Collaborative Mean Estimation over Intermittently Connected Networks with Peer-To-Peer Privacy

Rajarshi Saha
Stanford University
rajsaha@stanford.edu

Mohamed Seif
Princeton University
mseif@princeton.edu

Michal Yemini
Bar-Ilan University
michal.yemini@biu.ac.il

Andrea J. Goldsmith
Princeton University
goldsmith@princeton.edu

H. Vincent Poor
Princeton University
poor@princeton.edu

2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206910

***Abstract*—This work considers the problem of *Distributed Mean Estimation* (DME) over networks with intermittent connectivity, where the goal is to learn a global statistic over the data samples localized across distributed nodes with the help of a central server. To mitigate the impact of intermittent links, nodes can collaborate with their neighbors to compute local consensus which they forward to the central server. In such a setup, the communications between any pair of nodes must satisfy *local differential privacy* constraints. We study the tradeoff between collaborative relaying and privacy leakage due to the additional data sharing among nodes and, subsequently, propose a novel differentially private collaborative algorithm for DME to achieve the optimal tradeoff. Finally, we present numerical simulations to substantiate our theoretical findings.**

I. Introduction

Distributed Mean Estimation (DME) is a fundamental statistical problem that arises in several applications, such as model aggregation in federated learning [1], distributed K-means clustering [2], distributed power iteration [3], etc. DME presents several practical challenges, which prior research [4]–[8] has considered, including the problem of straggler nodes, where nodes cannot send their data to the parameter server (PS). Typically, there are two types of stragglers: (i) *computation stragglers*, in which nodes cannot finish their local computation within a deadline, and (ii) *communication stragglers*, in which nodes cannot transmit their updates due to communication blockage [9]–[14]. The problem of communication stragglers can be solved by relaying the updates/data to the PS via neighboring nodes. This approach was proposed and analyzed in [15]–[17], where it was shown that the proposed collaborative relaying scheme can be optimized to reduce the expected distance to optimality, both for DME [15] and federated learning [16], [17].

While the works [15]–[17] show that collaborative relaying reduces the expected distance to optimality, exchanging the individual data across nodes incurs privacy leakage caused by the additional estimates that are shared among the nodes. Nonetheless, this potential breach of privacy has not been addressed in the aforementioned works. To mitigate the privacy leakage in DME, we require a rigorous privacy notion. Within the context of distributed learning, *local differential privacy* (LDP) [18] has been adopted as a gold standard notion of privacy, in which a user can perturb and disclose a sanitized version of its data to an *untrusted* server. LDP ensures that the statistics of the user's output observed by adversaries are indistinguishable regardless of the realization of any input data. In this paper, we focus on *node-level* LDP where the neighboring nodes, as well as any eavesdropper that can observe the local node-node transmissions during collaborations, cannot infer the realization of the user's data.

There has been extensive research into the design of distributed learning algorithms that are both communication efficient and private (see [19] for a comprehensive survey and references therein). It is worth noting that LDP requires a significant amount of perturbation noise to ensure reasonable privacy guarantees. Nonetheless, the amount of perturbation noise can be significantly reduced by considering the intermittent connectivity of nodes in the learning process [20]. The intermittent connectivity in DME amplifies the privacy guarantees; it provides a boosted level of anonymity due to partial communication with the server. Various random node participation schemes have been proposed to further improve the utility-privacy tradeoff in distributed learning, such as Poisson sampling [21], importance sampling [22], [23], and sampling with/without replacement [20]. In addition, Balle et al. investigated in [24], the privacy amplification in federated learning via random check-ins and showed that the privacy leakage scales as $O(1/\sqrt{n})$, where n is the number of nodes. In other words, random node participation reduces the amount of noise required to achieve the same levels of privacy that are achieved without sampling.

So far, works in the privacy literature, such as [18]–[27], have not considered intermittent connectivity along with collaborative relaying, where nodes share their local updates to mitigate the randomness in network connectivity [15]–[17]. Thus, this paper aims to close this theoretical gap. To this end, we first show that there exists a tradeoff between collaborative relaying and privacy leakage due to data sharing among nodes for DME under intermittent connectivity assumption. We introduce our system model and proposed algorithm in §II, followed by its utility (MSE) and privacy analyses in §III and §IV respectively. We quantify the utility (MSE) and privacy tradeoff by formulating it as an optimization problem and solve it approximately due to its non-convexity. Finally, we demonstrate the efficacy of our private collaborative algorithm through numerical simulations. Due to space limitations, details can be found in the online extended version [28].

This work was supported by the AFOSR award #002484665, a Huawei Intelligent Spectrum grant, and NSF grants CCF-1908308 & CNS-2128448.

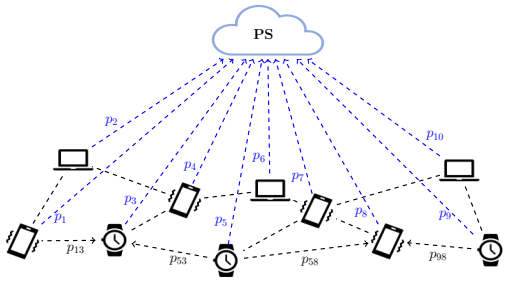

Fig. 1: An intermittently connected distributed learning network. Blue and black dotted lines denote intermittent node-PS and node-node connections. Communication between any two nodes must satisfy local differential privacy constraints.

II. System Model for Private Collaboration

Consider a distributed system with n nodes, each having a vector $\mathbf{x}_i \in \mathbb{R}^d$, $\|\mathbf{x}_i\|_2 \leq \mathrm{R}$ for some known $\mathrm{R} > 0$. The nodes communicate with a PS, as well as with each other over intermittent links with the goal of estimating their mean, $\overline{\mathbf{x}} \triangleq \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ at the PS (Fig. 1). For any estimate $\widehat{\overline{\mathbf{x}}}$ of the mean, the evaluation metric for any estimate is the MSE, given by $\mathcal{E} \triangleq \mathbb{E}\|\widehat{\overline{\mathbf{x}}} - \overline{\mathbf{x}}\|_2^2$.

A. Communication Model

As shown in Fig. 1, node i can communicate with the PS with probability p_i, with the link modeled using a Bernoulli random variable $\tau_i \sim \mathrm{Ber}(p_i)$. Similarly, node i can communicate with another node j with probability p_{ij}, i.e., $\tau_{ij} \sim \mathrm{Ber}(p_{ij})$. The links between different node pairs are assumed to be statistically independent, i.e., $\tau_i \perp \tau_j$ for $i \neq j$, $\tau_{ij} \perp \tau_{ml}$ for $(i,j) \neq (m,l)$, $(j,i) \neq (m,l)$, and $\tau_{ij} \perp \tau_l$ for $i,j,l \in [n]$. The correlation due to channel reciprocity between a pair of nodes $i,j \in [n]$ is denoted by $\mathrm{E}_{\{i,j\}} \equiv \mathbb{E}[\tau_{ij}\tau_{ji}]$. We assume that $\mathrm{E}_{\{i,j\}} \geq p_{ij}p_{ji}$, i.e., $\mathbb{P}(\tau_{ij}=1|\tau_{ji}=1) \geq \mathbb{P}(\tau_{ij}=1)$. Furthermore, $p_{ii} = 1 \ \forall \ i \in [n]$, and if node i can never transmit to j, we set $p_{ij} = 0$. We denote $\mathbf{p} \equiv (p_1, \ldots, p_n)$ and $\mathbf{P} \equiv (p_{ij})_{i,j\in[n]} \in [0,1]^{n\times n}$.

B. Privacy Model

The nodes are assumed to be honest but curious. They are *honest* because they faithfully compute the aggregate of the signals received; however, they are *curious* as they might be interested in learning sensitive information about nodes. Each node uses a local additive noise mechanism to ensure the privacy of its transmissions to neighboring nodes. We consider local privacy constraints, wherein node i trusts another node j to a certain extent and hence, randomizes its own data accordingly when sharing with node j by using a synthetic Gaussian noise (see [18]) to respect the privacy constraint while maintaining utility. We present a refresher on differential privacy and Gaussian mechanism in [28, App. A].

C. Private Collaborative Relaying for Mean Estimation

We now introduce our algorithm, **PriCER**: **Pri**vate **C**ollaborative **E**stimation via **R**elaying. PriCER is a two-stage semi-decentralized algorithm for estimating the mean. In the first stage, each node $j \in [n]$ sends a scaled and noise-added version of its data to a neighboring node $i \in [n]$ over the intermittent link τ_{ji}. The transmitted signal is given by

$$\widetilde{\mathbf{x}}_{ji} = \tau_{ji}(\alpha_{ji}\mathbf{x}_j + \mathbf{n}_{ji}) \tag{1}$$

Here, $\alpha_{ji} \geq 0$ is the weight used by node j while sending to node i, and $\mathbf{n}_{ji} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_d)$ is the multivariate Gaussian privacy noise added by node j. Here, σ^2 is the variance of each coordinate, and $\mathbf{I}_d \in \mathbb{R}^{d\times d}$ is the identity matrix. We denote the weight matrix by $\mathbf{A} \equiv (\alpha_{ij})_{i,j\in[n]}$. Consequently, node i computes the local aggregate of all received signals as

$$\widetilde{\mathbf{x}}_i = \sum_{j\in[n]} \tau_{ji}(\alpha_{ji}\mathbf{x}_j + \mathbf{n}_{ji}). \tag{2}$$

We quantify our privacy guarantees using the well-established notion of differential privacy [18]. By observing $\widetilde{\mathbf{x}}_{ji}$, node i should not be able to distinguish between the events when node j contains the data $\mathbf{x}_j$ versus when it contains some other data $\mathbf{x}'_j$. In other words, we are interested in protecting the local data of node j from a (potentially untrustworthy) neighboring node i. We assume that the privacy noise processes added by different nodes are uncorrelated, i.e., $\mathbb{E}[\mathbf{n}_{il}^\top \mathbf{n}_{jm}] = 0$ for all $i,j,l,m \in [n]$, when i,j,l,m are not all equal. In the second stage, each node i transmits $\widetilde{\mathbf{x}}_i$ to the PS over the intermittent link τ_i, and the PS computes the global estimate. Algs. 1 & 2 provide the pseudocode for PriCER.

Algorithm 1 PriCER-Stage 1 for local aggregation

Input: Non-negative weight matrix $\mathbf{A}$
Output: $\widetilde{\mathbf{x}}_i$ for all $i \in [n]$

1: **for** each $i \in [n]$ **do**
2: Locally generate $\mathbf{x}_i$
3: Transmit $\widetilde{\mathbf{x}}_{ij} = \alpha_{ij}\mathbf{x}_i + \mathbf{n}_{ij}$ to nodes $j \in [n] : j \neq i$
4: Receive $\widetilde{\mathbf{x}}_{ji} = \tau_{ji}(\alpha_{ji}\mathbf{x}_j + \mathbf{n}_{ji})$ from $j \in [n] : j \neq i$
5: Set $\widetilde{\mathbf{x}}_{ii} = \alpha_{ii}\mathbf{x}_i + \mathbf{n}_{ii}$
6: Locally aggregate available signals: $\widetilde{\mathbf{x}}_i = \sum_{j\in n} \widetilde{\mathbf{x}}_{ji}$
7: Transmit $\widetilde{\mathbf{x}}_i$ to the PS
8: **end for**

Algorithm 2 PriCER-Stage 2 for global aggregation

Input: $\tau_i\widetilde{\mathbf{x}}_i$ for all $i \in [n]$
Output: Estimate of the mean at the PS: $\widehat{\overline{\mathbf{x}}}$

1: **for** Each $i \in [n]$ **do**
2: Receive $\tau_i\widetilde{\mathbf{x}}_i$
3: **end for**
4: Aggregate the received signals: $\widehat{\overline{\mathbf{x}}} = \frac{1}{n}\sum_{i\in[n]} \tau_i\widetilde{\mathbf{x}}_i$

III. Mean Square Error Analysis

The goal of PriCER is to obtain an unbiased estimate of $\overline{\mathbf{x}}$ at the PS. Since each node sends its data to all other neighboring nodes, the PS receives multiple copies of the same data. Lemma 3.1, below gives a sufficient condition to ensure unbiasedness. This is the same condition as [15, Lemma 3.1], and holds true even for PriCER.

Lemma 3.1: Let the weights $\{\alpha_{ij}\}_{i,j\in[n]}$ satisfy

$$\sum_{j\in[n]} p_j p_{ij}\alpha_{ij} = 1, \tag{3}$$

for every $i \in [n]$. Then, $\mathbb{E}\left[\widehat{\overline{\mathbf{x}}} \mid \{\mathbf{x}_i\}_{i\in[n]}\right] = \overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$.

We prove this lemma in [28, App. B]. Under this unbiasedness condition, we derive a worst-case upper bound for the MSE in Thm. 3.2.

Theorem 3.2: Given $\mathbf{p}, \mathbf{P}$ and $\mathbf{A}$ such that (3) holds, and $\mathbf{n}_{ij} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_d)\ \forall\, i,j \in [n]$, the MSE with PriCER satisfies

$$\mathbb{E}\|\widehat{\overline{\mathbf{x}}} - \mathbf{x}\|_2^2 \le \mathrm{R}^2\sigma_{\mathrm{tv}}^2(\mathbf{p},\mathbf{P},\mathbf{A}) + \sigma_{\mathrm{pr}}^2(\mathbf{p},\mathbf{P},\sigma), \tag{4}$$

where $\sigma_{\mathrm{tv}}^2(\mathbf{p},\mathbf{P},\mathbf{A})$ is an upper bound on the variance induced by the stochasticity due to the intermittent topology given by

$$\sigma_{\mathrm{tv}}^2(\mathbf{p},\mathbf{P},\mathbf{A}) \triangleq \frac{1}{n^2}\Bigg[\sum_{i,j,l\in[n]} p_j(1-p_j)p_{ij}p_{lj}\alpha_{ij}\alpha_{lj} + \sum_{i,j\in[n]} p_{ij}p_j(1-p_{ij})\alpha_{ij}^2 + \sum_{i,l\in[n]} p_i p_l(\mathrm{E}_{\{i,l\}} - p_{il}p_{li})\alpha_{li}\alpha_{il}\Bigg],$$

and $\sigma_{\mathrm{pr}}^2(\mathbf{p},\mathbf{P},\sigma)$ is the variance due to the privacy given by

$$\sigma_{\mathrm{pr}}^2(\mathbf{p},\mathbf{P},\sigma) \triangleq \frac{1}{n^2}\sum_{i,j\in[n]} p_j p_{ij}\sigma^2 d. \tag{5}$$

Thm. 3.2 is derived in [28, App. C]. From (4), we see that $\sigma_{\mathrm{pr}}^2(\mathbf{p},\mathbf{P},\sigma)$ is the price of privacy. For a non-private setting, i.e., $\sigma = 0$, the privacy induced variance $\sigma_{\mathrm{pr}}^2(\mathbf{p},\mathbf{P},\sigma) = 0$, and Thm. 3.2 simplifies to [15, Thm. 3.2]. In the following section, we introduce our privacy guarantee and the corresponding constraints leading to a choice of weight matrix $\mathbf{A}$ for the optimal Utility (MSE) - Privacy tradeoff of PRICER.

IV. Privacy Analysis

PRICER yields privacy guarantees for two reasons: (i) the local noise added at each node, and (ii) the intermittent nature of the connections. We consider the local differential privacy when any eavesdropper (possibly including the receiving node) can observe the transmission from node i to node j in stage-1 of PRICER. Let us denote the local dataset of node i as $\mathcal{D}_i \subset \mathbb{R}^d$. In DME, $\mathcal{D}_i$ is a singleton set and by observing the transmission from node i to node j, the eavesdropper should not be able to differentiate between the events $\mathbf{x}_i \in \mathcal{D}_i$ and $\mathbf{x}_i' \in \mathcal{D}_i$, where $\mathbf{x}_i' \neq \mathbf{x}_i$. The following Thm. 4.1 (derived in [28, App. D]) formally states this guarantee.

Theorem 4.1: Given $\mathbf{n}_{ij} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_d)$, $\mathbf{x}_i, \mathbf{x}_i' \in \mathbb{R}^d$ with $\|\mathbf{x}_i\|_2, \|\mathbf{x}_i'\|_2 \le \mathrm{R}$, and any $\delta_{ij} \in (0,1]$, for pairs $(\epsilon_{ij}, p_{ij}\delta_{ij})$ satisfying

$$\epsilon_{ij} = \begin{cases} \left[2\ln\left(\frac{1.25}{\delta_{ij}}\right)\right]^{\frac{1}{2}} \frac{2\alpha_{ij}\mathrm{R}}{\sigma} & \text{if } p_{ij} > 0, \quad \text{and,} \\ 0 & \text{if } p_{ij} = 0, \end{cases} \tag{6}$$

the transmitted signal from node i to node j, $\widetilde{\mathbf{x}}_{ij} = \tau_{ij}(\alpha_{ij}\mathbf{x}_i + \mathbf{n}_{ij})$ is $(\epsilon_{ij}, p_{ij}\delta_{ij})$-differentially private, i.e., it satisfies

$$\Pr(\widetilde{\mathbf{x}}_{ij} \in \mathcal{S} | \mathbf{x}_i \in \mathcal{D}_i) \le e^{\epsilon_{ij}}\Pr(\widetilde{\mathbf{x}}_{ij} \in \mathcal{S} | \mathbf{x}_i' \in \mathcal{D}_i) + p_{ij}\delta_{ij}, \tag{7}$$

for any measurable set $\mathcal{S}$.

Setting $\delta := p_{ij}\delta_{ij}$, we can immediately see that intermittent connectivity inherently boosts privacy, since for the same δ for any pair $i,j \in [n]$, the privacy level ϵ_{ij} is proportional to $\ln(1.25p_{ij}/\delta)^{\frac{1}{2}}$, implying that a smaller p_{ij} leads to a stronger privacy guarantee. Additionally, from (6), the privacy guarantee ϵ_{ij} is directly related to the weight α_{ij}. That is, if node i trusts node j more, ϵ_{ij} can be relatively large, and consequently, node i can assign a higher weight to the data it sends to node j. On the other hand, if node i does not trust node j as much, a smaller value will be assigned to α_{ij}. In other words, for the same noise variance σ, node i will scale the signal α_{ij} so as to reduce the effective signal-to-noise ratio in settings where a higher privacy is required. Finally, when $p_{ij} = 0$, PRICER ensures $\alpha_{ij} = 0$, implying $\epsilon_{ij} = 0$, i.e., perfect privacy, albeit zero utility. Our weight optimization (§V) aims to minimize the MSE subject to the privacy constraints imposed by (6).

V. Privacy Constrained Weight Optimization

When deriving the utility-privacy tradeoff, our objective is to minimize the MSE at the PS subject to desired privacy guarantees, namely $(\underline{\epsilon}_{ij}, \underline{\delta}_{ij}p_{ij})$ node-node differential privacy. Here, $\underline{\epsilon}_{ij}, \underline{\delta}_{ij}$ are pre-designated system parameters that quantify the extent to which node i trusts node j (or alternatively, how much it trusts the communication link $i \to j$ against an eavesdropper). More specifically, we solve the following:

$$\begin{aligned} \min_{\mathbf{A},\sigma}\ & \mathrm{R}^2\sigma_{\mathrm{tv}}^2(\mathbf{p},\mathbf{P},\mathbf{A}) + \sigma_{\mathrm{pr}}^2(\mathbf{p},\mathbf{P},\sigma) \\ \text{s.t.:}\ & \alpha_{ij} \ge 0,\ \forall\, i,j \in [n], && \text{(non-negative weights)} \\ & \sum_{j\in[n]} p_j p_{ij}\alpha_{ij} = 1,\ \forall\, i \in [n], && \text{(unbiasedness)} \\ & \left[2\ln\left(\frac{1.25}{\underline{\delta}_{ij}}\right)\right]^{\frac{1}{2}} \frac{2\alpha_{ij}\mathrm{R}}{\sigma} \le \underline{\epsilon}_{ij}\ \ \forall i,j \in [n], && \text{(privacy)} \\ & \sigma \ge 0. && \text{(privacy).} \end{aligned} \tag{8}$$

The above optimization (8) is not necessarily convex. Furthermore, the objective is also not separable with respect to $\mathbf{A}$ and σ. Thus, in what follows, we propose an alternate minimization scheme, where we iteratively minimize with respect to $\mathbf{A}$ and σ; one variable at a time, keeping the other fixed. We tie up the components of §V-B and §V-A and present the complete PRICER weight and variance optimization algorithm in Alg. 3. For clarity of presentation, we assume that $p_i > 0$ for all $i \in [n]$, so we can have a simple initialization rule.

A. Optimize the variance σ for given weights $\mathbf{A}$

The non-negative weights, unbiasedness, and privacy constraints are present in (8) due to our problem formulation. However, when using alternating optimization we must choose a variance that can fulfill the unbiasedness condition in the weight optimization stage. In other words, PriCER needs to add a minimum amount of noise, σ_{thr}, in order to meet privacy constraints and unbiasedness conditions simultaneously. Thus, we introduce a necessary condition to ensure a non-empty feasible set when we optimize the weight for the chosen σ.

We visualize this in Fig. 2. Note that for a fixed $i \in [n]$, the unbiasedness constraint together with $\alpha_{ij} \ge 0$, defines a hyperplane $\mathcal{H}$ in the positive quadrant of $\mathbb{R}^n$ with respect to the

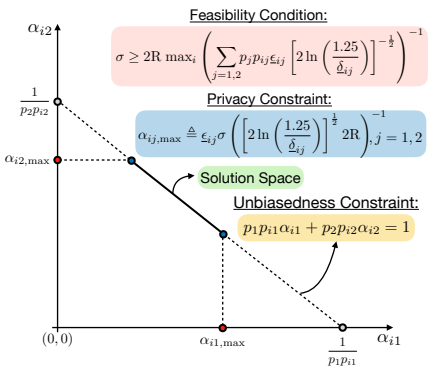

Fig. 2: Feasible solution for the optimization problem in (8) with $n = 2$ nodes and $i = 1, 2$. Note that $\alpha_{11} = \alpha_{22} = 1$.

optimization variables $\{\alpha_{ij}\}_{j\in[n]}$. Moreover, for $i \in [n]$, the constraints $\alpha_{ij} \geq 0$ and $\alpha_{ij} \leq \underline{\epsilon}_{ij}\sigma \cdot ([2\ln(1.25/\underline{\delta}_{ij})]^{\frac{1}{2}}2\mathrm{R})^{-1}$ $\forall\ j \in [n]$, together define a box $\mathcal{B}$ aligned with the standard basis of $\mathbb{R}^n$ with one of the vertices at the origin. The edge length of this box along any of the axes is proportional to σ When $\sigma = 0$, i.e., no privacy noise is added, $\mathcal{B} = \mathbf{0}_n$, where $\mathbf{0}_n$ denotes the origin of $\mathbb{R}^n$. Since $\mathcal{H}$ does not pass through $\mathbf{0}_n$, (11) is infeasible. This implies there is a minimum value of σ so that $\mathcal{B}$ is big enough to have a non-zero intersection with $\mathcal{H}$. More specifically, we require σ such that $\sigma \sum_{j\in[n]} p_j p_{ij} \underline{\epsilon}_{ij} \left(\left[2\ln\left(\frac{1.25}{\underline{\delta}_{ij}}\right)\right]^{\frac{1}{2}} 2\mathrm{R} \right)^{-1} \geq 1, \forall\, i \in [n]$, and hence, we have the last feasibility constraint in (9), where

$$\sigma_{\mathrm{thr}} \triangleq 2\mathrm{R}\, \max_{i\in[n]} \left(\sum_{j\in[n]} p_j p_{ij} \underline{\epsilon}_{ij} [2\ln(\frac{1.25}{\underline{\delta}_{ij}})]^{-\frac{1}{2}} \right)^{-1} \geq 0.$$

We now fix $\mathbf{A}$ in (9) and minimize the PIV, i.e.,

$$\begin{aligned} &\min_{\sigma}\ \frac{1}{n^2} \sum_{i,j\in[n]} p_j p_{ij} \sigma^2 d \\ &\text{s.t.:}\ \left[2\ln\left(\frac{1.25}{\underline{\delta}_{ij}}\right)\right]^{\frac{1}{2}} \frac{2\alpha_{ij}\mathrm{R}}{\sigma} \leq \underline{\epsilon}_{ij}\ \ \forall i, j \in [n], \\ &\quad \sigma \geq \sigma_{\mathrm{thr}}. \end{aligned} \tag{9}$$

It can be shown [28, App. E-A] that the update rule is:

$$\sigma = \max\left\{ \max_{i,j\in[n]} \left\{ \left[2\ln\left(\frac{1.25}{\underline{\delta}_{ij}}\right)\right]^{\frac{1}{2}} \frac{2\alpha_{ij}\mathrm{R}}{\underline{\epsilon}_{ij}} \right\}, \sigma_{\mathrm{thr}} \right\} \tag{10}$$

B. Optimize the weights $\mathbf{A}$ *for given variance* σ

Firstly, for a fixed σ, we optimize the weights $\mathbf{A}$, i.e.,

$$\begin{aligned} &\min_{\mathbf{A}}\ \mathrm{R}^2\sigma^2_{\mathrm{tv}}(\mathbf{p}, \mathbf{P}, \mathbf{A}) \\ &\text{s.t.:}\ \ \alpha_{ij} \geq 0,\ \forall\, i, j \in [n], \quad \sum_{j\in[n]} p_j p_{ij} \alpha_{ij} = 1,\ \forall\, i \in [n], \\ &\quad \left[2\ln\left(\frac{1.25}{\underline{\delta}_{ij}}\right)\right]^{\frac{1}{2}} \frac{2\alpha_{ij}\mathrm{R}}{\sigma} \leq \underline{\epsilon}_{ij}\ \ \forall i, j \in [n]. \end{aligned} \tag{11}$$

The objective function of problem (11) is not convex. With this in mind, we adopt an approach similar to [15], [17] wherein (11) is minimized in two iterative stages – (i) first, a convex relaxation of (11) is minimized using Gauss-Seidel method, and (ii) the outcome is subsequently fine-tuned again, using Gauss-Seidel on (11). The convex relaxation is chosen to be:

$$\min_{\mathbf{A}}\ \mathrm{R}^2\overline{\sigma}^2_{\mathrm{tv}}(\mathbf{p}, \mathbf{P}, \mathbf{A})\ \text{ s.t. the same constraints as (11),} \tag{12}$$

where the new objective function is,

$$\begin{aligned} \overline{\sigma}^2_{\mathrm{tv}}(\mathbf{p}, \mathbf{P}, \mathbf{A}) \triangleq \frac{1}{n^2} \Bigg[&\sum_{i,j,l\in[n]} p_j(1-p_j) p_{ij} p_{lj} \alpha_{ij} \alpha_{lj} \\ &+ \sum_{i,j\in[n]} p_{ij} p_j (1-p_{ij}) \alpha^2_{ij} + \sum_{i,l\in[n]} p_i p_l (\mathrm{E}_{\{i,l\}} - p_{il} p_{li}) \alpha^2_{il} \Bigg], \end{aligned} \tag{13}$$

We delegate the complete derivations to [28, App. E] and only mention the update rules here. Let us denote the i^{th} row of $\mathbf{A}$ as $\mathbf{A}_i$. Since the objective functions of both (11) and (13) are separable with respect to $\mathbf{A}_i$, we can apply Gauss-Seidel iterations on both (12) and subsequently on (11).

Minimizing the convex relaxation (12): Let us denote the iterate at the ℓ^{th} Gauss-Seidel iteration of the convex relaxation (13) as $\mathbf{A}^{(\ell)}$. Then, the update rule is given by

$$\mathbf{A}^{(\ell)}_i = \begin{cases} \widehat{\mathbf{A}}^{(\ell)}_i & \text{if } i = \ell \bmod n + n\, \mathbb{1}_{\{\ell \bmod n = 0\}}, \\ \mathbf{A}^{(\ell-1)}_i & \text{otherwise,} \end{cases} \tag{14}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. In one iteration, only the i^{th} row, i.e., the weights assigned by the i^{th} node for its neighbors, are updated. Since Gauss-Seidel performs block-wise descent, the update $\widehat{\mathbf{A}}^{(\ell)}_i \equiv \{\widehat{\alpha}_{ij}\}_{i,j\in[n]}$ can be obtained by formulating the Lagrangian [28, App. E-B]. Let us denote $\widetilde{w}_{ij} \triangleq \underline{\epsilon}_{ij}\sigma([2\ln(\frac{1.25}{\underline{\delta}_{ij}})]^{\frac{1}{2}}2\mathrm{R})^{-1}$ and

$$\overline{\alpha}_{ij}(\lambda_i) \triangleq \left(\frac{-2(1-p_j)\sum_{l\in[n]:l\neq i} p_{lj}\alpha^{(\ell-1)}_{lj} + \lambda_i}{2[(1-p_j p_{ij}) + p_i(E_{\{i,j\}}/p_{ij} - p_{ji})]} \right)^{+}. \tag{15}$$

We separate the solution into three scenarios:

a) $\boldsymbol{p_i < 1}$ **and** $\boldsymbol{p_j p_{ij} < 1}$ **for all** $\boldsymbol{j \in [n]}$**:** In this case,

$$\widehat{\alpha}_{ij} = \min\{\widetilde{\alpha}_{ij}(\lambda_i), \widetilde{w}_{ij}\}, \tag{16}$$

where $\widetilde{\alpha}_{ij}(\lambda_i)$ is given by

$$\widetilde{\alpha}_{ij}(\lambda_i) = \begin{cases} \overline{\alpha}_{ij}(\lambda_i) & \text{if } p_j p_{ij} > 0, \\ 0 & \text{if } p_j p_{ij} = 0. \end{cases} \tag{17}$$

Here, $(a)^+ \triangleq \max\{a, 0\}$, and $\lambda_i \geq 0$ is set such that $\sum_{j\in[n]} p_j p_{ij} \widehat{\alpha}_{ij} = 1$. λ_i is found using bisection search.

b) $\boldsymbol{p_i < 1}$ **and there exists** $\boldsymbol{j \neq i}$ **such that** $\boldsymbol{p_j p_{ij} = 1}$**:** Denote $S_i = \sum_{k\in[n]} \mathbb{1}_{\{p_k p_{ik} = 1\}} \widetilde{w}_{ik}$. If $S_i \geq 1$, then we choose, $\widehat{\alpha}_{ij} = \widetilde{w}_{ij}/S_i$ for all j such that $p_j p_{ij} = 1$, and $\widehat{\alpha}_{ij} = 0$ otherwise. If $S_i < 1$, we set $\widehat{\alpha}_{ij} = \widetilde{w}_{ij}$ for nodes j that satisfy $p_j p_{ij} = 1$, and subsequently allocate the residual $1 - S_i$ of the unbiasedness condition to minimize the objective function. Similar to (16), this will yield that $\widehat{\alpha}_{ij} = \min\{\widetilde{\alpha}_{ij}(\lambda_i), \widetilde{w}_{ij}\}$, where $\widetilde{\alpha}_{ij}(\lambda_i)$ is given by

$$\widetilde{\alpha}_{ij}(\lambda_i) = \begin{cases} \overline{\alpha}_{ij}(\lambda_i) & \text{if } p_j p_{ij} \in (0, 1), \\ 0 & \text{if } p_j p_{ij} = 0. \end{cases} \tag{18}$$

Here, $\lambda_i \geq 0$ is such that $\sum_{j: p_j p_{ij} \in (0,1)} p_j p_{ij} \widehat{\alpha}_{ij} = 1 - S_i$.

c) $\boldsymbol{p_i = 1}$**:** In this case, to preserve privacy we set

$$\widehat{\alpha}^{(\ell)}_{ij} = \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Algorithm 3 MSE-Privacy tradeoff: Joint opt. of $\boldsymbol{A}$ and σ^2

Input: Connection probabilities: $\mathbf{p} > 0$, $\mathbf{P}$, Pairwise privacy levels: $\{\underline{\epsilon}_{ij}, \underline{\delta}_{ij}\}_{i,j\in[n]}$, Maximal iterations: K, L_1, L_2.

Output: Weight matrix $\mathbf{A}^{(\mathrm{K})}$ and privacy noise variance $\sigma^{(\mathrm{K})}$ that approximately solve (8) .

Initialize: $\sigma^{(0)} = \sigma_{\mathrm{thr}}$ and $\mathbf{A}^{(0)} = \mathrm{diag}\left(\frac{1}{p_1}, \ldots, \frac{1}{p_n}\right)$.

for $k \leftarrow 0$ **to** $\mathrm{K}-1$ **do**
  $k \leftarrow k+1$.
  Set $\mathbf{A}^{(k,0)} \leftarrow \mathbf{A}^{(k-1)}$.
  Initialize $\ell \leftarrow 0$.
  for $\ell \leftarrow 0$ **to** $\mathrm{L}_1 - 1$ **minimize convex relaxation**, **do**
    $\ell \leftarrow \ell + 1$.
    $i \leftarrow \ell \mod n + n \cdot \mathbb{1}_{\{\ell \mod n = 0\}}$.
    Compute $\widehat{\mathbf{A}}_i^{(k,\ell)}$ according to (15)-(19).
    Set $\mathbf{A}_i^{(k,\ell)}$ according to (14).
  end for
  Warm initialize $\mathbf{A}^{(k,0)} \leftarrow \mathbf{A}^{(k,\mathrm{L})}$, re-initialize $\ell \leftarrow 0$.
  for $\ell \leftarrow 0$ **to** $\mathrm{L}_2 - 1$ **fine tune**, **do**
    $\ell \leftarrow \ell + 1$.
    $i \leftarrow \ell \mod n + n \cdot \mathbb{1}_{\{\ell \mod n = 0\}}$.
    Compute $\widehat{\mathbf{A}}_i^{(k,\ell)}$ according to (16)-(20).
    Set $\mathbf{A}_i^{(k,\ell)}$ according to (14).
  end for
  Set $\mathbf{A}^{(k)} \leftarrow \mathbf{A}^{(k,\mathrm{L})}$.
  For weights $\mathbf{A}^{(k)}$, set $\sigma^{(k)}$ according to (10).
end for

Fine tuning (11): We now fine tune the above solution by setting it as a warm start initialization and performing Gauss-Seidel on (11). Then, the update equation for fine tuning is of the same form as (14) and (16)-(19). However, we substitute the updated quantity $\widetilde{\alpha}_{ij}(\lambda_i)$ for this case, which is now

$$\widetilde{\alpha}_{ij}(\lambda_i) = \left(\frac{1}{2(1-p_j p_{ij})}\left(-2(1-p_j)\sum_{l\in[n]:l\neq i} p_{lj}\alpha_{lj}^{(\ell-1)} - 2p_i(E_{\{i,j\}}/p_{ij} - p_{ji})\alpha_{ji}^{(\ell-1)} + \lambda_i\right)\right)^+. \quad (20)$$

VI. Numerical Simulations

In Fig. 3, we consider a setup with $n = 10$ nodes that can collaborate over an Erdős-Rényi topology, i.e., $P_{ij} = \mathrm{p_c}$ for $j \neq i$ and $P_{ii} = 1$. The nodes can communicate to the PS with probabilities $\mathbf{p} = [0.1, 0.1, 0.8, 0.1, 0.1, 0.9, 0.1, 0.1, 0.9, 0.1]$, i.e., only three clients have good connectivity. Even though any node can communicate with any other node with a non-zero probability, they do not do so as they only trust a small number of immediate neighbors, which lies along the x-axis. If node i trusts node j, we set $\epsilon_{ij} = \epsilon_{\mathrm{high}} = 10^3$ (low privacy), otherwise, $\epsilon_{ij} = \epsilon_{\mathrm{low}} = 0.1$ (high privacy). Moreover, $\epsilon_{ii} = \epsilon_{\mathrm{high}}$. We also set $\delta_{ij} = \delta = 10^{-3}$. The y-axis shows the (optimized) objective value of (8). As is evident from Fig. 3, the MSE decreases as nodes trust more neighbors, as expected.

In Fig. 4, we consider that the data with $d = 1000$ at each node is generated from a Gaussian distribution $\mathcal{N}(0, 1)$, raised to the power of 3, and normalized – resulting in a heavy-tailed distribution. Consequently, if a node that has a vector with a few large coordinate values is unable to convey its data to the PS due to a failed transmission, this can incur a significant MSE. In this setup only some nodes have good connectivity to the PS, i.e., $p_i = p_{\mathrm{good}} = 0.9$, and the remaining have $p_i = p_{\mathrm{bad}} = 0.2$ In the naïve strategy, the PS averages whatever it successfully receives, i.e., it computes the mean estimate as $\frac{1}{n}\sum_{i\in[n]} \tau_i \mathbf{x}_i$. Whereas, in our collaborative strategy, each node trusts 6 other neighbors and can communicate with them with a probability $P_{ij} = 0.8$. Clearly, PriCER achieves a lower MSE than the naïve strategy. The plots are averaged over 50 realizations.

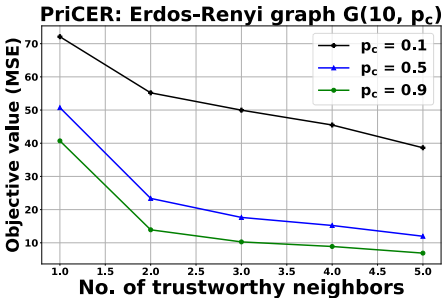

Fig. 3: Variation of worst-case MSE with trustworthy neighbors

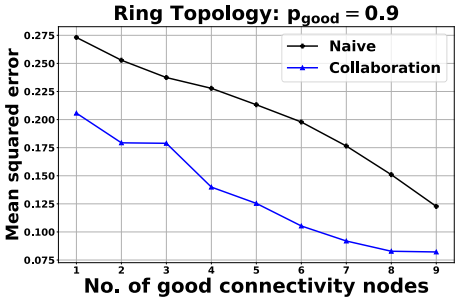

Fig. 4: Variation of MSE with number of good connectivity nodes

VII. Conclusions

In this paper, we have considered the problem of mean estimation over intermittently connected networks with collaborative relaying subject to peer-to-peer local differential privacy constraints. The nodes participating in the collaboration do not trust each other completely and, in order to ensure privacy, they scale and perturb their local data when sharing with others. We have proposed a two-stage consensus algorithm (PriCER), that takes into account these peer-to-peer privacy constraints to jointly optimize the scaling weights and noise variance so as to obtain an unbiased estimate of the mean at the PS that minimizes the MSE. Numerical simulations have shown the improvement of our algorithm relative to a non-collaborative strategy in MSE for various network topologies. Although this work considers peer-to-peer privacy, there can be other sources of privacy leakage such as at the PS. Moreover, adding correlated privacy noise may help reduce the MSE even further. In future work, we plan to include investigating these questions in more detail.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 2017, pp. 1273–1282.

[2] M.-F. F. Balcan, S. Ehrlich, and Y. Liang, "Distributed k-means and k-median clustering on general topologies," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 26, 2013.

[3] A. T. Suresh, X. Y. Felix, S. Kumar, and H. B. McMahan, "Distributed mean estimation with limited communication," in *Proceedings of the International Conference on Machine Learning (ICML)*. PMLR, 2017, pp. 3329–3337.

[4] D. Jhunjhunwala, A. Mallick, A. Gadhikar, S. Kadhe, and G. Joshi, "Leveraging spatial and temporal correlations in sparsified mean estimation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 14 280–14 292, 2021.

[5] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, 2009.

[6] R. Tandon, Q. Lei, A. G. Dimakis, and N. Karampatziakis, "Gradient coding: Avoiding stragglers in distributed learning," in *Proceedings of the International Conference on Machine Learning (ICML)*. PMLR, 2017, pp. 3368–3376.

[7] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3315–3326, 2008.

[8] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[9] M. Gapeyenko, A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, M. R. Akdeniz, E. Aryafar, N. Himayat, S. Andreev, and Y. Koucheryavy, "On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 124–10 138, Nov 2017.

[10] Y. Yan and Y. Mostofi, "Co-optimization of communication and motion planning of a robotic operation under resource constraints and in fading environments," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1562–1572, April 2013.

[11] M. M. Zavlanos, M. B. Egerstedt, and G. J. Pappas, "Graph-theoretic connectivity control of mobile robot networks," *Proceedings of the IEEE*, vol. 99, no. 9, pp. 1525–1540, Sep. 2011.

[12] N. Michael, M. M. Zavlanos, V. Kumar, and G. J. Pappas, "Maintaining connectivity in mobile robot networks," in *Experimental Robotics*, 2009.

[13] M. Yemini, S. Gil, and A. J. Goldsmith, "Exploiting local and cloud sensor fusion in intermittently connected sensor networks," in *Proceedings of the 2020 IEEE Global Communications Conference (Globecom)*, December 2020.

[14] M. Yemini, S. Gil, and A. J. Goldsmith, "Cloud-cluster architecture for detection in intermittently connected sensor networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 2, pp. 903–919, 2023.

[15] R. Saha, M. Yemini, E. Ozfatura, D. Gündüz, and A. Goldsmith, "ColRel: Collaborative relaying for federated learning over intermittently connected networks," in *Proceedings of the Workshop on Federated Learning: Recent Advances and New Challenges (in Conjunction with NeurIPS 2022)*, 2022. [Online]. Available: https://openreview.net/forum?id=8b0RHdh2Xd0

[16] M. Yemini, R. Saha, E. Ozfatura, D. Gündüz, and A. J. Goldsmith, "Semi-decentralized federated learning with collaborative relaying," in *Proceedings of the 2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1471–1476.

[17] M. Yemini, R. Saha, E. Ozfatura, D. Gündüz, and A. J. Goldsmith, "Robust federated learning with connectivity failures: A semi-decentralized framework with collaborative relaying," arXiv preprint arXiv:2202.11850, 2022.

[18] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[19] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[20] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018.

[21] Y. Zhu and Y.-X. Wang, "Poisson subsampled rényi differential privacy," in *Proceedings of the International Conference on Machine Learning (ICML)*. PMLR, 2019, pp. 7634–7642.

[22] B. Luo, W. Xiao, S. Wang, J. Huang, and L. Tassiulas, "Tackling system and statistical heterogeneity for federated learning with adaptive client sampling," in *Proceedings of the 2022 IEEE International Conference on Computer Communications (INFOCOM)*, 2022, pp. 1739–1748.

[23] E. Rizk, S. Vlaski, and A. H. Sayed, "Federated learning under importance sampling," *IEEE Transactions on Signal Processing*, vol. 70, pp. 5381–5396, 2022.

[24] B. Balle, P. Kairouz, B. McMahan, O. Thakkar, and A. Guha Thakurta, "Privacy amplification via random check-ins," *Advances in Neural Information Processing Systems(NeurIPS)*, vol. 33, pp. 4623–4634, 2020.

[25] H. Asi, V. Feldman, and K. Talwar, "Optimal algorithms for mean estimation under local differential privacy," in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 17–23 Jul 2022, pp. 1046–1056.

[26] M. Gaboardi, R. Rogers, and O. Sheffet, "Locally private mean estimation: z-test and tight confidence intervals," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 89. PMLR, 16–18 Apr 2019, pp. 2545–2554.

[27] Q. Xue, Y. Zhu, and J. Wang, "Mean estimation over numeric data with personalized local differential privacy," *Frontiers of Computer Science*, vol. 16, 06 2022.

[28] R. Saha, M. Seif, M. Yemini, A. J. Goldsmith, and H. V. Poor, "Collaborative mean estimation over intermittently connected networks with peer-to-peer privacy," arXiv preprint arXiv:2303.00035, 2023.