Interaction-Aware Parameter Privacy-Preserving Data Sharing in Coupled Systems via Particle Filter Reinforcement Learning

Haokun Yu YUHAOKUN@U.NUS.EDU
Institute of Operations Research and Analytics, National University of Singapore, Singapore 117602

Jingyuan Zhou JINGYUANZHOU@U.NUS.EDU
Department of Civil and Environmental Engineering, National University of Singapore, Singapore 119077

Kaidi Yang KAIDI.YANG@NUS.EDU.SG
Department of Civil and Environmental Engineering, National University of Singapore, Singapore 119077

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

This paper addresses the problem of parameter privacy-preserving data sharing in coupled systems, where a data provider shares data with a data user but wants to protect its sensitive parameters. The shared data affects not only the data user's decision-making but also the data provider's operations through system interactions. To trade off control performance and privacy, we propose an interaction-aware privacy-preserving data sharing approach. Our approach generates distorted data by minimizing a combination of (i) mutual information, quantifying privacy leakage of sensitive parameters and (ii) the impact of distorted data on the data provider's control performance, considering the interactions between stakeholders. The optimization problem is formulated into a Bellman equation and solved by a particle filter reinforcement learning (RL)-based approach. Compared to existing RL-based methods, our formulation significantly reduces history dependency and efficiently handles scenarios with continuous state space. Validated in a mixed-autonomy platoon scenario, our method effectively protects sensitive driving behavior parameters of human-driven vehicles (HDVs) against inference attacks while maintaining negligible impact on fuel efficiency. Detailed proofs and experiment setup can be found in this supplementary material.

Keywords: Data privacy; parameter privacy; information theory; coupled system; data sharing

1. Introduction

In the era of big data, data sharing between stakeholders has become a cornerstone for improving the operations of various cyber-physical systems such as transportation (Zheng et al., 2018) and power systems (Larsen et al., 2014). However, since operational data can contain sensitive information about customers and system operations, data sharing can raise significant privacy concerns. For example, although trajectory data of a ride-hailing company can enhance transportation practice and research, it contains sensitive information about individual mobility patterns (e.g., destinations and routing preferences) and about the company's sensitive operational parameters. Such privacy concerns can deter stakeholders from sharing data, thereby undermining the value of big data.

Despite the recent development of privacy-preserving data-sharing mechanisms, such as differential privacy (DP) (Shokri, 2014; Zhao et al., 2014; Chen et al., 2023), most research focuses on protecting individual-level data, while overlooking the privacy of sensitive parameters. However, sensitive parameters, often treated as business secrets, serve as one of the key reasons that can prevent stakeholders from data-sharing. For example, the algorithmic parameters of ride-hailing

companies, if leaked, can compromise the competitive advantages of the company. In control systems, the leakage of feedback gains (Nekouei et al., 2021) and internal states (Nekouei et al., 2022; Weng et al., 2023) can allow malicious actors to exploit vulnerabilities, causing system failures. These examples underscore the critical need to protect sensitive parameters, as their exposure can result in operational disruptions, competitive disadvantages, and security threats.

The research considering parameter privacy is sparse. Popular techniques such as differential privacy and k-anonymity seek to protect individual privacy by hiding them in a herd. Nevertheless, there is no herd to hide when it comes to protecting the parameter privacy. To address this issue, information-theoretic methods have attracted increasing attention, which use metrics like mutual information and entropy to precisely quantify the privacy leakage of parameters by sharing data. For example, Bassi et al. (2018) uses stochastic kernels to protect statistical properties involving private parameters, and Ziemann and Sandberg (2020) designs Gaussian-based privacy filters for linear Gaussian systems to prevent inference of system dynamics. However, these methods can be computationally inefficient in high-dimensional spaces. To address this, Nekouei et al. (2022) proposes a privacy filter using nonlinear transformations that retain the measurement distribution family, though it doesn't guarantee optimality in privacy metrics. More recently, Weng et al. (2023) introduces an approach that formulates privacy protection as a dynamic optimization problem, improving parameter privacy in dynamic systems through adaptive, state-dependent strategies.

Nevertheless, existing methods face two major limitations. First, the frameworks in Erdemir et al. (2020); Weng et al. (2023) can be inefficient in handling systems with high-dimensional or continuous state space, as they involve enumeration of past observations, which becomes exponentially complex with increasing state dimensions. Second, these methods generally assume that the data provider's system operations are independent of its data-sharing mechanisms, implying that distorted data has no impact on the provider's system. However, in reality, data sharing often occurs between coupled systems, where the data provider's operations are influenced by the decisions of the data user, which depend on the shared data. For instance, mobility data shared with traffic authorities may lead to changes in traffic signal settings, affecting the mobility provider's operations.

Statement of Contribution. This paper addresses the aforementioned challenges by proposing an interaction-aware parameter privacy-preserving data-sharing method. We make two main contributions. First, we propose an interaction-aware privacy-preserving approach based on information theory to protect operational parameters from inference attacks in a coupled system. Our approach generates distorted data by minimizing a combination of (i) mutual information that quantifies the privacy leakage of sensitive parameters and (ii) the impact of distorted data on the data provider's control performance, considering the interactions between stakeholders. Our approach can successfully balance privacy preservation with control performance. Second, we formulate the optimization problem into a Bellman equation and propose a particle filter reinforcement learning (RL)-based approach to solve the formulated optimization problem. Compared to existing RL-based methods (Weng et al., 2023; Erdemir et al., 2020; Zhang et al., 2022), our formulation significantly reduces history dependency and efficiently handles scenarios with continuous state space.

2. Problem Statement

We consider the data sharing between a data provider (denoted by A) and a data user (denoted by B) over a discretized time period $\mathcal{T} = \{0, 1, \cdots, T\}$ with time interval ΔT. The operations of both parties are modeled as discrete-time Markov systems. Specifically, the states of the data provider

and the data user at time step t are denoted by $X_t^A \in \mathbb{R}^{N_A}$ and $X_t^B \in \mathbb{R}^{N_B}$, respectively. The state transition of A follows $p(X_{t+1}^A \mid X_t^A, U_t^A)$, where $U_t^A \in \mathbb{R}^{M_A}$ is an action determined by the policy $p(U_t^A \mid \Theta, X_t^A, W_t)$. Here, $\Theta \in R^\theta$ is a sensitive, time-invariant parameter treated as business secrets, and W_t is the external input from B as a function of X_t^B. The state transition of B follows $p(X_{t+1}^B \mid X_t^B, U_t^B)$, where the action $U_t^B \in \mathbb{R}^{M_B}$ is determined by the policy $p(U_t^B \mid X_t^B, Y_t)$. To support B's decision-making, A agrees to share information about its state X_t^A. Nevertheless, to protect the sensitive parameter Θ, A employs a data-sharing policy π_t to generate a slightly distorted version of the data Y_t. Such distortion safeguards Θ from being inferred by B. Let Π denote the set of feasible policies. The interaction dynamics are illustrated in Fig. 1.

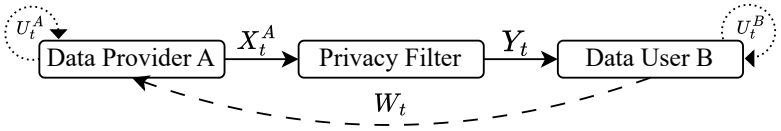


Figure 1: Interaction Between System A and System B.

We make the following assumptions regarding the knowledge of the data provider and the data user. First, all conditional probabilities represented by $p(\cdot \mid \cdot)$ are assumed to be public knowledge. This allows us to simplify the system transitions of A and B to $p(X_{t+1}^A \mid \Theta, X_t^A, X_t^B)$ and $p(X_{t+1}^B \mid X_t^B, Y_t)$, respectively. Second, data user B is assumed to have prior knowledge about Θ, represented by a prior distribution $p(\Theta)$, but does not know its true value θ^*. Three, we conservatively assume that data user B has full knowledge of the data-sharing policy π_t at any time t. If B does not know π_t, the privacy protection of our mechanism is only further enhanced.

We aim to design a data-sharing policy π_t at any time t for data provider A that satisfies three key requirements: (i) protecting the privacy of the sensitive parameter Θ, (ii) maintaining control performance of data provider A , and (iii) ensuring the usability of the shared data. We make the following remarks regarding these requirements. First, continuous access to shared data allows data user B to update its belief $\beta_t(\Theta)$ on Θ via Bayesian inference, so the policy must prevent B from inferring the exact value of Θ over time. Second, due to the interactions between A and B, the distorted data Y_t generated by A will influence the future states of B and, in turn, the future states of A. Therefore, the data-sharing policy π can affect the system operations of the data provider A. Three, to ensure meaningful collaboration, the shared data must remain useful to B, requiring that the deviation between Y_t and X_t^A be upper bounded.

To satisfy these requirements, we formulate the data-sharing policy design into the following optimization problem

$$\min_{\boldsymbol{\pi}=\{\pi_t\}_{t=1}^T} \rho I^{\boldsymbol{\pi}}(\Theta; Y_{1:T}, X_{1:T}^B) + \sum_{t=1}^{T} \mathbb{E}^{\boldsymbol{\pi}}\left[r(\theta^*, X_t^A, X_{t+1}^A, Y_t)\right] \tag{1a}$$

$$\text{s.t.} \quad \mathbb{E}^{\pi_t}[d(X_t^A, Y_t)] \le \hat{D}, \quad t = 1, 2, \cdots, T, \tag{1b}$$

where the objective function (1a) includes a combination of a privacy measure and system performance measure, weighted by ρ. The first term is the privacy measure represented by the mutual information (MI) between the sensitive parameter Θ and $(Y_{1:T}, X_{1:T}^B)$, i.e., the data sequence shared by A and the state data sequence of B. The MI quantifies the amount of information about Θ contained in $(Y_{1:T}, X_{1:T}^B)$. The second term represents the system cost over the data-sharing period, where $r(\theta^*, X_t^A, X_{t+1}^A, Y_t)$ quantifies the impact of the distorted data Y_t when transitioning from

state X_t^A to X_{t+1}^A, considering the true parameter θ^*. Constraint (1b) ensures that at each time step t, the expected distortion $d(X_t^A, Y_t)$ between the true state X_t^A and the distorted state Y_t under policy π_t is upper-bounded by $\hat{D}$.

It is challenging to compute $I(\Theta; Y_{1:T}, X_{1:T}^B)$ due to the high dimensionality of variables and the complexity of estimating joint distributions of $(Y_{1:T}, X_{1:T}^B)$ across the entire data-sharing period. However, as proved in Appendix A, $I(\Theta; Y_{1:T}, X_{1:T}^B)$ can be simplified as (2) by applying the chain rule of MI and considering the dependencies between variables:

$$I^{\boldsymbol{\pi}}(\Theta; Y_{1:T}, X_{1:T}^B) = \sum_{t=1}^{T} I^{\boldsymbol{\pi}}(\Theta; Y_t \mid Y_{1:t-1}, X_{1:t}^B). \tag{2}$$

3. Characterization of Interaction-Aware Privacy-Preserving Data-Sharing Policy

In this section, we devise an interaction-aware privacy-preserving data-sharing policy of data provider A by solving optimization problem (1). A general form of the data-sharing policy is given by $\pi_t(y_t \mid \Theta, X_{1:t}^A, Y_{1:t-1}, X_{1:t}^B)$, which generates the distorted data Y_t given the parameter Θ, histories of observations $X_{1:t}^A$ and $X_{1:t}^B$, and the history of shared data $Y_{1:t-1}$. Such a general form exploits all information available at time t, making the solution of Eq. (1) intractable due to the curse of dimensionality. To address this issue, we equivalently simplify this policy such that it only depends on the current system state, the sensitive parameter, and the belief state.

We first show that the optimal policy for optimization problem (1) can be simplified to $\pi_t^s(Y_t \mid \Theta, X_t^A, X_{1:t}^B, Y_{1:t-1})$, eliminating dependency on A's past states. Let $\boldsymbol{\pi^s} = \{\pi_t^s\}_{t=1}^T$ be the sequence of simplified policies over the entire time horizon and $\boldsymbol{\Pi^s}$ the set of all feasible simplified policies. Such a simplification is formalized in Theorem 1 and proved in Appendix B.1.

Theorem 1 *There exists a sequence of policy $\boldsymbol{\pi}^{\boldsymbol{s},*} \in \boldsymbol{\Pi}^{\boldsymbol{s}}$ such that $\boldsymbol{\pi}^{\boldsymbol{s},*}$ is an optimal solution to the optimization problem (1). Specifically, let $L(\cdot)$ denote the optimization problem (1) and $\boldsymbol{\pi}^*$ be the optimal sequence of policies, we have*
(i) $L(\boldsymbol{\pi}^{\boldsymbol{s},*}) = L(\boldsymbol{\pi}^*)$,
(ii) $\mathbb{E}^{\pi_t^{s,*}}[d(X_t^A, Y_t)] \leq \hat{D}, \quad i = 1, 2, \cdots, T.$

According to Theorem 1, the optimal simplified policy can be obtained from the following optimization problem

$$\min_{\boldsymbol{\pi^s}} \sum_{t=1}^{T} \left(\rho I^{\boldsymbol{\pi^s}}(\Theta; Y_t \mid Y_{1:t-1}, X_{1:t}^B) + \mathbb{E}^{\boldsymbol{\pi^s}}[r(\theta^*, X_t^A, X_{t+1}^A, Y_t)] + \lambda\Big(\mathbb{E}^{\boldsymbol{\pi^s}}[d(X_t^A, Y_t)] - \hat{D}\Big)\right) \tag{3}$$

where λ is the Lagrangian multiplier. To solve optimization problem (3), we introduce the cost-to-go function $V_t(h_t)$, which represents the minimum expected cost starting at time t given the history $h_t = (x_{1:t}^B, y_{1:t-1})$. This function $V_t(h_t)$ allows us to recursively express the optimization problem and solve it using dynamic programming. Specifically, the optimal cost-to-go at time t is defined as

$$V_t^*(h_t) = \min_{\{\pi_k^s\}_{k=t}^T} \Bigg[\sum_{k=t}^{T} \Bigg(\rho I^{\boldsymbol{\pi^s}}(\Theta; Y_k \mid Y_{1:k-1}, X_{1:k}^B) + \mathbb{E}^{\boldsymbol{\pi^s}}[r(X_k^A, X_{k+1}^A, Y_k, \theta^*)] + \lambda\Big(\mathbb{E}^{\boldsymbol{\pi^s}}[d(X_k^A, Y_k)] - \hat{D}\Big)\Bigg) \;\Big|\; h_t \Bigg] \tag{4}$$

with terminal condition $V^*_{T+1}(\cdot) = 0$. Accordingly, the Bellman optimality equation of (3) can be represented as

$$V^*_t(h_t) = \min_{\pi^s_t(Y^t|\Theta,X^A_t,h_t)} \left[C_t(h_t, \pi^s_t(Y_t \mid \Theta, X^A_t, h_t)) + \mathbb{E}\left(V^*_{t+1}(h_{t+1} \mid h_t)\right)\right], \tag{5}$$

where $C_t(h_t, \pi^s_t(Y^t \mid \Theta, X^A_t, h_t))$ represents the immediate cost when employing data-sharing policy set $\pi^s_t(Y^t \mid \Theta, X^A_t, h_t)$, with detailed formulation in Appendix B.3.

Although the simplified policy π^s reduces complexity by eliminating dependence on A's past states $X^A_{1:t-1}$, it still relies on the histories h_t. This dependence complicates the optimization problem as the time horizon T increases. To overcome this challenge and efficiently generate the policy set π^s while utilizing information from the histories h_t, we propose an alternative formulation. We encode the histories into a belief state β_t, which condenses all past observational information into a probability distribution over the possible current system states $\beta_t(\Theta, X^A_t) = p(\Theta, X^A_t \mid h_t)$. The belief state is updated recursively using Bayes' rule whenever new observations become available. Specifically, upon receiving the distorted data y_t at time t, the data user updates the belief state from β_t to β_{t+1} using the following formula:

$$\beta_{t+1}(\Theta, X^A_{t+1}) = \frac{\int_{x^A_t} p(X^A_{t+1} \mid \Theta, x^A_t, x^B_t) a(y_t \mid \Theta, x^A_t) \beta_t(\Theta, x^A_t) dx^A_t}{\int_{x^A_t, x^A_{t+1}, \theta} p(x^A_{t+1} \mid \theta, x^A_t, x^B_t) a(y_t \mid \theta, x^A_t) \beta_t(\theta, x^A_t) d\theta dx^A_{t+1} dx^A_t}, \tag{6}$$

Where $a(y_t \mid \Theta, x^A_t) = \pi^s_t(Y_t \mid \Theta, x^A_t, X^B_{1:t}, Y_{1:t-1})$.

For a detailed derivation of this update rule, please refer to the supplementary material, Appendix B.2.

According to (6), we can see the belief state β_t relies solely on the observations Y_t, policy set π^s_t and the prior belief β_t. Then, similar to the idea in Weng et al. (2023), we establish Lemma 2 to show that the optimal cost-to-go function, derived using the belief state β_t, is equivalent to that derived using the full history h_t. The proof is in the supplementary material (Appendix B.3).

Lemma 2 *At any time step t, given histories h_t, the optimal cost-to-go function $V^*_t(h_t)$ depends only on the belief state β_t, i.e., $V^*_t(h_t) = V^*_t(\beta_t)$*

Based on Lemma 2, given histories h_t, the data-sharing policy $\pi^s_t(Y_t \mid \Theta, X^A_t, h_t)$ defines a Markov kernel $\mathcal{K}_t$ that maps the state variables (Θ, X^A_t) to a probability distribution over Y_t. For any measurable spaces (Θ, X^A_t) and Y_t, $\mathcal{K}_t$ assigns a conditional distribution to Y_t based on π^s_t. Accordingly, the Bellman optimality equation can be expressed as:

$$V^*_t(\beta_t) = \min_{\mathcal{K}_t} \left[C_t(\beta_t, \mathcal{K}_t) + \mathbb{E}[V^*_{t+1}(\beta_{t+1}) \mid h_t]\right], \tag{7}$$

which is derived from (5) by replacing h_t by β_t and π^s_t by $\mathcal{K}_t$.

We deduce from (7) that the optimal Markov kernel $\mathcal{K}^*_t$ depends only on the belief state β_t, i.e., $\mathcal{K}^*_t = \arg\min V^*_t(\beta_t)$. Therefore, we consider $\mathcal{K}_t$ as a function of β_t, denoted by $\mathcal{K}^{\beta_t}_t$. Consequently, the data-sharing policy can be determined solely by β_t, Θ, and X^A_t, summarized in Theorem 3.

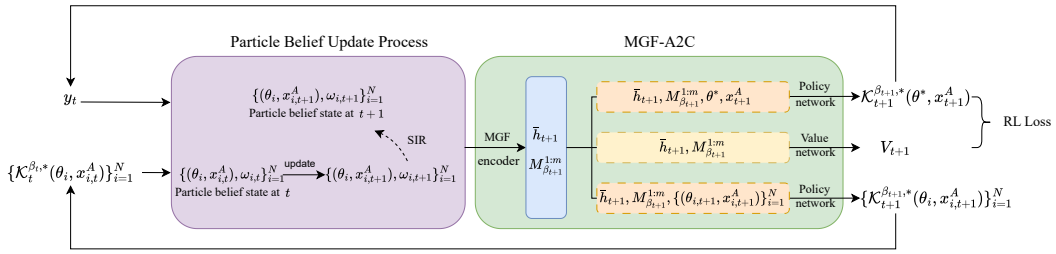


Figure 2: Overview of the proposed privacy-preserving data-sharing framework.

Theorem 3 *Let $\mathcal{K}^{\beta_t,*}_t$ denote the solution to Eq. (7). Given histories h_t, the optimal data distortion policy satisfies:*

$$\pi^{s,*}_t(Y_t \mid \Theta, X^A_t, h_t) = \mathcal{K}^{\beta_t,*}_t, \tag{8}$$

where β_t is the belief state associated with h_t.

Theorem 3 simplifies the data-sharing policy by demonstrating that it is fully determined by the current state (Θ, X^A_t) and the belief state β_t. Therefore, for brevity, we use $\mathcal{K}^{\beta_t,*}_t(Y_t \mid \Theta, X^A_t)$ to represent the policy $\pi^{s,*}_t(Y_t \mid \Theta, X^A_t, h_t)$ in the rest of the article.

4. Particle Filter Reinforcement Learning-Based Solution Method

In this section, we devise a solution method to solve the interaction-aware privacy-preserving data-sharing policy characterized in Section 3. Unlike existing works (Weng et al., 2023; Erdemir et al., 2020) that focus on low-dimensional discrete state space, our method combines particle filters (PF) and reinforcement learning (RL) to handle continuous state space with relatively high dimensions. As shown in Fig. 2, we employ a PF to characterize the evolution of the belief state β_t, which is then processed by a Particle Advantage Actor-Critic (A2C) algorithm with a Moment-Generating Function (MGF) encoder, hereafter named MGF-A2C, to generate the data-sharing policy.

4.1. Particle formulation of belief state

The belief state β_t is characterized by the posterior distribution of (Θ, X^A_t), which is challenging to compute for continuous (Θ, X^A_t). To address this challenge, we describe the evolution of the belief state β_t, represented by (6), using a PF, which approximates the posterior distribution of (Θ, X^A_t) with weighted sampled particles. As the system evolves, particles move according to the dynamic model, and their weights are updated based on how well they match new observations.

Specifically, we use a set of N weighted particles $\{(\theta_i, x^A_{i,t}), \omega_{i,t}\}_{i=1}^N$ to estimate β_t. Each particle $(\theta_i, x^A_{i,t})$ represents a sampled possible state from $p(\Theta, X^A_t \mid Y_{1:t}, X^B_{1:t})$ at time t, and $\omega_{i,t}$ is the associated weight, reflecting the importance of this particle, with the condition $\sum_{i=1}^N \omega_{i,t} = 1$. Thus, the estimated belief state $\hat{\beta}_t$ is given by the set $\hat{\beta}_t = \{(\theta_i, x^A_{i,t}), \omega_{i,t}\}_{i=1}^N$. The weight update process can be formalized in Theorem 4. The proof is in Appendix C in the supplementary material.

Theorem 4 *In the PF approximation of the belief state $\beta_t(\Theta, X^A_t)$, the weights $\omega_{i,t}$ can be updated recursively using the observation likelihood as*

$$\omega_{i,t} \propto \omega_{i,t-1} p(X^B_t \mid X^B_{t-1}, y_{t-1}) p(y_{t-1} \mid \theta_i, x^A_{i,t-1}, X^B_{1:t-1}, Y_{1:t-2}). \tag{9}$$

These updated weights are normalized as $\omega_{i,t} = \frac{\omega_{i,t}}{\sum_{j=1}^{N} \omega_{j,t}}$ to ensure that they form a valid probability distribution. Accordingly, we can easily derive Lemma 5, which further simplifies Theorem 4 to be independent of the dynamics of system B.

Lemma 5 *The weight update* $\omega_{i,t} \propto \omega_{i,t-1} p(X_t^B \mid X_{t-1}^B, y_{t-1}) p(y_{t-1} \mid \theta_i, x_{i,t-1}^A, X_{1:t-1}^B, Y_{1:t-2})$ *is equivalent to* $\omega_{i,t} \propto \omega_{i,t-1} p(y_{t-1} \mid \theta_{i,t-1}, x_{i,t-1}^A, X_{1:t-1}^B, Y_{1:t-2})$.

Sequential Importance Resampling (SIR) (Rubin, 1981) is adopted to mitigate the degeneracy phenomenon. The details can be found in Appendix D.1.2 in the supplementary material.

With this recursive weight update rule, given shared data y_t (observation) and the data-sharing policy $\mathcal{K}_t^{\beta_t,*}(Y_t \mid \Theta, X_t^A)$, we can calculate $\hat{\beta}_t$ using $\{(\theta_i, x_{i,t}^A), \omega_{i,t}\}_{i=1}^N$ with weight update process $\omega_{i,t+1} = \mathcal{K}_t^{\beta_t,*}(y_t \mid \theta_i, x_{i,t}^A)\omega_{i,t}$, and the state of each particle will evolve through the system dynamics described in Section 2.

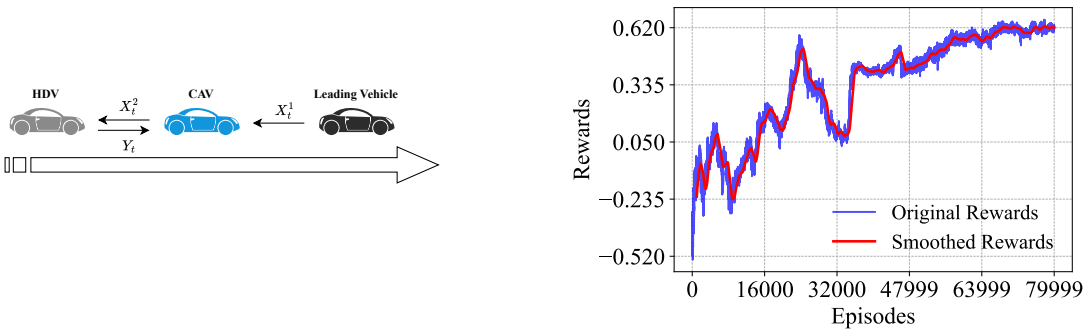
4.2. A2C solution algorithm with MGF encoder

With the particle representation of the belief state, we solve the optimal policy $\mathcal{K}_t^{\beta_t,*}(Y_t \mid \Theta, X_t^A)$ for each particle using an A2C algorithm. As illustrated in Fig. 2, our proposed algorithm comprises three main components: an MGF encoder, an actor network, and a critic network. The MGF encoder extracts features from the estimated particle-based belief state $\hat{\beta}_t$. This is because high-precision belief estimation often requires a large number of particles, leading to high-dimensional inputs to the actor and critic networks, which can make training computationally inefficient or even unstable. Additionally, SIR introduces particle permutation, which further complicates convergence. To address these issues, we adopt the MGF method (Ma et al., 2020) to efficiently encode higher-order moments into a low-dimensional vector. The MGF captures all moments of a distribution, fully characterizing it while reducing dimensionality (bul, 2012). Details of the mapping from particle beliefs to MGF features are provided in Ma et al. (2020). The actor network takes features from MGF features $(\bar{h}_t, M_{\beta_t}^{1:m})$ (detailed definition can be found in Appendix D.1.2), θ, and x_t^A as input, and outputs the parameters of a Dirichlet distribution, from which the policy $\mathcal{K}_t^{\beta_t}(Y_t \mid \theta, x_t^A)$ is sampled. The critic network takes the features from MGF as input and estimates a value function, which is used to update the actor network.

Training Procedure. The actor and critic networks, parameterized by ξ and θ_{critic}, respectively, are trained using standard A2C methods. The MGF encoder's parameters, $\phi \in \Phi$, are optimized jointly with the actor's policy loss, $\mathcal{L}_a(\xi)$, and the critic's value loss, $\mathcal{L}_c(\theta_{\text{critic}})$. This joint optimization enables the encoder to learn features that are beneficial for both action selection and value estimation. The MGF encoder's parameters are updated as $\phi_{t+1} = \phi_t - \eta_a \nabla_\phi \mathcal{L}_a(\xi_t) - \eta_c \nabla_\phi \mathcal{L}_c(\theta_{\text{critic},t})$, where $\mathcal{L}_a(\xi_t) = -\ln(q_{\xi_t}(a_t|\beta_t))\delta_t$ is the actor's loss with Temporal Difference (TD) error δ_t, $\mathcal{L}_c(\theta_{\text{critic},t}) = \delta_t^2$ is the critic's loss, minimizing the squared TD error, and η_a and η_c are the learning rates for the actor and critic, respectively. This dual-gradient update scheme allows the MGF encoder to optimize feature representations that benefit from both the actor's policy improvement and the critic's value estimation.

5. Experiments

In this section, we validate our proposed interaction-aware privacy-preserving data-sharing method in a mixed-autonomy platoon control scenario (Zhou and Yang, 2024).



(*a*) Mixed-autonomy platoon. (*b*) Training reward per episode.

Figure 3: Experiment settings and training results

5.1. Experimental setup

Our experimental setup includes scenario settings, adversaries, training, and evaluation. For a detailed description, please refer to the supplementary material (i.e., Appendix D).

Scenario settings. As shown in Fig. 3(*a*), we consider a mixed-autonomy platoon comprising a leading vehicle, a connected and automated vehicle (CAV) in the middle (data user B), and a following human-driven vehicle (HDV) at the tail (data provider A). The CAV gathers state information (speed and spacing) from all vehicles through vehicle-to-vehicle communications and computes its optimal control inputs accordingly. Therefore, this historical data shared by HDV could allow inference by adversaries of sensitive driving behavior parameters, which we aim to protect. The modeling of these vehicles is as follows. Similar to Zhou et al. (2024), the leading vehicle seeks to maintain its speed around a desired speed, and the speed tracking error follows a Gaussian distribution with mean of 0 and variance of 0.1. The details of system dynamics of the CAV and HDV can be referred to Wang et al. (2021); Zhou et al. (2024). The CAV control action $u(t)$ is determined using a distributed linear controller following Wang et al. (2021). The following HDV's car-following behavior is parameterized by the sensitive parameter Θ, an intrinsic property of a human driver.

Adversaries. Adversaries seek to infer the sensitive car-following parameters from the HDV using the system state of the CAV and the shared data of the HDV. Specifically, we assume that the adversaries can use Bayesian Inference (BI) and Recursive Least Squares (RLS) (Haykin, 2002) to infer the sensitive parameters. It is noted that BI is modeled as a PF that updates beliefs about system parameters using an observation model with Gaussian noise.

Training and evaluation: The proposed MGF-A2C algorithm is applied to learn the optimal data-sharing policy. The network architecture and hyperparameters are given in Appendix D in the supplementary material. The data sharing is performed at a frequency of 5 Hz. The training process considers data sharing of 40 s, whereby the belief states are characterized by 324 particles (4 for Θ and 81 for X_1^A), which are initially uniformly distributed within the parameter and state space with equal weight. The training rewards per episode are presented in Fig. 3(*b*), showing convergence after 60,000 episodes. To evaluate the trained model, the simulation was run for 1200 time steps (equivalent to 10 minutes), utilizing 5184 particles to ensure comprehensive assessment. We adopt the evaluation settings with longer and finer particle representation to more accurately measure the privacy leakage. For each value of Θ, experiments were conducted 10 times to evaluate privacy (i.e., the accuracy of RLS-estimated parameters) and control performance (i.e., fuel consumption).

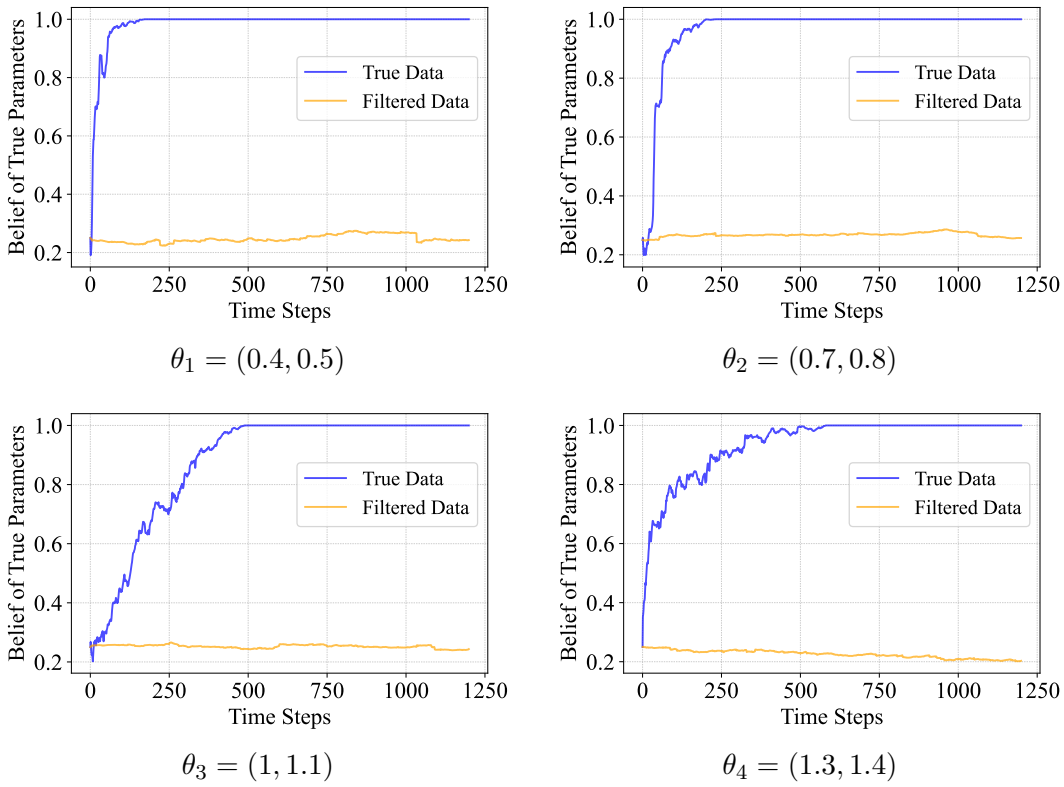


Figure 4: Comparison of Bayesian attacker's belief over time for different values of θ. Each figure shows the evolution of the attacker's belief when observing true data versus distorted (filtered) data.

5.2. Privacy-preserving performance

We evaluate the privacy-preserving performance of our proposed data-sharing algorithm (F) by comparing it against the benchmark policy that shares true data (R) without privacy protection.

Inference Attack using BI: BI is applied to quantify the privacy leakage represented by the belief of true parameters over time. The resulting evolution of belief is shown in Fig. 4 for different values of true theta. As we can see from Figure 4, the posterior probability resulting from our data-sharing algorithm remains largely similar to the prior probability which is uniform (i.e.,$p(\theta^*)$ is around 0.25). This indicates that the attacker is unable to significantly improve their belief about the θ^* from the distorted data sequence, demonstrating the effectiveness of the privacy filter in preventing sensitive information leakage. In contrast, when BI estimates θ^* only based on the true data sequence it observed, its belief about θ^* quickly converges to a value close to 1, showing that they can accurately infer the true parameters.

Inference Attack using RLS: We compare the parameters estimated by RLS using the log-transformed Root Mean Square Error (RMSE, σ_e) between θ^* and the estimated parameters. This metric, referred to as the success rate (SR), maps the RMSE to a range between 0 and 1, with values closer to 1 indicating a higher success rate of the θ^* estimation. The results are illustrated in Table. 1. We can clearly observe that the benchmark policy yields a small RMSE of the estimated parameter,

which means the RLS can accurately estimate the exact value of driving behavior. In contrast, the RMSE of the estimated parameter from the distorted data is larger, indicating that our framework effectively reduces the RLS's ability to accurately estimate the exact values of Θ.

These results confirm that our framework effectively safeguards driving behavior information against both BI and RLS-based attacks, consistently limiting the attacker's ability to deduce the true value of Θ and ensuring privacy protection under various conditions.

Table 1: Comparison of Parameter Estimation Accuracy and Fuel Consumption

Θ	**Privacy Measures for Parameters**				**Fuel Consumption**		
	σ_e **(R)**	**SR (R)**	σ_e **(F)**	**SR (F)**	**R**	**F**	$\Delta\%$
θ_1	0.0019	0.9981	2.928	0.0535	1.254	1.273	1.54
θ_2	0.0075	0.9925	3.581	0.0278	1.260	1.291	2.47
θ_3	0.0024	0.9976	2.280	0.1023	1.268	1.307	3.14
θ_4	0.0030	0.9970	2.454	0.0859	1.283	1.322	3.03

5.3. Tradeoff between privacy and control performance

To assess the impact of the proposed data-sharing policy, we compare the HDV's fuel consumption performance (mL/s) before and after applying the filter and calculate the difference Δ. The experimental results are summarized in Table 1. The results indicate that for different driving behaviors Θ, our data-sharing policy results in only a 1.5% to 3% increase in fuel consumption compared to directly sharing real state data, which is negligible. This demonstrates that our data-sharing policy effectively protects sensitive information with minimum impact on the control performance.

6. Conclusion

In this paper, we propose an interaction-aware privacy-preserving data-sharing approach to facilitate data sharing between coupled systems with continuous state space, where the mutual impact between systems introduces additional complexity. Our approach generates distorted data by minimizing a combination of (i) mutual information that quantifies the privacy leakage of sensitive parameters and (ii) the impact of distorted data on the data provider's control performance, considering the interactions between stakeholders. The optimization problem is formulated into a Bellman equation and solved by a particle filter reinforcement learning (RL)–based approach. Case studies on mixed-autonomy platoon control demonstrate the ability of our approach to protect sensitive parameters, such as HDV driving behaviors, while ensuring minimal impact on control performance. Future work includes extending the proposed approach to the sharing of high-dimensional operational data (e.g., vehicle trajectories). It would also be interesting to investigate the data-sharing incentives between multiple stakeholders using game-theoretic approaches.

Acknowledgments

This research was supported by the Singapore Ministry of Education (MOE) under its Academic Research Fund Tier 2 (A-8003064-00-00).

References

Principles of statistics. Courier Corporation, 2012.

Germán Bassi, Mikael Skoglund, and Pablo Piantanida. Lossy communication subject to statistical parameter privacy. In *2018 IEEE international symposium on information theory (ISIT)*, pages 1031–1035. IEEE, 2018.

Darrell P Bowyer, Rahmi Akçelik, and DC Biggs. *Guide to fuel consumption analyses for urban traffic management*. Number 32. 1985.

Chenxi Chen, Xianbiao Hu, Yang Li, and Qing Tang. Optimization of privacy budget allocation in differential privacy-based public transit trajectory data publishing for smart mobility applications. *IEEE Transactions on Intelligent Transportation Systems*, 2023.

Ecenaz Erdemir, Pier Luigi Dragotti, and Deniz Gündüz. Privacy-aware time-series data sharing with deep reinforcement learning. *IEEE Transactions on Information Forensics and Security*, 16: 389–401, 2020.

Simon Haykin. *Adaptive filter theory*. Prentice Hall, Upper Saddle River, NJ, 4th edition, 2002.

Rui Jiang, Qingsong Wu, and Zuojin Zhu. Full velocity difference model for a car-following theory. *Physical Review E*, 64(1):017101, 2001.

Gunn KH Larsen, Nicky D Van Foreest, and Jacquelien MA Scherpen. Power supply–demand balance in a smart grid: An information sharing model for a market mechanism. *Applied Mathematical Modelling*, 38(13):3350–3360, 2014.

Xiao Ma, Peter Karkus, David Hsu, Wee Sun Lee, and Nan Ye. Discriminative particle filter reinforcement learning for complex partial observations. *arXiv preprint arXiv:2002.09884*, 2020.

Ehsan Nekouei, Mohammad Pirani, Henrik Sandberg, and Karl H Johansson. A randomized filtering strategy against inference attacks on active steering control systems. *IEEE Transactions on Information Forensics and Security*, 17:16–27, 2021.

Ehsan Nekouei, Henrik Sandberg, Mikael Skoglund, and Karl Henrik Johansson. A model randomization approach to statistical parameter privacy. *IEEE Transactions on Automatic Control*, 68 (2):839–850, 2022.

Donald B Rubin. The bayesian bootstrap. *The annals of statistics*, pages 130–134, 1981.

Reza Shokri. Privacy games: Optimal user-centric data obfuscation. *arXiv preprint arXiv:1402.3426*, 2014.

Jiawei Wang, Yang Zheng, Chaoyi Chen, Qing Xu, and Keqiang Li. Leading cruise control in mixed traffic flow: System modeling, controllability, and string stability. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):12861–12876, 2021.

Chuanghong Weng, Ehsan Nekouei, and Karl H Johansson. Optimal privacy-aware dynamic estimation. *arXiv preprint arXiv:2311.05896*, 2023.

Wenjing Zhang, Bo Jiang, Ming Li, and Xiaodong Lin. Privacy-preserving aggregate mobility data release: An information-theoretic deep reinforcement learning approach. *IEEE Transactions on Information Forensics and Security*, 17:849–864, 2022.

Jing Zhao, Taeho Jung, Yu Wang, and Xiangyang Li. Achieving differential privacy of data disclosure in the smart grid. In *IEEE INFOCOM 2014-IEEE conference on computer communications*, pages 504–512. IEEE, 2014.

Jianfeng Zheng, Weili Sun, Shihong Huang, Shengyin Shen, Chunhui Yu, Jinqing Zhu, Bingbing Liu, and Henry X Liu. Traffic signal optimization using crowdsourced vehicle trajectory data. Technical report, 2018.

Jingyuan Zhou and Kaidi Yang. A parameter privacy-preserving strategy for mixed-autonomy platoon control. *Transportation Research Part C: Emerging Technologies*, 169:104885, 2024. ISSN 0968-090X. doi: https://doi.org/10.1016/j.trc.2024.104885.

Jingyuan Zhou, Longhao Yan, and Kaidi Yang. Enhancing system-level safety in mixed-autonomy platoon via safe reinforcement learning. *IEEE Transactions on Intelligent Vehicles*, 2024.

Ingvar Ziemann and Henrik Sandberg. Parameter privacy versus control performance: Fisher information regularized control. In *2020 American Control Conference (ACC)*, pages 1259–1265. IEEE, 2020.