QUICKEST CHANGE DETECTION IN STRUCTURED DATA WITH INCOMPLETE INFORMATION

Javad Heydari and Ali Tajer

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

ABSTRACT

This paper considers a network of agents generating correlated data according to a known kernel. The correlation structure might undergo a change at an unknown time instant, where the post-change kernel is not fully known. Moreover, due to the data processing and communication costs, only a subset of agents can be observed at any time instant. The objective is to detect the change-point with minimum average delay, while the rate of false alarms is controlled. This paper proposes a coupled data acquisition and decision-making process for change detection and establishes its optimality properties.

***Index Terms*—** Change detection, Chernoff, CUSUM

1. INTRODUCTION

Real-time monitoring of a system or process for detecting a change of behavior arises in many application domains such as detecting faults or security breaches in networks, and performing quality control in production lines. It is often of interest to detect abrupt changes with minimal delay after they occur. At the same time designing detection rules that are too sensitive to changes in observations are susceptible to raising frequent false alarms. This creates an inherent tension between the quickness and the quality of the decisions.

In this paper we focus on a network of agents generating correlated data streams and aim to detect abrupt changes in the correlation structure of the data. Such change detection problems are studied in the literature in two different directions. In one direction, the measurements are collected sequentially such that at each time instant one *complete* set of measurements are made from all the nodes in the network [1, 2, 3]. While being effective, such approaches lack efficiency when facing large networks and high dimensional data, in which data acquisition incurs substantial communication, sensing, and decision delay costs. To circumvent this issue, in the second direction the measurements are first quantized (e.g., with one bit) and then communicated to the decision-making units [4, 5, 6, 7, 8]. These approaches, while incurring substantially lower sensing and communication costs, are shown to achieve asymptotic optimality properties.

In this paper, we take a different approach and impose a constraint on the total number of measurements that can be made at any time instant. Hence, at each time only a subset of nodes with cardinality below a set threshold can be sampled. Therefore, quickest change detection in this setting involves coupled data-acquisition and decision-making processes, in which at each time instant one needs to decide whether to stop collecting data and declare a change, or to continue collecting more data, and in the latter case also to identify the subset of nodes to be observed. We assume that the pre-change distribution is fully known, while the post-change distribution follows a composite model. The composite post-change distribution with *fixed* node selection rules and on-off selection rules have been studied in [2] and [9], respectively. These studies are fundamentally different from the problem investigated in this paper since they examine one sequence of data.

Designing coupled data-acquisition and decision-making processes is closely related to controlled sensing, in which the Chernoff rule [10] is widely used. Chernoff rule was originally posed for designing binary hypothesis tests that involve controlling an action for gathering information. Under the assumption of uniformly distinguishable hypotheses, it is shown that selecting the action with the best immediate return, according to proper information measures, achieves optimal performance in the asymptote of diminishing rate of erroneous decisions. Generalizations and extensions of the Chernoff rule for various settings are studied in [11, 12, 13, 14, 15]. Specifically, in [15] the Chernoff rule is modified in a way that the uniformly distinguishable assumption is relaxed for multi-hypothesis setting by introducing randomized actions into the selection rule where it is shown that selecting actions randomly at certain time instants accepts the same asymptotic performance. The results are also extended to the setting in which available data belong to a discrete alphabet and follow a stationary Markov model [16]. In this paper, a modified Chernoff rule is proposed, which combined with parallel cumulative sum (CUSUM) test, it minimizes the average delay in detecting a change in the correlation structure, while ensuring that the rate of false alarms is controlled below a pre-specified level.

2. PROBLEM FORMULATION

2.1. Data Model

Consider a network of n nodes indexed by $\mathcal{V} \triangleq \{1, \dots, n\}$, in which each node generates a discrete-time sequence of random variables. We denote the random variable generated by node i at time t by X_t^i. Accordingly, we define $\mathcal{X}_t \triangleq \{X_t^i \ : \ i \in \mathcal{V}\}$ as the set of random variables generated by the network at time t. Prior to an unknown time instant γ, referred to as the change-point, the set of random variables are assumed to be correlated according to a known structure (kernel). Specifically, we assume that random variables $\mathcal{X}_t$ are generated according to a known joint distribution with cumulative distribution function (CDF) F_0. After the change point γ, the correlation structure governing $\mathcal{X}_t$ changes to one of the $M \in \mathbb{N}$ possible correlation structures. In order to account for such a change, we define F_θ as the joint CDF of $\mathcal{X}_t$ after the change-point, where $\theta \in \{\theta_1, \dots, \theta_M\}$. Hence, we have a composite model for the post-change distribution. Therefore,

$$\begin{array}{ll} \mathcal{X}_t \sim F_0, & t = 1, \dots, \gamma - 1 \\ \mathcal{X}_i \sim F_\theta, & t = \gamma, \gamma + 1, \dots \end{array} . \quad (1)$$

This research was supported in part by the U. S. National Science Foundation under Grants ECCS-1455228 and the CAREER Award ECCS-1554482.

We also assume that there exist well-defined probability density functions (pdfs) corresponding to F_0 and F_θ, which we denote by f_0 and f_θ, respectively. Also, for any $A \subseteq \mathcal{V}$ we denote the joint pdf of $X_A \triangleq \{X_i \,:\, i \in A\}$ corresponding to the pre-change and post-change events by $f_0(X_A; A)$ and $f_\theta(X_A; A)$, respectively. We denote the probability measure governing sequence $\{\mathcal{X}_t \,:\, t \in \mathbb{N}\}$ and the expectation with respect to this measure by $\mathbb{P}_\gamma^\theta$ and $\mathbb{E}_\gamma^\theta$, respectively. We also use $\mathbb{P}_\infty^\theta$ and $\mathbb{E}_\infty^\theta$ for the case that no change occurs and the distribution is always F_0.

2.2. Sampling Model

The ultimate objective of the quickest change-point detection (QCD) in the correlation structure is to sequentially collect measurements from the network and detect the change-point with minimal delay after it occurs, while, in parallel, controlling the rate of false alarms. The setting in which at each time we collect measurements from *all* nodes in the network, i.e., at time t observing $\mathcal{X}_t$ entirely, is studied extensively in the literature (c.f. [1, 2, 3]). In contrast, in this paper we assume that only a limited number of measurements can be taken at any time instant. This could be due to a variety of practical reasons, such as controlling the cost of sensing, the cost of communication, and computational complexity. We define $m < n$ as the maximum number of measurements we afford to make at any time t. Under this setting, the sampling process sequentially collects m measurements at-a-time, until the change-point can be detected with sufficient confidence. Hence, by denoting the index of the nodes observed at time $t \in \mathbb{N}$ by

$$\psi_t \triangleq \{\psi_{t,1}, \dots, \psi_{t,m}\} , \tag{2}$$

and the corresponding samples by

$$Y_t \triangleq \{Y_{t,1}, \dots, Y_{t,m}\} , \tag{3}$$

we can abstract the information accumulated sequentially by the filtration $\mathcal{F}_t$, where

$$\mathcal{F}_t \triangleq \sigma\big((Y_i, \psi_i) \,:\, i \in \{1, \dots, t\}\big) . \tag{4}$$

The sampling process continues until the stopping time, denoted by τ, after which no further measurements are made and a change is detected. Both the stopping time τ and the selection rule ψ_t are $\mathcal{F}_t$-measurable functions. A sampling strategy is completely characterized by $\Phi \triangleq (\tau, \psi_1, \dots, \psi_\tau)$.

2.3. Problem Formulation

We are interested in determining an optimal sampling policy Φ, which involves dynamically making two intertwined decisions at each time. Specifically, at each time t, and based on the information accumulated up to time t, we need to decide whether to terminate the sampling process and declare a change (i.e., $\tau = t$), or to continue and collect more measurements. In the latter case, we need to also determine ψ_{t+1}, which specifies the set of nodes to be measured at the subsequent time instant. Two relevant performance measures for evaluating the quality of the sampling process are the *average* delay between the change-point and the stopping time, and the frequency of false alarms. To account for the average delay we investigate two cases in which we adopt Pollak's and Lorden's definitions of the average delay. Specifically, Pollak's conditional average decision delay is defined as [17]

$$\mathsf{CADD}^\theta(\Phi) \triangleq \sup_{\gamma \geq 1} \mathbb{E}_\gamma^\theta\{\tau - \gamma \,|\, \tau \geq \gamma\} , \tag{5}$$

and Lorden's worst case average decision delay is defined as [18]

$$\mathsf{WADD}^\theta(\Phi) \triangleq \sup_{\gamma \geq 1} \operatorname*{esssup}_{\mathcal{F}_{\gamma-1}} \mathbb{E}_\gamma^\theta\{(\tau - \gamma)^+ \,|\, \mathcal{F}_{\gamma-1}\} . \tag{6}$$

It can be readily verified that

$$\mathsf{CADD}^\theta(\Phi) \leq \mathsf{WADD}^\theta(\Phi) . \tag{7}$$

Pollak's criterion is more natural for modeling an unknown change-point with no random mechanism specifying it (in contrast to Bayesian settings), while Lorden's criterion is more apt when the change-point mechanism depends on the history of the observations. In order to account for the frequency of the false alarms, we define [19]

$$\mathsf{FAR}(\Phi) \triangleq \frac{1}{\mathbb{E}_\infty\{\tau\}} , \tag{8}$$

which captures the average number of false alarms in a sufficiently long observation interval.

There exists an inherent tension between the average delay and the rate of false alarms as improving one penalizes the other one. The optimal sampling strategy can be obtained by controlling the false alarm rate and minimizing the average decision delay. Hence, under Pollak's setting, it is the solution to:

$$\begin{array}{ll} \inf\limits_{\Phi} & \mathsf{CADD}^\theta(\Phi) \\ \text{s.t.} & \mathsf{FAR}(\Phi) \leq \alpha \end{array} , \tag{9}$$

and under Lorden's setting it is the solution to:

$$\begin{array}{ll} \inf\limits_{\Phi} & \mathsf{WADD}^\theta(\Phi) \\ \text{s.t.} & \mathsf{FAR}(\Phi) \leq \alpha \end{array} , \tag{10}$$

where $\alpha \in (0, 1)$ controls the false alarm rate. We aim to characterize sampling procedures that, uniformly for $\theta \in \{\theta_1, \dots, \theta_M\}$, are asymptotically optimal. It is noteworthy that even under the setting without a dynamic selection rule the non-asymptotic optimal solutions to problems (9) and (10) are unknown.

3. QUICKEST CHANGE-POINT DETECTION

In order to treat the QCD problems in (9) and (10), we first briefly review the relevant procedures that treat one change-point detection in networks with one node (i.e., one sequence). Such settings do not involve the dynamic node selection process that we face in the setting of this paper, and are concerned with only determining the optimal stopping.

3.1. QCD in Networks with One Node

QCD in one sequence of random variables is studied extensively in the literature. When the post-change distribution is simple ($M = 1$), the CUSUM test is shown to be optimal for problem (10), and asymptotically optimal for (9) as α approaches zero [20]. Specifically, as α tends to zero under CUSUM we have

$$\mathsf{CADD}^{\theta_1}(\Phi) \leq \mathsf{WADD}^{\theta_1}(\Phi) = \frac{|\log \alpha|}{D_{\rm KL}(f_{\theta_1} \| f_0)}(1 + o(1)) . \tag{11}$$

where $D_{\rm KL}(f_{\theta_1} \| f_0)$ denotes the Kullback-Leibler (KL) divergence between pdfs f_{θ_1} and f_0. When the post-change distribution is composite, a multiple CUSUM algorithm is asymptotically optimal uniformly over θ for both problems [2]. Multiple CUSUM algorithm consists in M parallel CUSUM tests, and its stopping time is the minimum of those of the M constituent CUSUM tests. Under such a multiple CUSUM test, $\mathsf{CADD}^\theta(\Phi)$ and $\mathsf{WADD}^\theta(\Phi)$ satisfy (11) for every possible post-change model $\theta \in \{\theta_1, \dots, \theta_M\}$.

3.2. QCD in Networks with Complete Data

QCD in a network with $n = m$ is studied extensively in the literature. When all measurements are collected perfectly (infinite-rate precision) it is shown that the CUSUM test achieves the same optimality performance as in the single sequence setting [1, 2, 3], i.e.,

$$\mathsf{CADD}^{\theta_1}(\Phi) \leq \mathsf{WADD}^{\theta_1}(\Phi) = \frac{|\log \alpha|(1+o(1))}{D_{\mathrm{KL}}(f_{\theta_1}(\cdot;\mathcal{V})\|f_0(\cdot;\mathcal{V}))} \ .$$

Also, when the measurements have finite-rate precision, e.g., are quantized, it is shown that some variations of the CUSUM test applied locally achieves certain asymptotic optimality properties for the expected delay [4, 5, 6, 7, 8].

3.3. QCD in Networks with Incomplete Data

The fundamental distinction between the setting of this paper and those of sections 3.1 and 3.2 is the additional dynamic decision to be made about the set of m nodes to observe at any time, where $m \in \{1, \dots, n-1\}$. To this end, corresponding to each post-change model θ_i, for $i \in \{1, \dots, M\}$ we define set $\mathcal{S}_i \subseteq \mathcal{V}$ such that $|\mathcal{S}_i| = m$ and it maximizes the KL divergence between f_{θ_i} and f_0, i.e.,

$$\mathcal{S}_i \triangleq \underset{\{\mathcal{S}:|\mathcal{S}|=m\}}{\arg\max} \ D_{\mathrm{KL}}(f_{\theta_i}(\cdot;\mathcal{S})\|f_0(\cdot;\mathcal{S})) \ . \tag{12}$$

Also, we denote the corresponding KL divergence by

$$D^*_{\mathrm{KL}}(f_{\theta_i}\|f_0) \triangleq \max_{\{\mathcal{S}:|\mathcal{S}|=m\}} D_{\mathrm{KL}}(f_{\theta_i}(\cdot;\mathcal{S})\|f_0(\cdot;\mathcal{S})) \ . \tag{13}$$

Based on these definitions we provide the following lower bounds on the average delay metrics.

Theorem 1 *Corresponding to any dynamic selection rule Φ, in the asymptote of α approaching 0, $\forall\theta \in \{\theta_1, \dots, \theta_M\}$ we have*

$$\mathsf{WADD}^{\theta}(\Phi) \geq \mathsf{CADD}^{\theta}(\Phi) \geq \frac{|\log \alpha|}{D^*_{\mathrm{KL}}(f_\theta\|f_0)}(1+o(1)) \ . \tag{14}$$

Next, we propose an algorithm and prove that it achieves the delay lower bound provided in Theorem 1.

3.4. Simple Post-change Model

When the post-change distribution is simple, i.e., $M = 1$, an optimal node selection involves simply making measurements from nodes $\mathcal{S}_1$, defined in (12) at all times $t \in \{1 \dots, \tau\}$. This selection rule in conjunction with the CUSUM test on the observed nodes constitutes an asymptotically optimal sampling process, as formalized in the following theorem.

Theorem 2 *When $M = 1$, the CUSUM test combined with a selection rule that selects $\mathcal{S}_1$ at any time instant until the stopping time is asymptotically optimal as α approaches zero, i.e.,*

$$\mathsf{CADD}^{\theta_1}(\Phi) \leq \mathsf{WADD}^{\theta_1}(\Phi) \leq \frac{|\log \alpha|}{D^*_{\mathrm{KL}}(f_{\theta_1}\|f_0)}(1+o(1)) \ . \tag{15}$$

3.5. Composite Post-change Model

The main goal of this paper is to treat a composite post-change model, in which case the set $\mathcal{S}_i$ varies for different values of $i \in \{1, \dots, M\}$, and the sampling strategy for simple models (Section 3.4) cannot be adopted. In fact selecting the set $\mathcal{S}_i$ is coupled with the decision about the true post-change model. Making such coupled decisions is relevant to the notion of *controlled sensing* in which the Chernoff rule and its variations are used under different assumptions, and exhibit different optimality properties. In the context of this paper, Chernoff rule at each time t identifies the maximum likelihood (ML) decision about the true value of i, denoted by $\hat{i}(t)$, and selects the subset of nodes that have the largest KL divergence under $\theta_{\hat{i}(t)}$, i.e., $\psi_t = \mathcal{S}_{\hat{i}(t)}$. To ensure that the ML decision converges to the true i in finite time and, consequently, the final decision is asymptotically optimal, the Chernoff rule requires the distributions to be distinguishable for all possible actions [10], i.e., for any $i \in \{1, \dots, M\}$ and $\mathcal{S} \subseteq \mathcal{V}$ such that $|\mathcal{S}| = m$,

$$D_{\mathrm{KL}}(f_{\theta_i}(\cdot;\mathcal{S})\|f_0(\cdot;\mathcal{S})) > 0 \ . \tag{16}$$

In [15], this assumption is relaxed for a multi-hypothesis testing problem by adopting a randomized selection rule at exponentially-spaced time instants, and the asymptotic performance guarantees are provided. Specifically, it is shown that by adopting a randomized selection at time instants $t_\ell = \lceil a^\ell \rceil$ for $\ell \in \mathbb{Z}_+$ and a constant $a > 1$ that is sufficiently close to 1, and using the Chernoff rule at other time instants, the asymptotically optimal average delay is achieved.

In our setting, however, we cannot directly apply the modified Chernoff rule with randomization since the change-point is unknown and the delay between the randomized time instants grows exponentially. This means that if the change-point time is large, i.e., $\gamma \gg 1$, randomization fails to ensure the convergence of ML decision about the true i in finite time. To circumvent this issue, we exploit the renewal structure of the CUSUM test to ensure that the starting point of the randomization process is sufficiently close to the change-point γ. For this purpose, we run M CUSUM tests in parallel, one for each θ_i:

$$\mathsf{C}_0(\theta_i) = 0 \ , \tag{17}$$

$$\mathsf{C}_t(\theta_i) = \Big[\mathsf{C}_{t-1}(\theta_i) + \log\frac{f_{\theta_i}(Y_t;\psi_t)}{f_0(Y_t;\psi_t)}\Big]^+ \ , \tag{18}$$

and define an indicator vector

$$\boldsymbol{\zeta}(t) \triangleq [\zeta_1(t), \dots, \zeta_M(t)] \ , \tag{19}$$

where $\zeta_i(t)$ corresponds to the CUSUM test associated with θ_i at time t, and initialize it according to $\boldsymbol{\zeta}(0) = \mathbf{0}_{1\times M}$. We start the sampling process according to the Chernoff rule with a randomization. When the CUSUM value associated with model θ_i is zero (i.e., $\mathsf{C}_t(\theta_i) = 0$) we set its indicator to $\zeta_i(t) = 1$. This process continues until time T which is the first time that indicator values corresponding to all models are set to 1, i.e., $\boldsymbol{\zeta}(T) = \mathbf{1}_{1\times M}$, then we reset $\boldsymbol{\zeta}(T) = \mathbf{0}$, revise the randomization starting point by setting $\ell = 0$ and $t_\ell = T + \lceil a^\ell \rceil$, and resume the sampling process until a change-point is detected. This procedure is summarized in Algorithm 1, and its asymptotic optimality properties are formalized in the following theorem.

Theorem 3 *The dynamic CUSUM test of Algorithm 1, which follows the ML decision for its selection rule, except at time instants $t_\ell = T + \lceil a^\ell \rceil$ for $\ell = \mathbb{Z}_+$ at which it applies a uniform selection rule, is an asymptotically optimal solution to* (9) *and* (10) *when the false alarm rate approaches zero, and $\forall\theta \in \{\theta_1, \dots, \theta_M\}$*

$$\mathsf{WADD}^{\theta}(\tau) = \frac{|\log \alpha|}{D^*_{\mathrm{KL}}(f_\theta\|f_0)}(1+o(1)) \ . \tag{20}$$

Proof: If we show the convergence of ML decision in finite time, the proof follows a similar procedure as in [16, Theorem 4.1]. We define ν as the time instant after which the ML decision of i is always true, and show that randomization guarantees the finiteness of $(\nu - \gamma)$,

Algorithm 1. Proposed sampling strategy

Set $t = 0$, $\ell = 0$, $T = 0$, $\zeta(0) = \mathbf{0}$, and $\mathsf{C}_0(\theta_i) = 0$
While $\max_i C_t(\theta_i) < \log \frac{M}{\alpha}$
  $t \leftarrow t + 1$
  If $t = T + \lceil a^j \rceil$ for some $j \geq \ell$
    $\psi_t \leftarrow m$ nodes randomly
    $\ell \leftarrow \ell + 1$
  Else
    $\hat{i} \leftarrow \arg\max_i \mathsf{C}_t(\theta_i)$
    $\psi_t \leftarrow \mathcal{S}_{\hat{i}}$
  End if
  For $i = 1, \dots, M$
    $\mathsf{C}_t(\theta_i) \leftarrow \left[\mathsf{C}_{t-1}(\theta_i) + \log \frac{f_{\theta_i}(Y_t;\psi_t)}{f_0(Y_t;\psi_t)}\right]^+$
    If $\mathsf{C}_t(\theta_i) = 0$ Then $\zeta_i(t) \leftarrow 1$
  End for
  If $\zeta(t) = \mathbf{1}$
    $T \leftarrow t$ and $\ell \leftarrow 0$
  End if
End while
Stop sampling, declare a change, and $\tau = t$

i.e., when a is selected sufficiently close to 1, for some $K > 0$ and any $\beta > 2$, ν satisfies:

$$\mathbb{P}_\gamma^{\theta_i}(\nu - \gamma > t) \leq K t^{-\beta} , \quad \forall i \in \{1, \dots, M\} . \tag{21}$$

We denote the last time instant at which $\zeta(t)$ is reset to zero by η and show that $\mathbb{P}_\gamma^{\theta_i}(\gamma - \eta = j)$ for $j \in \{1, \dots, \gamma\}$, which is the probability of the event that at least one of the CUSUM values has remained positive, is upper bounded by

$$\begin{aligned}\mathbb{P}_\gamma^{\theta_i}(\gamma - \eta = j) &= \mathbb{P}_\gamma^{\theta_i}\Big(\bigcup_{i=1}^{M} (\mathsf{C}_{\eta+k}(\theta_i) > 0 \text{ for } k \in \{1, \dots, j\})\Big) \\ &\overset{(a)}{\leq} \sum_{i=1}^{M} \mathbb{P}_\gamma^{\theta_i}\Big(\mathsf{C}_{\eta+k}(\theta_i) > 0 \text{ for } k \in \{1, \dots, j\}\Big) \\ &\overset{(b)}{\leq} \sum_{i=1}^{M} \rho_i^j \overset{(c)}{\leq} M \rho_{\max}^j ,\end{aligned} \tag{22}$$

where (a) holds owing to the union bound, (b) results from applying the Markov and Cauchy-Schwartz inequalities on

$$\rho_i \triangleq \mathbb{P}_\infty\Big(\log \frac{f_{\theta_i}(X)}{f_0(X)} > 0\Big) < 1 , \tag{23}$$

and (c) is due to definition $\rho_{\max} \triangleq \max_i \rho_i$. From the analysis in [16, Lemma 6.4], it is clear that when the starting point of the randomization and the change-point are the same, i.e., $\eta = \gamma$, we have

$$\mathbb{P}_\gamma^{\theta_i}(\nu - \gamma > t \mid \eta = \gamma) \leq K_1 t^{-\beta_1} , \tag{24}$$

for some $\beta_1 > 2$. Following the same line of argument yields

$$\begin{aligned}\mathbb{P}_\gamma^{\theta_i}(\nu - \gamma > t) &= \sum_{j=0}^{\infty} \mathbb{P}_\gamma^{\theta_i}(\nu - \gamma > t \mid \gamma - \eta = j)\mathbb{P}(\gamma - \eta = j) \\ &\leq M K_1 \sum_{j=0}^{\infty} (t - j)^{-\beta_1} \rho_{\max}^j \leq K t^{-\beta} ,\end{aligned} \tag{25}$$

for some positive constant K and β, which concludes the proof. ■

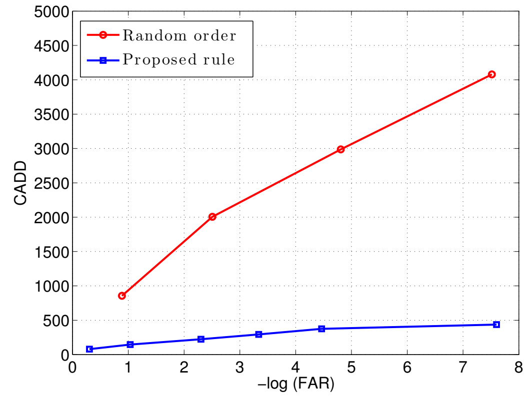

Fig. 1. Comparison of random and the proposed selection rules.

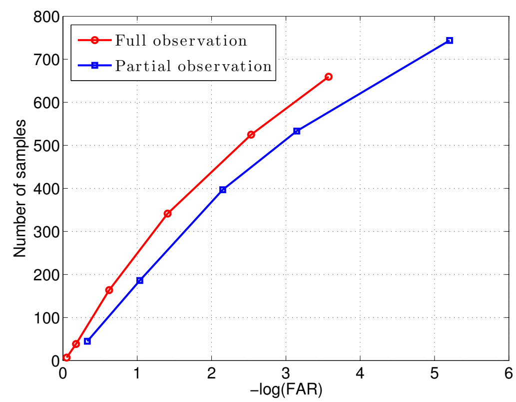

Fig. 2. Comparison of full and partial observation.

4. SIMULATION RESULTS

We consider a network consisting of $n = 10$ nodes generating zero-mean Gaussian random variables. Prior to the change-point γ the generated data set at each time are independent with covariance matrix $\boldsymbol{I}$, and after γ they have a covariance matrix $\boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3\}$. In Fig. 1 we set $m = 2$ and compare the conditional average decision delay of the proposed selection rule and a random selection rule. It is observed that the proposed selection rule shows a significant gain compared to the random selection of nodes. Figure 2 compares the number of measurements required for the proposed approach and the setting that observes the entire network. It is observed that the proposed approach requires fewer measurements to achieve the same false alarm rate since it identifies the most relevant nodes and collects their measurements.

5. CONCLUSION

We have analyzed the problem of quickest change-point detection over a network of data streams with sampling constraint. We have considered a setting in which the number of measurements that can be collected at each time is controlled, and the subset of the nodes for observation can be dynamically selected based on a data-adaptive strategy. Also we have assumed that the pre-change data stream has a known distribution, while the post-change distribution follows a composite model. A sequential sampling strategy has been proposed for identifying a change in the distribution for the setting in which the post-change distribution takes one of the finite number of possible forms. We have shown that the CUSUM test combined with a modified Chernoff rule is asymptotically optimal for minimizing the expected delay as the false alarm rate approaches zero.

6. REFERENCES

[1] Y. C. Chen, T. Banerjee, A. D. Dominguez-Garcia, and V. V. Veeravalli, “Quickest line outage detection and identification,” *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 749–758, Jan. 2016.

[2] A. G. Tartakovsky and A. S. Polunchenko, “Quickest change-point detection in distributed multisensor systems under unknown parameters,” in *Proc. 11th International Conference on Information Fusion*, Jun. 2008, pp. 1–8.

[3] A. G. Tartakovsky, “Multidecision quickest change-point detection: Previous achievements and open problems,” *Sequential Analysis*, vol. 27, no. 2, pp. 201–231, May 2008.

[4] V. V. Veeravalli, “Decentralized quickest change detection,” *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[5] X. Ren, K. H. Johansson, D. Shi, and L. Shi, “Quickest change detection in adaptive censoring sensor networks,” *arXiv:1503.04999v2*, Mar. 2015.

[6] O. Hadjiliadis, H. Zhang, and H. V. Poor, “One shot schemes for decentralized quickest change detection,” *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3346–3359, Jul. 2009.

[7] G. Fellouris and G. V. Moustakides, “Bandwidth and energy efficient decentralized sequential change detection,” *arXiv:1210.2029v2*, Oct. 2012.

[8] S. Banerjee and G. Fellouris, “Decentralized sequential change detection with ordered CUSUMs,” in *Proc. IEEE International Symposium on Information Theory*, Jul. 2016, pp. 36–40.

[9] T. Banerjee and V. V. Veeravalli, “Data-efficient minimax quickest change detection with composite post-change distribution,” *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 5172–5184, Sep. 2015.

[10] H. Chernoff, “Sequential design of experiments,” *The Annals of Mathematical Statistics*, vol. 30, no. 3, pp. 755–770, Sep. 1959.

[11] S. Bessler, “Theory and applications of the sequential design of experiments, k-actions and infinitely many experiments: Part I Theory,” Tech. Rep., Department of Statistics, Stanford University, Mar. 1960.

[12] A. E. Albert, “The sequential design of experiments for infinitely many states of nature,” *The Annals of Mathematical Statistics*, vol. 32, no. 3, pp. 774–799, Sep. 1961.

[13] G. E. P. Box and W. J. Hill, “Discrimination among mechanistic models,” *Technometrics*, vol. 9, no. 1, pp. 57–71, Feb. 1967.

[14] W. J. Blot and D. A. Meeter, “Sequential experimental design procedures,” *Journal of the American Statistical Association*, vol. 68, no. 343, pp. 586–593, Sep. 1973.

[15] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, “Controlled sensing for multihypothesis testing,” *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2451–2464, May 2013.

[16] S. Nitinawarat and V. V. Veeravalli, “Controlled sensing for sequential multihypothesis testing with controlled Markovian observations and non-uniform control cost,” *Sequential Analysis*, vol. 34, no. 1, pp. 1–24, Feb. 2015.

[17] M. Pollak, “Optimal detection of a change in distribution,” *The Annals of Statistics*, vol. 13, no. 1, pp. 206–227, Mar. 1985.

[18] G. Lorden, “Procedures for reacting to a change in distribution,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.

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

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