Higher-Order Moment-Based Anomaly Detection

Venkatraman Renganathan, Navid Hashemi, Justin Ruths, and Tyler H. Summers

Abstract—The identification of anomalies is a critical component of operating complex, large-scale and geographically distributed cyber-physical systems. While designing anomaly detectors, it is common to assume Gaussian noise models to maintain tractability; however, this assumption can lead to the actual false alarm rate being significantly higher than expected. Here we design a distributionally robust threshold of detection using finite and fixed higher-order moments of the detection measure data such that it guarantees the actual false alarm rate to be upper bounded by the desired one. Further, we bound the states reachable through the action of a stealthy attack and identify the trade-off between this impact of attacks that cannot be detected and the worst-case false alarm rate. Through numerical experiments, we illustrate how knowledge of higher-order moments results in a tightened threshold, thereby restricting an attacker's potential impact.

I. INTRODUCTION

From critical infrastructures and industrial process control to autonomous driving and various biomedical applications, dynamical control systems are increasingly able to be instrumented with new sensing and actuation capabilities. These cyber-physical systems (CPS) comprise growing webs of interconnected feedback loops and must operate efficiently and resiliently in dynamic and uncertain environments. As these systems become large, devising both model-based [1], [2] and data-driven [3]–[5] methods for detecting anomalies (such as component failures or malicious attacks) are critical for their robust and efficient operation. Such critically important cyber-physical networks have become an attractive target to attackers. These systems are large and complex and are often not monitored well enough, enabling attackers to manipulate the system without being detected and cause damage [6], [7].

To simplify the analysis and design, often such complex cyber-networks are modeled as a discrete-time linear time invariant system with Gaussian noises. However, this can lead to a significant miscalculation of probabilities and risk if the underlying processes behave differently, for example due to various nonlinearities or malicious attacks. In the context of attacks, it is possible for an attacker to modify the sensor outputs and effectively generate aggressive and strategic noise profiles to sabotage the operation of the system. With stochastic optimization techniques, particularly using the emerging area of distributionally robust optimization (DRO) approaches [8], these limitations can be recognized and addressed. DRO enables modelers to explicitly incorporate inherent ambiguity

in probability distributions into optimization problems. DRO approaches can be categorized based on the form of the ambiguity set. There are several different parameterizations, including those based on moments, support, directional derivatives [9], and Wasserstein balls [10]. In practice, we have access to only a finite amount of historical data. However, it is possible to use finite historical data and guarantee resiliency in such critical cyber-physical networks. Here, we propose to use moment-based DRO methods to improve modeling and reduce false alarm rates in cyber-physical networks.

In the context of attacks, the detector tuning has a direct implication on the effect an attacker can have while still remaining stealthy [2], [11]. A model-based approach to attack detection uses a detector that raises alarms when there is a large enough discrepancy between the actual and predicted measurements, a statistic termed the residual. The detector's sensitivity can be increased by decreasing the threshold of detection, but there is an inherent trade-off between sensitivity and the rate at which false alarms are generated. Keeping false alarms to a manageable level requires adjusting sensitivity and the tuning of the detector threshold is typically informed by the distribution of the residual. Authors in [12] used Gaussian Mixture Model to approximate the arbitrary noise distributions and obtained a detector threshold corresponding to a desired false alarm rate.

However, when noise distributions are only known to an ambiguity set, traditional tools and approximations no longer suffice to select the threshold and so we turn to a distributionally robust approach. Interest in a DRO-informed perspective on detector tuning is supported by recent work on using a Wasserstein metric [5]. Our moment-based ambiguity set formulation includes all distributions with fixed moments up to some order. The problem of designing anomaly detector thresholds subject to moment constraints of system uncertainties can be addressed using Generalized Moment Problems described in [13], [14]. Further, the conditions for a truncated (finite) moment sequence to represent a probability measure were studied in [15], [16]. Authors in [17] proposed semidefinite programs to compute a probability bound for a random variable lying in a set with known moments up to some order. These techniques can be utilized to design detector thresholds for residual distributions consistent with finite fixed moments up to some order.

Contributions: This paper is a significant extension of our previous work [3] where we used a moment-based ambiguity set formulation with fixed first two moments (Proposition 2) to obtain a detector threshold via generalized Chebyshev inequality. The main contributions of the present paper are:

- We propose an approach to construct moment-based ambiguity sets with fixed moments up to k^{th} order for the anomaly detection measure and design an anomaly

detector threshold for CPSs that exhibit non-Gaussian uncertainties. The approach can utilize either residual moments obtained through a dynamic model or estimated directly from residual data.

- We use a semidefinite program (SDP) defined using higher-order moments of the detection measure to find a sharper probability bound for classifying the residual with an improved detector threshold (Theorem 4).
- We illustrate empirically that inclusion of higher-order moments results in tightened threshold (Lemma 3), thereby restricting an attacker's potential impact, and also prove that the volume of the attack-reachable set shrinks with a tightened threshold (Corollary 5).

While anomaly detection is widely studied in CPS literature, our distributionally robust approach using higher-order moments of the detection measure data marks the novel contribution of this paper. The rest of the paper is organized as follows. Section II formulates the problem using moment-based ambiguity sets. In Section III, the design of anomaly detector threshold is discussed. Section IV describes the procedure to find the boundary of the reachable sets obtained using distributionally robust tuned detector. Section V presents the numerical results with inferences. Finally, Section VI concludes and summarizes the future research directions.

II. PROBLEM FORMULATION USING HIGHER-ORDER MOMENT-BASED AMBIGUITY SETS

In this section, we propose a framework for designing an anomaly detector threshold for cyberphysical systems. The approach can utilize either model-based propagation of residual moments or data-driven estimation of detection measure moments directly from data.

A. Model-based Problem Formulation

Here, we model an uncertain cyber-physical system as a stochastic discrete-time linear system

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad t \in \mathbb{N} \quad (1)$$
$$y_t = Cx_t + v_t, \quad (2)$$

where $x_t \in \mathbb{R}^n, u_t \in \mathbb{R}^m$ are the system state and input respectively at time t. The matrices A and B denote the system matrix and control input matrix, respectively. The output $y_t \in \mathbb{R}^p$ aggregates a linear combination of the states with the observation matrix $C \in \mathbb{R}^{p \times n}$. We assume that the pair (A, C) is detectable and (A, B) is stabilizable. The process noise $w_t \in \mathbb{R}^n$ and the sensor noise $v_t \in \mathbb{R}^p$ are modeled as zero-mean random vectors independent and identically distributed across time with covariance matrix Σ_w, Σ_v respectively. Let $\kappa = (k_1, \ldots, k_n)^\top$ with $k_j \in \mathbb{Z}_+$ non-negative integers and $J_k = \{\kappa | \sum_{i=1}^n k_i \le k\}$. Further, assume that the feasible first k moment sequence of the distributions of the w_t, v_t namely $M_w^\kappa, M_v^\kappa, \kappa \in J_k$ respectively are known. The distributions P_w of w_t and P_v of v_t are unknown (and not necessarily Gaussian and possibly heavy-tailed[1]) and will be assumed to belong to

[1]For the purposes of this paper, we consider heavy-tailed distributions as those whose moments above a certain order may be infinite, in which case their tails are heavier than a Gaussian. We assume the moments up to order k are finite.

the k-moments-based ambiguity sets of distributions $\mathcal{P}_k^w$ and $\mathcal{P}_k^v$ respectively defined as follows

$$\mathcal{P}_k^w = \{P_w \mid \mathbb{E}[w_{t_1}^{k_1} w_{t_2}^{k_2} \ldots w_{t_n}^{k_n}] = M_w^\kappa, \; \kappa \in J_k\}, \quad (3)$$
$$\mathcal{P}_k^v = \{P_v \mid \mathbb{E}[v_{t_1}^{k_1} v_{t_2}^{k_2} \ldots v_{t_p}^{k_p}] = M_v^\kappa, \; \kappa \in J_k\}. \quad (4)$$

When the actual measurement y_t is corrupted by an additive attack, $\delta_t \in \mathbb{R}^p$, the true output of the system fed to the controller becomes

$$\bar{y}_t = y_t + \delta_t = Cx_t + v_t + \delta_t. \quad (5)$$

We utilize a steady-state Kalman filter to construct a state estimate $\hat{x}_t$ to minimize the squared norm of the estimation error $e_t = x_t - \hat{x}_t$ where,

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + L(\bar{y}_t - C\hat{x}_t), \quad (6)$$

and the estimation error evolves as

$$e_{t+1} = (A - LC)e_t + w_t - Lv_t - L\delta_t. \quad (7)$$

In the absence of attacks (that is, $\delta_t = 0$) and when the covariance matrices of the noises are fixed and known, the Kalman gain $L = PC^\top(CPC^\top + \Sigma_v)^{-1}$ minimizes the steady state covariance matrix

$$P := \lim_{t \to \infty} P_t := \mathbf{E}[e_t e_t^\top]. \quad (8)$$

Since (A, C) is assumed to be detectable, the existence of P is guaranteed and it can be found through the solution of an algebraic Ricatti equation. We define a residual sequence r_t as the difference between the actual received output $\bar{y}_t$ and the predicted output $C\hat{x}_t$ as,

$$r_t = \bar{y}_t - C\hat{x}_t = Ce_t + v_t + \delta_t, \quad (9)$$

and in the attack free setting, r_t falls according to a zero mean distribution with covariance

$$\Sigma_r = \mathbf{E}[r_t r_t^\top] = CPC^\top + \Sigma_v. \quad (10)$$

Since (9) is linear, it is possible to obtain the fixed and first k moments of the random variable r_t by propagating the corresponding moments of the primitive random variables w_t, v_t. Thus, the distribution P_r of r_t (not necessarily Gaussian) belongs to an ambiguity set $\mathcal{P}_k^r$ given by

$$\mathcal{P}_k^r = \{P_r \mid \mathbb{E}[r_{t_1}^{k_1} r_{t_2}^{k_2} \ldots r_{t_p}^{k_p}] = M_r^\kappa, \; \kappa \in J_k\}. \quad (11)$$

We define detection measure q_t as a quadratic function of r_t

$$q_t = r_t^\top \Sigma_r^{-1} r_t, \quad (12)$$

which will be compared to a threshold for anomaly detection.

B. Data-Driven Moment Estimation from Residual Data

An alternative to obtaining residual moments by propagating the moments of primitive random variables through the system model is to instead collect residual data r_t (from attack-free operation) and estimate residual moments or the moments of q_t directly from the data. Such a data-driven approach allows our proposed tuning approaches to be used in much broader settings where it is difficult to propagate moments through a model (or even to obtain a model), but where residual

data is easily generated from sensors and a state estimator. Higher-order moments require increasingly more data to obtain accurate estimates. Determining the required amount of data is possible using finite-sample measure concentration results given as in [18], but we leave such an analysis for future work. The proposed methods for setting thresholds for anomaly detection can be used together with ambiguity sets built upon data-driven residual estimates, although the false alarm rates will also be affected by sampling errors. Moment estimation uncertainty could be accommodated using the same generalized moment problem computations we propose here, but with assumed *bounds* on moment estimates rather than having them fixed to exact known values. Such problems can also be reformulated in a computationally tractable manner, as described in, e.g., Section 3 of [19], which incorporate uncertainty sets for estimated moments. These formulations would simply impose further constraints on our primal problem defined in subsection III.B of our revised manuscript. It is possible and would be interesting to explore how to use statistical confidence intervals for data-driven moment estimates to inform bounds used in the SDP and obtain end-to-end guarantees on false alarm rates associated with certain thresholds. This will be pursued in future work.

In either setting, the *feasible* first k moment sequence of the distribution of q_t denoted by M_q^k is assumed to be known either from its primitive variables through the model-based approach or estimated from data. Then, the k-moments-based ambiguity set of the scalar random variable q_t is defined as

$$\mathcal{P}_k^q = \{P_q \mid \mathbb{E}[q_t^k] = M_q^k\}. \tag{13}$$

Proposition 1. *[17] Consider a univariate random variable X defined on $\Omega = \mathbb{R}_+$, endowed with its Borel sigma algebra of events. A sequence $\bar{\sigma} = (M_1, M_2, \ldots, M_k)^\top$ is a **feasible** $(1, k, \Omega)-$moment vector of the random variable X, if and only if the matrices $R_k \succeq 0$ and $R_{k-1} \succeq 0$, where for any integer $l \geq 0$, the matrices are defined as*

$$R_{2l} = \begin{bmatrix} 1 & M_1 & \ldots & M_l \\ M_1 & M_2 & \ldots & M_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ M_l & M_{l+1} & \ldots & M_{2l} \end{bmatrix}, R_{2l+1} = \begin{bmatrix} M_1 & \ldots & M_{l+1} \\ M_2 & \ldots & M_{l+2} \\ \vdots & \ddots & \vdots \\ M_{l+1} & \ldots & M_{2l+1} \end{bmatrix}. \tag{14}$$

III. DESIGN OF ANOMALY DETECTOR THRESHOLDS

Given a threshold [2]$\alpha_{q,k} > 0$ and the distance measure q_t, alarm time(s) $t^\star$ are produced according to the following rules,

$$\begin{cases} q_t \leq \alpha_{q,k}, & \text{no alarm} \\ q_t > \alpha_{q,k}, & \text{alarm: } t^\star = t. \end{cases} \tag{15}$$

Even in the absence of attacks, the detector is expected to generate false alarms due to the infinite support of v_t, because some values drawn from P_q will exceed the threshold $\alpha_{q,k}$. If P_q is known, then it is possible to extract an optimum threshold value α_q^* from the corresponding cumulative distribution function F_q for a desired false alarm rate, $\mathcal{A}$. For example,

---

[2]The first and second subscripts in the threshold separated by a comma denote the random variable and number of moments respectively.

if r_t is Gaussian, q_t would be a chi-squared random variable, and the optimum threshold $\alpha_q^\star$ corresponding to the desired false alarm rate $\mathcal{A} = \mathcal{A}^\star$ is then

$$\alpha_q^\star = \alpha_{\chi^2} := 2\mathbb{P}^{-1}\left(1 - \mathcal{A}^\star, \frac{p}{2}\right), \tag{16}$$

where $\mathbb{P}^{-1}(\cdot, \cdot)$ denotes the inverse regularized lower incomplete gamma function [1]. When the complete distribution is not available, tuning methods using (16) may design thresholds that generate actual false alarm rates significantly higher than what is desired. With the distributionally robust approach, we aim to achieve a false alarm rate less than $\mathcal{A}$, and the detector threshold $\alpha_{r,k}^\star$ is selected such that

$$\sup_{P_r \in \mathcal{P}_k^r} P_r(r_t^\top \Sigma_r^{-1} r_t \leq \alpha_{r,k}^\star) = 1 - \mathcal{A}. \tag{17}$$

When the first two moments $(k = 2)$ of r_t are known, the following proposition using the generalized Chebyshev inequality explained in [20], [21] can be used to obtain the worst case detector threshold $\alpha_{r,2}^\star$ satisfying (17).

Proposition 2. *[3] Given a desired false alarm rate $\mathcal{A}$ and $r_t \sim P_r \in \mathcal{P}_k^r$ with $k = 2$, the optimal distributionally robust threshold $\alpha_{r,2}^\star$ satisfying (17) is*

$$\alpha_{r,2}^\star = \frac{p}{\mathcal{A}}. \tag{18}$$

A. Improved Detector Threshold With Higher-Order Moments

It is possible to obtain a sharpened detector threshold than $\alpha_{r,2}^\star$, if higher-order moments are taken into account. For example, the skewness and kurtosis parameters convey asymmetry and heaviness of tails of the distribution, respectively. We can leverage such information about the true but unknown distribution revealed by the higher-order moments to tighten the required probability bound and thereby obtain an improved detector threshold. It is possible to use r_t with higher order moments (first k moments) to obtain a sharpened detector threshold $\alpha_{r,k}^\star$. However, it was shown in [17] that for $r_t \in \mathbb{R}^p$ with support $\Omega = \mathbb{R}^p$ and $k \geq 4$, it is NP-hard to find tight bounds for the corresponding moment bound problem with rational problem data. On the other hand, they provide a semidefinite optimization problem in $k + 1$ dimension for the case of a univariate random variable with k moments. Hence, rather looking for higher-order moments of r_t, instead we look for higher-order moments of scalar random variable q_t. Subsequently, we use a bisection algorithm to obtain a sharpened detection threshold $\alpha_{q,k}^\star$ for a given $\mathcal{A}$.

B. Estimating Probability Using $(1, k, \Omega)-$Moment Bound

Given the first k moments of random variable q_t with support $\Omega = \mathbb{R}_{\geq 0}$ and the set $\mathcal{S}_k = \mathbb{R}_{>\alpha_{q,k}}$ representing an alarm event as shown in Figure 1, the infinite dimensional $(1, k, \Omega)-$moment bound primal problem is given by

$$\sup_{P_q \in \mathcal{P}_k^q} P_q(q_t \in \mathcal{S}_k)$$
$$\text{s.t.} \quad \int_\Omega q^r d\mu = M_q^r, \quad r = 0, 1, \ldots, k. \tag{19}$$

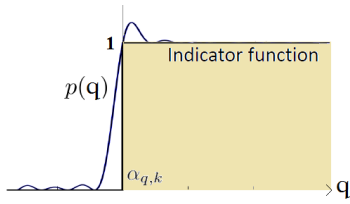Fig. 1: A moment based polynomial bounding the indicator function $\mathbf{1}_{\mathcal{S}_k}$ representing the set $\mathcal{S}_k = \mathbb{R}_{>\alpha_{q,k}}$ is shown here.

Algorithm 1: `Bisection Subroutine`

---

while $\alpha_u - \alpha_l > \epsilon$ **do**
  $\alpha_{q,k}^\star \leftarrow (\alpha_u + \alpha_l)/2$
  $p_{min} \leftarrow$ optimal value of (23) with $\alpha_{q,k} = \alpha_{q,k}^\star$
  if $p_{min} > \mathcal{A}$ **then**
    $\alpha_l \leftarrow \alpha_{q,k}^\star$
  else
    $\alpha_u \leftarrow \alpha_{q,k}^\star$

Since (19) is infinite dimensional, it is difficult to solve efficiently. However, we can use the linear programming duality theory to associate a dual variable $y_r, r = 0, 1, \ldots, k$ with each equality constraint of (19) to get the corresponding dual problem

$$\min \quad \sum_{r=0}^{k} y_r M_q^r$$

$$\text{s.t.} \quad p(q) = \sum_{r=0}^{k} y_r q^r \geq 1, \quad \forall q \in \mathcal{S}_k, \qquad (20)$$

$$p(q) = \sum_{r=0}^{k} y_r q^r \geq 0, \quad \forall q \in \Omega.$$

The probability obtained as the solution to (20) is an upper bound to the probability associated with (19). In general, if the moment vector $\bar{\sigma}_k = (M_q^0, M_q^1, \ldots M_q^k)$ is an interior point of the set $\mathcal{M}_k$ of all feasible moment vectors, then strong duality exists between (19) and (20) enabling us to obtain a *tight* bound on $P_q(q_t \in \mathcal{S}_k)$. To achieve a desired false alarm rate $\mathcal{A}$, we can tune the threshold $\alpha_{q,k}$ defining the set $\mathcal{S}_k$ such that solution to (20) is $\mathcal{A}$. The following lemma establishes the general trend observed between the values of the tuned threshold $\alpha_{q,k}$ for increasing values of k.

Lemma 3. *Let $\mathcal{A}$ be the desired false alarm rate and $q_t \sim P_q \in \mathcal{P}_k^q$ as in (13). Then, $\forall j, k \in \mathbb{Z}_{\geq 1}$ with $j < k$, we have*

$$\alpha_{q,k}^\star \leq \alpha_{q,j}^\star \qquad (21)$$

when the associated moments agree up to order j.

Proof. When the associated moments up to order $j < k$ agree, clearly we have $\mathcal{P}_k^q \subseteq \mathcal{P}_j^q$, since additional moment constraints restrict the set of distributions. For a fixed threshold $\alpha_{q,j}^\star$ tuned for the false alarm rate $\mathcal{A}$, it follows that

$$\sup_{P_q \in \mathcal{P}_k^q} P_q(q_t > \alpha_{q,j}^\star) \leq \sup_{P_q \in \mathcal{P}_j^q} P_q(q_t > \alpha_{q,j}^\star) = \mathcal{A}. \qquad (22)$$

Thus, to achieve a desired false alarm rate $\mathcal{A}$ under the constraint that $P_q \in \mathcal{P}_k^q$, the threshold $\alpha_{q,k}^\star$ must be non-increasing in k, satisfying $\alpha_{q,k}^\star \leq \alpha_{q,j}^\star$ for $j < k$. $\qquad \square$

Theorem 4. *Let $\epsilon > 0$, $k \in \mathbb{Z}_{\geq 2}$ and consider q_t with its first k moments of distribution $(M_q^1, M_q^2, \ldots, M_q^k)$ (we let $M_q^0 = 1$) defined on $\mathbb{R}_+$ being known (or estimated from the given data) and the associated anomaly detector threshold $\alpha_{q,k}$ introduced in (15), which is intended to achieve a desired false alarm rate $\mathcal{A}$. Suppose that $\alpha_{q,k}$ is obtained by solving the following bisection algorithm (knowing that the desired $\alpha_{q,k}^\star \in [\alpha_l, \alpha_u]$*

and $\alpha_u = \alpha_{q,k-1}^\star$ by Lemma 3) where the SDP that gives the tight upper bound on $P_q(q_t \geq \alpha_{q,k})$ in the third line is

$$\min \quad \sum_{r=0}^{k} y_r M_q^r$$

$$\text{s.t.} \quad \sum_{i,j:i+j=2l-1} x_{ij} = 0, \qquad\qquad l = 1, \ldots, k,$$

$$(y_0 - 1) + \sum_{r=1}^{k} y_r \alpha_{q,k}^r = x_{00},$$

$$\sum_{r=l}^{k} y_r \binom{r}{l} \alpha_{q,k}^{r-l} = \sum_{i,j:i+j=2l} x_{ij}, \qquad l = 1, \ldots, k,$$

$$X \succeq 0,$$

$$\sum_{i,j:i+j=2l-1} z_{ij} = 0, \qquad\qquad l = 1, \ldots, k,$$

$$\sum_{r=0}^{l} y_r \binom{k-r}{l-r} \alpha_{q,k}^{r-l} = \sum_{i,j:i+j=2l} z_{ij}, \quad l = 0, \ldots, k,$$

$$Z \succeq 0.$$
$$(23)$$

with variables $X, Z \in \mathbb{R}^{(k+1)\times(k+1)}, y_r, r = 0, 1, \ldots, k$. Then with this optimal detector threshold $\alpha_{q,k}^\star$, the false alarm rate under the worst-case distribution of the q_t specified by the k-moments-based ambiguity set (13) is at most $\mathcal{A}$.

Proof. To get an upper bound on required probability in (19), it suffices to check for the polynomials defined in (20) to be non-negative in their respective sets. That is, the polynomial $p(q)$ satisfies $p(q) - 1 \geq 0, \forall q \in \mathcal{S}_k$ if and only if there exists a positive semidefinite matrix $X \in \mathbb{R}^{(k+1)\times(k+1)}$ that satisfies the first four constraints defined in (23). Similarly, the same polynomial $p(q)$ satisfies $p(q) \geq 0, \forall q \in [0, \alpha_{q,k}]$ if and only if there exists a positive semidefinite matrix $Z \in \mathbb{R}^{(k+1)\times(k+1)}$ that satisfies the last three constraints defined in (23). Then using the results given in Theorem 3.2 of [17], for a fixed threshold $\alpha_{q,k}$, the optimal value of the semidefinite program in (23) yields the worst-case false alarm rate generated by the worst-case $P_q \in \mathcal{P}_k^q$ in (13). To ensure that the optimal value of (23) is equal to the desired false alarm rate $\mathcal{A}$, it suffices to tune the threshold $\alpha_{q,k}$ using the bisection algorithm 1, to get $\alpha_{q,k}^\star$. $\qquad \square$

C. Discussion

For $k = 1, 2, 3$, the solution to (23) can be obtained in closed-form using Theorem 3.3 of [17] with the corresponding

moments data without explicitly solving the SDP in (23). For $k = 1, 2$, we recover the Markov bound and a strictly improved Chebyshev bound respectively as the solution of (23). Assuming that $\hat{\delta}_k > 0$, $\alpha_{q,k}^\star = (1 + \hat{\delta}_k)M_q^1$, the resulting closed form solutions for the optimal detector threshold $\alpha_{q,k}^\star$ for a desired false alarm rate $\mathcal{A}$ with $k = 1, 2$ are summarized in Table I with $C_M^2 = \frac{M_q^2 - (M_q^1)^2}{(M_q^1)^2}$ and these values can be used as α_u towards computing $\alpha_{q,k}^\star$ for any $k > 2$. We omit the expression for the threshold $\alpha_{q,3}^\star$ for the sake of brevity. Further, the optimal threshold $\alpha_{q,k}^\star$ is tight for a given $\mathcal{P}_k^q$ up

k	$\alpha_{q,k}^\star$ value
1	$\frac{M_q^1}{\mathcal{A}}$
2	$\left(1 + \sqrt{\frac{1-\mathcal{A}}{\mathcal{A}}}\, C_M\right) M_q^1$

TABLE I: Closed-form solutions for $\alpha_{q,k}^\star$ with $k = 1, 2$.

to a tolerance $\epsilon > 0$ specified by the bisection in Algorithm 1.

The exact detector threshold (which we call $\alpha_q^\star$) corresponding to P_q is obtained only asymptotically (or equivalently when true P_q is known exactly). That is,

$$\alpha_q^\star = \begin{cases} \lim_{k \to \infty} \alpha_{q,k}^\star, & \text{if all moments are known,} \\ F_q^{-1}(\mathcal{A}), & \text{if true } P_q, F_q \text{ are known.} \end{cases} \quad (24)$$

However, while determining how many moments are needed to get a close approximation of the exact threshold is difficult in general, we find that significant improvements can be obtained from a small number of moments. Given k moments of q_t, the complexity of this higher-order moment based approach involves solving the SDP given by (23) with variables $X, Z \in \mathbb{R}^{(k+1)\times(k+1)}$, $y_r, r = 0, 1, \ldots, k$. For large k, it is advisable to use, e.g., the Legendre polynomial basis instead of the standard polynomial basis for obtaining the moment-based polynomial in (23) as the former has nice orthogonal properties that improve numerical stability.

IV. ATTACK-REACHABLE SET BOUNDS

With the attacker assumed to have perfect knowledge of the system dynamics, the Kalman filter, control inputs, measurements along with read and write access to all the sensors at each time step, we show that the volume of the attack-reachable set shrinks with the tightened threshold. We define a zero-alarm attack, which generates attack sequences so that no alarms are raised during attack. With $\Sigma_r^{\frac{1}{2}}$ being the symmetric square root of Σ_r and an attack input $\bar{\delta}_t$ such that $\bar{\delta}_t^\top \bar{\delta}_t \leq \alpha$ (here $\alpha = \alpha_{q,k}^\star$), the attack sequence $\delta_t = -Ce_t - v_t + \Sigma_r^{\frac{1}{2}} \bar{\delta}_t$ guarantees that no alarm is raised. Then, using a static estimator feedback $u_t = K\hat{x}_t$ with $\bar{\delta}_t$, the evolution of the system dynamics with the joint state $\xi_t = [x_t, e_t]^\top$ with input $\zeta_t = [w_t, \bar{\delta}_t]^\top$ is studied. We define a reachable set of interest, driven by the ellipsoidally bounded inputs w_t and $\bar{\delta}_t$, as

$$\mathcal{R}_x = \left\{ x_t = [I_n, \ 0_{n\times n}]\xi_t \ \middle| \ \begin{array}{l} \xi_{t+1} = \hat{A}\xi_t + \hat{B}\zeta_t, \\ \xi_1 = \mathbf{0}, \ \bar{\delta}_t^\top \bar{\delta}_t \leq \alpha, \\ w_t^\top \Sigma_w^{-1} w_t \leq \bar{w}, \ \forall t \in \mathbb{N} \end{array} \right\},$$
$$(25)$$

where the noise threshold $\bar{w}$ obtained using (18) satisfies,

$$\sup_{P_w \in \mathcal{P}_k^w} P_w(w_t^\top \Sigma_w^{-1} w_t \leq \bar{w}) = 1 - \mathcal{A}, \quad (26)$$

and $\hat{A} = \begin{bmatrix} A + BK & -BK \\ 0 & A \end{bmatrix}$, $\hat{B} = \begin{bmatrix} I & 0 \\ I & -L\Sigma_r^{1/2} \end{bmatrix}$. Using the geometric approach presented in [22] with $H_i = A_{cl}^i - A^i$, the reachable set of states is the Minkowski sum of the following ellipsoidal bound

$$\mathcal{R}_{x,t}(\bar{w}, \alpha) = \bigoplus_{i=0}^{t-2} \mathcal{E}\left(\bar{w}A^i\Sigma_w A^{i\top}\right) \oplus \mathcal{E}\left(\alpha H_i L\Sigma_r L^\top H_i^\top\right). \quad (27)$$

Specifically, theorem 1 of [22] provides us the exact boundary of Minkowski sum in (27) using an analytical formula. The following corollary highlights the effects of the tightened threshold on the size of the reachable set.

Corollary 5. *The volume of the reachable set $\mathcal{R}_{x,t}(\bar{w}, \alpha)$ shrinks with the tightened threshold $\alpha_{q,k}^\star$ for all $k \geq 1$. Further, $\forall k_1, k_2 \geq 1$ and $k_1 < k_2$, the following inclusion holds $\mathcal{R}_{x,t}(\bar{w}, \alpha_{q,k_1}^\star) \supseteq \mathcal{R}_{x,t}(\bar{w}, \alpha_{q,k_2}^\star)$.*

Proof. Apply (27) with $k_1, k_2 \geq 1$ & $k_1 < k_2$ along with $\alpha_{q,k_1}^\star > \alpha_{q,k_2}^\star$ (by Lemma 3), to see that volume$(\mathcal{R}_{x,t}(\bar{w}, \alpha_{q,k_2}^\star)) \leq$ volume$(\mathcal{R}_{x,t}(\bar{w}, \alpha_{q,k_1}^\star))$. Inclusion follows from Minkowski sum with $\alpha_{q,k_1}^\star > \alpha_{q,k_2}^\star$. $\square$

V. NUMERICAL SIMULATION

We consider an empirical system under study with the detector tuned to a false alarm rate $\mathcal{A} = 0.05$ (5%). We demonstrate here simulation results when the uncertainties are zero-mean Gaussian. We compare the size of the reachable set boundary computed using (27) for thresholds $\alpha_{q,1}^\star, \alpha_{q,2}^\star, \alpha_{q,4}^\star$. We assume that the modeler is unaware of the functional form of the uncertainties and has to arrive at a detector threshold satisfying the desired false alarm rate. The SDP in (23) was solved with $\epsilon = 10^{-4}$ using SOSToolbox in Matlab with SeDuMi solver.

$$A = \begin{bmatrix} 0.84 & 0.23 \\ -0.47 & 0.12 \end{bmatrix}, B = \begin{bmatrix} 0.07 & -0.32 \\ 0.23 & 0.58 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$
$$K = \begin{bmatrix} 1.404 & -1.402 \\ 1.842 & 1.008 \end{bmatrix}, L = \begin{bmatrix} 0.0276 & 0.0448 \\ -0.01998 & -0.0290 \end{bmatrix},$$
$$\Sigma_v = 6I_p, \Sigma_w = \begin{bmatrix} 0.0225 & -0.0055 \\ -0.0055 & 0.0100 \end{bmatrix}.$$

When the noises w_t and v_t are truly Gaussian, it is evident from Fig. 2 that the reachable set corresponding to the thresholds $\alpha_{q,1}^\star = \alpha_{r,2}^\star = 40$ is conservative and ensures that the false alarm does not exceed 5% but this also provides the attacker with the ability to launch a larger attack. However, with the knowledge of additional moments, the detector thresholds $\alpha_{q,2}^\star = 10.7684$, $\alpha_{q,4}^\star = 9.1315$ get tightened with $\alpha_{q,4}^\star \leq \alpha_{q,2}^\star \leq \alpha_{q,1}^\star$ as shown in Fig. 3. Further, this threshold tightening limits the attacker's ability to launch a larger attack which is depicted through the reachable sets corresponding to the thresholds $\alpha_{q,1}^\star, \alpha_{q,2}^\star, \alpha_{q,4}^\star$ as shown in Fig. 2. Subsequently, the false alarm rate corresponding to the threshold α_{χ^2} was 5% as expected and with the thresholds $\alpha_{q,4}^\star, \alpha_{q,2}^\star, \alpha_{q,1}^\star$, it
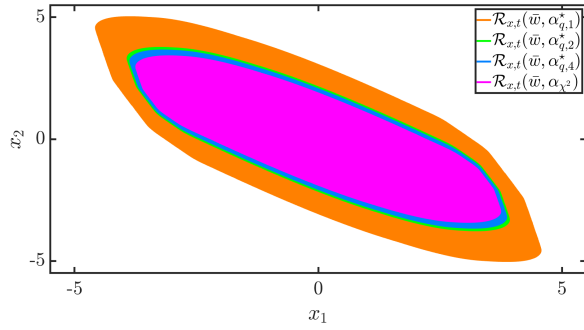
Fig. 2: The reachable sets associated with the detector thresholds $\alpha_{q,1}^\star, \alpha_{q,2}^\star, \alpha_{q,4}^\star, \alpha_{\chi^2}$ are shown in orange, green, blue and magenta colors respectively. It is evident from the size of the reachable set that given a desired false alarm rate $\mathcal{A}$, the knowledge of higher-order moments results in a tightened detector threshold and thereby restricting the attacker's ability to launch a larger attack.
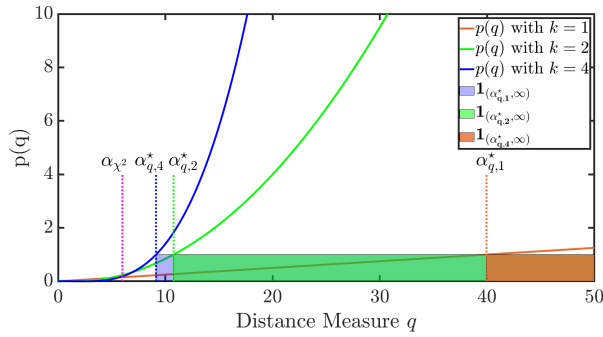

Fig. 3: The moment based polynomials in orange, green and blue bounding their respective indicator functions $\mathbf{1}_{(\alpha_{q,1}^\star,\infty)}, \mathbf{1}_{(\alpha_{q,2}^\star,\infty)}, \mathbf{1}_{(\alpha_{q,4}^\star,\infty)}$ in shaded orange, green and blue colors are shown here. Clearly, $\alpha_{q,1}^\star$ is very conservative and with $k = 4$, the threshold $\alpha_{q,4}^\star$ starts getting closer to the true threshold α_{χ^2} given by (16).

dropped to $1\%, 0.45\%, 0\%$ respectively. When noises are truly multi-variate Laplacian (which has heavier tails than normal distribution with same mean and covariance), it resulted in thresholds $\alpha_{q,1}^\star = 39.83$, $\alpha_{q,2}^\star = 17.23, \alpha_{q,4}^\star = 16.54$ with false alarm rates $0\%, 0.9\%, 1\%$ respectively. Thus, no matter what distributions satisfying (3), (4) govern the noises w_t, v_t respectively, the inclusion of higher-order moments restricts the attacker's potential impact through a tightened detector threshold.

VI. CONCLUSION & FUTURE OUTLOOK

We have proposed a distributionally robust approach to form the k moments based ambiguity set for the detection measure data and used it to tune the anomaly detectors for a desired false alarm rate. We found a detector threshold which guaranteed that the false alarm rate did not exceed a desired value using a semidefinite program. We have demonstrated the effectiveness of our proposed approach with a numerical example. Further, our approach using higher-order moments

restricted the attacker's potential impact. Future works include addressing the problems associated with the data-driven formulation with moment estimation uncertainty and securing nonlinear cyberphysical systems with distributionally robust unscented Kalman filter based state estimation.

REFERENCES

[1] C. Murguia and J. Ruths, "On model-based detectors for linear time-invariant stochastic systems under sensor attacks," *IET Control Theory & Applications*, vol. 13, no. 8, pp. 1051–1061, 2019.

[2] D. Umsonst and H. Sandberg, "Anomaly detector metrics for sensor data attacks in control systems," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 153–158.

[3] V. Renganathan, N. Hashemi, J. Ruths, and T. H. Summers, "Distributionally robust tuning of anomaly detectors in cyber-physical systems with stealthy attacks," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 1247–1252.

[4] M. J. Khojasteh, A. Khina, M. Franceschetti, and T. Javidi, "Learning-based attacks in cyber-physical systems," *IEEE Transactions on Control of Network Systems*, 2020.

[5] D. Li and S. Martínez, "High-confidence attack detection via wasserstein-metric computations," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 379–384, 2021.

[6] K. Zetter, *Countdown to Zero Day: Stuxnet and the launch of the world's first digital weapon*. Broadway books, 2014.

[7] A. Cherepanov, "Win32/industroyer, a new threat for industrial control systems," *White Paper. ESET*, 2017.

[8] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Operations Research*, vol. 62, no. 6, pp. 1358–1376, 2014.

[9] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Operations research*, vol. 58, no. 4-part-1, pp. 902–917, 2010.

[10] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1-2, pp. 115–166, 2018.

[11] N. Hashemi, C. Murguia, and J. Ruths, "A comparison of stealthy sensor attacks on control systems," in *2018 Annual American Control Conference (ACC)*, June 2018, pp. 973–979.

[12] N. Hashemi and J. Ruths, "Generalized chi-squared detector for lti systems with non-gaussian noise," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 404–410.

[13] J.-B. Lasserre, *Moments, positive polynomials and their applications*. World Scientific, 2010, vol. 1.

[14] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[15] R. E. Curto, L. A. Fialkow, and H. M. Möller, "The extremal truncated moment problem," *Integral Equations and Operator Theory*, vol. 60, no. 2, pp. 177–200, 2008.

[16] J. W. Helton and J. Nie, "A semidefinite approach for truncated k-moment problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 851–881, 2012.

[17] D. Bertsimas and I. Popescu, "Optimal inequalities in probability theory: A convex optimization approach," *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 780–804, 2005.

[18] J. Shawe-Taylor and N. Cristianini, "Estimating the moments of a random vector with applications," in *19th Colloque sur le traitement du signal et des images, FRA, 2003*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2003.

[19] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations research*, vol. 58, no. 3, pp. 595–612, 2010.

[20] L. Vandenberghe, S. Boyd, and K. Comanor, "Generalized chebyshev bounds via semidefinite programming," *SIAM review*, vol. 49, no. 1, pp. 52–64, 2007.

[21] X. Chen, "A New Generalization of Chebyshev Inequality for Random Vectors," *preprint arXiv:0707.0805*, Jul. 2007.

[22] N. Hashemi and J. Ruths, "Co-design for security and performance: Geometric tools," *preprint arXiv:2006.08739*, 2020.