HIERARCHICAL CENSORING SENSORS FOR CHANGE DETECTION

*Neal Patwari and Alfred O. Hero III**

University of Michigan, Dept. of EECS
Ann Arbor, MI 48109, USA
E-mail: [npatwari, hero]@eecs.umich.edu

Brian M. Sadler

Army Research Laboratory
Adelphi, MD 20783, USA
bsadler@arl.army.mil

ABSTRACT

Devices in energy-limited wireless sensor networks remain in a low-communication 'sleep' mode until an alarm event is detected. It has been proposed to use 'censoring sensors' to reduce the probability that a sensor must transmit in this mode, thereby minimizing energy consumption when alarm events are not occurring, and lengthening sensor lifetime. Further, since devices in multi-hop networks are not usually in single-hop range of a fusion center, hierarchical distributed detection can lead to further energy efficiency. We report on a system that applies censoring in a hierarchical network to the CUSUM test of Page and Lorden, an online abrupt change detector. In this paper, we explore via simulation an example change detection problem and demonstrate that significant reduction in the number of sensor transmissions can be achieved at the cost of a small increase in mean detection delay compared to uncensored change detection performance.

1. INTRODUCTION

Large-scale wireless sensor networks can be deployed to monitor wide areas for changing conditions in applications such as monitoring of seismic or acoustic activity, inventory location and tracking, medical monitoring, and intrusion detection. In large scale deployments, which may include thousands of sensors, change detection algorithms must be distributed to avoid data bottlenecks at a central decision point. Since broad coverage wireless sensor networks will be multi-hop, decentralizing the decision process using a hierarchy can be effective. Furthermore, as opposed to capacity-constrained networks in which detection is distributed to minimize the bit rate, in energy-limited wireless sensor networks a more appropriate goal is to minimize the probability that sensors must transmit. In this paper, we introduce a hierarchical 'censored' implementation similar to [8] for change detection. We use simulation to illustrate the performance of the distributed censored implementation vs. a centralized uncensored implementation.

1.1. Energy Constraint

Energy-limited devices in wireless sensor networks must use low duty cycles, ie. low percentage of device 'on' time (eg., on the order of 0.01% to 1%) in order to achieve long lifetimes of operation. When necessary, a device wakes up its sensor, processor, transmitter or receiver in order to sense, process, receive or transmit a message. Each wake-up consumes significant energy. Specifically, for the transmitter circuitry, it has been reported that there is a tradeoff between the time required for (and thus energy expended during) wake-up and the energy used during sleep mode [1]. Due to the large percentage of time spent in sleep mode, sleep mode energy is minimized, and, as a result, wake-up energy is high. It has been reported that wake-up energy can be significantly higher than the energy used during transmission [2].

Much distributed detection research has focused on capacity-constrained networks. Research has addressed quantization of sensor data [3] and exploiting source correlation [4] to reduce sensor bit rate. In particular cases, it has been shown that for a R-sensor network with a capacity constraint of R bits per unit time, having each sensor send one bit is optimal [5]. However, from the perspective of energy, the cost of transmitting one bit involves wake-up energy and packet overhead such as synchronization and 'id' bits. Considering all energy costs in an energy budget as in [6] shows that sending one bit of data consumes only marginally less energy than sending many bits. In fact, it can be argued that the appropriate constraint to bound energy consumption for many wireless sensor networks is the probability of transmission from each sensor, as used in [7].

1.2. Hierarchical Networks

In broad coverage wireless sensor networks, the devices' limited range makes it necessary to route communication to a fusion center through intermediate devices in the network. Such multi-hop routing is used for energy-efficiency and reducing device cost; long-range transmission energy is decreased due to lower $1/r^n$ losses. Network-wide power savings in multi-hop systems can be significant, especially for large-scale networks when reception energy costs are small compared to transmission costs. Technology scaling should reduce receiver energy consumption, while transmission costs will remain constant [1]. This projection underscores the importance of both multi-hop routing and minimization of the probability of transmission.

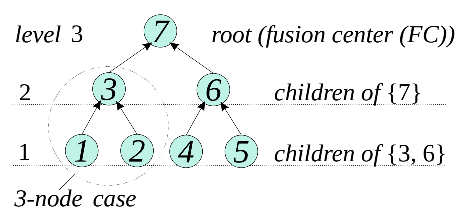

Fig. 1. Diagram of an example hierarchical network of sensors. Subset of sensors $\{1, 2, 3\}$ is also used as an example.

*This work was supported in part by a U. of Michigan Dept. of Electrical Engineering & Computer Science Graduate Fellowship and by ARO-DARPA MURI Grant #DAAD19-02-1-0262.

The use of multi-hop routing in wireless sensor networks can also be exploited to improve detection performance. In the distributed detection literature, it is often assumed that all sensors communicate directly with a fusion center. Intermediate nodes can perform much of the data aggregation and make local decisions, distributing the computation load across the network. In this paper, we consider a hierarchical 'spanning tree' topology (eg., Fig. 1) for the purpose of change detection.

1.3. Previous Work

Censoring sensors was presented in [8] and [9] as a means to reduce the number of sensor transmissions by requiring sensors to transmit only 'informative' observations. Observations are considered informative only if their local likelihood ratio (LLR) falls within a 'send' region R. In [8], the optimal censoring region was shown to be a single interval, $\bar{R} = [\phi^{(i)}, \lambda^{(i)})$. Moreover, in cases where the prior probability of H_1 is sufficiently small and limited communication is allowed, it is optimal to set $\phi^{(i)} = 0$. Sufficient conditions for the optimality of $\phi^{(i)} = 0$ are given in [9]. In previous work [10], we applied censoring sensors to static distributed simple hypothesis test in a hierarchical sensor network such as shown in Fig. 1. In a series of examples, we compared ROC curves for hierarchies of 3 and 7 sensors, in which sensors censored from 70% to 97% of their observations. Results showed that close to optimal (ie., centralized) detection performance could be achieved with a fraction of the transmissions. We expand the results in this paper by applying hierarchical censoring to a distributed dynamic online abrupt change detection problem.

2. CUSUM CHANGE DETECTION

In online abrupt change detection, we measure data $\mathbf{y}_i$ at each time i for $i \in 1 \ldots k$. The measured data $\mathbf{y}_i$ is a vector of measurements from N sensors, $\mathbf{y}_i = [y_i^{(1)}, \ldots, y_i^{(N)}]^T$. While $i < t_0$, $\mathbf{y}_i \sim p(\mathbf{y}_i|\theta_0)$, and $\mathbf{y}_i \sim p(\mathbf{y}_i|\theta_1)$ for $i \geq t_0$. Online abrupt change detectors seek to detect a change as quickly as possible using $\mathbf{y}_1, \ldots, \mathbf{y}_k$, where $k \geq t_0$, typically by thresholding a function g_k of the past observations. Specifically the detector is defined by its stopping time t_a, given by,

$$t_a = \min\{k \geq 1 : g_k(\mathbf{y}_1, \ldots, \mathbf{y}_k) > \lambda\} \qquad (1)$$

where λ is a detection threshold. The stopping time t_a can be viewed as an estimate of t_0. When $t_a \geq t_0$ the detector is said to have correctly detected with delay $t_a - t_0$; for $t_a < t_0$, the detection is a false alarm. We consider specifically the case in which $\{y_i^{(n)}\}$ are statistically independent $\forall i = 1 \ldots k$ and $n = 1 \ldots N$. The log-likelihood ratio (LLR) of the data at time k is thus

$$\Lambda_k = \sum_{i=1}^{k} \sum_{n=1}^{N} s_i^{(n)}, \quad \text{where} \quad s_i^{(n)} = \ln \frac{p(y_i^{(n)}|\theta_1)}{p(y_i^{(n)}|\theta_0)}. \qquad (2)$$

2.1. The N-Sensor Centralized CUSUM Test

The typical behavior of Λ_k is to exhibit a negative drift while $k < t_0$, and then a positive drift for $k \geq t_0$. Thus the relevant information can be seen to be presented by the difference between the current value Λ_k and the minimum value of Λ_i, $i \in 1 \ldots k$ [11]. This is the approach of the CUSUM change detector proposed by Page [12] and further analyzed by Lorden [13]. In the CUSUM test, g_k of (1) is given by,

$$g_k = \Lambda_k - m_k, \quad \text{where} \quad m_k = \min_{j \leq k} \Lambda_j. \qquad (3)$$

The CUSUM test has optimality properties as the detection threshold approaches infinity [11]. We discuss two figures of merit when discussing the CUSUM test: the average mean time between false alarms, $\mathrm{E}_0[t_a]$, and the mean delay for detection, $\mathrm{E}_1[t_a]$, where E_0 denotes the expectation when H_0 is always true ($t_0 = \infty$), and E_1 denotes the expectation when $t_0 = 0$. By raising the threshold λ, we can increase $\mathrm{E}_0[t_a]$ at the expense of increasing $\mathrm{E}_1[t_a]$. The tradeoff between $\mathrm{E}_0[t_a]$ and $\mathrm{E}_1[t_a]$ defines the performance of an online change detector [11].

2.2. The Hierarchical Censoring CUSUM Test

In a hierarchical network of N sensors, each sensor performs its own local CUSUM test at each time instant. If sensor n is the fusion center (root node in Fig. 1) FC, then its CUSUM test decides whether or not a change to H_1 has occurred. If $n \neq FC$ then the CUSUM test decides whether or not to transmit to the parent of n. We denote the set of children of node n as K_n (eg. in Fig. 1,$K_7 = \{3, 6\}$). At time i, sensor n records data $y_i^{(n)}$ and receives (or doesn't receive) data reported (or censored) by its children $m \in K_n$. Let $k_0^{(m)}$ denote the last time that sensor m transmitted its data to parent n, or $k_0^{(m)} = 0$ if m hasn't yet transmitted. At time i, if $i > k_0^{(m)}$ for any $m \in K_n$, then sensor n must form its decision on whether to transmit to its parent (or to decide H_1 if it is the fusion center) on incomplete information. The decision at sensor n is a threshold test based on its own sensor data, data its children have reported, and estimates of the data its non-reporting children have censored.

If sensor $n \neq FC$ decides to transmit at time i, it sends its parent a full report of previously censored data, ie., $y_i^{(n)}$ for $i = k_0^{(n)} + 1 \ldots i$; resets $k_0^{(n)} := i$; and restarts its test. Since the bit rate during transmission is not constrained by (9), full data reporting is possible although wasteful, since, data from the distant past is largely uninformative to the parent. For practical implementations, a sensor would likely significantly compress this data series. Here, full reporting is used in order to explore the capabilities of hierarchical censoring. Future work should compare these results to those in which sensors report compressed time series data.

Since a parent node has only imperfect or delayed knowledge of its children's data, its LLR as a function of time i can only be estimated and updated over time. Thus for the censored case, we adopt the somewhat bulky but necessary notation, that $\Lambda_i^{(n)}(k)$ is the estimate of the LLR at time k by sensor n of what its LLR was at time i. Then the hierarchical test at node n is defined by its stopping time $t_a^{(n)}$,

$$t_a^{(n)} = \min\{k > k_0^{(n)} : g_k^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_k) > \lambda^{(n)}\}, \qquad (4)$$

where $g_k^{(n)}$ takes the form,

$$g_k^{(n)} = \Lambda_k^{(n)}(k) - m_k^{(n)}, \quad \text{where} \quad m_k^{(n)} = \min_{k_0^{(n)} \leq j \leq k} \Lambda_j^{(n)}(k). \qquad (5)$$

At time k, the estimate of sensor n's LLR, $\Lambda_i^{(n)}(k)$, is

$$\Lambda_j^{(n)}(k) = \sum_{i=1}^{j} S_i^{(n)}(k), \tag{6}$$

$$S_i^{(n)}(k) = s_i^{(n)} + \sum_{\substack{m\in K_n:\\ i\le k_0^{(m)}}} S_i^{(m)}(k_0^{(m)}) + \sum_{\substack{m\in K_n:\\ i> k_0^{(m)}}} \bar{S}_i^{(m)}. \tag{7}$$

The three terms in (7) correspond to:

1. The log-likelihood ratio $s_i^{(n)}$ of the data recorded at time i by node n's own sensor, given previously in (2),
2. The combined log-likelihood ratio $S_i^{(m)}(k_0^{(m)})$ for each child node m which has transmitted data more recently than i, and
3. An estimate $\bar{S}_i^{(m)}$ of the (unknown) combined log-likelihood ratio $S_i^{(m)}$ for each child node m which has *not* transmitted data more recently than i. More specifically,

$$\bar{S}_i^{(m)} = \mathrm{E}_0\big[s_i^{(m)}\big] + \sum_{l\in K_m} \bar{S}_i^{(l)} \tag{8}$$

In words, $\bar{S}_i^{(m)}$ is the mean LLR under H_0 summed over all sensors l that are 'descendants' (children, children of children, etc.) of non-reporting sensor m.

We select the thresholds, $\lambda^{(n)}, \forall n \neq FC$, so that the average probability of sensor transmission under H_0 is bounded by $\rho \leq 1$. Specifically,

$$\frac{1}{N-1} \sum_{\substack{n=1\\ n\neq FC}}^{N} \left(\mathrm{E}_0[t_a^{(n)}]\right)^{-1} \leq \rho \leq 1. \tag{9}$$

Note that $t_a^{(n)}$ as defined in (4) is always ≥ 1. We separately set the threshold at the fusion center, $\lambda^{(FC)}$, to achieve the desired mean time between false alarms $\mathrm{E}_0[t_a^{(FC)}]$ and mean delay for detection $\mathrm{E}_1[t_a^{(FC)}]$.

3. SIMULATION RESULTS

In the following examples, we explore via simulation the performance of the hierarchical censored CUSUM detector described in Section 2.2 for the Gaussian change-of-mean case. For Gaussian data, the LLR $s_i^{(n)}$ is given by,

$$s_i^{(n)} = \frac{\mu_1 - \mu_0}{\sigma_y^2}\left(y_i^{(n)} - \frac{\mu_1+\mu_0}{2}\right).$$

To calculate the performance of a centralized CUSUM test, we use either the Seigmund approximation given in Chapter 5 of [11], or simulation results, since at low SNR or low thresholds, the Siegmund approximation can be inaccurate.

3.1. Three-Sensor Example

We first study a simple three sensor network (the three-sensor subnetwork shown in Fig. 1) where nodes 1 and 2 are children and node 3 is the parent and fusion center. Our first simulation studies the case when $\mu_0 = 0$, $\mu_1 = 1$, and $\sigma = 1$, ie., the case when the SNR is 1 at all sensors. The communications constraint is that

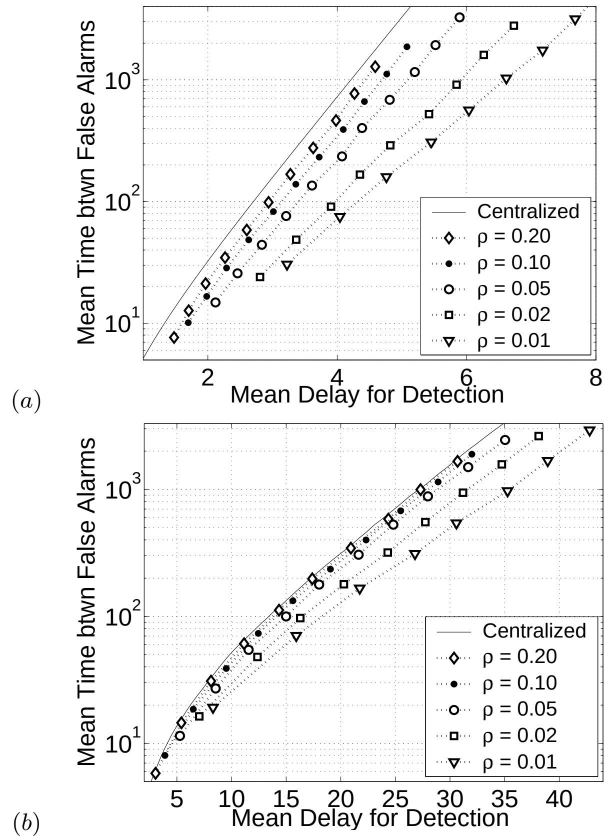

Fig. 2. Simulated performance of a 3-sensor hierarchical censoring change detector for values of $\rho = \{0.20, 0.10, 0.05, 0.02, 0.01\}$ in the Gaussian change-of-mean example with (a) SNR = 1 and (b) SNR = 0.1. Centralized ($\rho = 1$) detector is shown (—–) in (a) using Siegmund approximation and in (b) using simulation results.

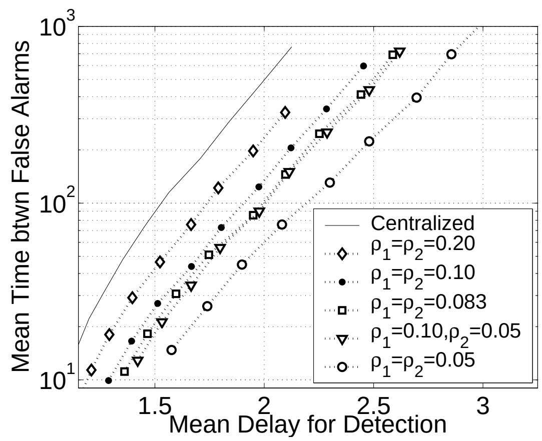

Fig. 3. Simulated performance of a 7-sensor hierarchical censoring change detector for values of $\rho = \{0.20, 0.10, 0.083, 0.05\}$ in the Gaussian change-of-mean example with SNR = 1. Simulated centralized ($\rho = 1$) detector performance is also shown (—–).

$(1/\mathrm{E}_0[t_a^{(1)}] + 1/\mathrm{E}_0[t_a^{(2)}])/2 \leq \rho$. Although it is not necessarily optimal, we meet the communications constraint by setting $\lambda^{(n)}$, $n = \{1, 2\}$, such that

$$\mathrm{E}_0[t_a^{(1)}] = \mathrm{E}_0[t_a^{(2)}] = 1/\rho. \tag{10}$$

We calculate using the Siegmund approximation [11] the correct $\lambda^{(1)} = \lambda^{(2)}$ so that (10) is met, first for $\rho = 0.20$, and in subsequent simulations, $\rho = 0.10, 0.05, 0.02$, and 0.01. Next, for a range of fusion center thresholds, ie., $\lambda^{(3)}$ from 0.5 to 5.5 in 0.5 increments, we simulate the hierarchical censored CUSUM detector, first, under H_0, and then, under H_1. In each run, we save the value of the stopping time, $t_a^{(3)}$. We run this simulation for 10^4 trials. We estimate $\bar{T}^{(3)}$ to be the average of all of $t_a^{(3)}$ simulated under H_0 and $\tau^{(3)}$ to be the average of all of $t_a^{(3)}$ simulated under H_1, and plot the pair in Fig. 2(a). We note that at worst, the hierarchical censored CUSUM test with only 5% probability of transmission incurs a delay of about one additional sample compared to the centralized CUSUM test. Also, at $\rho = 0.01$, the mean detection delay increases by a factor of 50% compared to the centralized case.

We also test the case where SNR = 0.1 (-10 dB) by setting $\sigma = \sqrt{10}$. Due to the inaccuracy of the Siegmund approximation at low SNR, we use simulation results to set $\lambda^{(1)} = \lambda^{(2)}$ so that (10) is met, for the same values of ρ as the previous case. Here, $\lambda^{(3)}$ ranges from 0.5 to 5 in 0.5 increments. The results are shown in Fig. 2(b).

3.2. Seven-Sensor Example

Next, we consider the 7-sensor case depicted in Fig. 1). Although it may not be the optimal solution, we constrain sensors on the same the level of the tree to have the same $\mathrm{E}_0[t_a^{(n)}]$. Thus the communications constraint in (9) becomes

$$\frac{1}{6}\left(\frac{4}{\mathrm{E}_0[t_a^{(L_1)}]} + \frac{2}{\mathrm{E}_0[t_a^{(L_2)}]}\right) \leq \rho$$

for any sensor L_1 on level 1 ($\{1, 2, 4, 5\}$ in Fig. 1) and sensor L_2 on level 2 ($\{3, 6\}$ in Fig. 1). First, we test two combinations of $\mathrm{E}_0[t_a^{(L_1)}]$ and $\mathrm{E}_0[t_a^{(L_2)}]$ which result in $\rho = 1/12$: (a) $\mathrm{E}_0[t_a^{(L_1)}] = 10$ and $\mathrm{E}_0[t_a^{(L_2)}] = 20$; and (b) $\mathrm{E}_0[t_a^{(L_1)}] = \mathrm{E}_0[t_a^{(L_2)}] = 12$. Here, $\lambda^{(7)}$ ranges from 0.5 to 4.5 in 0.5 increments. The results, shown in Fig. 3, show combination (b) slightly outperforms combination (a). Similarly, in [10] we found that setting the probability of transmission the same on all levels was nearly optimal for a wide range of parameters. This is also intuitively desirable in order to ensure constant energy consumption across all sensors in the network. Thus for our further simulations, we set $\lambda^{(L_1)}$ and $\lambda^{(L_2)}$ so that $\mathrm{E}_0[t_a^{(L_1)}] = \mathrm{E}_0[t_a^{(L_2)}] = 1/\rho$, for $\rho \in \{0.20, 0.05\}$. The value of $\lambda^{(L_1)}$ is calculated from the Siegmund approximation, while $\lambda^{(L_2)}$ is estimated from simulations. Again, we run 10^4 trials each under H_0 and H_1. Since $FC = 7$, we plot $\mathrm{E}_0[t_a^{(7)}]$ and $\mathrm{E}_1[t_a^{(7)}]$ for each ρ in Fig. 3.

4. CONCLUSION

We have demonstrated the potential of hierarchical censoring to dramatically reduce transmissions in a quickest detection sensor network at the price of a minor increase in overall detection delay. This research motivates further study of censoring in wireless sensor networks for detecting changes. Future research must consider the amount of data that is transmitted when a child node does decide to report to its parent. It may be possible to significantly reduce the amount of data transmitted without impacting performance. Furthermore, although results are already promising, no attempt has been made to derive an optimal distributed test. Analytical results would greatly enhance the portability of results to non-Gaussian change detection problems.

5. REFERENCES

[1] Jan M. Rabaey, M. Josie Ammer, Julio L. da Silva, Jr., Danny Patel, and Shad Roundy, "Picorodio supports ad hoc ultra-low power wireless networking," *IEEE Computer*, vol. 33, no. 7, pp. 42–48, July 2000.

[2] Anantha Chandrakasan, Rajeevan Amirtharajah, SeongHwan Cho, James Goodman, Gangadhar Konduri, Joanna Kulik, Wendi Rabiner, and Alice Wang, "Design considerations for distributed microsensor systems," in *IEEE Custom Integrated Circuits Conf.*, May 1999, pp. 279–286.

[3] Maurizio Longo, Tom D. Lookabaugh, and Robert M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. on Inform. Theory*, vol. 36, no. 2, pp. 241–255, Mar. 1990.

[4] S. Sandeep Pradhan, Julius Kusuma, and Kannan Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Mag.*, pp. 51–60, Mar. 2002.

[5] Jean-Francois Chamberland and Venugopal V. Veeravalli, "Decentralized detection in wireless sensor networks," in *2002 Conf. on Info. Sciences and Systems*, Princeton University, Mar. 2002.

[6] Vijay Raghunathan, Curt Schurgers, Sung Park, and Mani B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Mag.*, pp. 40–50, Mar. 2002.

[7] Teerasit Kasetkasem and Pramod K. Varshney, "Communication structure planning for multisensor detection systems," *IEE Proc. Radar, Sonar Navig.*, vol. 148, no. 1, pp. 2–8, Feb. 2001.

[8] Constantino Rago, Peter Willett, and Yaakov Bar-Shalom, "Censoring sensors: a low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.

[9] Swaroop Appadwedula, Venugopal V. Veeravalli, and Douglas L. Jones, "Robust and locally-optimum decentralized detection with censoring sensors," in 5^{th} *Int. Conf. on Information Fusion*, July 2002, vol. 1, pp. 56–63.

[10] Neal Patwari and Alfred O. Hero III, "Hierarchical censoring for distributed detection in wireless sensor networks," in *2003 IEEE ICASSP*, April 2003, vol. 4, pp. 848–851.

[11] Michèle Basseville and Igor V. Nikiforov, *Detection of Abrupt Changes: theory and application*, Prentice Hall, New Jersey, 1993.

[12] E.S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[13] G. Lorden, "Procedures for reacting to a change in distribution," *Annals Math. Statistics*, vol. 42, pp. 1897–1908, 1971.