Optimization-based attack against control systems with CUSUM-based anomaly detection

Gabriele Gualandi
Mälardalen University, Sweden
gabriele.gualandi@mdu.se

Martina Maggio
Saarland University, Germany
maggio@cs.uni-saarland.de

Alessandro Vittorio Papadopoulos
Mälardalen University, Sweden
alessandro.papadopoulos@mdh.se

***Abstract*—Security attacks on sensor data can deceive a control system and force the physical plant to reach an unwanted and potentially dangerous state. Therefore, attack detection mechanisms are employed in cyber-physical control systems to detect ongoing attacks, the most prominent one being a threshold-based anomaly detection method called CUSUM. Literature defines the maximum impact of stealth attacks on a plant as the maximum deviation in the plant's state an undetectable attack can introduce and formulates it as the solution to an optimization problem. This paper proposes an optimization-based attack with different saturation models, and it investigates how the attack duration significantly affects the impact of the attack on the state of the control system. We show that more dangerous attacks can be discovered when allowing saturation of the control system actuators. The proposed approach is compared with the geometric attack, showing how longer attack durations can lead to a greater impact of the attack while keeping the attack stealthy.**

I. Introduction

Cyber-physical systems realize critical infrastructure control (e.g., electric power, water resources), distributed robotics (telepresence, telemedicine, automated manufacturing), healthcare systems, assisted living, environmental control, traffic control, advanced automotive systems, unmanned vehicles, and more. Infamous cyber-attacks on nuclear centrifuges (Stuxnet in 2010) and power networks (Havex/Dragonfly in 2014) raised awareness on the vulnerability of control systems, and the risks for society. According to Kaspersky Lab, 41.2% of factories were attacked by malicious software at least once in the first half of 2018. Hence, many different companies tried to secure their industrial control systems [1], [2].

This led to the prominence of *intrusion detection* methods for cyber-physical systems [3], [4]. The detection process is usually based on the concept of anomaly, and it involves some form of verification through a model. A way to detect anomalies is to verify the controller calculations using redundancy, e.g., [5]. However, if sensors or actuators are attacked, redundant execution is not enough to guarantee detection, because the compromised component is not the execution of the control software. To detect anomalies at the interfaces (sensors and actuators), there is a need for *physics-based detection* [6]. The anomaly detection strategy, in this case, usually consists of comparing the signals produced and consumed by sensors and actuators with a physical system model, i.e., a digital twin that mimics the behavior of the physical process. In control terms, this component is often realized with a state estimator, that determines the expected state of the system and compares it with the state that is apparent via measurements. The difference between the signals of physical plant and those of the digital twin are called *residuals*. Due to the existence of unknown disturbances, model uncertainties, and numeric approximation, the detection process is necessarily probabilistic, and at run-time, a *threshold* is employed to discriminate the null hypothesis (absence of attacks) from the presence of an attack in progress.

Many researches on cyber-physical security use a *stateless* anomaly detector based on the current residual e.g., [7], [8]. Other approaches investigate *stateful* approaches, which sum up the residual, therefore, considering the history of the residual signal. Usually, these systems result in a Cumulative Sum (CUSUM)-based intrusion detection method [9], [10], [11].

Attacks that take into account the presence of anomaly detection strategies are called *stealth*, and they are designed to remain undetected [12]. This work focuses precisely on stealth sensor attacks or stealth attacks that consist of injecting false measurements in sensors. Many stealth attack functions have been experimented with [12]. However, the literature usually does not take into account implementation elements like the saturation of the control signal (i.e., physical limitations that prevent the control signal from exceeding maximum values or being below prescribed minimum values).

We devise a new optimisation-based method to deviate as much as possible the state of the control plant from its nominal value. An optimisation problem is solved once, to produce a vector of sensor attack values that can be injected into the system at every time step. We show that the way actuators' saturations are modeled can produce significantly different results.

Related work: The literature explored simulated stealth attacks, based on specific attack functions and empirically evaluated the minimum time required to cause damage [13] or the maximum impact that can be caused [12]. The takeaway message is that an exponentially-shaped attack on the sensors, also called *geometric attack*, is the most

effective attack function because it is capable of introducing the most significant deviation from the nominal plant conditions. For example, stealth-by-design geometric attacks have been applied to a robot [14]. In the works mentioned above, there are two main limitations. First, the definition of stealthiness for an attack is limited to physical auto-regressive models, i.e., control systems that do not employ a state estimator. Second, considering only specific attack functions like the geometric attack may hinder more dangerous unknown attack functions.

Recent works like Urbina et al. [15], Umsonst et al. [11], and Teixeira et al. [7] avoid considering specific attack functions and solve an optimization problem to obtain the "worst case" stealth attack, i.e., the one delivering maximum distance between the state and the setpoint (also called *maximal impact*). The works mentioned above have two major limitations. First, the proposed optimization problem does not take actuators' saturations into account. Secondly, the attack duration is not considered a parameter for an attack.

In this work, we show that one can discover more dangerous attacks by using different actuators' saturation models. Moreover, we show that the duration of an attack is crucial to determine more dangerous attacks. The attack we propose addresses control systems with and without state estimators and correctly handles control signal saturation.

The first work that took into account the limits of the actuator is [16], where the authors propose to limit the saturation thresholds to ensure that no attack can reach a harmful state. While this is in principle a viable defense against attacks, in our opinion the actuators' limits are imposed mostly by physical requirements (it is not possible to further increase the flow rate of a liquid through a pipe when the valves are already completely open), and we consider their minimum and maximum values as a parameter of the problem.

Contribution: This paper provides the following contributions: (i) it shows that considering actuators' saturations is necessary for the definition of a meaningful attack, (ii) it proposes an attack based on solving an optimization problem that explicitly includes actuators limits, and it compares two different ways of expressing the actuators' saturations.

Outline: In Section II we present the model of the control and the detection system, together with our assumptions. In Section III we present our attack and in Section IV we evaluate its effectiveness. Finally, in Section V we conclude the paper.

II. Model and assumptions

A. Closed-loop: plant and controller

Consider a discrete-time feedback control system composed of a physical plant and a controller, subject to a "sensor attack" [17], [18]. The plant is sampled at prescribed times, indicated with $k \in \mathbb{Z}$. At every sampling instant k, the controller receives a measurement $\tilde{y}_k$ of the plant output y_k, and produces an actuation signal u_k, to drive the future output towards a reference value w_k, that is passed to the controller as input. The value of the control signal u_k is saturated to belong to the interval $[u_{\min}, u_{\max}]$ due to physical limitations, generating $\tilde{u}_k$. Furthermore, we denote with x_k the (internal) state of the system at time k. We assume that an attacker can tamper with the system behavior by producing an attack signal a_k that is added to the actual measurement of the output,

$$\tilde{y}_k = y_k + a_k. \tag{1}$$

According to our threat model, an attacker forges a_k to divert the output y_k of the system from its desired value w_k to reach a dangerous state.

We assume that the plant is linear, time-invariant, controllable, and observable. The discrete-time dynamic equations of the plant under attack are

$$\text{Plant} = \begin{cases} x_{k+1} &= A\,x_k + B\,\tilde{u}_k \\ y_k &= C\,x_k + D\,\tilde{u}_k \end{cases} \tag{2}$$

where $A \in \mathbb{R}^n$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, n is the number of states, m is the number of inputs, and p is the number of outputs.

The controller is a Linear Quadratic Regulator (LQR) that is designed to stabilize the system, it employs a Luenberger observer to produce an estimate $\hat{x}_k$ of the state of the plant at time k, and a state-feedback control law K to drive the system toward w_k. Specifically, the controller equations are

$$\text{Controller} = \begin{cases} \hat{x}_{k+1} &= A\,\hat{x}_k + B\,\tilde{u}_k + L\,(\tilde{y}_k - \hat{y}_k) \\ \hat{y}_k &= C\,\hat{x}_k + D\,\tilde{u}_k \\ u_k &= -K\,(w_k - \hat{y}_k) \\ \tilde{u}_k &= \mathrm{sat}\,(u_k) \end{cases} \tag{3}$$

where $\hat{x}_k$ and $\hat{y}_k$ are respectively the estimated plant state and output, and u_k and $\tilde{u}_k$ are the control signals respectively produced by the controller and received by the plant, due to saturation levels. The function $\mathrm{sat}(\cdot)$, saturates the control signal within the values $[u_{\min}, u_{\max}]$.

B. Attack detection system

State-of-the-art defenses for control systems typically include an Attack Detection System (ADS) that determines whether an attack is occurring [19]. ADS usually employ models of a system that, under the null hypothesis of absence of attacks, receives the same input as the real system, and produces as output a prediction of the system output. The difference between the predicted output and the real one (under the null hypothesis) is due to disturbances like noise and model inaccuracies. If the null hypothesis is false, however, the predicted system output diverges with respect to the measured one, and the difference can be used to detect an ongoing attack. Our controller already includes an observer, that determines $\hat{y}_k$, see Eq. (3), which we can reuse for attack detection.

Stateful ADS accumulate the difference between the predicted output $\hat{y}_k$ and the measured output $\tilde{y}_k$ over multiple periods k, and fire an alert once the cumulative

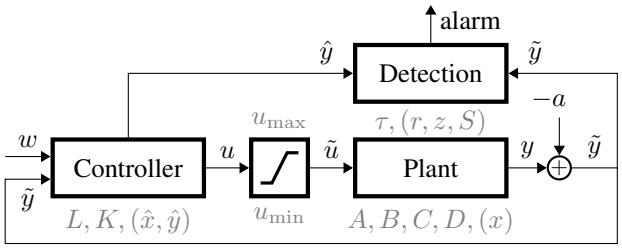

Fig. 1. Control system, under sensor attack a, protected by an attack detection system.

difference exceeds a specific threshold τ. Figure 1 shows the closed-loop and detection system. Around each block, in gray, we highlight specific block parameters and, in parenthesis, internal block states.

We consider a CUSUM-based ADS in non-parametric form [20], because this formulation can identify anomalies without any *a priori* hypothesis about a particular attack and attack function [13]. The equations for the detection mechanisms are

$$\text{Detection} = \begin{cases} r_k &= \tilde{y}_k - \hat{y}_k \\ z_k &= \|r_k\| \\ S_k &= \max\left(0, S_{k-1} + z_k - b\right) \\ S_1 &= 0 \\ \text{alarm} &= \text{logic}\left(S_k > \tau\right) \end{cases} \quad (4)$$

where r_k is the residual between the measured output $\tilde{y}_k$ and the predicted output $\hat{y}_k$. Here, z_k is a norm of the residual and S_k is the nonparametric CUSUM statistic at time k (initialized to zero).

Finally, $b \in \mathbb{R}$ is a small nonnegative constant, chosen to implement a forgetting factor. A common choice is $b = d\,\sigma$, where σ is the standard deviation of the output noise, and d is a positive constant determining an arbitrary low percentage of outliers when the attack signal $a = 0$ (e.g., $\approx 0.3\%$ when $d = 2$). For our purposes, we assume there is no output noise and hence we select $b = 0$.

At run-time, an *alarm* is fired if $S_k > \tau$. When this happens, S_{k+1} is set to 0 and the detection mechanism is reinitialized.

C. State-of-the-art attacks

Cardenas et al. [13] define various (sensor) attack functions. A *surge attack* is a constant signal, a *bias attack* is a constantly increasing signal, and a *geometric attack* is an exponentially increasing signal. Literature studies [13], [21], [14], [22] report that the geometric attack is one of the most effective, i.e., the one that provokes the maximum state deviation. As a consequence, we compare our proposed attack against geometric attacks. A geometric attack consists of defining

$$a_k = -\beta\,\alpha^{1+h-k},\ k \in (1, 2, \ldots, h) \quad (5)$$

where $\alpha \in \mathbb{R}, 0 < \alpha < 1$, determines the shape of an exponential function, $\beta \in \mathbb{R}^+$ is a positive scaling factor, and h is the duration of the attack. In the literature, the geometric attack is performed on systems that do not use a state estimator [13], and hence can be defined as stealthy by construction. In this case, the attack consumes exactly τ, independently of the values selected for α, β, and h. However, this is not the case when a state estimator is included in the controller. To properly compare our proposed attack (Section III) with the best geometric attack, we enforce stealthiness as a constraint in an optimization problem and determine the optimal values for α and β (and various values of h among which the optimal one) with respect to the deviation provoked to the plant's state (see Section IV).

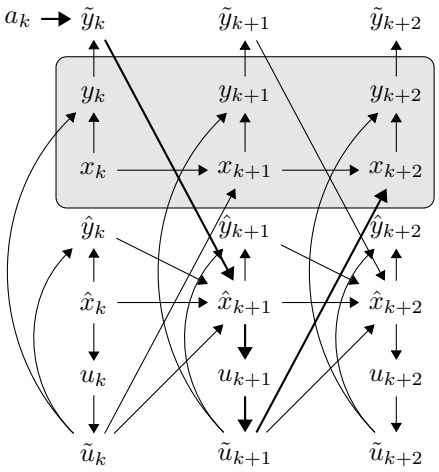

Fig. 2. Variables influence unfolding two steps dynamics. The gray background indicates purely physical quantities.

Works like [7] search for the maximum deviation from a stealth attack through an optimization problem without restricting to a particular function (like the geometric attack). However, no works include an explicit model of the actuators' saturation in the optimization problem.

To the best of our knowledge, the literature selects the attack duration of an attack h using thumb-rule criteria.

III. Optimization-based attack

In order to study the security of a control system, we want to assess if at least one dangerous state is reachable by a stealth attack, starting from the steady-state condition. The state deviation is a map $\mathcal{D} : \mathbb{R}^{n\times 1} \to \mathbb{R}$ that expresses how much the state x deviates from 0, defined as:

$$\mathcal{D}\left(x\right) = \|x\|_2^2. \quad (6)$$

We aim at performing a stealth attack, and ensuring that the system state x reaches a deviation as large as possible from the origin, i.e., the attack that maximizes Eq. (6), while remaining undetected.

A. Optimization problems

Figure 2 shows in a graph the influences of an attack a_k for two steps of the dynamic evolution of the plant and controller from Eqs. (2) and (3). For generality, we set up the optimization problem for h steps (i.e., for the entire attack duration).

We aim to find the *maximal impact* from a stealth attack to sensors. Therefore we search for the attack with duration h maximizing the state deviation with respect to the setpoint. We assume that the attacker has prior knowledge of the system matrices (A, B, C, D), the controller parameters (L, K) and the ADS threshold (τ). Moreover, we assume that when the attack begins the state is on the setpoint. The problem is formulated as:

$$\underset{(a_k,\dots,a_{k+h})}{\text{maximize}} \quad \mathcal{D}\left(x_{k+h}\right) \tag{7a}$$

subjected to:

$$x_{i+1} = A\,x_i + B\,\tilde{u}_i \qquad i = k,\dots,k+h-1 \tag{7b}$$

$$y_i = C\,x_i + D\,\tilde{u}_i, \qquad i = k,\dots,k+h \tag{7c}$$

$$\tilde{y}_i = y_i + a_i, \qquad i = k,\dots,k+h \tag{7d}$$

$$\hat{x}_{i+1} = A\,\hat{x}_i + C\,\tilde{u}_i + L\left(r_i\right), \quad i = k,\dots,k+h-1 \tag{7e}$$

$$\hat{y}_i = C\,\hat{x}_i + D\,\tilde{u}_i, \qquad i = k,\dots,k+h \tag{7f}$$

$$u_i = -K\,\hat{x}_i, \qquad i = k,\dots,k+h \tag{7g}$$

$$\tilde{u}_i = \mathrm{sat}(u_i), \qquad i = k,\dots,k+h \tag{7h}$$

$$r_i = \tilde{y}_i - \hat{y}_i, \qquad i = k,\dots,k+h \tag{7i}$$

$$S_i = S_{i-1} + \|r_i\|, \qquad i = k+1,\dots,k+h \tag{7j}$$

$$S_k = 0 \tag{7k}$$

$$S_{k+h} = \tau \tag{7l}$$

where (7b)–(7c) are the plant equations, (7d) models the sensor attack, (7e)–(7g) are the controller equations, (7h) models the saturated control signal, and finally (7i)–(7l) model the CUSUM dynamics, setting that the attack should consume the whole available CUSUM, while remaining stealthy. The solution of the optimization problem is the optimal attack sequence $[a_k,\dots a_{k+h}]$ of duration h, which provides maximum final state deviation, see Eq. (10).

Eq. (7h) models the saturation of the system from Eq. (3), but depending on how it is modeled, the resulting attack can be more or less effective. In the following we discuss two different ways to model the saturation in the optimization problem:

1) *Overflow-prevent* (Opt-P) constraint:

$$\begin{aligned} &u_{\min} \le \tilde{u}_i \le u_{\max}, \quad i = k,\dots,k+h \\ &\tilde{u}_i = u_i, \qquad\qquad\quad i = k,\dots,k+h \end{aligned} \tag{8}$$

The constraint imposes that the attack vector $(a_k,\dots a_{k+h})$ by construction ensures that u does not exceed the saturation values (it prevents overflowing the saturations). Such an approach is similar to the way saturations are typically included in optimal control problems, e.g., in Model Predictive Control (MPC).

2) *Overflow-allow* (Opt-A) constraint:

$$\tilde{u}_k = \max\left(\min\left(u_k, u_{\max}\right), u_{\min}\right), \quad i = k,\dots,k+h \tag{9}$$

The constraint does not impose any constraint on u, but it clamps $\tilde{u}$, i.e., the attack vector $(a_k,\dots a_{k+h})$ may produce u which exceeds the actuators' saturation values $u_{\max}$ and $u_{\min}$, but only their saturated value is applied to the plant (it allows overflowing the saturations).

For the geometric attack we solve the following optimization problem (G-A):

$$\underset{\alpha,\beta}{\text{maximize}}\ \mathcal{D}\left(x_{k+h}\right) \tag{10}$$

subjected to constraints (7b)–(7l), where (7h) is implemented as in Equation (9), and the constraint:

$$a_k = -\beta\,\alpha^{1+h-k}, \ \ i = k,\dots,k+h-1$$

IV. Experimental results

This Section compares the numerical solutions for the maximal impact when overflow-prevent and overflow-allow constraints are employed. More specifically, we compare the impact from Opt-A, Opt-P and the geometric attack (G-A). Results show that the most dangerous attack is discovered with Opt-A, and that both Opt-A and Opt-P are more dangerous than G-A for the considered system.

The simulated plant (in continuous time) is the mass-spring-damper system,

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\frac{K}{M}\,x_1(t) - \frac{D}{M}\,x_2(t) + \frac{1}{M}\,u(t) \\ y(t) = x_1(t), \end{cases} \tag{11}$$

with mass $M = 1\,\mathrm{kg}$, elastic constant $K = 1\,\mathrm{N/m}$, and damping $D = 0.01\,\mathrm{Ns/m}$. The model is discretized trough zero-order hold with sampling period $T_s = 0.05\,\mathrm{s}$. The actuator limits are $u_{\max} = 1$, $u_{\min} = -0.5$. The controller's parameters obtained with a LQR optimization are $K = [2.015,\ 3.52]$, and $L = [0.49,\ 1.06]^\top$.

For the ADS we set $\tau = 10$. We employ YALMIP and Gurobi to solve the optimization problems defined in Section III-A. For G-A we optimize Eq. (III-A) using the Matlab function `fmincon`. As we are dealing with time-invariant systems, we assume without the loss of generality that an attack starts at $k = 0$, i.e., the last sample of an attack is when $k = h$.

Effects of the attack duration.

Figure 3 shows on the top diagram the maximal impact as the attack duration h varies, for the two optimization problems Opt-A and Opt-P, and G-A. Each point in the figures is the impact of the optimal attack sequence, given h. For both Opt-A and Opt-P we iteratively increase h of a fixed quantity $h_\Delta = 50$, and we use as a stop criteria for h the condition $[a_0 \cdots a_{h_\Delta}] \approx 0$, because this means that the attack becomes significant only after $k = h_\Delta$ (i.e., the attack is delayed). We noticed that further increasing h results in extremely long computation time without significantly improving the solutions. The bottom diagram of Figure 3 shows the computation time for Opt-A and Opt-P (time to optimize G-A is negligible). We enforce

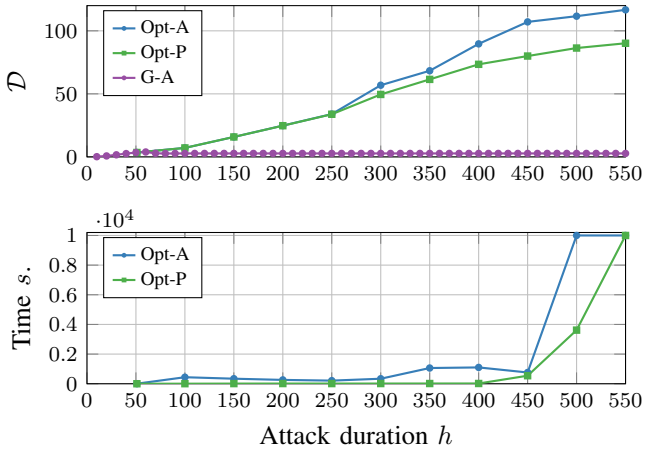

Fig. 3. (top) Solutions for the optimization problems in attack durations h; (bottom) Time to found a solution on a 12 thread I7-8700K CPU.

a maximum computation time of 10^4 seconds for each optimization problem, which is sufficient to render true $[a_0 \cdots a_{h_\triangle}] \approx 0$.

The results show that both Opt-A and Opt-P have a monotonically increasing impact with the attack duration, while G-A has a maximum when $h = 60$. The G-A has been considered as one of the most effective attacks in the literature [13], [21], [14], [22], and for a short duration of attack ($h < 60$) its maximum impact is the same from Opt-A and Opt-P. However, for longer attack duration, the impact of G-A is orders of magnitude smaller than the one from Opt-A and Opt-P. In particular, Opt-A provides the overall greater maximal when $h = 550$. On the other hand, Opt-P always performs worse than (or at most equal to) the Opt-A. Therefore Opt-A is the preferable attack. In our simulations, a difference in the maximal impact between Opt-A and Opt-P manifests only if the attack duration is long enough (i.e., $h = 300$).

Notice that Opt-A can outperform the Opt-P approach, as we are considering a sensor attack (1), that can influence the observer dynamics directly (3). Decoupling the constraint of $\tilde{u}$ from u means that even $\hat{x}$ and $\hat{y}$ are not constrained, yielding a higher degree of freedom to change $\hat{y}$ in a way that does not increase the CUSUM state, while keep attacking the system.

Solutions from Opt-A and Opt-P.

Figure 4 shows the comparison of the results of two optimization problems Opt-A and Opt-P with the highest maximal impact (i.e., $h = 550$, corresponding to 27.5 seconds). Both Opt-A and Opt-P produces a quasi-periodic signal on $\tilde{u}$ having the same frequency of the resonance frequency of the plant, i.e., 1.91 rad/s.

Figures 5 and 6 details the signals respectively from Opt-A and Opt-P. The graphs of the control signal u and $\tilde{u}$ highlight how the choice of the saturation model of Opt-A allows creating a value of u that manages to keep $\tilde{y}$ and $\hat{y}$ close to each other, without significantly increasing

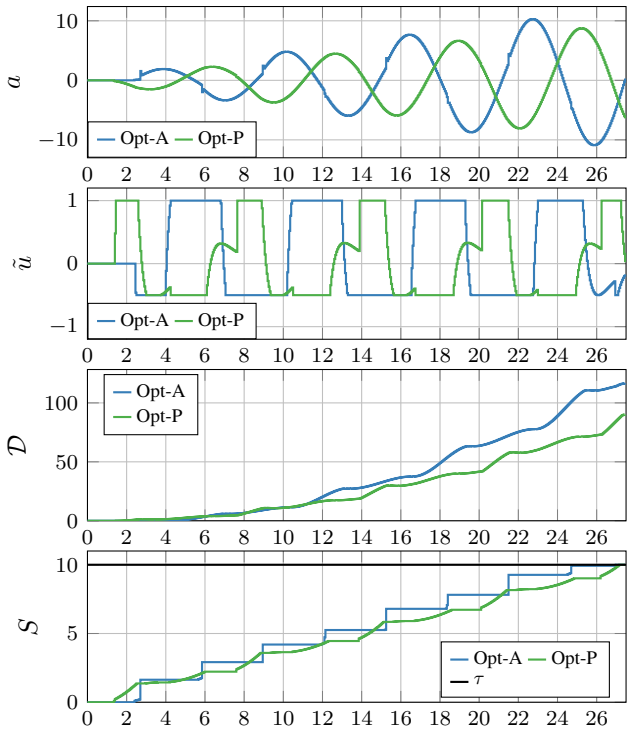

Fig. 4. Comparison of Opt-A and Opt-P when $h = 550$, with the mass-spring-damper system.

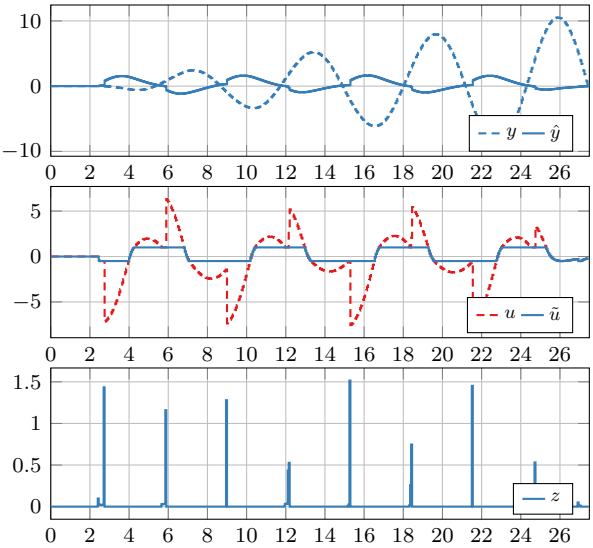

Fig. 5. Optimal solution for the Opt-A problem ($h = 550$), with the mass-spring-damper system.

the CUSUM S. However, Opt-A produces instantaneous increases of S (i.e., z) that are an order of magnitude greater than Opt-P. Therefore, a *stateless* detector could detect more easily an attack from Opt-A (e.g., in the considered system, a stateless threshold of 1 detects the attack of Opt-A at time $2.3s$, while it cannot detect the attack of Opt-P).

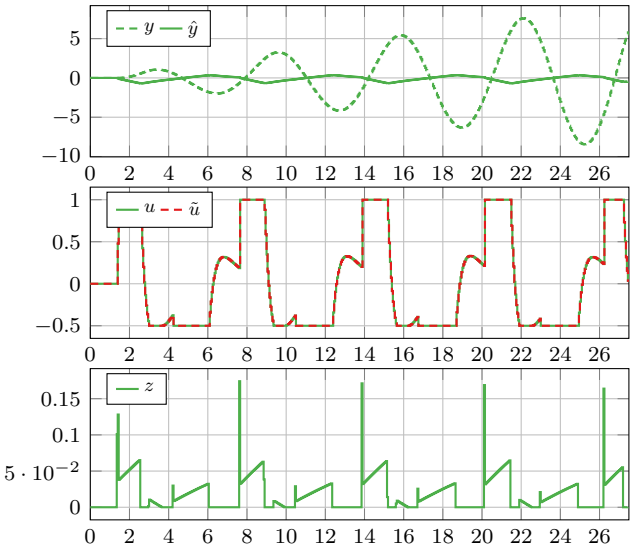

Fig. 6. Optimal solution for the Opt-P problem ($h = 550$), with the mass-spring-damper system.

V. Conclusion

In this paper, we present a novel formulation for a maximum impact attack on a cyber-physical control system. We show that parameters like attack duration and the choice of how to implement actuator saturation are crucial variables for determining the maximum impact of a stealth sensor attack. In particular, we compare two optimization-based approaches implementing the saturation function. The optima could be found using trivial stop criteria, i.e., increasing the attack duration and stopping when no further improvement is experienced.

We found out that allowing actuators' saturation (overflow-allow constraints and the corresponding Opt-A problem) results in more dangerous attacks, as shown for a mass-spring-damper system. The second way to implement actuator saturation prevents attack signals to induce corresponding actuator commands that exceed the saturation limits (overflow-prevent constraints and the corresponding Opt-P problem). While Opt-P results in less dangerous attacks, it produces attack sequences that could be harder to detect using a *stateless* anomaly detection system. The takeway messages is that security assessment of a control system should be performed using the Opt-A model, while the overall security can benefit by combining a stateful ADS (like the CUSUM) with a stateless ADS.

References

[1] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—a survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802–1831, 2017.

[2] N. Tuptuk and S. Hailes, "Security of smart manufacturing systems," *Journal of manufacturing systems*, vol. 47, pp. 93–106, 2018.

[3] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, 2019.

[4] Y. Zacchia Lun, A. D'Innocenzo, F. Smarra, I. Malavolta, and M. D. Di Benedetto, "State of the art of cyber-physical systems security: An automatic control perspective," *Journal of Systems and Software*, vol. 149, pp. 174–216, 2019.

[5] E. Casalicchio and G. Gualandi, "Asimov: A self-protecting control application for the smart factory," *Future Generation Computer Systems*, vol. 115, pp. 213–235, 2020.

[6] T. K. Das, S. Adepu, and J. Zhou, "Anomaly detection in industrial control systems using logical analysis of data," *Computers & Security*, vol. 96, p. 101935, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404820302121

[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] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 24–45, 2015.

[9] M. Basseville and I. V. N. Nikiforov, *Detection of abrupt changes: theory and application*. prentice Hall Englewood Cliffs, 1993, vol. 104.

[10] G. Carl, G. Kesidis, R. R. Brooks, and S. Rai, "Denial-of-service attack-detection techniques," *IEEE Internet computing*, vol. 10, no. 1, pp. 82–89, 2006.

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

[12] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Quantifying cyber-security for networked control systems," in *Control of cyber-physical systems*. Springer, 2013, pp. 123–142.

[13] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *ACM Symp. on Information, Computer and Comm. Security*, 2011, pp. 355–366.

[14] G. Sabaliauskaite, G. S. Ng, J. Ruths, and A. Mathur, "A comprehensive approach, and a case study, for conducting attack detection experiments in cyber–physical systems," *Robotics and Autonomous Systems*, vol. 98, pp. 174–191, 2017.

[15] D. I. Urbina, J. A. Giraldo, A. A. Cardenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, and H. Sandberg, "Limiting the impact of stealthy attacks on industrial control systems," in *ACM SIGSAC Conf. on Computer and Comm. Security*, 2016, pp. 1092–1105.

[16] J. Giraldo, S. H. Kafash, J. Ruths, and A. A. Cardenas, "Daria: Designing actuators to resist arbitrary attacks against cyber-physical systems," in *IEEE European Symp. on Security and Privacy (IEEE Euro S&P)*, 2020.

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

[18] L. F. Cómbita, A. A. Cárdenas, and N. Quijano, "Mitigation of sensor attacks on legacy industrial control systems," in *IEEE Colombian Conf. on Automatic Control (CCAC)*, 2017, pp. 1–6.

[19] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–36, 2018.

[20] H. Wang, D. Zhang, and K. G. Shin, "Detecting syn flooding attacks," in *IEEE Int. Conf. on Computer Comm.*, vol. 3, 2002, pp. 1530–1539.

[21] G. Sabaliauskaite, G. S. Ng, J. Ruths, and A. Mathur, "Experimental evaluation of stealthy attack detection in a robot," in *IEEE Pacific Rim Int. Symp. on Dependable Computing (PRDC)*, 2015, pp. 70–79.

[22] S. Chen, Z. Wu, and P. D. Christofides, "Cyber-attack detection and resilient operation of nonlinear processes under economic model predictive control," *Computers & Chemical Engineering*, vol. 136, p. 106806, 2020.