ACTIVE ANOMALY DETECTION IN HETEROGENEOUS PROCESSES

Boshuang Huang[⋆]*, Kobi Cohen*[‡]*, Qing Zhao*[⋆]

[⋆]School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853
Email: {bh467, qz16}@cornell.edu
[‡]Department of Electrical and Computer Engineering Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: yakovsec@bgu.ac.il

ABSTRACT

An active inference problem of detecting an anomalous process among M heterogeneous processes is considered. At each time, a subset of processes can be probed. The objective is to design a sequential probing strategy that dynamically determines which processes to observe at each time and when to terminate the search so that the expected detection time is minimized under a constraint on the probability of misclassifying any process. This problem falls into the general setting of sequential design of experiments pioneered by Chernoff in 1959, in which a randomized strategy, referred to as the Chernoff test, was proposed and shown to be asymptotically optimal as the error probability approaches zero. For the problem considered in this paper, a low-complexity deterministic test is shown to enjoy the same asymptotic optimality while offering significantly better performance in the finite regime and faster convergence to the optimal rate function, especially when the number of processes is large. Furthermore, the proposed test offers considerable reduction in implementation complexity.

***Index Terms*—** Active hypothesis testing, sequential design of experiments, anomaly detection, dynamic search.

1. INTRODUCTION

We consider the problem of detecting an anomalous process (referred to as the target) among M heterogeneous processes (referred to as the cells). At each time, K ($1 \leq K < M$) cells can be probed simultaneously to search for the target. Each search of cell i generates a noisy observation drawn i.i.d. over time from two different distributions f_i and g_i, depending on whether the target is absent or present. The objective is to design a sequential search strategy that dynamically determines which cells to probe at each time and when to terminate the search so that the expected detection time is minimized under a constraint on the probability of declaring a wrong location of the target.

The above problem is prototypical of searching for rare events in a large number of data streams or a large system. The rare events could be opportunities (e.g., financial trading opportunities or transmission opportunities in dynamic spectrum access [1]), unusual activities in surveillance feedings, frauds in financial transactions, attacks and intrusions in communication and computer networks, anomalies in infrastructures such as bridges, buildings, and the power grid that may indicate catastrophes. Depending on the application, a cell may refer to an autonomous data stream with a continuous data flow or a system component that only generates data when probed.

1.1. Main Results

The anomaly detection problem considered in this paper is a special case of active hypothesis testing originated from Chernoff's seminal work on sequential design of experiments in 1959 [2]. In [2], Chernoff proposed a *randomized* strategy, referred to as the Chernoff test, and established its asmyptotic (as the error probability diminishes) optimality. This randomized test chooses, at each time, a probability distribution that governs the selection of the experiment to be carried out at this time. This distribution is obtained by solving a minimax problem so that the next observation generated under the random action can best differentiate the current maximum likelihood estimate of the true hypothesis (using all past observations) from its closest alternative, where the closeness is measured by the Kullback-Liebler (KL) divergence. Due to the complexity in solving this minimax problem at each time, the Chernoff test can be expensive to compute and cumbersome to implement, especially when the number of hypotheses or the number of experiments is large.

In this paper, we show that the anomaly detection problem considered here exhibits sufficient structures to admit a low-complexity *deterministic* policy with strong performance. In particular, we develop a deterministic test that *explicitly* specifies which K cells to search at each given time and show that this test enjoys the same asymptotic optimality as the Chernoff test. Furthermore, extensive simulation examples have demonstrated a significant performance gain over the Chernoff test in the finite regime and faster convergence to the optimal rate function, especially when M is large. In contrast to the Chernoff test, the proposed test requires little offline or online computation.

We point out that proving the asymptotic optimality of the deterministic policy is much more involved comparing with the Chernoff test, due to the time dependency in the test statistics, namely, the log-likelihood ratios (LLRs), introduced by deterministic actions. In particular, since the distribution of the random action chosen by the Chernoff test depends only on the current maximum likelihood estimate of the underlying hypothesis which becomes time-invariant after an initial phase with a bounded duration, the stochastic behaviors of the LLRs are independent over time, resulting in a much easier analysis of the detection delay. The deterministic actions of the proposed test, however, lead to complex time dependencies in LLRs that make the analysis much more involved.

The work of B. Huang and Q. Zhao was supported by the U.S. Army Research Office under Grant W911NF-17-1-0464.

The work of K. Cohen was partially supported by the Cyber Security Research Center at Ben-Gurion University of the Negev.

1.2. Related Work

Chernoff's pioneering work on sequential design of experiments focuses on sequential binary composite hypothesis testing [2]. Variations and extensions of the problem were studied in [3–7], where the problem was referred to as controlled sensing for hypothesis testing in [4, 5] and active hypothesis testing in [6, 7]. As variants of the Chernoff test, the tests developed in [3–7] are all randomized tests.

There is an extensive literature on dynamic search and target whereabout problems under various scenarios, (see [8–15] and references therein). In [16, 17], the problem of quickly detecting anomalous components under the objective of minimizing system-wide cost incurred by all anomalous components was studied. The objective of minimizing operational cost as opposed to detection delay led to a different problem from the one considered in this paper. The readers are also referred to [18] for a comprehensive survey on the problem of detecting outlying sequences.

A prior study by Cohen and Zhao considered the problem for homogeneous processes (i.e., $f_i \equiv f$ and $g_i \equiv g$) [19]. This work builds upon this prior work and addresses the problem in heterogeneous systems where the absence distribution f_i and the presence distribution g_i are different across processes. Allowing heterogeneity significantly complicates the design of the test and the establishment of asymptotic optimality. Specifically, since each process has different observation distributions, the rate at which the state of a cell can be inferred is different across processes. Hence, the decision maker must balance the search time effectively among the observed processes, which makes both the algorithm design and the performance analysis much more involved under the heterogeneous case. In terms of algorithm design, when dealing with homogeneous processes, the search strategy is often static in nature [9, 11, 13, 19]. In contrast, the asymptotically optimal search strategy developed here for heterogeneous processes dynamically changes based on the current belief about the location of the target. In terms of performance analysis, handling heterogeneity adds new challenges and difficulties for establishing asymptotic optimality. When searching over homogeneous processes, the resulting rate function (which is inversely proportional to the search time) always obeys a certain averaging over the KL divergences between normal and abnormal distributions of all process. This observation follows from the fact that the decision maker completes gathering the required information from all the processes at approximately the same time due to the homogeneity. In contrast, when searching over heterogeneous processes, the overall rate function does not always obey a simple averaging across the KL divergences of all processes. Our approach to circumvent this difficulty is to analyze the detection time by considering two separate scenarios, referred to as the balanced and the unbalanced cases. The balanced case holds when a judicious allocation of probing resources can ensure the information gathering from all the processes be completed at approximately the same time, in which case the rate function is a weighted average among the heterogeneous processes. The unbalanced case occurs when there is a process with a sufficiently small KL divergence that dominates the overall rate function of the search.

Besides the active inference approach to anomaly detection considered in this paper, there is a growing body of literature on various approaches to the general problem of anomaly detection. We refer the readers to [20, 21] for comprehensive surveys on this topic.

2. PROBLEM FORMULATION

We consider the problem of detecting a single target located in one of M cells. Extensions to detecting multiple targets can be found in [22].

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.

When cell m is observed at time n, an observation $y_m(n)$ is drawn, independent of previous observations. If cell m contains a target, $y_m(n)$ follows distribution $g_m(y)$. Otherwise, $y_m(n)$ follows distribution $f_m(y)$. Let $\mathbf{P}_m$ be the probability measure under hypothesis H_m and $\mathbf{E}_m$ the operator of expectation with respect to the measure $\mathbf{P}_m$.

An active search strategy Γ consists of a stopping rule τ governing when to terminate the search, a decision rule δ for determining the location of the target at the time of stopping, and a sequence of selection rules $\{\phi(n)\}_{n\geq 1}$ governing which K cells to probed at each time n. Here we consider the case where only a single process can be observed at a time, i.e., $K = 1$ (See [22] for the extension to $K > 1$). Let $\mathbf{y}(n)$ be the set of all cell selections and observations up to time n. A deterministic selection rule $\phi(n)$ at time n is a mapping from $\mathbf{y}(n-1)$ to $\{1, 2, ..., M\}$. A randomized selection rule $\phi(n)$ is a mapping from $\mathbf{y}(n-1)$ to probability mass functions over $\{1, 2, ..., M\}$.

The error probability under policy Γ is defined as $P_e(\Gamma) = \sum_m \pi_m \alpha_m(\Gamma)$, where $\alpha_m(\Gamma) = \mathbf{P}_m(\delta \neq m|\Gamma)$ is the probability of declaring $\delta \neq m$ when H_m is true. Let $\mathbf{E}(\tau|\Gamma) = \sum_{m=1}^{M} \pi_m \mathbf{E}_m(\tau|\Gamma)$ be the average detection delay under Γ.

We adopt a Bayesian approach as in Chernoff's original study [2] by assigning a cost of c for each observation and a loss of 1 for a wrong declaration. Note that c represents the ratio of the sampling cost to the cost of wrong detections. The Bayes risk under strategy Γ when hypothesis H_m is true is given by:

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

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

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

A strategy Γ^* is *asymptotically optimal* if

$$\lim_{c\to 0} \frac{R(\Gamma^*)}{\inf_\Gamma R(\Gamma)} = 1, \tag{4}$$

which is denoted as

$$R(\Gamma^*) \sim \inf_{\Gamma} R(\Gamma). \tag{5}$$

3. THE DETERMINISTIC DGFi POLICY

In this section we propose a deterministic policy, referred to as the DGFi policy.

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

$$\ell_m(n) \triangleq \log \frac{g_m(y_m(n))}{f_m(y_m(n))}\ , \tag{6}$$

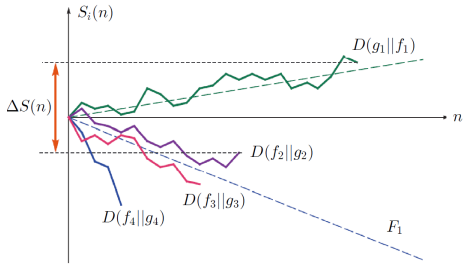

Fig. 1: Typical sample paths of sum LLRs.

and

$$S_m(n) \triangleq \sum_{t=1}^{n} \ell_m(t)\mathbf{1}_m(t) \tag{7}$$

be the log-likelihood ratio (LLR) and the observed sum LLRs of cell m at time n, respectively. Let $D(g||f)$ denote the KL divergence between two distributions g and f given by

$$D(g||f) \triangleq \int_{-\infty}^{\infty} \log \frac{g(x)}{f(x)} g(x)\, dx. \tag{8}$$

Illustrated in Fig. 1 are typical sample paths of the sum LLRs of $M = 4$ cells, where, without loss of generality, we assume that cell 1 is the target. Note that the sum LLR of cell 1 is a random walk with a positive expected increment $D(g_1||f_1)$, whereas the sum LLR of cell i is a random walk with a negative expected increment $-D(f_i||g_i)$ for $i = 2, 3, 4$. Thus, when the gap between the largest sum LLR and the second largest sum LLR is sufficiently large, we can declare with sufficient accuracy that the cell with the largest sum LLR is the target. This is the intuition behind the stopping rule and the decision rule. Specifically, we define $m^{(i)}(n)$ as the index of the cell with the i^{th} largest observed sum LLRs at time n. Let

$$\Delta S(n) \triangleq S_{m^{(1)}(n)}(n) - S_{m^{(2)}(n)}(n) \tag{9}$$

denote the difference between the largest and the second largest observed sum LLRs at time n. The stopping rule and the decision rule under the DGFi policy are given by:

$$\tau = \inf\{n \ : \ \Delta S(n) \geq -\log c\} \ , \tag{10}$$

and

$$\delta = m^{(1)}(\tau) \ . \tag{11}$$

We now specify the selection rule of the DGFi policy. The intuition behind the selection rule is to select a cell from which the observation can increase $\Delta S(n)$ at the fastest rate. The selection rule is thus given by comparing the rate at which $S_{m^{(1)}}(n)$ increases with the rate at which $S_{m^{(2)}(n)}(n)$ decreases. If $S_{m^{(1)}(n)}$ is expected to increase faster than $S_{m^{(2)}(n)}(n)$ decreases, cell $m^{(1)}(n)$ is chosen. Otherwise, cell $m^{(2)}(n)$ is chosen. This leads to the following selection rule:

$$\phi(n) = \begin{cases} m^{(1)}(n), & \text{if } D(g_{m^{(1)}(n)}||f_{m^{(1)}(n)}) \geq F_{m^{(1)}(n)} \\ m^{(2)}(n), & \text{otherwise} \end{cases} , \tag{12}$$

where

$$F_m \triangleq \frac{1}{\sum_{j\neq m} \frac{1}{D(f_j||g_j)}}. \tag{13}$$

The selection rule in (12) can be intuitively understood by noticing that $D(g_{m^{(1)}(n)}||f_{m^{(1)}(n)})$ is the asymptotic increasing rate of $S_{m^{(1)}}(n)$ when cell $m^{(1)}$ is probed at each time. This is due to the fact that $m^{(1)}(n)$ is the true target after an initial phase (defined by the last passage time that $m^{(1)}(n)$ is an empty cell) which can be shown to have a bounded expected duration. Similarly, even though much more involved to prove, $F_{m^{(1)}(n)}$ is the asymptotic rate at which $S_{m^{(2)}(n)}(n)$ decreases when cell $m^{(2)}(n)$ is probed at each time. To see the expression of F_m for any m as given in (13), consider the following analogy. Consider $M-1$ cars being driven by a single driver from 0 to $-\infty$. Car j ($j = 1, \ldots, M$, $j \neq m$) has a constant speed of $D(f_j||g_j)$. At each time, the car closest to the origin is chosen by the driver and driven by one unit of time. We are interested in the average moving speed of the position of the closest car to the origin. It is not difficult to see that it is given by F_m in (13). This analogy, concerned with deterministic processes, only serves as an intuitive explanation for the expression of F_m. As detailed in Sec. 4, proving $F_{m^{(1)}(n)}$ to be the asymptotic decreasing rate of $S_{m^{(1)}(n)}(n)$ requires analyzing the trajectories of the M sum LLRs $\{S_m(n)\}_{m=1}^M$, which are stochastic processes with complex dependencies both in time and across processes.

4. PERFORMANCE ANALYSIS

In this section, we establish the asymptotic optimality of the DGFi policy. While the intuitive exposition of DGFi given in Sec. 3 may make its asymptotic optimality seem expected, constructing a proof is much more involved. In particular, bounding the detection time of DGFi requires analyzing the trajectories of the M stochastic processes $\{S_m(n)\}_{m=1}^M$ which exhibit complex dependencies both over time and across processes as induced by the deterministic selection rule.

Define

$$I_m \triangleq \max\{D(g_m||f_m), F_m\}, \tag{14}$$

which is the increasing rate of $\Delta S(n)$ under hypothesis H_m. For a given *a priori* distribution $\{\pi_m\}_{m=1}^M$ of the true hypothesis, define

$$I^* \triangleq \frac{1}{\sum_{m=1}^{M} \frac{\pi_m}{I_m}}. \tag{15}$$

As shown in Theorem 1 below, I^* is the optimal rate function of the Bayes risk.

Theorem 1. *Let R^* and $R(\Gamma)$ be the Bayes risks under the DGFi policy and an arbitrary policy Γ, respectively. Then,*

$$R^* \ \sim \ \frac{-c\log c}{I^*} \ \sim \ \inf_{\Gamma} \ R(\Gamma) \tag{16}$$

Proof. Here we provide a sketch of the proof. The detailed proof can be found in [22].

We first show that the proposed DGFi policy achieves a Bayes risk $-c\log c/I^*$ asymptotically. First, we show that when $\Delta S(\tau)$ is large, the probability of error is small, i.e. $P_e = O(c)$. As a result, by the definition of the Bayes risk, it suffices to show that the detection time is upper bounded by $-\log c/I^*$. By the definition of I^* in (15), it suffices to show that the detection time is upper bounded by $-\log c/I_m$ under hypothesis H_m. This analysis is carried out by considering the balanced and the unbalanced cases separately.

The balanced case holds when a judicious allocation of probing resources can ensure the information gathering from all the processes be completed at approximately the same time, in which case the rate function is a weighted average among the heterogeneous

processes. The unbalanced case occurs when there is a process with a sufficiently small KL divergence that dominates the overall rate function of the search. Combining this analysis with a lower bound on the Bayes risk $-c\log c/I^*$ as we show in [22] completes the proof.

□

5. COMPARISON WITH THE CHERNOFF TEST

In this section, we compare the performance of the proposed DGFi policy and the Chernoff test in terms of both computational complexity and sample complexity.

5.1. The Chernoff Test

The Chernoff test has a randomized selection rule. Specifically, let $q = (q_1, ..., q_\kappa)$ be a probability mass function over a set of κ available experiments $\{u_i\}_{i=1}^{\kappa}$ that the decision maker can choose from, where q_i is the probability of choosing experiment u_i. Note that in our case, $\kappa = \binom{M}{K}$. For a general M-ary active hypothesis testing problem, the action at time n under the Chernoff test is drawn from a distribution $q^*(n) = (q_1^*(n), ..., q_\kappa^*(n))$ that depends on the past actions and observations:

$$q^*(n) = \arg\max_{q} \min_{j\in\mathcal{M}\setminus\{\hat{i}(n)\}} \sum_{u_i} q_i D(p_{\hat{i}(n)}^{u_i} || p_j^{u_i}) \ , \qquad (17)$$

where $\mathcal{M}$ is the set of the M hypotheses, $\hat{i}(n)$ is the ML estimate of the true hypothesis at time n based on past actions and observations, and $p_j^{u_i}$ is the observation distribution under hypothesis j when action u_i is taken. The stopping rule and the decision rule are the same as in (10), (11).

5.2. Comparison in computational complexity

Here we compare the computational complexity of the proposed DGFi policy with the Chernoff test. We show that the Chernoff test can be expensive to compute especially when the number of processes or the number of experiments is large. In contrast to the Chernoff test, the DGFi policy requires little computation.

For the case of detecting a single target, computing the selection rule of Chernoff test defined in (17) requires solving M minimax problems, each corresponding to a particular value of the ML estimate $\hat{i}(n) \in \{1, \ldots, M\}$. One efficient way of solving minimax problems is through linear programming which takes polynomial time with respect to the number of variables and constraints. For this problem, however, the number of variables is $\binom{M}{K}$, which is not polynomial and can be exponential in M in the worst case.

The only computation involved in the selection rule of DGFi is (13), which requires M summations each with $M-1$ elements. As a result, the computational time is $O(M^2)$, which is polynomial in M and independent of K.

5.3. Comparison in sample complexity

Although both the Chernoff test and the DGFi policy are asymptotically optimal[1], we show below via simulation examples the significant performance gain of DGFi over the Chernoff test in the finite regime (i.e., when the sample cost c is bounded away from 0).

[1]While the assumption of positive KL divergence between every pair of hypotheses under every probing action as required in Chernoff's proof of asymptotic optimality does not hold here, it can be shown that Chernoff test preserves its asymptotic optimality for the problem at hand.

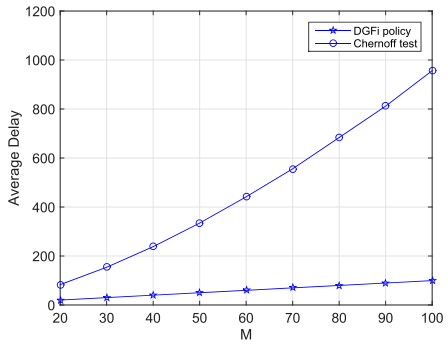

Fig. 2: Performance comparison ($K = 1, \lambda_g^{(m)} = 9 + m, \lambda_f^{(m)} = 0.0188, c = 10^{-5}$).

Consider a uniform prior and exponentially distributed observations: $f_m \sim \exp(\lambda_f^{(m)})$ and $g_m \sim \exp(\lambda_g^{(m)})$. The KL divergences can be easily computed as follows.

$$D(g_m||f_m) = \log(\lambda_g^{(m)}) - \log(\lambda_f^{(m)}) + \frac{\lambda_f^{(m)}}{\lambda_g^{(m)}} - 1 \ ,$$
$$D(f_m||g_m) = \log(\lambda_f^{(m)}) - \log(\lambda_g^{(m)}) + \frac{\lambda_g^{(m)}}{\lambda_f^{(m)}} - 1 \ .$$

Shown in Fig. 2 is the performance comparison between DGFi policy and Chernoff test for $K = 1$. More simulation examples for general cases with $K > 1$ and detecting multiple targets can be found in [22]. The figure clearly demonstrates the significant reduction in detection delay offered by the DGFi policy as compared with the Chernoff test. The performance gain increases drastically as M increases.

Next, we provide an intuition argument for the better finite-time performance of DGFi. Consider a special case where $K = 1$ and all f_i and g_i are identical, i.e., $f_i \equiv f$ and $g_i \equiv g$ and we assume $D(f||g) > (M-1)D(g||f)$. In this case, the DGFi policy chooses, at each time, the cell with the second largest sum LLR whereas the Chernoff test randomly and uniformly chooses a cell from all but the one with the largest sum LLR at each time. Consider a short horizon scenario where the sampling cost c is sufficiently high such that $D(f||g) > -\log c$. This means each empty cell only need one observation (with high probability) to distinguish from the true cell. We can formulate this as coupon collectors problem, where each empty cell is a coupon and the goal is to collect all $M-1$ coupons.

Since Chernoff test employs a randomized strategy that chooses empty cells with equal probability, based on results in coupon collectors problem, the expected probing time will be roughly $M\log M$. However, the proposed DGFi policy is deterministic and guaranteed to collect a new coupon at each time, therefore the expected probing time will only be M.

6. CONCLUSION

The problem of detecting anomalies among a large number of heterogeneous processes was considered. A low-complexity deterministic test was developed and shown to be asymptotically optimal. Its finite-time performance and computational complexity were shown to be superior to the classic Chernoff test for active hypothesis testing, especially when the problem size is large.

7. REFERENCES

[1] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE signal processing magazine*, vol. 24, no. 3, pp. 79–89, 2007.

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

[3] S. A. Bessler, "Theory and applications of the sequential design of experiments, k-actions and infinitely many experiments. part i. theory," tech. rep., DTIC Document, 1960.

[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] S. Nitinawarat and V. V. Veeravalli, "Controlled sensing for sequential multihypothesis testing with controlled markovian observations and non-uniform control cost," *Sequential Analysis*, vol. 34, no. 1, pp. 1–24, 2015.

[6] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, 2013.

[7] M. Naghshvar and T. Javidi, "Sequentiality and adaptivity gains in active hypothesis testing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 768–782, 2013.

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

[9] N. K. Vaidhiyan and R. Sundaresan, "Learning to detect an oddball target," *arXiv preprint arXiv:1508.05572*, 2015.

[10] K. Leahy and M. Schwager, "Always choose second best: Tracking a moving target on a graph with a noisy binary sensor," in *Control Conference (ECC), 2016 European*, pp. 1715–1721, IEEE, 2016.

[11] S. Nitinawarat and V. V. Veeravalli, "Universal scheme for optimal search and stop," in *Information Theory and Applications Workshop (ITA), 2015*, pp. 322–328, IEEE, 2015.

[12] B. Hemo, K. Cohen, and Q. Zhao, "Asymptotically optimal search of unknown anomalies," in *Proc. of the 16th IEEE Symposium on Signal Processing and Information Technology (ISSPIT)*, (Limassol, Cyprus), Dec. 2016.

[13] 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.

[14] 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.

[15] J. Heydari, A. Tajer, and H. V. Poor, "Quickest linear search over correlated sequences," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5786–5808, 2016.

[16] K. Cohen and Q. Zhao, "Asymptotically optimal anomaly detection via sequential testing," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2929–2941, 2015.

[17] K. Cohen, Q. Zhao, and A. Swami, "Optimal index policies for anomaly localization in resource-constrained cyber systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4224–4236, 2014.

[18] A. Tajer, V. V. Veeravalli, and H. V. Poor, "Outlying sequence detection in large data sets: A data-driven approach," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 44–56, 2014.

[19] K. Cohen and Q. Zhao, "Active hypothesis testing for anomaly detection," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1432–1450, 2015.

[20] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[21] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2014.

[22] B. Huang, K. Cohen, and Q. Zhao, "Active anomaly detection in heterogeneous processes," *arXiv preprint arXiv:1704.00766v2*.