A Privacy-Preserving Diffusion Strategy Over Multitask Networks

Chengcheng Wang, *Member, IEEE*, Wee Peng Tay, *Senior Member, IEEE*, Yuan Wang, *Student Member, IEEE*, and Ye Wei

Abstract

We develop a privacy-preserving distributed strategy over multitask diffusion networks, where each agent is interested in not only improving its local inference performance via in-network cooperation, but also protecting its own individual task against privacy leakage. In the proposed strategy, at each time instant, each agent sends a noisy estimate, which is its local intermediate estimate corrupted by a zero-mean additive noise, to its neighboring agents. We derive a sufficient condition to determine the amount of noise to add to each agent's intermediate estimate to achieve an optimal trade-off between the steady-state network mean-square-deviation and an inference privacy constraint. We show that the proposed noise powers are bounded and convergent, which leads to mean-square convergence of the proposed privacy-preserving multitask diffusion scheme. Simulation results demonstrate that the proposed strategy is able to balance the trade-off between estimation accuracy and privacy preservation.

Index Terms

Distributed strategies, diffusion strategies, multitask networks, privacy preservation, additive noises.

I. Introduction

In multitask diffusion networks, a set of interconnected agents work collaboratively to estimate different but related parameters of interest [1]–[9]. In order to make use of the relationship between different tasks for better inference performance, local estimates are exchanged amongst agents within the same neighborhood. However, each agent may wish to protect its own local

C. Wang, W. P. Tay and Y. Wang are with Nanyang Technological University, Singapore. Y. Wei is with Harbin Engineering University, China. E-mails: wangcc@ntu.edu.sg; wptay@ntu.edu.sg; wangyuan@ntu.edu.sg; weiye@hrbeu.edu.cn.

parameters of interest and prevent other agents in the network from accurately inferring these parameters. Sharing its local estimate may raise privacy concerns. For example, in an Internet of Things (IoT) network, sensors are deployed in smart grids, traffic monitoring, health monitoring, home monitoring and other applications [10]–[15]. Although different IoT or edge computing devices may have their local objectives, they can exchange information with each other or service providers [16]–[19] to improve inferences and services. This may leak to unnecessary privacy leakage.

To protect the privacy of the data being exchanged between agents in a distributed network, the works [20]–[25] propose local differential privacy mechanisms, while [26]–[29] develop privacy-preserving distributed data analytics. However, these approaches may lead to a significant trade-off in estimation accuracy as they do not specifically protect the privacy of the parameters of interest. To achieve inference privacy in a decentralized IoT network, [30]–[34] proposes nonparametric approaches with information privacy guarantees, while [35]–[38] propose to map the agents' raw observations into a lower dimensional subspace. These inference privacy works assume that all agents in the network are interested in inferring the same parameters or hypothesis of interest.

Existing works on multitask diffusion networks mainly focus on developing new distributed strategies for automatic clustering [1]–[3] and handling different relationships among the tasks [4]–[9]. Other works like [3], [39]–[41] focus on evaluating the performance of different distributed schemes. Few works have considered protecting the privacy of each agent's local parameters. The reference [42] considers data privacy of the agents' local measurements in a single-task network.

The objective of this work is to develop a privacy-preserving diffusion strategy over multitask networks, which balances the trade-off between estimation accuracy and privacy preservation of agents' local parameters or tasks. Specifically, we consider multitask estimation problems where the unknown parameters of interest within each neighborhood are linearly related with each other [8]. Such problems widely exist in applications such as electrical networks, telecommunication networks, and pipeline networks [8]. Different from the strategy in [8], which does not take privacy preservation into consideration, we propose to sanitize each agent's intermediate estimate before sharing it with its neighbors by adding an appropriate zero-mean noise to the intermediate estimate. We study how to design the power of the noise added to optimize the trade-off between the network mean-square-deviation (MSD) and the inference privacy of each agent's local parameters, measured by its neighbors' mean-square error in estimating the agent's local parameters. In addition, different from existing works on diffusion strategies in the presence of link noises [40], [41], [43], [44], which examine the effect of link noises on the performance of diffusion strategies, this work mainly focuses on developing agent-specific time-varying variances of the additive noises that enable agents to benefit from in-network cooperation as well as protect their own individual tasks against privacy leakage.

The rest of this paper is organized as follows. In Section II, we formulate the multitask estimate problem considered in this paper. A privacy-preserving multitask diffusion scheme is then proposed to solve the problem in Section III, where a zero-mean additive noise is added to the intermediate estimate that is communicated to the neighboring agents. We study the choice guideline for powers of the additive noises, and examine the boundedness and convergence of the proposed powers in Section IV. We present the simulation results in Section V. Section VI concludes the paper.

Notations: We use lowercase letters to denote vectors and scalars, uppercase letters for matrices, plain letters for deterministic variables, and boldface letters for random variables. We also use $(\cdot)^{\mathsf{T}}$ to denote transposition, $(\cdot)^{-1}$ for matrix inversion, $\mathrm{Tr}\,(\cdot)$ for the trace of a matrix, $\mathrm{diag}\{\cdot\}$ for a diagonal matrix, $\mathrm{col}\,\{\cdot\}$ for a column vector, $\mathrm{row}\,\{\cdot\}$ for a row vector, $\|\cdot\|$ for the two-induced norm of a matrix or the Euclidean norm of a vector, $\|x\|_{\Sigma}^{2}$ for the weighted square value $x^{\mathsf{T}}\Sigma x$, $\otimes$ for Kronecker product.

II. Linearly Related Multitask Network

In this section, we present our system model, and give a brief introduction to multitask networks, where neighboring agents' tasks are linearly related. Consider a strongly-connected network of N agents, where information can flow in either direction between any two connected agents [40]. At each time instant i, each agent k has access to a scalar observation $d_k(i)$, and an $M_k \times 1$ regression vector $u_k(i)$. The random data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_k(i)\}$ are related via the linear regression model

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_k^{\mathsf{T}}(i)\boldsymbol{w}_k^o + \boldsymbol{v}_k(i) \tag{1}$$

where the scalar $\boldsymbol{v}_k(i)$ is measurement noise, $\boldsymbol{w}_k^o$ is an $M_k \times 1$ unknown random vector, with mean $\mathbb{E}\boldsymbol{w}_k^o$ and covariance matrix

$$W_{kk} = \mathbb{E}\left[(\boldsymbol{w}_k^o - \mathbb{E}\boldsymbol{w}_k^o)(\boldsymbol{w}_k^o - \mathbb{E}\boldsymbol{w}_k^o)^{\mathsf{T}}\right]. \tag{2}$$

Note that although we assume that the parameter vector $\boldsymbol{w}_k^o$ is random instead of being a deterministic parameter vector, like most of the literature on diffusion strategies [1]–[8], [39]–[44], we assume that the parameter vector $\boldsymbol{w}_k^o$ is fixed at a certain realization w_k^o during the diffusion estimation process. Since our goal is to develop inference privacy mechanisms that lead to high estimation errors, on average, of agent k's local parameters $\boldsymbol{w}_k^o$ by other agents $\{\ell \neq k\}$, we adopt a Bayesian framework for the privacy criterion.

We make the following assumptions regarding model (1).

Assumption 1. *(Measurement noise) The measurement noise $\boldsymbol{v}_k(i)$ is white over time, with zero mean, and a variance of $\sigma_{v,k}^2$.*

Assumption 2. *(Regression data) The regression data $\{\boldsymbol{u}_k(i)\}$ are zero-mean, white over time and space with*

$$\mathbb{E}\boldsymbol{u}_k(i)\boldsymbol{u}_\ell^\intercal(j) = R_{u,k}\delta_{k,\ell}\delta_{i,j} \tag{3}$$

where $R_{u,k}$ is symmetric positive definite, and $\delta_{k,\ell}$ denotes the Kronecker delta sequence.

Assumption 3. *(Independence assumption) The random data $\{\boldsymbol{w}_k^o, \boldsymbol{u}_\ell(i), \boldsymbol{v}_m(j)\}$ are independent of each other for any agent k, ℓ, m and any time instant i, j.*

The objective of each agent k is to find the minimizer of the following mean-square-error cost function:

$$J_k(w_k) = \mathbb{E}\left[(\boldsymbol{d}_k(i) - \boldsymbol{u}_k^\intercal(i)w_k)^2 \,\middle|\, \boldsymbol{w}_k^o = w_k^o \right]. \tag{4}$$

Let $\mathcal{N}_k$ be the set of all neighboring agents of agent k, including agent k itself. Assume that neighboring tasks $\{\boldsymbol{w}_k^o, \boldsymbol{w}_\ell^o\}$ for any $\ell \in \mathcal{N}_k$ are involved in at least one linear equality of the form [8]:

$$\sum_{k \in \mathcal{I}_q} D_{qk}\boldsymbol{w}_k^o + b_q = 0, \tag{5}$$

where the subscript "q" is the index of the linear equality, the set $\mathcal{I}_q$ includes all agents involved in the q-th equality, the coefficient matrix D_{qk} and constant vector b_q are of size $L_q \times M_k$ and

$L_q \times 1$, respectively. Then, the objective for the entire network is to find the optimal solution to the following constrained optimization problem [8]:

$$\begin{aligned} \min_{w_1,\ldots,w_N} \; & J(w_1,\ldots,w_N) = \sum_{k=1}^{N} J_k(w_k) \\ \text{s.t.} \; & \sum_{k\in\mathcal{I}_q} D_{qk} w_k + b_q = 0, \text{ for } q = 1,\ldots,Q \end{aligned} \tag{6}$$

where individual costs $\{J_k(w_k)\}$ are defined by (4). Let

$$w = \text{col}\{w_1,\ldots,w_N\}.$$

We proceed to rewrite the constraints in (19) more compactly

$$\mathcal{D}w + b = 0,$$

where matrix $\mathcal{D}$ is a $Q \times N$ block matrix with blocks $\{D_{qk}\}$ for any $q = 1,\ldots,Q$ and $k = 1,\ldots,N$, and vector b is a $Q \times 1$ block vector with blocks $\{b_q\}$ for any $q = 1,\ldots,Q$. Let j_k be the total number of linear equalities that agent k is involved in. We make the same assumption as [8] below.

Assumption 4. *(Linear equality) Each agent k is involved in at least one linear equality constraint,* i.e., $j_k \geq 1$. *Each agent k has access to all j_k linear equalities that it is involved in. In addition, all agents involved in these j_k linear equalities are inside $\mathcal{N}_k$,* i.e., $\mathcal{I}_q \subset \mathcal{N}_k$ *for any $k \in \mathcal{I}_q$. Matrix $\mathcal{D}$ is full row-rank.*

As demonstrated in [8], each agent k can benefit through cooperation with neighboring agents by sharing their local parameter estimates with their neighbors, which enables it to leverage the linear relationships (5) and its neighbors' parameter estimates to improve its own inference accuracy. In this paper, we consider the scenario where agent k also wants to prevent other agents from inferring its own task $\boldsymbol{w}_k^o$. Thus, a privacy-preserving distributed solution is required to balance the trade-off between estimation accuracy and privacy protection of the individual tasks.

III. Privacy-preserving Diffusion Strategy

In this section, we propose a simple inference privacy mechanism to protect each agent's local task by adding noise to its intermediate estimate before sharing with its neighbors. We then propose a utility-privacy optimization trade-off to determine the amount of noise to add.

We start off with some definitions, which are required to describe our privacy-preserving diffusion strategy.

Let $i_q = |\mathcal{I}_q|$ be the number of agents that are involved in the q-th constraint. Let

$$\mathcal{D}_q = \text{row}\{D_{q\ell}\}_{\ell \in \mathcal{I}_q} \tag{7}$$

be a $1 \times i_q$ block matrix, which collects all the coefficient matrices that are defined by the q-th constraint. Let $M_q = \sum_{\ell \in \mathcal{I}_q} M_\ell$. Define

$$\mathcal{P}_q = I_{M_q} - \mathcal{D}_q^{\mathsf{T}}(\mathcal{D}_q\mathcal{D}_q^{\mathsf{T}})^{-1}\mathcal{D}_q \tag{8}$$

$$f_q = \mathcal{D}_q^{\mathsf{T}}(\mathcal{D}_q\mathcal{D}_q^{\mathsf{T}})^{-1}b_q \tag{9}$$

where I_M denotes an $M \times M$ identity matrix. Now, we rewrite the $i_q \times i_q$ block matrix $\mathcal{P}_q$

$$\mathcal{P}_q = \{[\mathcal{P}_q]_{k,\ell}\}_{k,\ell \in \mathcal{I}_q}$$

where the $M_k \times M_\ell$ (k, ℓ)-th block of $\mathcal{P}_q$, $[\mathcal{P}_q]_{k,\ell}$, is defined as

$$[\mathcal{P}_q]_{k,\ell} = \begin{cases} I_{M_k} - D_{qk}^{\mathsf{T}}(\mathcal{D}_q\mathcal{D}_q^{\mathsf{T}})^{-1}D_{qk}, & \text{if } k = \ell, \text{ and } k \in \mathcal{I}_q, \\ -D_{qk}^{\mathsf{T}}(\mathcal{D}_q\mathcal{D}_q^{\mathsf{T}})^{-1}D_{q\ell}, & \text{if } k \neq \ell, \text{ and } \{k, \ell\} \subset \mathcal{I}_q. \end{cases} \tag{10}$$

Likewise, we rewrite the $i_q \times 1$ block vector

$$f_q = \text{col}\{[f_q]_k\}_{k \in \mathcal{I}_q}$$

with the $M_k \times 1$ k-th block entry

$$[f_q]_k = D_{qk}^{\mathsf{T}}(\mathcal{D}_q\mathcal{D}_q^{\mathsf{T}})^{-1}b_q.$$

Then, each agent k in the network is expanded into a cluster of j_k virtual sub-agents, $\{k_m\}_{m=1}^{j_k}$, so that each sub-agent k_m is only involved in one constraint [8]. Let $\mathcal{I}_{e,q}$ be the set of sub-agent indices involved in the q-th constraint, for any $q = 1, \ldots, Q$. Then, if $\ell \in \mathcal{I}_q$ holds for some agent ℓ and constraint index q, it follows that there is a unique sub-agent ℓ_n, where $n \in \{1, \ldots, j_\ell\}$, such that $\ell_n \in \mathcal{I}_{e,q}$. Now, if a sub-agent $k_m \in \mathcal{I}_{e,q}$, for any $m = 1, \ldots, j_k$, we proceed to introduce the notations $\mathcal{P}_{k_m}$ and f_{k_m} as the k-th block row of $\mathcal{P}_q$ and f_q, respectively, namely,

$$\mathcal{P}_{k_m} = [\mathcal{P}_q]_k$$

$$f_{k_m} = [f_q]_k$$

where the notation $[\mathcal{P}_q]_k$ denotes the k-th block row in $\mathcal{P}_q$.

In our privacy-preserving diffusion strategy, we initialize $\boldsymbol{w}_k(-1) = 0$ for every agent k in the network. Given data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_k(i)\}$ for each time instant $i \geq 0$, and for each agent $k = 1, \ldots, N$, we perform the following steps iteratively:

1) Adaptation. Each agent k updates the current estimate $\boldsymbol{w}_k(i-1)$ with respect to (*w.r.t.*) $\boldsymbol{w}_k^o = w_k^o$ to an intermediate estimate $\boldsymbol{\psi}_k(i)$ by following the stochastic gradient descent (SGD) algorithm

$$\boldsymbol{\psi}_k(i) = \boldsymbol{w}_k(i-1) + \frac{\mu_k}{j_k}\boldsymbol{u}_k(i)\left(\boldsymbol{d}_k(i) - \boldsymbol{u}_k^{\mathsf{T}}(i)\boldsymbol{w}_k(i-1)\right) \tag{11}$$

where $\mu_k > 0$ is the step-size parameter at agent k.

2) Exchange. Each agent k collects estimates $\{\boldsymbol{\psi}'_\ell(i)\}$ from neighboring agents $\{\ell \in \mathcal{N}_k\}$

$$\boldsymbol{\psi}'_\ell(i) = \begin{cases} \boldsymbol{\psi}_\ell(i) + \boldsymbol{n}_\ell(i), & \text{if } \ell \in \mathcal{N}_k, \text{ and } \ell \neq k, \\ \boldsymbol{\psi}_k(i), & \text{if } \ell = k \end{cases} \tag{12}$$

where the random additive noise vector $\boldsymbol{n}_\ell(i)$ is of size $M_\ell \times 1$.

3) Projection. For each of the j_k linear equality constraints that agent k is involved in, do

$$\boldsymbol{\phi}_{k_m}(i) = \mathcal{P}_{k_m} \cdot \mathrm{col}\left\{\boldsymbol{\psi}'_\ell(i)\right\}_{\ell \in \mathcal{I}_q} - f_{k_m},\ k_m \in \mathcal{I}_{e,q}, m = 1, \ldots, j_k \tag{13}$$

which generates a total of j_k intermediate estimates $\{\boldsymbol{\phi}_{k_m}(i)\}_{m=1}^{j_k}$.

4) Combination. Each agent k takes the average over j_k intermediate estimates $\{\boldsymbol{\phi}_{k_m}(i)\}$, and obtains a new estimate, $\boldsymbol{w}_k(i)$, of the unknown parameter vector $\boldsymbol{w}_k^o = w_k^o$

$$\boldsymbol{w}_k(i) = \frac{1}{j_k}\sum_{m=1}^{j_k} \boldsymbol{\phi}_{k_m}(i). \tag{14}$$

Remark 1: The difference between the proposed privacy-preserving diffusion strategy (11) to (14) and the existing scheme in [8] is in the exchange step. Specifically, in order to protect each individual task $\boldsymbol{w}_k^o$ against privacy leakage, each agent k sends a noisy intermediate estimate $\boldsymbol{\psi}'_k(i)$, instead of the true estimate $\boldsymbol{\psi}_k(i)$ as in [8], to its neighboring agents. We call $\boldsymbol{n}_k(i)$ a *privacy mechanism noise*.

To allow a distributed implementation of the privacy mechanism, we make the following assumption.

Assumption 5. *(Privacy mechanism noise) The entries of $\boldsymbol{n}_k(i)$ at time i, for any $k = 1, \ldots, N$, are independent and identically distributed (i.i.d.), with zero mean and a time-varying variance of $\sigma_{n,k}^2(i)$. The random noises $\{\boldsymbol{n}_k(i)\}$ are white over time and space. The random process $\{\boldsymbol{n}_k(i)\}$ is independent of any other random processes.*

From Assumption 5, each agent k generates the noise $\boldsymbol{n}_k(i)$ independently of other agents in the network, and also independently over time instants i. We also have

$$R_{n,k}(i) \triangleq \mathbb{E}\left[\boldsymbol{n}_k(i)\boldsymbol{n}_k^{\mathsf{T}}(i)\right] = \sigma_{n,k}^2(i) I_{M_k}, \tag{15}$$

which is a time-varying matrix.

For the utility achieved by the network of agents, we consider the steady-state network MSD [40, p.583]

$$\text{MSD}_{\text{net}} = \lim_{i\to\infty} \frac{1}{N}\sum_{k=1}^{N} \mathbb{E}\|\boldsymbol{w}_k^o - \boldsymbol{w}_k(i)\|^2, \tag{16}$$

where a smaller MSD_{net} gives a better utility. Let

$$U_{kk}(i) = \mathbb{E}\left[(\boldsymbol{w}_k^o - \mathbb{E}\boldsymbol{w}_k^o)(\boldsymbol{\psi}_k'(i) - \mathbb{E}\boldsymbol{\psi}_k'(i))^{\mathsf{T}}\right] \tag{17}$$

$$R_{\psi',k}(i) = \mathbb{E}\left[(\boldsymbol{\psi}_k'(i) - \mathbb{E}\boldsymbol{\psi}_k'(i))\,(\boldsymbol{\psi}_k'(i) - \mathbb{E}\boldsymbol{\psi}_k'(i))^{\mathsf{T}}\right] \tag{18}$$

for any agent $k = 1, \ldots, N$. Let

$$\widehat{\boldsymbol{w}}_{k|\psi_k'}(i) = U_{kk}(i) R_{\psi',k}^{-1}(i)\,(\boldsymbol{\psi}_k'(i) - \mathbb{E}\boldsymbol{\psi}_k'(i)) + \mathbb{E}\boldsymbol{w}_k^o$$

be the linear least-mean-square estimator (l.l.m.s.e.) [45, p.66] of $\boldsymbol{w}_k^o$ at time instant i, given $\boldsymbol{\psi}_k'(i)$. Our goal is to determine the variances of the privacy mechanism noises $\{\sigma_{n,k}^2(i)\}$ to

$$\begin{aligned} &\min\ \text{MSD}_{\text{net}} \\ &\text{s.t.}\ \mathbb{E}\left\|\boldsymbol{w}_k^o - \widehat{\boldsymbol{w}}_{k|\psi_k'}(i)\right\|^2 \geq \delta_k, \text{ for } k = 1, \ldots, N, i \geq 0 \end{aligned} \tag{19}$$

for non-negative thresholds $\{\delta_k \geq 0\}$, which are chosen according to privacy requirements.

Remark 2: In (19), it is required that at each time instant $i \geq 0$, the *expected* squared distance $\left\|\boldsymbol{w}_k^o - \widehat{\boldsymbol{w}}_{k|\psi_k'}(i)\right\|^2$ over all realizations of $\boldsymbol{w}_k^o$ is no smaller than the predefined parameter δ_k. This provides an inference privacy constraint on the ability of a neighboring agent to agent k in accurately estimating $\boldsymbol{w}_k^o$ on average.

Remark 3: In (19), $\widehat{\boldsymbol{w}}_{k|\psi_k'}(i)$ is the l.l.m.s.e. of $\boldsymbol{w}_k^o$ based on the noisy estimate $\boldsymbol{\psi}_k'(i)$. For each neighboring agent $\ell \in \mathcal{N}_k$, it has access to not only the received noisy estimate $\boldsymbol{\psi}_k'(i)$ from agent k, but also its own intermediate estimate $\boldsymbol{\psi}_\ell(i)$ of $\boldsymbol{w}_\ell^o$. Both of these estimates can be used to infer $\boldsymbol{w}_k^o$ as the unknown parameters $\boldsymbol{w}_k^o$ and $\boldsymbol{w}_\ell^o$ are linearly related with each other through (5). In this paper, to simplify the analysis and to ensure that the sequence of noise variances $\{\sigma_{n,k}^2(i)\}$ at each agent k is bounded and convergent, we only consider the simplified case in (19). We illustrate the case where a neighboring agent uses its own estimates as additional information using simulations in Section V.

IV. Privacy Mechanism Noise Design and Convergence Analysis

In this section, we present an approximate solution to (19) by deriving a sufficient condition for the privacy constraint in (19). We first show that the steady-state network MSD, MSD_{net}, is a monotonically increasing function *w.r.t.* the steady-state variances of the privacy mechanism noises $\{\sigma_{n,k}^2(i)\}$ as $i \to \infty$. Then we derive a sufficient condition for $\sigma_{n,k}^2(i)$ to satisfy the privacy constraint in (19) for each agent k and each time instant $i \geq 0$. Finally, we propose to set the variance $\sigma_{n,k}^2(i)$ to the smallest value in the sufficient condition and show that the proposed variances $\{\sigma_{n,k}^2(i)\}$ converge as $i \to \infty$ for each agent k.

A. Mean-square Performance Analysis

Let

$$\begin{aligned} \boldsymbol{w}_e^o &= \text{col}\left\{\mathbb{1}_{j_k} \otimes \boldsymbol{w}_k^o\right\}_{k=1}^N \\ \boldsymbol{w}_e(i) &= \text{col}\left\{\mathbb{1}_{j_k} \otimes \boldsymbol{w}_k(i)\right\}_{k=1}^N \end{aligned} \tag{20}$$

where $\mathbb{1}_M$ denotes an $M \times 1$ vector with all its entries equal to one, and the subscript 'e' indicates the extended version of the corresponding quantity after the virtual sub-agents are introduced into the network. Then, we introduce

$$\widetilde{\boldsymbol{w}}_e(i) = \boldsymbol{w}_e^o - \boldsymbol{w}_e(i).$$

We also introduce these quantities

$$\begin{aligned} \mathcal{A}_k &= \frac{1}{j_k}\mathbb{1}_{j_k}\mathbb{1}_{j_k}^{\mathsf{T}} \otimes I_{M_k} \\ \mathcal{A} &= \text{diag}\left\{\mathcal{A}_k\right\}_{k=1}^N \end{aligned} \tag{21}$$

$$\boldsymbol{\mathcal{R}}_{u,e}(i) = \text{diag}\left\{I_{j_k} \otimes \frac{\mu_k}{j_k}\boldsymbol{u}_k(i)\boldsymbol{u}_k^{\mathsf{T}}(i)\right\}_{k=1}^N \tag{22}$$

$$\boldsymbol{g}(i) = \text{col}\left\{\mathbb{1}_{j_k} \otimes \frac{\mu_k}{j_k}\boldsymbol{u}_k(i)\boldsymbol{v}_k(i)\right\}_{k=1}^N. \tag{23}$$

Let

$$\boldsymbol{n}_e(i) = \text{col}\left\{\mathbb{1}_{j_k} \otimes \boldsymbol{n}_k(i)\right\}_{k=1}^N.$$

It then follows from Assumption 5 that

$$R_{n,e}(i) \triangleq \mathbb{E}\left[\boldsymbol{n}_e(i)\boldsymbol{n}_e^{\mathsf{T}}(i)\right] = \text{diag}\left\{(\mathbb{1}_{j_k}\mathbb{1}_{j_k}^{\mathsf{T}}) \otimes R_{n,k}(i)\right\}_{k=1}^N, \tag{24}$$

where $R_{n,k}(i)$ is defined by (15). Let

$$N_e = \sum_{k=1}^{N} j_k$$

be the total number of all sub-agents across the network. Let $\mathcal{P}_e$ be an $N_e \times N_e$ block matrix, whose $M_k \times M_\ell$ (k_m, ℓ_n)-th block equals

$$[\mathcal{P}_e]_{k_m,\ell_n} = \begin{cases} [\mathcal{P}_q]_{k,\ell}, & \text{if } \{k_m, \ell_n\} \subset \mathcal{I}_{e,q} \\ 0_{M_k \times M_\ell}, & \text{otherwise} \end{cases} \tag{25}$$

with the block entry $[\mathcal{P}_q]_{k,\ell}$ defined by (10). Let $[\mathcal{P}_e]_{k_m^-}$ be the k_m-th block row of $\mathcal{P}_e$ by setting $[\mathcal{P}_e]_{k_m,k_m} = 0_{M_k \times M_k}$. Then, we introduce

$$\boldsymbol{q}_{k_m}(i) = [\mathcal{P}_e]_{k_m^-} \boldsymbol{n}_e(i) \tag{26}$$

for $m = 1, \ldots, j_k$, and

$$\boldsymbol{q}(i) = \mathrm{col}\left\{ \{\boldsymbol{q}_{k_m}(i)\}_{m=1}^{j_k} \right\}_{k=1}^{N}. \tag{27}$$

It then follows from Assumption 5 that

$$\mathbb{E}\boldsymbol{q}(i) = 0. \tag{28}$$

Lemma 1. *(Network error dynamics) Consider a network of N interacting agents running the distributed strategy* (11) *to* (14). *The evolution of the error dynamics across the network relative to the reference vector $\boldsymbol{w}_e^o$ defined by (20) is described by the following recursion:*

$$\widetilde{\boldsymbol{w}}_e(i) = \mathcal{A}\mathcal{P}_e(I - \boldsymbol{\mathcal{R}}_{u,e}(i))\widetilde{\boldsymbol{w}}_e(i-1) - \mathcal{A}\mathcal{P}_e\boldsymbol{g}(i) - \mathcal{A}\boldsymbol{q}(i). \tag{29}$$

Let

$$\mathcal{G} \triangleq \mathbb{E}\left[\boldsymbol{g}(i)\boldsymbol{g}^\top(i)\right] = \mathrm{diag}\left\{ \left(\mathbb{1}_{j_k}\mathbb{1}_{j_k}^\top\right) \otimes \left(\frac{\mu_k^2}{j_k^2} R_{u,k}\sigma_{v,k}^2\right) \right\}_{k=1}^{N},$$

which follows from Assumptions 1 to 3. We also introduce an $N_e \times N_e$ block matrix

$$\Gamma(i) = \mathbb{E}\left[\boldsymbol{q}(i)\boldsymbol{q}^\top(i)\right],$$

whose $M_k \times M_\ell$ (k_m, ℓ_n)-th block entry equals

$$\begin{aligned} [\Gamma(i)]_{k_m,\ell_n} &= \mathbb{E}\left[\boldsymbol{q}_{k_m}(i)\boldsymbol{q}_{\ell_n}^\top(i)\right] \\ &\overset{(26)}{=} \mathbb{E}\left[[\mathcal{P}_e]_{k_m^-}\boldsymbol{n}_e(i)\boldsymbol{n}_e^\top(i)[\mathcal{P}_e]_{\ell_n^-}^\top\right] \\ &\overset{(24)}{=} [\mathcal{P}_e]_{k_m^-} R_{n,e}(i) [\mathcal{P}_e]_{\ell_n^-}^\top. \end{aligned} \tag{30}$$

Let Σ be a symmetric positive semi-definite matrix. Under Assumptions 1 to 5, it follows from (29) that

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_e(i)\|^2_\Sigma = \mathbb{E}\|\widetilde{\boldsymbol{w}}_e(i-1)\|^2_{\Sigma'} + \mathrm{Tr}\,(\mathcal{G}\mathcal{P}_e\mathcal{A}\Sigma\mathcal{A}\mathcal{P}_e) + \mathrm{Tr}\,(\Gamma(i)\mathcal{A}\Sigma\mathcal{A}), \tag{31}$$

where

$$\Sigma' = \mathbb{E}\left[(I_{M_e} - \boldsymbol{\mathcal{R}}_{u,e}(i))\mathcal{P}_e\mathcal{A}\Sigma\mathcal{A}\mathcal{P}_e(I_{M_e} - \boldsymbol{\mathcal{R}}_{u,e}(i))\right]$$

with

$$M_e = \sum_{k=1}^{N} j_k M_k.$$

Note that [40, p.762]

$$\mathrm{Tr}\,(AB) = \left[\mathrm{vec}(B)^\mathsf{T}\right]^\mathsf{T} \mathrm{vec}(A) \tag{32}$$

$$\mathrm{vec}(ACB) = (B^\mathsf{T} \otimes A)\,\mathrm{vec}(C) \tag{33}$$

for any matrices $\{A, B, C\}$ of compatible sizes. Let

$$\sigma = \mathrm{vec}(\Sigma).$$

Now, we introduce

$$\mathcal{F} = \mathbb{E}\left[((I_{M_e} - \boldsymbol{\mathcal{R}}_{u,e}(i))\mathcal{P}_e\mathcal{A}) \otimes ((I_{M_e} - \boldsymbol{\mathcal{R}}_{u,e}(i))\mathcal{P}_e\mathcal{A})\right]. \tag{34}$$

Then, it follows from (33) that

$$\sigma' \triangleq \mathrm{vec}(\Sigma') = \mathcal{F}\sigma.$$

Now, we rewrite the recursion (31)

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_e(i)\|^2_\sigma = \mathbb{E}\|\widetilde{\boldsymbol{w}}_e(i-1)\|^2_{\mathcal{F}\sigma} + \left[\mathrm{vec}\,(\mathcal{A}\mathcal{P}_e\mathcal{G}\mathcal{P}_e\mathcal{A})\right]^\mathsf{T}\sigma + \left[\mathrm{vec}\,(\mathcal{A}\Gamma(i)\mathcal{A})\right]^\mathsf{T}\sigma, \tag{35}$$

where we used the property (32). Assume that

Assumption 6. *(Sufficiently small step-sizes) The step-sizes $\{\mu_k\}$ are sufficiently small, such that the matrix $\mathcal{F}$ defined by (34) is stable.*

Under Assumptions 1 to 6, as $i \to \infty$, it follows from (35) that

$$\lim_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_e(i)\|^2_{(I_{M_e^2} - \mathcal{F})\sigma} = \left[\mathrm{vec}\,(\mathcal{A}\mathcal{P}_e\mathcal{G}\mathcal{P}_e\mathcal{A})\right]^\mathsf{T}\sigma + \lim_{i\to\infty}\left[\mathrm{vec}\,(\mathcal{A}\Gamma(i)\mathcal{A})\right]^\mathsf{T}\sigma, \tag{36}$$

provided that matrix $\Gamma(i)$ is convergent, *i.e.*, the sequence $\{\sigma_{n,k}^2(i)\}$ is convergent for each agent k in view of (30). Let [8]

$$\sigma_{ss} = (I_{M_e^2} - \mathcal{F})^{-1} \operatorname{vec}\left(\frac{1}{N} \operatorname{diag}\left\{\frac{1}{j_k} I_{j_k M_k}\right\}_{k=1}^{N}\right). \tag{37}$$

Then, we have the following theorem for the steady-state network MSD performance:

Theorem 1. *(Network MSD performance) Consider a network of N interacting agents running the distributed strategy* (11) *to* (14). *Under Assumptions 1 to 6, and that variance sequence $\{\sigma_{n,k}^2(i)\}$ is convergent for all k, it holds that*

$$\textit{MSD}_{\textit{net}} = \left[\operatorname{vec}\left(\mathcal{A}\mathcal{P}_e\mathcal{G}\mathcal{P}_e\mathcal{A}\right)\right]^{\mathsf{T}} \sigma_{ss} + \lim_{i\to\infty} \left[\operatorname{vec}\left(\mathcal{A}\Gamma(i)\mathcal{A}\right)\right]^{\mathsf{T}} \sigma_{ss}. \tag{38}$$

In (38), variances of the privacy mechanism noises, $\{\sigma_{n,k}^2(i)\}$, are involved in the second term on the right hand side (RHS) via matrix $\Gamma(i)$ as shown by (30). It follows that: (a) firstly, the mean-square convergence of the proposed privacy-preserving multitask diffusion algorithm is ensured *only if* variances of the privacy mechanism noises $\{\sigma_{n,k}^2(i)\}$ are convergent; and (b) secondly, the network MSD, MSD_{net}, is a monotonically increasing function *w.r.t.* variances of the privacy mechanism noises, $\{\sigma_{n,k}^2(i)\}$.

B. Privacy Mechanism Noise Power

We start with the following sufficient condition for the variance of the privacy mechanism noise $\sigma_{n,k}^2(i)$ to satisfy the privacy constraint in (19) for each agent k and each time instant $i \geq 0$.

Theorem 2. *(Sufficient condition) It holds that if*

$$\sigma_{n,k}^2(i) \geq \frac{\operatorname{Tr}\left(U_{kk}^{\mathsf{T}}(i) U_{kk}(i)\right)}{\operatorname{Tr}\left(W_{kk}\right) - \delta_k} \tag{39}$$

for any agent k and any time instant $i \geq 0$, and where the quantity $U_{kk}(i)$ is defined by (17)*, then the privacy constraint in (19) is satisfied.*

Proof. Let

$$R_{\psi,k}(i) = \mathbb{E}\left[\left(\boldsymbol{\psi}_k(i) - \mathbb{E}\boldsymbol{\psi}_k(i)\right)\left(\boldsymbol{\psi}_k(i) - \mathbb{E}\boldsymbol{\psi}_k(i)\right)^{\mathsf{T}}\right]. \tag{40}$$

Then, it follows from Assumption 5 that

$$R_{\psi',k}(i) = R_{\psi,k}(i) + R_{n,k}(i) \tag{41}$$

where the $R_{\psi',k}(i)$ is defined by (18). It holds that [45, p.66]

$$\mathbb{E}\big\|\boldsymbol{w}_k^o - \widehat{\boldsymbol{w}}_{k|\psi_k'}(i)\big\|^2 = \mathrm{Tr}\left(W_{kk} - U_{kk}(i)R_{\psi',k}^{-1}(i)U_{kk}^{\mathsf{T}}(i)\right),$$

where the matrix W_{kk} is defined by (2). Substituting into the left hand side (LHS) of the privacy constraint in (19), we have

$$\mathrm{Tr}\left(U_{kk}(i)R_{\psi',k}^{-1}(i)U_{kk}^{\mathsf{T}}(i)\right) \le \mathrm{Tr}\left(W_{kk}\right) - \delta_k. \tag{42}$$

Note that the $M_k \times M_k$ symmetric positive semi-definite matrix $R_{\psi,k}(i)$ admits a spectral decomposition of the form:

$$R_{\psi,k}(i) = T_k(i)\Lambda_k(i)T_k^{\mathsf{T}}(i),$$

where $T_k(i)$ is an orthogonal matrix, *i.e.*,

$$T_k(i)T_k^{\mathsf{T}}(i) = T_k^{\mathsf{T}}(i)T_k(i) = I_{M_k}, \tag{43}$$

and $\Lambda_k(i)$ is a diagonal matrix consisting of the spectrum of $R_{\psi,k}(i)$. Let

$$\Lambda_k(i) = \mathrm{diag}\left\{\lambda_k^m(i)\right\}_{m=1}^{M_k}, \tag{44}$$

where the $\{\lambda_k^m(i)\}_{m=1}^{M_k}$ are eigenvalues of $R_{\psi,k}(i)$. Then, in view of the fact that $R_{\psi,k}(i)$ is a symmetric positive semi-definite matrix, we conclude that $\lambda_k^m(i) \ge 0$ for any $m = 1, \ldots, M_k$. Let

$$\Lambda_k'(i) = \Lambda_k(i) + R_{n,k}(i) \tag{45}$$

which is a diagonal matrix with entries $\{\lambda_k^m(i) + \sigma_{n,k}^2(i)\}_{m=1}^{M_k}$, and where the $R_{n,k}(i)$ is defined by (15). Then, it follows from (41) that

$$R_{\psi',k}(i) = T_k(i)\Lambda_k'(i)T_k^{\mathsf{T}}(i) \tag{46}$$

where we used (15) and (43). Provided that $\{\sigma_{n,k}^2(i) > 0\}$, it holds that

$$R_{\psi',k}^{-1}(i) = T_k(i)(\Lambda_k'(i))^{-1}T_k^{\mathsf{T}}(i). \tag{47}$$

Let

$$\widetilde{U}_{kk}(i) = T_k^{\mathsf{T}}(i)U_{kk}^{\mathsf{T}}(i)U_{kk}(i)T_k(i) \tag{48}$$

which has entries $\{\widetilde{u}_{kk}^m(i)\}_{m=1}^{M_k}$ on its main diagonal. Now, we substitute (47) into the LHS of (42), and obtain

$$\begin{aligned}
\mathrm{Tr}\left(U_{kk}(i)R_{\psi',k}^{-1}(i)U_{kk}^{\mathsf{T}}(i)\right) &\overset{(47)}{=} \mathrm{Tr}\left(U_{kk}(i)T_k(i)(\Lambda_k'(i))^{-1}T_k^{\mathsf{T}}(i)U_{kk}^{\mathsf{T}}(i)\right) \\
&\overset{\text{(a)}}{=} \mathrm{Tr}\left((\Lambda_k'(i))^{-1}\widetilde{U}_{kk}(i)\right) \\
&= \sum_{m=1}^{M_k}\frac{\widetilde{u}_{kk}^m(i)}{\lambda_k^m(i)+\sigma_{n,k}^2(i)} \qquad (49) \\
&\overset{(b)}{\leq} \sum_{m=1}^{M_k}\frac{\widetilde{u}_{kk}^m(i)}{\sigma_{n,k}^2(i)} \\
&= \frac{\mathrm{Tr}\left(\widetilde{U}_{kk}(i)\right)}{\sigma_{n,k}^2(i)} \\
&\overset{(48)}{=} \frac{\mathrm{Tr}\left(T_k^{\mathsf{T}}(i)U_{kk}^{\mathsf{T}}(i)U_{kk}(i)T_k(i)\right)}{\sigma_{n,k}^2(i)} \\
&= \frac{\mathrm{Tr}\left(U_{kk}^{\mathsf{T}}(i)U_{kk}(i)\right)}{\sigma_{n,k}^2(i)} \qquad (50)
\end{aligned}$$

where in step (a) we used the property $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$ for any square matrices A and B of compatible sizes, (49) follows from (48) and (45), in step (b) we used the fact that $\{\lambda_k^m(i) \geq 0\}$, and (50) follows from property (43). Then, substituting the upper bound in (50) into the LHS of (42), we arrive at the desired result in Theorem 2.

In addition, a numerical solution for $\sigma_{n,k}^2(i)$ can be obtained if we continue to substitute the results in (49) into the LHS of (42), and solve the following inequality

$$\sum_{m=1}^{M_k}\frac{\widetilde{u}_{kk}^m(i)}{\lambda_k^m(i)+\sigma_{n,k}^2(i)} \leq \mathrm{Tr}\,(W_{kk}) - \delta_k.$$

□

Then, in view of the fact that the steady-state network MSD is a monotonically increasing function *w.r.t.* the steady-state variances of the privacy mechanism noises $\{\sigma_{n,k}^2(i)\}$ as $i \to \infty$, in practice, we set for all k and all $i \geq 0$

$$\sigma_{n,k}^2(i) = \frac{\mathrm{Tr}\left(U_{kk}^{\mathsf{T}}(i)U_{kk}(i)\right)}{\mathrm{Tr}\,(W_{kk}) - \delta_k}, \qquad (51)$$

which is the smallest value that satisfies the sufficient condition (39). Now, we proceed to formulate the quantities $\{U_{kk}(i)\}$. Let

$$\boldsymbol{\psi}_e(i) = \mathrm{col}\left\{\mathbb{1}_{j_k} \otimes \boldsymbol{\psi}_k(i)\right\}_{k=1}^N.$$

Then, we introduce these error quantities

$$\widetilde{\boldsymbol{\psi}}_k(i) = \boldsymbol{w}_k^o - \boldsymbol{\psi}_k(i),$$
$$\widetilde{\boldsymbol{\psi}}_e(i) = \boldsymbol{w}_e^o - \boldsymbol{\psi}_e(i).$$

We also introduce

$$V_{k\ell}(i) = \mathbb{E}\left[(\boldsymbol{w}_k^o - \mathbb{E}\boldsymbol{w}_k^o)(\widetilde{\boldsymbol{\psi}}_\ell(i) - \mathbb{E}\widetilde{\boldsymbol{\psi}}_\ell(i))^\mathsf{T}\right]$$

for any agent k, ℓ, and

$$\begin{aligned}\mathcal{V}_e(i) &= \mathbb{E}\left[(\boldsymbol{w}_e^o - \mathbb{E}\boldsymbol{w}_e^o)(\widetilde{\boldsymbol{\psi}}_e(i) - \mathbb{E}\widetilde{\boldsymbol{\psi}}_e(i))^\mathsf{T}\right] \\ &= \mathbb{E}\left[\boldsymbol{w}_e^o\widetilde{\boldsymbol{\psi}}_e^\mathsf{T}(i)\right] - \mathbb{E}\boldsymbol{w}_e^o\left(\mathbb{E}\widetilde{\boldsymbol{\psi}}_e(i)\right)^\mathsf{T}. \end{aligned} \tag{52}$$

Now, for each agent k in the network, we introduce a $1 \times N$ block matrix $\mathcal{S}_k$, whose ℓ-th block is of size $M_k \times j_\ell M_\ell$, and equal to

$$\mathcal{S}_k = [0, \dots, 0, \underbrace{[I_{M_k}, 0_{M_k \times M_k}, \dots, 0_{M_k \times M_k}]}_{j_k}, 0, \dots, 0] \tag{53}$$

with a non-zero matrix at the k-th block. Then, it holds that

$$\begin{aligned} U_{kk}(i) &= W_{kk} - V_{kk}(i) \\ &= W_{kk} - \mathcal{S}_k\mathcal{V}_e(i)\mathcal{S}_k^\mathsf{T}, \end{aligned} \tag{54}$$

where W_{kk} is defined by (2). Now, we proceed to calculate the quantity $\mathcal{V}_e(i)$ in (52). It follows from the iterative equations (11) to (14) that

$$\widetilde{\boldsymbol{\psi}}_e(i+1) = (I_{M_e} - \boldsymbol{\mathcal{R}}_{u,e}(i+1))\mathcal{A}\mathcal{P}_e\widetilde{\boldsymbol{\psi}}_e(i) - (I_{M_e} - \boldsymbol{\mathcal{R}}_{u,e}(i+1))\mathcal{A}\boldsymbol{q}(i) - \boldsymbol{g}(i) \tag{55}$$

for any time instant $i \geq 0$, and where the quantities $\mathcal{A}$, $\boldsymbol{\mathcal{R}}_{u,e}(i)$, $\boldsymbol{g}(i)$, $\mathcal{P}_e$, and $\boldsymbol{q}(i)$ are defined by (21) to (23), (25) and (27), respectively. Under Assumption 2, it holds that

$$\mathcal{R}_{u,e} \triangleq \mathbb{E}\left[\boldsymbol{\mathcal{R}}_{u,e}(i)\right] = \mathrm{diag}\left\{I_{j_k} \otimes \frac{\mu_k}{j_k}R_{u,k}\right\}_{k=1}^N$$

where the $R_{u,k}$ is defined by (3). Then, we introduce

$$\mathcal{B} = (I_{M_e} - \mathcal{R}_{u,e})\mathcal{A}\mathcal{P}_e. \tag{56}$$

Under Assumptions 1, 3 and 5, it follows from (55) that

$$\mathbb{E}\widetilde{\boldsymbol{\psi}}_e(i+1) = \mathcal{B}\mathbb{E}\widetilde{\boldsymbol{\psi}}_e(i) \tag{57}$$

for any $i \geq 0$. Initially, at $i = 0$, it holds that

$$\mathbb{E}\widetilde{\boldsymbol{\psi}}_e(0) = (I_{M_e} - \mathcal{R}_{u,e})\mathbb{E}\boldsymbol{w}_e^o. \tag{58}$$

Under Assumptions 1, 3 and 5, it follows from (55) after some algebraic operations that

$$\mathbb{E}\left[\boldsymbol{w}_e^o \widetilde{\boldsymbol{\psi}}_e^{\mathsf{T}}(i+1)\right] = \mathbb{E}\left[\boldsymbol{w}_e^o \widetilde{\boldsymbol{\psi}}_e^{\mathsf{T}}(i)\right]\mathcal{B}^{\mathsf{T}} \tag{59}$$

for any time instant $i \geq 0$, and where

$$\mathbb{E}\left[\boldsymbol{w}_e^o \widetilde{\boldsymbol{\psi}}_e^{\mathsf{T}}(0)\right] = \mathbb{E}\left[\boldsymbol{w}_e^o \left(\boldsymbol{w}_e^o\right)^{\mathsf{T}}\right](I - \mathcal{R}_{u,e}). \tag{60}$$

In view of (52), it follows from (57) and (59) that

$$\mathcal{V}_e(i+1) = \mathcal{V}_e(i)\mathcal{B}^{\mathsf{T}} \tag{61}$$

for any $i \geq 0$. Let

$$\mathcal{W}_e = \mathbb{E}\left[(\boldsymbol{w}_e^o - \mathbb{E}\boldsymbol{w}_e^o)(\boldsymbol{w}_e^o - \mathbb{E}\boldsymbol{w}_e^o)^{\mathsf{T}}\right].$$

Then, it follows from (58) and (60) that

$$\mathcal{V}_e(0) = \mathcal{W}_e(I_{M_e} - \mathcal{R}_{u,e}). \tag{62}$$

Substituting (61) and (62) into the RHS of (54), we obtain the quantities $\{U_{kk}(i)\}$ for any agent k and any time instant $i \geq 0$, which are then used to calculate variances $\{\sigma_{n,k}^2(i)\}$ by following (51).

C. Boundedness and Convergence

Let

$$\delta_k = \rho_k \operatorname{Tr}(W_{kk}). \tag{63}$$

Then, in order to ensure that $\{\sigma_{n,k}^2(i) > 0\}$, it is required that $\{\rho_k < 1\}$. This follows from (51), where the numerator $\operatorname{Tr}\left(U_{kk}^{\mathsf{T}}(i)U_{kk}(i)\right) > 0$ since the matrix $U_{kk}^{\mathsf{T}}(i)U_{kk}(i)$ is symmetric positive semi-definite. Then, we proceed to show that the quantity $\mathcal{V}_e(i)$ is bounded and convergent. We start by noting that matrix $\mathcal{B}$ defined in (56) is stable for sufficiently small step-sizes $\{\mu_k\}$ [8]. Initially, at $i = 0$, it holds that $\mathcal{V}_e(0)$ defined in (62) is bounded. Then, expression (61) is a Bounded-Input Bounded-Output (BIBO) stable recursion. In view of the boundedness of $\mathcal{V}_e(i)$, which results in the boundedness of $U_{kk}(i)$ via (54), the variances of the privacy mechanism noises $\{\sigma_{n,k}^2(i)\}$ shown by (51) are bounded. Then, in view of the fact that matrix $\mathcal{B}$ is stable, it follows from (61) that the quantity $\mathcal{V}_e(i)$ vanishes as $i \to \infty$. At steady state, we have

$$\lim_{i\to\infty} U_{kk}(i) \overset{(54)}{=} W_{kk}.$$

Substituting into the RHS of (51) gives

$$\lim_{i\to\infty} \sigma_{n,k}^2(i) = \frac{\mathrm{Tr}\left(W_{kk}^\mathsf{T} W_{kk}\right)}{\mathrm{Tr}\left(W_{kk}\right) - \delta_k}. \tag{64}$$

V. Simulation Results

In this section, we test performance of the proposed privacy-preserving multitask diffusion algorithm in terms of network inference privacy and network MSD. For comparison, we also test performance of the multitask diffusion algorithm [8] and the non-cooperative least-mean-squares (LMS) algorithm, where each agent k updates estimate of $\boldsymbol{w}_k^o$ from $\boldsymbol{w}_k(i-1)$ to $\boldsymbol{w}_k(i)$ by following the LMS algorithm [45, p.165]. Specifically, in the test of privacy-preserving performance, we consider at time instant $i \geq 0$: (a) for the multitask diffusion algorithm without privacy mechanism noises, a neighboring agent $\ell \in \mathcal{N}_k \backslash \{k\}$, where $\mathcal{N}_k \backslash \{k\}$ stands for the set after removing agent k from $\mathcal{N}_k$, uses both intermediate estimates $\{\boldsymbol{\psi}_\ell(i), \boldsymbol{\psi}_k(i)\}$ to infer $\boldsymbol{w}_k^o$; (b) for the proposed privacy-preserving multitask diffusion algorithm, a neighboring agent $\ell \in \mathcal{N}_k \backslash \{k\}$ uses estimates $\{\boldsymbol{\psi}_\ell(i), \boldsymbol{\psi}_k'(i)\}$ to infer $\boldsymbol{w}_k^o$; (c) for the non-cooperative LMS algorithm, a neighboring agent $\ell \in \mathcal{N}_k \backslash \{k\}$ uses its own estimate $\boldsymbol{w}_\ell(i)$ to infer $\boldsymbol{w}_k^o$. Let

$$n_k = |\mathcal{N}_k|$$

be the cardinality of $\mathcal{N}_k$. Now, we proceed to introduce the following steady-state mean-square errors to quantify the network inference privacy of local parameters $\{\boldsymbol{w}_k^o\}$:

$$\xi_{\text{net}}^{\text{coop, noise}} = \lim_{i\to\infty} \frac{1}{N} \sum_{k=1}^{N} \frac{1}{n_k - 1} \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} \mathbb{E}\left[\left\|\boldsymbol{w}_k^o - \widehat{\boldsymbol{w}}_{k|\{\psi_k', \psi_\ell\}}(i)\right\|^2\right] \tag{65}$$

$$\xi_{\text{net}}^{\text{coop, w/o noise}} = \lim_{i\to\infty} \frac{1}{N} \sum_{k=1}^{N} \frac{1}{n_k - 1} \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} \mathbb{E}\left[\left\|\boldsymbol{w}_k^o - \widehat{\boldsymbol{w}}_{k|\{\psi_k, \psi_\ell\}}(i)\right\|^2\right] \tag{66}$$

$$\xi_{\text{net}}^{\text{ncop}} = \lim_{i\to\infty} \frac{1}{N} \sum_{k=1}^{N} \frac{1}{n_k - 1} \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} \mathbb{E}\left[\left\|\boldsymbol{w}_k^o - \widehat{\boldsymbol{w}}_{k|w_\ell}(i)\right\|^2\right] \tag{67}$$

where the superscripts “coop, noise”, “coop, w/o noise”, and “ncop” denote the quantities for cooperative case with privacy mechanism noises, cooperative case without privacy mechanism noises and non-cooperative case, respectively.

TABLE I: The Steady-state Network Inference Privacy.

	Multitask Diffusion [8]	Proposed Algorithm	Non-coop. LMS
ξ_{net} (dB)	-1.750	-1.696	-1.493

As shown by Fig. 1a, we consider the case when there are $N = 6$ agents in the network. The random data $\{\boldsymbol{u}_k(i), \boldsymbol{v}_k(i)\}$ are independent, normally distributed with zero mean, and white over time and space. The lengths of the unknown parameter vectors $\{\boldsymbol{w}_k^o\}$ are $\{M_k = 2\}$. The agents in the network are involved in $Q = 5$ linear equality constraints, each of the form [8]:

$$\sum_{k \in \mathcal{I}_q} d_{qk} w_k + b_q = 0$$

with the scalar parameters $\{d_{qk}, b_q\}$ randomly selected from $[-3, -1] \cup [1, 3]$. Let

$$\text{SNR}_k = 10 \log_{10} \left(\mathbb{E} \left[(\boldsymbol{u}_k^\mathsf{T}(i) \boldsymbol{w}_k^o)^2 \right] / \sigma_{v,k}^2 \right)$$

be the signal-to-noise ratio (SNR) at agent k. Then, the parameters $\{R_{u,k}, W_{kk}, \mathbb{E}\boldsymbol{w}_k^o, \sigma_{v,k}^2\}$ are adjusted to make $\{\text{SNR}_k\}$ as shown by Fig. 1b. For the step-size parameters, we set $\{\mu_k / j_k = 0.02\}$ in the cooperative cases, and $\{\mu_k = 0.02\}$ in the non-cooperative case. In addition, we set the parameters $\{\rho_k = 0.1\}$, which are defined by (63). Fig. 2 shows the network MSD learning curves of the tested strategies, which are averaged over 1000 independent realizations of $\{\boldsymbol{w}_k^o\}$. In addition, Table I shows the steady-state mean-square errors defined by (65) to (67) in order to evaluate the privacy-preserving performance of the related schemes. It is clear from Fig. 2 and Table I that under the tested privacy requirement, the proposed privacy-preserving multitask diffusion strategy is able to balance the trade-off between estimation accuracy and privacy protection.

VI. Conclusion

We have developed a privacy-preserving diffusion strategy over multitask networks, which is able to protect each agent's local task by adding a privacy mechanism noise before sharing with its neighbors. We have proposed a utility-privacy optimization trade-off to determine the amount of noise to add. We have derived a sufficient condition for the powers of the privacy mechanism noises which satisfies the proposed privacy constraints. We have shown that the proposed powers

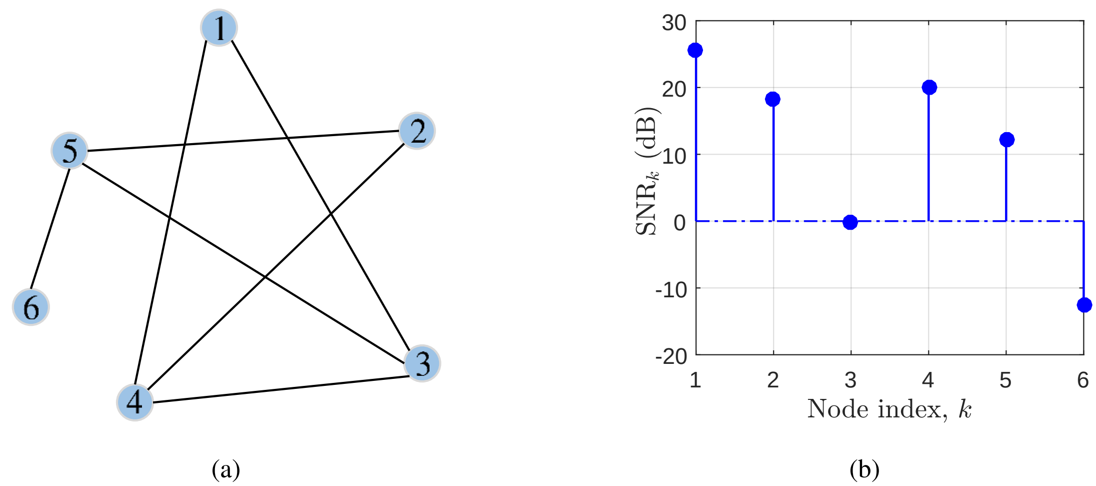

Fig. 1: Network topology consisting of $N = 6$ agents (left) and SNRs across the agents (right).

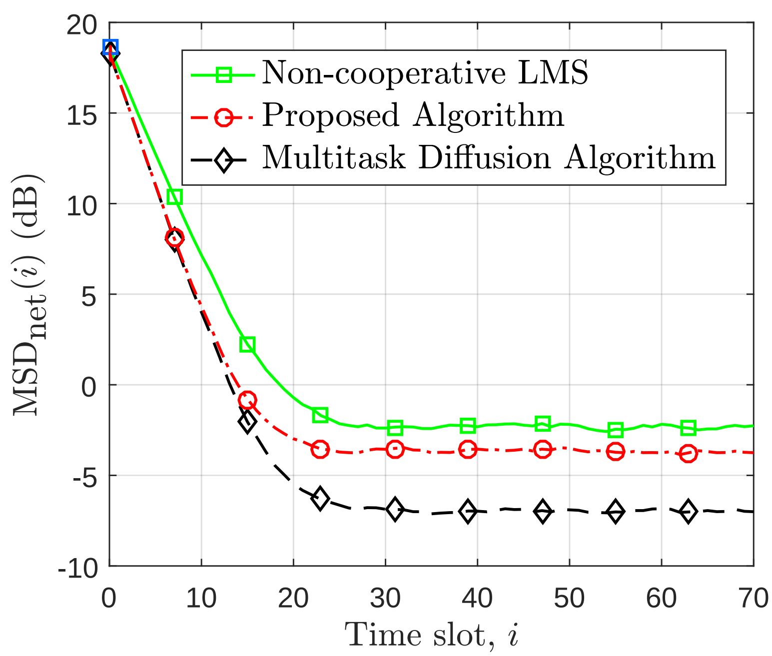

Fig. 2: Network MSD learning curves of the multitask diffusion algorithm [8], the non-cooperative LMS algorithm, and the proposed privacy-preserving multitask diffusion algorithm.

are bounded and convergent. We have presented simulation results to demonstrate that the proposed scheme is able to balance the trade-off between network MSD and network inference privacy.

REFERENCES

[1] X. Zhao and A. H. Sayed, "Clustering via diffusion adaptation over networks," in *Proc. Int. Workshop Cognitive Information Processing*, Baiona, Spain, May 2012.

[2] ——, "Distributed clustering and learning over networks," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3285 – 3300, Jul. 2015.

[3] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733 – 2748, Jun. 2015.

[4] ——, "Multitask diffusion adaptation over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129 – 4144, Aug. 2014.

[5] J. Plata-Chaves, N. Bogdanović, and K. Berberidis, "Distributed diffusion-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3448 – 3460, Jul. 2015.

[6] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Proximal multitask learning over networks with sparsity-inducing coregularization," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6329 – 6344, Dec. 2016.

[7] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks with common latent representations," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 563 – 579, Apr. 2017.

[8] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Diffusion LMS for multitask problems with local linear equality constraints," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 4979 – 4993, Oct. 2017.

[9] Y. Wang, W. P. Tay, and W. Hu, "A multitask diffusion strategy with optimized inter-cluster cooperation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 504 – 517, Apr. 2017.

[10] M. Chan, E. Campo, D. Estève, and J.-Y. Fourniols, "Smart homes – Current features and future perspectives," *Maturitas*, vol. 64, no. 2, pp. 90 – 97, Oct. 2009.

[11] L. D. Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233 – 2243, Jan. 2014.

[12] A. Whitmore, A. Agarwal, and L. Da Xu, "The Internet of Things – A survey of topics and trends," *Information Systems Frontiers*, vol. 17, no. 2, pp. 261 – 274, Apr. 2015.

[13] S. H. Shah and I. Yaqoob, "A survey: Internet of Things (IoT) technologies, applications and challenges," in *Proc. IEEE Smart Energy Grid Engineering*, Oshawa, ON, Canada, Aug. 2016.

[14] G. Misra, V. Kumar, A. Agarwal, and K. Agarwal, "Internet of Things (IoT) – A technological analysis and survey on vision, concepts, challenges, innovation directions, technologies, and applications (an upcoming or future generation computer communication system technology)," *American J. of Elect. and Electronic Engineering*, vol. 4, no. 1, pp. 23 – 32, 2016.

[15] S. Li, L. Da Xu, and S. Zhao, "5G Internet of Things: A survey," *J. of Ind. Inf. Integration*, vol. 10, pp. 1 – 9, Jun. 2018.

[16] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?" *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4155 – 4168, Sep. 2008.

[17] M. Leng, W. P. Tay, T. Q. S. Quek, and H. Shin, "Distributed local linear parameter estimation using gaussian SPAWN," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 244 – 257, Jan. 2015.

[18] W. P. Tay, "Whose opinion to follow in multihypothesis social learning? A large deviations perspective," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 344 – 359, Mar. 2015.

[19] J. Ho, W. P. Tay, T. Q. Quek, and E. K. Chong, "Robust decentralized detection and social learning in tandem networks," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5019 – 5032, Oct. 2015.

[20] Y. Wang, X. Wu, and H. Donghui, "Using randomized response for differential privacy preserving data collection," in *Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Washington, D.C., USA, Aug. 2003.

[21] S. Xiong, A. D. Sarwate, and N. B. Mandayam, "Randomized requantization with local differential privacy," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016.

[22] A. D. Sarwate and L. Sankar, "A rate-disortion perspective on local differential privacy," in *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello, IL, USA, Sep. 2014.

[23] J. Liao, L. Sankar, F. P. Calmon, and V. Y. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *Proc. IEEE Int. Symp. on Inform. Theory*, Aachen, Germany, Jun. 2017.

[24] H. Imtiaz and A. D. Sarwate, "Differentially private distributed principal component analysis," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Calgary, Alberta, Canada, Apr. 2018.

[25] X. Ren, C.-M. Yu, W. Yu, S. Yang, X. Yang, J. A. McCann, and S. Y. Philip, "LoPub: High-dimensional crowdsourced data publication with local differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2151 – 2166, Mar. 2018.

[26] B. Razeghi and S. Voloshynovskiy, "Privacy-preserving outsourced media search using secure sparse ternary codes," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Calgary, Alberta, Canada, Apr. 2018.

[27] F. D. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," in *Proc. ACM SIGMOD Inter. Conf. Management of data*, Providence, Rhode Island, USA, Jun. 2009.

[28] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Annual Inter. Conf. on the Theory and Applications of Cryptographic Techniques*, St. Petersburg, Russia, May 2006.

[29] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. IEEE Symp. on Foundations of Computer Science*, Monticello, IL, USA, Oct. 2013.

[30] M. Sun and W. P. Tay, "Privacy-preserving nonparametric decentralized detection," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016.

[31] X. He, W. P. Tay, and M. Sun, "Privacy-aware decentralized detection using linear precoding," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Rio de Janeiro, Brazil, Jul. 2016.

[32] M. Sun and W. P. Tay, "Inference and data privacy in IoT networks," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Commun.*, Sapporo, Japan, Jul. 2017.

[33] M. Sun, W. P. Tay, and X. He, "Toward information privacy for the Internet of Things: A non-parametric learning approach," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1734 – 1747, Apr. 2018.

[34] X. He, M. Sun, W. P. Tay, and Y. Gong, "Multilayer nonlinear processing for information privacy in sensor networks," *arXiv preprint arXiv:1711.04459*, 2018.

[35] K. Diamantaras and S. Kung, "Data privacy protection by kernel subspace projection and generalized eigenvalue decomposition," in *IEEE Int. Workshop Machine Learning for Signal Processing*, Vietri sul Mare, Italy, Sep. 2016.

[36] S. Y. Kung, "Compressive privacy from information estimation," *IEEE Signal Process. Mag.*, vol. 34, no. 1, pp. 94 – 112, Jan. 2017.

[37] ——, "A compressive privacy approach to generalized information bottleneck and privacy funnel problems," *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 1846 – 1872, Mar. 2018.

[38] M. Al, S. Wan, and S. Kung, "Ratio utility and cost analysis for privacy preserving subspace projection," *arXiv preprint arXiv:1702.07976*, 2017.

[39] J. Chen and A. H. Sayed, "Distributed Pareto optimization via diffusion strategies," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 2, pp. 205 – 220, Apr. 2013.

[40] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, pp. 311 – 801, 2014. [Online]. Available: http://dx.doi.org/10.1561/2200000051

[41] R. Nassif, C. Richard, J. Chen, A. Ferrari, and A. H. Sayed, "Diffusion LMS over multitask networks with noisy links," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016.

[42] I. E. K. Harrane, R. Flamary, and C. Richard, "Toward privacy-preserving diffusion strategies for adaptation and learning over networks," in *Proc. of European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, Aug. 2016.

[43] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 974 – 979, Feb. 2012.

[44] X. Zhao, S.-Y. Tu, and A. H. Sayed, "Diffusion adaptation over networks under imperfect information exchange and non-stationary data," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3460 – 3475, Jul. 2012.

[45] A. H. Sayed, *Adaptive Filters*. New York, USA: Wiley-IEEE Press, 2008.