Multi-Message Shuffled Privacy in Federated Learning

Antonious M. Girgis and Suhas Diggavi

We study the distributed mean estimation (DME) problem under privacy and communication constraints in the local differential privacy (LDP) and multi-message shuffled (MMS) privacy frameworks. The DME has wide applications in both federated learning and analytics. We propose a communication-efficient and differentially private algorithm for DME of bounded ℓ_2-norm and ℓ_∞-norm vectors. We analyze our proposed DME schemes showing that our algorithms have order-optimal privacy-communication-performance trade-offs. Our algorithms are designed by giving unequal privacy assignments at different resolutions of the vector (through binary expansion) and appropriately combining it with coordinate sampling. These results are directly applied to give guarantees on private federated learning algorithms. We also numerically evaluate the performance of our private DME algorithms.

***Index Terms*—Differential privacy, mean estimation, federated learning, shuffled model.**

I. Introduction

Federated learning (FL) is a distributed system approach to build machine learning models from multiple clients without directly sharing the local data [2], [3]. In standard FL algorithms, the central server sends the global model to a set of sampled clients at each round. The server aggregates the local updates (stochastic gradients) of the participated clients to update the global model towards the next round. In FL, communication becomes a bottleneck for training high dimensional model as the communication is performed over a limited-bandwidth networks [3]–[5]. To address this challenge, there are several work for designing communication-efficient FL algorithms [6]–[8]. Besides communication, the clients' data might contain sensitive information, and hence, each client wants to preserve privacy of her own local data. Although, the local data doesn't leave the client's device, FL algorithm cannot provide a provable privacy guarantees, where sensitive data can be reconstructed from observing the global model and/or the local updates [9]–[12]. Differential privacy (DP) [13] has become a standard definition of privacy in privacy-preserving data analysis. DP ensures that the participation of a single client in a database does not change the probability of an outcome by much. Thus, providing DP guarantees for FL algorithms has received a considerable attention from academia as well as industry [14]–[21].

To accommodate privacy of locally held data, a more appropriate notion is local differential privacy (LDP) [22], [23], where each client randomizes her own message before sending it to the (untrusted) server. However, LDP mechanisms suffers from poor performance comparing with the central DP mechanisms [22], [23]. To improve the performance of LDP mechanisms, an intermediate trusted model called *shuffled model* has been proposed [24]–[26]. In the shuffled model, there exists a trusted shuffler that randomly permutes the randomized messages of the clients before passing them to the server. The shuffled model amplifies the privacy guarantees of the LDP mechanism and achieves better privacy-utility performance in different statistical and learning problems [27]–[29]. The goal of this paper is to design communication-efficient and private mechanisms for federated learning in the LDP and the multi-message shuffled models.

A. Contributions and Techniques

At the core of FL algorithms, the server wants to estimate the mean of local update vectors at each round. Therefore, we study the problem of distributed mean estimation (DME) under privacy and communication constraints in both LDP and MMS privacy models.

We propose simple and effective mechanisms for DME of bounded ℓ_∞-norm and ℓ_2-norm vectors. We prove that our proposed mechanisms achieve order optimal mean squared error (MSE) for all privacy and communication regimes simultaneously for the multi-message shuffled (MMS) model. We show that there exists an (ε, δ)-DP mechanism in the MMS model that has MSE $\mathcal{O}\left(\frac{d}{n^2 \min\{\varepsilon^2,\varepsilon\}}\right)$, and requires $\mathcal{O}\left(d \log\left(\frac{n\min\{\varepsilon^2,\varepsilon\}}{d}\right)\right)$ bits per client when $d \leq n\min\{\varepsilon^2,\varepsilon\}$ and $\mathcal{O}\left(n\min\{\varepsilon^2,\varepsilon\}\log\left(\frac{d}{n\min\{\varepsilon^2,\varepsilon\}}\right)\right)$ bits per client when $d > n\min\{\varepsilon^2,\varepsilon\}$ to estimate the mean of n bounded ℓ_2-norm vectors. We believe our mechanism is the first scheme to achieve simultaneously the order optimal privacy-communication-accuracy trade-offs that which matches the best known lower bound.

Observe that our proposed scheme has significant savings in communication cost to achieve the same privacy and MSE in MMS model comparing to the best known results in literature. The results in [30] requires $\mathcal{O}\left(d\sqrt{n}\right)$-bits of communication per client to achieve order optimal MSE. In [31], Chang

Antonious M. Girgis and Suhas Diggavi are with the University of California, Los Angeles, USA.

Email: amgirgis@ucla.edu, suhas@ee.ucla.edu.

This work was supported in part by NSF grants 2139304, 2007714.

This work has been presented in part in [1].

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes Appendices for proofs and discussions referenced in the paper. Contact amgirgis@ucla.edu for further questions about this work.

et al. proposed a private mechanism in MMS model that requires $\mathcal{O}\left(d\log(n)\right)$-bits of communication per client. For example, our proposed scheme achieves a multiplicative gain of $\mathcal{O}\left(\frac{d}{n\min\{\varepsilon,\varepsilon^2\}}\right)$ in communication per client when $d < n\min\{\varepsilon,\varepsilon^2\}$. Furthermore, our MMS mechanism requires significantly less amount of communication per client than used in secure aggregation to achieve the same privacy and the same order of MSE, where secure aggregation requires at least $\mathcal{O}\left(n^2\varepsilon^2\right)$-bits of communication per client [32] (see Remark 7 for more details). A similar result is concurrently and independently proven in [33].

Our proposed schemes can be applied directly in the LDP model and achieve order-optimal privacy-communication-accuracy trade-offs. We use the results of communication-efficient and private DME to analyze privacy-communication-convergence trade-offs of the DP-SGD algorithm (similar to algorithms in [17], [30]). In addition, we evaluate the performance of our proposed algorithms for scalar and vector private DME showing that our proposed MMS mechanisms achieve a significant improvements comparing to single-message shuffle model.

The core technical idea of our proposed scheme consists of three stages as follows. Suppose that the i-client holds a vector $\mathbf{x}_i$. In the first stage, each vector $\mathbf{x}_i$ is represented as a weighted summation of m binary vectors: $\mathbf{x}_i \approx \sum_{k=1}^{m-1} 2^{-k}\mathbf{b}_i^{(k)} + 2^{-m+1}\mathbf{u}_i^{(m)}$ for communication efficiency, where the last binary vector is dedication for unbiasedness. In the second stage, for given privacy target ε, we apply *private-waterfilling* to privatize the binary vectors, where we allocate unequal privacy for different binary vectors $\{\mathbf{b}_i^{(k)}\}$. We assign lower privacy for most significant bits (MSBs) ($\varepsilon^{(k)} \approx 4^{-k/3}\varepsilon$ such that $\sum_{k=1}^{m}\varepsilon^{(k)} = \varepsilon$). Observe that lower privacy implies better accuracy. Thus, this gives better performance in terms of mean squared error (MSE), as MSBs has higher weight. Finally, we privatize each binary vector using coordinate sampling and binary randomized response mechanism, where coordinate sampling helps in reducing the total communication cost. Furthermore, we track the total privacy of our mechanism using the Rényi differential privacy (RDP) for careful accounting of composition.

B. Related Work

We discuss the most relevant work related to the paper and review their connections to our work.

a) Private DME: Distributed mean estimation (DME) in the local DP model is well-studied with a characterization of the optimal privacy-communication-utility trade-off (see [17], [34], [35] and reference therein). In [34], Chen *et al.* established the order optimal private DME under LDP constraints for bounded ℓ_2-norm vectors. In [17], Girgis *et al.* established order optimal private DME under LDP constraints for bounded ℓ_∞-norm and separately for bounded ℓ_2-norm vectors. It also extended its use in the single-shuffled model and private optimization framework. In [21], [36], a family of communication-efficient mechanisms are proposed under LDP constraints in federated learning. Our scheme also achieves the optimal privacy-communication-utility trade-offs for LDP framework (see Theorems 4 and 6).

However, LDP mechanisms suffer from high MSE comparing to the central DP mechanisms. To improve the performance of the LDP mechanism without a need for a trusted server, the shuffle model has been proposed [24]–[26], where a secure shuffler randomly permutes the private messages of the clients before sending them to the untrusted server. For single-message shuffle model, Balle *et al.* presented lower and matching upper bounds for the scalar private real summation, showing that the MSE is order $\Theta\left(n^{1/3}\right)$. his was further enhanced by using multi-message shufflers in [29], [37]. A MMS mechanism based on IKOS scheme [38] was proposed in [29], [37] for scalar summation in which each client needs to send only $\mathcal{O}(1)$ messages to the shuffler, each of size $\mathcal{O}(\log(n))$ bits. The private vector DME has received less attention in the shuffled model. In [30], a MMS mechanism for vector summation is proposed which has $\mathcal{O}(d\sqrt{n})$ communication bits per client, where d is the vector dimension. In [31], a MMS mechanism for vector summation in MMS model is proposed that requires $\mathcal{O}\left(d\log(n)\right)$-bits of communication per client. In this work, we establish the fundamental privacy-communication-performance trade-offs for computing *vector sum* in the multi-message shuffle (MMS) model. Our private vector DME results in Theorem 7 improves the privacy-communication-performance order-wise, see Table I for comparison.

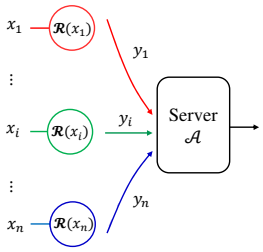
After the completion of our work and posting on arxiv [39], a closely related work [40] was posted. This work independently obtained similar results with different proof techniques. There are small differences in our results in details, but there are broad similarities in these independently obtained results.

b) Private optimization in the shuffled model: Local differentially private optimization has been studied in [17], [41] and references therein. Recently [17], [42], [43] have proposed DP-SGD algorithms for federated learning in the shuffled model. In [44], Girgis *et al.* studied a private optimization framework using RDP and additionally evaluated subsampling (of clients) in the shuffled model. The use of RDP for establishing tight composition bounds for interactive optimization in the shuffled model has been studied in [45], [46]. For the multi-message shuffled (MMS) model, private convex optimization has been studied in [30], which used at its core the private vector DME mechanism. We use our results of vector DME to analyze the convergence of DP-SGD in the MMS model showing that we can obtain optimal convergence rate with low communication costs per client per round.

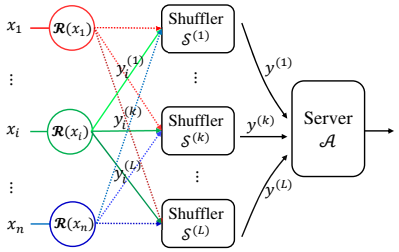
C. Organization

The remainder of the paper is organized as follows. We provide preliminary background on privacy definitions in Section II. We present the problem setup in Section III. We provide the main results of this paper and also present our proposed algorithms in Section IV. We give numerical results comparing our proposed multi-message shuffled schemes with the best known single-message shuffled schemes in Section V.

II. Preliminaries



(a) Local differential privacy (LDP) model of n clients.



(b) An L-message shuffled (MMS) model of n clients

In this section, we state some preliminary definitions that we use throughout the paper and also state some results from literature. We start by defining different privacy notions.

Definition 1 (Local Differential Privacy - LDP [22])**.** For $\varepsilon_0 \geq 0$, a randomized mechanism $\mathcal{R} : \mathcal{X} \to \mathcal{Y}$ is said to be ε_0-local differentially private (in short, ε_0-LDP), if for every pair of inputs $d, d' \in \mathcal{X}$, we have

$$\Pr[\mathcal{R}(d) \in \mathcal{S}] \leq e^{\varepsilon_0} \Pr[\mathcal{R}(d') \in \mathcal{S}], \qquad \forall \mathcal{S} \subset \mathcal{Y}. \tag{1}$$

Let $\mathcal{D} = \{d_1, \ldots, d_n\}$ denote a dataset comprising n points from $\mathcal{X}$. We say that two datasets $\mathcal{D} = \{d_1, \ldots, d_n\}$ and $\mathcal{D}' = \{d'_1, \ldots, d'_n\}$ are neighboring (and denoted by $\mathcal{D} \sim \mathcal{D}'$) if they differ in one data point, i.e., there exists an $i \in [n]$ such that $d_j = d'_j$ for all $j \neq i$.

Definition 2 (Central Differential Privacy - DP [47], [48])**.** For $\varepsilon, \delta \geq 0$, a randomized mechanism $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ is said to be (ε, δ)-differentially private (in short, (ε, δ)-DP), if for all neighboring datasets $\mathcal{D} \sim \mathcal{D}' \in \mathcal{X}^n$ and every subset $\mathcal{S} \subseteq \mathcal{Y}$, we have

$$\Pr\left[\mathcal{M}(\mathcal{D}) \in \mathcal{S}\right] \leq e^{\varepsilon} \Pr\left[\mathcal{M}(\mathcal{D}') \in \mathcal{S}\right] + \delta. \tag{2}$$

Definition 3 ($(\alpha, \varepsilon(\alpha))$-RDP (Rényi Differential Privacy) [49])**.** A randomized mechanism $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ is said to have $\varepsilon(\alpha)$-Rényi differential privacy of order $\alpha \in (1, \infty)$ (in short, $(\alpha, \varepsilon(\alpha))$-RDP), if for any neighboring datasets $\mathcal{D} \sim \mathcal{D}' \in \mathcal{X}^n$, we have that $D_\alpha(\mathcal{M}(\mathcal{D}) || \mathcal{M}(\mathcal{D}')) \leq \varepsilon(\alpha)$, where $D_\alpha(P||Q)$ denotes the Rényi divergence between two distributions P and Q defined by:

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \left(\mathbb{E}_{\theta \sim Q} \left[\left(\frac{P(\theta)}{Q(\theta)} \right)^\alpha \right] \right), \tag{3}$$

The RDP provides a tight privacy accounting of interactive mechanisms. The following results state the composition of RDP mechanisms and the conversion from RDP to approximate DP.

Lemma 1 (Adaptive composition of RDP [49])**.** *For any $\alpha > 1$, let $\mathcal{M}_1 : \mathcal{X} \to \mathcal{Y}_1$ be a $(\alpha, \varepsilon_1(\alpha))$-RDP mechanism and $\mathcal{M}_2 : \mathcal{Y}_1 \times \mathcal{X} \to \mathcal{Y}$ be a $(\alpha, \varepsilon_2(\alpha))$-RDP mechanism. Then, the mechanism defined by $(\mathcal{M}_1, \mathcal{M}_2)$ satisfies $(\alpha, \varepsilon_1(\alpha) + \varepsilon_2(\alpha))$-RDP.*

Lemma 2 (From RDP to DP [50], [51])**.** *Suppose for any $\alpha > 1$, a mechanism $\mathcal{M}$ is $(\alpha, \varepsilon(\alpha))$-RDP. For any $\delta > 0$, the mechanism $\mathcal{M}$ is $(\varepsilon_\delta, \delta)$-DP, where ε_δ is given by:*

$$\varepsilon_\delta = \min_\alpha \varepsilon(\alpha) + \frac{\log(1/\delta)}{\alpha - 1} + \log(1 - 1/\alpha)$$

In our algorithms, we use an unbiased version of the classical binary randomized response (*2RR*) [52] whose input is a bit $b \in \{0, 1\}$ and the output is $\frac{b-p}{1-2p}$ w.p. $1 - p$ and $\frac{1-b-p}{1-2p}$ w.p. p, where $p \in [0, 1/2)$ controls the privacy-utility trade-off (see Algorithm 7 in Section X).

Theorem 1. *For any $p \in [0, 1/2)$, the 2RR is ε_0-LDP, where $\varepsilon_0 = \log\left(\frac{1-p}{p}\right)$. The output y of the 2RR mechanism is an unbiased estimate of b with bounded MSE:*

$$\mathsf{MSE}^{2RR} = \sup_{b \in \{0,1\}} \mathbb{E}\left[\|b - y\|_2^2\right] = \frac{p(1-p)}{(1-2p)^2}. \tag{4}$$

Theorem 1 gives an upper bound on the mean square error (MSE) of the *2RR* mechanism. For completeness, we present its proof in Section X.

III. Problem Formulation

We study federated learning (FL) framework, where a set of n clients are connected to an untrusted server to solve the empirical risk minimization (ERM) problem:

$$\min_\theta F(\theta) = \frac{1}{n} \sum_{i=1}^{n} F_i(\theta), \tag{5}$$

where $\theta \in \mathbb{R}^d$ denotes the global model. $F_i(\theta) = \mathbb{E}_{d_i \sim \mathcal{D}_i}\left[F_i(\theta, d_i)\right]$ denotes the loss function of the i-th client, where $\mathcal{D}_i$ is the local dataset of the ith client. Our goal is to construct communication-efficient and private FL algorithm via stochastic gradient descent (SGD). At each round, the server updates the global model by aggregating the local updates. Therefore, at each round, the server applies distributed mean estimation (DME) of the local model updates. To isolate this problem we define DME under privacy and communication constraints.

a) Distributed Mean Estimation (DME):: Suppose we have a set of n clients. Each client has a d dimensional vector $\mathbf{x}_i \in \mathcal{X}$ for $i \in [n]$, where $\mathcal{X} \subset \mathbb{R}^d$ denotes a bounded subset of all possible inputs. For example, $\mathcal{X} \triangleq \mathbb{B}_2^d(r_2)$ denotes the d dimensional ball with radius r_2, i.e., each vector $\mathbf{x}_i$ satisfies $\|\mathbf{x}_i\|_2 \leq r_2$ for $i \in [n]$. Furthermore, each client has a communication budget of b-bits. The clients are connected to an (untrusted) server that wants to estimate the mean $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$. we consider two distributed privacy models, where the server is untrusted:(i) Local DP (LDP) model (ii) Multi-message shuffled (MMS) model.

LDP-model: In the LDP model, we design two mechanisms as depicted in Figure 1a: (i) Client-side mechanism $\mathcal{R} : \mathcal{X} \to \mathcal{Y}$

and (ii) Server aggregator $\mathcal{A}: \mathcal{Y}^n \to \mathbb{R}^d$. The local randomizer $\mathcal{R}$ takes an input $\mathbf{x}_i \in \mathcal{X}$ and generates a randomized output $\mathbf{y}_i \in \mathcal{Y}$. In LDP model, the local randomizer $\mathcal{R}$ satisfies privacy and communication constraints as follows. The output $\mathbf{y}_i = \mathcal{R}(\mathbf{x}_i)$ can be represented using only b-bits, as well as, it satisfies ε_0-LDP. Each client sends the output $\mathbf{y}_i$ directly to the server, which applies the aggregator $\mathcal{A}$ to estimate the mean $\hat{\mathbf{x}} = \mathcal{A}(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ such that the estimated mean $\hat{\mathbf{x}}$ is an unbiased estimate of the true mean $\overline{\mathbf{x}}$.

MMS-model: The multi-message shuffled model is similar to the LDP model but with secure shufflers which anonymize the clients' identities to the server. Precisely, the L-message shuffled model consists of three parameters $(\mathcal{R}, \mathcal{S}, \mathcal{A})$ as depicted in Figure 1b: (i) *Encode:* a local randomizer $\mathcal{R}: \mathcal{X} \to \mathcal{Y}^L$, where the output $\mathbf{y}_i = \mathcal{R}(\mathbf{x}_i) = (\mathbf{y}_i^{(1)}, \ldots, \mathbf{y}_i^{(L)})$ consists of L messages. The local randomizer satisfies communication constraints in which the output $\mathbf{y}_i$ can be represented using b communication bits. (ii) *Shuffle:* a single shuffler $\mathcal{S}^{(k)}: \mathcal{Y}^n \to \mathcal{Y}^n$, for $k \in [L]$, generates a random permutation of the received n reports: $\mathbf{y}^{(k)} = \mathcal{S}^{(k)}\left(y_1^{(k)}, \ldots, y_n^{(k)}\right)$, where the kth message of each client is sent to the kth shuffler. (iii) *Analyze:* a server aggregator $\mathcal{A}: \left(\mathcal{Y}^L\right)^n \to \mathbb{R}^d$ is applied to the received messages from the L shufflers to estimate the mean $\hat{\mathbf{x}} = \mathcal{A}\left(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L)}\right)$. We say that the shuffled model is (ε, δ)-DP if the view of the output of the shufflers $\{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L)}\}$ satisfies (ε, δ)-DP.

Remark 1 (parallel shufflers vs single shuffler)**.** Observe that we describe the multi-message shuffled model using L independent shufflers, where each shuffler receives a single message from each client. We can also represent the multi-message shuffled model with a single shuffler that receives the total nL messages from all clients by indexing the messages of each client with a slight increase of the communication cost, see [29, Sec. 2.4] for more details.

In the two privacy models, the performance of the estimator $\hat{\mathbf{x}}$ is measured by the expected mean squared error (MSE):

$$\mathsf{MSE} = \sup_{\{\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}\}} \mathbb{E}\left[\|\hat{\mathbf{x}} - \overline{\mathbf{x}}\|_2^2\right], \tag{6}$$

where the expectation is taken over the randomness of the private mechanisms. Hence, our goal is to design communication-efficient and private schemes to generate an unbiased estimate of the true mean $\overline{x}$ while minimizing the expected loss (6). We propose a local mechanism $\mathcal{R}$ and a server aggregator $\mathcal{A}$ showing that these mechanisms achieve simultaneously optimal privacy and communication efficiency in both privacy models (LDP and MMS models). We start by studying the DME problem of binary vectors, where $\mathcal{X} \triangleq \{0,1\}^d$. Next, we study the DME for bounded ℓ_∞-norm, *i.e.,* $\|\mathbf{x}_i\|_\infty \leq r_\infty$ for all $i \in [n]$, and for bounded ℓ_2-norm vectors where $\|\mathbf{x}_i\|_2 \leq r_2$.

IV. Overview and Main Theoretical Results

In this section we give an overview of our algorithmic solution for private DME and the theoretical guarantees of our proposed algorithms. We consider the private DME of binary vectors in Section IV-A, bounded ℓ_∞-norm vectors in Section IV-B, and bounded ℓ_2-norm vectors in Section IV-C. We will use these results to provide the guarantees for solving the trade-off for the ERM problem in Theorem 9 in Section IV-D.

Algorithm 1 : Local Randomizer $\mathcal{R}_{v,s}^{\text{Bin}}$

Public parameter: Privacy parameter v, and communication budget s.
Input: $\mathbf{b}_i \in \{0,1\}^d$.
If $\frac{d}{s}$ is not integer, add $(s\lceil \frac{d}{s} \rceil - d)$ dummy zeros to the binary vector $\mathbf{b}$. Let $a \leftarrow \frac{d}{s}$.
$p \leftarrow \frac{1}{2}\left(1 - \sqrt{\frac{v^2/s^2}{v^2/s^2+4}}\right)$
for $j \in [s]$ **do**
    Choose uniformly at random one coordinate $a_{ij} \leftarrow \mathsf{Unif}\left(\{(j-1)a+1, \ldots, ja\}\right)$.
    $y_{ij} \leftarrow a\mathcal{R}_p^{2RR}\left(\mathbf{b}_i[a_{ij}]\right)$
Return: The client sends s messages $\mathcal{Y}_i \leftarrow \{(a_{i1}, y_{i1}), \ldots, (a_{is}, y_{is})\}$.

A. Binary Vectors

In this section, we consider binary vectors: $\mathbf{b}_i \in \{0,1\}^d$. The server wants to estimate the mean $\overline{\mathbf{b}} = \frac{1}{n}\sum_{i=1}^n \mathbf{b}_i$. The binary vector mechanism is the main building block of the next algorithms. This problem is a generalization to the scalar binary summation problem studied in [26]. A straightforward solution is to apply the scalar mechanism in [26] per coordinate that requires d bits per client. Our private mechanisms require $\mathcal{O}\left(\min\{\varepsilon_0, d\}\right)$ and $\mathcal{O}\left(\min\{n\min\{\varepsilon^2, \varepsilon\}, d\}\right)$ communication bits per client in the LDP and shuffled models, respectively.

The client-side mechanism is presented in Algorithm 1, where the parameter s determines the communication budget per client and the parameter v determines the total privacy budget (see Theorem 2). For given $s \in \{1, \ldots, d\}$, each client splits the binary vector $\mathbf{b}_i$ into s sub-vectors, each with dimension $a = \lceil \frac{d}{s} \rceil$. Then, the client chooses uniformly at random one coordinate from each sub-vector and privatizes its bit using the binary randomized response (*2RR*) Algorithm 7 in Section X. Observe that the output of Algorithm 1 can be represented as a sparse d-dimensional vector with only s non-zero coordinates.

When $s = d$, each client applies the *2RR* mechanism on each coordinate separately. On the other hand, when $s = 1$, each client chooses uniformly at random one coordinate and applies the *2RR* mechanism. Thus, we get trade-offs between privacy-communication and accuracy. The server aggregator $\mathcal{A}^{\text{Bin}}$ is simply aggregating the received randomized bits. For completeness, we present the aggregator $\mathcal{A}^{\text{Bin}}$ in Algorithm 4 in Section VII.

Below, we state the bound on the MSE of the proposed mechanisms in the LDP and shuffled models. The proofs are presented in Section VII. Furthermore, we present RDP guarantees of our mechanisms for both LDP and shuffled models in the detailed proofs in Section VII.

Theorem 2 (LDP model)**.** *The output of the local mechanism $\mathcal{R}_{v,s}^{Bin}$ can be represented using $s\left(\log\left(\lceil d/s \rceil\right) + 1\right)$-bits. By*

choosing $v = \varepsilon_0$, the mechanism $\mathcal{R}^{Bin}_{v,s}$ satisfies ε_0-LDP. Let $\hat{\mathbf{b}}$ be the output of the analyzer $\mathcal{A}^{Bin}$. The estimator $\hat{\mathbf{b}}$ is an unbiased estimate of $\overline{\mathbf{b}}$ with MSE:

$$\mathsf{MSE}^{Bin}_{ldp} = \mathcal{O}\left(\frac{d^2}{n}\max\left\{\frac{1}{s}, \frac{s}{\varepsilon_0^2}\right\}\right). \tag{7}$$

Now, we move to the shuffled model, where we assume there exists a secure shuffler that randomly permutes the set of messages $\{\mathcal{Y}_i : i \in [n]\}$ from the n clients.

Theorem 3 (MMS model)**.** *The output of the local mechanism $\mathcal{R}^{Bin}_{v,s}$ can be represented using $s\left(\log\left(\lceil d/s\rceil\right)+1\right)$ bits. For every $n \in \mathbb{N}$, $\varepsilon \leq s$, and $\delta \in (0, 1/e)$, shuffling the outputs of n mechanisms $\mathcal{R}^{Bin}_{v,s}$ satisfies (ε,δ)-DP by choosing $v^2 = \frac{sn\min\{\varepsilon^2,\varepsilon\}}{c\log(1/\delta)}$, where $c = 2,304$ is constant. Let $\hat{\mathbf{b}}$ be the output of the analyzer $\mathcal{A}^{Bin}$. The estimator $\hat{\mathbf{b}}$ is an unbiased estimate of $\overline{\mathbf{b}}$ with MSE:*

$$\mathsf{MSE}^{Bin}_{shuffle} = \mathcal{O}\left(\frac{d^2}{n^2}\max\left\{n\left(\frac{1}{s}-\frac{1}{d}\right), \frac{\log(1/\delta)}{\min\{\varepsilon^2,\varepsilon\}}\right\}\right). \tag{8}$$

Observe that the MSE in (7) and (8) consists of two terms. The first term represents the communication cost for sending s coordinates out of d coordinates. The second term represents the cost of privacy to randomize the randomly chosen s coordinates. Theorem 2 shows that each client has to send $s = \min\{\lceil\varepsilon_0\rceil, d\}$ communication bits to achieve MSE $\mathcal{O}\left(\frac{d^2}{n\min\{\varepsilon_0,\varepsilon_0^2\}}\right)$ in the LDP model. Similarly, Theorem 3 shows that each client has to send $s = \mathcal{O}\left(\min\{n\{\varepsilon^2,\varepsilon\}, d\}\right)$ communication bits to achieve MSE $\mathcal{O}\left(\frac{d^2}{n^2\{\varepsilon^2,\varepsilon\}}\right)$ that matches the MSE of central DP mechanisms. For the scalar case when $d = 1$, our results in Theorem 3 match the optimal MSE as in [26].

B. Bounded ℓ_∞-norm vectors

In this Section, we consider the DME problem for bounded ℓ_∞-norm vectors, where $\|\mathbf{x}_i\|_\infty \leq r_\infty$ for $i \in [n]$. For ease of operation, we will scale each vector such that each coordinate becomes bounded in range $[0, 1]$, and then re-scale it at the server-side. Let $\mathbf{z}_i = \frac{\mathbf{x}_i + r_\infty}{2r_\infty}$, where the operations are done coordinate-wise. Thus, we have that $\mathbf{z}_i[j] \in [0, 1]$ for all $j \in [d]$ and $i \in [n]$, where $\mathbf{z}_i[j]$ denotes the jth coordinate of the vector $\mathbf{z}_i$. Observe that the vector $\mathbf{z}_i$ can be decomposed into a weighted summation of binary vectors $\mathbf{b}_i^{(k)} \in \{0, 1\}^d, \forall k \geq 1$ as follows:

$$\mathbf{z}_i = \sum_{k=1}^{\infty} \mathbf{b}_i^{(k)} 2^{-k}, \tag{9}$$

where $\mathbf{b}_i^{(k)} = \lfloor 2^k\left(\mathbf{z}_i - \mathbf{z}_i^{(k-1)}\right)\rfloor, k \geq 1$ such that $\mathbf{z}_i^{(0)} = \mathbf{0}$ and $\mathbf{z}_i^{(k)} = \sum_{l=1}^{k} \mathbf{b}_i^{(l)} 2^{-l}$. To make our mechanism communication efficient, each client approximates the vector $\mathbf{z}_i$ by using the first m binary vectors $\{\mathbf{b}_i^{(k)} : 1 \leq k \leq m\}$. Note that the first m binary vectors together give an approximation to the real vector $\mathbf{z}_i$ with error $\|\mathbf{z}_i - \mathbf{z}_i^{(m)}\|_2^2 \leq d/4^m$, where $\mathbf{z}_i^{(m)} = \sum_{k=1}^{m} \mathbf{b}_i^{(k)} 2^{-k}$. However, this mechanism creates a biased estimate of $\mathbf{z}_i$. Hence, to design an unbiased mechanism,

Algorithm 2 : Local Randomizer $\mathcal{R}^{\ell_\infty}_{v,m,s}$

Public parameter: Privacy budget v, communication levels m, and coordinate sampling per level s.
Input: $\mathbf{x}_i \in \mathbb{B}^d_\infty(r_\infty)$.
$\mathbf{z}_i \leftarrow (\mathbf{x}_i + r_\infty)/2r_\infty$
$\mathbf{z}_i^{(0)} \leftarrow 0$
for $k = 1, \ldots, m-1$ **do**
  $\mathbf{b}_i^{(k)} \leftarrow \lfloor 2^k(\mathbf{z}_i - \mathbf{z}_i^{(k-1)})\rfloor$
  $v_k \leftarrow \frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)} v$
  $\mathcal{Y}_i^{(k)} \leftarrow \mathcal{R}^{\text{Bin}}_{v_k,s}(\mathbf{b}_i^{(k)})$
  $\mathbf{z}_i^{(k)} \leftarrow \mathbf{z}_i^{(k-1)} + \mathbf{b}_i^{(k)} 2^{-k}$
Sample $\mathbf{u}_i \leftarrow \text{Bern}\left(2^{m-1}\left(\mathbf{z}_i - \mathbf{z}_i^{(m-1)}\right)\right)$
$v_m \leftarrow \frac{4^{\frac{-m+1}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)} v$
$\mathcal{Y}_i^{(m)} \leftarrow \mathcal{R}^{\text{Bin}}_{v_m,s}(\mathbf{u}_i)$
Return: The client sends $\mathcal{Y}_i \leftarrow \left\{\mathcal{Y}_i^{(1)}, \ldots, \mathcal{Y}_i^{(m)}\right\}$.

the client approximates the vector $\mathbf{z}_i$ using the first $m-1$ binary vectors $\{\mathbf{b}_i^{(k)} : 1 \leq k \leq m-1\}$ of the binary representation above and the last binary vector $(\mathbf{u}_i)$ is reserved for unbiasedness as follows:

$$\mathbf{u}_i[j] = \text{Bern}\left(2^{m-1}(\mathbf{z}_i[j] - \mathbf{z}_i^{(m-1)}[j])\right), \tag{10}$$

where $\mathbf{z}_i^{(m-1)} = \sum_{k=1}^{m-1} \mathbf{b}_i^{(k)} 2^{-k}$ and $\text{Bern}(p)$ denotes Bernoulli random variable with bias p. For completeness, we prove some properties of this quantization scheme in Section VI. Then, we estimate the mean of binary vectors $\{\mathbf{b}_i^{(k)} \in \{0, 1\}^d : i \in [n]\}$ using Algorithm 1 with different privacy guarantees for each level $k \in [m]$, where we allocate lower privacy (higher privacy parameter v_k) for the most significant bits (MSBs) (lower k) in order to get better performance in terms of the MSE.

The private DME mechanism is given in Algorithm 2, where v controls the total privacy of the mechanism. There are two communication parameters: m controls the number of levels for quantization and s controls the number of dimensions used to represent each binary vector. In Theorems 4 and 5, we present how the privacy and communication parameters v, m, s affects the accuracy of the mechanism. The server aggregator $\mathcal{A}^{\ell_\infty}$ is presented in Algorithm 5 in Section VIII, where the server first estimates the mean of each binary vectors $\{b_i^{(k)} : i \in [n]\}$ for $k \in [m-1]$ and decodes the messages to generate an estimate of the true mean $\overline{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i$. Then, the server scales the vector $\overline{\mathbf{z}}$ to generate an unbiased estimate of the mean $\overline{\mathbf{x}}$. We prove the bound on the MSE of our proposed mechanism for the LDP and MMS models in the following theorems. We defer the proofs to Section VIII.

Theorem 4 (Local DP model)**.** *The output of the local mechanism $\mathcal{R}^{\ell_\infty}_{v,m,s}$ can be represented using $ms\left(\log\left(\lceil d/s\rceil\right)+1\right)$ bits. By choosing $v = \varepsilon_0$, the mechanism $\mathcal{R}^{\ell_\infty}_{v,m,s}$ satisfies ε_0-LDP. Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_\infty}$. The estimator*

$\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ with bounded MSE:

$$\begin{aligned}\mathsf{MSE}_{LDP}^{\ell_\infty} &= \sup_{\{\mathbf{x}_i \in \mathbb{B}_\infty^d(r_\infty)\}} \mathbb{E}\left[\|\hat{\mathbf{x}} - \overline{\mathbf{x}}\|_2^2\right] \\ &= \mathcal{O}\left(\frac{r_\infty^2 d^2}{n}\max\left\{\frac{1}{d4^m}, \frac{1}{s}, \frac{s}{\varepsilon_0^2}\right\}\right).\end{aligned} \tag{11}$$

Theorem 5 (MMS model)**.** *The output of the local mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ can be represented using $ms\left(\log\left(\lceil d/s\rceil\right)+1\right)$ bits. For every $n \in \mathbb{N}$, $\varepsilon \le ms$, and $\delta \in (0, 1/e)$, shuffling the outputs of n mechanisms $\mathcal{R}_{v,m,s}^{\ell_\infty}$ satisfies (ε,δ)-DP by choosing $v^2 = \frac{sn\min\{\varepsilon^2,\varepsilon\}}{c\log(1/\delta)}$, where $c = 2,304$ is constant. Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_\infty}$. The estimator $\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ with bounded MSE:*

$$\begin{aligned}\mathsf{MSE}_{MMS}^{\ell_\infty} &= \sup_{\{\mathbf{x}_i \in \mathbb{B}_\infty^d(r_\infty)\}} \mathbb{E}\left[\|\hat{\mathbf{x}} - \overline{\mathbf{x}}\|_2^2\right] \\ &= \mathcal{O}\left(\frac{r_\infty^2 d^2}{n^2}\max\left\{\frac{n}{d4^m}, n\left(\frac{1}{s}-\frac{1}{d}\right), \frac{\log(1/\delta)}{\min\{\varepsilon^2,\varepsilon\}}\right\}\right).\end{aligned} \tag{12}$$

Observe that the MSE in (11) and (12) consists of three terms. The first term is the communication cost of quantizing the real vector $\mathbf{z}_i$ using m binary vectors. The second term represents the communication cost of sending s out of d coordinates from each binary vector. The third term is the privacy cost to randomize the binary vectors. Theorem 4 shows that each client has to set $m = 1$ and $s = \lceil\varepsilon_0\rceil$ of total $\mathcal{O}\left(\lceil\varepsilon_0\rceil\right)$ communication bits to achieve MSE $\mathcal{O}\left(\frac{d^2}{n\min\{\varepsilon_0,\varepsilon_0^2\}}\right)$ when $\varepsilon_0 \le d$. Similarly, by setting $m = \max\{1, \lceil\log\left(n\min\{\varepsilon^2,\varepsilon\}/d\right)\rceil\}$ and $s = \mathcal{O}\left(\min\{n\{\varepsilon^2,\varepsilon\}, d\}\right)$ in Theorem 5, the MSE is bounded by $\mathcal{O}\left(\frac{d^2}{n^2\min\{\varepsilon^2,\varepsilon\}}\right)$, which matches the MSE of central differential privacy mechanisms with total communication cost of $\mathcal{O}\left(d\log\left(\frac{n\min\{\varepsilon^2,\varepsilon\}}{d}\right)\right)$ when $d \le n\min\{\varepsilon^2,\varepsilon\}$ and $\mathcal{O}\left(n\{\varepsilon^2,\varepsilon\}\log\left(\frac{d}{n\{\varepsilon^2,\varepsilon\}}\right)\right)$ when $d > n\{\varepsilon^2,\varepsilon\}$.

Remark 2 (Scalar case)**.** When $d = 1$, i.e., scalar case, our MMS algorithm achieves the central DP error $\mathcal{O}\left(\frac{1}{n^2\min\{\varepsilon^2,\varepsilon\}}\right)$ using $m = \max\{1, \lceil\log\left(n\min\{\varepsilon^2,\varepsilon\}\right)\rceil\}$ bits per client. This result covers the private-communication trade-offs for all privacy regimes. For example, for $\varepsilon = \frac{1}{\sqrt{n}}$, each client needs only a single bit to achieve the central DP error. On the other hand, the multi-message shuffled mechanism based on IKOS protocol [38] proposed in [29], [37] requires $\mathcal{O}\left(\log\left(n\right)\right)$-bits of communication for all privacy regimes, where it doesn't provide any guarantees for any small communication cost [29, Sec. 1.2]. Even when particular regimes of order-optimality are achieved for the MMS, the communication bound is in expectation [53], whereas ours is deterministic.

Remark 3 (Scalar summation with sampling/sketching)**.** Observe that when $d < n\min\{\varepsilon^2,\varepsilon\}$, it is not possible to combine the scalar summation scheme [29], [37] with coordinate sampling due to the following. When each client independently chooses a set of s coordinates, we might loose the amplification gain from shuffling, as not all the n clients will choose the same set of s coordinates. When choosing the same s coordinates for all clients, the MSE is bounded below by $\Omega\left(r_\infty^2(d-s)\right)$. Thus, the scalar summation in MMS cannot be directly combined with coordinate sampling.

Algorithm 3 : Local Randomizer $\mathcal{R}_{v,m,s}^{\ell_2}$

Public parameter: Privacy budget v, communication levels m, coordinate sampling per level s, and confidence term β.
Input: $\mathbf{x}_i \in \mathbb{B}_2^d(r_2)$.
Let $U = \frac{1}{\sqrt{d}}\mathbf{H}D$, where $\mathbf{H}$ denotes a Hadamard matrix and D is a diagonal matrix with i.i.d. uniformly random $\{\pm 1\}$ entries.
$\mathbf{w}_i \leftarrow W\mathbf{x}_i$
$r_\infty \leftarrow 10r_2\sqrt{\frac{\log(dn/\beta)}{d}}$
for $j = 1, \ldots, d$ **do**
    $\mathbf{w}_i[j] = \min\left\{r_\infty, \max\left\{\mathbf{w}_i(j), -r_\infty\right\}\right\}$
$\mathcal{Y}_i \leftarrow \mathcal{R}_{v,m,s}^{\ell_\infty}(\mathbf{w}_i)$
Return: The client sends $\mathcal{Y}_i$.

Remark 4. After the completion of our work [39], we were directed to [54], where the idea of unequal privacy allocation had also been proposed to obtain pure DP for scalar summation in the shuffled model. There are some differences between our proposed scheme and the scheme in [54]. First, the focus in [54] was for pure privacy for scalar problems that requires a higher communication budget, whereas our overall scheme is for vector MMS problem. There is also a small difference in the privacy allocation strategies. In our proposed scheme, we assign privacy $\varepsilon^{(k)} \approx 4^{-k/3}\varepsilon$ for the k-th bit, while $\varepsilon^{(k)} \approx \max\{0.9^k, \varepsilon/m\}$ in [54]. The reason behind our privacy allocation $\varepsilon^{(k)} \approx 4^{-k/3}\varepsilon$ is that the k-th bit has a weight 2^{-k} in the real summation, and hence, the error in estimating this bit contributes with 4^{-k} term in the MSE. To the best of our knowledge, the unequal privacy allocation scheme in [54] is analyzed for pure-DP scalar summation in the shuffled model. On the other hand, our proposed scheme achieves order optimal MSE for (ε,δ)-DP scalar summation in the shuffled model. In addition, we show that our scheme can be exploited to achieve order optimal MSE in the LDP model and the MMS model for vector summation. The simple combination of coordinate sampling and the MMS scheme for scalar summation cannot achieve the optimal MSE (see Remark 3 for more details.). Our proposed scheme achieves order optimal MSE and saves communication cost by carefully applying coordinate sampling in the shuffled model.

C. Bounded ℓ_2-norm Vectors

In this section, we consider the DME problem for bounded ℓ_2-norm vectors, where $\|\mathbf{x}_i\|_2 \le r_2$ for $i \in [n]$. We first use the random rotation proposed in [19] to bound the ℓ_∞-norm of the vector with radius $r_\infty = \mathcal{O}\left(\frac{r_2}{\sqrt{d}}\right)$. Then, we apply the bounded ℓ_∞-norm algorithm in Section IV-B. The client-side scheme is presented in Algorithm 3 and the server-side scheme is presented in Algorithm 6 in Section IX.

Theorem 6 (LDP model)**.** *The output of the local mechanism $\mathcal{R}_{v,m,s}^{\ell_2}$ can be represented using $ms\left(\log\left(\lceil d/s\rceil\right)+1\right)$ bits. By choosing $v = \varepsilon_0$, the mechanism $\mathcal{R}_{v,m,s}^{\ell_2}$ satisfies ε_0-LDP.*

	MMS model (this work)	MMS (Cheu *et al.* [30])	MMS (Chang *et al.* [31])	SecAgg([18], [32])
$d < n\varepsilon^2$	$\mathcal{O}\left(d\log\left(\frac{n\varepsilon^2}{d}\right)\right)$	$\mathcal{O}\left(d\sqrt{n}\right)$	$\mathcal{O}\left(d\log(n)\right)$	$\mathcal{O}\left(d\log(n)\right)$
$n\varepsilon^2 < d < n^2\varepsilon^2$	$\mathcal{O}\left(n\varepsilon^2\log\left(\frac{d}{n\varepsilon^2}\right)\right)$	$\mathcal{O}\left(d\sqrt{n}\right)$	$\mathcal{O}\left(d\log(n)\right)$	$\mathcal{O}\left(d\log(n)\right)$
$d > n^2\varepsilon^2$	$\mathcal{O}\left(n\varepsilon^2\log\left(\frac{d}{n\varepsilon^2}\right)\right)$	$\mathcal{O}\left(d\sqrt{n}\right)$	$\mathcal{O}\left(d\log(n)\right)$	$\mathcal{O}\left(n^2\varepsilon^2\log(d)\right)$

TABLE I: Comparison on the communication cost of several schemes to design (ε,δ)-DP mechanism achieving MSE $\mathcal{O}\left(\frac{r_2^2 d}{n^2\varepsilon^2}\right)$ for $\varepsilon = \mathcal{O}(1)$.

Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_2}$. With probability at least $1-\beta$, the estimator $\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ with bounded MSE:

$$\mathsf{MSE}_{LDP}^{\ell_2} = \tilde{\mathcal{O}}\left(\frac{r_2^2 d}{n}\max\left\{\frac{1}{d4^m}, \frac{1}{s}, \frac{s}{\varepsilon_0^2}\right\}\right), \tag{13}$$

where $\tilde{\mathcal{O}}$ hides $\log(nd/\beta)$ factor.

Theorem 7 (MMS model)**.** *The output of the local mechanism $\mathcal{R}_{v,m,s}^{\ell_2}$ can be represented using $ms\left(\log\left(\lceil d/s\rceil\right)+1\right)$ bits. For every $n\in\mathbb{N}$, $\varepsilon \le ms$, and $\delta\in(0,1/e)$, the shuffling the outputs of n mechanisms $\mathcal{R}_{v,m,s}^{\ell_2}$ satisfies (ε,δ)-DP by choosing $v^2 = \frac{sn\min\{\varepsilon^2,\varepsilon\}}{c\log(1/\delta)}$, where $c = 2,304$ is constant. Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_2}$. With probability at least $1-\beta$, the estimator $\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ with bounded MSE:*

$$\mathsf{MSE}_{MMS}^{\ell_2} = \tilde{\mathcal{O}}\left(\frac{r_2^2 d}{n^2}\max\left\{\frac{n}{d4^m}, n\left(\frac{1}{s}-\frac{1}{d}\right), \frac{\log(1/\delta)}{\min\{\varepsilon^2,\varepsilon\}}\right\}\right), \tag{14}$$

where $\tilde{\mathcal{O}}$ hides $\log(nd\beta)$ factor.

Remark 5 (Kashin's represention)**.** Observe that the MSE in (13) and in (14) is achievable with probability $(1-\beta)$, and has a factor of $(\log(nd/\beta))$ due to the random rotation matrix. We can remove this factor by using the Kashin's representation [55] to transform the bounded ℓ_2-norm vector into a bounded ℓ_∞-norm vector with radius $r_\infty = \frac{cr_2}{\sqrt{d}}$, where c is constant (see e.g., [34], [56], [57]). However, Kashin's representation has large constants in practice [58].

Next we present a lower bound for the MSE of the DME under privacy and communication constraints.

Theorem 8 (Lower Bound For central DP model [[32]& [59]). *] Let $n,d\in\mathbb{N}$, $\varepsilon = \mathcal{O}(1)$, $r_2 \ge 1$, and $\delta = o(\frac{1}{n})$. For any $\mathbf{x}_1,\dots,\mathbf{x}_n \in \mathbb{B}_2^d(r_2)$, the MSE is bounded below by:*

$$\mathsf{MSE}_{central}^{\ell_2} = \Omega\left(\frac{r_2^2 d}{n^2}\max\left\{\frac{\log(1/\delta)}{\varepsilon^2}, \frac{n}{d4^{b/d}}\right\}\right) \tag{15}$$

for any unbiased algorithm $\mathcal{M}$ that is (ε,δ)-DP with $b > d$-bits of communication per client. Furthermore, when $b < d$ bits per client, the MSE is bounded below by:

$$\mathsf{MSE}_{central}^{\ell_2} = \Omega\left(\frac{r_2^2 d}{n^2}\max\left\{\frac{\log(1/\delta)}{\varepsilon^2}, \frac{n}{b}\right\}\right). \tag{16}$$

Remark 6. (Optimality of our mechanism) When the communication budget $b > d$, we can see that our MSE in Theorem 7 matches the lower bound in Theorem 8 (up to logarithmic factor) by choosing $s = d$ and $m = b/d$. Furthermore, when the communication budget $b < d$, our algorithm achieve the lower bound by choosing $s = b$ and $m = 1$. Thus, our algorithm for MMS is order optimal for all privacy-communication regimes.

Remark 7 (Comparison with SecAgg)**.** When $d < n\varepsilon^2$, our MMS algorithm requires $\mathcal{O}\left(d\log\left(\frac{n\varepsilon^2}{d}\right)\right)$ bits per client to achieve the central DP error $\mathcal{O}\left(\frac{d}{n^2\varepsilon^2}\right)$. Furthermore, it requires only $\mathcal{O}\left(n\varepsilon^2\log\left(\frac{d}{n\varepsilon^2}\right)\right)$-bits when $d > n\varepsilon^2$. In contrast, the DDG algorithm [18] needs $\mathcal{O}\left(d\log\left(n\right)\right)$-bits when $d < n^2\varepsilon^2$ and $\mathcal{O}\left(n^2\varepsilon^2\log\left(d\right)\right)$-bits when $d > n^2\varepsilon^2$ [32] to achieve the same order MSE. Hence, the MMS saves communication in comparison with SecAgg.

In Table I, we present comparison on the communication cost of several schemes in the literature to design (ε,δ)-DP mechanism and to achieve MSE $\mathcal{O}\left(\frac{r_2^2 d}{n^2\varepsilon^2}\right)$ that matches the optimal MSE of the central DP mechanisms. We can see that our proposed mechanism saves a significant amount of communication cost when $d > n\varepsilon^2$ comparing to the MMS schemes in [30], [31]. Furthermore, our MMS mechanism saves a gain of $\mathcal{O}\left(n\right)$ of communication cost comparing with the secure aggregation scheme [32] when $d > n\varepsilon^2$.

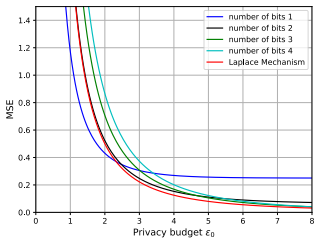
D. Application to Multi-Message Shuffled Federated Learning

In this section, we exploit our private mechanisms for DME to give convergence guarantees for DP-SGD algorithm in the multi-message shuffled model. We consider a standard SGD algorithm, where the server initialize the model by choosing $\theta^0 \in \mathcal{C}$. At the t-th round, the server sends the current model θ^t to the n clients. Each client computes the local gradient $\nabla F_i\left(\theta^t\right)$ for $i\in[n]$. Then, the client applies our private $\mathcal{R}_{v,m,s}^{\ell_2}$ mechanism before sending it to the shufflers. The sever receives the shuffled messages and aggregates the private gradients and updates the model as follows:

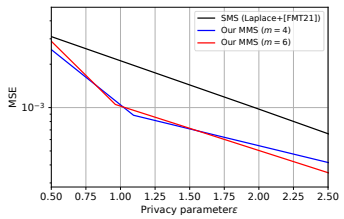
$$\theta^{t+1} = \theta^t - \eta g_t, \tag{17}$$

where $g_t = \mathcal{A}^{\ell_2}\left(\{\mathcal{Y}_i : i\in[n]\}\right)$ denotes the private estimate of the true average gradients $h_t = \frac{1}{n}\sum_{i\in[n]}\nabla F_i\left(\theta^t\right)$. In the following theorem, we derive the convergence of the DP-SGD algorithm described above.

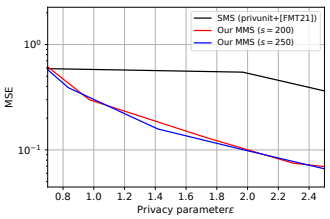
Theorem 9 (DP-SGD convergence in the MMS model)**.** *Let F be L-smooth and $\forall\theta\|\nabla F\left(\theta\right)\|_2 \le D$. Let θ^0 satisfies $F\left(\theta^0\right) - F\left(\theta^*\right) \le D_F$. Let $\mathcal{R}_{v,m,s}^{\ell_2}$ be our private-compression scheme and $\eta = \min\left\{L^{-1}, \sqrt{2D_F}\left(\sigma\sqrt{LT}\right)^{-1}\right\}$. By choosing $m = \max\{1, \log(\frac{n\min\{\varepsilon^2,\varepsilon\}}{Td})\}$, $s = \max\{1, \min\{d, \frac{n\min\{\varepsilon^2,\varepsilon\}}{T}\}\}$,*



(a) Comparison of our LDP mechanism $\mathcal{R}^{\ell\infty}_{v,m,s}$ with Laplace mechanism for $d = 1$, $n = 1$, and $m \in \{1, 2, 3, 4\}$.



(b) Comparison of our MMS mechanism $\mathcal{R}^{\ell\infty}_{v,m,s}$ with SMS (Laplace+[FMT21]) for $d = 1$, $n = 1000$, and $m \in \{4, 6\}$.



(c) Comparison of our MMS mechanism $\mathcal{R}^{\ell 2}_{v,m,s}$ with SMS (privunit+[FMT21]) for $d = 300$, $n = 1000$, and $s \in \{200, 250\}$.

and $v^2 = \frac{sn\min\{\varepsilon^2,\varepsilon\}}{cT\log(1/\delta)}$*, then after* T *rounds, the total algorithm is* (ε, δ)*-DP. Furthermore, we get:*

$$\mathbb{E}_{t\sim Unif(T)}\left[\nabla F\left(\theta^t\right)\right] \leq \mathcal{O}\left(L\sqrt{\frac{dD_F\log\left(1/\delta\right)}{n^2\varepsilon\min\{\varepsilon^2,\varepsilon\}}}\right), \quad (18)$$

for any $T > \sqrt{\frac{n\min\{\varepsilon^2,\varepsilon\}D_F}{d}}$ *with* $\mathcal{O}\left(ms\log\left(d/s\right)\right)$ *communication bits per client per round.*

The proof of Theorem 9 is presented in Appendix A. Note that in our DP-SGD algorithm, we assume that each client compute the full gradient $\nabla F_i\left(\theta^t\right)$ and then applies the private-compression mechanism $\mathcal{R}^{\ell 2}_{v,m,s}$. Hence, our privacy guarantees in Theorem 9 is user-level privacy, i.e., it satisfies (ε, δ)-DP by replacing the entire local dataset associated with certain client (not only a single datapoint in the item-level privacy). Also, we assume that all clients contribute in each round $t \in [T]$. We can extend the results in Theorem 9 to client sampling at each round by using the privacy amplification via sub-sampling [60] and shuffling [44].

In [61], Girgis *et al.* proposed a communication-efficient and private DP-SGD mechanism in the single-message shuffled model. However, the results in [61, Theorem 1] is order optimal for high privacy regime ($\varepsilon = \mathcal{O}(1)$). In the high privacy regime, the privacy parameter per round is of order $\varepsilon_t \approx \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, and hence, the single-message shuffled model is order optimal. Our result in Theorem 9 generalizes the result in [61] for all privacy regimes.

In [30], Cheu *et al.* studied the stochastic convex optimization in the multi-message shuffled model. However, their algorithm requires $\mathcal{O}\left(d\sqrt{n}\right)$ communication bits per client, while our algorithm requires much smaller amount of communication per client per round.

V. Numerical Results

In this section, we compare the performance of our proposed algorithm with the Laplace mechanism in the LDP model. Furthermore, we compare our algorithms for multi-message shuffled model with the best known algorithms for the single-message shuffled model for both scalar and vector summation.

Local DP model: We start by comparing the performance of our algorithm $\mathcal{R}^{\ell\infty}_{v,m,s}$ with the performance of the Laplace mechanism [32] in the local model for scalar case, i.e., $d = 1$, where the elements $\mathbf{x}_i \in [-r_\infty, r_\infty]$ and $r_\infty = 0.5$. Observe that the Laplace mechanism has infinite communication bits. In Figure 2a, we plot the MSE of our $\mathcal{R}^{\ell\infty}_{v,m,s}$ with different communication budget $s = 1$ and $m \in \{1, 2, 3, 4\}$ for a single client $n = 1$. We can observe that our mechanism achieves MSE closer to the MSE of the Laplace mechanism. Furthermore, we only need at most $m = 3$ bits to achieve similar performance as Laplace mechanism.

Shuffled model: We consider two cases in the shuffler model: 1) The scalar case when $d = 1$ to evaluate the performance of our $\mathcal{R}^{\ell\infty}_{v,m,s}$ mechanism in the multi-message shuffled model. 2) The vector case when $d = 1000$ to evaluate the performance of our $\mathcal{R}^{\ell 2}_{v,m,s}$ mechanism in the multi-message shuffled model.
Scalar: In Figure 2b, we plot the MSE of two different mechanisms versus the central privacy ε for fixed $\delta = 10^{-5}$. The first mechanism is single message shuffled (SMS) model obtained using Laplace mechanism with privacy amplification results in [62]. Observe that Laplace mechanism is the optimal LDP mechanism for $\varepsilon_0 = \mathcal{O}(1)$ and the privacy amplification results in [62] is approximately optimal for computing the (ε, δ)-DP of the shuffled model. Hence, we expect that this is the best that an SMS mechanism can achieve. The second mechanism is our multi-message shuffled (MMS) mechanism $\mathcal{R}^{\ell\infty}_{v,m,s}$ mechanism for $d = 1$ and $m \in \{4, 6\}$. Since we have MMS, we use the RDP results of privacy amplification by shuffling in [45] which is better for composition to compute the RDP of our mechanism. Then, we transform from RDP bound to approximate (ε, δ)-DP. We choose number of clients $n = 1000$. We can see that our multi-message shuffled algorithm achieve lower MSE than the single message shuffled especially for large value of central DP parameter ε.
Bounded ℓ_2-norm vectors Similar to the scalar case, we consider two mechanisms. The first mechanism SMS is obtained by using `privunit` mechanism with the privacy amplification results in [62], where `privunit` [63] is asymptotically optimal LDP mechanism [35]. We choose $n = 1000$ and $d = 300$. For our MMS $\mathcal{R}^{\ell 2}_{v,m,s}$, we choose $s \in \{200, 250\}$. It is clear from Figure 2c that our MMS mechanism has better performance compared to SMS mechanism.

VI. Properties of The Quantization Scheme

In this section, we prove some properties of the quantization scheme proposed in Section IV-B for vector $\mathbf{z}_i \in [0, 1]^d$. We first prove some properties for a scalar case when $x \in [0, 1]$,

and then, the results of the bounded ℓ_∞ will be obtained directly from repeating the scalar case on each coordinate.

Let $x \in [0,1]$ and $x^{(k)} = \sum_{l=1}^{s} b_l 2^{-l}$ for $k \geq 1$, where $x^{(0)} = 0$ and $b_k = \lfloor 2^k (x - x^{k-1}) \rfloor$. For given $m \geq 1$, we represent x using m bits as follows: $\tilde{x}^{(m)} = \sum_{k=1}^{m-1} b_k 2^{-k} + u 2^{-m+1}$, where $u = \text{Bern}\left(2^{m-1}(x - x^{(m-1)}[j])\right)$. This estimator needs only m communication bits.

Lemma 3. *For given $x \in [0,1]$, let $\tilde{x}^{(m)}$ be the quantization of x presented above. We have that $\tilde{x}^{(m)}$ is an unbiased estimate of x with bounded MSE:*

$$\mathsf{MSE}_{scalar}^{quan} = \sup_{x \in [0,1]} \mathbb{E}\left[\|x - \tilde{x}^{(m)}\|_2^2\right] \leq \frac{1}{4^m}, \tag{19}$$

where the expectation is taken over the randomness in the quantization scheme.

Proof. First, we show that $\tilde{x}^{(m)}$ is an unbiased estimate of x:

$$\begin{aligned} \mathbb{E}\left[\tilde{x}^m\right] &= \sum_{k=1}^{m-1} b_k 2^{-k} + \mathbb{E}\left[u\right] 2^{-m+1} \\ &\overset{(a)}{=} \sum_{k=1}^{m-1} b_k 2^{-k} + 2^{m-1}(x - x^{(m-1)}) 2^{-m+1} \\ &= x_i, \end{aligned} \tag{20}$$

where step (a) is obtained from the fact that u is a Bernoulli random variable with bias $p = 2^{m-1}(x - x^{(m-1)})$. We show that the estimator $\tilde{x}^{(m)}$ has a bounded MSE by 4^{-m}:

$$\begin{aligned} \mathsf{MSE}_{\text{scalar}}^{\text{quan}} &= \sup_{x \in [0,1]} \mathbb{E}\left[\|x - \tilde{x}^{(m)}\|_2^2\right] \\ &= \sup_{x \in [0,1]} \mathbb{E}\left[\|x - x^{(m-1)} - u 2^{-m+1}\|^2\right] \\ &= \sup_{x \in [0,1]} 4^{-(m-1)} \mathbb{E}\left[\|2^{-(m-1)}(x - x^{(m-1)}) - u\|^2\right] \\ &\overset{(a)}{\leq} \frac{1}{4^m}, \end{aligned} \tag{21}$$

where the inequality (a) is obtained from the fact that u is a Bernoulli random variable, and hence, it has a variance less that $1/4$. This completes the proof of Lemma 3. ■

Corollary 1. For given a vector $\mathbf{z}_i \in [0,1]^d$, let $\tilde{\mathbf{z}}_i^{(m)}$ be the quantization of $\mathbf{z}_i$ by applying the above scalar quantization scheme on each coordinate $\mathbf{z}_i[j]$ for $j \in [d]$. Then, $\tilde{\mathbf{z}}_i^{(m)}$ is an ubiased estimate of $\mathbf{z}_i$ with bounded MSE:

$$\mathsf{MSE}_{\text{vector}}^{\text{quan}} = \sup_{\mathbf{z}_i \in [0,1]^d} \mathbb{E}\left[\|\mathbf{z}_i - \tilde{\mathbf{z}}_i^{(m)}\|_2^2\right] \leq \frac{d}{4^m}, \tag{22}$$

where the expectation is taken over the randomness in the quantization scheme.

VII. Proofs of Theorem 2 and Theorem 3 (Binary vectors)

In this section, we prove Theorem 2 and Theorem 3 for estimating the mean of binary vectors in the LDP and MMS models, respectively.

Algorithm 4 : Analyzer $\mathcal{A}^{\text{Bin}}$

Inputs: $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, where $\mathcal{Y}_i$ consists of s messages of a pair (a_{ij}, y_{ij}) for $j \in [s]$ and $i \in [n]$.
$\hat{\mathbf{b}} \leftarrow \mathbf{0}_d$
for $i \in [n]$ **do**
  for $j \in [s]$ **do**
    $\hat{\mathbf{b}}[a_{ij}] \leftarrow \hat{\mathbf{b}}[a_{ij}] + y_{ij}$.
$\hat{\mathbf{b}} \leftarrow \frac{1}{n}\hat{\mathbf{b}}$
Return: The server returns $\hat{\mathbf{b}}$.

A. Communication Bound for Theorem 2 and Theorem 3

Observe that each client sends s messages; each message consists of a pair (a_{ij}, y_{ij}), where a_{ij} is drawn uniformly at random from $\lceil \frac{d}{s} \rceil$ values and y_{ij} is a binary element. Hence, each message requires $\log\left(\lceil \frac{d}{s} \rceil\right) + 1$ bits. As a result the total communication bits per client is given by $s\left(\log\left(\lceil \frac{d}{s} \rceil\right) + 1\right)$-bits.

B. Privacy of the LDP model in Theorem 2

In the mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$, each client sends s messages of the *2RR* mechanism $((a_{i1}, y_{i1}), \ldots, (a_{is}, y_{is}))$ with parameter $p = \frac{1}{2}\left(1 - \sqrt{\frac{\varepsilon_0^2/s^2}{\varepsilon_0^2/s^2 + 4}}\right)$. Hence, from Lemma 6, each message is $\frac{\varepsilon_0}{s}$-LDP. As a results, the total mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$ is ε_0-LDP from the composition of the DP mechanisms [48] when $v = \varepsilon_0$.

In addition, we can bound the RDP of the mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$ in the LDP model by using the composition of the RDP (see Lemma 1). From the proof of Theorem 1 in Section X, the *2RR* mechanism is $(\alpha, \varepsilon(\alpha))$-RDP, where $\varepsilon(\alpha)$ is bounded by:

$$\varepsilon(\alpha) = \frac{1}{\alpha - 1} \log\left(p^\alpha (1-p)^{1-\alpha} + p^{1-\alpha}(1-p)^\alpha\right), \tag{23}$$

In the mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$, each client sends s messages of the *2RR* mechanism. Hence, the mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$ is $(\alpha, s\varepsilon(\alpha))$-RDP, where $\varepsilon(\alpha)$ is given is (23).

C. Privacy of the MMS model in Theorem 3

In the mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$, each client sends s messages of the *2RR* mechanism $((a_{i1}, y_{i1}), \ldots, (a_{is}, y_{is}))$. For simplicity, assume that there exist s shufflers, where the kth shuffler randomly permutes the set of messages $\{(a_{ik}, y_{ik}) : i \in [n]\}$ from the n clients. Hence from composition of the RDP, it is sufficient to bound the RDP of shuffling n outputs of the *2RR* mechanism.

We use the recent results of privacy amplification by shuffling to analyze the RDP of the shuffled model, which states the following

Lemma 4. *[45] For any $n \in \mathbb{N}$, $\varepsilon_0 > 0$, and α such that $\alpha^4 e^{5\varepsilon_0} \leq \frac{n}{9}$, the output of shuffling n messages of an ε_0-LDP mechanism is $(\alpha, \varepsilon(\alpha))$-RDP, where $\varepsilon(\alpha)$ is bounded by:*

$$\begin{aligned} \varepsilon(\alpha) &\leq \frac{1}{\alpha - 1} \log\left(1 + \alpha(\alpha - 1)\frac{2\left(e^{\varepsilon_0} - 1\right)^2}{n}\right) \\ &\leq 2\alpha \frac{\left(e^{\varepsilon_0} - 1\right)^2}{n} \end{aligned} \tag{24}$$

Recently [46] improved the dependence on ε_0 of the result in [45] by showing the following.

Lemma 5. *[46, Corollary 7.2] For any $n \in \mathbb{N}$, $\varepsilon_0 > 0$, and $\alpha \leq \frac{n}{32\varepsilon_0 e^{\varepsilon_0}}$, the output of shuffling n messages of an ε_0-LDP mechanism is $(\alpha, \varepsilon(\alpha))$-RDP, where $\varepsilon(\alpha)$ is bounded by:*

$$\varepsilon(\alpha) \leq \alpha \frac{768\left(e^{\varepsilon_0}-1\right)^2}{ne^{\varepsilon_0}}. \tag{25}$$

The exact constants in Lemma 5 is obtained from [46, Appendix B].

From Theorem 1, each single message of the client is $\varepsilon_0 = \log\left(\frac{1-p}{p}\right)$-LDP. Hence, from Lemma 5, the output of a single shuffler is $(\alpha, \tilde{\varepsilon}(\alpha))$-RDP, where $\tilde{\varepsilon}(\alpha) \leq 768\alpha\frac{(1-2p)^2}{np(1-p)}$. Thus, from composition, the output of the s shufflers is $(\alpha, \varepsilon(\alpha))$-RDP, where $\varepsilon(\alpha)$ is bounded by:

$$\varepsilon(\alpha) \leq 768\alpha\frac{s(1-2p)^2}{np(1-p)}. \tag{26}$$

Observe that (26) gives a closed form bound on the RDP of the mechanism $\mathcal{R}^{\text{Bin}}_{v,s}$ in the shuffled model. However, we can numerically provide better bound on the RDP of the shuffle model using [45], [46]. By setting $p = \frac{1}{2}\left(1 - \sqrt{\frac{v^2/s^2}{v^2/s^2+4}}\right)$, we get that $\frac{(1-2p)^2}{p(1-p)} = v^2/s^2$, and hence, $\varepsilon(\alpha) \leq 768\alpha v^2/(sn)$. Now, we use Lemmas 7 in Section X to convert from RDP to approximate DP, where $\rho = 768v^2/(sn)$. For given $\delta > 0$, shuffling the outputs of n mechanisms $\mathcal{R}^{\text{Bin}}_{v,s}$ is (ε, δ)-DP, where ε is bounded by

$$\varepsilon \leq 3\max\left\{\frac{768v^2}{sn}\log(1/\delta), \sqrt{\frac{768v^2}{sn}\log(1/\delta)}\right\}. \tag{27}$$

By setting $v^2 = \frac{sn\min\{\varepsilon^2,\varepsilon\}}{c\log(1/\delta)}$, we can easily show that (27) is satisfied, where $c = 2,304$ is constant. Hence, the output of the shufflers is (ε, δ)-DP.

Observe that we choose $\alpha = 1 + \sqrt{\frac{\log(1/\delta)}{\rho}}$ when applying Lemma 7 to transform from RDP to approximate DP, where $\rho = 768\frac{(e^{\varepsilon_0}-1)^2}{ne^{\varepsilon_0}} = 768\frac{v^2}{s^2}$. On the other hand, the RDP in (26) is valid for $\alpha \leq \frac{n}{32\varepsilon_0 e^{\varepsilon_0}}$ from Lemma 5. Thus, we need $1 + \sqrt{\frac{\log(1/\delta)}{\rho}} \leq \frac{n}{32\varepsilon_0 e^{\varepsilon_0}}$ such that (27) is satisfied. Thus, we have

$$\begin{aligned}
1 + \sqrt{\frac{\log(1/\delta)}{\rho}} &\leq \frac{n}{32\varepsilon_0 e^{\varepsilon_0}} \\
1 + \sqrt{ne^{\varepsilon_0}\frac{\log(1/\delta)}{c_1(e^{\varepsilon_0}-1)^2}} &\leq \frac{n}{32\varepsilon_0 e^{\varepsilon_0}} \\
\sqrt{\frac{(e^{\varepsilon_0}-1)^2}{ne^{\varepsilon_0}}} + \sqrt{\frac{\log(1/\delta)}{c_1}} &\leq \sqrt{\frac{n(1-e^{-\varepsilon_0})^2}{c_2\varepsilon_0^2}},
\end{aligned} \tag{28}$$

where $c_1 = 768$ and $c_2 = 1,024$ are constants. Note that (28) is satisfied for $e^{\varepsilon_0} \leq n$. Thus, we need $e^{\varepsilon_0} \leq 2 + \frac{v^2}{s^2} \leq n$ which is valid for any $\min\{\varepsilon^2, \varepsilon\}/s \leq 1$ or equivalently, $\varepsilon \leq s$. Thus, the transformation in (27) is valid for any $\varepsilon \leq s$.

D. MSE bound of the local DP model (Theorem 2) and shuffle model (Theorem 3)

For ease of analysis, we assume in the remaining part that $\frac{d}{s}$ is integer, otherwise, we can add dummy $s\lceil\frac{d}{s}\rceil - d$ zeros to the vector $\mathbf{b}_i$ to make the size of the vector divisible by s.

Now, we show that the output of the mechanism $\mathcal{R}^{\text{Bin}}_{v,s}$ is unbiased estimate of $\mathbf{b}_i$. Let $\mathcal{Y}_i$ be the output of Algorithm 1 and $a = \frac{d}{s}$. We can represent the output $\mathcal{Y}_i$ as a vector of dimension d that has s non-zero elements as follows: $\mathbf{y}_i = [\mathbf{y}_{i1}, \ldots, \mathbf{y}_{is}]$, where $\mathbf{y}_{ij} = a\mathcal{R}^{2RR}_p(\mathbf{b}_i[a_{ij}])\,\mathbf{e}_{a_{ij}}$ is a sub-vector of a dimensions that has only one non-zero element. Then, we have

$$\begin{aligned}
\mathbb{E}[\mathbf{y}_{ij}] &= \frac{1}{a}\sum_{a_{ij}=(j-1)a+1}^{ja} a\mathbf{e}_{a_{ij}}\mathbb{E}\left[\mathcal{R}^{2RR}_p(\mathbf{b}_i[a_{ij}])\right] \\
&\overset{(a)}{=} \sum_{a_{ij}=(j-1)a+1}^{ja}\mathbf{e}_{a_{ij}}\mathbf{b}_i[a_{ij}] \\
&= \mathbf{b}_i[(j-1)a+1:ja],
\end{aligned} \tag{29}$$

where $\mathbf{e}_j$ denotes the jth basis vector and (a) follows from the fact that the mechanism $\mathcal{R}^{2RR}_p$ shown in Theorem 1 is unbiased. $\mathbf{b}_i[l:m]$ denotes the values of the coordinates $l, l+1, \ldots, m$. As a result, we have that $\mathbb{E}[\mathbf{y}_i] = [\mathbb{E}[\mathbf{y}_{i1}], \ldots, \mathbb{E}[\mathbf{y}_{is}]] = \mathbf{b}_i$. Hence, Algorithm 1 is an unbiased estimate of the input $\mathbf{b}_i$. Furthermore, the variance of Algorithm 1 is bounded by:

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{y}_i - \mathbf{b}_i\|_2^2\right] &= \sum_{j=1}^{s}\mathbb{E}\left[\|\mathbf{y}_{ij} - \mathbf{b}_i[(j-1)a+1:ja]\|_2^2\right] \\
&= \sum_{j=1}^{s}\frac{1}{a}\sum_{a_{ij}=(j-1)a+1}^{ja}\mathbb{E}\Bigg[\Big\|a\mathbf{e}_{a_{ij}}\mathcal{R}^{2RR}_p(\mathbf{b}_i[a_{ij}]) \\
&\quad - \mathbf{b}_i[(j-1)a+1:ja]\Big\|^2\Bigg] \\
&= \frac{1}{a}\sum_{j=1}^{s}\sum_{a_{ij}=(j-1)a+1}^{ja}\mathbb{E}\Bigg[\Big\|\mathbf{e}_{a_{ij}}a\mathcal{R}^{2RR}_p(\mathbf{b}_i[a_{ij}]) - \mathbf{e}_{a_{ij}} \\
&\quad a\mathbf{b}_i[a_{ij}] + \mathbf{e}_{a_{ij}}a\mathbf{b}_i[a_{ij}] - \mathbf{b}_i[(j-1)a+1:ja]\Big\|^2\Bigg] \\
&\overset{(a)}{=} \frac{1}{a}\sum_{j=1}^{s}\sum_{a_{ij}=(j-1)a+1}^{ja}\mathbb{E}\Bigg[\Big\|\mathbf{e}_{a_{ij}}a\mathcal{R}^{\text{Bin}}_p(\mathbf{b}_i[a_{ij}]) - \mathbf{e}_{a_{ij}} \\
&\quad a\mathbf{b}_i[a_{ij}]\Big\|^2\Bigg] + \|\mathbf{e}_{a_{ij}}a\mathbf{b}_i[a_{ij}] - \mathbf{b}_i[(j-1)a+1:ja]\|^2 \\
&\overset{(b)}{=} \frac{sa^2p(1-p)}{(1-2p)^2} + \frac{1}{a}\sum_{j=1}^{d}\left((a-1)^2 + (a-1)\right)\mathbf{b}_i^2[j] \\
&= \frac{sa^2p(1-p)}{(1-2p)^2} + \frac{(a-1)\left((a-1)+1\right)}{a}\sum_{j=1}^{d}\mathbf{b}_i^2[j] \\
&= \frac{a^2sp(1-p)}{(1-2p)^2} + (a-1)\|\mathbf{b}_i\|^2 \\
&\overset{(c)}{\leq} \frac{sa^2p(1-p)}{(1-2p)^2} + (a-1)d
\end{aligned}$$

$$\overset{\text{(d)}}{=} \frac{s^3a^2}{v^2} + (a-1)d = d^2\left(\frac{1}{s} - \frac{1}{d} + \frac{s}{v^2}\right), \tag{30}$$

where (a) follows from the fact that the *2RR* mechanism $\mathcal{R}_p^{2RR}$ is unbiased and (b) from the variance of the *2RR* mechanism $\mathcal{R}_p^{2RR}$ (see Theorem 1). Step (c) follows from the fact that $\|\mathbf{b}_i\|^2 \leq d$. Step (d) follows from the fact that $p = \frac{1}{2}\left(1 - \sqrt{\frac{v^2/s^2}{v^2/s^2+4}}\right)$. Hence, we can bound the MSE in the local DP model and the shuffle model as follows.

MSE for the local DP model (Theorem 2): Observe that the output of the server $\hat{b} = \mathcal{A}^{\text{Bin}}(\mathcal{Y}_1, \ldots, \mathcal{Y}_n)$ can be represented as $\hat{b} = \frac{1}{n}\sum_{i=1}^n \mathbf{y}_i$, where $\mathbf{y}_i$ is the sparse representation of the i-th client private message discussed above. By setting $v^2 = \varepsilon_0^2$, we have that:

$$\begin{aligned}
\mathsf{MSE}_{\text{ldp}}^{\text{Bin}} &= \sup_{\{\mathbf{b}_i \in \{0,1\}^d\}} \mathbb{E}\left[\|\hat{\mathbf{b}} - \overline{\mathbf{b}}\|_2^2\right] \\
&\overset{\text{(a)}}{=} \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[\|\mathbf{y}_i - \mathbf{b}_i\|_2^2\right] \\
&\overset{\text{(b)}}{\leq} \frac{d^2}{n}\left(\frac{1}{s} - \frac{1}{d} + \frac{s}{v^2}\right) \\
&\overset{\text{(c)}}{=} \frac{d^2}{n}\left(\left(\frac{1}{s} - \frac{1}{d}\right) + \frac{s}{\varepsilon_0^2}\right) \\
&= \mathcal{O}\left(\frac{d^2}{n}\max\left\{\frac{1}{s}, \frac{s}{\varepsilon_0^2}\right\}\right),
\end{aligned} \tag{31}$$

where (a) follows from the i.i.d of the random mechanisms $\mathcal{R}_{v,s}^{\text{Bin}}$. Step (b) follows from the variance of the mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$ in (30). Step (c) follows from substituting $v^2 = \varepsilon_0^2$. This completes the proof of Theorem 2.

MSE for the MMS model (Theorem 3): Observe that the output of the server $\hat{b} = \mathcal{A}^{\text{Bin}}(\mathcal{Y}_1, \ldots, \mathcal{Y}_n)$ can be represented as $\hat{b} = \frac{1}{n}\sum_{i=1}^n \mathbf{y}_i$, where $\mathbf{y}_i$ is the sparse representation of the i-th client private message discussed above. By setting $v^2 = \frac{sn\min\{\varepsilon^2,\varepsilon\}}{c\log(1/\delta)}$, we have that:

$$\begin{aligned}
\mathsf{MSE}_{\text{shuffle}}^{\text{Bin}} &= \sup_{\{\mathbf{b}_i \in \{0,1\}^d\}} \mathbb{E}\left[\|\hat{\mathbf{b}} - \overline{\mathbf{b}}\|_2^2\right] \\
&\overset{\text{(a)}}{=} \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[\|\mathbf{y}_i - \mathbf{b}_i\|_2^2\right] \\
&\overset{\text{(b)}}{\leq} \frac{d^2}{n}\left(\frac{1}{s} - \frac{1}{d} + \frac{s}{v^2}\right) \\
&\overset{\text{(c)}}{=} \frac{d^2}{n^2}\left(n\left(\frac{1}{s} - \frac{1}{d}\right) + \frac{c\log(1/\delta)}{\min\{\varepsilon^2,\varepsilon\}}\right) \\
&= \mathcal{O}\left(\frac{d^2}{n^2}\max\left\{n\left(\frac{1}{s} - \frac{1}{d}\right), \frac{\log(1/\delta)}{\min\{\varepsilon^2,\varepsilon\}}\right\}\right),
\end{aligned} \tag{32}$$

where (a) follows from the i.i.d of the random mechanisms $\mathcal{R}_{v,s}^{\text{Bin}}$. Step (b) follows from the variance of the mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$ in (30). Step (c) follows from substituting $v^2 = \frac{sn\min\{\varepsilon^2,\varepsilon\}}{c\log(1/\delta)}$, where $c = 2,304$ is constant. This completes the proof of Theorem 3.

VIII. Proofs of Theorem 4 and Theorem 5 (Bounded ℓ_∞-norm vectors)

Algorithm 5 : Analyzer $\mathcal{A}^{\ell_\infty}$

Inputs: $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, where $\mathcal{Y}_i = \left\{\mathcal{Y}_i^{(1)}, \ldots, \mathcal{Y}_i^{(m)}\right\}$ is a set of m sets.
for $k = 1, \ldots, m-1$ **do**
  $\hat{\mathbf{b}}^{(k)} \leftarrow \mathcal{A}^{\text{Bin}}\left(\mathcal{Y}_1^{(k)}, \ldots, \mathcal{Y}_n^{(k)}\right)$
$\hat{\mathbf{u}} \leftarrow \mathcal{A}^{\text{Bin}}\left(\mathcal{Y}_1^{(m)}, \ldots, \mathcal{Y}_n^{(m)}\right)$
$\hat{\mathbf{z}} \leftarrow \sum_{k=1}^{m-1} \hat{\mathbf{b}}^{(k)} 2^{-k} + \hat{\mathbf{u}} 2^{-m+1}$
Return: The server returns $\hat{\mathbf{x}} \leftarrow 2r_\infty \hat{\mathbf{z}} - r_\infty$.

In this section, we prove Theorem 4 and Theorem 5 for estimating the mean of bounded ℓ_∞-norm vectors in LDP and shuffle models, respectively.

A. Communication cost for Theorem 4 and Theorem 5

In the mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$, the client sends m binary vectors $\mathbf{b}_i^{(1)}, \ldots, \mathbf{b}_i^{(m-1)}, \mathbf{u}_i$ using the private mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$. From Theorem 2 and Theorem 3, the private mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$ needs $\log\left(\lceil\frac{d}{s}\rceil\right)+1$ bits for communication. Thus, the total communication of the private mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ is $ms\left(\log\left(\lceil\frac{d}{s}\rceil\right)+1\right)$-bits.

B. Privacy of the local DP model in Theorem 4

In the mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$, each client sends m messages from the private mechanism $\mathcal{R}_{v,s}^{\text{Bin}}$ as follows: $\left\{\mathcal{R}_{v_1,s}^{\text{Bin}}(\mathbf{b}_i^{(1)}), \ldots, \mathcal{R}_{v_{m-1},s}^{\text{Bin}}(\mathbf{b}_i^{(m-1)}), \mathcal{R}_{v_m,s}^{\text{Bin}}(\mathbf{u}_i)\right\}$, where $v_k = \frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)}v$ for $k \in [m-1]$ and $v_m = \frac{4^{\frac{-m+1}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)}v$. Hence, from Theorem 2, the k-th message $\mathcal{R}_{v_k,s}^{\text{Bin}}(\mathbf{b}_i^{(k)})$ is $\varepsilon_0^{(k)}$-LDP, where $\varepsilon_0^{(k)} = v_k$ for $k \in [m]$. As a results, the total mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ is bounded by:

$$\begin{aligned}
\varepsilon_0 = \sum_{k=1}^m \varepsilon_0^{(k)} = \sum_{k=1}^m v_k = \sum_{k=1}^{m-1}\left\{\frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)}v\right\} \\
+ \frac{4^{\frac{-m+1}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)}v = v,
\end{aligned} \tag{33}$$

from the composition of the DP mechanisms [48]. Observe that we choose $v = \varepsilon_0$, and hence, the bound in (33) is satisfied. In addition, we can bound the RDP of the mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ in the local DP model by using the composition of the RDP (see Lemma 1). From the proof of Theorem 4 in Section VII, the mechanism $\mathcal{R}_{v_k,s}^{\text{Bin}}$ is $\left(\alpha, \varepsilon^{(k)}(\alpha)\right)$-RDP, where $\varepsilon^{(k)}(\alpha)$ is bounded by:

$$\varepsilon^{(k)}(\alpha) = \frac{s}{\alpha-1}\log\left(p_k^\alpha(1-p_k)^{1-\alpha} + p_k^{1-\alpha}(1-p_k)^\alpha\right), \tag{34}$$

where $p_k = \frac{1}{2}\left(1 - \sqrt{\frac{v_k^2/s^2}{v_k^2/s^2+4}}\right)$. Hence, the mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ is $(\alpha, \varepsilon(\alpha))$-RDP, where $\varepsilon(\alpha) = \sum_{k=1}^m \varepsilon^{(k)}(\alpha)$.

C. Privacy of the MMS model in Theorem 5

In the mechanism $\mathcal{R}^{\ell_\infty}_{v,m,s}$, each client sends m messages from the private mechanism $\mathcal{R}^{\text{Bin}}_{p,s}$ as follows: $\left\{\mathcal{R}^{\text{Bin}}_{v_1,s}(\mathbf{b}_i^{(1)}),\ldots,\mathcal{R}^{\text{Bin}}_{v_{m-1},s}(\mathbf{b}_i^{(m-1)}),\mathcal{R}^{\text{Bin}}_{v_m,s}(\mathbf{u}_i)\right\}$, where $v_k = \frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1}4^{\frac{-l}{3}}+4^{\frac{-m+1}{3}}\right)}v$ for $k\in[m-1]$ and $v_m = \frac{4^{\frac{-m+1}{3}}}{\left(\sum_{l=1}^{m-1}4^{\frac{-l}{3}}+4^{\frac{-m+1}{3}}\right)}v$.

From the proof of Theorem 3 in Section VII, shuffling the outputs of n mechanisms $\mathcal{R}^{\text{Bin}}_{v_k,s}$ is $\left(\alpha,\varepsilon^{(k)}(\alpha)\right)$, where $\varepsilon^{(k)}(\alpha)$ is bounded by:

$$\varepsilon^{(k)}(\alpha)\leq 48\alpha\frac{v_k^2}{sn}, \tag{35}$$

from (26) by substituting $p_k=\frac{1}{2}\left(1-\sqrt{\frac{v_k^2/s^2}{v_k^2/s^2+4}}\right)$. From Lemma 1 of the RDP composition, we get that the total RDP of the mechanism $\mathcal{R}^{\ell_\infty}_{v,m,s}$ is bounded by:

$$\begin{aligned}\varepsilon(\alpha)&=\sum_{k=1}^m\varepsilon^{(k)}(\alpha)=\alpha\frac{48}{sn}\sum_{k=1}^m v_k^2\\&=\alpha\frac{48v^2}{sn}\sum_{k=1}^m f_k^2\leq\alpha\frac{48v^2}{sn},\end{aligned} \tag{36}$$

where $f_k=\frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1}4^{\frac{-l}{3}}+4^{\frac{-m+1}{3}}\right)}$ for $k\in[m]$ and $f_m=\frac{4^{\frac{-m+1}{3}}}{\left(\sum_{l=1}^{m-1}4^{\frac{-l}{3}}+4^{\frac{-m+1}{3}}\right)}$. The last inequality is obtained from the fact that $\sum_{k=1}^m f_k=1$ and hence $\sum_{k=1}^m f_k^2\leq 1$. Now, we use Lemma 7 in Section X to convert from RDP to approximate DP, where $\rho=48v^2/(sn)$. For given $\delta>0$, shuffling the outputs of n mechanisms $\mathcal{R}^{\ell_\infty}_{v,m,s}$ is (ε,δ)-DP, where ε is bounded by

$$\varepsilon\leq 3\max\left\{\frac{48v^2}{sn}\log(1/\delta),\sqrt{\frac{48v^2}{sn}\log(1/\delta)}\right\}. \tag{37}$$

By setting $v^2=\frac{sn\min\{\varepsilon^2,\varepsilon\}}{c\log(1/\delta)}$, we can easily show that (37) is satisfied, and hence, the output of the shufflers is (ε,δ)-DP, where $c=2,304$ is constant.

D. MSE bound of the local DP model (Theorem 4) and MMS model (Theorem 5)

We first present some notations to simplify the analysis. For given $\mathbf{x}_i\in\mathbb{B}^d_\infty(r_\infty)$, we define $\mathbf{z}_i=\frac{\mathbf{x}_i+r_\infty}{2r_\infty}$, where the operations are done coordinate-wise. Thus, we have that $\mathbf{z}_i\in[0,1]^d$. For given $\mathbf{z}_i\in[0,1]^d$ and $m\geq 1$, we define $\tilde{\mathbf{z}}_i^{(m)}=\sum_{k=1}^{m-1}\mathbf{b}_i^{(k)}2^{-k}+\mathbf{u}_i2^{-m+1}$, where $\mathbf{b}_i^{(k)}=\lfloor 2^k\left(\mathbf{z}_i-\mathbf{z}_i^{(k-1)}\right)\rfloor$ and $\mathbf{z}_i^{(0)}=\mathbf{0}$ and $\mathbf{z}_i^{(k)}=\sum_{l=1}^k\mathbf{b}_i^{(l)}2^{-l}$ for $k\geq 1$. Furthermore, $\mathbf{u}_i$ is a Bernoulli vector defined by $\mathbf{u}_i=\text{Bern}\left(2^{m-1}\left(\mathbf{z}_i-\mathbf{z}_i^{(m-1)}\right)\right)$. Let $\overline{\mathbf{b}}^{(k)}=\frac{1}{n}\sum_{i=1}^n\mathbf{b}_i^{(k)}$, $\overline{\mathbf{u}}=\frac{1}{n}\sum_{i=1}^n\mathbf{u}_i$, and $\bar{\tilde{\mathbf{z}}}^{(m)}=\frac{1}{n}\sum_{i=1}^n\tilde{\mathbf{z}}_i^{(m)}$.

MSE for the local DP model (Theorem 2): Observe that the output of the server $\hat{\mathbf{x}}=\mathcal{A}^{\ell_\infty}(\mathcal{Y}_1,\ldots,\mathcal{Y}_n)=2r_\infty\hat{\mathbf{z}}-r_\infty$, where $\hat{\mathbf{z}}=\sum_{k=1}^{m-1}\hat{\mathbf{b}}^{(k)}2^{-k}+\hat{\mathbf{u}}2^{-m+1}$. Thus, we have that:

$$\begin{aligned}\mathsf{MSE}^{\ell_\infty}_{\text{ldp}}&=\sup_{\{\mathbf{x}_i\in\mathbb{B}^d_\infty(r_\infty)\}}\mathbb{E}\left[\|\hat{\mathbf{x}}-\overline{\mathbf{x}}\|_2^2\right]\\&\overset{(a)}{=}4r_\infty^2\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\mathbb{E}\left[\|\hat{\mathbf{z}}-\overline{\mathbf{z}}\|_2^2\right]\\&=4r_{\infty^2}\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\mathbb{E}\left[\|\hat{\mathbf{z}}-\bar{\tilde{\mathbf{z}}}^{(m)}+\bar{\tilde{\mathbf{z}}}^{(m)}-\overline{\mathbf{z}}\|_2^2\right]\\&\overset{(b)}{=}4r_\infty^2\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\left(\mathbb{E}\left[\|\hat{\mathbf{z}}-\bar{\tilde{\mathbf{z}}}^{(m)}\|_2^2\right]+\mathbb{E}\left[\|\bar{\tilde{\mathbf{z}}}^{(m)}-\overline{\mathbf{z}}\|_2^2\right]\right)\\&\overset{(c)}{\leq}4r_\infty^2\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\mathbb{E}\left[\left\|\sum_{k=1}^{m-1}\hat{\mathbf{b}}^{(k)}2^{-k}+\hat{\mathbf{u}}2^{-m+1}\right.\right.\\&\quad\left.\left.-\sum_{k=1}^{m-1}\overline{\mathbf{b}}^{(k)}2^{-k}+\overline{\mathbf{u}}2^{-m+1}\right\|_2^2\right]+\frac{d}{n4^m}\\&\overset{(d)}{\leq}4r_\infty^2\left(\sum_{k=1}^{m-1}\frac{d^24^{-k}}{n}\left(\frac{1}{s}+\frac{s}{v_k^2}\right)+\frac{d^24^{-m+1}}{n}\left(\frac{1}{s}+\frac{s}{v_m^2}\right)\right.\\&\quad\left.+\frac{d}{n4^m}\right)\\&\overset{(e)}{\leq}4r_\infty^2\left(\frac{d^2}{ns}\left(\sum_{k=1}^{m-1}4^{-k}+4^{-m+1}\right)\right.\\&\quad\left.+\frac{d^2s}{nv^2}\left(\sum_{k=1}^{m-1}4^{-k/3}+4^{-(m-1)/3}\right)^3+\frac{d}{n4^m}\right)\\&\overset{(f)}{\leq}4r_\infty^2\left(\frac{4d^2}{3ns}+\frac{5d^2s}{n\varepsilon_0^2}+\frac{d}{n4^m}\right)\end{aligned} \tag{38}$$

$$=\mathcal{O}\left(\frac{r_\infty^2d^2}{n}\max\left\{\frac{1}{d4^m},\frac{1}{s},\frac{s}{\varepsilon_0^2}\right\}\right), \tag{39}$$

where (a) follows from the fact that $\mathbf{z}_i$ is a linear transformation of $\mathbf{x}_i$. Step (b) follows from the fact that $\bar{\tilde{\mathbf{z}}}^{(m)}$ is an unbiased estimate of $\overline{\mathbf{z}}$ from Corollary 1. Step (c) from the bound of the MSE of the quantization scheme $\bar{\tilde{\mathbf{z}}}^{(m)}$ in Corollary 1. Step (d) follows from the MSE of the private mean estimation of binary vectors in Theorem 2. Step (e) follows from substituting $v_k=\frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1}4^{\frac{-l}{3}}+4^{\frac{-m+1}{3}}\right)}v$. Step (f) follows from the geometric series bound. This completes the proof of Theorem 4.

MSE for the MMS model (Theorem 3): Observe that the output of the server $\hat{\mathbf{x}}=\mathcal{A}^{\ell_\infty}(\mathcal{Y}_1,\ldots,\mathcal{Y}_n)=2r_\infty\hat{\mathbf{z}}-r_\infty$, where $\hat{\mathbf{z}}=\sum_{k=1}^{m-1}\hat{\mathbf{b}}^{(k)}+\hat{\mathbf{u}}2^{-m+1}$. Thus, we have that:

$$\begin{aligned}\mathsf{MSE}^{\ell_\infty}_{\text{shuffle}}&=\sup_{\{\mathbf{x}_i\in\mathbb{B}^d_\infty(r_\infty)\}}\mathbb{E}\left[\|\hat{\mathbf{x}}-\overline{\mathbf{x}}\|_2^2\right]\\&\overset{(a)}{=}4r_\infty^2\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\mathbb{E}\left[\|\hat{\mathbf{z}}-\overline{\mathbf{z}}\|_2^2\right]\\&=4r_{\infty^2}\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\mathbb{E}\left[\|\hat{\mathbf{z}}-\bar{\tilde{\mathbf{z}}}^{(m)}+\bar{\tilde{\mathbf{z}}}^{(m)}-\overline{\mathbf{z}}\|_2^2\right]\\&\overset{(b)}{=}4r_\infty^2\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\mathbb{E}\left[\|\hat{\mathbf{z}}-\bar{\tilde{\mathbf{z}}}^{(m)}\|_2^2\right]+\mathbb{E}\left[\|\bar{\tilde{\mathbf{z}}}^{(m)}-\overline{\mathbf{z}}\|_2^2\right]\\&\overset{(c)}{\leq}4r_\infty^2\sup_{\{\mathbf{z}_i\in[0,1]^d\}}\mathbb{E}\left[\left\|\sum_{k=1}^{m-1}\hat{\mathbf{b}}^{(k)}2^{-k}+\hat{\mathbf{u}}2^{-m+1}\right.\right.\\&\quad\left.\left.-\sum_{k=1}^{m-1}\overline{\mathbf{b}}^{(k)}2^{-k}+\overline{\mathbf{u}}2^{-m+1}\right\|_2^2\right]+\frac{d}{n4^m}\\&\overset{(d)}{\leq}4r_\infty^2\left(\sum_{k=1}^{m-1}\frac{d^24^{-k}}{n}\left(\left(\frac{1}{s}-\frac{1}{d}\right)+\frac{s}{v_k^2}\right)\right.\end{aligned}$$

Algorithm 6 : Analyzer $\mathcal{A}^{\ell_2}$

Inputs: $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, where $\mathcal{Y}_i = \left\{\mathcal{Y}_i^{(1)}, \ldots, \mathcal{Y}_i^{(m)}\right\}$ is a set of m sets.
$\hat{\mathbf{w}} \leftarrow \mathcal{A}^{\ell_\infty}\left(\mathcal{Y}_1, \ldots, \mathcal{Y}_n\right)$
Return: The server returns $\hat{\mathbf{x}} \leftarrow U^{-1}\hat{\mathbf{w}}$.

$$
\begin{aligned}
&+ \frac{d^2 4^{-m+1}}{n}\left(\left(\frac{1}{s}-\frac{1}{d}\right)+\frac{s}{v_m^2}\right)+\frac{d}{n4^m}\Bigg)\\
&\overset{\text{(e)}}{\leq} 4r_\infty^2\Bigg(\frac{d^2}{n}\left(\frac{1}{s}-\frac{1}{d}\right)\left(\sum_{k=1}^{m-1}4^{-k}+4^{-m+1}\right)\\
&+\frac{d^2 s}{nv^2}\left(\sum_{k=1}^{m-1}4^{-k/3}+4^{-(m-1)/3}\right)^3+\frac{d}{n4^m}\Bigg)\\
&\overset{\text{(f)}}{\leq} 4r_\infty^2\left(\frac{4d^2}{3n}\left(\frac{1}{s}-\frac{1}{d}\right)+\frac{5d^2\log\left(1/\delta\right)}{n^2\min\{\varepsilon^2,\varepsilon\}}+\frac{d}{n4^m}\right) \qquad (40)\\
&= \mathcal{O}\left(\frac{r_\infty^2 d^2}{n^2}\max\left\{\frac{n}{d4^m}, n\left(\frac{1}{s}-\frac{1}{d}\right), \frac{\log\left(1/\delta\right)}{\min\{\varepsilon^2,\varepsilon\}}\right\}\right), \qquad (41)
\end{aligned}
$$

where (a) follows from the fact that $\mathbf{z}_i$ is a linear transformation of $\mathbf{x}_i$. Step (b) follows from the fact that $\bar{\tilde{\mathbf{z}}}^{(m)}$ is an unbiased estimate of $\bar{\mathbf{z}}$ from Corollary 1. Step (c) from the bound of the MSE of the quantization scheme $\bar{\tilde{\mathbf{z}}}^{(m)}$ in Corollary 1. Step (d) follows from the MSE of the private mean estimation of binary vectors in Theorem 3. Step (e) follows from substituting $v_k = \frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1}4^{\frac{-l}{3}}+4^{\frac{-m+1}{3}}\right)}v$. Step (f) follows from the geometric series bound. This completes the proof of Theorem 5.

IX. Proofs of Theorem 6 and Theorem 7 (Bounded ℓ_2-norm vectors)

In this section, we prove Theorem 6 and Theorem 7 for the mean estimation of bounded ℓ_2-norm vectors in local DP and shuffle models, respectively.

In the mechanism $\mathcal{R}_{v,m,s}^{\ell_2}$, each client applies random rotation to her vector $\mathbf{x}_i$ and then applies the private mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ to the bounded ℓ_∞-norm vector $\mathbf{w}_i$. Hence the communication and privacy are the same as the private mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$. Thus, it remains to prove the MSE bound for both local DP model and shuffle model.

A. MSE bound of the local DP model (Theorem 6) and shuffle model (Theorem 7)

The proofs are obtained directly from the MSE of the bounded ℓ_∞-norm vector in Theorem 4 and Theorem 5 with the following Theorem about the random rotation matrix.

Theorem 10. *[64] Let $U = \frac{1}{\sqrt{d}}\mathbf{H}D$, where $\mathbf{H}$ denotes Hadamard matrix and D is a diagonal matrix with i.i.d. uniformly random $\{\pm 1\}$ entries. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{B}_2^d(r_2)$ be bounded ℓ_2-norm vectors and $\mathbf{w}_i = U\mathbf{x}_i$. With probability at least $1-\beta$, we have that*

$$
\max_{i\in[n]}\|\mathbf{w}_i\|_\infty = \max_{i\in[n]}\|U\mathbf{x}_i\|_\infty \leq 10r_2\sqrt{\frac{\log(\frac{nd}{\beta})}{d}}. \qquad (42)
$$

From Lemma 10, the vector $\mathbf{w}_i = U\mathbf{x}_i$ is bounded ℓ_∞-norm of radius $r_\infty = 10r_2\sqrt{\frac{\log(\frac{nd}{\beta})}{d}}$ with probability at least $1-\beta$. Hence, by plugging the radius $r_\infty = 10r_2\sqrt{\frac{\log(\frac{nd}{\beta})}{d}}$ into Theorem 6, we obtained the MSE in Theorem 6. Similarly, by plugging the radius $r_\infty = 10r_2\sqrt{\frac{\log(\frac{nd}{\beta})}{d}}$ into Theorem 5, we obtained the MSE in Theorem 7.

B. Lower bounds

A lower bound for the LDP model was proposed in [34, Theorem 2.1] and [61, Lemma 1].

Theorem 11 (Lower Bound For LDP model [34], [61])**.** *Let $n, d \in \mathbb{N}$ and $\varepsilon_0 > 0$. For any $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{B}_2^d(r_2)$, the MSE is bounded below by:*

$$
\mathsf{MSE}_{LDP}^{\ell_2} = \Omega\left(\frac{r_2^2 d}{n\min\left\{\varepsilon_0^2, \varepsilon_0, b\right\}}\right) \qquad (43)
$$

for any unbiased algorithm $\mathcal{M}$ that is ε_0-LDP with b-bits of communication per client.

Our lower bound for the shuffle model in Theorem 8 is a combination of the lower bound on the DME with communication constraints proposed in [32] and the lower bound on the DME with central (ε, δ)-DP constraints proposed in [59].

X. Binary Randomized Response

In this section we review an unbiased version of the classical binary randomized response (*2RR* mechanism) presented in Algorithm 7. We also gather some results on the classical binary randomized response, which will be useful for our proofs.

Theorem (Restating Theorem 1)**.** *For any $p \in [0, 1/2)$, the 2RR is ε_0-LDP, where $\varepsilon_0 = \log\left(\frac{1-p}{p}\right)$. The output y of the 2RR mechanism is an unbiased estimate of b with bounded MSE:*

$$
\mathsf{MSE}^{2RR} = \sup_{b\in\{0,1\}}\mathbb{E}\left[\|b-y\|_2^2\right] = \frac{p(1-p)}{(1-2p)^2}. \qquad (44)
$$

Proof. First, we show that the output of Algorithm 7 is unbiased estimate of b. Let y be the output of the *2RR* Algorithm 7. Then, we have

$$
\begin{aligned}
\mathbb{E}\left[y\right] &= \frac{b-p}{1-2p}(1-p)+\frac{1-b-p}{1-2p}p\\
&= b\left(\frac{1-2p}{1-2p}\right)-\frac{p(1-p)}{1-2p}+\frac{p(1-p)}{1-2p} \qquad (45)\\
&= b.
\end{aligned}
$$

Algorithm 7 : Local Randomizer $\mathcal{R}_p^{2RR}$

Public parameter: p
Input: $b \in \{0,1\}$.
Sample $\gamma \leftarrow \text{Ber}(p)$
if $\gamma == 0$ **then**
  $y = \frac{b-p}{1-2p}$
else
  $y = \frac{1-b-p}{1-2p}$
Return: The client sends y.

Hence, the Algorithm 7 is an unbiased estimate of the input b. Furthermore, the MSE of the *2RR* is bounded by:

$$\begin{aligned}
\mathsf{MSE}^{2RR} &= \sup_{b\in\{0,1\}} \mathbb{E}\left[\|y-b\|^2\right] = \mathbb{E}\left[y^2\right] - b^2 \\
&= \sup_{b\in\{0,1\}} \frac{1}{(1-2p)^2}\left[(b-p)^2(1-p) + (1-b-p)^2 p\right] - b^2 \\
&= \sup_{b\in\{0,1\}} \frac{1}{(1-2p)^2}\left[b^2 - 4p(1-p)b + p(1-p)\right] - b^2 \\
&= \sup_{b\in\{0,1\}} \frac{1}{(1-2p)^2}\left[b^2 - 4p(1-p)b + p(1-p)\right] - b^2 \\
&= \frac{1}{(1-2p)^2}\left[b^2(4p(1-p)) - 4p(1-p)b + p(1-p)\right] \\
&= \frac{p(1-p)}{(1-2p)^2}.
\end{aligned} \tag{46}$$

The LDP guarantees of the *2RR* is obtained from the fact that $e^{-\varepsilon_0} \le 1 \le \frac{1-p}{p} \le e^{\varepsilon_0}$ for any $p \in (0, 1/2]$. Furthermore, we can prove that the *2RR* satisfies $(\alpha, \varepsilon(\alpha))$-RDP, where $\varepsilon(\alpha)$ is given by:

$$\varepsilon(\alpha) = \frac{1}{\alpha-1}\log\left(p^\alpha(1-p)^{1-\alpha} + p^{1-\alpha}(1-p)^\alpha\right), \tag{47}$$

where this bound is obtained from the definition of the RDP and also given in [49]. This completes the proof of Theorem 1. ■

Next we present the following lemma which is useful for bounding the privacy parameter (ε_0) of our mechanisms which depend on the binary randomized response.

Lemma 6. *(Privacy parameter) For any $v > 0$, by setting $p = \frac{1}{2}\left(1 - \sqrt{\frac{v^2}{v^2+4}}\right)$, the 2RR mechanism with parameter p satisfies ε_0-LDP, where $\varepsilon_0 \le v$.*

Proof. From Theorem 1, the *2RR* mechanism with parameter $p < 1/2$ is ε_0-LDP, where $\varepsilon_0 = \log\left(\frac{1-p}{p}\right)$. Hence, it is sufficient to prove that $\varepsilon_0 = \log\left(\frac{1-p}{p}\right) \le v$ when choosing $p = \frac{1}{2}\left(1 - \sqrt{\frac{v^2}{v^2+4}}\right)$ for any $v \ge 0$.

Observe that $1 - p = \frac{1}{2}\left(1 + \sqrt{\frac{v^2}{v^2+4}}\right)$ when $p = \frac{1}{2}\left(1 - \sqrt{\frac{v^2}{v^2+4}}\right)$. Let $f(v) = v - \log\left(\frac{\sqrt{v^2+4}+v}{\sqrt{v^2+4}-v}\right)$. We have that

$$\begin{aligned}
\frac{\partial f}{\partial v} &= 1 - \frac{\sqrt{v^2+4}-v}{\sqrt{v^2+4}+v}\frac{8}{\left(\sqrt{v^2+4}-v\right)^2\sqrt{v^2+4}} \\
&= 1 - \frac{8}{(v^2+4-v^2)\sqrt{v^2+4}} \\
&= 1 - \frac{2}{\sqrt{v^2+4}} \\
&\ge 0 \qquad \forall\ v \ge 0.
\end{aligned} \tag{48}$$

Hence the function $f(v)$ is a non-decreasing function for all $v \ge 0$. As a result $f(v) \ge f(0) = 0$ for all $v \ge 0$. Thus, we have $v \ge \log\left(\frac{1-p}{p}\right)$ for all $v \ge 0$. This completes the proof of Lemma 6. ■

Now, we prove a useful lemma for conversion from RDP to approximate DP.

Lemma 7. *(Conversion from RDP to approximate DP) For given $\rho > 0$, let a mechanism $\mathcal{M}$ be $(\alpha, \alpha\rho)$-RDP. For any $\delta \in (0, 1/e)$, the mechanism $\mathcal{M}$ satisfies (ε, δ)-DP, where ε is bounded by:*

$$\varepsilon \le 3\max\left\{\rho\log(1/\delta), \sqrt{\rho\log(1/\delta)}\right\}. \tag{49}$$

Proof. The proof is obtained from Lemma 2, where the ε is bounded by:

$$\varepsilon \le \min_\alpha \rho\alpha + \frac{\log(1/\delta)}{\alpha-1} + \log\left(1 - \frac{1}{\alpha}\right), \tag{50}$$

for given $\delta \in (0, 1/e)$. By setting $\alpha = 1 + \sqrt{\frac{\log(1/\delta)}{\rho}}$, we get that:

$$\begin{aligned}
\varepsilon &\le \rho + 2\sqrt{\rho\log(1/\delta)} \\
&\le \rho\log(1/\delta) + 2\sqrt{\rho\log(1/\delta)} \\
&\le 3\max\left\{\rho\log(1/\delta), \sqrt{\rho\log(1/\delta)}\right\}.
\end{aligned} \tag{51}$$

This completes the proof of Lemma 7. ■

REFERENCES

[1] A. M. Girgis and S. Diggavi, "Distributed mean estimation for multi-message shuffled privacy," in *Federated Learning and Analytics in Practice: Algorithms, Systems, Applications, and Opportunities, ICML Workshops*, 2023.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[3] 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.

[4] M. Li, D. G. Andersen, A. J. Smola, and K. Yu, "Communication efficient distributed machine learning with the parameter server," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[5] J. Bernstein, Y.-X. Wang, K. Azizzadenesheli, and A. Anandkumar, "signsgd: Compressed optimisation for non-convex problems," in *International Conference on Machine Learning*. PMLR, 2018, pp. 560–569.

[6] D. Alistarh, D. Grubic, J. Li, R. Tomioka, and M. Vojnovic, "Qsgd: Communication-efficient sgd via gradient quantization and encoding," *Advances in neural information processing systems*, vol. 30, 2017.

[7] S. U. Stich, J.-B. Cordonnier, and M. Jaggi, "Sparsified sgd with memory," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[8] V. Gandikota, D. Kane, R. K. Maity, and A. Mazumdar, "vqsgd: Vector quantized stochastic gradient descent," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 2197–2205.

[9] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," *Advances in neural information processing systems*, vol. 32, 2019.

[10] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients-how easy is it to break privacy in federated learning?" *Advances in Neural Information Processing Systems*, vol. 33, pp. 16 937–16 947, 2020.

[11] N. Carlini, F. Tramer, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, A. Roberts, T. Brown, D. Song, U. Erlingsson *et al.*, "Extracting training data from large language models," in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 2633–2650.

[12] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE symposium on security and privacy (SP)*. IEEE, 2017, pp. 3–18.

[13] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*. Springer, 2006, pp. 265–284.

[14] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization." *Journal of Machine Learning Research*, vol. 12, no. 3, 2011.

[15] R. Bassily, A. Smith, and A. Thakurta, "Private empirical risk minimization: Efficient algorithms and tight error bounds," in *2014 IEEE 55th annual symposium on foundations of computer science*. IEEE, 2014, pp. 464–473.

[16] P. Kairouz, B. McMahan, S. Song, O. Thakkar, A. Thakurta, and Z. Xu, "Practical and private (deep) learning without sampling or shuffling," in *International Conference on Machine Learning*. PMLR, 2021, pp. 5213–5225.

[17] A. M. Girgis, D. Data, S. Diggavi, P. Kairouz, and A. T. Suresh, "Shuffled model of differential privacy in federated learning," in *International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 2021, pp. 2521–2529.

[18] P. Kairouz, Z. Liu, and T. Steinke, "The distributed discrete gaussian mechanism for federated learning with secure aggregation," in *Proceedings International Conference on Machine Learning, ICML*, vol. 139, 2021, pp. 5201–5212.

[19] A. T. Suresh, X. Y. Felix, S. Kumar, and H. B. McMahan, "Distributed mean estimation with limited communication," in *International conference on machine learning*. PMLR, 2017, pp. 3329–3337.

[20] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[21] K. Chaudhuri, C. Guo, and M. Rabbat, "Privacy-aware compression for federated data analysis," in *Uncertainty in Artificial Intelligence*. PMLR, 2022, pp. 296–306.

[22] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.

[23] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 2013, pp. 429–438.

[24] A. Bittau, Ú. Erlingsson, P. Maniatis, I. Mironov, A. Raghunathan, D. Lie, M. Rudominer, U. Kode, J. Tinnes, and B. Seefeld, "Prochlo: Strong privacy for analytics in the crowd," in *Proceedings of the 26th symposium on operating systems principles*, 2017, pp. 441–459.

[25] Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta, "Amplification by shuffling: From local to central differential privacy via anonymity," in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2019, pp. 2468–2479.

[26] A. Cheu, A. Smith, J. Ullman, D. Zeber, and M. Zhilyaev, "Distributed differential privacy via shuffling," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2019, pp. 375–403.

[27] A. M. Girgis, D. Data, and S. Diggavi, "Differentially private federated learning with shuffling and client self-sampling," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 338–343.

[28] B. Ghazi, N. Golowich, R. Kumar, R. Pagh, and A. Velingker, "On the power of multiple anonymous messages: Frequency estimation and selection in the shuffle model of differential privacy," in *Advances in Cryptology - EUROCRYPT 2021 - Theory and Applications of Cryptographic Techniques*, vol. 12698, 2021, pp. 463–488.

[29] B. Balle, J. Bell, A. Gascón, and K. Nissim, "Private summation in the multi-message shuffle model," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 2020, pp. 657–676.

[30] A. Cheu, M. Joseph, J. Mao, and B. Peng, "Shuffle private stochastic convex optimization," in *International Conference on Learning Representations (ICLR)*, 2022.

[31] A. Chang, B. Ghazi, R. Kumar, and P. Manurangsi, "Locally private k-means in one round," in *International Conference on Machine Learning*. PMLR, 2021, pp. 1441–1451.

[32] W.-N. Chen, C. A. C. Choo, P. Kairouz, and A. T. Suresh, "The fundamental price of secure aggregation in differentially private federated learning," in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 162, 17–23 Jul 2022, pp. 3056–3089.

[33] W.-N. Chen, D. Song, A. Ozgur, and P. Kairouz, "Privacy amplification via compression: Achieving the optimal privacy-accuracy-communication trade-off in distributed mean estimation," *arXiv preprint arXiv:2304.01541*, 2023.

[34] W.-N. Chen, P. Kairouz, and A. Ozgur, "Breaking the communication-privacy-accuracy trilemma," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3312–3324, 2020.

[35] H. Asi, V. Feldman, and K. Talwar, "Optimal algorithms for mean estimation under local differential privacy," in *International Conference on Machine Learning*. PMLR, 2022, pp. 1046–1056.

[36] C. Guo, K. Chaudhuri, P. Stock, and M. Rabbat, "The interpolated mvu mechanism for communication-efficient private federated learning," *arXiv preprint arXiv:2211.03942*, 2022.

[37] B. Ghazi, R. Kumar, P. Manurangsi, and R. Pagh, "Private counting from anonymous messages: Near-optimal accuracy with vanishing communication overhead," in *International Conference on Machine Learning*. PMLR, 2020, pp. 3505–3514.

[38] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Cryptography from anonymity," in *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE, 2006, pp. 239–248.

[39] A. M. Girgis and S. N. Diggavi, "Multi-message shuffled privacy in federated learning," *CoRR*, vol. abs/2302.11152, 2023, posted 22nd February 2023, also presented at ICML FL workshop, June 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.11152

[40] W. Chen, D. Song, A. Özgür, and P. Kairouz, "Privacy amplification via compression: Achieving the optimal privacy-accuracy-communication trade-off in distributed mean estimation," *CoRR*, vol. abs/2304.01541, 2023, posted 4th April 2023, also presented at ICML FL workshop, June 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2304.01541

[41] N. Agarwal, A. T. Suresh, F. X. X. Yu, S. Kumar, and B. McMahan, "cpsgd: Communication-efficient and differentially-private distributed sgd," in *Advances in Neural Information Processing Systems*, 2018, pp. 7564–7575.

[42] Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, S. Song, K. Talwar, and A. Thakurta, "Encode, shuffle, analyze privacy revisited: Formalizations and empirical evaluation," *CoRR*, vol. abs/2001.03618, 2020.

[43] A. M. Girgis, D. Data, S. Diggavi, P. Kairouz, and A. T. Suresh, "Shuffled model of federated learning: Privacy, accuracy and communication trade-offs," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 464–478, 2021.

[44] A. Girgis, D. Data, and S. Diggavi, "Renyi differential privacy of the subsampled shuffle model in distributed learning," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, pp. 29 181–29 192, 2021.

[45] A. M. Girgis, D. Data, S. Diggavi, A. T. Suresh, and P. Kairouz, "On the renyi differential privacy of the shuffle model," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2021, pp. 2321–2341.

[46] V. Feldman, A. McMillan, and K. Talwar, "Stronger privacy amplification by shuffling for Renyi and approximate differential privacy," in *Proceedings of ACM-SIAM Symposium on Discrete Algorithms, SODA*. SIAM, 2023, pp. 4966–4981.

[47] C. Dwork, F. McSherry, K. Nissim, and A. D. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference (TCC)*, 2006, pp. 265–284.

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

[49] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 2017, pp. 263–275.

[50] C. L. Canonne, G. Kamath, and T. Steinke, "The discrete gaussian for differential privacy," in *Advances in Neural Information Processing Systems NeurIPS*, 2020.

[51] B. Balle, G. Barthe, M. Gaboardi, J. Hsu, and T. Sato, "Hypothesis testing interpretations and renyi differential privacy," in *International Conference*

on Artificial Intelligence and Statistics (AISTATS), ser. Proceedings of Machine Learning Research, S. Chiappa and R. Calandra, Eds., vol. 108. PMLR, 2020, pp. 2496–2506.
[52] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.
[53] B. Ghazi, R. Kumar, P. Manurangsi, R. Pagh, and A. Sinha, "Differentially private aggregation in the shuffle model: Almost central accuracy in almost a single message," in *International Conference on Machine Learning*. PMLR, 2021, pp. 3692–3701.
[54] B. Ghazi, N. Golowich, R. Kumar, P. Manurangsi, R. Pagh, and A. Velingker, "Pure differentially private summation from anonymous messages," in *1st Conference on Information-Theoretic Cryptography*, 2020.
[55] B. S. Kashin, "Diameters of some finite-dimensional sets and classes of smooth functions," *Math. USSR, Izv*, vol. 11, no. 2, pp. 317–333, 1977.
[56] Y. Lyubarskii and R. Vershynin, "Uncertainty principles and vector quantization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3491–3501, 2010.
[57] S. Caldas, J. Konečny, H. B. McMahan, and A. Talwalkar, "Expanding the reach of federated learning by reducing client resource requirements," *arXiv preprint arXiv:1812.07210*, 2018.
[58] V. Feldman and K. Talwar, "Lossless compression of efficient private local randomizers," in *International Conference on Machine Learning*. PMLR, 2021, pp. 3208–3219.
[59] M. Bun, J. Ullman, and S. Vadhan, "Fingerprinting codes and the price of approximate differential privacy," in *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 2014, pp. 1–10.
[60] Y.-X. Wang, B. Balle, and S. P. Kasiviswanathan, "Subsampled rényi differential privacy and analytical moments accountant," in *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 1226–1235.
[61] A. M. Girgis, D. Data, S. Diggavi, P. Kairouz, and A. T. Suresh, "Shuffled model of federated learning: Privacy, accuracy and communication trade-offs," *IEEE journal on selected areas in information theory*, vol. 2, no. 1, pp. 464–478, 2021.
[62] V. Feldman, A. McMillan, and K. Talwar, "Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling," in *2022 IEEE 62nd Annual Symposium on Foundations of Computer Science*. IEEE, 2022.
[63] A. Bhowmick, J. Duchi, J. Freudiger, G. Kapoor, and R. Rogers, "Protection against reconstruction and its applications in private federated learning," *arXiv preprint arXiv:1812.00984*, 2018.
[64] D. Levy, Z. Sun, K. Amin, S. Kale, A. Kulesza, M. Mohri, and A. T. Suresh, "Learning with user-level privacy," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 466–12 479, 2021.

Antonious M. Girgis is currently a research scientist at Google, Mountain View, CA, USA. He received the B.Sc. degree in electrical engineering from Cairo University, Egypt, in 2014, the M.Sc. degree in electrical engineering from Nile University, Egypt, in 2018, and the Ph.D. degree in the electrical and computer engineering from the University of California, Los Angeles (UCLA), in 2023. He was the receipt of the 2021 ACM Conference on Computer and Communications Security (CCS) best paper award, and the receipt of distinguished Ph.D. dissertation award in signals and systems from the ECE department, UCLA. He was an Exchange Research Assistant with Sabanci University, Turkey, from 2016 to 2017. He received the Masters Fellowship and a graduate Research Assistantship from Nile University for the years 2014-2018. He received the Electrical and Computer Engineering Department Fellowship from UCLA for the year 2018/2019, and the 2022 Amazon Ph.D fellowship. His research interests include privacy, machine learning, information theory, and optimization.

Suhas Diggavi is currently a Professor of Electrical and Computer Engineering at UCLA. His undergraduate education is from IIT, Delhi and his PhD is from Stanford University. He has worked as a principal member research staff at AT&T Shannon Laboratories and directed the Laboratory for Information and Communication Systems (LICOS) at EPFL. At UCLA, he directs the Information Theory and Systems Laboratory.

His research interests include information theory and its applications to several areas including machine learning, security & privacy, wireless networks, data compression, cyber-physical systems, bio-informatics and neuroscience; more information can be found at http://licos.ee.ucla.edu.

He has received several recognitions for his research from IEEE and ACM, including the 2013 IEEE Information Theory Society & Communications Society Joint Paper Award, the 2021 ACM Conference on Computer and Communications Security (CCS) best paper award, the 2013 ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) best paper award, the 2006 IEEE Donald Fink prize paper award among others. He was selected as a Guggenheim fellow in 2021. He also received the 2019 Google Faculty Research Award, 2020 Amazon faculty research award and 2021 Facebook/Meta faculty research award. He served as a IEEE Distinguished Lecturer and also served on board of governors for the IEEE Information theory society (2016-2021). He is a Fellow of the IEEE.

He is the Editor-in-Chief of the IEEE BITS Information Theory Magazine and has been an associate editor for IEEE Transactions on Information Theory, ACM/IEEE Transactions on Networking and other journals and special issues, as well as in the program committees of several IEEE conferences. He has also helped organize IEEE and ACM conferences including serving as the Technical Program Co-Chair for 2012 IEEE Information Theory Workshop (ITW), the Technical Program Co-Chair for the 2015 IEEE International Symposium on Information Theory (ISIT) and General co-chair for ACM Mobihoc 2018. He has 8 issued patents.