QUICKEST DETECTION OF STOCHASTIC FALSE DATA INJECTION ATTACKS WITH UNKNOWN PARAMETERS

Bettina D. Barros[*1], *Naveen K. D. Venkategowda*[1,2], *Stefan Werner*[1]

[1] Department of Electronic Systems, NTNU, Trondheim, Norway
[2] Department of Science and Technology, Linköping University, Sweden

ABSTRACT

This paper considers a multivariate quickest detection problem with false data injection (FDI) attacks in internet of things (IoT) systems. We derive a sequential generalized likelihood ratio test (GLRT) for zero-mean Gaussian FDI attacks. Exploiting the fact that covariance matrices are positive, we propose strategies to detect positive semi-definite matrix additions rather than arbitrary changes in the covariance matrix. The distribution of the GLRT is only known asymptotically whereas quickest detectors deal with short sequences, thereby leading to loss of performance. Therefore, we use a finite-sample correction to reduce the false alarm rate. Further, we provide a numerical approach to estimate the threshold sequences, which are analytically intractable to compute. We also compare the average detection delay of the proposed detector for constant and varying threshold sequences. Simulations showed that the proposed detector outperforms the standard sequential GLRT detector.

Index Terms— Sequential change detection, cybersecurity, false data injection attacks.

1. INTRODUCTION

With the modernization of cities, internet of things (IoT) systems will emerge in many applications such as smart grids, autonomous transportation, smart environments, and smart homes [1]. IoT systems have numerous sensors that interact with a central processor that governs decision-making processes. False data injection (FDI) attackers try to influence or disrupt such processes to cause adverse conditions such as hazards in autonomous transportation systems. Hence, given the importance of cybersecurity in IoT, there has been a growing interest in assessing the impact of FDI attacks in recent years [2, 3].

A simple approach to counteract FDI attacks is to employ detectors to identify malicious change of the data, which has received considerable attention in the literature [4–14]. In IoT, the sensors sequentially transmit their observations to a central processor. The central processor can analyze the sequential data stream from the sensors to detect FDI attacks by exploiting the fact that FDI attacks change the statistical properties of the data received at the central processor at an unknown time point. The goal is to detect attacks as soon as possible while keeping the false alarm rates at a low level, a problem known as quickest detection.

In [4, 5], the authors propose to use a χ^2 detector as an attack detector that raises an alarm if the energy of the Kalman filter innovation exceeds a certain threshold. Further, in [6], the authors use a consensus+innovation approach, which results in a resilient algorithm that locally detects FDI attacks and lets the system continue its operation. Last, in [7], the idea is to send local decisions as redundant information to the central processor, enhancing security at the cost of increased communication.

Although there are many relevant approaches to detect FDI attacks, most of the existing works on quickest detection use cumulative sum (CUSUM) algorithms or variations of them [8–11]. The standard CUSUM algorithm [15] is minimax optimal when the pre- and post-distributions are known [16]. However, we rarely know the post-distribution in real applications since the attackers can design their own injection sequences, trying to maximize the damage. When only the family of the distribution is known, we can use the generalized likelihood ratio test (GLRT) by replacing the distribution parameters with their maximum likelihood estimates [17]. More recently, in [12–14], the authors propose sequential GLRT algorithms for unknown, deterministic FDI attacks. In contrast, we assume attack sequences to be zero-mean Gaussian since it is proved to be the worst-case attack strategy [3].

When the FDI attack sequence is zero-mean Gaussian, the observation sequence admits a change in the covariance rather than the mean as in [12–14]. The works in [17–19] are considered state-of-art to detect changes in the variance or covariance of unknown pre- and post-Gaussian distributions. The multivariate approach in [18] is designed to detect arbitrary changes in the mean vector and the covariance matrix. However, when dealing with zero-mean Gaussian FDI attacks, the mean vector does not change, and the covariance matrix can only have positive semi-definite matrix additions, namely here as positive covariance-shifts. In light of this, we propose strategies that consider this extra information to im-

*This work was supported by the Research Council of Norway.

prove the detection performance.

In summary, in this paper, we derive a quickest detector for zero-mean multivariate Gaussian FDI attacks in IoT systems. First, we use the GLRT to estimate the unknown parameters of the system. Exploiting the fact that covariance matrices are positive, we propose strategies to detect positive covariance-shifts, which has received little attention in the literature. The distribution of the GLRT is only known asymptotically, whereas quickest detectors deal with short sequences, thereby leading to loss of performance. Therefore, we use a finite-sample correction to improve the detector performance. Further, we encounter an analytically intractable equation to compute the threshold sequences for the detector. Therefore, we provide simulations to estimate the thresholds. We also compare the average detection delay of the proposed detector for constant and varying threshold sequences.

2. PROBLEM FORMULATION

We consider a measurement model at discrete time t, given by

$$\mathbf{x}(t) = \mathbf{H}\boldsymbol{\theta}(t) + \mathbf{w}(t) + \mathbf{a}(t), \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^m$ is the observation vector, $\boldsymbol{\theta}(t) \in \mathbb{R}^n$ is the unknown time-variant and deterministic parameter, $\mathbf{H} \in \mathbb{R}^{m\times n}$ defines the system matrix, $\mathbf{w}(t) \in \mathbb{R}^m$ is white Gaussian noise with an unknown covariance matrix $\boldsymbol{\Sigma}_w \in \mathbb{R}^{m\times m}$, with $\boldsymbol{\Sigma}_w \succ 0$, and $\mathbf{a}(t) \in \mathbb{R}^m$ is the unknown FDI attack sequence which we consider independent of $\mathbf{w}(t)$.

When the system operates in its normal state, the attack sequence $\mathbf{a}(t)$ is zero. Otherwise, when the system is under attack, $\mathbf{a}(t)$ is a Gaussian sequence with zero mean and unknown covariance matrix $\boldsymbol{\Sigma}_a \in \mathbb{R}^{m\times m}$, with $\boldsymbol{\Sigma}_a \succeq 0$. In this paper, we consider that an attack starts at time t_a and continues indefinitely; thus, FDI attacks can be modelled as

$$\begin{cases} \mathbf{a}(t) = 0, & t < t_a \\ \mathbf{a}(t) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_a), & t \geq t_a, \end{cases} \tag{2}$$

which further implies

$$\mathbf{x}(t) \sim \begin{cases} \mathcal{N}(\mathbf{H}\boldsymbol{\theta}(t), \boldsymbol{\Sigma}_w), & t < t_a \\ \mathcal{N}(\mathbf{H}\boldsymbol{\theta}(t), \boldsymbol{\Sigma}_w + \boldsymbol{\Sigma}_a), & t \geq t_a. \end{cases} \tag{3}$$

This paper considers the problem of detecting FDI attacks given observations $\mathbf{x}(t)$. This problem can be formulated as a sequential hypothesis testing problem where the null hypothesis represents no attack. The goal is to minimize the average detection delay (ADD) subject to a constraint on the average run length to false alarm (ARLFA), known as a quickest detection problem. We define ARLFA $= \mathbb{E}_\infty[\hat{t}_a]$, and $\text{ADD}(t_a) = \mathbb{E}_{t_a}\left[\hat{t}_a - t_a \mid \hat{t}_a \geq t_a \wedge \mathbf{x}(t), \forall t < t_a\right]$, where $\hat{t}_a$ is the stopping time, and $\mathbb{E}_{t_a}$ is the expectation assuming that an attack starts at t_a, making $\mathbb{E}_\infty$ the expectation when there is no attack.

3. POSITIVE COVARIANCE-SHIFT DETECTOR

In this section, we construct a quickest detector for zero-mean multivariate Gaussian FDI attacks defined in (2). We use the well-known generalized likelihood ratio test (GLRT) [18] with the doubled-negative log-likelihood ratio computed as

$$\Lambda_{k,l} = -2\log \frac{\sup_{\boldsymbol{\theta},\boldsymbol{\Sigma}_w} \prod_{t=1}^{k} f(\mathbf{x}(t) \mid t_a > k)}{\sup_{\boldsymbol{\theta},\boldsymbol{\Sigma}_w,\boldsymbol{\Sigma}_a} \prod_{t=1}^{k} f(\mathbf{x}(t) \mid t_a = l)}, \tag{4}$$

where $f(\cdot)$ is a Gaussian density function. This test serves as a comparison between the likelihoods of having an attack starting at time l and not having an attack at all until the current time k. In a standard GLRT, the detection statistic is calculated as $\Lambda_k = \max_{1\leq l\leq k} \Lambda_{k,l}$, and then compared with a given threshold δ_k. The stopping time is defined as $\hat{t}_a = \min\{k : \max_{1\leq l\leq k} \Lambda_{k,l} \geq \delta_k\}$.

3.1. Test statistic

From the system model (1), we can see that, for the detector in (4), it is not possible to distinguish between the unknown vector $\boldsymbol{\theta}(t)$ and attack sequences that lie in the column space of $\mathbf{H}$. Therefore, the attack vector can be divided into two parts as $\mathbf{a}(t) = \mathbf{H}\bar{\mathbf{a}}(t) + \tilde{\mathbf{a}}(t)$, where $\tilde{\mathbf{a}}(t)$ is detectable and $\mathbf{H}\bar{\mathbf{a}}(t)$ is undetectable. The following lemma allows us to simplify our model and use only the detectable information from the observations.

Lemma 1. *Consider $\mathbf{H} \in \mathbb{R}^{m\times n}$ with rank n, then the projection matrix $\mathbf{P} = \mathbf{I}_m - \mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}$, where $\mathbf{I}_m \in \mathbb{R}^{m\times m}$ is the identity matrix, can be decomposed into $\mathbf{P} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{U} \in \mathbb{R}^{m\times r}$, $r = m - n$, is full-column rank.*

Proof. Since $\mathbf{P}$ is a projection matrix, we know that $\mathbf{P}$ is positive semi-definite with eigenvalues 0's and 1's and rank r. From the eigen-value decomposition of $\mathbf{P}$, we have

$$\mathbf{P} = \begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix} \begin{bmatrix}\mathbf{I}_r & 0 \\ 0 & 0\end{bmatrix} \begin{bmatrix}\mathbf{U}^{\mathrm{T}} \\ \mathbf{V}^{\mathrm{T}}\end{bmatrix} = \mathbf{U}\mathbf{U}^{\mathrm{T}} \tag{5}$$

which concludes the proof. □

Multiplying both sides in (1) with $\mathbf{U}^{\mathrm{T}}$ defined in Lemma 1, we obtain a simplified measurement model as

$$\tilde{\mathbf{x}}(t) = \tilde{\mathbf{w}}(t) + \tilde{\mathbf{a}}(t) \tag{6}$$

which further implies

$$\tilde{\mathbf{x}}(t) \sim \begin{cases} \mathcal{N}(0, \mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}_w\mathbf{U}), & t < t_a \\ \mathcal{N}(0, \mathbf{U}^{\mathrm{T}}(\boldsymbol{\Sigma}_w + \boldsymbol{\Sigma}_a)\mathbf{U}), & t \geq t_a \end{cases}. \tag{7}$$

Note that, unlike before, the unknown parameter $\boldsymbol{\theta}$ is not present in (6) and (7), resulting in a case where the mean of the transformed observation vector $\tilde{\mathbf{x}}(t)$ is time-invariant with

known value zero. Therefore, using the simplified observation model in (7), the test statistic in (4) can be written as

$$\tilde{\Lambda}_{k,l} = \log \frac{|\hat{\mathbf{\Sigma}}_{1,k}|^k}{|\hat{\mathbf{\Sigma}}_{1,l-1}|^{l-1}|\hat{\mathbf{\Sigma}}_{l,k}|^{k-l+1}}, \tag{8}$$

where $\hat{\mathbf{\Sigma}}_{i,j}$ is the covariance matrix estimated for GLRT as

$$\hat{\mathbf{\Sigma}}_{i,j} = \frac{1}{j-i+1}\sum_{t=i}^{j} \tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^{\mathrm{T}}(t). \tag{9}$$

Under the null hypothesis, i.e., no attack, the test statistic (8) has an asymptotic chi-squared distribution [20] with degrees of freedom $d = r(r+1), r = m-n$, which is also the number of parameters being estimated through the generalized approach; see [21] for other examples. However, quickest detection problems require that the tests are performed with short sequences. For this reason, it is common to use the so-called finite-sample correction to improve the convergence rate to the chi-squared distribution. This correction uses the fact that if Λ is a statistic with an asymptotic χ^2_d distribution, then the new statistic $\Lambda^c = d\Lambda/\mathbb{E}[\Lambda]$ converges at a faster rate [22]. Following similar steps as in [18], the expected value of $\tilde{\Lambda}_{k,l}$ can be computed as

$$\begin{aligned}\mathbb{E}[\tilde{\Lambda}_{k,l}] = r[&\log 2 + k\log(k) - (l-1)\log(l-1) \\ &- (k-l+1)\log(k-l+1)] \\ &+ \sum_{i=0}^{r-1}\left[k\psi\left(\frac{k-i}{2}\right) - (l-1)\psi\left(\frac{(l-1)-i}{2}\right)\right. \\ &\left.-(k-l+1)\psi\left(\frac{(k-l+1)-i}{2}\right)\right],\end{aligned} \tag{10}$$

where $\psi(\cdot)$ is the digamma function. From (10), the corrected test statistic is obtained as

$$\tilde{\Lambda}^c_{k,l} = r(r+1)\frac{\tilde{\Lambda}_{k,l}}{\mathbb{E}[\tilde{\Lambda}_{k,l}]}. \tag{11}$$

3.2. Unilateral detection statistics

The standard GLRT [18] considers any arbitrary change in the covariance matrix as a change-point. Our interest here, however, is to detect positive semi-definite matrix additions to the covariance, namely here as positive covariance-shifts. If a positive covariance-shift occurs, there is an increase in the generalized variance, defined as the determinant of the covariance matrix. From this idea, we derive two strategies, here called early-informed and late-informed strategies, depending on when we use the extra information that, for Gaussian FDI attacks, the covariance can only have positive shifts.

In the late-informed strategy, after calculating the standard detection statistic

$$\tilde{\Lambda}^{c,\text{late}}_k = \max_{1\le l\le k} \tilde{\Lambda}^c_{k,l}, \tag{12}$$

the alarm raises only if an increase in the generalized variance is signaled at $l^* = \arg\max_{1\le l\le k}\tilde{\Lambda}_{k,l}$, that is,

$$\hat{t}^{\,\text{late}}_a = \min\left\{k : \tilde{\Lambda}^{c,\text{late}}_k \ge \delta_k \wedge |\hat{\mathbf{\Sigma}}_{l^*,k}| > |\hat{\mathbf{\Sigma}}_{1,l^*-1}|\right\}.$$

In the early-informed strategy, the detection statistic is calculated only over the cases where an increase in the generalized variance happened, i.e.,

$$\tilde{\Lambda}^{c,\text{early}}_k = \max_{1\le l\le k}\left\{\tilde{\Lambda}^c_{k,l} : |\hat{\mathbf{\Sigma}}_{l,k}| > |\hat{\mathbf{\Sigma}}_{1,l-1}|\right\}, \tag{13}$$

and then the stopping time is directly

$$\hat{t}^{\,\text{early}}_a = \min\left\{k : \tilde{\Lambda}^{c,\text{early}}_k \ge \delta_k\right\}. \tag{14}$$

3.3. Threshold sequences

After calculating the detection statistic, a crucial part of constructing a sequential detector is to determine the threshold sequence δ_k, which can be constant [10–14] or time-varying [17–19]. Although constant thresholds are easy to compute in comparison to varying threshold sequences, they may not provide the best trade-off between detection delay and false alarm rate. One way of designing a varying threshold sequence is to fix the probability of false alarm at each observation to α, given that an alarm was not raised before the current time, i.e., for a detection statistic Λ_k,

$$Pr\left[\Lambda_k > \delta_k \mid \Lambda_t \le \delta_t, \forall t < k\right] = \alpha. \tag{15}$$

The average run length to false alarm is expected to be $1/\alpha$.

Because (15) is not known to be analytically tractable, we estimate δ_k through simulations, that even though are computationally expensive, must be performed only once. Moreover, the asymptotic distribution of $\tilde{\Lambda}^c_{k,l}$ in (11) is independent of the unknown covariance matrices (more details in [18]). Therefore, without loss of generality, we can estimate the threshold sequences using the transformed noise covariance as an identity matrix $\mathbf{U}^{\mathrm{T}}\mathbf{\Sigma}_w\mathbf{U} = \mathbf{I}_r$, and later on, use the results for any other covariance matrix.

4. SIMULATIONS

In this section, we illustrate the performance of the described detector, together with the effect of using constant or varying thresholds, and early- and late-informed testing strategies.

First step of the simulations is to estimate the threshold sequences. For this task, we created 1 million observation sequences with dimensions $m = 4, n = 2, r = 2$ and length 220 each. In order to get non-singular covariance estimations in (9), we should have at least $2(r+1)$ samples to perform a test. Based on the results in [18, 19], we use the initial 20 samples as learning observations and start the testing phase at sample 21.

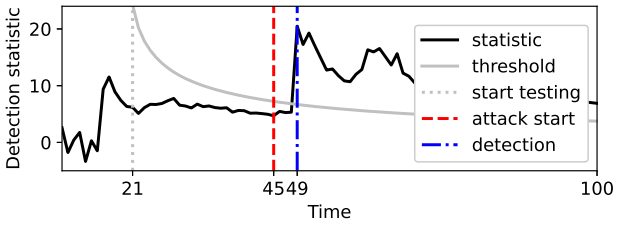

Fig. 1. Attack profile using varying threshold sequence, early-informed testing strategy, $\alpha = 0.02$.

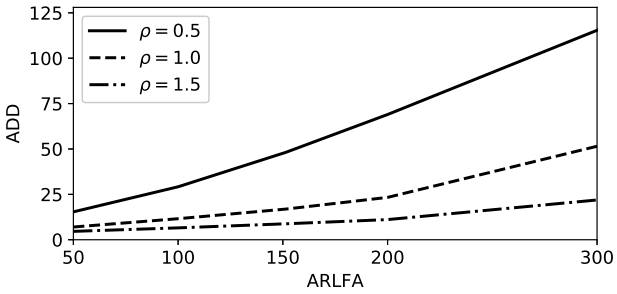

Fig. 2. ADD versus ARLFA using varying threshold sequence and early-informed testing strategy, with $t_a = 120$.

Figure 1 shows an example of the detector operation. The threshold sequence is decreasing, which is expected since the statistic has higher variations when the number of observations available is small. In Figure 1, when the attack starts, the statistic takes a few observations to change visually, and once it surpasses the threshold value, the attack is detected.

In this paper, for simplicity, we study the case where the transformed attack covariance is $\mathbf{U}^{\mathrm{T}}\mathbf{\Sigma}_a\mathbf{U} = \rho\mathbf{I}_2, \rho \in \mathbb{R}^+$. For each set of parameters, we created 10,000 observation sequences with dimension 2 and length 1,000 each. Fig. 2 shows that the greater the increase ρ, the faster the attacks are detected, as expected. We also see that there is a clear compromise between ADD and ARLFA. The lower we set the probability of false alarm, the more difficult it becomes to detect attacks, especially for small increases.

Table 1 compares the performance of the two types of testing strategies. The early-informed strategy performed better than the late one for all the scenarios considered, as expected since the late-informed strategy does not consider all the cases of generalized variance increase that the early one does.

Table 2 shows the detection delay for the two types of threshold sequences, constant and varying. Since the varying threshold is estimated by modeling the detection statistic distribution, we expect that it provides better performance than the constant one, which is verified through the results. The varying threshold performed better for all scenarios considered, especially for $t_a = 120$. When the sequences are still short, the variability of the detection statistic is higher. While the varying threshold can be higher at the beginning of the testing phase and then reduced later, the constant threshold, on the other hand, can not adapt to these scenarios. In order to avoid false alarms initially, the constant threshold must be higher than the desired value for longer sequences. Therefore, it results in a worse compromise between ADD and ARLFA when compared with the varying threshold.

Table 1. ADD for early- and late-informed testing strategies, and the relative difference in performance, with $\alpha = 0.005$.

	$t_a = 45$			$t_a = 120$		
ρ	late	early	diff%	late	early	diff%
0.5	142.8	133.7	6.4%	85.5	68.9	19.4%
1.0	101.6	89.2	12.2%	32.9	23.3	29.2%
1.5	40.9	32.5	20.5%	14.2	11.1	21.8%

Table 2. ADD for constant and varying threshold sequences, with $\alpha = 0.005$ and early-informed testing strategy.

	$t_a = 45$		$t_a = 120$	
ρ	constant	varying	constant	varying
0.5	162.4	133.7	192.3	68.9
1.0	113.3	89.2	134.2	23.3
1.5	52.1	32.5	74.3	11.1

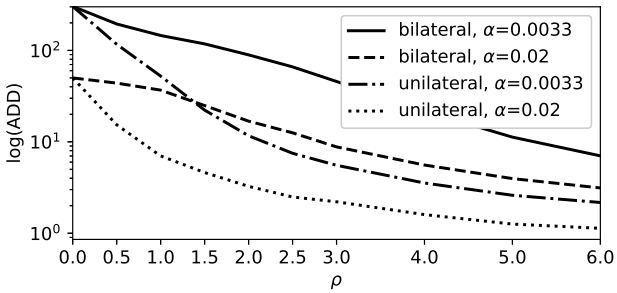

Fig. 3. ADD versus ρ using varying threshold, and bilateral or early-informed unilateral testing strategy, with $t_a = 120$.

Last, to visualize the performance of the proposed detector, we extend the simulations to higher values of ρ, and compare it to a standard, bilateral GLRT detector, which raises an alarm for any arbitrary change detected in the covariance. Figure 3 shows that the proposed unilateral detector outperformed the standard bilateral detector. Furthermore, the ADD of the proposed detector exponentially decreases at a faster rate compared to the bilateral detector, showing its advantage for Gaussian FDI attacks.

5. CONCLUSION

This paper considered a quickest detection problem with zero-mean multivariate Gaussian FDI attacks in IoT systems. Exploiting the fact that covariance matrices are positive, we proposed testing strategies to detect positive covariance-shifts. The proposed detector outperformed the standard GLRT detector, which raises an alarm for any arbitrary change detected in the covariance. We also compared the performance of constant and varying false alarm threshold sequences. Although the varying threshold sequences are computationally expensive to estimate, it has to be done only once, and the simulations showed a considerable performance improvement.

6. REFERENCES

[1] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, and B. Sikdar, "A survey on IoT security: Application areas, security threats, and solution architectures," *IEEE Access*, vol. 7, pp. 82 721–82 743, Jun. 2019.

[2] J. Zhang, R. S. Blum, and H. V. Poor, "Approaches to secure inference in the internet of things: Performance bounds, algorithms, and effective attacks on IoT sensor networks," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 50–63, Sept. 2018.

[3] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117–124, Mar. 2018.

[4] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, Sept. 2013.

[5] Y. Chen, S. Kar, and J. M. Moura, "Optimal attack strategies subject to detection constraints against cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1157–1168, Mar. 2017.

[6] ——, "Resilient distributed estimation through adversary detection," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2455–2469, Mar. 2018.

[7] W. Hashlamoun, S. Brahma, and P. K. Varshney, "Mitigation of Byzantine attacks on distributed detection systems using audit bits," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 18–32, Jul. 2017.

[8] E. Bayraktar and L. Lai, "Byzantine fault tolerant distributed quickest change detection," *SIAM Journal on Control and Optimization*, vol. 53, no. 2, pp. 575–591, 2015.

[9] A. S. Polunchenko and A. G. Tartakovsky, "State-of-the-art in sequential change-point detection," *Methodology and Computing in Applied Probability*, vol. 14, no. 3, pp. 649–684, 2012.

[10] D. Ye and T.-Y. Zhang, "Summation detector for false data-injection attack in cyber-physical systems," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2338–2345, Jun. 2019.

[11] V. V. Veeravalli and T. Banerjee, "Quickest change detection," in *Academic Press Library in Signal Processing*. Elsevier, 2014, vol. 3, pp. 209–255.

[12] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 106–115, Aug. 2012.

[13] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Dec. 2014.

[14] J. Zhang and X. Wang, "Low-complexity quickest change detection in linear systems with unknown time-varying pre-and post-change distributions," *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1804–1824, Jan. 2021.

[15] E. Page, "A test for a change in a parameter occurring at an unknown point," *Biometrika*, vol. 42, no. 3/4, pp. 523–527, Dec. 1955.

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

[17] D. M. Hawkins and K. Zamba, "Statistical process control for shifts in mean or variance using a changepoint formulation," *Technometrics*, vol. 47, no. 2, pp. 164–173, May 2005.

[18] K. Zamba and D. M. Hawkins, "A multivariate change-point model for change in mean vector and/or covariance structure," *Journal of Quality Technology*, vol. 41, no. 3, pp. 285–303, Jul. 2009.

[19] G. J. Ross, "Sequential change detection in the presence of unknown parameters," *Statistics and Computing*, vol. 24, no. 6, pp. 1017–1030, Nov. 2014.

[20] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, Mar. 1938.

[21] M. S. Bartlett, "A note on the multiplying factors for various $\chi 2$ approximations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 16, no. 2, pp. 296–298, Jul. 1954.

[22] J. L. Jensen, "A historical sketch and some new results on the improved log likelihood ratio statistic," *Scandinavian Journal of Statistics*, pp. 1–15, 1993.