This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2021.3077302, IEEE Transactions on Signal Processing

Ordering for Communication-Efficient Quickest Change Detection in a Decomposable Graphical Model

Yicheng Chen, Rick S. Blum, *Fellow, IEEE*, and Brian M. Sadler, *Life Fellow, IEEE*

***Abstract*—A quickest change detection problem is considered in a sensor network with observations whose statistical dependency structure across the sensors before and after the change is described by a decomposable graphical model (DGM). Distributed computation methods for this problem are proposed that are capable of producing the optimum centralized test statistic. The DGM leads to the proper way to collect nodes into local groups equivalent to cliques in the graph, such that a clique statistic which summarizes all the clique sensor data can be computed within each clique. The clique statistics are transmitted to a decision maker to produce the optimum centralized test statistic. In order to further improve communication efficiency, an ordered transmission approach is proposed where transmissions of the clique statistics to the fusion center are ordered and then adaptively halted when sufficient information is accumulated. This procedure is always guaranteed to provide the optimal change detection performance, despite not transmitting all the statistics from all the cliques. A lower bound on the average number of transmissions saved by ordered transmissions is provided and for the case where the change seldom occurs the lower bound approaches approximately half the number of cliques provided a well behaved distance measure between the distributions of the sensor observations before and after the change is sufficiently large. We also extend the approach to the case when the graph structure is different under each hypothesis. Numerical results show significant savings using the ordered transmission approach and validate the theoretical findings.**

***Index Terms*—Communication-efficient, CUSUM, decomposable graphical models, minimax, ordered transmissions, quickest change detection, sensor networking.**

I. Introduction

Sensor networks are critical for many applications such as disaster response, security, smart cities, enhanced building operation for optimized energy usage, health monitoring and assisted living, and smart transportation systems [1], [2]. A fundamental problem is to detect the occurrence of a change. This can be modeled as a quickest change detection (QCD) problem, see [3]–[19] and references therein.

The work is supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-17-1-0331, the National Science Foundation under Grant ECCS-1744129, and a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance (PITA). Some preliminary work was presented in Y. Chen, R. Blum, and B. Sadler, "Optimal quickest change detection in sensor networks using ordered transmissions," IEEE Workshop on Signal Processing Advances in Wireless Communications, 2020.

Yicheng Chen and Rick S. Blum are with Lehigh University, Bethlehem, PA 18015 USA (email: yic917@lehigh.edu, rblum@eecs.lehigh.edu).

Brian M. Sadler is with the Army Research Laboratory, Adelphi, MD 20878 USA (email: brian.m.sadler6.civ@mail.mil).

The classical centralized and unconstrained communication QCD problem in sensor networks is well investigated [13], [14], [16], [17] where each sensor monitoring the environment takes a sequence of observations. Based on the data received from the sensor nodes, a decision maker at a fusion center (FC) would like to detect the change as soon as possible subject to a false alarm constraint. Depending on knowledge of the change time distribution, minimax [3]–[6] and Bayesian [7]–[9] QCD formulations have been developed, and related theory and analyses on QCD are given in [10]–[19]. In some previous work, the distribution of the observations at all sensors changes simultaneously [13] at an unknown change time. In other work, the change time at different sensors is modeled to be different. In [20], a centralized Bayesian version of the QCD problem is considered where the change propagates across the sensors and its propagation is modeled as a Markov process. From a minimax point of view, [21], [22] have considered rapidly detecting a change in an unknown subset of sensors. While [21], [22] consider cases with statistically dependent sensor observations, we have not seen published work on distributed implementation of QCD for statistically dependent sensor observations, which is the topic of this paper. Here, we focus on the minimax formulation where we model the change time as a deterministic but unknown positive integer and minimize the worst case average detection delay (WADD) subject to a false alarm constraint. In many cases, each sensor in the network carries its own limited energy source, and thus the energy cost of sensor transmissions is significant. On the other hand, in many cases the FC may have much less concern about lowering its own energy usage. Hence, improving communication efficiency from sensors to the FC is an important topic for QCD in sensor networks.

A particularly popular approach called censoring has been shown to be an effective method to improve communication efficiency where sensors transmit only highly informative data [23]. In [23], upper and lower thresholds are set and sensors transmit only very large or small likelihood ratios because these values provide significant information about which hypothesis is most likely to be true. Censoring-based QCD is proposed in [11] where it is shown that censoring yields transmission savings but always increases detection delay (the accepted performance measure for QCD). Another communication efficient QCD approach called top-r local cumulated sum (CUSUM) [24], [25] only employs the largest r statistics. Different from this paper where the value of r is chosen in an adaptive manner, top-r in CUSUM [24], [25] has predefined r to save communications although this generally results in sub-optimal performance through increased detection delay.

In this paper we introduce an ordered transmission QCD method that will lower communications without increasing the detection delay. The ordered transmission approach (also called ordering) was first introduced for a distributed testing problem between two fixed hypotheses and employing an FC [26]. Using ordering the sensors with the most informative observations transmit first. Transmissions can be halted when sufficient information is accumulated for the FC to decide which hypothesis is true. In [26], it was shown that this ordered transmission approach can reduce the number of transmissions without losing any detection performance for cases with statistically independent observations. Under the restriction of independent observations, an upper bound on the transmission savings of ordering has been developed in [27] for a specific signal-in-noise detection problem. Under the assumption of known a priori probabilities, distributed processing for testing the mean or covariance matrix in statistically dependent Gaussian observations following a decomposable Gaussian graphical model (GGM) is considered in [28] and [29], respectively.

The previous work on ordered transmissions [26]–[29] focused entirely on distributed testing of the mean or covariance matrix with statistically independent or Gaussian dependent observations across the sensors where there is no change in the distributions over time. In practice, the possibly non-Gaussian and statistically dependent observations at different sensors might undergo a change at some unknown time and we might want to detect this change quickly, so QCD is needed. In this paper we provide a new communication efficient QCD algorithm for rapidly detecting a change with statistically dependent observations across the sensors using distributed processing. The performance metric used in this paper is based on minimizing the average delay, which is different from the performance metrics in [26]–[29]. This leads to a different ordering algorithm than the one employed in [26]–[29]. Given the different performance metric, a very different approach was required to develop the analytical lower bound on communications savings given in this paper as compared to the one in [26], [28], [29]. Here, the focus is on the case where the observations follow a decomposable graphical model (DGM) to characterize the dependence among sensor observations, completely subsuming the case with independent observations. This work is a highly nontrivial extension of the preliminary work in [30], which was limited to independent observations across sensors and did not provide a communication saving lower bound.

In this paper we focus on the case where the underlying graphical model is known. This can be obtained based on the association with some corresponding physical network or some system for which we know the model. A few examples of corresponding physical networks to which the graph may be related include the electrical grid, a natural gas delivery network, a water distribution network, and a telecommunication network. In the electrical grid example, the graph describing electrical measurements is directly determined by the known topology of the electrical grid and the components which connect the nodes [31] using the standard theory of electrical networks. Natural gas delivery networks are another case where the graph is determined by the known pipeline topology and the components which connect the nodes [32]. These same ideas apply for many other examples where a physical network is present for which a mathematical model exists. There are other examples where there may not be an associated physical network but a known model can be used to determine the graph. For example, economic theory might be employed to describe a model between several variables of interest. The work in [33] focuses on one such case where the related variables are the prime interest rate and the median house prices in several different cities. Any such model of a system can be used in a similar way. This opens up many more cases where the graph structure is essentially known.

A. Our Contributions

This work describes a new methodology/algorithm that provides a communication efficient distributed processing method for a sensor networking change detection problem where the observations follow a statistically dependent distribution before and after the change. The focus is on the case where both distributions are characterized by a decomposable graphical model (DGM) with common graph structure, which completely subsumes the case where observations follow a GGM or are independent. We prove that the methodology/algorithm provides identical performance, in terms of the worst case average detection delay subject to a false alarm constraint, to the optimum communication unconstrained performance, while reducing the number of required communications from the most energy constrained entities. We also prove that the communication savings will be significant in some important cases, saving more than half of the communications from the most energy constrained entities. In Section IV-D, extensions are described for the case where the graph structure of the distributions before and after the change are different. An approximate, but accurate, approach is described to extend these results to graphs that are not decomposable.

B. Paper Organization

The paper is organized as follows. In Section II, we first introduce the problem formulation of QCD and then describe a brief discussion on mathematical formulations of decomposable graph models. In Section III, we describe distributed computation of the optimum test statistic. Communication-efficient QCD using ordered transmissions is described in Section IV and a lower bound on the average number of transmissions saved via ordering is provided. Section V presents some numerical results to demonstrate the communication efficiency of the proposed algorithm. Finally, we conclude the paper in Section VI.

II. Problem Formulation and Decomposable Graphical Models

In this section, we first introduce the problem formulation of QCD which is modeled as a constrained optimization problem.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2021.3077302, IEEE Transactions on Signal Processing

Then we briefly describe mathematical formulations of DG which characterize the dependent sensor observations by decomposable undirected graph.

A. Problem Formulation

We consider a sensor network with M sensors anc FC. Sensor m for $m = 1, 2, .., M$ observes the seque $\{X_{n,m}\}_{n\geq 1}$ with n denoting the time slot index. The object is to detect a change as quickly as possible after the chai occurs which implies the goal is to minimize detection de if the change occurs. As long as no change is declared, sensors will continue observing data. Throughout this pa we make the following assumption.

Assumption 1: The distributions of the observations at all sensors change simultaneously at the change time τ. In particular, at the unknown time slot τ, the distribution of $X_{n,[1,M]} \triangleq (X_{n,1}, X_{n,2}, ..., X_{n,M})$ changes from f_0 to f_1 where f_0 and f_1 are the known probability density functions (pdfs) before and after the change time, respectively. The random variable $X_{n,m}$ is independent across the time slot index n but will generally assumed to be dependent across the sensor index m.

Without a prior on the distribution of the change time, we model the change time τ as a deterministic but unknown integer and we employ the constraint

$$\mathbb{E}_\infty(n') \geq \gamma \tag{1}$$

where n' denotes the time slot when the decision maker declares a change has occurred, $\mathbb{E}_\infty(n')$ is the average delay when the change does not occur, and γ is a pre-specified constant. If the change occurs, then one formulation to evaluate the detection delay is to employ the WADD defined in [4] as

$$\mathrm{WADD}(n') = \sup_{\tau\geq 1} \text{ ess } \sup \mathbb{E}_\tau \left[(n'-\tau)^+ | \mathcal{I}_{\tau-1}\right] \tag{2}$$

where $\operatorname{ess\,sup} X$ denotes essential supremum of X, $\mathbb{E}_\tau$ is the expectation when the change occurs at time τ, $(x)^+ \triangleq \max\{x, 0\}$, $\mathcal{I}_{\tau-1} \triangleq (X_{[1,\tau-1],1}, ..., X_{[1,\tau-1],M})$ denotes past global information at time slot τ, and $X_{[1,\tau-1],m} \triangleq (X_{1,m}, ..., X_{\tau-1,m})$ denotes past local information at sensor m. Thus, the QCD problem in a minimax setting can be formulated as a constrained optimization problem

$$\begin{aligned} &\min_n \mathrm{WADD}(n') \\ &s.t.\ \mathbb{E}_\infty(n') \geq \gamma. \end{aligned} \tag{3}$$

Before proposing the new energy-efficient QCD approach to solve (3), we review some basic properties of any DGMs in the next subsection.

B. Decomposable Graphical Models

DGMs have received extensive study in machine learning [34], [35], sensor networks [36] and electric power systems [37]. In this section, we briefly describe the basic theory of DGMs. Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with M vertices, where $\mathcal{V} = \{1, 2, ..., M\}$ is the set of vertices and $\mathcal{E} = \{(i_1, j_1), (i_2, j_2), ..., (i_{|\mathcal{E}|}, j_{|\mathcal{E}|})\}$ denotes the set of undirected edges of the graph. The graphical model for a random vector $X_{i,[1,M]}$ at each time slot i with graph $\mathcal{G}$ describes the statistical dependency model such that for each time slot i, $X_{i,[1,M]}$ follows a known distribution that obeys the pairwise Markov property with respect to $\mathcal{G}$. The distributed observation vector $X_{i,[1,M]}$ satisfies the pairwise Markov property with respect to $\mathcal{G}$ if, for any pair (m, m') of non-adjacent vertices, i.e., $(m, m') \notin \mathcal{E}$, the corresponding pair of elements of $X_{i,[1,M]}$, $X_{i,m}$ and $X_{i,m'}$, are conditionally independent when conditioned on the remaining elements. This can be expressed as

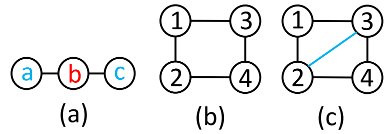

Fig. 1. Samples of three graphical models.

$$\begin{aligned} &f\left(X_{i,m}, X_{i,m'} \,\middle|\, X_{i,\mathcal{V}\backslash\{m,m'\}}\right) \\ &= f\left(X_{i,m} \,\middle|\, X_{i,\mathcal{V}\backslash\{m,m'\}}\right) f\left(X_{i,m'} \,\middle|\, X_{i,\mathcal{V}\backslash\{m,m'\}}\right) \end{aligned} \tag{4}$$

where we have used $f(\cdot)$ to denote the corresponding pdfs. To illustrate (4), let $X_{i,m}, X_{i,m'}$, and $X_{i,\mathcal{V}\backslash\{m,m'\}}$ in (4) denote the data at nodes $a, b,$ and c, respectively, in Fig. 1(a). The result in (4) implies that the data at nodes a and c are conditionally independent given the data at node b. This illustrates how the graph describes dependency.

An undirected graph is defined to be decomposable if the graph has the property that every cycle of length larger than 3 possesses a chord [38]. A cycle is a path which begins and ends at the same node. A chord is a connection that ensures that every two non-consecutive vertices in every cycle are neighbors. For example, the single cycle graph in Fig. 1(b) is not decomposable but the graph in Fig. 1(c) is decomposable due to the blue link which is a chord. This definition implies that a decomposable graph can be successively decomposed into its largest fully connected subgraphs called cliques.

Let K denote the number of cliques in the decomposable undirected graph $\mathcal{G}$. The sequence of cliques of the graph $\mathcal{G}$ is denoted by $\{\mathcal{C}_k\}_{k=1}^K$. Note that each clique $\mathcal{C}_k$ is a group of fully connected nodes. We denote the corresponding histories $\{\mathcal{H}_k\}_{k=1}^K$ and separators $\{\mathcal{S}_k\}_{k=2}^K$ as

$$\mathcal{H}_k = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_k,\ \forall k = 1, 2, ..., K, \tag{5}$$

and

$$\mathcal{S}_k = \mathcal{H}_{k-1} \cap \mathcal{C}_k,\ \forall k = 2, ..., K. \tag{6}$$

Note that from (5), the k-th history $\mathcal{H}_k$ contains all nodes in the first k cliques. The k-th separator $\mathcal{S}_k$ in (6) is the set of the common nodes between $\mathcal{H}_{k-1}$ and $\mathcal{C}_k$ such that all paths from the nodes in $\mathcal{H}_{k-1}\backslash\mathcal{S}_k$ to the nodes in $\mathcal{C}_k\backslash\mathcal{S}_k$ intersect $\mathcal{S}_k$. Every separator will share nodes with more than one clique and we say that these cliques are associated with that separator. In Fig. 2, the cliques labeled $C1$ and $C2$ are associated with the separator labeled $S2$.

Define $q(k)$ as the smallest index of any clique which is

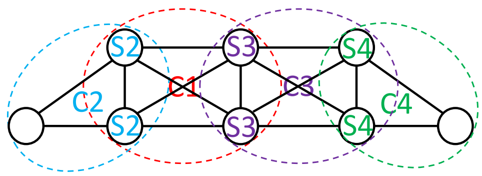

Fig. 2. The decomposable graphical model with 4 cliques and numbered separators.

associated with the k-th separator $\mathcal{S}_k$, that is,

$$q\left(k\right) \triangleq \min\left\{j \mid \mathcal{S}_k \subseteq \mathcal{C}_j\right\}, \ \forall k = 2, 3, ..., K. \tag{7}$$

We have freedom to number the cliques as we like and here we always use a popular numbering approach that is known to exist for any DGM (called perfect sequencing) which ensures the following inequality [38][1]:

$$q(k) < k. \tag{8}$$

Note that the k-th separator $\mathcal{S}_k$ is not only contained in the $q(k)$-th clique $\mathcal{C}_{q(k)}$ according to (7), but it is also contained in the k-th clique $\mathcal{C}_k$ based on (6), that is,

$$\mathcal{S}_k \subseteq \mathcal{C}_{q(k)}, \tag{9}$$

and

$$\mathcal{S}_k \subseteq \mathcal{C}_k. \tag{10}$$

The conclusions in (9) and (10) will be employed later in describing our distributed processing, as will the definitions.

Let $\mathcal{Q}_j$ denote the set of indices of separators which are associated with the j-th clique via the mapping q in (7), that is,

$$\mathcal{Q}_j \triangleq \left\{k \mid q\left(k\right) = j\right\}. \tag{11}$$

Note that by definition $\mathcal{Q}_j$ will not include the j-th separator. Thus, $\mathcal{Q}_j \cup \{j\}$ contains all the indices of the separators which are contained in the j-th clique. From (8), we know that the minimal element in $\mathcal{Q}_j$ satisfies

$$\min \mathcal{Q}_j > j, \tag{12}$$

which implies that

$$\mathcal{Q}_j \subseteq \left\{j+1, j+2, ..., K\right\}, \ \forall j = 1, 2, ..., K-1, \tag{13}$$

and

$$\mathcal{Q}_K = \emptyset. \tag{14}$$

1) Illustration of (7)–(14) using Fig. 2: Consider the example in Fig. 2. By employing (7), we observe that $q(2) = q(3) = 1$ and $q(4) = 3$ which implies $\mathcal{S}_2 \subseteq \mathcal{C}_1$, $\mathcal{S}_3 \subseteq \mathcal{C}_1$ and $\mathcal{S}_4 \subseteq \mathcal{C}_3$. By employing (10), we also obtain $\mathcal{S}_2 \subseteq \mathcal{C}_2$, $\mathcal{S}_3 \subseteq \mathcal{C}_3$ and $\mathcal{S}_4 \subseteq \mathcal{C}_4$. As per (8), $q(2) < 2$, $q(3) < 3$ and $q(4) < 4$. By employing (11), we obtain $\mathcal{Q}_1 = \{2, 3\}$ and $\mathcal{Q}_3 = \{4\}$ but $\mathcal{Q}_2 = \mathcal{Q}_4 = \emptyset$.

At each time slot i, let $X_{i,\mathcal{C}_k}$ denote the set of observations in $X_{i,[1,M]}$ that come from the nodes in the k-th clique. Let $X_{i,\mathcal{S}_k}$ denote the observations in $X_{i,[1,M]}$ that come from the nodes in the k-th separator set. For any DGM, from the fact that the ordered sequence of cliques $\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_K$ forms a perfect sequence, the joint distribution of $X_{i,[1,M]}$ follows the factorization [38][2]

$$f\left(X_{i,[1,M]}\right) = \frac{\prod_{k=1}^{K} f\left(X_{i,\mathcal{C}_k}\right)}{\prod_{k=2}^{K} f\left(X_{i,\mathcal{S}_k}\right)} \tag{15}$$

where $f(X)$ denotes the marginal pdf of X. Note that (15) can be derived using the pairwise Markov property in (4). The factorization in (15) provides us with a way to implement distributed processing among different cliques to solve the QCD problem which we will describe in detail in next section.

We note that there is extensive literature on manipulating decomposable graphs, which includes determining decomposability of a graph [39], determining their cliques [40], and constructing a decomposable version from a nondecomposable graph [41]. Throughout the paper, we concentrate on decomposable undirected graphical models.

III. Distributed Computation of the CUSUM Procedure in DGM

In this section, we begin by reviewing the well-known CUSUM procedure in the QCD problem [5], [10]. The QCD problem can be modeled as a hypothesis testing problem, given by

$$\begin{aligned} &H_0 : \text{no change occurs} \\ &H_1 : \text{change occurs at a finite unknown time slot } \tau. \end{aligned} \tag{16}$$

Note that when the change occurs, all sensors are assumed to be affected simultaneously as mentioned in *Assumption 1*.

The centralized CUSUM test statistic up to the current time slot n for (16) is

$$\tilde{W}_n = \max\left\{0, \max_{1 \le n' \le n} \log \frac{\prod_{i=1}^{n'-1} f_0\left(X_{i,[1,M]}\right) \prod_{i=n'}^{n} f_1\left(X_{i,[1,M]}\right)}{\prod_{i=1}^{n} f_0\left(X_{i,[1,M]}\right)}\right\} \tag{17}$$

$$= \max\left\{0, \max_{1 \le n' \le n} \sum_{i=n'}^{n} \log \frac{f_1(X_{i,[1,M]})}{f_0(X_{i,[1,M]})}\right\} \tag{18}$$

where $\tilde{W}_n = 0$ implies the decision maker does not declare change up to the current time slot n and will continue acquiring more observations. In this centralized QCD approach, at each time slot n, each sensor sends its observation to the FC. After receiving the data from all sensors, the FC calculates (18) and compares it to a threshold to decide whether to declare a change or continue to collect observations. In particular, the

[1] A perfect sequence of cliques can be found by using the method in Chapter 2.1.3 of [38]. We note that the existence of perfect sequences of cliques is a characteristic of decomposable graphs [38].

[2] $X_{i,[1,M]}$ following a decomposable graph is only a sufficient condition to guarantee (15). In other cases we can attempt to verify (15) directly.

classical centralized CUSUM procedure will raise an alarm at the time given by [10]

$$T_{GLR}(b) = \inf\left\{n \geq 1 : \tilde{W}_n \geq b\right\} \tag{19}$$

where the constant $b > 0$ needs to be chosen properly to satisfy the false alarm constraint in (1). The procedure in (19) is shown to be optimal for (3) in [5].

Next we introduce our distributed approach. We make the following additional assumptions throughout the paper.

Assumption 2: At each time slot i, we assume that $X_{i,[1,M]} \triangleq (X_{i,1}, X_{i,2}, ..., X_{i,M})$ satisfies the pairwise Markov property in (4) with respect to a given decomposable undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Assumption 3: Nodes in the same clique are close so that the energy cost of intra-clique communications is negligible compared to that of communications between the cliques to the FC. Additionally, the cliques have limited energy resources which are more greatly constrained than the FC. Thus, we focus on reducing communications from the cliques to the FC.

Assumption 4: The sets $\{\mathcal{S}_k\}_{k=2}^K$ and $\{\mathcal{C}_k\}_{k=1}^K$ do not change throughout the detection process. We generalize this in Section IV-D.

Next we develop our distributed computation approach. From (15), we have

$$\begin{aligned}
&\log \frac{f_1(X_{i,[1,M]})}{f_0(X_{i,[1,M]})} \\
&= \log \frac{\prod_{k=1}^{K} f_1\left(X_{i,\mathcal{C}_k}\right) \prod_{k=2}^{K} f_0\left(X_{i,\mathcal{S}_k}\right)}{\prod_{k=2}^{K} f_1\left(X_{i,\mathcal{S}_k}\right) \prod_{k=1}^{K} f_0\left(X_{i,\mathcal{C}_k}\right)} && (20)\\
&= \sum_{k=1}^{K} \log \frac{f_1\left(X_{i,\mathcal{C}_k}\right)}{f_0\left(X_{i,\mathcal{C}_k}\right)} - \sum_{k=2}^{K} \log \frac{f_1\left(X_{i,\mathcal{S}_k}\right)}{f_0\left(X_{i,\mathcal{S}_k}\right)} && (21)\\
&= \log \frac{f_1\left(X_{i,\mathcal{C}_1}\right)}{f_0\left(X_{i,\mathcal{C}_1}\right)} - \sum_{k\in\mathcal{Q}_1} \beta_k \log \frac{f_1\left(X_{i,\mathcal{S}_k}\right)}{f_0\left(X_{i,\mathcal{S}_k}\right)} \\
&\quad + \sum_{j=2}^{K} \left(\log \frac{f_1\left(X_{i,\mathcal{C}_j}\right)}{f_0\left(X_{i,\mathcal{C}_j}\right)} - \alpha_j \log \frac{f_1\left(X_{i,\mathcal{S}_j}\right)}{f_0\left(X_{i,\mathcal{S}_j}\right)} \right. \\
&\quad \left. - \sum_{k\in\mathcal{Q}_j} \beta_k \log \frac{f_1\left(X_{i,\mathcal{S}_k}\right)}{f_0\left(X_{i,\mathcal{S}_k}\right)} \right) && (22)\\
&= \sum_{k=1}^{K} L_k\left(X_{i,\mathcal{C}_k}\right) && (23)
\end{aligned}$$

where $\mathcal{Q}_j$ is defined in (11), and the set of non-negative coefficient pairs $\{(\alpha_k, \beta_k)\}_{k=2}^K$ satisfies

$$\alpha_k + \beta_k = 1,\ \forall k = 2, 3, ..., K. \tag{24}$$

Note that (23) expresses $\log f_1(X_{i,[1,M]})/f_0(X_{i,[1,M]})$ as a sum of the clique statistics $L_k\left(X_{i,\mathcal{C}_k}\right)$ for $k = 1, 2, ..., K$ that are computed at each clique. After (15) is used to obtain (20), we can group the separator terms into the associated clique terms based on the results in (9) and (10). In fact, each separator set $\mathcal{S}_k$ can be a member of several cliques. However, according to (9) and (10), we know $\mathcal{S}_k$ is always a member of two associated cliques $\mathcal{C}_k$ and $\mathcal{C}_{q(k)}$. For any term in (22) involving data coming from the k-th separator set $\mathcal{S}_k$, we allocate α_k percentage of that term to the k-th clique and β_k percentage to the $q(k)$-th clique[3] This allows us to obtain (22) from (21). The centralized change detection test statistic can always be expressed as the sum in (23) as long as (24) is satisfied. From (24), there are uncountably many choices of $\{\alpha_k, \beta_k\}_{k=2}^K$ which introduces flexibility in the definition of $L_k(X_{i,\mathcal{C}_k})$ in (23) while still ensuring local computation. In (23), $L_k(X_{i,\mathcal{C}_k})$ is defined as

$$L_1(X_{i,\mathcal{C}_1}) \triangleq \log \frac{f_1\left(X_{i,\mathcal{C}_1}\right)}{f_0\left(X_{i,\mathcal{C}_1}\right)} - \sum_{k\in\mathcal{Q}_1} \beta_k \log \frac{f_1\left(X_{i,\mathcal{S}_k}\right)}{f_0\left(X_{i,\mathcal{S}_k}\right)} \tag{25}$$

and for all $j = 2, 3, ..., K$,

$$\begin{aligned}
L_j(X_{i,\mathcal{C}_j}) \triangleq \log \frac{f_1\left(X_{i,\mathcal{C}_j}\right)}{f_0\left(X_{i,\mathcal{C}_j}\right)} - \alpha_j \log \frac{f_1\left(X_{i,\mathcal{S}_j}\right)}{f_0\left(X_{i,\mathcal{S}_j}\right)} \\
- \sum_{k\in\mathcal{Q}_j} \beta_k \log \frac{f_1\left(X_{i,\mathcal{S}_k}\right)}{f_0\left(X_{i,\mathcal{S}_k}\right)}.
\end{aligned} \tag{26}$$

Plugging (23) and (18) into (19) implies that the FC declares a change at time

$$T_{CS}(b) = \inf\left\{n \geq 1 : W_n \geq b\right\} \tag{27}$$

where the CUSUM statistic W_n is defined as

$$W_n \triangleq \max\left\{0, \quad \max_{1\leq n'\leq n} \sum_{i=n'}^{n} \sum_{k=1}^{K} L_k\left(X_{i,\mathcal{C}_k}\right)\right\}. \tag{28}$$

A nice property of the non-negative CUSUM statistic W_n is that it can be computed recursively as

$$W_n = \max\left\{0, \quad W_{n-1} + \sum_{k=1}^{K} L_k\left(X_{n,\mathcal{C}_k}\right)\right\} \tag{29}$$

with $W_0 = 0$. The above recursion is very useful because it requires little memory and is easily updated sequentially. Instead of directly sending all sensor observations to the FC, the distributed computation method provides the proper way to partition the sensor nodes into K local groups that correspond to the cliques. Each clique k will collect the information from the clique nodes to produce the clique statistic $L_k(X_{n,\mathcal{C}_k})$ using (25) or (26) and then transmit it to the FC. The FC will compute the CUSUM statistic W_n in (29) and compare it to the threshold b to decide whether to raise an alarm or continue the process. We note that when we employ (27) with $W_0 = 0$, WADD(n') in (2) is equal to [16]

$$\text{WADD}(n') = \mathbb{E}_1\left[n' - 1\right] \tag{30}$$

which implies that the average detection delay in the metric occurs at $\tau = 1$. The result in (30) makes the computation of the WADD in (2) straightforward.

While the distributed computation method takes advantage of the graph structure to aggregate the statistics and avoids unnecessary long range transmissions, it is also of interest to

[3] We do not allocate this term to other members that might also contain the k-th separator set.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2021.3077302, IEEE Transactions on Signal Processing

further reduce the number of transmissions by the cliques to the FC. This is addressed in the next section.

IV. Energy-Efficient QCD using Ordered transmissions

In the last section the proposed distributed computation method implements the optimum centralized CUSUM algorithm while taking advantage of the graph structure using cliques. In order to further reduce the number of long distance transmissions from the cliques to the FC, in this section we describe an ordered transmission approach, and a lower bound on the average number of transmissions saved is provided. The savings are shown to be large for cases of interest.

A. Ordered Transmissions for QCD

The idea of ordered transmissions for QCD is to order and then adaptively halt the transmissions of the clique statistics $\{L_k(X_{n,\mathcal{C}_k})\}_{k=1}^{K}$ during each time slot n. Specifically, after grouping the nodes into several cliques, the clique with the largest clique statistic magnitude transmits first and the cliques with smaller clique statistic magnitudes possibly transmit later. This process is repeated during each time slot. We will show that by sometimes halting transmissions before all K cliques have communicated their clique statistics, further transmissions can be saved while achieving the same detection delay as the optimal centralized CUSUM algorithm that requires all nodes to communicate their observations to the FC.

Algorithm 1 ordered-CUSUM.

Input: a positive constant b.
Initialize: $n = 0$, $W_0 = 0$ and a positive number η.

1: **while** $W_n < b$ **do**[4]
2: The FC updates time slot $n = n + 1$ and sets $j = 1$.
3: **for** $k = 1, 2, ..., K$ **do**
4: Clique $\mathcal{C}_k$ summarizes all the clique sensor data to produce $L_k(X_{n,\mathcal{C}_k})$ as per (24)–(26) at time t_n.
5: Clique $\mathcal{C}_k$ determines a time $t_{n,k} = t_n + \eta/|L_k(X_{n,\mathcal{C}_k})|$ to transmit $L_k(X_{n,\mathcal{C}_k})$ to the FC.
6: **end for**
7: Order cliques using $t_n < t_{n,k_1} \leq t_{n,k_2} \leq ... \leq t_{n,k_K} < t_{n+1}$ where k_j is the index of the clique which has the k_j-th largest $|\hat{L}_{n,k_j}|$ such that (31) holds.
8: **while** $j \leq K$ **do**
9: At time t_{n,k_j}, clique k_j transmits $\hat{L}_{n,k_j}$ to the FC.
10: **if** $W_{n,k_j} \leq \phi_{n,L}$ **then**
11: The FC decides $W_n = 0$.
12: break **while** loop (line 8).
13: **else**
14: The FC computes W_{n,k_j} via (32).
15: **end if**
16: The FC sets $j = j + 1$.
17: **end while**
18: **end while**
19: Declare the change occurs at the current time slot n and set $n' = n$.

---

[4]In practice it may be desired to stop at some large value of time slot n, even if $W_n \geq b$ has not been satisfied.

Our approach, which we call ordered-CUSUM, is summarized in Algorithm 1. At the beginning of the current time slot n (denoted time t_n) each clique k for $k = 1, ..., K$ determines a time $t_{n,k} = t_n + \eta/|L_k(X_{n,\mathcal{C}_k})|$ to transmit its local statistic $L_k(X_{n,\mathcal{C}_k})$ to the FC, where the positive number η can be made as small as the system will allow. Thus clique transmissions are time ordered using $t_n < t_{n,k_1} \leq t_{n,k_2} \leq ... \leq t_{n,k_K} < t_{n+1}$ where k_j is the index of the clique which has the k_j-th largest $|\hat{L}_{n,k_j}|$ such that

$$\left|\hat{L}_{n,1}\right| \geq \left|\hat{L}_{n,2}\right| \geq ... \geq \left|\hat{L}_{n,K}\right|. \tag{31}$$

In this way the cliques with larger-in-magnitude local test statistic transmit earlier. When the FC receives a new transmission from a clique k_j, it computes

$$W_{n,k_j} \triangleq W_{n-1} + \sum_{k=1}^{k_j} \hat{L}_{n,k} \tag{32}$$

where W_{n-1} is the CUSUM statistic at time slot $(n-1)$, and compares W_{n,k_j} from (32) with the updated threshold

$$\phi_{n,L} \triangleq -(K - k_j)\left|\hat{L}_{n,k_j}\right|. \tag{33}$$

By sending a message to all cliques, the FC stops any further clique transmission when $W_{n,k_j} \leq \phi_{n,L}$. When this occurs the FC declares $W_n = 0$ and the system progresses to the next time slot t_{n+1}. If all cliques transmit prior to $W_{n,k_j} \leq \phi_{n,L}$, then the current time slot is also ended and a decision is made using (27) and (32). If all transmission propagation delays are known and timing is synchronized, one can schedule all transmissions back to the FC so they arrive in the correct order. However, the process can easily be implemented with robustness to small timing errors. Even with inaccurate estimates of propagation delays or imperfect synchronization, since the FC receives the values to be ordered, the FC can put them back in order correctly as long as the FC waits a short period related to the uncertainty. By design, the ordered-CUSUM algorithm will always be optimal, as summarized in the following Theorem.

Theorem 1: Ordered-CUSUM achieves exactly the same performance metric as CUSUM, minimizing the WADD subject to a given false alarm constraint, while using a smaller number of transmissions.

Proof: In the ordered-CUSUM algorithm, during time slot n, when the FC receives a new clique statistic it updates the threshold $\phi_{n,L}$ via (33), and compares $\phi_{n,L}$ with $W_{n,k'}$ in (32). Let the most recent transmission be given by $\hat{L}_{n,k_j}$. Due to ordering it follows that $|\hat{L}_{n,k_j}|$ is an upper bound for those of the clique statistics which have not yet been transmitted. It further follows that the sum of the clique statistics that have not yet transmitted must be less than or equal to $(K - k_j)|\hat{L}_{n,k_j}|$, which is equal to $-\phi_{n,L}$. If $W_{n,k_j} \leq \phi_{n,L}$, then W_n has to be zero according to (29), regardless of the clique statistics that have not yet been transmitted. Hence, even without receiving further transmissions, the FC can implement the optimum centralized communication unconstrained approach and declare $W_n = 0$ at time slot n. On the other hand, the optimum centralized communication unconstrained approach continues to transmit at time slot n with nonzero probability. Thus, at

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2021.3077302, IEEE Transactions on Signal Processing

each time slot n, the average number of transmissions required by ordered-CUSUM is smaller than that of the optimum centralized communication unconstrained approach while the detection performance is the same. ■

While we have shown that the ordered-CUSUM algorithm built on ordering is more communication-efficient than the optimum centralized communication unconstrained approach, it is interesting to consider whether there exists a lower bound on the average number of transmissions saved by the ordered-CUSUM algorithm. This question will be addressed in the next subsection.

B. Lower Bound on the Average Number of Transmissions Saved

In this subsection, we derive a lower bound on the average number of transmissions saved by ordered-CUSUM. The following theorem formally describes the communication saving lower bound for each time slot n.

Theorem 2: Consider the QCD problem described in (3). When using ordered-CUSUM, for any choice of the pairs $\{(\alpha_k, \beta_k)\}_{k=2}^{K}$ with $K \geq 2$ and $\alpha_k + \beta_k = 1$, the average number of transmissions saved S_n for time slot n is bounded from below by

$$S_n > \left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \Pr\left(W_{n-1} = 0 \text{ and } L_k(X_{n,\mathcal{C}_k}) < 0, \forall k\right). \tag{34}$$

Proof: According to ordered-CUSUM, if W_{n,k_j} defined in (32) is smaller than the threshold $\phi_{n,L}$ in (33), then transmissions will be stopped during time slot n and the algorithm proceeds to the next time slot $(n+1)$. Let

$$k_n^* \triangleq \min \left\{ 1 \leq k' < K : W_{n-1} + \sum_{k=1}^{k'} \hat{L}_{n,k} < -(K - k') \left| \hat{L}_{n,k'} \right| \right\} \tag{35}$$

denote the number of necessary transmissions during time slot n using ordered-CUSUM, and define

$$S_n \triangleq \mathbb{E}\left[K - k_n^*\right] \tag{36}$$

as the average number of transmissions saved at time slot n. Then from (36), S_n can be bounded from below by

$$\begin{aligned} S_n &= \sum_{k=1}^{K} (K - k) \Pr(k_n^* = k) && (37) \\ &\geq \sum_{k=1}^{\lfloor K/2 \rfloor + 1} (K - k) \Pr(k_n^* = k) && (38) \\ &> \left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \sum_{k=1}^{\lfloor K/2 \rfloor + 1} \Pr(k_n^* = k) && (39) \\ &= \left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \Pr\left(k_n^* \leq \left\lfloor \frac{K}{2} \right\rfloor + 1 \right). && (40) \end{aligned}$$

The result in (38) is obtained by dropping some non-negative terms in (37). In going from (38) to (39), we bound $(K - k)$ by using $(\lceil K/2 \rceil - 1)$ for $k = 1, ..., \lfloor K/2 \rfloor + 1$. Plugging the definition of k_n^* from (35) into (40), we obtain

$$\begin{aligned} S_n &> \left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \Pr\left(W_{n-1} + \sum_{k=1}^{\lfloor K/2 \rfloor + 1} \hat{L}_{n,k} \leq -\left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \left| \hat{L}_{n, \lfloor \frac{K}{2} \rfloor + 1} \right| \right) && (41) \\ &\geq \left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \Pr\left(W_{n-1} = 0, \ \hat{L}_{n,k} < 0, \forall k, \text{ and } W_{n-1} + \sum_{k=1}^{\lfloor K/2 \rfloor + 1} \hat{L}_{n,k} \leq -\left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \left| \hat{L}_{n, \lfloor \frac{K}{2} \rfloor + 1} \right| \right) && (42) \\ &\geq \left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \Pr\left(W_{n-1} = 0 \text{ and } \hat{L}_{n,k} < 0, \forall k \right) && (43) \\ &= \left(\left\lceil \frac{K}{2} \right\rceil - 1 \right) \Pr\left(W_{n-1} = 0 \text{ and } L_k(X_{n,\mathcal{C}_k}) < 0, \forall k \right). && (44) \end{aligned}$$

In going from (41) to (42), we add two extra constraints which will maintain or reduce the probability. In (42), when $W_{n-1} = 0$ and $\hat{L}_{n,k} < 0, \forall k$ are true, $W_{n-1} + \sum_{k=1}^{\lfloor K/2 \rfloor + 1} \hat{L}_{n,k} \leq -(\lceil K/2 \rceil - 1)|\hat{L}_{n,\lfloor K/2 \rfloor + 1}|$ must be true which implies the result in (43). In going from (43) to (44), we use the fact that all of the ordered statistics $\hat{L}_{n,k}$ being negative implies all of the original unordered statistics $L_k(X_{n,\mathcal{C}_k})$ are negative. This completes the proof. ■

We point out that the lower bound in (44) is very general and is valid for any DGM and any choice of the set of non-negative coefficient pairs $\{(\alpha_k, \beta_k)\}_{k=2}^{K}$ with $\alpha_k + \beta_k = 1$. The result in (44) indicates that the lower bound on S_n depends on the number of cliques and the joint statistics of W_{n-1} and $L_k(X_{n,\mathcal{C}_k})$. In the next subsection we show that the savings can be large for several cases of interest.

C. Large Saving Gains for Several Cases of Interest

Consider a distance measure s between the distributions of the sensor observations before and after the change time τ. The distance measure s is assumed to satisfy the following mild condition.

Assumption 5: For the hypothesis testing problem considered in (16) with $L_k(X_{n,\mathcal{C}_k})$ as per (24)–(26), we assume that the probability $\Pr(L_k(X_{n,\mathcal{C}_k}) < 0) \to 1$ as $s \to \infty$ for all $k = 1, ..., K$ and $n < \tau$.

Intuitively, with a large distance between the distributions of the sensor observations before and after the change time, it should be easy for the FC to decide when the change occurs. At the end of this subsection, we provide two popular general QCD problems and the corresponding distance measure to illustrate that *Assumption 5* is reasonable.

Under *Assumption 5*, the following theorem describes the limiting behavior of the lower bound on the total number of transmissions saved by ordered-CUSUM.

Theorem 3: Under *Assumptions 1–5*, consider the approach in Algorithm 1 for the QCD problem in (3). With a sufficiently large s, the total number of transmissions saved over the optimum centralized communication unconstrained approach increases at least as fast as proportional to K while the detection delay is not affected. In particular, the total number of transmissions saved is lower bounded by $(\lceil K/2\rceil-1)(\tau-1)$.

Proof: From (34) in *Theorem 2*, we have

$$S_n > \left(\left\lceil \frac{K}{2} \right\rceil - 1\right) \Pr\left(W_{n-1}=0 \text{ and } L_k(X_{n,\mathcal{C}_k})<0, \forall k\right) \tag{45}$$

Next we employ induction to show that as $s\to\infty$, for $n<\tau$, we have

$$\Pr(W_{n-1}=0 \text{ and } L_k(X_{n,\mathcal{C}_k})<0,\forall k)\to 1. \tag{46}$$

Throughout this proof, we only consider the time slots before the change occurs which means we focus on the case when $n<\tau$. Specifically, we set $W_0=0$. For a sufficiently large s, *Assumption 5* implies that all clique statistics are negative for $n<\tau$. Thus, the probability $\Pr(W_0=0$ and $L_k(X_{1,\mathcal{C}_k})<0,\forall k)\to 1$ implies $\Pr(W_1=0)\to 1$. If we assume $\Pr(W_{n-2}=0)\to 1$ at time slot $(n-1)$, then under *Assumption 5* we have $\Pr(W_{n-2}=0$ and $L_k(X_{n-1,\mathcal{C}_k})<0,\forall k)\to 1$ for $n<\tau$ which implies $\Pr(W_{n-1}=0)\to 1$. Thus, at time slot n, we obtain $\Pr(W_{n-1}=0$ and $L_k(X_{n,\mathcal{C}_k})<0,\forall k)\to 1$ for $n<\tau$. From (45), for all $n<\tau$, we have

$$S_n > \left\lceil \frac{K}{2} \right\rceil - 1. \tag{47}$$

Finally, it follows that the total number of transmissions saved K_s can be bounded by

$$K_s > \left(\left\lceil \frac{K}{2} \right\rceil - 1\right)(\tau-1). \tag{48}$$

■

As illustrated by *Theorem 3*, the total number of transmissions saved by ordering the communications from the cliques to the FC increases at least as fast as linearly proportional to the number of cliques K while achieving the same detection delay as the optimum centralized communication unconstrained approach. *Theorem 3* also states that more transmissions can be saved as the change time increases. In the following, we provide two general problems and the corresponding distance measure.

Example 1: Consider detecting a change in the mean of a sequence of sensor observation vectors following a multivariate Gaussian distribution as[5]

$$\begin{array}{ll} X_{n,[1,M]} \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma}) & \text{when } n<\tau \\ X_{n,[1,M]} \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{\Sigma}) & \text{when } n\ge\tau \end{array}, \tag{49}$$

where $\boldsymbol{\mu}\neq\mathbf{0}$ and the known covariance matrix $\mathbf{\Sigma}$ is assumed to be positive definite. In this problem, we can choose $s=\min_k\|\boldsymbol{\mu}_{\mathcal{C}_k}\|$ as the distance measure where $\boldsymbol{\mu}_{\mathcal{C}_k}$ denotes the mean vector of the nodes in the k-th clique with ℓ_2-norm $\|\boldsymbol{\mu}_{\mathcal{C}_k}\|$.

Example 2: Consider detecting a change in the covariance matrix of a sequence of sensor observation vectors following a multivariate Gaussian distribution as

$$\begin{array}{ll} X_{n,[1,M]} \sim \mathcal{N}(\mathbf{0},\mathbf{I}) & \text{when } n<\tau \\ X_{n,[1,M]} \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma}) & \text{when } n\ge\tau \end{array}, \tag{50}$$

where $\mathbf{I}$ is an identity matrix and the known covariance matrix $\mathbf{\Sigma}$ is assumed to be positive definite. In this problem, $s=\min_k\lambda_{\min,k}$ where $\lambda_{\min,k}$ is the minimum eigenvalue of $\mathbf{\Sigma}_{\mathcal{C}_k}$ which denotes the covariance matrix associated with $X_{n,\mathcal{C}_k}$ for $n\ge\tau$. The following theorem shows that *Assumption 5* is valid for *Examples 1* and *2*.

Theorem 4: Consider the problems in *Example 1* and *Example 2* and the ordered transmission approach described in the ordered-CUSUM algorithm which employs (23) with

$$\alpha_k = 1-2^{K-k}\xi, \tag{51}$$

and

$$\beta_k = 2^{K-k}\xi, \tag{52}$$

for all $k=2,3,...,K$ using any ξ which satisfies

$$\xi \in \left(0,\frac{1}{2^{K-1}-1}\right). \tag{53}$$

For any $k=1,2,...,K$ with $K>2$, with sufficiently large $\min_k\|\boldsymbol{\mu}_{\mathcal{C}_k}\|$ for (49) or sufficiently large $\min_k\lambda_{\min,k}$ for (50), we have for all $k=1,2,...,K$ and $n<\tau$,

$$\Pr(L_k(X_{n,\mathcal{C}_k})<0)\to 1. \tag{54}$$

Proof: The proof of this theorem is omitted, since it follows from the proof in [28] and [29]. Specifically, for the problem in (49), as $s\to\infty$ with $s=\min_k\|\boldsymbol{\mu}_{\mathcal{C}_k}\|$, the result in (54) can be obtained by following *Theorem 2* in [28]. For the problem in (50), as $s\to\infty$ with $s=\min_k\lambda_{\min,k}$, then (54) can be obtained by following *Theorem 3* in [29]. ■

D. Extensions: Graph Structure Change and Nondecomposable

In this subsection, we generalize *Assumption 4* and consider the case where the graph is fixed under each hypothesis but the graph structure of $f_0\left(X_{n,[1,M]}\right)$ (relevant for $n<\tau$) is not the same as $f_1\left(X_{n,[1,M]}\right)$ (relevant for $n\ge\tau$). Suppose the sensor observations $X_{n,[1,M]}$ obey the pairwise Markov property with respect to a decomposable graph $\mathcal{G}_1=(\mathcal{V}_1,\mathcal{E}_1)$ before the change $(n<\tau)$ and a decomposable graph $\mathcal{G}_2=(\mathcal{V}_2,\mathcal{E}_2)$ after the change $(n\ge\tau)$ with $\mathcal{G}_1\neq\mathcal{G}_2$. In order to implement distributed computation and ordered transmissions for this case, the sequence of cliques $\{\mathcal{C}_k\}_{k=1}^K$ is derived based on $\mathcal{G}=\mathcal{G}_1\cup\mathcal{G}_2$ instead of $\mathcal{G}_1$ or $\mathcal{G}_2$. Compared to $\mathcal{G}_1$ and $\mathcal{G}_2$, the graph structure $\mathcal{G}$ possibly increases the size of some cliques, implying extra node data needs to be collected in these larger cliques. However, when we employ the known pdf $f_0\left(X_{n,[1,M]}\right)$ or $f_1\left(X_{n,[1,M]}\right)$ in these computations, we use some of the extra data in computations involving $f_0\left(X_{n,[1,M]}\right)$ and the rest in computations involving $f_1\left(X_{n,[1,M]}\right)$. Thus, the graph $\mathcal{G}$ allows the computations that either $\mathcal{G}_1$ or $\mathcal{G}_2$ require.

Consider the example illustrated in Fig. 3. Before the change, the graph structure is indicated by graph $\mathcal{G}_1$ which has two clique sets $\mathcal{C}_1=\{1,2,3\}$ and $\mathcal{C}_2=\{2,4\}$ along with one separator set $\mathcal{S}_2=\{2\}$. After the change occurs, the graph

[5] $\mathcal{N}(\mathbf{a},\mathbf{A})$ denotes the multivariate Gaussian distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$.

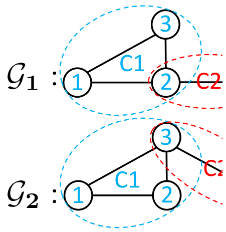

Fig. 3. Choice of cliques and

structure illustrated in $\mathcal{G}$
and $\mathcal{C}_2 = \{3,4\}$ along
order to keep the clique
the detection process, w
graph $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2$ whc
$\mathcal{C}_2 = \{2,3,4\}$. We also
These sets are respectiv

the separator set through the whole detection process. Note that compared to $\mathcal{C}_2 = \{2,4\}$ in $\mathcal{G}_1$, $\mathcal{G}$ defines a new larger clique $\mathcal{C}_2 = \{2,3,4\}$ to indicate that the data at node 3 will be collected at clique $\mathcal{C}_2$. However, when we compute $f_0\left(X_{n,[1,4]}\right)$ in a distributed way, we do not really use the data from node 3 in $\mathcal{C}_2$ to compute $f_0\left(X_{n,\mathcal{C}_2}\right)$. Similarly, when we compute $f_1\left(X_{n,[1,4]}\right)$, we do not really use the data from node 2 in $\mathcal{C}_2$ to compute $f_1\left(X_{n,\mathcal{C}_2}\right)$. After implementing the distributed computation using the clique and separator sets of $\mathcal{G}$, ordered transmissions can be developed according to Section IV. It is worth mentioning that the union operation might decrease the number of cliques which can degrade the gains of distributed processing and ordering.

A nondecomposable graphical model. Throughout the paper we have assumed the graph to be decomposable. However, we can apply our algorithm to nondecomposable graphs as follows. If we start with a nondecomposable graph, it is always possible to add edges to make it decomposable. Then, if we make the added edge weights sufficiently small, we will get results very close to those for the original problem. In the next section, we illustrate these ideas with an example.

V. Numerical Results

In this section, numerical examples for two representative classes of decomposable graphical models (chain structure and tree structure) are presented in order to illustrate the communication saving performance using the proposed ordered transmission approach. Chain structure and tree structure graphs have been employed in studies on feature representation [42], topology identification [43], structure learning [44] and electrical power systems [37].

A. Total number of Transmissions Saved versus the Distance Measure

In this subsection, the lower bound in (34) is compared with the actual number of transmissions saved by ordered-CUSUM from Monte Carlo simulations (1000 runs). Consider a graph with chain structure as illustrated in Fig. 4 where we set the number of cliques $K = 50$. As indicated in Fig. 4, each clique

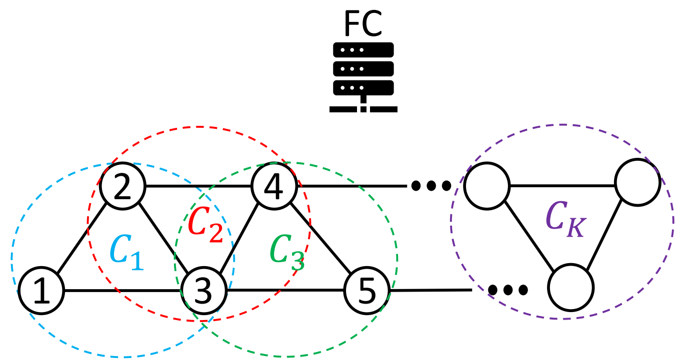

Fig. 4. The decomposable graphical model with chain structure. Clique 1: Nodes 1,2,3; Separator: Nodes 2,3; Clique 2: Nodes 2,3,4; Separator: Nodes 3,4; and so on down the chain. Each clique communicates with the fusion center (FC).

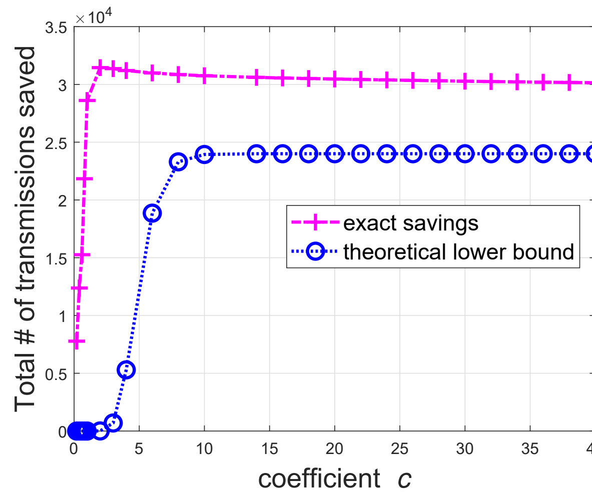

Fig. 5. Impact of mean shift on the total number of transmissions saved when the change does not occur during these 10^3 time slots.

has 3 nodes, and every two-connected clique pair are coupled through a 2-sensor separator set. We first consider the change detection problem in (49) and generate a covariance matrix which satisfies the conditional independence specified by the graph structure in Fig. 4. In the simulation results of Fig. 5, we set $\tau = 1$, $\xi = 0.5/(2^{49}-1)$ and $\boldsymbol{\mu} = c[1,1,...,1]^\top$. In order to satisfy the false alarm constraint $\mathbb{E}_\infty(n) \geq \gamma = 10^3$, the minimum value of the positive constant b is found using grid search with the grid points spaced 0.01 apart. Note that hereafter in the simulation results, we only count the number of transmissions from the cliques to the FC for the first 10^3 time slots. In Fig. 5, we plot the total number of transmissions saved versus c when the change does not occur during these 10^3 time slots. Fig. 5 indicates that our theoretical lower bound in (34) is valid and its value increases as c increases which means the distance measure s increases. As expected from our analysis, Fig. 5 shows that the lower bound on the total number of transmissions saved nearly equals 24000 when $c = 40$ which is consistent with *Theorem 3* since $(\lceil K/2 \rceil - 1) \times 10^3 = 24000$ when $K = 50$.

B. Total Number of Transmissions Saved versus the Number of Cliques

In this subsection, using Monte Carlo simulations (1000 runs), we investigate the total number of transmissions saved for the first 10^3 time slots by ordering the communications from the cliques to the FC for different number of cliques K for the case when no change occurs during these 10^3 time slots. We consider the testing problem in (49) with the same class of graph structures as in Fig. 4. We plot the total number of transmissions saved when no change occurs during these 10^3 time slots versus K in Fig. 6 for the parameters $\tau = 1$, $\gamma = 10^3$, $\xi = 0.5/(2^{K-1} - 1)$ and $\boldsymbol{\mu} = c[1, 1, ..., 1]^\top$. For comparison, the limiting theoretical lower bound on communication savings in *Theorem 3* is also provided. For the specific cases considered, Fig. 6 indicates that the total number of transmissions saved by Ordered-CUSUM increases approximately linearly with K for every value of c. It also indicates that the rate of increase with K increases with increasing c for smaller c but eventually, the rate of increase saturates as c becomes large, corresponding to a large and easily detectable change.

With the same parameter setting as above, Fig. 7 shows the WADD, which is computed according to (30), as a function of the number of cliques K for ordered-CUSUM and CUSUM with $c = 0.1$ and $c = 0.5$. It indicates that our ordered-CUSUM algorithm provides the same detection delay as classical CUSUM. It also shows that a larger value of c results in a smaller WADD which follows because a change with a larger value of c is easier to detect. The result in Fig. 7 illustrates that the WADD decreases as we increase the number of cliques K which is consistent with our intuition that more observations per time slot can help us detect the change quickly.

Next, we consider a different class of graphs with the tree structure as illustrated in Fig. 8 where each clique contains 4 nodes and every two-connected clique pair are coupled through a 1-sensor separator set. Here we consider the testing problem in (50). We set $\tau = 1$, $\gamma = 10^3$ and $\xi = 0.5/(2^{K-1} - 1)$. The diagonal elements of $\boldsymbol{\Sigma}_{\mathcal{C}_k}$ for all k are set to be x^2 and the other elements of $\boldsymbol{\Sigma}_{\mathcal{C}_k}$ are set to equal to $x/10$ where the minimum value of $\boldsymbol{\Sigma}_{\mathcal{C}_k}$ is $x^2 - x/10$, so its value may be changed by varying x. In Fig. 9, we plot the total number of transmissions saved by ordered-CUSUM when the change does not occur during 10^3 time slots versus K for different values of x. Fig. 9 implies that the total number of transmissions saved increases approximately linearly with K for every value of x. Fig. 9 also indicates that when x is relatively small then increasing x increases the slope which is very similar to the result in Fig. 6.

C. Performance of Ordering with Imprecise Model

Now we apply the proposed ordered-CUSUM algorithm to implement QCD in a real-world Abilene network [45] for the case when the model is imprecise. The Abilene network describes internet traffic between universities in the United States with its topology given in Fig. 10 where we have 11 routers and 30 links (each edge in Fig. 10 corresponds to two links, one in each direction). Here we are interested in the relationship between the traffic on the different links. Similar to [45], we use the links as the vertices of the conditional independence graph. The graphical model describing the conditional independence looks different from the physical network graph in Fig. 10, and is given in Fig. 11. In the resulting conditional independence graph, two separators $\mathcal{S}_2 = \{\text{DNVR-KSCY}, \text{SNVA-KSCY}, \text{LOSA-HSTN}\}$ and $\mathcal{S}_3 = \{\text{KSCY-IPLS}, \text{HSTN-ATLA}\}$ separate the other three link subsets, which correspond to three cliques. The cardinalities of the three cliques and two separator sets are $|\mathcal{C}_1| = 14$, $|\mathcal{C}_2| = 12$, $|\mathcal{C}_3| = 14$, $|\mathcal{S}_2| = 6$, and $|\mathcal{S}_3| = 4$.

We consider the change detection problem where $X_{n,[1,M]} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$ when $n < \tau$ and $X_{n,[1,M]} \sim \mathcal{N}(\mathbf{0}, 3\boldsymbol{\Sigma}_1)$ when $n \geq \tau$, and randomly generate a covariance matrix $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{30\times 30}$ which satisfies the conditional independence specified by the graphical model in Fig. 10. In order to mimic an imprecise model, we add independent uniform random variables, uniformly distributed over $(-e, 0)$, to those entries of $\boldsymbol{\Sigma}_1^{-1}$ describing a precision[6] involving nodes in $\mathcal{C}_1$ to obtain a distorted precision matrix $\tilde{\boldsymbol{\Sigma}}_1^{-1}$. We generate $X_{n,[1,M]} \sim \mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}_1)$ when $n < \tau$ to test ordered-QCD, while using an undistorted model when $n \geq \tau$. In the simulation results of Fig. 12, we set $\tau = 1$, $\xi = 0.5/(2^2 - 1)$ and $\gamma = 10^3$. In Fig. 12, we plot the Kullback–Leibler (KL)-divergence [46] between two distributions $\mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}_1)$ and $\mathcal{N}(\mathbf{0}, 3\boldsymbol{\Sigma}_1)$ and that between $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}(\mathbf{0}, 3\boldsymbol{\Sigma}_1)$, denoted as $D_{KL}(\tilde{\boldsymbol{\Sigma}}_1 \| 3\boldsymbol{\Sigma}_1)$ and $D_{KL}(\boldsymbol{\Sigma}_1 \| 3\boldsymbol{\Sigma}_1)$ respectively, versus the model error parameter e. In Fig. 12, we also plot the WADD versus e for ordered-CUSUM with and without model error. Fig. 12 indicates the WADD increases as we increase the model error parameter e, which is possible since as model error e is increased, the KL divergence between $\tilde{\boldsymbol{\Sigma}}_1$ and $3\boldsymbol{\Sigma}_1$ in this particular case is decreased which implies that the QCD problem becomes more difficult. The result in Fig. 12 indicates that even when the model is imprecise, ordered-CUSUM can still detect the change but possibly with a larger detection delay.

D. Nondecomposable Graphical Model

Now we investigate the performance of ordering with a nondecompoasble graphical model when using the approximate approach suggested at the end of Section IV. Specifically, we consider the change detection problem in (49) with $\boldsymbol{\mu} = [0.02, 0.02]^\top$ and generate a covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{2\times 2}$ which is specified by the nondecomposable graph in Fig. 1(b) where all nodes need to transmit their data to the FC. By adding a chord as illustrated in Fig. 1(c) with precision (edge weight) δ, we can implement ordering and obtain an approximated version of $\boldsymbol{\Sigma}$, denoted as $\tilde{\boldsymbol{\Sigma}}$, when $|\delta|$ is small. In Fig.13, we plot the WADD versus $|\delta|$ for distributed and centralized processing for the parameters $\tau = 1$, $\gamma = 10^3$, and $\alpha_2 = \beta_2 = 0.5$ where $\tilde{\boldsymbol{\Sigma}}$ (distributed) and $\boldsymbol{\Sigma}$ (centralized) are employed in (49), respectively. The result in Fig.13 illustrates that the WADD in the modified decomposable graph case is

[6]The inverse of the covariance matrix is called the precision matrix whose entries describe the precision between two variables.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2021.3077302, IEEE Transactions on Signal Processing

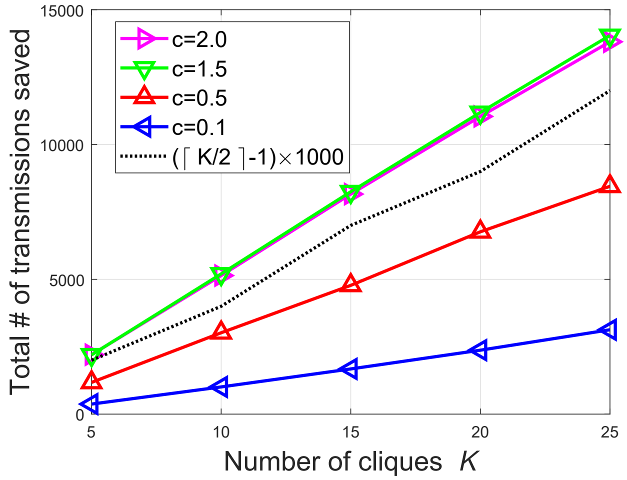

Fig. 6. The total number of transmissions saved when the change does not occur during these 10^3 time slots versus K for the model illustrated in Fig. 4.

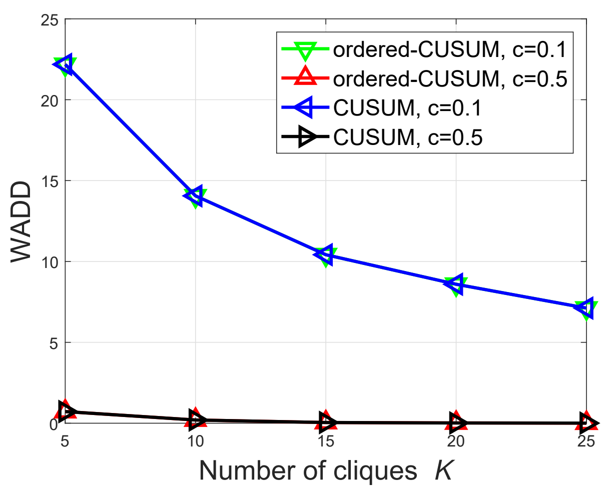

Fig. 7. The WADD versus the number of cliques K for $c = 0.1$ and $c = 0.5$ for the model illustrated in Fig. 4.

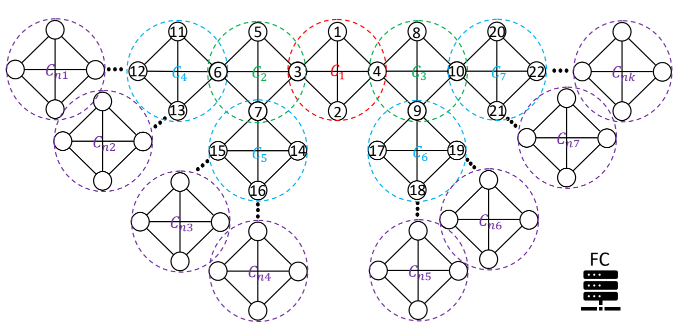

Fig. 8. The decomposable graphical model with tree structure. Clique 1: Nodes 1,2,3,4; Separator: Nodes 3,4; Clique 2: Nodes 3,5,6,7; Separator: Nodes 6,7; and so on. Each clique communicates with the fusion center (FC).

very close to that for the original problem as $|\delta|$ is sufficiently small (regardless whether δ is positive or negative). Note that the sign of δ can result in an increase or decrease in the

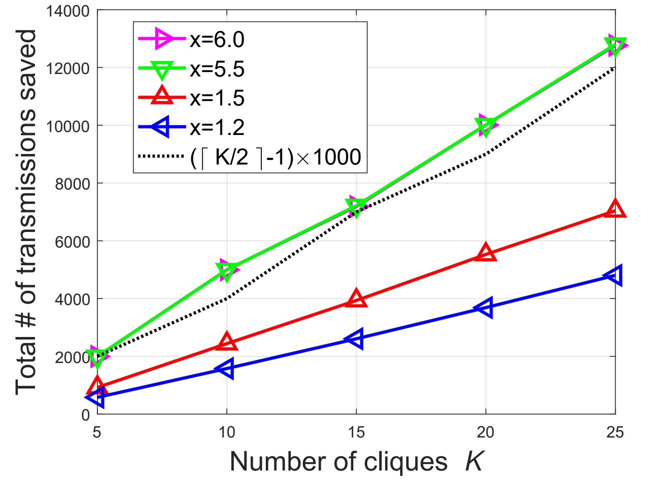

Fig. 9. The total number of transmissions saved when the change does not occur during these 10^3 time slots versus K for the model illustrated in Fig. 8.

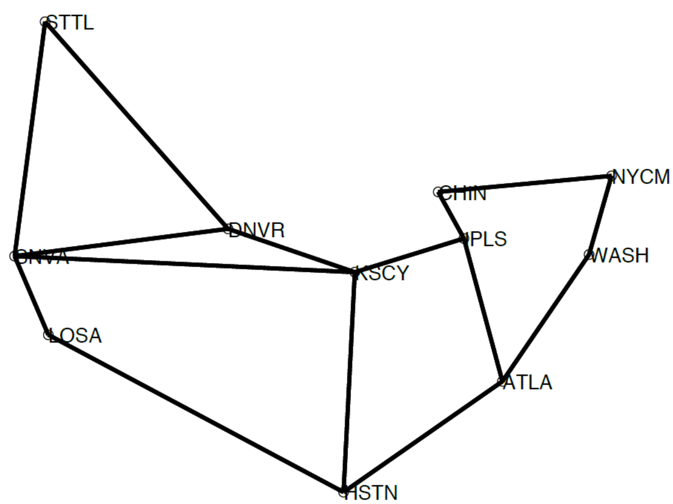

Fig. 10. Topology of the Abilene network [45] where each edge corresponds to two links.

WADD, because the perturbation can make the test slightly easier or more difficult.

VI. Conclusion

In this paper a new class of communication-efficient QCD schemes for sensor networks have been developed that reduce the number of transmissions without any impact on detection delay when compared to the optimum centralized communication unconstrained QCD approach. It is assumed that the observations follow a decomposable graphical model (DGM), which is a very broad class of network topologies, and the observations between sensors may be dependent. For a QCD problem with sensor observations following any DGM, we write the optimum centralized change detection test statistic as a sum of clique statistics where each clique statistic can be computed only using local data available at the corresponding clique. To complete the computation of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2021.3077302, IEEE Transactions on Signal Processing

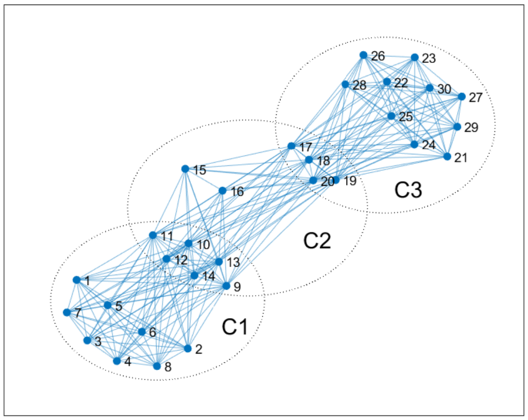

Fig. 11. Graphical model of the Abilene network [45] where the links in Fig. 10 are used as the vertices.

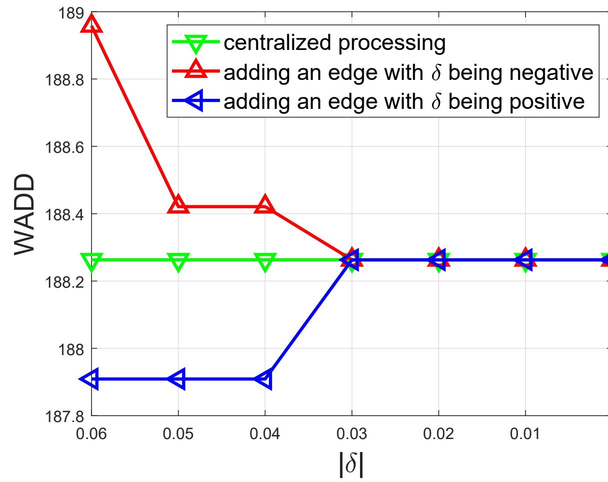

Fig. 13. The WADD versus $|\delta|$ (from large to small) for the model illustrated in Fig. 1(b) with its decomposable version given in Fig. 1(c).

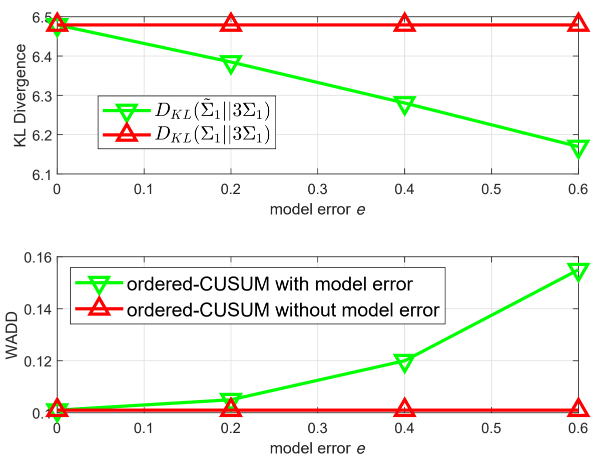

Fig. 12. The KL divergence and the WADD versus the model error e in a real-world Abilene network in Fig. 10. The KL divergence between two distributions $\mathcal{N}(\mathbf{0}, \Sigma_1)$ and $\mathcal{N}(\mathbf{0}, 3\Sigma_1)$ is denoted as $D_{KL}(\Sigma_1\|3\Sigma_1)$.

the optimum centralized test, each clique forwards its clique statistic to the FC.

In order to further improve the communication efficiency, we have applied the ordered transmission approach over the cliques to reduce the number of transmissions from the cliques to the FC without performance loss. In the ordered transmission approach, the cliques with more informative observations transmit their clique statistics to the FC first. Transmissions are halted after sufficient evidence is accumulated to save transmissions, and a new round of sensing is initiated. Furthermore, a lower bound on the average number of transmissions saved has been provided. When a well-behaved distance measure between the pdfs of the sensor observations before and after the change becomes sufficiently large, the lower bound approaches approximately half the number of cliques. Extensions to the case where the graph structure changes have been discussed. In order to illustrate our theoretical analysis, two popular general QCD problems with sensor observations following a multivariate Gaussian distribution have been considered and numerical results have been provided which are consistent with the analytical findings.

Our results lead to some interesting open questions and opportunities for further work. First, showing the ordered transmission approach is optimal in terms of communication savings or further improving it would be interesting. Second, this paper only develops a lower bound on communication savings, but more exact categorization would be useful. Finally, our study focuses on the case where the distributions of the observations at all sensor change simultaneously, and generalizations would be useful.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] C.-Y. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.

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

[4] 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.

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

[6] Y. Ritov, "Decision theoretic optimality of the cusum procedure," *The Annals of Statistics*, pp. 1464–1469, 1990.

[7] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.

[8] ——, *Optimal stopping rules*. Springer Science & Business Media, 2007, vol. 8.

[9] ——, "The detection of spontaneous effects," in *Doklady Akademii Nauk*, vol. 138, no. 4. Russian Academy of Sciences, 1961, pp. 799–801.

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

[11] 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.

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

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

[14] 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.

[15] D. Siegmund, *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media, 2013.

[16] V. V. Veeravalli and T. Banerjee, "Quickest change detection," in *Academic Press Library in Signal Processing*, 2014, vol. 3, pp. 209–255.

[17] 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.

[18] M. Basseville, I. V. Nikiforov *et al.*, *Detection of abrupt changes: theory and application*. Prentice hall Englewood Cliffs, 1993, vol. 104.

[19] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press, 2008.

[20] V. Raghavan and V. V. Veeravalli, "Quickest change detection of a markov process across a sensor array," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1961–1981, 2010.

[21] 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.

[22] H. Zhang, O. Hadjiliadis, T. Schafer, and H. V. Poor, "Quickest detection in coupled systems," *SIAM Journal on Control and Optimization*, vol. 52, no. 3, pp. 1567–1596, 2014.

[23] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 554–568, 1996.

[24] Y. Mei, "Quickest detection in censoring sensor networks," in *2011 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2011, pp. 2148–2152.

[25] K. Liu, Y. Mei, and J. Shi, "An adaptive sampling strategy for online high-dimensional process monitoring," *Technometrics*, vol. 57, no. 3, pp. 305–319, 2015.

[26] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3229–3235, 2008.

[27] E. Ayeh, K. Namuduri, and X. Li, "Statistical ranking of sensor observations for centralized detection with distributed sensors," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 6337–6341.

[28] J. Zhang, Z. Chen, R. S. Blum, X. Lu, and W. Xu, "Ordering for reduced transmission energy detection in sensor networks testing a shift in the mean of a gaussian graphical model," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2178–2189, 2017.

[29] Y. Chen, R. S. Blum, B. M. Sadler, and J. Zhang, "Testing the structure of a gaussian graphical model with reduced transmissions in a distributed setting," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5391–5401, 2019.

[30] Y. Chen, R. Blum, and B. Sadler, "Optimal quickest change detection in sensor networks using ordered transmissions," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (IEEE SPAWC 2020)*, Atlanta, USA, May 2020.

[31] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 342–351, 2011.

[32] Z. Wang and R. S. Blum, "A statistical learning-based algorithm for topology verification in natural gas networks based on noisy sensor measurements," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3653–3666, 2020.

[33] D. Sun and X. Sun, "Estimation of the multivariate normal precision and covariance matrices in a star-shape model," *Annals of the Institute of Statistical Mathematics*, vol. 57, no. 3, pp. 455–484, 2005.

[34] Y. Xiang, S. K. M. Wong, and N. Cercone, "A "microscopic" study of minimum entropy search in learning decomposable markov networks," *Machine Learning*, vol. 26, no. 1, pp. 65–92, 1997.

[35] K. Mohan, P. London, M. Fazel, D. Witten, and S.-I. Lee, "Node-based learning of multiple gaussian graphical models," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 445–488, 2014.

[36] M. Cetin, L. Chen, J. W. Fisher, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42–55, 2006.

[37] Y. Weng, R. Negi, and M. D. Ilić, "Graphical model for state estimation in electric power systems," in *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2013, pp. 103–108.

[38] S. L. Lauritzen, *Graphical models*. Clarendon Press, 1996, vol. 17.

[39] S. L. Lauritzen, T. Speed, and K. Vijayan, "Decomposable graphs and hypergraphs," *Journal of the Australian Mathematical Society*, vol. 36, no. 1, pp. 12–29, 1984.

[40] A. M. Childs, E. Farhi, J. Goldstone, and S. Gutmann, "Finding cliques by quantum adiabatic evolution," *arXiv preprint quant-ph/0012104*, 2000.

[41] D. Barber, *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[42] J. Liu and J. Ye, "Moreau-yosida regularization for grouped tree structure learning," in *Advances in neural information processing systems*, 2010, pp. 1459–1467.

[43] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2682–2694, 2016.

[44] V. Y. Tan, A. Anandkumar, and A. S. Willsky, "Learning gaussian tree models: Analysis of error exponents and extremal structures," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2701–2714, 2010.

[45] Z. Meng, A. Wiesel, and A. O. Hero, "Distributed principal component analysis on networks via directed graphical models," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 2877–2880.

[46] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.