Work in Progress: Emerging From Shadows: Optimal Hidden Actuator Attack to Cyber-Physical Systems

Md Kausar Hamid Miji[†], Mengyu Liu[‡], Francis E Akowuah[†], Fanxin Kong[‡]
[†]*Department of Electrical Engineering and Computer Science, The South Dakota School of Mines & Technology, Rapid City, SD*
[‡]*Department of Computer Science and Engineering, University of Notre Dame, South Bend, IN*
md.miji@mines.sdsmt.edu, mliu9@nd.edu, francis.akowuah@sdsmt.edu, fkong@nd.edu

***Abstract*—Industries are embracing information technology and constructing more robust machines known as Cyber-Physical Systems(CPS) to automate processes. CPSs are envisioned to be pervasive, coordinating, and integrating computation, sensing, actuation, and physical processes. CPSs have various applications in life-critical scenarios, where their performance and reliability can have direct impacts on human safety and well-being. However, CPSs are vulnerable to malicious attacks, and researchers have developed detectors to identify such attacks in different contexts. Surprisingly, little work has been done to detect attacks on the actuators of CPS. Furthermore, actuators face a high risk of optimal hidden attacks designed by powerful attackers, which can push them into an unsafe state without detection. To the best of our knowledge, no such attacks on actuators have been developed yet. In this paper, we design an optimal hidden attack for actuators and evaluate its effectiveness. First, we develop a mathematical model for actuators and then create a linear program for convex optimization. Second, we solve the optimization problem and simulate the optimal attack.**

***Index Terms*—Cyber-physical Systems, Stealthy Attack, Optimal Hidden Attack.**

I. Introduction

In the 4th Industrial Revolution, cyber-physical systems (CPS) enable effective and efficient task performance by integrating sensing, actuation, computation, and networking. This integration reduces risks in manufacturing and finished products. For example, the automotive industry's CPS advancements detect obstacles for forward collision, lane, blind-spot detection, and parking assistance. Many other realms are advancing rapidly with the integration of CPS including energy, healthcare, defense, and smart cities.

Many widely used sensors in the automotive industries, like MEMS, ultrasonic, LiDARs, cameras, radars, etc. are highly vulnerable to physical invariant-based attacks. These attacks are also referred to as transduction attacks, in which the attacker focuses on altering how sensors capture real-world data. This involves injecting false sensor readings or even manipulating the physical environment around the sensor to produce a deceptive actuation action [1]–[4]. In [5], authors devised a context-aware attack called the frustum attack and demonstrated its stealthiness. By formulating the attack generation as an optimization problem, two attack scenarios that could potentially compromise road safety and mobility were constructed and evaluated by the authors in [6]. Additionally, researchers in [7] explored the adverse impact of injecting out-of-band acoustic signals into MEMS inertial sensors. They formulated non-invasive attacks, manipulated the sensor output, and used the derived inertial information to deceive control systems. Authors in [8] conducted a similar attack, targeting the drones' gyroscope and disrupting it with intentional sound noise. Authors in [9]–[11] illustrated spoofing, jamming, and acoustic cancellation attacks successfully on ultrasonic sensors. In most cases, the attack detection models compare the sensor measurement with the predicted value and test the residual through stateless methods, like Chi-Square, or stateful ones, such as cumulative sum (CUSUM) [12], [13].

While the detectors mentioned above demonstrate promising detection capabilities, hidden attacks can still bypass them and remain stealthy. These concealed attacks are intentionally formulated by malicious actors who possess full knowledge of the system and the deployed detectors [13]–[15].

In recent studies referenced in [15]–[17], researchers avoided specific attack functions. Instead, these works formulate an optimization problem, with its solution representing the ”worst-case” stealth attack. Authors in [18] considered optimization-based attacks on sensors that reduced a state's safe distance from an unsafe zone, but their work did not consider actuators for such attack formulation.

This study demonstrates that a malicious actor can devise sophisticated attacks that pose significant threats to the system, causing maximum deviation in its state while evading detection. The CPS could be linear or non-linear, and it is possible to develop sophisticated attacks that will evade both stateful and stateless detectors. In our research, we propose an optimal hidden actuator attack that takes into account a system with an estimator and CUSUM score-based stateful attack detector, aiming to manipulate the control input after the controller. Our work includes the following major contributions:

- We define and propose a novel optimal hidden actuator attack based on the system's full knowledge including the parameters of the detector.
- We evaluate the proposed optimal hidden actuator attack on a numerical benchmark.

2642-7346/24/$31.00 ©2024 IEEE
DOI 10.1109/RTAS61025.2024.00044

The rest of the paper is organized as follows. Section II includes the definitions and basic terms we use in this work. Section III presents the attack generation methodologies. Section IV shows the experimental evaluation of our approach. The paper concludes with Section V.

II. Background

1) Attack Vector: Figure 1 illustrates the general structure of a CPS where it integrates machines and information technology where the electro-mechanical parts include controllers, actuators, physical processes, sensors, etc. CPS also includes an element management system called supervisor or configuration management, used for making configuration changes to any element in the CPS network. For any given target, the controller provides control input to the actuator, and the actuator runs the physical process. The sensor measures the output state of the physical process and feeds it back to the controller. Based on this feedback, the controller generates new control inputs to keep the plant under control. A malicious actor can compromise each element in the CPS or the communication path between two elements to forge an attack. The defender observes the whole CPS and keeps track of its states. It calculates the CUSUM score for each state and compares it with its predefined threshold to detect abnormal behavior of the CPS. In this work, we chose to work with actuator attacks.

2) System Modeling: A physical process, which is also known as the plant, is considered the CPS model in this work. This plant is controlled by a computer program or controller, which operates at every constant time, known as a control step. The desired output state of the plant is given by x_{ref} and the output state of the plant is $\tilde{x}$. We assume that the transfer function of the filter and sensor is 1, which yields $y = \tilde{x}$. The difference between these two is known as error, e, where $e = \tilde{x}_{ref} - \tilde{x}$. The controller reads the sensor measurement at the onset of each step t and computes the state estimate of the plant. These estimates are represented by the values of a set of real-valued variables $[\tilde{x} = \tilde{x}[1]_t, \tilde{x}[2]_t], \tilde{x}[3]_t, \ldots, \tilde{x}[n]_t$, where n is the number of states in the system. Subsequently, the controller computes the control input $u_t = u[1]_t, u[2]_t, u[3]_t, \ldots, u[m]_t$, with m denoting the number of actuators in the system. The actuators then act upon the control inputs to drive the system toward the specified reference or target point. For the sake of clarity in presentation, it is assumed that the plant is entirely observable to the sensors, implying that the sensors can provide all the state estimates of the plant. It is important to keep in mind that, the system encompasses three distinct types of states, i.e., (1) the actual physical states of the system, denoted by x, (2) the sensor-measured states of the system, denoted as $\tilde{x}$, which are not necessarily governed by the dynamics f, and (3) estimated states, denoted as $\hat{x}$, are calculated from preceding measurements through f. In a concise notation, the actual state, expected state, and the state estimate at time t_k are also expressed as x_k, $\hat{x_k}$, $\tilde{x_k}$, where $k \in \mathbb{N}_0$

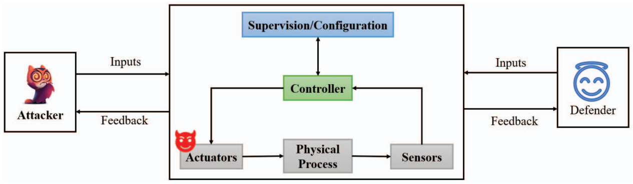

Fig. 1. Expanding the scope of the attack vector in CPS, the depicted model illustrates two competing players engaging with a shared CPS [19], [20].

III. Attack Design

1) The Plant: We utilized a state-space control design technique to develop our model, incorporating dynamic compensation by directly engaging with the state-variable representation of the system. The state equations can be represented in the state-variable form as the vector equation:

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} \tag{1}$$

$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du} \tag{2}$$

where x is the state of the system, u is the control input, and y is the output of the system respectively. Here, the state of the system is represented by the column vector $\mathbf{x}$, containing elements for an nth-order system. The matrix $\mathbf{A}$ is an $(n \times n)$ system matrix, $\mathbf{B}$ is an $(n \times 1)$ input matrix, $\mathbf{C}$ is a $(1 \times n)$ row matrix denoted as the output matrix, and $\mathbf{D}$ is a scalar known as the direct transmission term.

2) The Controller: To steer the system to the desired state, we used an optimal Linear Quadratic Regulator (LQR) controller. The infinite-horizon cost function of LQR is given by below equation:

$$J = \int_0^\infty [x^T \mathbf{Q} x + u^T \mathbf{R} u] \, dx \tag{3}$$

where, $x^T \mathbf{Q} x \geq 0, \forall x$ and $u^T \mathbf{R} u \geq 0, \forall u$. Here, $\mathbf{Q}$ and $\mathbf{R}$ denote the state cost matrix and the control cost matrix respectively. The optimal control law for the LQR controller is given by [21]:

$$u^* = K \, (x_{ref} - \tilde{x}). \tag{4}$$

A. Mathematical Model of Optimal Hidden Attack

We examine the actuator attack scenario, wherein a malicious attacker manipulates the control input after the controller transmits it to the actuator. To analyze the optimal hidden actuator attack aiming to deviate the system states further from reference states, we assume that the attacker possesses complete knowledge of:

- The system dynamics, given by $\dot{x} = f(x, u)$, induce a finite change in each dimension of the state per unit of time, expressed as $|\dot{x}| \leq \Delta\overline{x}$, where the inequality holds dimension-wise.
- The system is equipped with a state estimator that predicts the state $\hat{x}$ by forward propagation from a (not necessarily trustworthy) cached state on dynamics f.
- The system incorporates a detector $g(\tilde{x}, \hat{x})$, taking both a state estimation and a physical measurement as input. An attack is identified at time t when $g(\tilde{x}_t, \hat{x}_t) \geq 0$.

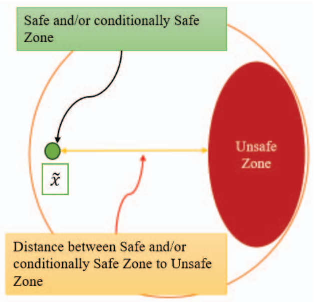
Safe and/or conditionally Safe Zone
$\tilde{x}$
Unsafe Zone
Distance between Safe and/or conditionally Safe Zone to Unsafe Zone

Fig. 2. Overview of System State Deviation

- The control signal u is confined within the interval u_{lower}, u_{upper}.

Figure 2 illustrates a simplified concept of optimal hidden actuator attack. The total state of the system consists of a safe zone, a conditionally safe zone, and an unsafe zone. The attacker aims to reduce the state of the system between the current state $\tilde{x}$ and an unsafe set shown in the red zone. As this work aims to reduce the distance between the system state and the unsafe state, we can choose a target function that measures the distance between a point and a line. For a one-dimensional system, the shortest distance d from a point (state) to the unsafe state is defined below equation:

$$distance = \frac{|(\sum d_i * (\tilde{x}_n[i])) - g|}{\sqrt{\sum (d_i{}^2)}} \tag{5}$$

where the unsafe set is $D^T\tilde{x} \leq g, D = [d_1, d_2, d_3, \ldots, d_k]$ This can be achieved with a convex optimization technique that minimizes the convex functions over convex sets. To minimize this function, we must bind it with some constraints.

The state estimator takes the measurement of the state $\tilde{x}$ and the control input $\tilde{u}$, for timestamp (n), and forecasts the state for timestamp $n+1$. The difference equations of the system and the state estimate are given by:

$$\tilde{x}_{n+1} = \mathbf{A}\tilde{x}_n + \mathbf{B}\tilde{u}_n \tag{6}$$

$$\hat{x}_{n+1} = \mathbf{A}\tilde{x}_n + \mathbf{B}u_n \tag{7}$$

where $\tilde{x}, \hat{x}, u,$ and $\tilde{u}$ represents the sensor measurement of the ground truth, state estimate, control input from the controller, and forged control input respectively. Without the appropriate system dynamics, it will be difficult to get an accurate estate estimation. Therefore, there is no guarantee that the attack will not trigger the detector.

By substituting the values of $\tilde{x}_n$ and control law, $u_n = -K(\tilde{x}_{ref} - \tilde{x}_n)$ in equation 5, we can formulate the final distance function.

Now, the attacker aims to conceal himself from the detector. So, the attack will be stealthy, and the deviated state will remain below the detector threshold τ. The detector calculates the CUSUM (cumulative sum) score for the state. For each control step, the estimator will estimate the state $(\hat{x})$ and the sensor will provide the actual state of the plant $(\tilde{x})$. The residue, r is the difference between these two states. The CUSUM (cumulative sum) statistic computes the CUSUM score S by comparing the state estimates with expected states over time. The CUSUM can be represented by the below equation:

$$S_n = \hat{x}_n - \tilde{x}_n - nd \leq \tau \tag{8}$$

where n is the number of control steps, and d is a parameter representing the drift that can avoid the increase of the CUSUM score when there is no attack.

The attacker also chooses the control input in such a way, that it remains between the control limit. Typically, a control limit is established based on the physical characteristics of the actuator. Formally,

$$\tilde{u}_{lower} \leq \tilde{u}_i \leq \tilde{u}_{upper} \tag{9}$$

Now, we have our target function and constraints and our convex optimization can be formulated as the following:

$$\begin{aligned} &\text{Minimize} && (5) \\ &\text{Subject to} && (8) \wedge (9) \end{aligned} \tag{10}$$

Here, the target function (equation 5) and all the constraints (equation 8 and 9) are linear. Therefore, this is a linear programming problem and there are many efficient solvers (Gurobi, LP_Solve, GLPK, MOSEK, CONELP, etc.) [22]. It is essential to note that, as indicated by the authors in [18], it is possible to design a real-time alert system that can more effectively defend our system against the optimal actuator attack outlined.

IV. Results

Throughout this chapter, we justify our hypothetical analysis by using one linear simulator of CPS and provide a detailed experimental result analysis.

1) Simulation Setting: The experiments were conducted on a PC with 8GB memory and an Intel(R) core(TM) i7-8056U CPU @ 1.90GHz 2.10 GHz. All the results were produced by the GLPK solver and CVXOPT library in Python.

To explore the possibility and its adverse effect on the CPS, we considered the vehicle turning equation as our plant. Authors in [23], [24] modeled the turning of a vehicle changing the speed of each wheel differently. The physical dynamics of the system are given by:

$$\dot{x} = -\frac{25}{3}x + 5u \tag{11}$$

$$\text{Output}, y = \tilde{x} \tag{12}$$

Here, $\dot{x}$ represents the change in speed difference, x signifies the speed difference between the wheels, while u denotes the control input, representing the voltage difference applied to the motors that control the two wheels. The objective is to sustain the speed difference at a reference value of 1 meter per second. Comparing equations 11, 12, 1 and 2, we find the value of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$ matrices as following: $\mathbf{A} = [-\frac{25}{3}]$, $\mathbf{B} = [5]$, $\mathbf{C} = [1]$, and $\mathbf{D} = [0]$. We used the Python control library to calculate the LQR gain for the system using the penalty matrices $\mathbf{Q} = \begin{bmatrix}1\end{bmatrix}$ and $\mathbf{R} = \begin{bmatrix}1\end{bmatrix}$.

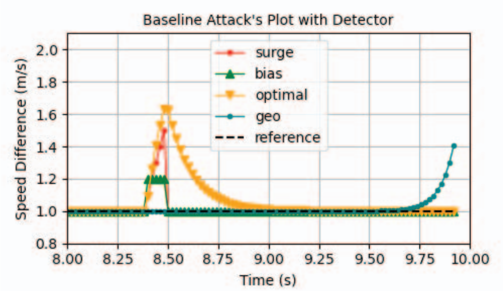

Fig. 3. Plot of the attacked state against time for Vehicle Turning

2) Evaluation: We employed CVXOPT and GurobiPy to solve the optimization problem defined in equation 10. Using the Python control library, we simulated the vehicle's turning system and analyzed the closed-loop system's response to the surge, bias, geometric, and optimal hidden attacks, presenting the results in Figure 3. We initiated all attacks at the same time (at $t = 420$), except for geometric attacks (started at $t = 0$), and observed the behavior of the state. The bias attack led to a constant state increase with minimal deviation. The surge attack caused rapid state deviation without fully pushing the system into the unsafe zone. Conversely, the optimal control input, derived from convex optimization, most significantly and quickly deviated from the reference point. After the surge attack, the state quickly realigned with the reference point, while the optimal hidden actuator attack took longer to reset, making it the most severe compared to the baseline attacks.

V. Conclusion

In this paper, we propose an optimal hidden attack on actuators, evaluate the impact of such attacks, and compare the result with baseline attacks. The experimental results show that the optimal hidden attack can deviate the system most within the shortest possible time while the attacker remains stealthy. Since this work focused on one-dimensional systems for optimal hidden actuator attacks, it will be interesting to explore such attacks for multidimensional linear and non-linear systems. In future studies, we will also investigate how various types of noise affect the state space model. Also, enthusiastic researchers can consider developing similar attacks based on stateless detectors.

Acknowledgement

This work was supported in part by NSF CNS-2333980 and SDBOR CRG award MA2300014. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the National Science Foundation (NSF) or the South Dakota Board of Reagents.

References

[1] K. Fu and W. Xu, "Risks of trusting the physics of sensors," *Communications of the ACM*, vol. 61, no. 2, pp. 20–23, 2018.

[2] I. Giechaskiel and K. Rasmussen, "Taxonomy and challenges of out-of-band signal injection attacks and defenses," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 645–670, 2019.

[3] M. Liu, L. Zhang, V. V. Phoha, and F. Kong, "Learn-to-respond: Sequence-predictive recovery from sensor attacks in cyber-physical systems," in *2023 IEEE Real-Time Systems Symposium (RTSS)*. IEEE, 2023, pp. 78–91.

[4] L. Zhang, Z. Wang, M. Liu, and F. Kong, "Adaptive window-based sensor attack detection for cyber-physical systems," in *Proceedings of the 59th ACM/IEEE design automation conference*, 2022, pp. 919–924.

[5] R. S. Hallyburton, Y. Liu, Y. Cao, Z. M. Mao, and M. Pajic, "Security analysis of {Camera-LiDAR} fusion against {Black-Box} attacks on autonomous vehicles," in *31st USENIX Security Symposium (USENIX Security 22)*, 2022, pp. 1903–1920.

[6] Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu, and Z. M. Mao, "Adversarial sensor attack on lidar-based perception in autonomous driving," in *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*, 2019.

[7] Y. Tu, Z. Lin, I. Lee, and X. Hei, "Injected and delivered: Fabricating implicit control over actuation systems by spoofing inertial sensors," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018.

[8] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim, "Rocking drones with intentional sound noise on gyroscopic sensors," in *24th USENIX Security Symposium (USENIX Security 15)*, 2015.

[9] B. S. Lim, S. L. Keoh, and V. L. Thing, "Autonomous vehicle ultrasonic sensor vulnerability and impact assessment," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE, 2018, pp. 231–236.

[10] C. Yan, W. Xu, and J. Liu, "Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle," *Def Con*, vol. 24, no. 8, p. 109, 2016.

[11] W. Xu, C. Yan, W. Jia, X. Ji, and J. Liu, "Analyzing and enhancing the security of ultrasonic sensors for autonomous vehicles," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 5015–5029, 2018.

[12] R. Quinonez, J. Giraldo, L. Salazar, E. Bauman, A. Cardenas, and Z. Lin, "Savior: Securing autonomous vehicles with robust physical invariants," in *Usenix Security*, 2020.

[13] C. Murguia and J. Ruths, "Cusum and chi-squared attack detection of compromised sensors," in *2016 IEEE Conference on Control Applications (CCA)*. IEEE, 2016, pp. 474–480.

[14] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *2013 American control conference*. IEEE, 2013, pp. 3344–3349.

[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 *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 1092–1105.

[16] 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.

[17] 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.

[18] M. Liu, L. Zhang, P. Lu, K. Sridhar, F. Kong, O. Sokolsky, and I. Lee, "Fail-safe: Securing cyber-physical systems against hidden sensor attacks," in *2022 IEEE Real-Time Systems Symposium (RTSS)*. IEEE, 2022, pp. 240–252.

[19] F. Xia, A. Vinel, R. Gao, L. Wang, and T. Qiu, "Evaluating ieee 802.15. 4 for cyber-physical systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, pp. 1–14, 2011.

[20] A. Cardenas, *The Cyber Security Body of Knowledge v1.0, 2019*. University of Bristol, 2019, ch. Cyber-Physical Systems Security, kA Version 1.0. [Online]. Available: https://www.cybok.org/

[21] R. Tedrake, "Underactuated robotics," 2023. [Online]. Available: https://underactuated.csail.mit.edu

[22] B. Meindl and M. Templ, "Analysis of commercial and free and open source solvers for linear optimization problems," *Eurostat and Statistics Netherlands within the project ESSnet on common tools and harmonised methodology for SDC in the ESS*, vol. 20, 2012.

[23] F. Kong, M. Xu, J. Weimer, O. Sokolsky, and I. Lee, "Cyber-physical system checkpointing and recovery," in *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 2018.

[24] L. Zhang, X. Chen, F. Kong, and A. A. Cardenas, "Real-time attack-recovery for cyber-physical systems using linear approximations," in *IEEE Real-Time Systems Symposium*. IEEE, 2020.