Side-channel Attacks and New Principles in the Shuffle Model of Differential Privacy

Anonymous Author(s)

ABSTRACT

The shuffle model employs a shuffler to anonymize and permute user messages, thereby enhancing privacy/utility trade-offs compared to the local model. Ideally, it assumes perfect message anonymity protection against adversaries, such as analyzers or aggregation servers, allowing each user to blend into a large population. However, in contexts like mobile and edge computing or telemetry data collection across cable and wireless networks, this assumption is frequently unrealistic. In this study, we demonstrate the vulnerability of the shuffle model to communication side-channel attacks, which substantially compromise privacy amplification via shuffling.

We categorize side-channel information in the shuffle model into three types: (i) in-out information, revealing the victim user's participation and timing, (ii) message-cardinality information, indicating the victim's message count, and (iii) message-length information, disclosing the victim's message length(s). Numerical results indicate these attacks increase privacy loss by 200% to 4100%, revealing secret value with probability more than 90%. After a theoretical analysis of the remaining privacy amplification effects, we suggest several countermeasures and principles to alleviate degradation caused by these attacks: (a) appending padding bits to each message to counter message-length attacks, (b) maximizing query parallelization to elude in-out attacks and increase the population for privacy amplification, and (c) introducing dummy messages to exchange communication costs for improved privacy amplification effects. These principles are also applicable to scenarios without side-channel attacks and result in significant privacy budget savings compared to existing intra-batch privacy amplification methods. Through representative experiments, we corroborate our analyses of novel attacks and new guidelines, achieving a 75% reduction in privacy budget relative to current models.

CCS CONCEPTS

• **Security and privacy** → **Data anonymization and sanitization**; **Privacy-preserving protocols**;

KEYWORDS

differential privacy, side channels, shuffle model

1 INTRODUCTION

The shuffle model of differential privacy [15, 39] has emerged as a compelling approach to data privacy protection, combining the benefits of the classical centralized model [36] (i.e., relatively high data utility) and the local model [35, 57] (i.e., minimal trust in other parties). In the shuffle model, an intermediary shuffler anonymizes and randomly permutes messages from a user population before forwarding them to the server (e.g., data analysts and statisticians). As the shuffling operation is data-agnostic and can be executed over ciphertext space, numerous semi-trusted parties can assume the shuffler's role in practice, including anonymous communication channels [2, 18, 31, 85], secure hardware [15, 29, 79], and cryptographic tools [4, 13, 14, 28, 54]. Moreover, since potential privacy adversaries (e.g., the server/analyzer) only observe anonymized and shuffled messages, each user can hide within a large population. Consequently, adding a minimal amount of noise to local messages sufficiently protects data privacy in the released view of shuffled messages. This phenomenon is known as *privacy amplification via shuffling* [39]. For instance, in the single-message shuffle model where each user sends one message satisfying local ϵ_0-differential privacy, the shuffled messages from n users effectively preserve $\tilde{O}(\sqrt{e^{\epsilon_0}/n})$-differential privacy [9, 42, 43]; in the multi-message shuffle model where each user sends multiple messages without a direct local privacy guarantee, as long as these messages are carefully calibrated, the overall shuffled messages can achieve (ϵ, δ)-DP. Due to its potential for achieving excellent privacy-utility trade-offs in decentralized settings, the shuffle model has been applied across various domains, including count/summation queries [6, 7, 9–11, 19, 23, 39, 47–49, 56, 66, 69, 84, 93, 94, 96] and machine learning [20, 24, 46, 51–53, 64, 68, 88, 101].

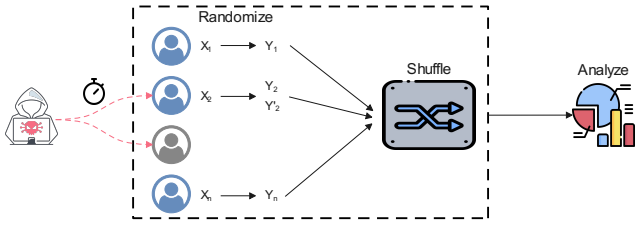

Figure 1: An illustration of the shuffle model under user-side communication side-channel attacks.

1.1 Our Contributions

Side-Channel Attacks in the Shuffle Model. Despite the shuffle model's success in decentralized private data analysis, it relies on several unrealistic security assumptions in such environments. Ideally, privacy adversaries in the shuffle model are assumed to lack intermediate information about a victim user during protocol execution, only having access to the victim's shuffled messages. We argue that this assumption is easily violated in mobile computing [45], edge computing [71], or cable networking environments where the shuffle model is applied (see Figure 1 for an illustration). For example, in wireless/cellular networks, privacy adversaries can precisely infer a victim user's communication activities at a negligible cost [32, 41, 86, 89, 102]. Adversaries can effortlessly deduce the victim's communication timing with the shuffler, potentially ascertain the message payload size, and infer the number of messages contributed by the victim.

Table 1: Comparison of the amplification population size and other properties of various shuffle models under in-out and message-length attacks. The variables include: n, the full population size; n_k, the size of the population in the same k-th batch as the victim user; $n_{*,l}$, the size of the population with the same message length as the victim user; $n_{k,l}$, the size of the population with the same message length and in the same k-th batch as the victim user; and $n_{(m)}$, the size of the population that selected the same m-th bin as the victim user.

shuffle models	normal setting	in-out attacks	in-out & length atk.	adaptive queries	extra costs
shuffle-then-randomize [39, 42, 43][Definition 3.5]	n	1	1	✓	N.A.
divide-randomize-shuffle [20, 24, 46, 64, 68, 88] [Definition 3.6]	n_k	n_k	$n_{k,l}$	✓	N.A.
subsample-randomize-shuffle [50, 51] [Definition 3.7]	n	n_k	$n_{k,l}$	✓	N.A.
parallel-randomize-shuffle [39, 91][Algorithm 3]	n	n	$n_{*,l}$	✗	N.A.
multinomial-randomize-shuffle **this work** [Definition 6.1]	n	n_k	$n_{k,l}$	✓	N.A.
rectified parallel-randomize-shuffle **this work** [Algorithm 4]	n	n	n	✗	padding bits per message
rectified multinomial-randomize-shuffle **this work** [Algorithm 5]	n	n	n	✓	padding bits per message K-time dummy messages
rectified bin-randomize-shuffle **this work** [Algorithm 6]	$n_{(m)}$	$n_{(m)}$	$n_{(m)}$	✓	padding bits per message few dummy messages

Shuffle Privacy Degradation Due to Side-Channel Information. In this study, we demonstrate that side-channel information, easily inferred by privacy adversaries, significantly impacts privacy amplification effects. Intuitively, a victim user's communication timing reveals the specific task (e.g., gradient descent iteration of federated learning [60, 65]) they participated in, allowing the victim to hide only among the sub-population of that particular task, which comprises several tens or hundreds of users. If privacy adversaries know the length of the victim's contributed message, they can filter out many other messages with different lengths in the shuffled messages, substantially reducing the number of messages/users the victim can hide among. In some multi-message protocols, the number of messages sent by a user correlates with the user's true value, meaning the message cardinality information possessed by privacy adversaries entirely undermines privacy amplification via shuffling. Numerical results on typical settings (e.g., a dozen queries) show that communication timing (termed as in-out attacks) and message-length attacks increase the victim user's privacy loss by 200% to 4100%, revealing the victim's secret value with probability more than 90%; for several state-of-the-art multi-message protocols, the message-cardinality attack increases privacy loss to $+\infty$, revealing the victim's secret value with probability more than 95%. It is important to note that these side-channel attacks are specific to the shuffle model. In contrast, in the local model of differential privacy, revealing the timing, payload length, or the number of messages incurs no additional privacy loss, as a user's local messages are publicly releasable and, consequently, the side-channel information is also releasable.

Defending against Side-Channel Attacks. We conducted a theoretical analysis of privacy degradation in widely-used shuffle models (e.g., shuffle-then-randomize model [39, 42, 43] and randomize-then-shuffle model [9, 51, 91]) under side-channel attacks and developed new shuffle models resistant to such attacks. In the proposed model, users make their own participation choices, reducing trust assumptions regarding the shuffler/server. To minimize privacy loss due to participation timing, the original data analysis task is first represented as a directed acyclic graph (DAG). Based on this, multiple independent queries are issued simultaneously, while sequential dependent queries are addressed using dummy messages from users, ensuring indistinguishable participation timings. To counteract privacy loss resulting from message payload size/length, each user's message is padded to a fixed length. These techniques form new principles for private data analysis shuffle models under side-channel attacks. Furthermore, even in private data analysis without side-channel attacks, our proposed shuffle models: multinomial-randomize-shuffle and bin-randomize-shuffle models, provide much stronger privacy amplification than existing methods (e.g., commonly-used divide-randomize-shuffle model [20, 68, 88], subsample-randomize-shuffle model [50, 51] and random check-in model [11]), saving 50% to 80% of privacy budgets in typical settings. In single-message shuffle models, we demonstrate the feasibility of achieving full population privacy amplification across multiple adaptive queries; in multi-message shuffle models, we also show the potential to achieve full population privacy amplification across multiple (non-adaptive or adaptive) queries.

We summarize in-out/message-length attack and defense results in Table 1. The *normal* column denotes the amplification population size within which the victim user can hide when there are no side-channel attacks; the *in-out attacks* column represents the amplification population size under in-out attacks; the *in-out & length atk.* column signifies the amplification population size under joint in-out and message-length attacks. The *adaptive query* column indicates whether the model supports sequentially adaptive queries. The *extra costs* column highlights the additional communication costs compared to the normal model.

1.2 Organization

The remainder of this paper is organized as follows. Section 2 reviews related work on the shuffle model and side-channel attacks. Section 3 provides background information and defines relevant

terminology. Section 4 identifies several side-channel attacks in various shuffle models and analyzes their impact on privacy. Section 6 proposes rectified shuffle models to defend against side-channel attacks. Section 7 summarizes new principles for the shuffle model and presents several discussions. Finally, Section 8 concludes the paper and offers directions for future research.

2 RELATED WORK

2.1 Side-channel attacks

Side-channel attacks exploit unintended information leakage through a system's physical properties or observable characteristics, such as electromagnetic radiation [16], power consumption [59], and communication patterns [58, 67].

Side-channel attacks on anonymous channels. Anonymous channels [34] aim to protect users' identities and relationships between communicating parties. Researchers have investigated various side-channel attacks that target anonymous channels, including traffic analysis attacks [74], timing attacks [77], and intersection attacks [30]. These attacks exploit information leaks from communication patterns, message timings, or user behavior to undermine the anonymity guarantees provided by the channels.

In this work, we demonstrate that side-channel attacks can significantly impact privacy guarantees in the shuffle model, even when the internal shuffling/anonymous channel is perfectly secure. To the best of our knowledge, this is the first study to apply side-channel attacks/defenses to joint systems of cryptographic tools (i.e., the shuffling/anonymous channel) and statistical privacy tools (e.g., differential privacy). We note that the local model of differential privacy, where each user trusts no other party, has limited vulnerability to side-channel attacks, as the plaintext data leaving the user/device is publishable, and thus side-channel information related to it is also publishable.

2.2 Shuffle model of differential privacy

The shuffle model of differential privacy anonymizes user messages through semi-trusted shufflers before sending them to a server for analysis. The model's foundation lies in privacy amplification analysis, ensuring the global privacy of shuffled messages. In the seminal work, Erlingsson et al. [39] employed privacy amplification via subsampling [8, 57] to analyze privacy amplification in shuffling. They demonstrated that n shuffled messages satisfy $(\epsilon_0\sqrt{144\log(1/\delta)/n}, \delta)$-DP. This bound is further improved by recent works [9, 42, 43, 91] (see Section 3.4 for more detail). Depending on the number of messages a user can send, the shuffle privacy model can be categorized as either multi-message [10, 23, 47, 69] or single-message [9, 39, 42, 61].

Various shuffle model variants exist in the literature. For convenient theoretical analysis of privacy amplification, studies such as [39, 42, 43] utilize the shuffle-then-randomize model, in which user data are first shuffled and then fed into adaptive local randomizers sequentially. Practical approaches to support multiple adaptive queries in decentralized setting is to divide the user population into non-overlapping parts (e.g., in [20, 88], referred to as the divide-randomize-shuffle model), or letting the shuffler to perform user subsampling (e.g., in [50, 51], referred to as the subsample-randomize-shuffle model). Another approach is to let each user randomly select (at most) one query to check in, and then choose one user (from checked-in users) for each query slot.

Table 2: List of notations.

Notation	Description
$[i]$	$\{1, 2, ..., i\}$
$[i:j]$	$\{i, i+1, ..., j\}$
n	the number of users (data owners)
K	the number of sequential queries
U_k	the users participated in the k-th round query
$\mathbb{X}$	the domain of input data
$\mathbb{Y}$	the domain of a single message
ϵ_0	the local privacy budget
$\mathcal{S}$	the shuffling algorithm/procedure
$\mathcal{R}$	the randomization algorithm
io_v	the participation information of user v
num_v	the number of messages from user v
$len(y)$	the message length (bits) of message y

We demonstrate that existing shuffle models, including the shuffle-then-randomize model, divide-randomize-shuffle model, subsample-randomize-shuffle model, and random check-in, are all susceptible to side-channel attacks. Their privacy amplification degrades to the intra-batch level or even to the local level (i.e., no privacy amplification). Worse still, for several protocols in these shuffle models, such as SOTA multi-message summation protocols [6, 47–49], the privacy loss under side-channel attack increases to $+\infty$.

2.3 Security attacks in differential privacy

Differential privacy, as the de facto standard for data privacy, is widely adopted in industry for sensitive databases [3, 70, 76] and decentralized data collection/analysis [33, 40, 87]. Although differential privacy provides rigorous data privacy in theory, its practical implementation may encounter unexpected security issues. For example, an improper implementation of floating-point numbers for the prevalent Laplace mechanism compromises the intended differential privacy [72]. Besides, a small portion of adversarial users may completely undermine the aggregation results of locally private protocols [17, 21, 92, 95, 100] and shuffle private protocols [7, 10, 69]. To alleviate the (semi)-trustness on a single shuffler, it might be necessary to introduce multiple shufflers in the shuffle model [97]. To the best of our knowledge, this work is the first to study side-channel security issues of differential privacy.

3 PRELIMINARIES

In this section, we present definitions of differential privacy and shuffle models. Commonly used notations are listed in Table 2.

3.1 Divergences and Differential Privacy

Definition 3.1 (Hockey-stick divergence). The Hockey-stick divergence between two random variables P and Q is defined as:

$$D_{e^\epsilon}(P\|Q) = \int \max\{0, P(x) - e^\epsilon Q(x)\}\mathrm{d}x,$$

where P and Q denote both the random variables and their probability density functions.

Two variables P and Q are (ϵ, δ)-indistinguishable if $\max\{D_{e^\epsilon}(P\|Q), D_{e^\epsilon}(Q\|P)\} \leq \delta$. For two datasets of equal size that differ only by a single individual's data, they are referred to as *neighboring datasets*. Differential privacy limits the divergence of query results on neighboring datasets (see Definition 3.2). Similarly, in the local setting that accepts a single individual's data as input, we introduce the local (ϵ, δ)-differential privacy in Definition 3.3. When $\delta = 0$, the concept is abbreviated as ϵ-LDP.

Definition 3.2 (Differential privacy [38]). A protocol $\mathcal{R} : \mathbb{X}^n \mapsto \mathbb{Z}$ satisfies (ϵ, δ)-differential privacy if, for all neighboring datasets $X, X' \in \mathbb{X}^n$, $\mathcal{R}(X)$ and $\mathcal{R}(X')$ are (ϵ, δ)-indistinguishable.

Definition 3.3 (Local differential privacy [57]). A protocol $\mathcal{R} : \mathbb{X} \mapsto \mathbb{Y}$ satisfies local (ϵ, δ)-differential privacy if, for all $x, x' \in \mathbb{X}$, $\mathcal{R}(x)$ and $\mathcal{R}(x')$ are (ϵ, δ)-indistinguishable.

3.2 The Classic Shuffle Model

This part reviews the classic single-/multi-message shuffle model with one single round.

Single-message shuffle model. Following the conventions of the randomize-then-shuffle model [9, 22], we define a single-message protocol $\mathcal{P}$ as a list of algorithms $\mathcal{P} = (\{\mathcal{R}_i\}_{i\in[n]}, \mathcal{A})$, where $\mathcal{R}_i : \mathbb{X} \to \mathbb{Y}$ is user i's local randomizer, and $\mathcal{A} : \mathbb{Y}^n \to \mathbb{Z}$ is the analyzer on the data collector's side. We refer to $\mathbb{Y}$ as the protocol's *message space* and $\mathbb{Z}$ as the *output space*. The overall protocol implements a mechanism $\mathcal{P} : \mathbb{X}^n \to \mathbb{Z}$ as follows. User i holds a data record x_i and a local randomizer $\mathcal{R}_i$, then computes a message $y_i = \mathcal{R}_i(x_i)$. The messages $y_1, \ldots, y_n$ are shuffled and submitted to the analyzer. We denote the shuffling step as $\mathcal{S}(y_1, \ldots, y_n)$, where $\mathcal{S} : \mathbb{Y}^n \to \mathbb{Y}^n$ is a *shuffler* that applies a uniform-random permutation to its inputs. In summary, the output of $\mathcal{P}(x_1, \ldots, x_n)$ is represented by $\mathcal{A} \circ \mathcal{S} \circ \mathcal{R}_{[n]}(X) = \mathcal{A}(\mathcal{S}(\mathcal{R}_1(x_1), \ldots, \mathcal{R}_n(x_n)))$.

Multi-message shuffle model. In contrast to sending a single message, the multi-message shuffle model allows each user to release multiple messages to the shuffler. The $\mathcal{R}_i : \mathbb{X} \to \mathbb{Y}^*$ is user i's local randomizer. The output $\mathcal{P}(x_1, \ldots, x_n)$ of the overall protocol is $\mathcal{A} \circ \mathcal{S} \circ \mathcal{R}_{[n]}(X) = \mathcal{A}(\mathcal{S}(\mathcal{R}_1(x_1) \cup \mathcal{R}_2(x_2) \cup \cdots \cup \mathcal{R}_n(x_n)))$.

The shuffle model strives to ensure the privacy of $\mathcal{P}(x_1, \ldots, x_n)$ for any analyzer $\mathcal{A}$. Owing to the post-processing property of Hockey-stick divergence, guaranteeing that the shuffled messages $\mathcal{S} \circ \mathcal{R}_{[n]}(X)$ exhibit differential privacy suffices. We formally delineate differential privacy in the shuffle model in Definition 3.4.

Definition 3.4 (Differential privacy in the shuffle model). A protocol $\mathcal{P} = (\{\mathcal{R}_i\}_{i\in[n]}, \mathcal{P})$ satisfies (ϵ, δ)-differential privacy in the shuffle model iff for all neighboring datasets X and X', the $\mathcal{S} \circ \mathcal{R}_{[n]}(X)$ and $\mathcal{S} \circ \mathcal{R}_{[n]}(X')$ are (ϵ, δ)-indistinguishable.

3.3 Variants of the Shuffle Model

Several variants of shuffle model exist in the literature, depending on the organization of users to respond to adaptive queries.

3.3.1 Shuffle-then-randomize Model. We revisit the ideal (single-message) shuffle model based on shuffle-then-randomize [39, 42, 43]. Given an input dataset $X = \{x_1, \ldots, x_n\}$, a uniform-random permutation $\pi : [n] \mapsto [n]$ is first applied to obtain $\mathcal{S}(X) = \{x_{\pi^{-1}(1)}, ..., x_{\pi^{-1}(n)}\}$, followed by a series of adaptive randomizers $\{\mathcal{R}_i\}_{i\in[n]}$. The i-th randomizer $\mathcal{R}_i$ takes the i-th datum $x_{\pi^{-1}(i)}$ in $\mathcal{S}(X)$ and the previous $i-1$ randomization results as input (see Definition 3.5). Since π is not revealed to the server, each user's message can hide among all n messages. Specifically, when every $\mathcal{R}_i$ satisfies ϵ_0-LDP (not unnecessarily identical), the messages $\mathcal{R}_{[n]} \circ \mathcal{S}(X)$ satisfy $(\tilde{O}(e^{\epsilon_0/2}/\sqrt{n}), \delta)$-DP [42, 43], and privacy amplification increases with the number of users n. The model serves as an ideal one that supports fully adaptive queries while providing the strongest privacy amplification effects (amplified by n users).

Definition 3.5 (Shuffle-then-randomize model [39, 42, 43]). Let $\mathcal{R}_i : \mathbb{Z}_0 \times \mathbb{Y}_1 \times \cdots \times \mathbb{Y}_{i-1} \times \mathbb{X} \to \mathbb{Y}_i$ for $i \in [n]$ (where $\mathbb{Y}_i$ is the range space of $\mathcal{R}_i$) be a sequence of algorithms such that $\mathcal{R}_i(z_{[0:i-1]}, \cdot)$ is the i-th local randomizer, where $\mathbb{Z}_0$ denotes the range space of global information and $z_0 \in \mathbb{Z}_0$ is the concrete global information. A protocol $\mathcal{P}_{s\text{-}r} : \mathbb{Z}_0 \times \mathbb{X}^n \to \mathbb{Y}_0 \times \cdots \times \mathbb{Y}_n$ in the shuffle-then-randomize model proceeds as follows: given a dataset $x_{[n]} \in \mathbb{X}^n$, it samples a permutation π uniformly at random and then sequentially computes $z_i = \mathcal{R}_i(z_{[0:i-1]}, x_{\pi^{-1}(i)})$ for $i \in [n]$ before outputting $z_{[0:n]}$.

3.3.2 Divide-randomize-shuffle Model. In decentralized settings, a more realistic approach is letting the analyzer divide users into multiple groups and employ each group for one query sequentially, see Definition 3.6. This model is adopted by many existing works, such as for federated learning [20, 101], and for online bandit/reinforcement learning [24, 46, 64, 68, 88]. Since the division is known to the analyzer (i.e., a potential adversary), a user $i \in U_k$ can only hide among users in the same group U_k, thus privacy can only be amplified by $|U_k|$ (see analyses in Appendix A). This is considerably weaker than the ideal shuffle model.

Definition 3.6 (Divide-randomize-shuffle model [20, 68, 88]). Let $U_{[K]}$ denote a division of set $[n]$ such that $U_k \cap U_{k'} = \emptyset$ for all $k, k' \in [K]$ that $k \neq k'$, and $U_1 \cup \cdots \cup U_K = [n]$. Let $\mathcal{R}_{(k)} : \mathbb{Z}_0 \times \mathbb{Z}_1 \times \cdots \times \mathbb{Z}_{k-1} \times \mathbb{X} \to \mathbb{Y}_{(k)}$ denote the randomizer in the k-th round, where $\mathbb{Z}_k = \mathbb{Y}_{(k)}^{|U_k|}$ and $\mathbb{Y}_{(k)}$ is the range space of $\mathcal{R}_{(k)}$. The $z_{(0)} \in \mathbb{Z}_0$ is global information. A protocol $\mathcal{P}_{d\text{-}r\text{-}s} : \mathbb{Z}_0 \times \mathbb{X}^n \to \mathbb{Z}_0 \times \cdots \times \mathbb{Z}_K$ in the divide-randomize-shuffle model operates as follows: given a dataset $x_{[n]} \in \mathbb{X}^n$, it sequentially computes $z_{(i)} = \mathcal{S}(\mathcal{R}_{(k)}(z_{(0:k-1)}, x_i)_{i\in U_k})$ and outputs $z_{(0)}, z_{(1)}, \ldots, z_{(K)}$ along with $U_{[K]}$.

3.3.3 Subsample-randomize-shuffle Model. Rather than letting the analyzer (i.e., a potential adversary) to perform user division, a more secure and efficient method involves having the shuffler (or another non-colluding party) randomly sample a subpopulation of users in each round independently, as illustrated in Definition 3.7. This approach is also commonly employed for adaptive queries in the shuffle model, as seen in [50, 51].

Definition 3.7 (Subsample-randomize-shuffle model [50, 51]). Let $\{U_k\}_{k\in[K]}$ denote a list of subsets that each has size s and is uniform-randomly sampled from $[n]$ without replacement. Let $\mathcal{R}_{(k)} : \mathbb{Z}_0 \times \mathbb{Z}_1 \times \cdots \times \mathbb{Z}_{k-1} \times \mathbb{X} \to \mathbb{Y}_{(k)}$ denote the randomizer in the k-th, where $\mathbb{Z}_k = \mathbb{Y}_{(k)}^{|U_k|}$ and $\mathbb{Y}_{(k)}$ is the range space of $\mathcal{R}_{(k)}$. The $z_{(0)} \in \mathbb{Z}_0$ is global information. A protocol $\mathcal{P}_{d\text{-}r\text{-}s} : \mathbb{Z}_0 \times \mathbb{X}^n \to \mathbb{Z}_0 \times \cdots \times \mathbb{Z}_K$ in the subsample-randomize-shuffle model operates

as follows: given a dataset $x_{1:n} \in \mathbb{X}^n$, it independently samples $\{U_k\}_{k\in[K]}$ as described previously, then sequentially computes $z_{(i)} = \mathcal{S}(\mathcal{R}_{(k)}(z_{(0:k-1)}, x_i)_{i\in U_k})$ and outputs $z_{(0)}, \ldots, z_{(K)}$.

Given that only a relatively small batch of users, with a size of s, is randomly selected for each round, privacy is amplified by subsampling [8]. Assuming that local randomizers satisfy ϵ_0-LDP, and when combined with privacy amplification via shuffling and the advanced composition theorem of differential privacy, the overall privacy loss is $(\tilde{O}(\frac{s}{n}\sqrt{Ke^{\epsilon_0}/s}), \delta)$ [50, 51].

3.4 Privacy Amplification via Shuffling

Privacy amplification of single-message protocols. In the single-message model, local randomizers often satisfy ϵ_0-LDP, latest results show that the shuffled messages $\mathcal{S} \circ \mathcal{R}_{[n]}(X)$ satisfy $(\tilde{O}(\sqrt{e^{\epsilon_0}/n}, \delta)$-DP [42, 43], where the number of users/messages n that a victim can hide among play a vital role.

Privacy amplification of multi-message protocols. Based on the correlation of local messages, multi-message protocols can be classified into two categories: correlated ones [10, 47–49] and non-correlated ones [6, 7, 23, 66, 69]. In non-correlated protocols, each user i sends a message set $\mathcal{R}_i(x_i)$, in which blanket messages are typically uniformly sampled from the message space. Consequently, non-correlated protocols resemble single-message protocols, allowing for privacy amplification across multiple rounds (see Section 6.3.3). In correlated multi-message protocols, the local messages from each user are carefully designed with dependence, and designated privacy analyses are conducted accordingly.

4 SIDE-CHANNEL ATTACKS IN THE SHUFFLE MODEL

Privacy amplification in the shuffle model relies on various security assumptions. In decentralized settings, one challenging assumption is that privacy adversaries lack knowledge of the message subset contributed by the victim. We show that such information can be readily divulged to privacy adversaries via side-channel information, thereby undermining privacy guarantees.

4.1 Threat model of side-channel attacks

We assume that the internal shuffling process $\mathcal{S}$ is perfectly secure, with privacy adversaries only able to observe the victim users' communication activities (with the shuffler or the analyzer) and the output of the shuffle model $\mathcal{P}(X)$. Typically, every user in the shuffle model encrypts their messages $\mathcal{R}_i(x_i)$ using the analyzer's public key before sending them to the shuffler. Privacy adversaries do not have access to the values of these encrypted messages, but only to their side-channel information, such as communication timing, time duration, and the number of messages.

The aim of privacy adversaries is to deduce the secret value x_i of victim users, utilizing either the success probability of inferring x_i or the differential privacy loss of x_i as a metric.

4.2 Categories of Side-channel Information

In the shuffle models described above, we define the following three types of side-channel information about the victim user v ($v \in [n]$):

I. (*In-out information*) Privacy adversaries know whether or not the victim user participated in one or more round(s) of the shuffle protocol. We denote this information as $io_v \in \{0, 1, *\}^K$, where the k-th value $io_v(k)$ is 1 if privacy adversaries know the victim user participated in query k; the value is 0 if privacy adversaries know the victim user did not participate in query k; the value is $*$ if privacy adversaries are unaware of the participation information in the round.

II. (*Message-cardinality information*) In a multi-message protocol, privacy adversaries might know the number of messages from the victim. We denote this information as $num(\mathcal{R}_v(x_v))$, representing the number of messages outputted by $\mathcal{R}_v(x_v)$.

III. (*Message-length information*) Privacy adversaries know the length of the victim user's message. We denote this information as $len(\mathcal{R}_v(x_v))$, representing the number of bits in $\mathcal{R}_v(x_v)$. In a multi-message protocol, privacy adversaries might know the message-length of every message from $\mathcal{R}_v(x_v)$.

We argue that in decentralized settings, such as mobile computing and wireless networks, privacy adversaries can easily infer in-out, message-cardinality, and message-length information. The in-out information is strongly correlated with the communication patterns of the victim user, which are almost public information in wireless networks or cellular networks [5, 90]. Privacy adversaries can sniff the number of packets sent from the victim to the shuffler through network activities and by examining packet headers, even when packets are delivered over the prevalent HTTPS [78]. Furthermore, the length of TCP/IP packets from the victim can be monitored by sniffing network traffic [25, 75, 82]. Relatively, the in-out information is easier to be inferred. The side-channel information about message number or length may often be imprecise since modern wireless or cellular networking protocols employ techniques such as packet padding and packet slicing. However, from the defender's perspective, we assume that the exact information is at risk of being attacked.

Note that in-out information is especially vulnerable in interactive queries with substantial download overheads, such as federated machine learning. Users download the latest model parameters only when participating in a round. The correlated heavy-loaded downloading traffic, along with uploading traffic to the shuffler, significantly increases the risk of in-out information exposure.

4.3 Privacy under Message-cardinality Attacks

Message-cardinality attacks occur in multi-message protocols, where the number of messages a user releases might depend on the true value the user holds (e.g., in [6, 19, 20, 47–49]). When privacy adversaries observe the message cardinality of a victim, significant privacy leakage may occur due to this side-channel information. In the following theorem, we define the (local) privacy loss resulting from message-cardinality information.

Definition 4.1 (Local privacy loss of message cardinality). For any local randomizer $\mathcal{R} : \mathbb{X} \mapsto \mathbb{Y}^*$, let num_x denote the random variable indicating number of messages in $\mathcal{R}(x)$ given an input $x \in \mathbb{X}$, then the privacy loss of message cardinality is (ϵ, δ) if:

$$D_{e^\epsilon}(num_x \parallel num_{x'}) \geq \delta$$

holds for some $x, x' \in \mathbb{X}$.

Algorithm 1: Δ-Summation Randomizer [49]

if $x_i \neq 0$ **then**
  Send x_i
Sample $z_i^{+1}, z_i^{-1} \sim \mathsf{NB}(1, e^{-(1-\gamma)\epsilon})_{/n}$
Send z_i^{+1} copies of +1, and z_i^{-1} copies of −1
for $\mathbf{s} \in S$ **do**
  Sample $z_i^{\mathbf{s}} \sim \mathsf{NB}(3(1+\log(2/\delta)), e^{-0.1\min(1,\gamma\epsilon)/4})_{/n}$
  for $m \in \mathbf{s}$ **do**
    Send $z_i^{\mathbf{s}}$ copies of m

Algorithm 2: Δ-Summation Analyzer [49]

$T \leftarrow$ multiset of messages received
return $\sum_{y \in T} y$

we now analyze a state-of-the-art multi-message Δ-summation protocol [49], which satisfies (ϵ, δ)-DP in the shuffle model with a mean squared error less than $Var[Laplace(0, \Delta/(\epsilon(1-\gamma))]$ and averagely sends $1 + \tilde{O}(\log(1/\delta)/\sqrt{n})$ messages per user. In this protocol, each user holds a value $x_i \in [0, \Delta]$ and releases several copies of $-\Delta, \ldots, -1, +1, \ldots, \Delta$ (see Algorithm 1); the analyzer simply adds up all messages to get an unbiased estimation of $\sum_{i\in[n]} x_i$ (see Algorithm 2). Considering the case $\Delta = 1$, which is a fundamental building block in various applications, the zero-sum messages collection simply contains one message set: $S = \{\{-1, 1\}\}$. Therefore, when $x_i = 0$, the number of +1s follows $Z^1 + Z^3$ and the number of −1s follows $Z^2 + Z^3$, where $Z^1 \sim \mathsf{NB}(1, e^{-(1-\gamma)\epsilon})_{/n}$, $Z^2 \sim \mathsf{NB}(1, e^{-(1-\gamma)\epsilon})_{/n}$, and $Z^3 \sim \mathsf{NB}(3(1+\log(2/\delta)), e^{-0.1\min(1,\gamma\epsilon)/4})_{/n}$ and $*_{/n}$ meaning the noise is decomposed into n parts. As comparison, when $x_i = 1$, the number of +1s follows variable $1 + Z^1 + Z^3$ and the number of −1s follows variable $Z^2 + Z^3$. In summary, depending on the private information x_i, the number of messages of $\mathcal{R}(x_i)$ is either $Z^1 + Z^2 + 2Z^3$ or $1 + Z^1 + Z^2 + 2Z^3$.

According to Definition 4.1, the local privacy loss due to message-cardinality is then $(+\infty, 0)$, as $\frac{\mathbb{P}[Z^1+Z^2+2Z^3=0]}{\mathbb{P}[1+Z^1+Z^2+2Z^3=0]} = +\infty$. In Figure 2(a), we plot the probability mass distributions of $num_0 = \mathbb{P}[Z^1+Z^2+2Z^3 = m]$, $num_1 = \mathbb{P}[Z^1+Z^2+2Z^3 = m]$ under a typical setting: $n = 10^4$, $\epsilon = 1.0$, $\delta = 0.01/n$, and $\gamma = 0.1$. It is observed that $\mathbb{P}[num_b = b] \geq 95.1\%$ for $b \in \{0, 1\}$, meaning that message cardinality num_{x_v} nearly expose x_v with certain. In Figure 2(b), we plot two corresponding privacy curves: $D_{e^\epsilon}(Z^1+Z^2+2Z^3 \| 1+Z^1+Z^2+2Z^3)$, and $D_{e^\epsilon}(1+Z^1+Z^2+2Z^3 \| Z^1+Z^2+2Z^3)$. These numerical results demonstrate that message-cardinality information completely ruins privacy protection, as adversaries can infer the true value $x_i \in \{0, 1\}$ almost with certainty from the side-channel information. Similar conclusion holds for multi-message protocols in [6, 19, 20, 47, 48].

4.4 Privacy under In-out Attacks

This section focuses on the privacy degradation of shuffle models under in-out attacks. We start with the ideal shuffle model:

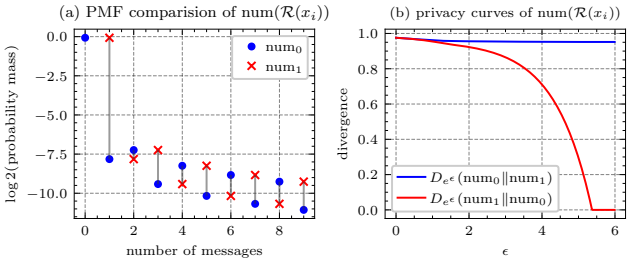

Figure 2: The probability mass distributions of message numbers $num_0 = num(\mathcal{R}(0))$, $num_1 = num(\mathcal{R}(1))$ and privacy curves due to message-cardinality information in the multi-message summation protocol [49].

shuffle-then-randomize, and then give results about the prevalent subsample-randomize-shuffle model.

Shuffle-then-randomize model under in-out attacks [no privacy amplification]. The shuffle-then-randomize model, introduced by the seminal work [39] and refined by recent studies [42, 43], provides the best amplification guarantees to date. In a typical setting from [42, 43], $n = 10,000$ users contribute binary data $x_i \in \{0, 1\}$ using randomized response [99], aiming to achieve a relatively rigid $(\epsilon = 0.2, \delta = 10^{-6})$-DP, such as for surveys on sensitive topics. The amplification guarantee from [42] recommends each user to utilize a budget $\epsilon_0 = 2.29$ and flip x_i with probability $p = \frac{1}{e^{\epsilon_0}+1} \approx 0.092$ as follows (i.e., lie with probability p):

$$\mathcal{R}_i(x_i) = \begin{cases} x_i, & \text{with probability } 1-p; \\ 1-x_i, & \text{with probability } p. \end{cases}$$

The latest amplification guarantee [43] suggests each user to use budget $\epsilon_0 = 2.81$ and flip x_i with probability $p \approx 0.057$. In a decentralized setting, if a victim user's in-out information io_v is leaked, privacy adversaries can link the victim to the k_v-th output message z_{k_v} from the protocol, where k_v denotes the non-zero index in io_v. Consequently, the victim can no longer hide their message among a crowd, as privacy adversaries directly observe $z_{k_v} = \mathcal{R}_v(x_v)$, leaving x_v only protected by the local randomizer $\mathcal{R}_v$ (i.e., no privacy amplification remains). Given the low flip probability, when in-out attacks occur, the victim cannot plausibly deny their answer, as the message $\mathcal{R}_v(x_v)$ reveals the secret x_v with high probability $1 - p$ (e.g., $1 - p \geq 90\%$ in the examples). Similar conclusions apply to binary randomized response used in [39] for continual/streaming data aggregation, generalized randomized response used by [42, 43] for histogram estimation, or any other local randomizers in the shuffle-then-randomize model.

Subsample-randomize-shuffle model under in-out attacks [intra-batch amplification]. The subsample-randomize-shuffle model, commonly used for sequentially adaptive queries (e.g., federated learning [50, 51]), is highly susceptible to in-out information leakage due to significant download and upload traffic. When adversaries possess the victim user's in-out information io_v, they can identify $\mathcal{R}_v(x_v)$ in the k_v-th shuffled messages $z(k_v)$, where k_v denotes the user's participation round. The victim's message can only hide among a smaller sub-population, eliminating privacy amplification via subsampling [8]. Considering an experimental setting from [51] for MNIST image classification, $s = 1000$ users

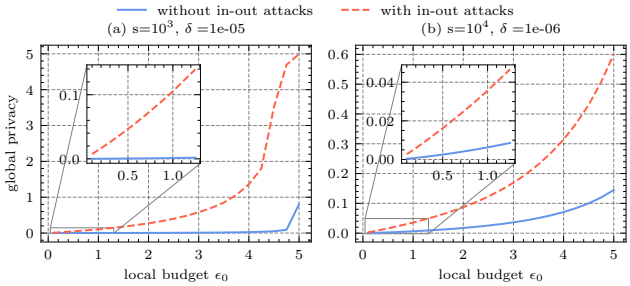

Figure 3: The damage of in-out attack on privacy amplification of the single-message protocol [51] in subsample-randomize-shuffle model.

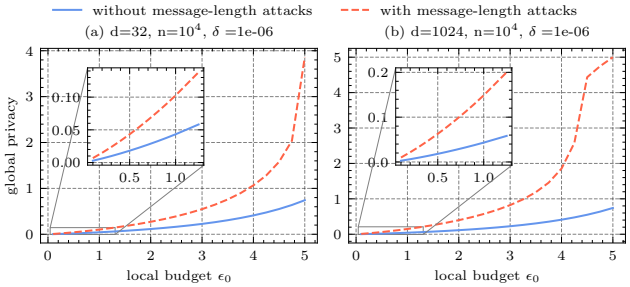

Figure 4: The damage of message length attack on privacy amplification of the single-message protocol [39].

are selected from a total $n = 60000$ population for the k_v-th round, and each user holds a gradient vector $x_i \in \{-0.01, 0.01\}^d$, where $d = 13170$. The vectors are sanitized using randomized response on a randomly selected dimension with budget $\epsilon_0 = 2$. In benign environments, [51] implies that single-round gradient aggregation incurs a privacy loss of $(0.0064, 10^{-5})$-DP (using the near-optimal shuffle amplification upper bound in [43, Theorem 3.1] and subsampling amplification bound in [51, Lemma 3]). However, when adversaries possess in-out information, the victim's privacy loss increases by over 41 times to $(0.27, 10^{-5})$ (using the shuffle amplification lower bound for randomized response in [91, Theorem 5.1]). Figure 3 presents more privacy loss comparisons with and without in-out attacks. For a worst-case user v selected for $\Omega(K)$ rounds, the accumulated privacy loss is $\Omega(\sqrt{Ke^{\epsilon_0}}/s)$, creating a gap of s/n from the preconfigured privacy loss in benign settings.

4.5 Privacy under Message-length Attacks

This section studies the impact of message-length leakage on privacy in the shuffle model, analyzing both local privacy and shuffle privacy amplification impacts.

Privacy Amplification Degradation. Intuitively, if privacy adversaries know the victim's message length $\mathcal{R}_i(x_i)$, they can filter out messages in $\mathcal{S} \circ \mathcal{R}_{[n]}(X)$ with unmatched lengths. Thus, the victim's message can only hide among a subset users with the same length. Specifically, message length information reduces the population size for privacy amplification to:

$$\#\{\mathcal{R}_i(x_i) \mid i \in [n_k] \backslash \{v\} \text{ and } len(\mathcal{R}_i(x_i)) = len(\mathcal{R}_v(x_v))\}. \quad (1)$$

For instance, the seminal work [39] on shuffle privacy amplification considers longitude data $x_i \in \{0, 1\}^d$ aggregation over a period of time $[1, d]$. To avoid $\Theta(d)$ errors in estimators, a common practice is representing x_i by its hierarchical residue and having users report one hierarchy level. Assuming periods time as $d = 2^H$, the k-th value in the h-th hierarchy is given by $V_{h,k} = x_i(k \cdot 2^h) - x_i((k-1) \cdot 2^h)$, where $h \in [0, H-1]$, $k \in [1, d/2^h]$. The study lets users uniformly select one hierarchy level $h \in [0, H-1]$ and employ binary randomized response with full ϵ_0 to sanitize ternary vector $V_{h,*} \in \{-1, 0, 1\}^{d/2^h}$, obtaining a message $Y_{h,*} \in \{-1, 1\}^{d/2^h}$. Depending on the hierarchy level h selected, the message length of $Y_{h,*}$ varies significantly, ranging from 2 to d. Consequently, the victim can only hide among approximately $n/\log_2 d$ users. In Figure 4, we compare remaining privacy protection under message-length attacks with claimed privacy levels in [39], varying d and local budget ϵ_0. The actual privacy loss increases by over 200%, and the gap grows with d. Similar issues exist in other prevalent aggregation tasks (e.g., for range queries [27] and marginal queries [26]) employing sampling-based query selection or compressed binary randomized response [35]. In these protocols, while $len(\mathcal{R}_i(x_i))$ follows the same distribution P_{len} for any input data x_i, message length may vary widely with high entropy. When adversaries possess the message length $len(\mathcal{R}_v(x_v)) = l$ of the victim user v, the user can only hide among a much smaller randomized sub-population (of size $|U_{k,l}| \approx 1 + (n-1) \cdot P_{len}[l]$).

Local Privacy Loss. In some commonly-used local randomizers (e.g., RAPPOR [40], set-valued data randomizer [81], key-value data randomizer [55], high-dimensional gradient randomizers [**?**]) combined with practical techniques (e.g., list representation of sparse vector [20], data compression), severe local privacy loss may occur due to message-length. The message length might probabilistically reveal information about the message value. For instance, in RAPPOR, the secret value x_i is hashed into a Bloom filter of length b using h hash functions $\{H_j\}_{j \in [h]}$, and each bit is randomly flipped with probability $p \in [0, 0.5]$. For transmission efficiency, the randomized Bloom filter is compressed, but the compressed message length leaks information on the number of 0s and 1s. Consequently, given different input values x_i, the number of collisions among hash functions may differ, allowing adversaries to infer the secret value from the message length with negligible success probability. Despite the randomized Bloom filter satisfying $\epsilon_0 = 2m \log((1-p)/p)$-LDP, the message length in the shuffle model may still incur local privacy loss. The privacy budget ϵ_0 is often large, making message length information highly sensitive. We define local privacy loss due to message-length information in Definition 4.2. In RAPPOR, considering $x, x' \in \mathbb{X}$ such that $1 = \#\{H_j(x)\}_{j \in [h]}$ and $h = \#\{H_j(x)\}_{j \in [h]}$, the privacy loss due to message length is $((h-1)\log((1-p)/p), 0) = (\frac{h-1}{2h}\epsilon_0, 0)$, which may be unacceptable in the shuffle model. Similar phenomena arise in local randomizers for set-valued or key-value data, where the number of items or non-empty keys affects output message length.

Definition 4.2 (Local privacy loss due to message length). Considering a local randomizer $\mathcal{R} : \mathbb{X} \mapsto \mathbb{Y}$, let $len(\mathcal{R}(x))$ denote the number of bits of $\mathcal{R}(x)$ for a given input $x \in \mathbb{X}$. The local privacy loss of message length information is (ϵ, δ) if following condition

holds for some $x, x' \in \mathbb{X}$:

$$D_{e^\epsilon}(len(\mathcal{R}(x)) \parallel len(\mathcal{R}(x'))) \geq \delta.$$

Joint In-out and Message-length Attack. Given that both timing of message transmission (in-out information) and the size of transmitted messages (message-length information) are often simultaneously leaked, privacy amplification under joint in-out and message-length attacks can be significantly diminished (e.g., private frequent itemset mining [98], marginal queries [26], data synthesis [103], and decision tree models [44]). Consequently, the pool of users a victim can hide among is further limited to those sharing the same message length in the same round.

5 THEORETICAL JUSTIFICATION OF SIDE-CHANNEL ATTACKS

This section presents formal analyses of the detrimental impact of side-channel information on privacy guarantees in shuffle models. To facilitate theoretical characterization, we introduce some properties of distance measures employed in differential privacy.

The data processing inequality of a privacy measure asserts that the privacy guarantee cannot be weakened by further analysis of a private mechanism's output.

Definition 5.1 (Data processing inequality). A distance measure $D : \Delta(\mathbb{T}) \times \Delta(\mathbb{T}) \to [0, \infty]$ on the space of probability distributions satisfies the data processing inequality if, for all distributions P and Q in $\Delta(\mathbb{T})$ and for all (possibly randomized) functions $g : \mathbb{T} \to \mathbb{T}'$,

$$D(g(P)\|g(Q)) \leq D(P\|Q).$$

In the shuffle model, a plethora of sources of randomness exists (e.g., randomness of users' participation choices and randomness of query results from previous rounds). We put forth two instrumental tools for scrutinizing the distance measures under intricate sources of randomness: separability property (refer to Definition 5.2), and conditioning increasing property (refer to Definition 5.3).

Definition 5.2 (Separability property). A distance measure $D : \Delta(\mathbb{T}) \times \Delta(\mathbb{T}) \to [0, \infty]$ on the space of probability distributions satisfies separability property if, for all distributions P and Q that are joint densities over $\mathbb{T} = \mathbb{T}_1 \times \mathbb{T}_2$ with the same marginal density with respect to $\mathbb{T}_1$, i.e. $P = P_{T_1} \cdot P_{T_2|T_1}$ and $Q = P_{T_1} \cdot Q_{T_2|T_1}$,

$$D(P\|Q) = \mathbb{E}_{t_1 \sim P_{T_1}}[D(P|t_1\|Q|t_1)],$$

where $P|t_1$ and $Q|t_1$ denote the conditional variables $P_{T_2|t_1}$ and $Q_{T_2|t_1}$, respectively.

Definition 5.3 (Conditioning increasing property). A distance measure $D : \Delta(\mathbb{T}) \times \Delta(\mathbb{T}) \to [0, \infty]$ on the space of probability distributions satisfies condition increase property if, for all distributions P and Q over $\mathbb{T}$ that are generated by $P = \int P_{T|T_1} \mathrm{d}T_1$ and $Q = \int Q_{T|t_1} \mathrm{d}T_1$ with the same variable T_1,

$$D(P\|Q) \leq \mathbb{E}_{t_1 \sim P_{T_1}}[D(P|t_1\|Q|t_1)].$$

It is crucial to note that all f-divergence measures, including the Hockey-stick divergence and Rényi divergence, adhere to these three properties [80, Proposition 7.1, Theorem 7.2]. When the distribution of the marginal variable or the randomness source variable T_1 is the same for two variables P and Q, we use $D(P\|Q|t_1)$ to represent $D(P|t_1\|Q|t_1)$ for notational simplicity.

5.1 Privacy Damage of In-out Attacks

In Theorem 5.4, we formally demonstrate that the privacy guarantee of the shuffle-then-randomize model under in-out attack deteriorates to the local level (no privacy amplification). In Theorem 5.5, we reveal that the privacy guarantee of the subsample-randomize-shuffle model under in-out attack degrades to the intra-batch level (amplification by $|U_k| = s$ users). The proofs for these theorems can be found in Appendices C and D, respectively. These two theorems are conditioned on fixed U_k and $io_v(k) = 1$. For the unconditional case, one can use the conditioning increasing property to derive lower bounds (see Appendix J for details). For similar results on other variants of the shuffle model, interested readers are encouraged to refer to Appendix E.

Theorem 5.4 (Privacy amplification of shuffle-then-randomize model under in-out attack). *Given two neighboring datasets $X = \{x_0, \ldots, x_v = a, \ldots, x_n\}, X' = \{x_0, \ldots, x_v = b, \ldots, x_n\} \in \mathbb{X}^n$ that differ at the v-th user data. Let io_v denote the in-out information about user v that $io_v(k) = 1$, then for any distance measure D that satisfies the data processing inequality and the separability property:*

$$\begin{aligned} &D(\mathcal{P}_{s\text{-}r}(X)\|\mathcal{P}_{s\text{-}r}(X')|io_v(k) = 1) \\ \geq &\min_{z_{[0:k-1]} \in \mathbb{Z}_0 \times \cdots \times \mathbb{Y}_{k-1}} D(\mathcal{R}_k(z_{[0:k-1]}, a)\|\mathcal{R}_k(z_{[0:k-1]}, b)). \end{aligned}$$

Theorem 5.5 (Privacy amplification of subsample-randomize-shuffle model under in-out attack). *Given user subsets $\{U_k\}_{k\in[K]}$ sampled as in the subsample-randomize-shuffle model, and two neighboring datasets $X = \{x_0, \ldots, x_v = a, \ldots, x_n\}, X' = \{x_0, \ldots, x_v = b, \ldots, x_n\} \in \mathbb{X}^n$ that differ at the v-th user data. Let io_v denote the in-out information about user i that $io_v(k) = 1$, and let $S = \{X(i)\}_{i\in U_k}, S' = \{X'(i)\}_{i\in U_k} \in \mathbb{X}^{|U_k|}$ denote neighboring datasets w.r.t. U_k, then for any distance measure D that satisfies the data processing inequality:*

$$\begin{aligned} &D(\mathcal{P}_{s\text{-}r\text{-}s}(X)\|\mathcal{P}_{s\text{-}r\text{-}s}(X')|U_k, io_v(k) = 1, z_{(0:k-1)}) \\ \geq &D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0 = z_{(0:k-1)}), \end{aligned}$$

where $\mathcal{R}_{[|U_k|]} = \mathcal{R}_{(k)}$ are the local randomizers in $\mathcal{P}_{s\text{-}r}$ and z_0 is the global information in $\mathcal{P}_{s\text{-}r}$.

5.2 Privacy Damage of Message-length Attacks

Using the divide-randomize-shuffle model as an example, we illustrate the destructive potential of message-length attacks in Theorem 5.6. The victim user can only conceal themselves among users with the same message length in the same division. For the shuffle-then-randomize and subsample-randomize-shuffle models under message-length attacks, the privacy amplification population is limited to $U_{*,l}$, which represents users with identical message lengths across all rounds. When joint in-out and message-length attacks occur, the privacy amplification population further deteriorates to the same level as the divide-randomize-shuffle model: $U_{k,l}$ (refer to Appendix G).

Theorem 5.6 (Privacy amplification of divide-randomize-shuffle model under message-length attack). *Consider non-overlapping and complete user divisions $\{U_k\}_{k\in[K]}$ in the divide-randomize-shuffle model, and two neighboring datasets $X = \{x_0, \ldots, x_v = a, \ldots, x_n\}, X' = \{x_0, \ldots, x_v = b, \ldots, x_n\} \in \mathbb{X}^n$ that differ at the*

v-th user data. Assuming that $v \in U_k$ and $len(\mathcal{R}_{(k)}(z_{(0:k-1)}, a)) \overset{d}{=} len(\mathcal{R}_{(k)}(z_{(0:k-1)}, b))$, let l denote the observed message-length information about user i. Define $U_{k,l}$ as the set of users with the same message length (i.e., $U_{k,l} = \{i \mid for\ i \in U_k \text{ and } len(\mathcal{R}_{(k)}(x_i)) = l\}$), and let $S = \{X(i)\}_{i \in U_{k,l}}, S' = \{X'(i)\}_{i \in U_{k,l}} \in \mathbb{X}^{|U_{k,l}|}$ denote neighboring datasets w.r.t. $U_{k,l}$ then for any distance measure D that satisfies the data processing inequality and the separability property:

$$D(\mathcal{P}_{d\text{-}r\text{-}s}(X) \| \mathcal{P}_{d\text{-}r\text{-}s}(X') | U_{k,l}, len_v = l)$$
$$\in \Big[\min_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0), \max_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0) \Big],$$

where $\mathcal{R}_{[|U_{k,l}|]} = \mathcal{R}_{(k)}$ are the local randomizers in $\mathcal{P}_{s\text{-}r}$ and z_0 is the global information in $\mathcal{P}_{s\text{-}r}$.

It is important to note that when the condition that message lengths are distributionally equal given whether $x_v = a$ or $x_v = b$:

$$len(\mathcal{R}_{(k)}(z_{(0:k-1)}, a)) \overset{d}{=} len(\mathcal{R}_{(k)}(z_{(0:k-1)}, b))$$

does not hold in the theorem, there is also local privacy loss due to message length (see Section 4.5).

Privacy implications of message-cardinality attacks: The message-cardinality information primarily results in local privacy loss, as illustrated in Section 4.3. The effects of this information on the degradation of privacy amplification are intriguing and warrant more sophisticated analyses. We reserve such investigations for future research.

6 DEFENDING AGAINST SIDE-CHANNEL ATTACKS

In this section, we present countermeasures for defending against side-channel attacks. These countermeasures give rise to new principles in the shuffle model that offer robustness to potential attacks and yield stronger privacy amplification effects. Prior to discussing the specifics, we introduce a novel model: the multinomial-randomize-shuffle model, which possesses several advantages in decentralized settings and serves as a foundation for defense against side-channel attacks. For other variants of the shuffle model, the defense techniques are essentially the same.

6.1 Multinomial-randomize-shuffle Model

In this model, each user randomly selects one query k_i from $[K]$ according to a distribution $P_K \in \Delta_K$ and then responds with $\mathcal{R}_{k_i}(z_{(0:k_i-1)}, x_i)$ in the k_i-th round. The shuffler uniformly permutes messages received in the k_i-th round and releases them to the analyzer. The query selection of each user is private and is not exposed other parties. Consequently, this model has similar privacy amplification effects as the ideal model (the privacy amplification population is n, see formal analyses in Appendix B).

Definition 6.1 (Multinomial-randomize-shuffle (MRS) model). Let P_K denote a public probability distribution over $[K]$, and let $k_1, \ldots, k_n$ denote n independent samples following P_K. Define $U_k = \{i \mid for\ i \in [n] \text{ and } k_i = k\}$ as the k-subgroup of users. Let $\mathcal{R}_{(k)} : \mathbb{Z}_0 \times \cdots \times \mathbb{Z}_{k-1} \times \mathbb{X} \to \mathbb{Y}_{(k)}$ denote the randomizer in the k-th round, where $\mathbb{Z}_k = \mathbb{Y}^*_{(k)}$ and $\mathbb{Y}_{(k)}$ is the range space of $\mathcal{R}_{(k)}$. The $z_{(0)} \in \mathbb{Z}_0$ is global information. A protocol $\mathcal{P}_{m\text{-}r\text{-}s} : \mathbb{Z}_0 \times \mathbb{X}^n \to \mathbb{Z}_0 \times \cdots \times \mathbb{Z}_K$ in the MRS model proceeds as follows: given a dataset $x_{[1:n]} \in \mathbb{X}^n$, it samples $k_1, \ldots, k_n \sim P_K$ to obtain $\{U_k\}_{k \in [K]}$, then sequentially compute $z_{(i)} = \mathcal{S}(\{\mathcal{R}_{(k)}(z_{(0:k-1)}, x_i)\}_{i \in U_k})$, finally outputs $z_{(0)}, \ldots, z_{(K)}$ along with the distribution P_K.

The MRS model bears some resemblance to the random check-in approach [11]. However, the random check-in involves complex dummy & uniform-selection operations on received messages in each round and only provides privacy amplified by K users, rendering it significantly weaker than the privacy guarantees of MRS. In benign decentralized settings, the MRS is also preferable to the divide-randomize-shuffle model as it offers greater potential for privacy amplification, yielding asymptotic savings of $(1-\sqrt{1/K}) \cdot 100\%$ in budgets when all local randomizers comply with LDP. The MRS exhibits similar asymptotic privacy consumption behavior as the subsample-randomize-shuffle model. Assuming there are $K \approx \frac{n}{s}$ queries (i.e., one epoch in federated learning), both models consume $(\tilde{O}(\sqrt{e^{\epsilon_0}/n}), \delta)$-DP. Since the MRS ensures each user participates in at most one query while the subsample-randomize-shuffle model has higher sampling variance, the constant factor in privacy loss of the MRS model is smaller. Specifically, under the same settings as Figure 3: $n = 60000$, $s = 1000$, $\epsilon_0 = 2$, and $K = n/s = 60$, the subsample-randomize-shuffle model consumes privacy of $(0.12, 10^{-5})$-DP (using the near-optimal shuffle amplification [43] and tight numerical composition with subsampling [62]), while the MRS model consumes only $(0.036, 10^{-5})$-DP (using [43] as well). Similarly, when $s = 10000$ and $K = n/s = 6$, the subsample-randomize-shuffle model consumes privacy of $(0.116, 10^{-5})$-DP. MRS results in a more than 3× improvements. Moreover, the subsample-randomize-shuffle model relies on additional trust in the shuffler for subsampling, whereas the MRS grants control to the user. We believe the MRS model holds independent interest even under benign environments.

6.2 Defending Against Message-length Attack

In this subsection, we present a countermeasure to defend against potential message-length attacks. Recall that message-length information possessed by adversaries threatens both local privacy and privacy amplification. To avoid local privacy loss, it is necessary to ensure that the message-length distribution is independent of the true value x_i held by user i; to avoid degradation in privacy amplification as shown in Equation 1, it is necessary to ensure that the (sampled) message-length is the same across all users. The latter requirement is stricter as it demands identical concrete message-lengths for all users with various randomizers and true values.

Padding messages. We employ a straightforward approach, message padding, to defend against message-length attacks. Specifically, we let len_{max} denote the maximum possible message length outputted from the original local randomizers across all users:

$$len_{max} = \max_{i \in [n]} \sup_{x \in \mathbb{X}} len(\mathcal{R}_i(x)).$$

Then, for all messages released from each user, the message payload is padded to len_{max} bits.

After implementing the aforementioned padding, the message length becomes global information shared among all parties. Consequently, possessing the victim user's message length information no longer provides an advantage for privacy attacks.

Algorithm 3: Parallel local randomizer [39, 91]

Params: A distribution $P_K : [K] \mapsto [0,1]$, base randomizers $\{\mathcal{M}_k : \mathbb{X} \mapsto \mathbb{Y}_k\}_{k\in[K]}$ each satisfies ϵ_0-LDP.
Input: An input $x \in \mathbb{X}$.
Output: An output that satisfies ϵ_0-LDP.

1 sample $k \sim P_K$
2 $y \leftarrow \mathcal{M}_k(x)$
3 **return** y

Algorithm 4: Rectified parallel local randomizer

Params: A distribution $P_K : [K] \mapsto [0,1]$, base randomizers $\{\mathcal{M}_k : \mathbb{X} \mapsto \mathbb{Y}_k\}_{k\in[K]}$ each satisfies ϵ_0-LDP, the maximum possible length len_{max}.
Input: An input $x \in \mathbb{X}$.
Output: An output that satisfies ϵ_0-LDP.

1 sample $k \sim P_K$
2 $y \leftarrow \mathcal{M}_k(x)$
3 add padding bits to y to form a len_{max}-length bit vector y'
4 encrypts y' with the public key of the analyzer and get $Enc_a(y')$
5 **return** $Enc_a(y')$

6.3 Defending Against In-out Attacks

In this subsection, we present several strategies for defending against in-out attacks. For non-adaptive queries in the single-message shuffle model, there is a simple strategy: *parallelizing queries*; for adaptive queries in the single-message shuffle model, users could sacrifice communication overheads to regain privacy amplification from in-out attacks by sending *dummy messages*.

6.3.1 Non-adaptive queries in the single-message shuffle model. Many data analysis tasks involve multiple non-adaptive estimation queries. In the local model of differential privacy, a common practice for achieving better utility (compared to dividing the privacy budget ϵ_0) is to separate the entire user population into multiple non-overlapping subsets and assign each subset to accomplish one query with the full budget ϵ_0. For example, this approach is used in heavy hitter estimation [12], multi-dimensional data publication [83], frequent itemset mining [98], range queries [27], marginal queries [26], data synthesis [103], and machine learning [44]. This complies with the parallel composition theorem of differential privacy in the central model [36]. In the shuffle model, an (almost) equivalent approach is to have each user randomly choose one query among all K queries with a fixed probability distribution $P_K \in \Delta_K$, and contribute to the chosen query with the full budget [91]. We illustrate this approach in Algorithm 3. Since all mechanisms M_k ($k \in [K]$) are ϵ_0-private, the overall algorithm is ϵ_0-LDP.

Given that every user follows the same distribution P_K, this implies all users are adopting an identical randomization algorithm, which ensures that privacy amplification via shuffling still holds. Denote Algorithm 3 as $\mathcal{R}$, one straightforward conclusion is that:

$$D(\mathcal{S} \circ \mathcal{R}(X) \parallel \mathcal{S} \circ \mathcal{R}(X')) \leq D(\mathcal{P}_{s\text{-}r}(X) \| \mathcal{P}_{s\text{-}r}(X')),$$

where the local randomizers in the shuffle-then-randomize model are all $\mathcal{R}$.

Defend in-out attacks. We emphasize that parallel local randomizers are extremely effective techniques in the shuffle model, especially when facing in-out attacks. As all K queries are performed together in one single round (with the parallel local randomizer in Algorithm 3), the in-out information $io_v = [*, ..., *]$ is trivial, thus maintaining the population size for shuffle privacy amplification at n and leading to better utility.

Defend joint in-out & message-length attacks. We note that message-length information must be protected in the parallel local randomizer, as each base randomizer $\mathcal{M}_k$ often has a different output space $\mathbb{Y}_k$, and the message length leaks information about each query the victim participated in. We present an implementation that resists joint in-out and message-length attacks in Algorithm 4, where messages are padded. Now for the privacy properties of Algorithm 4, since it can be considered as post-processing upon Algorithm 3 and both in-out & message-length information are trivial, it enjoys the same local and shuffle privacy guarantees as the shuffle-then-randomize model (with identical local randomizers).

6.3.2 Adaptive queries in single-message shuffle model. For adaptive queries, parallelizing queries is not applicable since the k-th query relies on previous querying results $z_{(0)}, \ldots, z_{(k-1)}$. We propose contributing dummy messages at every round, at the cost of $(K-1)$-times more communication overhead (see Algorithm 5). We then show how to trade-off communication and amplification.

To obscure the information about which round a user participated in and prevent adversaries from obtaining meaningful in-out information, we let every user contribute message(s) at all K rounds. In the true participating round k_i (sampled as in the normal MRS model), user i contributes a true message $\mathcal{R}(k_i)(z_{(0:k-1)}, x_i)$ (line 10); in the other $K-1$ rounds, user i contributes a dummy/empty message (line 12). To further prevent adversaries from launching message-length attacks, all messages are padded (line 14) and encrypted before transmission.

Since every user participates in all rounds, the in-out information $io_v = [1, ..., 1]$ becomes trivial in the new model presented in Algorithm 5; since all messages have the same length, there is no degradation due to side-channel message-length information. Moreover, the observed shuffled messages $E_1, \ldots, E_K$ can be viewed as adding padded empty messages at every round to the outputs of $\mathcal{P}_{m\text{-}r\text{-}s}$ according to the total number of users n (a public piece of information). Therefore, the new model resists in-out & message-length attacks and has the same privacy amplification effects as the normal MRS model in Theorem B.1.

A practical issue in the above new model is that the total number of messages grows to K times when compared to the original model. This might pose significant communication burdens on users with scarce resources (e.g., in mobile devices or sensors). To provide a flexible balance between communication costs and privacy amplification (data utility), we introduce a new participation model that generalizes the multinomial one, as in Definition 6.1.

Binned participation paradigm. The model first defines a series of bar points $\{b_0, b_1, \ldots, b_M\}$ in $[K+1]$. Specifically, $b_0 \equiv 1$, $b_M \equiv K+1$, and $b_m < b_{m+1}$ for $m \in [M-1]$. The model then defines a set $Q = \{Q_{(1)}, \ldots, Q_{(M)}\}$ of non-overlapping, consecutive, and complete subsets (bins) of $[K]$ where $Q_{(m)} = [b_{m-1}, b_m - 1]$ for $m \in [M]$. It is obvious that $Q_{(m)} \neq \emptyset, Q_{(m)} \cap Q_{(m')} = \emptyset$ holds for all

Algorithm 5: Rectified MRS model.

Params: A probability distribution $P_K : [K] \mapsto [0,1]$, adaptive local randomizers $\{\mathcal{R}_{(k)}\}_{k\in[K]}$, the maximum possible length len_{max}, global information $z_{(0)}$.
Input: Inputs $x_1, \ldots, x_n \in \mathcal{X}$ from n users.
Output: The querying results of K adaptive queries.

▷ Sample participation choices on the user side
for *users* $i \in [n]$ **do**
  sample $k_i \sim P_K$
▷ Run randomization & shuffling
for $k \in [K]$ **do**
  for *users* $i \in [n]$ **do**
    ▷ Randomize, pad, and encrypt on the user side
    if $k = k_i$ **then**
      $y_{i,k} \leftarrow \mathcal{R}_{(k)}(z_{(0:k-1)}, x_i)$
    else
      let $y_{i,k}$ be an empty message
    pad $y_{i,k}$ to form a len_{max}-length bit vector $y'_{i,k}$
    encrypts $y'_{i,k}$ with the analyzer's public key to get $Enc_a(y'_{i,k})$
  ▷ Uniform-randomly permute on the shuffler side
  $E_k = \mathcal{S}(Enc_a(y'_{1,k}), \ldots, Enc_a(y'_{n,k}))$
  ▷ Decrypt and analyze on the server side
  $Y'_k = \{Dec_a(ciphertext) \mid cihpertext \in E_k\}$
  $Y_k = \{y \mid y \in Y'_k \text{ and } y \text{ is not empty}\}$
  $z_{(k)} = \mathcal{P}(Y_k)$
return $z_{(0)}, \ldots, z_{(K)}$

Algorithm 6: Rectified bin-randomize-shuffle model

Params: Participation bins $Q_{(1:M)}$, probability distributions $P_{(m)} : Q_{(m)} \mapsto [0,1]$ for $m \in [M]$, adaptive local randomizers $\{\mathcal{R}_{(k)}\}_{k\in[K]}$, the maximum possible length len_{max}, global information $z_{(0)}$.
Input: Inputs $x_1, \ldots, x_n \in \mathcal{X}$ from n users.
Output: The querying results of K adaptive queries.

▷ Compute participation choices on the user side
for *users* $i \in [n]$ **do**
  choose $m_i \in [M]$ with an arbitrary (personalized) rule
  $k_i \sim P_{(m_i)}$
▷ Run randomization & shuffling
for $k \in [K]$ **do**
  for *users* $i \in [n]$ **do**
    ▷ Randomize, pad, and encrypt on the user side
    if $k \in Q_{(m_i)}$ **then**
      if $k = k_i$ **then**
        $y_{i,k} \leftarrow \mathcal{R}_{(k)}(z_{(0:k-1)}, x_i)$
      else
        let $y_{i,k}$ be an empty message
      pad $y_{i,k}$ to form a len_{max}-length bit vector $y'_{i,k}$
      encrypts $y'_{i,k}$ with the analyzer's public key to get $Enc_a(y'_{i,k})$
  ▷ Uniform-randomly permute on the shuffler side
  $E_k = \mathcal{S}(\{Enc_a(y'_{i,k})\}_{i\in U_{l_{(k)}}})$
  ▷ Decrypt and analyze on the server side
  $Y'_k = \{Dec_a(ciphertext) \mid cihpertext \in E_k\}$
  $Y_k = \{y \mid y \in Y'_k \text{ and } y \text{ is not empty}\}$
  $z_{(k)} = \mathcal{P}(Y_k)$
return $z_{(0)}, \ldots, z_{(K)}$

$m, m' \in [M]$ when $m \neq m'$, and $Q_{(1)} \cup Q_{(2)} \cup \cdots \cup Q_{(M)} = [K]$. For each $Q_{(m)} \in Q$, it is associated with a probability distribution $P_{(m)} : Q_{(m)} \mapsto [0,1] \in \Delta_{|Q_{(m)}|}$. In the binned multinomial participating model, each user $i \in [n]$ select $m_i \in [M]$ with an *arbitrary rule*. For example, if a user becomes available online after round k, the user might select a bin $Q_{(m_i)}$ such that $Q_{(m_i)} \subseteq [k+1:K]$; the user might also randomly select a bin $Q_{(m_i)}$ from Q (with an arbitrary probability distribution). After choosing the bin $Q_{(m_i)}$, the user i samples a true participation round k_i from $Q_{(m_i)}$ according to $P_{(m_i)}$. The corresponding shuffle model with the binned participation paradigm is termed the bin-randomize-shuffle model. Specifically, when $Q = [n]$, the bin-randomize-shuffle model is equivalent to the MRS model.

We let $U_{(m)} \subseteq [n]$ denote the users that selected bin $Q_{(m)}$ in the bin-randomize-shuffle model, $U_k \subseteq [n]$ denote the users that selected query k, and let $l_{(k)}$ denote the bin where k belongs to (i.e., $k \in Q_{(l_{(k)})}$). Then, by combining the defensive techniques in the MRS model, we show that user $i \in U_{(m)}$ can hide among $U_{(m)}$, even under in-out and message-length attacks. We present the overall procedure in Algorithm 6, and formally state the privacy amplification guarantee in Theorem 6.2 (see Appendix I for proof). Since $U_{(m)}$ lies between U_k and $[n]$, the privacy amplification and communication costs can be flexibly traded off.

Theorem 6.2 (Privacy amplification of rectified bin-randomize-shuffle model under in-out and message-length attack). *Given non-overlapping and complete bins $Q = \{Q_{(1)}, \ldots, Q_{(m)}\}$, non-overlapping and complete user divisions $\{U_{(m)}\}_{m\in[M]}$ of the binned-randomize-shuffle model in Algorithm 6, and two neighboring datasets $X = \{x_0, \ldots, x_v = a, \ldots, x_n\}$, $X' = \{x_0, \ldots, x_v = b, \ldots, x_n\} \in \mathbb{X}^n$ that differ at the i-th user data. Let l denote the message-length information about user i that $i \in U_k \subseteq U_{(m)}$, and let $S = \{X(i)\}_{i'\in U_{(m)}}, S' = \{X'(i)\}_{i'\in U_{(m)}} \in \mathbb{X}^{|U_{(m)}|}$ denote neighboring datasets w.r.t. $U_{(m)}$ then for any distance measure D that satisfies the data processing inequality and the separability property:*

$$D(\mathcal{P}_{b\text{-}r\text{-}s}(X) \| \mathcal{P}_{b\text{-}r\text{-}s}(X') | U_{(m)}, len_v = l) \leq \max_{z_0} D(\mathcal{P}_{m\text{-}r\text{-}s}(S) \| \mathcal{P}_{m\text{-}r\text{-}s}(S') | z_0),$$

where $\{\mathcal{R}_{(k')}\}_{k'\in Q_m}$ are the local randomizers in $\mathcal{P}_{m\text{-}r\text{-}s}$ that has $|Q_m|$ rounds, z_0 is the global information in $\mathcal{P}_{m\text{-}r\text{-}s}$ and $P_{(m)}$ is the query selection distribution of the m-th bin.

We note that there are other participation paradigms that might enjoy similar privacy guarantees as the binned multinomial paradigm, such as those with non-consecutive queries in each bin or overlapping bins, among others. We have chosen to focus on such a practical but simplified paradigm to maintain the succinctness of the privacy guarantees in Theorem 6.2.

6.3.3 Adaptive queries in multi-message shuffle model. In some SOTA multi-message protocols (e.g., [6, 7, 23, 66]), each user generates multiple uncorrelated messages, each of which satisfies a

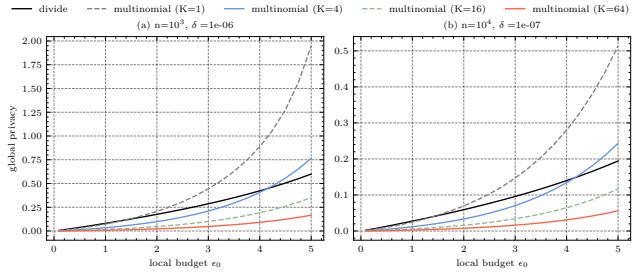

Figure 5: Comparison of amplification effects, on the multi-message protocol [23] (two messages per user as suggested) used for K adaptive queries. In the divide-random-shuffle model, the number of users for amplification is $n = 10^3$ or $n = 10^4$. In the rectified MRS model, the number of virtual users for amplification is $n \cdot K \cdot m$.

certain level of ϵ_0-LDP. Therefore, they act like multiple users in the single-message model. For these protocols, their privacy guarantees under side-channel attacks can be the same as single-message protocols, provided that minor modifications are made.

The basic idea involves having each user in these multi-message protocols act exactly the same way as multiple (independent) users in Algorithm 4, 5, or 6 for non-adaptive/adaptive queries. Taking the recent work [23] for histogram estimation as a concrete example, the original one-round protocol uses binary randomized response (with local budget ϵ_0) to sanitize the secret one-hot vector $x_i \in \{0,1\}^d$ and releases $m-1$ messages, each sanitized in the same way on the zero vector $\{0\}^d$. To precisely mimic m virtual users, one virtual user is supposed to hold the secret one-hot x_i, and others are supposed to hold secret zero vectors $\{0,1\}^d$. Then, m virtual users participate independently in Algorithm 4, 5, or 6 with the K-round protocol of [23]. Assuming that the number of users in the k-th round protocol of [23] is $|U_k|$, since there are $\sum_{k\in[K]} |U_k| \cdot m$ virtual users in the K-round single-message protocol, and one virtual user holds the true secret value v_i, the local privacy ϵ_0 is amplified by $\sum_{k\in[K]} |U_k| \cdot m$ users. In Figure 5, we compare the privacy guarantees of this approach with divide-randomize-shuffle model under varying K (see Appendix H for more detail). As K grows, the privacy amplification effects of our approach become much stronger, saving about 75% of the budget.

We note that such a virtual user transformation is not universally applicable. For instance, the multi-message protocols in [10, 48, 49] achieve near-central accuracy by correlating local messages. This correlation renders the virtual user technique, which relies on message independence, inapplicable.

6.4 Defending Against Message-cardinality Attacks

Section 4.3 highlights that message-cardinality attacks may result in significant privacy loss for some SOTA protocols [6, 47–49]. In this section, we demonstrate the difficulty in securing the Δ-summation protocol [49] against message-cardinality attacks.

Recall that in the Δ-summation protocol (see Algorithms 1 and 2) with $\Delta = 1$, if a user has $x_i = 1$, they send a 1 and a random number of $-1, 1$ messages to the shuffler. Conversely, if a user has $x_i = 0$, they send only a random number of $-1, 1$ messages following the same distribution as when $x_i = 1$. A straightforward remedy would involve sending an extra 0 to the shuffler when $x_i = 0$ to ensure identical message cardinality distributions for both $x_i = 1$ and $x_i = 0$. However, while this addresses the message-cardinality issue, the number of 0s observed by the analyzer (i.e., potential adversaries) in the shuffled messages directly reveals the number of users with 0, leading to an infinite privacy loss. It is worth noting that sending dummy messages (for users with $x_i = 0$) and allowing the shuffler to filter out these dummy messages is also not a viable solution, as it places greater trust in the shuffler and exposes the exact counts to them.

One feasible yet ineffective approach is to let the covering messages collection be $S = \{\{0\}, \{-1, 1\}\}$ and allocate a small budget (e.g., 0.1ϵ, 0.1δ) for the special zero-sum message set $\{0\}$, such as sending an additional $\mathrm{NB}(3(1 + \log(10/\delta)), e^{-0.1\epsilon})$ of 0s per user (there is no $*_{/n}$ decomposition as we aim to protect privacy locally). Then, according to [49, Theorem 10], the local privacy loss due to message-cardinality information is upper bounded by $(0.1\epsilon, 0.1\delta)$. However, this method does not entirely address the message-cardinality issue (non-zero local privacy loss) and substantially increases message complexity. The expected number of messages per user grows by $3(1 + \log(10/\delta))\frac{e^{-0.1\epsilon}}{1-e^{-0.1\epsilon}}$, which can be on the scale of hundreds in typical settings (e.g., sending more than 228 messages per user when $\delta = 10^{-6}$ and $\epsilon = 1$).

Based on the aforementioned attempts, we find that it might be challenging to rectify state-of-the-art protocols [20, 47–49] concerning the message-cardinality issue, and we consider finding a perfect solution (with minimal extra utility/communication costs) for this line of work an open problem. Alternatively, practitioners could employ other available protocols (e.g., in [10, 23, 69]) with the same functionality but without message-cardinality issues.

7 DISCUSSIONS

In this section, we summary new principles for the shuffle model, and account privacy under sequential composition.

7.1 New Principles in the Shuffle Model

To maximize privacy amplification effects and minimize attack risks in the shuffle model, we propose three new principles for the shuffle model:

- **(a) Parallelize queries.** When handling multiple queries with no sequential dependence, the best practice in the shuffle model is to pack them into one parallel query (as described in Sections 6.3.1 and 6.3.3). This allows every user to benefit from the full privacy amplification population without incurring extra communication costs.
- **(b) Pad every message.** As demonstrated in Section 4.5, message-length information can significantly compromise user privacy in both the local and shuffle models. A simple but effective solution is to pad each message to at least the maximum possible message length, len_{max}. Consequently, users in the shuffle model become immune to message-length attacks and avoid degradation of local and shuffle privacy.

(c) **Contribute dummy messages.** The current shuffle model fails to amplify privacy across multiple queries under side-channel attacks when dealing with adaptive queries that have sequential dependence, leading to unfavorable privacy-utility trade-offs. To address this issue, a more effective approach is to have each user contribute extra dummy messages to the shuffler at every round (as shown in Algorithm 5) to allow users to hide among the overall user population. Alternatively, users may adopt the bin-randomize-shuffle model (as shown in Algorithm 6) to flexibly control the trade-off between communication costs, privacy amplification, and utility. This principle can work in conjunction with the first principle by decomposing data analysis tasks into multiple parallel and sequential queries guided by a directed acyclic graph (DAG).

By combining these principles, users can achieve an optimized privacy-utility-communication-security trade-off. Although these principles and new paradigms stem from a side-channel attack setting, they can also be applied to normal benign settings (where padding and dummy messages are unnecessary), resulting in much stronger privacy amplification effects. Specifically, this work achieve full population privacy amplification has been achieved across multiple adaptive queries in single-message randomize-then-shuffle models (in Algorithm 5), and for the first time achieve full population privacy amplification across multiple (non-adaptive or adaptive) queries in the multi-message shuffle model (in Section 6.3.3).

7.2 Sequential Composition

We emphasize that our analyses are conducted under the assumption that each user participates in at most one query (except the subsample-randomize-shuffle model), whether among multiple non-adaptive queries in data mining or multiple adaptive queries within a single epoch of federated learning. In scenarios where users may be involved in multiple queries, such as multiple epochs in federated machine learning, various privacy accounting tools from existing literature can be employed, drawing on our single-participation analyses in Theorems B.1 and 6.2. Examples of these tools include the strong composition theorem [37], Rényi privacy [73], and the Fourier accountant [63, 91, 104].

In line with the updated principles for the shuffle model, we discourage the use of analyzer-, third-party-, or shuffler-conducted user subsampling [1, 8]. This approach assumes a high level of trust in the party conducting the sampling and has a much larger constant factor (due to the variance of subsampling) in privacy loss than the normal MRS model. Instead, we advocate for the adoption of our proposed paradigms, which maximize amplification and minimize risk in each epoch. Consequently, the privacy accounting tools mentioned earlier can be directly applied to analyze the accumulated privacy loss across multiple epochs.

8 CONCLUSION

This study presents a novel identification of communication side-channel attacks associated with the shuffle model of differential privacy. We classify these attacks into three categories: in-out, message-length, and message-cardinality attacks. Additionally, we investigate the resulting degradation of privacy due to these attacks. Our findings reveal that these attacks cause a significant or even infinity increase in privacy loss, potentially nullifying the benefits of privacy amplification through shuffling. To counteract these vulnerabilities, we propose two new variants of the shuffle model: the MRS model and the bin-randomize-shuffle model. We introduce new principles within these models, such as parallelizing queries, padding messages, and sending dummy messages. As a result, the privacy amplification effects are preserved with minimal additional costs, such as sending padding bits and dummy messages. Furthermore, these new models and principles apply not only to defend against attacks but also to normal benign environments, leading to stronger privacy amplification effects than existing models such as the divide-randomize-shuffle model, subsampling-randomize-shuffle model (by constant factor) and random check-in.

Future Research. In light of the severe privacy degradation demonstrated in this study due to side-channel attacks, it is essential to integrate side-channel resistance into the design of shuffle private protocols. Specifically, current SOTA protocols for binary summation [20, 48], counting distinct elements [19] and histogram estimation [49] that achieving central accuracy, and the only available protocol [47] in the literature with $(\epsilon, 0)$-DP guarantee, are susceptible to message-cardinality attacks and not easily repaired. Redesigning these protocols to withstand side-channel attacks represents a promising research avenue. Theoretical questions also arise, such as:

- *Are there $(\epsilon, 0)$-DP protocols in the shuffle model that can resist side-channel attacks?*
- *Are there multi-message protocols that can simultaneously resist side-channel attacks and achieve central accuracy with vanishing message complexity (e.g., $1 + \tilde{O}(1/\sqrt{n})$ messages per user)?*
- *Can non-correlated multi-message protocols achieve central accuracy?*

REFERENCES

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 308–318.

[2] Masayuki Abe. 1998. Universally verifiable mix-net with verification work independent of the number of mix-servers. In *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 437–447.

[3] John M Abowd. 2018. The US Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2867–2867.

[4] Ittai Abraham, Benny Pinkas, and Avishay Yanai. 2020. Blinder–Scalable, Robust Anonymous Committed Broadcast. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. 1233–1252.

[5] John S Atkinson, John E Mitchell, Miguel Rio, and George Matich. 2018. Your WiFi is leaking: What do your mobile apps gossip about you? *Future Generation Computer Systems* 80 (2018), 546–557.

[6] Victor Balcer and Albert Cheu. 2020. Separating Local & Shuffled Differential Privacy via Histograms. In *1st Conference on Information-Theoretic Cryptography*.

[7] Victor Balcer, Albert Cheu, Matthew Joseph, and Jieming Mao. 2021. Connecting robust shuffle privacy and pan-privacy. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2384–2403.

[8] Borja Balle, Gilles Barthe, and Marco Gaboardi. 2018. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in Neural Information Processing Systems* 31 (2018).

[9] Borja Balle, James Bell, Adria Gascón, and Kobbi Nissim. 2019. The privacy blanket of the shuffle model. *CRYPTO* (2019).

[10] Borja Balle, James Bell, Adria Gascon, and Kobbi Nissim. 2020. Private summation in the multi-message shuffle model. *CCS* (2020).

[11] Borja Balle, Peter Kairouz, Brendan McMahan, Om Dipakbhai Thakkar, and Abhradeep Thakurta. 2020. Privacy Amplification via Random Check-Ins. *NeurIPS* (2020).

[12] Raef Bassily, Kobbi Nissim, Uri Stemmer, and Abhradeep Guha Thakurta. 2017. Practical locally private heavy hitters. *Advances in Neural Information Processing Systems* 30 (2017).

[13] Stephanie Bayer and Jens Groth. 2012. Efficient zero-knowledge argument for correctness of a shuffle. In *Advances in Cryptology–EUROCRYPT 2012: 31st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cambridge, UK, April 15-19, 2012. Proceedings 31*. Springer, 263–280.

[14] James Henry Bell, Kallista A Bonawitz, Adrià Gascón, Tancrède Lepoint, and Mariana Raykova. 2020. Secure single-server aggregation with (poly) logarithmic overhead. In *CCS*. ACM.

[15] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, and Bernhard Seefeld. 2017. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th symposium on operating systems principles*. 441–459.

[16] Eric Brier, Christophe Clavier, and Francis Olivier. 2004. Correlation power analysis with a leakage model. In *Cryptographic Hardware and Embedded Systems-CHES 2004: 6th International Workshop Cambridge, MA, USA, August 11-13, 2004. Proceedings 6*. Springer, 16–29.

[17] Xiaoyu Cao, Jinyuan Jia, and Neil Zhenqiang Gong. 2021. Data poisoning attacks to local differential privacy protocols. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*.

[18] David L Chaum. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM* 24, 2 (1981), 84–90.

[19] Lijie Chen, Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. 2021. On Distributed Differential Privacy and Counting Distinct Elements. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[20] Albert Cheu, Matthew Joseph, Jieming Mao, and Binghui Peng. 2022. Shuffle Private Stochastic Convex Optimization. In *International Conference on Learning Representations*. https://openreview.net/forum?id=DrZXuTGg2A-

[21] Albert Cheu, Adam Smith, and Jonathan Ullman. 2021. Manipulation attacks in local differential privacy. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 883–900.

[22] Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. 2019. Distributed differential privacy via shuffling. *EUROCRYPT* (2019).

[23] Albert Cheu and Maxim Zhilyaev. 2022. Differentially private histograms in the shuffle model from fake users. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 440–457.

[24] Sayak Ray Chowdhury and Xingyu Zhou. 2022. Shuffle Private Linear Contextual Bandits. In *International Conference on Machine Learning*. PMLR, 3984–4009.

[25] Mauro Conti, Qian Qian Li, Alberto Maragno, and Riccardo Spolaor. 2018. The dark side (-channel) of mobile devices: A survey on network traffic analysis. *IEEE communications surveys & tutorials* 20, 4 (2018), 2658–2713.

[26] Graham Cormode, Tejas Kulkarni, and Divesh Srivastava. 2018. Marginal release under local differential privacy. In *Proceedings of the 2018 International Conference on Management of Data*. ACM, 131–146.

[27] Graham Cormode, Tejas Kulkarni, and Divesh Srivastava. 2019. Answering range queries under local differential privacy. *Proceedings of the VLDB Endowment* 12, 10 (2019), 1126–1138.

[28] Henry Corrigan-Gibbs and Bryan Ford. 2010. Dissent: accountable anonymous group messaging. In *Proceedings of the 17th ACM conference on Computer and communications security*. 340–350.

[29] Victor Costan and Srinivas Devadas. 2016. Intel SGX explained. *Cryptology ePrint Archive* (2016).

[30] George Danezis and Andrei Serjantov. 2005. Statistical disclosure or intersection attacks on anonymity systems. In *Information Hiding: 6th International Workshop, IH 2004, Toronto, Canada, May 23-25, 2004, Revised Selected Papers 6*. Springer, 293–308.

[31] Jean Paul Degabriele and Martijn Stam. 2018. Untagging Tor: a formal treatment of onion encryption. In *Advances in Cryptology–EUROCRYPT 2018: 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29-May 3, 2018 Proceedings, Part III 37*. Springer, 259–293.

[32] Jing Deng, Richard Han, and Shivakant Mishra. 2005. Countermeasures against traffic analysis attacks in wireless sensor networks. In *First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM'05)*. IEEE, 113–126.

[33] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. 2017. Collecting telemetry data privately. In *NeurIPS*.

[34] Roger Dingledine, Nick Mathewson, and Paul Syverson. 2004. *Tor: The second-generation onion router*. Technical Report. Naval Research Lab Washington DC.

[35] John C Duchi, Michael I Jordan, and Martin J Wainwright. 2013. Local privacy and statistical minimax rates. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*. IEEE, 429–438.

[36] Cynthia Dwork. 2006. Differential privacy. In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33*. Springer, 1–12.

[37] Cynthia Dwork. 2008. Differential privacy: A survey of results. In *Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings 5*. Springer, 1–19.

[38] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. 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, 265–284.

[39] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. 2019. Amplification by shuffling: From local to central differential privacy via anonymity. *SODA* (2019).

[40] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. ACM, 1054–1067.

[41] Saman Feghhi and Douglas J. Leith. 2016. A Web Traffic Analysis Attack Using Only Timing Information. *IEEE Transactions on Information Forensics and Security* 11, 8 (2016), 1747–1759. https://doi.org/10.1109/TIFS.2016.2551203

[42] Vitaly Feldman, Audra McMillan, and Kunal Talwar. 2022. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 954–964.

[43] Vitaly Feldman, Audra McMillan, and Kunal Talwar. 2023. Stronger privacy amplification by shuffling for Rényi and approximate differential privacy. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 4966–4981.

[44] Sam Fletcher and Md Zahidul Islam. 2019. Decision tree classification with differential privacy: A survey. *ACM Computing Surveys (CSUR)* 52, 4 (2019), 1–33.

[45] George H. Forman and John Zahorjan. 1994. The challenges of mobile computing. *Computer* 27, 4 (1994), 38–47.

[46] Evrard Garcelon, Kamalika Chaudhuri, Vianney Perchet, and Matteo Pirotta. 2022. Privacy amplification via shuffling for linear contextual bandits. In *International Conference on Algorithmic Learning Theory*. PMLR, 381–407.

[47] Badih Ghazi, Noah Golowich, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. 2020. Pure Differentially Private Summation from Anonymous Messages. In *1st Conference on Information-Theoretic Cryptography (ITC 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[48] Badih Ghazi, Ravi Kumar, Pasin Manurangsi, and Rasmus Pagh. 2020. Private counting from anonymous messages: Near-optimal accuracy with vanishing communication overhead. In *International Conference on Machine Learning*. PMLR, 3505–3514.

[49] Badih Ghazi, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Amer Sinha. 2021. Differentially private aggregation in the shuffle model: Almost central accuracy in almost a single message. In *International Conference on Machine Learning*. PMLR, 3692–3701.

[50] Antonious Girgis, Deepesh Data, and Suhas Diggavi. 2021. Renyi differential privacy of the subsampled shuffle model in distributed learning. *Advances in Neural Information Processing Systems* 34 (2021), 29181–29192.

[51] Antonious Girgis, Deepesh Data, Suhas Diggavi, Peter Kairouz, and Ananda Theertha Suresh. 2021. Shuffled Model of Differential Privacy in Federated Learning. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2521–2529.

[52] Antonious M Girgis, Deepesh Data, Suhas Diggavi, Ananda Theertha Suresh, and Peter Kairouz. 2021. On the rényi differential privacy of the shuffle model. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 2321–2341.

[53] Antonious M Girgis and Suhas Diggavi. 2023. Multi-Message Shuffled Privacy in Federated Learning. *arXiv preprint arXiv:2302.11152* (2023).

[54] Dov Gordon, Jonathan Katz, Mingyu Liang, and Jiayu Xu. 2022. Spreading the privacy blanket: Differentially oblivious shuffling for differential privacy. In *Applied Cryptography and Network Security: 20th International Conference, ACNS 2022, Rome, Italy, June 20–23, 2022, Proceedings*. Springer, 501–520.

[55] Xiaolan Gu, Ming Li, Yueqiang Cheng, Li Xiong, and Yang Cao. 2020. PCKV: Locally Differentially Private Correlated Key-Value Data Collection with Optimized Utility. *USENIX Security* (2020).

[56] Jacob Imola, Takao Murakami, and Kamalika Chaudhuri. 2022. Differentially Private Triangle and 4-Cycle Counting in the Shuffle Model. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. 1505–1519.

[57] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What can we learn privately? *SIAM J. Comput.* 40, 3 (2011), 793–826.

[58] Paul Kocher, Jann Horn, Anders Fogh, Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, et al. 2020. Spectre attacks: Exploiting speculative execution. *Commun. ACM* 63, 7 (2020), 93–101.

[59] Paul Kocher, Joshua Jaffe, and Benjamin Jun. 1999. Differential power analysis. In *Advances in Cryptology—CRYPTO'99: 19th Annual International Cryptology Conference Santa Barbara, California, USA, August 15–19, 1999 Proceedings 19*. Springer, 388–397.

[60] Jakub Konečnỳ, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. 2016. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527* (2016).

[61] Antti Koskela, Mikko A Heikkilä, and Antti Honkela. 2021. Tight accounting in the shuffle model of differential privacy. *arXiv preprint arXiv:2106.00477* (2021).

[62] Antti Koskela, Mikko A Heikkilä, and Antti Honkela. 2023. Numerical Accounting in the Shuffle Model of Differential Privacy. *Transactions on Machine Learning Research* (2023).

[63] Antti Koskela, Joonas Jälkö, and Antti Honkela. 2020. Computing tight differential privacy guarantees using fft. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2560–2569.

[64] Fengjiao Li, Xingyu Zhou, and Bo Ji. 2022. Differentially private linear bandits with partial distributed feedback. In *2022 20th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*. IEEE, 41–48.

[65] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. 2020. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine* 37, 3 (2020), 50–60.

[66] Xiaochen Li, Weiran Liu, Hanwen Feng, Kunzhe Huang, Yuke Hu, Jinfei Liu, Kui Ren, and Zhan Qin. 2023. Privacy enhancement via dummy points in the shuffle model. *IEEE Transactions on Dependable and Secure Computing* (2023).

[67] Moritz Lipp, Michael Schwarz, Daniel Gruss, Thomas Prescher, Werner Haas, Jann Horn, Stefan Mangard, Paul Kocher, Daniel Genkin, Yuval Yarom, et al. 2020. Meltdown: Reading kernel memory from user space. *Commun. ACM* 63, 6 (2020), 46–56.

[68] Andrew Lowy and Meisam Razaviyayn. 2021. Private federated learning without a trusted server: Optimal algorithms for convex losses. *arXiv preprint arXiv:2106.09779* (2021).

[69] Qiyao Luo, Yilei Wang, and Ke Yi. 2022. Frequency Estimation in the Shuffle Model with Almost a Single Message. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. 2219–2232.

[70] Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. 2008. Privacy: Theory meets practice on the map. In *2008 IEEE 24th international conference on data engineering*. IEEE, 277–286.

[71] Yuyi Mao, Changsheng You, Jun Zhang, Kaibin Huang, and Khaled B Letaief. 2017. A survey on mobile edge computing: The communication perspective. *IEEE communications surveys & tutorials* 19, 4 (2017), 2322–2358.

[72] Ilya Mironov. 2012. On significance of the least significant bits for differential privacy. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 650–661.

[73] Ilya Mironov. 2017. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 263–275.

[74] Steven J Murdoch and George Danezis. 2005. Low-cost traffic analysis of Tor. In *2005 IEEE Symposium on Security and Privacy (S&P'05)*. IEEE, 183–195.

[75] Diala Naboulsi, Marco Fiore, Stephane Ribot, and Razvan Stanica. 2015. Large-scale mobile traffic analysis: a survey. *IEEE Communications Surveys & Tutorials* 18, 1 (2015), 124–161.

[76] Joe Near. 2018. Differential privacy at scale: Uber and berkeley collaboration. In *Enigma 2018 (Enigma 2018)*.

[77] Lasse Overlier and Paul Syverson. 2006. Locating hidden servers. In *2006 IEEE Symposium on Security and Privacy (S&P'06)*. IEEE, 15–pp.

[78] Eva Papadogiannaki and Sotiris Ioannidis. 2021. A survey on encrypted network traffic analysis applications, techniques, and countermeasures. *ACM Computing Surveys (CSUR)* 54, 6 (2021), 1–35.

[79] Sandro Pinto and Nuno Santos. 2019. Demystifying arm trustzone: A comprehensive survey. *ACM computing surveys (CSUR)* 51, 6 (2019), 1–36.

[80] Yury Polyanskiy and Yihong Wu. 2022. *Information Theory: From Coding to Learning*. Cambridge University Press. https://people.lids.mit.edu/yp/homepage/data/itbook-export.pdf.

[81] Zhan Qin, Yin Yang, Ting Yu, Issa Khalil, Xiaokui Xiao, and Kui Ren. 2016. Heavy hitter estimation over set-valued data with local differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 192–203.

[82] Jingjing Ren, Ashwin Rao, Martina Lindorfer, Arnaud Legout, and David Choffnes. 2016. Recon: Revealing and controlling pii leaks in mobile network traffic. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. 361–374.

[83] Xuebin Ren, Chia-Mu Yu, Weiren Yu, Shusen Yang, Xinyu Yang, Julie A McCann, and S Yu Philip. 2018. LoPub: high-dimensional crowdsourced data publication with local differential privacy. *IEEE Transactions on Information Forensics and Security* 13, 9 (2018), 2151–2166.

[84] Mary Scott, Graham Cormode, and Carsten Maple. 2022. Aggregation and Transformation of Vector-Valued Messages in the Shuffle Model of Differential Privacy. *IEEE Transactions on Information Forensics and Security* 17 (2022), 612–627.

[85] Elaine Shi and Ke Wu. 2021. Non-interactive anonymous router. In *Advances in Cryptology–EUROCRYPT 2021: 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17–21, 2021, Proceedings, Part III*. Springer, 489–520.

[86] Yi-Sheng Shiu, Shih Yu Chang, Hsiao-Chun Wu, Scott C-H Huang, and Hsiao-Hwa Chen. 2011. Physical layer security in wireless networks: A tutorial. *IEEE wireless Communications* 18, 2 (2011), 66–74.

[87] Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. 2017. Privacy loss in apple's implementation of differential privacy on macos 10.12. *arXiv preprint arXiv:1709.02753* (2017).

[88] Jay Tenenbaum, Haim Kaplan, Yishay Mansour, and Uri Stemmer. 2021. Differentially private multi-armed bandits in the shuffle model. *Advances in Neural Information Processing Systems* 34 (2021), 24956–24967.

[89] Yinxin Wan, Kuai Xu, Feng Wang, and Guoliang Xue. 2020. Characterizing and mining traffic patterns of IoT devices in edge networks. *IEEE Transactions on Network Science and Engineering* 8, 1 (2020), 89–101.

[90] Qinglong Wang, Amir Yahyavi, Bettina Kemme, and Wenbo He. 2015. I know what you did on your smartphone: Inferring app usage over encrypted data traffic. In *2015 IEEE conference on communications and network security (CNS)*. IEEE, 433–441.

[91] Shaowei Wang. 2023. Privacy Amplification via Shuffling: Unified, Simplified, and Tightened. *arXiv preprint arXiv:2304.05007* (2023).

[92] Shaowei Wang, Jiachun Du, Wei Yang, Xinrong Diao, Zichun Liu, Yiwen Nie, Liusheng Huang, and Hongli Xu. 2019. Aggregating votes with local differential privacy: Usefulness, soundness vs. indistinguishability. *arXiv preprint arXiv:1908.04920* (2019).

[93] Shaowei Wang, Jin Li, Yuntong Li, Wei Yang, and Hongyang Yan. 2023. Differentially Private Numerical Vector Analyses in the Local and Shuffle Model. *arXiv preprint arXiv:2304.04410* (2023).

[94] Shaowei Wang, Jin Li, Yuqiu Qian, Jiachun Du, Wenqing Lin, and Wei Yang. 2021. Hiding Numerical Vectors in Local Private and Shuffled Messages.. In *IJCAI*. 3706–3712.

[95] Shaowei Wang, Xuandi Luo, Yuqiu Qian, Jiachun Du, Wenqing Lin, and Wei Yang. 2022. Analyzing Preference Data With Local Privacy: Optimal Utility and Enhanced Robustness. *IEEE Transactions on Knowledge and Data Engineering* (2022).

[96] Shaowei Wang, Xuandi Luo, Yuqiu Qian, Youwen Zhu, Kongyang Chen, Qi Chen, Bangzhou Xin, and Wei Yang. 2023. Shuffle Differential Private Data Aggregation for Random Population. *IEEE Transactions on Parallel and Distributed*

Systems (2023).

[97] Tianhao Wang, Bolin Ding, Min Xu, Zhicong Huang, Cheng Hong, Jingren Zhou, Ninghui Li, and Somesh Jha. 2020. Improving utility and security of the shuffler-based differential privacy. *Proceedings of the VLDB Endowment* 13, 13 (2020), 3545–3558.

[98] Tianhao Wang, Ninghui Li, and Somesh Jha. 2018. Locally differentially private frequent itemset mining. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 127–143.

[99] Stanley L Warner. 1965. Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* 60, 309 (1965), 63–69.

[100] Yongji Wu, Xiaoyu Cao, Jinyuan Jia, and Neil Zhenqiang Gong. 2022. Poisoning Attacks to Local Differential Privacy Protocols for {Key-Value} Data. In *31st USENIX Security Symposium (USENIX Security 22)*. 519–536.

[101] Shui Yu and Lei Cui. 2022. Anonymous Communication and Shuffle Model in Federated Learning. In *Security and Privacy in Federated Learning*. Springer, 109–114.

[102] Fan Zhang, Wenbo He, and Xue Liu. 2011. Defending against traffic analysis in wireless networks through traffic reshaping. In *2011 31st International Conference on Distributed Computing Systems*. IEEE, 593–602.

[103] Zhikun Zhang, Tianhao Wang, Ninghui Li, Shibo He, and Jiming Chen. 2018. Calm: Consistent adaptive local marginal for marginal release under local differential privacy. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 212–229.

[104] Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. 2022. Optimal accounting of differential privacy via characteristic function. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 4782–4817.

A PRIVACY AMPLIFICATION OF DIVIDE-RANDOMIZE-SHUFFLE MODEL

Theorem A.1 (Privacy amplification of divide-randomize-shuffle model). *Given non-overlapping and complete user divisions $\{U_k\}_{k\in[K]}$, and two neighboring datasets $X = \{x_0, \ldots, x_v = a, \ldots, x_n\}$, $X' = \{x_0, \ldots, x_v = b, \ldots, x_n\} \in \mathbb{X}^n$ that differ at the v-th user data, if $v \in U_k$, then for any distance measure D that satisfies the data processing inequality:*

$$\begin{aligned}&D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{[K]}, z_{(0:k-1)})\\ =&D(\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})\|\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k}))\end{aligned}$$

where the local randomizers in $\mathcal{P}_{s\text{-}r}$ are all $\mathcal{R}_{(k)}$, and the public information in $\mathcal{P}_{s\text{-}r}$ is $z_0 = \{z_{(0)}, \ldots, z_{(k-1)}\}$.

Proof. Let $left$ denote $D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{[K]}, z_{(0:k-1)})$, our proof contains two parts:

$$\begin{aligned}left \le D(\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})\|\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})),\\ left \ge D(\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})\|\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})).\end{aligned}$$

Considering the first part, we define the following post-processing function for the $\mathcal{P}_{s\text{-}r}$ protocol with global information $z_0 = z_{(0:k-1)}$ and with local randomizers $\mathcal{R}_{k'} \equiv \mathcal{R}_{(k)}$ for $k' \in [|U_k|]$:

step (1): The output of $\mathcal{P}_{s\text{-}r}$ given inputs $z_0, \{x_i\}_{i\in U_k}$ is:

$$\{\mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(1)}), \ldots, \mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(|U_k|)})\},$$

where $\pi : U_k \mapsto [|U_k|]$ is a uniform-random permutation sampled by $\mathcal{P}_{s\text{-}r}$. Note that since the local randomizers are identical, altering ordering of the shuffling and randomize operations leads to equivalent output distributions. We denote the output as $z_{(k)}$.

step (2): Compute $z_{(k')} = \mathcal{S}_{(k')}(\{\mathcal{R}_{(k')}(z_{(0:k'-1)}, x_i)\}_{i\in U_{k'}})$ for $k' \in [k+1:K]$ sequentially.

step (3): Return $z_{(0)}, z_{(1)}, \ldots, z_{(K)}$ and $\{U_k\}_{k\in[K]}$.

Then, for the input dataset X where $x_v = a$, the (joint) output distribution of $\mathcal{P}_{d\text{-}r\text{-}s}(X), \{U_{k'}\}_{k'\in[K]}, z_{(0)}, \ldots, z_{(k-1)}$ is equal to the output in the above step (3) with $x_v = a$; for the input dataset X' where $x_v = b$, the output distribution $\mathcal{P}_{d\text{-}r\text{-}s}(X'), \{U_{k'}\}_{k'\in[K]}, z_{(0:k-1)}$ is equal to the output in the above step (3) with $x_v = b$. Based on the data processing property of distance measure D, we then have

$$\begin{aligned}&D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{[K]}, z_{(0:k-1)})\\ \le&D(\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})\|\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})).\end{aligned}$$

Considering the second part, we define a post-processing function for the output of $\mathcal{P}_{d\text{-}r\text{-}s}$ conditional on $\{U_{k'}\}_{k'\in[K]}, z_{(0:k-1)}$:

step (1): Remove $z_{(k+1)}, \ldots, z_{(K)}$ from the output list;

step (2): Return $z_{(0)}, \ldots, z_{(k-1)}$ and $z_{(k)}$.

It is clear that for the input dataset X where $x_v = a$, the output distribution of the above step (2) is equal to the $\mathcal{P}_{s\text{-}r}$ protocol with global information $z_0 = \{z_{(0)}, \ldots, z_{(k-1)}\}$, local randomizers $\mathcal{R}_{k'} \equiv \mathcal{R}_{(k)}$ for $k' \in [|U_k|]$, and inputs $\{x_i\}_{i\in U_k}$ that $x_v = a$; for the input dataset X where $x_v = b$, the output distribution of the above step (2) is equal to the $\mathcal{P}_{s\text{-}r}$ protocol with global information $z_0 = \{z_{(0)}, \ldots, z_{(k-1)}\}$, local randomizers $\mathcal{R}_{k'} \equiv \mathcal{R}_{(k)}$ for $k' \in [|U_k|]$, and inputs $\{x_i\}_{i\in U_k}$ that $x_v = b$. According to the data processing property of distance measure D, we then have:

$$\begin{aligned}&D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{[K]}, z_{(0:k-1)})\\ \ge&D(\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})\|\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})).\end{aligned}$$

Combining the two inequalities, we conclude that the equation holds. □

By amalgamating Theorem A.1 and the separability property of divergence measures, and leveraging the fact that $z_{(0:k-1)}$ are discernible in the output of $\mathcal{P}_{d\text{-}r\text{-}s}$ and maintain distributional equivalence given either X or X' as input, we can effortlessly deduce:

$$\begin{aligned}&D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{[K]})\\ =&\mathbb{E}_{z_{(0:k-1)}} D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{[K]}, z_{(0:k-a)})\\ =&\mathbb{E}_{z_0\sim z_{(0:k-1)}} D(\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})\|\mathcal{P}_{s\text{-}r}(\{x_i\}_{i\in U_k})|z_0).\end{aligned}$$

B PRIVACY AMPLIFICATION OF MULTINOMIAL-RANDOMIZE-SHUFFLE MODEL

Theorem B.1 (Privacy amplification of multinomial-randomize-shuffle model). *Given non-overlapping and complete user divisions $\{U_k\}_{k\in[K]}$ sampled as in the multinomial-randomize-shuffle model, and two neighboring datasets $X = \{x_0, \ldots, x_v = a, \ldots, x_n\}, X' = \{x_0, \ldots, x_v = b, \ldots, x_n\} \in \mathbb{X}^n$ that differ at the v-th user data, then for any distance measure D that satisfies data processing inequality:*

$$\begin{aligned}&D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|\{|U_k|\}_{k\in[K]})\\ \le&D(\mathcal{P}_{s\text{-}r}(X)\|\mathcal{P}_{s\text{-}r}(X'))\end{aligned}$$

where $\mathcal{R}_{[n]} = \mathcal{R}_{(1)}^{|U_1|} \times \cdots \times \mathcal{R}_{(K)}^{|U_K|}$ are the local randomizers in $\mathcal{P}_{s\text{-}r}$ and the global information in $\mathcal{P}_{s\text{-}r}$ is the same as in $\mathcal{P}_{m\text{-}r\text{-}s}$.

Proof. We utilize the data processing inequality to prove the theorem. Specifically, we show there exists a shuffle-then-randomize protocol $\mathcal{P}_{s\text{-}r}$ (where $\mathcal{R}_{[n]} \in \{\mathcal{R}_{(1)}, \ldots, \mathcal{R}_{(K)}\}^n$) and a post-processing function, such that the output from multinomial model (distributionally) equals to post-processed output of $\mathcal{P}_{s\text{-}r}$.

Consider a protocol $\mathcal{P}_{s\text{-}r}$ that $\mathcal{R}_i \equiv \mathcal{R}_{(k)}$ for $i \in [\sum_{k\in[k-1]} |U_k| : \sum_{k'\in[k]} |U_k|]$ and the global information z_0 is the same as in $\mathcal{P}_{m\text{-}r\text{-}s}$. Furthermore, we assume there is no adaptivity within U_k (the $k \in [K]$), meaning that every $\mathcal{R}_i = \mathcal{R}_{(k)}$ are independent from $\{z_{i'}\}_{i'\in U_k \text{ and } i'\le i}$. Let $\mathbf{z} = \{z_0, \ldots, z_n\}$ denote the output variables of such an special algorithm. Since all users follow the same multinomial distribution P_K, and both $\mathcal{P}'_{s\text{-}r}$ and $\mathcal{P}_{m\text{-}r\text{-}s}$ uses uniform-random shuffling, due to the uniformity of all $\{x_1, \ldots, x_n\}$ in the input, we have the output distribution of $\mathbf{z}$ from the $\mathcal{P}_{s\text{-}r}$ equal to the $\mathcal{P}_{m\text{-}r\text{-}s}$ with the same input. Therefore, according to the post-processing inequality (on an identical map), we have the conclusion. □

We examine the conditional cases of the multinomial-randomize-shuffle model in Theorem B.1, wherein the sub-population sizes for all rounds $\{|U_k|\}_{k\in[K]}$ are fixed. Regarding non-conditional case $D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X'))$, one can effortlessly employ the

conditioning increasing property of divergence measures to derive:

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')) \leq \mathop{\mathbb{E}}_{|U_{[K]}|\sim multinomial(n,P_K)} D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|\{|U_k|\}_{k\in[K]}).$$

C PROOF OF THEOREM 5.4 FOR SHUFFLE-THEN-RANDOMIZE MODEL

To prove that:

$$D(\mathcal{P}_{s\text{-}r}(X)\|\mathcal{P}_{s\text{-}r}(X')|io_v(k)=1) \geq \min_{z_{[0:k-1]}\in\mathbb{Z}_0\times\cdots\mathbb{Y}_{k-1}} D(\mathcal{R}_k(z_{[0:k-1]},a)\|\mathcal{R}_k(z_{[0:k-1]},b)),$$

we consider the output of $\mathcal{P}_{s\text{-}r}$ with fixed variables $io_v(k)=1$. We then define a post-processing function:

step (1): Remove the $z_{k'}$ for $k'\in[k+1:K]$ from the output list;
step (2): Return $z_0=z_{[0:k-1]}$ and z_k.

When $z_{[0:k-1]}$ is fixed, given X or X', the output distribution of the above step (2) is equivalent to $\mathcal{R}_k(z_{[0:k-1]},a)$ or $\mathcal{R}_k(z_{[0:k-1]},b)$, respectively. Therefore, according to the data processing inequality, we have:

$$\begin{aligned}&D(\mathcal{P}_{s\text{-}r}(X)\|\mathcal{P}_{s\text{-}r}(X')|io_v(k)=1,z_{[0:k-1]}=z))\\ \geq&D(\mathcal{R}_k(z_{[0:k-1]},a)\|\mathcal{R}_k(z_{[0:k-1]},b)).\end{aligned}$$

Further since X and X' differ only at x_i and user i appears at the k-th round, in two independent runs, $\mathcal{P}_{s\text{-}r}(X)|io_v(k)=1,z_{[0:k-1]}$ and $\mathcal{P}_{s\text{-}r}(X')|io_v(k)=1,z_{[0:k-1]}$, the distributions of $z_{[0:k-1]}$ are identical. We let P_z denote this distribution. Then, using the separability property of the distance measure over observable $z_{[0:k-1]}$ in the shuffle-then-randomize model, we have:

$$\begin{aligned}&D(\mathcal{P}_{s\text{-}r}(X)\|\mathcal{P}_{s\text{-}r}(X')|io_v(k)=1)\\ =&\mathbb{E}_{z\sim P_z}D(\mathcal{P}_{s\text{-}r}(X)\|\mathcal{P}_{s\text{-}r}(X')|io_v(k)=1,z_{[0:k-1]}=z))\\ \geq&\mathbb{E}_{z\sim P_z}D(\mathcal{R}_k(z_{[0:k-1]},a)\|\mathcal{R}_k(z_{[0:k-1]},b))\\ \geq&\min_z D(\mathcal{R}_k(z_{[0:k-1]},a)\|\mathcal{R}_k(z_{[0:k-1]},b)).\end{aligned}$$

D PROOF OF THEOREM 5.5 FOR SUBSAMPLE-RANDOMIZE-SHUFFLE MODEL

To prove that:

$$\begin{aligned}&D(\mathcal{P}_{s\text{-}r\text{-}s}(X)\|\mathcal{P}_{s\text{-}r\text{-}s}(X')|U_k,io_v(k)=1,z_{(0:k-1)})\\ \geq&D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0=z_{(0:k-1)}),\end{aligned}$$

we consider the output of $\mathcal{P}_{s\text{-}r\text{-}s}$ with observed/fixed variables $U_k, io_v(k)=1$ and $z_{(0:k-1)}$. We then define the following post-processing function:

step (1): Remove the $z_{(k')}$ for $k'\in[k+1:K]$ from the output list;
step (2): Return $z_0=z_{(0:k-1)}$ and $z_{(k)}$.

Since $z_{(0:k-1)}$ is fixed, the output distribution of the above step (2) is equivalent to an algorithm $\mathcal{P}_{s-r}$ in the shuffle-randomize model where global information is $z_{(0:k-1)}$. Therefore, according to the data processing inequality, we have:

$$\begin{aligned}&D(\mathcal{P}_{s\text{-}r\text{-}s}(X)\|\mathcal{P}_{s\text{-}r\text{-}s}(X')|U_k,io_v(k)=1,z_{(0:k-1)})\\ \geq&D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0=z_{(0:k-1)}).\end{aligned}$$

E OTHER VARIANTS OF SHUFFLE MODEL UNDER IN-OUT ATTACKS

It is obvious that the divide-randomize-shuffle model is not affected by the in-out attacks, as all users' in-out information are public information in the model. While in the multinomial-randomize-shuffle model, since the victim can only conceal themselves among users participating in the same query, the privacy amplification degrades to a level similar to the divide-randomize-shuffle model. See Theorem E.1 for formal statements.

Theorem E.1 (Privacy amplification of MRS model under in-out attack). *Given non-overlapping and complete user divisions $\{U_k\}_{k\in[K]}$ sampled as in the multinomial-randomize-shuffle model, and two neighboring datasets $X=\{x_0,\ldots,x_v=a,\ldots,x_n\}$, $X'=\{x_0,\ldots,x_v=b,\ldots,x_n\}\in\mathbb{X}^n$ that differ at the v-th user data. Let io_v denote the in-out information about user i that $io_v(k)=1$, and let $S=\{X(i)\}_{i\in U_k}, S'=\{X'(i)\}_{i\in U_k}\in\mathbb{X}^{|U_k|}$ denote neighboring datasets w.r.t. U_k, then for any distance measure D that satisfies the data processing inequality and the separability property:*

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k,io_v(k)=1) \in\Big[\min_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0),\max_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0)\Big],$$

where $\mathcal{R}_{[|U_k|]}=\mathcal{R}_{(k)}$ are the local randomizers in $\mathcal{P}_{s\text{-}r}$ and z_0 is the global information in $\mathcal{P}_{s\text{-}r}$.

Proof. We first establish the privacy amplification lower bound, which indicates the destructive power of in-out information attacks. To prove that:

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k,io_v(k)=1)\geq\min_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')),$$

we consider the output of $\mathcal{P}_{m\text{-}r\text{-}s}$ with observed/fixed variables $U_k, io_v(k)=1$. We then define a post-processing function:

step (1): Remove the $z_{(k')}$ for $k'\in[k+1:K]$ from the output list;
step (2): Return $z_0=z_{(0:k-1)}$ and $z_{(k)}$.

When $z_{(0:k-1)}$ is fixed, the output distribution of the above step (2) is equivalent to an algorithm $\mathcal{P}_{s-r}$ in the shuffle-randomize model where global information is $z_{(0:k-1)}$, thus we have:

$$\begin{aligned}&D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k,io_v(k)=1,z_{(0:k-1)})\\ \geq&D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0=z_{(0:k-1)}).\end{aligned}$$

Since X and X' differ only at x_v and $v\in U_k$, in two independent runs, $\mathcal{P}_{m\text{-}r\text{-}s}(X)|U_k,io_v(k)=1,z_{(0:k-1)}$ and $\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k,io_v(k)=1,z_{(0:k-1)}$, the distributions of $z_{(0:k-1)}$ are identical. We let P_{z_0} denote this distribution. Then, using the separability property of the distance measure over observable $z_{(0:k-1)}$, we have:

$$\begin{aligned}&D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k,io_v(k)=1)\\ =&\mathbb{E}_{z_0\sim P_{z_0}}D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k,io_v(k)=1,z_{(0:k-1)}=z_0))\\ \geq&\mathbb{E}_{z_0\sim P_{z_0}}D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0)\\ \geq&\min_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0).\end{aligned}$$

We now prove the upper bound on privacy amplification, which indicates the remaining amplification effects. Given a fixed $z_0=$

$z_{(0:k-1)}$, we want to show that:

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k, io_v(k)=1, z_{(0:k-1)}) \geq \min_{z_0} D(\mathcal{P}_{s\text{-}r}(S)|z_0\|\mathcal{P}_{s\text{-}r}(S')|z_0).$$

We define the following post-processing function over the output of $\mathcal{P}_{s\text{-}r}$ with local randomizers $\mathcal{R}_{k'} \equiv \mathcal{R}_{(k)}$ for $k' \in [|U_k|]$:

step (1): The output of $\mathcal{P}_{s\text{-}r}$ given input $z_0, \{x_i\}_{i\in U_k}$ is

$$\{\mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(1)}), \ldots, \mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(|U_k|)})\},$$

where $\pi : U_k \mapsto [|U_k|]$ is a uniform-random permutation sampled by $\mathcal{P}_{s\text{-}r}$. We denote this output as $z_{(k)}$.

step (2): Compute $z_{(k')} = \mathcal{S}_{(k')}(\{\mathcal{R}_{(k')}(z_{(0:k'-1)}, x_i)\}_{i\in U_{k'}})$ for $k' \in [k+1:K]$ sequentially.

step (3): Return $z_{(0)}, z_{(1)}, \ldots, z_{(K)}$.

The output distributions of step (3) with $x_v = a$ or $x_v = b$ are equal to the output distributions of $\mathcal{P}_{m\text{-}r\text{-}s}(X)|U_k, io_v(k)=1, z_{(0:k-1)}$ and $\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k, io_v(k)=1, z_{(0:k-1)}$, respectively. We use the data processing inequality and the separability property of the distance measure to obtain:

$$\begin{aligned}
&D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k, io_v(k)=1)\\
=&\mathbb{E}_{z_0\sim P_{z_0}} D(\mathcal{P}_{m\text{-}r\text{-}s}(X)\|\mathcal{P}_{m\text{-}r\text{-}s}(X')|U_k, io_v(k)=1, z_{(0:k-1)}=z_0)\\
\leq&\mathbb{E}_{z_0\sim P_{z_0}} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0)\\
\leq&\max_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0).
\end{aligned}$$

□

The theoretical results suggest that when privacy adversaries can perform the easiest in-out attack, the privacy amplification effects can degrade significantly. For local randomizers that satisfy ϵ_0-LDP, the amplified privacy almost increases by a factor of $\sqrt{K}$. However, if the number of queries K is of the same order as n, for instance, in the shuffle-then-randomize model, there is no privacy amplification remaining.

For simplicity in notation, Theorem E.1 considers cases where the victim user participated in the k-th query and user IDs in the same group U_k are leaked to privacy adversaries. To obtain an unconditional divergence bound, one may use the conditioning increasing property over the randomness of U_k.

F PROOF OF THEOREM 5.6

We first prove the privacy amplification lower bound, which indicates the destructive power of message-length information attacks. To prove that $D(\mathcal{P}_{d\text{-}r\text{-}s}(X)|U_{k,l}, len_v = l\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l) \geq \min_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S'))$, we consider the output of $\mathcal{P}_{d\text{-}r\text{-}s}$ with observed/fixed variables $U_{k,l}, len_v = l$, and define the following post-processing function:

step (1): Remove the $z_{(k')}$ for $k' \in [k+1:K]$ from the output;

step (2): Remove $\mathcal{R}_{(k)}(x_{i'})$ from $z_{(k)}$ for all $i' \in U_k\backslash U_{k,l}$ to get $z'_{(k)}$;

step (3): Returns $z_0 = z_{(0:k-1)}$ and $z'_{(k)}$.

When $z_{(0:k-1)}$ is fixed, the output distribution of step (2) is equal to an algorithm $\mathcal{P}_{s-r}$ in the shuffle-randomize model where global information is $z_{(0:k-1)}$. Thus, we have:

$$D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l, z_{(0:k-1)}) \geq D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0 = z_{(0:k-1)}).$$

Since X and X' differ only at x_v and $v \in U_k$, in two independent runs: $\mathcal{P}_{d\text{-}r\text{-}s}(X)|U_{k,l}, len_v = l, z_{(0:k-1)}$ and $\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l, z_{(0:k-1)}$, the distributions of $z_{(0:k-1)}$ are identical. Let P_{z_0} denote this distribution. Then, using the saprability property of distance measure over observable $z_{(0:k-1)}$, we have:

$$\begin{aligned}
&D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l)\\
=&\mathbb{E}_{z_0\sim P_{z_0}} D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l, z_{(0:k-1)} = z_0)\\
\geq&\mathbb{E}_{z_0\sim P_{z_0}} D(\mathcal{P}_{s\text{-}r}(S)|z_0\|\mathcal{P}_{s\text{-}r}(S')|z_0)\\
\geq&\min_{z_0} D(\mathcal{P}_{s\text{-}r}(S)|z_0\|\mathcal{P}_{s\text{-}r}(S')|z_0).
\end{aligned}$$

We then prove the privacy amplification upper bound, indicating the remaining amplification effects. Considering fixed $z_0 = z_{(0:k-1)}$ and fixed $U_{k,l}$, to prove

$$D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l, z_{(0:k-1)}) \geq \min_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0),$$

for the output of $\mathcal{P}_{s\text{-}r}$ with local randomizers $\mathcal{R}_{k'} \equiv \mathcal{R}_{(k)}$ and $k' \in [|U_{k,l}|]$, we define the following post-processing function:

step (1): The output of $\mathcal{P}_{s\text{-}r}$ given input $z_0, \{x_i\}_{i\in U_k}$ is

$$\{\mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(1)}), \ldots, \mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(|U_k|)})\},$$

where $\pi : U_{k,l} \mapsto [|U_{k,l}|]$ is a uniform-random permutation sampled by $\mathcal{P}_{s\text{-}r}$. Now initialize $z_{(k)}$ as the output of $\mathcal{P}_{s\text{-}r}$, for every $i' \in U_k\backslash U_{k,l}$, compute $\mathcal{R}_{(k)}(z_{(0:k-1)}, x_i)$, append it to $z_{(k)}$. Then, uniform-randomly permutes the $z_{(k)}$.

step (2): Compute $z_{(k')} = \mathcal{S}(\{\mathcal{R}_{(k')}(z_{(0:k'-1)}, x_{i'})\}_{i'\in U_{k'}})$ for $k' \in [k+1:K]$ sequentially.

step (3): Return $z_{(0)}, z_{(1)}, \ldots, z_{(K)}$.

The output distributions of step (3) with $x_v = a$ or $x_v = b$ are equal to the output distributions of $\mathcal{P}_{d\text{-}r\text{-}s}(X)|U_{k,l}, len_v = l, z_{(0:k-1)}$ and $\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l, z_{(0:k-1)}$, respectively. According to the data processing inequality and the separability property of distance measure, we have:

$$\begin{aligned}
&D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l)\\
=&\mathbb{E}_{z_0\sim P_{z_0}} D(\mathcal{P}_{d\text{-}r\text{-}s}(X)\|\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l, z_{(0:k-1)} = z_0)\\
\leq&\mathbb{E}_{z_0\sim P_{z_0}} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0)\\
\leq&\max_{z_0} D(\mathcal{P}_{s\text{-}r}(S)\|\mathcal{P}_{s\text{-}r}(S')|z_0).
\end{aligned}$$

G OTHER VARIANTS OF SHUFFLE MODEL UNDER MESSAGE LENGTH ATTACKS

Theorem G.1 (Privacy amplification of MRS model under in-out and message-length attack). *Given non-overlapping and complete user divisions $\{U_k\}_{k\in[K]}$ sampled as in the multinomial-randomize-shuffle model, and two neighboring datasets $X = \{x_0, \ldots, x_v = a, \ldots, x_n\}$, $X' = \{x_0, \ldots, x_v = b, \ldots, x_n\} \in \mathbb{X}^n$ that differ at the v-th user data. Assuming that $v \in U_k$ and $len(\mathcal{R}_{(k)}(z_{(0:k-1)}, a)) \overset{d}{=} len(\mathcal{R}_{(k)}(z_{(0:k-1)}, b))$, let l denote the observed message-length information about user v. Define $U_{k,l}$ as the set of users having the message*

length (i.e., $U_{k,l} = \{i \mid for i \in U_k \text{ and } len(\mathcal{R}_{(k)}(x_i)) = l\}$), and let $S = \{X(i)\}_{i' \in U_{k,l}}$, $S' = \{X'(i)\}_{i \in U_{k,l}} \in \mathbb{X}^{|U_{k,l}|}$ denote neighboring datasets w.r.t. $U_{k,l}$, then for any distance measure D that satisfies the data processing inequality and the separability property:

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l) \geq \min_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0), \max_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0) \Big],$$

where $\mathcal{R}_{[|U_{k,l}|]} = \mathcal{R}_{(k)}$ are the local randomizers in $\mathcal{P}_{s\text{-}r}$ and z_0 is the global information in $\mathcal{P}_{s\text{-}r}$.

Proof. First, we prove the privacy amplification lower bound, which highlights the destructive power of message-length information attacks. To prove that:

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l) \geq \min_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S')),$$

we consider the output of $\mathcal{P}_{m\text{-}r\text{-}s}$ with observed/fixed variables $U_k, U_{k,l}, len_v = l$, and define the following post-processing function:

step (1): Remove the $z_{(k')}$ for $k' \in [k+1:K]$ from the output;
step (2): Remove $\mathcal{R}_{(k)}(x_i)$ from $z_{(k)}$ for all $i \in U_k \backslash U_{k,l}$ to get $z'_{(k)}$;
step (3): Return $z_0 = z_{(0:k-1)}$ and $z'_{(k)}$.

When $z_{(0:k-1)}$ is fixed, the output distribution of step (2) equals to an algorithm $\mathcal{P}_{s-r}$ in the shuffle-randomize model where global information is $z_{(0:k-1)}$. Consequently, we obtain

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l, z_{(0:k-1)}) \geq D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0 = z_{(0:k-1)}).$$

Since X and X' differ only at x_v and $v \in U_k$, in two independent runs: $\mathcal{P}_{d\text{-}r\text{-}s}(X) | U_k, U_{k,l}, len_v = l, z_{(0:k-1)}$ and $\mathcal{P}_{d\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l, z_{(0:k-1)}$, the distributions of $z_{(0:k-1)}$ are identical. We denote this distribution as P_{z_0}. Then, using the separability property of distance measure, we obtain:

$$\begin{aligned} & D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l) \\ =& \mathbb{E}_{z_0 \sim P_{z_0}} D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l, z_{(0:k-1)} = z_0) \\ \geq & \mathbb{E}_{z_0 \sim P_{z_0}} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0) \\ \geq & \min_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0). \end{aligned}$$

We then prove the privacy amplification upper bound indicating the remaining amplification effects. Considering fixed $z_0 = z_{(0:k-1)}$ and fixed $U_k, U_{k,l}$, to prove

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l, z_{(0:k-1)}) \geq \min_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0),$$

we consider the output of $\mathcal{P}_{s\text{-}r}$ with local randomizers $\mathcal{R}_{k'} \equiv \mathcal{R}_{(k)}$ for $k' \in [|U_{k,l}|]$, and define the following post-processing function:

step (1): The output of $\mathcal{P}_{s\text{-}r}$ given input $z_0, \{x_i\}_{i \in U_k}$ is:

$$\{\mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(1)}), \ldots, \mathcal{R}_{(k)}(z_{(0:k-1)}, x_{\pi^{-1}(|U_k|)})\},$$

where $\pi : U_{k,l} \mapsto [|U_{k,l}|]$ is a uniform-random permutation sampled by $\mathcal{P}_{s\text{-}r}$. Now initialize $z_{(k)}$ as the output of $\mathcal{P}_{s\text{-}r}$, for every $i \in U_k / U_{k,l}$, compute $\mathcal{R}_{(k)}(z_{(0:k-1)}, x_i)$, append it to $z_{(k)}$. Then, uniform-randomly permutes the $z_{(k)}$.

step (2): Compute $z_{(k')} = \mathcal{S}(\{\mathcal{R}_{(k')}(z_{(0:k'-1)}, x_i)\}_{i \in U_{k'}})$ for $k' \in [k+1:K]$ sequentially.
step (3): Return $z_{(0)}, z_{(1)}, \ldots, z_{(K)}$.

The output distributions of step (3) with $x_v = a$ or $x_v = b$ are equal to the output distributions of $\mathcal{P}_{m\text{-}r\text{-}s}(X) | U_k, U_{k,l}, len_v = l, z_{(0:k-1)}$ and $\mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l, z_{(0:k-1)}$, respectively. Utilizing the data processing inequality and the separability property of distance measure, we obtain:

$$\begin{aligned} & D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_k, U_{k,l}, len_v = l) \\ =& \mathbb{E}_{z_{(0:k-1)}} D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | U_{k,l}, len_v = l, z_{(0:k-1)}) \\ \leq & \mathbb{E}_{z_{(0:k-1)}} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0 = z_{(0:k-1)}) \\ \leq & \max_{z_0} D(\mathcal{P}_{s\text{-}r}(S) \| \mathcal{P}_{s\text{-}r}(S') | z_0). \end{aligned}$$

□

We note that Theorem G.1 considers simplified conditional cases with known $U_k, U_{k,l}, len(\mathcal{R}_{(k)}(x_v)) = l$. For unconditional cases aiming at analyze the divergence:

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X') | len(\mathcal{R}_{(k)}(x_v = b)),$$

if $len(\mathcal{R}_{(k)}(x_v = a))$ follows a different distribution as $len(\mathcal{R}_{(k)}(x_v = b))$, then there will be additional local privacy loss (described earlier); if $len(\mathcal{R}_{(k)}(x_v = a))$ follows the same distribution as $len(\mathcal{R}_{(k)}(x_v = b)) = l$, then for distance measures satisfying the linearity property, since the $\mathbb{P}[U_k, U_{k,l}, len_v = l, z_{(0:k-1)}]$ are identical in two independent runs: $\mathcal{P}_{m\text{-}r\text{-}s}(X) | U_{k,l}, len_v = l, z_{(0:k-1)}$ and $\mathcal{P}_{m\text{-}r\text{-}s}(X') | U_{k,l}, len_v = l, z_{(0:k-1)}$, the overall divergence can be upper bounded by an expectation of the formulas in Theorem G.1, according to the separability property (for observable variable $len_v = l, z_{(0:k-1)}$) and the conditioning increasing property (for unobserved prior distribution of $U_{k,l}$) of the distance measure.

H DETAIL OF INTER-ROUND PRIVACY AMPLIFICATION IN MULTI-MESSAGE SHUFFLE MODEL

Intra-batch amplification in the divide-randomize-shuffle model. Recall that the base histogram protocol we use in each round is [23], which utilizes binary randomized response with budget ϵ_0 for the true data $x_i \in \{0,1\}^d$ (in the one-hot form) and uses binary randomized response on $\{0\}^d$ as a blanket message. Each user sends one blanket message. Assuming each round has n_0 users, according to Theorem A.1 for the divide-randomize-shuffle model, the divergence of one round, $D(\mathcal{P}_{d\text{-}r\text{-}s}(X) \parallel \mathcal{P}_{d\text{-}r\text{-}s}(X'))$, is lower bounded by a one-round shuffle model with n_0 users. Then, according to the variation-ratio reduction [91, Theorem 4.2, Table 3], the divergence $D(\mathcal{P}_{s\text{-}r}(X) \parallel \mathcal{P}_{s\text{-}r}(X')$ is upper bounded by

$D\Big(P^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0/2}} \,\|\, Q^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0/2}}\Big)$, where $P^q_{\beta,p}$ and $Q^q_{\beta,p}$ are defined with $n = n_0$ users as follows:

For $p > 1, \beta \in [0, \frac{p-1}{p+1}], q \geq 1$, let $C \sim Binomial(n-1, \frac{2\beta p}{(p-1)q})$, $A \sim Binomial(C, 1/2)$ and $\Delta_1 \sim Bernoulli(\frac{\beta p}{p-1})$ and $\Delta_2 \sim Bernoulli(1-\Delta_1, \frac{\beta}{p-1-\beta p})$; let $P^q_{\beta,p}$ denote $(A+\Delta_1, C-A+\Delta_2)$ and $Q^q_{\beta,p}$ denote $(A+\Delta_2, C-A+\Delta_1)$.

Inter-batch amplification in the rectified multinomial-randomize-shuffle model. Now consider the rectified multinomial-randomize-shuffle model in Algorithm 5. Each user acts like two virtual users in the single-message protocol where binary randomized response with budget ϵ_0 is used as the local randomizer. One virtual user holds the true secret value x_i, and the other virtual user holds 0^d. Assuming there are a total of K adaptive histogram queries, then according to Theorem B.1 or Theorem G.1, the divergence is upper bounded by a shuffle-then-randomize model with $n \cdot K \cdot 2$ users and identical local randomizers (i.e., the binary randomized response with ϵ_0-LDP). Meanwhile, according to the amplification bound of the shuffle-then-randomize model with binary randomized response [43, Theorem 3.1], the parameter β of binary randomized response is $\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1}$, and the divergence $D(\mathcal{P}_{m\text{-}s\text{-}r}(X) \,\|\, \mathcal{P}_{m\text{-}s\text{-}r}(X')$ is upper bounded by $D\Big(P^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0}} \,\|\, Q^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0}}\Big)$, where $P^q_{\beta,p}$ and $Q^q_{\beta,p}$ are defined with $n = n_0 \cdot K \cdot 2$ users as in the previous paragraph.

Both the numerical values of $D\Big(P^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0/2}} \,\|\, Q^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0/2}}\Big)$ and $D\Big(P^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0}} \,\|\, Q^{e^{\epsilon_0}}_{\frac{e^{\epsilon_0/2}-1}{e^{\epsilon_0/2}+1},e^{\epsilon_0}}\Big)$ can be efficiently computed with D as the Hockey-stick divergence [91]. The presented results in Figure 5 are the corresponding numerical results.

I PROOF OF THEOREM 6.2

Since $len(\mathcal{R}(x)) \equiv len_{max}$ in Algorithm 6, it is suffice to prove

$$\begin{aligned} &D(\mathcal{P}_{b\text{-}r\text{-}s}(X) \| \mathcal{P}_{b\text{-}r\text{-}s}(X') | U_{(m)}) \\ \leq &\max_{z_0} D(\mathcal{P}_{m\text{-}r\text{-}s}(S) \| \mathcal{P}_{m\text{-}r\text{-}s}(S') | z_0). \end{aligned}$$

Considering fixed $z_0 = z_{(0:k-1)}$, we define the following post-processing function for the output of $\mathcal{P}_{m\text{-}r\text{-}s}$ with local randomizers $\{\mathcal{R}_{(k')}\}_{k' \in Q_m}$ and query selection distribution $P_{(l)}$:

step (1): Let $z_{(b_{m-1})}, \ldots, z_{(b_m-1)}$ denote the output from $\mathcal{P}_{m\text{-}r\text{-}s}$ with $|Q_m|$ rounds.

step (2): Compute $z_{(k')} = \mathcal{S}_{(k')}(\{\mathcal{R}_{(k')}(z_{(0:k'-1)}, x_i)\}_{i \in U_{k'}})$ for $k' \in [b_m : K]$ sequentially, where $U_{k'}$ is chosen by users with the binned multinomial participation paradigm.

step (3): Return $z_{(0)}, z_{(1)}, \ldots, z_{(K)}$.

The output distributions of step (3) with $x_v = a$ or $x_v = b$ are equal to the output distributions of $\mathcal{P}_{b\text{-}r\text{-}s}(X)|U_{(m)}, z_{(0:k-1)}$ and $\mathcal{P}_{b\text{-}r\text{-}s}(X')|U_{(m)}, z_{(0:k-1)}$, respectively. According to the data processing inequality and the separability property of distance measure, we have:

$$\begin{aligned} &D(\mathcal{P}_{b\text{-}r\text{-}s}(X)U_{(m)} \| \mathcal{P}_{b\text{-}r\text{-}s}(X')|U_{(m)} \\ =&\mathbb{E}_{z_{(0:k-1)}} D(\mathcal{P}_{b\text{-}r\text{-}s}(X) \| \mathcal{P}_{b\text{-}r\text{-}s}(X')|U_{(m)}, z_{(0:k-1)} \\ \leq&\mathbb{E}_{z_{(0:k-1)}} D(\mathcal{P}_{m\text{-}r\text{-}s}(S) \| \mathcal{P}_{m\text{-}r\text{-}s}(S')|z_0 = z_{(0:k-1)}) \\ \leq &\max_{z_0} D(\mathcal{P}_{m\text{-}r\text{-}s}(S) \| \mathcal{P}_{m\text{-}r\text{-}s}(S')|z_0). \end{aligned}$$

J UNCONDITIONAL DIVERGENCES UNDER SIDE-CHANNEL ATTACKS

We note that Theorem 5.4 considers conditional cases with $io_v(k) = 1$. For the unconditional case $D(\mathcal{P}s\text{-}r(X) \| \mathcal{P}s\text{-}r(X')|io_v)$, the fact that random variables io_v are distributed identically, regardless of whether X or X' is provided as input, allows for the derivation of a lower bound using the conditioning increasing property of D. Specifically, we have:

$$\begin{aligned} &D(\mathcal{P}_{s\text{-}r}(X) \| \mathcal{P}_{s\text{-}r}(X')|io_v) \\ \geq &\mathbb{E}_{k \sim \text{uniform}[n]} D(\mathcal{P}_{s\text{-}r}(X) \| \mathcal{P}_{s\text{-}r}(X')|io_v(k) = 1) \\ \geq &\mathbb{E}_{k \sim \text{uniform}[n]} \min_z D(\mathcal{R}_k(z_{[0:k-1]}, a) \| \mathcal{R}_k(z_{[0:k-1]}, b)). \end{aligned}$$

Theorem 5.5 considers conditional cases where $U_k, io_v(k) = 1$, and $z_{(0:k-1)}$ is fixed. For unconditional U_k, io_v, the conditioning increasing property of distance measure D can be applied similarly. It is important to note that $z_{(0:k-1)}$ is observable in the subsample-randomize-shuffle model, hence $z_{(0:k-1)}$ consistently appears as a condition.

Theorem 5.6 considers simplified conditional cases with known $U_{k,l}, len_v = l$. Regarding the unconditional case:

$$D(\mathcal{P}_{m\text{-}r\text{-}s}(X) \| \mathcal{P}_{m\text{-}r\text{-}s}(X')|len(\mathcal{R}_{(k)}(b)),$$

if $len(\mathcal{R}_{(k)}(a))$ follows a different distribution as $len(\mathcal{R}_{(k)}(b))$, there is an additional local privacy loss (as described in the previous paragraph). If $len(\mathcal{R}_{(k)}(x_v = a))$ follows the same distribution as $len(\mathcal{R}_{(k)}(x_v = b))$, then for distance measures satisfying conditional composition, since the probabilities $\mathbb{P}[U_{k,l}, len(\mathcal{R}_{(k)}(x_v) = l), z_{(0:k-1)}]$ are identical in the following two independent runs: $\mathcal{P}_{d\text{-}r\text{-}s}(X)|U_{k,l}, len_v = l, z_{(0:k-1)}$ and $\mathcal{P}_{d\text{-}r\text{-}s}(X')|U_{k,l}, len_v = l, z_{(0:k-1)}$, the overall divergence can be upper bounded by an expectation of the formulas presented in the theorem, according to the conditioning increasing property.