2022 IEEE Conference on Control Technology and Applications (CCTA)
August 22-25, 2022. Trieste, Italy

Risk assessment and optimal allocation of security measures under stealthy false data injection attacks

Sribalaji C. Anand[1], André M. H. Teixeira[2], and Anders Ahlén[1]

Abstract— **This paper firstly addresses the problem of risk assessment under false data injection attacks on uncertain control systems. We consider an adversary with complete system knowledge, injecting stealthy false data into an uncertain control system. We then use the Value-at-Risk to characterize the risk associated with the attack impact caused by the adversary. The worst-case attack impact is characterized by the recently proposed output-to-output gain. We observe that the risk assessment problem corresponds to an infinite non-convex robust optimization problem. To this end, we use dissipative system theory and the scenario approach to approximate the risk-assessment problem into a convex problem and also provide probabilistic certificates on approximation. Secondly, we consider the problem of security measure allocation. We consider an operator with a constraint on the security budget. Under this constraint, we propose an algorithm to optimally allocate the security measures using the calculated risk such that the resulting Value-at-risk is minimized. Finally, we illustrate the results through a numerical example. The numerical example also illustrates that the security allocation using the Value-at-risk, and the impact on the nominal system may have different outcomes: thereby depicting the benefit of using risk metrics.**

I. Introduction

Critical infrastructures describe the assets that are vital for the normal operation of society. In general, control systems are an integral part of critical infrastructures. Examples include pH control systems in a bioreactor, frequency control of power generating systems, etc. Due to the vitality of their operation, and partly due to advances in technology, these control systems are monitored regularly through wireless digital communication channels [1]. And due to the increased use of non-secure communication channels, the control systems are prone to cyber-attacks such as the attack on the Ukrainian power grid, and the Kemuri cyber attack to name a few [2]. Thus there is an increased research interest in the cyber-security of control systems [3].

One of the common recommendations for improving the security of control systems is to follow the risk management cycle: Risk assessment, risk response, and risk monitoring [4]. The risk assessment step involves careful consideration of risk sources, their likelihood, and their consequences. The consequence can be quantified in terms of impact which can be obtained through simulation or optimization-based methods. The risk response step involves implementing additional measures to minimize the risk if and when necessary. The risk response step can involve either (i) re-designing of the system controller/detectors to be robust against attacks, or (ii) allocating additional security measures such as encrypted communication channels. Finally, the risk monitoring step involves constant monitoring of the risk at acceptable intervals of time. This paper studies the risk assessment and risk response step of the risk management cycle.

Although the risk assessment and the risk response steps have been studied in the literature [4, Chapter 2], there are some research gaps that are outlined next. Firstly, the majority of the literature considers a deterministic system [5], [6]. Secondly, the risk frameworks are mostly application-specific. For instance, [7], [8] and [9] determine the risk of cyber-attacks on automatic generation control, power systems, and energy storage systems in smart grids respectively.

To address these limitations, we consider the following setup. A discrete-time (DT) linear time-invariant (LTI) process with uncertainties, an output feedback controller, and an anomaly detector. A stealthy adversary with complete system knowledge injects false data into the sensor or actuator channels for a long but finite amount of time. With this setup, we provide the following contributions.

1) We formulate the risk assessment problem using the Value-at-Risk (VaR) as a risk metric and the Output-to-Output Gain (OOG) as an impact metric.
2) We observe that the risk-assessment problem is NP-hard in general. To this end, we propose an approximate risk assessment problem that is computationally tractable.
3) We show that the approximate risk assessment problem can be solved by an equivalent convex semi-definite program (SDP). We provide the necessary and sufficient conditions for the (approximate) risk to be bounded.
4) We provide a preliminary algorithm to optimally allocate security measures using the calculated risk.
5) We numerically illustrate that the security measure allocation using the Value-at-risk, and the impact on the nominal system may have different outcomes. We thereby depict the advantage of using risk metrics as suggested in this paper.

The remainder of the paper is organized as follows. The control system and the adversary are described in Section II. The risk assessment problem (RAP) is formulated in Section III. We approximate the RAP and convert it to a convex SDP in Section IV. In Section V we formulate the security measure allocation problem (SMAP) and provide an algorithm which solves the SMAP for small-scale systems.

*This work is supported by the Swedish Research Council under the grant 2018-04396 and by the Swedish Foundation for Strategic Research.

[1] Sribalaji C. Anand and Anders Ahlén are with the Department of Electrical Engineering, Uppsala University, PO Box 65, SE-75103, Uppsala, Sweden. sribalaji.anand@angstrom.uu.se

[2] André M. H. Teixeira is with the Department of Information Technology, Uppsala University, PO Box 337, SE -75105, Uppsala, Sweden. andre.teixeira@it.uu.se

2022 IEEE Conference on Control Technology and Applications (CCTA) | 978-1-6654-7338-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/CCTA49430.2022.9966025

The results are illustrated through a numerical example in Section VI. Finally, we conclude the paper in Section VII.

Notation: Throughout this paper, $\mathbb{R}, \mathbb{C}, \mathbb{Z}$ and $\mathbb{Z}^+$ represent the set of real numbers, complex numbers, integers and non-negative integers respectively. A positive semi-definite matrix A is denoted by $A \succeq 0$. Let $x : \mathbb{Z} \to \mathbb{R}^n$ be a discrete-time signal with $x[k]$ as the value of the signal x at the time step k. Let the time horizon be $[0, N] = \{k \in \mathbb{Z}^+ | \ 0 \leq k \leq N\}$. The ℓ_2-norm of x over the horizon $[0, N]$ is represented as $||x||^2_{\ell_2,[0,N]} \triangleq \sum_{k=0}^{N} x[k]^T x[k]$. Let the space of square summable signals be defined as $\ell_2 \triangleq \{x[k] : \mathbb{Z}^+ \to \mathbb{R}^n | \ ||x||^2_{\ell_2,[0,\infty]} < \infty\}$ and the extended signal space be defined as $\ell_{2e} \triangleq \{x[k] : \mathbb{Z}^+ \to \mathbb{R}^n | \ ||x||^2_{[0,N]} < \infty, \forall N \in \mathbb{Z}^+\}$. For the sake of simplicity, we represent $||x||^2_{\ell_2,[0,\infty]}$ as $||x||^2_{\ell_2}$. For $x \in \mathbb{R}, \lceil x \rceil$ represents the nearest integer $\geq x$. For any finite set $\mathcal{Q}$, and element of $\mathcal{Q}$ is represented by $q_{(\cdot)}$, and the cardinality of the set is represented by $|\mathcal{Q}|$.

II. Problem background

In this section, we describe the control system structure and the goal of the adversary. Consider a closed-loop DT LTI system with a process ($\mathcal{P}$), output feedback controller ($\mathcal{C}$) and an anomaly detector ($\mathcal{D}$) represented by

$$\mathcal{P} : \begin{cases} x_p[k+1] & = A^{\Delta} x_p[k] + B^{\Delta} \tilde{u}[k] \\ y[k] & = C^{\Delta} x_p[k] \\ y_p[k] & = C_J^{\Delta} x_p[k] + D_J^{\Delta} \tilde{u}[k] \end{cases} \tag{1}$$

$$\mathcal{C} : \begin{cases} z[k+1] & = A_c z[k] + B_c \tilde{y}[k] \\ u[k] & = C_c z[k] + D_c \tilde{y}[k] \end{cases} \tag{2}$$

$$\mathcal{D} : \begin{cases} s[k+1] & = A_e s[k] + B_e u[k] + K_e \tilde{y}[k] \\ y_r[k] & = C_e s[k] + D_e u[k] + E_e \tilde{y}[k] \end{cases} \tag{3}$$

where $A^{\Delta} \triangleq A + \Delta A(\delta)$ with A representing the nominal system matrix and $\delta \in \Omega$. Additionally we assume Ω to be closed, bounded and to include the zero uncertainty yielding $\Delta A(0) = 0$. The other matrices are similarly expressed. The state of the process is represented by $x_p[k] \in \mathbb{R}^{n_x}$, $z[k] \in \mathbb{R}^{n_z}$ is the state of the controller, $s[k] \in \mathbb{R}^{n_s}$ is the state of the observer, $\tilde{u}[k] \in \mathbb{R}^{n_u}$ is the control signal received by the process, $u[k] \in \mathbb{R}^{n_u}$ is the control signal generated by the controller, $y[k] \in \mathbb{R}^{n_m}$ is the measurement output produced by the process, $\tilde{y}[k] \in \mathbb{R}^{n_m}$ is the measurement signal received by the controller and the detector, $y_p[k] \in \mathbb{R}^{n_p}$ is the virtual performance output, and $y_r[k] \in \mathbb{R}^{n_r}$ is the residue generated by the detector. The closed-loop system is also shown in Fig. 1. The reason to adopt uncertainty only in the process is that the parameters of the controller and the detector are chosen by the system operator. However the parameters of the process may not be known to the operator due to a variety of reasons such as, e.g. modelling errors.

In general, the system is considered to have a good performance when the energy of the performance output $||y_p||^2_{\ell_2}$ is small and an anomaly is considered to be detected when the detector output energy $||y_r||^2_{\ell_2}$ is greater than a predefined threshold, say ϵ_r. Without loss of generality (w.l.o.g.), we assume $\epsilon_r = 1$ in the sequel.

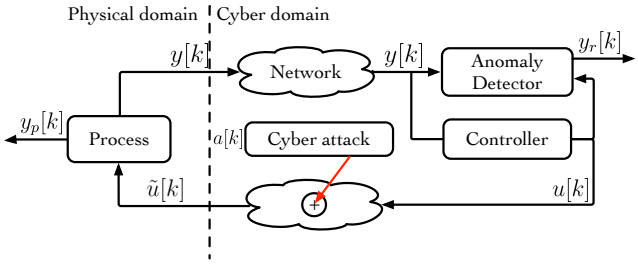

Fig. 1. Control system under data injection attack

A. Data injection attack scenario

In the closed-loop system described in (1)-(3), we consider that an adversary is injecting false data into the sensors or actuators of the plant but not both. Given this setup, we now discuss the resources the adversary has access to.

1) Disruption and disclosure resources: The adversary can access (eavesdrop) the control and sensor channels and can inject data. This is represented by $\begin{bmatrix} \tilde{u}^T[k] & \tilde{y}^T[k] \end{bmatrix}^T =$

$$\begin{bmatrix} u[k] \\ y[k] \end{bmatrix} + \begin{bmatrix} B_a \\ F_a \end{bmatrix} a[k], \left[\begin{array}{c|c} B_a^T & D_a^T \end{array}\right] \triangleq \left[\begin{array}{c|c} E_a^T & 0 \\ 0^T & F_a^T \end{array}\right]$$

where $a[k] \in \mathbb{R}^{n_a}$ is the data injected by the adversary. The matrix $E_a(F_a)$ is a diagonal matrix with $E_a(i,i) = 1$ ($F_a(i,i) = 1$), if the actuator (sensor) channel i is under attack and zero otherwise.

2) System knowledge: We assume that, the system operator knows the bounds of the set Ω and the nominal system matrices. Next, we assume that the adversary has full system knowledge, i.e., s/he knows the system matrices (1),(2), and (3). In reality, it is hard for the adversary to know the system matrices, but this assumption helps to study the worst case.

Defining $x[k] \triangleq [x_p[k]^T \ \ z[k]^T \ \ s[k]^T]^T$, the closed-loop system under attack with the performance output and detection output as system outputs becomes

$$\begin{aligned} x[k+1] &= A_{cl}^{\Delta} x[k] + B_{cl}^{\Delta} a[k] \\ y_p[k] &= C_p^{\Delta} x[k] + D_p^{\Delta} a[k] \\ y_r[k] &= C_r^{\Delta} x[k] + D_r^{\Delta} a[k], \end{aligned} \tag{4}$$

where the nominal matrices are given by $\left[\begin{array}{c|c} A_{cl} & B_{cl} \end{array}\right] \triangleq$

$$\left[\begin{array}{ccc|c} A + BD_cC & BC_c & 0 & BB_a + BD_cD_a \\ B_cC & A_c & 0 & B_cD_a \\ (B_eD_c + K_e)C & B_eC_c & A_e & (B_eD_c + K_e)D_a \end{array}\right]$$

$$C_p \triangleq \begin{bmatrix} C_J + D_J D_c C & D_J C_c & 0 \end{bmatrix}$$
$$C_r \triangleq \begin{bmatrix} (D_e D_c + E_e)C & D_e C_c & C_e \end{bmatrix}$$
$$D_p \triangleq D_J(D_c D_a + B_a), D_r \triangleq (D_e D_c + E_e) D_a.$$

In this paper, we consider the adversarial setup where the adversary is omniscient.

Definition 2.1 (Omniscient adversary): An adversary is defined to be omniscient if it knows the matrices in (4). ◁

In reality, it is hard for an adversary to know the system matrices of (4) due to the uncertainty. Thus, such an adversarial setup is far from reality but it can help us study a worst-case scenario. For clarity, we assume the following.

Assumption 2.1: The control system (4) is stable $\forall \delta \in \Omega$.$\triangleleft$

Assumption 2.2: The input matrix has full column rank i.e., $\nexists\ s \in \mathbb{R}^{n_a} \neq 0$ such that $B_{cl}^{\Delta} s = 0$. $\triangleleft$

3) Attack goals and constraints.: Given the resources the adversary has access to, it aims at disrupting the system's behavior while staying stealthy. The system disruption is evaluated by the increase in energy of the performance output whereas, the adversary is stealthy if the energy of the detection output is below a predefined threshold (ϵ_r). We discuss the attack policy for a deterministic system next.

B. Optimal attack policy for the nominal system

From the previous discussions, it can be understood that the goal of the adversary is to maximize the performance cost while staying undetected. When the system (4) is deterministic, [10] formulates that the attack policy of the adversary as the following non-convex optimization problem

$$\begin{aligned} ||\Sigma||^2_{\ell_{2e}, y_p \leftarrow y_r} \triangleq \sup_{a \in \ell_{2e}} & ||y_p||^2_{\ell_2} \\ \text{s.t.} & ||y_r||^2_{\ell_2} \leq 1, x[0] = 0, x[\infty] = 0, \end{aligned} \tag{5}$$

where $||\Sigma||^2_{\ell_{2e}, y_p \leftarrow y_r}$ represents the OOG that characterizes the disruption caused by the attack signal a. In (5), the constraint $x[0] = 0$ is introduced because the system is assumed to be at equilibrium before the attack.

Assumption 2.3: The closed-loop system (4) is at equilibrium $x[0] = 0$ before the attack commences. $\triangleleft$

We also assume that the adversary has finite amount of energy (similar to H_∞ control). Thus, the adversary does not attack the system for an infinite amount of time but stops after a very long time, say T. And since the attack stops, the state is brought back to equilibrium. To this end, we introduce the constraint $x[\infty] = 0$ in (5).

In the literature, such characterization of the impact of stealthy attacks (5) has only been studied for fully known deterministic systems, but not for an uncertain system. Thus, the first goal of the paper is to quantify the impact in terms of risk on the uncertain system (4). We later describe, in Section V, as to how the attack impact determined can be used for the benefit of the system operator.

III. Problem Formulation

To quantify the risks of data injection attacks on an uncertain control system, we start by defining a random variable that characterizes the impact as a function of the system uncertainty and the attack vector.

Definition 3.1 (Impact random variable): Let the random variable $X^A(\cdot)$ be defined as

$$X^A(a, \delta) \triangleq \|y_p(\delta)\|^2_{\ell_2} \times \mathbb{I}\Big(\|y_r(\delta)\|^2_{\ell_2} \leq 1, x(\delta)[\infty] = 0 \Big)$$

where $X^A(\cdot)$ is the impact caused on the system (4) with the uncertainty $\delta \in \Omega$ by the attack vector $a \in \ell_{2e}$, $\mathbb{I}$ is the indicator function, $y_p(\delta), y_r(\delta)$ and $x(\delta)$ are the performance, residue output and state of the system with the isolated uncertainty δ. Here, the signals $y_p(\delta), y_r(\delta)$ and $x(\delta)$ are also functions of the attack vector a. $\triangleleft$

With the random variable defined in *Definition 3.1*, we next formulate the risk assessment problem. Consider the data injection attack scenario where the parametric uncertainty $\delta \in \Omega$ of the system is known to the adversary but not to the system operator. The system operator has knowledge only about the bounds of the set Ω. Recall that such a scenario is far from reality, but such a setup helps us study the worst case. Under this setup, the adversary can cause high disruption by remaining stealthy because the adversary will be able to inject attacks by solving the optimization problem

$$||\tilde{\Sigma}(\delta)||^2_{\ell_{2e}, y_p \leftarrow y_r} \triangleq \sup_{a_\delta \in \ell_{2e}} X^A(a_\delta, \delta), \tag{6}$$

where a_δ represents the attack vector corresponding to the uncertainty δ. Since the system operator does not know the uncertainty δ, $||\tilde{\Sigma}(\delta)||^2_{\ell_{2e}, y_p \leftarrow y_r}$ can be interpreted as a random variable. Thus, for the system operator, the best option is to assess the risk associated with the impact based on a risk metric. In this paper, we adopt the risk metric VaR [11].

Definition 3.2 (Value-at-Risk (VaR)): Given a random variable X and $\beta \in (0, 1)$, the VaR is defined as

$$\text{VaR}_\beta(X) \triangleq \inf\{x | \mathbb{P}[X \leq x] \geq 1 - \beta\}.$$

With a specified probability level $\beta \in (0, 1)$, VaR_β is the lowest amount of x such that with probability $1 - \beta$, the random variable, X, does not exceed x. $\triangleleft$

Therefore, by calculating VaR_β, one can ensure that the probability that the value (X) exceeds VaR_β is less than or equal to β. In our setting, the system operator is interested in determining the VaR_β given a small β such that the impact rarely exceeds VaR_β. Given that the impact caused by the adversary on (4) is characterized by $||\tilde{\Sigma}(\delta)||^2_{\ell_{2e}, y_p \leftarrow y_r}$, the $\text{VaR}_\beta(\cdot)$ can be obtained, using *Definition 3.2*, by solving

$$\begin{aligned} \gamma_{OA} \triangleq \inf_{\gamma} & \ \gamma \\ \text{s.t. } & \mathbb{P}_\Omega(||\tilde{\Sigma}(\delta)||^2_{\ell_{2e}, y_p \leftarrow y_r} \leq \gamma) \geq 1 - \beta, \end{aligned} \tag{7}$$

where γ_{OA} represents the VaR associated with the impact caused by an **O**mniscient **A**dversary.

Although VaR is not extensively used in the literature [11], it is used here to only assess the worst-case risk even though this setup may be deemed unrealistic. Since $\mathbb{P}$ in (7) operates over the continuous space Ω, the optimization problem is computationally intensive or in general NP-hard [12, Section 3]. Besides (6) is a non-convex [10]. In Section IV we discuss a method to solve (7) approximately and efficiently.

IV. Risk assessment

To recall, (7) is computationally intensive since Ω is a continuum. To this end, in this section we determine an approximate solution to (7) and also provide some probabilistic certificates using the scenario approach introduced in [13].

A. Discrete uncertainty set

In this section, we consider a discrete uncertainty set Ω. The results of this section is the basis for addressing a continuous uncertainty set next. For brevity, given a sampled uncertainty $\delta_i \in \Omega$, we define $\tilde{\Sigma}_{p,i} \triangleq (A_{cl,i}, B_{cl,i}, C_{p,i}, D_{p,i})$ and $\tilde{\Sigma}_{r,i} \triangleq (A_{cl,i}, B_{cl,i}, C_{r,i}, D_{r,i})$ with $y_p(\delta_i) = y_{pi}, y_r(\delta_i) = y_{ri}$ and $x(\delta_i) = x_i$ as the outputs and states of $\tilde{\Sigma}_{p,i}$ and $\tilde{\Sigma}_{r,i}$ correspondingly. For such an isolated uncertainty, (6) can be rewritten as

$$||\tilde{\Sigma}(\delta_i)||^2_{\ell_{2e},y_p \leftarrow y_r} \triangleq \sup_{a_i \in \ell_{2e}} X^A(a_i, \delta_i) = \left\{ \begin{array}{l} \sup\limits_{a_i \in \ell_{2e}} \|y_{p,i}\|^2_{\ell_2} \\ \text{s.t. } \|y_{r,i}\|^2_{\ell_2} \leq 1, x_i[\infty] = 0 \end{array} \right\} \quad (8)$$

where a_i is the attack vector corresponding to the uncertainty δ_i. The optimization problem (8) has two disadvantages: it is non-convex and intractable (since the optimizer is infinite-dimensional). To this end, we can use the Lagrange dual function to reformulate the non-convex problem into its dual-counterpart. Furthermore, we can use dissipative system theory to convert the intractable non-convex problem to a convex problem with LMI constraints which is tractable. This reformulation is presented in *Lemma 4.1*.

Lemma 4.1: The optimization problem (8) is equivalent to the convex SDP (9).

$$\min_{\gamma_i \geq 0, P_i = P_i^T} \left\{ \gamma_i \,\middle|\, M(\gamma_i, \delta^i, P_i) \preceq 0 \right\} \quad (9)$$

where $M(\gamma_i, \delta^i, P_i) \triangleq \begin{bmatrix} A_{cl,i}^T P_i A_{cl,i} - P_i & A_{cl,i}^T P_i B_{cl,i} \\ B_{cl,i}^T P_i A_{cl,i} & B_{cl,i}^T P_i B_{cl,i} \end{bmatrix} + \begin{bmatrix} C_{p,i}^T \\ D_{p,i}^T \end{bmatrix} \begin{bmatrix} C_{p,i} & D_{p,i} \end{bmatrix} - \gamma_i \begin{bmatrix} C_{r,i}^T \\ D_{r,i}^T \end{bmatrix} \begin{bmatrix} C_{r,i} & D_{r,i} \end{bmatrix}$.

Proof: The result is similar to [10, Theorem 1] and thus the proof is omitted. ■

Next, we discuss the the conditions for boundedness of $||\tilde{\Sigma}(\delta_i)||^2_{\ell_{2e},y_p \leftarrow y_r}$ in the *Lemma 4.2*.

Lemma 4.2 (Boundedness): Consider the closed-loop system (4) with an uncertainty δ_i which is known to the adversary. Then, the optimal value of (9) is bounded if and only if one of the following conditions hold:

1) The system $\tilde{\Sigma}_{r,i}$ has no zeros on the unit circle.
2) The zeros on the unit circle of the system $\tilde{\Sigma}_{r,i}$ (including multiplicity and input direction) are also zeros of $\tilde{\Sigma}_{p,i}$.

Proof: See appendix. ■

Lemma 4.2 states that the attack impact caused by an omniscient adversary on (4) with an uncertainty δ_i is bounded if either, there does not exist an attack vector which makes the output y_r identically zero, or all attack vectors which yields y_r identically 0 also yields y_p identically zero.

In this section, we formulated the results on characterizing the attack impact by a convex SDP and the condition for its boundedness for an isolated discrete uncertainty. Next, the approach discussed in this section for risk assessment is extended when considering a continuum of uncertainties.

B. Continuous uncertainty set

The optimization problem (9) can be directly extended to solve (7) only when we consider a discrete set Ω. But (7) that we intend to solve operates over a continuous set Ω. Using the framework of scenario-based reliability estimation [13], Ω can be approximated with a finite set. With this scenario-based framework we revisit (7) in *Theorem 4.3*.

Theorem 4.3: Let $\epsilon_1 \in (0, 1)$ represent the accuracy with which the probability operator $\mathbb{P}_\Omega$ in (7) is to be approximated. Let $\beta_1 \in (0, 1)$ represent the confidence with which the accuracy ϵ_1 is guaranteed, i.e.,

$$\mathbb{P}\{|\mathbb{P}_\Omega(||\tilde{\Sigma}(\delta)||^2_{\ell_{2e},y_p \leftarrow y_r} \leq \gamma) - \hat{\mathbb{P}}_{N_1}| \geq \epsilon_1\} \leq \beta_1.$$

Here $\hat{\mathbb{P}}_{N_1}$ represents the approximation of the probability operator $\mathbb{P}_\Omega$ in (7) defined as

$$\hat{\mathbb{P}}_{N_1} \triangleq \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbb{I}\left(||\tilde{\Sigma}(\delta_i)||^2_{\ell_{2e},y_p \leftarrow y_r} \leq \gamma \right), \text{where}$$

$$N_1 \geq \frac{1}{2\epsilon_1^2} \log \frac{2}{\beta_1}. \quad (10)$$

Then, the VaR_β defined in (7) can be obtained with an accuracy ϵ_1 and confidence β_1 by solving

$$\hat{\gamma}_{OA} = \left\{ \begin{array}{ll} \min & \gamma \\ \text{s.t.} & \frac{1}{N_1} \sum\limits_{i=1}^{N_1} \mathbb{I}(\gamma_i \leq \gamma) \geq 1 - \beta, \end{array} \right\} \quad (11)$$

where $\hat{\gamma}_{OA}$ represents the VaR_β with an accuracy ϵ_1, and $\mathbb{I}$ is the indicator function. The value of $\gamma_i, i = 1, \ldots, N_1$ is obtained by solving the convex SDP (9).

Proof: In the interest of space, we omit the proof. ■

Theorem 4.3 states that, to solve (11), one could solve the optimization problem (9) for N_1 unique realizations of the uncertainty. Since strong duality holds between (8) and (9), the optimal value of the dual optimization problem (11) indeed provides the VaR_β with an accuracy ϵ_1 and confidence β_1. Thus *Theorem 4.3* provides a method to determine the risk through a sampled uncertainty set, and provides apriori probabilistic certificates on the accuracy and confidence of the operator $\mathbb{P}$. Next, by extending *Lemma 4.2*, the condition for boundedness of (11) is stated in *Lemma 4.4*.

Lemma 4.4 (Boundedness): Consider N_1 independent and identically distributed realizations of (4), each with an uncertainty δ_i. The optimal value of (11) with these N_1 system realizations is bounded iff the optimal value of (9) is bounded for at least $\lceil N_1(1 - \beta) \rceil$ system realizations.

Proof: In the interest of space, we omit the proof. ■

The interpretation of *Lemma 4.4* is that the system operator tolerates a fraction ($\beta \times 100\%$) of cases where the impact (9) is unbounded. Conversely, even-though the impact is unbounded for certain realization of uncertainty, the risk will still be bounded. This allows the system operator to be less pessimistic: In the sense that, even though the attack impact in certain scenarios can be very high, the risk evaluated by the operator will be bounded. In the next section, we briefly discuss the benefit of determining the risk and how it can be used by the system operator.

V. Optimal allocation of security measures

In this section, we discuss how the RAP discussed in Section IV can be used for the benefit of the operator. That is, after determining the risk, the operator might be interested in the question "If the risk value is not acceptable what actions steps can be taken?". To this end, the determined risk can be used in two ways. On one hand, the operator can use the risk as a metric to design the controllers/detectors of the system optimally [14]. On the other hand, the risk can be used to allocate the security measures optimally [4, Chapter 5] which is the problem considered here.

Let n_w be the number of secure resources. In reality, secure resources refer to some form of secure communication channels for the sensors and actuators such that an attack cannot occur. If the number of secure resources (n_w) is equal to the number of actuators and sensors ($n_u + n_y$), then the SMAP is solved trivially. However, in general $n_w << n_u + n_y$ since secure communication channels are expensive[1]. Thus, we discuss a method to optimally allocate the security measures when they are limitedly available.

To formulate the problem, we define the following. The set of all sensors (actuators) is represented by $\mathcal{S}(\mathcal{A})$, where $|\mathcal{S}| = n_y(|\mathcal{A}| = n_u)$. The set of all vulnerabilities is represented by $\mathcal{V}$. Any sensor (actuator) is a vulnerability if the operator believes that an adversary might be able to access the sensor (actuator) channels. Thus $|\mathcal{V}| = n_v \leq n_y + n_u$. Let the set of secure resources be represented by $\mathcal{W}$ where $|\mathcal{W}| = n_w$. And as discussed before $n_w << n_v \leq n_y + n_u$. Then the SMAP has the following structure.

Firstly, the operator aims at optimizing the risk metric. Secondly, if an actuator (sensor) $i \in \mathcal{V}$ is secure, then the corresponding actuator (sensor) channels cannot be accessed by the adversary (C1). Recall from Section II, that the matrix $E_a(F_a)$ is a diagonal matrix with $E_a(i,i) = 1$ ($F_a(i,i) = 1$), if the actuator (sensor) channel i is under attack and zero otherwise. And, as discussed before, the operator can only secure n_w actuators (sensors) at most (C2). To this end, the optimal SMAP under actuator attacks can be formulated as

$$\{\hat{\gamma}^*_{OA}, W^*\} = \left\{ \begin{array}{l} \inf\limits_{z_i} \hat{\gamma}_{OA}(z) \\ \text{s.t. (C1)} E_a(i,i) = z_i,\ \forall i \in \mathcal{V} \\ \quad \text{(C2)} \sum\limits_{i \in \mathcal{V}} (1 - E_a(i,i)) \leq n_w \\ \quad z_i \in \{0,1\},\ \forall i \end{array} \right\} \quad (12)$$

where, $\hat{\gamma}^*_{OA}$ is the optimal risk after the security measures are allocated, the corresponding optimal vulnerabilities to be protected are represented by W^*, and where the constraint C2 considers that a vulnerable actuator is protected if and only if the corresponding actuator has $E_a(i,i) = 0$. Similarly, when the sensors are under attack, the optimal SMAP can be formulated by replacing $E_a(\cdot)$ by $F_a(\cdot)$ in (12).

The optimization problem (12) is hard to solve since it is a combinatorial problem. That is, the operator has to search through the whole set of $\mathcal{V}$ to secure n_w vulnerabilities. And it is well known that combinatorial problems with a large search space ($\mathcal{V}$) are NP-hard in general. Thus, providing a heuristic to solve (12) is beyond the scope of this paper and is left for future work. Interested readered are referred to [4, Chapter 5]. However, we provide a scheme to solve (12) when $|\mathcal{V}|$ is small in **Algorithm 1**.

Algorithm 1: Algorithm to solve SMAP

Initialization: $\beta, \epsilon, \Omega, \epsilon, \mathcal{V}, n_w$ an empty list γ_l
Step 1: Determine $\mathcal{Q}$ which is the set of all subsets of $\mathcal{V}$ with maximum cardinality n_w.
Step 2: For all $q_{(\cdot)} = \mathcal{Q}$, do:
1) Set $E_a(i,i) = 0$ if $i \in q_{(\cdot)}$ and 1 otherwise.
2) Determine $\hat{\gamma}_{OA}$ with the new E_a.
3) Append γ_l with $\hat{\gamma}_{OA}$
Step 3: Determine the minimum of γ_l which is γ^*_{OA}.
Step 4: Determine the corresponding $q^*_{(\cdot)}$ which yields γ^*_{OA}.
Step 5: Set $W^* \triangleq q^*_{(\cdot)}$.
Result: $\hat{\gamma}^*_{OA}, W^*$

The algorithm first determines all possible subsets of the vulnerabilities with the maximum cardinality of n_w. Then, the operator determines the maximum attack impact caused by the adversary when these various subsets of vulnerabilities are protected. It then selects the set of vulnerabilities (W^*) which yields the minimum attack impact (γ^*_{OA}).

In this section we discussed how the risk determined in Section IV can be used by the system operator to optimally allocate the security measures. In the next section, we will illustrate what results can be obtained by a simple example.

VI. Numerical example

In this section, the effectiveness of the proposed **Algorithm 1** is illustrated through a numerical example. Consider the system described in (4) with $C = C_J^T = I_3, E_a = I_2$

$$\left[\begin{array}{c|cc} A & B^\Delta \end{array}\right] = \left[\begin{array}{ccc|cc} 1 & 0 & 1 & 1.5+\delta & 0 \\ 1 & 0.5 & 0 & 0.3 & 0 \\ 0 & 1 & -0.5 & 0 & 1 \end{array}\right],$$

$\Omega \triangleq [-0.5,\ \ 0.5], A_e = A, B_e = B, C_e = C, \left[\begin{array}{c|c} D_c^T & K_e \end{array}\right] =$

$$\left[\begin{array}{cc|ccc} -0.066 & 0.178 & 0.393 & 0 & 1 \\ 0.047 & 0.940 & 1 & -0.048 & 0 \\ 0.524 & -1.346 & 0 & 1 & -0.996 \end{array}\right],$$

and all the other unspecified matrices are zero. In the system description, only the matrix B^Δ is a function of the uncertain variable. And the system has no uncertain zeros on the unit circle, which makes the condition of *Lemma 4.1* hold $\forall \delta \in \Omega$. Thus, for a nominal system with $\delta = 0$, the OOG obtained by solving (5) is $||\Sigma||^2_{\ell_{2e}, y_p \leftarrow y_r} = 197.76$.

Let $\epsilon_1 = 0.05, \beta_1 = \beta = 0.1$ and $N_1 = 235$ which satisfies (10). Here $1 - \epsilon_1$ represents the accuracy of the approximation of the probability operator in determining the

[1] by expensive we here mean encryption and processing costs

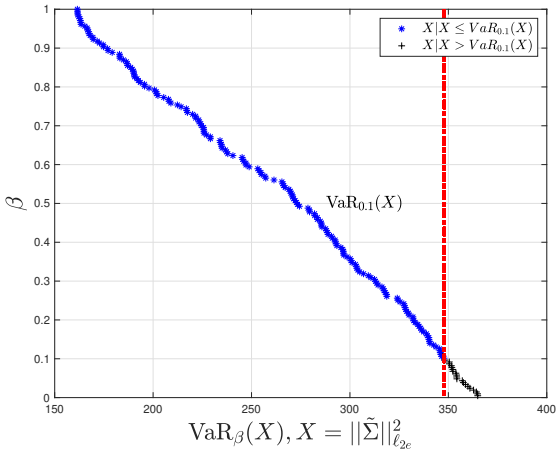

Fig. 2. The parameter β is shown on the Y-axis and the corresponding VaR_β on the X-axis. The red line indicates $\mathrm{VaR}_{0.1}$. The blue dots denotes the value of the impact random variable $X|X < VaR_{0.1}$, whereas the black dots denotes the value of the impact random variable $X|X > VaR_{0.1}$. It can be seen that the probability that $X > VaR_{0.1}$ is low when β is small.

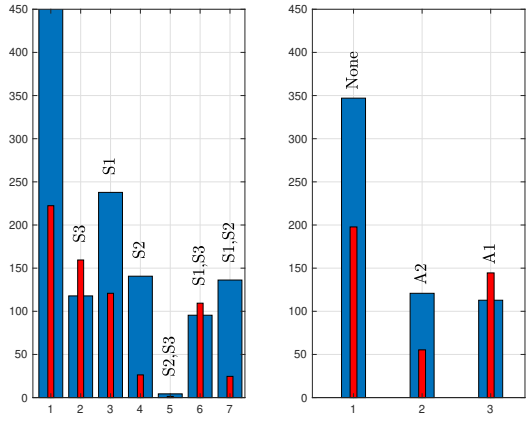

Fig. 3. The $\mathrm{VaR}_{0.1}$ after protecting various combination of sensors (actuators) are depicted on the left (right) figure in blue. The text on the top of each bar denotes the sensor (actuator) that is protected. For instance, "None" represents that none of the sensors are protected. The bar at position "1" of the figure in left corresponds to the risk when none of the sensors are protected and the corresponding risk was found to be 9081.4. The Y axis of the figure is not extended to show this value since it would affect clarity of the figure. The plots in red represent the impact on the nominal system after the security measures are allocated using the impact on the nominal system as a metric.

VaR_β. And β_1 represents the guarantee. We then solve (11) and obtain $\gamma_{OA} = 347.15$.

The optimization problem (11) is solved as follows. The set Ω is sampled for N_1 samples. The value of γ_i $\forall i = \{1, \ldots, N_1\}$ is obtained by solving the SDP (9). From these values of γ_i, we choose γ_{OA} such that the $\mathbb{P}(\gamma_i \geq \gamma_{OA}) = \beta$. Maintaining ϵ_1 and β_1 constant, for varying values of β, the value of $\gamma_{OA} = \mathrm{VaR}_\beta(X)$ is shown in Fig. 2.

Fig. 2 depicts that the value of the impact is greater than the VaR with a probability β and confidence $1-\epsilon_1$. To recall, by determining VaR_β for a given β, the operator can ensure that the impact of any stealthy attack impact is greater than VaR_β with probability β. After determining the VaR, the operator decides if the risk is acceptable or not. If the risk is not acceptable, the risk can be minimized for a given β by implementing additional security measures.

We next use the risk metric determined to allocate the security measures. Let $\mathcal{V} = \mathcal{S} = \{S1, S2, S3\}$. That is, we assume that all sensors communication channels are vulnerable to attacks. Then, we determine the risk when there are no security measures available ($|\mathcal{X}| = 0$). Next, we determine the risks when there is only one security measure available ($|\mathcal{X}| = 1$). That is, we determine the risks corresponding to the setup where either $S1, S2$ or $S3$ is protected. And finally, we determine the risks when there are two security measure available ($|\mathcal{X}| = 2$). That is, we determine the risk corresponding to the setup where either $\{S1, S2\}, \{S2, S3\}$ or $\{S3, S1\}$ are protected. The risks are depicted in the left diagram of Fig. 3 in blue, where the text on the top of the bar depicts the sensors that are protected. From Fig 3, we can conclude that (i) when $|\mathcal{X}| = 1$, it is optimal to secure $S3$ since it minimizes the risk the most, and (ii) when $|\mathcal{X}| = 2$, it is optimal to secure $S2$, and $S3$.

We repeat the procedure for the actuators where $\mathcal{V} = \mathcal{A} = \{A1, A2\}$. That is, we assume that all actuator communication channels are vulnerable to attacks. The risks are depicted in the right diagram of Fig. 3 in blue. From Fig. 3, we can conclude that (i) when $|\mathcal{X}| = 1$, it is optimal to secure $A1$ (actuator 1), and (ii) it is much more riskier to leave the sensors unprotected since the risk of unprotected sensors is higher than unprotected actuators.

Finally, we show the advantage of using the risk metric. We use the impact on the nominal system as a metric to allocate the protection resources. That is, instead of solving the SMAP with the risk determined from (11), we simply solve the optimization problem (9) for the nominal system and use it as a metric to allocate the security measure. To this end, we determine the impact on the nominal system when $|\mathcal{X}| = 0, |\mathcal{X}| = 1$, and $|\mathcal{X}| = 2$. The corresponding impact are shown in Fig. 3 in red. It can be seen that the conclusion that we drew using the risk metric are not reproducible when we use the nominal impact as a metric. For instance, when the sensors are under attack and $|\mathcal{X}| = 1$, the conclusion from the risk metric is to protect $S3$, whereas if we use the nominal impact, we end up protecting $S2$. Similarly, when the actuators are under attack and $|\mathcal{X}| = 1$, the conclusion from the risk metric is to protect $A1$, whereas if we use the nominal impact, we end up protecting $A2$.

VII. Conclusion

In this paper, we first addressed the RAP of false data attacks injected by an omniscient adversary on uncertain control systems. We formulated the RAP and observed that it is a non-convex infinite robust optimization problem. Using the theory of dissipative systems and scenario approach, we approximated the RAP as a convex SDP with probabilistic certificates. The necessary and sufficient conditions for the risk to be bounded were also formulated. Secondly, we

consider the optimal SMAP. We used the risk determined as a metric to formulate the SMAP. We provide a preliminary algorithm to solve the allocation problem. The results were depicted through a numerical example. Future works include (i) considering a process with process and measurement noise, and (ii) providing a more detailed solution for the allocation problem.

APPENDIX

PROOF OF *Lemma 4.2*

Proof: To recall, the optimization problem (9) was formulated using 3) in [15, Proposition 2] where $y_1 = \sqrt{\gamma} y_r$ and $y_2 = y_p$. Due to the equivalency between 3) and 4) of [15, Proposition 2], the Frequency Domain Inequality (FDI) $G_1(\bar{z})^T G_1(z) - G_2(\bar{z})^T G_2(z) \succeq 0$ should hold $\forall\, |z| = 1$. Since we know that $y_1 = \sqrt{\gamma} y_r$ and $y_2 = y_p$, we can deduce that $G_1(z)$ corresponds to $\sqrt{\gamma}\tilde{G}_{r,i}(z)$ and $G_2(z)$ to $\tilde{G}_{p,i}(z)$ in FDI, where $\tilde{G}_{r,i}(z_1) \triangleq C_{r,i}(z_1 I - A_{cl,i})^{-1} B_{cl,i} + D_{r,i}$ and $\tilde{G}_{p,i}(z_1) \triangleq C_{p,i}(z_1 I - A_{cl,i})^{-1} B_{cl,i} + D_{p,i}$. Thus, (9) can be rewritten as

$$\inf_{\gamma_i \geq 0} \left\{ \gamma_i \left| \underbrace{\gamma_i \tilde{G}_{r,i}(\bar{z})^T \tilde{G}_{r,i}(z) - \tilde{G}_{p,i}^T(\bar{z}) \tilde{G}_{p,i}(z)}_{H(z,\gamma_i)} \succeq 0, \forall |z| = 1 \right. \right\}$$

which is equivalent to

$$\inf_{\gamma_i \geq 0} \left\{ \gamma_i \left| x^H H(z, \gamma_i) x \geq 0, \forall |z| = 1 \right. \right\} \tag{13}$$

Next, let us define the following sets

$$\begin{aligned}
\mathcal{Z}_{pr} &\triangleq \{x \in \mathbb{C}^{n_a} \mid \tilde{G}_{r,i}(z)x = 0, \tilde{G}_{p,i}(z)x = 0\},\\
\mathcal{Z} &\triangleq \{x \in \mathbb{C}^{n_a} \mid \tilde{G}_{r,i}(z)x \neq 0, \tilde{G}_{p,i}(z)x \neq 0\},\\
\mathcal{Z}_r &\triangleq \{x \in \mathbb{C}^{n_a} \mid \tilde{G}_{r,i}(z)x = 0, \tilde{G}_{p,i}(z)x \neq 0\},\\
\mathcal{Z}_p &\triangleq \{x \in \mathbb{C}^{n_a} \mid \tilde{G}_{r,i}(z)x \neq 0, \tilde{G}_{p,i}(z)x = 0\}.
\end{aligned}$$

In the above definitions, each set corresponds to a combination of two logical conditions of $\tilde{G}_{r,i}(z)x$ and $\tilde{G}_{p,i}(z)x$. Therefore, the union of all four sets explores all possible combinations of the two logical conditions, and thus their union corresponds to the entire set $\mathbb{C}^{n_a}$.

For any given z such that $|z| = 1$, if $x \in \mathcal{Z}_p$, choosing $\gamma = 0$ satisfies the constraint of (13). Similarly, if $x \in \mathcal{Z}$, then $\gamma = \sup_{|z|=1, x \in \mathcal{Z}} \frac{x^H \left\{ \tilde{G}_{r,i}^T(\bar{z}) \tilde{G}_{r,i}(z) \right\} x}{x^H \left\{ \tilde{G}_{p,i}^T(\bar{z}) \tilde{G}_{p,i}(z) \right\} x}$. This ratio is well defined since the denominator cannot become zero (since $x \in \mathcal{Z}$), and the ratio is bounded since we have assumed that the transfer functions $\tilde{G}_{r,i}(z)$ and $\tilde{G}_{p,i}(z)$ are always stable (*Assumption 2.1*). Therefore, we have proven that the value of (13) is bounded whenever $x \in \mathcal{Z}_p \cup \mathcal{Z}$. We next begin by proving that the lemma statements are sufficient for (13) to be bounded whenever $x \in \mathcal{Z}_r \cup \mathcal{Z}_{pr}$.

Sufficiency: Assume that condition (1) of the lemma statement holds. By definition of a zero $\forall |z| = 1, \nexists s \neq 0 \in \mathbb{C}^{n_a}$ s.t. $\tilde{G}_{r,i}(z)s = 0$. Thus it follows that $\mathcal{Z}_r = \mathcal{Z}_{pr} = \emptyset$.

Assume that condition (2) of the lemma statement holds. Then by definition of a zero $\forall |z| = 1, \nexists s \neq 0$ such that $\tilde{G}_{r,i}(z)s = 0$ and $\tilde{G}_{p,i}(z)s \neq 0$. Thus it follows that $\mathcal{Z}_r = \emptyset$. And if $x \in \mathcal{Z}_{pr}$, then picking $\gamma = 0$ simply satisfies the constraint of (13). This concludes the proof on sufficiency.

Necessity: We prove by contradiction. Assume that there exists a bounded γ which solves the optimization problem (13). And we also assume that there exists a complex number z_1 on the unit circle which is a zero of the system $\tilde{\Sigma}_{r,i}$ (including multiplicity and input direction) but are not zeros of $\tilde{\Sigma}_{p,i}$. By definition of a zero, it holds that $\exists s \neq 0$ such that $\tilde{G}_{r,i}(z_1)s = 0,\ \tilde{G}_{p,i}(z_1)s \neq 0$. Thus, $\mathcal{Z}_{rp} \neq \emptyset$ and becomes a part of the feasible set for x. Then, if $z = z_1$ and $x = s$, the constraint of (13) can be rewritten as $-s^H \tilde{G}_{p,i}^T(\bar{z}_1) \tilde{G}_{p,i}(z_1)s \geq 0$ which cannot hold since $\tilde{G}_{p,i}(z_1)s \neq 0$. That is, the feasibility set of (13) is empty which contradicts our assumption. In terms of the primal problem (9), it means that there cannot exist a bound to its optimal value. This concludes the proof. ■

REFERENCES

[1] T. Samad, P. McLaughlin, and J. Lu, "System architecture for process automation: Review and trends," *Journal of Process Control*, vol. 17, no. 3, pp. 191–201, 2007.

[2] K. E. Hemsley, E. Fisher, *et al.*, "History of industrial control system cyber incidents," tech. rep., Idaho National Lab.(INL), Idaho Falls, ID (United States), 2018.

[3] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.

[4] J. Milošević, *Security metrics and allocation of security resources for control systems*. PhD thesis, KTH Royal Institute of Technology, 2020.

[5] J. Milošević, H. Sandberg, and K. H. Johansson, "Estimating the impact of cyber-attack strategies for stochastic networked control systems," *IEEE Transactions on Control of Network Systems*, 2019.

[6] C. Murguia, I. Shames, J. Ruths, and D. Nešić, "Security metrics and synthesis of secure control systems," *Automatica*, vol. 115, p. 108757, 2020.

[7] Y. W. Law, T. Alpcan, and M. Palaniswami, "Security games for risk minimization in automatic generation control," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 223–232, 2014.

[8] E. Bompard, C. Gao, R. Napoli, A. Russo, M. Masera, and A. Stefanini, "Risk assessment of malicious attacks against power systems," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 39, no. 5, pp. 1074–1085, 2009.

[9] X. Liu, M. Shahidehpour, Y. Cao, L. Wu, W. Wei, and X. Liu, "Microgrid risk analysis considering the impact of cyber attacks on solar PV and ESS control systems," *IEEE transactions on smart grid*, vol. 8, no. 3, pp. 1330–1339, 2016.

[10] A. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: the output-to-output ℓ_2-gain," in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 2582–2587, IEEE, 2015.

[11] H. Mausser and D. Rosen, "Beyond VaR: from measuring risk to managing risk," in *Proceedings of the IEEE/IAFE 1999 Conference on Computational Intelligence for Financial Engineering (CIFEr)(IEEE Cat. No. 99TH8408)*, pp. 163–178, IEEE, 1999.

[12] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of operations research*, vol. 23, no. 4, pp. 769–805, 1998.

[13] G. C. Calafiore and F. Dabbene, "Probabilistic robust control," in *2007 American Control Conference*, pp. 147–158, IEEE, 2007.

[14] M. I. Müller, J. Milošević, H. Sandberg, and C. R. Rojas, "A risk-theoretical approach to $\mathcal{H}_2$-optimal control under covert attacks," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 4553–4558, 2018.

[15] A. M. Teixeira, "Optimal stealthy attacks on actuators for strictly proper systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 4385–4390, IEEE, 2019.