Quickest Anomaly Detection: A Case of Active Hypothesis Testing

Kobi Cohen, Qing Zhao
Department of Electrical and Computer Engineering,
University of California, Davis, CA 95616
{yscohen, qzhao}@ucdavis.edu

***Abstract*— The problem of quickest detection of an anomalous process among M processes is considered. At each time, a subset of the processes can be observed, and the observations follow two different distributions, depending on whether the process is normal or abnormal. The objective is a sequential search strategy that minimizes the expected detection time subject to an error probability constraint. This problem can be considered as a special case of active hypothesis testing first considered by Chernoff in 1959, where a randomized test was proposed and shown to be asymptotically optimal. For the special case considered in this paper, we show that a simple deterministic test achieves asymptotic optimality and offers better performance in the finite regime.**

***Index Terms*—** Sequential detection, hypothesis testing, dynamic search.

I. Introduction

We consider the problem of detecting a single anomalous process among M processes. Borrowing terminologies from target search, we refer to these processes as cells and the anomalous process as the target which can locate in any of the M cells. The decision maker is allowed to search the target over K cells at a time ($1 \leq K \leq M$). The observations from searching a cell are realizations drawn from two different distributions f or g, depending on whether the target is absent or present. The objective is a sequential search strategy that dynamically determines which cells should be searched at each time and when to terminate the search so that the expected detection time is minimized under a constraint on the error probability.

The problem under study applies to intrusion detection in cyber-systems in cases where an intrusion to a subnet has been detected and the probability of each component being compromised is small (thus with high probability, there is only one abnormal component). It also finds applications in target search, fraud detection, and spectrum scanning in cognitive radio networks.

A. A Case of Active Hypothesis Testing

The above anomaly detection problem can be considered as a special case of the sequential design of experiments problem first studied by Chernoff in 1959 [1], in which the decision maker chooses and dynamically changes the experiment (thus the observation model) among a set of available experiments. Chernoff focused on the case of binary hypotheses and showed that a randomized strategy (referred to as the Chernoff test) is asymptotically optimal as the maximum error probability diminishes. Specifically, the Chernoff test chooses the current experiment based on a distribution that depends on the past actions and observations. Variations and extensions of the problem and the Chernoff test were studied in [2]–[5], where the problem was referred to as controlled sensing for hypothesis testing in [3], [4] and active hypothesis testing in [5] (see Sec. I-C for a more detailed discussion on [3]–[5]).

It is not difficult to see that the quickest anomaly detection problem considered in this paper is a special case of the active hypothesis testing problem considered in [1]–[5]. In particular, under each hypothesis that the target is located in the i^{th} ($i = 1, 2, \ldots, M$) cell, the distribution (either f or g) of the next observation depends on the action of which cell to search (i.e., which experiment to carry out). The Chernoff test (with the extension proposed in [3], [4] to allow indistinguishable hypotheses under some but not all available experiments) thus directly applies to our problem. However, in contrast to the randomized nature of the Chernoff test, we show in this paper that a simple deterministic test achieves asymptotic optimality and offers better performance in the finite regime.

B. Main Results

Similar to [1]–[5], we focus on asymptotically optimal policies in terms of minimizing the detection time as the error probability approaches zero. A Bayesian approach is adopted, assigning a cost of c per observation and a loss of 1 for wrong decisions. Differing from the randomized Chernoff test, we propose a deterministic closed-loop selection rule wherein the selected cells are determined (and are not drawn randomly) by the past observations and actions. Specifically, the selection rule $\phi(n)$ which indicates which K cells should be observed at time n is given by:

This work was supported by Army Research Lab under Grant W911NF1120086.

$$\phi(n) = \begin{cases} \left(m^{(1)}(n), m^{(2)}(n), ..., m^{(K)}(n)\right) \quad , \\ \qquad\qquad \text{if} \;\; D(g||f) \geq \frac{D(f||g)}{(M-1)} \text{ or } K = M \\ \left(m^{(2)}(n), m^{(3)}(n), ..., m^{(K+1)}(n)\right) \quad , \\ \qquad\qquad \text{if} \;\; D(g||f) < \frac{D(f||g)}{(M-1)} \text{ and } K < M \end{cases}$$

where $m^{(i)}(n)$ denotes the cell index with the i^{th} highest sum of log-likelihood ratio collected from the cell up to time n, and $D(\cdot||\cdot)$ is the Kullback-Liebler (KL) divergence between two distributions.

It is shown that the proposed policy is asymptotically optimal in terms of achieving the lower bound on the Bayes risk as c approaches zero. Furthermore, it is simple to implement and achieves significant performance gain over the Chernoff test as illustrated via simulations.

C. Related Work

As discussed in Sec. I-A, the above anomaly detection problem can be considered as a special case of an active hypothesis testing first studied by Chernoff [1]. Chernoff focused on the case of sequential binary composite hypothesis testing and showed the asymptotic optimality of the Chernoff test. Extensions to M-ary hypotheses (where M is fixed) were done in [2]–[4]. Specifically, by applying the Chernoff test, the action u at time n is drawn from a distribution $q(u)$ that depends on the past actions and observations:

$$q(u) = \arg \max_{\overline{q}(u)} \min_{j \in \mathcal{M} \setminus \{\hat{i}(n)\}} \sum_u \overline{q}(u) D(p^u_{\hat{i}(n)} || p^u_j) \; ,$$

where $\mathcal{M}$ is the set of the M hypotheses, $\hat{i}(n)$ is the ML estimate of the hypothesis at time n, and p^u_j is the observation distribution under hypothesis j when action u is taken. The original Chernoff test proposed in [1] (and the extended test in [2]) requires that under any action, any pair of hypotheses are distinguishable (i.e., has positive KL divergence). In [3], [4], an extended Chernoff test was proposed to allow zero KL divergence between hypotheses under some but not all actions. In [5], the problem was studied under a non-zero information rate, where the number of hypotheses approaches infinity.

Optimal solutions for the sequential target search problem were studied in [6]–[9] for some special cases. Optimal policies were derived for the problem of quickest search over Weiner processes under the model that a single process is observed at a time [6]–[8]. It was shown in [6], [7] that the optimal policy is to select the process with the highest posterior probability of being the target at each given time. In [8], a simple SPRT-based solution was derived, which is equivalent to the optimal policy in the case of searching over Weiner processes. However, the optimal policy for general distributions or when multiple processes are observed at a time remained an open question. In this paper we address these questions under the asymptotic regime, as the error probability approaches zero.

Another related problem is the whereabouts search, which is often considered under the setting of fixed sample size as in [10]–[13]. In [10], [11], [13], searching in a specific location provides a binary-valued measurement regarding the presence or absent of the target. In [12], Castanon considered the dynamic search problem under continuous observations: the observations from a location without the target and with the target have distributions f and g, respectively. The optimal policy was established under a symmetry assumption that $f(x) = g(b - x)$ for some b. The problem of universal outlier hypothesis testing was studied in [14]. Under this setting, a vector of observations containing coordinates with an outlier distribution is observed at each given time. The goal is to detect the coordinates with the outlier distribution based on a sequence of n independent and identically distributed (i.i.d) vectors of observations.

Differing from the search problem, sequential detection involving independent processes have been considered in [15]–[22]. In [15], [16], the problem of quickly detecting an idle period over multiple independent ON/OFF processes was considered. An optimal threshold policy was derived in [16]. In [17], the problem of quickest detection of idle channels over K independent channels with fixed idle/busy state was studied. It was shown that the optimal policy is to carry out an independent SPRT over each channel, irrespective of the testing order. In [18], [19], optimal index probing strategies were derived for the anomaly localization problem, where the objective is to minimize the expected cost incurred by the abnormal sequences. In [20], the problem of identifying the first abnormal sequence among an infinite number of i.i.d sequences was considered. An optimal cumulative sum (CUSUM) test was established under this setting. Further studies on this model can be found in [21], [22].

II. Problem Formulation

Consider the following anomaly detection problem. A decision maker is required to detect the location of a single anomalous object (referred as a target) located in one of M cells. If the target is in cell m, we say that hypothesis H_m is true. The *a priori* probability that H_m is true is denoted by π_m, where $\sum_{m=1}^{M} \pi_m = 1$. To avoid trivial solutions, it is assumed that $0 < \pi_m < 1$ for all m.

At each time, only K $(1 \leq K \leq M)$ cells can be observed. When cell m is observed at time n, an observation (or a vector of observations) $y_m(n)$ is independently drawn from a distribution $f_m(y)$ in a one-at-a-time manner. If hypothesis m is false, $y_m(n)$ follows distribution $f_m(y) = f(y)$; if hypothesis m is true, $y_m(n)$ follows distribution $f_m(y) = g(y)$. Let $\mathbf{P}_m$ be the probability measure, corresponding to distribution f_m and $\mathbf{E}_m$ be the operator of expectation with respect to the measure $\mathbf{P}_m$.

We define the stopping rule τ as the time when the decision maker finalizes the search (i.e., delay) by declaring the location of the target. Let $\delta \in \{1, 2, ..., M\}$ be a decision rule, where $\delta = m$ if the decision maker declares that H_m is true. Let $\phi(n) \in \{1, 2, ..., M\}^K$ be a selection rule indicates which K cells are chosen to be observed at time n. The time series vector of selection rules is denoted by $\boldsymbol{\phi} = (\phi(n), n = 1, 2, ...)$. Let $\mathbf{y}(t) = \left\{\phi(i), \mathbf{y}_{\phi(i)}\right\}_{i=1}^{t}$ be the set of all the available observations and the cell indices up to time t. A selection rule is a mapping from $\mathbf{y}(t-1)$ to $\{1,2,...,M\}^K$.

Definition 1: An admissible strategy Γ for the sequential search problem is given by the tuple $\Gamma=(\tau,\delta,\boldsymbol{\phi})$.

Let $P_e(\Gamma)=\sum_{m=1}^{M}\pi_m\alpha_m(\Gamma)$ be the probability of error, achieved by Γ, where $\alpha_m(\Gamma)=\mathbf{P}_m(\delta\neq m|\Gamma)$ is the probability to decide $\delta\neq m$ when H_m is true. Let $\mathbf{E}(\tau|\Gamma)=\sum_{m=1}^{M}\pi_m\mathbf{E}_m(\tau|\Gamma)$ be the average delay, achieved by Γ.

We adopt a Bayesian approach, as was done in [1], [3], by assigning a cost of c for each observation and a loss of 1 for wrong decisions and 0 otherwise. The Bayes risk under $\mathbf{P}_m$, achieved by policy Γ, is defined by:

$$R_m(\Gamma)\triangleq\alpha_m(\Gamma)+c\mathbf{E}_m(\tau|\Gamma)\ . \tag{1}$$

Note that c represents the ratio of the sampling cost to the cost due to wrong decisions.
The average Bayes risk is given by:

$$R(\Gamma)=\sum_{m=1}^{M}\pi_m R_m(\Gamma)=P_e(\Gamma)+c\mathbf{E}(\tau|\Gamma)\ . \tag{2}$$

The problem is to find a strategy Γ that minimizes the Bayes risk $R(\Gamma)$:

$$\inf_{\Gamma}\ \ R(\Gamma)\ . \tag{3}$$

III. An Asymptotically Optimal Deterministic Index Policy

In this section we propose a deterministic index policy to solve (3). In subsequent sections we show that the proposed policy is asymptotically optimal in terms of minimizing the Bayes risk (2) as $c\rightarrow 0$.

Let $\mathbf{1}_m(n)$ be the indicator function, where $\mathbf{1}_m(n)=1$ if cell m is observed at time n, and $\mathbf{1}_m(n)=0$ otherwise. Let

$$\ell_m(n)\triangleq\mathbf{1}_m(n)\log\frac{g(y_m(n))}{f(y_m(n))}\ , \tag{4}$$

and

$$S_m(n)\triangleq\sum_{i=1}^{n}\ell_m(i) \tag{5}$$

be the observed log-likelihood ratio (LLR) at time n and the sum of the observed LLRs up to time n of cell m, respectively.
Let

$$\Delta S_{m,j}(n)\triangleq S_m(n)-S_j(n)\ , \tag{6}$$

be the difference between the the observed sum of LLRs of cells m and j.
We further define

$$\Delta S_m(n)\triangleq\min_{j\neq m}\Delta S_{m,j}\ , \tag{7}$$

and

$$\Delta S(n)\triangleq\max_{m}\Delta S_m(n)\ . \tag{8}$$

The following recursive formula for $m^{(i)}(n)$ describes the index of the cell with the i^{th} highest sum of LLRs:

$$\begin{aligned}&m^{(1)}(n)\triangleq\arg\ \max_{m}\ S_m(n)\ ,\\&m^{(i)}(n)\triangleq\arg\max_{m\notin\left\{m^{(r)}(n)\right\}_{r=1}^{i-1}}S_m(n)\ \ ,\ \ i=2,...,M\ .\end{aligned} \tag{9}$$

A. The policy

At each time n, the proposed policy applies a simple deterministic index selection rule, where the observed cells are determined according to their indices $S_j(n)$. Specifically, the selection rule is given by:

$$\phi(n)=\begin{cases}\left(m^{(1)}(n),m^{(2)}(n),...,m^{(K)}(n)\right)\ \ ,\\ \qquad\qquad\text{if}\ \ D(g||f)\geq\frac{D(f||g)}{(M-1)}\ \text{or}\ K=M\\ \left(m^{(2)}(n),m^{(3)}(n),...,m^{(K+1)}(n)\right)\ \ ,\\ \qquad\qquad\text{if}\ \ D(g||f)<\frac{D(f||g)}{(M-1)}\ \text{and}\ K<M\end{cases}\ , \tag{10}$$

where the KL divergences $D(g||f)>0, D(f||g)>0$ are assumed to be strictly positive (otherwise, the error probability does not approach zero).
The stopping rule and terminal decision rule are given by:

$$\tau=\inf\left\{n\ :\ \Delta S(n)\geq-\log c\right\}\ , \tag{11}$$

and

$$\delta=m^{(1)}(\tau)\ . \tag{12}$$

B. Performance Analysis

In this section we analyze the asymptotic performance of the proposed deterministic index policy as $c\rightarrow 0$. Define

$$I^*(M,K)\triangleq\begin{cases}D(g||f)+D(f||g)\ \ ,\ \ \text{if}\ \ K=M\\ \max\left[\dfrac{KD(f||g)}{M-1},D(g||f)+\dfrac{(K-1)D(f||g)}{M-1}\right]\ \ ,\\ \qquad\qquad\qquad\qquad\text{if}\ K<M\end{cases} \tag{13}$$

The following theorem shows that the proposed policy is asymptotically optimal in terms of minimizing the Bayes risk as c approaches zero:

Theorem 1: Let R^* and $R(\Gamma)$ be the Bayes risks, achieved by the proposed deterministic index policy and any other policy Γ, respectively. Then,

$$R^*\ \sim\ \frac{-c\log c}{I^*(M,K)}\ \sim\ \inf_{\Gamma}\ R(\Gamma)\quad\text{as}\quad c\rightarrow 0\ . \tag{14}$$

The proof is given in the extended version of this paper [23].

C. Comparison to Chernoff's Test

As discussed in Section I-A, the search problem can be considered as a special case of an active hypothesis testing studied in [1]–[5], and the Chernoff test and its variations considered in [1]–[5] directly apply. Specifically, the Chernoff test when applied to the anomaly detection problem with $K = 1$ works as follows: select cell $\phi(n) = m^{(1)}(n)$ if $D(g||f) \geq D(f||g)/(M-1)$ (as the proposed deterministic index policy does) or to draw $\phi(n) = m^{(j)}(n)$ for $j \neq m^{(1)}(n)$ from a uniform distribution (with probability $1/(M-1)$ for each cell) if $D(g||f) < D(f||g)/(M-1)$. On the other hand, the proposed deterministic index policy surely selects $\phi(n) = m^{(2)}(n)$ if $D(g||f) < D(f||g)/(M-1)$.

IV. Numerical Examples

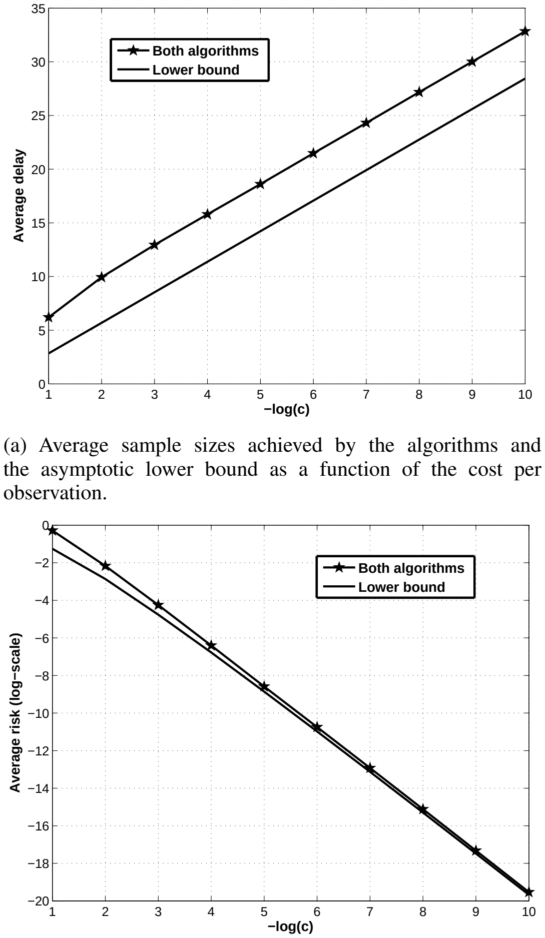
In this section we present numerical examples to illustrate the performance of the proposed deterministic index policy. We simulated a single anomalous object (i.e., target) located in one of M cells with the following parameters: The *a priori* probability that the target is present in cell m was set to $\pi_m = 1/M$ for all $1 \leq i \leq M$. When cell m is observed at time n, an observation $y_m(n)$ is independently drawn from a distribution $f \sim \exp(\lambda_f)$ or $g \sim \exp(\lambda_g)$, depending on whether the target is absent or present, respectively. It can be verified that:

$$D(g||f) = \log(\lambda_g) - \log(\lambda_f) + \frac{\lambda_f}{\lambda_g} - 1$$

$$D(f||g) = \log(\lambda_f) - \log(\lambda_g) + \frac{\lambda_g}{\lambda_f} - 1 \; .$$

We compared two schemes: 1) The proposed deterministic index policy; 2) The Chernoff test as discussed in Sec. III-C. We consider the case where $M = 5$ and $K = 1$. Note that the proposed deterministic index policy and the Chernoff test select cell $m^{(1)}(n)$ only if $D(g||f) \geq D(f||g)/(M-1)$. Otherwise, the proposed policy selects cell $m^{(2)}(n)$, while the Chernoff test selects a cell $j \neq m^{(1)}(n)$ randomly at each given time n. We expect both schemes to approach the asymptotic lower bound as $c \to 0$. First, we set $\lambda_f = 2, \lambda_g = 10$. In this case $D(g||f) \approx 0.8, D(f||g)/(M-1) \approx 0.6$. As a result, both algorithms select cell $m^{(1)}(n)$ at each given time n. The performance of the algorithms (which perform the same in this case) are presented in Fig. 1(a), 1(b). In Fig. 1(a), the asymptotic lower bound on the expected sample size and the average sample sizes achieved by the algorithms are presented as a function of the cost per observation c (log-scale). In Fig. 1(b), the asymptotic lower bound on the Bayes risk and the average Bayes risks achieved by the algorithms are presented as a function of c. It can be seen that the performance of the algorithms approach the lower bounds as $c \to 0$.

Next, we set $\lambda_f = 0.5, \lambda_g = 10$. In this case $D(g||f) \approx 2.05, D(f||g)/(M-1) \approx 4$. As a result, the Chernoff test and the proposed policy have different cell selection rules. The performance of the Algorithms are presented in Fig. 2(a), 2(b). In Fig. 2(a), the asymptotic lower bound on the expected sample size and the average sample sizes achieved by the algorithms are presented as a function of c. In Fig. 2(b), the asymptotic lower bound on the Bayes risk and the average Bayes risks achieved by the algorithms are presented as a function of c. It can be seen that the proposed policy significantly outperforms the Chernoff test in the finite regime. These results demonstrate the advantage of using a deterministic closed-loop selection rule applied by the proposed policy instead of randomization for the sequential search problem.



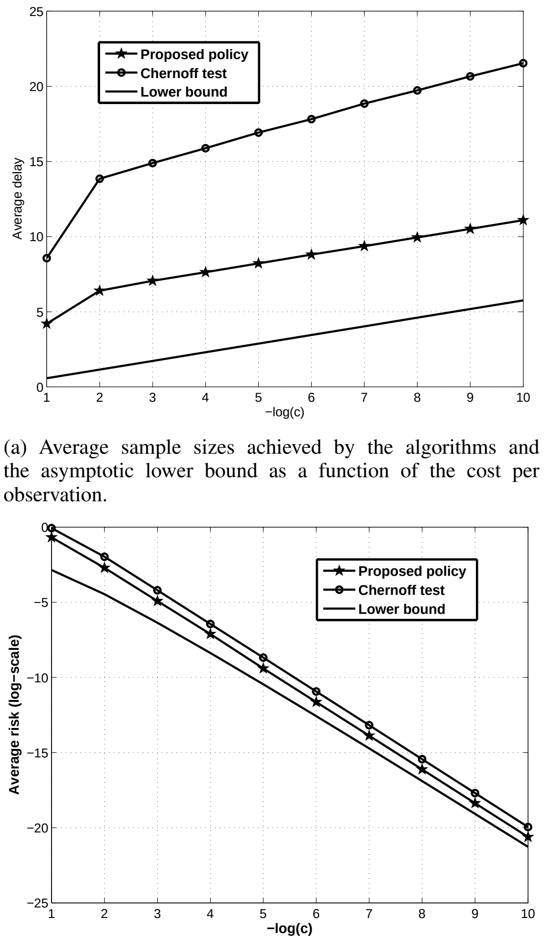
(a) Average sample sizes achieved by the algorithms and the asymptotic lower bound as a function of the cost per observation.

(b) Average Bayes risk achieved by the algorithms and the asymptotic lower bound as a function of the cost per observation.

Fig. 1. Algorithms' performance for $M = 5$, $K = 1$, $\lambda_f = 2, \lambda_g = 10$

V. Conclusion

The problem of sequential detection of a single anomalous object located in one of M cells was investigated. Due to resource constraints, only a subset of the cells can be observed at a time, The objective is a search strategy that minimizes the expected detection time subject to an error probability constraint. The observations from searching a cell are realizations drawn from two different distributions f, g, depending on whether the target is absent or present. A simple deterministic index policy was established to solve the Bayesian formulation of the search problem, where a cost of c per observation and a loss of 1 for wrong decisions are assigned. It was shown



(a) Average sample sizes achieved by the algorithms and the asymptotic lower bound as a function of the cost per observation.

(b) Average Bayes risk achieved by the algorithms and the asymptotic lower bound as a function of the cost per observation.

Fig. 2. Algorithms' performance for $M = 5$, $K = 1$, $\lambda_f = 0.5, \lambda_g = 10$

that the proposed policy is asymptotically optimal in terms of minimizing the Bayes risk as c approaches zero. Simulation results show significant performance gain of the proposed policy over existing methods.

References

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

[2] S. Bessler, "Theory and applications of the sequential design of experiments, k-actions and infinitely many experiments: Part I–Theory," *Tech. Rep. Applied Mathematics and Statistics Laboratories, Stanford University*, no. 55, 1960.

[3] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for hypothesis testing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5277–5280, 2012.

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

[5] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *Annals of Statistics*, vol. 41, no. 6, 2013.

[6] K. S. Zigangirov, "On a problem in optimal scanning," *Theory of Probability and Its Applications*, vol. 11, no. 2, pp. 294–298, 1966.

[7] E. Klimko and J. Yackel, "Optimal search strategies for Wiener processes," *Stochastic Processes and their Applications*, vol. 3, no. 1, pp. 19–33, 1975.

[8] V. Dragalin, "A simple and effective scanning rule for a multi-channel system," *Metrika*, vol. 43, no. 1, pp. 165–182, 1996.

[9] L. D. Stone and J. A. Stanshine, "Optimal search using uninterrupted contact investigation," *SIAM Journal on Applied Mathematics*, vol. 20, no. 2, pp. 241–263, 1971.

[10] K. P. Tognetti, "An optimal strategy for a whereabouts search," *Operations Research*, vol. 16, no. 1, pp. 209–211, 1968.

[11] J. B. Kadane, "Optimal whereabouts search," *Operations Research*, vol. 19, no. 4, pp. 894–904, 1971.

[12] D. A. Castanon, "Optimal search strategies in dynamic hypothesis testing," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 7, pp. 1130–1138, 1995.

[13] Y. Zhai and Q. Zhao, "Dynamic search under false alarms," *in Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2013.

[14] Y. Li, S. Nitinawarat, and V. V. Veeravalli, "Universal outlier hypothesis testing," *available at http://arxiv.org/abs/1302.4776*, 2013.

[15] H. Li, "Restless watchdog: Selective quickest spectrum sensing in multichannel cognitive radio systems," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, 2009.

[16] Q. Zhao and J. Ye, "Quickest detection in multiple on–off processes," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 5994–6006, 2010.

[17] R. Caromi, Y. Xin, and L. Lai, "Fast multiband spectrum scanning for cognitive radio systems," *IEEE Transaction on Communications*, vol. 61, no. 1, pp. 63–75, 2013.

[18] K. Cohen, Q. Zhao, and A. Swami, "Optimal index policies for quickest localization of anomaly in cyber networks," *in Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2013.

[19] K. Cohen, Q. Zhao, and A. Swami, "Optimal index policies for anomaly localization in resource-constrained cyber systems," *submitted to the IEEE Transaction on Signal Proccessing*, 2013.

[20] L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis, "Quickest search over multiple sequences," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5375–5386, 2011.

[21] M. L. Malloy, G. Tang, and R. D. Nowak, "Quickest search for a rare distribution," *IEEE Annual Conference on Information Sciences and Systems*, pp. 1–6, 2012.

[22] A. Tajer and H. V. Poor, "Quick search for rare events," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4462–4481, 2013.

[23] K. Cohen and Q. Zhao, "Active hypothesis testing for quickest anomaly detection," *to be submitted to the IEEE Transaction on Information Theory, 2014, available at http://www.ece.ucdavis.edu/~yscohen/.*