Stability-Guaranteed Defense Mechanism for Detecting Integrity Attacks on Cyber-Physical Systems

M. Kaheni, V. De Iuliis, C. Manes, and A. V. Papadopoulos

Cyber-physical systems (CPS) integrate computational algorithms with the physical world to achieve various goals, such as enhancing efficiency and performance across different sectors. The combination of computational intelligence and physical processes makes CPSs increasingly favorable in automation, particularly in complex and embedded control systems. Alongside CPSs' benefits to our lives, the necessity for communication channels to connect the cyber and physical worlds makes them vulnerable to various cyber-attacks. Attackers can disrupt or even arbitrarily manipulate the data transferred from the physical plant to the computational algorithm or the control signals computed by the algorithm that need to be communicated back to the physical plant. Several notable examples of cyber-attacks on real-world CPSs include Stuxnet [1], which targeted Iran's nuclear enrichment facilities; BlackEnergy [2], which aimed at Ukraine's power grid; and Trisis [2], which attacked a petrochemical plant in Saudi Arabia. Therefore, despite the numerous benefits CPSs offer, ensuring they are sufficiently secure against cyber-attacks before implementation is crucial.

The first step in safeguarding CPSs against cyber-attacks is *anomaly detection*. The primary objective of this step is to ensure that the system operates as expected. If the system's behavior, based on the measurement data received by the anomaly detection unit, deviates from expectations beyond a certain threshold, an alarm is triggered, indicating a potential issue. The control center can then execute predesigned scenarios to address the detected attack.

The defense mechanism we pursue in this abstract falls under the umbrella of active defenses and could be considered a Moving Target Defense (MTD). If an attacker has knowledge of the system dynamics, they can arbitrarily and stealthily perturb a system using a covert attack [3]. To diminish adversaries' knowledge about control systems, the concept of MTD has recently garnered significant interest within the research community. Initially proposed in [4] and later expanded upon in additional studies [5], [6], [7], this approach to anomaly detection involves the defender introducing dynamic changes through time-varying parameters. These parameters are known to the defender but remain concealed from the attacker. The ongoing changes create a *moving target*, preventing adaptive adversaries from effectively identifying the system. The application of MTD in practical settings, such as power grids [8], [9], transportation [10], and cloud computing [11], has yielded promising results. Therefore, further investigation into this approach's specifics and potential enhancements is highly advantageous.

To achieve dynamic movements in closed-loop systems, one can consider maintaining fixed plant dynamics while assigning variation duties to the controller, as suggested in [12]. This approach is generally more practical since the controller operates using a cyber algorithm, whereas the plant is a physical system. This abstract aims to introduce a novel attack detection method, named *Variable Control Gain Defense* (VCGD), that functions by applying a bounded arbitrary sequence $v_k \in \mathscr{V}$ to scale the control signal. This parameter remains concealed from attackers but is known to the anomaly detection unit. To summarize, the information available to the defender and the attacker at time step k, denoted as Ψ_k^D and Ψ_k^A, respectively, is as follows:

$$\begin{aligned} \Psi_k^D &= \{\mathbf{A},\mathbf{B},\mathbf{C},v_{\{0:k\}},\mathbf{u}^a_{\{0:k\}},\mathbf{y}^a_{\{0:k\}},f(\omega_k,\nu_k)\},\\ \Psi_k^A &= \{\mathbf{A},\mathbf{B},\mathbf{C},\mathscr{V},\mathbf{u}^a_{\{0:k\}},\mathbf{y}^a_{\{0:k\}},\xi_{\{0:k\}},f(\omega_k,\nu_k)\}, \end{aligned} \tag{1}$$

Here, $f(\omega_k,\nu_k)$ represents the Probability Density Function (PDF) of measurement noise and process disturbance.$\mathbf{u}^a$ is the computed control signal, which may be manipulated by the adversary in the communication channel. Similarly, the measured output, $\mathbf{y}^a$, is subject to potential adversarial manipulation. Lastly, ξ denotes an $n \times m$ vector containing the injected values introduced by the adversary into the input and output. The block diagram of our proposed setting in this abstract is depicted in Fig. 1. The main findings of this work can be divided into two categories:

Detection: We provide mathematical guarantees that any integrity attack that lays in the general setting depicted in Fig. 1, potentially can be revealed by our anomaly detection mechanism, in the sense that there exists $k \geq 0$ where $\|\Delta\mathbf{z}_k\| \neq 0$, in which $\Delta\mathbf{z}$ is the difference between residues of the attacked and benign CPS. In other words, if there exists some $k \geq 0$ such that $\|\xi_k\| \neq 0$, the residues of the attacked, and the attack-free systems cannot be the same, bringing room for detecting abnormal behaviors. Our finding can be summarized in the following Theorem.

Theorem 1: Consider a LTI CPS protected by VCGD, and subject to both input and sensor integrity attacks. Given the information available to both the defender and the attacker, as outlined in (1), the probability of the adversary successfully crafting a stealthy attack is zero. □

Stability: Since applying a time-varying gain on the control signal turns the system into a Linear Time-Varying (LTV) one, assuring stability becomes more challenging. In this work, we mathematically prove that there exists an interval around 1, within which we can randomly select the control gain at each time step while ensuring the stability

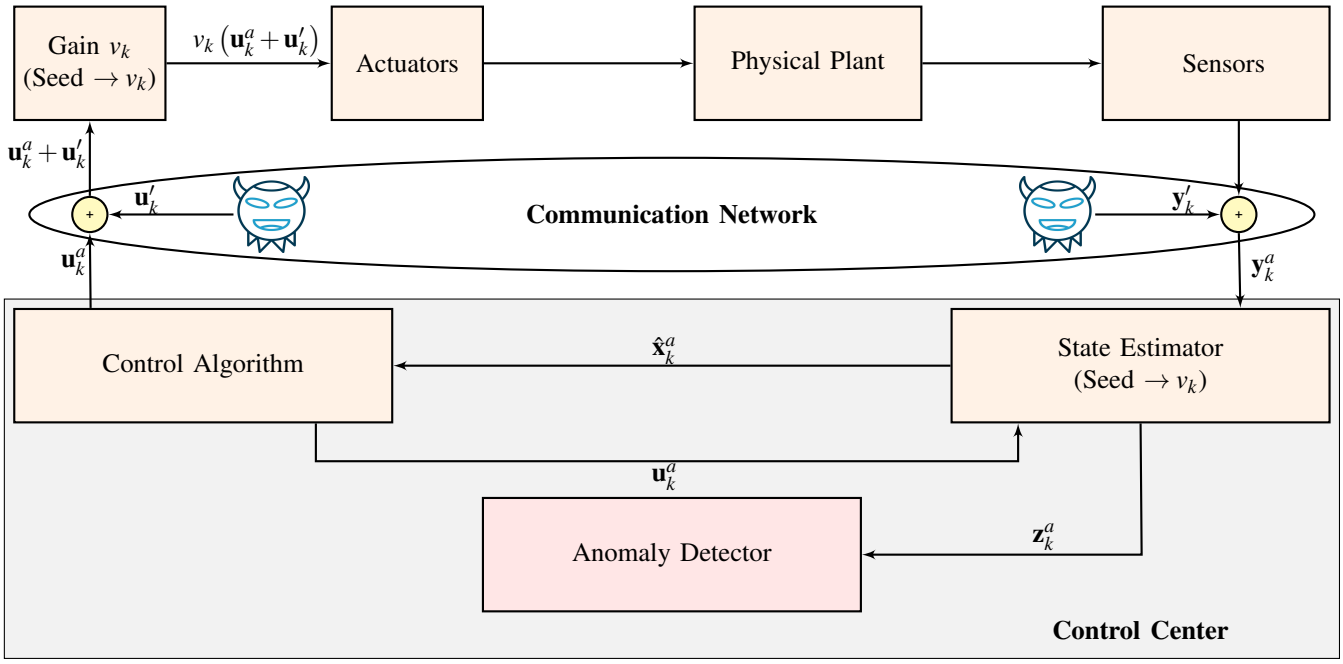

Fig. 1. Block diagram of a CPS, under both input and sensors integrity attacks, and safeguarded by VCGD.

of the system. A mathematical relation finds the upper and lower bounds of this interval in Single Input (SI) systems, while a criterion is introduced to check whether a proposed interval jeopardizes stability. Our finding can be summarized in the following Lemma and Theorem.

Lemma 1: Consider matrices $\mathbf{B}\in\mathbb{R}^{n\times m}$ and $\mathbf{K}\in\mathbb{R}^{m\times n}$, and symmetric positive definite matrices $\mathbf{Q}\in\mathbb{R}^{n\times n}$ and $\mathbf{R}\in\mathbb{R}^{m\times m}$. Then, there exists an interval $\mathscr{V}=[\underline{v},\bar{v}]$, where $0<\underline{v}<1<\bar{v}$, such that

$$\mathbf{Q}-\mathbf{K}^{\mathrm{T}}\big((v-1)^2\mathbf{B}^{\mathrm{T}}\mathbf{H}\mathbf{B}+(1-2v)\mathbf{R}\big)\mathbf{K}\geq 0,\quad \forall v\in[\underline{v},\bar{v}]. \tag{2}$$

□

Theorem 2: Consider a LTI CPS, whose pair $(\mathbf{A},\mathbf{B})$ is controllable. For given $\mathbf{Q}\in\mathbb{R}^{n\times n}$ and $\mathbf{R}\in\mathbb{R}^{m\times m}$, symmetric and positive definite matrices, and given $\beta\in(0,1)$, consider the matrix $\mathbf{K}\in\mathbb{R}^{m\times n}$ defined as

$$\mathbf{K}=(\mathbf{R}+\mathbf{B}^{\mathrm{T}}\mathbf{H}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{H}\mathbf{A}, \tag{3}$$

where $\mathbf{H}\in\mathbb{R}^{n\times n}$ is the unique solution of the modified DARE

$$\mathbf{A}^{\mathrm{T}}\mathbf{H}\mathbf{A}-\beta^2\mathbf{H}-\mathbf{A}^{\mathrm{T}}\mathbf{H}\mathbf{B}(\mathbf{R}+\mathbf{B}^{\mathrm{T}}\mathbf{H}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{H}\mathbf{A}+\mathbf{Q}=\mathbf{0}_{n\times n},$$

and consider the interval $[\underline{v},\bar{v}]$, defined in Lemma 1, such that the inequality (2) is satisfied. Then, the time-varying state feedback $\mathbf{u}_k=-v_k\mathbf{K}\mathbf{x}_k$, where v_k is any sequence taking values in the interval $[\underline{v},\bar{v}]$, is such that the origin of the closed-loop system

$$\mathbf{x}_{k+1}=(\mathbf{A}-v_k\mathbf{B}\mathbf{K})\mathbf{x}_k,\quad k=0,1,\dots \tag{4}$$

is exponentially stable, satisfying

$$\|\mathbf{x}_k\|\leq\beta^k\sqrt{\frac{h_{\max}}{h_{\min}}}\|\mathbf{x}_0\|,\quad k=0,1,\dots, \tag{5}$$

where $h_{\max}=\lambda_{\max}(\mathbf{H})$ and $h_{\min}=\lambda_{\min}(\mathbf{H})$. □

References

[1] T. M. Chen and S. Abu-Nimeh, "Lessons from stuxnet," *Computer*, vol. 44, no. 4, pp. 91–93, 2011.

[2] M. Geiger, J. Bauer, M. Masuch, and J. Franke, "An analysis of black energy 3, crashoverride, and trisis, three malware approaches targeting operational technology systems," in *Proc. 25th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA), 2020*, vol. 1, 2020, pp. 1537–1543.

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

[4] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *2015 54th IEEE Conf. on Decis. and Control (CDC)*, 2015, pp. 5820–5826.

[5] P. Griffioen, S. Weerakkody, and B. Sinopoli, "An optimal design of a moving target defense for attack detection in control systems," in *2019 Amer. Control Conf. (ACC)*, 2019, pp. 4527–4534.

[6] S. Weerakkody and B. Sinopoli, "A moving target approach for identifying malicious sensors in control systems," in *2016 54th Annu. Allerton Conf. on Commun., Control, and Comput. (Allerton)*, 2016, pp. 1149–1156.

[7] P. Griffioen, S. Weerakkody, and B. Sinopoli, "A moving target defense for securing cyber-physical systems," *IEEE Trans. Autom. Control.*, vol. 66, no. 5, pp. 2016–2031, 2021.

[8] J. Tian, R. Tan, X. Guan, Z. Xu, and T. Liu, "Moving target defense approach to detecting stuxnet-like attacks," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 291–300, 2020.

[9] A. Abdelwahab, W. Lucia, and A. Youssef, "Decoy-based moving target defense against cyber-physical attacks on smart grid," in *2020 IEEE Electr. Power Energy Conf. (EPEC)*, 2020, pp. 1–5.

[10] J. Giraldo and A. A. Cardenas, *Moving Target Defense for Attack Mitigation in Multi-Vehicle Systems*. Cham: Springer International Publishing, 2019, pp. 163–190.

[11] L. Santos, C. Brito, I. Fé, J. Carvalho, M. Torquato, E. Choi, D. Min, J.-W. Lee, T. A. Nguyen, and F. A. Silva, "Event-based moving target defense in cloud computing with vm migration: A performance modeling approach," *IEEE Access*, pp. 1–1, 2024.

[12] M. Kaheni and A. V. Papadopoulos, "Hybrid moving controller: Modified hybrid moving target defense with stability guarantees," in *2024 Eur. Control Conf. (ECC)*, 2024, pp. 1698–1703.