Generalized Analysis of a Distributed Energy Efficient Algorithm for Change Detection

Taposh Banerjee, Vinod Sharma, Veeraruna Kavitha, and A. K. JayaPrakasam

***Abstract*—We propose an energy efficient distributed cooperative Change Detection scheme called DualCUSUM based on Page's CUSUM algorithm. In the algorithm, each sensor runs a CUSUM and transmits only when the CUSUM is above some threshold. The transmissions from the sensors are fused at the physical layer. The channel is modeled as a Multiple Access Channel (MAC) corrupted with noise. The fusion center performs another CUSUM to detect the change. The algorithm performs better than several existing schemes when energy is at a premium. We generalize the algorithm to also include nonparametric CUSUM and provide a unified analysis. Our results show that while the false alarm probability is smaller for observation distribution with a lighter tail, the detection delay is asymptotically the same for any distribution. Consequently, we provide a new viewpoint on why parametric CUSUM performs better than nonparametric CUSUM. In the process, we also develop new results on a reflected random walk which can be of independent interest.**

***Index Terms*—Nonparametric CUSUM, decentralized change detection, reflected random walk.**

I. Introduction

THE detection of an abrupt change in the distribution of a sequence of random variables is a classical problem in statistics. In this problem, a decision maker observes a sequence of random variables. At some point of time, unknown to the decision maker, the distribution of these observations changes. The decision maker has to detect this change of law as soon as possible subject to some false alarm constraint. This is also called the *centralized* version of the *change detection* problem and has been well studied. When the observations are independent and identically distributed (*iid*) conditioned on the time of change and the distribution of the change time T is known (this is called the Bayesian setting), the optimal algorithm was obtained by Shiryaev ([29]). When distribution of T is not known, the CUSUM algorithm, first proposed by Page in [23], was shown by Lorden ([17]) and Moustakides ([22]) to minimize the worst case delay (Min-Max optimality).

In the *distributed* version of the change detection problem, multiple geographically distributed sensors take observations and send the processed information to a decision maker (fusion center) for the detection of change. This model finds application in biomedical signal processing, intrusion detection in computer and sensor networks ([35], [33]), finance, quality control engineering, and recently, distributed detection of spectrum holes in cognitive radio networks ([16], [28]). The distribution of the observations of all the sensors changes simultaneously at some random point of time. While this model is slightly restrictive, it is the most widely studied model in the literature. As evident from the work in [3], [37] and in this paper, even this model is not well explored.

In the absence of communication and resource constraints, the sensors can send the raw observations to the fusion center and the problem reduces to the centralized one discussed above. However, in applications like sensor networks, the sensors are low power, battery operated devices and thus there are severe constraints on their communication and processing capabilities. Therefore, it is suggested that sensors send processed version (e.g., quantized) of their observations to the fusion center and the fusion center fuses the information from various sensors to make the decision (see [7], [33], [35], [36] for various processing possibilities).

Several distributed algorithms have been proposed for detection of change. When sensors send quantized version of their observations to the fusion center, the author in [35] has obtained asymptotically optimal algorithms in the Bayesian setting. In [20], a CUSUM based algorithm is proposed which is shown to be asymptotically Min-Max optimal. In this algorithm, each sensor runs the CUSUM algorithm and sends a '1' if it detects a change and a '0' otherwise. The fusion center declares a change when all the sensors transmit a '1' simultaneously. In [33] and [34], various distributed change detection algorithms are compared.

The above problem formulations do not explicitly take energy consumption in to account. Furthermore, these algorithms ignore the unreliability of the communication channel. Recently, a Bayesian formulation of the decentralized change detection problem with energy constraints was considered in [37]. The problem is solved using dynamic programming by restricting the solution to a class of algorithms.

In this paper we propose a CUSUM based algorithm called DualCUSUM and show that, for given constraints on false alarm probability and energy, its mean detection delay is much less than that in [37]. Also, DualCUSUM is computationally much less complex and requires no feedback from the fusion node. We also provide the false alarm and delay analysis of our algorithm.

DualCUSUM uses physical layer fusion to reduce transmission delays from different nodes. Physical layer fusion requires phase, frequency and time synchronization of different nodes.

Manuscript received August 5, 2009; revised June 24, 2010; accepted August 14, 2010. The associate editor coordinating the review of this paper and approving it for publication was D. Zeghlache.

V. Sharma and A. K. J. are with the Dept. of Electrical Communication Engineering, IISc, Bangalore, India (e-mail: vinod@ece.iisc.ernet.in, jarun.research@gmail.com).

T. Banerjee is with the University of Illinois at Urbana-Champaign (e-mail: taposh@gmail.com).

V. Kavitha is with the University of Avignon, France (e-mail: Kavitha.Voleti_Veeraruna@sophia.inria.fr).

Preliminary versions of this paper have been presented at ICASSP 2008 and MSWiM 2009.

Digital Object Identifier 10.1109/TWC.2010.12.091177

This is feasible in sensor networks ([19], [30]). However, if one does not provide for such synchronization, DualCUSUM can be used without physical layer fusion (using other MAC layer protocols, e.g., TDMA). Due to other features mentioned above, it still provides good performance (compared to the algorithms available in literature). Even in the absence of an energy constraint, our preliminary investigations indicate that DualCUSUM performs better than other distributed algorithms, some of which have been identified to be asymptotically optimal ([20], [33], [35]).

DualCUSUM has been used for cooperative spectrum sensing in Cognitive Radio Systems in [28] and shown to provide better performance than other algorithms available in literature not only in delay but also in saving energy.

Although this algorithm has many desirable features, there is one practical limitation: to use CUSUM one needs the distribution of observations before and after change at each sensor node. This may not be a realistic assumption in many cases. For example, there can be random time varying fading in the wireless channels in sensor networks (see other articles in [30]), and in the Cognitive Radio Systems ([16], [28]). See also [7] for other practical examples. Thus in this paper we also extend DualCUSUM to a nonparametric set up.

We analyze a generalized version of DualCUSUM of which parametric and nonparametric versions are special cases. A few interesting facts emerge from this analysis: mean detection delay is insensitive to the distribution of the observations but the false alarm probability crucially depends on the tail behavior of the distributions at least for the nonparametric CUSUM. The lighter the tail, the lower the false alarm probability. We also show that the log likelihood function converts a heavy tailed distribution to a light tail distribution. Since, parametric CUSUM uses log likelihood and nonparametric CUSUM does not, the former performs better than the latter for a given distribution of observations.

Since CUSUM is, or will be a fundamental element of many distributed algorithms for detection of change, the tools and techniques used here can be of general interest. Also, since the CUSUM algorithm is essentially a reflected random walk, during our analysis, we obtain new results on passage times, overshoot distribution and excursion above a level for reflected random walks. Despite extensive studies on random walks, there are comparatively few results on reflected random walks ([10]).

The paper is organized as follows. We explain the model and introduce the algorithm in Section II. Section III analyzes the performance of the algorithm and provides comparison with simulations. Section IV concludes the paper.

II. Model and Algorithm

Let there be L sensors in a sensor field, sensing observations and transmitting to a fusion node. The transmissions from the sensor nodes to the fusion node are over a MAC. In our system we assume that all the sensor nodes can transmit at the same time. There is physical layer fusion at the fusion node (commonly studied Gaussian MAC is a special case). The fusion node receives data over time and decides if there is a change in distribution of the observations at the sensors.

Let $X_{k,l}$ be the observation made at sensor l at time k. Sensor l transmits $Y_{k,l}$ at time k after processing $X_{k,l}$ and its past observations. The fusion node receives $Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_{MAC}$, where $\{Z_{MAC}\}$ is iid receiver noise. The distribution of the observations at each sensor changes simultaneously at a random time T, with a known distribution. The $\{X_{k,l}, l \geq 1\}$ are independent and identically distributed (iid) over l and are independent over k, conditioned on change time T. Before the change $X_{k,l}$ have density f_0 and after the change the density is f_1. The expectation under f_i will be denoted by E_i, $i = 0, 1$, and P_∞ and P_1 denote the probability measure under no change and when change happens at time 1, respectively.

These assumptions are commonly made in the literature (see e.g., [7], [35], [36]). Physical layer fusion is considered in [21] and [37].

The objective of the fusion center is to detect this change as soon as possible at time τ (say) using the messages transmitted from all the L sensors, subject to an upper bound α on the probability of False Alarm $P_{FA} \stackrel{\triangle}{=} P[\tau < T]$ and $\mathscr{E}_0$ on the average energy used. Often the desired α is quite low, e.g., $\leq 10^{-6}$ in intrusion detection in sensor networks. Then, the general problem is:

$$\min E_{DD} \stackrel{\triangle}{=} E[(\tau - T)^+],$$
$$\text{Subj to } P_{FA} \leq \alpha \text{ and } \mathscr{E}_{avg} = E\left[\sum_{k=1}^{\tau} Y_{k,l}^2\right] \leq \mathscr{E}_0, 1 \leq l \leq L \quad (1)$$

For this distributed optimization problem there is no optimal solution available so far although asymptotically optimal solution have been identified ([33] - [35]). In the following instead of solving the optimization problem directly, we develop an efficient parametric class of algorithms. We also analyze its performance. This analysis can be used to optimize its parameter. Our algorithm has several desirable features to provide better performance than the algorithms we are aware of (including the asymptotically optimal solutions). This algorithm is called DualCUSUM and is as follows:

1) Sensor l uses CUSUM,
$$W_{k,l} = \max\left\{0, W_{k-1,l} + \log\left[f_1\left(X_{k,l}\right)/f_0\left(X_{k,l}\right)\right]\right\}, \quad (2)$$
where, $W_{0,l} = 0, 1 \leq l \leq L$.
Remark: One can see that the CUSUM is a reflected random walk.
2) Sensor l transmits $Y_{k,l} = b1_{\{W_{k,l} > \gamma\}}$. Here 1_A denotes the indicator function of set A.
Remark: This is the energy saving step. The parameter b is chosen offline based on the energy constraint.
3) Physical layer fusion (as in [37]) is used to reduce transmission delay, i.e., $Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_{MAC,k}$, where $Z_{MAC,k}$ is the receiver noise.
4) Finally, Fusion center runs CUSUM:
$$F_k = \max\left\{0, F_{k-1} + \log\frac{g_I(Y_k)}{g_0(Y_k)}\right\}; \quad F_0 = 0, \quad (3)$$
where g_0 is the density of $Z_{MAC,k}$, the MAC noise at the fusion node, and g_I is the density of $Z_{MAC,k} + bI$, I being a design parameter.
Remark: In the absence of MAC noise, it is Min-Max

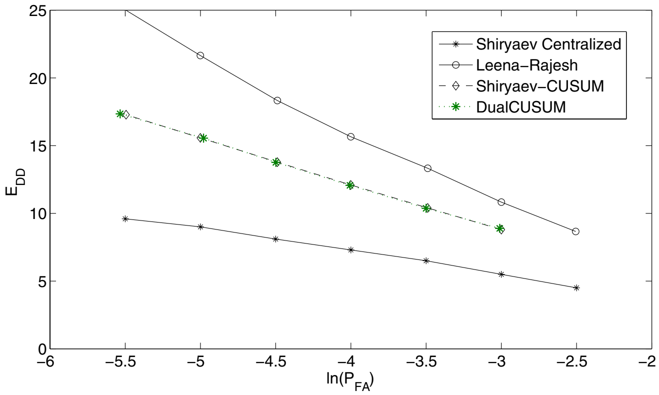


Fig. 1. $\ln(P_{FA})$ (x axis) vs E_{DD} comparison with [37].

optimal for the fusion center to declare change when $Y_k = Lb$. In the presence of noise, such a decision is not possible and hence we use another CUSUM to detect the change. Before the change, sensors transmit rarely and hence Y_k can be approximated by $N(0, \sigma_M^2)$. Also, well after the change has taken place, when all the sensors are transmitting, $Y_k \sim N(Lb, \sigma_M^2)$. But the number of sensors transmitting evolves from 1 to L after the change and hence, we represent Y_k, post change, by $N(Ib, \sigma_M^2)$ and optimize over the choice of I ($1 \le I \le L$).

5) The fusion center declares a change at time $\tau(\beta, \gamma, b, I)$ when F_k crosses a threshold β:

$$\tau(\beta, \gamma, b, I) = \inf\{k : F_k > \beta\}.$$

Remark: After the change, when the mean of Y_k is Lb, the drift of F_k will be positive (because $1 \le I \le L$) and change will be detected with probability 1.

Multiple values of (β, γ, b, I) will satisfy both the false alarm and the energy constraint. One can minimize $E_{DD} \triangleq E[(\tau - T)^+]$ over this parameter set. In Section III we obtain the performance of DualCUSUM for given values (β, γ, b, I) which then can be used to solve the optimization problem:

$$(\beta^*, \gamma^*, b^*, I^*) = \arg \min_{(\beta,\gamma,b,I)} E_{DD}(\beta, \gamma, b, I) \tag{4}$$

subj to $P[\tau(\beta, \gamma, b, I) < T] \le \alpha$, energy $\mathscr{E}_{avg}(\beta, \gamma, b, I) \le \mathscr{E}_0$. For the case of Gaussian distribution and Geometric T and explicit optimization algorithm is provided in [3] to solve (4).

Figure (1) compares the optimal DualCUSUM (obtained via the optimization algorithm in [3]) with the scheme in [37] and the optimal centralized Shiryaev scheme via simulation. We use the parameters: $L = 2, I = 1, f_0 \sim N(0, 1), f_1 \sim N(0.75, 1)$, $Z_{MAC,k} \sim N(0, 1)$, $T \sim \text{Geom}(\rho = 0.05)$ and $\mathscr{E}_0 = 7.61$. Clearly DualCUSUM performs better than [37] and the performance tends to improve as P_{FA} decreases.

If the distribution of T is known, then for a single node, Shiryaev algorithm is optimal ([15], [35]). One could possibly use that also in our setup at the secondary or fusion nodes. However, especially, in cooperative setup its performance analysis may become intractable. DualCUSUM itself has been difficult to analyze. Thus for cooperative Shiryaev algorithm getting optimal parameters will be almost impossible except via simulations. Furthermore, surprisingly DualCUSUM performs as well as an algorithm where Shiryaev algorithm is used at the local nodes and CUSUM at the fusion node. This comparison is also shown in Figure 1. Surprisingly our initial investigations also show that DualCUSUM may work better than the algorithm which uses Shiryaev algorithm at both the local nodes as well as at the fusion center. In addition, DualCUSUM can be used in the non-Bayesian setup. Most of the analysis remains the same.

More recently we have used DualCUSUM for spectrum sensing and shown in [28] that it performs better than several recently proposed algorithms. This motivates us to study DualCUSUM further.

DualCUSUM, as the original CUSUM itself, has a strong limitation. It requires exact knowledge of f_0 and f_1. This information will be available apriori to varying degrees in a practical scenario. Depending upon the type of uncertainty in f_0, f_1, different algorithms/variations on CUSUM are available ([9], [15]). One common algorithm, called nonparametric CUSUM is to replace (2) by

$$W_{k+1,l} = \max\{0, W_{k,l} + X_{k+1,l} - D\}, \tag{5}$$

where, D is an appropriate constant such that $E[X_{k,l} - D]$ is negative before change and positive after change. If the mean of $X_{k,l}$ is known before and after the change, D can be chosen as the average of the two means. For Gaussian and exponential distributions, nonparametric CUSUM becomes CUSUM for some appropriate D and scaling. If at the fusion node g_0 is not known (in our CUSUM algorithm (3) at the fusion node, $g_I(x) = Ib + g_0(x)$), then one can use (5) even at the fusion node.

In the following we will compute P_{FA} and E_{DD} for a generalized class of algorithms where at the sensor nodes and at the fusion node we use the algorithm,

$$W_{k+1} = \max\{0, W_k + Z_{k+1}\}, \tag{6}$$

where, $\{Z_k\}$ is an iid sequence with different distributions before and after the change. (At the fusion node the situation is more complicated; we will comment on it as and when needed). We will assume that $E[Z_k] < 0$ before the change and $E[Z_k] > 0$ after the change. We will denote by f_Z, F_Z and P_Z the density, cdf and probability measure for Z_k.

Algorithm (6) contains CUSUM and nonparametric CUSUM as special cases. In the next section we analyze the generalized DualCUSUM with (6). We emphasize that unlike DualCUSUM, this algorithm *may not* require knowledge of f_0 and f_1, (e.g., we only need to choose D appropriately for nonparametric CUSUM). But, the performance of this algorithm, as we show in the next section, does depend on the underlying distribution. This is typical of such algorithms.

III. Analysis

In this section, we first compute the false alarm probability P_{FA} and then the delay E_{DD}. The idea is to model the times at which the CUSUM $\{W_k\}$ at the local sensors, crosses the threshold γ (we drop subscript l for convenience) and the local nodes transmit to the fusion node (Fig. 2).

Computing P_{FA} requires finding (when Z_k has distribution f_0) the distribution of τ_γ, the first time W_k crosses γ, the

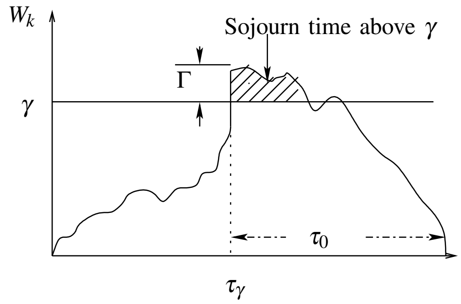


Fig. 2. Excursions of W_k above γ can be approximated by a compound Poisson process. A local node transmits to the fusion node during these excursions.

amount of time it stays above γ (excursion time above γ), and the probability that the fusion node declares a change during an excursion time. These are computed in Sections III-A-III-E. The delay E_{DD} is computed in Section III-G.

We will need the following notations and definitions. Let X be a random variable with distribution F. Then F^{*n} denotes the n-fold convolution of F and $\bar{F}(x) = 1 - F(x)$. A function l is *slowly varying* if for all $\lambda > 0$, $l(\lambda x)/l(x) \to 1$ as $x \to \infty$.

Definition 1: ([1]) F is *heavy tailed* if for any $\varepsilon > 0$, $E[e^{\varepsilon|X|}] = \infty$. F is *subexponential* if $\bar{F}^{*2}(x)/\bar{F}(x) \to 2$ as $x \to \infty$. If F is not heavy tailed, we call it *light tailed*. If $1 - F(x) = l(x)x^{-\alpha}, \alpha > 0$ where l is slowly varying then F is regularly varying with index $-\alpha$.

Gaussian, Exponential and Laplace distributions are light tailed. Pareto, Lognormal and Weibull distributions are subexponential. Subexponential distributions are a subclass of heavy tailed distributions and regularly varying distributions are a subclass of subexponential distributions. We will also be concerned with a sub family $\mathscr{S}^*$ of subexponential distributions defined in [11] which contains all the above members of subexponential family if they have a finite mean.

Often it is said that light tailed distributions may provide better system behavior than the heavy tailed ([8]). We demonstrate this for the probability of false alarm. In particular we will show that if the positive tail of F_Z is light then P_{FA} is much less than if it is heavy tailed. Interestingly, we will also show that E_{DD} is largely insensitive to the tail behavior of F_Z.

CUSUM has the interesting property that it transforms a large class of heavy tailed distributions into light tailed distributions. This important property of log likelihood seems to have escaped the attention of investigators before. This makes CUSUM perform better than the nonparametric CUSUM. Lemma 1 (in Appendix A) states this property in reasonable generality.

A. Behavior of W_k under P_∞

The process $\{W_k\}$ is a reflected random walk with negative drift under P_∞. Figure (2) shows a typical sample path for $\{W_k\}$. The process visits 0 (regenerates) a finite number of times before it crosses the threshold γ at,

$$\tau_\gamma \stackrel{\triangle}{=} \inf\{k \geq 1 : W_k \geq \gamma\}. \tag{7}$$

We call τ_γ the *First Passage Time (FPT)*. The *overshoot* Γ is defined as $W_{\tau_\gamma} - \gamma$. The time between two regenerations is inter-regeneration time τ. Under P_∞, $E[\tau] < \infty$. Let

$$\begin{aligned} \tau_0 &\stackrel{\triangle}{=} \inf\{k : k > \tau_\gamma; W_k \leq 0\} - \tau_\gamma \textit{ and,} \\ \eta &= \#\{k : W_k \geq \gamma; \tau_\gamma \leq k \leq \tau_\gamma + \tau_0\}. \end{aligned} \tag{8}$$

During time η (called a batch) a local node transmits to the fusion node. Thus, these are the times during which the fusion node will most likely declare a change. The overshoot Γ can have significant impact on η.

It has been shown in [24] that the point process of exceedances of γ by W_k, converges to a compound Poisson process as $\gamma \to \infty$. The points appear as clusters. The intervals between the clusters have the same distribution as that of τ_γ in (7) and the distribution of η in (8) gives the distribution of the size of the cluster, i.e., the batch of the compound Poisson process. Since, one has to choose large values of γ to keep P_{FA} small, a batch Poisson process provides a good approximation in our scenario.

In the next few sections we give results on the distribution of τ_γ, overshoot Γ, and the distribution of the batch η which will be used in computing P_{FA}.

B. First Passage Time under P_∞

From the compound Poisson process approximation mentioned above,

$$\lim_{\gamma \to \infty} P_\infty[\tau_\gamma > x] = \exp(-\lambda_\gamma x), \; x > 0, \tag{9}$$

where, λ_γ a positive constant. In [3] a formula for λ_γ was used which is computable for Gaussian distribution only. However, by solving integral equations obtained via renewal arguments ([25]), one can obtain the mean of FPT for any distribution. Epochs when $W_k = 0$ are renewal epochs for this process. Let $L(s)$ be the mean FPT with $W_0 = s \geq 0$. Hence $\lambda_\gamma = 1/L(0)$. Then from renewal arguments:

$$\begin{aligned} L(s) &= F_Z(-s)(L(0)+1) \\ &\quad + \int_{-s}^{\gamma - s} (L(s+z)+1) dF_Z(z) dz + P[Z > \gamma - s]. \end{aligned} \tag{10}$$

This equation is obtained by conditioning on $Z_0 = z$. If $Z_0 \leq -s$, then $W_1 = 0$, providing the first term on the right. If $Z > \gamma - s$, then the threshold is approached in one step only, providing the last term.

Equation (10) can be shown to be a Fredholm integral equation of second kind ([26]). Theorem 1 shows that Equation (10) has a unique continuous solution in our set up under weak conditions.

Theorem 1: If F_Z is continuous and $F_Z(\gamma) < 1$ then (10) has a unique continuous solution L.

Proof: See Appendix B. ■

Equation (10) can be solved recursively on $L(s), 0 \leq s \leq \gamma$. An efficient algorithm is provided in [18].

From (10) not only we can compute the $E[\tau_\gamma]$ exactly, but can also get some asymptotic rates. Taking $s = 0$ in (10), and writing $L(0)$ as $L_\gamma(0)$ to make dependence on γ explicit, we get (since $L_\gamma(0) \geq L_\gamma(y)$ for $0 \leq y \leq \gamma$)

$$\begin{aligned} L_\gamma(0) &= 1 + F_Z(0)L_\gamma(0) + \textstyle\int_0^\gamma L_\gamma(y) f_Z(y) dy \\ &\leq 1 + F_Z(0)L_\gamma(0) + L_\gamma(0)(F_Z(\gamma) - F_Z(0)). \end{aligned}$$

TABLE I
MEAN FPT $E[\tau_\gamma]$ FOR PARETO ($K = 2.1$) AND GAUSSIAN WITH $EZ_k = -0.5$.

γ	$E[\tau_\gamma]$ Gauss	$E[\tau_\gamma]$ Pareto
5	930	800
6	2551	1100
7	6950	1455
8	19020	1880

Thus,

$$L_\gamma(0) \leq \frac{1}{1-F_Z(\gamma)}. \quad (11)$$

Equation (11) provides the dependence of $E[\tau_\gamma]$ on the tail of distribution of F_Z. For example, if $1-F_Z(\gamma) \sim \gamma^{-\alpha}, \alpha > 0$, then $E[\tau_\gamma] \leq \gamma^\alpha$ and if F_Z is light tailed ($\sim e^{-\alpha\gamma}$) then $E[\tau_\gamma] \sim e^{\alpha\gamma}$.

Although (11) only gives an upper bound on the growth of $E[\tau_\gamma]$ with γ, it turns out that this upper bound in fact gives the exact rate of growth. This can be seen from the following facts (due to lack of space we will be brief). Let p_γ be the probability of W_k exceeding γ in one regeneration length. Let $E[\tau]$ be the mean regeneration length. From [14],

$$\frac{p_\gamma E[\tau_\gamma]}{E[\tau]} \to 1 \text{ as } \gamma \to \infty. \quad (12)$$

From [1], if Z has subexponential distribution then for γ large, $p_\gamma \approx P[Z > \gamma]E[\tau]$ and hence from (12), $E[\tau_\gamma] \sim 1/P[Z > \gamma]$. If there exists an $\bar{\gamma} > 0$ with $E[e^{\bar{\gamma}Z}] = e^{\bar{\gamma}}$ then from [27], $\gamma^{-1}\log p_\gamma \to -\bar{\gamma}$ and hence from (12) we get $\gamma^{-1}\log E[\tau_\gamma] \to \bar{\gamma}$.

Table I provides $E[\tau_\gamma]$ for Pareto distribution with $K = 2.1$ and Gaussian distribution with $E[Z_k] = -0.5$ and $var(Z_k) = 1$. We see that as γ increases $E[\tau_\gamma]$ for Gaussian distribution becomes much larger than for the Pareto distribution. This implies that P_{FA} for the Gaussian distribution should be much less than for Pareto, $K = 2.1$ if γ is large.

C. Distribution of overshoot

Next we consider the mean and the distribution of the overshoot Γ. From renewal equations as in (10), we can exactly compute $E[\Gamma]$ for any distribution. If $R(x) = E[\Gamma]$ with $W_0 = x$,

$$\begin{aligned} R(x) &= E[Z_k - (\gamma - x)|Z_k > (\gamma - x)]P_Z[Z_k > \gamma - x] + \\ &\int_{y=0}^{\gamma} R(y) f_Z(y-x)dy + R(0)F_Z(-x). \end{aligned} \quad (13)$$

The mean overshoot $E[\Gamma]$ equals $R(0)$. Similar to the equation (10), this is also a Fredholm integral equation of second kind. Thus, we can obtain existence of a unique continuous solution of this equation as in Theorem 1 under the same conditions. For light tails $E[\Gamma]$ converges quickly to a constant value as $\gamma \to \infty$. Thus for light tails (13) can be evaluated for a much smaller value of γ which can then be used for all higher values of threshold as well.

As in case of (10), (13) also provides some asymptotic rates and dependence of $E[\Gamma]$ on the tails of Z. Taking $x = 0$ in (13) and denoting $R(0)$ as $R_\gamma(0)$, we get

$$\begin{aligned} R_\gamma(0) &= E[Z - \gamma|Z > \gamma]P[Z > \gamma] + R_\gamma(0)F_Z(0) \\ &+ \int_{y=0}^{\gamma} R_\gamma(y) f_Z(y)dy, \end{aligned}$$

and hence $R_\gamma(0) \geq \frac{E[Z-\gamma)|Z>\gamma]P[Z>\gamma]}{1-F_Z(0)}$.

If $1 - F_Z$ is of regular variation with index $-\alpha$ then

$$E[Z-\gamma)|Z > \gamma]P[Z > \gamma] = \int_\gamma^\infty z dP_Z(z) - \gamma P[Z > \gamma]$$

is of regular variation with index $-\alpha + 1$ and hence $R_\gamma(0) \geq l(\gamma)\gamma^{-\alpha+1}$ for slowly varying function l.

If Z is of exponential type, i.e., $\lim_{x\to\infty} \frac{f_Z(x+\gamma)}{f_Z(x)} = e^{-\lambda\gamma}$ for all $\gamma > 0$ for some $\lambda > 0$, then $R_\gamma(0) \geq \beta e^{-\lambda\gamma}$ for large γ. This suggests that for heavy tailed Z mean overshoot will be much more. The following results further strengthen this. Let $M(\tau) = \max\{W_k, 0 \leq k \leq \tau - 1\}$.

Theorem 2: The following hold:

(a) If $Z \in \mathscr{S}^*$ then for $x > 0$, $P[\Gamma(\gamma) > x] \leq P[M(\tau) > \gamma + x|M(\tau) > \gamma] \to 1$ as $\gamma \to \infty$ and $M(\tau)$ is subexponential.

(b) If Z is regular with index $-\alpha$, $\alpha > 1$, then $M(\tau)$ is regular with index $-\alpha$ and for any $\varepsilon > 0$, $\Gamma(\gamma)\gamma^{\frac{-1}{(\alpha-\varepsilon)}} \to 0$ a.s. and $E[\Gamma(\gamma)]\gamma^{\frac{-1}{(\alpha-\varepsilon)}} \to 0$ as $\gamma \to \infty$.

(c) If there is an $\alpha > 0$ such that $E[e^{\alpha Z}] = 1$ then Γ is light tailed and $E[\Gamma(\gamma)] \lessapprox e^{-\alpha\gamma}$.

Proof: See Appendix C. ■

Theorem 2(c) states that if Z is light tailed, $E[\Gamma(\gamma)]$ decays exponentially with γ. The following discussion suggests that $\Gamma(\gamma)$ has an exponential distribution as $\gamma \to \infty$.

To express the results related to distribution of the overshoot, we need the concept of *Maximum Domain of Attraction (MDH)*. Let $M_n = \max\{W_1, \ldots, W_n\}$. Since, $\{W_k\}$ is Harris ergodic and hence strongly mixing, (see [1]), $a_n(M_n - b_n) \xrightarrow{d} H$, where $\xrightarrow{d}$ denotes convergence in distribution and a_n, b_n are appropriate positive constants. Here H is either a Frechet distribution, $H(x) = \exp(-x^{-\alpha}), x \geq 0$, for some $\alpha > 0$ or the Gumbel distribution, $H(x) = \exp(-e^{-x}), -\infty < x < \infty$. The distribution of W_k is said to belong to the *MDA* of H. The *MDA* of Subexponential distributions is a Frechet distribution, while light tailed distributions belong to the *MDA* of the Gumbel distribution.

For subexponential distributions, with Z_k in *MDA* of an H with parameter α, ([1]),

$$\lim_{x\to\infty} \bar{F}^{(\gamma)}(\omega(\gamma)y) = P_\alpha(y), \quad (14)$$

where, $\bar{F}^{(\gamma)}(x) = 1 - (F_0(x+\gamma) - F_0(x))/\bar{F}_0(x)$, $\omega(\gamma) = E[Z_k - \gamma|Z_k > \gamma]$ and P_α is the generalized Pareto distribution, with

$$P_\alpha(y) = \begin{cases} (1 + y/(\alpha-1))^{-\alpha}, & \alpha < \infty, \\ e^{-y} & \alpha = \infty, \end{cases} \quad y > 0. \quad (15)$$

Here, $\alpha < \infty$ corresponds to the Frechet case and $\alpha = \infty$ to the Gumbel case.

We plot the distribution of overshoot for Pareto distribution with $K = 2.1$ in Figure (3). The mean overshoot $\omega(\gamma)$ was obtained using equation (13). We observe that equation (15) gives a very good estimate of the overshoot distribution. We have verified that (15) is a good approximation even when Z_k is Lognormal (with $\omega(\gamma)$ obtained from (13)).

The above arguments suggest that even for the light tailed distributions, the overshoot converges to exponential distribution where the mean can be obtained from (13). We plot this

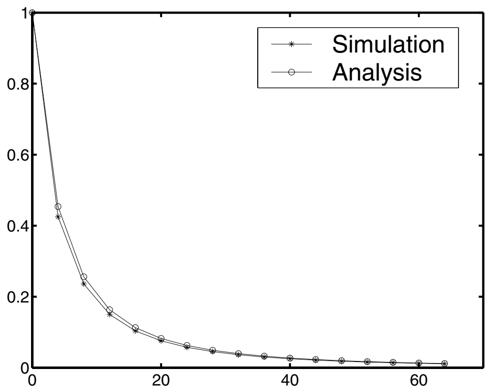


Fig. 3. Complementary CDF of Γ for Pareto $K = 2.1$, $EZ_k = -0.3$ and $var(Z_k) = 1$ and $\gamma = 8$.

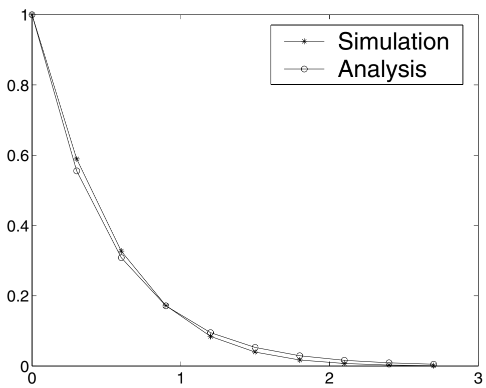


Fig. 4. Complementary CDF of Γ for $Z_k \sim N(-0.3, 1)$ and $\gamma \geq 6$.

approximation for Gaussian distribution in Figure (4) and find an excellent match with simulations. We have verified this for Laplace distribution also.

Comparing Figures (3) and (4), we see that the overshoot for Pareto distribution is much more than for the Gaussian distribution.

D. Distribution of the Batch

In this section we give the distribution of the batch. Although the distribution of batch for sub-exponential tails is given in [1], the one for light tails is not previously available in the literature (for example, it is not explicitly provided in [24]).

1) Distribution of batch for heavy tail: From Theorem 2.4 of [1], the batch size distribution for subexponential Z (belonging to the *MDA* of a Frechet distribution H with parameter α) satisfies

$$\frac{E[Z]}{\omega(\gamma)}\eta \xrightarrow{d} Y_\alpha, \tag{16}$$

as $\gamma \to \infty$, where $\omega(\gamma) = E[Z - \gamma | Z > \gamma]$ and Y_α has distribution P_α.

Figure (5) shows the plot of Batch complementary CDF for Pareto distribution with parameters $K = 2.1$. One sees a good match with simulations.

2) Distribution of batch for light tail: Let $G_j(x)$ be the conditional batch distribution,

$$G_j(x) = P[\eta \leq j | W_{\tau_\gamma} = \gamma + x],$$

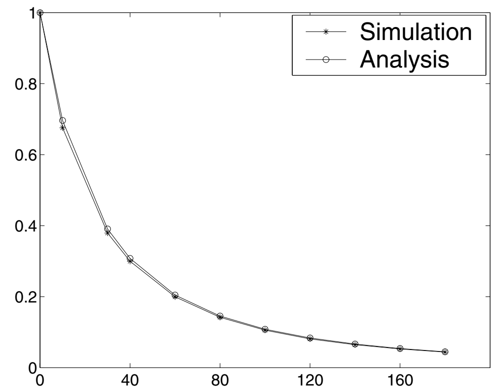


Fig. 5. Complementary CDF of Batch η for Pareto $K = 2.1$, $EZ_k = -0.3$ and $var(Z_k) = 1$ and $\gamma = 15$.

when the overshoot is x. We now obtain $G_j(x)$ using Brownian Motion (BM) approximation of $\{W_k\}$.

The reflected random walk $\left\{W_{k+\tau_\gamma}\right\}_{k\geq 0}^{\tau_0}$ is given by an ordinary random walk. Further, with large values of γ (needed for large P_{FA}), τ_0 is sufficiently large. Thus, using Donsker's theorem [6] we approximate (with large N):

$$\begin{aligned}\left\{W_{k+\tau_\gamma}\right\}_{k\geq 0}^{\tau_0} &\sim \left\{W_{\tau_\gamma} + S_{k,l}\right\}_{k\geq 0}^{\tau_0} \\ &\sim \left\{W_{\tau_\gamma} + \sigma_S\sqrt{N}\zeta\left(\frac{k}{N}\right) + k\mu\right\}_{k\geq 0}^{\tau_0}\end{aligned}$$

where $\zeta(t), t \geq 0$ is a standard Brownian motion (BM), $\mu = EZ_k$ and $\sigma_S = var(Z_k)$. Given $W_{\tau_\gamma} = \gamma + x$, τ_0 is approximated by the time taken by the above BM to reach 0 starting with $\gamma_{ov} = \gamma + x$. This is given by ([13]):

$$P[\tau_0 > i] = \Phi\Big(\frac{\gamma_{ov} - \mu i}{\sigma_S\sqrt{i}}\Big) - e^{\frac{2\mu\gamma_{ov}}{\sigma_S^2}}\Phi\Big(\frac{-\gamma_{ov} - \mu i}{\sigma_S\sqrt{i}}\Big), \tag{17}$$

where Φ denotes the CDF of the standard Gaussian distribution.

We obtain the batch distribution using occupation measure, above γ, of the BM till time τ_0 ([32]). Choose time t_B such that for some small enough $\varepsilon > 0$, $P[\tau_0 \leq t_B] > 1 - \varepsilon$ and $P[\tau_\gamma \geq t_B] > 1 - \varepsilon$. This is possible if, $P[\tau_0 << \tau_\gamma 2]$ is close to 1, which is true for small P_{FA} (and hence large γ).

Define $\delta = (\gamma + x)\big/(\sigma_S\sqrt{t_B})$, and $m = \mu\sqrt{t_B}\big/\sigma_S$ The conditional batch size distribution is approximated using, [32], as

$$\begin{aligned}G_j(x) = 2\int_0^j &\Big[\frac{\varphi(m\sqrt{1-u})}{\sqrt{1-u}} + m\Phi(m\sqrt{1-u})\Big] \\ &\Big[\varphi\Big(\frac{\delta - mu}{\sqrt{u}}\Big)\frac{1}{\sqrt{u}} - me^{2m\delta}\Phi\Big(\frac{-\delta - m}{\sqrt{u}}\Big)\Big]du,\end{aligned} \tag{18}$$

where, φ represents the standard Gaussian pdf. Since the overshoot distribution is exponential, for light tailed Z_k,

$$P[\eta \leq j] = \int_0^\infty G_j(x)\frac{1}{E[\Gamma]}\exp(-\frac{x}{E[\Gamma]})dx \tag{19}$$

The mean overshoot $E[\Gamma] = R(0)$, where $R(0)$ is obtained from equation (13). Figure (6) plots the distribution of η for Z_k with Laplace distribution via (19) and via simulations.

For Lognormal distribution, which can be approximated via both heavy tailed and light tailed approximations provided

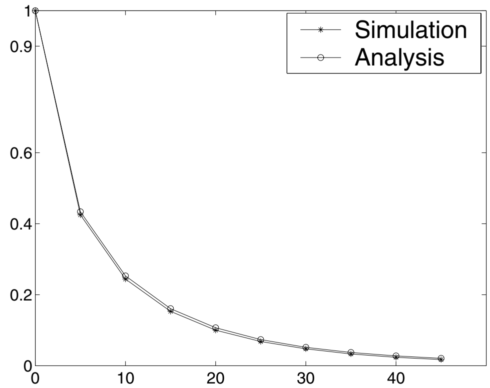


Fig. 6. Complementary CDF of Batch η for Laplace Z_k with $EZ_k = -0.3$ and $var(Z_k) = 1$ and $\gamma \geq 7$.

TABLE II
P_{FA} FOR VARIOUS DISTRIBUTIONS USING (5) AT THE LOCAL NODE AND (3) AT THE FUSION NODE: $EZ_k = -0.3$, $var(Z_k) = 1$, $\rho = 0.005$ AND $b = 1$.

	L	I	γ	β	P_{FA} Anal. $\times 10^{-4}$	P_{FA} Sim. $\times 10^{-4}$
Gauss	5	2	15	18	1.22	1.1
	10	2	15	18	2.43	2.28
Laplace	6	2	16	16	2.57	2.06
	12	3	16	16	0.66	0.55
Log-normal	5	2	25	20	1.47	1.76
	10	2	25	20	2.97	3.5
Pareto K=2.1	5	3	30	30	1.93	1.77
	5	3	50	50	0.23	0.25

above, (19) provides a better approximation.

Comparing Figures (5) and (6) one sees that the batch size for a Pareto distribution is larger than for a Laplace distribution even when they have same mean and variance. This is a direct consequence of having larger overshoots.

E. False Alarm Analysis

The false alarm in DualCUSUM can happen in two ways: one within a batch (we denote its probability by $\tilde{p}$) and another outside it, i.e., due to $\{Z_{MAC,k}\}$. We will compute these later on. First, we compute the P_{FA} from these quantities.

From the assumptions made and the above approximation, the inter-arrival time of the batches in the system (at the fusion center) is exponentially distributed with rate $L\lambda_\gamma$ (because the processes $\{W_{k,l}\}$ are independent for different nodes each generating batches as Poisson processes with rate λ). Then, the number of batches appearing before the time of change is a Poisson random variable with parameter $L\lambda_\gamma i$, when $T = i$. In the following, we will show that the time to FA outside a batch is exponentially distributed with parameter λ_0 (to be defined below). Therefore, if $T \sim \text{Geom}(\rho)$, then one can show that:

$$P_{FA} = 1 - \frac{e^{-(\lambda_0 + \lambda_\gamma L\tilde{p})}\rho}{1 - e^{-(\lambda_0 + \lambda_\gamma L\tilde{p})}(1-\rho)}. \tag{20}$$

Similarly, one can obtain expression for P_{FA} when T is not geometric.

False Alarm within a Batch:

We have seen above that for light tailed Z_k, the $E[\tau_\gamma]$ is large and the batch sizes are small. Thus, the batches by different local nodes do not overlap. However, it is not true for heavy tailed distributions. Thus we compute the $\tilde{p}$ for the two cases separately.

Light Tailed

The false alarm probability $\tilde{p}$ within a batch, can be computed as, $\tilde{p} \approx \sum_{i=1}^{\infty} P[\eta = i] P[\text{FA} \mid \eta = i]$, where $P[\text{FA} \mid \eta = i]$ represents the probability of FA (CUSUM at the fusion center crossing β) in i transmissions when one local node is already transmitting, i.e., $Y_k = b + Z_{MAC,k}$. If τ_β is the FPT variable at the fusion center, then, $P[\text{FA} \mid \eta = i] = P[\tau_\beta \leq i]$. Since η is small for negative drift under f_0 (since D in (5) is chosen that way) we use integral equations to compute the distribution of τ_β for observations Y_k given in this paragraph.

Table II gives the comparison of the P_{FA} values obtained via (20) and simulations for light tailed distributions (Gaussian and Laplace). It turns out that the expression is also valid for heavy tailed distributions like Lognormal (also shown in Table II). One can see a good match.

Heavy Tailed

Now, we use different arguments to compute $\tilde{p}$ and then use it in (20). For simplicity, in the following, the fusion center is assumed to use (3) for detection and not nonparametric CUSUM. From [3], the optimal choice of I is found to be always greater than 1.

Let m be the minimum number of sensors required to make drift of F_k positive. We denote by μ_m the drift with m nodes transmitting. Then we approximate $\tilde{p}$ by the probability that F_k will have positive drift during a batch and that the batch lasts for β/μ_m time (the time needed for F_k to cross β when the drift is μ_m) after m sensors start transmitting. We compute this in the following.

Within a batch of size η, let T_1 be the time at which one out of the remaining $L-1$ nodes transmit. Let the second transmission (one out of $L-2$) happens at $T_1 + T_2$, and so on. Since τ_γ is exponential, T_i are also exponential with parameter $(L-i)\lambda_\gamma$ if $\mu_{i+1} < \mu_m$. Then,

$$\tilde{p} \approx P\left[T_1 + T_2 + \ldots + T_{m-1} + \frac{\beta}{\mu_m} < \eta\right].$$

We use this approximation to compute P_{FA} for Pareto $K = 2.1$ distribution. This is also provided in Table II. We see that the approximation is indeed good for Pareto $K = 2.1$.

False Alarm outside a Batch

In the absence of any transmission from the sensors, $Y_k \sim N(0, \sigma_{MAC}^2)$ if $Z_{MAC} \sim N(0, \sigma_{MAC}^2)$, where $N(0, \sigma_{MAC}^2)$ denotes Gaussian distribution with mean 0 and variance σ_{MAC}^2. Hence, F_k has negative drift. Thus the time to first reach β, i.e., time till FA, is approximately exponentially distributed with parameter λ_0 which can be obtained from Section III-B.

F. Comparative overall performance

The effect of tail of Z_k on FPT, overshoot and batch size was shown in the previous sections. This causes much larger P_{FA}

TABLE III
COMPARISON OF E^*_{DD} OF GAUSSIAN ($I^*=3$) AND PARETO ($I^*=4$) WITH $L=5$, $\mathscr{E}_0=5$, $EZ_k=-0.5$, $var(Z_k)=1$ AND $Z_{MAC}=N(0,1)$.

ρ	P_{FA}	E^*_{DD} Gauss	E^*_{DD} Pareto K=2.1
5e-4	e-2	29	36
5e-4	e-3	33	49
1e-4	e-3	41	95

TABLE IV
COMPARATIVE PERFORMANCE OF PARAMETRIC AND NONPARAMETRIC DUALCUSUM FOR $P_{FA}=0.01$ WITH $\rho=0.05$, $\mathscr{E}_0=7.61$, AND $Z_{MAC}=N(0,1)$.

$f_0 \to f_1$	L/I	E^*_{DD} nonparametric	E^*_{DD} parametric
Pareto $x_m=1$ $K=7$ to $K=3$	5/4	54.8	4.4
Pareto $x_m=1$ $K=40$ to $K=30$	5/4	69.1	24.9
Gaussian $\sigma=1$ $EZ_k=0$ to $EZ_k=0.6$	5/3	10.1	10.1

for heavy tailed Z_k compared to the light tailed distributions for same mean and variance. This gets reflected into large E_{DD} for heavy tailed distributions for a given P_{FA}. Table III confirms these conclusions as the E_{DD} for a light tailed system is much smaller as compared to the one from a heavy tailed system. The individual systems are optimized to make sure that each performs at its best.

Table IV shows the comparative performance of parametric and nonparametric DualCUSUM's for given f_0 and f_1. The difference in performance is most pronounced when the tail of f_0 is heaviest, i.e. for $K=7$, while the performance is same for Gaussian distributions on which log likelihood function has no effect.

Note that in Table IV, the variance of Z_k is different for parametric and nonparametric CUSUMs. The overall effect is thus a combination of the effect of tails and that of the variances. However, as can be seen from the table, the effect of tail dominates and the general conclusion that light tailed systems are better, still holds.

E_{DD} in Tables III and IV is computed via simulations. However in the next section we theoretically evaluate E_{DD} and then compare with the simulated values.

G. Computation of E_{DD}

The mean detection delay, E_{DD}, at the fusion node, after the change has occurred, can be written as,

$$E_{DD} = E\left[(\tau-T)^+\right] = E[\tau-T|\tau\geq T](1-P_{FA}). \qquad (21)$$

When $\mu=E[Z_k]>0$, the time τ_γ for W_k at a local node to cross threshold γ satisfies $E[\tau_x]/x \to 1/\mu$ as $x\to\infty$. Thus for large γ, $E[\tau_\gamma]\sim\gamma/\mu$.

Let μ_l be the drift of fusion CUSUM F_k when l local nodes are transmitting.

TABLE V
COMPARISON OF E_{DD} FOR VARIOUS DISTRIBUTIONS: $L=10, I=1, \beta=\gamma$ $EZ_k=-0.3$, VAR(Z_k)=1 AND $b=1$.

γ	E_{DD} $Anal.$	E_{DD} $Gauss$	E_{DD} $Laplace$	E_{DD} $Log-normal$	E_{DD} $Pareto$ $K=3$
5	5.3	9.1	9.3	9.3	10.7
8	11.4	16.6	16.8	16.9	18.7
15	30.3	36.3	36.5	36.7	38.5
50	146.7	146.8	147.1	147.6	150.5

Let $L=1$. Let the change take place at $k=0$. After approximately τ_γ slots the local node will start transmitting signal level b to the fusion center. Hence, after τ_γ slots the drift of F_k is μ_1. Since $L=1$, μ_1 has to be positive for reasonable system performance. Then, the mean time for fusion center to touch threshold β, for large β is approximately β/μ_1. Therefore a reasonable asymptotic estimate of E_{DD}, for large γ and β is, $E_{DD}\approx\gamma/\mu+\beta/\mu_1$. We have verified that this is a good approximation even for small positive drifts μ,μ_1.

For $L\geq 2$, $\gamma/\mu+\beta/\mu_1$ is not a good approximation for E_{DD}. This is because of three reasons. First, when there is more than one node running CUSUM W_k, any one of them can cross γ, and the time for the first among them to cross is much less than $\frac{\gamma}{\mu}$, especially when L is large. Second, as the number of nodes crossing γ increases, the drift at the fusion node changes from μ_0 through μ_L. Finally, depending on the choice of I or D (based on whether (3) or (5) is used at the fusion center), some of the μ_l's can be negative or zero. Taking these factors into account, we have developed an approximation for E_{DD} which works quite well for $L>1$ (see [4]). However, in the following we use a somewhat different approach which is useful in more general scenario also.

Using the above approximation via LLN and via central limit theorem approximation, we can show that for each node, $\tau_\gamma\sim N(\frac{\gamma}{\mu},\frac{\sigma^2\gamma}{\mu^3})$. Thus, to compute the time $\tau(l)$ when l nodes start transmitting one can compute the lth order statistics of L i.i.d. random variables with the distribution of τ_γ. Let I nodes need to transmit before the CUSUM at the fusion node has drift μ_I positive. Then we approximate E_{DD} by $E[\tau(I)]+\beta/\mu_I$ where β/μ_I approximate the time CUSUM at the fusion node takes to cross β with drift μ_I.

Since, the strong law of large number and the central limit theorem suffice to build the approximations, the E_{DD} is independent of the distribution of Z_k but depends only on its mean and variance. The results are shown in Table V for different distributions. The second column is our approximation developed above and the rest are obtained via actual system simulations. It can be seen that, as γ reaches 50, the E_{DD} of all the distributions considered is nearly 147.

IV. CONCLUSIONS

We have proposed an energy efficient distributed change detection scheme which uses the physical layer fusion technique and CUSUM at the sensors as well as at the fusion center. We have shown that it performs better than various algorithms available in literature. We also extended the algorithm to also include the nonparametric CUSUM. We have theoretically computed the probability of false alarm and mean delay in change detection for the general algorithm. The analytical results provide good approximations for different distributions. Our analysis provides interesting conclusions and insights. One is that the tail of the distribution has significant effect on the performance of nonparametric CUSUM. We also show why parametric CUSUM is relatively insensitive to the tails. In the process we obtain new results on the reflected random walk which can be of independent interest.

Appendix A

This section states and proves Lemma 1 which shows that log likelihood converts a large class of distributions into light tailed distributions. Let $Z = \log\frac{f_1(x)}{f_0(x)}$ and $g(x) = \frac{f_1(x)}{f_0(x)}$. Then, the following hold:

Lemma 1: (a) If $g(x) \le x^{\beta}$ for some $\beta > 0$, and all x large enough and $1-F_0(x) \le x^{-\alpha_0}$ for all large x then the positive tail of distribution of Z decays exponentially with parameter α_0/β.

(b) If $g(x) \le \exp(\alpha x^{\beta})$ for some $\alpha,\beta > 0$ and all x large and $1-F_0(x) \le \exp(-\alpha_0 x^{\beta_0})$ for $\alpha_0,\beta_0 > 0$, then $P[z > x] < \exp(-\alpha_0 \frac{x}{\alpha}^{\beta_0/\beta})$.

Proof: (a) For $x > 0$, $P_0[Z > x] = P_0[g(X) > e^x] \le P_0[X^{\beta} > e^x] \le e^{-x\alpha_0/\beta}$.

(b) For $x > 0$, $P_0[Z > x] = P_0[g(X) > e^x] \le P_0[\exp(\alpha X^{\beta}) > e^x] = P_0[\alpha X^{\beta} > x] \le \exp(-\alpha_0 \frac{x}{\alpha}^{\beta_0/\beta})$. ■

The above Lemma covers a large number of cases as we illustrate now. Part (a) of the theorem shows that if f_0 and f_1 are heavy tailed, Z can become light tailed. Part (b) of theorem shows that light tailed f_0, f_1 will keep Z light tailed. For example let F_1 and F_0 be of regular variation with parameters $-\alpha_0$ and $-\alpha_1$, i.e., $1-F_i(x) = l_i(x)x^{-\alpha_i}, i = 1,2$ for $x > 0$, where l_i are slowly varying functions, and $\alpha_i > 0$ with $\alpha_0 \ne \alpha_1$. Then, Theorem 1(a) applies. If $\alpha_0 > \alpha_1$, $g(x) = l(x)x^{\alpha_0-\alpha_1} \le x^{\alpha_0-\alpha_1+\beta_1}$ for any $\beta_1 > 0$ for all large x. Also, $1-F_0(x) \le x^{\alpha_0+\beta_2}$ for any $\beta_2 > 0$ for x large enough. Chose $0 < \beta_2 < \alpha_0$. Hence, $P_0[Z > x] < \exp(-x(-\alpha_0+\beta_2)/(\alpha_0-\alpha_1+\beta_1))$ for all x large enough providing Z with light tail under P_0. If $\alpha_0 < \alpha_1$, then $g(x) < x^{\beta_1}$ for any $\beta_1 > 0$ and we get $P_0[Z > x] < \exp(\frac{-x(\alpha_0-\beta_2)}{\beta_1})$.

Next consider exponential distributions: $f_i(x) = \lambda_i \exp(-\lambda_i x)$, $\lambda_i > 0$, $x > 0$, $\lambda_0 \ne \lambda_1$. Then, $g(x) \le e^{|\lambda_1-\lambda_0|x}$ and $P_0[Z > x] < e^{-\lambda_0(x-\log\lambda_1/\lambda_0)/|\lambda_1-\lambda_0|}$. Thus Z is light tailed under P_0.

Now we show the versatility of the above result by considering $f_1(x) = \beta_1 x^{-\alpha_1}$ and $f_0(x) = \exp(-(x-\mu_0)^2/2\sigma^2)/\sqrt{2\pi}\sigma_0$. Then, $g(x) = \beta_1 x^{-\alpha_1}\exp((x-\mu_0)^2/2\sigma^2)\sqrt{2\pi}\sigma_0 \le \exp(\alpha x^2)$ for appropriately chosen $\alpha > 0$ for all large x. Thus, since

$$
\begin{aligned}
1-F_0(x) &\le \exp(-(x-\mu_0)^2/4\sigma_0^2) \\
&\le \exp(\alpha_0 x^2), \\
P_0[Z > x] &\le \exp(\frac{-\alpha_0 x}{\alpha_1}).
\end{aligned}
$$

Appendix B

This appendix provides the proof of Theorem 1.

Proof: Obtain an equation for $L(0)$ by substituting 0 for s in (10) and then plug in the expression for $L(0)$ in (10). We obtain

$$
\begin{aligned}
L(s) = &\left(1+\frac{F_Z(-s)}{1-F_Z(0)}\right) + \\
&\int_0^{\gamma} L_{\gamma}(y)\left(f_Z(y-s)+\frac{f_Z(y)F_Z(-s)}{1-F_Z(0)}\right)dy \quad (22)
\end{aligned}
$$

which is Fredholm integral equation of second kind with kernel $k(s,y) = f_Z(y-s) + \frac{f_Z(y)F_Z(-s)}{1-F_Z(0)}$ and we consider the mapping $f \mapsto g$ defined by

$$
g(s) = \int_0^{\gamma} k(s,y)f(y)dy \quad (23)
$$

on the space of functions $L_2([0,\gamma])$. From [26] (pp. 269-270), to show that (22) has a continuous solution, for F_Z continuous we need to show that $\phi(s) = \int_0^{\gamma} k(s,y)\phi(y)dy$ has a unique solution which then is the trivial solution $\phi(s) = 0$.

For this we show that (23) is a contraction mapping on the space of continuous functions on $[0,\gamma]$ (with sup norm). We have, for $\|f\| \stackrel{\triangle}{=} \sup_{0\le y\le\gamma}|f(y)|$,

$$
\begin{aligned}
\|g\| &= \sup_{0\le s\le\gamma} |\int_0^{\gamma} k(s,y)f(y)dy| \\
&\le \|f\| \sup_{0\le s\le\gamma}\left[\int_0^{\gamma} f_Z(y-s)dy + \int_0^{\gamma} f_Z(y)\frac{F_Z(-s)}{1-F_Z(0)}dy\right] \\
&= \|f\| \sup_{0\le s\le\gamma}\left[F_Z(\gamma-s) - F_Z(-s) + \frac{F_Z(-s)(F_Z(\gamma)-F_Z(0))}{1-F_Z(0)}\right] \\
&= \|f\| \sup_{0\le s\le\gamma}\left[\frac{(F_Z(\gamma-s)(1-F_Z(0)) + F_Z(-s)(F_Z(\gamma)-1)}{1-F_Z(0)}\right] \\
&< \|f\| \sup_{0\le s\le\gamma} F_Z(\gamma-s) \\
&\le \|f\| F_Z(\gamma).
\end{aligned}
$$

Thus if $F_Z(\gamma) < 0$, this operator is a contractor and hence has a unique fixed point which is the trivial function $\phi(s) = 0$. ■

Appendix C

This appendix provides the proof of Theorem 2.

Proof: (a) If $Z \in \mathscr{S}^*$ then from [1], Theorem 2.1

$$
P[M(\tau) > x] \approx E[\tau]P[Z > x] \quad (24)
$$

for large x. Thus $M(\tau)$ is subexponential if Z is. Let $\{Y_k\}$ be i.i.d. with the distribution of $M(\tau)$. Let $N(\gamma) = \inf\{n : Y_n > \gamma\}$. Then $W_{\tau_{\gamma}} \le_{st} Y_{N(\gamma)}$ ($X \le_{st} Y$ denotes $P[X \le x] \ge P[Y \le x]$ for all x). Thus for $x > 0$,

$$
\begin{aligned}
P[W_{\tau_{\gamma}} - \gamma > x] &\le P[Y_{N(\gamma)} - \gamma > x] \\
&= \sum_{n=1}^{\infty} P[Y_n > \gamma + x, N(\gamma) = n] \\
&= \sum_{n=1}^{\infty} P[Y_n > \gamma + x, \max_{1\le k\le n-1} Y_k \le \gamma] \\
&= P[Y > \gamma + x]\sum_{n=1}^{\infty}(P[Y \le \gamma])^{n-1} \\
&= \frac{P[Y > \gamma + x]}{P[Y > \gamma]}. \quad (25)
\end{aligned}
$$

Because Y is subexponential $P[Y > \gamma + x]/P[Y > \gamma] \to 1$ as $Y \to \infty$. Also taking expectations in above inequality $E[\Gamma] \le E[Y - \gamma | Y > \gamma]$.

(b) From (24), if Z is of regular variation with index $-\alpha$, $\alpha > 1$ then $M(\tau)$ is of regular variation with index $-\alpha$. Also then $E[Z] < \infty$ and $E[M(\tau)] < \infty$. Therefore, $E[Y^{\alpha-\varepsilon}] < \infty$ for any $\varepsilon > 0$. Hence, from Gut [12], Chapter 1, Th.2.3, $(\Gamma(\gamma))\gamma^{\frac{-1}{(\alpha-\varepsilon)}} \le Y_{N(\gamma)}\gamma^{\frac{-1}{(\alpha-\varepsilon)}} \to 0$ a.s. because $N(\gamma) \to \infty$ a.s. as $\gamma \to \infty$ and $N(\gamma)/\gamma \to 1/E[\gamma]$ a.s. Also since $\{N(\gamma)/\gamma\}$ is uniformly integrable (Gut [12], P.54), we get from Gut [12], Chapter 1, Thm.7.2, $\lim_{\gamma\to\infty} \frac{E[\Gamma(\gamma)]}{\gamma^{1/(\alpha-\varepsilon)}} = 0$.

(c) From [2], Chapter 7, $P[M(\tau) > x] \approx ce^{-u\alpha}$ for $u > 0$ for some $c > 0$. Thus from (25)

$$\begin{aligned} P[\Gamma(\gamma) > x] &= P[W_{\tau_\gamma} - \gamma > x] \\ &\le \frac{P[Y > \gamma + x]}{P[Y > \gamma]} \approx e^{-\alpha x} \text{ as } \gamma \to \infty. \quad (26) \end{aligned}$$

■

Acknowledgements

This work was supported in part by a grant from the Aerospace and Networking Research Consortium and by DRDO sensor network project.

References

[1] S. Asmussen, "Subexponential asymptotics for stochastic processes: extremal behavior, stationary distribution and first passage probabilities," *Ann. Appl. Prob*. vol. 8, no. 2, pp. 354-374, 1998.

[2] S. Asmussen, *Applied Probability and Queues*, 2nd edition. Springer, 2003.

[3] T. Banerjee, V. Kavitha, and V. Sharma, "Energy efficient change detection over a MAC using physical layer fusion," in *Proc. IEEE ICASSP*, Las Vegas, Apr. 2008.

[4] T. Banerjee and V. Sharma, "Generalized analysis of a distributed energy efficient algorithm for change detection," *12th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM09)*, Spain, Oct. 2009.

[5] H. M. Barakat and Y. H. Abdul Kader, "Computing the moments of order statistics from nonidentical random variables," *Statistical Methods and Applications*, vol. 13. Springer, 2003.

[6] P. Billingsley, *Weak Convergence of Measures: Applications in Probability*. John Wiley and Sons, Inc., 1987.

[7] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors—part II: advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64-79, 1997.

[8] O. J. Boxma and J. W. Cohen, *The Single Server Queue: Heavy Tails and Heavy Traffic in self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger (editors.) Wiley, 2000.

[9] B. E. Brodsky and B. S. Darkhovsky, *Nonparametric Methods in Change Point Problems*. Kluwer Academic Publishers, 1993.

[10] R Doney, R. Maller, and M. Savov, "Renewal theorems and stability for the reflected process," *Stochastic Processes and Their Applications*, vol. 119, pp. 1290–1297, 2009.

[11] C. M. Goldie and C. Kliippelberg, "Subexponential distributions," in *Practical Guide to Heavy Tails*, R. Adler, R. Feldman and M. S. Taqqu, editors. Birkhauser, 1998.

[12] A. Gut, *Stopped Random Walks: Limit Theorems and Applications*. Springer, 1988.

[13] J. M. Harrison, *Brownian Motion and Stochastic Flow Systems*. John Wiley and Sons, 1985.

[14] J. Keilson, *Markov Chain Models—Rarity and Exponentiality*. Springer, 1979.

[15] T. L. Lai, "Sequential change point detection in quality control and dynamical systems," *J. Royal Statistical Society, Series B (Methodological)*, vol. 57, no. 4, pp. 613-658, 1995.

[16] K. B. Lataief and W. Zhang, *Cooperative Spectrum Sensing in Cognitive Wireless Communication Networks*, E. Hossain and V. K. Bhargava, editors. Springer, 2007.

[17] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 41, pp. 1897–1908, 1971.

[18] A. Luceno and J. Puig-Pey, "Evaluation of the run-length probability distribution for CUSUM charts: assessing chart performance," *Technometrics*, vol. 42, no. 4, pp. 411-416, 2000.

[19] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks, *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754-1763, May 2007.

[20] Y. Mei, "Information bounds and quickest change detection in decentralized decision systems," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2669-2681, July 2005

[21] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Process.*, vol. 54, no. 2, Feb. 2006.

[22] G. V. Moustakides, "Optimal stopping times for detecting changes in distribution," *Ann. Statist*., vol. 14, pp. 1379–1387, 1986.

[23] E. S. Page, "Continuous inspection schemes," *Biometrica*, vol. 41, no. 1/2, pp. 100-115, June 1954.

[24] H. Rootzen, "Maxima and exceedances of stationary Markov chains," *Adv. in Appl. Prob.*, vol. 20, no. 2, pp. 371-390, June 1998.

[25] S. Ross, *Stochastic Processes*, 2nd edition. Wiley, 1996.

[26] T. L. Saaty, *Nonlinear Integral Equations*. Dover Publications, 1981.

[27] J. Sadowsky, "Large deviation theory and efficient simulation of excessive backlogs in a GI/G/m Queue," *IEEE Trans. Auto. Cont.*, vol. 36, no. 12, pp. 1383-1394, 1991.

[28] V. Sharma and A. K. Jaya Prakasam, "An efficient algorithm for cooperative spectrum sensing in cognitive radio networks," in *Proc. National Conf. on Comm. (NCC)*, Jan. 2009, Guwahati, India.

[29] A. N. Shiryaev, "On optimal methods in quickest detection problems," *Theory Probab. Appl.*, vol. 8, pp. 22-46, 1963.

[30] W. Su, *Time-Synchronization Challenges and Techniques in Wireless Sensor Networks and Applications*, Y. Li, M. T. Thai, and W. Wu, editors. Springer, 2008.

[31] R. Tantra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, pp. 4-17, Feb. 2008.

[32] L. Takacs, "On a generalization of the ARC-SINE law," *Ann. Appl. Probab.*, vol. 6, no. 3, pp. 1035-1040, 1996.

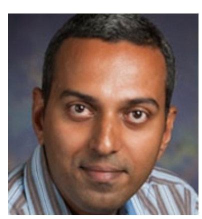
[33] A. Tartakovsky and V. V. Veeravalli, "Quickest change detection in distributed sensor systems," in *Proc. 6th Int. Sym. on Inf. Fusion*, Australia, July 2003, pp. 756-763.

[34] A. G. Tartakovsky and H. Kim, "Performance of certain decentralized distributed change detection procedures," in *Proc. Int. Sym. on Inf. Fusion*.

[35] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1657-1665, May 2001.

[36] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors–part I: fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54-63, 1997.

[37] L. Zacharias and R. Sundaresan, "Decentralized sequential change detection using physical layer fusion," in *Proc. ISIT*, France, June 2007.



Taposh Banerjee received the B.E. degree in electronics engineering from the Nagpur University, Nagpur, India in 2000 and the M.E. degree in telecommunication from the Indian Institute of Science, Bangalore, India in 2003. He is currently pursuing his Ph.D. from the University of Illinois at Urbana-Champaign. He has worked for three years with Samsung India Software Operations, Bangalore from 2003-2005 and worked as a Project Associate in the ECE department of the Indian Institute of Science, Bangalore from 2007-2009. His research interests include detection theory, applied probability and communication systems.

Vinod Sharma completed his B Tech in 1978 from IIT Delhi and Ph.D. in 1984 from CMU in Pittsburgh in Electrical Engg. He worked at Northeastern Univ., Boston and UCLA before joining Indian Institute of Science in 1988. Currently he is a Professor and Chairman of Electrical Communication Engg Dept. He is on the editorial boards of *International Journal of Information and Coding Theory* (Inderscience), *Journal of Electrical and Computer Engineering* (Hindawai), and *International Journal of Advanced Computer Engineeringg* (Serials Publications). Prof Sharma's research interests are in Wireless communication, Information Theory and Communication Networks.

Veeraruna Kavitha received her B.E. degree in Electronics from UVCE, Bangalore in 1994 and the M.Sc (Engg) and Ph.D. degree from Dept. of ECE, Indian Institute of Science (IISc), Bangalore in 2002 and 2007 respectively. From 1994-2000, She was involved in the design and development of GPS, CDMA and Voice band modems at Accord S/W and Systems, Bangalore. She was a NBHM (National Board for Higher Mathematics) scholar at Tata Insititue of Fundamental Research (TIFR), Bangalore, during 2007-08. From 2008 onwards, She has been a post doctoral researcher with MAESTRO, INRIA, Sophia Antipolis, France and LIA, University of Avignon, France. Her research interests span communication theory, wireless networks, signal processing, game theory and optimization.

Arunkumar Jayaprakasam received his B.E. degree from Regional Engineering College, Tiruchirapalli in 2002. He recieved his M.Sc (Engg) from Indian Institute of Science in 2010. He was with Texas Instruments (India) Pvt. Ltd., Bangalore from 2002 - 2008 where he was involved in ADSL and GSM/GPRS/EDGE modem development. He is currently with Nokia (India) Pvt. Ltd. His research interests include Wireless Communication and Communication Networks.