Asymptotic Random Distortion Testing and Application to Change-in-Mean Detection

Guillaume Ansel, Dominique Pastor, Frédéric Cuppens, Nora Boulahia Cuppens

Asymptotic Random Distortion Testing and Application to Change-in-Mean Detection

Guillaume Ansel
Mathematical and Electrical Engineering
IMT Atlantique
Brest, France
guillaume.ansel@imt-atlantique.fr

Dominique Pastor
Mathematical and Electrical Engineering
IMT Atlantique
Brest, France
dominique.pastor@imt-atlantique.fr

Frédéric Cuppens
Computer Engineering and Software Engineering
Polytechnique Montréal
Montréal, Canada
frederic.cuppens@polymtl.ca

Nora Boulahia Cuppens
Computer Engineering and Software Engineering
Polytechnique Montréal
Montréal, Canada
nora.boulahia-cuppens@polymtl.ca

***Abstract*—We introduce an extension of the Random Distortion Testing (RDT) framework which allows its use when the noise variance is estimated. This asymptotic extension, named AsympRDT, shows that we asymptotically retain the level of the RDT test as the estimate of the noise variance converges to its real value. The validity of this approach is justified through both theoretical and simulation results. We make use of AsympRDT to develop a change-in-mean detection method for time series. It features three parameters: the size of the processed blocks, the maximum desired false alarm rate and a tolerance. We then show a use-case for this method in cybersecurity for Industrial Control Systems (ICS) as part of an anomaly and cyberattack detection system, where it can be used for segmenting signals and learning normal behaviors.**

***Index Terms*—Random Distortion Testing, Change point detection, anomaly detection, time series analysis, cybersecurity**

I. Introduction

A frequently encountered problem in statistical decision theory is the issue of deciding whether a parameter θ is equal to a given value θ_0 or not. This problem is often formulated as testing the simple null hypothesis $\mathcal{H}_0 \colon \theta = \theta_0$ against the composite alternative hypothesis $\mathcal{H}_1 \colon \theta \neq \theta_0$ given some observation that depends on θ. In general there is no Uniformly Most Powerful (UMP) test to solve this problem, but it is sometimes possible to find optimal tests among a restricted class of tests (e.g. the Wald test [1]).

Testing whether $\theta = \theta_0$ may be sometimes too restrictive, and it can be more reasonable to instead test if θ lies close enough to θ_0 given some user-controlled tolerance. This problem is addressed with Random Distortion Testing (RDT) [2], which consists in deciding whether an unknown random signal Θ lies close enough to a known deterministic model θ_0. The signal is assumed to be observed in presence of independent additive Gaussian noise X with known covariance matrix C. For this problem, [2] exhibits an optimal test for any desired level $\gamma \in (0, 1)$ among the tests which respect the invariance properties of the noise X. In practice however, it is common not to have access to the covariance matrix of the noise. An estimate is often used instead of the real matrix, and we do not necessarily retain optimality in this case. To address this, we establish an extension of the RDT framework named AsympRDT that takes the noise variance estimation into account. We show that the desired false alarm rate γ is asymptotically respected as the noise variance estimate improves. The study of the asymptotic detection probability and its optimality is in-progress and will be presented in a forthcoming paper. In the following, we only consider the case where the noise components are independent and identically distributed (i.i.d.), hence the noise covariance matrix can be reduced to a single parameter. This asymptotic result is confirmed through simulations, which also hint at a potential further extension where both the noise variance and the model θ_0 are estimated. This extension, along with the study of the asymptotic power of the test and the case of a general noise covariance matrix, are postponed to future work.

Using this result, we then present a new method for change-in-mean detection in time series. The method proposed can be regarded as a continuation and a formalization of works initiated and presented in [3]. The method relies on estimating the current mean and noise variance of the signal, and using the RDT test to detect whether the mean of the next block of samples has changed or not. The tolerance defined in the RDT test is used to specify the minimum amplitude of changes that should be detected, allowing the user to only detect relevant changes. The change-in-mean detection method we propose asymptotically inherits the false-alarm rate guarantee of the RDT test. Using this method, we then show a potential application in cybersecurity for industrial systems to help segment signals obtained from sensors. This segmentation step is part of a work in progress to build a model of the normal behavior of a system that can then be used to detect anomalies and attacks. Learning and using this model is beyond the scope

This work was done at the chair Cyber CNI with support of the FEDER development fund of the Brittany region.

of this paper; we will only demonstrate that we can use this change-in-mean detection method to segment signals measured on an industrial system.

Section II will introduce the RDT framework and the main results as presented in [2]. We then present the AsympRDT extension in Section III with both theoretical and simulation results. Section IV will then introduce our change-in-mean detection algorithm based on these results and an application to cybersecurity for Industrial Control Systems (ICS) in the context of water treatment, where it is used to help learn the normal behavior of the system.

Notations and terminology: All random variables and vectors are defined on the same probability space $(\Omega, \Sigma, \mathbb{P})$, and $\mathcal{M}(\Omega, \mathbb{R}^d)$ denotes the set of all $\mathbb{R}^d$-valued random vectors defined on Ω. For any positive definite $d \times d$ matrix C, the Mahalanobis norm defined on $\mathbb{R}^d$ with respect to C is denoted by ν_C. Hence, for any vector $y \in \mathbb{R}^d$, we have $\nu_C(y) = \sqrt{y^T C^{-1} y}$. $\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^d$. For any $\tau > 0$, $Q_{d/2}(\tau, \cdot)$ denotes the generalized Marcum function [4], which is the complementary cumulative distribution function (cdf) of the square root of any random variable that follows the non-central χ^2 distribution with d degrees of freedom and non-centrality parameter τ^2. Given $\theta_0 \in \mathbb{R}^d$ and a positive definite $d \times d$ matrix C, for any $\rho > 0$, we define Υ_ρ as the set $\Upsilon_\rho = \{y \in \mathbb{R}^d : \nu_C(y - \theta_0) = \rho\}$, and set $\mathfrak{F} = \{\Upsilon_\rho : \rho > 0\}$. For any $t > 0$, we define the *thresholding test* $\mathcal{T}_t$ by:

$$\begin{array}{rccl} \mathcal{T}_t : & \mathbb{R}^d & \longrightarrow & \{0,1\} \\ & y & \longmapsto & \begin{cases} 1 & \text{if } \nu_C(y-\theta_0) > t \\ 0 & \text{otherwise} \end{cases} \end{array} \tag{1}$$

II. RDT Framework

A. Problem statement

We briefly introduce the RDT problem and the main related results. A full presentation of the RDT framework can be found in [2]. Consider an unknown random vector $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^d)$ representing a quantity of interest, for example some physical phenomenon, of which we have a measurement Y in presence of independent additive Gaussian noise $X \sim \mathcal{N}(0, C)$ with known covariance matrix C. Given $\omega \in \Omega$ and the observation $Y(\omega)$, the RDT problem consists in deciding whether the quantity $\Theta(\omega)$ is close enough to a given model $\theta_0 \in \mathbb{R}^d$ with respect to a specified tolerance $\tau > 0$:

$$\text{RDT}: \begin{cases} \textbf{Observation}: \begin{cases} Y(\omega) = \Theta(\omega) + X(\omega) \\ X \sim \mathcal{N}(0, C) \\ \Theta \text{ and } X \text{ independent} \end{cases} \\ \textbf{Null event}: \nu_C(\Theta(\omega) - \theta_0) \leq \tau \\ \textbf{Alternative event}: \nu_C(\Theta(\omega) - \theta_0) > \tau \end{cases} \tag{2}$$

Unlike many other statistical decision problems, this decision problem is not about the model which generates the observation, but instead about a property of the realization $\Theta(\omega)$ of the unknown random vector Θ.

B. The Deterministic Case: a UMPI Test

We define the *power function* $\beta_{\mathcal{T}} : \mathbb{R}^d \to [0,1]$ and the *size* $\alpha_{\mathcal{T}}$ of any given test $\mathcal{T} : \mathbb{R}^d \to \{0,1\}$ as:

$$\begin{aligned} &\forall \theta \in \mathbb{R}^d,\ \beta_{\mathcal{T}}(\theta) = \mathbb{P}[\mathcal{T}(\theta + X) = 1] \\ &\alpha_{\mathcal{T}} = \sup_{\theta \in \mathbb{R}^d :\ \nu_C(\theta - \theta_0) \leq \tau} \beta_{\mathcal{T}}(\theta) \end{aligned} \tag{3}$$

Note that θ is deterministic in these definitions. Using these definitions of size and power, we can attempt to find an optimal test in the sense of a Uniformly Most Powerful (UMP) test of level $\gamma \in (0,1)$, i.e. a test $\mathfrak{T}^*$ such that $\alpha_{\mathfrak{T}^*} \leq \gamma$ and which verifies for any other test $\mathcal{T}$:

$$\forall \theta \in \mathbb{R}^d, \nu_C(\theta - \theta_0) > \tau \Rightarrow \beta_{\mathfrak{T}^*}(\theta) \geq \beta_{\mathcal{T}}(\theta)$$

Unfortunately, there exists no UMP test among all tests defined on $\mathbb{R}^d$. Since we have no knowledge on the random vector Θ, we consider the invariance properties of the noise X, whose probability density function (pdf) can be written as a function of the Mahalanobis norm ν_C. The RDT problem is invariant by any transformation preserving the Mahalanobis norm. Therefore, we restrict our search for an optimal test to the set of tests that respect this invariance. Among these tests, which can be written as functions of ν_C [5], it has been shown that the thresholding test $\mathcal{T}_{\lambda_\gamma(\tau)}$ is Uniformly Most Powerful Invariant (UMPI) with size γ, where $\lambda_\gamma(\tau) > 0$ is the unique positive real number verifying the equation $Q_{d/2}(\tau, \lambda_\gamma(\tau)) = 1 - \gamma$.

C. The Random Case: a γ-MCCP Test

We can go further and show that this test is optimal among a larger class of tests. The UMPI criterion relies on our definition of size and power function, which involve a deterministic parameter θ only, and do not take into account the fact that the RDT problem concerns a random vector Θ. Appropriate notions of size and power, involving conditional probabilities, can be introduced to bypass this limitation. These notions can then be linked to those introduced in Eq. (3), before recalling the main result of [2]. In this regard, the following lemma will prove useful in the sequel. In particular, the left hand side in Eq. (4) below can be regarded as a natural definition for the size of a test with respect to the RDT problem of Eq. (2). Lemma 1 establishes a direct connection between this notion of size and the size as defined in Eq. (3).

Lemma 1. *For any test $\mathcal{T} : \mathbb{R}^d \to \{0,1\}$, we have:*

$$\sup_{\Xi \in \mathcal{M}(\Omega, \mathbb{R}^d) :\ \mathbb{P}[\nu_C(\Xi - \theta_0) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\Xi + X) = 1 \mid \nu_C(\Xi - \theta_0) \leq \tau] = \alpha_{\mathcal{T}} \tag{4}$$

Definition 1 (Constant conditional power function given $\Xi \in \Upsilon_\rho$)**.** *Let $\rho > 0$, and let $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^d)$ be a random vector independent of $X \sim \mathcal{N}(0, C)$. A test $\mathcal{T}$ is said to have constant conditional power function given $\Xi \in \Upsilon_\rho$ if:*

$$\forall \theta \in \Upsilon_\rho, \mathbb{P}[\mathcal{T}(\Xi + X) = 1 \mid \Xi \in \Upsilon_\rho] = \beta_{\mathcal{T}}(\theta)$$

Lemma 2. *A test $\mathcal{T}$ has constant power function on every $\Upsilon_\rho \in \mathfrak{F}$ if and only if, for any $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^d)$, the test $\mathcal{T}$ has constant conditional power given $\Xi \in \Upsilon_\rho$ for $\mathbb{P}\nu_C(\Xi - \theta_0)^{-1}$-almost every $\rho \in \mathbb{R}$.*

The criterion used to find an optimal test for the RDT problem is the following:

Definition 2 (γ-MCCP test)**.** *Given $\tau \in \mathbb{R}$ and $\gamma \in (0,1)$, a test $\mathfrak{T}^*$ is said to have level γ and Maximum Conditional Constant Power (MCCP) — and we say that $\mathfrak{T}^*$ is γ-MCCP — if:*

- *(Level) $\alpha_{\mathcal{T}} \leq \gamma$;*
- *(Constant conditional power function) For any $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^d)$ and for $\mathbb{P}(\nu_C(\Xi - \theta_0))^{-1}$-almost every $\rho > \tau$, $\mathfrak{T}^*$ has constant conditional power function given $\Xi \in \Upsilon_\rho$;*
- *(MCCP) For any $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^d)$, for $\mathbb{P}\nu_C(\Xi - \theta_0)^{-1}$-almost every $\rho > \tau$, and for any test $\mathcal{T} \in \mathcal{K}_\gamma$ with constant conditional power function given $\Xi \in \Upsilon_\rho$ we have:*

$$\mathbb{P}[\mathfrak{T}^*(\Xi + X) = 1 \mid \Xi \in \Upsilon_\rho] \geq \mathbb{P}[\mathcal{T}(\Xi + X) = 1 \mid \Xi \in \Upsilon_\rho] \quad (5)$$

We now state the main result of [2]:

Theorem 1. *The test $\mathcal{T}_{\lambda_\gamma(\tau)}$ is γ-MCCP.*

III. Asymptotic RDT

In the following, we consider the case when the covariance matrix C of the noise X can be written $C = \sigma_0^2 I_d$. We present here an extension to the RDT framework in which the noise variance σ_0^2 is estimated instead of being perfectly known.

A. Problem statement

The problem statement remains the same as in Eq. (2), but unlike the previous section, we do not have access to σ_0^2 and we have to rely on an estimate instead. For example, we may have access to a representative time series $(R_n)_{n\in\mathbb{N}}$ that can be used to estimate the noise variance via the usual maximum likelihood estimator:

$$\forall n \in \mathbb{N}, \widehat{\sigma}_n = \frac{1}{d\,n} \sum_{i=1}^{n} \left(\|R_i - m_R\|_2^2 \right) \quad (6)$$

with m_R the mean of $(R_n)_{n\in\mathbb{N}}$. For the asymptotic RDT problem, we assume that we have a sequence of random variables $(\widehat{\sigma}_n)_{n\in\mathbb{N}} \in \mathcal{M}(\Omega, (\mathbb{R}))^{\mathbb{N}}$ that converges in distribution to σ_0. The objective is to show that the test is asymptotically of level γ when we consider the sequence $\left(\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\cdot, \theta_0, \widehat{\sigma}_n)\right)_{n\in\mathbb{N}}$, where $\widetilde{\mathcal{T}}_t$ is defined for any $t > 0$ by:

$$\begin{aligned} \widetilde{\mathcal{T}}_t \colon \quad \mathbb{R}^d \times \mathbb{R}^d \times (0,\infty) &\longrightarrow \{0,1\} \\ (y, \theta, \sigma) &\longmapsto \begin{cases} 1 & \text{if } \dfrac{\|y-\theta\|_2}{\sigma} > t \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (7)$$

One can note that, for any $\theta \in \mathbb{R}^d$ and any $\sigma \in (0,\infty)$, the application $\widetilde{\mathcal{T}}_t(\cdot, \theta, \sigma)$ is a test.

B. Theoretical results

Let $(\widehat{\sigma}_n) \in \mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}}$ be a sequence of random variables that converges in distribution to σ_0^2 such that for all $n \in \mathbb{N}$, Y and $\widehat{\sigma}_n$ are independent. Let $\mathcal{S} = \{\Xi \in \mathcal{M}(\Omega, \mathbb{R}^d) \colon \forall n \in \mathbb{N}, \Xi \text{ and } \widehat{\sigma}_n \text{ are independent}\}$.

Theorem 2.

$$\limsup_n \sup_{\Xi\in\mathcal{S}\colon \mathbb{P}[\|\Xi-\theta_0\|_2 \leq \sigma_0\tau] \neq 0} \mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi + X, \theta_0, \widehat{\sigma}_n) = 1 \,|\, \|\Xi - \theta_0\|_2 \leq \sigma_0\tau\right] \leq \gamma \quad (8)$$

Proof. Let $\Xi_0 \in \mathcal{S}$ such that $\mathbb{P}[\|\Xi_0 - \theta_0\|_2 \leq \sigma_0\tau] \neq 0$.

$$\begin{aligned} &\mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi_0 + X, \theta_0, \widehat{\sigma}_n) = 1 \mid \|\Xi_0 - \theta_0\|_2 \leq \sigma_0\tau\right] \\ &= \int \mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi_0 + X, \theta_0, \sigma) = 1 \,|\, \|\Xi_0 - \theta_0\|_2 \leq \sigma_0\tau\right] \mathbb{P}\widehat{\sigma}_n^{-1}(\mathrm{d}\sigma) \\ &\leq \int \sup_{\Xi\in\mathcal{M}(\Omega,\mathbb{R}^d)\colon \mathbb{P}[\|\Xi-\theta_0\|_2\leq\sigma_0\tau]\neq 0} \mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi + X, \theta_0, \sigma) = 1 \,|\, \|\Xi - \theta_0\|_2 \leq \sigma_0\tau\right] \mathbb{P}\widehat{\sigma}_n^{-1}(\mathrm{d}\sigma) \end{aligned}$$

From Lemma 1, since $\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\cdot, \theta_0, \sigma)$ is a test, we have:

$$\begin{aligned} &\sup_{\Xi\in\mathcal{M}(\Omega,\mathbb{R}^d)\colon \mathbb{P}[\|\Xi-\theta_0\|_2\leq\sigma_0\tau]\neq 0} \mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi + X, \theta_0, \sigma) = 1 \mid \|\Xi - \theta_0\|_2 \leq \sigma_0\tau\right] \\ &\qquad = \sup_{\theta\in\mathbb{R}^d\colon \|\theta-\theta_0\|_2\leq\sigma_0\tau} \mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\theta + X, \theta_0, \sigma) = 1\right] \end{aligned}$$

Therefore:

$$\begin{aligned} &\mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi_0 + X, \theta_0, \widehat{\sigma}_n) = 1 \mid \|\Xi_0 - \theta_0\|_2 \leq \sigma_0\tau\right] \\ &\qquad \leq \int \sup_{\theta\in\mathbb{R}^d\colon \|\theta-\theta_0\|_2\leq\sigma_0\tau} \mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\theta + X, \theta_0, \sigma) = 1\right] \mathbb{P}\widehat{\sigma}_n^{-1}(\mathrm{d}\sigma) \\ &\qquad \leq \int \sup_{\theta\in\mathbb{R}^d\colon \|\theta-\theta_0\|_2\leq\sigma_0\tau} \mathbb{P}\left[\frac{\|\theta + X - \theta_0\|_2}{\sigma} > \lambda_\gamma(\tau)\right] \mathbb{P}\widehat{\sigma}_n^{-1}(\mathrm{d}\sigma) \end{aligned}$$

Let $W = \theta + X - \theta_0 \sim \mathcal{N}(\theta - \theta_0, \sigma_0^2 I_d)$. We have:

$$\begin{aligned} \mathbb{P}\left[\frac{\|W\|_2}{\sigma} > \lambda_\gamma(\tau)\right] &= \mathbb{P}\left[\frac{\|W\|_2}{\sigma_0} > \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\right] \\ &= Q_{d/2}\left(\frac{\|\theta - \theta_0\|_2}{\sigma_0}, \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\right) \end{aligned}$$

The function $Q_{d/2}$ is continuous and increases with its first argument, therefore:

$$\sup_{\theta\in\mathbb{R}^d\colon \|\theta-\theta_0\|_2\leq\sigma_0\tau} Q_{d/2}\left(\frac{\|\theta - \theta_0\|_2}{\sigma_0}, \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\right) = Q_{d/2}\left(\tau, \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\right)$$

Hence:

$$\begin{aligned} &\mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi_0 + X, \theta_0, \widehat{\sigma}_n) = 1 \mid \|\Xi_0 - \theta_0\|_2 \leq \sigma_0\tau\right] \\ &\qquad \leq \int Q_{d/2}\left(\tau, \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\right) \mathbb{P}\widehat{\sigma}_n^{-1}(\mathrm{d}\sigma) \end{aligned}$$

This inequality is valid for any $\Xi_0 \in \mathcal{S}$ such that $\mathbb{P}[\|\Xi_0 - \theta_0\|_2 \leq \sigma_0\tau] \neq 0$, and the right-hand side does not depend on Ξ_0. Therefore:

$$\begin{aligned} &\sup_{\Xi\in\mathcal{S}\colon \mathbb{P}[\|\Xi-\theta_0\|_2\leq\sigma_0\tau]\neq 0} \mathbb{P}\left[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi + X, \theta_0, \widehat{\sigma}_n) = 1 \mid \|\Xi - \theta_0\|_2 \leq \sigma_0\tau\right] \\ &\qquad \leq \int Q_{d/2}\left(\tau, \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\right) \mathbb{P}\widehat{\sigma}_n^{-1}(\mathrm{d}\sigma) \end{aligned}$$

The function $\sigma \mapsto Q_{d/2}\left(\tau, \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\right)$ defined on $[0,\infty)$ is uniformly continuous, since it is continuous and has a finite limit as σ tends to infinity. This function is also bounded by

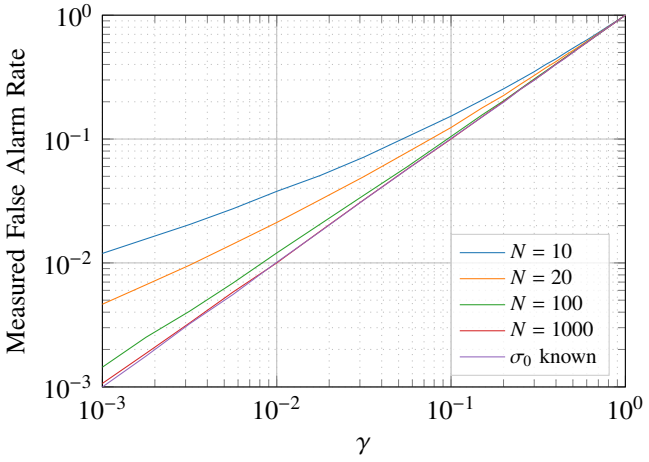

Fig. 1. Measured false alarm rate vs. parameter γ using N samples to estimate σ_0 ($\tau = 0$, $d = 1$).

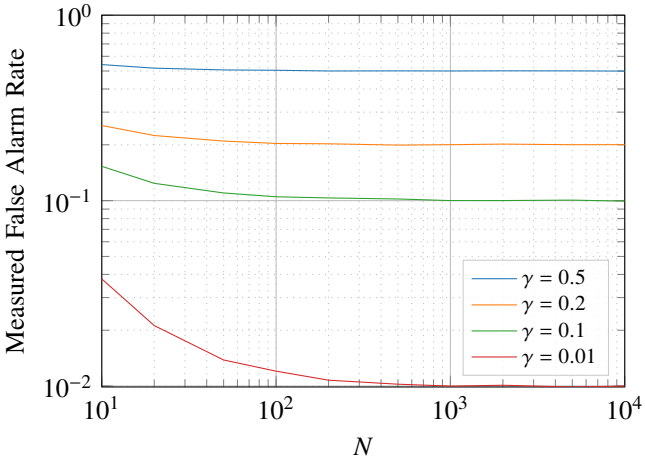

Fig. 2. Measured false alarm rate vs. number of samples N used to estimate σ_0 ($\tau = 0$, $d = 1$).

definition of $Q_{d/2}$. Therefore, from the Portmanteau theorem [6], since $(\widehat{\sigma}_n)_{n\in\mathbb{N}}$ converges in distribution to σ_0^2, and by definition of $\lambda_\gamma(\tau)$, we have:

$$\lim_n \int Q_{d/2}\Big(\tau, \frac{\sigma}{\sigma_0}\lambda_\gamma(\tau)\Big)\mathbb{P}\widehat{\sigma}_n^{-1}(\mathrm{d}\sigma) = Q_{d/2}\big(\tau, \lambda_\gamma(\tau)\big) = \gamma$$

Hence:

$$\limsup_n \sup_{\Xi\in\mathcal{S}:\, \mathbb{P}[\|\Xi-\theta_0\|_2\le\sigma_0\tau]\neq 0} \mathbb{P}\big[\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}(\Xi+X, \theta_0, \widehat{\sigma}_n) = 1 \mid \|\Xi-\theta_0\|_2 \le \sigma_0\tau\big] \le \gamma$$

which is the desired result. □

C. Simulation results

We now illustrate the result stated in Theorem 2 with simulation results. These simulations were conducted with mono-dimensional signals only ($d = 1$). The objective is to show empirically that, as we get a better estimate of the noise variance, the measured false alarm rate tends to become lower than the specified false alarm rate γ. In these simulations, we first generate N independent Gaussian distributed samples $(R_1, ..., R_N)$ with variance ${\sigma_0}^2$. These samples are used to estimate ${\sigma_0}^2$ via the maximum likelihood ratio estimator of Eq. (6). This yields an estimate $\widehat{{\sigma_0}^2}$ of the noise variance. Then we generate one sample $Y \sim \mathcal{N}(\theta_0, \sigma_0^2)$ for a given value of θ_0 and the detection result is given by $\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}\big(Y, \theta_0, \widehat{{\sigma_0}^2}\big)$.

Figure 1 shows the measured false alarm rate for different values of the parameter γ and with different number of samples N used for estimating σ_0^2. We can see that for every value of γ, the measured false alarm rate decreases and tends to γ as N increases. Figure 2 confirms this behavior, where we see the evolution of the measured false alarm rate with N for several values of γ.

We also considered an extension of AsympRDT to the case when, in addition to estimating the noise variance, we also estimate the model θ_0 by using the samples (R_n). The detection is then performed using the estimated model $\widehat{\theta}_0 = \frac{1}{N}\sum_{i=1}^N R_i$ instead of θ_0. The simulation results when both θ_0 and σ_0 are estimated are presented in Fig. 3. We observe the same behavior as in Fig. 1: the measured false alarm rate tends to γ as N increases. As can be expected since two parameters were estimated instead of one only, we can notice a slight increase in the measured false alarm rate in comparison to the results of Fig. 1.

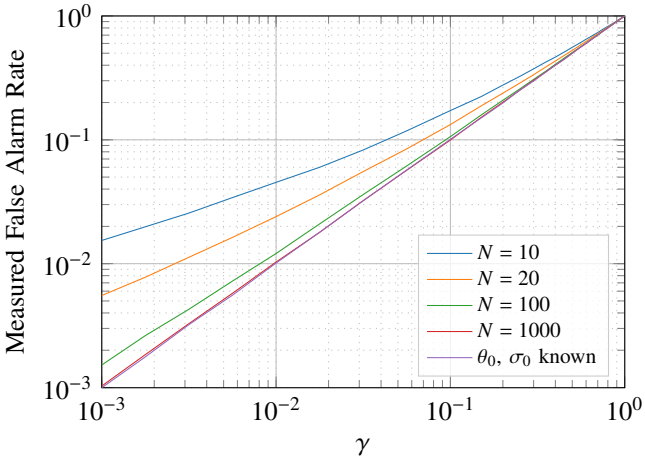

Fig. 3. Measured false alarm rate vs. parameter γ using N samples to estimate both θ_0 and σ_0 ($\tau = 0$, $d = 1$).

These results motivate an extension of the theoretical framework of Section III above to the case when both θ_0 and σ_0 are estimated and strongly suggest that Theorem 2 is likely to remain valid in that context.

IV. Change-in-Mean Detection Method

A. Method description

We now introduce our change-in-mean detection algorithm based on AsympRDT presented before. We consider a d-dimensional time series $(Y_n)_{n\in\mathbb{N}} \in \mathcal{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$ where every sample Y_n can be written as $Y_n = \Theta_n + X_n$ with $X_n \sim \mathcal{N}(0, \sigma_0^2 I_d)$ independent of $\Theta_n \in \mathcal{M}(\Omega, \mathbb{R}^d)$. The principle of our method is to process the signal sequentially in blocks of B samples and test if the mean of each of these blocks is close enough to the current estimated model ($\widehat{\theta}_0$ and $\widehat{{\sigma_0}^2}$). If

no change has been detected, we can then update the estimated model to improve our estimates of θ_0 and σ_0^2.

This algorithm has three parameters to be set by the user:

- the block-size B,
- the tolerance τ,
- the desired false alarm probability γ,

and can be decomposed in three steps:

1) model initialization;
2) test;
3) model update.

We start by estimating the initial model of the signal on a first block of B samples $(Y_n)_{0\leq n<B}$, yielding a first model $(\widehat{\mu}, \widehat{\sigma_0}^2)$. Then we test if the empirical mean of the next block $(Y_n)_{B\leq n<2B}$ lies close enough to $\widehat{\mu}$ given the tolerance τ.

Since the value of interest to be tested in this algorithm is the mean of a block of samples, the testing is a Block-RDT problem [7], that is, an RDT problem where the observation is the empirical mean of a given block of samples and the model is the mean θ_0 of the process. The optimal test for the Block-RDT problem is thus $\mathcal{T}_{\lambda_\gamma(\tau\sqrt{B})/\sqrt{B}}$ applied to the empirical mean of the observed samples. However this test is only usable when θ_0 and σ_0 are known, which is not the case here. AsympRDT and the pertaining simulations above show that we can replace the test $\mathcal{T}_{\lambda_\gamma(\tau\sqrt{B})/\sqrt{B}}$ with the test $\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau\sqrt{B})/\sqrt{B}}(\cdot, \widehat{\mu}, \widehat{\sigma_0}^2)$, which guarantees the false-alarm rate γ as long as the estimated model $(\widehat{\mu}, \widehat{\sigma_0}^2)$ is good enough.

Consequently, we use the following test to decide on the presence or the absence of a change:

$$\frac{\left|\frac{1}{B}\sum_{i=B}^{2B-1} Y_i - \widehat{\mu}\right|}{\widehat{\sigma_0}} \underset{\text{change}}{\overset{\text{no change}}{\lessgtr}} \frac{\lambda_\gamma(\tau\sqrt{B})}{\sqrt{B}} \tag{9}$$

If the threshold is not exceeded, we update the model estimate using the samples $(Y_n)_{B\leq n<2B}$. We then repeat the test phase with the next B samples of (Y_n). Otherwise, if the threshold is exceeded, a change in the mean of the signal has been detected. In this case, we forget the model estimated up to this point and estimate the new mean and variance from the next B samples. We then resume the testing with the block following these samples. It is worth noticing that we do not estimate the new model with the samples in which the change has been detected, since all the samples in this block do not necessarily have the same statistical properties. Figure 4 recapitulates the different steps of this algorithm.

It is also worth mentioning that by definition (see Eq. (2)), τ represents the maximum deviation allowed from the model normalized by the noise variance σ_0. One may instead want to choose a tolerance that does not depend on the noise variance. If we denote this desired tolerance as τ', the only required change in the algorithm is to replace the threshold $\lambda_\gamma(\tau\sqrt{B})/\sqrt{B}$ with $\widehat{\sigma_0}\lambda_\gamma(\tau'\sqrt{B}/\widehat{\sigma_0})/\sqrt{B}$, and recompute the latter whenever the noise variance estimate is updated.

Input: Time series Y_n, block size B, tolerance τ, desired false alarm probability γ
Output: Estimated instants of change

$T \leftarrow \lambda_\gamma(\tau\sqrt{B})/\sqrt{B}$ (Threshold)
$\widehat{\mu} \leftarrow \text{mean}(Y_0, ..., Y_{B-1})$ (Initial mean estimate)
$\widehat{\sigma_0}^2 \leftarrow \text{var}(Y_0, ..., Y_{B-1})$ (Initial variance estimate)
$n \leftarrow B$ (Number of samples in the current segment)
$i_s \leftarrow 0$ (Index of the first sample of the current segment)
repeat
  $s \leftarrow \text{mean}(Y_{i_s+n}, ..., Y_{i_s+n+B-1})$ (Current block mean)
  $z \leftarrow |s - \widehat{\mu}|/\widehat{\sigma_0}$ (Test statistic)
  if $z \leq T$ **then**
    $n \leftarrow n + B$ (Extend the current segment)
    $\widehat{\mu} \leftarrow \text{mean}(Y_{i_s}, ..., Y_{i_s+n+B-1})$ (Update the mean estimate)
    $\widehat{\sigma_0}^2 \leftarrow \text{var}(Y_{i_s}, ..., Y_{i_s+n+B-1})$ (Update the variance estimate)
  else
    Notify that a change has been detected between indices $i_s + n$ and $i_s + n + B - 1$
    $i_s \leftarrow i_s + n + B$ (Start the next segment after the end of the current block)
    $\widehat{\mu} \leftarrow \text{mean}(Y_{i_s}, ..., Y_{i_s+B-1})$ (Initial mean estimate)
    $\widehat{\sigma_0}^2 \leftarrow \text{var}(Y_{i_s}, ..., Y_{i_s+B-1})$ (Initial variance estimate)
    $n \leftarrow B$
  end if
until end of Y_n reached

Fig. 4. Proposed change-in-mean detection algorithm

B. Cybersecurity Application: Industrial Control Systems

In this section, we present an application of our algorithm for cybersecurity. We focus here on industrial control systems, comprising a physical process controlled by programmable logic controllers with sensors and actuators to monitor and affect its physical components [8]. These systems are at the core of many critical infrastructures and can be high-value targets for attackers. Stuxnet [9] is one of the first case of such an attack and demonstrated that many industrial systems are potentially vulnerable to these types of attacks as these systems are increasingly complex and connected, whereas security is not always part of their design.

In this respect we present some results of experiments conducted on the SWaT (Secure Water Treatment) dataset [10], which was recorded on a test bed aimed at reproducing a water treatment plant. Figure 5 displays the evolution during six hours under normal conditions of a water level sensor situated in a tank. The sampling rate is 1 Hz. We can see that this signal is piecewise linear and can be decomposed in four different phases repeated over time. The state of the water level (slow increase, fast increase, stationary, or fast decrease) depends directly on the state of the inflow and outflow pumps connected to this tank.

We want to build a reference model of the system normal operating conditions, for further use to detect anomalies and attacks. A first step towards this goal is to identify the different phases of each signal involved in the water treatment process. In the case of Fig. 5, this can be done by applying our change-

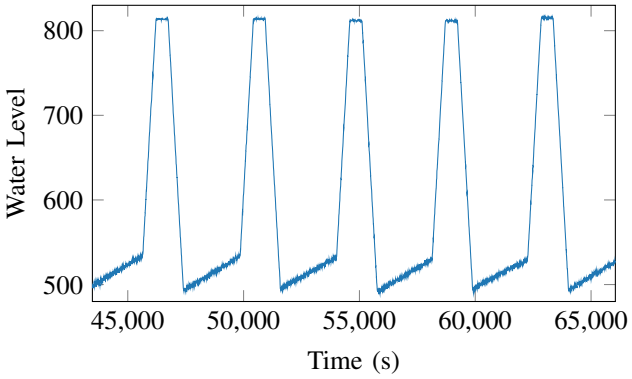

Fig. 5. Water level measurement in a water tank of the SWaT system under normal operating conditions (sensor LIT101).

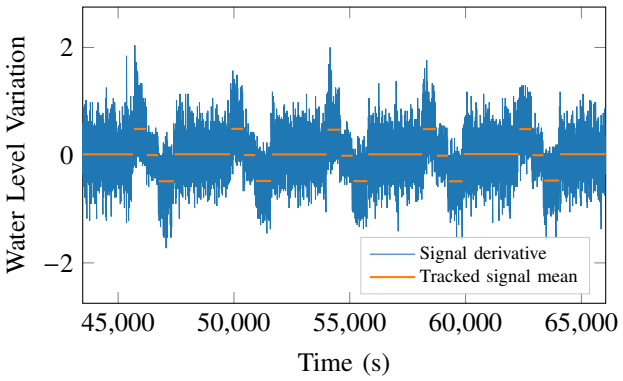

Fig. 6. Sensor LIT101 derivative and tracked mean using the algorithm described in Fig. 4 ($\gamma = 0.01$, $\tau = 0.1$, $B = 40$).

detection method to its derivative, which is piecewise constant as can be seen in Fig. 6. The orange signal is the output of our algorithm and exhibits the different phases that have been detected and the mean value of each phase. Comparing this output with the original signal in Fig. 5, we can see that it matches the variations of the signal. Out of 484 changes present in this signal, 466 have been detected and no false alarm occurred. Missed changes seem to occur when the amplitude of the change is small and if the change is temporary, i.e. the signal quickly returns to its previous value.

V. Conclusion and Perspectives

Theorem 2 and our simulation results show that we can get asymptotic performance guarantees when replacing the known noise variance σ_0^2 with an estimate. Based on this, we have built a change-detection method that inherits this property by design and can therefore asymptotically respect a given false-alarm rate. This approach opens many prospects, notably in cybersecurity. Indeed, it makes it possible to perform signal segmentation to describe the behavior of the system, for example by computing the expected variation or length of each phase.

However, a few points still need to be addressed to design a fully asymptotically optimal method. First, as mentioned in Section III-C, we can extend our theoretical study to the case where both θ_0 and σ_0 are estimated. Second, the result proved in Theorem 2 only concerns the size of the sequence of tests $\widetilde{\mathcal{T}}_{\lambda_\gamma(\tau)}$. In order to have optimality, we also need to study its asymptotic power. Third, the performance in terms of mean-time between false alarms and mean-time before correct change detection remains an open issue because of the estimators involved. Finally, regarding cybersecurity applications, although we require more data to refine our analysis and performance measurements in practice, a full anomaly and attack detection system based on the results presented in this paper is currently in the works.

References

[1] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, vol. 54, no. 3, pp. 426–482, 1943.

[2] D. Pastor and Q.-T. Nguyen, "Random Distortion Testing and Optimality of Thresholding Tests," *IEEE Transactions on Signal Processing*, vol. 61, pp. 4161–4171, Aug. 2013.

[3] Q.-T. Nguyen, *Contributions to Statistical Signal Processing with Applications in Biomedical Engineering*. PhD thesis, Télécom Bretagne, Université de Bretagne Occidentale, Nov. 2012.

[4] Y. Sun, Á. Baricz, and S. Zhou, "On the Monotonicity, Log-Concavity, and Tight Bounds of the Generalized Marcum and Nuttall Q-Functions," *IEEE Transactions on Information Theory*, vol. 56, pp. 1166–1186, Mar. 2010.

[5] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer Texts in Statistics, New York, NY: Springer, 3. ed ed., 2005. OCLC: 249833198.

[6] P. Billingsley, *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, New York, NY: Wiley, 2. ed ed., 1999.

[7] D. Pastor and Q.-T. Nguyen, "Robust statistical process control in Block-RDT framework," pp. 3896–3900, IEEE, Apr. 2015.

[8] V. L. Do, "Statistical detection and isolation of cyber-physical attacks on SCADA systems," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 3524–3529, Oct. 2017.

[9] N. Falliere, L. O. Murchu, and E. Chien, "W32.Stuxnet Dossier," Feb. 2011. Symantec.

[10] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A Dataset to Support Research in the Design of Secure Water Treatment Systems," in *Critical Information Infrastructures Security* (G. Havarneanu, R. Setola, H. Nassopoulos, and S. Wolthusen, eds.), Lecture Notes in Computer Science, pp. 88–99, Springer International Publishing, 2017.