entropy

MDPI

Article

Privacy-Preserving Design of Scalar LQG Control

Edoardo Ferrari [1,2], Yue Tian [1], Chenglong Sun [1], Zuxing Li [1,*] and Chao Wang [1]

[1] School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China; edoardo.ferrari2@studio.unibo.it (E.F.); 2132995@tongji.edu.cn (Y.T.); sunchenglong@tongji.edu.cn (C.S.); chaowang@tongji.edu.cn (C.W.)

[2] School of Electrical, Electronic, and Information Engineering "Guglielmo Marconi"—DEI, University of Bologna, 40136 Bologna, Italy

* Correspondence: zuxing@tongji.edu.cn

Abstract: This paper studies the agent identity privacy problem in the scalar linear quadratic Gaussian (LQG) control system. The agent identity is a binary hypothesis: Agent A or Agent B. An eavesdropper is assumed to make a hypothesis testing the agent identity based on the intercepted environment state sequence. The privacy risk is measured by the Kullback–Leibler divergence between the probability distributions of state sequences under two hypotheses. By taking into account both the accumulative control reward and privacy risk, an optimization problem of the policy of Agent B is formulated. This paper shows that the optimal deterministic privacy-preserving LQG policy of Agent B is a linear mapping. A sufficient condition is given to guarantee that the optimal deterministic privacy-preserving policy is time-invariant in the asymptotic regime. It is also shown that adding an independent Gaussian random process noise to the linear mapping of the optimal deterministic privacy-preserving policy cannot improve the performance of Agent B. The numerical experiments justify the theoretic results and illustrate the reward–privacy trade-off.

Keywords: control–privacy trade-off; hypothesis testing; Kullback–Leibler divergence; optimal control policy; privacy risk analysis

Citation: Ferrari, E.; Tian, Y.; Sun, C.; Li, Z.; Wang, C. Privacy-Preserving Design of Scalar LQG Control. *Entropy* **2022**, *24*, 856. https://doi.org/10.3390/e24070856

Academic Editor: Eduard Jorswieck

Received: 17 April 2022
Accepted: 20 June 2022
Published: 22 June 2022

1. Related Work

During the last decades, control technologies have been widely employed and significantly improved the industry productivity, management efficiency, and life convenience. The breakthrough of the deep reinforcement learning (DRL) technology [1] enables the control systems to be intelligent and applicable for more complicated tasks. Along with the increasing concerns about information security and privacy, adversarial problems in control systems have also attracted increasing attentions recently.

The related works and literature are introduced and discussed in the following. There are two types of adversarial problems considered in these works: active attacks and privacy problems.

1.1. Research on Active Adversarial Attacks

Most previous works focus on studying the active adversarial attacks on the control systems, which aim to degenerate the control efficiency, or even worse, to lead the system to an undesired state, and developing the corresponding defense mechanisms. Depending on their methodologies, these works can be divided into two classes. One class aims to develop the adversarial reinforcement learning algorithm under attack. The other class makes a theoretic study on the adversarial problem in the standard control model.

DRL takes advantage of the deep network to represent a complex non-linear value function or policy function. Similar to the deep network, DRL is also vulnerable to the adversarial example attack, i.e., the DRL-trained policy can be misled to take a wrong action by adding a minor distortion to the observation of the agent [2]. In [2–5], the optimal

generation of adversarial examples has been studied for given DRL algorithms. As a countermeasure, the mechanism of adversarial training uses adversarial examples in the training phase to enhance the robustness of control policy under attack [6–8]. In [9,10], attack/robustness-related regularization terms are added in the optimization objective to improve the robustness of the policy.

In most theoretic studies, adversarial attack problems are modeled from the game theoretic perspective. Stochastic game (SG) [11] and partially observable SG (POSG) can model the indirect (In SG or POSG, players indirectly interact with each other by feeding their actions back to the dynamic environment.) interactions between multiple players in the dynamic control system and have been employed in the robust or adversarial control studies [12–14]. Cheap talk game [15] models direct (In the cheap talk game, the sender with private information sends a message to the receiver and the receiver takes an action based on the received message and a belief on the inaccessible private information.) interactions between a sender and a receiver. In [16–19], the single-step cheap talk game has been extended to dynamic cheap talk games to model the adversarial example attacks in the multi-step control systems. With uncertainty about the environment dynamics in a partially observable Markov decision process (POMDP), the robust POMDP is formulated as a Stackelberg game in [20], where the agent (leader) optimizes the control policy under the worst-case assumption of the environment dynamics (follower). Another kind of adversarial attack maliciously falsifies the agent actions and feeds the falsified actions back to the dynamic environment to degrade the control performance. The falsified action attack can be modeled by Stackelberg games [21,22], where the dynamic environment is the leader and the adversarial agent is the follower. In our previous work [23], the falsified action attack on the linear quadratic regulator control is modeled by a dynamic cheap talk game and the adversarial attack is evaluated by the Fisher information between the random agent action and the falsified action.

Optimal stealthy attacks have also been studied. In [24,25], Kullback–Leibler divergence is used to measure the stealthiness of the attacks on the control signal and the sensing data, respectively; then the optimal attacks against LQG control system are developed with the objective of maximizing the quadratic cost while maintaining a degree of attack stealthiness.

1.2. Research on Privacy Problems

Besides the active attacks, passive eavesdropping in control systems leads to privacy problems. Most works focus on preserving the privacy-sensitive environment states. The design of agent actions in the Markov decision process has been investigated when the equivocation of states given system inputs and outputs is imposed as the privacy-preserving objective [26]. In [27–30], the notion of differential privacy [31] is introduced in the multi-agent control, where each agent adds privacy noise to his states before sharing them with other agents while guaranteeing the whole control system network to operate well. The reward function is a succinct description of the control task and is strongly relevant with the agent actions. The DRL-learned value function can reveal the privacy-sensitive reward function. Regarding this privacy problem, functional noise is added to the value function in the Q-learning such that the neighborhood reward functions are indistinguishable [32]. As a promising computational secrecy technology, labeled homomorphic encryption has been employed to encrypt the private states, gain matrices, control inputs, and intermediary steps in the cloud-outsourced LQG [33].

2. Introduction

2.1. Motivation

In this paper, we consider the agent identity privacy problem in the LQG control, which is motivated by the inverse reinforcement learning (IRL). IRL algorithms [34] can reconstruct the reward functions of agents and therefore can also be maliciously exploited to identify the agents. Similar to many other privacy problems in the big data era, such as

the smart meter privacy problem, the agent identity of a control system is privacy-sensitive. When the agent identity is leaked, an adversary can further employ the corresponding optimal attacks on the control system.

2.2. Content and Contribution

We model the agent identity privacy problem as an adversarial binary hypothesis testing and employ the Kullback–Leibler divergence between the probability distributions of environment state sequences under different hypotheses as the privacy risk measure. We formulate a novel optimization problem and study the optimal privacy-preserving LQG policy. This work is compared with the previous research on privacy problems in Table 1.

Table 1. Comparison of research on privacy problems.

	Private Information	**Privacy Model/Measure**	**Privacy Mechanism**
[26]	State	Equivocation	Privacy-preserving policy design
[27–30]	State	Differential privacy	Adding privacy noise to state
[32]	Reward function	Differential privacy	Adding privacy noise to value function
[33]	The whole LQG system	Computational secrecy	Labeled homomorphic encryption
This work	Agent identity	Kullback–Leibler divergence	Privacy-preserving policy design

The rest of this paper is organized as follows. In Section 3, we formulate the agent identity privacy problem in the LQG control system. In Section 4, we optimize the deterministic privacy-preserving LQG policy and give a sufficient condition for time-invariant optimal deterministic policy in the asymptotic regime. In Section 5, we discuss the random privacy-preserving LQG policy and show that the optimal linear Gaussian random policy reduces to the optimal deterministic privacy-preserving LQG policy. In Section 6, we present and analyze the numerical experiment results. Section 7 concludes this paper.

2.3. Notation

Unless otherwise specified, we denote a random scalar by a capital letter, e.g., X, its realization by the corresponding lower case letter, e.g., x, the Gaussian distribution with mean μ and variance σ^2 by $\mathcal{N}(\mu, \sigma^2)$, the expectation operation by $\mathbb{E}(\cdot)$, the Kullback–Leibler divergence between two probability distributions by $\mathbb{D}(\cdot||\cdot)$, and the natural logarithm by $\log(\cdot)$.

3. Agent Identity Privacy Problem in LQG Control

We consider an N-step LQG control in the presence of an eavesdropper as shown in Figure 1. There are two possible agents, Agent A and Agent B, which are with respect to a hypothesis $H = 0$ and an alternative hypothesis $H = 1$. We assume that the agents and the eavesdropper have perfect observations of the environment states. Based on the intercepted state sequence, the eavesdropper makes a binary hypothesis testing (A binary hypothesis is considered in this paper for simplification and can be extended to a multi-hypothesis.) to identify the current agent, which results in an agent identity privacy problem. To have a better understanding of the privacy problem, we give an example in the emerging application of autonomous vehicle. An autonomous vehicle can be controlled by a human driver (Agent A) or an autonomous driving system (Agent B). An adversary, who can be a compromised manager of the vehicle to everything (V2X) network, has access to the sensing data (environment state) of the autonomous vehicle and aims to attack the

autonomous vehicle, e.g., to mislead the autonomous vehicle off the lane. To this end, the adversary needs to first identify if the current driver is the autonomous driving system by the intercepted sensing data sequence. The agent identity privacy problem commonly exists in intelligent autonomous systems, e.g., unmanned aerial vehicles and robots, where the autonomous control agents depending strongly on the sensing data are vulnerable to injection attacks and therefore the agent identities are privacy-sensitive.

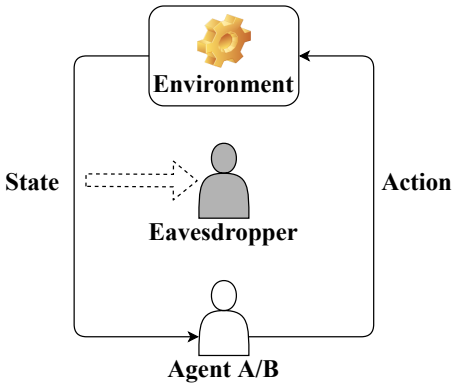

Figure 1. LQG control in the presence of an eavesdropper.

The LQG control model for each agent is given as follows: For $H = 0$ or $H = 1$, $1 \leq i \leq N$,

$$s_{i+1}^{(H)} = \alpha s_i^{(H)} + \beta a_i^{(H)} + z_i, \tag{1}$$

$$a_i^{(H)} = F_i^{(H)}\left(s_i^{(H)}\right), \tag{2}$$

$$r_i^{(H)} = R^{(H)}\left(s_i^{(H)}, a_i^{(H)}\right) = -\theta^{(H)}\left(s_i^{(H)}\right)^2 - \phi^{(H)}\left(a_i^{(H)}\right)^2, \tag{3}$$

$$S_1^{(H)} \sim b_1^{(H)} \triangleq \mathcal{N}(\mu_1, \sigma_1^2), \tag{4}$$

$$Z_i \sim \mathcal{N}(0, \omega^2), \tag{5}$$

where the parameters $\alpha \neq 0$, $\beta \neq 0$, $\theta^{(H)} > 0$, $\phi^{(H)} > 0$, μ_1, $\sigma_1^2 > 0$, and $\omega^2 > 0$ are given. The initial environment state $s_1^{(H)}$ is randomly generated following an *independent* Gaussian distribution. In the i-th time step, on observing the environment state $s_i^{(H)}$, the agent with respect to the hypothesis H employs the control policy $F_i^{(H)}$ to (randomly) determine an action $a_i^{(H)}$ as (2); the instantaneous control reward $r_i^{(H)}$ is jointly determined by the current state $s_i^{(H)}$ and action $a_i^{(H)}$ as (3); the next state $s_{i+1}^{(H)}$ is jointly determined by the current state $s_i^{(H)}$, the current action $a_i^{(H)}$, and z_i randomly generated following an *independent* zero-mean Gaussian distribution as (1). In the standard LQG problem, the agent with respect to the hypothesis H only aims to maximize the expected accumulative reward by optimizing the control policies $F_{1:N}^{(H)}$:

$$F_{1:N}^{(H)*} = \arg\max_{F_{1:N}^{(H)}} \mathbb{E}\left(\sum_{i=1}^{N} R^{(H)}\left(S_i^{(H)}, A_i^{(H)}\right)\right). \tag{6}$$

The optimal LQG control policy has been well established [35] and can be described as follows. For $H = 0$ or $H = 1$, $1 \leq i \leq N$,

$$\tilde{\theta}_{N+1}^{(H)} = 0, \tag{7}$$

$$\tilde{\theta}_i^{(H)} = L^{(H)}\left(\tilde{\theta}_{i+1}^{(H)}\right) = \theta^{(H)} + \tilde{\theta}_{i+1}^{(H)}\alpha^2 - \frac{\left(\tilde{\theta}_{i+1}^{(H)}\right)^2\alpha^2\beta^2}{\phi^{(H)} + \tilde{\theta}_{i+1}^{(H)}\beta^2} > 0, \tag{8}$$

$$\kappa_i^{(H)*} = -\frac{\tilde{\theta}_{i+1}^{(H)}\alpha\beta}{\phi^{(H)} + \tilde{\theta}_{i+1}^{(H)}\beta^2}, \tag{9}$$

$$F_i^{(H)*}\left(s_i^{(H)}\right) = \kappa_i^{(H)*}s_i^{(H)}. \tag{10}$$

For $H = 0$ or 1, it can be easily verified that the mapping $L^{(H)}$ is order-preserving, i.e., $L^{(H)}(x) \leq L^H(x')$ if $0 \leq x \leq x'$. From the Kleene's fixed point theorem [36], it follows that

$$\begin{aligned} \tilde{\theta}^{(H)} &= \lim_{N\to\infty} \underbrace{L^{(H)}(L^{(H)}(\cdots(L^{(H)}(L^{(H)}}_{N \text{ iterations}}(\tilde{\theta}_{N+1}^{(H)})))\cdots)) \\ &= \theta^{(H)} + \tilde{\theta}^{(H)}\alpha^2 - \frac{\left(\tilde{\theta}^{(H)}\right)^2\alpha^2\beta^2}{\phi^{(H)}+\tilde{\theta}^{(H)}\beta^2} \\ &= \frac{\sqrt{(\phi^{(H)}-\theta^{(H)}\beta^2-\phi^{(H)}\alpha^2)^2+4\theta^{(H)}\phi^{(H)}\beta^2}-(\phi^{(H)}-\theta^{(H)}\beta^2-\phi^{(H)}\alpha^2)}{2\beta^2}. \end{aligned} \tag{11}$$

Therefore, if we consider the asymptotic regime as $N \to \infty$, the optimal control polices are *time-invariant*: For $H = 0$ or $H = 1$, $i \geq 1$,

$$\kappa^{(H)*} = -\frac{\tilde{\theta}^{(H)}\alpha\beta}{\phi^{(H)} + \tilde{\theta}^{(H)}\beta^2}, \tag{12}$$

$$F_i^{(H)*}\left(s_i^{(H)}\right) = \kappa^{(H)*}s_i^{(H)}. \tag{13}$$

For the agent identity privacy problem, we assume that the eavesdropper collects a sequence of environment states and carries out a binary hypothesis testing on the agent identity. Thus, the privacy risk can be measured by the hypothesis testing performance. In information theory, Kullback–Leibler divergence measures the "distance" between two probability distributions. When the value of the Kullback–Leibler divergence $\mathbb{D}\left(p_{S_{1:N}^{(1)}}||p_{S_{1:N}^{(0)}}\right)$ is smaller, the random environment state sequences $S_{1:N}^{(0)}$ and $S_{1:N}^{(1)}$ are statistically "closer" to each other and it is more difficult for the eavesdropper to identify the current agent, i.e., a poorer hypothesis testing performance and a lower privacy risk. In this paper, we employ the Kullback–Leibler divergence $\mathbb{D}\left(p_{S_{1:N}^{(1)}}||p_{S_{1:N}^{(0)}}\right)$ as the privacy risk measure.

Furthermore, we assume that both agents aim to improve their own expected accumulative rewards while only Agent B considers to reduce the privacy risk. This assumption makes sense in a lot of scenarios. In the aforementioned autonomous vehicle example, Agent A denotes the human driver and does not need to change the optimal driving style; Agent B denotes the autonomous driving system and can be reconfigured with respect to the human's optimal driving style to improve the driving efficiency and to reduce the privacy risk. Under the assumption, Agent A takes the optimal LQG control policy as described by (7)–(10) with $H = 0$. In the following, we focus on the privacy-preserving

LQG control policy of Agent B. Taking into account the two design objectives of Agent B, we formulate the following optimization problem:

$$F_{1:N}^{(1)\star} = \arg\max_{F_{1:N}^{(1)}} \mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda \mathbb{D}\left(p_{S_{1:N}^{(1)}} || p_{S_{1:N}^{(0)*}}\right), \tag{14}$$

where $\lambda \geq 0$ denotes the privacy-preserving design weight; the random environment state sequence $S_{1:N}^{(0)*}$ is induced by the optimal LQG policy $F_{1:N}^{(0)*}$ of Agent A. It follows from the chain rule of Kullback–Leibler divergence and the Markovian property of the state sequences that the privacy risk measure can be further decomposed as

$$\begin{aligned} \mathbb{D}\left(p_{S_{1:N}^{(1)}} || p_{S_{1:N}^{(0)*}}\right) &= \mathbb{D}\left(p_{S_1^{(1)}} || p_{S_1^{(0)}}\right) + \textstyle\sum_{i=2}^{N} \mathbb{D}\left(p_{S_i^{(1)}|S_{i-1}^{(1)}} || p_{S_i^{(0)*}|S_{i-1}^{(0)*}}\right) \\ &= \textstyle\sum_{i=2}^{N} \mathbb{D}\left(p_{S_i^{(1)}|S_{i-1}^{(1)}} || p_{S_i^{(0)*}|S_{i-1}^{(0)*}}\right). \end{aligned} \tag{15}$$

It is obvious that the optimal privacy-preserving LQG control policy of Agent B depends on the value of λ. In the following two remarks, the optimal privacy-preserving LQG control policies are characterized for two special cases, $\lambda = 0$ and $\lambda \to \infty$, respectively.

Remark 1. *When $\lambda = 0$, Agent B only aims to maximize the expected accumulative reward $\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right)$. In this case, the optimal privacy-preserving LQG policy of Agent B reduces to the optimal LQG policy of Agent B, i.e., $F_i^{(1)\star}\left(s_i^{(1)}\right) = F_i^{(1)*}\left(s_i^{(1)}\right) = \kappa_i^{(1)*} s_i^{(1)}$ for all $1 \leq i \leq N$.*

Remark 2. *When $\lambda \to \infty$, Agent B only aims to minimize the privacy risk, which is measured by the Kullback–Leibler divergence $\mathbb{D}\left(p_{S_{1:N}^{(1)}} || p_{S_{1:N}^{(0)*}}\right)$. In this case, the optimal privacy-preserving LQG policy of Agent B reduces to the optimal LQG policy of Agent A, i.e., $F_i^{(1)\star}\left(s_i^{(1)}\right) = F_i^{(0)*}\left(s_i^{(1)}\right) = \kappa_i^{(0)*} s_i^{(1)}$ for all $1 \leq i \leq N$, and the minimum privacy risk is achieved, i.e., $\mathbb{D}\left(p_{S_{1:N}^{(1)\star}} || p_{S_{1:N}^{(0)*}}\right) = 0$.*

When $0 < \lambda < \infty$, we characterize the optimal privacy-preserving LQG control policies of Agent B in different forms in the following sections. For ease of reading, we list the parameters and their meanings in Table 2.

Table 2. Parameters.

Parameter	Meaning	Parameter	Meaning
N	Number of steps	H	Agent identity binary hypothesis
α, β	Time-invariant linear coefficients in the linear Gaussian dynamic model	z_i, ω^2	Independent zero-mean Gaussian-distributed disturbance noise in the i-th step and its variance
$s_i^{(H)}$	State of the agent (H) in the i-th step	$a_i^{(H)}$	Action of the agent (H) in the i-th step
$F_i^{(H)}$	Policy of the agent (H) in the i-th step	$\kappa_i^{(H)}$	State feedback gain of a linear policy of the agent (H) in the i-th step
$r_i^{(H)}$	Instantaneous control reward of the agent (H) in the i-th step	$R^{(H)}$, $\theta^{(H)}$, $\phi^{(H)}$	Time-invariant instantaneous quadratic control reward function of the agent (H) and its coefficients
μ_1, σ_1^2	Mean and variance of the Gaussian-distributed initial state	λ	Privacy-preserving design weight

4. Deterministic Privacy-Preserving LQG Policy

When the privacy risk is not considered, as shown in (10), the optimal LQG control policy of Agent B is a deterministic linear mapping. In this section, we study the optimal deterministic privacy-preserving LQG policy of Agent B. Therefore, the policy of Agent B can be specified as: For $1 \leq i \leq N$,

$$F_i^{(1)} : \mathbb{R} \to \mathbb{R}. \tag{16}$$

In the following theorem, we characterize the optimal deterministic privacy-preserving LQG policy of Agent B.

Theorem 1. *At each step, the optimal deterministic privacy-preserving LQG policy of Agent B with respect to the optimization problem* (14) *is a linear mapping as: For* $1 \leq i \leq N$,

$$\hat{\theta}_{N+1}^{(1)} = 0, \tag{17}$$

$$\hat{\theta}_i^{(1)} = J_{N+1-i}\left(\hat{\theta}_{i+1}^{(1)}\right) = \theta^{(1)} + \hat{\theta}_{i+1}^{(1)}\alpha^2 + \frac{\lambda}{2\omega^2}\beta^2\left(\kappa_i^{(0)*}\right)^2 - \frac{\left(\frac{\lambda}{2\omega^2}\beta^2\kappa_i^{(0)*} - \hat{\theta}_{i+1}^{(1)}\alpha\beta\right)^2}{\phi^{(1)} + \hat{\theta}_{i+1}^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2} > 0, \tag{18}$$

$$\kappa_i^{(1)\star} = \frac{\frac{\lambda}{2\omega^2}\beta^2\kappa_i^{(0)*} - \hat{\theta}_{i+1}^{(1)}\alpha\beta}{\phi^{(1)} + \hat{\theta}_{i+1}^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2}, \tag{19}$$

$$F_i^{(1)\star}\left(s_i^{(1)}\right) = \kappa_i^{(1)\star}s_i^{(1)}. \tag{20}$$

Then, the maximum achievable weighted design objective of Agent B is

$$\max_{F_{1:N}^{(1)}} \mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda \mathbb{D}\left(p_{S_{1:N}^{(1)}} \middle|\middle| p_{S_{1:N}^{(0)*}}\right) = -\hat{\theta}_1^{(1)}(\mu_1^2 + \sigma_1^2) - \omega^2 \sum_{i=1}^{N-1} \hat{\theta}_{i+1}^{(1)}. \quad (21)$$

The proof of Theorem 1 is presented in Appendix A.

Remark 3. *When $\lambda = 0$, it is easy to show that $\kappa_i^{(1)\star} = \kappa_i^{(1)*}$ for all $1 \leq i \leq N$, i.e., the optimal deterministic privacy-preserving LQG policy is consistent with the optimal privacy-preserving LQG policy shown in Remark 1.*

Remark 4. *It is easy to show that $\lim_{\lambda\to\infty} \kappa_i^{(1)\star} = \kappa_i^{(0)*}$ for all $1 \leq i \leq N$, i.e., the optimal deterministic privacy-preserving LQG policy is consistent with the optimal privacy-preserving LQG policy shown in Remark 2.*

Remark 5. *Although the objective in (14) is a linear combination of the expected accumulative reward and the privacy risk measured by the Kullback–Leibler divergence, the optimal linear coefficient $\kappa_i^{(1)\star}$ is a non-linear function of $\kappa_i^{(1)*}$ (the optimal linear coefficient with respect to only maximize the expected accumulative reward) and $\kappa_i^{(0)*}$ (the optimal linear coefficient with respect to only minimize the privacy risk) when we consider the deterministic privacy-preserving LQG control policy of Agent B.*

Remark 6. *When Agent B employs the optimal deterministic privacy-preserving LQG policy at each step, the random state-action sequence is jointly Gaussian distributed.*

In the asymptotic regime as $N \to \infty$, the optimal LQG control policy is time-invariant. In this case, the design of the optimal policy becomes an easier task. Theorem 2 gives a sufficient condition such that the optimal deterministic privacy-preserving LQG policy of Agent B is time-invariant in the asymptotic regime.

Theorem 2. *When the model parameters satisfy the following inequality*

$$\frac{\left|\frac{\lambda}{2\omega^2}\beta^4\left(\kappa^{(0)*}\right)^2\phi^{(1)} - \left(\phi^{(1)}\alpha^2 + \frac{\lambda}{2\omega^2}\beta^2\left(\alpha + \beta\kappa^{(0)*}\right)^2\right)\left(\phi^{(1)} + \frac{\lambda}{2\omega^2}\beta^2\right)\right|}{\left(\phi^{(1)} + \frac{\lambda}{2\omega^2}\beta^2\right)^2} < 1, \quad (22)$$

the optimal deterministic privacy-preserving LQG policy of Agent B is time-invariant in the asymptotic regime. More specifically, $J_N(J_{N-1}(\cdots(J_2(J_1(\hat{\theta}_{N+1}^{(1)})))\cdots))$ converges to the unique fixed point $\hat{\theta}^{(1)}$ as

$$\begin{aligned}\hat{\theta}^{(1)} &= \lim_{N\to\infty} J_N(J_{N-1}(\cdots(J_2(J_1(\hat{\theta}_{N+1}^{(1)})))\cdots)) \\ &= \theta^{(1)} + \alpha^2\hat{\theta}^{(1)} + \frac{\lambda}{2\omega^2}\beta^2\left(\kappa^{(0)*}\right)^2 - \frac{\left(\frac{\lambda}{2\omega^2}\beta^2\kappa^{(0)*} - \alpha\beta\hat{\theta}^{(1)}\right)^2}{\phi^{(1)} + \beta^2\hat{\theta}^{(1)} + \frac{\lambda}{2\omega^2}\beta^2};\end{aligned} \quad (23)$$

and the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B can be described by

$$\kappa^{(1)\star} = \frac{\frac{\lambda}{2\omega^2}\beta^2\kappa^{(0)*} - \hat{\theta}^{(1)}\alpha\beta}{\phi^{(1)} + \hat{\theta}^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2}, \tag{24}$$

$$F_i^{(1)\star}\left(s_i^{(1)}\right) = \kappa^{(1)\star}s_i^{(1)}. \tag{25}$$

Under this condition, the asymptotic weighted design object rate of Agent B achieved by the time-invariant optimal deterministic privacy-preserving LQG policy is

$$\lim_{N\to\infty}\frac{1}{N}\max_{F_{1:N}^{(1)}}\mathbb{E}\left(\sum_{i=1}^{N}R^{(1)}\left(S_i^{(1)},A_i^{(1)}\right)\right) - \lambda\mathbb{D}\left(p_{S_{1:N}^{(1)}}\middle|\middle|p_{S_{1:N}^{(0)*}}\right) = -\omega^2\hat{\theta}^{(1)}. \tag{26}$$

The proof of Theorem 2 is given in Appendix B.

5. Random Privacy-Preserving LQG Policy

As shown in Theorem 1, the optimal deterministic privacy-preserving LQG policy of Agent B is a linear mapping. In this section, we first discuss the optimal random privacy-preserving LQG policy and then consider a particular random policy by extending the deterministic linear mapping to the linear Gaussian random policy for Agent B. Here, the random policy of Agent B can be specified as: For $1 \le i \le N$,

$$F_i^{(1)} : \mathbb{R}\times\mathbb{R} \to \mathbb{R}_{\ge 0}. \tag{27}$$

With slight abuse of notation, we denote the condition probability (density) of taking the action $a_i^{(1)} \in \mathbb{R}$ given the state $s_i^{(1)} \in \mathbb{R}$ and the random policy $F_i^{(1)}$ by $F_i^{(1)}\left(a_i^{(1)}\middle|s_i^{(1)}\right) \in \mathbb{R}_{\ge 0}$.

It can be easily shown that the optimal random privacy-preserving LQG policy of Agent B in the final step $F_N^{(1)\star}$ reduces to the deterministic linear mapping in (A2). For $1 \le i \le N-1$, it follows from the backward dynamic programming that the optimal random privacy-preserving LQG policy of Agent B in the i-th step does not reduce to a deterministic linear mapping in general. That is because the conditional probability distribution $p_{S_{i+1}^{(1)}|S_i^{(1)}}$ given a random policy $F_i^{(1)}$ is a Gaussian mixture model and then the Kullback–Leibler divergence $\mathbb{D}\left(p_{S_{i+1}^{(1)}|S_i^{(1)}}\middle|\middle|p_{S_{i+1}^{(0)*}|S_i^{(0)*}}\right)$ between a Gaussian mixture model and a Gaussian distribution generally does not reduce to the quadratic mean of $A_i^{(1)} - \kappa_i^{(0)*}S_i^{(1)}$ as (A5). To the best of our knowledge, there is no analytically tractable formula for Kullback–Leibler divergence between Gaussian mixture models and only approximations are available [37–39]. Therefore, we do not give the close-form solution of the optimal random privacy-preserving LQG policy in this paper.

In what follows, we focus on the linear Gaussian random policy: For $1 \le i \le N$,

$$F_i^{(1)}\left(s_i^{(1)}\right) = \kappa_i^{(1)}s_i^{(1)} + w_i^{(1)}, \tag{28}$$

where $w_i^{(1)}$ is the realization of an *independent* zero-mean Gaussian random process noise $W_i^{(1)} \sim \mathcal{N}(0,\delta_i^2)$. Thus, a linear Gaussian random policy $F_i^{(1)}$ can be completely described by the parameters $\left(\kappa_i^{(1)},\delta_i^2\right)$. Theorem 3 characterizes the optimal linear Gaussian random privacy-preserving LQG policy of Agent B.

Theorem 3. *At each step, the optimal linear Gaussian random privacy-preserving LQG policy of Agent B with respect to the optimization problem (14) is the same deterministic linear mapping as in Theorem 1.*

The proof of Theorem 3 is presented in Appendix C.

Remark 7. *Adding an independent zero-mean Gaussian random process noise to the linear mapping of the optimal deterministic privacy-preserving LQG policy cannot improve the performance of Agent B.*

6. Numerical Experiments

6.1. Convergence of the Sequence $\left(\hat{\theta}^{(1)}_{N+1}, \hat{\theta}^{(1)}_{N}, \hat{\theta}^{(1)}_{N-1}, \dots\right)$

When the constraint (22) in Theorem 2 is satisfied, we first illustrate the convergence of the sequence $\left(\hat{\theta}^{(1)}_{N+1}, \hat{\theta}^{(1)}_{N}, \hat{\theta}^{(1)}_{N-1}, \dots\right)$. In addition to the default model parameters in Table 3, we set $\theta^{(1)} = 8$, $\phi^{(1)} = 1$, and let the privacy-preserving design weight $\lambda = 1$, 5 or 10. By using these parameters, it can be easily verified that the constraint (22) is satisfied. Figure 2 shows that $\hat{\theta}^{(1)}_{N+1-k} = J_k(J_{k-1}(\cdots(J_2(J_1(\hat{\theta}^{(1)}_{N+1})))\cdots))$ converges after $k = 20$ iterations for different values of λ. Furthermore, different convergence patterns can be observed for different values of λ.

Table 3. Default model parameters.

Parameter	μ_1	σ_1^2	α	β	ω^2	$\theta^{(0)}$	$\phi^{(0)}$
Value	1	1	1	0.5	0.5	1	16

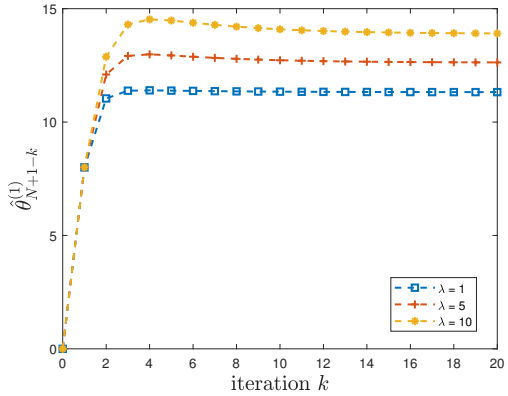

Figure 2. For $\lambda = 1$, 5 or 10, the convergence of $\hat{\theta}^{(1)}_{N+1-k} = J_k(J_{k-1}(\cdots(J_2(J_1(\hat{\theta}^{(1)}_{N+1})))\cdots))$.

6.2. Impact of the Privacy-Preserving Design Weight λ

Here, we show the impact of the privacy-preserving design weight λ on the trade-off between the control reward of Agent B and the privacy risk. We use the same parameters as in Section 6.1, but allow $0 \leq \lambda \leq 10{,}000$. Then, Theorem 2 is applicable and therefore the optimal deterministic privacy-preserving LQG policy of Agent B is time-invariant in the asymptotic regime. Figures 3 and 4 show that both the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)\star}, A_i^{(1)\star}\right)\right)$ and the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S^{(1)\star}_{1:N}} \middle|\middle| p_{S^{(0)*}_{1:N}}\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B decrease as λ increases, i.e., the control reward of Agent B is degraded while the privacy is enhanced. When the privacy risk is not considered, the best control reward of Agent B is achieved at the cost of the highest privacy risk.

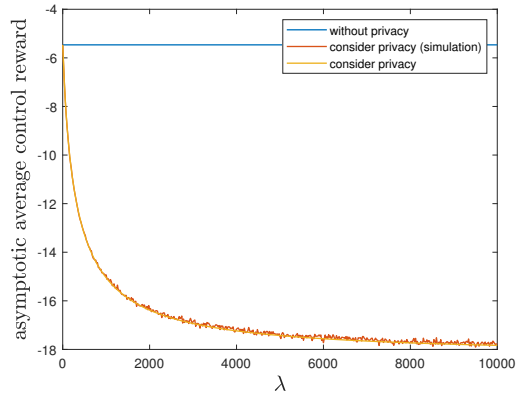

Figure 3. When $0 \leq \lambda \leq 10{,}000$, comparison of the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)*}, A_i^{(1)*}\right)\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)\star}, A_i^{(1)\star}\right)\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

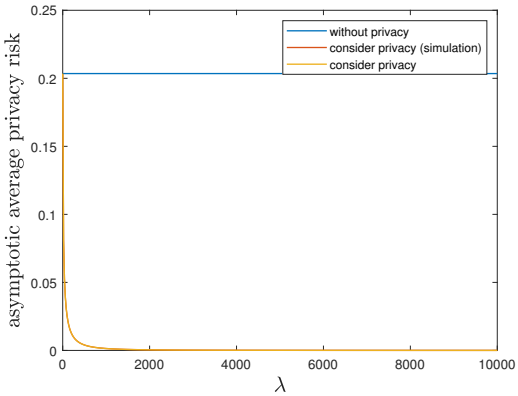

Figure 4. When $0 \leq \lambda \leq 10{,}000$, comparison of the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)*}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)\star}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

In addition to the analytical results, we also present the simulation results by considering privacy in Figures 3 and 4. Given $0 \leq \lambda \leq 10{,}000$, we employ the corresponding time-invariant optimal deterministic privacy-preserving LQG policy of Agent B and run the 10,000-step privacy-preserving LQG control with 100 randomly generated initial states. Then, the average control reward and the average privacy risk are evaluated and compared with the analytical results of asymptotic average control reward and asymptotic average privacy risk, respectively. As shown in Figures 3 and 4, the simulation results match quite well with the analytical results, which validates our analytical results.

6.3. Impact of Parameter $\theta^{(1)}$

Here, we study the impact of the parameter $\theta^{(1)}$ on the control reward of Agent B and the privacy risk. In addition to the default model parameters in Table 3, we set $\phi^{(1)} = \phi^{(0)} = 16$ and allow $0.01 \leq \theta^{(1)} \leq 8$, $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000. It can be verified that Theorem 2 holds for those model parameters. For all $0.01 \leq \theta^{(1)} \leq 8$ and by increasing the value of λ, Figures 5 and 6 show a trade-off between the control reward of Agent B and the privacy risk, which is consistent with the previous observations. For $\lambda = 0$, 10, 100, 1000 or 10,000, Figure 5 shows that the asymptotic average

control reward of Agent B decreases as $\theta^{(1)}$ increases. This is reasonable since $-\theta^{(1)}$ is the quadratic coefficient in the instantaneous reward function $R^{(1)}$. For $\lambda = 0$, 10, 100, 1000 or 10,000, Figure 6 shows that the asymptotic average privacy risk has a pattern to decrease first, then to increase, and to achieve the minimum value 0 when $\theta^{(1)} = \theta^{(0)} = 1$. When $\theta^{(1)} = \theta^{(0)} = 1$, both agents have the same instantaneous reward function and employ the same optimal LQG control policy, which leads to the same state sequence distribution under both hypotheses and the minimum value 0 of the Kullback–Leibler divergence. As $\theta^{(1)}$ deviates from the value of $\theta^{(0)}$, the agents have more different instantaneous reward functions, which lead to more different state sequence distributions under both hypotheses and a larger value of the Kullback–Leibler divergence.

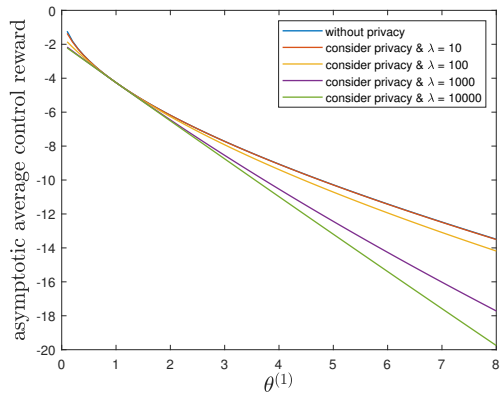

Figure 5. For $0.01 \leq \theta^{(1)} \leq 8$ and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)*}, A_i^{(1)*}\right)\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)\star}, A_i^{(1)\star}\right)\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

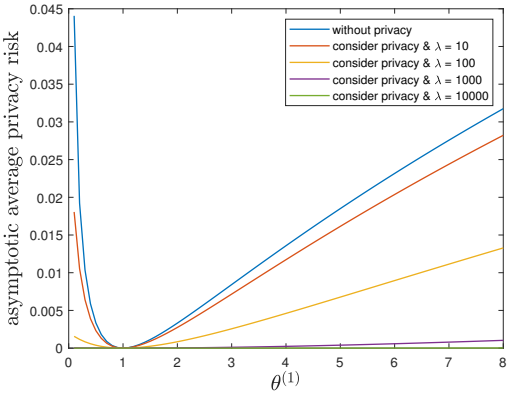

Figure 6. For $0.01 \leq \theta^{(1)} \leq 8$ and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)*}} \middle|\middle| p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)\star}} \middle|\middle| p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

6.4. Impact of Parameter $\phi^{(1)}$

Here, we show the impact of the parameter $\phi^{(1)}$ on the control reward of Agent B and the privacy risk. In addition to the default model parameters in Table 3, we set $\theta^{(1)} = \theta^{(0)} = 1$ and allow $0.01 \leq \phi^{(1)} \leq 40$, $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000. It can be verified that Theorem 2 holds for those model parameters. For all $0.01 \leq \phi^{(1)} \leq 40$ and by increasing the value of λ, Figures 7 and 8 also show a trade-off between the control

reward of Agent B and the privacy risk. For $\lambda = 0$, 10, 100, 1000 or 10,000, Figure 7 shows that the asymptotic average control reward of Agent B decreases as $\phi^{(1)}$ increases. This is because $-\phi^{(1)}$ is the other quadratic coefficient in the instantaneous reward function $R^{(1)}$. For $\lambda = 0$, 10, 100, 1000 or 10,000, Figure 8 shows that the asymptotic average privacy risk has a similar pattern to decrease first, then to increase, and to achieve the minimum value 0 when $\phi^{(1)} = \phi^{(0)} = 16$. This pattern can be similarly explained as Section 6.3.

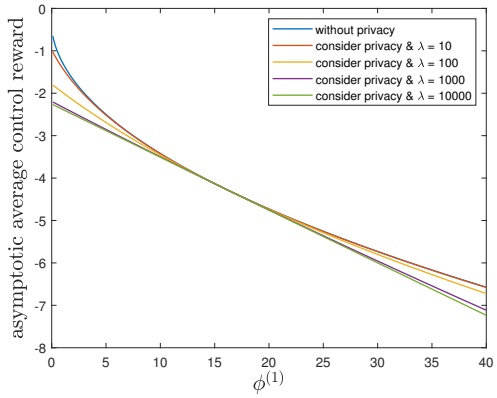

Figure 7. For $0.01 \leq \phi^{(1)} \leq 40$ and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)*}, A_i^{(1)*}\right)\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)\star}, A_i^{(1)\star}\right)\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

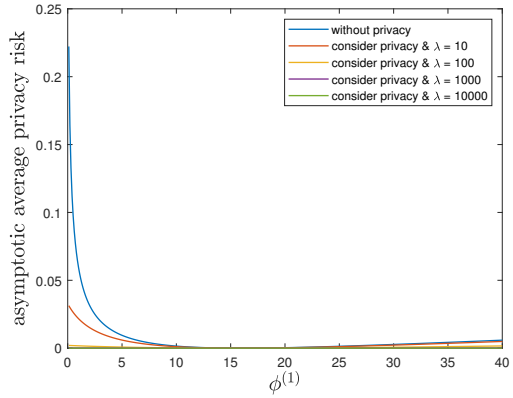

Figure 8. For $0.01 \leq \phi^{(1)} \leq 40$ and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)*}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)\star}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

6.5. Impact of Parameter $\theta^{(0)}$

By fixing $\theta^{(1)} = 1$ and $\phi^{(1)} = \phi^{(0)} = 16$, we study the impact of the parameter $\theta^{(0)}$ on the control reward of Agent B and the privacy risk. In addition to the default model parameters in Table 3, we allow $0.01 \leq \theta^{(0)} \leq 8$ and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000. It can be verified that Theorem 2 holds for those model parameters. For all $0.01 \leq \theta^{(0)} \leq 8$ and by increasing the value of λ, Figures 9 and 10 show a trade-off between the control reward of Agent B and the privacy risk. For $\lambda = 0$, 10, 100, 1000 or 10,000, Figures 9 and 10 show that the asymptotic average control reward of Agent B achieves the maximum value while the asymptotic average privacy risk achieves the minimum value 0 when $\theta^{(1)} = \theta^{(0)} = 1$. In this case, both agents have the same instantaneous reward

function and employ the same optimal LQG control policy, which maximizes their control rewards, leads to the same state sequence distribution under both hypotheses, and therefore achieves the minimum value 0 of the Kullback–Leibler divergence.

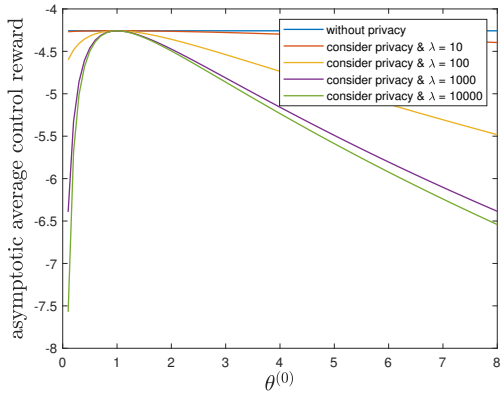

Figure 9. For $\theta^{(1)} = 1$, $\phi^{(1)} = \phi^{(0)} = 16$, $0.01 \leq \theta^{(0)} \leq 8$, and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)*}, A_i^{(1)*}\right)\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)\star}, A_i^{(1)\star}\right)\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

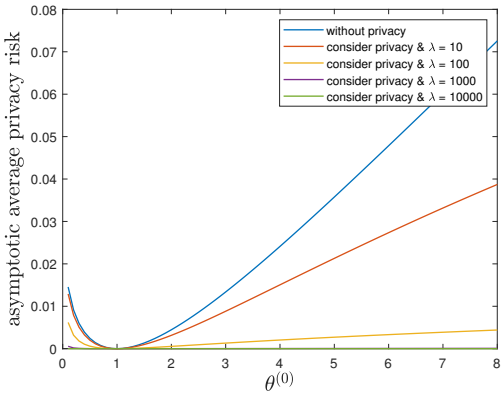

Figure 10. For $\theta^{(1)} = 1$, $\phi^{(1)} = \phi^{(0)} = 16$, $0.01 \leq \theta^{(0)} \leq 8$, and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)*}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)\star}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

6.6. Impact of Parameter $\phi^{(0)}$

By fixing $\phi^{(1)} = 16$ and $\theta^{(1)} = \theta^{(0)} = 1$, we study the impact of the parameter $\phi^{(0)}$ on the control reward of Agent B and the privacy risk. In addition to the default model parameters in Table 3, we allow $0.01 \leq \phi^{(0)} \leq 40$ and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000. From Figures 11 and 12, we have similar observations of the impact of $\phi^{(0)}$ as in Section 6.5. These observations here can be similarly explained as well.

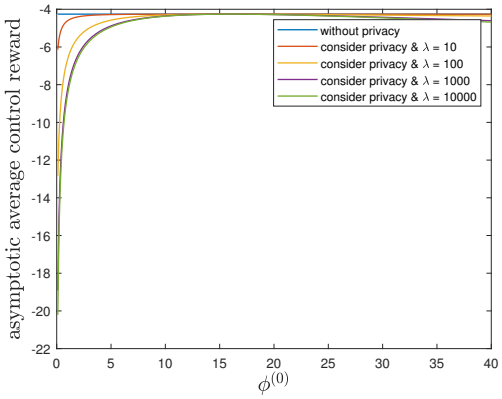

Figure 11. For $\theta^{(1)} = \theta^{(0)} = 1$, $\phi^{(1)} = 16$, $0.01 \leq \phi^{(0)} \leq 40$, and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)*}, A_i^{(1)*}\right)\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average control reward $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^{N} R^{(1)}\left(S_i^{(1)\star}, A_i^{(1)\star}\right)\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

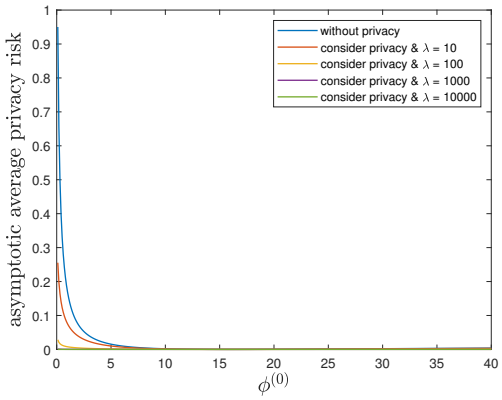

Figure 12. For $\theta^{(1)} = \theta^{(0)} = 1$, $\phi^{(1)} = 16$, $0.01 \leq \phi^{(0)} \leq 40$, and $\lambda = 0$ (without privacy), 10, 100, 1000 or 10,000, comparison of the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)*}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal LQG policy of Agent B and the asymptotic average privacy risk $\lim_{N\to\infty} \frac{1}{N}\mathbb{D}\left(p_{S_{1:N}^{(1)\star}} || p_{S_{1:N}^{(0)*}}\right)$ achieved by the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B.

7. Conclusions

In this paper, we consider the agent identity privacy problem in the scalar LQG control. Regarding this novel privacy problem, we model it as an adversarial binary hypothesis testing and employ the Kullback–Leibler divergence to measure the privacy risk. We then formulate a novel privacy-preserving LQG control optimization by taking into account both the accumulative control reward of Agent B and the privacy risk. We prove that the optimal deterministic privacy-preserving LQG control policy of Agent B is a linear mapping, which is consistent with the standard LQG. We further show that the random policy formulated by adding an independent Gaussian random process noise to the optimal deterministic privacy-preserving LQG policy cannot improve the performance. We also give a sufficient condition to guarantee the time-invariant optimal deterministic privacy-preserving LQG policy in the asymptotic regime.

This research can be extended in our future works. Studying the general random policy of Agent B is an interesting extension. This theoretic study can be extended to develop privacy-preserving reinforcement learning algorithms. The problem can also be

extended and formulated as a non-cooperative game of multiple agents with conflicting objectives, where some agents only aim to optimize their own accumulative control rewards while the other agents consider the agent identity privacy risk in addition to their own accumulative control rewards.

Author Contributions: Conceptualization, E.F. and Z.L.; methodology, E.F., Y.T. and Z.L.; validation, E.F., Y.T. and C.S.; formal analysis, E.F., Y.T. and Z.L.; experiment, C.S.; writing—original draft preparation, E.F. and Y.T.; writing—review and editing, Z.L. and C.W.; supervision, Z.L. and C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Natural Science Foundation of China (62006173, 62171322) and the 2021-2023 China-Serbia Inter-Governmental S&T Cooperation Project (No. 6). We are also grateful for the support of the Sino-German Center of Intelligent Systems, Tongji University.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Theorem 1. The proof is based on the backward dynamic programming.

We first consider the sub-problem of the final step. Given a probability distribution $p_{S_N^{(1)}}$, the final step optimization problem of the deterministic control policy $F_N^{(1)}$ is

$$\begin{aligned} F_N^{(1)\star} &= \arg\max_{F_N^{(1)}} \mathbb{E}\left(R^{(1)}\left(S_N^{(1)}, A_N^{(1)}\right)\right) \\ &= \arg\max_{F_N^{(1)}} -\theta^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(F_N^{(1)}\left(S_N^{(1)}\right)\right)^2. \end{aligned} \tag{A1}$$

Since $p_{S_N^{(1)}}$ is given, the first term $-\theta^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2$ is fixed. Note the upper bound on the second term $-\phi^{(1)}\mathbb{E}\left(F_N^{(1)}\left(S_N^{(1)}\right)\right)^2 \leq 0$. The upper bound can be achieved by the optimal deterministic privacy-preserving LQG policy:

$$F_N^{(1)\star}\left(s_N^{(1)}\right) = \kappa_N^{(1)\star} s_N^{(1)} = 0, \tag{A2}$$

where

$$\kappa_N^{(1)\star} = \frac{\frac{\lambda}{2\omega^2}\beta^2\kappa_N^{(0)*} - \hat{\theta}_{N+1}^{(1)}\alpha\beta}{\phi^{(1)} + \hat{\theta}_{N+1}^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2} = 0 = \kappa_N^{(1)*}. \tag{A3}$$

Then, the maximum achievable objective of the final step is

$$\max_{F_N^{(1)}} \mathbb{E}\left(R^{(1)}\left(S_N^{(1)}, A_N^{(1)}\right)\right) = -\theta^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2 = -\hat{\theta}_N^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2. \tag{A4}$$

We then consider the sub-problem from the $(N-1)$-th step until the final step. Given a probability distribution $p_{S_{N-1}^{(1)}}$ and the optimal deterministic privacy-preserving LQG policy in the final step $F_N^{(1)\star}$, the sub-optimization problem of the deterministic control policy $F_{N-1}^{(1)}$ is

$$
\begin{aligned}
F_{N-1}^{(1)\star} = & \arg\max_{F_{N-1}^{(1)}} \mathbb{E}\left(\sum_{i=N-1}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda \mathbb{D}\left(p_{S_N^{(1)}|S_{N-1}^{(1)}} \,||\, p_{S_N^{(0)*}|S_{N-1}^{(0)*}}\right) \\
= & \arg\max_{F_{N-1}^{(1)}} -\theta^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(A_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2 \\
& \quad -\lambda \mathbb{D}\left(p_{S_N^{(1)}|S_{N-1}^{(1)}} \,||\, p_{S_N^{(0)*}|S_{N-1}^{(0)*}}\right) \\
= & \arg\max_{F_{N-1}^{(1)}} -\theta^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(F_{N-1}^{(1)}\left(S_{N-1}^{(1)}\right)\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2 \\
& \quad -\lambda \mathbb{E}\left(\log \frac{\frac{1}{\sqrt{2\pi\omega^2}} \exp\left(-\frac{\left(S_N^{(1)} - \alpha S_{N-1}^{(1)} - \beta F_{N-1}^{(1)}\left(S_{N-1}^{(1)}\right)\right)^2}{2\omega^2}\right)}{\frac{1}{\sqrt{2\pi\omega^2}} \exp\left(-\frac{\left(S_N^{(1)} - \alpha S_{N-1}^{(1)} - \beta \kappa_{N-1}^{(0)*} S_{N-1}^{(1)}\right)^2}{2\omega^2}\right)}\right) \\
= & \arg\max_{F_{N-1}^{(1)}} -\theta^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(F_{N-1}^{(1)}\left(S_{N-1}^{(1)}\right)\right)^2 \\
& \quad -\hat{\theta}_N^{(1)}\mathbb{E}\left(\alpha S_{N-1}^{(1)} + \beta F_{N-1}^{(1)}\left(S_{N-1}^{(1)}\right) + Z_{N-1}\right)^2 \\
& \quad -\tfrac{\lambda}{2\omega^2}\beta^2\mathbb{E}\left(F_{N-1}^{(1)}\left(S_{N-1}^{(1)}\right) - \kappa_{N-1}^{(0)*} S_{N-1}^{(1)}\right)^2 \\
= & \arg\max_{F_{N-1}^{(1)}} -\left(\theta^{(1)} + \hat{\theta}_N^{(1)}\alpha^2 + \tfrac{\lambda}{2\omega^2}\beta^2\left(\kappa_{N-1}^{(0)*}\right)^2\right)\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}(Z_{N-1})^2 \\
& \quad -\left(\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\mathbb{E}\left(F_{N-1}^{(1)}\left(S_{N-1}^{(1)}\right)\right)^2 \\
& \quad -\left(2\hat{\theta}_N^{(1)}\alpha\beta - \tfrac{\lambda}{\omega^2}\beta^2\kappa_{N-1}^{(0)*}\right)\mathbb{E}\left(S_{N-1}^{(1)} F_{N-1}^{(1)}\left(S_{N-1}^{(1)}\right)\right) \\
= & \arg\max_{F_{N-1}^{(1)}} \int_{\mathbb{R}} \Bigg[-\left(\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\left(F_{N-1}^{(1)}\left(s_{N-1}^{(1)}\right)\right)^2 \\
& \quad -\left(2\hat{\theta}_N^{(1)}\alpha\beta - \tfrac{\lambda}{\omega^2}\beta^2\kappa_{N-1}^{(0)*}\right) s_{N-1}^{(1)} F_{N-1}^{(1)}\left(s_{N-1}^{(1)}\right)\Bigg] p_{S_{N-1}^{(1)}}\left(s_{N-1}^{(1)}\right) \mathrm{d}s_{N-1}^{(1)} \\
= & \int_{\mathbb{R}} \arg\max_{F_{N-1}^{(1)}} \Bigg[-\left(\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\left(F_{N-1}^{(1)}\left(s_{N-1}^{(1)}\right)\right)^2 \\
& \quad -\left(2\hat{\theta}_N^{(1)}\alpha\beta - \tfrac{\lambda}{\omega^2}\beta^2\kappa_{N-1}^{(0)*}\right) s_{N-1}^{(1)} F_{N-1}^{(1)}\left(s_{N-1}^{(1)}\right)\Bigg] p_{S_{N-1}^{(1)}}\left(s_{N-1}^{(1)}\right) \mathrm{d}s_{N-1}^{(1)}.
\end{aligned} \tag{A5}
$$

Since $\hat{\theta}_N^{(1)} = \theta^{(1)} > 0$, it follows that

$$
-\left(\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2\right) < 0. \tag{A6}
$$

Given any $s_{N-1}^{(1)} \in \mathbb{R}$, the objective of the inner optimization in (A5) is a concave quadratic function of $F_{N-1}^{(1)}(s_{N-1}^{(1)})$. Therefore, we can obtain the optimal deterministic privacy-preserving LQG policy as

$$
F_{N-1}^{(1)\star}\left(s_{N-1}^{(1)}\right) = \kappa_{N-1}^{(1)\star} s_{N-1}^{(1)}, \tag{A7}
$$

where

$$
\kappa_{N-1}^{(1)\star} = \frac{\frac{\lambda}{2\omega^2}\beta^2\kappa_{N-1}^{(0)*} - \hat{\theta}_N^{(1)}\alpha\beta}{\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2}. \tag{A8}
$$

By using the optimal deterministic policies $F_{N-1:N}^{(1)\star}$, the maximum achievable objective of the sub-problem is

$$
\begin{aligned}
& \max_{F_{N-1:N}^{(1)}} \mathbb{E}\left(\sum_{i=N-1}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda \mathbb{D}\left(p_{S_N^{(1)}|S_{N-1}^{(1)}} \,||\, p_{S_N^{(0)*}|S_{N-1}^{(0)*}}\right) \\
& \quad = -\hat{\theta}_{N-1}^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}(Z_{N-1})^2.
\end{aligned} \tag{A9}
$$

The coefficient $\hat{\theta}_{N-1}^{(1)}$ can be specified as

$$\hat{\theta}_{N-1}^{(1)} = \theta^{(1)} + \frac{\phi^{(1)}\hat{\theta}_N^{(1)}\alpha^2 + \frac{\lambda}{2\omega^2}\beta^2\left(\kappa_{N-1}^{(0)*}\right)^2\phi^{(1)} + \frac{\lambda}{2\omega^2}\beta^2\hat{\theta}_N^{(1)}\left(\alpha+\beta\kappa_{N-1}^{(0)*}\right)^2}{\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2}. \tag{A10}$$

It can be easily justified that $\hat{\theta}_{N-1}^{(1)} > 0$ since $\hat{\theta}_N^{(1)} > 0$.

We now consider the sub-problem from the $(N-2)$-th step until the final step. Given a probability distribution $p_{S_{N-2}^{(1)}}$ and the optimal deterministic privacy-preserving LQG policies $F_{N-1:N}^{(1)\star}$, the sub-optimization problem of the deterministic control policy $F_{N-2}^{(1)}$ is

$$\begin{aligned}
F_{N-2}^{(1)\star} = & \arg\max_{F_{N-2}^{(1)}} \mathbb{E}\left(\sum_{i=N-2}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda\sum_{i=N-1}^{N}\mathbb{D}\left(p_{S_i^{(1)}|S_{i-1}^{(1)}}||p_{S_i^{(0)*}|S_{i-1}^{(0)*}}\right) \\
= & \arg\max_{F_{N-2}^{(1)}} -\theta^{(1)}\mathbb{E}\left(S_{N-2}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(A_{N-2}^{(1)}\right)^2 - \hat{\theta}_{N-1}^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}(Z_{N-1})^2 \\
& -\lambda\mathbb{D}\left(p_{S_{N-1}^{(1)}|S_{N-2}^{(1)}}||p_{S_{N-1}^{(0)*}|S_{N-2}^{(0)*}}\right) \\
= & \arg\max_{F_{N-2}^{(1)}} -\theta^{(1)}\mathbb{E}\left(S_{N-2}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(F_{N-2}^{(1)}\left(S_{N-2}^{(1)}\right)\right)^2 \\
& -\hat{\theta}_{N-1}^{(1)}\mathbb{E}\left(\alpha S_{N-2}^{(1)} + \beta F_{N-2}^{(1)}\left(S_{N-2}^{(1)}\right) + Z_{N-2}\right)^2 \\
& -\tfrac{\lambda}{2\omega^2}\beta^2\mathbb{E}\left(F_{N-2}^{(1)}\left(S_{N-2}^{(1)}\right) - \kappa_{N-2}^{(0)*}S_{N-2}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}(Z_{N-1})^2 \\
= & \int_{\mathbb{R}} \arg\max_{F_{N-2}^{(1)}} \Bigg[-\left(\phi^{(1)} + \hat{\theta}_{N-1}^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\left(F_{N-2}^{(1)}\left(s_{N-2}^{(1)}\right)\right)^2 \\
& -\left(2\hat{\theta}_{N-1}^{(1)}\alpha\beta - \tfrac{\lambda}{\omega^2}\beta^2\kappa_{N-2}^{(0)*}\right)s_{N-2}^{(1)}F_{N-2}^{(1)}\left(s_{N-2}^{(1)}\right)\Bigg] p_{S_{N-2}^{(1)}}\left(s_{N-2}^{(1)}\right)\mathrm{d}s_{N-2}^{(1)}.
\end{aligned} \tag{A11}$$

Note that the objective functions in (A5) and (A11) have the same form. We have also proved that $\hat{\theta}_{N-1}^{(1)} > 0$. Therefore, we can use the same arguments to obtain the optimal deterministic privacy-preserving LQG policy as

$$F_{N-2}^{(1)\star}\left(s_{N-2}^{(1)}\right) = \kappa_{N-2}^{(1)\star}s_{N-2}^{(1)}, \tag{A12}$$

where

$$\kappa_{N-2}^{(1)\star} = \frac{\frac{\lambda}{2\omega^2}\beta^2\kappa_{N-2}^{(0)*} - \hat{\theta}_{N-1}^{(1)}\alpha\beta}{\phi^{(1)} + \hat{\theta}_{N-1}^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2}, \tag{A13}$$

the maximum achievable objective of the sub-problem as

$$\begin{aligned}
&\max_{F_{N-2:N}^{(1)}} \mathbb{E}\left(\sum_{i=N-2}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda\sum_{i=N-1}^{N}\mathbb{D}\left(p_{S_i^{(1)}|S_{i-1}^{(1)}}||p_{S_i^{(0)*}|S_{i-1}^{(0)*}}\right) \\
&\quad = -\hat{\theta}_{N-2}^{(1)}\mathbb{E}\left(S_{N-2}^{(1)}\right)^2 - \hat{\theta}_{N-1}^{(1)}\mathbb{E}(Z_{N-2})^2 - \hat{\theta}_N^{(1)}\mathbb{E}(Z_{N-1})^2,
\end{aligned} \tag{A14}$$

and $\hat{\theta}_{N-2}^{(1)} > 0$.

We can further prove the optimal deterministic privacy-preserving LQG policies in the remaining steps and the maximum achievable weighted design objective of Agent B in Theorem 1 using the same arguments. □

Appendix B

Proof of Theorem 2. The proof is based on the optimal deterministic privacy-preserving LQG policy of Agent B in Theorem 1.

For all $1 \leq i \leq \left\lceil \frac{N}{2} \right\rceil$ and $x \geq 0$, let

$$J(x) = \lim_{N\to\infty} J_{N+1-i}(x) = \theta^{(1)} + \alpha^2 x + \frac{\lambda}{2\omega^2}\beta^2\left(\kappa^{(0)*}\right)^2 - \frac{\left(\frac{\lambda}{2\omega^2}\beta^2\kappa^{(0)*} - \alpha\beta x\right)^2}{\phi^{(1)} + \beta^2 x + \frac{\lambda}{2\omega^2}\beta^2}, \quad \text{(A15)}$$

where the second equality follows from

$$\begin{aligned}
\lim_{N\to\infty} \kappa_i^{(0)*} &= \lim_{N\to\infty} -\frac{\tilde{\theta}_{i+1}^{(0)}\alpha\beta}{\phi^{(0)} + \tilde{\theta}_{i+1}^{(0)}\beta^2} \\
&= -\frac{\alpha\beta \lim_{N\to\infty} \underbrace{L^{(0)}(L^{(0)}(\cdots(L^{(0)}(L^{(0)}}_{N-i \text{ iterations}}(\tilde{\theta}_{N+1}^{(0)})))\cdots))}{\phi^{(0)} + \beta^2 \lim_{N\to\infty} \underbrace{L^{(0)}(L^{(0)}(\cdots(L^{(0)}(L^{(0)}}_{N-i \text{ iterations}}(\tilde{\theta}_{N+1}^{(0)})))\cdots))} \\
&= -\frac{\tilde{\theta}^{(0)}\alpha\beta}{\phi^{(0)} + \tilde{\theta}^{(0)}\beta^2} \\
&= \kappa^{(0)*}.
\end{aligned} \quad \text{(A16)}$$

When the model parameters satisfy the condition in (22), $J(x)$ is a contraction mapping, i.e., there exists $0 < \gamma < 1$ such that $|J(x) - J(x')| \leq \gamma|x - x'|$ for all $x \geq 0$ and $x' \geq 0$. From the Banach's fixed point theorem, there is a unique fixed point $\hat{\theta}^{(1)}$ with respect to the contraction mapping J such that

$$\begin{aligned}
\hat{\theta}^{(1)} &= \lim_{N\to\infty} J_N(J_{N-1}(\cdots(J_2(J_1(\hat{\theta}_{N+1}^{(1)})))\cdots)) \\
&= \lim_{N\to\infty} J_N(J_{N-1}(\cdots(J_{N+2-\lceil \frac{N}{2} \rceil}(J_{N+1-\lceil \frac{N}{2} \rceil}(\hat{\theta}_{\lceil \frac{N}{2} \rceil+1}^{(1)})))\cdots)) \\
&= \lim_{N\to\infty} \underbrace{J(J(\cdots(J(J}_{\lceil \frac{N}{2} \rceil \text{ iterations}}(\hat{\theta}_{\lceil \frac{N}{2} \rceil+1}^{(H)})))\cdots)) \\
&= \theta^{(1)} + \alpha^2\hat{\theta}^{(1)} + \frac{\lambda}{2\omega^2}\beta^2\left(\kappa^{(0)*}\right)^2 - \frac{\left(\frac{\lambda}{2\omega^2}\beta^2\kappa^{(0)*} - \alpha\beta\hat{\theta}^{(1)}\right)^2}{\phi^{(1)} + \beta^2\hat{\theta}^{(1)} + \frac{\lambda}{2\omega^2}\beta^2}.
\end{aligned} \quad \text{(A17)}$$

From (19)–(21), (A16), and (A17), it is easy to verify the time-invariant optimal deterministic privacy-preserving LQG policy of Agent B in (24)–(25) and the asymptotic weighted design objective rate in (26). □

Appendix C

Proof of Theorem 3. The proof is similar as that of Theorem 1.

We first consider the sub-problem of the final step. Given a probability distribution $p_{S_N^{(1)}}$, the final step optimization problem of the linear Gaussian random policy $F_N^{(1)}$ with parameters $\kappa_N^{(1)}$ and δ_N^2 is

$$\begin{aligned}
\left(\kappa_N^{(1)\star}, \delta_N^{2\star}\right) &= \arg\max_{\kappa_N^{(1)}\in\mathbb{R}, \delta_N^2\in\mathbb{R}_{\geq 0}} \mathbb{E}\left(R^{(1)}\left(S_N^{(1)}, A_N^{(1)}\right)\right) \\
&= \arg\max_{\kappa_N^{(1)}\in\mathbb{R}, \delta_N^2\in\mathbb{R}_{\geq 0}} -\theta^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(A_N^{(1)}\right)^2 \\
&= \arg\max_{\kappa_N^{(1)}\in\mathbb{R}, \delta_N^2\in\mathbb{R}_{\geq 0}} \left(-\theta^{(1)} - \phi^{(1)}\left(\kappa_N^{(1)}\right)^2\right)\mathbb{E}\left(S_N^{(1)}\right)^2 - \phi^{(1)}\delta_N^2.
\end{aligned} \quad \text{(A18)}$$

It is obvious the optimal parameters are $\kappa_N^{(1)\star} = 0$ and $\delta_N^{2\star} = 0$, i.e., the optimal linear Gaussian random policy reduces to the optimal deterministic privacy-preserving LQG policy in the final step.

Similarly, we then consider the sub-problem from the $(N-1)$-th step until the final step. Given a probability distribution $p_{S_{N-1}^{(1)}}$ and the optimal linear Gaussian random policy

in the final step $F_N^{(1)\star}$, the sub-optimization problem of the linear Gaussian random policy $F_{N-1}^{(1)}$ with parameters $\kappa_{N-1}^{(1)}$ and δ_{N-1}^2 is

$$
\begin{aligned}
&\left(\kappa_{N-1}^{(1)\star}, \delta_{N-1}^{2\star}\right) \\
&= \arg\max_{\kappa_{N-1}^{(1)}\in\mathbb{R}, \delta_{N-1}^2\in\mathbb{R}_{\geq 0}} \mathbb{E}\left(\sum_{i=N-1}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda\mathbb{D}\left(p_{S_N^{(1)}|S_{N-1}^{(1)}} \middle|\middle| p_{S_N^{(0)*}|S_{N-1}^{(0)*}}\right) \\
&= \arg\max_{\kappa_{N-1}^{(1)}\in\mathbb{R}, \delta_{N-1}^2\in\mathbb{R}_{\geq 0}} -\theta^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(A_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2 \\
&\qquad -\lambda\mathbb{D}\left(p_{S_N^{(1)}|S_{N-1}^{(1)}} \middle|\middle| p_{S_N^{(0)*}|S_{N-1}^{(0)*}}\right) \\
&= \arg\max_{\kappa_{N-1}^{(1)}\in\mathbb{R}, \delta_{N-1}^2\in\mathbb{R}_{\geq 0}} -\theta^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(A_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\mathbb{E}\left(S_N^{(1)}\right)^2 \\
&\qquad -\lambda\mathbb{E}\left(\log \frac{\frac{1}{\sqrt{2\pi\left(\omega^2+\beta^2\delta_{N-1}^2\right)}}\exp\left(-\frac{\left(S_N^{(1)}-\alpha S_{N-1}^{(1)}-\beta\kappa_{N-1}^{(1)}S_{N-1}^{(1)}\right)^2}{2\left(\omega^2+\beta^2\delta_{N-1}^2\right)}\right)}{\frac{1}{\sqrt{2\pi\omega^2}}\exp\left(-\frac{\left(S_N^{(1)}-\alpha S_{N-1}^{(1)}-\beta\kappa_{N-1}^{(0)*}S_{N-1}^{(1)}\right)^2}{2\omega^2}\right)}\right) \\
&= \arg\max_{\kappa_{N-1}^{(1)}\in\mathbb{R}, \delta_{N-1}^2\in\mathbb{R}_{\geq 0}} -\theta^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(\kappa_{N-1}^{(1)}S_{N-1}^{(1)} + W_{N-1}^{(1)}\right)^2 \\
&\qquad -\hat{\theta}_N^{(1)}\mathbb{E}\left(\alpha S_{N-1}^{(1)} + \beta\kappa_{N-1}^{(1)}S_{N-1}^{(1)} + \beta W_{N-1}^{(1)} + Z_{N-1}\right)^2 \\
&\qquad -\tfrac{\lambda}{2\omega^2}\beta^2\left(\kappa_{N-1}^{(1)} - \kappa_{N-1}^{(0)*}\right)^2\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \tfrac{\lambda}{2\omega^2}\beta^2\delta_{N-1}^2 \\
&\qquad -\tfrac{\lambda}{2}\log\tfrac{\omega^2}{\omega^2+\beta^2\delta_{N-1}^2} \\
&= \arg\max_{\kappa_{N-1}^{(1)}\in\mathbb{R}, \delta_{N-1}^2\in\mathbb{R}_{\geq 0}} -\left(\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\mathbb{E}\left(S_{N-1}^{(1)}\right)^2\left(\kappa_{N-1}^{(1)}\right)^2 \\
&\qquad -\left(2\hat{\theta}_N^{(1)}\alpha\beta - \tfrac{\lambda}{\omega^2}\beta^2\kappa_{N-1}^{(0)*}\right)\mathbb{E}\left(S_{N-1}^{(1)}\right)^2\kappa_{N-1}^{(1)} \\
&\qquad -\left(\theta^{(1)} + \hat{\theta}_N^{(1)}\alpha^2 + \tfrac{\lambda}{2\omega^2}\beta^2\left(\kappa_{N-1}^{(0)*}\right)^2\right)\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\omega^2 \\
&\qquad -\left(\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\delta_{N-1}^2 - \tfrac{\lambda}{2}\log\tfrac{\omega^2}{\omega^2+\beta^2\delta_{N-1}^2}.
\end{aligned}
\tag{A19}
$$

(A19) consists of two independent optimizations: the optimization of $\kappa_{N-1}^{(1)} \in \mathbb{R}$ and the optimization of $\delta_{N-1}^2 \in \mathbb{R}_{\geq 0}$. Since $\hat{\theta}_N^{(1)} > 0$, it follows that

$$
-\left(\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2\right)\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 < 0. \tag{A20}
$$

The optimization of $\kappa_{N-1}^{(1)} \in \mathbb{R}$ has a concave quadratic objective. Then we can obtain the optimal linear coefficient as

$$
\kappa_{N-1}^{(1)\star} = \frac{\frac{\lambda}{2\omega^2}\beta^2\kappa_{N-1}^{(0)*} - \hat{\theta}_N^{(1)}\alpha\beta}{\phi^{(1)} + \hat{\theta}_N^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2}. \tag{A21}
$$

The optimization of $\delta_{N-1}^2 \in \mathbb{R}_{\geq 0}$ has a decreasing objective. Then, the optimal variance is $\delta_{N-1}^{2\star} = 0$. Therefore, the optimal linear Gaussian random policy reduces to the optimal deterministic privacy-preserving LQG policy in the $(N-1)$-th step.

We then consider the sub-problem from the $(N-2)$-th step until the final step. Given a probability distribution $p_{S_{N-2}^{(1)}}$ and the optimal linear Gaussian random policies $F_{N-1:N}^{(1)\star}$, the sub-optimization problem of the linear Gaussian random policy $F_{N-2}^{(1)}$ with parameters $\kappa_{N-2}^{(1)}$ and δ_{N-2}^2 is

$$
\begin{aligned}
&\left(\kappa_{N-2}^{(1)\star}, \delta_{N-2}^{2\star}\right) \\
&= \arg\max_{\kappa_{N-2}^{(1)}\in\mathbb{R}, \delta_{N-2}^2\in\mathbb{R}_{\geq 0}} \mathbb{E}\left(\sum_{i=N-2}^{N} R^{(1)}\left(S_i^{(1)}, A_i^{(1)}\right)\right) - \lambda \sum_{i=N-1}^{N} \mathbb{D}\left(p_{S_i^{(1)}|S_{i-1}^{(1)}} \middle|\middle| p_{S_i^{(0)*}|S_{i-1}^{(0)*}}\right) \\
&= \arg\max_{\kappa_{N-2}^{(1)}\in\mathbb{R}, \delta_{N-2}^2\in\mathbb{R}_{\geq 0}} -\theta^{(1)}\mathbb{E}\left(S_{N-2}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(A_{N-2}^{(1)}\right)^2 - \hat{\theta}_{N-1}^{(1)}\mathbb{E}\left(S_{N-1}^{(1)}\right)^2 - \hat{\theta}_N^{(1)}\omega^2 \\
&\qquad -\lambda\mathbb{D}\left(p_{S_{N-1}^{(1)}|S_{N-2}^{(1)}} \middle|\middle| p_{S_{N-1}^{(0)*}|S_{N-2}^{(0)*}}\right) \\
&= \arg\max_{\kappa_{N-2}^{(1)}\in\mathbb{R}, \delta_{N-2}^2\in\mathbb{R}_{\geq 0}} -\theta^{(1)}\mathbb{E}\left(S_{N-2}^{(1)}\right)^2 - \phi^{(1)}\mathbb{E}\left(\kappa_{N-2}^{(1)}S_{N-2}^{(1)} + W_{N-2}^{(1)}\right)^2 \\
&\qquad -\hat{\theta}_{N-1}^{(1)}\mathbb{E}\left(\alpha S_{N-2}^{(1)} + \beta\kappa_{N-2}^{(1)}S_{N-2}^{(1)} + \beta W_{N-2}^{(1)} + Z_{N-2}\right)^2 - \hat{\theta}_N^{(1)}\omega^2 \\
&\qquad -\tfrac{\lambda}{2\omega^2}\beta^2\left(\kappa_{N-2}^{(1)} - \kappa_{N-2}^{(0)*}\right)^2\mathbb{E}\left(S_{N-2}^{(1)}\right)^2 - \tfrac{\lambda}{2\omega^2}\beta^2\delta_{N-2}^2 \\
&\qquad -\tfrac{\lambda}{2}\log\tfrac{\omega^2}{\omega^2+\beta^2\delta_{N-2}^2} \\
&= \arg\max_{\kappa_{N-2}^{(1)}\in\mathbb{R}, \delta_{N-2}^2\in\mathbb{R}_{\geq 0}} -\left(\phi^{(1)} + \hat{\theta}_{N-1}^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\mathbb{E}\left(S_{N-2}^{(1)}\right)^2\left(\kappa_{N-2}^{(1)}\right)^2 \\
&\qquad -\left(2\hat{\theta}_{N-1}^{(1)}\alpha\beta - \tfrac{\lambda}{\omega^2}\beta^2\kappa_{N-2}^{(0)*}\right)\mathbb{E}\left(S_{N-2}^{(1)}\right)^2\kappa_{N-2}^{(1)} \\
&\qquad -\left(\theta^{(1)} + \hat{\theta}_{N-1}^{(1)}\alpha^2 + \tfrac{\lambda}{2\omega^2}\beta^2\left(\kappa_{N-2}^{(0)*}\right)^2\right)\mathbb{E}\left(S_{N-2}^{(1)}\right)^2 - \hat{\theta}_{N-1}^{(1)}\omega^2 - \hat{\theta}_N^{(1)}\omega^2 \\
&\qquad -\left(\phi^{(1)} + \hat{\theta}_{N-1}^{(1)}\beta^2 + \tfrac{\lambda}{2\omega^2}\beta^2\right)\delta_{N-2}^2 - \tfrac{\lambda}{2}\log\tfrac{\omega^2}{\omega^2+\beta^2\delta_{N-2}^2}.
\end{aligned}
\tag{A22}
$$

Note that the objective functions in (A19) and (A22) have the same form. Therefore, we can use the same arguments to show that the optimal linear Gaussian random policy reduces to the optimal deterministic privacy-preserving LQG policy in the $(N-2)$-th step, i.e., $\delta_{N-2}^{2\star} = 0$ and

$$
\kappa_{N-2}^{(1)\star} = \frac{\frac{\lambda}{2\omega^2}\beta^2\kappa_{N-2}^{(0)*} - \hat{\theta}_{N-1}^{(1)}\alpha\beta}{\phi^{(1)} + \hat{\theta}_{N-1}^{(1)}\beta^2 + \frac{\lambda}{2\omega^2}\beta^2}. \tag{A23}
$$

We can further prove the optimal linear Gaussian random policies in the remaining steps reduce to the optimal deterministic privacy-preserving LQG policies based on the same arguments. □

References

1. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjel, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533.
2. Huang, S.; Papernot, N.; Goodfellow, I.; Duan, Y.; Abbeel, P. Adversarial attacks on neural network policies. *arXiv* **2016**, arXiv:1702.02284.
3. Lin, Y.C.; Hong, Z.W.; Liao, Y.H.; Shih, M.L.; Liu, M.Y.; Min, S. Tactics of adversarial attack on deep reinforcement learning agents. In Proceedings of the 2017 International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 3756–3762.
4. Behzadan, V.; Munir, A. Vulnerability of deep reinforcement learning to policy induction attacks. In Proceedings of the MLDM 2017, New York, NY, USA, 15–20 July 2017; pp. 262–275.
5. Russo, A.; Proutiere, A. Optimal attacks on reinforcement learning policies. *arXiv* **2019**, arXiv:1907.13548.
6. Goodfellow, I.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. In Proceedings of the ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
7. Tramer, F.; Kurakin, A.; Papernot, N.; Goodfellow, I.; Boneh, D.; McDaniel, P. Ensemble adversarial training: Attacks and defenses. In Proceedings of the ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
8. Sinha, A.; Namkoong, H.; Duchi, J. Certifying some distributional robustness with principled adversarial training. In Proceedings of the ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
9. Zheng, S.; Song, Y.; Leung, T.; Goodfellow, I. Improving the robustness of deep neural networks via stability training. In Proceedings of the CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016.
10. Yan, Z.; Guo, Y.; Zhang, C. Deep defense: Training DNNs with improved adversarial robustness. In Proceedings of the NIPS 2018, Montréal, QC, Canada, 3–8 December 2018.
11. Shapley, L. Stochastic games. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 1095–1100.

12. Gleave, A.; Dennis, M.; Wild, C.; Kant, N.; Levine, S.; Russell, S. Adversarial policies: Attacking deep reinforcement learning. In Proceedings of the ICLR 2020, Addis Ababa, Ethiopia, 26–30 April 2020.
13. Pinto, L.; Davidson, J.; Sukthankar, R.; Gupta, A. Robust adversarial reinforcement learning. In Proceedings of the ICML 2017, Sydney, NSW, Australia, 6–11 August 2017; pp. 2817–2826.
14. Horak, K.; Zhu, Q.; Bosansky, B. Manipulating adversary's belief: A dynamic game approach to deception by design for proactive network security. In Proceedings of the GameSec 2017, Vienna, Austria, 23–25 October 2017; pp. 273–294.
15. Crawford, V.P.; Sobel, J. Strategic information transmission. *Econometrica* **1982**, *50*, 1431–1451.
16. Saritas, S.; Yuksel, S.; Gezici, S. Nash and Stackelberg equilibria for dynamic cheap talk and signaling games. In Proceedings of the ACC 2017, Seattle, WA, USA, 24–26 May 2017; pp. 3644–3649.
17. Saritas, S.; Shereen, E.; Sandberg, H.; Dán, G. Adversarial attacks on continuous authentication security: A dynamic game approach. In Proceedings of the GameSec 2019, Stockholm, Sweden, 30 October–1 November 2019; pp. 439–458.
18. Li, Z.; Dán, G. Dynamic cheap talk for robust adversarial learning. In Proceedings of the GameSec 2019, Stockholm, Sweden, 30 October–1 November 2019; pp. 297–309.
19. Li, Z.; Dán, G.; Liu, D. A game theoretic analysis of LQG control under adversarial attack. In Proceedings of the IEEE CDC 2020, Jeju Island, Korea, 14–18 December 2020; pp. 1632–1639.
20. Osogami, T. Robust partially observable Markov decision process. In Proceedings of the ICML 2015, Lille, France, 6–11 July 2015.
21. Sayin, M.O.; Basar, T. Secure sensor design for cyber-physical systems against advanced persistent threats. In Proceedings of the GameSec 2017, Vienna, Austria, 23–25 October 2017; pp. 91–111.
22. Sayin, M.O.; Akyol, E.; Basar, T. Hierarchical multistage Gaussian signaling games in noncooperative communication and control systems. *Automatica* **2019**, *107*, 9–20.
23. Sun, C.; Li, Z.; Wang, C. Adversarial linear quadratic regulator under falsified actions. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022.
24. Zhang, R.; Venkitasubramaniam, P. Stealthy control signal attacks in linear quadratic Gaussian control systems: Detectability reward tradeoff. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 1555–1570.
25. Ren, X.X.; Yang, G.H. Kullback-Leibler divergence-based optimal stealthy sensor attack against networked linear quadratic Gaussian systems. *IEEE Trans. Cybern.* **2021**. doi:10.1109/TCYB.2021.3068220.
26. Venkitasubramaniam, P. Privacy in stochastic control: A Markov decision process perspective. In Proceedings of the Allerton 2013, Monticello, IL, USA, 2–4 October 2013; pp. 381–388.
27. Ny, J.L.; Pappas, G.J. Differentially private filtering. *IEEE Trans. Autom. Control.* **2013**, *59*, 341–354.
28. Hale, M.T.; Egerstedt, M. Cloud-enabled differentially private multiagent optimization with constraints. *IEEE Trans. Control. Netw. Syst.* **2017**, *5*, 1693–1706.
29. Hale, M.; Jones, A.; Leahy, K. Privacy in feedback: The differentially private LQG. In Proceedings of the ACC 2018, Milwaukee, WI, USA, 27–29 June 2018; pp. 3386–3391.
30. Hawkins, C.; Hale, M. Differentially private formation control. In Proceedings of the IEEE CDC 2020, Jeju Island, Korea, 14–18 December 2020; pp. 6260–6265.
31. Dwork, C. Differential privacy. In Proceedings of the ICALP 2006, Venice, Italy, 10–14 July 2006, pp. 1–12.
32. Wang, B.; Hegde, N. Privacy-preserving Q-learning with functional noise in continuous spaces. In Proceedings of the NeurIPS 2019, Vancouver, BC, Canada, 8–14 December 2019.
33. Alexandru, A.B.; Pappas, G.J. Encrypted LQG using labeled homomorphic encryption. In Proceedings of the ACM/IEEE ICCPS 2019, Montreal, QC, Canada, 16–18 April 2019.
34. Arora, S.; Doshi, P. A survey of inverse reinforcement learning: Challenges, methods and progress. *Artif. Intell.* **2021**, *297*, 103500.
35. Soderstrom, T. *Discrete-Time Stochastic Systems*; Springer: Berlin/Heidelberg, Germany, 2002.
36. Baranga, A. The contraction principle as a particular case of Kleene's fixed point theorem. *Discret. Math.* **1991**, *98*, 75–79.
37. Hershey, J.R.; Olsen, P.A. Approximating the Kullback Leibler divergence between Gaussian mixture models. In Proceedings of the IEEE ICASSP 2007, Honolulu, HI, USA, 15–20 April 2007.
38. Durrieu, J.; Thiran, J.; Kelly, F. Lower and upper bounds for approximation of the Kullback-Leibler divergence between Gaussian mixture models. In Proceedings of the IEEE ICASSP 2012, Kyoto, Japan, 25–30 March 2012.
39. Cui, S.; Datcu, M. Comparison of Kullback-Leibler divergence approximation methods between Gaussian mixture models for satellite image retrieval. In Proceedings of the IEEE IGARSS 2015, Milan, Italy, 26–31 July 2015.