

Algorithms for Change Detection with Unknown Number of Affected Sensors

P Sarath Kumar*, B Sai Kiran†, Arun Pachai Kannu‡, Srikrishna Bhashyam§

Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai - 600036, India

Email: *sarathkumar.hi@gmail.com, †bs.kiran.bulsai@gmail.com, ‡arunpachai@ee.iitm.ac.in, §skrishna@ee.iitm.ac.in

Abstract— In this paper, we consider change detection in a sensor network where an unknown subset of sensor nodes are affected by the change. We consider two models for the channel between the sensors and the fusion center: (1) parallel non-interfering channels, and (2) physical layer fusion. For both these models, we propose quantized transmission schemes for the sensors and corresponding fusion rules at the fusion center. The proposed fusion rules are based on an adaptive version of CUSUM. The detection delay performance of the proposed schemes is studied as a function of the number of affected sensors for a given false alarm constraint. Simulation results show that the proposed schemes can work well for a wide range of the fraction of affected sensors.

Keywords—change detection, CUSUM, Adaptive CUSUM, sensor networks

I. INTRODUCTION

Detection of abrupt changes based on statistical tests has been studied for several decades [1], [2]. Recently, applications in sensor networks have led to further advancements (See [3] and references therein). In this paper, we consider change detection in a sensor network where an unknown subset of sensor nodes are affected by the change.

Consider a network of K sensors where each sensor makes observations $\{X_{k,n}\}$ with discrete time index $n = 1, \dots, \infty$ and sensor index $k = 1, \dots, K$. Observations are assumed to be statistically independent across time and across sensors. Initially, the observations $X_{k,n}$ follow the distribution f_k at the k^{th} sensor. A change event happens at an unknown time $\nu \in \mathbb{N}$. Only a subset of sensors $\mathcal{S} \subset \{1, \dots, K\}$ are affected by the change. After the change event, the observations at the affected sensors follow a different distribution g_k . To summarize,

$$X_{k,n} \sim \begin{cases} f_k & : n < \nu \\ f_k & : n \geq \nu \text{ and } k \notin \mathcal{S} \\ g_k & : n \geq \nu \text{ and } k \in \mathcal{S} \end{cases} \quad (1)$$

The set of affected sensors \mathcal{S} and its size $|\mathcal{S}|$ are unknown a priori. Based on their observations, sensors send signals $U_{k,n}$ to the fusion center. Let \mathbf{Y}_n denote the signal received at the fusion center at time n . We consider two models: (1) where the fusion center receives noiselessly through parallel channels

This work was done at the Department of Electrical Engineering, IIT Madras. P Sarath Kumar is now with Cisco Systems, Bangalore and B Sai Kiran is with Samsung, Bangalore.

such that observation vector is given by

$$\mathbf{Y}_n = [U_{1,n}, \dots, U_{K,n}], \quad (2)$$

and (2) the noisy physical layer fusion model [4] where the scalar observation is given by

$$Y_n = \sum_{k=1}^K U_{k,n} + V_n, \quad (3)$$

where V_n is white gaussian noise with variance σ^2 . The models are depicted in Figure 1.

Based on its observations, say at time T , the fusion center declares that change has happened. A typical problem formulation is to minimize the worst-case detection delay subject to a constraint on the false alarm rate. Using \mathcal{F}_n to denote the set of all the observations up to time n , the worst case detection delay τ_w is obtained as

$$\tau_w = \sup_{\nu \geq 1} \text{ess sup } E[(T - \nu + 1)^+ / \mathcal{F}_{\nu-1}] \quad (4)$$

where the supremum is taken over all possible change points and all possible pre-change observations and expectation is taken over the post change observations. The average false alarm rate also referred as the average run length is given by

$$\phi_l = E[T / \nu = \infty]. \quad (5)$$

ϕ_l gives the average time taken to *falsely* declare the change when there was no change event. In [5], under the noiseless parallel channel model, the problem of minimizing τ_w under the constraint that $\phi_l \geq \gamma$ for a prespecified γ is considered. Several asymptotically optimal schemes when $\gamma \rightarrow \infty$ were proposed in [5]. However, the schemes assume availability of infinite bandwidth parallel channels between sensors and the fusion center. The goal of our work is to study the problem under practical considerations of finite bandwidth and noisy channels. Specifically, we study the following scenarios.

- 1) Using finite number of levels for summary messages $U_{k,n}$ from sensors to the fusion center.
- 2) Using variants of CUSUM algorithms at the fusion center for the noisy physical layer fusion model.
- 3) Using prior information on the change event for ON/OFF observation control and message quantization for the sensors.

We develop new processing algorithms for the above scenarios and make relevant comparison with several algorithms in the

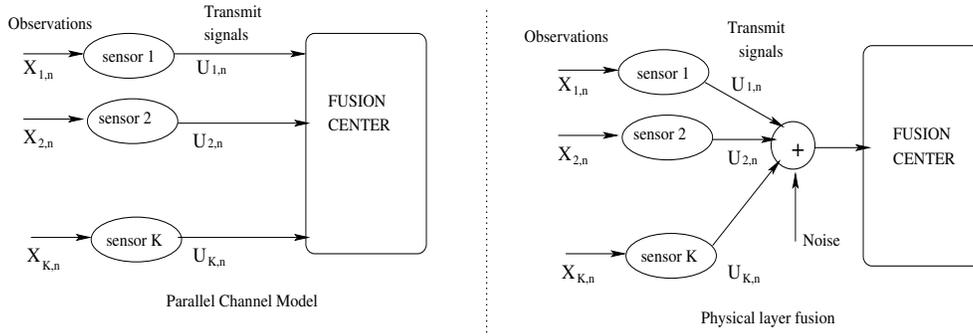


Fig. 1. Observation Model at sensors and fusion center.

existing literature.

II. PARALLEL CHANNEL MODEL

In this section, we consider the model in (2) where the fusion center gets the noiseless observations from the sensors. For the single sensor case, the problem of minimizing the worst case detection delay under the average run length constraint is well studied [2]. It is well known that CUSUM algorithm is asymptotically optimal as average run-length approaches ∞ . At each sensor, the log-likelihood ratio (LLR) at time instant n is computed as,

$$L_{k,n} = \log \frac{g_k(X_{k,n})}{f_k(X_{k,n})}, \quad (6)$$

and the CUSUM metric at each sensor is calculated as

$$W_{k,n} = \max(W_{k,n-1} + L_{k,n}, 0) \quad (7)$$

with the initialization $W_{k,0} = 0 \forall k$.

A. Hard thresholding scheme

For reference, we present the hard thresholding scheme presented in [5]. The summary message sent by the sensors are given by

$$U_{k,n} = \begin{cases} W_{k,n} & \text{if } W_{k,n} \geq b \\ 0 & \text{else} \end{cases} \quad (8)$$

where b is a suitably chosen threshold. The fusion center gets the noiseless observations \mathbf{Y}_n as given in (2). The decision at the fusion center is given by

$$\delta_h(\mathbf{Y}_n) = \begin{cases} \text{declare change at time } n & \text{if } \sum_{k=1}^K U_{k,n} \geq a \\ \text{Continue} & \text{else} \end{cases} \quad (9)$$

The constant a is chosen to satisfy the run-length constraint $\phi_l \geq \gamma$. This hard thresholding scheme is shown to be asymptotically optimal in minimizing τ_w in (4) as $\phi_l \rightarrow \infty$. One practical problem with this scheme is that the sensors need to transmit the CUSUM metric $W_{k,n}$ (once it goes above threshold) and this may consume significant bandwidth.

B. Binary Quantization at Sensors

In this section, we propose a binary message quantization at the sensors and subsequent detection at the fusion center. The

summary message $U_{k,n}$ sent from the sensors to the fusion center is the quantized version of the LLR $L_{k,n}$ as given by

$$U_{k,n} = \begin{cases} 1 & \text{if } L_{k,n} \geq t_k \\ 0 & \text{else} \end{cases} \quad (10)$$

In order to optimize the performance we have to choose t_k such that $E[L_{k,n}^{BQ}]$ is maximized [6] where

$$L_{k,n}^{BQ} = \log \frac{P_k^{(g)}(U_{k,n})}{P_k^{(f)}(U_{k,n})} = a_k^{(g)} U_{k,n} + a_k^{(f)} \quad (11)$$

with $P_k^{(g)}(\cdot)$ and $P_k^{(f)}(\cdot)$ denoting the probability induced on $U_{k,n}$ given that the observation $X_{k,n}$ is distributed as f_k and g_k respectively. The scalars a_k 's are given as

$$a_k^{(g)} = \log \frac{\beta_k^{(g)}(1 - \beta_k^{(f)})}{\beta_k^{(f)}(1 - \beta_k^{(g)})}, \quad a_k^{(f)} = \log \frac{(1 - \beta_k^{(f)})}{(1 - \beta_k^{(g)})},$$

where $\beta_k^{(f)} = P_k^{(f)}\{U_{k,n} = 1\}$ and $\beta_k^{(g)} = P_k^{(g)}\{U_{k,n} = 1\}$. Now, $E[L_{k,n}^{BQ}] = \beta_k^{(g)} a_k^{(g)} + a_k^{(f)} = \mathcal{I}_k^{BQ}(t_k)$ which is the Kullback-Leibler (KL) distance between $P_k^{(g)}$ and $P_k^{(f)}$ and

$$t_k^{opt} = \arg \max_{t_k > 0} \mathcal{I}_k^{BQ}(t_k). \quad (12)$$

Now each sensor transmits $U_{k,n}$ to the fusion center through the dedicated channel. At the fusion center, we calculate

$$R_{k,n} = \max\{0, R_{k,n-1} + L_{k,n}^{BQ}\}$$

and compute the SUM (S_n) and MAX (M_n)

$$S_n = \sum_{k=1}^K R_{k,n}, \quad M_n = \max_{1 \leq k \leq K} R_{k,n}.$$

The decision rule at the fusion center is given by

$$\delta_b = \begin{cases} \text{declare change} & \text{if } S_n \geq a_1 \text{ or } M_n \geq a_2 \\ \text{Continue} & \text{else} \end{cases} \quad (13)$$

and the constants a_1 and a_2 are chosen to satisfy the constraint $\phi_l \geq \gamma$. Through simulations, we compare the performance of binary quantization and hard thresholding schemes.

III. NOISY PHYSICAL LAYER FUSION MODEL

Here we consider the physical layer fusion model (3) and develop suitable signal processing algorithms at the sensor and at the fusion center.

A. Adaptive Quantization at the Sensors

In this section we propose an adaptive quantization approach for the physical layer fusion model. The summary message $U_{k,n}$ is obtained by multi-level quantization of CUSUM metric $W_{k,n}$ as

$$U_{k,n} = \begin{cases} 0 & \text{if } W_{k,n} < \Delta\rho_k b \\ s & \text{if } \Delta\rho_k b \leq W_{k,n} < 2\Delta\rho_k b \\ \dots & \dots \\ (L-1)s & \text{if } W_{k,n} \geq (L-1)\Delta\rho_k b \end{cases} \quad (14)$$

where $s > 0$ and $\Delta \geq 1$ are pre-specified constants and threshold b is suitably chosen. Threshold scaling factor ρ_k is chosen as $\rho_k = \frac{\mathcal{I}(g_k||f_k)}{\sum_{k=1}^K \mathcal{I}(g_k||f_k)}$, where $\mathcal{I}(g_k||f_k) = \int g_k(x) \log \frac{g_k(x)}{f_k(x)} dx$. Intuitively the sensors with lower KL distance between the pre change and the post change distributions will have a lower threshold b_k and vice versa. Ideally, the observation at the fusion center given by (3) follows $\mathcal{N}(0, \sigma^2)$ before change and $\mathcal{N}(m, \sigma^2)$ after the change and the mean $m > 0$ is unknown as the number of affected sensors is unknown. In Section IV, we discuss the algorithms which can be employed at the fusion center for this model.

B. ON-OFF Observation control and Quantization at Sensors

For the single sensor case, using a Bayesian prior model for the change point, an ON-OFF observation control policy is designed in [7] to minimize the average number of observations obtained at the sensor and the detection delay subject to a constraint on the false alarm probability. In this section, we use this ON-OFF observation control scheme at each sensor in the multi-sensor model with unknown number of affected sensors and physical layer fusion.

The change point is considered to follow geometric distribution with parameter ρ as given by

$$\pi_n = P[\nu = n] = \rho(1 - \rho)^{(n-1)}, \quad n \geq 1.$$

When the change happens, the observations follow the model in (1). Let $S_{k,n} \in \{0, 1\}$ denote the binary variable which indicates whether observation is made or not at time instant n by sensor k . Let T_k denote the time instant at which k^{th} sensor decides that change has happened. The average number of observations (ANO) made by sensors prior to change point is defined as $\text{ANO}(k) = E \left[\sum_{n=1}^{\min\{T_k, \nu-1\}} S_{k,n} \right]$, and the average detection delay is given by $E_{DD}(k) = E(T_k - \nu)^+$. We consider the problem of

$$\begin{aligned} & \text{minimize } E_{DD}(k) \\ & \text{subject to } P\{T_k < \nu\} \leq \alpha, \text{ and ANO}(k) \leq \beta \end{aligned}$$

where α and β are given constraints. Using results from the single sensor case [7], we obtain the following optimal policy

at each sensor.

The sufficient statistic for the above problem is

$$p_{k,n} = P\{\nu \leq n | X_{k,1}(S_{k,1}), \dots, X_{k,n}(S_{k,n})\} \quad (15)$$

where $\{X_{k,i}(S_{k,i}), 1 \leq i \leq n\}$ represents the history of measurements made upto time n . $p_{k,n}$ can be updated as

$$p_{k,n+1} = \begin{cases} \Phi^{(0)}(p_{k,n}) & \text{if } S_{k,n} = 0 \\ \Phi^{(1)}(X_{k,n}, p_{k,n}) & \text{if } S_{k,n} = 1 \end{cases} \quad (16)$$

where

$$\Phi^{(0)}(p_{k,n}) = p_{k,n} + (1 - p_{k,n})\rho \quad (17)$$

$$\Phi^{(1)}(X_{k,n}, p_{k,n}) = \frac{\Phi^{(0)}(p_{k,n})L_{k,n}}{\Phi^{(0)}(p_{k,n})L_{k,n} + (1 - \Phi^{(0)}(p_{k,n}))} \quad (18)$$

with $L_{k,n}$ is the LLR given in (6) and $p_0 = \pi_0$.

The optimal ON-OFF control policy is

$$S_{k,n+1} = \begin{cases} 0 & \text{if } p_{k,n} < B \\ 1 & \text{if } p_{k,n} \geq B \end{cases} \quad (19)$$

and the optimal sensor change detection policy is

$$\delta_o = \begin{cases} \text{declare change at time } n & \text{if } p_{k,n} \geq A \\ \text{Continue} & \text{else} \end{cases} \quad (20)$$

for suitably chosen thresholds A and B . Once the sensors decide on the change, the sensor sends signals

$$U_{k,n} = \begin{cases} s & \text{after declaring change} \\ 0 & \text{else} \end{cases} \quad (21)$$

The fusion center receives the signals $U_{k,n}$ from the sensors and makes its decision. Ideally, fusion center can use the prior information as well as the signals from the sensors to make its decision. However, the computation of sufficient statistics at the fusion center is highly involved and hence we resort to simple non-bayesian CUSUM algorithms at the fusion center. We consider the physical layer fusion model (3) at the fusion center and the detection algorithms at the fusion center is discussed in the following section.

IV. DETECTION ALGORITHMS AT THE FUSION CENTER

From the physical layer fusion model (3), it is clear that, when the sensors do not send any signal (before they detect the change), the fusion center simply observes the noise. On the other hand, after the sensors identify the change, their signals coherently add up at the fusion center. When the number of affected sensors is *known*, a dual CUSUM approach has been proposed in [4]. As the number of affected sensors is unknown in our model, we develop alternative approaches which are discussed below.

For convenience of exposition, we assume that $T_k = \Gamma$ for all $k \in \mathcal{S}$, i.e., all the affected sensors identify the change at the same instant. This may sound impractical but the adaptive algorithms proposed in this paper are robust and do not rely on this assumption. Denoting Gaussian distribution of mean

m and variance σ^2 by \mathcal{N}_m , we have the following model

$$Y_n \sim \begin{cases} \mathcal{N}_0 & n < \Gamma \\ \mathcal{N}_{\tilde{m}} & n \geq \Gamma \end{cases}$$

where the mean \tilde{m} is unknown since the number of affected sensors is unknown. The mean \tilde{m} belongs to a finite discrete set $\mathcal{M} = \{0, s, \dots, Ms\}$ for some M which depends on the number of sensors K and the number of quantization levels used by the sensors (14). When the unknown post-change mean takes a values from a continuum, a parallel recursive test (parallel CUSUM) has been proposed and studied in [8]. Its extension (adaptive CUSUM) to a time-varying unknown mean case has been studied in [9]. Here, we adapt these algorithms to our problem in which the unknown mean belongs to a discrete set.

A. Parallel CUSUM

Here we run a CUSUM algorithm for each possible mean in \mathcal{M} as follows. For each $m \in \mathcal{M}$, compute the LLR

$$L_n^m = \log \frac{\mathcal{N}_m(Y_n)}{\mathcal{N}_0(Y_n)} \quad (22)$$

and the CUSUM metric is computed as

$$W_n^m = \max(W_{n-1}^m + L_n^m, 0). \quad (23)$$

Let us look at characteristics of the post-change CUSUM metrics. For $n > \Gamma$, define $F(m) = \mathbb{E} L_n^m$ where the expectation is taken over the distribution of Y_n . Note that, for $n > \Gamma$, the true distribution of Y_n is $\mathcal{N}_{\tilde{m}}$. We have

$$\begin{aligned} F(\tilde{m}) &= \mathbb{E} \log \frac{\mathcal{N}_{\tilde{m}}(Y_n)}{\mathcal{N}_0(Y_n)} = \mathbb{E} \log \frac{\mathcal{N}_{\tilde{m}}(Y_n)}{\mathcal{N}_m(Y_n)} + \mathbb{E} \log \frac{\mathcal{N}_m(Y_n)}{\mathcal{N}_0(Y_n)} \\ &= \mathcal{I}(\mathcal{N}_{\tilde{m}} || \mathcal{N}_m) + F(m) \end{aligned}$$

and since the KL distance $\mathcal{I}(\cdot || \cdot)$ is non-negative, we have $F(\tilde{m}) \geq F(m)$ with equality if and only if $m = \tilde{m}$. Hence the post change CUSUM metric W_n^m has a non-negative drift with the average slope of the drift $F(m)$ being maximum when $m = \tilde{m}$. So, among the parallel CUSUM metrics W_n^m , the one corresponding to $m = \tilde{m}$ is expected to be maximum after the change point. So, the fusion center detection rule for the parallel CUSUM is

$$\delta_p = \begin{cases} \text{declare change} & \text{if } \max_m W_n^m \geq a \\ \text{Continue} & \text{else} \end{cases} \quad (24)$$

where a is a suitably chosen threshold to satisfy the average run length constraint.

B. Adaptive CUSUM

Parallel CUSUM computes the CUSUM metrics W_n^m for all $m \in \mathcal{M}$. In order to reduce the computational load, we employ the adaptive CUSUM approach developed in [9]. We consider an adaptive approach which estimates the unknown mean \tilde{m} . Allowing m to take continuum of values, we note that $F(m)$ is a strictly concave function which achieves global maximum when $m = \tilde{m}$. So, given a small value of ϵ , we can always find two values p, q such that $q = p + \epsilon$ and $F(p) = F(q)$ and

\tilde{m} lies in the interval (p, q) . Once we find such p and q , we have narrowed down the value of \tilde{m} . Given an ϵ , to find the values for p and q , we proceed as follows. Choose arbitrarily the value of p_0 and set $q_0 = p_0 + \epsilon$ and then subsequent values are obtained as

$$p_{n+1} = \max(0, p_n + \xi \mathcal{D}^n), q_{n+1} = \min(Ms, p_{n+1} + \epsilon) \quad (25)$$

where ξ is the step size and \mathcal{D}^n is the difference $F(q_n) - F(p_n)$. Since, we do not have the values $F(\cdot)$ as they involve expectation over unknown distribution, we replace the ensemble average \mathcal{D}^n by the time average given as

$$\hat{\mathcal{D}}^n = \log \frac{\mathcal{N}_{q_n}(Y_n)}{\mathcal{N}_{p_n}(Y_n)} \quad (26)$$

The convergence analysis is straight forward. If $\hat{\mathcal{D}}^n > 0$, p_{n+1} and q_{n+1} will increase according to (25), so that $\hat{\mathcal{D}}^{n+1}$ will decrease due to the concavity of the function. Similarly, if $\hat{\mathcal{D}}^n < 0$, p_{n+1} and q_{n+1} will decrease so that $\hat{\mathcal{D}}^{n+1}$ will increase. In both the cases, $\hat{\mathcal{D}}^n$ will converge to zero. The choice of ϵ is crucial. Higher value of ϵ leads to faster convergence but it also increases the interval range of the estimated parameter.

Based on the current values p_n and q_n , the mean value is estimated as $\hat{m}_n = \frac{p_n + q_n}{2}$ and the CUSUM is computed based on the current estimate of the mean as

$$W_n = \max\left(0, W_{n-1} + \log \frac{\mathcal{N}_{\hat{m}_n}(Y_n)}{\mathcal{N}_0(Y_n)}\right) \quad (27)$$

and the decision rule is

$$\delta_a = \begin{cases} \text{declare change} & \text{if } W_n \geq a \\ \text{Continue} & \text{else} \end{cases} \quad (28)$$

where a is a suitably chosen threshold to satisfy the average run length constraint.

V. SIMULATION RESULTS

Number of sensors $K = 100$. All the sensors are assumed to be homogeneous. The observations at each sensor are i.i.d. with pre-change distribution $\mathcal{N}(0, 1)$ and post-change distribution $\mathcal{N}(0.5, 1)$. We use the detection delay at the fusion center as the performance metric of interest and keep the same average run length constraint for different schemes.

For the noiseless parallel channel model, the Figure 2 compares the delay performance of hard thresholding scheme [5] with the likelihood binary quantization scheme for the same average run length constraint. We see that even with binary messages from the sensors, the delay performance is close to the hard thresholding scheme.

In Fig. 3, we compare the performance of adaptive and parallel CUSUM methods for the noisy physical layer channel model. The simulations are performed with Signal to Noise ratio (SNR) $10 \log \frac{s^2}{\sigma^2}$ being 0dB and binary quantization (14) with threshold $b = 460.52$. From the plot we can observe that the delay performance is nearly the same for the two schemes.

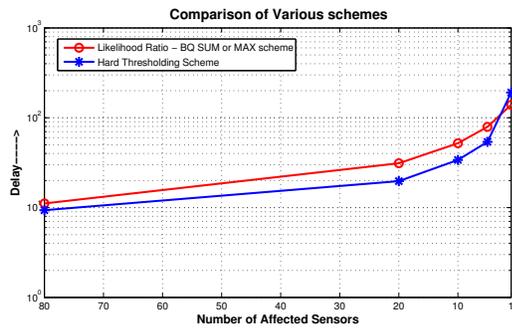


Fig. 2. Comparison of Hard thresholding with Binary Quantization

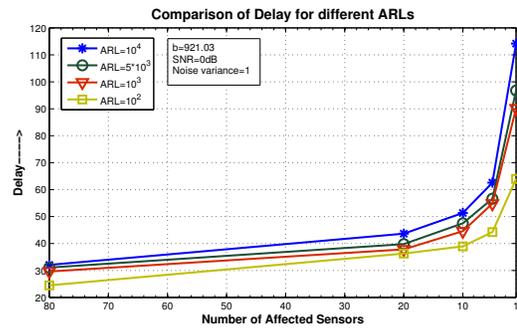


Fig. 5. Delay for various average run lengths

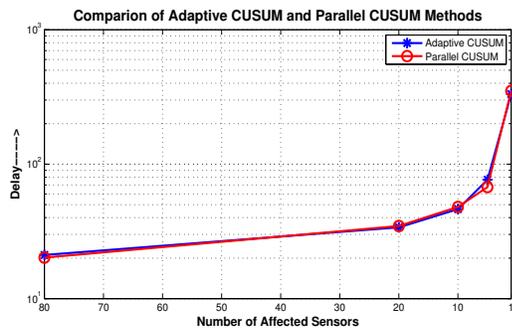


Fig. 3. Comparison of Parallel and Adaptive CUSUM

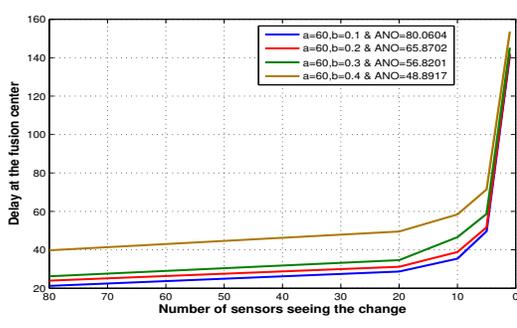


Fig. 6. Delay performance with ON/OFF observation control

Figure 4 illustrates a realization of the convergence behavior of adaptive CUSUM with the change point being $\nu = 100$ and it can be seen that the estimated mean converges to the true mean.

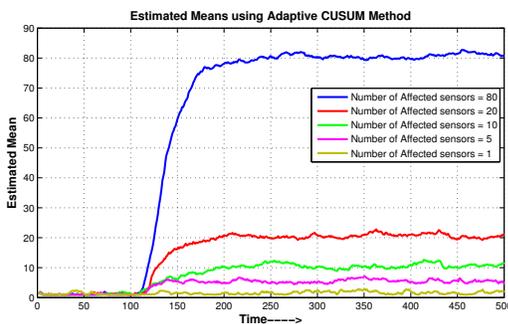


Fig. 4. Convergence of Adaptive CUSUM

Figure 5 compares the delay performance of Adaptive CUSUM with binary quantization (14) for various values of average run lengths.

Figure 6 compares the delay for various ON/OFF observation policy thresholds with adaptive CUSUM at the fusion center. We see that as the thresholds A and B are changed, the average number of observations (ANO) changes and hence the delay at the fusion center is affected.

VI. SUMMARY

We proposed and studied variants of CUSUM when the number of affected sensors is unknown. We also studied the effect of using finite quantization levels for sensor messages. An interesting future direction will be to study the robustness of the proposed methods to the pre-change and post-change distributions, and number of quantization levels.

REFERENCES

- [1] A. N. Shiryaev, *Optimal Stopping Rules*. Springer-Verlag, 1978.
- [2] M. Basseville and I. Nikiforov, "Detection of abrupt changes: theory and application," 1993.
- [3] V. V. Veeravalli and T. Banerjee, "Quickest Change Detection," *ArXiv e-prints*, Oct. 2012.
- [4] T. Banerjee, V. Sharma, V. Kavitha, and A. JayaPrakasam, "Generalized analysis of a distributed energy efficient algorithm for change detection," *Wireless Communications, IEEE Transactions on*, vol. 10, pp. 91–101, january 2011.
- [5] Y. Mei, "Quickest detection in censoring sensor networks," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 2148–2152, 31 2011-aug. 5 2011.
- [6] A. Tartakovsky and A. Polunchenko, "Quickest changepoint detection in distributed multisensor systems under unknown parameters," in *Information Fusion, 2008 11th International Conference*, pp. 1–8, 30 2008-july 3 2008.
- [7] T. Banerjee and V. V. Veeravalli, "Data-Efficient Quickest Change Detection with On-Off Observation Control," *ArXiv e-prints*, May 2011.
- [8] I. Nikiforov, "A suboptimal quadratic change detection scheme," *Information Theory, IEEE Transactions on*, vol. 46, pp. 2095–2107, sep 2000.
- [9] C. Li, H. Dai, and H. Li, "Adaptive quickest change detection with unknown parameter," in *ICASSP 2009*, pp. 3241–3244, april 2009.