledge in the distributions and has been shown empirically to improve the average detection delay in the finite sample regime over the GLR test. From a two-sample hypothesis testing perspective, the ULR test has shown comparable (even better in certain scenarios) performance to the GLR test, both empirically and theoretically [27]. Thus, incorporating the ULR test into the network QCD framework may generalize the problem to allow for uncertainty in the distributions.

An empirical evaluation of the network ULR tests show that the average detection delay is comparable (even better) to the network GLR tests when no training data is available to learn the distributions, while the detection delay approaches the optimal network CuSum tests as the amount of training data becomes large. Furthermore, it is seen that the amount of training data available affects the network tests ability to identify the set of agents affected by the change, which alters the detection delay. Our results provide an empirical understanding of the benefits of different fused network ULR statistics, while proposing future theoretical research directions to improve QCD in complex networked systems.

This paper is organized as follows. Section II formulates the problem of QCD, while Section III discusses the existing fusion rules for known and unknown distributions. Then, the network ULR test statistic is proposed in Section IV. Section V provides an empirical study and discusses future theoretical directions. Finally, Section VI concludes the paper.

II. Problem Formulation

Consider the problem of determining whether a complex networked system is operating in either a nominal or anomalous state. Initially, the system operates in the nominal state and at some *unknown* deterministic time $\gamma > 0$, the system switches to the anomalous state. We assume that the change is persistent and effects the entire network at the same time γ. The goal is to use sequential observations of the system to detect the change in the operating state as quickly as possible subject to a false alarm constraint. This problem is known as the network *Quickest Change Detection* (QCD) problem.

Let the system contain a set of K heterogeneous agents (sensors) $\mathcal{A} = \{a^1, ..., a^K\}$ that collectively monitor its state, where each agent $a^i \in \mathcal{A}$ collects an unbounded set of i.i.d. observations $\mathbf{X}^i = \{x_1^i, x_2^i, ...\}$ sequentially. When the system is operating in the nominal state, the observations x_t^i for $t < \gamma$ are drawn from a (pre-change) probability distribution $p^i(\cdot|\boldsymbol{\phi}_0^i)$ within the parametric family $\mathcal{P}^i$, where $\boldsymbol{\phi}_0^i \in \Phi^i$ are the set of parameters. Once the anomaly occurs, the observations x_t^i for $t \geq \gamma$ are drawn from the (post-change) probability distribution $p^i(\cdot|\boldsymbol{\phi}_1^i)$ within $\mathcal{P}^i$ with parameters $\boldsymbol{\phi}_1^i \in \Phi^i$. We assume that the observations across each agent are independent and that the parametric family and/or set of parameters can vary between agents due to their heterogeneous nature, i.e., for some agents a^i and a^j s.t. $i \neq j$, possibly $\mathcal{P}^i \neq \mathcal{P}^j$ and/or $\boldsymbol{\phi}_0^i \neq \boldsymbol{\phi}_0^j$ and $\boldsymbol{\phi}_1^i \neq \boldsymbol{\phi}_1^j$.

From a practical perspective, it is common to assume that the change does not affect every agent in the system and the subset of affected agents is unknown *a prior*. Therefore, each agent may not be able to distinguish the anomalous state from the nominal state individually. We define the unknown set of *distinguishable* agents as $\mathcal{A}_D = \{a^i \in \mathcal{A} | \boldsymbol{\phi}_1^i \neq \boldsymbol{\phi}_0^i, i = 1, ..., K\}$ s.t. $|\mathcal{A}_D| > 0$ and the unknown set of *indistinguishable* agents as $\bar{\mathcal{A}}_D = \mathcal{A} \setminus \mathcal{A}_D$. Thus, during each time $t \geq 1$, the observation of each agent a^i is drawn from $x_t^i \sim p^i(\cdot|\boldsymbol{\phi}_t^i)$, where

$$\begin{aligned} &\boldsymbol{\phi}_t^i = \boldsymbol{\phi}_0^i \ \forall \ a^i \in \mathcal{A} \text{ and } t < \gamma, \\ &\boldsymbol{\phi}_t^i = \begin{cases} \boldsymbol{\phi}_1^i & a^i \in \mathcal{A}_D \\ \boldsymbol{\phi}_0^i & a^i \in \bar{\mathcal{A}}_D \end{cases} \quad t \geq \gamma. \end{aligned} \tag{1}$$

Under this formulation, individual agents may not be able to identify if the operating condition has changed. However, since $|\mathcal{A}_D| > 0$, the collective group of agents are able to identify the change.

We consider the *centralized* fusion problem where during each time step $t \geq 1$, each agent transmits information to a central fusion center that makes the overall decision on whether a change has occurred. The information transmitted can either be (i) the raw observations or (ii) *beliefs* that summarize their local knowledge of the possible change points.[1] The fusion center sequentially computes a *test statistic* S_t for each $t \geq 1$ and identifies a *stopping time* t_c based on a *stopping criteria* λ as follows.

$$t_c = \inf\{t \geq 1 | S_t > \lambda\}. \quad (2)$$

In this setting, the test statistic S_t is based on a fusion rule (defined later), while the stopping criteria λ is a designed threshold based on a false alarm constraint p_{fa}. The goal is to minimize the *Conditional Average Detection Delay* (CADD) [17] defined as

$$\text{CADD}(t_c) = \sup_{\gamma \geq 1} \mathbb{E}_\gamma[t_c - \gamma | t_c \geq \gamma], \quad (3)$$

subject to the *False Alarm Rate* (FAR) defined as

$$\text{FAR}(t_c) = \frac{1}{\mathbb{E}_\infty[t_c]}, \quad (4)$$

which is the reciprocal of the *Average Run Length* (ARL). The expectation in (3) is taken over the joint anomalous distribution when the change occurs at time γ, while the ARL in (4) is the expected value of the stopping time conditioned on a change never occurring, i.e., $\gamma = \infty$. Therefore, the network QCD problem is to solve the following objective

$$\text{minimize CADD}(t_c) \text{ subject to FAR}(t_c) \leq p_{fa}. \quad (5)$$

III. Review of Fusion Rules for Network QCD

A. Known Probability Distribution

We first consider the network QCD formulation where it is assumed that the pre- and post-change distributions are known exactly. The optimal fusion rule in this setting is the network CuSum test [5], which fuses the agents' local cumulative log-likelihood ratios at each hypothesized change point $1 \leq \tau \leq t$ while preserving the maximum value as the overall test statistic. This is defined as

$$S_t^{\text{CuSum}} = \max_{1 \leq \tau \leq t} \sum_{a^i \in \mathcal{A}} \log \Lambda_{\tau,t}^i(\mathbf{X}_{\tau,t}^i) \quad (6)$$

where $\mathbf{X}_{t_1:t_2}^i = \{x_{t_1}^i, ..., x_{t_2}^i\} \subseteq \mathbf{X}^i$ and

$$\Lambda_{\tau,t}^i(\mathbf{X}_{\tau,t}^i) = \frac{p^i(\mathbf{X}_{\tau:t}^i | \boldsymbol{\phi}_1^i)}{p^i(\mathbf{X}_{\tau:t}^i | \boldsymbol{\phi}_0^i)} \quad (7)$$

is the cumulative likelihood ratio of agent a^i at the hypothesized change point τ. An alternative approach, known as the Shiryaev-Roberts (SR) procedure, fuses both the agents' local cumulative log-likelihood ratios and the possible hypothesized change points as follows

$$S_t^{\text{SR}} = \sum_{\tau=1}^{t} \prod_{a^i \in \mathcal{A}} \Lambda_{\tau,t}^i(\mathbf{X}_{\tau,t}^i). \quad (8)$$

[1]Beliefs can take the form of likelihood ratio statistics, quantized observations/statistics [5], [28], or local decisions [7]. This work focuses on likelihood ratio statistics. Note that this formulation can be extended to decision fusion rules, but is beyond the scope of this paper.

The SR test is also asymptotically optimal for the minimax formulation of the network QCD problem.

When the distributions are known, the set of distinguishable agents are also known, and the tests (6) and (8) implicitly filter out the indistinguishable agents since the likelihood ratio $\Lambda_{\tau,t}^i(\mathbf{X}_{\tau:t}^i) = 1$ for all $\tau = 1, ..., t$ and agent $a^i \in \bar{\mathcal{A}}_D$. Furthermore, these rules can be recursively computed based on the latest set of observation $\mathcal{X}_t = \{x_t^i | \forall a^i \in \mathcal{A}\}$ or set of likelihood ratios Λ_t [7], where $\Lambda_t = \{p^i(x_t^i|\boldsymbol{\phi}_1^i)/p^i(x_t^i|\boldsymbol{\phi}_0^i) | \forall a^i \in \mathcal{A}\}$. In the former case where observations are transmitted, the fusion center needs to know each agents distributions exactly, while the latter case does not require distributional knowledge. The former results in a total of K transmissions across the network at each time step t, while the latter requires Kt transmissions at each time step t.

B. Unknown Probability Distributions

In many situations, not only are the set of distinguishable agents unknown, it is possible that the individual distribution $p^i(\cdot|\boldsymbol{\phi}_0^i)$ and $p^i(\cdot|\boldsymbol{\phi}_1^i)$ for each agent a^i may be unknown or partially known. The classical approach to handling unknown probability distributions considers two conditions: (i) the pre-change distribution is known, while the post-change distribution is unknown, and (ii) both the pre- and post-change distributions are unknown. One approach to handle parametric uncertainty in the distribution is to use an estimator $\hat{\boldsymbol{\phi}}_{t_1,t_2}^i = \hat{\boldsymbol{\phi}}^i(\mathbf{X}_{t_1,t_2}^i)$. Under this formulation, the cumulative likelihood ratio of agent a^i for the hypothesized change point τ is defined as

$$\hat{\Lambda}_{\tau,t}^i(\mathbf{X}_{\tau:t}^i) = \frac{p^i(\mathbf{X}_{\tau:t}^i | \hat{\boldsymbol{\phi}}_{\tau,t}^i)}{p^i(\mathbf{X}_{\tau:t}^i | \boldsymbol{\phi}_0^i)} \quad (9)$$

for condition (i) and the following for condition (ii) [20]

$$\hat{\Lambda}_{\tau,t}^i(\mathbf{X}_{1:t}^i) = \frac{p^i(\mathbf{X}_{1:\tau-1}^i | \hat{\boldsymbol{\phi}}_{1,\tau-1}^i) p^i(\mathbf{X}_{\tau:t}^i | \hat{\boldsymbol{\phi}}_{\tau,t}^i)}{p^i(\mathbf{X}_{1:t}^i | \hat{\boldsymbol{\phi}}_{1,t}^i)}, \quad (10)$$

Typically, these tests are referred to as *Generalized Likelihood Ratio* (GLR) tests when the parameters are estimated using Maximum Likelihood Estimates (MLE), i.e., $\hat{\boldsymbol{\phi}}_{t_1,t_2}^i = \arg\max_{\boldsymbol{\phi}^i \in \boldsymbol{\Phi}^i} p^i(\mathbf{X}_{t_1,t_2}^i | \boldsymbol{\phi}^i)$. Other approaches to estimate the parameters include Method Of Moments (MOM), adaptive estimates from an uncertainty set [18], shrinkage-estimates [19], and piece-wise estimates based on the possible change point [29]. These approaches have been extensively studied for the single agent case, while fusion techniques have been limited and restricted to condition (i).

In this paper, we focus on the GLR network fusion rules $\hat{S}_t^{\text{CuSum}}$ and $\hat{S}_t^{\text{SR}}$ defined as (6) and (8), respectively, where the likelihood ratios are replaces with $\hat{\Lambda}_{\tau,t}^i$ defined in (9) and (10). Mixture-based variations of $\hat{S}_t^{\text{CuSum}}$ were considered in [21], [23], while $\hat{S}_t^{\text{SR}}$ was considered in [19] with shrinkage-estimators. Note that in this case, having each agent transmit their latest observation significantly increases the computational and memory requirements of the fusion center. Otherwise, each agent must transmit their likelihood ratios for each hypothesized change point requiring $\sum_{\tau=1}^{t} K\tau$ transmission up to time t.

A critical problem with these fusion rules are that the indistinguishable agents cannot be implicitly filtered out, unlike the known distribution case. To circumvent this, one could identify the maximum fused test statistics over the $2^K - 1$ combinations of distinguishable agents at a high computational cost. However, Mei [22] found that when the set of distinguishable agents is unknown, it is better to fuse local CuSum statistics. Specifically, when the total number of distinguishable agents $|\mathcal{A}_D|$ is small compared to the network size, it is best to use the maximum over the local tests. The GLR version of this is defined as follows

$$\hat{S}_t^{\max} = \max_{a^i \in \mathcal{A}} \hat{Z}_t^i, \tag{11}$$

where $\hat{Z}_t^i = \max_{1 \le \tau \le t} \log \hat{\Lambda}_{\tau,t}^i(\mathbf{X}_{\tau,t}^i)$ is defined as the local GLR-based CuSum test at agent a^i. However, when the number of distinguishable agents is $1 < |\mathcal{A}_D| < K$, the detection delay may increase and it was shown that the summation of local CuSum statistics is more robust. The GLR version of this test is defined as

$$\hat{S}_t^{\text{sum}} = \sum_{a^i \in \mathcal{A}} \hat{Z}_t^i. \tag{12}$$

In addition to being robust to different combinations of distinguishable agents, (12) is also robust to different change points affecting each individual agent. Note that these approaches only require each agent to transmit their local test statistic $\hat{Z}_t^i$ to the fusion center at each time step t, resulting in a total of Kt transmissions up to time step t.

Overall, these approaches have evaluated various fusion rules for specific problem formulations. However, a detailed comparison of the fusion rules for the proposed formulation is lacking. Additionally, these approaches assume the extremes, i.e., either the distributions are known, or completely unknown, while a general procedure for QCD when limited information is available for the distributions is lacking.

IV. Uncertain Likelihood Ratio Test for Network QCD

To generalize the QCD procedure, we consider that the probability distributions of each agent a^i, i.e., $p^i(\cdot|\boldsymbol{\phi}_0^i)$ and $p^i(\cdot|\boldsymbol{\phi}_1^i)$, are unknown *a priori*. However, it has access to a set of training data, $\boldsymbol{\tau}_0^i = \{\tau_{1,0}^i, ..., \tau_{n_{\tau_0}^i,0}^i\}$ and $\boldsymbol{\tau}_1^i = \{\tau_{1,1}^i, ..., \tau_{n_{\tau_1}^i,1}^i\}$ from both the pre- and post-change distributions, respectively, to learn the distributions, where $\tau_{m,0}^i \sim p^i(\cdot|\boldsymbol{\phi}_0^i)$, $\tau_{m,1}^i \sim p^i(\cdot|\boldsymbol{\phi}_1^i)$, $0 \le n_{\tau_0}^i < \infty$, and $0 \le n_{\tau_1}^i < \infty$. In addition, each agent assumes an uninformative prior distribution $f(\boldsymbol{\phi}^i)$ for the possible parameters of the pre- and post-change distributions. In this setting, we define the *Uncertain Likelihood Ratio* (ULR) test for each agent as follows

$$\tilde{\Lambda}_{\tau,t}^i = \frac{\hat{p}_0^i(\mathbf{Y}_{1,\tau-1,\boldsymbol{\tau}_0}^i)\tilde{p}_1^i(\mathbf{Y}_{\tau,t,\boldsymbol{\tau}_1}^i)}{\tilde{p}_0^i(\mathbf{Y}_{1,t,\boldsymbol{\tau}_0}^i)} = \frac{\tilde{p}_1^i(\mathbf{Y}_{\tau,t,\boldsymbol{\tau}_1}^i)}{\tilde{p}_{0|\mathbf{Y}_{1,\tau-1,\boldsymbol{\tau}_0}^i}^i(\mathbf{X}_{\tau,t}^i)}, \tag{13}$$

where $\mathbf{Y}_{t_1,t_2,\boldsymbol{\tau}_z}^i = \{\mathbf{X}_{t_1:t_2}^i, \boldsymbol{\tau}_z^i\}$;

$$\tilde{p}_1^i(\mathbf{Y}_{\tau,t,\boldsymbol{\tau}_1}^i) = \int_{\boldsymbol{\phi}^i \in \boldsymbol{\Phi}^i} p^i(\mathbf{Y}_{\tau,t,\boldsymbol{\tau}_1}^i|\boldsymbol{\phi}^i) f(\boldsymbol{\phi}^i) d\boldsymbol{\phi}^i, \tag{14}$$

is a prior predictive distribution;

$$\tilde{p}_{0|\mathbf{Y}_{1,\tau-1,\boldsymbol{\tau}_0}^i}^i(\mathbf{X}_{\tau,t}^i) = \int_{\boldsymbol{\phi}^i \in \boldsymbol{\Phi}^i} p^i(\mathbf{X}_{\tau:t}^i|\boldsymbol{\phi}^i) f(\boldsymbol{\phi}^i|\mathbf{Y}_{1,\tau-1,\boldsymbol{\tau}_0}^i) d\boldsymbol{\phi}^i, \tag{15}$$

is a posterior predictive distribution, and the right hand side of (13) is an application of Bayes rule. Then, the network test statistics $\tilde{S}_t^{\text{CuSum}}$, $\tilde{S}_t^{\text{SR}}$, $\tilde{S}_t^{\max}$, and $\tilde{S}_t^{\text{sum}}$ are defined as (6), (8), (11), (12), respectively, where the likelihood ratios are replaces with $\tilde{\Lambda}_{\tau,t}^i$ in (13).

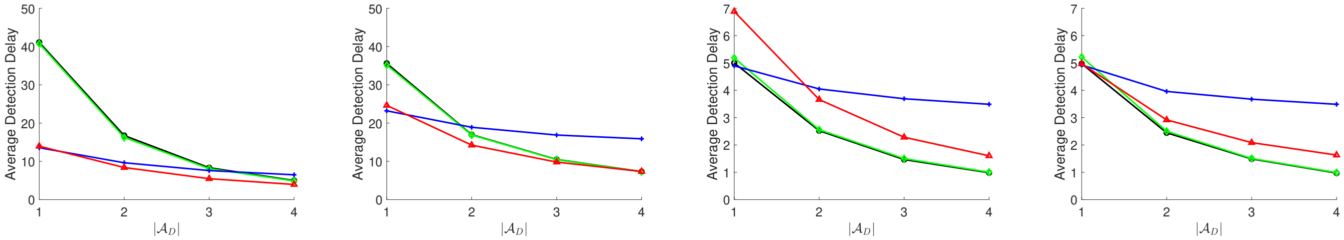
The intuition behind the ULR test is as follows. First, the ULR test, from a composite hypothesis testing standpoint, is optimal under the Neyman-Pearson setting when the parameters are drawn from the priors [27]. When there is no change, the parameters $\boldsymbol{\phi}_0^i$ are drawn from $\boldsymbol{\phi}_0^i \sim f(\boldsymbol{\phi}^i)$ and $\boldsymbol{\phi}_1^i = \boldsymbol{\phi}_0^i$, while under a change, the parameters are drawn independently from $\boldsymbol{\phi}_0^i \sim f(\boldsymbol{\phi}^i)$ and $\boldsymbol{\phi}_1^i \sim f(\boldsymbol{\phi}^i)$. Then, marginalizing the likelihoods over the assumed prior translates the composite hypothesis testing problem into a simple test, where the Neyman-Pearson lemma can be directly applied. This does not provide any analysis into the sequential hypothesis testing problem, making it necessary to study the performance of the ULR test for the QCD problem.

Second, it is known that the ULR test outperforms the GLR test for the two-sample hypothesis testing problem when we average over the assumed priors. However, the QCD literature evaluates the detection delay for specific choices of parameters and does not consider averaging over the assumed prior distributions. Therefore, it is unclear how the choice of the parameters may affect the detection delay of the ULR as compared to the GLR in the QCD framework.

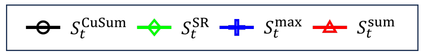
Third, when $|\boldsymbol{\tau}_0^i| \to \infty$, the ULR (13) for an agent a^i reduces to the Mixture rule [24]. This is because the posterior distribution $f(\boldsymbol{\phi}^i|\mathbf{Y}_{1,\tau-1,\boldsymbol{\tau}_0}^i)$ converges to a Dirac delta function at $\boldsymbol{\phi}_0^i$ according to the Bernstein-Von Misses theorem when the prior adheres to certain regularity conditions [30].

Finally, preliminary experimental studies of the ULR test for QCD show the following [25], [27]. (i) The training data act as additional observations drawn from the pre- and post-change distributions, which can improve the detection delay as compared to the mixture rule with the same prior. (ii) When the parameters are drawn from the true prior distributions, the ULR test can have a smaller detection delay than the GLR test when we average over the prior distribution. (iii) As the number of training data becomes larger, i.e., as $|\boldsymbol{\tau}_0^i| \to \infty$ and $|\boldsymbol{\tau}_1^i| \to \infty$, the average detection delay of the ULR becomes equivalent to the CUSUM. Thus, the ULR test generalizes the QCD problem to allow for completely unknown distributions, imprecise knowledge in the distributions, and complete knowledge in the distributions.

Preliminary studies of the ULR test lack theoretical analysis on its optimality for the QCD problem. Taking the results of



(a) ULR with $|\boldsymbol{\tau}_1^i| = 0$ (b) GLR with unknown post-change (c) ULR with $|\boldsymbol{\tau}_1^i| = 1000$ (d) Known distributions



(e) Legend

Figure 1: The Average Detection Delay vs. the number of distinguishable agents $|\mathcal{A}_D|$ for (a) the ULR tests with known pre-change distributions and $|\boldsymbol{\tau}_1| = 0$ training samples for the post-change distribution, (b) the GLR tests with known pre-change, (c) the ULR tests with known pre-change distribution and $|\boldsymbol{\tau}_1| = 1000$, and (d) the tests with known distributions.

Polak [24], when $|\boldsymbol{\tau}_0^i| \to \infty$, the ULR test holds the martingale properties and is optimal for the QCD problem. However, when $|\boldsymbol{\tau}_0^i| < \infty$, these properties do not hold, making it difficult to prove optimality. Furthermore, the ULR test has only been studied from a single agent perspective, and it is unclear as to the best global fusion rule in the network setting.

V. Experimental Study

In this section, we experimentally study the global average detection delay for the various fusion rules presented. Without loss of generality, we assume that the network consists of $K = 4$ agents and that the observations and training data for each agent $a^i \in \mathcal{A}$ are drawn from a two-dimensional Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_0^i, \mathbf{I})$ for the pre-change distribution and $\mathcal{N}(\boldsymbol{\mu}_1^i, \mathbf{I})$ for the post-change distribution, where $\mathbf{I}$ is the identify matrix. The pre-change means are set to $\boldsymbol{\mu}_0^i = \mathbf{0}$ for each agent $a^i \in \mathcal{A}$, while the post-change means are $\boldsymbol{\mu}_1^i = \mathbf{0}$ for agents $a^i \in \bar{\mathcal{A}}_D$ and $\boldsymbol{\mu}_1^i$ for $a^i \in \mathcal{A}_D$ is chosen s.t. $\sum_{a^i \in \mathcal{A}_D} D_{KL}(\mathcal{N}(\boldsymbol{\mu}_1^i, \mathbf{I}) \| \mathcal{N}(\boldsymbol{\mu}_0^i, \mathbf{I})) = 2$. We assume that the prior is the Normal Inverse Wishart (NIW) distribution, i.e., the natural conjugate prior, s.t. $f(\boldsymbol{\phi}^i) = INW(\mathbf{m}, \kappa, \nu, \mathbf{S})$ for all agents. The hyperparameters are chosen as $\mathbf{m} = \mathbf{0}$, $\kappa = 1.01$, $\nu = 1.01$, $\mathbf{S} = 0.1\mathbf{I}$ which are close to a Jeffreys prior to provide minimal prior knowledge. For details on how to compute the closed form of the ULR statistic for each agent, see [27].

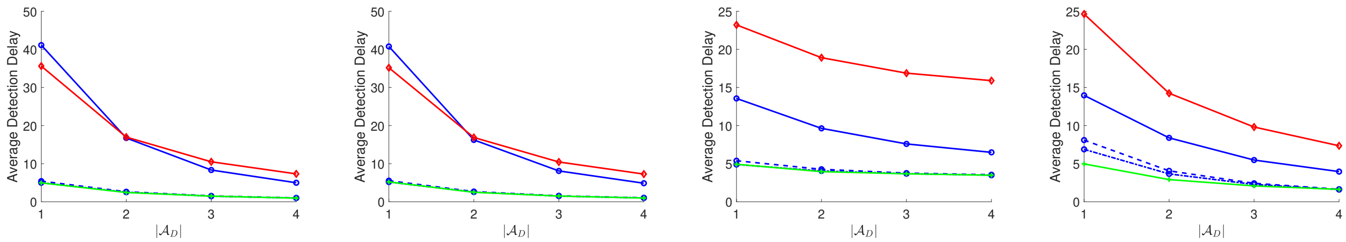
A. Known Pre-change Distribution

First, we studied the average detection delay of each fusion rule vs. the number of distinguishable agents when the pre-change distributions are known and the post-change distributions are unknown. We consider that the change occurs at time $\gamma = 1$ to estimate the worst case detection delay and we consider that the number of training samples for the post-change distribution for each agent are within the set $|\boldsymbol{\tau}_1^i| \in \{0, 100, 1000\}$.

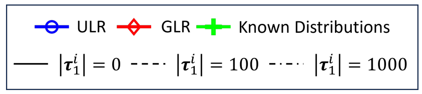
Figures 1(a) and 1(b) show the average detection delay vs. number of distinguishable agents for the ULR test with $|\boldsymbol{\tau}_1^i| = 0$ and the GLR test. Overall, the curves for each fusion rule have similar trends. Initially, the max and sum tests, i.e., $\tilde{S}_t^{\max}$, $\tilde{S}_t^{\text{sum}}$, $\hat{S}_t^{\max}$, and $\hat{S}_t^{\text{sum}}$, perform the best when only one agent is distinguishable, while the ULR and GLR variants of the CuSum and SR perform the worst. As $|\mathcal{A}_D|$ increases, the sum test performs best, followed by max, CuSum, and SR tests. Finally, when all agents are distinguishable, the sum, CuSum, and SR variant tests are equal with the minimum average detection delay followed by the max test. These results are consistent with the observations made by Mei [22] when the set of indistinguishable agents is not implicitly known. Thus, the summation variants of the tests are the best procedure for network QCD with known pre-change distributions and completely unknown post-change distributions.

Furthermore, Figure 1(c) shows the average detection delay for the ULR tests when $|\boldsymbol{\tau}_1^i| = 1000$ for each agent a^i. In this setting, the post-change distribution is almost known and we can see that the CuSum and SR variants with the ULR tests are the best choice. The results for the CuSum, SR, and max variants provide the same performance as the optimal CuSum procedure presented in Figure 1(d) with known distributions. This is because the ULR test is able to implicitly filter out indistinguishable agents and the ULR test performance is converging toward the performance of the CuSum test. However, the ULR variant of the sum test has a larger detection delay than the sum test with known distributions, requiring further investigation.

Next, we compare the global tests for each likelihood ratio formulation to understand the difference in the average detection delay as seen in Figure 2. Figures 2(a) and 2(b) show that the ULR tests with no training data for each agent initially has a larger average detection delay as compared to the GLR test for $|\mathcal{A}_D| = 1$. As $|\mathcal{A}_D|$ increases, the ULR begins to outperform the GLR. The benefits of the ULR-based test with no training samples are seen in the max and sum tests in Figures 2(c) and 2(d), where the average detection delay is much smaller for the ULR test than the GLR test. Thus, when the distributions are completely unknown and the set of distinguishable agents is unknown, the use of

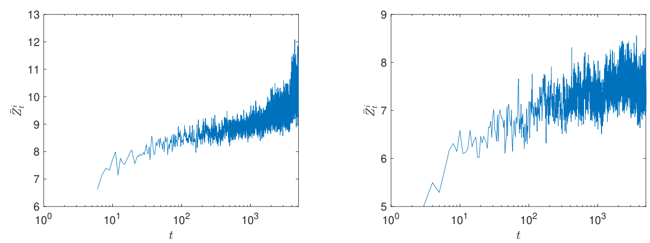


(a) S_t^{CuSum} (b) S_t^{SR} (c) $S_t^{\max}$ (d) S_t^{sum}

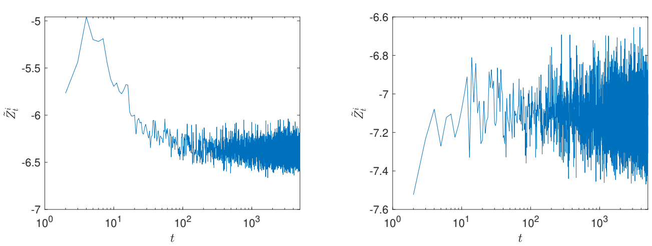


(e) Legend

Figure 2: The Average Detection Delay vs. the number of distinguishable agents $|\mathcal{A}_D|$ for the test (a) (6), (b) (8), (c) (11), and (d) (12) where the pre-change distribution is known.



(a) (b)



(c) (d)

Figure 3: Example of the local test statistic of the GLR with (a) unknown pre- and post-change distributions, (b) known pre-change distribution and unknown post-change distribution, and the ULR test with (c) $|\boldsymbol{\tau}_0^i| = 0$ and $|\boldsymbol{\tau}_1^i| = 0$, and (d) $|\boldsymbol{\tau}_0^i| = 100$ and $|\boldsymbol{\tau}_1^i| = 0$, when no change is present.

the summation procedure allows the ULR to outperform the GLR. Furthermore, when the ULR has one thousand training samples for each agent, the ULR tests in Figure 2 are consistent with the tests that assume the distributions are known (except for the sum test). This shows that the ULR test can bridge the gap between completely unknown distributions and completely known distributions.

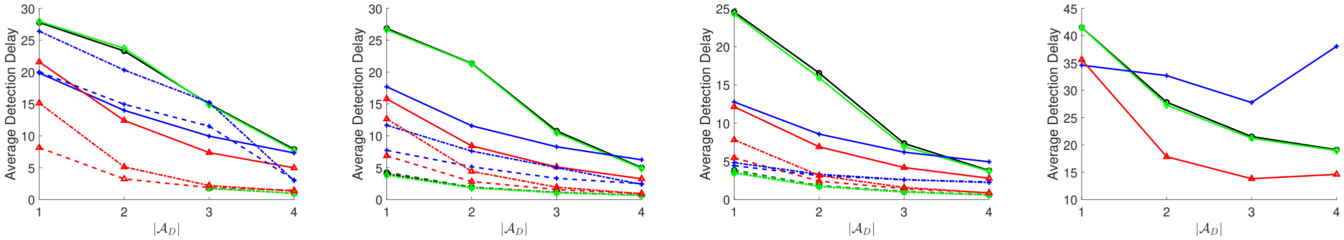
B. Unknown Pre- and Post-Change Distributions

Next, we consider that both the pre- and post-change distributions are unknown. Under this condition and discussed by Lai [20], we cannot evaluate the worst-case detection delay based on a change occurring at time $\gamma = 1$. This is because the GLR test require multiple samples to estimate the ML estimates of the parameters of the pre-change distributions. Furthermore, when no change is present, the local GLR test statistic $\hat{Z}_t^i$ increases logarithmically with t, as seen in Figure 3(a), as opposed to eventually fluctuating around a constant value ($t > 400$) when the pre-change distributions are known, as seen in Figure 3(b). Thus, it was suggested in [20] to evaluate the individual average detection delay for a change point $\gamma \geq n_c$, where n_c is chosen to allow for a sufficient estimate of the pre-change parameters.

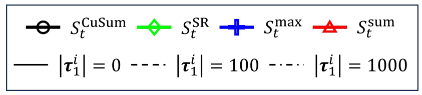
A similar effect is observed with the ULR test. When there is zero training data for the pre-change distribution of agent i, the ULR test at time $t = 1$ is equal to $\Lambda_t^i = 1$, and thus requires at least two samples in order to generate a response. Additionally, as seen in Figure 3(c), the local test statistic $\tilde{Z}_t^i$ with the ULR likelihood (13) initially increases for a burn-in period, and then decreases and fluctuates around a constant value. As the amount of training data for the post-change distribution increases, the test statistics eventually decrease and fluctuate around a smaller constant value. Therefore, constructing a threshold λ to achieve a high ARL is difficult and restricts the computation of the average detection delay.

When there is training data for the pre-change distribution, e.g., one hundred samples, the ULR test increases toward a constant value as seen in Figure 3(d). In this condition, we are able to construct a threshold λ to achieve a desired ARL. Thus, the ULR test requires a sufficient number of training samples in order to provide a meaningful evaluation of the average detection delay. Note that shifting the change point from $\gamma = 1$ to $\gamma > n_c$ provides greater than n_c training samples of the pre-change distribution. Theoretically quantifying the amount required is left for future work.

Another issue with the ULR test is that it does not have the martingale property that allows the test statistic to increase toward infinity after a change occurs. Our previous work [30] has shown that the amount of training data collected for the pre-change distribution governs how high the ULR statistic can



(a) ULR with $|\boldsymbol{\tau}_0^i| = 0$ (b) ULR with $|\boldsymbol{\tau}_0^i| = 10$ (c) ULR with $|\boldsymbol{\tau}_0^i| = 100$ (d) GLR completely unknown



(e) Legend

Figure 4: The Average Detection Delay vs. the number of distinguishable agents $|\mathcal{A}_D|$ for the ULR with (a) $|\boldsymbol{\tau}_0^i| = 0$, (b) $|\boldsymbol{\tau}_0^i| = 10$, (c) $|\boldsymbol{\tau}_0^i| = 100$, and (d) the GLR with unknown pre- and post-change distributions.

increase before converging to a finite value. As the amount of training data increases, the convergence point increases toward infinity, because the ULR test with known pre-change holds the martingale property. This means that the ULR test statistic can only increase to a finite value with finite pre-change training data. When the threshold is chosen to be larger than the convergence point, the ULR tests can lead to missed detections. Therefore, future QCD formulations with unknown distributions must minimize the detection delay subject to both the false alarm rate and the missed detection probability.

Next, we evaluated the average detection delay of the ULR and GLR tests vs. the number of distinguishable agents when the change occurs at $\gamma = 50$, as seen in Figure 4. We varied the number of training samples for the ULR tests within the set $|\tau_0^i| \in \{0, 10, 100\}$ for the pre-change distributions and $|\tau_1^i| \in \{0, 100, 1000\}$ for the post-change distributions.

Overall, the ULR and GLR tests show similar trends as the known pre-change condition except with a larger detection delay. Furthermore, the ULR test generalizes between the GLR and CUSUM by incorporating imprecise knowledge in the distributions. For the ULR tests, the max and sum tests outperforms the CuSum and SR variant tests when there is no post-change training data. Once the agents collect training data for the post-change distribution or the number of distinguishable agents equals K, the max test has a larger detection delay then the sum test followed by the CuSum and SR variants. While the crossover for between the GLR tests occurs with less distinguishable agents.

When there is zero training data for the pre-change distribution, the detection delay for the max test increases as the amount of training data increases for the post-change distribution as seen in Figure 4(a). This is because the test statistic requires additional samples after the change to overcome the initial burn-in period. However, once the pre-change training data increases to ten and one thousand in Figures 4(b) and 4(c), the detection delay decreases as the post-change training data increases.

The detection delay for the CuSum and SR variant tests decreases as the post-change training data increases as seen in Figures 4(b) and 4(c). In Figure 4(a), the detection delay is only available when there is no post-change training data. As the training data increases for one and two distinguishable agents, the test statistics cannot overcome the initial increase due to the lack of pre-change training data, causing the tests to always miss the change or detect a change before the change point. However, when the number of distinguishable agents is greater than two, the tests produce a small delay.

C. Discussion

Based on the experiments conducted, the following theoretical questions arise for the ULR tests. When the pre-change distribution is known, can we quantify the number of distinguishable agents required for the CuSum and SR variant tests to outperform the max rule conditioned on the amount of training data available? Furthermore, a detailed proof of the ULR test asymptotically converging to the global CUSUM and SR tests seems promising.

When both the pre- and post-change distributions are unknown, does the amount of pre-change training samples required to overcome in initial "burn-in" significantly hinder the ULR tests? Additionally, does the ratio between pre- and post-change training data govern whether the ULR test statistics provide similar detection delay trends as the known pre-change distribution condition?

Overall, the optimal fusion rule requires that the indistinguishable agents are filter out of the overall test. Since increasing the training data cause the agents to converge toward known distributions, would a mixture-based fusion rule improve the detection delay of the ULR tests by exploiting the amount of training data to properly weight the local ULRs? Furthermore, our experimental study considers a special case where the distributions are normal with Jeffreys priors and that the amount of training data across agents is balanced. Do the detection delay trends observed translate across diverse local distributions with varying dimensions, varying training data, and unique priors for the pre- and post-change distributions?

VI. Conclusion

This paper studies the problem of QCD in a complex networked system consisting of a set of heterogeneous agents that sequentially feed information to a centralized fusion center. We assume that at any unknown time, a persistent anomaly occurs that changes the distribution of observations from a pre-change distribution to a post-change distribution for a subset of distinguishable agents within the network. We consider that the agents are uncertain and have imprecise knowledge of the distribution based on training data. We propose the use the Uncertain likelihood Ratio within existing fusion frameworks for QCD to generalize between completely known and completely unknown distributions. The tests are empirically evaluated to study the benefits of each fusion rule and to provide theoretical insights for future study.

We found that when the pre-change distributions are known and there is limited training data for the post-change distribution, it is best to sum over the individual ULR test statistics. When the pre- and post-change distributions are unknown, the detection delay provides similar trends, but requires special attention in order to accurately evaluate its performance. As future work, we will investigate this empirical study in more detail and aim to answer the theoretical questions posed in our results section.

Acknowledgement

This work was supported by the DEVCOM Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196 (ARL IoBT CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the DEVCOM Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] T. Abdelzaher, N. Ayanian, T. Basar, S. Diggavi, J. Diesner, D. Ganesan, R. Govindan, S. Jha, T. Lepoint, B. Marlin *et al.*, "Will distributed computing revolutionize peace? the emergence of battlefield IoT," in *IEEE International Conference on Distributed Computing Systems*. IEEE, 2018, pp. 1129–1138.

[2] G. Wen, X. Yu, W. Yu, and J. Lu, "Coordination and control of complex network systems with switching topologies: A survey," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 10, pp. 6342–6357, 2020.

[3] X.-M. Zhang, Q.-L. Han, X. Ge, D. Ding, L. Ding, D. Yue, and C. Peng, "Networked control systems: A survey of trends and techniques," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 1, pp. 1–17, 2019.

[4] 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.

[5] Y. Mei, "Information bounds and quickest change detection in decentralized decision systems," *IEEE Transactions on Information theory*, vol. 51, no. 7, pp. 2669–2681, 2005.

[6] A. G. Tartakovsky and V. V. Veeravalli, "Change-point detection in multichannel and distributed systems with applications," *Statistics Textbooks and Monographs*, vol. 173, pp. 339–370, 2004.

[7] ——, "Asymptotically optimal quickest change detection in distributed sensor systems," *Sequential Analysis*, vol. 27, no. 4, pp. 441–475, 2008.

[8] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3727–3735, 2015.

[9] X. Ren, K. H. Johansson, and L. Shi, "Quickest change detection with observation scheduling," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2635–2647, 2016.

[10] A. G. Tartakovsky, "Asymptotically optimal quickest change detection in multistream data—part 1: General stochastic models," *Methodology and Computing in Applied Probability*, vol. 21, no. 4, pp. 1303–1336, 2019.

[11] Z. Sun, S. Zou, R. Zhang, and Q. Li, "Quickest change detection in anonymous heterogeneous sensor networks," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1041–1055, 2022.

[12] Z. Sun and S. Zou, "Quickest anomaly detection in sensor networks with unlabeled samples," *IEEE Transactions on Signal Processing*, vol. 71, pp. 873–887, 2023.

[13] W.-H. Li and Y.-C. Huang, "Bandwidth-constrained distributed quickest change detection in heterogeneous sensor networks: Anonymous vs non-anonymous settings," in *2023 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2023, pp. 1384–1389.

[14] 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.

[15] 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.

[16] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley, "Quickest detection of dynamic events in networks," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2280–2295, 2019.

[17] G. Lorden, "Procedures for reacting to a change in distribution," *The annals of mathematical statistics*, pp. 1897–1908, 1971.

[18] A. G. Tartakovsky and A. S. Polunchenko, "Decentralized quickest change detection in distributed sensor systems with applications to information assurance and counter terrorism," in *Proceedings of the 13th Annual Army Conference on Applied Statistics*. Citeseer, 2007.

[19] Y. Wang and Y. Mei, "Large-scale multi-stream quickest change detection via shrinkage post-change estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6926–6938, 2015.

[20] T. L. Lai and H. Xing, "Sequential change-point detection when the pre- and post-change parameters are unknown," *Sequential analysis*, vol. 29, no. 2, pp. 162–175, 2010.

[21] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," in *Information Theory and Applications Workshop*. IEEE, 2013, pp. 1–20.

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

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

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

[25] J. Z. Hare, L. Kaplan, and V. V. Veeravalli, "Toward uncertainty aware quickest change detection," in *IEEE International Conference on Information Fusion*. IEEE, 2021, pp. 1–8.

[26] J. Z. Hare and L. Kaplan, "Uncertainty-aware quickest change detection: An experimental study," in *IEEE International Conference on Information Fusion*. IEEE, 2022, pp. 1–8.

[27] J. Z. Hare and L. M. Kaplan, "Improved small sample hypothesis testing using the uncertain likelihood ratio," in *IIEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2023, pp. 1–5.

[28] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Transactions on Information theory*, vol. 47, no. 4, pp. 1657–1665, 2001.

[29] G. Lorden and M. Pollak, "Nonanticipating estimation applied to sequential analysis and changepoint detection," *The Annals of Statistics*, vol. 33, no. 3, pp. 1422–1454, 2005.

[30] J. Z. Hare, C. A. Uribe, L. Kaplan, and A. Jadbabaie, "A general framework for distributed inference with uncertain models," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 7, pp. 392–405, 2021.