

Necessary Conditions in Multi-Server Differential Privacy

Albert Cheu   

Department of Computer Science, Georgetown University, Washington D. C., USA

Chao Yan  

Department of Computer Science, Georgetown University, Washington D. C., USA

Abstract

We consider protocols where users communicate with multiple servers to perform a computation on the users' data. An adversary exerts semi-honest control over many of the parties but its view is differentially private with respect to honest users. Prior work described protocols that required multiple rounds of interaction or offered privacy against a computationally bounded adversary. Our work presents limitations of non-interactive protocols that offer privacy against unbounded adversaries. We prove that these protocols require exponentially more samples than centrally private counterparts to solve some learning, testing, and estimation tasks. This means sample-efficiency demands interactivity or computational differential privacy, or both.

2012 ACM Subject Classification Security and privacy → Information-theoretic techniques; Mathematics of computing → Probabilistic algorithms; Theory of computation → Distributed algorithms; Theory of computation → Online algorithms; Theory of computation → Sample complexity and generalization bounds; Security and privacy; Security and privacy → Privacy protections

Keywords and phrases Differential Privacy, Parity Learning, Multi-server

Digital Object Identifier 10.4230/LIPIcs.ITCS.2023.36

Related Version *Full Version:* <https://arxiv.org/abs/2208.08540>

Funding The authors are members of the Data Co-Ops project (<https://datacoopslab.org>). This work was supported in part by a gift to Georgetown University.

Acknowledgements We would like to thank Matthew Joseph for correspondence that refined our understanding of Bayesian re-sampling. We also thank Kobbi Nissim for suggestions for our sample complexity analysis.

1 Introduction

Following the seminal work by Dwork, McSherry, Nissim, and Smith [11], much research in differential privacy takes place in the central model. This assumes owners of data are willing to give their data to a central analysis server – an *analyst* for short – who runs a differentially private algorithm and reports the output to the (adversarial) world. Such an algorithm will guarantee that, loosely speaking, its output will not leak much information about any individual who gave their data. But the analyst could run other algorithms, so leaks and data misuse can still occur.

To keep data out of an analyst's hands, prior research has produced a variety of alternative models. A well-studied example is the local model. Here, *users* of mobile phones or web browsers run differentially private algorithms on their data and send the resulting messages to the analyst [22, 18]. The local nature of the randomization ensures privacy for any user even when the analyst and all other users are corrupted by the adversary. An inherent limitation of such a protocol is that the privacy noise from the users significantly weakens



© Albert Cheu and Chao Yan;

licensed under Creative Commons License CC-BY 4.0

14th Innovations in Theoretical Computer Science Conference (ITCS 2023).

Editor: Yael Tauman Kalai; Article No. 36; pp. 36:1–36:21

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

the signals they are meant to send; Kasiviswanathan, Lee, Nissim, Raskhodnikova, and Smith showed that locally private parity learning demands exponentially more samples than centrally private parity learning [18].

A line of work augments local protocols with a shuffler, an intermediary that applies a random permutation on user messages before sending the result to the analyst [6, 9]. The anonymity offered by the shuffler acts as a second layer of protection, atop the local randomization. The shuffler can be securely instantiated via anonymous broadcast protocols (e.g. Eskandarian & Boneh [13]) or classic mixnets (see Chaum [7]).

We focus on an alternative relaxation of the local model that appears in various forms in prior work [20, 2, 21, 5]: instead of just one analysis server, we assume there are $k \geq 2$ servers who share responsibility in processing user messages. An adversary can corrupt all but one of the users (as in the local model) and a large and unknown subset of the servers.¹ The adversary observes the messages received by parties it corrupts; we would like this view to change only slightly when an honest user changes their contribution. We remark that this multi-server model subsumes the shuffle model, since Eskandarian & Boneh create a multi-server protocol that performs anonymous broadcast [13].

That construction relies on the assumption that the adversary is computationally bounded. As noted by Steinke [20], such an assumption also implies accurate simulation of centrally private algorithms via secure multiparty computation. But classic work in central and local privacy allow *unbounded adversaries*. The natural line of thought is to explore the power of multi-server protocols when playing against that stronger class of adversary.

We are also interested in practical protocols: they should ideally be computationally lightweight, consume low bandwidth, and take place over few rounds of interaction. We focus on interactivity. Specifically, we would like to know how many rounds are needed to solve problems with few samples. There is precedent for such results in distributed differential privacy: in the local model, Joseph, Mao, and Roth showed exponential separations between r -round and $r + 1$ -round protocols [17].

As a starting point for adapting this to the multi-server setting, we can consider 1-round or *non-interactive* protocols. Here, each party produces one batch of outputs and never receives feedback. Refer to Figure 1 for a visualization.

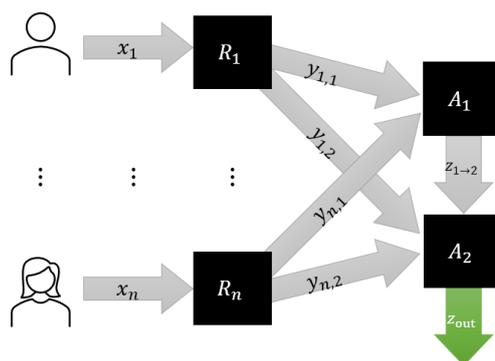
In Appendix B, we sketch a non-interactive multi-server protocol that ensures differential privacy against an unbounded adversary. It accurately estimates the sum of bits held by users. But aside from this basic task, *what can multi-server protocols compute if they must be non-interactive and ensure privacy against an unbounded adversary?*

1.1 Our Results

We show that the protocols of interest cannot perform some learning, testing, and estimation tasks without *exponentially* more samples than centrally private algorithms. An example is parity learning.

► **Theorem 1.1** (Informal). *Consider the family of non-interactive multi-server protocols that offer ϵ -differential privacy against unbounded adversaries that control $\leq \lceil k/2 \rceil$ servers. If each user samples d binary features and a binary label from the same distribution, any member of the protocol family requires $\Omega(2^{d/2}/\epsilon)$ samples to find a parity function that predicts the label of a fresh sample.*

¹ Steinke [20] and Talwar [21] describe protocols that ensure privacy holds when $k - 1$ servers are corrupt. Our results apply to protocols with a weaker guarantee.



■ **Figure 1** Diagram of a non-interactive two-server protocol. Users input their data into local randomizers, which send one message to each server. The first server produces an intermediary value, which the second server uses to produce the output. No party responds to messages (there are no cycles in the graph).

For comparison, Kasiviswanathan et al. show that $O(d/\epsilon)$ samples suffice under ϵ -central privacy [18]. In the full version of this work, we present similar exponential gaps in sample complexity for feature selection, simple hypothesis testing, and sparse mean estimation. Finally, we present a lower bound for uniformity testing, which is polynomially larger than the upper bound in the central model.

Our lower bounds have the following implication:

To solve some tasks with as few samples as the central model, multi-server protocols must be interactive or only ensure computational differential privacy, or both.

We note that our result holds in the case where the adversaries are honest but curious, meaning that the adversaries follow the protocol but they use the information they have access to in arbitrary ways. Our result implies a lower bound on sample complexity for more general adversaries, because the adversary can still choose to follow protocol.

1.1.1 Techniques

To arrive at our lower bounds, we take two high-level steps. First, we transform a multi-server protocol Π into an *internally private online algorithm* A_Π . Such an algorithm reads its input in a single for-loop and ensures that its internal state at the end of any single iteration respects differential privacy. This ensures that an adversary does not gain much advantage by intruding into the algorithm's memory. In our second step, we invoke lower bounds for internally private algorithms. As noted in the thesis by Cheu [8], these lower bounds are implied by the work of Cheu and Ullman [10] and Amin, Joseph, and Mao [1].

Section 3.2 describes the complete transformation for the two-server case; we briefly sketch the main ideas here. A_Π processes its input stream x_1, \dots, x_m in two batches. The first batch, labeled $[n] = \{1, \dots, n\}$, serves as the input to an execution of Π (on n samples): A_Π simulates the construction of messages from every user $i \in [n]$ to server 1. It is tempting to also simulate the messages to server 2, but the pair of messages may be non-private.² This motivates the second batch of samples, labeled $\{n+1, \dots, m\}$.

² Consider additive secret sharing amongst two servers: one share reveals nothing but the joint distribution is wholly dependent on the data value.

■ **Table 1** A non-exhaustive sampling of work on multi-server differential privacy. Most existing protocols are interactive and assume a bounded adversary. We are the first to give lower bounds.

	No. Servers	No. Corrupt Servers	Bounded Adversary?	> 1 round?	Notes
[2]	3	≤ 2	Yes	Yes	A protocol for histograms. Users prove inputs are valid.
[5]	2	≤ 1			A protocol for histograms with optimal ℓ_∞ error
[20]	k	$\leq k - 1$	Yes, for some		Protocols for counting, heavy-hitters, feature selection
[21]			Not if servers agree on coins		A protocol for vector sum. Rejects a user's vector if too big.
This Work		$\leq \lceil k/2 \rceil$	No	No	Lower bounds for feat. selection, parity learning, and more

Naively, we could use the second batch to sample from the marginal distributions of the messages to server 2. But the message a user sends to server 2 could depend on the one it sent to server 1. So we instead rely on a technique by Joseph, Mao, Neel, and Roth [16]: for each user i , A_Π obtains a fresh sample from \mathbf{D} *conditioned on having seen* the message from i to server 1. This proxy for i 's data is then used to produce the message from i to server 2. A technical hurdle arises from the fact that the desired conditional distribution requires the specification of \mathbf{D} , which is unknown to the algorithm. But A_Π approximates a sample from the conditional distribution by performing rejection sampling on the second batch of samples from \mathbf{D} .

Joseph et al. developed their technique in the context of local protocols, where the adversary can see all of the messages that an honest user generates [16]. In our model, the adversary can only see a subset of them. As a consequence, the simulator A_Π will purge old messages from memory when they are no longer needed: once A_Π simulates the message from i to server 2, it erases the message from i to server 1.

► **Remark 1.2.** It is conceivable that one could prove our transformations satisfy *pan-privacy*. An online algorithm satisfying this notion ensures differential privacy for any pairing of internal state and output. The lower bounds by Cheu & Ullman and Balcer et al. were originally stated for pan-privacy [10, 3]. But proofs of pan-privacy require an extra layer of case analysis, so we choose to prove only internal privacy to ease readability.

1.2 Related Work

The whitepaper by Apple & Google presents an interactive protocol for exposure notification analytics [2]. It offers protection against bounded adversaries. Bell, Gascon, Ghazi, Kumar, Manurangsi, Raykova, and Schoppmann also describe an interactive protocol offering computational DP [5]. It provably achieves asymptotically optimal ℓ_∞ error for histogram estimation. Talwar [21] and Steinke [20] both describe information-theoretically secure protocols, though the vector summation protocol in [21] assumes shared coins. Computational guarantees are offered by some protocols in [20]. See Table 1 for a summary of the above works. To our knowledge, we are the first to prove lower bounds in the multi-server model.

Other works, while not in the same model that we study, are of sufficient similarity to warrant mentioning below.

McGregor, Mironov, Pitassi, Reingold, Talwar, and Vadhan [19] present lower bounds in the *two-party* model, as do Haitner, Mazon, Silbak, and Tsfadia [14]. In that model, each of two servers has direct access to half of all user data and an honest server interacts with its peer in a way that ensures differential privacy for its half of the users. The model captures the scenario where users are evenly divided on which server to trust. In contrast, users in the two-server case of our model are all uncertain as to which server is honest. This distinction has an important consequence: a server in the two-party model can simply run a centrally private algorithm on their half of the samples, so the sample complexity in the central and two-party models are asymptotically identical. This holds for parity learning but also for any of the statistical estimation and testing problems we consider.

Beimel, Nissim, and Omri [4] study protocols that ensure privacy against an unbounded adversary like we do, but they focus on a user-to-user setting without dedicated servers. More significantly, the authors limit their constructions to secure function evaluation of differentially private algorithms, while we do not impose that constraint.

Finally, we remark that our online algorithms differ from the type found in the continual release literature (see Jain, Raskhodnikova, Sivakumar, and Smith [15] & citations within). There, the online algorithm outputs a value after every read. The stream of outputs must respect differential privacy and also serve as a good estimate of some function *as applied to each prefix* (e.g. sum). In contrast, the online algorithms we construct only produce output at the end of the stream. Moreover, we define privacy with respect to an arbitrary internal state chosen by the adversary but a private continual release algorithm could conceivably maintain a non-private internal state.

1.3 Open Questions

An intriguing open question is whether interactivity alone suffices to learn parity with low sample complexity. We can also ask the more fine-grained question that motivated our work: are there strong separations between r and $r + 1$ round protocols like in the local model?

Our work also leaves open the possibility of sample-efficient, non-interactive parity learning with computational differential privacy. We do not know the minimal assumptions needed to perform parity learning with polynomial sample complexity. For example, it is unknown if one-way functions suffice.

2 Preliminaries

For any (possibly randomized) algorithm M and distribution \mathbf{D} over inputs of M , $M(\mathbf{D})$ is shorthand for the distribution of $M(x)$ when $x \sim \mathbf{D}$. For any pair of distributions \mathbf{P}, \mathbf{Q} , the expression $SD(\mathbf{P}, \mathbf{Q})$ denotes the statistical (total variation) distance between the two.

We write $\mathbf{P} \approx_{\epsilon, \delta} \mathbf{Q}$ if, for all events Y , both of the following are true:

$$\mathbb{P}[\mathbf{P} \in Y] \leq e^\epsilon \cdot \mathbb{P}[\mathbf{Q} \in Y] + \delta$$

$$\mathbb{P}[\mathbf{Q} \in Y] \leq e^\epsilon \cdot \mathbb{P}[\mathbf{P} \in Y] + \delta$$

In the case where $\delta = 0$, we simply write $\mathbf{P} \approx_\epsilon \mathbf{Q}$.

► **Fact 2.1** (Post-Processing). *If $\mathbf{P} \approx_{\epsilon, \delta} \mathbf{Q}$, then for any algorithm M , $M(\mathbf{P}) \approx_{\epsilon, \delta} M(\mathbf{Q})$*

Throughout this work, *user* refers to a party who holds a single input value and *server* refers to a party who does not hold any input.

2.1 Non-interactive Multi-Server Protocols

Here, we take the number of users to be n and the number of servers to be $k > 1$. A protocol Π in the non-interactive multi-server model is specified by a tuple $(\{R_i\}_{i \in [n]}, \{A_j\}_{j \in [k]}, G)$. Each R_i is a local randomizer run by user i on their data; if all randomizers are identical, we simply write R . Meanwhile, A_j is an algorithm run by server $j \in [k]$. Finally, G is a k -node directed acyclic graph that determines the communication pattern between servers. We assume nodes (servers) are named according to a topological ordering. For brevity, we will drop the term “non-interactive” when discussing these protocols. Algorithm 1 describes how Π is executed.

■ **Algorithm 1** The execution of a multi-server protocol $\Pi = (\{R_i\}_{i \in [n]}, \{A_j\}_{j \in [k]}, G)$ on input \vec{x} .

```

For  $i \in [n]$ 
  User  $i$  samples  $(y_{i,1}, \dots, y_{i,k})$  from  $R_i(x_i)$ 
  User  $i$  sends  $y_{i,j}$  to server  $j$ , for all  $j \in [k]$ 
For  $j \in [k]$ 
  Server  $j$  receives  $y_{1,j}, \dots, y_{n,j}$  from users and  $z_{j' \rightarrow j}$  from servers  $j'$  where
     $(j', j) \in G$ 
  If  $j = k$  :
    The protocol's output is  $z_{\text{out}} \leftarrow A_j(\{y_{i,j}\}_{i \in [n]}, \{z_{j' \rightarrow j}\}_{(j',j) \in G})$ 
  Else
    Server  $j$  samples  $\{z_{j \rightarrow j''}\}_{(j,j'') \in G}$  from  $A_j(\{y_{i,j}\}_{i \in [n]}, \{z_{j' \rightarrow j}\}_{(j',j) \in G})$ 
    Server  $j$  sends  $z_{j \rightarrow j''}$  to server  $j''$ 

```

An attack Φ is specified by a tuple (C_u, C_s) . $C_u \subset [n]$ is the set of corrupted users while $C_s \subset [k]$ is the set of corrupted servers. For any multi-server protocol Π , input \vec{x} , and attack Φ , an adversary's *view* in execution $\Pi_\Phi(\vec{x})$ is the random variable

$$\text{View}_\Phi^\Pi(\vec{x}) := (z_{\text{out}}, \{z_{j \rightarrow j'}\}_{(j,j') \notin \overline{C_s} \times \overline{C_s}}, \{y_{i,j}\}_{(i,j) \notin \overline{C_u} \times \overline{C_s}}, \{x_i\}_{i \in C_u})$$

That is, an adversary attacking Π with Φ can observe the output of the protocol, all messages except those between honest parties, and the data of corrupted users.

For differential privacy to be satisfied, the adversary's view must be insensitive to any one user.

► **Definition 2.2** (Multi-Server Differential Privacy). Π is (ε, δ) -differentially private against c corrupted servers if, for all Φ where $|C_s| \leq c$ and for every neighboring pair $\vec{x} \sim \vec{x}'$ differing on $i \notin C_u$,

$$\text{View}_\Phi^\Pi(\vec{x}) \approx_{\varepsilon, \delta} \text{View}_\Phi^\Pi(\vec{x}')$$

Our work relies on a variety of constructions involving local randomizers, so we close this subsection with some relevant notation. For any subset of servers S , let $R_{i,S}$ be the algorithm that, on input x , computes $(y_{i,1}, \dots, y_{i,k}) \leftarrow R_i(x)$ and reports only $\{y_{i,j}\}_{j \in S}$. For any event E and disjoint subsets of servers S, S' , let $R_{i,S}(x) \mid R_{i,S'}(x) \in E$ denote the distribution of $\{y_{i,j}\}_{j \in S}$ conditioned on $\{y_{i,j}\}_{j \in S'} \in E$, where $\{y_{i,j}\}_{j \in [k]}$ are jointly drawn from $R_i(x)$. We use $R_{i,S}(\mathbf{D})$ and $R_{i,S}(\mathbf{D}) \mid R_{i,S'}(\mathbf{D}) \in E$ to denote the distributions when x is first sampled from \mathbf{D} .

2.2 Online Algorithms

An online algorithm M is specified by three algorithms $(M_{\text{init}}, M_{\text{update}}, M_{\text{out}})$. Algorithm 2 depicts how M is executed on a stream \vec{x} of length n : after initializing state, it repeatedly updates the state based upon the input stream.

■ **Algorithm 2** The execution of an online algorithm $M = (M_{\text{init}}, M_{\text{update}}, M_{\text{out}})$.

```

Initialize internal state  $S_0 \leftarrow M_{\text{init}}(\cdot)$ 
For  $i \in [n]$ 
   $\lfloor$  Update internal state  $S_i \leftarrow M_{\text{update}}(i, S_{i-1}, x_i)$ 
  Compute output  $z_{\text{out}} \leftarrow M_{\text{out}}(S_n)$ 
Return  $z_{\text{out}}$ 

```

To define privacy in this model, we mirror the previous section and define adversarial views. We make two assumptions: M_{update} is atomic and the privacy adversary can only view one internal state.

► **Definition 2.3** (Internally Private Online Algorithms). *For any online algorithm M , input \vec{x} , and time of intrusion t , let $\text{View}_t^M(\vec{x}) := S_t$ where S_t is generated as in Algorithm 2. M is (ε, δ) -internally private if and only if the following holds for every neighboring pair $\vec{x} \sim \vec{x}'$ and intrusion time t :*

$$\text{View}_t^M(\vec{x}) \approx_{\varepsilon, \delta} \text{View}_t^M(\vec{x}')$$

The above definition originates in the thesis by Cheu [8]. It is a relaxation of pan-privacy, wherein the adversary's view also includes the output z_{out} [12, 1, 3, 10].

3 From Multi-Server Protocols to Online Algorithms

► **Theorem 3.1.** *Suppose Π is a k -server protocol that takes n inputs and offers (ε, δ) -privacy against $\lceil k/2 \rceil$ corrupt servers. There exists a $(7\varepsilon, O(e^{5\varepsilon}\delta))$ -pan-private algorithm A_Π that takes $m = O(e^{4\varepsilon}n + e^{2\varepsilon} \log(1/\beta))$ inputs with the following property: for any distribution \mathbf{D} over inputs,*

$$SD(\Pi(\mathbf{D}^n), A_\Pi(\mathbf{D}^m)) \leq n\delta + \beta$$

We proceed in three stages. First, we prove some essential technical lemmas regarding local randomizers. Next, we describe how to simulate Π with an online algorithm in the case where $k = 2$. Finally, we argue that any protocol with larger k can be simulated by a two-server protocol with the same privacy parameters.

3.1 Properties of Local Randomizers

Although Π 's privacy guarantee does not imply any R_i is differentially private, it is straightforward to show that any strict subset of the randomizer's outputs is (ε, δ) -private.

▷ **Claim 3.2.** For any user $i \in [n]$ and subset of servers $S \subset [k]$ where $|S| \leq \lceil k/2 \rceil$, $R_{i,S}$ is (ε, δ) -differentially private.

Proof. Consider any attack Φ where $C_s = S$ and $i \notin C_u$. Fix any $\vec{x} \sim \vec{x}'$ that differ on i . Both $\text{View}_\Phi^\Pi(\vec{x})$ and $\text{View}_\Phi^\Pi(\vec{x}')$ contain messages from user i to servers in S in the same positions; closure under post-processing (Fact 2.1) implies $R_{i,S}(x_i) \approx_{\varepsilon, \delta} R_{i,S}(x'_i)$. ◁

It is easier to work with local randomizers that satisfy pure differential privacy than those that only satisfy approximate differential privacy. For this reason, we present the following technical lemma:

► **Lemma 3.3.** *If $R_L : \mathcal{X} \rightarrow \mathcal{Y}$ is (ε, δ) -differentially private, there exists an algorithm \tilde{R}_L that is 2ε -differentially private such that, for any $x \in \mathcal{X}$, $SD(R_L(x), \tilde{R}_L(x)) \leq \delta$.*

Proofs of Lemma 3.3 can be found in prior work; see e.g. Lemma 3.7 in Cheu and Ullman [10]. By combining this lemma with Claim 3.2, we obtain a very useful corollary:

► **Lemma 3.4.** *For any user $i \in [n]$ and subset of servers $S \subset [k]$ where $|S| \leq \lceil k/2 \rceil$, there exists a 2ε -differentially private algorithm $\tilde{R}_{i,S}$ such that for any $x \in \mathcal{X}$, $SD(R_{i,S}(x), \tilde{R}_{i,S}(x)) \leq \delta$.*

3.2 The Two-server Case

► **Theorem 3.5.** *Suppose Π is a two-server protocol that takes n inputs and offers (ε, δ) -privacy against one corrupt server. There exists a $(7\varepsilon, O(e^{5\varepsilon}\delta))$ -internally-private online algorithm A_Π that takes $m = O(e^{4\varepsilon}n + e^{2\varepsilon} \log(1/\beta))$ inputs with the following property: for any distribution \mathbf{D} over inputs,*

$$SD(\Pi(\mathbf{D}^n), A_\Pi(\mathbf{D}^m)) \leq n\delta + \beta$$

We construct a sequence of algorithms M_1, M_2, M_3 . Each approximates its predecessor and we show that M_3 , by erasing unnecessary random variables, is our desired internally private algorithm A_Π .

► **Remark 3.6.** M_1 and M_3 are online algorithms but to enhance readability, we avoid explicitly decomposing them into initialization, update, and output sub-routines as done in Section 2.2.

3.2.1 Step One: Shifting to Pure Differential Privacy

The pseudocode of M_1 is given in Algorithm 3. The sole difference between M_1 and the correct execution of Π (Algorithm 1) is swapping $R_{i,1}$ with $\tilde{R}_{i,1}$, the $(2\varepsilon, 0)$ -d.p. version of $R_{i,1}$. We do this to ease downstream analysis. Note that this step can be skipped if Π already guarantees $\delta = 0$.

■ **Algorithm 3** $M_1(\vec{x})$, an online algorithm that approximates $\Pi(\vec{x})$.

```

For  $i \in [n]$ 
   $y_{i,1} \sim \tilde{R}_{i,1}(x_i)$  /* Refer to Lemma 3.4 */
   $y_{i,2} \sim R_{i,2}(x_i) \mid R_{i,1}(x_i) = y_{i,1}$ 
 $z_{1 \rightarrow 2} \sim A_1(y_{1,1}, \dots, y_{n,1})$ 
 $z_{\text{out}} \leftarrow A_2(z_{1 \rightarrow 2}, y_{1,2}, \dots, y_{n,2})$ 
Return  $z_{\text{out}}$ 

```

▷ **Claim 3.7.** For any protocol inputs \vec{x} , $SD(M_1(\vec{x}), \Pi(\vec{x})) \leq n\delta$

Proof. Lemma 3.4 implies that swapping out $R_{i,1}$ for $\tilde{R}_{i,1}$ changes the distribution only by δ at each of the n sample points. A union bound completes the proof. ◁

3.2.2 Step Two: Generating Messages to Server 2 via Bayesian Re-Sampling

The pseudocode of M_2 is given in Algorithm 4. It proceeds in two phases, each dedicated to simulating a server's inputs. Like M_1 , it creates the messages to server 1 by running $\tilde{R}_{1,1}, \dots, \tilde{R}_{n,1}$ on the input. Unlike M_1 , M_2 does not generate the messages to server 2 $\{y_{i,2}\}_{i \in [n]}$ directly from the input. Instead, it performs *Bayesian re-sampling* as done by Joseph, Mao, Neel, and Roth [16]: to produce $y_{i,2}$, it runs the local randomizer on a fresh sample from \mathbf{D} conditioned on having seen $y_{i,1}$.

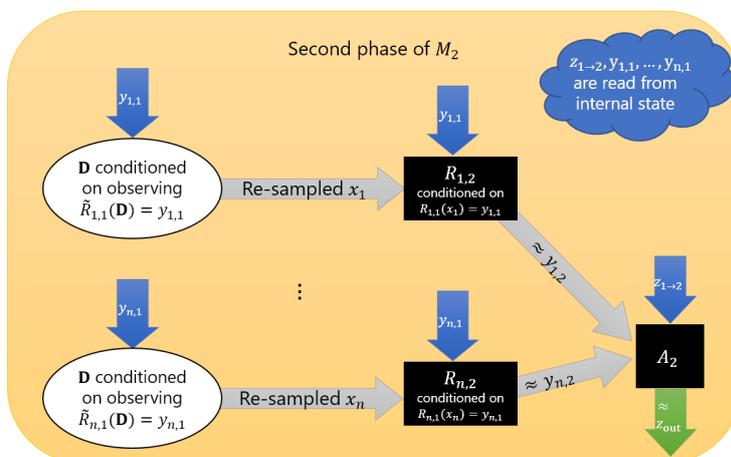
■ **Algorithm 4** $M_2(\vec{x})$, an algorithm that uses Bayesian re-sampling to simulate $M_1(\vec{x})$ when $\vec{x} \sim \mathbf{D}^n$.

```

/* First Phase: Create messages to server 1 */
For  $i \in [n]$ 
   $y_{i,1} \sim \tilde{R}_{i,1}(x_i)$ 
 $z_{1 \rightarrow 2} \sim A_1(y_{1,1}, \dots, y_{n,1})$ 
/* Second Phase: Create messages to server 2 by re-sampling data */
For  $i \in [n]$ 
   $\hat{x}_i \sim \mathbf{D} \mid \tilde{R}_{i,1}(\mathbf{D}) = y_{i,1}$ 
   $y_{i,2} \sim R_{i,2}(\hat{x}_i) \mid R_{i,1}(\hat{x}_i) = y_{i,1}$ 
 $z_{\text{out}} \leftarrow A_2(z_{1 \rightarrow 2}, y_{1,2}, \dots, y_{n,2})$ 
Return  $z_{\text{out}}$ 

```

We visualize the second phase of M_2 in Figure 2.



■ **Figure 2** Visualization of how M_2 simulates the messages to server 2.

▷ **Claim 3.8.** The distribution $M_2(\mathbf{D}^n)$ is identical to $M_1(\mathbf{D}^n)$

For brevity, we defer the proof to Appendix A.

3.2.3 Step Three: Implementing Bayesian Re-Sampling via Rejection Sampling

M_2 requires us to sample from \mathbf{D} conditioned on $\tilde{R}_{i,1}(\mathbf{D}) = y_{i,1}$, for every i . Since \mathbf{D} is the unknown distribution that is the subject of study, M_2 can only be a thought experiment. But we approximate the desired conditional distribution by performing private *rejection*

36:10 Necessary Conditions in Multi-Server Differential Privacy

sampling on independent samples from \mathbf{D} , as done by Joseph et al. [16]. This modification is presented in M_3 (Algorithm 5). It takes in $m > n$ samples, the excess being used for the rejection sampling. The updated second phase is visualized in Figure 3.

■ **Algorithm 5** $M_3(\mathbf{D}^m)$, an online algorithm that approximates $M_2(\mathbf{D}^n)$.

```

/* First Phase: Create messages to server 1 */
For  $i \in [n]$ 
   $y_{i,1} \sim \tilde{R}_{i,1}(x_i)$ 
 $z_{1 \rightarrow 2} \sim A_1(y_{1,1}, \dots, y_{n,1})$ 
/* Second Phase: Create messages to server 2 by re-sampling data */
/* Re-sampling approximated by rejection sampling */
 $i \leftarrow 1$ 
For  $h \in \{n+1, \dots, m\}$ 
  Compute acceptance rate  $rate_h \leftarrow \frac{\mathbb{P}[\tilde{R}_{i,1}(x_h) = y_{i,1}]}{2 \cdot \max_u \mathbb{P}[\tilde{R}_{i,1}(u) = y_{i,1}]}$ 
   $a_h \sim \text{Ber}(rate_h)$ 
  Erase  $rate_h$  from internal state
  If  $a_h = 1$  :
     $y_{i,2} \sim R_{i,2}(x_h) \mid R_{i,1}(x_h) = y_{i,1}$ 
    Erase  $y_{i,1}$  from internal state
     $i \leftarrow i + 1$ 
    If  $i = n + 1$  :
      Break loop
 $z_{out} \leftarrow A_2(z_{1 \rightarrow 2}, y_{1,2}, \dots, y_{n,2})$ 
Return  $z_{out}$ 

```

For brevity, Appendix A contains the proofs of the statements in this subsection.

▷ **Claim 3.9.** For any $\beta \in (0, 1)$, there is some $m = O(e^{4\varepsilon}n + e^{2\varepsilon} \log(1/\beta))$ where $SD(M_3(\mathbf{D}^m), M_2(\mathbf{D}^n)) \leq \beta$

We briefly sketch where this claim comes from. Because each $\tilde{R}_{i,1}$ is 2ε private (Lemma 3.4), we can see that $rate_h$ – the rate at which we accept a sample in the second batch as a proxy for a sample in the first batch – is at least $\exp(-2\varepsilon)/2$. The expected number of consumed samples is therefore $O(e^{2\varepsilon}n)$. We note that number of consumed samples is a sum of geometrically distributed random variables; a tail bound accounts for the extra $e^{2\varepsilon}$ factor and the $e^{2\varepsilon} \ln(1/\beta)$ term.

Our last step is to prove M_3 ensures internal differential privacy.

▷ **Claim 3.10.** If Π is (ε, δ) -private against 1 corrupt server, then M_3 is $(7\varepsilon, O(e^{5\varepsilon}\delta))$ -internally private.

It is conceivable that our transformation is pan-private, meaning that the pairing of internal state and output is differentially private, but proving so would double the number of cases in our analysis. We state our cases as separate sub-claims.

▷ **Claim 3.11.** If Π is (ε, δ) -private against 1 corrupt server and \vec{x}, \vec{x}' differ only on $i^* \in [n]$, then for any intrusion time $t > i^*$

$$\text{View}_t^{M_3}(\vec{x}) \approx_{2\varepsilon} \text{View}_t^{M_3}(\vec{x}')$$

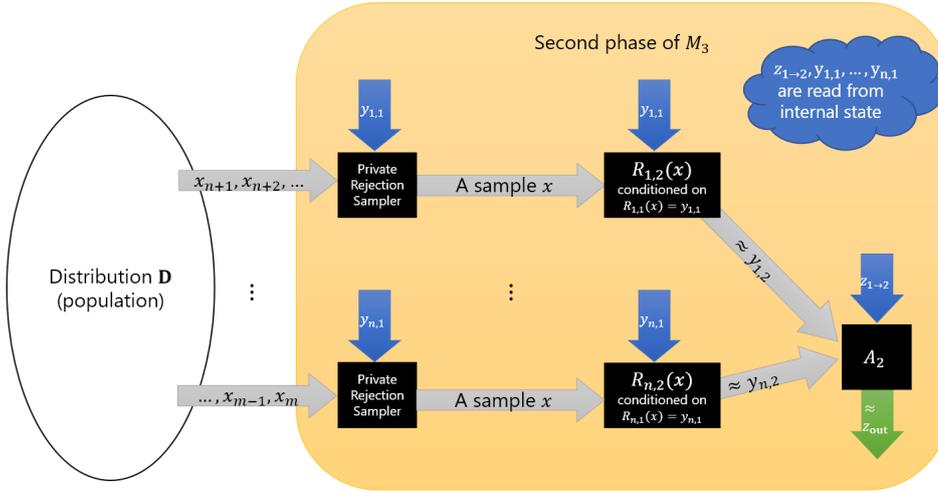


Figure 3 Visualization of how M_3 simulates the messages to server 2, assuming it has sample access to \mathbf{D} .

At a high level, the above comes from the fact that intrusions will either yield only information accessible to server 1 (which is 2ϵ -private), or yield only a post-processing of $z_{1 \rightarrow 2}$ (which comes from server 1’s view).

▷ Claim 3.12. If Π is (ϵ, δ) -private against 1 corrupt server and \vec{x}, \vec{x}' that differ only on $i^* \in [n + 1, m]$, then for any intrusion time $t > i^*$,

$$\text{View}_t^{M_3}(\vec{x}) \approx_{7\epsilon, 3e^{5\epsilon}\delta} \text{View}_t^{M_3}(\vec{x}')$$

To see where the above comes from, note that an adversary intruding at time t can only obtain a_t and precisely one of $y_{i,1}$ and $y_{i,2}$ (determined by a_t). Privacy of a_t follows from privacy of each $\tilde{R}_{i,1}$. If the adversary views $y_{i,1}$, we again have privacy via $\tilde{R}_{i,1}$. Otherwise, we argue that the marginal distribution of $y_{i,2}$ is close to that of $\tilde{R}_{i,2}(x_t)$ which is private.

Why are Interactive Protocols Difficult to Transform?

One can imagine a two-server protocol where server 2 sends some $z_{2 \rightarrow 1}$ to server 1, who then produces the output z_{out} . This is challenging to transform into a private online algorithm. To see why, recall that we need to simulate the view of the server who produces output, which is here the joint random variable $(y_{1,1}, \dots, y_{n,1}, z_{1 \rightarrow 2}, z_{2 \rightarrow 1})$. Our method allows us to privately simulate a different random variable $(y_{1,2}, \dots, y_{n,2}, z_{1 \rightarrow 2}, z_{2 \rightarrow 1})$. We could attempt to iteratively replace each $y_{i,2}$ with $y_{i,1}$ by again using Bayesian re-sampling, but the construction must now involve $z_{1 \rightarrow 2}$. This random variable is obtained from n independent samples from \mathbf{D} , unlike one in Algorithm 4. Rejection sampling is now quite difficult: we would need a way to compute an acceptance rate for a batch of n users, all while trying to maintain privacy of the internal state after reading each user’s data.

3.3 The Multi-server Case

Here, we construct a reduction from the k -server case to the two-server case.

▷ Claim 3.13. If the k -server protocol $\Pi = (\{R_i\}_{i \in [n]}, \{A_j\}_{j \in [k]}, G)$ is (ϵ, δ) -differentially private against $\lceil k/2 \rceil$ corrupted servers, then there exists a two-server protocol $\Pi' = (\{R'_i\}_{i \in [n]}, \{A'_j\}_{j \in [2]}, G')$ which is (ϵ, δ) -differentially private against 1 corrupted server and $\Pi(\vec{x}) = \Pi'(\vec{x})$ for any input $\vec{x} = (x_1, \dots, x_n)$.

36:12 Necessary Conditions in Multi-Server Differential Privacy

We provide the proof sketch here. The formal proof is deferred to Appendix A

Proof Sketch. In the two-server protocol Π' , the first server A'_1 plays the role of servers $A_1, \dots, A_{\lceil k/2 \rceil}$ – meaning it simulates all their code and interactions – while the second server A'_2 plays the role of $A_{\lceil k/2 \rceil + 1}, \dots, A_k$. The message from each user to A'_1 (resp. A'_2) is a tuple consisting of messages from that user to $A_1, \dots, A_{\lceil k/2 \rceil}$ (resp. $A_{\lceil k/2 \rceil + 1}, \dots, A_k$). Next, the message from A'_1 to A'_2 is a tuple consisting of messages from the first half of servers in Π to the second half. Finally, the output of A'_2 is the output of A_k .

For the accuracy, every computation in the servers of Π is done in A'_1 and A'_2 , so the output of A'_2 is the same with the output of A_k . For the privacy, since the protocol Π can defend the corruption of $\lceil k/2 \rceil$ servers, it can keep private when the first half of servers are corrupted or when the second half of servers are corrupted. Thus the protocol Π' can keep private when A'_1 or A'_2 is corrupted. \triangleleft

Combining Theorem 3.5 and Claim 3.13, we finally arrive at Theorem 3.1. We restate it below for convenience:

► **Theorem 3.1.** *Suppose Π is a k -server protocol that takes n inputs and offers (ε, δ) -privacy against $\lceil k/2 \rceil$ corrupt servers. There exists a $(7\varepsilon, O(e^{5\varepsilon}\delta))$ -pan-private algorithm A_Π that takes $m = O(e^{4\varepsilon}n + e^{2\varepsilon} \log(1/\beta))$ inputs with the following property: for any distribution \mathbf{D} over inputs,*

$$SD(\Pi(\mathbf{D}^n), A_\Pi(\mathbf{D}^m)) \leq n\delta + \beta$$

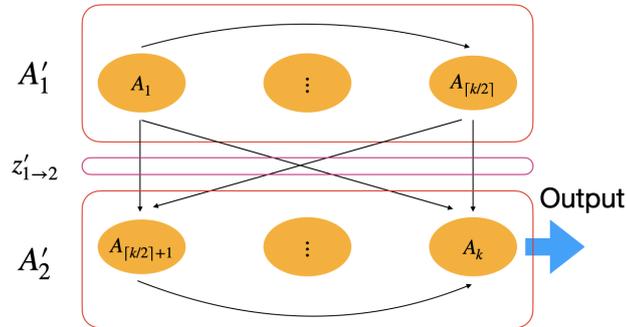
4 Application: A Lower Bound for Parity Learning

Given Theorem 3.1, we can invoke any one of the lower bounds for pan-privacy. For brevity, we only present a parity learning lower bound (and accompanying definitions). Other lower bounds can be found in the full version of our work.

To streamline the presentation, we assume $\varepsilon = O(1)$ and $\delta \ll 1/n$.

Let $\mathcal{X} = \{\pm 1\}^{d+1}$ be the domain; we will treat the last bit of every member string as the label of the string. The error of a parity function $(\ell, b) \in 2^{[d]} \times \{\pm 1\}$ with respect to a distribution \mathbf{D} over \mathcal{X} is

$$\text{err}_{\mathbf{D}}(\ell, b) := \mathbb{P}_{x \sim \mathbf{D}} \left[b \cdot \prod_{j \in \ell} x_j \neq x_{d+1} \right]$$



■ **Figure 4** Visualization of how to use two servers to simulate the execution of k servers.

► **Definition 4.1** (Parity Learning). *An algorithm M performs (d, t, α) -parity learning with sample complexity n if it takes n independent samples from a distribution \mathbf{D} over \mathcal{X} and reports a tuple $(L, B) \in 2^{[d]} \times \{\pm 1\}$ such that, with probability $99/100$, $|L| \leq t$ and*

$$\text{err}_{\mathbf{D}}(L, B) \leq \min_{|\ell| \leq t, b} \text{err}_{\mathbf{D}}(\ell, b) + \alpha$$

The following lower bound can be found in the thesis by Cheu [8]. $\binom{d}{\leq t}$ is shorthand for $\sum_{j=0}^t \binom{d}{j}$.

► **Theorem 4.2.** *If online algorithm M performs (d, t, α) -parity learning with sample complexity n and is (ε, δ) -pan-private for $\delta = 0$ or $\delta \log \left(\binom{d}{\leq t} / \delta \right) \ll \alpha^2 \varepsilon^2 / \binom{d}{\leq t}$, then $n = \Omega \left(\sqrt{\binom{d}{\leq t}} / \alpha \varepsilon \right)$.*

We remark that the above still holds if we had set a smaller success probability in the problem definition (e.g. $9/10$ instead of $99/100$). Such flexibility allows us to combine the above theorem with Theorem 3.1 and obtain the following bound on the sample complexity of any parity learner in the (non-interactive) multi-server model.

► **Theorem 4.3.** *If Π is a k -server protocol that solves (d, t, α) -parity learning with sample complexity n and offers (ε, δ) -differential privacy against $\lceil k/2 \rceil$ corrupt servers for $\delta = 0$ or $\delta \log \left(\binom{d}{\leq t} / \delta \right) \ll \alpha^2 \varepsilon^2 / \binom{d}{\leq t}$, then $n = \Omega \left(\sqrt{\binom{d}{\leq t}} / \alpha \varepsilon \right)$*

In contrast, Kasiviswanathan et al. show that $O(d/\alpha\varepsilon)$ samples suffice under ε -central privacy [18].

Proof of Theorem 4.3. Fix β to be a small constant. From Theorem 3.1 and our bound on ε , we can create an $(O(\varepsilon), O(\delta))$ -pan-private algorithm that takes $O(n)$ inputs and produces output that is within $n\delta + \beta$ of the parity learner in statistical distance. Note that this gap is bounded by a small constant (e.g. $9/100$) due to the magnitude of δ and our choice of β . So the pan-private algorithm learns parity with only a slightly smaller success probability and $O(n)$ samples. Theorem 4.2 will still apply. ◀

References

- 1 Kareem Amin, Matthew Joseph, and Jieming Mao. Pan-private uniformity testing. *CoRR*, abs/1911.01452, 2019. [arXiv:1911.01452](https://arxiv.org/abs/1911.01452).
- 2 Apple and Google. Exposure notification with privacy-preserving analytics (enpa) white paper. URL: <https://github.com/google/exposure-notifications-android/blob/master/doc/ENPA.pdf>.
- 3 Victor Balcer, Albert Cheu, Matthew Joseph, and Jieming Mao. Connecting robust shuffle privacy and pan-privacy. *CoRR*, abs/2004.09481, 2020. [arXiv:2004.09481](https://arxiv.org/abs/2004.09481).
- 4 Amos Beimel, Kobbi Nissim, and Eran Omri. Distributed private data analysis: Simultaneously solving how and what. In David A. Wagner, editor, *Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings*, volume 5157 of *Lecture Notes in Computer Science*, pages 451–468. Springer, 2008. doi:10.1007/978-3-540-85174-5_25.
- 5 James Bell, Adria Gascon, Badih Ghazi, Ravi Kumar, Pasin Manurangsi, Mariana Raykova, and Phillipp Schoppmann. Distributed, private, sparse histograms in the two-server model. *IACR Cryptology ePrint Archive*, 2022. URL: <https://eprint.iacr.org/2022/920>.
- 6 Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnés, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles, Shanghai, China, October 28-31, 2017*, pages 441–459. ACM, 2017. doi:10.1145/3132747.3132769.

- 7 David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24(2):84–88, 1981. doi:10.1145/358549.358563.
- 8 Albert Cheu. Differential privacy in the shuffle model. URL: <http://hdl.handle.net/2047/D20409473>.
- 9 Albert Cheu, Adam D. Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part I*, volume 11476 of *Lecture Notes in Computer Science*, pages 375–403. Springer, 2019. doi:10.1007/978-3-030-17653-2_13.
- 10 Albert Cheu and Jonathan R. Ullman. The limits of pan privacy and shuffle privacy for learning and estimation. *CoRR*, abs/2009.08000, 2020. arXiv:2009.08000.
- 11 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006. doi:10.1007/11681878_14.
- 12 Cynthia Dwork, Moni Naor, Toniann Pitassi, Guy N Rothblum, and Sergey Yekhanin. Pan-private streaming algorithms. In *Innovations in Computer Science (ICS)*, 2010.
- 13 Saba Eskandarian and Dan Boneh. Clarion: Anonymous communication from multiparty shuffling protocols. *IACR Cryptol. ePrint Arch.*, page 1514, 2021. URL: <https://eprint.iacr.org/2021/1514>.
- 14 Iftach Haitner, Noam Mazor, Jad Silbak, and Eliad Tsfadia. On the complexity of two-party differential privacy. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1392–1405. ACM, 2022. doi:10.1145/3519935.3519982.
- 15 Palak Jain, Sofya Raskhodnikova, Satchit Sivakumar, and Adam D. Smith. The price of differential privacy under continual observation. *CoRR*, abs/2112.00828, 2021. arXiv:2112.00828.
- 16 Matthew Joseph, Jieming Mao, Seth Neel, and Aaron Roth. The role of interactivity in local differential privacy. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 94–105. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00015.
- 17 Matthew Joseph, Jieming Mao, and Aaron Roth. Exponential separations in local differential privacy. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 515–527. SIAM, 2020. doi:10.1137/1.9781611975994.31.
- 18 Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 531–540. IEEE Computer Society, 2008. doi:10.1109/FOCS.2008.27.
- 19 Andrew McGregor, Ilya Mironov, Toniann Pitassi, Omer Reingold, Kunal Talwar, and Salil P. Vadhan. The limits of two-party differential privacy. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 81–90. IEEE Computer Society, 2010. doi:10.1109/FOCS.2010.14.
- 20 Thomas Steinke. Multi-central differential privacy. *CoRR*, abs/2009.05401, 2020. arXiv:2009.05401.
- 21 Kunal Talwar. Differential secrecy for distributed data and applications to robust differentially secure vector summation. *CoRR*, abs/2202.10618, 2022. arXiv:2202.10618.
- 22 Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

A

 Deferred Proofs

▷ **Claim 3.8.** The distribution $M_2(\mathbf{D}^n)$ is identical to $M_1(\mathbf{D}^n)$

Proof. Because the outputs of M_1, M_2 are determined by running A_2 on $(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]})$, it will suffice to show that $(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]})$ has the same distribution in both $M_2(\mathbf{D}^n)$ and $M_1(\mathbf{D}^n)$.

$$\begin{aligned}
& \mathbb{P}_{M_2(\mathbf{D}^n)} [(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]}) = (z, \vec{u})] \\
&= \sum_{\vec{v}} \mathbb{P}_{M_2(\mathbf{D}^n)} [(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]}) = (z, \vec{u}), \forall i y_{i,1} = v_i] \\
&= \sum_{\vec{v}} \mathbb{P}_{M_2(\mathbf{D}^n)} [(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]}) = (z, \vec{u}) \mid \forall i y_{i,1} = v_i] \cdot \mathbb{P}_{M_2(\mathbf{D}^n)} [\forall i y_{i,1} = v_i] \\
&= \sum_{\vec{v}} \mathbb{P}_{M_2(\mathbf{D}^n)} [(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]}) = (z, \vec{u}) \mid \forall i y_{i,1} = v_i] \cdot \mathbb{P}_{y_{i,1} \sim \tilde{R}_{i,1}(\mathbf{D})} [\forall i y_{i,1} = v_i] \\
&\hspace{15em} \text{(By construction)} \\
&= \sum_{\vec{v}} \mathbb{P}[A_1(\vec{v}) = z] \cdot \prod_{i=1}^n \underbrace{\mathbb{P}_{M_2(\mathbf{D}^n)} [y_{i,2} = u_i \mid y_{i,1} = v_i]}_T \cdot \mathbb{P}_{y_{i,1} \sim \tilde{R}_{i,1}(\mathbf{D})} [\forall i y_{i,1} = v_i] \tag{1}
\end{aligned}$$

(1) follows from the fact that $y_{1,2}, \dots, y_{n,2}$ are mutually independent and also independent of $z_{1 \rightarrow 2}$.

We know that the term T is equal to $\mathbb{P}_{M_1(\mathbf{D}^n)} [y_{i,2} = u_i \mid y_{i,1} = v_i]$ because M_2 merely changes the sampling order from “data, message 1, message 2” to “message 1, data, message 2.” Thus,

$$\begin{aligned}
(1) &= \sum_{\vec{v}} \mathbb{P}[A_1(\vec{v}) = z] \cdot \mathbb{P}_{M_2(\mathbf{D}^n)} [\forall i y_{i,2} = u_i \mid \forall i y_{i,1} = v_i] \cdot \mathbb{P}_{y_{i,1} \sim \tilde{R}_{i,1}(\mathbf{D})} [\forall i y_{i,1} = v_i] \\
&= \sum_{\vec{v}} \mathbb{P}_{M_1(\mathbf{D}^n)} [(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]}) = (z, \vec{u}) \mid \forall i y_{i,1} = v_i] \cdot \mathbb{P}_{M_1(\mathbf{D}^n)} [\forall i y_{i,1} = v_i] \\
&= \mathbb{P}_{M_1(\mathbf{D}^n)} [(z_{1 \rightarrow 2}, \{y_{i,2}\}_{i \in [n]}) = (z, \vec{u})] \quad \blacktriangleleft
\end{aligned}$$

▷ **Claim 3.9.** For any $\beta \in (0, 1)$, there is some $m = O(e^{4\epsilon}n + e^{2\epsilon} \log(1/\beta))$ where $SD(M_3(\mathbf{D}^m), M_2(\mathbf{D}^n)) \leq \beta$

Proof. As before, our analysis will condition on an arbitrary realization of $(z_{1 \rightarrow 2}, \{y_{i,1}\}_{i \in [n]})$.

We will in fact spend much of the proof studying $M_3^\infty(\mathbf{D}^\infty)$, a version of M_3 which takes in an *unbounded* stream of samples (replace the second for-loop with a while-loop). After establishing basic facts, we show that this alternate algorithm correctly simulates $M_2(\mathbf{D}^n)$. Then we show that this alternate algorithm consumes only m samples with $\geq 1 - \beta$ probability. Therefore, stopping at the m -th sample changes the overall distribution by at most β .

Facts about $M_3^\infty(\mathbf{D}^\infty)$: Let η_1 be the number of iterations of the while-loop until we obtain $y_{1,2}$ and move pointer i to 2. For $i > 1$, let η_i be number of iterations between sampling $y_{i-1,2}$ and sampling $y_{i,2}$.

36:16 Necessary Conditions in Multi-Server Differential Privacy

We now characterize the distribution of every η_i . To do so, note that, for any step h in the while-loop before $y_{i,2}$ is sampled,

$$\begin{aligned} \mathbb{P}_{x_h \sim \mathbf{D}} [a_h = 1] &= \int_{v=0}^1 v \cdot \mathbb{P}_{x_h \sim \mathbf{D}} [\text{rate}_h = v] dv \\ &= \int_{v=0}^1 v \cdot \mathbb{P}_{x_h \sim \mathbf{D}} \left[\frac{\mathbb{P}[\tilde{R}_{i,1}(x) = y_{i,1}]}{2 \cdot \max_u \mathbb{P}[\tilde{R}_{i,1}(u) = y_{i,1}]} = v \right] dv \\ &=: p_i \in \left[\frac{1}{2e^{2\varepsilon}}, \frac{1}{2} \right] \end{aligned} \quad (2)$$

The last step comes from Lemma 3.4. The immediate corollary is that η_i is drawn from $\mathbf{Geo}(p_i)$, the geometric distribution characterizing the number of $\mathbf{Ber}(p_i)$ trials until success.

Correctness of $M_3^\infty(\mathbf{D}^\infty)$: Because we have shown that the acceptance rate of a sample is nonzero ($p_i > 0$), the algorithm *eventually* samples $y_{i,2}$ for every i . We claim this implies correct simulation. Specifically, conditioned on $a_h = 1$, we claim that x_h is drawn from the correct posterior $\mathbf{D} \mid \tilde{R}_{i,1}(\mathbf{D})y_{i,1}$. The calculation below is adapted from an equivalent step in Joseph et al. [16]:

$$\begin{aligned} \mathbb{P}[x_h = x \mid a_h = 1] &= \mathbb{P}[a_h = 1 \mid x_h = x] \cdot \frac{\mathbb{P}[x_h = x]}{\mathbb{P}[a_h = 1]} \\ &= \frac{\mathbb{P}[\tilde{R}_{i,1}(x) = y_{i,1}]}{2 \cdot \max_u \mathbb{P}[\tilde{R}_{i,1}(u) = y_{i,1}]} \cdot \frac{\mathbb{P}[x_h = x]}{\sum_{\hat{x}} \mathbb{P}[x_h = \hat{x}] \cdot \frac{\mathbb{P}[\tilde{R}_{i,1}(\hat{x}) = y_{i,1}]}{2 \cdot \max_u \mathbb{P}[\tilde{R}_{i,1}(u) = y_{i,1}]}} \\ &= \frac{\mathbb{P}[\tilde{R}_{i,1}(x) = y_{i,1}] \cdot \mathbb{P}[x_h = x]}{\sum_{\hat{x}} \mathbb{P}[x_h = \hat{x}] \cdot \mathbb{P}[\tilde{R}_{i,1}(\hat{x}) = y_{i,1}]} \\ &= \mathbb{P}_{v \sim \mathbf{D}} [v = x \mid \tilde{R}_{i,1}(v) = y_{i,1}] \end{aligned}$$

Distance between $M_3^\infty(\mathbf{D}^\infty)$ and $M_3(\mathbf{D}^m)$: The total number of samples consumed by $M_3^\infty(\mathbf{D}^\infty)$ is $n + \sum_{i=1}^n \eta_i$. Since we know each η_i is a geometric random variable, the following lemma is useful:

► **Lemma A.1** (Tail Bound for Geometric Convolutions). *Fix any $p_1, \dots, p_n, \ell \in (0, 1/2]$ such that $\ell \leq p_i$ for every $i \in [n]$. If we sample η_i from $\mathbf{Geo}(p_i)$ for every $i \in [n]$, then*

$$\sum_{i=1}^n \eta_i = O\left(\frac{n}{\ell} \ln \frac{1}{\ell} + \frac{1}{\ell} \ln \frac{1}{\beta}\right)$$

with probability at least $1 - \beta$, for any $\beta \in (0, 1)$.

Refer to the full version of our paper for a proof. From (2), we have that $\ell = 1/2e^{2\varepsilon}$ so that substitution implies

$$\begin{aligned} \sum_{i=1}^n \eta_i &= O\left(e^{2\varepsilon} n \ln(2e^{2\varepsilon}) + e^{2\varepsilon} \ln \frac{1}{\beta}\right) \\ &= O\left(e^{4\varepsilon} n + e^{2\varepsilon} \ln \frac{1}{\beta}\right) \quad (\ln(2e^{2\varepsilon}) < 1 + 2\varepsilon) \end{aligned}$$

So setting m to the above (plus n) ensures $M_3(\mathbf{D}^m)$ is within β of $M_3^\infty(\mathbf{D}^\infty)$ in statistical distance. Because we know $M_3^\infty(\mathbf{D}^\infty)$ matches $M_2(\mathbf{D}^n)$, the proof is complete. \triangleleft

▷ Claim 3.10. If Π is (ε, δ) -private against 1 corrupt server, then M_3 is $(7\varepsilon, O(e^{5\varepsilon}\delta))$ -internally private.

▷ Claim 3.11. If Π is (ε, δ) -private against 1 corrupt server and \vec{x}, \vec{x}' differ only on $i^* \in [n]$, then for any intrusion time $t > i^*$

$$\text{View}_t^{M_3}(\vec{x}) \approx_{2\varepsilon} \text{View}_t^{M_3}(\vec{x}')$$

Proof. We proceed with case analysis, first with $t \in [n]$. The state S_t (resp. S'_t) consists of all random variables created by $M_3(\vec{x})$ (resp. $M_3(\vec{x}')$) up to time t , which includes the messages $\{y_{i,1}\}$ (resp. $\{y'_{i,1}\}$) for $i \in [t]$. When $t = n$, the state also includes $z_{1 \rightarrow 2}$ (resp. $z'_{1 \rightarrow 2}$), the message between the servers. But these random variables are obtained from post-processing the messages. Thus, it will suffice to show

$$\{y_{i,1}\}_{i \in [n]} \approx_{2\varepsilon} \{y'_{i,1}\}_{i \in [n]}$$

Because $\tilde{R}_{i^*,1}$ is $(2\varepsilon, 0)$ -private, we have that $y_{i^*,1} \approx_{2\varepsilon} y'_{i^*,1}$. And by construction, every other $y_{i,1}$ is identically distributed with $y'_{i,1}$. The claim follows by the mutual independence of the messages.

For other intrusion times $t \in [n+1, m]$, observe that x_{i^*} (resp. x'_{i^*}) is never read again so that the claim again holds by post-processing. ◁

▷ Claim 3.12. If Π is (ε, δ) -private against 1 corrupt server and \vec{x}, \vec{x}' that differ only on $i^* \in [n+1, m]$, then for any intrusion time $t > i^*$,

$$\text{View}_t^{M_3}(\vec{x}) \approx_{7\varepsilon, 3e^{5\varepsilon}\delta} \text{View}_t^{M_3}(\vec{x}')$$

Proof. In this case, notice that the view on input \vec{x} is³

$$\text{View}_t^{M_3}(\vec{x}) = (z_{1 \rightarrow 2}, y_{1,2}, \dots, y_{i-1,2}, y_{i,1}, \dots, y_{n,1}, a_{n+1}, \dots, a_t)$$

where each a_h is a bit sampled from $\mathbf{Ber}(\text{rate}_h)$ and $\hat{i} = 1 + \sum_{h=n+1}^t a_h$ is the index of the user in the first batch whose data we are re-sampling. Likewise,

$$\text{View}_t^{M_3}(\vec{x}') = (z'_{1 \rightarrow 2}, y'_{1,2}, \dots, y'_{\hat{i}'-1,2}, y'_{\hat{i}',1}, \dots, y'_{n,1}, a'_{n+1}, \dots, a'_t)$$

where each a'_h is sampled from $\mathbf{Ber}(\text{rate}'_h)$ and $\hat{i}' = 1 + \sum_{h=n+1}^t a'_h$.

Let $v(i^*) \in [n]$ (resp. $v'(i^*)$) be the value of the pointer i at the beginning of iteration i^* in $M_3(\vec{x})$ (resp. $M_3(\vec{x}')$).

We will argue that $(a_{i^*}, y_{v(i^*),2}) \approx_{7\varepsilon, 3e^{5\varepsilon}\delta} (a'_{i^*}, y_{v'(i^*),2})$. This suffices because the other pairs of variables in $\text{View}_t^{M_3}(\vec{x})$ are either (a) identically distributed with counterparts in $\text{View}_t^{M_3}(\vec{x}')$ or (b) obtained by post-processing $(a_{i^*}, y_{v(i^*),2})$.

By the privacy of every $\tilde{R}_{i,1}$, we have that $\mathbb{P}[a_h = 1], \mathbb{P}[a'_h = 1] \in [1/2e^{2\varepsilon}, 1/2]$ for any $h \in [n+1, m]$. Then,

$$\begin{aligned} \frac{\mathbb{P}[a_h = 1]}{\mathbb{P}[a'_h = 1]} &\leq \frac{1/2}{1/2e^{2\varepsilon}} = e^{2\varepsilon} \\ \frac{\mathbb{P}[a_h = 0]}{\mathbb{P}[a'_h = 0]} &\leq \frac{1 - 1/2e^{2\varepsilon}}{1/2} = 2 - e^{-2\varepsilon} < e^{2\varepsilon} \end{aligned} \quad (\text{Taylor series})$$

³ Similar to before, the views include z_{out} and z'_{out} when $t = m$ but closure under post-processing will again ensure that this does not affect the proof.

Consequently,

$$\begin{aligned}
 & \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E] \\
 &= \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E, a_{i^*} = 0] + \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E, a_{i^*} = 1] \\
 &\leq e^{2\varepsilon} \cdot \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid a_{i^*} = 0] \cdot \mathbb{P}[a'_{i^*} = 0] \\
 &\quad + e^{2\varepsilon} \cdot \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid a_{i^*} = 1] \cdot \mathbb{P}[a'_{i^*} = 1] \\
 &= e^{2\varepsilon} \cdot \mathbb{P}[(a'_{i^*}, y'_{v'(i^*),2}) \in E \mid a'_{i^*} = 0] \cdot \mathbb{P}[a'_{i^*} = 0] \\
 &\quad + e^{2\varepsilon} \cdot \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid a_{i^*} = 1] \cdot \mathbb{P}[a'_{i^*} = 1] \tag{3}
 \end{aligned}$$

The last step comes from two facts. First, x_{i^*} (resp. x'_{i^*}) will not be used in the creation of $y_{v(i^*),2}$ (resp. $y'_{v'(i^*),2}$) when $a_{i^*} = 0$ (resp. $a'_{i^*} = 0$) because the bit indicates rejection. Second, $v(i^*)$ is identically distributed with $v'(i^*)$ because the inputs \vec{x}, \vec{x}' are by definition identical prior to i^* .

If $a_{i^*} = 1$, x_{i^*} will be used to generate $y_{v(i^*),2}$. Let \mathcal{Y}_j be the range of $R_{v(i^*),j}$ and we assume without loss of generality it is discrete.

$$\begin{aligned}
 & \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid a_{i^*} = 1] \\
 &= \sum_{y \in \mathcal{Y}_1} \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid y_{v(i^*),1} = y, a_{i^*} = 1] \cdot \mathbb{P}[y_{v(i^*),1} = y \mid a_{i^*} = 1] \\
 &= \sum_{y \in \mathcal{Y}_1} \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid y_{v(i^*),1} = y, a_{i^*} = 1] \cdot \mathbb{P}[\tilde{R}_{v(i^*),1}(x_{v(i^*)}) = y] \\
 &\leq e^{2\varepsilon} \cdot \sum_{y \in \mathcal{Y}_1} \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid y_{v(i^*),1} = y, a_{i^*} = 1] \cdot \mathbb{P}[\tilde{R}_{v(i^*),1}(x_{i^*}) = y] \\
 &\hspace{15em} \text{(Lemma 3.4)} \\
 &\leq e^{2\varepsilon} \cdot \left(\delta + \sum_{y \in \mathcal{Y}_1} \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid y_{v(i^*),1} = y, a_{i^*} = 1] \cdot \mathbb{P}[R_{v(i^*),1}(x_{i^*}) = y] \right) \\
 &\hspace{15em} \text{(Lemma 3.4)} \\
 &= e^{2\varepsilon} \cdot \mathbb{P}[R_{v(i^*),2}(x_{i^*}) \in F] + e^{2\varepsilon} \delta \tag{4}
 \end{aligned}$$

where F be the subset of \mathcal{Y}_2 such that $y \in F$ if and only if $(1, y) \in E$.

Symmetric reasoning yields

$$\mathbb{P}[(a'_{i^*}, y'_{v'(i^*),2}) \in E \mid a'_{i^*} = 1] \geq e^{-2\varepsilon} \cdot \mathbb{P}[R_{v'(i^*),2}(x'_{i^*}) \in F] - e^{-2\varepsilon} \delta \tag{5}$$

We also know that $v(i^*)$ is identically distributed with $v'(i^*)$ and, for any i ,

$$\mathbb{P}[R_{i,2}(x_{i^*}) \in F] \leq e^\varepsilon \mathbb{P}[R_{i,2}(x'_{i^*}) \in F] + \delta, \tag{6}$$

Combining (4), (5), and (6), we obtain

$$\mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E \mid a_{i^*} = 1] < e^{5\varepsilon} \cdot \mathbb{P}[(a'_{i^*}, y'_{v'(i^*),2}) \in E \mid a'_{i^*} = 1] + 3e^{3\varepsilon} \delta$$

When we substitute the above into (3), we have

$$\begin{aligned}
 & \mathbb{P}[(a_{i^*}, y_{v(i^*),2}) \in E] \\
 &< e^{2\varepsilon} \cdot \mathbb{P}[(a'_{i^*}, y'_{v'(i^*),2}) \in E \mid a'_{i^*} = 0] \cdot \mathbb{P}[a'_{i^*} = 0] \\
 &\quad + e^{2\varepsilon} \cdot \left(e^{5\varepsilon} \cdot \mathbb{P}[(a'_{i^*}, y'_{v'(i^*),2}) \in E \mid a'_{i^*} = 1] + 3e^{3\varepsilon} \delta \right) \cdot \mathbb{P}[a'_{i^*} = 1] \\
 &< e^{7\varepsilon} \cdot \mathbb{P}[(a'_{i^*}, y'_{v'(i^*),2}) \in E] + 3e^{5\varepsilon} \delta
 \end{aligned}$$

which is what we wanted to prove. \triangleleft

▷ **Claim 3.13.** If the k -server protocol $\Pi = (\{R_i\}_{i \in [n]}, \{A_j\}_{j \in [k]}, G)$ is (ε, δ) -differentially private against $\lceil k/2 \rceil$ corrupted servers, then there exists a two-server protocol $\Pi' = (\{R'_i\}_{i \in [n]}, \{A'_j\}_{j \in [2]}, G')$ which is (ε, δ) -differentially private against 1 corrupted server and $\Pi(\vec{x}) = \Pi'(\vec{x})$ for any input $\vec{x} = (x_1, \dots, x_n)$.

Proof. Recall that the i -th local randomizer in Π has the form $R_i(x_i) = (y_{i,1}, \dots, y_{i,k})$. $R'_i(x_i)$ (Algorithm 6) constructs $(y'_{i,1}, y'_{i,2})$, where the first element is $(y_{i,1}, \dots, y_{i, \lceil k/2 \rceil})$ and the second is $(y_{i, \lceil k/2 \rceil + 1}, \dots, y_{i,k})$.

■ **Algorithm 6** R'_i , the i -th local randomizer in the two-servers protocol.

Let R_i be the i -th local randomizer in the k -server protocol.

Input: x_i

- 1 Sample $(y_{i,1}, \dots, y_{i,k}) \sim R_i(x_i)$.
 - 2 Let $y'_{i,1} = (y_{i,1}, \dots, y_{i, \lceil k/2 \rceil})$ and $y'_{i,2} = (y_{i, \lceil k/2 \rceil + 1}, \dots, y_{i,k})$.
 - 3 **Return** $(y'_{i,1}, y'_{i,2})$
-

A'_1 (Algorithm 7) simulates the execution of $A_1 \dots A_{\lceil k/2 \rceil}$ and A'_2 (Algorithm 8) simulates $A_{\lceil k/2 \rceil + 1} \dots A_k$. Since we assume that all servers in Π are named according to a topological ordering, A'_2 will not produce a message destined for a server simulated by A'_1 : there is no interaction between the two servers. Let $z'_{1 \rightarrow 2}$ be the collection of messages from the first half of the k servers to the second half.

■ **Algorithm 7** A'_1 , the first server in the two-server protocol.

Let $A_1, \dots, A_{\lceil k/2 \rceil}$ be the 1-st to $\lceil k/2 \rceil$ -th servers in the k -server protocol.

Input: $y'_{1,1}, \dots, y'_{n,1}$, where $y'_{i,1} = (y_{i,1}, \dots, y_{i, \lceil k/2 \rceil + 1})$ for $i \in [n]$

- 1 Sample $(z_{1 \rightarrow 2}, \dots, z_{1 \rightarrow k}) \sim A_1(y_{1,1}, \dots, y_{n,1})$.
 - 2 **For** $j \in \{2, \dots, \lceil k/2 \rceil\}$
 - 3 | Sample $(z_{j \rightarrow j+1}, \dots, z_{j \rightarrow k}) \sim A_j(y_{1,j}, \dots, y_{n,j}, z_{1 \rightarrow j}, \dots, z_{j-1 \rightarrow j})$.
 - 4 Let $z'_{1 \rightarrow 2} = (z_{1 \rightarrow \lceil k/2 \rceil + 1}, \dots, z_{1 \rightarrow k}, \dots, z_{\lceil k/2 \rceil \rightarrow \lceil k/2 \rceil + 1}, \dots, z_{\lceil k/2 \rceil \rightarrow k})$.
 - 5 **Return** $z'_{1 \rightarrow 2}$
-

■ **Algorithm 8** A'_2 , the second server in the two-server protocol.

Let $A_{\lceil k/2 \rceil + 1}, \dots, A_k$ be the $\lceil k/2 \rceil + 1$ -th to k -th servers in the k -server protocol.

Input: $y'_{1,2}, \dots, y'_{n,2}, z'_{1 \rightarrow 2}$, where $y'_{i,2} = (y_{i, \lceil k/2 \rceil + 1}, \dots, y_{i,k})$ for $i \in [n]$ and

$$z'_{1 \rightarrow 2} = (z_{1 \rightarrow \lceil k/2 \rceil + 1}, \dots, z_{1 \rightarrow k}, \dots, z_{\lceil k/2 \rceil \rightarrow \lceil k/2 \rceil + 1}, \dots, z_{\lceil k/2 \rceil \rightarrow k})$$

- 1 **For** $j \in \{\lceil k/2 \rceil + 1, \dots, k - 1\}$
 - 2 | Sample $(z_{j \rightarrow j+1}, \dots, z_{j \rightarrow k}) \sim A_j(y_{1,j}, \dots, y_{n,j}, z_{1 \rightarrow j}, \dots, z_{j-1 \rightarrow j})$.
 - 3 Sample $z_{out} \sim A_k(y'_{1 \rightarrow k}, \dots, y'_{n \rightarrow k}, z_{1 \rightarrow k}, \dots, z_{k-1 \rightarrow k})$
 - 4 **Return** z_{out}
-

The new protocol preserves accuracy because the output of A'_2 is identical to the output of A_k .

We now argue that Π' is (ε, δ) -differentially private against 1 corrupted server if Π is (ε, δ) -differentially private against $\lceil k/2 \rceil$ corrupted servers. This is done by arguing any attack $\Phi' = (C'_u, C'_s)$ against Π' corresponds to an attack Φ against Π . Specifically, let the set of corrupted users in Φ to be $C_u = C'_u$. If the set of corrupted servers in Φ' is $C'_s = \{1\}$,

let $C_s = \{1, \dots, \lceil k/2 \rceil\}$. If $C'_u = \{2\}$, let $C_s = \{\lceil k/2 \rceil + 1, \dots, k\}$. Note that $|C_s| \leq \lceil k/2 \rceil$ and the view of the adversary in both attacks are identical. Since Π is (ε, δ) -differentially private against Φ , Π' must be (ε, δ) -differentially private against Φ' . \triangleleft

B A Two-Server Protocol for Robust Count Estimates

Here, we sketch an example of a protocol in our model. It performs differentially private counting (summation of $\{0, 1\}$ values). Steinke [20] gave a simple protocol for this problem but a malicious user can greatly skew the count estimate.⁴ The protocol by Talwar [21] is designed with such attacks in mind. It performs high-dimensional addition (summation of values in the unit ℓ_2 ball). The one-dimensional nature of counting admits a greatly simpler construction.

Before we define the algorithms that make up the protocol, we give some preliminary notation. For predicate p , $\mathbb{I}\{p\}$ is 1 if p is true and 0 if it is false. For natural number t , let \mathbf{D}_t be the distribution over $[t]$ such that $\mathbb{P}_{\eta \sim \mathbf{D}_t}[\eta = v] \propto \exp(-\varepsilon \cdot |t/2 - v|)$.

On input x_i , the local randomizer R samples η_i from \mathbf{D}_t and reports

$$(y_{i,1} \leftarrow x_i + \eta_i, y_{i,2} \leftarrow \eta_i)$$

The first server samples α_1 from \mathbf{D}_t and reports

$$z_{1 \rightarrow 2} := \alpha_1 + \sum_{i \in [n]} y_{i,1} \cdot \mathbb{I}\{y_{i,1} \in [0, t + 1]\}$$

to the second server, who then reports

$$z_{\text{out}} := z_{1 \rightarrow 2} + \alpha_2 - t - \sum_{i \in [n]} y_{i,2} \cdot \mathbb{I}\{y_{i,2} \in [0, t + 1]\}$$

where α_2 is yet another sample from \mathbf{D}_t .

In our analysis, we make the simplifying assumption that $\delta \ll \varepsilon < 1$.

\triangleright **Claim B.1.** For $t = \Theta(\frac{1}{\varepsilon} \log \frac{1}{\delta})$, the protocol is (ε, δ) -differentially private against 1 (semi-honest) corrupted server.

Proof Sketch. Without loss of generality, we will ensure privacy for user 1 and assume the adversary corrupts all other users.

If server 1 is corrupted but server 2 is honest, the only variables pertaining to user 1 the adversary can obtain are $y_{1,1}$ and $\alpha_2 + y_{1,1} - y_{1,2} = \alpha_2 + x_1$. The former is received directly from user 1. The latter is obtainable by subtracting the other user's messages from z_{out} . $y_{1,1}$ is the result of adding noise from a truncated discrete Laplace distribution to x_1 , a value with sensitivity 1. The same is true for $\alpha_2 + x_1$. Hence we have (ε, δ) -differential privacy from composition (and the right choice of parameters).

If server 2 is corrupted (but server 1 is honest), the adversary can observe $y_{1,2}$ and the value $\alpha_1 + y_{1,1}$. The former is direct from user 1 while the latter is obtained by subtracting the other user's messages from $z_{1 \rightarrow 2}$. The adversary can compute $\alpha_1 + y_{1,1} - y_{1,2} = \alpha_1 + x_1$ which is (ε, δ) -differentially private for the same reason that $\alpha_2 + x_1$ is private. \triangleleft

⁴ Each honest user secret-shares their value across servers, so the modulus must be at least n . But a single malicious user can send shares that encode a $\Omega(n)$ value instead of a $\{0, 1\}$ value and honest servers cannot detect this.

▷ **Claim B.2.** If all parties are honest and t is set as above, then the protocol produces an unbiased estimate of the count such that, with 90% probability, the error is $O\left(\frac{1}{\varepsilon}\right)$. If there are m malicious users and no malicious servers, the error is $O\left(\frac{m}{\varepsilon} \log \frac{1}{\delta}\right)$.

Proof Sketch. We first argue that there is no bias in the honest execution. In this case, we equate z_{out} with the sum of $h_1 = \alpha_1 + \alpha_2 - t$ and $h_2 = \sum_{i \in [n]} y_{i,1} \cdot \mathbb{I}\{y_{i,1} \in [0, t + 1]\} - y_{i,2} \cdot \mathbb{I}\{y_{i,2} \in [0, t + 1]\}$. The random variable h_1 has mean 0 because $t = \mathbb{E}[\alpha_1 + \alpha_2]$. And observe that the construction of $y_{i,1}, y_{i,2}$ implies $h_2 = \sum x_i$.

Now we argue the error is likely low. By manipulating geometric series, it can be shown that $|\alpha_1 - t/2|$ is at most k with probability $(e^\varepsilon + 1 - e^{-\varepsilon k}) / (e^\varepsilon + 1 - \Theta(\delta))$. Invoking our bounds on δ and ε , this probability is at least 0.95 when $k = \Theta(1/\varepsilon)$. The same goes for α_2 . A union bound completes the proof.

We conclude with the analysis of the manipulation case. The honest servers limit the influence of any user on the output to be $O(t)$ because they only add messages that belong in the range $[0, t + 1]$. Hence, no coalition of m users can introduce more than $O\left(\frac{m}{\varepsilon} \log \frac{1}{\delta}\right)$ bias. ◁