Quickest Detection of a Dynamic Anomaly in a Sensor Network

Georgios Rovatsos[1], George V. Moustakides[2], and Venugopal V. Veeravalli[1]
University of Illinois at Urbana-Champaign[1]
University of Patras[2]
Emails: rovatso2@illinois.edu, moustaki@upatras.gr, vvv@illinois.edu

***Abstract*— We study the problem of detecting the emergence of a dynamic anomaly in sensor networks. The generated observations initially follow a pre-change distribution. At some unknown time, an anomaly appears, affecting a different set of nodes at each instant. The affected nodes generate data according to a post-change distribution. It is assumed that the trajectory of the anomaly is unknown. We propose a test that is optimal with respect to a measure of the expected delay for the worst-case trajectory. We compare the optimal test numerically with a test that uses the knowledge of the path of the anomaly and a heuristic test.**

***Index Terms*— Sensor networks, dynamic anomaly, quickest change detection, worst-path approach.**

I. Introduction

Quickest change detection (QCD) theory has been widely used as a tool to model and theoretically analyze detection problems where the goal is to quickly detect events that lead to a change in the distribution of sequentially observed processes. A selection of applications that can be modeled by this framework is given in [1]– [6]. In QCD, detection problems of interest are framed in an optimization framework where the goal is to design tests that minimize rigorously defined detection delays, subject to *false alarm* (FA) constraints. In the classical QCD problem [7], [8] two standard formulations are used i) the *minimax* setting [9]— [11], where the changepoint is considered to be deterministic but unknown and the goal is to minimize a worst-case average detection delay subject to a lower bound on the mean time to false alarm; and ii) the *Bayesian* setting [12], [13], where the changepoint is modeled as a random variable with a known distribution, and the goal is to minimize the average detection delay, subject to a bound on the probability of false alarm.

Timely detection of events of interest in sensor networks has been a topic of detailed study in the QCD literature. Significant work has been done in the case of detecting anomalies that are *static* [4], [13]–[18]. By the word static we refer to events that affect a fixed set of nodes after the changepoint, which might be unknown. Problems of detecting events that affect different sets of nodes as time progresses have attracted less attention. Some works that may be considered to fall in such a category include [19]–[23]. In these works, the sensors affected by the anomaly have their data generating distributions altered at different time instants, and this change in distribution is persistent at each node.

The crucial difference between these works and our work is that in our setting the anomaly is assumed to be *dynamic*, i.e., its effect may not be persistent in any specific node, but it is persistent if we view the entire network as a whole. This means that the anomaly is moving, implying that different sets of sensors may be affected at different time instants. This also implies that a sensor can change between the pre- and post-change generating mode as time progresses. In this work, we study the case of a dynamic anomaly of constant size, i.e., affecting a fixed number of nodes which is known to the decision maker. We study this QCD problem under Lorden's minimax framework [9]. To this end, we assume that the identities of the anomalous nodes are unknown but deterministic. To account for the lack of a specific model for the trajectory of the anomaly we modify Lorden's detection delay to evaluate candidate detection procedures with respect to the path of the anomaly that corresponds to the worst delay. Our work is related to [24] and [25] where the movement of the anomaly is modeled using a *discrete time Markov chain*.

For the specific case of homogeneous sensors, i.e., when the pre-change and post-change distributions are the same accross sensors, we establish that a Cumulative Sum (CuSum) test that detects a transition to a post-change mixture model that arises when the anomalous nodes are chosen uniformly at random is exactly optimal with respect to Lorden's [9] delay-FA formulation when the delay considers the worst-case anomaly path. Furthermore, we derive a first-order asymptotic approximation of the detection delay as the mean time to false alarm goes to infinity. Finally, we compare our proposed procedure to a heuristic detection scheme that is based on a detection statistic that grows after the anomaly emerges, and to an oracle test that exploits complete knowledge of the identities of the anomalous nodes.

II. Problem Model

Consider a network of L nodes denoted by $[L] \triangleq \{1,\dots,L\}$. Define by $\{\boldsymbol{X}[k]\}_{k=1}^{\infty}$ the measurements that are sampled from the sensors of the network, which become available sequentially to the decision maker. Here, $\boldsymbol{X}[k] = [X_1[k],\dots,X_L[k]]^\top$ denotes the observation vector at time k, where $X_\ell[k]$ is the measurement obtained by sensor $\ell \in [L]$ at time k. Let $\boldsymbol{X}[k_1,k_2] = [\boldsymbol{X}[k_1],\dots,\boldsymbol{X}[k_2]]^\top$ denote the observations sampled from time instant k_1 to

This work was sponsored in part by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196 (IoBT CRA), and in part by the National Science Foundation (NSF) under grant CCF 16-18658 and Grant CIF1513373, through Rutgers University.

$k_2 \geq k_1$. At some *unknown* but *deterministic* changepoint $\nu \geq 1$, a dynamic anomaly emerges in the network, affecting a set of $1 \leq m \leq L$ nodes which may be different at different time instants. The affected nodes generate data according to a post-change probability density function (pdf). We assume that m is known to the decision maker. We define by $\boldsymbol{S} \triangleq \{\boldsymbol{S}[k]\}_{k=1}^{\infty}$ the sequence characterizing the location of the anomalous nodes. Here, $\boldsymbol{S}[k]$ is a vector containing the m nodes which are anomalous at time k. Note that there can be $\binom{L}{m}$ different possible values for $\boldsymbol{S}[k]$. We define the set of these vector values by $\text{comb}([L], m)$. For notational convenience, $\boldsymbol{S}[k]$ is defined for all $k \geq 1$ and not simply for $k \geq \nu$.

Define by $g(\cdot)$ and $f(\cdot)$ the pre-change and post-change pdfs at a sensor. By post-change (pre-change) pdf we mean the pdf of a particular sensor's observations when it is (not) affected by the anomaly. For simplicity of the analysis, we assume that the pre- and post-change pdfs are the same for each sensor. Before the changepoint, it is assumed that the observations are independent and identically distributed (i.i.d.) across time and across sensors. As a result, we have that for $k < \nu$ the joint pdf of $\boldsymbol{X}[k]$ is given by

$$g(\boldsymbol{X}[k]) \triangleq \prod_{\ell=1}^{L} g(X_\ell[k]). \tag{1}$$

After the changepoint, all the affected sensors follow the post-change pdf and the observations are independent across sensors and time, conditioned on $\boldsymbol{S}$. Hence, we have that for $k \geq \nu$ the joint pdf of $\boldsymbol{X}[k]$ conditioned on $\boldsymbol{S}$ is given by

$$p_{\boldsymbol{S}[k]}(\boldsymbol{X}[k]) \triangleq \left(\prod_{\ell \in \boldsymbol{S}[k]} f(X_\ell[k]) \right) \cdot \left(\prod_{\ell \notin \boldsymbol{S}[k]} g(X_\ell[k]) \right), \tag{2}$$

where for $\boldsymbol{A} \in \text{comb}([L], m)$, $p_{\boldsymbol{A}}(\cdot)$ denotes the joint pdf of a vector measurement at a time instant when the underlying distribution is the one induced when the anomalous nodes are given by $\boldsymbol{A}$. As a result, for fixed $\boldsymbol{S}$ we have the following statistical model for the observations:

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]) & 1 \leq k < \nu \\ p_{\boldsymbol{S}[k]}(\boldsymbol{X}[k]) & k \geq \nu. \end{cases} \tag{3}$$

In addition, for $k_2 \geq k_1$ we have that

$$\boldsymbol{X}[k_1, k_2] \sim \begin{cases} \prod_{j=k_1}^{k_2} g(\boldsymbol{X}[j]) & 1 \leq k_2 < \nu \\ \left(\prod_{j=k_1}^{\nu-1} g(\boldsymbol{X}[j]) \right) \cdot \left(\prod_{j=\nu}^{k_2} p_{\boldsymbol{S}[j]}(\boldsymbol{X}[j]) \right) & k_2 \geq \nu. \end{cases} \tag{4}$$

Furthermore, the likelihood ratio between the hypothesis that the change has occurred at ν with a trajectory described by $\boldsymbol{S}$, and that there is no anomaly is given by

$$\Gamma_{\boldsymbol{S}}(k, \nu) \triangleq \prod_{j=\nu}^{k} \left(\prod_{\ell \in \boldsymbol{S}[j]} \frac{f(X_\ell[j])}{g(X_\ell[j])} \right). \tag{5}$$

In this work, our goal is to design algorithms in the form of *stopping times* that will detect the abrupt change described in (3). Denote by $\mathcal{F}_k = \sigma(\boldsymbol{X}[1], \ldots, \boldsymbol{X}[k])$ the σ-algebra generated by $\boldsymbol{X}[1, k]$. A stopping time τ, adapted to the filtration $\{\mathcal{F}_k\}_{k=1}^{\infty}$, is a positive integer-valued random variable such that for all $k \geq 1$ we have that $\{\tau \leq k\} \in \mathcal{F}_k$, i.e., knowledge of $\boldsymbol{X}[1, k]$ is sufficient to decide whether or not to stop at time k. Define by $\mathbb{E}_\infty[\cdot]$ the expectation when no anomaly is present. To quantify the frequency of *false alarm* (FA) events we use the *mean time to false alarm* (MTFA) denoted by $\mathbb{E}_\infty[\tau]$ for stopping time τ. Furthermore, to account for the lack of a specific model for $\boldsymbol{S}$, we use a detection delay metric based on a modification of Lorden's delay [9]. Our detection delay evaluates the performance of stopping times by considering the worst locations of the anomalous nodes. In particular, denote by $\mathbb{E}_\nu^{\boldsymbol{S}}[\cdot]$ the expectation when the anomaly emerges at ν, and the identities of the anomalous nodes are completely specified by the trajectory process $\boldsymbol{S}$. To evaluate our candidate detection procedures, we use the following delay metric:

$$\text{WADD}(\tau) = \sup_{\boldsymbol{S}} \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu^{\boldsymbol{S}} \left[\tau - \nu + 1 | \tau \geq \nu, \mathcal{F}_{\nu-1} \right], \tag{6}$$

where we use the convention that $\mathbb{E}_\nu^{\boldsymbol{S}} \left[\tau - \nu + 1 | \tau \geq \nu, \mathcal{F}_{\nu-1} \right] \triangleq 1$ when $\mathbb{P}_\nu^{\boldsymbol{S}}(\tau \geq \nu) = 0$. For $\gamma > 0$ a pre-determined constant, define the class of stopping times

$$\mathcal{C}_\gamma \triangleq \{\tau : \mathbb{E}_\infty[\tau] \geq \gamma\}. \tag{7}$$

Our goal is to solve the following optimization problem

$$\begin{aligned} \min_{\tau} \quad & \text{WADD}(\tau) \\ \text{s.t.} \quad & \tau \in \mathcal{C}_\gamma. \end{aligned} \tag{8}$$

To proceed with our theoretical analysis, it is important to introduce another observation model that arises when the m anomalous nodes are chosen from $\text{comb}([L], m)$ uniformly at random at each time $k \geq \nu$. According to this model, we have that the observations before the changepoint are independent and identically distributed across time and across sensors and are generated according to (1). After the changepoint, the data are distributed i.i.d. according to the mixture pdf that arises when the m anomalous nodes are chosen uniformly at random. This joint pdf is given by

$$\overline{p}(\boldsymbol{X}[k]) \triangleq \sum_{\boldsymbol{A} \in \text{comb}([L], m)} \frac{1}{\binom{L}{m}} p_{\boldsymbol{A}}(\boldsymbol{X}[k]). \tag{9}$$

As a result, the underlying statistical model when the anomalous nodes are chosen uniformly at random is given by

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]) & 1 \leq k < \nu \\ \overline{p}(\boldsymbol{X}[k]) & k \geq \nu. \end{cases} \tag{10}$$

and the likelihood rate between the hypotheses that a change occurs at ν and that no change occurs is given by

$$\mathcal{L}(k, \nu) \triangleq \prod_{j=\nu}^{k} \left(\sum_{\boldsymbol{A} \in \text{comb}([L], m)} \frac{1}{\binom{L}{m}} \prod_{\ell \in \boldsymbol{A}} \frac{f(X_\ell[j])}{g(X_\ell[j])} \right). \tag{11}$$

Note that the statistical model in (10) corresponds to an instance of the classic QCD problem studied in [9]– [11]. For the underlying QCD problem arising from (10), define the corresponding detection delay of stopping time τ by

$$\overline{\mathrm{WADD}}(\tau) = \sup_{\nu \geq 1} \operatorname{ess\,sup} \overline{\mathbb{E}}_\nu[\tau - \nu + 1 | \tau \geq \nu, \mathcal{F}_{\nu-1}] \quad (12)$$

where $\overline{\mathbb{E}}_\nu[\cdot]$ denotes the expectation when the underlying statistical model is that of (10) and the change happens at ν. Here also, it is assumed that $\overline{\mathbb{E}}_\nu[\tau - \nu + 1 | \mathcal{F}_{\nu-1}, \tau \geq \nu] \triangleq 1$ when $\overline{\mathbb{P}}_\nu(\tau \geq \nu) = 0$. Furthermore, define the Kullback-Leibler (KL) number corresponding to the QCD problem of (10) by

$$I \triangleq \overline{\mathbb{E}}_1 \left[\log \frac{\overline{p}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])} \right]. \quad (13)$$

III. Mixture-CuSum Test

For $\gamma > 0$, define the following *mixture-CuSum* statistic:

$$W[k] \triangleq \max_{1 \leq j \leq k} \mathcal{L}(k, j), \quad (14)$$

with the corresponding stopping time

$$\tau_C = \inf \{k \geq 1 : W[k] \geq b\}, \quad (15)$$

where b is the threshold that guarantees that $\mathbb{E}_\infty[\tau_C] = \gamma$. It can be seen that the test statistic in (14) can be expressed recursively as

$$W[k] = \max\{W[k-1], 1\} \mathcal{L}(k, k), \quad (16)$$

with $W[0] \triangleq 0$.

IV. Main Results

Our main theoretical result in this work is to establish the exact optimality of the mixture-CuSum (M-CuSum) test introduced in eqs. (14)-(16) with respect to the QCD problem defined in (3), (6), (8). In addition to the results, we also present sketches of the proofs. We start by presenting a universal lower bound on the detection delay of stopping times.

***Lemma** 1:* For any stopping time τ adapted to the filtration $\{\mathcal{F}_k\}_{k=1}^\infty$ and $N > 0$ define its truncated version by $\tau^{(N)} = \min\{\tau, N\}$. If $\mathbb{E}_\infty[\tau] < \infty$ we have that

$$\mathrm{WADD}(\tau^{(N)}) \geq \frac{\sup_{\boldsymbol{S}[1,N-1]} \mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1 - W[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1, \nu) \right]}{\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} (1 - W[\nu-1])^+ \right]}. \quad (17)$$

Proof: By doing a change of measure [26] it can seen that for any path $\boldsymbol{S}$ and for any $\nu \geq 1$ we have that almost surely under $\mathbb{P}_\infty(\cdot)$

$$\begin{aligned} \mathrm{WADD}(\tau^{(N)}) &\geq \mathbb{E}_\nu^{\boldsymbol{S}} \left[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \mathcal{F}_{\nu-1} \right] \\ &= \mathbb{E}_\nu^{\boldsymbol{S}} \left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)} \geq j\}} \middle| \tau^{(N)} \geq \nu, \mathcal{F}_{\nu-1} \right] \\ &= \mathbb{E}_\infty \left[\sum_{j=\nu}^{\infty} \Gamma_{\boldsymbol{S}}(j-1, \nu) \mathbb{1}_{\{\tau^{(N)} \geq j\}} \middle| \tau^{(N)} \geq \nu, \mathcal{F}_{\nu-1} \right]. \end{aligned} \quad (18)$$

By multiplying both sides of the inequality (18) with $\mathbb{1}_{\{\tau^{(N)} \geq \nu\}}(1 - W[\nu-1])^+$, summing over ν, and taking the expected value under $\mathbb{E}_\infty[\cdot]$ we have that after using the tower property of expectations the result follows. ■

Now we are ready to establish our first theorem that connects the delay metrics introduced in (6) and (12).

***Theorem** 1:* For any $\gamma > 0$ and for τ_C defined in Sec. III we have that

$$\mathrm{WADD}(\tau_C) \geq \inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}(\tau) \geq \overline{\mathrm{WADD}}(\tau_C). \quad (19)$$

Proof: By using the fact that for any set $E = \{e_1, \dots, e_M\}$ where $1 < M < \infty$ and $e_j \geq 0$ for all $1 \leq j \leq M$ we have that $\sup(E) \geq \sum_{j=1}^{M} \frac{e_j}{M}$, and by using Lemma 1 after taking the limit as $N \to \infty$ we can easily establish that for any $\gamma > 0$ and $b' \geq b$ such that $b' \geq 1$

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}(\tau) \geq \frac{\inf_{\tau \in \mathcal{C}_\gamma} \mathbb{E}_\infty \left[\sum_{j=0}^{\tau-1} \min\{\max\{W[j], 1\}, b'\} \right]}{\sup_{\tau \in \mathcal{C}_\gamma} \mathbb{E}_\infty \left[\sum_{\nu=0}^{\tau-1} (1 - W[\nu])^+ \right]}. \quad (20)$$

By using Theorem 1 of [11] and since $W[j] < b \leq b'$ for $0 \leq j < \tau_C$ we can then easily show that

$$\frac{\inf_{\tau \in \mathcal{C}_\gamma} \mathbb{E}_\infty \left[\sum_{j=0}^{\tau-1} \min\{\max\{W[j], 1\}, b'\} \right]}{\sup_{\tau \in \mathcal{C}_\gamma} \mathbb{E}_\infty \left[\sum_{\nu=0}^{\tau-1} (1 - W[\nu])^+ \right]} = \frac{\mathbb{E}_\infty \left[\sum_{j=0}^{\tau_C-1} \max\{W[j], 1\} \right]}{\mathbb{E}_\infty \left[\sum_{\nu=0}^{\tau_C-1} (1 - W[\nu])^+ \right]} = \overline{\mathrm{WADD}}(\tau_C). \quad (21)$$

Finally, from (20) and (21) the result is established. ■

To proceed with our analysis we investigate the relationship between $\mathrm{WADD}(\tau_C)$ and $\overline{\mathrm{WADD}}(\tau_C)$.

***Theorem** 2:* For any $\gamma > 0$ and for τ_C defined in Sec. III we have that

$$\mathrm{WADD}(\tau_C) = \overline{\mathrm{WADD}}(\tau_C) \quad (22)$$

Proof: Note that the observation model in (3) and the M-CuSum test of Sec. III is symmetric with respect to the different nodes of the network. This implies that the delay of the M-CuSum test is independent of $\boldsymbol{S}$. The equality in (22) can then be easily established by induction and by a change of measure argument similar to that in the proof of Lemma 1. ■

We now establish the optimality of our proposed test.

Theorem 3: For any $\gamma > 0$ and for τ_C defined in Sec. III we have that

$$\mathrm{WADD}(\tau_C) = \inf_{\tau \in C_\gamma} \mathrm{WADD}(\tau). \tag{23}$$

Furthermore, we have that as $\gamma \to \infty$

$$\mathrm{WADD}(\tau_C) \sim \frac{\log \gamma}{I}. \tag{24}$$

Proof: The theorem follows directly from Theorems 1 and 2 and by the asymptotic approximation of the delay of the CuSum test [9], [26]. ■

V. Heuristic and Oracle Tests

In this section, we design a heuristic test and an oracle test that can be used as a comparison to our proposed detection procedure described in (14) - (16).

Define by $J \triangleq D(f\|g)$ the KL divergence between $f(\cdot)$ and $g(\cdot)$. Note that for all $\boldsymbol{S}$ we have that

$$\mathbb{E}_\infty \left[\sum_{\ell=1}^{L} \log \frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)J \right] = -mJ < 0$$

$$\mathbb{E}_1^{\boldsymbol{S}} \left[\sum_{\ell=1}^{L} \log \frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)J \right] = mJ > 0,$$

This suggests that the following naive-CuSum (N-CuSum) test may be a candidate test for detecting the distribution change described in (3). In particular, consider the test described by the following recursion:

$$W_N[k+1] \triangleq \left(W_N[k] + \sum_{\ell=1}^{L} \log \frac{f(X_\ell[k+1])}{g(X_\ell[k+1])} + (L-m)J \right)^+ \tag{25}$$

with $W_N[0] \triangleq 0$ and corresponding stopping time

$$\tau_N = \inf \{k \geq 1 : W_N[k] \geq b\}.$$

Although the N-CuSum test can be employed to detect the anomaly because of having a positive expected drift, it does not necessarily solve the QCD problem in (3), (6), (8).

We also compare our proposed M-CuSum procedure to an oracle-CuSum (O-CuSum) test, which is a CuSum test that exploits the knowledge of the location of the anomalous modes. That is, to define this test we assume that at time k we do not know whether a change has occured, but we know which set of sensors would be affected if an anomaly had already emerged in the network. In particular, consider the statistic calculated by using the following recursion:

$$W_O[k+1] = \left(W_O[k] + \log \left(\prod_{\ell \in \boldsymbol{S}[k+1]} \frac{f(X_\ell[k+1])}{g(X_\ell[k+1])} \right) \right)^+ \tag{26}$$

with $W_O[0] \triangleq 0$ and with corresponding stopping time

$$\tau_O = \inf \{k \geq 1 : W_O[k] \geq b\}. \tag{27}$$

Since this O-CuSum test uses the knowledge of the location of the anomalous nodes, it is expected to perform better than our proposed test. However, such a test is not tractable since in practice such location information will not be available to the decision maker.

VI. Simulation Results

In this section, we conduct numerical simulations for the studied dynamic anomaly QCD problem for the case of $m = 1$ and when $g = \mathcal{N}(0,1)$ and $f = \mathcal{N}(1,1)$. We consider different values of network size L, and compare all the algorithms discussed in this paper. We also investigate how network size affects the performance of our proposed test for a fixed anomaly size. In particular, in Figs. 1, 2 and 3 we compare the M-CuSum test, the N-CuSum test and the O-CuSum test for $m = 1$ and network sizes $L = 5$, $L = 10$ and $L = 20$. Note that due to the symmetry of the M-CuSum and the N-CuSum test, WADD is equal to the delay for any arbitrary path of the anomaly. By looking at Figs. 1, 2 and 3 we note that the M-CuSum test outperforms the heuristic N-CuSum test, which is expected since the M-CuSum test is optimal with respect to (8). In addition, we note that the O-CuSum test performs better than the other detection schemes, since it uses the knowledge of the path of the anomaly. We also note that as L increases the performance gap between the O-CuSum test and the M-CuSum test increases. This

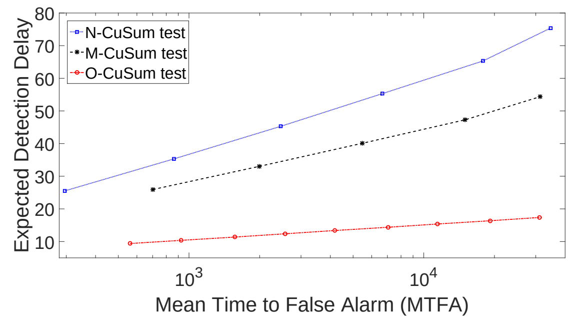

Fig. 1. WADD versus MTFA for $L = 5$, $m = 1$.

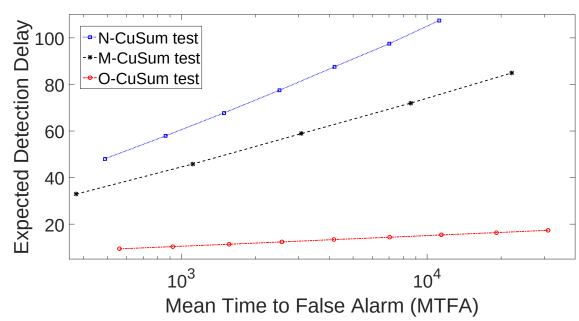

Fig. 2. WADD versus MTFA for $L = 10$, $m = 1$.

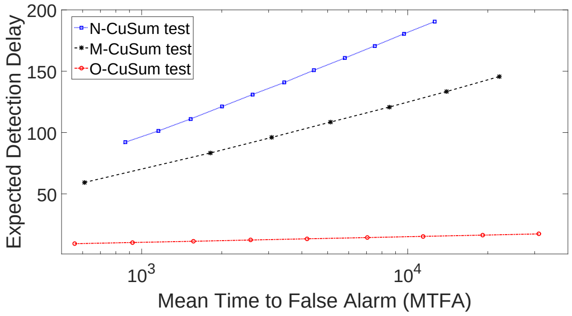

Fig. 3. WADD versus MTFA for $L = 20$, $m = 1$.

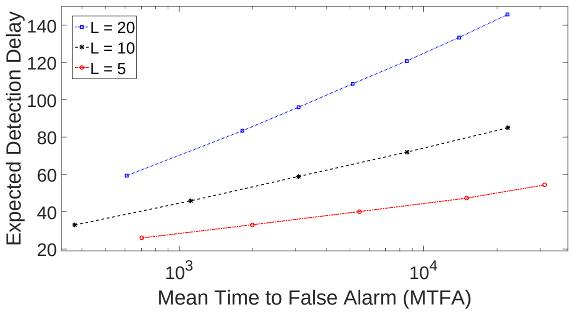

Fig. 4. WADD versus MTFA for the M-CuSum when $m = 1$ and for different L values.

is expected since as the network size increases the noise that is introduced in the M-CuSum test due to nodes that are not anomalous also increases. This is not the case for the O-CuSum test, since this scheme inherently assumes complete knowledge of the anomalous nodes. Finally, in Fig. 4, we compare the performance of the optimal M-CuSum test for $m = 1$ and for different values of L. We note that as L increases our proposed test performs worse, which is expected since the algorithm is affected by more noise from non-anomalous nodes for larger network sizes.

VII. Conclusion

In this paper, we introduced the problem of detecting a dynamic anomaly, i.e., an anomaly that may affect a sensor network without affecting each node of the network persistently. We studied this detection problem within the framework of quickest change detection (QCD) theory. We established that a Cumulative Sum test can be used to exactly solve the resulting QCD problem when the detection delay of candidate stopping procedures is evaluated according to the worst-path of the anomaly for the case of homogeneous sensors. Some potential directions of future interest include generalizing to the case of non-homogeneous sensors, and to the case of anomalies that grow in size.

References

[1] G. Rovatsos, X. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Comparison of statistical algorithms for power system line outage detection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 2946–2950.

[2] G. Rovatsos, J. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Statistical power system line outage detection under transient dynamics," *IEEE Trans. Signal Proc.*, vol. 65, no. 11, pp. 2787–2797, Jun. 2017.

[3] J. A. Rice, K. Mechitov, S. H. Sim, T. Nagayama, S. Jang, R. Kim, B. F. Spencer, G. Agha, and Y. Fujino, "Flexible smart sensor framework for autonomous structural health monitoring," *Smart Structures and Systems*, vol. 6, no. 5-6, pp. 423–438, Jul. 2010.

[4] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blazek, and H. Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," *IEEE Trans. Signal Proc.*, vol. 54, no. 9, pp. 3372–3382, Sept. 2006.

[5] S. E. Fienberg and G. Shmueli, "Statistical issues and challenges associated with rapid detection of bio-terrorist attacks," *Statistics in Medicine*, vol. 24, no. 4, pp. 513–529, Feb. 2005.

[6] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. Glob. Tel. Conf. (GLOBECOM)*, Nov. 2008, pp. 1–5.

[7] V. V. Veeravalli and T. Banerjee, *Quickest Change Detection*, Elsevier: E-reference Signal Processing, 2013.

[8] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, 2009.

[9] G. Lorden, "Procedures for reacting to a change in distribution," *Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.

[10] M. Pollak, "Optimal detection of a change in distribution," *Annals of Statistics*, vol. 13, no. 1, pp. 206–227, Mar. 1985.

[11] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.

[12] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.

[13] A. G. Tartakovsky and V. V. Veeravalli, "Change-point detection in multichannel and distributed systems," *Applied Sequential Methodologies: Real-World Examples with Data Analysis*, vol. 173, pp. 339–370, Jan. 2004.

[14] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, Apr. 2010.

[15] Y. Mei, "Quickest detection in censoring sensor networks," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2011.

[16] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *Annals of Statistics*, vol. 41, no. 2, pp. 670–692, 2013.

[17] G. Fellouris and A. G. Tartakovsky, "Multichannel sequential detection — part I: Non-i.i.d. data," *IEEE Trans. Inform. Theory*, vol. 63, no. 7, pp. 4551–4571, Jul. 2017.

[18] G. Fellouris and G. Sokolov, "Second-order asymptotic optimality in multisensor sequential change detection," *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 3662–3675, 2016.

[19] O. Hadjiliadis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3346–3359, 2009.

[20] V. Raghavan and V. V. Veeravalli, "Quickest change detection of a Markov process across a sensor array," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1961–1981, 2010.

[21] M. Ludkovski, "Bayesian quickest detection in sensor arrays," *Sequential Analysis*, vol. 31, no. 4, pp. 481–504, 2012.

[22] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley, "Quickest detection of dynamic events in networks," *arXiv e-prints*, p. arXiv:1807.06143, Jull 2018.

[23] R. Zhang and Y. Mei, "Asymptotic statistical properties of communication-efficient quickest detection schemes in sensor networks," *Sequential Analysis*, vol. 37, no. 3, pp. 375–396, 2018.

[24] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Sequential algorithms for moving anomaly detection in networks," *Sequential Analysis*, (to appear).

[25] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Quickest detection of a moving target in a sensor network," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2019.

[26] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2917–2929, Nov. 1998.