A Risk-Theoretical Approach to $\mathcal{H}_2$-Optimal Control under Covert Attacks

Matias I. Müller, Jezdimir Milošević, Henrik Sandberg and Cristian R. Rojas

***Abstract*— We consider the control design problem of optimizing the $\mathcal{H}_2$ performance of a closed-loop system despite the presence of a malicious covert attacker. It is assumed that the attacker has incomplete knowledge on the true process we are controlling. To account for this uncertainty, we employ different measures of risk from the so called family of coherent measures of risk. In particular, we compare the closed-loop performance when a nominal value is used, with three different measures of risk: average risk, worst-case scenario and *conditional value-at-risk* (CVaR). Additionally, applying the approach from a previous work, we derive a convex formulation for the control design problem when CVaR is employed to quantify the risk. A numerical example illustrates the advantages of our novel approach.**

I. INTRODUCTION

Control-systems represent the fusion of computing and communication resources that interact with some physical process. Although designed to improve operational performance, decrease costs, and make these systems less prone to failures, the integration of cyber and physical worlds opened the possibility for malicious cyber-attacks that can endanger the physical world [2]. In fact, several such cyber-physical attacks have already been conducted [3], [4], [5]. The most well known of these attacks was the Stuxnet malware [3], which was specifically designed to damage a targeted physical system, while staying undetected by the system operators. Thus, it is not surprising that the topic of cyber-security has attracted considerable attention within the control community.

Significant effort has been dedicated towards analyzing intelligent cyber-attacks based on a physical model of a control system [6], [7], [8]. It has been recognized that an attacker that uses the model knowledge can construct an attack that is hard, or sometimes even impossible, to detect from the collected sensor measurements [6]. Examples of these attacks include replay [9], zero dynamics [10], and covert attacks [11].

As stated in a recent survey [12], the problem of designing a control system that works in the presence of stealthy attacks has not received much attention. However, in order to protect a control system against these stealthy attacks, novel detection approaches have been proposed [9], [13], [14]. This problem is much more difficult than the well-studied fault tolerant control problem, since the attacker will always try to design the worst case attack based on the available resources, while trying to stay undetected. Motivated by this challenge, we aim to design a control algorithm that mitigates the impact of covert attacks. To the best of our knowledge, this problem has not been addressed in the literature so far.

In the covert attack strategy, the attacker uses specially designed filters to construct additive measurement and control signals in order to stay undetected. If these filters are designed based on the full model knowledge of the plant, the covert attack is perfectly stealthy and, moreover, it is impossible to design a controller that mitigates the influence of this attack [11]. However, assuming that the attacker has full model knowledge can be quite conservative. For instance, the attacker may have an outdated system model, or an inaccurate model due to identification error [15]. The defender could also manipulate the plant intentionally (*e.g.*, through sensors and actuation gain), introducing fictitious uncertainty as a defensive measure, as proposed in [13]. In this work, we consider an attacker possessing only a partial knowledge of the plant, with different levels of accuracy. Naturally, the more knowledge the attacker possesses, the harder it is for the defender to control the system. The question we aim to answer is then how to design a controller that *performs well in most of the feasible attacker scenarios* in the presence of such a covert attack.

In the control design problem, the lack of knowledge of the attacker can be modeled as *uncertainty*. More precisely, since the defender does not know what the attacker's knowledge is, the filters designed by the latter are, in fact, uncertain. The problem of measuring uncertainty has been widely studied in *theory of risk* in finance [16], where *risk* is known as the impact of decisions made under uncertainty. Here, a family of measures known as *coherent measures of risk* [17] have became popular due to their attractive properties, such as convexity.

Along this work, we discuss how different measures of risk can be employed to account for the uncertainty in the attacker's filters design. In particular, we compare the nominal design case, *i.e.*, when the defender assumes only one single realization of the attacker's filters, to different risk-oriented designs, such as average behavior and worst-case scenario. In addition, we propose to use a third measure of risk called *conditional value-at-risk* (CVaR) [18], which has gained popularity in the control community: [19] used this notion of risk to derive quadratic constraints in LQG (linear-quadratic Gaussian) control, while [1] used it to account for the modeling error when maximizing the closed-loop disturbance rejection.

*This work was partially supported by the Swedish Research Council under contracts 2015-04393 and 2016-06079, and by the Swedish Civil Contingencies Agency through the CERCES project.

All the authors are with the Department of Automatic Control, KTH Royal Institute of Technology, SE 100 44 Stockholm, Sweden. Corresponding author e-mail: mimr2@kth.se.

The specific contributions of this paper are:

1) To apply a novel approach to control design developed in [1] to cybersecurity in the presence of covert attacks.
2) To discuss how the notion of coherent measures of risk can be employed in this problem, providing a systematic approach to account for the uncertainty in the control design problem under attack.
3) We prove that the problem of designing a controller mitigating covert attacks can be reduced to a convex optimization problem.

The goal is to obtain a controller that minimizes the risk of falling into low closed-loop tracking performance, by minimizing its conditional value-at-risk, and to compare it with traditional designs such as average risk, worst-case scenario, and a nominal design.

The remainder of the paper is organized as follows. In Section II, we model and formulate the problem. Section III presents different coherent measures of risk accounting for the uncertainty in the attacker's filters. Section IV exposes how the filter design problem can be casted as quadratically-constrained linear program, while Section V derives the equivalence to the control design problem. Section VI presents a numerical example comparing the closed-loop performance for different designs, and Section VII, presents the conclusions and possible extensions of this work.

A. Notation

The one-step forward shift operator, for discrete time, is denoted as q. $\mathcal{H}_2$ is the set composed by all complex functions $F(z)$ of a complex variable z which are analytic in $|z| \geq 1$ satisfying $\|F\|_2^2 := \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|F(\mathrm{e}^{j\omega})\right|^2 d\omega = \lim_{t\to\infty} \frac{1}{N}\sum_{t=1}^{N} |f_t|^2 < \infty$, where $(f_t)_{t\in\mathbb{N}}$ satisfies $F(z) = \sum_{i=0}^{\infty} f_t z^{-i}$, and $\|F\|_2 = \|f\|_2$. For transfer functions, arguments q and z are interchangeable if needed. $\mathbb{L}^2$ is the Hilbert space composed of all random variables with bounded first and second moments. The positive part function is defined as $[X]_+ := \max\{X, 0\}$. The set of natural and real numbers are $\mathbb{N}$ and $\mathbb{R}$, respectively. A transfer function $F_\theta(q)$ represents a function in q parametrized by $\theta \in \mathbb{R}^{n_\theta}$. We define $\mathbf{1}_N := [1\, 1\, \cdots\, 1]^\top \in \mathbb{R}^N$. The expectation operator is $\mathbb{E}\{\cdot\}$, $\mathbb{E}_\theta\{\cdot\}$ denotes the expectation with respect to a random vector θ, and $\mathbb{E}\{\cdot|\theta\}$ represents conditional expectation given θ. The support of a random vector Y is denoted as $\operatorname{supp}\{Y\}$. The transposed and conjugate transposed operators are $(\cdot)^\top$ and $(\cdot)^{\mathsf{H}}$, respectively.

II. Model Setup

We use an appropriate variant of the framework introduced in [11], where the process consists of a physical plant $G(q)$, a controller $C(q)$, and attacker's blocks $\Pi_0(q)$ and $K_0(q)$ as shown in Fig. 1. In what follows, we introduce these blocks in more detail, and then formulate the problem of designing a controller that mitigates the impact of covert attacks.

A. Plant and Feedback Controller

The process is modeled as $y(t) = G(q)u(t)$, where $G(q)$ represents the stable LTI (linear and time-invariant) SISO

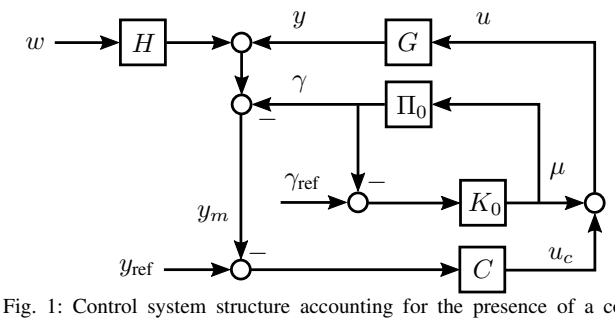

Fig. 1: Control system structure accounting for the presence of a covert agent.

(single-input single-output) transfer function of the system, $u(t)$ denotes the control signal applied to the process, $y(t)$ represents the system output, and $t \in \{0 \cup \mathbb{N}\}$. The output is perturbed, in general, by colored noise $H(q)w(t)$, where $H(q)$ is a biproper stable LTI filter, while $w(t)$ is unit-variance white noise.

A one-degree-of-freedom feedback control architecture is employed, consisting of a controller $C = C(q)$, whose output, the *control signal*, is given by $u_c(t) = C(q)\big[y_{\text{ref}}(t) - y_m(t)\big]$, where $y_m(t)$ denotes the measurable signal, and $y_{\text{ref}}(t)$ is the reference signal with spectral factor $R(z)$.

B. Attacker

Due to the noise and attack signals, the signal received by the controller, $y_m(t)$, and the input to the plant, $u(t)$, differ from the true plant output $y(t)$ and the control signal $u_c(t)$, respectively. In particular, we assume that

$$y_m(t) = y(t) + H(q)w(t) - \gamma(t), \quad u(t) = u_c(t) + \mu(t),$$

as depicted in Fig. 1. In contrast to the noise $w(t)$, that has random nature, the signals $\gamma(t)$ and $\mu(t)$ are outputs of a specially designed closed loop system consisting of blocks $\Pi_0(q)$ and $K_0(q)$. System $\Pi_0(q)$ represents the attacker's estimate of the physical plant $G(q)$, while K_0 corresponds to the attacker's controller, which is assumed to be an LTI filter designed to track the attacker's reference signal $\gamma_{\text{ref}}(t)$, of spectral factor $S(z)$. We additionally assume that $K_0/(1 + K_0\Pi_0)$ is stable. Then, signals $\gamma(t)$ and $\mu(t)$ can be expressed in terms of the attacker reference $\gamma_{\text{ref}}(t)$ as

$$\begin{aligned} \gamma(t) &= \frac{K_0(q)\Pi_0(q)}{1 + K_0(q)\Pi_0(q)}\gamma_{\text{ref}}(t), \\ \mu(t) &= \frac{K_0(q)}{1 + K_0(q)\Pi_0(q)}\gamma_{\text{ref}}(t). \end{aligned} \tag{1}$$

C. Problem Formulation

We assume that the defender knows $G(q)$ and $H(q)$ accurately enough for control purposes. From the defender's perspective, the attacker's blocks $\Pi_0(q)$ and $K_0(q)$ are unknown. Thus, we adopt the following assumption.

Assumption 1. *The defender's uncertainty about the attacker's filters is represented by a random vector $\theta \in \mathbb{L}^2$ with bounded support $\Theta := \operatorname{supp}\{\theta\}$, such that $K_0(q) = K_\theta(q)$ and $\Pi_0(q) = \Pi_\theta(q)$, for some known structures $\Theta \to K_\theta(q)$ and $\Theta \to \Pi_\theta(q)$. Additionally, we assume that $G(q)$ and $\Pi_\theta(q)$ are strictly proper, that $K_\theta(q)$ is proper, and that*

$K_\theta(q)$, $\Pi_\theta(q)$ *and* $(1+K_\theta(q)\Pi_\theta(q))^{-1}$ *are stable transfer functions, for every realization of* θ.

Remark 1. *The reasoning behind why* θ *is random can be understood within a Bayesian framework. The probability density function (pdf)* $p(\theta)$ *denotes the defender's knowledge about* θ*, coming from prior information and possibly experimental data. For instance, the attacker can obtain a model of the plant by using system identification tools [15], where* $p(\theta)$ *is obtained from the accuracy of the identification method. Nevertheless, the problem of obtaining* $p(\theta)$ *might be difficult, and it will be treated in future work. The key observation here is that* $p(\theta)$ *is assumed given.*

Remark 2. *The presented approach can also be employed when the attacker has full model knowledge,* i.e.*, when* $\Pi_\theta(q) = G(q)$*. In this case, the defender needs to use a detection mechanism by intentionally perturbing the plant* $G(q)$ *[13] (*e.g.*, by perturbing sensors' gain) with an additive known perturbation* $\Delta_G(q)$*, introducing fictitious uncertainty.*

The difference between the desired output $y_{\text{ref}}(t)$ and the true output of the system $y(t)$ is known as the *tracking error* $e(t) := y_{\text{ref}}(t) - y(t)$, and it is commonly employed to evaluate the closed-loop performance. Assuming that all external signals $(\gamma_{\text{ref}}, y_{\text{ref}}, w)$ are independent, and defining the closed-loop performance, using the $\mathcal{H}_2$ norm, as $J_C(\theta) := \|e\|_2^2$, it holds that

$$J_C(\theta) = \left\| \left(1 - \frac{[G-\Pi_\theta]\,C}{1+GC}\right)\frac{GSK_\theta}{1+K_\theta\Pi_\theta} \right\|_2^2 + \left\| \left(1 - \frac{GC}{1+GC}\right) R \right\|_2^2 + \left\| \frac{HGC}{1+GC} \right\|_2^2. \quad (2)$$

In the above equation, each term represents the effect on the tracking error due to each of the external signals $(\gamma_{\text{ref}}, y_{\text{ref}}, w)$. We notice here that, due to Assumption 1, (2) is a random cost function depending on the random vector θ, parametrized by controller $C \in \mathcal{C}(\Theta)$, where $\mathcal{C}(\Theta)$ is the set of all stabilizing controllers for any realization of θ. The later, together with the compactness of Θ, imply that $J_C(\theta) \in \mathbb{L}^2$. To quantify the uncertainty in $J_C(\theta)$ due to the lack of knowledge of the defender about the attacker's plant model and controller (captured in θ), we consider a functional measure $\mathcal{R} : \mathbb{L}^2 \to (-\infty, \infty]$. This measure will account for the risk of making bad decisions in the face of the uncertainty. We are now ready to formalize our problem.

Problem 1 ($\mathcal{H}_2$-risk optimal control under covert attack (H2RCA))**.** *Under the set-up of this section, find the optimal controller* $C^\star$ *such that*

$$C^\star := \arg\min_{C\in\mathcal{C}(\Theta)} \mathcal{R}\{J_C(\theta)\}, \quad (3)$$

where θ *is distributed according to the known pdf* $p(\theta)$*, and where* $S(z)$ *and* $R(z)$ *are the spectral factors of* $\gamma_{\text{ref}}(t)$ *and* $y_{\text{ref}}(t)$*, respectively.*

Remark 3. *Since* $p(\theta)$ *is fixed, different decisions* $C \in \mathcal{C}(\Theta)$ *define (possibly) different probability density functions of* $J_C(\theta)$*. Thus,* $C^\star(q)$ *is such that* $J_C(\theta)$ *is distributed according to a pdf with the lowest risk, where the risk is measured by* $\mathcal{R}$*.*

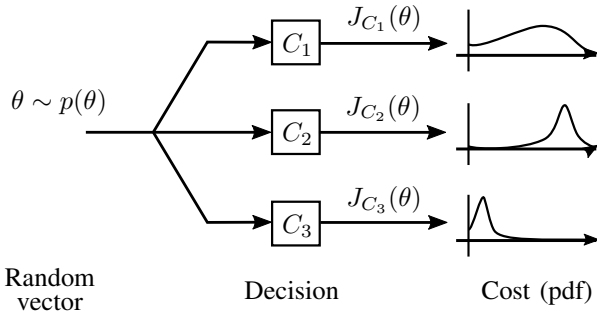

Fig. 2: In control, the risk framework is composed by a random vector θ condensing the uncertainty, a decision $C \in \mathcal{C}(\Theta)$, and a random cost function parametrized by a decision $J_C(\theta)$.

As discussed in [1], the main problem here is how to properly choose $\mathcal{R}$. In the next section, we describe several measures of risk that may be used to design a controller.

III. Measures of Risk

The choice of $\mathcal{R}$ in Problem 1 is essential for the controller design, since it defines the notion of *optimality*. The natural motivation of this choice is that the cost function in (2) is itself random, but its pdf is subjected to the designed controller $C^\star$. In fact, to minimize the risk of falling into poor closed-loop performances, we aim to concentrate the pdf mass of $J_C(\theta)$ towards zero.

Example 1. *Fig. 2 illustrates this problem, where different decisions* C_1, C_2, C_3 *achieve* 3 *different cost functions* $J_{C_1}, J_{C_2}, J_{C_3}$*. A reasonable decision to choose is* C_3*, since it concentrates more mass of the cost pdf towards zero.*

In this section, we discuss how *coherent measures of risk* [17] can be used to account for the uncertainty captured by θ. Coherent measures of risk are notions that have been developed in the financial theory of risk [20], whose objective is, in fact, to determine an optimal decision minimizing a cost under uncertainty. Together with the average performance and the worst-case performance, we introduce a special measure called *Conditional Value-at-Risk* (CVaR), all of them belonging to this family of measures.

A. Average Performance

This measure accounts for the uncertainty by considering the average performance of the closed loop, *i.e.*, considering the expected value over θ of the tracking error. Formally, when we choose this index, the risk is measured as $\mathcal{R}(J_C) = \mathbb{E}_\theta\{J_C(\theta)\}$. However, this measure will not account for how large the tracking error becomes for some realizations of θ. This is particularly dangerous for systems that do not tolerate having large values of $\|e\|_2^2$.

B. Worst-case Performance

In this case, we measure the risk as the maximum possible value that J_C can take. In other words, the measure is indexed by $\mathcal{R}(J_C) = \sup_{\xi\in\Theta} J_C(\xi)$. Analogously, this measure does not account for the average closed-loop performance, which might be conservative in most of the cases.

C. Conditional Value-at-Risk (CVaR)

Consider a random variable $Y \in \mathbb{L}^2$ with bounded support $\operatorname{supp}\{Y\}$. For a design parameter α, CVaR is defined [21] as

$$\text{CVaR}_\alpha(Y) := \frac{1}{1-\alpha}\int_{\{z:\text{Prob}\{Y\le z\}\ge\alpha\}} yp(y)dy, \quad (4)$$

where $Y \sim p(y)$, and where $\min\{z : \text{Prob}\{Y \le z\} \ge \alpha\}$ is known as the value at risk (VaR) of Y [22] (VaR is not a coherent measure of risk, since it lacks convexity [22]). From (4), CVaR corresponds to the mean of the α-tail distribution. Additionally, $\text{CVaR}_\alpha(Y) = \beta$ implies that $Y \le \beta$ at least $\alpha \times 100\%$ of the time, hence $\text{VaR}_\alpha(Y) \le \text{CVaR}_\alpha(Y)$.

For simplicity, we rely on an alternative way of computing the CVaR of a random quantity, proposed by [18], since (4) might be hard to directly optimize. We then compute CVaR as the solution of an optimization problem:

$$\text{CVaR}_\alpha(Y) = \min_{\nu\in\mathbb{R}} \nu + \frac{1}{1-\alpha}\mathbb{E}\left\{[Y-\nu]_+\right\}. \quad (5)$$

In our problem, $Y = J_C(\theta)$, and then the problem of computing CVaR in (5) might still be hard since the pdf of $J_C(\theta)$ is not easy to obtain in general. However, we can circumvent this issue if we are allowed to sample from the pdf $p(\theta)$. In that case, we can approximate (5) as

$$\begin{aligned}\text{CVaR}_\alpha(J_C(\theta)) &\approx \min_{\nu\in\mathbb{R}} \nu + \frac{1}{(1-\alpha)}\frac{1}{N}\sum_{i=1}^{N}[J_C(\theta_i)-\nu]_+ \\ &=: \overline{\text{CVaR}}_\alpha(\{J_C(\theta_i)\}_{i=1}^N). \quad (6)\end{aligned}$$

We finish this section by stating the following remark:

Remark 4. *α can be seen as a tuning parameter, since*

$$\lim_{\alpha\to 0^+}\text{CVaR}_\alpha(Y) = \mathbb{E}\{Y\} \quad (7)$$

$$\lim_{\alpha\to 1^-}\text{CVaR}_\alpha(Y) = \sup\operatorname{supp}\{Y\}. \quad (8)$$

CVaR then provides a trade-off between $\mathbb{E}\{Y\}$ and $\sup\operatorname{supp}\{Y\}$, as it is not possible in general to reduce both quantities at the same time.

IV. CVaR Filter Design

In this section we derive a quadratically-constrained linear program (QLCP) to the filter design problem

$$Q^\star := \arg\min_{Q\in\mathcal{H}_2} \text{CVaR}_\alpha \underbrace{\left(\sum_{m=1}^{M}\|A_\theta^{(m)} - B_\theta^{(m)}Q\|_2^2\right)}_{=:V_Q(\theta)}, \quad (9)$$

which represents a generalization of the problem stated in [1]. Here, $V_Q(\theta)$ is a new cost function that will be shown to be equivalent to $J_C(\theta)$ under some mild conditions, and where $\{A_\theta^{(m)}, B_\theta^{(m)}\}_{m=1}^M$, $M \in \mathbb{N}$, is a set of parametrized transfer functions that are stable for all $\theta \in \Theta$.

Remark 5. *The optimization problem (9) is convex since the cost function $\|A_\theta^{(m)} - B_\theta^{(m)}Q\|$ is convex in Q and CVaR is monotonic and convex [23]. The search space is convex since $\mathcal{H}_2$ is a linear space. Additionally, $\mathbb{E}\{\cdot\}$ and $\sup\{\cdot\}$ are special cases of CVaR (for $\alpha \to 0$, and $\alpha \to 1$, resp.), and then both formulations are convex as well.*

Let us define the family of L-length FIR (finite impulse response) filters as $\mathcal{Q}_L := \left\{Q : Q(q) = \sum_{\ell=0}^{L} x_\ell q^{-\ell}\right\}$, with $\boldsymbol{x} := [x_0 \ \dots \ x_L]^\top \in \mathbb{R}^{L+1}$. Notice that $\lim_{L\to\infty}\mathcal{Q}_L = \mathcal{H}_2$, in the sense that functions $g_i(z) = z^{-i}$ are a complete orthonormal set in $\mathcal{H}_2$. Then, with $\Gamma(q) := [1 \ q^{-1} \ \dots \ q^{-L}]^\top$, every controller $Q \in \mathcal{Q}_L$ can be written as

$$Q(q) = \boldsymbol{x}^\top\Gamma(q). \quad (10)$$

The facts that $\mathcal{Q}_L$ is dense in $\mathcal{H}_2$ and that CVaR_α is convex, allow us to state the following result:

Lemma 1. *Let $r^\star := \min_{Q\in\mathcal{H}_2}\text{CVaR}_\alpha(V_Q)$, for a fixed value of $\alpha \in (0,1)$, and let $Q^\star \in \mathcal{H}_2$ be the minimizer under this measure. Then, with probability 1,*

$$\lim_{L\to\infty}\lim_{N\to\infty}\min_{Q\in\mathcal{Q}_L}\overline{\text{CVaR}}_\alpha(\{V_Q(\theta_i)\}_{i=1}^N) = r^\star$$

$$\lim_{L\to\infty}\lim_{N\to\infty}\arg\min_{Q\in\mathcal{Q}_L}\overline{\text{CVaR}}_\alpha(\{V_Q(\theta_i)\}_{i=1}^N) = Q^\star(q).$$

Proof. See [1, Theorem 1]. □

We now state the main result of this section where, by means of Lemma 1, $Q^\star \in \mathcal{H}_2$ and $r^\star$ can be approximated arbitrarily well by solving a QLCP.

Lemma 2. *Let $\{A_\theta^{(m)}, B_\theta^{(m)}\}_{m=1}^M$, $M \in \mathbb{N}$, be a set of stable transfer functions parametrized by θ. Let $\{\theta_i\}_{i=1}^N$ be a set of N samples from a known pdf $p(\theta)$, and let $\boldsymbol{t} := \begin{bmatrix}t_1 & t_2 & \cdots & t_N\end{bmatrix}^\top \in \mathbb{R}^N$. The problem of finding $Q^\star$ as in (9) can be approximated arbitrarily well (as $N, L \to \infty$) by $\bar{Q}^\star_{N,L} \in \mathcal{Q}_L \subset \mathcal{H}_2$ and $\overline{\text{CVaR}}_\alpha(\{V_{\bar{Q}^\star_{N,L}}(\theta_i)\}_{i=1}^N)$, respectively, where $\bar{Q}^\star_{N,L}(q) := (\boldsymbol{x}^\star_N)^{\mathsf{T}}\Gamma(q)$, and*

$$\begin{aligned}[\boldsymbol{x}^\star_N \quad \nu^\star \quad \boldsymbol{t}^\star]^{\mathsf{T}} := \arg\min_{[\boldsymbol{x}\,\nu\,\boldsymbol{t}]^{\mathsf{T}}\in\mathbb{R}^{L+N+2}} \ & \nu + \frac{1}{N(1-\alpha)}\mathbf{1}_N^{\mathsf{T}}\boldsymbol{t} \\ \text{s.t.} \quad & t_i \ge k_i + \boldsymbol{x}^{\mathsf{T}}\boldsymbol{M}_i\boldsymbol{x} - 2\boldsymbol{c}_i^{\mathsf{T}}\boldsymbol{x} - \nu, \\ & t_i \ge 0, \quad i = 1,\dots,N,\end{aligned}$$

where $k_i \in \mathbb{R}$, $\boldsymbol{c}_i \in \mathbb{R}^{L+1}$, and $\boldsymbol{M}_i \in \mathbb{R}^{(L+1)\times(L+1)}$, are

$$\boldsymbol{M}_i = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Gamma(\mathrm{e}^{j\omega})\Gamma^{\mathsf{H}}(\mathrm{e}^{j\omega})\left(\sum_{k=1}^{M}\left|B_{\theta_i}^{(m)}(\mathrm{e}^{j\omega})\right|^2\right)d\omega,$$

$$\boldsymbol{c}_i^\top = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{k=1}^{M}A_{\theta_i}^{(m)}(\mathrm{e}^{j\omega})(B_{\theta_i}^{(m)}(\mathrm{e}^{j\omega}))^{\mathsf{H}}\right)\Gamma^{\mathsf{H}}d\omega,$$

$$k_i = \sum_{k=1}^{M}\left\|A_{\theta_i}^{(k)}(\mathrm{e}^{j\omega})\right\|_2^2.$$

Proof. We notice that, for a fixed m and for the i-th sample of θ, say θ_i, it holds that

$$\|A_{\theta_i}^{(m)} - B_{\theta_i}^{(m)}Q\|_2^2 = \underbrace{\|A_{\theta_i}^{(m)}\|_2^2}_{:=k_i^{(m)}} + \|B_{\theta_i}^{(m)}Q\|_2^2 - 2\left\langle A_{\theta_i}^{(m)}, B_{\theta_i}^{(m)}Q\right\rangle,$$

where, by means of (10), it follows that

$$\|B_{\theta_i}^{(m)}Q\|_2^2 = \boldsymbol{x}^\top\underbrace{\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\Gamma(\mathrm{e}^{j\omega})\Gamma^{\mathsf{H}}(\mathrm{e}^{j\omega})\left|B_{\theta_i}^{(m)}(\mathrm{e}^{j\omega})\right|^2 d\omega\right)}_{:=\boldsymbol{M}_i^{(m)}}\boldsymbol{x},$$

and analogously,

$$\left\langle A_{\theta_i}^{(m)}, B_{\theta_i}^{(m)}Q\right\rangle = \underbrace{\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}A_{\theta_i}^{(m)}(\mathrm{e}^{j\omega})(B_{\theta_i}^{(m)}(\mathrm{e}^{j\omega}))^{\mathsf{H}}\Gamma^{\mathsf{H}}(\mathrm{e}^{j\omega})d\omega\right)}_{:=(\boldsymbol{c}_i^{(m)})^\top}\boldsymbol{x}.$$

Then, for each θ_i, $i = 1, \dots, N$, the i-th sample of the cost function is given by

$$\begin{aligned} V_Q(\theta_i) &= \sum_{m=1}^{M} \|A_{\theta_i}^{(m)} - B_{\theta_i}^{(m)} Q\|_2^2 \\ &= \sum_{m=1}^{M} \left(k_i^{(m)} + \boldsymbol{x}^\top \boldsymbol{M}_i^{(m)} \boldsymbol{x} - 2(\boldsymbol{c}_i^{(m)})^\top \boldsymbol{x} \right) \\ &= \underbrace{\sum_{m=1}^{M} k_i^{(m)}}_{:=k_i} + \boldsymbol{x}^\top \underbrace{\left(\sum_{m=1}^{M} \boldsymbol{M}_i^{(m)} \right)}_{:=\boldsymbol{M}_i} \boldsymbol{x} - 2 \underbrace{\sum_{m=1}^{M} (\boldsymbol{c}_i^{(m)})^\top}_{:=\boldsymbol{c}_i^\top} \boldsymbol{x}. \end{aligned}$$

On the other hand, the problem of finding

$$\bar{Q}_{N,L}^\star = \arg \min_{Q \in \mathcal{Q}_L} \overline{\text{CVaR}}_\alpha(\{V_Q(\theta_i)\}_{i=1}^N)$$

is equivalent to finding

$$\boldsymbol{x}^\star = \arg \min_{\boldsymbol{x} \in \mathbb{R}^{L+1}} \overline{\text{CVaR}}_\alpha(\{V_Q(\theta_i)\}_{i=1}^N),$$

when Q is parametrized as in (10).

Finally, using (6), the latter expression is equivalent to

$$\begin{aligned} \boldsymbol{x}^\star = \arg \min_{[\boldsymbol{x}^\top \ \nu \ \boldsymbol{t}^\top]^\top \in \mathbb{R}^{L+N+2}} \ & \nu + \frac{1}{N(1-\alpha)} \mathbf{1}^\top \boldsymbol{t} \\ \text{subject to} \quad & t_i \geq V_Q(\theta_i) - \nu, \\ & t_i \geq 0, \quad i = 1, \dots, N, \end{aligned}$$

since the nonlinearity $\sum_{i=1}^N [V_Q(\theta_i) - \nu]_+$ can be replaced by its upper bound $\mathbf{1}^\top \boldsymbol{t}$, provided $t_i \geq 0$ or $t_i \geq V_Q(\theta_i) - \nu$. □

V. CVaR Controller Design

This section presents the main contribution of this article, connecting the filter design problem with the $\mathcal{H}_2$ control problem under covert-attacks, given there is uncertainty in the attacker design.

Theorem 1. *Solving Problem 1 (H2RCA),* i.e.*, the problem of designing a controller C such that the risk of falling into poor performances $\mathcal{R}(J_C)$ is minimized, is equivalent to*

$$\min_{Q \in \mathcal{H}_2} \mathcal{R} \left(\sum_{m=1}^{M} \|A_\theta^{(m)} - B_\theta^{(m)} Q\|_2^2 \right), \tag{11}$$

whose solution can be approximated by the approach described in Lemma 2. Here, $\{A_\theta^{(m)}, B_\theta^{(m)}\}_{m=1}^M$ are stable transfer functions parametrized by θ.

Proof. Define the affine parametrization

$$Q(q) := \frac{C(q)}{1 + G(q)C(q)} \tag{12}$$

in (2), also known as Youla parametrization [24], with Q being the Youla parameter. The new cost function is then

$$\begin{aligned} V_Q(\theta) = & \left\| (1 - [G - \Pi_\theta] Q) \frac{GSK_\theta}{1 + K_\theta \Pi_\theta} \right\|_2^2 \\ & + \|(1 - GQ)R\|_2^2 + \|HGQ\|_2^2, \end{aligned} \tag{13}$$

which is exactly in the form of the argument of (11). The equivalence between $V_Q(\theta)$ and $J_C(\theta)$ follows from G being a stable and proper transfer function. Moreover, $C(q) = Q(q)/(1 - G(q)Q(q))$ is guaranteed to stabilize the closed loop whenever $Q \in \mathcal{H}_2$ [24]. Then, solving Problem 1 in terms of $C^\star$ is equivalent to solving (11) in terms of $Q^\star$. □

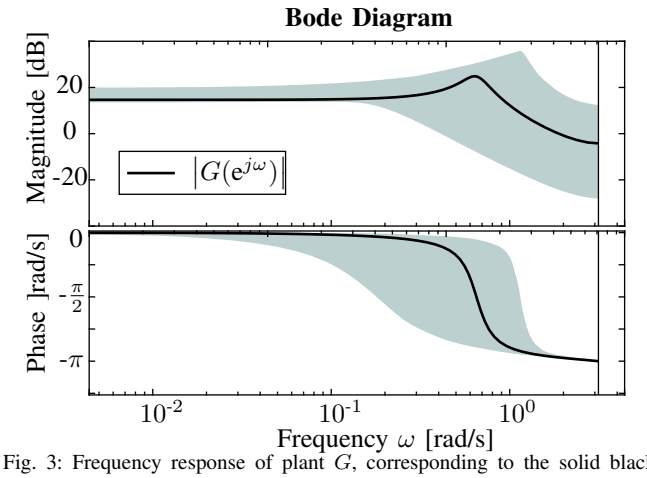

Fig. 3: Frequency response of plant G, corresponding to the solid black line. The shaded areas correspond to all possible frequency responses Π_θ parametrized by $\theta \in \Theta$.

We remark the versatility of the approach exposed here, where different signals can be added into the loop depicted in Fig. 1 by just stacking one more term of the form $\|A - BQ\|_2^2$ to the cost function, as long as these signals are mutually uncorrelated. Additionally, uncertainty can also be assumed to be present in block H of Fig. 1, addressing a more complex and robust problem.

VI. Illustrative Example

Here we provide insight into how different coherent measures of risk achieve different density functions for $J_C(\theta)$. In particular, we compare the closed-loop performance for different controllers, each of them designed under different risk measures, such as CVaR, expected value, and worst-case scenario. We also show how the nominal design performs.

Example 2. *Consider a second-order process*

$$G(q) = \frac{2q}{(q - 0.9\,\mathrm{e}^{j\omega_0})(q - 0.9\,\mathrm{e}^{-j\omega_0})} \tag{14}$$

with resonance frequency $\omega_0 = 0.6435$ [rad/s].

We assume that the attacker knows the process structure, but it does not know the resonance frequency nor the static gain of the process exactly. More precisely, the structure for the attacker is given by

$$\Pi_\theta(q) = \frac{\theta_1 \left| 1 - 0.9\,\mathrm{e}^{j\theta_2} \right| q}{(q - 0.9\,\mathrm{e}^{j\theta_2})(q - 0.9\,\mathrm{e}^{-j\theta_2})}, \tag{15}$$

where θ_1 is distributed uniformly over $[5, 10]$ and θ_2 is distributed uniformly over $[\omega_0 - 0.5, \omega_0 + 0.5]$, $\theta := \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}^\top$. For clarity, Fig. 3 shows the frequency response of the plant G compared to the possible outcomes of the frequency response of Π_θ.

We also assume that the attacker's controller K_θ is optimally designed to minimize the tracking error of signal $\gamma_{\text{ref}}(t)$, i.e., for a given θ, we assume that

$$K_\theta(q) := \arg \min_{K \in \mathcal{H}_2} \left\| \left(1 - \frac{\Pi_\theta K}{1 + \Pi_\theta K} \right) S \right\|_2^2. \tag{16}$$

The spectral factors of signals y_{ref} and γ_{ref} are $R(z) = \frac{z}{z^2 - 1.52z + 0.9025}$, $S(z) = \frac{z}{z - 0.4}$, respectively, and the noise

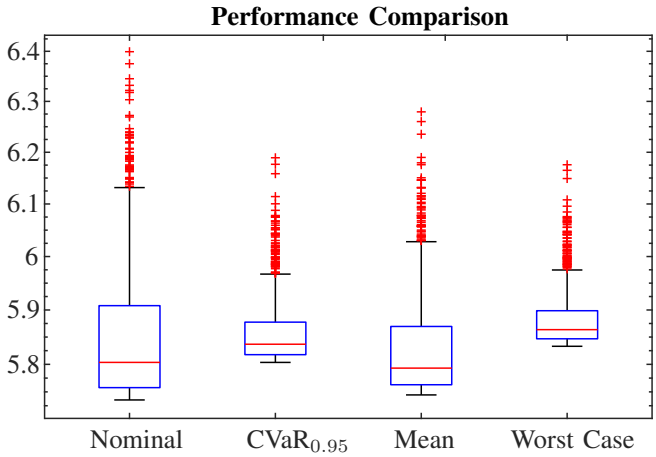

Fig. 4: Tracking performance comparison when the uncertainty is captured by $\theta \sim p(\theta)$, when $N = 1000$ samples are drawn from this distribution for design purposes, and when N_{val}=1000 samples are used for validation.

transfer function is $H(q) = \frac{q^2}{q^2+0.81q+0.81}$.

Under 4 different measures $\{\mathcal{R}_{\text{nom}}, \mathcal{R}_{\text{CVaR}}, \mathcal{R}_{\text{mean}}, \mathcal{R}_{\text{wc}}\}$, *the parametrized controllers* $\{C^\star_{\text{nom}}, C^\star_{\text{CVaR}}, C^\star_{\text{mean}}, C^\star_{\text{wc}}\}$ *are designed as each corresponding minimizer. Measure* $\mathcal{R}_{\text{nom}}(J_C(\theta))$ *corresponds to the nominal value* $J_C(\hat{\theta})$, *where* $\hat{\theta} = \mathbb{E}\{\theta\} = \begin{bmatrix}0.75 & 0.6435\end{bmatrix}^\top$. *On the other hand,* $\mathcal{R}_{\text{CVaR}} = \text{CVaR}_{0.95}$, $\mathcal{R}_{\text{mean}} = \mathbb{E}\{\cdot\}$ *(mean) and* $\mathcal{R}_{\text{wc}} = \sup_{\xi\in\Theta} J_C(\xi)$ *(worst-case scenario). The latter are designed upon* $N = 1000$ *samples from the pdf* $p(\theta)$, *and* $N_{val} = 1000$ *samples for validation purposes.*

The results are presented in Fig. 4, for each design, where the top and the bottom of the blue boxes represent the 25th and 75th percentiles of the samples, respectively, the red line inside the box stands for the sample median, and observations beyond 1.5 times the interquartile length are marked as outliers, displayed with red crosses.

We first notice how poorly the nominal case performs in terms of the outliers, showing how bad is to omit the information contained in the pdf of θ, giving the poorest results among the four designs. On the other hand, the CVaR and worst-case scenario controllers perform similarly in terms of outliers, however, CVaR outperforms the latter in the average behavior, pushing a big mass of the pdf of $J_{C^\star_{\text{CVaR}}}$ towards zero. As expected, the controller designed under $\mathcal{R}_{\text{mean}}$ performs better than the later two in terms of average performance. However, the closed loop under $C^\star_{\text{mean}}$ falls into poor performance for some samples generating higher outliers.

VII. Conclusions and Future Work

In this paper, we have considered the problem of designing a controller that mitigates the impact of covert attacks. To capture the lack of knowledge about the attacker, we model the attacker's blocks as uncertain, depending on a random vector θ. We have introduced several measures of risk, for which the controller design can be formulated as a convex optimization program. The tracking performance achieved by controllers designed under different risk measures were compared in a numerical simulation. The simulation showed that CVaR, a risk measure recently introduced in [1] for optimal control, can be used to balance the trade-off between the average performance and the worst case performance. Future work includes the problem of accurately estimating the pdf of the parameter condensing the uncertainty in the attacker's design.

References

[1] M. I. Müller, P. E. Valenzuela, and C. R. Rojas, "Risk-coherent $\mathcal{H}_2$-optimal disturbance rejection under model uncertainty," *20th IFAC World Congress*, vol. 50, no. 1, pp. 15 530 – 15 535, 2017.

[2] K. Stouffer, J. Falco, and K. Scarfone, "Guide to industrial control systems (ICS) security," *NIST special publication*, vol. 800, no. 82, 2011.

[3] D. Kushner, "The real story of STUXNET," *IEEE Spectrum*, vol. 50, no. 3, pp. 48–53, March 2013.

[4] J. Slay and M. Miller, *Lessons Learned from the Maroochy Water Breach*. Boston, MA: Springer US, 2008, pp. 73–82.

[5] *Analysis of the Cyber Attack on the Ukrainian Power Grid*, SANS ICS, Washington, DC, 2016.

[6] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, Nov 2013.

[7] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, June 2014.

[8] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 5967–5972.

[9] 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, July 2014.

[10] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[11] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," *IEEE Control Systems*, vol. 35, no. 1, pp. 82–92, Feb 2015.

[12] J. Giraldo, E. Sarkar, A. A. Cardenas, M. Maniatakos, and M. Kantarcioglu, "Security and privacy in cyber-physical systems: A survey of surveys," *IEEE Design Test*, vol. 34, no. 4, pp. 7–17, Aug 2017.

[13] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 1806–1813.

[14] A. Hoehn and P. Zhang, "Detection of covert attacks and zero dynamics attacks in cyber-physical systems," in *American Control Conference (ACC), 2016*, pp. 302–307.

[15] A. O. de Sá, L. F. R. d. C. Carmo, and R. C. S. Machado, "Covert attacks in cyber-physical control systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1641–1651, Aug 2017.

[16] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. John Wiley and Sons, 1959.

[17] P. Artzner, F. Delbaen, J. Eber, and D. Heath, "Thinking coherently," *Risk*, pp. 68–71, 1999.

[18] R. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, pp. 21–42, 2000.

[19] B. P. G. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari, "Distributionally robust control of constrained stochastic systems," *IEEE Trans. Autom. Control*, vol. 61, no. 2, pp. 430–442, Feb 2016.

[20] J.-P. Bouchaud and M. Potters, *Theory of Financial Risks and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2000.

[21] C. Acerbi, "A coherent representation of subjective risk aversion," *Journal of Banking and Finance*, vol. 26, pp. 1505–1518, 2002.

[22] G. A. Holton, *Value-at-Risk: Theory and Practice*. Academic Press, 2014.

[23] R. T. Rockafellar, "Coherent approaches to risk in optimization under uncertainty," *Tutorials in Operations Research, Informs*, 2007.

[24] G. C. Goodwin, S. F. Graebe, and M. E. Salgado, *Control System Design*. Prentice Hall, 2000.