2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

On the Computation of Worst Attacks: a LP Framework

Nabil H. Hirzallah and Petros G. Voulgaris

***Abstract*— We consider the problem of false data injection attacks modeled as additive disturbances in various parts of a general LTI feedback system and derive necessary and sufficient conditions for the existence of stealthy unbounded attacks. We also consider the problem of characterizing the worst, bounded and stealthy attacks. This problem involves a maximization of a convex function subject to convex constraints, and hence, in principle, it is not easy to solve. However, by employing a ℓ_∞ framework, we show how tractable Linear Programming (LP) methods can be used to obtain the worst attack design. Moreover, we provide a controller synthesis iterative method to minimize the worst impact of such attacks and test its efficacy in a power system component.**

I. Introduction

Real world incidents and research papers have shown that stealthy attacks can be carefully designed to cause significant damage in control systems. Recent work on security of cyber-physical systems from a control-theoretic perspective has been focused on the characterization of feasible attacks and proposing ways for detection and/or improving the resiliency of the control system subject to such attacks. The type of attacks studied can be generally split into two categories: static attacks (attacks that do not take into account the dynamics of the system and/or do not affect the states of the system directly) or dynamic attacks. Attacks under each category can be classified as stealthy or not stealthy depending on the assumptions and the detection methods used. Examples of static attacks include attacks on the power system state estimators [1], where a carefully designed bias can be added to the sensor measurements without being detected by the commonly used statistical detection methods. Another work on static attacks is by [2] and [3] where they showed that the states of the system cannot be accurately reconstructed if half of the sensors are attacked. Both papers propose computationally intensive methods to reconstruct the states when less than half of the sensors are attacked. Their work was extended by [4] where the authors provide a framework to to reconstruct the states that is robust to additive and multiplicative modeling errors. On the other hand, research work related to dynamic attacks include [5] where the authors provide necessary and sufficient conditions for the existence of unbounded stealthy actuator and/or sensor attacks. In addition they proposed dual rate control to detect unbounded stealthy actuator attacks (zero dynamics attacks). In [6] the authors inject a random signal (unknown to the attacker) into the system to detect replay attacks at the expense of increasing the cost of the LQG controller. In [7] the case for finding the worst bias constant (steady state) attack has been considered and a tractable procedure to compute it has been developed where the energy of the detection signal was considered as a measure of stealthiness. In [8] optimal attacks are computed on a LQG systems that minimize the K-L divergence between the true and falsified state estimates such that the attack impact is above a specified a limit, showing that the optimal attacks are additive white noise. In [9] optimal actuator attacks are designed using the minimum principle that maximizes a quadratic cost related to the error between the healthy (un-attacked) system and the attacked system while minimizing the attack cost, without including any stealthiness requirement.

In this work, we consider signal attacks where the general problem from the attacker's perspective is to find the attack input $d = \{d(k)\}$ so that it is stealthy while inflicting the maximum damage on the performance variable $z = \{z(k)\}$. We showed in our previous work [5] that unbounded attacks for LTI systems are related to the unstable zeros and/or poles of the open loop system. However, in this paper we consider the problem of characterizing the worst, bounded and stealthy attacks. This problem involves a maximization of a convex function subject to convex constraints. We propose different attack resource constraints to make the problem more practical. More specifically, we assume that the attacker has a finite time window $\{0, 1, \ldots, t_a\}$ to attack the system and inflict the maximum damage before the attack is over, and we attempt to solve the following three attack scenarios: Scenario 1 : Attacker can attack in a finite time window up to $t = t_a$, his goal is to inflict the maximum damage anywhere (before or after t_a) while remaining stealthy for all t. Scenario 2 : Attacker can attack in a finite interval up to $t = t_a$, his goal is to inflict the maximum damage anywhere (before or after t_a) while remaining stealthy for $t \leq t_a$ (does not care if detected after the attack is over). Scenario 3 : Attacker can attack in a finite interval up to $t = t_a$, his goal is to inflict the maximum damage at $t \leq t_a$ while remaining stealthy for $t \leq t_a$. We show that by employing a ℓ_∞ framework, tractable Linear Programming (LP) methods can be used to compute the worst attack for the above three scenarios.

Our work is closely related to [7], [10], [8], [9], [11], [12] and [13]. However, we don't assume a constant d such as in [7] where they assume the system is in steady state. In addi-

N. H. Hirzallah is a PhD candidate with the Electrical and Computer Engineering Department, University of Illinois, Urbana, IL, USA. hirzall2@illinois.edu

P. G. Voulgaris is with the Aerospace Engineering Department and the Coordinated Science Laboratory, University of Illinois, Urbana, IL, USA, and with Khalifa University, Abu Dhabi, UAE. voulgari@illinois.edu

This work was supported in part by the National Science Foundation under NSF awards CMMI-1663460, ECCS-1739732.

tion, the work in the mentioned references does not address attack impact and stealthiness after the attack is concluded, and relate to either a specific detection method (e.g. residual detectors) which assumes certain thresholding mechanisms that may be stochastic, or to a specific controller in use. We plan to investigate these problems in a more general input-output fashion that does not depend on the particular controller used, and in a totally deterministic worst case scenario. In other words, the assumed noise thresholds are based on the existence of a worst case magnitude bounded noise. In this sense, the noise is allowed to "conspire" with the attacker to keep the detection signals within what is assumed normal operation.

In the second part of this paper, we build on the worst attack design problem and provide a novel K-d controller synthesis iterative method to minimize the performance cost without increasing the impact of the worst attack. Each iteration is a LP and alternates between finding the worst attack d for a given controller K, and finding the next K that minimizes the performance cost while keeping a non-increasing upper bound on the worst case impact inflicted by d.

Some standard notation we use is as follows: $\mathbb{Z}_+$, $\mathbb{R}^n$, $\mathbb{C}^n$ and $\mathbb{R}^{n\times m}$ denote the sets of non-negative integers, n-dimensional real vectors, n-dimensional complex vectors and $n\times m$ dimensional real matrices, respectively. For any $\mathbb{R}^n$ or $\mathbb{C}^n$ vector x we denote x' its transpose and $|x| := \max_i \sqrt{x_i^2}$ where $x' = [x_1, x_2, ..., x_n]$; for a sequence of real n-dimensional vectors, $x = \{x(k)\}_{k\in\mathbb{Z}_+}$ we denote $||x||_\infty := \sup_k |x(k)|$; for a sequence of real $n \times m$ dimensional real matrices $G = \{G_k\}_{k\in\mathbb{Z}_+}$ we denote its z-transform $G(z) := \sum_{k=0}^{\infty} G_k z^{-k}$; and if viewed as the pulse response of the LTI system G then $||G||_1 = \sup_{||x||_\infty\le 1} ||Gx||_\infty$. We will also be using the standard notions for zeros and coprime factorizations of a LTI system G (e.g., [14], [15], [16].)

II. Problem Setup

We consider the case of a general signal attack d on a closed loop system of Figure 1. Let $\Phi(K)$ describe the effect of d on the performance variable z and on the monitoring signal ψ, i.e. let $\Phi = \begin{bmatrix}\Phi_{zd}\\ \Phi_{\psi d}\end{bmatrix} =: d \mapsto \begin{bmatrix} z\\ \psi\end{bmatrix}$. The monitoring signal ψ consists of the measured output y and the control signal u; it can however contain any other information that is recorded and measured, e.g., reference inputs. In this setup, we assume that there may be other external disturbances and noise inputs which are "normal", i.e., not malicious attackers, which are not shown in the figure. Also, all the formulation deals with discrete-time systems and signals.

The attacker's goal can be stated in general as

$$\begin{aligned} &\max_d \|z\|_\infty \\ &\text{s.t. } \|\psi\|_\infty \le \theta, \end{aligned} \tag{1}$$

where θ is an alarm threshold, associated with the afore mentioned normal set of disturbances. In our previous work [5], we established exact conditions for stealthiness of unbounded actuator and sensor attacks which can totally destroy the system. These attacks are ultimately related to the open loop plant P, and for LTI systems in particular, to the non-minimum phase (unstable) zeros and unstable poles of P. We note, as pointed in [5], that unstable zeros can also be due to the sampled data implementation of controllers.

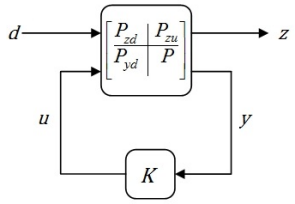


Fig. 1: General setup of input-output maps.

In this general setup of Figure 1, we elaborate on the existence of stealthy unbounded attacks using an input-output approach. In particular, considering a left coprime factorization ([14], [15], [16]) for the part of the generalized system that connects inputs to the measured output $y = P_{yd}d + Pu$ in the open loop, we have

$$[P_{yd}\ P] = \tilde{M}^{-1}[\tilde{N}_{yd}\ \tilde{N}]$$

Using a left coprime factorization for the stabilizing controller $K = YX^{-1}$ we can express

$$\psi = \begin{bmatrix} y\\ u\end{bmatrix} = \begin{bmatrix} X\\ Y\end{bmatrix} W^{-1}\tilde{M}P_{yd}d = \begin{bmatrix} X\\ Y\end{bmatrix} W^{-1}\tilde{N}_{yd}d.$$

where $W = \tilde{M}X - \tilde{N}Y$. Since W is stable and, by stability of the closed loop, has a stable inverse W^{-1} we have that the detectability of d depends on the unstable zeros of $\tilde{N}_{yd}$: unbounded stealthy attacks d are possible if and only if $\tilde{N}_{yd}$ has unstable zeros.

For actuator only attacks

$$P_{yd} = P = \tilde{M}^{-1}\tilde{N} \Longrightarrow \tilde{N}_{yd} = \tilde{N}$$

while for sensor only attacks

$$P_{yd} = I = \tilde{M}^{-1}\tilde{M} \Longrightarrow \tilde{N}_{yd} = \tilde{M}.$$

Hence, this shows how the unstable zeros of P (which are the unstable zeros of $\tilde{N}$) and the unstable poles of P (which are the unstable zeros of $\tilde{M}$) relate to the actuator and sensor attacks considered in [5]. Multirate sampling can potentially remove unstable zeros of $\tilde{N}_{yd}$ as it was shown in [5] for unbounded actuator attacks, but it cannot work for total sensor unbounded attacks.

In the following we consider the case of bounded in magnitude (and time) attacks with various levels of stealth. The question we want to address is how to compute the worst possible bounded attacks and how to defend against such attacks by a suitable controller design.

III. Computation of Worst Attack

We consider the problem of computing the worst case attack in (1) when the attacker has a finite time window $\{0, 1, \dots, t_a\}$ to attack the system. In addition, we require the attack to remain stealthy after the attack is over. This

allows for repeatedly attacking the system without triggering monitoring signal alarm.

Specifically, consider the optimization problem in (1), we are interested in finding the worst, stealthy, bounded (in magnitude and time) attack. Assume the LTI closed loop system $\Phi(K)$ is stable and let t_{zd} and $t_{\psi d}$ be design parameters related to the decay rate of the pulse responses of of Φ_{zd} and $\Phi_{\psi d}$ respectively. These parameters determine the time windows that the attacker cares for impact and stealthiness respectively. Suppose the intruder can only attack the system during a finite interval $\{0, 1, \ldots, t_a\}$, with attack magnitude less than or equal to α. Then, a corresponding problem of interest can be formulated as

$$\begin{aligned} &\max_{d} \|z\|_{\infty}^{[0,t_a+t_{zd}]} \\ &\text{s.t. } \|\psi\|_{\infty}^{[0,t_a+t_{\psi d}]} \le \theta, \\ &\quad |d(k)| \le \alpha, \ k = 0, 1, \ldots, t_a, \\ &\quad d(k) = 0, \quad k = t_a + 1, \ldots. \end{aligned} \tag{2}$$

where $\|z\|_{\infty}^{[0,t_a+t_{zd}]} = \max_{0\le k\le t_a+t_{zd}} = |z(k)|$, and similarly $\|\psi\|_{\infty}^{[0,t_a+t_{\psi d}]} = \max_{0\le k\le t_a+t_{\psi d}} = |\psi(k)|$.

The system of equations governing the output z when subjected to the attack input d for each instance of time are given by

$$\begin{bmatrix} z(0) \\ z(1) \\ \vdots \\ z(t_a) \\ z(t_a+1) \\ \vdots \\ z(t_a+t_{zd}) \end{bmatrix} = \begin{bmatrix} \Phi_{zd}(0) & 0 & 0 & \cdots \\ \Phi_{zd}(1) & \Phi_{zd}(0) & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_{zd}(t_a) & \Phi_{zd}(t_a-1) & \Phi_{zd}(t_a-2) & \cdots \\ \Phi_{zd}(t_a+1) & \Phi_{zd}(t_a) & \Phi_{zd}(t_a-1) & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_{zd}(t_a+t_{zd}) & \Phi_{zd}(t_a+t_{zd}-1) & \Phi_{zd}(t_a+t_{zd}-2) & \cdots \end{bmatrix} \begin{bmatrix} d(0) \\ d(1) \\ \vdots \\ d(t_a) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3}$$

where $d(k) = 0$ for $t > t_a$. The objective is to find the sequence $\{d(k)\}$, $k = \{0, \ldots, t_a\}$ that maximizes $\|z\|_{\infty}^{[0,t_a+t_{zd}]}$ such that $\|\psi\|_{\infty}^{[0,t_a+t_{\psi d}]} \le \theta$. This corresponds to selecting the optimal row in (3) to be maximized and finding the optimal d that would maximize this row. In view of the above, the following proposition is obvious.

Proposition 1: Problem (2) can be formulated as the following optimization problem:

$$\begin{aligned} &\max_{d,n\in\{0,1,\ldots,t_a+t_{zd}\}} \sum_{k=0}^{n} \Phi_{zd}(n-k)d(k) \\ &\text{s.t. } \Big| \sum_{k=0}^{\tau} \Phi_{\psi d}(\tau-k)d(k) \Big| \le \theta, \ \tau = 0, 1, \ldots, t_a + t_{\psi d}, \\ &\quad |d(k)| \le \alpha, \ k = 0, 1, \ldots, t_a, \\ &\quad d(k) = 0, \quad k = t_a + 1, \ldots \ . \end{aligned} \tag{4}$$

After finding the worst case attack $\hat{d}$, the worst case impact can be obtained by computing $\left\|\Phi_{zd}\hat{d}\right\|_{\infty}$.

Note also that an optimal $\hat{d}$ can always be selected so that $\Big| \sum_{k=0}^{n} \Phi_{zd}(n-k)\hat{d}(k) \Big| = \sum_{k=0}^{n} \Phi_{zd}(n-k)\hat{d}(k)$, thus the expression for the cost in (4).

Remark 2: The objective function looks for the optimal row in the set $\{0, \ldots, t_a + t_{zd}\}$. We can always choose a sufficiently long t_{zd}, determined by the decay rate of Φ_{zd} and the bound α on d, to ensure that we capture the worst case $\|z\|_{\infty} = \sup_{t_{zd}} \|z\|_{\infty}^{[0,t_a+t_{zd}]}$.

Remark 3: Note that the first set of constraints ensures the monitoring signal ψ is below a threshold level ($\|\psi\|_{\infty} \le \theta$) during and after the attack interval. Since we assume that $\Phi_{\psi d}$ is stable and that $d(k) = 0$ for $t > t_a$, if $t_{\psi d}$ is chosen long enough, depending on the decay rate of $\Phi_{\psi d}$ and the bound α, one can guarantee that d is undetectable for all t. Therefore, to guarantee stealthiness for all t it is sufficient to enforce the monitoring constraints up to $t_a + t_{\psi d}$. The last set of constraints ensures the attack is bounded and decays to zero at the end of the attack interval.

Remark 4: Remarks 2 and 3 basically state that for a priori computable t_{zd} and $t_{\psi d}$, problems (2) and (4) solve the following problem

$$\begin{aligned} &\max_{d} \|z\|_{\infty} \\ &\text{s.t. } \|\psi\|_{\infty} \le \theta, \\ &\quad |d(k)| \le \alpha, \ k = 0, 1, \ldots, t_a, \\ &\quad d(k) = 0, \quad k = t_a + 1, \ldots \ . \end{aligned} \tag{5}$$

Remark 5: Problem (4) is LP for a fixed n (fixed row) which can be solved efficiently. Fixing n transforms the objective function to a linear function under linear (polytopic) constraints. However, one has to solve (in principle) $t_a + t_{zd}$ LPs.

In the sequel, we consider certain cases which simplify further the computations. Specifically, we consider the problem of computing the worst case attack when the attacker has a finite time window $k = \{0, \ldots, t_a\}$ to attack the system such as in Proposition 1. However, in this case we assume that the intruder does not mind being detected after the attack is over, i.e., stealthiness constraints are checked up to $t = t_a$ only. The following corollary describes how to construct the optimal d.

Corollary 6: Consider the optimization Problem in (2) with $t_{\psi d} = 0$ (finite stealthiness interval). Then, its solution can be obtained by solving

$$\begin{aligned} &\max_{d,n\in\{t_a,\ldots,t_a+t_{zd}\}} \sum_{k=0}^{n} \Phi_{zd}(n-k)d(k) \\ &\text{s.t. } \Big| \sum_{k=0}^{\tau} \Phi_{\psi d}(\tau-k)d(k) \Big| \le \theta, \ \tau = 0, 1, \ldots, t_a, \\ &\quad |d(k)| \le \alpha, \ k = 0, 1, \ldots, t_a, \\ &\quad d(k) = 0, \quad k = t_a + 1, \ldots \ . \end{aligned} \tag{6}$$

Proof: We will prove that the optimal row to be maximized is in the set $\{z(t_a), \ldots, z(t_a + t_{zd})\}$. Let $\hat{d}$ be the worst attack that maximizes $\mu =: \|z\|_{\infty}^{[0,t_a+t_{zd}]}$ found by solving for the maximum impact over all the rows of (3) where the stealthiness constraints are enforced up to $t = t_a$. Assume that $\hat{d}$ was found by maximizing any row before $z(t_a)$ calling it row i. Since the stealthiness constraints are imposed only up to t_a and $\Phi(K)$ is LTI, we can delay $\hat{d}$ by

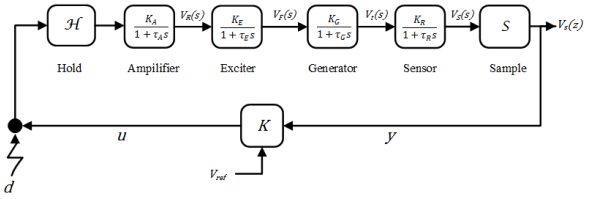


Fig. 2: A simplified automatic voltage regulator block diagram.

$t_a - i$ steps (shift $\hat{d}$ to the right) so that $\|z\|_\infty^{[0,t_a+t_{zd}]} = \mu$ is achieved by maximizing the row $z(t_a)$ without violating the stealthiness constraints. In addition, we cannot shift the attack beyond $z(t_a)$ since $\hat{d}(k) = 0$ for $t > t_a$. As a result, maximizing $\|z\|_\infty^{[0,t_a+t_{zd}]}$ is equivalent to maximizing $\|z\|_\infty^{[t_a,t_a+t_{zd}]}$ for $t_{\psi d} = 0$. ∎

Remark 7: The optimization problem in (6) differs from the problem in (4) in two ways: First, the stealthiness constraints set in (6) is a subset of the set in (4), since in (6) the objective is to remain stealthy only during the attack interval, where in (4) the stealthiness condition is enforced at all times. Therefore, the attack designed using Corollary 6 yields worse impact in the ℓ_∞ sense than the attack designed using Proposition 1. The second difference is in the objective function where in (4) we have to maximize each row in (3) to find the worst attack (i.e., $t_a + t_{zd}$ LPs), while in (6) we only need to maximize the last rows associated with $[z(t_a), \dots, z(t_a + t_{zd})]'$ (i.e., $t_{zd} + 1$ LPs). An immediate corollary is as follows.

Corollary 8: Let $t_{zd} = t_{\psi d} = 0$, i.e., the attacker cares to inflict maximum damage in the window up to t_a while does not care for stealthiness after t_a. Then, the optimal d is obtained by solving the following single LP

$$\begin{aligned} &\max_d \sum_{k=0}^{t_a} \Phi_{zd}(t_a - k)d(k) \\ &\text{s.t. } \left| \sum_{k=0}^{\tau} \Phi_{\psi d}(\tau - k)d(k) \right| \le \theta, \ \tau = 0, 1, \dots, t_a, \\ &\quad |d(k)| \le \alpha, \ k = 0, 1, \dots, t_a. \end{aligned} \tag{7}$$

The above corollary states that computing the worst attack when the attack impact and stealthiness constraints are desired to be inside the attack interval only is equivalent to solving (6) for $n = t_a$.

Remark 9: If $\Phi_{\psi d}$ is non-minimum phase and α is not specified, then the optimization problems in Corollary 6 and Corollary 8 will yield unbounded zero dynamics attacks [5].

IV. Examples - Worst Attack Computation

In this section we work on an example of a real power system component and compute the worst attack for different scenarios.

A. Automatic Voltage Regulator

The automatic voltage regulator (AVR) or the generator excitation control, specifies the terminal voltage magnitude of a synchronous generator by controlling the reactive power. A simplified block diagram of a linearized AVR is shown in Figure 2 [17]. An increase in the reactive power load of the generator results in a drop in the voltage magnitude across its terminals. The voltage drop is sensed by a potential transformer which then is rectified and compared to the reference voltage magnitude. The error signal is then amplified and raises the generator terminal voltage by controlling the excitation field. For a set of typical system parameters $K_A = 10, \tau_A = 0.1, K_E = 1, \tau_e = 0.4, K_G = 1, \tau_G = 1, K_R = 1, \tau_R = 0.05$ as in Figure 2. We consider actuator attacks as depicted in Figure 2 and seek to find the attack with the worst impact on V_F (excitation voltage) while keeping the monitoring vector $\psi = \begin{bmatrix} y \\ u \end{bmatrix}$ below a noise level threshold θ. Let K be a suitable controller for the system and let

$$\begin{aligned} P &= \mathcal{S}\frac{K_A}{1+\tau_A s}\frac{K_E}{1+\tau_E s}\frac{K_G}{1+\tau_G s}\frac{K_R}{1+\tau_R s}\mathcal{H} \text{ and} \\ P_F &= \mathcal{S}\frac{K_A}{1+\tau_A s}\frac{K_E}{1+\tau_E s}\mathcal{H} \end{aligned} \tag{8}$$

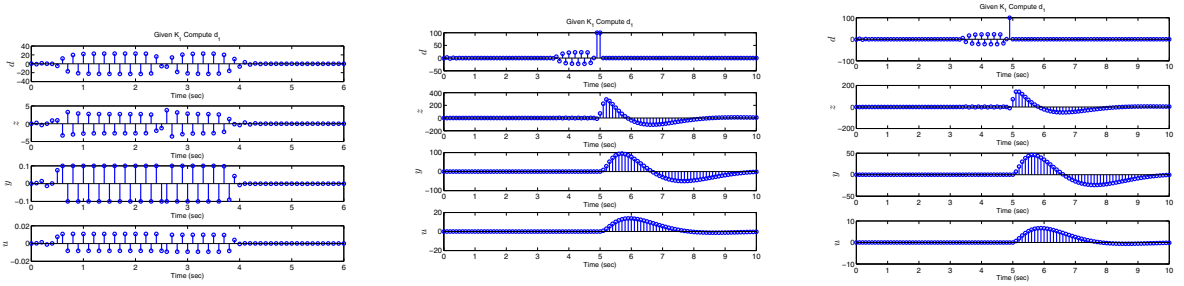
Then closed loop system $\Phi(K)$ describing the effect of d on $z = V_F$ and the monitoring vector ψ is given by

$$\Phi(K) = \begin{bmatrix} \Phi_{zd} \\ \Phi_{\psi d} \end{bmatrix} =: d \mapsto \begin{bmatrix} z \\ y \\ u \end{bmatrix} = \begin{bmatrix} \frac{P_F}{1+PK} \\ \frac{P}{1+PK} \\ \frac{PK}{1+PK} \end{bmatrix}$$

Given $K = \frac{0.1z-0.09}{z-1}$, then $\Phi(K)$ becomes

$$\begin{bmatrix} \Phi_{zd} \\ \Phi_{\psi d} \end{bmatrix} = \begin{bmatrix} \frac{0.8423z^4-1.162z^3-0.1551z^2+0.5433z-0.06808}{z^5-3.186z^4+3.794z^3-2.043z^2+0.4705z-0.03522} \\ \frac{0.01114z^4+0.05639z^3-0.03266z^2-0.03337z-0.001502}{z^5-3.186z^4+3.794z^3-2.043z^2+0.4705z-0.03522} \\ \frac{(1.11z^4+5.75z^3-2.59z^2-2.99z-0.135)\times 10^{-3}}{z^5-3.186z^4+3.794z^3-2.043z^2+0.4705z-0.03522} \end{bmatrix}$$

sampled at $T = 0.1$ seconds. We note that the has $\Phi_{\psi d}$ an unstable zero at $z = 1.42$. We compute the attack for 3 cases. In the first case we employ Proposition 1 to compute the worst attack for an attack interval $\{0, \dots, t_a\}$ that is stealthy for all t. In the second case, we compute the worst attack for an attack interval $\{0, \dots, t_a\}$ using Corollary 6, i.e., stealthiness requirement for $t \le t_a$ only. In the third case, we compute the worst attack using Corollary 8, i.e., $\max_d z(t_a)$ where the stealthiness requirement holds for $t \le t_a$ only. For all cases, we fix $t_a = 500$ (corresponding to 5 seconds), $\theta = 0.1$, $\alpha = 100$. Figures 3a, 3b and 3c show the computed worst attack signals for cases 1, 2 and 3 with their impact on the performance variable z and and monitoring signal ψ. Case 1 was obtained by maximizing $z(260)$ (corresponding to 2.6 seconds), case 2 was obtained by maximizing $z(520)$ (corresponding to 5.2 seconds) and case was obtained by maximizing $z(500)$ (corresponding to 2.6 seconds). We note that the maximum impact on z in case 2 is larger than in case 3 which in turn is larger than in case 1, confirming remark 7. We also show in Figure 4a a plot for the maximum impact on z for all t_a for case 1. This is obtained by iterating $t_a \ge 0$ and solving (4) until $\|z\|_\infty$ stops increasing. We note from Figure 4a that $\|z\|_\infty$ stops increasing after $t_a = 200$ (corresponding to 2 seconds). As a result, for this example and for $t_a \ge 200$, solving (4) is equivalent

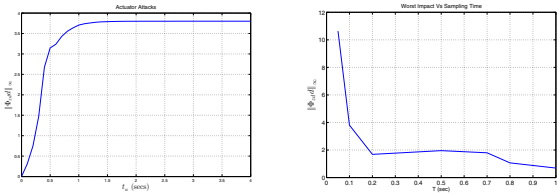


(a) Case 1 - Stealthy for all t.

(b) Case 2 - Stealthy for $t \leq t_a$ ($t_{\psi d} = 0$) and $t_{zd} \neq 0$.

(c) Case 3 - Stealthy for $t \leq t_a$ ($t_{\psi d} = 0$) and $t_{zd} = 0$.

Fig. 3: Worst attack computation with the effect of d on z, y and u.



(a) Different attack intervals t_a.

(b) Different sample and hold rates T.

Fig. 4: Worst impact, stealthy for all t, i.e., case 1.

to solving (1). Furthermore, we show in Figure 4b how of the worst impact yielded by the optimization problem in 4 changes with changing the sampling and hold time (T). It is not clear from the figure if a direct relationship between $\|z\|_\infty$ and T exists. This is because although for a faster a rate, the cardinality of the attack sequence for a fixed time interval increases allowing for extra optimization variables, however the number of stealthiness constraints also increases reducing the set of feasible solutions.

V. Controller Design for Resiliency - K-d iteration

In view of the previous discussion, a controller design procedure can be formulated based on LP. In particular, given a desired ℓ_1 performance level γ for attacks d, find K such that $\|\Phi_{zd}(K)\|_1 \leq \gamma$, and to ensure that for a given attack level characterized by $\|d\|_\infty \leq \alpha$, where α is an attack resource parameter, the "undetected loss" of the closed loop given by

$$\mu_\alpha := \max_d \|\Phi_{zd}(K)d\|_\infty \text{ s.t. } \|\Phi_{\psi d}(K)d\|_\infty \leq \theta, \ \|d\|_\infty \leq \alpha$$

remains below a desired level μ. Computing μ_α for a given K corresponds to the problem of computing the worst d of the previous section. A synthesis procedure can be developed by a "K-d" type of iteration:

- Given K_i with $\|\Phi_{zd}(K_i)\|_1 = \gamma_i$ find d_i from:

$$\mu_i := \max_d \|\Phi_{zd}(K_i)d\|_\infty$$
$$\text{s.t. } \|\Phi_{\psi d}(K_i)d\|_\infty \leq \theta, \ \|d\|_\infty \leq \alpha.$$

- Given d_i find K_{i+1} from:

$$\gamma_{i+1} := \min_K \|\Phi_{zd}(K)\|_1 \text{ s.t. } \|\Phi_{zd}(K)d_i\|_\infty \leq \mu_i$$

- At each iteration i the problem is a LP with

$$\gamma_i \leq \gamma_{i-1} \leq \gamma_0, \ \mu_i \leq \gamma_i \|d_i\|_\infty, \ \|d_i\|_\infty \leq \alpha.$$

The above formulation guarantees that the upper bound on the attack impact (i.e. μ_i) is non-increasing with each iteration.

VI. Examples - Controller Design for Resiliency

In this section we build on the AVR example in Section IV-A sampled at $T = 0.1$ seconds and seek to design a controller that minimizes the performance variable z while possibly minimizing the the impact of the worst attack d. Similar to section IV-A we start with a simple PI controller represented by the transfer function $K_1 = \frac{0.1z-0.09}{z-1}$. We use controller parametrization for stable transfer functions to set up the controller optimization problem. As a result, The maps Φ_{zd} and $\Phi_{\psi d}$ are given by

$$\begin{bmatrix}\Phi_{zd}\\ \Phi_{\psi d}\end{bmatrix} =: d \mapsto \begin{bmatrix}z\\ y\\ u\end{bmatrix} = \begin{bmatrix}\frac{P_F}{1+PK}\\ \frac{P}{1+PK}\\ \frac{PK}{1+PK}.\end{bmatrix} = \begin{bmatrix}P_F(1-PQ)\\ P(1-PQ)\\ PQ\end{bmatrix}$$

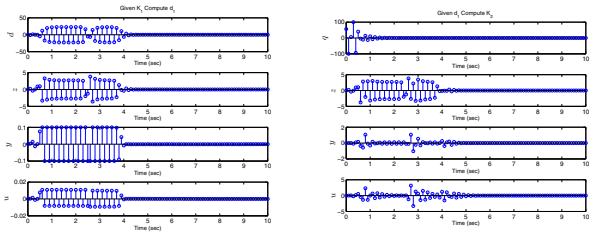
where $Q = \dfrac{K}{1+PK}$ and P is the open loop transfer function of the AVR system given in (8) along with P_F, both sampled at $T = 0.1$ seconds. The controller synthesis problem is carried on the affine parameter Q in the time domain using the following formulation:

- Given K_i with $\|\Phi_{zd}(K_i)\|_1 = \gamma_i$ find d_i from:

$$\mu_i = \max_{d, n \in \{0,1,\ldots,t_a+t_{zd}\}} \sum_{k=0}^{n} \Phi_{zd}(n-k)d(k)$$
$$\text{s.t. } \left|\sum_{k=0}^{\tau} \Phi_{\psi d}(\tau-k)d(k)\right| \leq \theta, \ \tau = 0,1,\ldots,t_a+t_{\psi d},$$
$$|d(k)| \leq \alpha, \ k = 0,1,\ldots,t_a,$$
$$d(k) = 0, \quad k = t_a+1,\ldots \ .$$

TABLE I: γ_i and μ_i for each iteration.

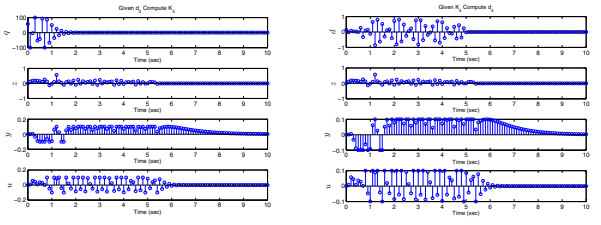
Iteration i	$\gamma_i = \|\Phi_{zd}(K_i)\|_1$	$\mu_i = \|\Phi_{zd}(K_i)d_i\|_\infty$
1	64.6449	3.8008
2	37.0244	2.0107
3	36.9109	0.5704
4	36.9109	0.5704



(a) Computation of d_1 given K_1, and the effect of d_1 on z, y and u.

(b) Computation of q_2, and the effect of d_1 on z, y and u controlled by K_2.

Fig. 5: Controller synthesis using K-d iteration. Iteration 1.



(a) Computation of q_4, and the effect of d_3 on z, y and u controlled by K_4.

(b) Computation of d_4 given K_4, and the effect of d_4 on z, y and u.

Fig. 6: Controller synthesis using K-d iteration. Iteration 4.

- Given d_i find K_{i+1} from:

$$\begin{aligned} \gamma_{i+1} := \quad & \min_q \|p_F * (1 - p * q)\|_1 \\ \text{s.t.} \quad & \|p_F * (1 - p * q) * d_i\|_\infty \le \mu_i \\ & \|q\|_\infty \le \beta \\ & q(t) = 0, \quad t \ge t_q \end{aligned}$$

where $p = \{p(k)\}$, $p_F = \{p_F(k)\}$ and $q = \{q(k)\}$ are the pulse responses of P, P_F and Q respectively, and t_q and β are design constraints for shaping the controller. The problem is solved for the following parameters: $t_a = 500$, $t_q = 500$, $\theta = 0.1$, $\alpha = 100$, and $\beta = 100$. Table I shows the outcome of the controller synthesis iterative procedure. From the table we see that at each iteration we improved the performance and reduced the impact of the worst d until no further improvement is feasible. Figures 5 and 6 show the results of the first and last iteration of the controller design process. Figures 5a, 6b plot the computed worst attack d and it's impact on the variables z, y and u. While Figures 5b, 6a plot the optimized controller parameter impulse response q, and the effect of the previous d on the variables z, y and u governed by the new controller (i.e. $\Phi_{zd}(K_{i+1})d_i$ and $\Phi_{\psi d}(K_{i+1})d_i$). We note that for iteration 1 although $\|\Phi_{zd}(K_1)d_1\|_\infty = \|\Phi_{zd}(K_2)d_1\|_\infty$, however d_1 is no longer optimal for the next iteration because $\|\Phi_{\psi d}(K_2)d_1\|_\infty \ge \theta$ as seen in Figure 5b.

VII. Conclusions

We considered the problem of computing worst case bounded stealthy false data injection attacks for LTI systems. We considered different attack resource constraints and stealthiness intervals. This problem involves a maximization of a convex function subject to convex constraints, and it was shown that it can be cast as a series of LP problems under ℓ_∞ framework. A search algorithm is constructed to solve the set of LPs (not provided here for space limitations) and was used to compute the worst stealthy attacks on AVR systems. Furthermore, we provided an iterative controller synthesis procedure that alternates between computing worst attacks and designing optimal controllers that enhance performance and minimize the impact of worst attacks. We used this method to design a controller for the AVR system that resulted in a substantial decrease in the worst impact inflicted by the worst attack.

References

[1] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *49th IEEE Conference on Decision and Control*, December 2010, pp. 5991–5998.

[2] 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.

[3] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *2015 American Control Conference (ACC)*, July 2015, pp. 2439–2444.

[4] S. Z. Yong, M. Q. Foo, and E. Frazzoli, "Robust and resilient estimation for cyber-physical systems under adversarial attacks," in *American Control Conference*, July 2016, pp. 308–315.

[5] M. Naghnaeian, N. Hirzallah, and P. G. Voulgaris, "Dual rate control for security in cyber-physical systems," in *2015 54th IEEE Conference on Decision and Control*, December 2015, pp. 1415–1420.

[6] 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.

[7] 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.

[8] R. Zhang and P. Venkitasubramaniam, "Stealthy control signal attacks in vector lqg systems," in *2016 American Control Conference*, July 2016, pp. 1179–1184.

[9] G. Wu and J. Sun, "Optimal data integrity attack on actuators in cyber-physical systems," in *2016 American Control Conference (ACC)*, July 2016, pp. 1160–1164.

[10] C. Kwon, W. Liu, and I. Hwang, "Analysis and design of stealthy cyber attacks on unmanned aerial systems," *Journal of Aerospace Information Systems*, vol. 11, no. 8, pp. 525–539, August 2014.

[11] A. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: The output-to-output l2-gain," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 2582–2587.

[12] D. Umsonst, H. Sandberg, and A. A. Crdenas, "Security analysis of control system anomaly detectors," in *2017 American Control Conference (ACC)*, May 2017, pp. 5500–5506.

[13] S. D. Bopardikar, A. Speranzon, and J. P. Hespanha, "An h-infinity approach to stealth-resilient control design," in *2016 Resilience Week (RWS)*, Aug 2016, pp. 56–61.

[14] P. Antsaklis and A. N. Michel, *Linear Systems*. McGraw-Hill, 2006.

[15] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice-Hall, 1995.

[16] M. A. Dahleh and I. Diaz-Bobillo, *Control of Uncertain Systems: A Linear Programming Approach*. Prentice-Hall, 1995.

[17] H. Saadat, *Power system analysis*. McGraw-Hill, 2009.