https://doi.org/10.20965/jrm.2025.p1392

Paper:

Directed Poisoning Attacks on FRIT in Adaptive Cruise Control

Taichi Ikezaki*, Kenji Sawada, and Osamu Kaneko*****

*Faculty of Environmental, Life, Natural Science and Technology, Okayama University
3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan
E-mail: t.ikezaki@okayama-u.ac.jp
**Graduate school of Mechanical Engineering, The University of Osaka
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
E-mail: knj.sawada@mech.eng.osaka-u.ac.jp
***Graduate School of Informatics and Engineering, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan
E-mail: o.kaneko@uec.ac.jp
[Received April 4, 2025; accepted July 26, 2025]

Recent advances in connected-vehicle technologies have enabled the large-scale collection of driving data, facilitating the deployment of data-driven control schemes. Although these methods offer advantages by eliminating the need for explicit modeling, they also introduce vulnerabilities due to their reliance on stored data. This study investigates a class of targeted data poisoning attacks on fictitious reference iterative tuning, a widely used data-driven controller tuning approach. We present a method that allows an adversary to influence closed-loop dynamics by manipulating the training data so that the resulting controller behavior matches a maliciously defined reference response. This strategy differs from conventional poisoning attacks, which aim only to the degrade control performance. Instead, it enables deliberate alteration of control characteristics such as overshoot and convergence time. The proposed attack is formulated as a constrained optimization problem under bounded tampering signals. Through a numerical study involving adaptive cruise control with stop functionality, we show that minor data modifications, indistinguishable from sensor noise, can cause significant degradation in control behavior. These findings highlight the need for robust security mechanisms in data-driven control implementation.

Keywords: cyberattack, data-driven control, cruise control, FRIT, poisoning attack

1. Introduction

In recent years, cyberattacks have increasingly targeted important systems such as power grids, factories, and other control systems that support our daily lives [1]. Because of this growing risk, many researchers are studying methods to protect control systems from cyber threats, especially from an engineering perspective [2].

One area of particular concern is the automotive industry. As more vehicles adopt Internet of Things (IoT) technologies, their measurement and control systems have become more vulnerable to attackers. In self-driving cars, ensuring the security of control algorithms is therefore a critical concern.

Data-driven control (DDC) has emerged as a promising framework for controller design, in which control parameters are derived directly from input–output time-series data obtained during actual system operation [3–6]. Unlike traditional model-based approaches, DDC does not require an explicit plant model such as a transfer function or state-space representation, thereby allowing more efficient controller tuning.

A well-studied example is fictitious reference iterative tuning (FRIT), which enables the design of high-performance controllers from a single experimental trial [3]. The practicality of DDC has led to its application in domains such as vehicle motion control [7]. Moreover, several studies have applied FRIT in various contexts, including industrial equipment [8] and theoretical extensions [9].

However, DDC is inherently dependent on the quality and integrity of the recorded input–output data. Since the control law is obtained by minimizing a cost function defined over these data, any modification or corruption may directly affect the resulting system behavior. Recent studies have shown that DDC is susceptible to "poisoning attacks," in which adversaries deliberately contaminate the data to degrade the performance or stability of the resulting control system [10–16].

A major challenge in addressing such attacks is that many existing studies use the same cost function for both attack design and the DDC method [10, 11]. Consequently, it is often unclear how an attack actually changes the system's behavior. In addition, DDC algorithms are especially weak against small changes in data, as such changes can be concealed within normal sensor noise. This presents

a serious safety issue, especially in systems such as adaptive cruise control (ACC), where the vehicle must stop at correct location. For example, excessive overshoot during stopping may cause the vehicle to pass the stop point and lead to an accident [17]. Previous studies investigating cyberattacks on ACC systems have already revealed significant safety concerns [18].

In this study, we focus on a type of directed poisoning attack against FRIT, applied to stop-motion control in automotive ACC systems [17, 19]. Unlike typical poisoning attacks that simply reduce control performance, our method allows an attacker to intentionally alter the system's behavior for example, by increasing the time constant or causing an overshoot. Because the attacker can control system behavior, the effects of the attack become clearer and more dangerous, particularly for safety-critical systems.

Although this study focuses on ACC stop control, the proposed attack formulation is applicable to control systems that can be tuned using least-squares-based FRIT.

The remainder of this paper is organized as follows. Section 2 describes the problem formulation of ACC control systems and FRIT. Section 3 introduces the poisoning attack scenario and reviews related studies. Section 4 presents the formulation of the proposed attack. Finally, Section 5 provides two numerical examples. Section 6 concludes the paper.

<Notation> Let $\mathbb{R}^n$ denote the set of n-dimensional real column vectors, and let $a^\top$ denote the transpose of a vector $a \in \mathbb{R}^n$.

For a continuous-time signal w, the value of the signal at time t is denoted by $w(t)$. The finite-time sequence of this signal, sampled with period t_s, from $k = k_1$ to $k = k_2$, is represented by $\boldsymbol{w}_{[k_1,k_2]} \in \mathbb{R}^{k_2-k_1+1}$. For brevity, we simply write $\boldsymbol{w} \in \mathbb{R}^{k_2-k_1+1}$ when the context is clear.

The norm of this finite-time signal is defined as

$$\|w\|_{[0,N]} := \sqrt{\sum_{i=0}^{N} w(t_s i)^2}. \tag{1}$$

Let $\mathcal{R}(s)$ denote the set of rational functions in s with real coefficients.

The output of a system with transfer function $G(s) \in \mathcal{R}(s)$ in response to the input signal $w(t)$ is, strictly speaking, given by the convolution of the Markov parameters of $G(s)$ and the signal $w(t)$. However, for notational simplicity, we denote this as $G(s)[\boldsymbol{w}](t)$ throughout the study.

2. Preliminary

2.1. Vehicle Dynamics and Control Objective

The control system is illustrated in **Fig. 1**. This system is designed to decelerate a vehicle cruising at a constant speed v_c [m/s] so that it can stop at a designated target position using the deceleration mode.

The vehicle is modeled as a first-order lag system in which the control input is torque $u(t)$ [Nm], and the output is the vehicle velocity $v(t)$ [m/s]. The corresponding

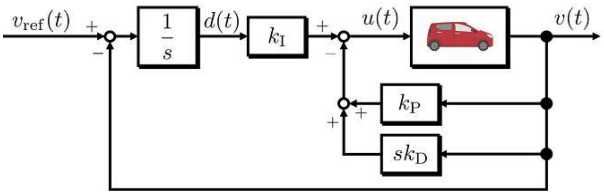

Fig. 1. Block diagram of automatic stop system.

transfer function is

$$P(s) := \frac{1}{s(\tau s + 1)}, \tag{2}$$

where τ is the time constant. Only straight-line motion is considered, and lateral dynamics and interactions with other vehicles are not modeled. Although this model is used throughout the study, the exact value of τ is assumed to be unknown, as it varies depending on vehicle type and operating conditions.

The vehicle position and acceleration are denoted by $x(t)$ [m] and $a(t)$ [m/s^2], respectively. Defining the state vector as $X(t) := [x(t), v(t), a(t)]^\top$, the continuous-time state-space representation becomes:

$$\dot{X}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\frac{1}{\tau} \end{bmatrix} X(t) + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\tau} \end{bmatrix} u(t), \tag{3}$$

$$y(t) := [0, 1, 0] X(t), \tag{4}$$

$$X(0) = [0, v_c, 0]^\top. \tag{5}$$

2.2. Stopping Strategy and Control Law

In **Fig. 1**, $d(t)$ [m] denotes the distance to the stop position d_{stop} at time t, which is defined as

$$d(t) = d_{\mathrm{stop}} - x(t). \tag{6}$$

The required stopping distance is given by:

$$d_{\mathrm{stop}} = -\frac{v_c^2}{2a_{\mathrm{ref}}}, \tag{7}$$

as shown in [17]. When the condition $d(t) \le d_{\mathrm{stop}}$ is satisfied, the system switches to the deceleration mode.

During deceleration, an I-PD controller was used instead of a standard PID controller to suppress sudden changes in the control input [19]:

$$u(t) = k_{\mathrm{I}} d(t) - k_{\mathrm{P}} v(t) - k_{\mathrm{D}} a(t), \tag{8}$$

$$= k_{\mathrm{I}} \int_0^t \left\{ v_{\mathrm{ref}}(\tau) - v(\tau) \right\} d\tau - k_{\mathrm{P}} v(t) - k_{\mathrm{D}} \frac{dv(t)}{dt}, \tag{9}$$

where, $k_{\mathrm{P}}, k_{\mathrm{I}}$, and k_{D} are the proportional, integral, and derivative gains, respectively. The reference velocity $v_{\mathrm{ref}}(t)$ [m/s] was obtained by integrating the reference deceleration a_{ref} [m/s^2], $a_{\mathrm{ref}} < 0$ with the cruise speed v_c.

2.3. FRIT for the Stop Controller

We now describe the use of FRIT to tune the controller parameters $\rho = [k_\mathrm{I}, k_\mathrm{P}, k_\mathrm{D}]^\top$ in Eq. (8). Suppose that experimental driving data over the interval $t = 0$ to $t = t_\mathrm{s}N$ has been collected as $\mathcal{D}_0 := \{u_{0[0,N]}, v_{0[0,N]}\}$. The controller tuning objective is to align the system output $v(t, \rho)$ to the reference response $v_\mathrm{des}(t)$ defined by a desired transfer function $T_\mathrm{des}(s)[v_\mathrm{ref}](t)$. Here, we denote the signal $w(t)$ generated by some system with tunable parameter ρ as $w(t, \rho)$.

The tuning problem can be formulated as:

$$\min_{\rho} \left\| v_\mathrm{des}(t) - v(t, \rho) \right\|. \tag{10}$$

As the solution to this problem, this paper considers the use of FRIT.

FRIT for I-PD type ACC systems [12]

We consider the I-PD control system shown in **Fig. 1**. Given the desired property $T_\mathrm{des}(s)$ and the initial feedback controller $C(\rho^0)$, and the experimental data set $\{u_{0[0,N]}, v_{0[0,N]}\}$, we define $\theta = \theta^*$ as the parameter choice that minimizes the following cost function:

$$J_\mathrm{FRIT}(\theta) := \left\| v_0 - T_\mathrm{des}(s)\left[\tilde{r}(\theta)\right] \right\|_{[0,N]}, \tag{11}$$

where

$$\begin{aligned}\tilde{r}(\theta) :=\ & v_{0[0,N]} + s\theta_1 \left[u_{0[0,N]}\right] \\ & + s\theta_2 \left[v_{0[0,N]}\right] + s^2\theta_3 \left[v_{0[0,N]}\right], \end{aligned} \tag{12}$$

$$\theta := \left[\theta_1, \theta_2, \theta_3\right]^\top = \left[\frac{1}{k_\mathrm{I}}, \frac{k_\mathrm{P}}{k_\mathrm{I}}, \frac{k_\mathrm{D}}{k_\mathrm{I}}\right]^\top. \tag{13}$$

Then, we can get a parameter $\rho = \rho^*$ as

$$\rho^* := \left[\frac{1}{\theta_1}, \frac{\theta_2}{\theta_1}, \frac{\theta_3}{\theta_1}\right]^\top, \tag{14}$$

which can be minimized Eq. (10).

Since J_FRIT is linear in θ, the globally optimal solution can be obtained by least squares:

$$\theta^* := \left(\Phi^\top \Phi\right)^{-1} \Phi^\top \eta, \tag{15}$$

where

$$\begin{aligned}\Phi := \Big[& sT_\mathrm{des}(s)\left[u_{0[0,N]}\right], sT_\mathrm{des}(s)\left[v_{0[0,N]}\right], \\ & s^2 T_\mathrm{des}(s)\left[v_{0[0,N]}\right]\Big], \end{aligned} \tag{16}$$

$$\eta := \left(1 - T_\mathrm{des}(s)\right)\left[v_{0[0,N]}\right]. \tag{17}$$

If any signal has a non-zero initial offset, it is corrected to zero for accurate parameter estimation.

3. Threat Model and Attack Formulation

This section describes the threat model for poisoning attacks in DDC, focusing on industrial controller tuning scenarios, as illustrated in **Fig. 2**. We adopt the "maximum-impact, minimum-resource" attacker model proposed in [20].

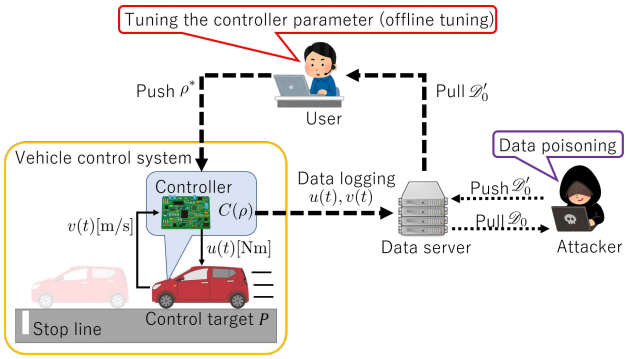

Fig. 2. Conceptual diagram of a poisoning attack against DDC with connected ACC system. The user believes they are accessing the original data $\mathcal{D}_0$, but the stored data has been tampered with by the attacker, resulting in $\mathcal{D}_0'$, which is used unintentionally for controller tuning.

3.1. Settings

We define the system entities and roles of the user and attacker as illustrated in **Fig. 2**.

3.1.1. Control System

The control system consists of a plant P and a controller $C(\theta)$, as shown in **Fig. 2**. The controller regulates the plant using a model reference approach, such as a PID controller implemented on an electronic control unit (ECU). During operation, the system exchanges input–output data to follow a reference signal $v_\mathrm{ref}(t)$. The ECU transmits the operational data $\{u(t), v(t)\}$ to a central data server.

The control objective is to stop appropriately at stop lines specified by law such as those in the Road Traffic Act [a]. In this case, crossing the stop line was not permitted.

3.1.2. Data Server

The data server is located in a control center or cloud environment, and stores the received data as time-series records: $\mathcal{D}_\mathrm{s} := \{\boldsymbol{u}, \boldsymbol{v}\}$. The data can be accessed and modified by authorized operators.

3.1.3. User

The user is responsible for tuning the controller to improve system performance. Without access to the plant model, the user retrieves historical data $\mathcal{D}_0 \subseteq \mathcal{D}_\mathrm{s}$. The user then minimizes the FRIT cost function $J_\mathrm{FRIT}(\boldsymbol{\theta})$ to obtain an optimal parameter set $\boldsymbol{\rho}^*$ using Eqs. (11)–(14). The result $\boldsymbol{\rho}^*$ is uploaded to the controller. This parameter tuning process can be viewed as a mapping from data to parameters: $\mathcal{D}_0 \to \boldsymbol{\rho}^*$.

3.1.4. Attacker

The attacker aims to degrade system performance by tampering with the stored data before it is used for tuning. In recent years, fileless malware has become increasingly prevalent due to its high stealth, often behaving like legitimate operating system processes [21–23]. Such malware can execute DDC-based tuning without requiring explicit knowledge of the desired control specifications.

We assume that the attacker knows the DDC algorithm used by the user; however, like the user, the attacker lacks knowledge of the true plant dynamics. The attacker accesses the data $\mathcal{D}_0$ from the server, modifies it into $\mathcal{D}_0'$, and uploads it before the user performs tuning. Consequently, the user unknowingly tunes the controller using corrupted data. The tampered input and output signals are represented as $\boldsymbol{w}_u \in \mathbb{R}^{N+1}$ and $\boldsymbol{w}_v \in \mathbb{R}^{N+1}$, respectively, and the poisoned dataset is defined as

$$\mathcal{D}_0' := \left\{\boldsymbol{u}_0', \boldsymbol{v}_0'\right\}, \tag{18}$$

$$\boldsymbol{u}_0' := \boldsymbol{u}_0 + \boldsymbol{w}_u, \tag{19}$$

$$\boldsymbol{v}_0' := \boldsymbol{v}_0 + \boldsymbol{w}_v. \tag{20}$$

The attacker's objective is to induce overshoot behavior in an automatic driving vehicle, causing it to pass beyond the legally designated stop line instead of coming to a complete stop at the prescribed position.

3.2. Baseline Attack for ACC

Prior studies, such as [12], assume an attacker who optimizes data perturbations to maximize the FRIT cost function, which is formulated as follows:

Poisoning attack against FRIT [12]

$$\max_{\boldsymbol{w}_u, \boldsymbol{w}_v} \; J_{\mathrm{FRIT}}\left(\hat{\boldsymbol{\theta}}\left(\boldsymbol{w}_u, \boldsymbol{w}_v\right)\right), \tag{21}$$

subject to

$$\hat{\boldsymbol{\theta}}(\boldsymbol{w}_u, \boldsymbol{w}_v) := \left(\hat{\Phi}^\top \hat{\Phi}\right)^{-1} \hat{\Phi}^\top \hat{\eta}, \tag{22}$$

$$\begin{aligned}\hat{\Phi} := \big[&sT_{\mathrm{des}}(s)\left[u_{0[0,N]} + w_{u[0,N]}\right], \\ &sT_{\mathrm{des}}(s)\left[v_{0[0,N]} + w_{v[0,N]}\right], \\ &s^2 T_{\mathrm{des}}(s)\left[v_{0[0,N]} + w_{v[0,N]}\right]\big],\end{aligned} \tag{23}$$

$$\hat{\eta} := \left(1 - T_{\mathrm{des}}(s)\right)\left[v_{0[0,N]} + w_{v[0,N]}\right], \tag{24}$$

$$\max\left|\boldsymbol{w}_{u[0,N]}\right| \le \delta_u, \tag{25}$$

$$\max\left|\boldsymbol{w}_{v[0,N]}\right| \le \delta_v. \tag{26}$$

The poisoned data used by the user becomes:

$$\mathcal{D}_0' := \left\{u'_{0[0,N]}, v'_{0[0,N]}\right\} \tag{27}$$

$$= \left\{u_{0[0,N]} + w^*_{u[0,N]}, v_{0[0,N]} + w^*_{v[0,N]}\right\}, \tag{28}$$

where $w^*_{u[0,N]}$ and $w^*_{v[0,N]}$ are the optimal solutions obtained from Eq. (21). However, this type of attack does not explicitly steer system behavior in a targeted manner, resulting in unpredictable or suboptimal outcomes.

4. Main Result

4.1. Directed Poisoning Attacks on FRIT

We propose a directed poisoning attack, in which the attacker defines an adversarial reference model $T_{\mathrm{a}}(s)$ (e.g., with an undesirable overshoot) and manipulates the tuning process, such that the resulting controller aligns with $T_{\mathrm{a}}(s)$ instead of the original $T_{\mathrm{des}}(s)$.

Replacing $T_{\mathrm{des}}(s)$ with $T_{\mathrm{a}}(s)$ in the FRIT cost function yields:

$$J_{\mathrm{a}}(\theta) := \left\|v_0 - T_{\mathrm{a}}(s)\left[\tilde{r}(\theta)\right]\right\|_{[0,N]}, \tag{29}$$

where $\tilde{r}(\theta)$ is similarly redefined using $T_{\mathrm{a}}(s)$.

The optimal parameter is then obtained to achieve $v(\rho^*) \to T_{\mathrm{a}}(s)[r]$ by minimizing Eq. (29).

Here, we consider obtaining the parameters through data poisoning. Hence, an attacker can solve the following problem:

Directed poisoning attack against FRIT

$$\min_{w_u, w_v} J_{\mathrm{a}}\left(\hat{\theta}(w_u, w_v)\right), \tag{30}$$

subject to:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{w}_u, \boldsymbol{w}_v) := \left(\hat{\Phi}^\top \hat{\Phi}\right)^{-1} \hat{\Phi}^\top \hat{\eta}, \tag{31}$$

$$\begin{aligned}\hat{\Phi} := \big[&sT_{\mathrm{des}}(s)\left[u_{0[0,N]} + w_{u[0,N]}\right], \\ &sT_{\mathrm{des}}(s)\left[v_{0[0,N]} + w_{v[0,N]}\right], \\ &s^2 T_{\mathrm{des}}(s)\left[v_{0[0,N]} + w_{v[0,N]}\right]\big],\end{aligned} \tag{32}$$

$$\hat{\eta} := \left(1 - T_{\mathrm{des}}(s)\right)\left[v_{0[0,N]} + w_{v[0,N]}\right], \tag{33}$$

$$\max\left|\boldsymbol{w}_{u[0,N]}\right| \le \delta_u, \tag{34}$$

$$\max\left|\boldsymbol{w}_{v[0,N]}\right| \le \delta_v. \tag{35}$$

The formulated optimization problem is a min-min problem, where the solution to the inner optimization problem is embedded within the main problem. Due to this hierarchical structure and the presence of constraints, the overall problem becomes nonconvex with respect to the decision variables w_u and w_v. Hence, solving the problem is challenging, and the obtained solution corresponds to a local optimum computed using a nonlinear optimization algorithm.

4.2. Interpretation of the Attack Strategy

Next, we explain why solving the proposed optimization problem produces the desired attack signals. For simplicity, we temporarily ignore tampering-bound constraints. Since the estimated parameter vector $\hat{\theta}(w_u, w_v)$ depends on the tampering signals w_u and w_v, the resulting parameter θ^* changes according to the choice of these signals.

The attacker selects w_u^* and w_v^* to minimize the cost function J_{a}, where J_{a} is defined based on the estimated parameter $\hat{\theta}(w_u, w_v)$ obtained from the tampered data. In this formulation, minimizing J_{a} effectively aligns the

Table 1. Simulation parameters.

Symbol	Value	Unit
a_{ref}	-1.15	m/s^2
τ	0.54	s
v_{c}	8.0	m/s
t_{s}	0.001	s

closed-loop system response with the attacker-defined target model $T_{\mathrm{a}}(s)$, owing to the structure of the FRIT-based cost function.

Therefore, in the absence of constraints, the optimal tampering signals w_u^* and w_v^* yield a parameter $\theta^*(w_u^*, w_v^*)$ (the actual control parameter obtained is $\boldsymbol{\rho}^*$ made by $\boldsymbol{\theta}^*$) such that the resulting system output $v(t, \rho^*(w_u^*, w_v^*))$ approximates the attacker's desired response $T_{\mathrm{a}}(s)[v_{\mathrm{ref}}](t)$.

When tampering constraints are applied, the tracking performance depends on the magnitude of the signals the attacker is allowed to modify. If the tampering bound is large, the system can follow the adversarial reference model $T_{\mathrm{a}}(s)$ more closely, but the attack becomes easier to detect. Conversely, if the tampering bound is small, the attack is harder to notice, but the attacker has less ability to alter the system behavior.

5. Numerical Example

In this section, the adversarial effects of poisoning attacks on FRIT are investigated using two numerical examples. Example 1 compares the proposed method with the previous study in [12] whereas Example 2 analyzes the performance with respect to the tampering widths δ_v and δ_u.

5.1. Common Simulation Setup

The same conditions were used for both simulations. The control system corresponds to that shown in **Fig. 1**, with parameters listed in **Table 1**, based on values reported in [17, 19]. These experiments were conducted using MATLAB/Simulink [b].

Here, t_{s} denotes the sampling period. Sensor noise was modeled as additive white Gaussian noise with zero mean and variance 0.1 applied to each state of the vehicle model.

The desired reference transfer function for FRIT is defined as:

$$T_{\mathrm{des}}(s) = \left(\frac{1}{1.5s+1}\right)^2. \tag{36}$$

FRIT is performed using the least-squares approach.

The attacker designs an adversarial reference model as follows:

$$T_{\mathrm{a}}(s) = \frac{0.2}{s^2 + 0.3s + 0.2}. \tag{37}$$

To generate the poisoning signals, the attacker solves the optimization problem using `fmincon` in MATLAB with an interior-point algorithm.

Table 2. Controller parameters before and after FRIT tuning.

Case	k_{I}	k_{P}	k_{D}
Initial setting	0.50	1.00	1.00
After FRIT (normal)	2.98	8.78	6.57
After FRIT [12]	0.00	0.08	-0.14
After FRIT (proposed)	0.19	0.30	-0.19

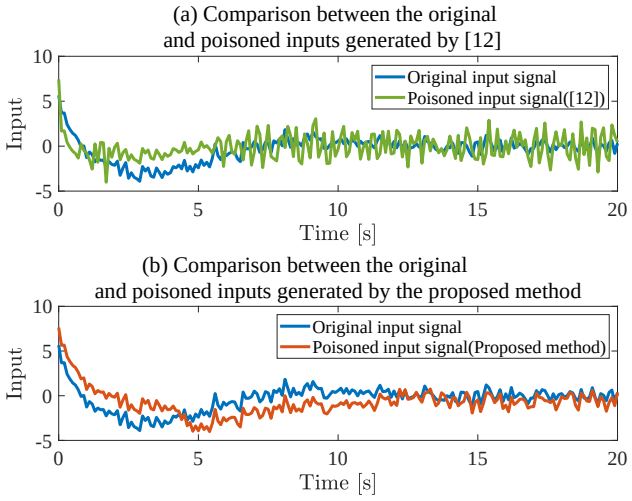

Fig. 3. Comparison of input torque signals (original vs. tampered).

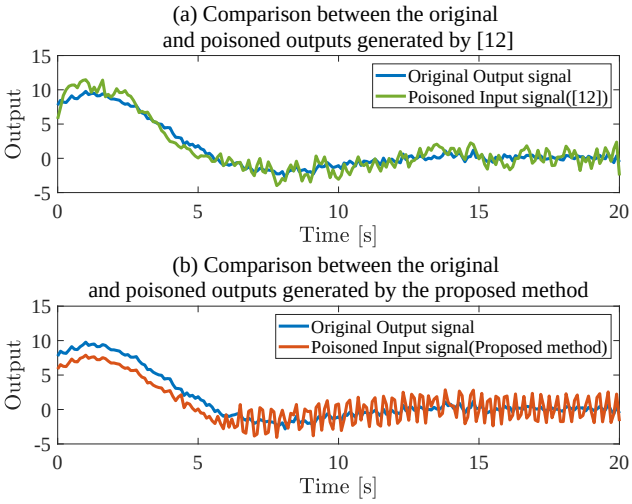

Fig. 4. Comparison of vehicle speed signals (original vs. tampered).

5.2. Example 1

We first evaluated the effectiveness of the proposed method by comparing it to the previous method in [12]. The controller parameters $\boldsymbol{\rho}^0$ used to obtain the experimental data $\{u_{0[0,N]}, v_{0[0,N]}\}$ are listed in **Table 2**. **Figs. 3** and **4** illustrate the original input torque and output speed signals used for controller tuning by FRIT, as well as the signals after poisoning.

In each figure, the original data were obtained using the initial (pre-tuned) control parameter $\boldsymbol{\rho}^0$ (shown in **Table 2**).

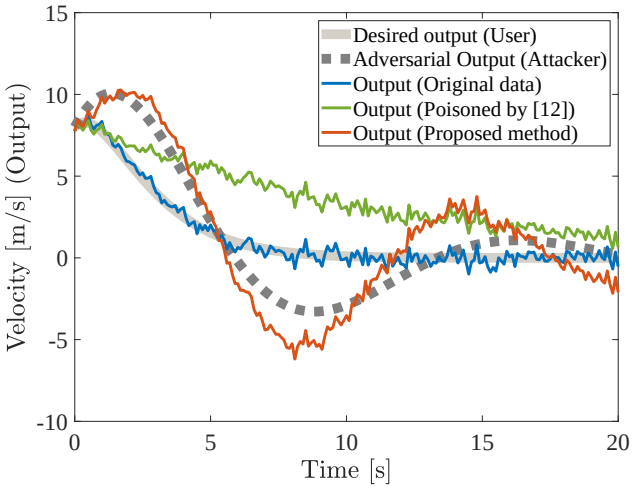

Fig. 5. Comparison of vehicle speed after FRIT tuning.

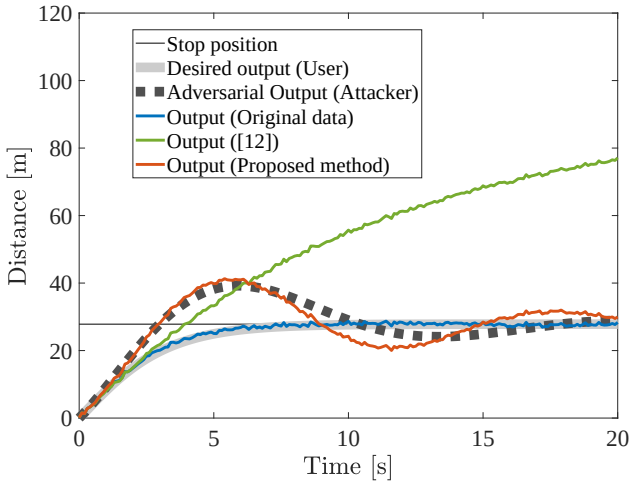

Fig. 6. Comparison of vehicle position relative to the stop line after FRIT tuning.

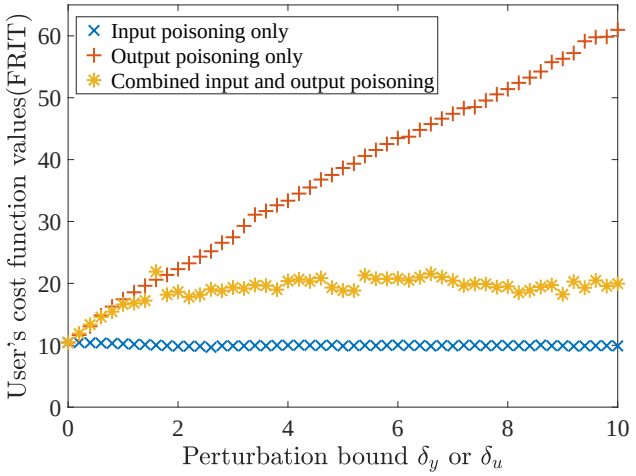

Fig. 7. The cost function value of FRIT.

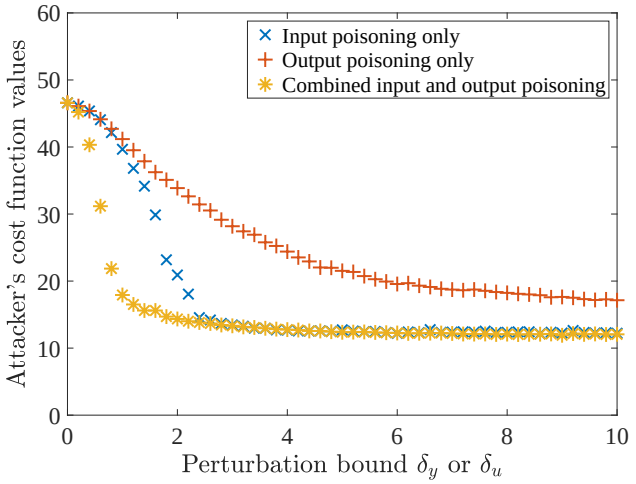

Fig. 8. The attacker's cost function value.

We tampered with the original data using both the proposed method and the previous method, with $\delta_v = \delta_u = 2$.

The tampered input and output signals closely resemble the original signals, indicating that the poisoning attack is subtle and difficult to detect.

We then performed controller tuning using FRIT on both the original and the tampered datasets. The resulting controller parameters are listed in **Table 2**.

The vehicle responses obtained using the tuned controllers are shown in **Figs. 5** and **6**.

As shown in **Fig. 5**, the vehicle using the original controller successfully tracked the desired output without overshooting. In contrast, the controller tuned using the tampered dataset exhibited a significant overshoot. The output obtained using the proposed method closely matches the attacker's adversarial reference trajectory $v_a(t)$. As shown in **Fig. 6**, this overshoot caused the vehicle to stop beyond the designated stop line. In the presented example, the vehicle was cruising at 36 km/h and stopped approximately 12 m beyond the intended stop position. This level of overshoot not only violates traffic rules related to stop-line compliance but is also sufficient to enter critical zones such as crosswalks at intersections or railway crossings, poses a serious risk of accidents.

Next, we compared these results with those obtained using an existing method [12]. In the case of the existing method, the tampering was successful and caused a significant overshoot. However, the attacker's objective in this approach is to maximize the FRIT evaluation error, which does not necessarily produce the specific system behavior the attacker intends. By contrast, the proposed method explicitly considers the attacker's desired adversarial response as the target. In this experiment, we confirmed that the system behavior closely followed the attacker-defined reference characteristics, demonstrating the effectiveness of the proposed approach.

5.3. Example 2

Next, we inverted the width of the data tampering. We conducted the directed poisoning attack by changing δ_v and δ_u from 0.2 to 10 in increments of 0.2.

The results are presented in **Figs. 7** and **8**. The parameters calculated using Eq. (11) with the poisoned data gen-

erated by the proposed method, along with the attacker's cost function value Eq. (30) are shown in **Figs. 7** and **8**, respectively.

Figures 7 and **8** illustrate how FRIT and the attacker's cost function respond to varying tampering widths. From **Fig. 7**, we observe that tampering only with the input signal has little effect on the FRIT's cost value, whereas tampering only with the output signal results in a larger increase in the cost value, which is proportional to the tampering width. In both the input-only and output-only cases, the FRIT's cost function value eventually levels off.

In **Fig. 8**, we analyze the changes in the attacker's cost function. The cost decreases more quickly when only the input signal is tampered with than when only the output signal is tampered with. Moreover, tampering with both the input and output signals simultaneously leads to the fastest reduction in cost value.

These findings indicate that the FRIT of the I-PD system is more sensitive to input tampering. Furthermore, the joint modification of both signals can enhance the effectiveness of the attack, particularly when the allowed tampering range is small.

6. Conclusion

In this study, we examined how poisoning attacks can affect controller tuning using FRIT, focusing on stop control in automotive ACC systems. We proposed a directed poisoning attack that forces the system to behave in a specific manner chosen by the attacker. In our simulations, this attack caused the vehicle to stop late by making the system follow an adversarial reference model. Such attacks are particularly dangerous in stop control, where precise stopping is critical for safety. For example, if a car overshoots a stop point, it may enter a crosswalk or intersection, creating a risk of accident. Although this study focused on ACC as an example, the proposed method is not limited to automotive systems. The same concept can be applied to any system using least-squares-based FRIT, making these findings relevant for a wide range of DDC applications. This highlights the importance of implementing protection mechanisms when deploying DDC in real systems.

In future work, we plan to test the proposed attack in real vehicle environments, investigate the sensitivity of system performance to the attacker's design, and develop methods for detecting and preventing such attacks.

Acknowledgments

This research was supported by the JST CREST Grant Number JPMJCR23M4. The authors gratefully acknowledge this support.

References:

[1] D. Bhamare, M. Zolanvaric, A. Erbad, R. Jain, K. Khan, and N. Meskin, "Cybersecurity for industrial control systems: A survey," Computers & Security, Vol.89, Article No.101677, 2020. https://doi.org/10.1016/j.cose.2019.101677

[2] K. Sawada, "Model-based cybersecurity for control systems: Modeling, design and control," Proc. of the 2017 56th Annual Conf. of the Society of Instrument and Control Engineers of Japan (SICE), pp. 724-727, 2017. https://doi.org/10.23919/SICE.2017.8105750

[3] O. Kaneko, "Introduction to Data-Driven Control," Corona Publishing Co., 2024 (in Japanese).

[4] H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin, "Iterative Feedback Tuning: Theory and Applications," IEEE Control Systems Magazine, Vol.18, No.4, pp. 26-41, 1998. https://doi.org/10.1109/37.710876

[5] M. C. Campi, A. Lecchini, and S. M. Savaresi, "Virtual Reference Feedback Tuning: A Direct Method for the Design of Feedback Controllers," Automatica, Vol.38, No.8, pp. 1337-1346, 2002. https://doi.org/10.1016/S0005-1098(02)00032-8

[6] T. Ikezaki and O. Kaneko, "A New Approach of Data-Driven Controller Tuning Method by Using Virtual IMC Structure—Virtual Internal Model Tuning—," Proc. of the 13th IFAC Workshop on Adaptive and Learning Control Systems (ALCOS 2019), pp. 344-349, 2019. https://doi.org/10.1016/j.ifacol.2019.12.699

[7] R. Yamamoto and O. Kaneko, "Application and experimental verification of FRIT to vehicle steering systems," Proc. of the IEEJ Conf. on Electronics, Information and Systems, pp. 1192-1195, 2022 (in Japanese).

[8] M. Kozui, T. Yamamoto, M. Akiyama, K. Koiwai, and Y. Yamazaki, "Application of a MIMO-PID Controller for a Hydraulic Excavator Considering the Velocity of CoM," J. Robot. Mechatron., Vol.32, No.3, pp. 643-651, 2020. https://doi.org/10.20965/jrm.2020.p0643

[9] H. Si and O. Kaneko, "FRIT of Internal Model Controllers for Poorly Damped Linear Time Invariant Systems: Kautz Expansion Approach," J. Robot. Mechatron., Vol.28, No.5, pp. 745-751, 2016. https://doi.org/10.20965/jrm.2016.p0745

[10] A. Russo and A. Proutiere, "Poisoning attack against data-driven control methods," Proc. of the American Control Conf. (ACC), pp. 3234-3241, 2021. https://doi.org/10.23919/ACC50511.2021.9482992

[11] T. Ikezaki, O. Kaneko, K. Sawada, and J. Fujita, "Poisoning attack on VIMT and its adverse effect," Artificial Life and Robotics, Vol.29, pp. 168-176, 2024. https://doi.org/10.1007/s10015-023-00914-7

[12] T. Ikezaki, K. Sawada, and O. Kaneko, "A Study of Data-Driven Control and Poisoning Attack for Vehicle Cruise Control Systems," Proc. of the 11th Multi-Symp. on Control Systems (MSCS2024), 3A6-2, 2024 (in Japanese).

[13] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar, "Can machine learning be secure?," Proc. of the 2006 ACM Symp. on Information, Computer and Communications Security, pp. 16-25, 2006. https://doi.org/10.1145/1128817.1128824

[14] B. Biggio, B. Nelson, and P. Laskov, "Poisoning attacks against support vector machines," Proc. of the 29th Int. Conf. on Machine Learning (ICML), pp. 1467-1474, 2012.

[15] A. Russo, M. Molinari, and A. Proutiere, "Data-driven control and data-poisoning attacks in buildings: The KTH Live-In Lab case study," Proc. of the 29th Mediterranean Conf. on Control and Automation (MED), pp. 53-58, 2021. https://doi.org/10.1109/MED51440.2021.9480238

[16] H. Sasahara, "Adversarial attacks to direct data-driven control for destabilization," Proc. of the IEEE Conf. on Decision and Control (CDC), pp. 7094-7099, 2023. https://doi.org/10.1109/CDC49753.2023.10383531

[17] P. Raksincharoensak, K. Tsuchiya, A. Yamasaki, H. Mouri, and M. Nagai, "Study on automated driving system for two-stage stop and start operation for intersection collision avoidance in unsignalized intersections," Trans. of the JSME, Vol.82, No.834, Article No.15-00475, 2016 (in Japanese). https://doi.org/10.1299/transjsme.15-00475

[18] T. Fujimoto, K. Sawada, Y. Minami, and K. Sando, "Filtering Function to Mitigate the Impact of Cyber Attacks in Cooperative Adaptive Cruise Control," J. Robot. Mechatron., Vol.36, No.3, pp. 669-679, 2024. https://doi.org/10.20965/jrm.2024.p0669

[19] T. Fujimoto, H. Matsushita, K. Sawada, and K. Yamafuji, "Design of ACC considering sensor error using predictive governor," Proc. of the 66th Annual Conf. of the Institute of Systems, Control and Information Engineers (SCI'22), Article No.342-4, 2022 (in Japanese).

[20] 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. https://doi.org/10.1109/MCS.2014.2364709

[21] Sudhakar and S. Kumar, "An emerging threat: Fileless malware-a survey and research challenges," Cybersecurity, Vol.3, Article No.1, 2020. https://doi.org/10.1186/s42400-019-0043-x

[22] S. Liu, G. Peng, H. Zeng, and J. Fu, "A survey on the evolution of fileless attacks and detection techniques," Computers & Security, Vol.137, Article No.103653, 2024. https://doi.org/10.1016/j.cose.2023.103653
[23] J. Lee and S. Hong, "Host-Oriented Approach to Cyber Security for the SCADA Systems," Proc. of the 2020 6th IEEE Congress on Information Science and Technology (CiSt), pp. 151-155, 2020. https://doi.org/10.1109/CiSt49399.2021.9357299

Supporting Online Materials:

[a] "Road Traffic Act" (in Japanese). https://www.japaneselawtranslation.go.jp/ja/laws/view/2962 [Accessed June 6, 2025]
[b] "The MathWorks, Inc. Website" (in Japanese). https://jp.mathworks.com/ [Accessed June 6, 2025]

Name:
Taichi Ikezaki

Affiliation:
Assistant Professor, Faculty of Environmental, Life, Natural Science and Technology, Okayama University

Address:
3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan

Brief Biographical History:
2023- Assistant Professor, Okayama University.

Main Works:
• T. Ikezaki, O. Kaneko, K. Sawada, and J. Fujita, "Poisoning attack on VIMT and its adverse effect," Artificial Life and Robotics, Vol.29, pp. 168-176, 2024.

Membership in Academic Societies:
• Institute of Electrical and Electronics Engineers (IEEE)
• The Society of Instrument and Control Engineers (SICE)
• The Institute of Electrical Engineers of Japan (IEEJ)

Name:
Kenji Sawada

Affiliation:
Professor, Graduate School of Engineering, The University of Osaka

Address:
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Brief Biographical History:
2009-2015 Assistant Professor, The University of Electro-Communications
2015-2025 Associate Professor, Info-Powered Energy System Research Center, The University of Electro-Communications
2025- Professor, The University of Osaka

Main Works:
• Y. Mochizuki and K. Sawada, "Demand-for Graph and Its State Transition Expression Evaluating Traffic Congestion Due to CAVs Control," IEEE Access, Vol.12, pp. 139837-139849, 2024.

Membership in Academic Societies:
• Institute of Electrical and Electronics Engineers (IEEE)
• The Society of Instrument and Control Engineers (SICE)
• The Institute of Systems, Control and Information Engineers (ISCIE)

Name:
Osamu Kaneko

Affiliation:
Professor, Graduate School of Informatics and Engineering, The University of Electro-Communications

Address:
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

Brief Biographical History:
1994-1996 Ishikawajima System Technology Co., Ltd.
1996-1999 Graduate School of Engineering Science, Osaka University
1996-2007 Research Associate, Osaka University
2007-2009 Assistant Professor, Osaka University
2009-2015 Associate Professor, Kanazawa University
2015- Professor, The University of Electro-Communications

Main Works:
• Y. Tanaka and O. Kaneko, "Mode Determination in the Data Informativity," Trans. of the Society of Instrument and Control Engineers, Vol.61, No.2, pp. 55-64, 2025 (in Japanese).

Membership in Academic Societies:
• Institute of Electrical and Electronics Engineers (IEEE)
• The Society of Instrument and Control Engineers (SICE)
• The Institute of Electrical Engineers of Japan (IEEJ)