

Received April 2, 2022, accepted April 19, 2022, date of publication April 25, 2022, date of current version May 2, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3170463

Change Detection for Large Distributed Sensor Networks With Multitriggered Local Sensors

MATTHEW W. QUARISA^{ID}, (Member, IEEE),
AND STEVEN D. BLOSTEIN^{ID}, (Senior Member, IEEE)

Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada

Corresponding author: Steven D. Blostein (steven.blostein@queensu.ca)

This work was supported by the Natural Research Council of Canada Discovery under Grant RGPIN-2019-06237.

ABSTRACT The emergence of large sensor networks and the Internet of Things has reinvigorated interest into distributed quickest change detection. Important shortcomings of existing approaches are ease of design, flexibility in communication, and applicability to larger networks. The new approach proposed in this work features local sensors that can be triggered multiple times, i.e., can reset and continue monitoring after transmitting their decisions. With larger sensor networks as a focus, the system allows for multiple simultaneous transmissions to a fusion center within bandwidth limitations. The proposed system uses the cumulative-sum procedure at local sensors to binarize local decisions, which are then transmitted to the fusion center that also employs cumulative-sum quickest detection. Test overshoot due to sequential test overshoot is avoided, and global and local thresholds are chosen to meet a desired false alarm rate constraint using numerical computation of expected delay performance. The system compares favourably to several existing methods while offering greater flexibility in the amount of fusion center communication.

INDEX TERMS Sequential change detection, distributed detection, sequential probability ratio test, CUSUM, sensor networks, decision fusion.

I. INTRODUCTION

Continuous advancement in sensor technology and wireless networking has made it feasible to widely deploy distributed decision-making systems [1] allowing for decisions to be made from vast amounts of data with increased efficiency, robustness and accuracy. The Internet of Things (IoT) is an embodiment of this trend that seeks to establish itself in everyday life and is impacting industries such as manufacturing, healthcare, transportation, retail and utilities [2], [3]. A critical consideration in the design of IoT systems leveraging large sensor networks lies in achieving efficient decision fusion [4], [5] in terms of decision performance relative to communication cost.

A. MOTIVATION

In order to realize the potential brought about by distributed decision making systems, renewed attention has been directed towards the application of sequential detection [6], whereby at each time slot, a decision making system either stops and declares a decision on the process observed, or continues to observe measurements. The two

main types of sequential detection problems are hypothesis testing, addressed by the sequential probability ratio test (SPRT) [7], and change detection, addressed by the cumulative sum (CUSUM) algorithm, which can be expressed as a repeated SPRT [8]–[11]. Problems categorized as sequential hypothesis testing seek to distinguish the correct hypotheses consistent with a set of observed measurements. In these types of problems, the objective is to take as few observations as possible and thereby minimize delay while meeting an acceptable prespecified probability of misclassification.

Sequential change detection problems, on the other hand, seek to identify a change of events from a set of measurements. In change detection problems, measurements are typically assumed to be initially generated according to a specified probability distribution. At an unknown time, a change in probability distribution abruptly occurs. Similar to sequential hypothesis testing, tests are designed to determine as quickly as possible when a change occurs are subject to meet some acceptable error rate, i.e., the goal of such tests are to minimize decision delay within a given false alarm constraint.

Specific use-cases of sequential detection procedures within distributed, decentralized systems are in cognitive radio including jammer detection and localization [12],

The associate editor coordinating the review of this manuscript and approving it for publication was Mohamad Afendee Mohamed^{ID}.

monitoring and data collection systems for critical infrastructure such as power grids, oil and gas networks, roadways [6], malicious user detection for mobility tracking in autonomous driving [13], as well as optimized sensor networks under communication constraints for control applications [5]. In these applications, large geographical areas are monitored, thereby requiring data from a large number of sensors at different locations. Typically, due to bandwidth restrictions, cost or reliability, it is expensive to have each sensor forward its observations to the fusion center (FC) to make a decision requiring the decision making process to be decentralized. [14]. Besides outdoor uses, there is significant potential for wireless IoT networks in a healthcare setting [15], [16]. Distributed sensing can facilitate remote health monitoring, fitness programs, chronic disease management and elderly care. Various medical and non-medical devices, such as sensors installed within the home and imaging are being proposed to be used as smart devices to provide clinicians with valuable information that assist in making insightful decisions about a patient's care. The integration of IoT within healthcare is expected to reduce medical costs, increase quality of life and enrich the user's experience [17].

As evident from the use-cases discussed above, future IoT systems are evolving to employ many sensors. Current approaches to sequential detection in distributed systems do not scale well for systems with many sensors because they neglect test threshold overshoot and exhibit an increase in design complexity. In addition, in certain applications, local sensor data may be continuously received after a local decision is made. This motivates investigating new distributed sequential detection procedures that overcome these limitations.

B. RELATED WORK

Early research first treated local sensors as physical extensions of the FC, enabling detection decisions to be centralized. Extensions to monitoring the quickest detection of changes appear in [18] and [19]. Specific solutions have also been proposed e.g., two sensors [20], or based on nonparametric statistical models [21]. Alternatively, [22] addresses a decentralized quickest detection scenario where only an unknown subset of local sensors experience the change in distribution. The motivation behind distributing decision-making to local sensors is, in large part, due to bandwidth constraints, as wireless spectrum is a limited resource, as well as energy constraints [23]. An extreme example is the fully decentralized ad hoc IoT network without a FC which was proposed for sequential change detection in [24], [25].

Bayesian formulations to optimize the distributed quickest change detection problem have been investigated for certain cases. Examples include the ad hoc network in [24], as well as [26] where communication is highly restricted wherein only a single report can be received at the FC in each time slot. Additional reports are ignored by the FC discarding valuable information. The approach in [19] assumes a

known change-time distribution as well as full feedback from the FC to local sensors, which is resource demanding. Bayesian quickest detection also has practical issues that limit applicability. First, required a priori probability of change as well as a priori change-time distribution are often not available. In addition, the dynamic programming solution to Bayesian formulations has exponential complexity as well as very high sensitivity to the assumed a priori probability of change.

Alternatively, minimax quickest change detection criteria have led to the CUSUM algorithm [8]–[11] with optimality for conventional sensing, although globally optimal minimax solutions have eluded distributed networks thus far. Several minimax approaches have been shown to possess desirable asymptotic properties. In [27], a decentralized solution uses CUSUM asynchronously at local sensors. A one-shot scheme that informs the FC of a change after the first local sensor decision is made, also known as minimal combining, is shown to perform well asymptotically as the mean time between false alarms tends to infinity. However, its performance advantage is less clear at specific false alarm rates, and explicit design procedures to satisfy maximum system error rates were not provided. Recently, Biswas et. al. consider a non-Bayesian quickest change detection approach that addresses quantization in transmission as well as energy harvesting [28].

In [29], a simultaneous change in distribution across all local sensors is monitored. Two one-shot procedures that utilize CUSUM at local sensors are proposed and analyzed. In the first, each local sensor may communicate with the FC only once with no test reinitialization. The FC concludes a change has occurred as soon as m local CUSUM messages out of a total of K local sensors are received by the FC. The second procedure considered makes a change decision when m out of K local sensor CUSUMs agree simultaneously that a change has occurred [29]. For each case, non-asymptotic lower bounds on expected time to false alarm, and asymptotic upper bounds to worst case detection delay are obtained. Its limitations are (1) that it lacks a design method to achieve a given false alarm rate, (2) local decisions are done on a one-shot basis, and (3) its performance has not been investigated for more than 4 sensors.

In the systems considered in [30] and [31] local sensors process observations using local log likelihood ratio statistics, similar to the SPRT, with a set of predefined constant thresholds. A single bit is transmitted to the FC when the local statistic crosses above or below the set of thresholds, which is termed *level triggering*. The FC uses a SPRT on the asynchronously arriving bits. Through an asymptotic performance analysis, the performance in using additional bits to encode the sequential test overshoot at local sensors is assessed. A threshold design procedure to achieve a given error performance was not presented. Instead, offline simulation is employed.

In each of [26], [27], [29]–[31], the scalability to networks with large numbers of sensors is challenging. With the

exception of [30], the occurrence of sequential test overshoot at local sensors is ignored. In networks with many local sensors, the accumulated impact of overshoot can be significant. System designs [32] and [33] account for local sensor overshoot but address the distributed hypothesis testing problem rather than quickest change detection. In [32], overshoot is encoded into a time delay between a local sensor's decision time and its report transmission time. This time delay is then decoded by the FC to determine the amount of overshoot. While accounting for overshoot, this scheme adds delay to the system. In [33] quantization strategies used by local sensors to summarize their observations are examined. Such strategies may have potential for capturing overshoot in summary messages, but complicated quantization schemes incur extra overhead and complexity for low-power sensors. The issue of how to efficiently scale up these quantization strategies for large sensor networks was not addressed.

The more recent proposed approaches are categorized in summary form in Table 1. Not all address quickest change detection as opposed to sequential hypothesis testing [30], [31] or decision fusion [4]. These previous schemes have local sensors that are one-shot, i.e., shut off after they make a decision and ignore future local observations. The system in [24] is distinct by not having a fusion center. Both [26] and [24] have the drawbacks of requiring assumed a priori change time statistics. None of the networks for quickest change detection permit simultaneous reports from multiple local sensors. Most previous approaches are motivated by asymptotic properties rather than explicit design procedures with finite sample analysis to determine suitable sensor thresholds. Given a maximum error constraint, test thresholds that closely meet this constraint are desired. Over- and under-designed tests are especially problematic in large networks, where impacts on performance accumulate. Finally, none of the approaches listed in Table 1 are tailored specifically to bandlimited fusion center based networks with large numbers of local sensors.

C. CONTRIBUTIONS AND ORGANIZATION

This paper fills a gap in the literature by addressing the design of minimax quickest change detection based on independent observations for distributed systems with large numbers of sensors that can each be reset multiple times and transmit multiple reports to a FC. By extending direct numerical performance analysis methods, joint local sensor and FC test thresholds can be designed to closely meet overall specifications. A key element of the direct computational approach is the incorporation of the effect of overshoot that occurs for resetting CUSUM tests on observations at local sensors as well as overshoot for CUSUM tests on reports at the FC. In addition, direct performance analysis is applicable to arbitrary discrete probability density functions to accommodate desired quantization schemes besides the binary one proposed here. The main contributions are summarized as follows:

TABLE 1. Property summary of recent related approaches: Yes (Y) /No (N).

Property	Wang (GC15) [26]	Yilmaz (SP13) [31]	Banerjee (ISIT16) [29]	Li (IoT18) [24]	Al-Jarrah (IoT20) [4]
quickest ch. detector	Y	N	Y	Y	N
a priori change time pdf reqd	Y	N	N	Y	N
local+FC architecture	Y	Y	Y	N	Y
multiple FC reports	N	N	Y	N	Y
finite sample analysis	N	N	N	N	Y
test design procedure	Y	Y	N	N	Y
large network scalability	N	N	Y	Y	N

- 1) By generalizing a recursive method proposed for Gaussian-based SPRT design in [34] and [35], numerical techniques are presented to analyze the performance of SPRT, CUSUM and repeating-CUSUM procedures that account for test overshoot, greatly improving accuracy over Wald's approximations. These are applied in tandem to continuous distributions at local sensors which then feed sensor reports to the FC that obey discrete distributions.
- 2) The proposed distributed system is based on per-node optimal minimax quickest change detection that allows for sensors that reset and continue monitoring after transmitting their local decisions within an overall bandwidth restriction at the FC.
- 3) The false alarm probability and average delay performance analysis is applied to test threshold design to create a methodology that trades off delay and bandwidth while closely meeting a specified overall false alarm rate constraint.

The rest of the paper is organized as follows: Section II presents the system model and formulates the minimax distributed quickest change detection problem. Section III presents numerical techniques used to analyze the performance of the SPRT, CUSUM and distributed systems. Sections IV and V present a distributed, decentralized system based on a minimax (CUSUM) quickest change detection framework with the restriction of FC communication bandwidth. Section VI presents a threshold design methodology for the proposed system. In Section VII, the accuracy of the design method, performance tradeoffs, and comparisons with existing systems are quantified.

D. NOTATION USED

By default, capital letters, X , denote random variables and associated probability density and cumulative distribution functions are described by $f(x)$ and $F(x)$, respectively, whether discrete or continuous, where x is a realization of X . Subscripted random variables by lower case k denote discrete time. On occasion probability distributions under

hypothesis H_0 and H_1 are referred to as Q_0 and Q_1 , respectively. Observations at local sensors are typically continuous-valued whereas reports observed at the fusion center are discrete-valued, e.g., r_k is a realization of the number of reports observed by the fusion center at time k . The expressions $P_\theta[A]$ and $E_\theta[A]$ denote conditional probability and expectation of event A , given parameter θ . Superscripts are used only when absolutely necessary to distinguish time and sensor: X_k^i refers to a random variable observed at sensor i at time k . Otherwise sensor indices are dropped. All other specific functions are defined where used.

II. SYSTEM MODEL

A distributed system of M sensors is considered. The i -th sensor observes a sequence $X_1^i, X_2^i, \dots, i \in \{1, 2, \dots, M\}$. It is assumed that observations between sensors are mutually independent conditioned on hypothesis H_0 or H_1 . This would occur, for example, when sensors are physically well separated in space. Observations at the same sensor are also assumed to be independent and identically distributed (IID) before and after a single abrupt change from H_0 to H_1 , i.e., that they are monitoring a statistically homogeneous environment under H_0 . At some point in time, the observations at all sensors abruptly and simultaneously change from H_0 to H_1 .

This change can be described as

$$H_0: X_1^i, X_2^i, \dots, X_{\Gamma-1}^i, \sim Q_0 \quad (1)$$

$$H_1: X_\Gamma^i, X_{\Gamma+1}^i, \dots, X_k^i, \sim Q_1 \quad (2)$$

where Γ represents an unknown deterministic change time, and Q_0 and Q_1 are the probability distributions before and after the change. A CUSUM test, with known optimality properties [10], is performed at local sensors. In the following, when referring to an arbitrary sensor, superscript notation is omitted.

When a local sensor decides that a change in distribution occurs, it transmits a binary-valued report indicator to the FC, which, upon collecting information from local sensors, acts as the final global decision maker. Over time, the FC thus observes a sequence of numbers of reports, $r_k \in \{0, 1, 2, \dots, C\}$, where a value of $r_k = j$ means that j sensors have made a decision and sent reports to the FC in time slot k . If a report is not received, the value of r_k is zero. To reflect network bandwidth limitations, the number of reports received at a given time slot is limited to a value no larger than C , the number of available channels, where $C \leq M$. The time sequence of reports r_k , at times $k = 0, 1, 2, \dots$, updates a cumulative statistic at the FC.

Let the FC's expected decision delay and false alarm rate be expressed, respectively, as

$$E_{DD}^{FC} = E[(\tau_F - \Gamma)^+], \quad (3)$$

$$R_{FA}^{FC} = \frac{1}{\text{ARL-to-FA}}, \quad (4)$$

where τ_F is the FC's decision time, Γ is the actual change time, $(x)^+ = \max(x, 0)$, and ARL-to-FA is the average run

length to false alarm. The FC's decision policy, δ , aims to minimize expected decision delay subject to a false alarm constraint:

$$\min_{\delta} E_{DD}^{FC} \quad (5)$$

$$\text{subject to } R_{FA}^{FC} \leq \alpha_{FC}, \quad (6)$$

where α_{FC} is the desired maximum rate of false alarms. A network architecture is assumed where local sensors only have one-way error-free communication with the fusion center and no communication with one another.

III. NUMERICAL COMPUTATION OF FINITE-SAMPLE SPRT AND CUSUM PERFORMANCE

In order to analyze the delay and false alarm performance of CUSUM tests used in the distributed sensing system as a function of time, core recursive performance computations are first introduced for the SPRT, which satisfy a Fredholm equation [35], and are generalizations of expressions found in [34], [36]. Unlike the traditional Wald approximations [37], these expressions account for overshoot in the design of test thresholds. These expressions are then extended to analyze CUSUM tests.

A. SPRT DIRECT ANALYSIS

Consider the sequential probability ratio test (SPRT) for binary hypothesis testing for IID observations proposed by Wald [38] with well-known optimality properties [7]. This test tracks the likelihood ratio of observations X_i observed sequentially at times $i = 1, 2, \dots$ by the accumulating likelihood statistic at time k ,

$$G_k = \sum_{i=1}^k L_i \quad (7)$$

where per-sample log-likelihood ratio $L_i \equiv \ln\left(\frac{q_1(X_i)}{q_0(X_i)}\right)$, for probability density functions (or probability mass functions if discrete), $q_0 \sim Q_0$ and $q_1 \sim Q_1$. This test, a realization, g_k , of G_k is compared with two precomputed thresholds a and b after each observation. The SPRT decision rule is as follows [37]:

$$\phi_{SPRT}(g_k) = \begin{cases} \text{Decide } H_1 & \text{if } g_k \geq b \\ \text{Decide } H_0 & \text{if } g_k \leq a \\ k \leftarrow k + 1 & \text{else,} \end{cases} \quad (8)$$

and continuing to the next stage indefinitely until a decision is reached.

The objective is to determine the probabilities of error of (8) as well as the expected number of samples required to reach a decision. The notation $f(\cdot)$ refers to the common probability density function (PDF) (or probability mass function (PMF) if discrete), assuming it exists, of the L_i , which form a statistically independent sequence in time.

Without loss of generality, in the following it is assumed that H_1 is true. If H_0 is true, the derivation is similar. Denote $f_k(x)$ as the conditional PDF (or PMF) of SPRT statistic (8)

in stage k given that the test reaches stage k . Noting that the test reaches stage one with probability one, the probability of reaching stage k can be expressed in terms of the conditional probabilities

$$d_k(H_1) = \prod_{j=2}^k \text{P}(\text{test reaches stage } j \mid \text{test reaches stage } j-1), \quad (9)$$

where $k \geq 2$. From (8) and the definition of $f_k(x)$,

$$\begin{aligned} &\text{P}(\text{test reaches stage } k+1 \mid \text{test reaches stage } k) \\ &= \int_a^b f_k(x) \mu(dx), \end{aligned} \quad (10)$$

where measure notation $\mu(\cdot)$ may refer to Riemann integration or summation, depending on whether $f_k(x)$ is continuous or discrete,

$$f_k(x) = \frac{(f \star f_{k-1}^{trunc})(x)}{\int_{-\infty}^{\infty} (f \star f_{k-1}^{trunc})(t) \mu(dt)} \quad (11)$$

and

$$f_{k-1}^{trunc}(x) = \begin{cases} f_{k-1}(x) & a < x < b \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

where \star denotes convolution and the initial term of Eq. (11), $f_1(x)$, is the PDF of a single log-likelihood ratio observation. This follows from that fact that the PDF of a sum of independent random variables is a convolution. Eq. (11) performs the normalization transformation required so that $f_k(x)$ is a PDF after truncation. In summary, the probability of reaching each SPRT stage is obtained by successive convolutions using numerical integration, and integrating the result over the region between the SPRT thresholds.

The probability of choosing H_1 at stage k given that stage k is reached is computed as follows:

$$\begin{aligned} \gamma_k(H_1) &\equiv \text{P}(\text{choose } H_1 \text{ at stage } k \mid \text{test reaches stage } k) \\ &= \int_b^{\infty} f_k(x) \mu(dx) \\ &= 1 - \int_{-\infty}^b f_k(x) \mu(dx) \\ &= 1 - \int_{-\infty}^b \int_a^b f(x-w) f_{k-1}(w) \mu(dw) \mu(dx) \\ &= 1 - \int_a^b \int_{-\infty}^b f(x-w) f_{k-1}(w) \mu(dx) \mu(dw) \\ &= 1 - \int_a^b F(b-w) f_{k-1}(w) \mu(dw) \end{aligned} \quad (13)$$

using (11), where $F(\cdot)$ denotes the cumulative distribution function of $f(\cdot)$. Using (9) and (13), the probability of choosing H_1 at any stage is then

$$\text{P}(\text{choose } H_1) = \sum_{k=1}^{\infty} \gamma_k(H_1) d_k(H_1). \quad (14)$$

The average test length of an SPRT given H_1 , $E_1[T]$, can be straightforwardly shown to be

$$E_1[T] = \sum_{k=1}^{\infty} d_k(H_1). \quad (15)$$

When (14) is truncated to a finite number of stages, N , a lower bound for $\text{P}(\text{choose } H_1)$ is obtained. An upper bound can be obtained by adding the probability of reaching stage $N+1$ to this

$$\begin{aligned} &\sum_{k=1}^N \gamma_k(H_1) d_k(H_1) \\ &< \text{P}(\text{choose } H_1) \\ &< \sum_{k=1}^N \gamma_k(H_1) d_k(H_1) + d_{N+1}(H_1). \end{aligned} \quad (16)$$

These bounds can be used to determine at which stage it is appropriate to terminate execution of the recursive procedure. A value of N is found that, if exceeded, will insignificantly affect the computed delay and error probabilities. In numerical computations, a value of N is selected such that continuing on to stage $N+1$ would change the computed error probability by less than three orders of magnitude. The magnitude difference between delay stages is defined as ϵ and this method of termination is used in Algorithm 1.

B. CUSUM DIRECT ANALYSIS

The above analysis procedure can be extended to analyze the CUSUM test for change detection that tracks statistic S_k , the maximum of log-likelihood ratio tests for all possible change times up to time k ,

$$S_k = \max_{1 \leq j \leq k} \sum_{i=j}^k L_i \quad (17)$$

starting at $S_0 = 0$ until it exceeds CUSUM log threshold h . Page [8] showed that S_k can be expressed as a repeating sequence of SPRTs with boundaries 0 and h . Each time the SPRT exceeds the lower boundary, it is reset to zero. This can be seen by rewriting (17) as

$$S_k = (S_{k-1} + L_k)^+ \text{ for } k \geq 1. \quad (18)$$

Two different types of methods are proposed for analysis:

1) ONE-SIDED SPRT CUSUM ANALYSIS

A CUSUM test has lower and upper log-thresholds 0 and h , respectively, where detection threshold h is set by the system designer. The CUSUM test can be analyzed using a similar procedure as was shown for the SPRT in Section III-A. As in the case of the SPRT, the PDF of the CUSUM statistic in stage k given that it reaches stage k can be computed using the fact that the PDF of each log-likelihood ratio observation is IID.

A CUSUM test will only terminate if the statistic S_k exceeds threshold h . The expression in (10) becomes

$$\text{P}(\text{test reaches stage } k+1 \mid \text{test reaches stage } k)$$

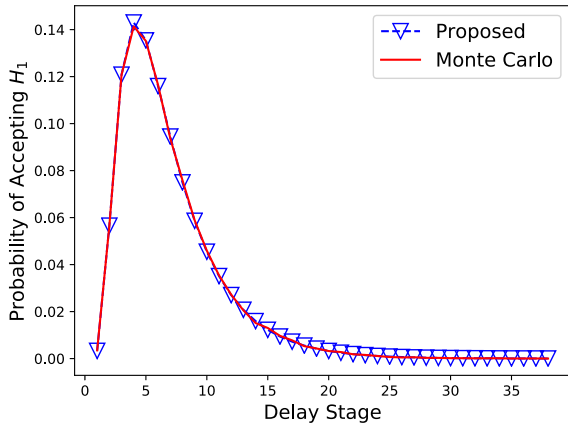


FIGURE 1. Probability of accepting H_1 in a $N(0, 1)$ vs $N(1, 1)$ test with H_1 true and $h = \ln(25)$. 10^6 Monte Carlo trials are distributed over 38 delay stages to compute the probabilities.

$$= \int_{-\infty}^h f_k(x)\mu(dx) \quad (19)$$

where $f_k(x)$ is given by the convolution in Eq. (11) with

$$f_{k-1}^{trunc}(x) = \begin{cases} f_{k-1}(x) & 0 \leq x < h \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

When a CUSUM test statistic at time, or *test stage* k , S_k , falls below the lower threshold 0, the test reinitializes itself by setting $S_k = 0$ and continuing operation. At CUSUM time (or *stage*) k , the probability that the test resets,

$$P(\text{reset at stage } k) = \int_{-\infty}^{0^-} f_k(x)\mu(dx) \quad (21)$$

is added to $f_k^{trunc}(0)$ of the current stage's truncated PDF prior to the next observation. This simple computation accounts for reinitialization in the subsequent CUSUM stage's PDF, which motivates this direct numerical computation approach. The probability of choosing H_1 in stage $k + 1$ given that stage $k + 1$ is reached and the decision delay, or *average run length* given either hypothesis are similarly found using (13) and (15) with this new PDF. The unconditional probability of accepting H_1 at a given stage k is computed as

$$P(\text{choose } H_1 \text{ at stage } k) = \gamma_k(H_1)d_k(H_1). \quad (22)$$

The above CUSUM analysis is illustrated by an example of an IID standard Gaussian shift-in-mean from zero to one, i.e., from $N(0, 1)$ to $N(1, 1)$. The probability of accepting H_1 at a given test stage is computed using the proposed procedure, Eq. (22), and compared to a Monte Carlo simulation of CUSUM in Figs. 1 and 2. Figure 1 shows close agreement between the numerical computations and Monte Carlo simulation over the plotted 38 delay stages when a change to H_1 occurs immediately at delay stage 0.

Figure 2 shows the probability of CUSUM false alarm, i.e., when H_0 is true. Statistical agreement between Monte Carlo and numerical results is observed to within a 90% confidence

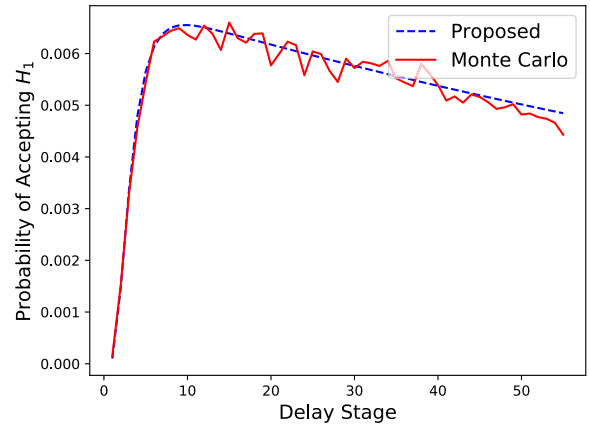


FIGURE 2. Probability of accepting H_1 in a $N(0, 1)$ vs $N(1, 1)$ test with H_0 true and $h = \ln(25)$. 10^6 Monte Carlo trials are used over 55 delay stages in simulating these probabilities.

interval, given by

$$CI_{90\%} = 1.645\sqrt{\frac{P(1-P)}{W}}, \quad (23)$$

where $P \equiv P(\text{choose } H_1 \text{ at stage } k)$ in Eq. (22) and W is the number of Monte Carlo trials [39]. In Figure 2, a total of 10^6 Monte Carlo trials are used over the 55 delay stages, resulting in $W = \frac{10^6}{55}$. A value of $P = 0.005$, for example, corresponds to $CI_{90\%} = 8.6 \times 10^{-4}$ which encompasses the Monte Carlo curve's oscillations.

2) REPEATED SPRT ANALYSIS

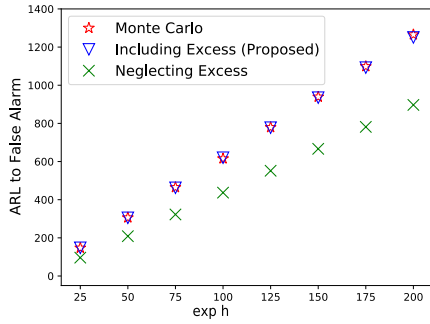
The CUSUM test (17) may be alternatively expressed as a sequence of repeating SPRTs [8]. This suggests a method for estimating a CUSUM's average run length (ARL) to false alarm and its ARL to detection.

Consider a SPRT with initial statistic $G_k = 0$ and lower and upper log-thresholds 0 and h , respectively. Let $P_\theta(H_1)$ represent the probability that an SPRT terminates at upper threshold h when H_θ is true, $\theta \in \{0, 1\}$ and let $E_\theta[T_0]$ and $E_\theta[T_h]$ denote the average numbers of samples for a single SPRT conditional on the test ending on 0 and h , respectively, given H_θ . Page [8] showed that the ARL of a CUSUM test given H_θ being the true hypothesis is

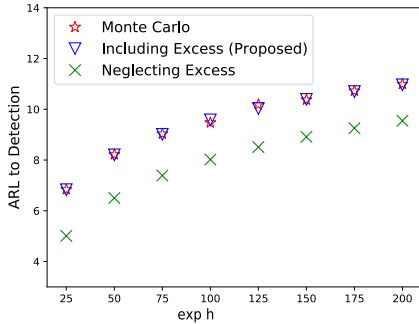
$$\begin{aligned} ARL_\theta &= \frac{1 - P_\theta(H_1)}{P_\theta(H_1)} E_\theta[T_0] + E_\theta[T_h] \\ &= \frac{E_\theta[T]}{P_\theta(H_1)}, \theta \in \{0, 1\}. \end{aligned} \quad (24)$$

In (24), $P_\theta(H_1)$, $\theta = 0, 1$ represent the probability of false alarm and correct detection of a SPRT, respectively. Eq. (24) provides the ARL to false alarm and detection, respectively. The numerator of (24) is computed using (15) and the denominator is computed using (14).

In Figures 3 - 6, the analytical ARL to false alarm and detection of CUSUM tests are shown, respectively, for varying detection thresholds and compared to Monte Carlo simulations of CUSUM for Gaussian shift in mean, Gaussian



(a) ARL to FA



(b) Expected Detection Delay

FIGURE 3. Comparison of proposed ARL between proposed method, Monte Carlo simulation, and (25), (26) for a shift in mean $N(0,1)$ vs $N(1,1)$ CUSUM test.

shift in variance, Bernoulli, and the discrete PMF described in the caption of Fig. 6. In each case, the ARL values from the proposed numerical computations show much closer agreement to those obtained from Monte Carlo simulation than those resulting from the use of [40]

$$E_0[T] \approx \frac{|e^h - h - 1|}{|\mu_0|} \quad (25)$$

$$E_1[T] \approx \frac{|e^h + h - 1|}{|\mu_1|} \quad (26)$$

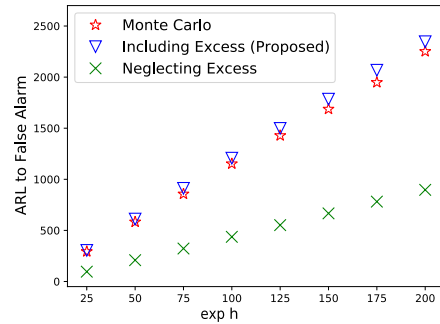
where $\mu_\theta = E_\theta[L_1]$ for $\theta = 0, 1$ which are derived from the well-known Wald approximations that neglect overshoot and undershoot from test threshold boundaries [37].

It is worth noting that *staircase* discontinuities observed in Figures 5 and 6 are due to discrete distributions (PMFs), which are encountered at the fusion center described later.

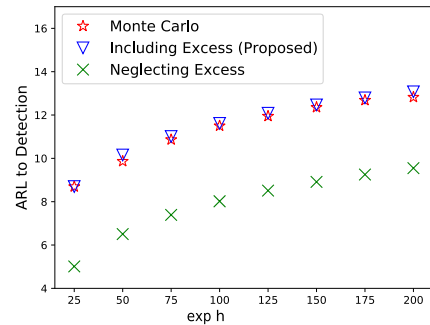
C. APPLICATION TO DISTRIBUTED SYSTEMS

The CUSUM analysis of Section III-B can be applied to the decentralized systems described in Section II. In such systems, local sensor detection procedures offload FC processing by only transmitting detected changes to the FC. The FC instead receives indicator reports from the M sensors and makes a final decision on the state of the system.

When a local sensor sends a report, rather than shutting off, the local sensor is assumed to reinitialize itself and continue processing observations since they are still arriving. The previous CUSUM analysis can be easily extended to

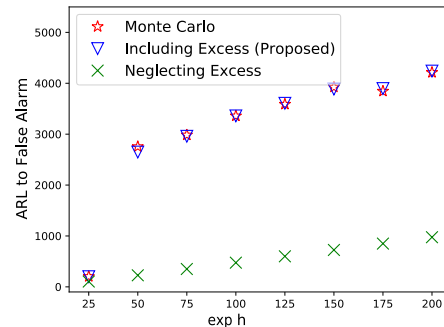


(a) ARL to FA

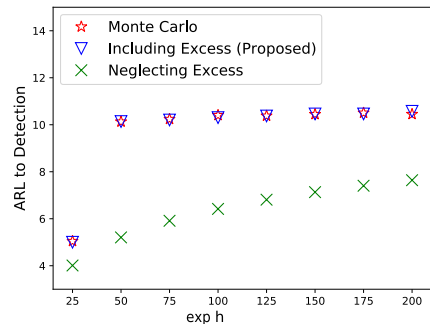


(b) Expected Detection Delay

FIGURE 4. Comparison of ARL between proposed method, Monte Carlo simulation and (25), (26) for a shift in variance $N(0,1)$ vs $N(0,3)$ CUSUM test.



(a) ARL to FA



(b) Expected Detection Delay

FIGURE 5. A comparison of ARL between proposed method, Monte Carlo, and (25), (26) for a Bernoulli $p_1=0.005$ vs $p_0=0.2$ CUSUM test.

analyze local sensor CUSUM procedures that reinitialize after reporting. The focus here is on the case of reinitialization

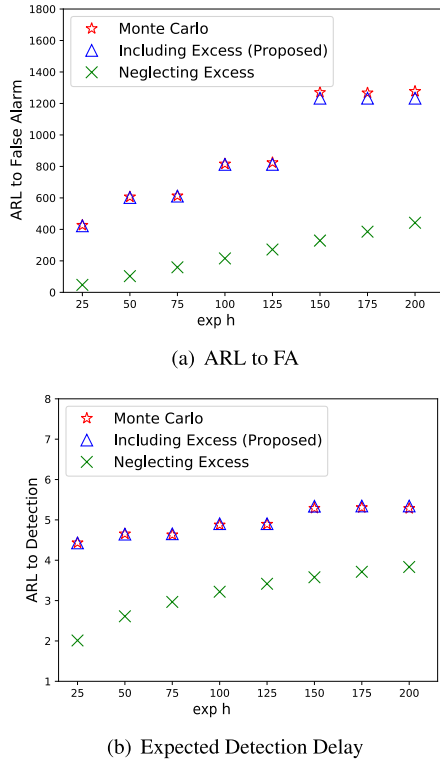


FIGURE 6. A comparison of ARL between proposed method, Monte Carlo, and (25), (26) for discrete pmfs $p_{4|H_1}(x) = 0.587\delta(x) + 0.325\delta(x - 1) + 0.078\delta(x - 2) + 0.011\delta(x - 3)$ vs. $p_{4|H_0}(x) = 0.977\delta(x) + 0.023\delta(x - 1) + 2.215 \times 10^{-4}\delta(x - 2) + 1.210 \times 10^{-6}\delta(x - 3)$.

to zero, but a similar procedure follows for any reset strategy. To extend the one-sided SPRT CUSUM analysis of the preceding section, the reinitialization probability

$$P(\text{reinitialize to zero at stage } k - 1) = \int_h^\infty f_{k-1}(x)\mu(dx) \tag{27}$$

is added to the PDF at $f_{k-1}(0)$ prior to observing the next observation. This allows for the possibility of a local sensor's CUSUM terminating and reinitializing itself to zero to be accounted for in the subsequent stage's PDF. The probability of choosing H_1 in delay stage k given that stage k is reached, $\gamma_k(H_\theta)$, is then computed using Equation (13) for a CUSUM test that after each observation has some probability of transmitting a report and reinitializing itself to zero. The analysis of the repeating local CUSUM is summarized in Algorithm 1.

Algorithm 1 is applied to compute the probability mass function of received reports at the FC as follows: define the term $P(r_{k+1} = n|H_{\theta,k+1})$ $\theta \in \{0, 1\}$ as the probability of receiving n reports simultaneously at the fusion center given hypothesis θ at time slot $k + 1$. Under steady-state conditions, and when $n \ll M$, these probabilities can be approximated as

$$P(r_{k+1} = n|H_{\theta,k+1}) \approx \binom{M}{n} \gamma_{k+1}(H_\theta)^n (1 - \gamma_{k+1}(H_\theta))^{M-n}. \tag{28}$$

Algorithm 1 Repeating CUSUM Analysis With Reinitialization to Zero

```

1:  $k = 1$ 
2: while true:
3:   generate pdf  $f(x)$ 
4:   if  $k = 1$ :
5:      $f_1(x) = f(x)$ 
6:      $P(\text{reach stage } 1) = 1$ 
7:      $P(\text{reach stage } 2) = \int_{-\infty}^h f_1(x)\mu(dx)$ 
8:      $P(\text{accept } H_1 \text{ at stage } 1) = \int_h^\infty f_1(x)\mu(dx)$ 
9:     if  $\frac{P(\text{accept } H_1 \text{ at stage } 1) + P(\text{reach stage } 2)}{P(\text{reach stage } 2)} \geq \epsilon$ :
10:      break
11:   if  $k \geq 2$ :
12:      $P(\text{reset at stage } k-1) = \int_{-\infty}^0 f_{k-1}(x)$ 
13:      $P(\text{reinitialize to zero at stage } k - 1) = \int_h^\infty f_{k-1}(x)\mu(dx)$ 
14:      $f_{k-1}^{trunc} = \text{truncate}(f_{k-1}, 0, h)$ 
15:      $f_{k-1}^{trunc}(0) = f_{k-1}^{trunc}(0) + P(\text{reset at stage } k-1)$ 
16:      $f_k = (f_{k-1}^{trunc} \star f)(x)$ 
17:      $P(\text{reach stage } k + 1) = \int_{-\infty}^h f_k(x)\mu(dx)$ 
18:      $P(\text{accept } H_1 \text{ at stage } k) = \int_h^\infty f_k(x)\mu(dx)$ 
19:     if  $\frac{P(\text{accept } H_1 \text{ at stage } k) + P(\text{reach stage } k + 1)}{P(\text{reach stage } k + 1)} \geq \epsilon$ :
20:      break
21:      $k = k + 1$ 
22: end while
23:  $E_\theta[T] = \text{sum}(P(\text{reach stage } k))$ 

```

The steady-state condition approximation is due to ignoring the small subset of sensors that may have recently reset after having reported H_1 to the FC. The resets introduce time-variation to the FC report sequence, $r_k, k = 1, 2, \dots$ violating the IID assumption implicit in (28). In the scenarios of interest where numbers of sensors, M , are large and false alarm rates are low, the number of simultaneous reports, $n \ll M$, and the low proportion of reset sensors therefore has a negligible effect on the overall reporting probabilities. This will be investigated later on.

From (28), the probability that the number of received reports exceeds a specified bandwidth constraint of $C \in [1, M]$ sensor-to-FC communication channels can be estimated.

To illustrate the above, Figures 7(a)-(c) use (28) to plot the probabilities of the FC simultaneously receiving one, two, and three reports, respectively, given H_1 for a Gaussian shift-in-mean scenario. Fig. 7(d) plots the probability of receiving one report given H_0 . These are compared against Monte Carlo simulations of the 13-sensor distributed system. These probabilities are plotted as a function of absolute time, k , from the start of system operation. The change is assumed to have occurred at time $k = 0$ under H_1 , and to have never occurred under H_0 . It is observed that after an initial transient period where the values of

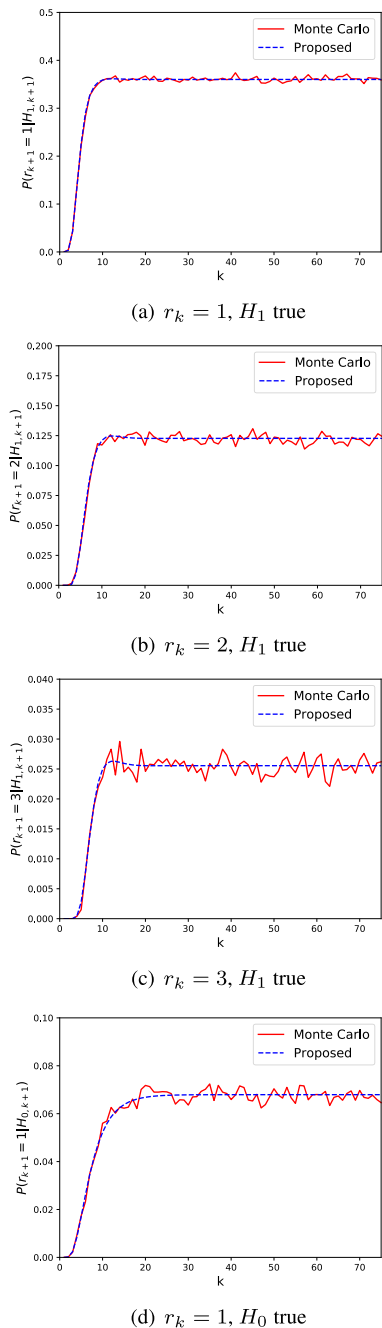


FIGURE 7. Comparison between numerically computed and Monte Carlo generated $P(r_{k+1} = 1|H_{1,k+1})$, $P(r_{k+1} = 2|H_{1,k+1})$, $P(r_{k+1} = 3|H_{1,k+1})$ and $P(r_{k+1} = 1|H_{0,k+1})$ for $N(0, 1)$ vs $N(0.5, 1)$ with $h_l = \ln(15)$, and $M = 13$.

$P(r_{k+1} = n|H_\theta)$ are time varying, a steady state value is quickly reached. Transient periods either occur immediately after a change between H_0 and H_1 as in Fig. 7(a)-(c) as well as at the start of system operation as in Fig. 7(d). As expected, under conditions where numbers of simultaneous reports are small relative to the number of sensors, the frequency of occurrence of the transient reset periods is negligible and the proposed numerical computations closely approximate the reporting probabilities of local sensors observed by Monte Carlo simulation.

IV. LOCAL SENSING ALGORITHM

To design the distributed decentralized system, both FC and local sensors need to ensure that the global false alarm rate constraint is satisfied. Local sensors also require a decision policy to minimize delay in detecting a change in distribution. Without specifying an a priori change time distribution, the CUSUM algorithm (17), with its *minimax* optimality [10], and simplicity by having a single design threshold parameter, h_l , is proposed for local sensing. If the test statistic S_k^i exceeds this threshold at time k , a binary report is generated by local sensor i and transmitted to the fusion center,

$$r_k^i = \begin{cases} 1 & S_k^i \geq h_l \\ 0 & S_k^i < h_l. \end{cases} \quad (29)$$

The generation of one-bit local sensor reports quantizes observations accumulated by local sensors into indicators of whether a local sensor CUSUM has terminated at each time k .

Dropping superscripts, when an arbitrary local sensor's CUSUM terminates, it transmits a report to the FC and its statistic, S_k , is reinitialized to zero. It is worth noting that reinitialization to $S_k = 0$ favours H_0 a priori, while reinitialization to $S_k = h_l$ favours H_1 a priori. Reinitializing to $S_k = 0$ is therefore the most conservative reinitialization strategy for detecting a change, resulting in the smallest possible FC false alarm rate and the largest FC detection delay among all other possible resetting strategies.

It is also noted that from CUSUM properties, the design choice of reinitialization to $S_k = 0$ is equivalent to performing likelihood ratio tests for all possible change times in the future, i.e., a union of likelihood ratio tests, which serves as a method to guarantee a lower bound on the ARL to FA when designing local sensing tests.

Computationally, each local sensor, in deciding to generate a report via Eq. (29), computes Eq. (18) for each time slot, with per-sample likelihood ratio computable by a simple table lookup. Thus, the per-sensor computation is minimal. Moreover, the frequency of communication incurred when $r_k^i = 1$ can be adjusted by h_l based on performance requirements by the proposed test design, which is also dependent on the fusion center algorithm, described next.

V. FUSION CENTER ALGORITHM

At the fusion center, local sensor reports are received sequentially in time. At each time slot, the fusion center receives $r_k = \sum_{i=1}^M r_k^i \in (0, 1, 2, \dots, C)$ local sensor reports, where $C \leq M$ is the maximum number of simultaneous reports allowed in the system. When the report generation policy at local sensors is fixed, the numbers of reports arriving at the FC can be described by a probability mass function (PMF) under H_0 or H_1 .

The case of bandwidth constraints preventing $r_{k+1} \geq C$ reports from being received at the FC in a single time slot requires simultaneous arrivals of C to M reports to be capped at C . This constraint is realistic for large sensor networks,

where the number of reports received in a single time slot could exceed the communication bandwidth to the FC.

In Section III-C Algorithm 1 was proposed to compute the probabilities of receiving multiple reports at the FC of a distributed system based on a repeating CUSUM with reinitialization to zero after choosing H_1 . These probabilities, though time varying, approximated steady state values quickly after startup and after a change in distribution.

The repeating CUSUM analysis method, Algorithm 1, is also applied to compute PMFs describing report arrivals at the FC, denoted by $R_\theta(k, n)$, $\theta \in \{0, 1\}$ for H_0 and H_1 , using the steady state approximation (28) whenever the number of simultaneous reports is small compared to M , and assume that the probabilities are zero (negligible) otherwise. Specifically, let J be a constant chosen such that $J \ll M$, e.g., $J < M/5$. The steady-state report arrival PMF is expressed as

$$R_\theta(k, n) = \begin{cases} \frac{1}{N_0} P(r_{k+1} = n | H_\theta), & n = 1, \dots, \min(J, C) \\ 0 & \text{otherwise,} \end{cases} \quad (30)$$

where $\theta \in \{0, 1\}$ and normalizing factor $N_0 = \sum_{i=1}^J P(r_{k+1} = i | H_\theta)$. It is worth noting that the extreme case of a bandwidth constraint of $C = 1$ on the system results in a Bernoulli distribution.

Reports arriving at the FC described by the above PMFs are processed using CUSUM, with fusion center test statistic

$$Z_k = \max_{1 \leq j \leq k} \sum_{i=j}^k \ln \left(\frac{R_1(k, n)}{R_0(k, n)} \right), \quad (31)$$

and the FC makes the decision that a change has occurred at the first time k where the realization of observed reports, z_k , of Z_k in (31) exceeds FC detection threshold h_f . This procedure is described by Algorithm 2, which also includes local sensing.

A remark here is that computationally, at each time slot, the fusion center algorithm accumulates the total number of received reports up to the network's bandwidth limit and performs the CUSUM test Eq. (31), which can be performed recursively in a similar manner to the local tests via Algorithm 2, Step 9.

Threshold Design Tradeoffs: In order to determine FC CUSUM threshold h_f , the repeated SPRT analysis from Section III-B2 can be applied. Given a (h_l, h_f) (local sensor, global FC) threshold pair, the FC's expected detection delay and false alarm rate are then determined, noting that the report arrival PMFs under H_1 and H_0 input to the FC are determined by h_l . In principle, therefore, a target global system false alarm rate can be achieved by iteratively adjusting h_l and/or h_f . The 2 degrees of freedom to specify overall false alarm rate enables the following flexibility in system design: when h_l is small, communication cost and bandwidth is cheap and the system behaves like a *centralized* detector (with quantization), performance is determined by h_f and detection delay would be low. Alternatively, h_l set to a large

Algorithm 2 Distributed CUSUM FC / Local Sensing

```

1:  $k = z_k = 0$ 
2: while  $z_k < h_f$ :
3:    $n = 0$ 
4:   for  $i = 1$  to  $M$ : % local sensors
5:      $g_k^i = (g_k^i + \ln(\frac{q_1(X_k)}{q_0(X_k)}))^+$ 
6:     if  $g_k^i \geq h_l$ :
7:        $n = n + 1$  % add report
8:   end for
9:    $z_k = (z_k + \ln(\frac{R_1(k, n)}{R_0(k, n)}))^+$ 
10:   $k = k + 1$  % step time
11: end while
12: % where  $(y)^+ \equiv \max(y, 0)$ 

```

value corresponds to high bandwidth cost and, if sufficiently large, results in *or-ing* (fully distributed) detection behaviour with identical global and local false alarm probabilities and maximum detection delay.

VI. SYSTEM DESIGN PROCEDURE

Consider a distributed system with M sensors. The system design requires that $R_{FA}^{FC} \leq \alpha_{FC}$, a given FC false alarm probability, with the objective of minimizing delay E_{DD}^{FC} . In other words, the design objective, (5) becomes

$$\min_{h_l, h_f} E_{DD}^{FC} \quad (32)$$

$$\text{subject to } R_{FA}^{FC} \leq \alpha_{FC}. \quad (33)$$

For M sensors, α_{FC} can be related to a local sensor minimum false alarm rate constraint, α_l , via

$$\frac{\alpha_{FC}}{M} = \alpha_l. \quad (34)$$

To avoid over design and keep delay minimum, local sensors should achieve a local false alarm rate R_{FA}^l no less than the constraint, i.e.,

$$R_{FA}^l \geq \alpha_l. \quad (35)$$

If a local sensing threshold h_l is chosen such that $R_{FA}^l \approx \alpha_l$ in (35) the FC should terminate and choose H_1 after only one report in order to ensure $R_{FA}^{FC} = \alpha_{FC}$. This value of h_l serves as the upper bound. For h_l that result in $R_{FA}^l > \alpha_l$, more than one report is required to be processed before choosing H_1 at the FC in order to satisfy $R_{FA}^{FC} = \alpha_{FC}$. If a value of h_l is selected that causes local sensors to violate (35), then for any choice of h_f the proposed distributed system will be over designed in terms of false alarm rate, resulting in greater delay.

A suboptimal solution is therefore proposed: given the value of α_l from (34), the maximum possible value of h_l that satisfies (35) is computed using the CUSUM analysis of Section III-B2 within a discretized search procedure. Once this value of h_l is determined, its associated reporting PMFs are computed using Algorithm 1 and then the FC threshold h_f is determined that designs the system to achieve the closest R_{FA}^{FC} to α_{FC} as possible by treating it as a shift in PMF test,

as described in Section V. A 1-D search is then performed as follows: compute and store the E_{DD}^{FC} and R_{FA}^{FC} of the given (h_l, h_f) pair, decrease h_l and repeat.

The choice of a final (h_l, h_f) pair to design local sensors and the FC is based on system requirements. For example, if the only design goal is to minimize expected detection delay while meeting a maximum false alarm rate, the threshold pair with minimum E_{DD}^{FC} is selected. If, for example, energy consumption at the sensors is a secondary concern compared to bandwidth within the network, it may be desirable to reduce the number of report transmissions between local sensors and the FC. In this case, the threshold pair with largest possible h_l is chosen so that the reports are most informative, and thus fewer reports are needed for a decision.

Given a system with M sensors, C parallel channels and false alarm probability constraint α_{FC} , the design procedure can be summarized by the following steps:

- 1) Using Eq. (34), convert α_{FC} to a lower bound on the local sensor false alarm rate constraint, α_l .
- 2) Until R_{FA}^l is sufficiently close to α_l , compute R_{FA}^l for different local sensor thresholds, h_l , using the procedure in Section III-B2. The result is the maximum value of h_l to be used in Step 3.
- 3) With the current h_l , perform Algorithm 1 and (28) to determine the reporting PMFs at the FC, $R_0(r_k)$ and $R_1(r_k)$, $r_k = 0, 1, \dots, C$, corresponding to H_0 and H_1 , respectively.
- 4) Solve FC report arrival discrete shift in PMF problems: Until R_{FA}^{FC} is sufficiently close to α_{FC} , vary h_f and compute E_{DD}^{FC} and R_{FA}^{FC} using PMFs $R_1(r_k)$ and $R_0(r_k)$ from Step 3 using the analysis in Section III-B2.
- 5) Record $(E_{DD}^{FC}, R_{FA}^{FC})$ for threshold pair (h_l, h_f) .
- 6) Decrease h_l and go to Step 3.

It is worth remarking that the proposed system design has the following advantages in scaling to large sensor networks over existing approaches:

- The CUSUM analysis of Section III-B accounts for sequential test overshoot in the design. This accuracy avoids over designing error specifications which result in excess detection delay. As the number of sensors, M , in the distributed system increases, the larger number of overshoot events has a cumulative effect.
- The design complexity is determined by the desired accuracy of the convolutions, which is proportional to the square of the number of bins in the discretized PDFs, typically 100, multiplied by the number of stages, N in Eq. (16), with typical values in the range from 10-50. Thus, test design does not increase appreciably with network size M . Increasing M only increases the complexity of computing Eq. (28), which can be approximated for large M by a closed form expression using Stirling's formula.

VII. NUMERICAL RESULTS

The proposed test design procedure for distributed detection presented in Section VI is assessed for its accuracy

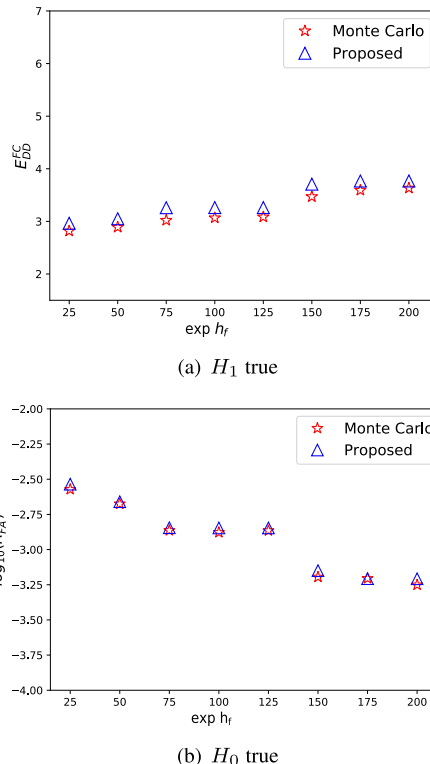


FIGURE 8. Expected detection delay and false alarm rate vs. FC threshold h_f for $N(0, 1)$ vs $N(0.5, 1)$ test with $M = 25$, $h_l = \ln(60)$, $C = 3$.

of expected detection delay and false alarm rate in Section VII-A. Trade offs among choices of fusion center thresholds, number of sensors and numbers of parallel channels (bandwidth) are provided in Section VII-B, and comparisons to existing schemes are discussed in Section VII-C. Application to change-in-variance detection is presented in Section VII-D, and performance in a multiple-alternative change scenario is presented in Section VII-E.

A. DESIGN ACCURACY

To illustrate the accuracy of the proposed design procedure, a Gaussian shift in mean test of $N(0, 1)$ vs $N(0.5, 1)$ is investigated under different combinations of parameters h_l , M , and C . Each combination resulted in a unique report arrival PMF from local sensors which are computed using Algorithm 1. The procedure in Section III-B2 was then used to compute E_{DD}^{FC} and R_{FA}^{FC} for each candidate threshold h_f . Monte Carlo trials simulating the distributed networks were performed to evaluate the accuracy of the proposed procedure in predicting E_{DD}^{FC} and R_{FA}^{FC} . This comparison is shown for 25, 15, and 7-sensor networks and for a bandwidth limit of $C = 3$ simultaneous reports in Figures 8, 9, and 10, respectively. Closer agreement is observed between the proposed analysis and Monte Carlo simulation in Fig. 8 for $M = 25$ sensors compared to Figs. 9 and 10 for $M = 15$ and 7 sensors, respectively. This is expected since the discrepancy is due to the previously described steady-state approximation in Section III-C, noting that for $M = 25$ there

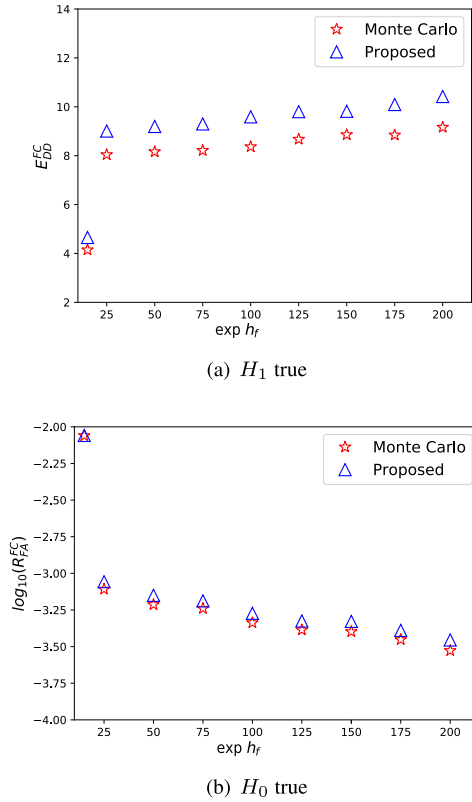


FIGURE 9. Expected detection delay and false alarm rate as a function of FC threshold h_f for a $N(0, 1)$ vs $N(0.5, 1)$ with $M = 15$, $h_f = \ln(60)$, $C = 3$.

is a lower occurrence of reset transients in the report arrival process. Examining the performance tradeoffs among these three scenarios, increasing the numbers of sensors from 7 to 15 to 25 decreases expected delay from 8.5 to 5.0 to 3.9, respectively at a log false alarm rate of -3.5 , or 3.16×10^{-4} .

Significant *staircase discontinuities* mentioned in Section III-B2 are also observed in all these results arising from the discrete report arrival distributions (PMFs) with at most $C + 1$ non-zero terms. When C is small, the discretization of observation distributions causes discontinuities in achievable false alarm rates for FC design under different values of h_f . This is clearly seen in Figs. 8-9. At discontinuities, randomization could be used at the FC detector in order to achieve intermediate desired false alarm rates. The largest discontinuity in these scenarios occurs at low h_f values, corresponding to FCs that terminate after one report versus those that terminate after more than one report.

B. BANDWIDTH / DELAY PERFORMANCE TRADEOFFS

The performance of the proposed distributed quickest detection framework as characterized by expected detection delay when achieving a given false alarm rate is investigated in Figure 11 that considers an $N(0, 1)$ vs $N(0.5, 1)$ test with $M = 13$ sensors as a function of bandwidth constraint C . The difference in system performance between $C = 1$ versus $C = 2$ is significant since two reports are often received simultaneously at the FC in this scenario, and there is

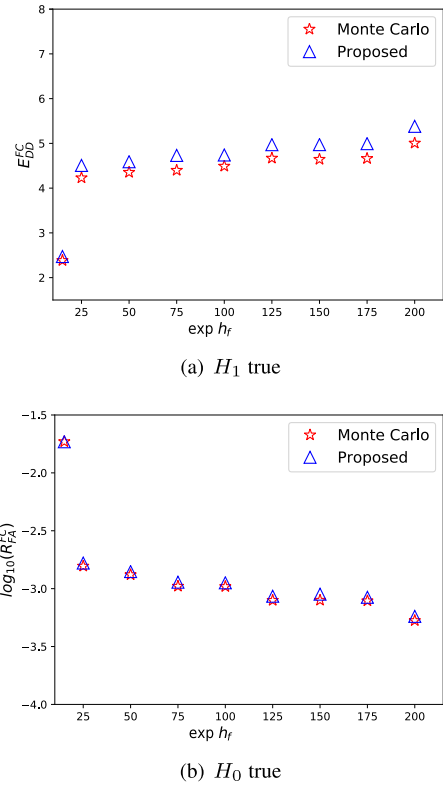


FIGURE 10. Expected delay and false alarm rate vs. FC threshold h_f for $N(0, 1)$ vs $N(0.5, 1)$, $M = 7$, $h_f = \ln(60)$, $C = 3$. Here $M \gg C$ is not met.

noticeable gain in the proposed framework that exploits distinguishing between numbers of reports being received. The difference between $C = 2$ and 3, in contrast, is marginal because of the rarity of receiving three reports simultaneously for this scenario. From this analysis, it can be concluded that a bandwidth constraint of $C = 2$ is sufficient to obtain most of the performance benefits.

A greater improvement from $C = 1$ to 2 is observed as the false alarm rate constraint R_{FA}^{FC} decreases. A tighter false alarm rate constraint results, on average, in a greater number of reports required by the FC to decide on the change to H_1 . Figure 12 shows results for the scenario described above as the number of sensors, M , is varied with bandwidth fixed at $C = 3$. As expected, for a given false alarm rate, increasing the number of sensors reduces delay due to the greater observation rate. These differences magnify as the false alarm rate decreases. For example, at a global false alarm rate of 10^{-3} increasing the number of sensors from 4 to 13 can reduce expected delay by a factor of nearly 3.

C. COMPARISONS TO EXISTING SCHEMES

The proposed distributed detection scheme is not directly comparable to others in the literature due to differing network and/or communication assumptions between local sensors and FC. A recent approach for a similar type of sensor network is the one-shot type that has severe communication constraints. In Fig. 11 are Monte Carlo simulation results for the m -th alarm scheme proposed in [29] that makes an H1

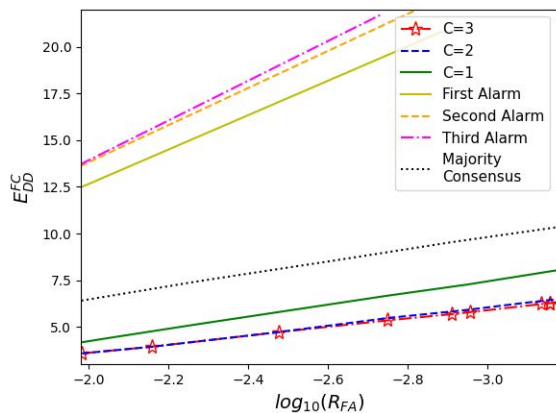


FIGURE 11. Delay performance of decentralized CUSUM system for $N(0, 1)$ vs $N(0.5, 1)$ with $h_l = \ln(15)$ and $M = 13$ sensors as a function of false alarm rate under varying bandwidth constraints C as well as comparison to the m -th alarm scheme for $m=1,2$ and 3 and majority consensus schemes [29].

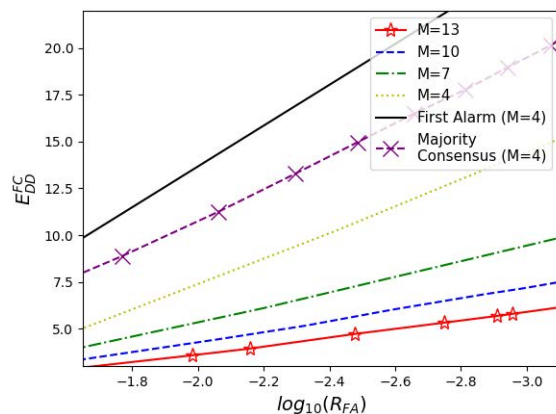
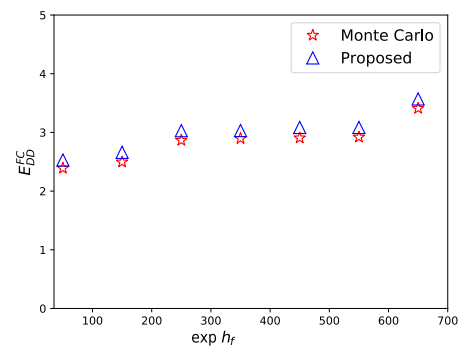


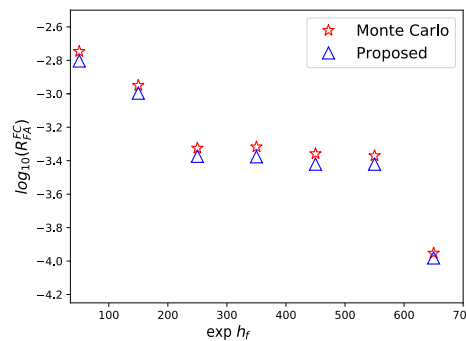
FIGURE 12. Delay performance of the decentralized CUSUM system for $N(0, 1)$ vs $N(0.5, 1)$ with $h_l = \ln(15)$ with different number of sensors as well as in comparison to First Alarm and Majority Consensus schemes with $M=4$ sensors in [29] vs false alarm rate.

decision upon receipt of m local CUSUM change detection reports at the FC for the case of $M=13$ sensors. In Fig. 11, the same false alarm rate results in m -th alarm's significantly higher expected delay curves as shown for $m = 1, 2$, and 3. This is expected since m -th alarm minimizes total number of reports transmitted to the FC whereas the proposed scheme limits communication bandwidth (available channels), i.e., number of simultaneous reports rather than the total number.

For the case of fewer sensors, $M = 4$, Fig. 12 reveals that for both the m -th alarm and majority consensus schemes in [29], expected delays are somewhat higher than those of the proposed scheme. This would be expected since communication cost limits performance more significantly as the number of sensors increases. Previously proposed one-shot schemes are also limited by less information. Upon decisions, local sensors are shut off in [29] whereas the proposed scheme utilizes observations that continue to arrive at local sensors. Local sensing can instead be viewed as filtering and/or quantization. An alternative strategy is



(a) H_1 true



(b) H_0 true

FIGURE 13. Expected detection delay and false alarm rate as a function of FC threshold h_f for $N(0, 1)$ vs $N(0, 2)$ with $M = 25$, $h_l = \ln(60)$, $C = 3$.

centralized detection where all information is sent to the FC. A centralized system with CUSUM FC (not shown in Fig. 11) has expected delay ranging from 1.8 to 2.5 samples for $\log_{10}(R_{FA})$ in the range -2.0 to -3.0 [41], far less than the proposed scheme. This is expected since centralized detection does not reduce information communicated from local sensors to FC.

In summary, the proposed approach allows for flexibility in communication, controlled by design thresholds h_l and h_f . A combination of large h_l and small h_f yields less communication and greater expected delay at a given overall false alarm rate, resulting in performance similar to methods in [29]. Conversely, small h_l and large h_f increase communication and lower expected delay, with performance approaching that of a centralized system with no communication restriction.

D. CHANGE-IN-VARIANCE SCENARIO

Next, the same design procedure is applied to the problem of detecting a change in variance from 1 to 2 of a zero-mean Gaussian signal using a 25-sensor network under a bandwidth restriction of 3 simultaneous reports. In contrast to a shift in mean of a deterministic signal featured in the previous section, this scenario models an inherently stochastic signal. This represents detection of a change in observed energy, which may model signal prediction error in monitoring the behaviour of a stochastic system with independent prediction

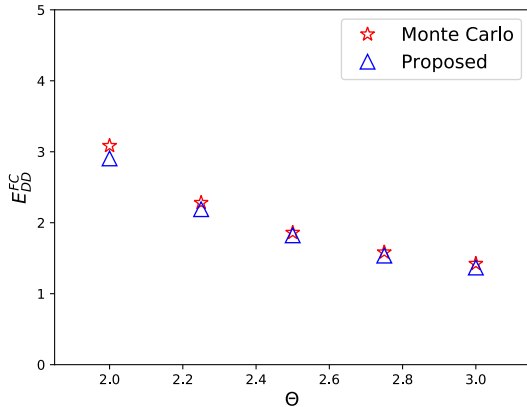


FIGURE 14. Expected detection delay as a function of variance θ for $N(0, 1)$ vs $N(0, \theta)$ with $M = 25$, $h_l = \ln(60)$, $h_f = \ln(450)$, $C = 3$.

errors. As observed in Fig. 13, Monte Carlo simulation closely matches the analytical performance of the proposed design. As expected, the low value of $C = 3$ reports results in significant discretization (staircase) effects on the achievable false alarm rates for the reasons explained at the end of Section VII-A.

E. MULTIPLE-ALTERNATIVE SCENARIO

To illustrate a multi-alternative change scenario, shown in Fig. 14 is the performance of a change-of-variance test of $N(0, 1)$ versus $N(0, \theta)$ for $\theta \in [2, 3]$ with $M = 25$ sensors and a limit of $C = 3$ simultaneous reports. The test thresholds $h_l = \ln(60)$ and $h_f = \ln(450)$ were obtained by designing for the worst-case false alarm value corresponding to $\theta = 2$, though other possible criteria based on the set of H_1 hypotheses may be used. Note that the false alarm probability in this example is unaffected by θ , and the design for $\theta = 2$ meets the specifications for $\theta > 2$ alternatives. As illustrated by the agreement with Monte Carlo simulation, the proposed analysis is applicable to hypotheses with arbitrary PDFs that are independent over time that may not match ones used in the nominal design.

VIII. CONCLUSION AND FUTURE WORK

A distributed quickest change detection system is proposed that features sensors that reset after making decisions to process future data rather than operating on a one-shot basis. The proposed minimax approach, based on CUSUM for independent observations, is shown to provide greater communication constraint flexibility than existing systems such as one-shot or centralized detection. Performance analysis applied to test design closely meets global false alarm specifications while satisfying bandwidth restrictions on numbers of simultaneous reports received by the FC from local sensors in any time slot. The numerical procedures developed in Section III analyze local sensor operation by computing probability distributions of numbers of arriving summary reports at the FC. This information is used for designing FC and local sensor thresholds. The proposed approach enables quantitative tradeoffs of network size and

sensor communication bandwidth against delay performance among systems with the same global false alarm rate. Comparisons with existing systems show that the proposed system outperforms low communication one-shot detection and is outperformed by high communication centralized detection as expected. In the future, the proposed method could be extended to include the design of alternative quantization schemes employed by local sensors other than the binary one used here.

REFERENCES

- [1] C. Roff, J. R. Henderson, D. Clarke, M. C. Walden, and S. Fitz, "Low-cost millimeter-wave radio-frequency sensors: New applications enabled by developments in low cost chipsets," in *Proc. IEEE SENSORS*, Nov. 2017, pp. 1–3.
- [2] K. Routh and T. Pal, "A survey on technological, business and societal aspects of Internet of Things by Q3, 2017," in *Proc. 3rd Int. Conf. Internet Things, Smart Innov. Usages (IoT-SIU)*, Feb. 2018, pp. 1–4.
- [3] Y. Shaikh, V. K. Parvati, and S. R. Biradar, "Survey of smart healthcare systems using Internet of Things (IoT): (Invited paper)," in *Proc. Int. Conf. Commun., Comput. Internet Things (IC3IoT)*, Feb. 2018, pp. 508–513.
- [4] M. A. Al-Jarrah, M. A. Yaseen, A. Al-Dweik, O. A. Dobre, and E. Alsusa, "Decision fusion for IoT-based wireless sensor networks," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1313–1326, Feb. 2020.
- [5] A. Nayyar and D. Teneketzis, "Signaling in sensor networks for sequential detection," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 1, pp. 36–46, Mar. 2015.
- [6] M. N. Kurt and X. Wang, "Multisensor sequential change detection with unknown change propagation pattern," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, pp. 1498–1518, Jun. 2019.
- [7] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Statist.*, vol. 19, no. 3, pp. 326–339, Nov. 1947.
- [8] A. R. Kamat, "Continuous inspection schemes," *Biometrika*, vol. 41, nos. 1–2, pp. 100–115, Apr. 1954.
- [9] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Stat.*, vol. 42, pp. 1897–1908, Dec. 1971.
- [10] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Stat.*, vol. 14, no. 4, pp. 1379–1387, 1986.
- [11] L. Xie, S. Zou, Y. Xie, and V. V. Veeravalli, "Sequential (quickest) change detection: Classical results and new directions," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 2, pp. 494–514, Jun. 2021.
- [12] J. Choi, M. O. Mughal, Y. Choi, D. Kim, J. A. Lopez-Salcedo, and S. Kim, "CUSUM-based joint jammer detection and localization," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Oct. 2018, pp. 1–5.
- [13] W. Pi, P. Yang, D. Duan, C. Chen, X. Cheng, L. Yang, and H. Li, "Malicious user detection for cooperative mobility tracking in autonomous driving," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 4922–4936, Jun. 2020.
- [14] N. Maleki and A. Vosoughi, "On bandwidth constrained distributed detection of a known signal in correlated Gaussian noise," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11428–11444, Oct. 2020.
- [15] S. Anand and S. K. Routray, "Issues and challenges in healthcare narrowband IoT," in *Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Mar. 2017, pp. 486–489.
- [16] P. A. Laplante and N. Laplante, "The Internet of Things in healthcare: Potential applications and challenges," *IT Prof.*, vol. 18, no. 3, pp. 2–4, May 2016.
- [17] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The Internet of Things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.
- [18] R. W. Crow and S. C. Schwartz, "Quickest detection for sequential decentralized decision systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 1, pp. 267–283, Jan. 1996.
- [19] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.
- [20] B. Jiang and L. Lai, "Distributed quickest detection with optional observations at the fusion center," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, 2016, pp. 112–116.

- [21] D. Yang and H. Qi, "An effective decentralized nonparametric quickest detection approach," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2278–2281.
- [22] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3727–3735, Jul. 2015.
- [23] N. Sriranga, K. G. Nagananda, R. S. Blum, A. Saucan, and P. K. Varshney, "Energy-efficient decision fusion for distributed detection in wireless sensor networks," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, 2018, pp. 1541–1547.
- [24] D. Li, S. Kar, and S. Cui, "Distributed quickest detection in sensor networks via two-layer large deviation analysis," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 930–942, Apr. 2018.
- [25] J. Li, D. Towsley, S. Zou, V. V. Veeravalli, and G. Ciocarlie, "A consensus-based approach for distributed quickest detection of significant events in networks," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 1–4.
- [26] H. Wang and S. D. Blostein, "Communication-efficient decentralized change detection for cognitive wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
- [27] O. Hadjiladis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3346–3359, Jul. 2009.
- [28] S. Biswas, S. Dey, S. Knorn, and A. Ahlen, "On optimal quantized non-Bayesian quickest change detection with energy harvesting," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 433–447, Jun. 2020.
- [29] S. Banerjee and G. Fellouris, "Decentralized sequential change detection with ordered CUSUMs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 36–40.
- [30] Y. Yilmaz, G. V. Moustakides, and X. Wang, "Cooperative sequential spectrum sensing based on level-triggered sampling," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4509–4524, Sep. 2012.
- [31] Y. Yilmaz, G. V. Moustakides, and X. Wang, "Channel-aware decentralized detection via level-triggered sampling," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 300–315, Jan. 2013.
- [32] Y. Yilmaz and X. Wang, "Sequential distributed detection in energy-constrained wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3180–3193, Jun. 2014.
- [33] A. G. Tartakovsky and V. V. Veeravalli, "An efficient sequential procedure for detecting changes in multichannel and distributed systems," in *Proc. 5th Int. Conf. Inf. Fusion*, vol. 1, Jul. 2002, pp. 41–48.
- [34] L. A. Arojan and D. E. Robison, "Direct methods for exact truncated sequential tests of the mean of a normal distribution," *Technometrics*, vol. 11, no. 4, pp. 661–675, 1969.
- [35] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Boca Raton, FL, USA: CRC Press, 2015.
- [36] S. D. Blostein and T. S. Huang, "Detecting small, moving objects in image sequences using sequential hypothesis testing," *IEEE Trans. Signal Process.*, vol. 39, no. 7, pp. 1611–1629, Jul. 1991.
- [37] A. Wald, *Sequential Analysis*. Hoboken, NJ, USA: Wiley, 1947.
- [38] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, 1945.
- [39] E. J. Dudewicz and S. N. Mishra, *Modern Mathematical Statistics*. New York, NY, USA: Wiley, 1988.
- [40] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York, NY, USA: Springer-Verlag, 1985.
- [41] M. Quarisa, "Procedures to analyze and design decentralized, distributed sensor networks for change detection," M.S. thesis, Dept. Elect. Comput. Eng., Queen's Univ., Kingston, Ontario, Canada, 2019.



MATTHEW W. QUARISA (Member, IEEE) received the B.A.Sc. and M.A.Sc. degrees in electrical engineering from Queen's University, Kingston, ON, Canada, in 2017 and 2019, respectively.

Since 2019, he has been employed with Cisco Systems, Toronto, ON, Canada, as a Systems Engineer working with both customers and partners. His current research interests include wireless communications and network programmability.

This includes LoRaWAN-based sensor networks, private 5G, and network infrastructure automation and telemetry. He is also interested in technologies and trends that are converging information technology and operational technology systems, processes, and insights.



STEVEN D. BLOSTEIN (Senior Member, IEEE) received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, USA, in 1983, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois Urbana-Champaign, IL, USA, in 1985 and 1988, respectively.

Since 1988, he has been with the Faculty of the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON, Canada, and currently a Professor. From 2004 to 2009, he was the Department Head. He has also been a consultant to industry and government in the areas of image compression, target tracking, radar imaging, and wireless communications. His current research interests include wireless communications systems, including detection and estimation, signal processing, energy efficiency, MIMO, dynamic access, and dense deployments. He was the Chair of the IEEE Kingston Section, the Chair of the Biennial Symposium on Communications, the Publications Chair of IEEE ICASSP, an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and served on numerous technical program committees for the IEEE Communications Society conferences. He is a Registered Professional Engineer in Ontario.

• • •