Online Stealthy Attack for Cyber-Physical Systems with Applications in Autonomous Driving

by

CHEN Xingzhou
Department of Electronic and Computer Engineering,
Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong
Email: xchenfk@connect.ust.hk

1 Introduction

Cyber Physical Systems (CPSs), a term [1] firstly proposed by Professor Raj Rajkumar of Carnegie Mellon University USA in 2006, represents the system that integrates the ability of computing and communications [2][3]. With the aim of monitoring or controlling the entities in the physical world [4][5], CPSs typically combine multifarious cyber and physical components, such as sensors, actuators, communication networks and computational controllers. Unlike the traditional embedded systems operating as standalone devices, most full-fledged CPSs [6][7] are designed as a network of interacting elements with physical input and output.

In the last two decades, ongoing advances in communication and computing capabilities have promoted a vigorous development of CPSs and have seen a variety of relevant applications [8]. With the advancement of wireless communications, it has become possible for CPSs to transmit real-time signals on large-scale physical networks [9]. Moreover, equipped with powerful computational resources, CPSs are capable to find a feasible control strategy within the time limit to obtain the desired levels of performance [10]. These extend the potential of CPSs in several areas, including medical devices [11], smart grids [12] and autonomous vehicles [13].

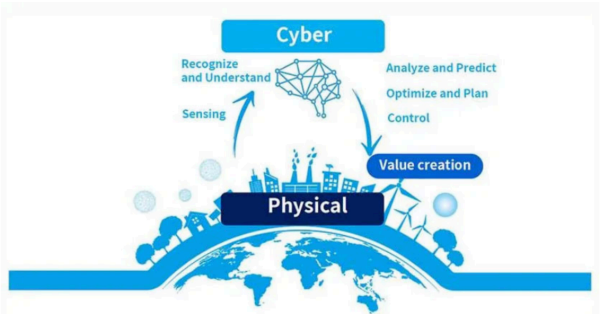

Figure 1: Cyber Physical System

However, in design and utilization of CPSs, opportunities and challenges [14][15] always go hand in hand. While previous research has mitigated the impact of lost packets [16] and time delays [17] in control networks, security has become the most concerned topic of CPSs in recent years. For many control systems utilize the unprotected networks to transmit sensor data and actuation commands, attackers can damage the vulnerable system by injecting malicious data and commands [18]. The security concerns are further motivated by several consequential attack scenarios that have occurred in real cyber physcal systems. **Stuxnet** was a severe CPS attack on an Iranian uranium enrichment plant in late 2009 [19][20]. By inserting a malware, the attacker are capable to observe the key outputs of the system and inject malicious signal into critical infrastructures, resulting in damage to amount of centrifuges. Additionally, **Maroochy attack** occurred in Australia water services in 2000 [21]. Infiltrating the SCADA network, the attacker took control of 150 sewage pumping stations and viciously discharged one million liters of untreated sewage. **RQ-170**, an USA drone, lost control in Iran [22][23]. The reason for this attack was that Iranian forces spoofed the GPS signal and deceive the UAV to land in the desired location.

The remainder of this report is organized as follows: The relevant work is presented in Section 1.1. In Section 2, the Kalman filter and Linear–Quadratic–Gaussian (LQG) control are briefly introduced as preliminary knowledge. In Section 3, some existing attack models and model-based detection mechanisms are reviewed. In Section 4, the design of attack strategy is formulated as an optimal control problem with constraints. In Section 5, an algorithm is

proposed to address this problem and demonstrated using numerical examples. Several open problems and future directions on CPS security are summarized and discussed in Section 6.

1.1 Literature review

For CPS security issues, there are several works that focus on a specific domain, for instance, the security and privacy in smart grids [12][24][25], software and hardware vulnerability in medical devices [26][11][27], and a surge of interest in autonomous vehicles [28][29][13][30]. Apart from the interest in a specific area, other researchers study the general CPSs security issues in three ways, i.e., attack strategies, under attack system performance analysis and defense countermeasures.

The attack strategies are determined by the purpose and method of the attack. On the one hand, attacks aim to degrade the CPS's control performance. Mo et al. [18] constructed a false data injection(FDI) attack to destabilize the system, and [31] defined a replay attack on a control system evading the classical detectors. Moreover, the optimal Denial-of-Service (DoS) attacks are scheduled to maximize the Linear-Quadratic-Gaussian (LQG) cost in [32] and degrade the robust control performance in [33], respectively. To induce the CPS to approach a specific target state, optimal stealthy attacks are designed in [34][35] against different type of detectors. On the another hand, how to degrade the CPS's estimate performance is also studied. Shi et al. [36] provided the optimal attack strategy to schedule DoS attacks, maximizing the expected average estimation error. Besides, an optimal linear integrity attack, which can maximize estimation error covariance, is proposed by Guo et al. [37] with a more general optimality criterion in [38]. Additionally, in most literature, attack strategies are developed on the full system knowledge and Gaussian noises model, but it is unpractical in many real scenarios and such premises greatly limit their application. To handle these difficulties, [18, 39, 40] study the attack strategies with imperfect system knowledge and observability.

Following the attack strategy design, some papers analyze the performance degradation of CPS and quantify the attack's potential impact. Mo et al. [41] studied the performance of Kalman filter under attack and provided a quantitative measure of the resilience to adversarial attacks. Besides, Mo et al. [39] formulated the attacker's action as a constrained control problem, and characterized the maximum perturbation using the concept of reachable set. Moreover, based on information theoretic analysis, the author of [40][42] worked on the performance bound and limitations induced by an ϵ-stealthy attack. Motivated by [43], where a feedback structure that enables to take over control of a system while remaining hidden, [44] illustrated the magnitude of the potential vulnerability by covert attack. In [45], it derived the necessary and sufficient condition for insecurity, which means there exist malicious stealthy attacks that can lead to unbounded estimation errors. What's more, the author of [46] provided necessary and sufficient conditions for the vulnerability, describing when the CPS can be destabilized by stealthy attacks.

Against various attack strategies, there are some existing defense mechanisms based on different analysis tools, i.e., control theory, game theory, machine learning and cryptography. Model-based detectors, which solve an optimal state estimation problem and analyze the residue sequences, are widely used in CPS, especially windows χ^2 detector [31], non-parametric CUSUM detector [47], MEWMA detector [48]. The framework of game theory also enables the design of the optimal defense policies. For instance, by using game-theoretic models, [10] yielded the system balance between resilience and robustness, [49] obtained the optimal DoS defense strategy by solving the associated Bellman equations, and [50] improved the efficiency of synthesized robust supervisors against sensor deception attacks. Furthermore, the data-based

models, which has been successfully applied to object detection problems, are presented as a natural methodology to detect attacks in [51][52][53]. Additionally, with the idea of cryptography, Mo et al. [54] proposed the watermarking approach to authenticate the correct operation of a control system, while the authors of [55] incorporated the public key encryption scheme into the design of remote controller or observers.

However, advances in research and requirements for real-world applications result in more challenges. Although stealthy attack [31][56] has been widely discussed, most ignore the duration limitation and prioritize covertness over the effect of attacks. Studying the trade-off between stealthiness and efficiency of CPS attacks is challenging. Furthermore, the optimal attack strategy from previous work, in particular [32][37], relies on random system information in the process, which makes it challenging to evaluate the specific attack effect on the system state. Thus, much room remains to study the attack policy, which has the desired impact on the system state. Moreover, it should be pointed out that current attack strategies [34][35] only consider cases, where countermeasures are known but cannot handle an unexpected issue, e.g., some attack signals are blocked accidentally. Therefore, how to design an online attack to meet complex situations has yet to be worked out. These challenges provide directions and motivations for our research.

2 Cyber Physical System Model

In the literature on CPS, e.g., [31][37], the system is often modeled as a linear discrete-time stochastic system in state-space form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_t \\ y_t &= Cx_t + v_t \end{aligned} \tag{1}$$

where the system state $x_t \in \mathbb{R}^n$, and the system input $u_t \in \mathbb{R}^m$, the sensor output $y_t \in \mathbb{R}^p$. Moreover, the process noise $\{w_t \in \mathbb{R}^n\}$ has (i.i.d) distribution $\mathbb{N}(0, \Sigma_w)$ with $\Sigma_w \succ 0$, the sensor noise $\{v_t \in \mathbb{R}^p\}$ has (i.i.d) distribution $\mathbb{N}(0, \Sigma_v)$ with $\Sigma_v \succ 0$, and w_t is independent of v_t. The pair (A, B) is controllable, and the pair (A, C) is observable.

Estimation and control performance are fundamental issues in closed-loop control system. As a linear time-invariant model with Gaussian noise, the most effective control approach is to design proper Kalman filter and linear–quadratic–Gaussian (LQG) controller, seen in Figure 2. Following the separation principle, the optimal Kalman gain K and controller gain L can be obtained by solving algebraic Riccati equations, respectively.

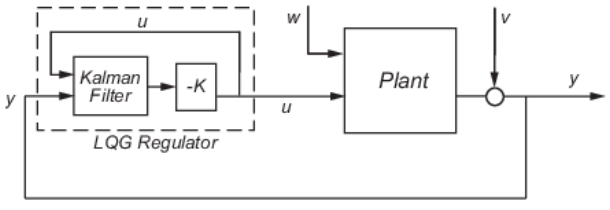

Figure 2: LQG controller with Kalman filter

2.1 Kalman filter

To compute a state estimate $\widehat{x}_t$, the CPS is equipped with a Kalman filter, which is the MMSE(minimum mean-square error) estimate of the states for Gaussian noises models [57][58]. In this case,

$$\widehat{x}_t^{MMSE} = E[x_t|x_0, y_1, ..., y_t] \tag{2}$$

If the system satisfies that (A, C) is observable and $(A, \sqrt{\Sigma_w})$ is controllable, the Kalman filter converges to a fixed gain estimator:

$$\widehat{x}_t = \widehat{x}_{t|t-1} + K(y_t - C\widehat{x}_{t|t-1}) \tag{3}$$

$$\widehat{x}_{t+1|t} = A\widehat{x}_t + Bu_t \tag{4}$$

where $P = APA^T + \Sigma_w - APC^T(CPC^T + \Sigma_v)^{-1}CPA^T$, and $K = PC^T(CPC^T + \Sigma_v)^{-1}$. Moreover, to represent the random error generated at the step t, we denote $r_t = y_t - C\widehat{x}_{t|t-1}$ as the Kalman innovation. Under nominal conditions, $\{r_t\}$ are (i.i.d) distribution $\mathbb{N}(0, \Sigma_r)$, where $\Sigma_r = CPC^T + \Sigma_v$.

2.2 LQG controller

The LQG framework considers the problem of regulating a linear dynamical system perturbed by environmental noise. For the LTI system with known Gaussian noise, an analytic solution of the optimal control policy can be derived by minimizing the cost function $\mathbb{J}$:

$$\mathbb{J} = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}(x_t^T Q x_t + u_t^R U u_t)\right] \tag{5}$$

where Q and R are positive semi-definite penalty matrices on state and control, respectively.

When (A, B) is controllable and $(A, \sqrt{Q})$ is observable, the optimal linear feedback gain converges to a time-invariant matrix L:

$$u_t = L\widehat{x}_t \tag{6}$$

where $S = A^TSA - A^TSB(B^TSB + R)^{-1}BSA + Q$, and $L = -(B^TSB + R)^{-1}B^TSA$.

Based on our assumptions, the operation of CPS utilizes the pre-designed estimation strategy and control law. In the other word, the Kalman gain K and controller gain L are constant, even under attack.

3 Attack models and Detection Method

As we mentioned in Section 1, the combination of cyber and physical components increases the vulnerability of the system. Before reviewing the attack models, we first discuss how attacks damage the system. As seen in Figure 3, when attacks happen, the attackers are capable to monitor, jam and alter the sensor data and actuation commands, thus degrading system performances via feedback control loop in CPS.

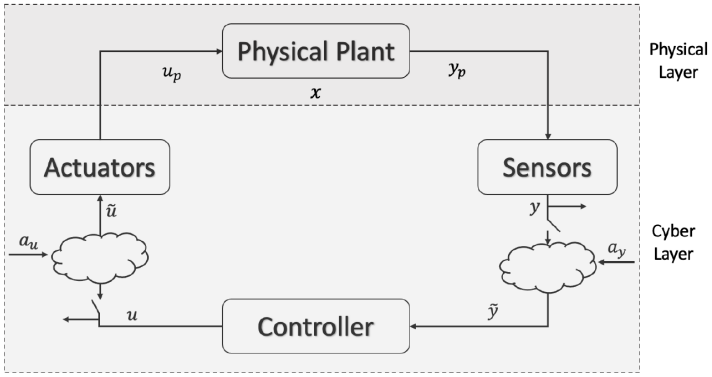

Figure 3: Architecture of general CPS attacks

3.1 Attack models

There are three commonly used attack models in CPS, named disclosure attacks, deception attacks, and disruption attacks [59][60]. Categorized by different attack methods and consequences, they violate the system's confidentiality, integrity, and availability, respectively.

Disclosure attack A disclosure attack refers to disclosing sensitive information to an unauthorized entity [61]. By monitoring the sensory data $\{y_t\}$ and control input $\{u_t\}$, attackers infer the private information of the system, such as its transfer function and current system states [62]. Although the disclosure attack, known as a passive attack, has little direct impact on the system, it provides key information for other attackers and then launch further serious attacks. For instance, replay attacks [63] are such complex attacks involving disclosure attacks and deception attacks.

Deception attack A deception attack aims to alter the sensory data and control signals, and causes misbehaving of the system [64, 65]. Spoofing attacks [66], false data injection attacks [67] and replay attacks [63] are typical deception attacks. In particular, it replaces the system output y_t by $\widetilde{y}_t$:

$$\widetilde{y}_t = y_t + y_t^a \;\; or \;\; \widetilde{y}_t = y_{t-T} \tag{7}$$

where y_t^a is an false data injected by the attacker, and T is the replay time interval. And attacks to system input u_t is similar. Deception attacks from previous work, in particular [37][63], maximize the LQG cost and state estimate error.

Disruption attack A disruption attack destroys the stability of the system by hindering critical information in CPSs. Denial-of-Service attacks [36] and jamming attacks are typical disruption attacks. These attacks can be modeled as follows:

$$\widetilde{y}_t = \gamma_t y_t \;,\; \widetilde{u}_t = \delta_t u_t \tag{8}$$

where binary variables γ_t and δ_t represent the process of adversarial attacks, such that $\gamma_t = 0$ or $\delta_t = 0$ if the communication channel is interrupted and $\gamma_t = \delta_t = 1$ otherwise. In most scenarios of disruption attack, Lyapunov methods [68][69] are used to design attack strategy and analyze the input-to-state stability of the system.

3.2 Model-based detection methods

In response to the various attacks, some existing detection approaches are implemented to the CPSs, e.g., [10][51][55]. Among them, most common and effective detection method is

based on analysis of Kalman innovation. There are two main reasons for using Kalman innovation as a detection target. One is that Kalman innovation sequences $\{r_t\}$ are (i,i,d) distribution, which are easy to analyze. Another reason is that the distribution are known in advance, and anomalies can be detected while attacks change the distribution of $\{r_t\}$. In the following part, we present three typical residue detectors.

3.2.1 χ^2 detector

As we mentioned in Section 2.1, the Kalman innovation $r_t = y_t - C\widehat{x}_{t|t-1}$ of Kalman filter are (i.i.d) Gaussian $\mathbb{N}(0, \Sigma_r)$. The probability of the sequence $y_{t-\tau+1}, ..., y_t$ under nominal conditions is

$$\mathbb{P}(y_{t-\tau+1}, ..., y_t) = \left[\frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_r|}\right]^{\tau} e^{-\frac{1}{2}g_t} \tag{9}$$

where

$$g_t = \sum_{i=t-\tau+1}^{t} r_t^T \Sigma_r^{-1} r_t \gtrless threshold \tag{10}$$

When this probability is low, it means that the system is likely to be attacked. In order to check the probability, we only need to compute g_t, and then compare it to a fixed threshold.

3.2.2 CUSUM detector

For this residual $r_t = y_t - C\widehat{x}_{i|i-1}$, CUSUM detector identifies two hypothesis to be tested: $\mathcal{H}_0$ the normal mode (no attacks) and $\mathcal{H}_1$ the faulty mode (with attacks). The paper [70] proposes the absolute value of every entry of the residual sequence as distance measure, that is,

$$z_{t,i} := |r_{t,i}| = |y_{t,i} - C_i\widehat{x}_{i|i-1}|. \tag{11}$$

If there is no CPS attacks, the value of $r_{t,i}$ follows the normal distribution $\mathcal{N}(0, \sigma_i^2)$, where σ_i denotes the i-th diagonal entry of Σ_r. Therefore, under mode H_0, the non-negative number $z_{k,i}$ follows a half-normal distribution with $\mathbb{E}[|z_{t,i}|] = \frac{\sqrt{2}}{\sqrt{\pi}}\sigma_i$ and $var[|z_{t,i}|] = \sigma_i^2(1 - \frac{2}{\pi})$.

Based on the above fact, we can design the bias $b_i > 0$ and threshold $\tau_i > 0$. As a sign of attack detection, the CUSUM values can be initialized with $S_{1,i} = 0$, and be calculated as follows:

$$S_{t,i} = max(0, S_{t-1,i} + z_{t,i} - b_i) \gtrless threshold \tag{12}$$

3.2.3 KLD detector

In mathematical statistics, the Kullback–Leibler divergence (KLD) is a type of statistical distance to mearsure how one probability distribution P is different from a reference probability distribution Q. In continuous case, the relative entropy from Q to P on a measurable space χ is defined as KLD:

$$\mathbb{D}_{KL}(P \parallel Q) = \int_{\chi} p(x) log\left(\frac{p(x)}{q(x)}\right)\mu(x) \tag{13}$$

where μ is a measure on χ with density p and q. Relative entropy is always non-negative, with $\mathbb{D}_{KL}(P \parallel Q) = 0$ if and only if $P = Q$ as measures. In general, the lower the KL divergence value, the closer the two distributions are to one another.

Based on above properties, the $\mathbb{D}_{KL}$ between Kalman innovations with and without attacks is usually adopted to measure the stealthiness of the attack. Following [71][72], KLD detector is applied with a threshold:

$$\mathbb{D}_{KL}(r_t^a \parallel r_t^0) \gtrless threshold \tag{14}$$

where r_t^0 is the nominal Kalman innovations, and r_t^a is under attack.

4 Problem Formulation

In a real-world scenario, many unprotected sensors are scattered in various places, making it vulnerable to be attacked. For sensor attackers, efficiency and effectiveness are equally important. The opportunity is always fleeting, so it is necessary to deploy the attack in a short period of time. Besides, rather than degrading general system performance, such sensor attacks tend to have a specific impact on the system state. A related example is the GPS spoofing attack in autonomous vehicles. The attacker changes the GPS signal so that the autonomous car will deviate from the expected trajectory and crash into the obstacles.

Motivated by GPS spoofing, our research topic is to design an optimal attack strategy to drive the CPS's state to a specific target within a fixed time. This attack, as a deception attack, degrades the system performance by injecting malicious false data to the system output. Moreover, it should be stealthy, which means that it will not be detected during the attack.

In this section, we present the attack model with assumptions, and then evaluate the attack effect to system states and detection residue.

4.1 Problem description and assumption

Consider the CPS model mentioned in (1). The system dynamics under attack has the following form:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_t \\ y_t &= Cx_t + E\xi_{t-1} + v_t \end{aligned} \tag{15}$$

where ξ_t is the attack in the sensors. Without loss of generality, we assume the matrix E is injective.

We assume the CPS, equipped with Kalman filter and LQG, begins running at $t = -\infty$ and the attack starts at time $t = 0$ lasting until $t = N$. With this assumption, when the attack occurs, CPS already has fixed filter gain K and feedback gain L.

For the attacker, it knows the full information of the system model, including the matrices $A, B, C, E, L, K, \Sigma_w, \Sigma_v$. From its knowledge of the system, the attacker can compute the system's state estimate $\widehat{x}_t$ and the control input $u_t = L\widehat{x}_t$. The attacker's information set, $\mathcal{I}$, at instant $t \geq 0$ is given follows:

$$\mathcal{I}_0 = \{y_0\}\ ,\ \mathcal{I}_{t+1} = \{\mathcal{I}_t, y_{t+1}, \xi_t\} \tag{16}$$

Based on the information $\mathcal{I}_t$, the attacker's aim is to find an attack policy, i.e., $\xi_t = \mathcal{F}(\mathcal{I}_t)$, which misleads the system to a certain target x^* while being stealthy from the KLD detector. The author of [34] studied this problem by combining the control error cost J_c and the detection cost of the whole process J_d as the objective function. However, in his algorithm, it is difficult to choose the fixed penalty parameter of J_d or guarantee the detection cost to be small at each

step. By developing his problem, we consider the control error as the cost function J and the detection cost of each step as the constraint. The formulation of Problem 1 allows us to adjust constraints to avoid detection directly.

Problem 1.

$$\begin{array}{cc} \min\limits_{\mathcal{F}(\mathcal{I}_0),..,\mathcal{F}(\mathcal{I}_N)} & J = \mathbb{E}\left\{\sum_{t=1}^{N+1} \|x_t - x^*\|_{Q_t}^2\right\} \\ \text{s.t.} & \mathbb{D}_{KL}(r_t^a \parallel r_t^0) \leq \delta, t = 1, ..., N+1 \end{array} \tag{17}$$

4.2 Attack effect

An attack sequence $\{\xi_0, ..., \xi_t\}$ will induce a bias to the Kalman innovation sequence $\{r_0, ..., r_t\}$ and system state $\{x_1, ..., x_t\}$, which are the critical variables to our problem. In this part, we evaluate the effect of an attack sequence.

Define $\widehat{x}_t^a$ and $\widehat{x}_t^0$ to be the CPS's estimate with and without attack. Denote the estimation error and the innovation of the Kalman filter as $e_t^a = x_t - \widehat{x}_t^a$ and $r_t^a = y_t - C\widehat{x}_{t|t-1}^a$, respectively. Similarly, let e_t^0 and r_t^0 be the estimation error and the Kalman innovation under no attack. Although the system input u_t is unknown to attacker, it can be implicitly calculated by $u_t = L\widehat{x}_t^a$. With the information $\mathcal{I}_t$, the attacker can give its own estimate, denoted as $\widetilde{x}_t$ with innovation $\widetilde{r}_t$. It is easy to prove that $\widetilde{r}_t = r_t^0$ and $\widetilde{e}_t = e_t^0$.

Combining the dynamics, we obtain

$$\begin{aligned} r_{t+1}^a &= y_{t+1} - C\widehat{x}_{t+1}^a \\ &= CAe_t + E\xi_t + Cw_k + v_{t+1} \\ &= CA\Delta e_t + E\xi_t + r_{t+1}^0 \end{aligned} \tag{18}$$

where $\Delta e_t = e_t^a - e_t^0$ is the estimation error induced by the attack, and $\{r_{t+1}^0\}$ has (i.i.d) distribution $\mathbb{N}(0, \Sigma_r)$.

Moreover, we also derive the dynamic system of Δe_t:

$$\begin{aligned} e_{t+1} = x_{t+1} - \widehat{x}_{t+1}^a = (Ax_t + Bu_t + w_t) - [A\widehat{x}_t^a + Bu_t + K(CAe_t + E\xi_t + Cw_t + v_{t+1})] \\ \Rightarrow \quad \Delta e_{t+1} = (A - KCA)\Delta e_t - KE\xi_t \end{aligned} \tag{19}$$

Hence, we can summarize the effect of the attack on the CPS using the virtual state $\theta_t = [\widetilde{x}_t^T \ \widehat{x}_t^{aT} \ \Delta e_t^T]^T$ and the dynamical system

$$\theta_{t+1} = \mathcal{A}\theta_t + \mathcal{B}\xi_t + \mathcal{H}\widetilde{r}_{t+1} \tag{20}$$

where

$$\mathcal{A} = \begin{bmatrix} A & BL & 0 \\ 0 & A + BL & KCA \\ 0 & 0 & A - KCA \end{bmatrix}, \mathcal{B} = \begin{bmatrix} 0 \\ KE \\ -KE \end{bmatrix}, \mathcal{H} = \begin{bmatrix} K \\ K \\ 0 \end{bmatrix}$$

and $\widetilde{x}_0 - \widehat{x}_0^a = \Delta e_0 = 0$.

4.3 Detection constraint

From previous knowledge, we know $\{r_t^0\}$ has (i.i.d) distribution $\mathbb{N}(0, \Sigma_r)$. Given the condition $\mathcal{I}_k$, we can derive the distribution of $\{r_t^a | t > k\}$:

$$r_t|\mathcal{I}_k \sim \mathbb{N}(\beta_{k,t}, \Sigma_{k,t}) \tag{21}$$

where

$$\beta_{k,t} = CA(A-KCA)^{t-k-1}\Delta e_k + E\xi_{t-1} - \sum_{j=k}^{t-2} CA(A-KCA)^{t-2-j}KE\xi_j \tag{22}$$

and $\Sigma_{k,t} = \Sigma_r$. Hence, we calculate the Kullback–Leibler divergence in r_t^a and r_t^0

$$\begin{aligned}\mathbb{D}_{KL}(r_t^a||r_t^0\ |\mathcal{I}_k) &= \frac{1}{2}\left[Tr(\Sigma_r^{-1}\Sigma_{k,t}) + log\frac{|\Sigma_r|}{|\Sigma_{k,t}|} - p + \beta_{k,t}^T\Sigma_r^{-1}\beta_{k,t}\right] \\ &= \frac{1}{2}\beta_{k,t}^T\Sigma_r^{-1}\beta_{k,t}\end{aligned} \tag{23}$$

Since the value of $\mathbb{D}_{KL}(r_t^a||r_t^0)$ is related to the false alarm rate, we bound it to keep stealthy

$$\beta_{k,t}^T\Sigma_r^{-1}\beta_{k,t} \leq \delta, t = k+1, ..., N+1 \tag{24}$$

4.4 Reformulate the problem

In this subsection, we reformulate the original problem to a quadratic programming.

From the knowledge of Kalman filter, it is well-known that the attacker's estimation error $\widetilde{e}_t \sim \mathbb{N}(0, (I-KC)P)$, and is orthogonal to $\widetilde{x}_t$, which means $\mathbb{E}[(x_t - \widetilde{x}_t)^T\widetilde{x}_t] = 0$. Therefore, the cost function becomes

$$J = \mathbb{E}\left\{\sum_{t=1}^{N+1}\|x_t - x^*\|_{Q_t}^2\right\} \Longleftrightarrow J = \mathbb{E}\left\{\sum_{t=1}^{N+1}[Tr((I_n - KC)PQ_t) + \|\widetilde{x}_t - x^*\|_{Q_t}^2]\right\} \tag{25}$$

Add the target state x^* to the virtual state $\theta_t = [\widetilde{x}_t^T\ \widehat{x}_t^{aT}\ \Delta e_t^T]^T$, and denote $\bar{\theta}_t = [\theta_t^T\ (x^*)^T]^T$, $F_1 = [I_n\ 0\ 0\ -I_n]$, $F_2 = [0\ 0\ I_n\ 0]$. Applying $\widetilde{x}_t - x^* = F_1\bar{\theta}_t$ and $\Delta e_t = F_2\bar{\theta}_t$, the original problem is expressed in the form

Problem 2.

$$\begin{aligned}\min_{\mathcal{F}(\mathcal{I}_0),..,\mathcal{F}(\mathcal{I}_N)}\quad & \bar{J} = \mathbb{E}\left\{\sum_{t=1}^{N+1}\left\|F_1\bar{\theta}_t\right\|_{Q_t}^2\right\} \\ \text{s.t.}\quad & \bar{\theta}_{t+1} = \bar{\mathcal{A}}\bar{\theta}_t + \bar{\mathcal{B}}\xi_t + \bar{\mathcal{H}}\widetilde{r}_{t+1},\ \ t = 0, ..., N \\ & \beta_{k,t}^T\Sigma_r^{-1}\beta_{k,t} \leq \delta,\ \ k = 1, ..., N, t = k+1, ..., N+1 \\ & \beta_{k,t} = CA(A-KCA)^{t-k-1}F_2\bar{\theta}_k + E\xi_{t-1} - \sum_{j=k}^{t-2}CA(A-KCA)^{t-2-j}KE\xi_j\end{aligned} \tag{26}$$

where $\bar{\mathcal{A}} = \begin{bmatrix}\mathcal{A} & 0\\ 0 & I\end{bmatrix}$, $\bar{\mathcal{B}} = \begin{bmatrix}\mathcal{B}\\ 0\end{bmatrix}$, and $\bar{\mathcal{H}} = \begin{bmatrix}\mathcal{H}\\ 0\end{bmatrix}$. Knowing the initial state $\bar{\theta}_t = [\widetilde{x}_0^T\ \widehat{x}_0^{aT}\ 0\ (x^*)^T]$, the attacker aims to design the optimal attack sequences $\{\xi_t\}$ based on the measurement of the

stochastic variable $\widetilde{r}_t$.

5 Preliminary Results

In this section, we introduce an algorithm about how to calculate the optimal attack policy, and demonstrate the attack strategy using numerical examples.

5.1 Algorithm

First, we discuss the attack strategy at time k, which means the attacker has deployed the attack signals $\{\xi_0, ..., \xi_{k-1}\}$ and known the Kalman innovation $\{\widetilde{r}_0, ..., \widetilde{r}_k\}$. Given the condition $\mathcal{F}(\mathcal{I}_k)$, the term $\bar{\theta}_k$ is also known.

Define the system state sequence $\{\bar{\theta}_{k+1}, ..., \bar{\theta}_{N+1}\}$ as a vector Θ_{k+1}^{N+1}, the attack sequences $\{\xi_k, ..., \xi_N\}$ as Ξ_k^N, the residue sequence $\{\widetilde{r}_{k+1}, ..., \widetilde{r}_{N+1}\}$ as $\widetilde{R}_{k+1}^{N+1}$, the sequence $\{\beta_{k,k+1}, ..., \beta_{k,N+1}\}$ as Γ_{k+1}^{N+1}, and then we have

$$\begin{aligned} \Theta_{k+1}^{N+1} &= \mathbb{A}_k \bar{\theta}_k + \mathbb{B}_k \Xi_k^N + \mathbb{H}_k \widetilde{R}_{k+1}^{N+1} \\ \Gamma_{k+1}^{N+1} &= \mathbb{C}_k \bar{\theta}_k + \mathbb{D}_k \Xi_k^N \end{aligned} \tag{27}$$

where

$$\mathbb{A}_k = \begin{bmatrix} \bar{\mathcal{A}} \\ \bar{\mathcal{A}}^2 \\ \vdots \\ \bar{\mathcal{A}}^{N+1-k} \end{bmatrix}, \mathbb{B}_k = \begin{bmatrix} \bar{\mathcal{B}} & 0 & \cdots & 0 \\ \bar{\mathcal{A}}\bar{\mathcal{B}} & \bar{\mathcal{B}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathcal{A}}^{N-k}\bar{\mathcal{B}} & \bar{\mathcal{A}}^{N-k-1}\bar{\mathcal{B}} & \cdots & \bar{\mathcal{B}} \end{bmatrix}, \mathbb{H}_k = \begin{bmatrix} \bar{\mathcal{H}} & 0 & \cdots & 0 \\ \bar{\mathcal{A}}\bar{\mathcal{H}} & \bar{\mathcal{H}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathcal{A}}^{N-k}\bar{\mathcal{H}} & \bar{\mathcal{A}}^{N-k-1}\bar{\mathcal{H}} & \cdots & \bar{\mathcal{H}} \end{bmatrix},$$

$$\mathbb{C}_k = \begin{bmatrix} CAF_2 \\ CA(A-KCA)F_2 \\ \vdots \\ CA(A-KCA)^{N-k}F_2 \end{bmatrix}, \mathbb{D}_k = \begin{bmatrix} E & \cdots & 0 & 0 \\ \vdots & \ddots & \cdots & \cdots \\ -CA(A-KCA)^{N-2-k}KE & \cdots & E & 0 \\ -CA(A-KCA)^{N-1-k}KE & \cdots & -CAKE & E \end{bmatrix}$$

Substituting (27) into (26), and after some algebraic manipulations, the original problem become a convex problem.

Problem 3.

$$\begin{aligned} \min_{\Xi_k^N} \quad & \bar{J}_k = 2\bar{\theta}_k^T \mathbb{A}_k^T \mathbb{Q}_k \mathbb{B}_k \Xi_k^N + \left\| \mathbb{B}_k \Xi_k^N \right\|_{\mathbb{Q}_k}^2 \\ \text{s.t.} \quad & \left\| \mathbb{C}_k \bar{\theta}_k + \mathbb{D}_k \Xi_k^N \right\|_{\mathbb{S}_i}^2 \leq \delta, \ i = 1, .., N-k \end{aligned} \tag{28}$$

where

$$\begin{aligned} \mathbb{Q}_k &= diag(F_1^T Q_{k+1} F_1, ..., F_1^T Q_{N+1} F_1) \\ \mathbb{S}_i &= diag(0, ..., 0, {\Sigma_r^{-1}}_{-i_{th}}, 0, ..., 0), \ i = 1, .., N-k \end{aligned}$$

Obviously, Problem 3 is a standard QCQP problem and can be effectively solved by CVX toolbox. We then obtain Algorithm 1 of the attacker policy by solving Problem 3 repeatedly.

Algorithm 1 Optimal attack policy $\mathcal{F}(\mathcal{I}_k)$

$\mathcal{I}_0 = \{u_{-\infty}, .., u_{-1}, y_{-\infty}, ..., y_0\}$, $\xi_k = 0$ and $k = 0$
while $1 \leq k \leq N$ **do**
    Attack the sensor with optimal attack ξ_{k-1}
    Measure the output y_k and update $\mathcal{I}_k = \{\mathcal{I}_{k-1}, u_{k-1}, y_k, \xi_{k-1}\}$
    Calculate the θ_k from $\mathcal{F}(\mathcal{I}_k)$ and find optimal Ξ_k^N by solving problem 3
    $k \leftarrow k + 1$
end while

Although the attack strategy $\mathcal{F}(\mathcal{I}_k)$ is optimal at each instant k, it is not sufficient to illustrate the optimality for the entire process. It is necessary to prove that Algorithm 1 provides the optimal attack sequences $\{\xi_1, ..., \xi_N\}$ that solves Problem 1, which is shown it in the appendix.

5.2 Experiment results

In this numerical experiment, we consider a scenario that autonomous vehicles are following the lane line and driving at a uniform speed in parallel when GPS spoofing attack occurs. The GPS acts as an observer, providing the car's location to calculate the distance from the centerline of lane, while the controller manages to keep it at zero. The experiment is implemented in Python, and the code is available at https://github.com/Eric-hkust/cps_attack.git.

To illustrate the effectiveness of the algorithm succinctly, we use the one-dimensional kinematic model, where the state x_t is the car's position in the vertical direction of the lane line. The parameters of system (15) are given by $A = B = C = E = 1$, $\Sigma_w = 0.001$ and $\Sigma_v = 0.002$. The controller gain is chosen to be $L = -0.5$, and the initial state is assumed to be $x_0 = 0$. Moreover, the attacker's goal is to move the state of the car to $x^* = 0.6$ within a given step $N = 19$.

Experiment 1: attack effect

This experiment illustrates the attack policy and how it changes the original trajectory. As it shown in Figure 4, under the attack, the vehicle's trajectory gradually deviates from his estimates and original trajectory, and finally crashes to an obstacle at instant $t = 19$. In Figure 5, we compare the χ^2 in the normal case with the attacked situation, where the attack deployed at $t = 0$. The result shows that the attack policy has little effect on χ^2, thus it is difficult to be detected.

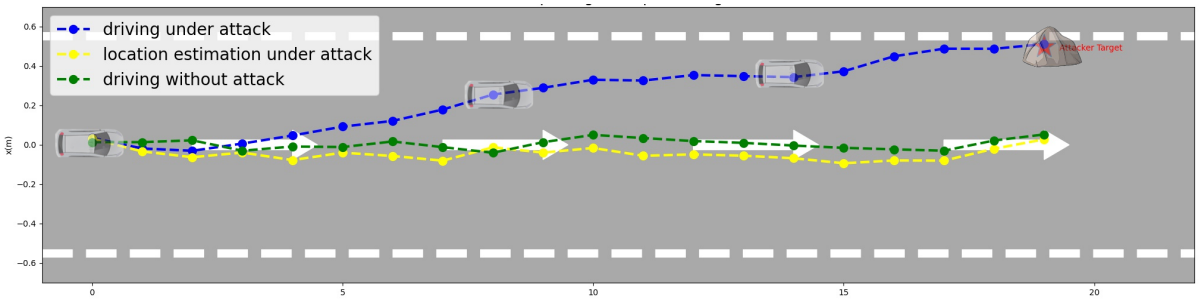

Figure 4: Autonomous vehicle's trajectory under attack

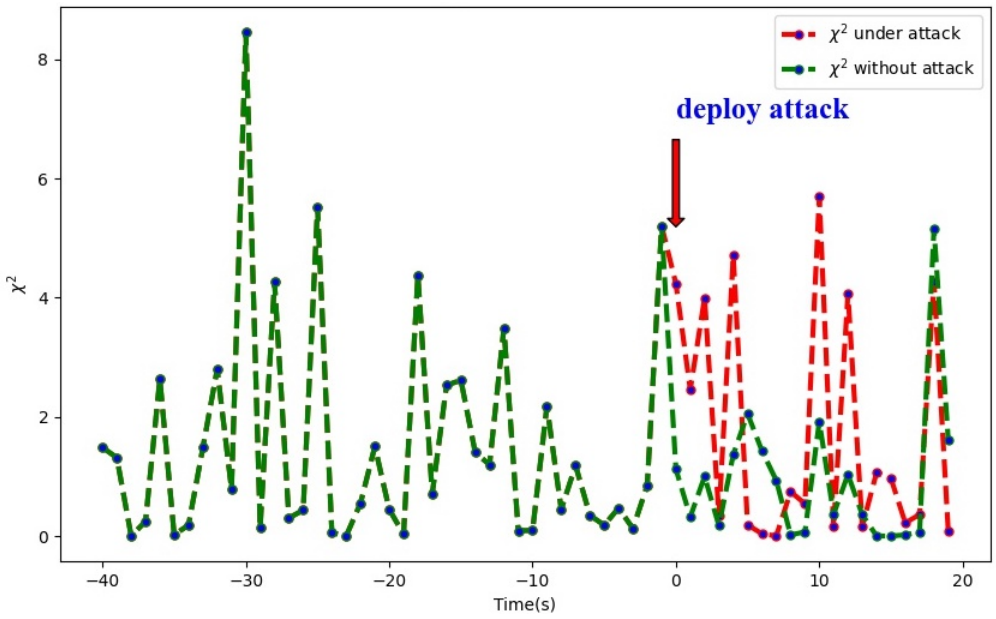

Figure 5: The impact of attack on the detector

Experiment 2: the detector with exceptions handling

In practice, considering the false alarm rate, the system is always equipped with relatively large threshold for attack alerts, like the alarm threshold $\delta_{attack} = 20$ for χ^2 detector. However, there are also some common anomalies, e.g., large measurement noises by accident. To handle the gap between the attack alarms and common anomalies, the detector sets a relatively small threshold $\delta_{anomaly}$ for common anomalies. When the χ^2 is larger than $\delta_{anomaly}$ but smaller than δ_{attack}, the system's Kalman filter ignores the measurement at this moment and only uses the prediction to update the state estimate.

$$\text{detector}(\chi^2) = \begin{cases} \text{update state estimation}, & \chi^2 \leq \delta_{anomaly} \\ \text{ignore this measurement}, & \delta_{anomaly} < \chi^2 \leq \delta_{attack} \\ \text{sent an attack warning}, & \delta_{attack} < \chi^2 \end{cases}$$

This experiment illustrates that our attack strategy works well while confronting the detector equipped with exceptions handling capability. We show the impact of different value of anomaly thresholds to our attack sequences in Figure 7. As seen in Figure 6, the result shows that even though some of the attack signals are blocked, our algorithm can have an adaptive adjustment and finally induce the car to crash into the target obstacle.

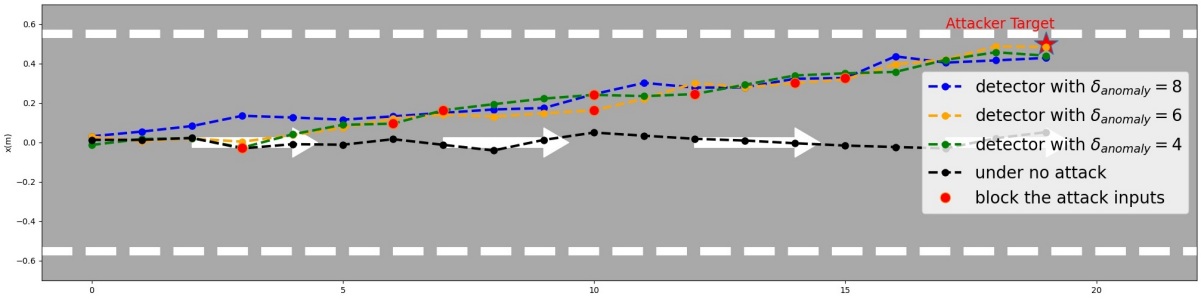

Figure 6: Autonomous vehicle's trajectory under attack

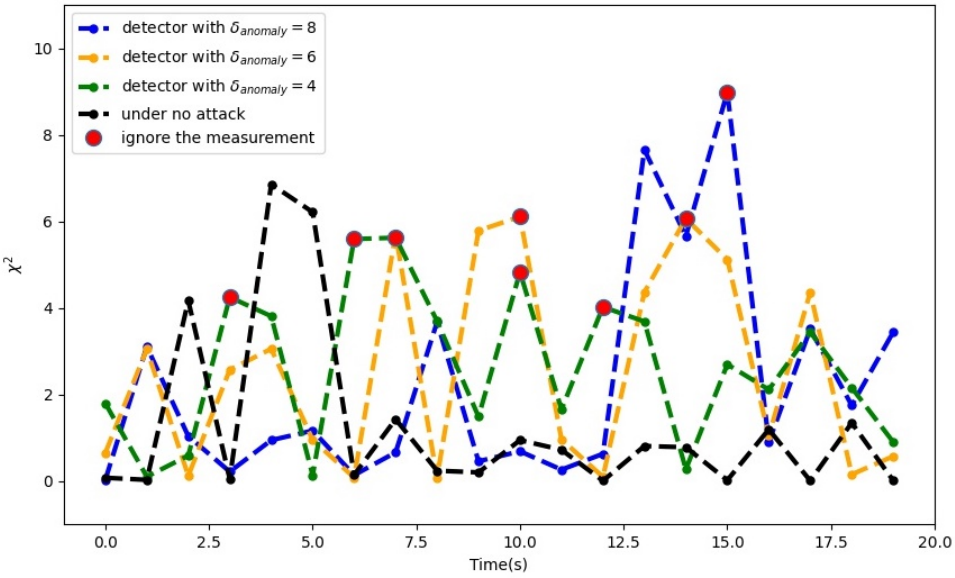

Figure 7: The detector equipped with different $\delta_{anomaly}$

6 Future Work

With sensor attacks, we aim to drive the CPS's state to a specific target. The optimal attack strategy is obtained by solving a series of QCQP problems, and it works well even if some attack signals are blocked unexpectedly. However, there are some unresolved issues of this research topic, i.e., nonlinear system model, multi-sensor case, and imperfect system knowledge. Future works addressing these limitations are summarized as follows.

System modeling and analysis In the published literature, most of the work is on time-invariant linear system. The reason is that the existing analysis approach cannot completely handle the complex system dynamics, e.g., time-varying system behaviors and nonlinear system model. Additionally, when faced with partial information, how to design the attack strategies and defense mechanisms without full knowledge and observability is challenging. Although some researchers have noticed this issue zhang2022design, the research results are relatively few. Moreover, there is an obvious gap between the analysis tools based on system models and CPS design requirement. In other words, computing and decision-making should be modeled at different levels of communication and system dynamics.

Multi-sensor and large-scale network system To ensure system reliability, a modern system is always equipped with different type of sensors, measuring partial information of system states respectively. For instance, the autonomous vehicle utilizes cameras, Lidar and GPS for self-localization, but it increases the risk of sensor attacks. It is essential to study attacks in multi-sensor case as it determines whether one can put a large number of sensors in a system. The auther of [73] proposed a related framework, but there remains much room to study this topic. Furthermore, in large-scale CPS network, physical and cyber components are usually deployed in a spatially distributed way. Any attack signal can be propagated over communication topology and affect the dynamical behavior of the entire physical system. As such, a challenge is to accurately locate the attack sources and isolate the attack.

The practical applications CPS security issue is generalized from the real world. However, how to apply existing CPS theory and model-based methods remain a challenging issue. Most of them need too-strong assumptions, e.g., system are supposed to be simple and work

perfectly, which are unlikely for most real-world industrial CPSs. Besides, most algorithms' performance is sensitive to key parameters, while a perfect parameter is difficult to know in most case. For example, the attack policy heavily relies on guesses about the control law and Kalman gain. Furthermore, the evaluation criterion of CPS is different from control theory. Rather than stability or reachable set, attackers are more interested in how to maximize damage with limited information in a short time ,and systems are interested in how to detect anomalies. From an application perspective, this new evaluation of system is still in its infancy and deserves considerable attention.

References

[1] Raj Rajkumar, Dionisio De Niz, and Mark Klein. *Cyber-physical systems*. Addison-Wesley Professional, 2016.

[2] Edward Ashford Lee and Sanjit Arunkumar Seshia. *Introduction to embedded systems: A cyber-physical systems approach*. Mit Press, 2016.

[3] Edward A Lee. Cyber physical systems: Design challenges. In *Proceedings of the 11th international symposium on object and component-oriented real-time distributed computing (ISORC)*, pages 363–369. IEEE, 2008.

[4] Jeff C Jensen, Danica H Chang, and Edward A Lee. A model-based design methodology for cyber-physical systems. In *Proceedings of the 7th international wireless communications and mobile computing conference*, pages 1666–1671. IEEE, 2011.

[5] Jairo Giraldo, David Urbina, Alvaro Cardenas, Junia Valente, Mustafa Faisal, Justin Ruths, Nils Ole Tippenhauer, Henrik Sandberg, and Richard Candell. A survey of physics-based attack detection in cyber-physical systems. *ACM Computing Surveys (CSUR)*, 51(4):1–36, 2018.

[6] Juliza Jamaludin and Jemmy Mohd Rohani. Cyber-physical system (cps): State of the art. In *Proceedings of the International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, pages 1–5. IEEE, 2018.

[7] Yifei Tan, Wenhe Yang, Kohtaroh Yoshida, and Soemon Takakuwa. Application of iot-aided simulation to manufacturing systems in cyber-physical system. *Machines*, 7(1):2, 2019.

[8] Hong Chen. Applications of cyber-physical system: a literature review. *Journal of Industrial Integration and Management*, 2(03):1750012, 2017.

[9] Xinghuo Yu and Yusheng Xue. Smart grids: A cyber–physical systems perspective. *Proceedings of the IEEE*, 104(5):1058–1070, 2016.

[10] Quanyan Zhu and Tamer Basar. Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: games-in-games principle for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine*, 35(1):46–65, 2015.

[11] Ovunc Kocabas, Tolga Soyata, and Mehmet K Aktas. Emerging security mechanisms for medical cyber physical systems. *IEEE/ACM transactions on computational biology and bioinformatics*, 13(3):401–416, 2016.

[12] Haibo He and Jun Yan. Cyber-physical attacks and defences in the smart grid: a survey. *IET Cyber-Physical Systems: Theory & Applications*, 1(1):13–27, 2016.

[13] Kyounggon Kim, Jun Seok Kim, Seonghoon Jeong, Jo-Hee Park, and Huy Kang Kim. Cybersecurity for autonomous vehicles: Review of attacks and defense. *Computers & Security*, 103:102150, 2021.

[14] Eric Ke Wang, Yunming Ye, Xiaofei Xu, Siu-Ming Yiu, Lucas Chi Kwong Hui, and Kam-Pui Chow. Security issues and challenges for cyber physical system. In *Proceedings of the Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*, pages 733–738. IEEE, 2010.

[15] Wenjin Yu, Tharam Dillon, Fahed Mostafa, Wenny Rahayu, and Yuehua Liu. Implementation of industrial cyber physical system: Challenges and solutions. In *Proceedings of the International Conference on Industrial Cyber Physical Systems (ICPS)*, pages 173–178. IEEE, 2019.

[16] Ling Shi, Lihua Xie, and Richard M Murray. Kalman filtering over a packet-delaying network: A probabilistic approach. *Automatica*, 45(9):2134–2140, 2009.

[17] Kun Liu, Anton Selivanov, and Emilia Fridman. Survey on time-delay approach to networked control. *Annual Reviews in Control*, 48:57–79, 2019.

[18] Yilin Mo and Bruno Sinopoli. False data injection attacks in control systems. In *Proceedings of the Preprints of the 1st workshop on Secure Control Systems*, volume 1, 2010.

[19] David Kushner. The real story of stuxnet. *ieee Spectrum*, 50(3):48–53, 2013.

[20] Ralph Langner. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3):49–51, 2011.

[21] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In *Proceedings of the International conference on critical infrastructure protection*, pages 73–82. Springer, 2007.

[22] Alexander Ruegamer, Dirk Kowalewski, et al. Jamming and spoofing of gnss signals–an underestimated risk?! *Proc. Wisdom Ages Challenges Modern World*, 3:17–21, 2015.

[23] Kim Hartmann and Christoph Steup. The vulnerability of uavs to cyber attacks-an approach to the risk assessment. In *Proceedings of the 5th international conference on cyber conflict*, pages 1–23. IEEE, 2013.

[24] Nikos Komninos, Eleni Philippou, and Andreas Pitsillides. Survey in smart grid and smart home security: Issues, challenges and countermeasures. *IEEE Communications Surveys & Tutorials*, 16(4):1933–1954, 2014.

[25] Jing Liu, Yang Xiao, Shuhui Li, Wei Liang, and CL Philip Chen. Cyber security and privacy issues in smart grids. *IEEE Communications surveys & tutorials*, 14(4):981–997, 2012.

[26] Riham AlTawy and Amr M Youssef. Security tradeoffs in cyber physical systems: A case study survey on implantable medical devices. *Ieee Access*, 4:959–979, 2016.

[27] Michael Rushanan, Aviel D Rubin, Denis Foo Kune, and Colleen M Swanson. Sok: Security and privacy in implantable medical devices and body area networks. In *Proceedings of the symposium on security and privacy*, pages 524–539. IEEE, 2014.

[28] Kui Ren, Qian Wang, Cong Wang, Zhan Qin, and Xiaodong Lin. The security of autonomous driving: Threats, defenses, and future directions. *Proceedings of the IEEE*, 108(2):357–372, 2019.

[29] Minh Pham and Kaiqi Xiong. A survey on security attacks and defense techniques for connected and autonomous vehicles. *Computers & Security*, 109:102269, 2021.

[30] Jin Cui, Lin Shen Liew, Giedre Sabaliauskaite, and Fengjun Zhou. A review on safety failures, security attacks, and available countermeasures for autonomous vehicles. *Ad Hoc Networks*, 90:101823, 2019.

[31] Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. Detecting integrity attacks on scada systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, 2013.

[32] Heng Zhang, Peng Cheng, Ling Shi, and Jiming Chen. Optimal dos attack scheduling in wireless networked control system. *IEEE Transactions on Control Systems Technology*, 24(3):843–852, 2015.

[33] Liwei An and Guang-Hong Yang. Data-based optimal denial-of-service attack scheduling against robust control based on q-learning. *International Journal of Robust and Nonlinear Control*, 29(15):5178–5194, 2019.

[34] Yuan Chen, Soummya Kar, and José MF Moura. Cyber-physical attacks with control objectives. *IEEE Transactions on Automatic Control*, 63(5):1418–1425, 2017.

[35] Qirui Zhang, Kun Liu, André MH Teixeira, Yuzhe Li, Senchun Chai, and Yuanqing Xia. An online kullback-leibler divergence-based stealthy attack against cyber-physical systems. *IEEE Transactions on Automatic Control*, 2022.

[36] Heng Zhang, Peng Cheng, Ling Shi, and Jiming Chen. Optimal denial-of-service attack scheduling with energy constraint. *IEEE Transactions on Automatic Control*, 60(11):3023–3028, 2015.

[37] Ziyang Guo, Dawei Shi, Karl Henrik Johansson, and Ling Shi. Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*, 4(1):4–13, 2016.

[38] Shuang Wu, Ziyang Guo, Dawei Shi, Karl Henrik Johansson, and Ling Shi. Optimal innovation-based deception attack on remote state estimation. In *Proceedings of the American Control Conference (ACC)*, pages 3017–3022. IEEE, 2017.

[39] Yilin Mo and Bruno Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61(9):2618–2624, 2015.

[40] Cheng-Zong Bai, Fabio Pasqualetti, and Vijay Gupta. Security in stochastic control systems: Fundamental limitations and performance bounds. In *Proceedings of the American Control Conference (ACC)*, pages 195–200. IEEE, 2015.

[41] Yilin Mo, Emanuele Garone, Alessandro Casavola, and Bruno Sinopoli. False data injection attacks against state estimation in wireless sensor networks. In *Proceedings of the 49th Conference on Decision and Control (CDC)*, pages 5967–5972. IEEE, 2010.

[42] Enoch Kung, Subhrakanti Dey, and Ling Shi. The performance and limitations of stealthy attacks on higher order systems. *IEEE Transactions on Automatic Control*, 62(2):941–947, 2016.

[43] Roy S Smith. A decoupled feedback structure for covertly appropriating networked control systems. *IFAC Proceedings Volumes*, 44(1):90–95, 2011.

[44] Roy S Smith. Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems Magazine*, 35(1):82–92, 2015.

[45] Liang Hu, Zidong Wang, Qing-Long Han, and Xiaohui Liu. State estimation under false data injection attacks: Security analysis and system protection. *Automatica*, 87:176–183, 2018.

[46] Tianju Sui, Yilin Mo, Damián Marelli, Ximing Sun, and Minyue Fu. The vulnerability of cyber-physical system under stealthy attacks. *IEEE Transactions on Automatic Control*, 66(2):637–650, 2020.

[47] Carlos Murguia and Justin Ruths. Characterization of a cusum model-based sensor attack detector. In *Proceedings of the 55th Conference on Decision and Control (CDC)*, pages 1303–1309. IEEE, 2016.

[48] Cynthia A Lowry, William H Woodall, Charles W Champ, and Steven E Rigdon. A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1):46–53, 1992.

[49] Yuzhe Li, Daniel E Quevedo, Subhrakanti Dey, and Ling Shi. Sinr-based dos attack on remote state estimation: A game-theoretic approach. *IEEE Transactions on Control of Network Systems*, 4(3):632–642, 2016.

[50] Rômulo Meira-Góes, Stéphane Lafortune, and Hervé Marchand. Synthesis of supervisors robust against sensor deception attacks. *IEEE Transactions on Automatic Control*, 66(10):4990–4997, 2021.

[51] Xiaokang Zhou, Wei Liang, Shohei Shimizu, Jianhua Ma, and Qun Jin. Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems. *IEEE Transactions on Industrial Informatics*, 17(8):5790–5798, 2020.

[52] Beibei Li, Yuhao Wu, Jiarui Song, Rongxing Lu, Tao Li, and Liang Zhao. Deepfed: Federated deep learning for intrusion detection in industrial cyber–physical systems. *IEEE Transactions on Industrial Informatics*, 17(8):5615–5624, 2020.

[53] Felix O Olowononi, Danda B Rawat, and Chunmei Liu. Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for cps. *IEEE Communications Surveys & Tutorials*, 23(1):524–552, 2020.

[54] Yilin Mo, Sean Weerakkody, and Bruno Sinopoli. Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems Magazine*, 35(1):93–109, 2015.

[55] Kiminao Kogiso and Takahiro Fujita. Cyber-security enhancement of networked control systems using homomorphic encryption. In *Proceedings of the 54th Conference on Decision and Control (CDC)*, pages 6836–6843. IEEE, 2015.

[56] Aditya Ashok, Pengyuan Wang, Matthew Brown, and Manimaran Govindarasu. Experimental evaluation of cyber attacks on automatic generation control using a cps security testbed. In *Proceedings of the Power & Energy Society General Meeting*, pages 1–5. IEEE, 2015.

[57] Richard J Meinhold and Nozer D Singpurwalla. Understanding the kalman filter. *The American Statistician*, 37(2):123–127, 1983.

[58] Gary Bishop, Greg Welch, et al. An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8(27599-23175):41, 2001.

[59] Algirdas Avizienis, J-C Laprie, Brian Randell, and Carl Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE transactions on dependable and secure computing*, 1(1):11–33, 2004.

[60] Jason Andress. *The basics of information security: understanding the fundamentals of InfoSec in theory and practice*. Syngress, 2014.

[61] Michael Hylkema. A survey of database inference attack prevention methods. *Educational Technology Research*, 2009.

[62] Ye Yuan and Yilin Mo. Security in cyber-physical systems: Controller design against known-plaintext attack. In *Proceedings of the 54th Conference on Decision and Control (CDC)*, pages 5814–5819. IEEE, 2015.

[63] Yilin Mo and Bruno Sinopoli. Secure control against replay attacks. In *Proceedings of the 47th annual Allerton conference on communication, control, and computing (Allerton)*, pages 911–918. IEEE, 2009.

[64] Yao Liu, Peng Ning, and Michael K Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, 14(1):1–33, 2011.

[65] Xuan Liu, Zhen Bao, Dan Lu, and Zuyi Li. Modeling of local false data injection attacks with reduced network information. *IEEE Transactions on Smart Grid*, 6(4):1686–1696, 2015.

[66] Xiao Wei and Biplab Sikdar. Impact of gps time spoofing attacks on cyber physical systems. In *Proceedings of the international conference on industrial technology (ICIT)*, pages 1155–1160. IEEE, 2019.

[67] Qingyu Yang, Jie Yang, Wei Yu, Dou An, Nan Zhang, and Wei Zhao. On false data-injection attacks against power system state estimation: Modeling and countermeasures. *IEEE Transactions on Parallel and Distributed Systems*, 25(3):717–729, 2013.

[68] Eduardo D Sontag. Input to state stability: Basic concepts and results. In *Nonlinear and optimal control theory*, pages 163–220. Springer, 2008.

[69] Claudio De Persis and Pietro Tesi. Input-to-state stabilizing control under denial-of-service. *IEEE Transactions on Automatic Control*, 60(11):2930–2944, 2015.

[70] Carlos Murguia and Justin Ruths. Cusum and chi-squared attack detection of compromised sensors. In *Proceedings of the Conference on Control Applications (CCA)*, pages 474–480. IEEE, 2016.

[71] Ziyang Guo, Dawei Shi, Karl Henrik Johansson, and Ling Shi. Worst-case stealthy innovation-based linear attack on remote state estimation. *Automatica*, 89:117–124, 2018.

[72] Cheng-Zong Bai, Fabio Pasqualetti, and Vijay Gupta. Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs. *Automatica*, 82:251–260, 2017.

[73] Yuzhe Li, Ling Shi, and Tongwen Chen. Detection against linear deception attacks on multi-sensor remote state estimation. *IEEE Transactions on Control of Network Systems*, 5(3):846–856, 2017.