Differential Privacy in Cooperative Multiagent Planning (Supplementary Material)

Bo Chen[*1] **Calvin Hawkins**[*1] **Mustafa O. Karabag**[*2] **Cyrus Neary**[*2] **Matthew Hale**[1] **Ufuk Topcu**[2]

[1] The University of Florida
[2] The University of Texas at Austin

A PROOFS FOR THEORETICAL RESULTS

The Kullback-Leibler (KL) divergence Cover and Thomas [1991] between discrete probability distributions Q^1 and Q^2 with supports $\boldsymbol{Q}^1$ and $\boldsymbol{Q}^2$, respectively, is

$$KL\left(Q^1 \| Q^2\right) = \sum_{q \in \boldsymbol{Q}^1} Q^1(q) \log\left(\frac{Q^1(q)}{Q^2(q)}\right).$$

Notations We first define some notations that will be used for the proofs. Let $\boldsymbol{S}_t$ be a random variable denoting the joint state of the agents at time t under the joint policy with no privatization, $\boldsymbol{A}_t$ be a random variable denoting the joint action of the agents at time t, S_t^i be a random variable denoting the state of agent i at time t, A_t^i be a random variable denoting the action of Agent i at time t. $\boldsymbol{S}_t^{-i}$ be a random variable denoting the state of agent i's teammate exclude agent i itself at time t, and $\boldsymbol{A}_t^{-i}$ be a random variable denoting the action of agent i's teammate exclude agent i itself at time t. The total correlation $C_{\boldsymbol{\pi}}$ of a joint policy $\boldsymbol{\pi}$ is

$$C_{\boldsymbol{\pi}} = \sum_{i=1}^{N} H(S_0^i A_0^i \dots S_\eta^i) - H(\boldsymbol{S}_0 \boldsymbol{A}_0 \dots \boldsymbol{S}_\eta)$$

where η denotes the random hitting time to $\boldsymbol{\mathcal{S}}_{\mathcal{T}} \cup \boldsymbol{\mathcal{S}}_{\mathcal{D}}$, i.e., the effective end of the trajectory in terms of the reach-avoid specification Karabag et al. [2022].

Let W denote all trajectory fragments that end at a state in $\boldsymbol{\mathcal{S}}_{\mathcal{T}} \cup \boldsymbol{\mathcal{S}}_{\mathcal{D}}$, i.e., $W = \{\boldsymbol{v} = \boldsymbol{s_0 a_0} \dots \boldsymbol{s_T} | \boldsymbol{s_T} \in \boldsymbol{\mathcal{S}}_{\mathcal{T}} \cup \boldsymbol{\mathcal{S}}_{\mathcal{D}} \text{ and } \forall t < T, \boldsymbol{s_t} \notin \boldsymbol{\mathcal{S}}_{\mathcal{T}} \cup \boldsymbol{\mathcal{S}}_{\mathcal{D}}\}$, and W' denote all trajectories that never reach $\boldsymbol{\mathcal{S}}_{\mathcal{T}} \cup \boldsymbol{\mathcal{S}}_{\mathcal{D}}$, i.e., $W' = \{w = \boldsymbol{v} = \boldsymbol{s_0 a_0} \dots | \forall t \geq 0, \boldsymbol{s_t} \notin \boldsymbol{\mathcal{S}}_{\mathcal{T}} \cup \boldsymbol{\mathcal{S}}_{\mathcal{D}}\}$. Note that every trajectory either starts with a trajectory fragment from W or is in W'. Also, let $R \subseteq W \cup W'$ denote all trajectory fragments that end at a state in $\boldsymbol{\mathcal{S}}_{\mathcal{T}}$, i.e., $R = \{\boldsymbol{v} = \boldsymbol{s_0 a_0} \dots \boldsymbol{s_T} | \boldsymbol{s_T} \in \boldsymbol{\mathcal{S}}_{\mathcal{T}} \text{ and } \forall t < T, \boldsymbol{s_t} \notin \boldsymbol{\mathcal{S}}_{\mathcal{T}} \cup \boldsymbol{\mathcal{S}}_{\mathcal{A}}\}$.

Let Γ^{tr} be the distribution of joint trajectories induced by the joint policy executed with truthful communications (i.e., no privacy). Also, let Γ^{pr} be the distribution of joint trajectories with privacy enforced. Let v^{tr} be the probability of success under truthful communications and v^{pr} be the probability of success under private communications.

We use μ^{tr} to denote the probability measure over the actual (finite or infinite) state-action process under the joint policy with truthful communications. μ^{pr} denotes the probability measure over the actual (finite or infinite) state-action process under joint policy with private communications. With abuse of notation, we also use μ_ϵ to denote the conditional probability measure over private state trajectories given the actual state trajectory.

Let $\boldsymbol{v} = \boldsymbol{s}_0 \boldsymbol{a}_0 \boldsymbol{s}_1 \boldsymbol{a}_1 \dots \boldsymbol{s}_T \in (\boldsymbol{\mathcal{S}} \times \boldsymbol{\mathcal{A}})^T$ be a joint trajectory fragment and $\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{s}}_0 \tilde{\boldsymbol{s}}_1 \dots \tilde{\boldsymbol{s}}_T \in \boldsymbol{\mathcal{S}}^T$ be a private joint state trajectory fragment. We use $\hat{\boldsymbol{s}}_{\boldsymbol{t}}^{\boldsymbol{j}} = \{\tilde{s}_t^0, \dots, \tilde{s}_t^{j-1}, s_t^j, \tilde{s}_t^{j+1}, \dots, \tilde{s}_t^N\}$ to denote agent j's copy of private joint state.

[*] Indicates equal contribution.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

The Kleene star applied to a set V of symbols is the set $V^* = \cup_{i\geq 0} V^i$ of all finite-length words where $V^0 = \{\Lambda\}$ and Λ is the empty string. The set of all infinite-length words is denoted by V^ω.

We introduce the following lemma, to be used in the proof of other theoretical results.

Lemma 1.

$$\mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\log \mu_\epsilon(\tilde{\boldsymbol{z}} = \boldsymbol{v}|\boldsymbol{v})\right] \geq -N \log\left((\rho_{\max} - 1)\exp(-\frac{\epsilon}{k}) + 1\right) l^{tr}$$

where $\rho(s^i_{t-1})$ is the out degree of s^i_{t-1} and $\rho_{\max} = \max_{s\in\cup_{i=1}^N \mathcal{S}^i} \rho(s)$.

Proof of Lemma 1. Due to the Markovianity of the online privacy mechanism (Algorithm 1), independence between the agents, and the acyclic property of the policy graph G, we have

$$\mu_\epsilon(\tilde{\boldsymbol{z}} = \boldsymbol{v}|\boldsymbol{v} = \boldsymbol{s}_0 \dots \boldsymbol{s}_T) = \prod_{t=0}^{T-1}\prod_{i=1}^{N} \mu_\epsilon(\tilde{s}^i_t = s^i_t | s^i_t, \tilde{s}^i_{t-1})$$

We note that if $\tilde{\boldsymbol{z}} = \boldsymbol{v}$, then for all $t \geq 0$ and $j \in [N]$, we have $\hat{\boldsymbol{s}}^j_t = \boldsymbol{s}_t$, i.e., the copy of the private state for every agent always matches the actual joint state. Hence,

$$\mu_\epsilon(\tilde{\boldsymbol{z}} = \boldsymbol{v}|\boldsymbol{v} = \boldsymbol{s}_0 \dots \boldsymbol{s}_T) = \prod_{t=0}^{T-1}\prod_{i=1}^{N} \mu_\epsilon(\tilde{s}^i_t = s^i_t | s^i_t, \tilde{s}^i_{t-1} = s^i_{t-1})$$

From [Chen et al., 2023, Theorem 7], we have

$$\mu_\epsilon(\tilde{s}^i_t = s^i_t | s^i_t, \tilde{s}^i_{t-1} = s^i_{t-1}) = \frac{1}{\left(\rho(s^i_{t-1}) - 1\right)\exp(-\frac{\epsilon}{k}) + 1},$$

where $\rho(s^i_{t-1})$ is the out degree of s^i_{t-1}. Let $\rho_{\max} = \max_{s\in\cup_{i=1}^N \mathcal{S}^i} \rho(s)$ which gives

$$\mu_\epsilon(\tilde{s}^i_t = s^i_t | s^i_t, \tilde{s}^i_{t-1} = s^i_{t-1}) \geq \frac{1}{(\rho_{\max} - 1)\exp(-\frac{\epsilon}{k}) + 1}.$$

Using this, we get

$$\begin{aligned}
\log \mu_\epsilon(\tilde{\boldsymbol{z}} = \boldsymbol{v}|\boldsymbol{v} = s_0 \dots s_T) &= \log\left(\prod_{t=0}^{T-1}\prod_{i=1}^{N} \mu_\epsilon(s^i_t | s^i_t, \tilde{s}^i_{t-1} = s^i_{t-1})\right) \\
&= \sum_{t=0}^{T-1}\sum_{i=1}^{N} \log \mu_\epsilon(s^i_t | s^i_t, \tilde{s}^i_{t-1} = s^i_{t-1}) \\
&\geq \sum_{t=0}^{T-1}\sum_{i=1}^{N} \log \frac{1}{(\rho_{\max} - 1)\exp(-\frac{\epsilon}{k}) + 1} \\
&= \sum_{t=0}^{T-1} -N \log\left((\rho_{\max} - 1)\exp(-\frac{\epsilon}{k}) + 1\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\log \mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right] &= \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log\left(\prod_{t=0}^{T-1}\prod_{i=1}^{N} \mu_\epsilon(s_t^i|s_t^i, s_{t-1}^i)\right) \\
&\geq \sum_{\boldsymbol{v}\in W} -\mu^{tr}(\boldsymbol{v}) \sum_{t=0}^{T-1} N \log\left((\rho_{\max}-1)\exp(-\frac{\epsilon}{k})+1\right) \\
&= E\left[\sum_{t=0}^{\tau-1} -N\log\left((\rho_{\max}-1)\exp(-\frac{\epsilon}{k})+1\right)|\mu^{tr}\right] \\
&= -N\log\left((\rho_{\max}-1)\exp(-\frac{\epsilon}{k})+1\right) E\left[\sum_{t=0}^{\tau-1} 1|\mu^{tr}\right] \\
&= -N\log\left((\rho_{\max}-1)\exp(-\frac{\epsilon}{k})+1\right) l^{tr}.
\end{aligned}$$

□

Proof of Theorem 1. For any agent i, let $\boldsymbol{v}=(v^1, v^2, \ldots, v^n) \in D_{\boldsymbol{\pi},T}$ and $\boldsymbol{w}=(w^1, w^2, \ldots, w^n) \in D_{\boldsymbol{\pi},T}$ be two adjacent joint trajectories and by Definition 1 we have $d(v^i, w^i) \leq k$ and $v^j = w^j$ for all $j \neq i$. Let $\tilde{z}^i$ denote a private output of the online mechanism M for agent i. We will show that the probability that the online mechanism generates $\tilde{z}^i$ satisfies Definition 2.

Let $\{p_1, p_2, \ldots, p_m\}$ be the agents that have directed paths to node i in the directed policy dependency graph G. We note that

$$\Pr(M(\boldsymbol{v})=\tilde{z}^i) = \Pr(M(\boldsymbol{v})=\tilde{z}^i|v^i, v^{p_1}, \ldots, v^{p_m}),$$

since the private path $\tilde{z}^i$ of agent i is conditionally independent from v^j for all $j \in [N]\backslash\{p_1, p_2, \ldots, p_m\}$. This independence occurs because 1) the policy dependency graph is acyclic, and $\tilde{z}^i$ is generated from a distribution that does not depend on $\tilde{z}^j$ for all $j \in [N]\backslash\{p_1, p_2, \ldots, p_m\}$, and 2) the online mechanism of agent i depends on its own private and true state trajectories. Next, we note that

$$\begin{aligned}
\Pr(M(\boldsymbol{v})=\tilde{z}^i|v^i, v^{p_1}, \ldots, v^{p_m}) &= \frac{\Pr(M(\boldsymbol{v})=\tilde{z}^i, v^i, v^{p_1}, \ldots, v^{p_m})}{\Pr(v^i, v^{p_1}, \ldots, v^{p_m})} \\
&= \frac{\Pr(v^{p_1}, \ldots, v^{p_m}|M(\boldsymbol{v})=\tilde{z}^i, v^i)\Pr(M(\boldsymbol{v})=\tilde{z}^i, v^i)}{\Pr(v^{p_1}, \ldots, v^{p_m}|v^i)\Pr(v^i)} \\
&= \frac{\Pr(v^{p_1}, \ldots, v^{p_m})\Pr(M(\boldsymbol{v})=\tilde{z}^i, v^i)}{\Pr(v^{p_1}, \ldots, v^{p_m})\Pr(v^i)} \\
&= \Pr(M(\boldsymbol{v})=\tilde{z}^i|v^i) \\
&= \prod_{t=1}^{T} \mu_\epsilon^i(\tilde{s}_t^i \mid s_t^i, \tilde{s}_{t-1}^i),
\end{aligned}$$

since the probability of v^p, $\forall p \in \{p_1, p_2, \ldots, p_m\}$, does not depend on agent i's private and true local state trajectories and the online mechanism is Markovian.

Similarly, for $w^i = y_1^i y_2^i \ldots y_T^i$,

$$\Pr(M(\boldsymbol{w})=\tilde{z}^i|w^i, w^{p_1}, \ldots, w^{p_m}) = \Pr(M(\boldsymbol{w})=\tilde{z}^i|w^i) = \prod_{t=1}^{T} \mu_\epsilon^i(\tilde{s}_t^i \mid y_t^i, \tilde{s}_{t-1}^i)$$

Consequently,

$$\frac{\Pr(M(\boldsymbol{v})=\tilde{z}^i)}{\Pr(M(\boldsymbol{w})=\tilde{z}^i)} = \frac{\Pr(M(\boldsymbol{v})=\tilde{z}^i|v^i)}{\Pr(M(\boldsymbol{w})=\tilde{z}^i|w^i)} = \frac{\prod_{t=1}^{T} \mu_\epsilon^i(\tilde{s}_t^i \mid s_t^i, \tilde{s}_{t-1}^i)}{\prod_{t=1}^{T} \mu_\epsilon^i(\tilde{s}_t^i \mid y_t^i, \tilde{s}_{t-1}^i)}$$

The rest of the proof immediately follows from the proof of Theorem 7 from [Chen et al., 2023].

□

Proof of Theorem 2. Due to the causality property of the online mechanism (Algorithm 1) and the joint policy execution (Algorithm 2), we have

$$\begin{aligned}\mu^{pr}(\boldsymbol{v}) &= \sum_{\tilde{\boldsymbol{z}}\in\boldsymbol{\mathcal{S}}^T} \mu^{pr}(\boldsymbol{v},\tilde{\boldsymbol{z}}) \\ &= \sum_{\tilde{\boldsymbol{z}}\in\boldsymbol{\mathcal{S}}^T} \prod_{t=0}^{T-1} \Pr(\boldsymbol{a}_t\boldsymbol{s}_{t+1},\tilde{\boldsymbol{s}}_t|\boldsymbol{a}_{t-1}\boldsymbol{s}_t\dots\boldsymbol{a}_0\boldsymbol{s}_1,\tilde{\boldsymbol{s}}_{t-1}\dots\tilde{\boldsymbol{s}}_0),\end{aligned}$$

where,

$$\begin{aligned}\Pr(\boldsymbol{a}_t\boldsymbol{s}_{t+1},\tilde{\boldsymbol{s}}_t|\boldsymbol{a}_{t-1}\boldsymbol{s}_t\dots\boldsymbol{a}_0\boldsymbol{s}_1,\tilde{\boldsymbol{s}}_{t-1}\dots\tilde{\boldsymbol{s}}_0) &= \Pr(\boldsymbol{a}_t\boldsymbol{s}_{t+1},\tilde{\boldsymbol{s}}_t|\boldsymbol{s}_t,\tilde{\boldsymbol{s}}_{t-1}) && (4)\\ &= \prod_{i=1}^{N}\Pr(a_t^i s_{t+1}^i,\tilde{\boldsymbol{s}}_t|\boldsymbol{s}_t,\tilde{\boldsymbol{s}}_{t-1}) && (5)\\ &= \prod_{i=1}^{N}\Pr(a_t^i s_{t+1}^i|\boldsymbol{s}_t,\tilde{\boldsymbol{s}}_t,\tilde{\boldsymbol{s}}_{t-1})\Pr(\tilde{\boldsymbol{s}}_t|\boldsymbol{s}_t,\tilde{\boldsymbol{s}}_{t-1}) \\ &= \prod_{i=1}^{N}\Pr(a_t^i s_{t+1}^i|\boldsymbol{s}_t,\tilde{\boldsymbol{s}}_t,\tilde{\boldsymbol{s}}_{t-1})\left(\prod_{k=1}^{N}\mu_\epsilon(\tilde{s}_t^k|s_t^k,\tilde{s}_{t-1}^k)\right) && (6)\\ &= \prod_{i=1}^{N}\Pr(a_t^i s_{t+1}^i|\hat{\boldsymbol{s}}_t^i)\left(\prod_{k=1}^{N}\mu_\epsilon(\tilde{s}_t^k|s_t^k,\tilde{s}_{t-1}^k)\right) && (7)\\ &= \prod_{i=1}^{N}\mathcal{T}(s_t^i,a_t^i,s_{t+1}^i)\pi^i(\hat{\boldsymbol{s}}_t^i,a_t^i)\left(\prod_{k=1}^{N}\mu_\epsilon(\tilde{s}_t^k|s_t^k,\tilde{s}_{t-1}^k)\right).\end{aligned}$$

Equation (4) is because of the Markovian property. Equation (5) is because the each agent are choosing its next action and state independently. Equation (6) is due to each state is generating its private state independently. Equation (7) is because for each agent i, its true next state $\tilde{s}_{t+1}^i$ is independent of other states' true states and the private state itself.

Therefore,

$$\begin{aligned}\mu^{pr}(\boldsymbol{v}) &= \sum_{\tilde{\boldsymbol{z}}\in\boldsymbol{\mathcal{S}}^T}\prod_{t=0}^{T-1}\prod_{i=1}^{N}\mathcal{T}(s_t^i,a_t^i,s_{t+1}^i)\pi^i(\hat{\mathbf{s}}_t^i,a_t^i)\left(\prod_{i=1}^{N}\mu_\epsilon(\tilde{s}_t^i|s_t^i,\tilde{s}_{t-1}^i)\right)\\ &\geq \prod_{t=0}^{T-1}\prod_{i=1}^{N}\mathcal{T}(s_t^i,a_t^i,s_{t+1}^i)\pi^i(\hat{\mathbf{s}}_t^i,a_t^i)\left(\prod_{i=1}^{N}\mu_\epsilon(\tilde{s}_t^i|s_t^i,\tilde{s}_{t-1}^i)\right),\forall\tilde{\boldsymbol{z}}\in\boldsymbol{\mathcal{S}}^T, \qquad (8)\end{aligned}$$

where Equation (8) is because the probability of all possible private state trajectories has to be greater than any single private state trajectory. We only consider the case when $\tilde{\boldsymbol{s}}_t = \boldsymbol{s}_t$, which means the private online mechanism will make the correct decision at every time t. Therefore,

$$\begin{aligned}\mu^{pr}(\boldsymbol{v}) &\geq \prod_{t=0}^{T-1}\prod_{i=1}^{N}\mathcal{T}(s_t^i,a_t^i,s_{t+1}^i)\pi^i(\boldsymbol{s}_t,a_t^i)\left(\prod_{i=1}^{N}\mu_\epsilon(s_t^i|s_t^i,s_{t-1}^i)\right)\\ &= \mu^{tr}(\boldsymbol{v})\left(\prod_{i=1}^{N}\mu_\epsilon(s_t^i|s_t^i,s_{t-1}^i)\right)\end{aligned}$$

Now we look at the following KL divergence:

$$
\begin{aligned}
KL(\mathbf{\Gamma}^{tr}||\mathbf{\Gamma}^{pr}) &= \sum_{\boldsymbol{v}\in W\cup W'} \mu^{tr}(\boldsymbol{v}) \log\left(\frac{\mu^{tr}(\boldsymbol{v})}{\mu^{pr}(\boldsymbol{v})}\right) \\
&= \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log\left(\frac{\mu^{tr}(\boldsymbol{v})}{\mu^{pr}(\boldsymbol{v})}\right) \qquad (9)\\
&\leq \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log\left(\frac{\mu^{tr}(\boldsymbol{v})}{\mu^{tr}(\boldsymbol{v})\left(\prod_{t=0}^{\infty}\prod_{i=1}^{N}\mu_\epsilon(s_t^i|s_t^i,s_{t-1}^i)\right)}\right) \\
&= \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log(\mu^{tr}(\boldsymbol{v})) - \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log(\mu^{tr}(\boldsymbol{v})) \\
&\quad - \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log\left(\prod_{t=0}^{\infty}\prod_{i=1}^{N}\mu_\epsilon(s_t^i|s_t^i,s_{t-1}^i)\right) \\
&= H(\boldsymbol{S}_0\boldsymbol{A}_0\dots\boldsymbol{S}_\eta) - H(\boldsymbol{S}_0\boldsymbol{A}_0\dots\boldsymbol{S}_\eta) - \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log\left(\prod_{t=0}^{T-1}\prod_{i=1}^{N}\mu_\epsilon(s_t^i|s_t^i,s_{t-1}^i)\right) \\
&\leq \sum_{i=1}^{N} H(S_0^iA_0^i\dots S_\eta^i) - H(\boldsymbol{S}_0\boldsymbol{A}_0\dots\boldsymbol{S}_\eta) - \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log\left(\prod_{t=0}^{T-1}\prod_{i=1}^{N}\mu_\epsilon(s_t^i|s_t^i,s_{t-1}^i)\right) \qquad (10)\\
&= C_{\boldsymbol{\pi}} - \sum_{\boldsymbol{v}\in W} \mu^{tr}(\boldsymbol{v}) \log\left(\prod_{t=0}^{T-1}\prod_{i=1}^{N}\mu_\epsilon(s_t^i|s_t^i,s_{t-1}^i)\right) \qquad (11)\\
&= C_{\boldsymbol{\pi}} - \mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right]
\end{aligned}
$$

where (9) is due to $\sum_{\boldsymbol{v}\in W'}\mu^{tr}(\boldsymbol{v})=0$, (10) is due to the subadditivity of entropy, and (11) is due to the definition of $C_{\boldsymbol{\pi}}$. Using Lemma 1 in (11) gives

$$
KL(\mathbf{\Gamma}^{tr}||\mathbf{\Gamma}^{pr}) \leq C_{\boldsymbol{\pi}} + N\log\left((\rho_{\max}-1)\exp(-\frac{\epsilon}{\ell})+1\right)l^{tr}. \qquad (12)
$$

Finally, we show that $v^{pr} \geq v^{tr} - 1 + \exp(-C_{\boldsymbol{\pi}})\left((\rho_{max}-1)\exp\left(-\frac{\epsilon}{\ell}\right)+1\right)^{Nl^{tr}}/2$. Let $R'\subseteq W\cup W'$ be an arbitrary set.

$$
\begin{aligned}
\boldsymbol{v}^{tr} - \boldsymbol{v}^{pr} &= \sum_{\boldsymbol{v}\in R} \mu^{tr}(\boldsymbol{v}) - \mu^{pr}(\boldsymbol{v}) \\
&\leq \left|\sum_{\boldsymbol{v}\in R} \mu^{tr}(\boldsymbol{v}) - \mu^{pr}(\boldsymbol{v})\right| \\
&\leq \sup_{R'}\left|\sum_{\boldsymbol{v}\in R'} \mu^{tr}(\boldsymbol{v}) - \mu^{pr}(\boldsymbol{v})\right| \\
&\leq \sqrt{1-\exp(-KL(\mathbf{\Gamma}^{tr}||\mathbf{\Gamma}^{pr}))} \qquad (13a)
\end{aligned}
$$

where (13a) is due to Bretagnolle-Huber inequality Bretagnolle and Huber [1979]. Rearranging the terms of (13a) and using (12) yields to the desired result.

□

We note that apart from Theorem 2, we can derive a tighter lower bound on v^{pr}.

Theorem 3. *Given $\epsilon > 0$, for N agents, we have*

$$
v^{pr} \geq v^{tr} - 1 + \left((\rho_{\max}-1)\exp(-\frac{\epsilon}{k})+1\right)^{Nl^{tr}}. \qquad (14)
$$

Proof of Theorem 3. As shown in the proof of Theorem 2, we have

$$\begin{aligned} v^{pr} &= \sum_{\boldsymbol{v}=\boldsymbol{s}_0\boldsymbol{a}_0\boldsymbol{s}_1\boldsymbol{a}_1\ldots\boldsymbol{s}_T\in W} \mu^{pr}(\boldsymbol{v})\mathbb{1}(\boldsymbol{v}\in R) \\ &\geq \sum_{\boldsymbol{v}=\boldsymbol{s}_0\boldsymbol{a}_0\boldsymbol{s}_1\boldsymbol{a}_1\ldots\boldsymbol{s}_T\in W} \mu^{tr}(\boldsymbol{v})\mathbb{1}(\boldsymbol{v}\in R)\left(\prod_{t=0}^{T-1}\prod_{k=1}^{N}\mu_\epsilon(s_t^k|s_t^k,s_{t-1}^k)\right) \\ &= \Pr(\boldsymbol{v}\in R\wedge\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v}\sim\mu^{tr},\tilde{\boldsymbol{z}}\sim\mu^{\epsilon}(\cdot|\boldsymbol{v})). \end{aligned}$$

By the union bound, we have

$$\begin{aligned} v^{pr} &\geq \Pr(\boldsymbol{v}\in R|\boldsymbol{v}\sim\mu^{tr}) + \mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right] - 1 \\ &= v^{tr} + \mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right] - 1 \end{aligned}$$

Then with

$$\mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right] = \sum_{\boldsymbol{v}\in W}\mu^{tr}(\boldsymbol{v})\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})$$

and Jensen's inequality, we have

$$\begin{aligned} \mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right] &= \exp\left(\log\sum_{\boldsymbol{v}\in W}\mu^{tr}(\boldsymbol{v})\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right) \\ &\geq \exp\left(\sum_{\boldsymbol{v}\in W}\mu^{tr}(\boldsymbol{v})\log\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right) \\ &= \exp\left(\mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\log\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right]\right). \end{aligned}$$

Using Lemma 1, we get

$$\begin{aligned} v^{pr} &\geq v^{tr} - 1 + \exp\left(\mathbb{E}_{\boldsymbol{v}\sim\mu^{tr}}\left[\log\mu_\epsilon(\tilde{\boldsymbol{z}}=\boldsymbol{v}|\boldsymbol{v})\right]\right) \\ &\geq v^{tr} - 1 + \left((\rho_{\max}-1)\exp(-\frac{\epsilon}{k})+1\right)^{Nl^{tr}}, \end{aligned}$$

which completes the proof.

□

Compared to (2), (14) does not take the total correlation $C_{\boldsymbol{\pi}}$ into account and only focuses on the success probability when the private state trajectories are the same with the original state trajectories. As a result, a joint policy $\boldsymbol{\pi} = \{\pi^i\}_{i=1}^N$ synthesized by minimizing the lower bound in (14) does not enjoy the robustness brought by minimizing (2). The inclusion of total correlation in the objective function increases the team performance under private communications since the agents' policies are less sensitive to each other's state trajectories.

References

Jean Bretagnolle and Catherine Huber. Estimation des densités: risque minimax. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 47(2):119–137, 1979.

Bo Chen, Kevin Leahy, Austin Jones, and Matthew Hale. Differential privacy for symbolic systems with application to markov chains. *Automatica*, 152:110908, 2023.

Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

Mustafa O. Karabag, Cyrus Neary, and Ufuk Topcu. Planning not to talk: Multiagent systems that are robust to communication loss. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, 2022.