2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Multi-Owner Multi-User Privacy

Fragkiskos Koufogiannis and George J. Pappas

Abstract— **In the realm of Internet of Things, sensitive information is distributed among several data owners, while multiple data users wish to access different aspects of this information. This paper presents an approach for a multi-owner multi-user (MOMU) system where data owners require privacy guarantees before offering their private data. In such a setting each owner has different privacy needs against each user, whereas, users may seek to collaborate in order to violate owners' privacy.**

Using approximate differential privacy, we focus on the case where n data owners possess a real-valued private data and m data users wish to learn a linear query of this data. We consider a Gaussian mechanism, derive the constraints on the covariance matrix for the mechanism to be multi-owner multi-user private, and propose a convex semi-definite relaxation to design the covariance. Finally, we illustrate our approach to a synthetic scenario where n agents act both as data owners and data users and we evaluate the privacy and the accuracy of the resulted mechanism.

I. Introduction

The paradigm of Internet of Things envisions that numerous *things* sense the environment, interact with their neighbors, and take actions towards their goals [1]. Since *things* —or agents— are considered highly interconnected, information gathered by a set of agents is used by a different group of agents. Although these agents use information gathered by other *things* to more efficiently perform their tasks, this flow of information raises privacy concerns.

For example, a collaborative recommender system for merchandise uses the buying history of customers in order to propose an item for future transactions [2]. In such a setting, a system that uses the purchase history of Alice to propose new products to her does not harm Alice's privacy. However, whenever Bob's history is used in this process, then, his privacy is compromised. The level of interaction between Alice and Bob is determined by the following two factor: (i) How useful is Bob's data for Alice? In fact, Alice is greatly benefit only if the two agents have similar shopping habits. (ii) How much does Bob trusts Alice? Specifically, if the two agents are close friends, there are little privacy concerns when recommending products to each other.

More generally, the underlying private data curation process is viewed as follows. Multiple agents, called data users, are interested in different aspects of a collection of private data, where the private data is distributed across a set of agents, called data owners—the two sets of agents are not necessarily disjoint. A key observation is that data users may benefit from colluding by either performing their tasks more efficiently or violating owners' privacy.

Differential privacy [3], which is employed here, injects noise to the responses such that an adversary cannot confidently infer the original private data. This technique provides formal privacy guarantees against a powerful and unmodeled adversary, where the strength of these guarantees is controlled by the *privacy level*. Notably, compared to cryptographic approaches, differentially privacy protects the private data against the intended recipient of the response [4]. The literature of differential privacy was initiated in a database context [3], [5], such as recommender systems [4], and was later extended to Kalman filtering [6], optimization problems [7], consensus algorithms [8] etc. The underpinning approach of these works consider a single, possibly high-dimensional, private database and propose a differential private algorithm which produces a noisy response approximating a quantity of interest. Next, this noisy response is published and, thus, everyone observes the same noisy response. In this case of a single private data and a single noisy response, there is a single associated privacy level. In the multi-owner scenario [9], [10], the private data is owned by multiple agents and a single noisy response is generated, where each owner requires a different privacy level. Earlier work [11] introduced the multi-user case, where multiple data users are interested in a single private data. In this case, the single data owner required different privacy levels against each user and, thus, each user receives a different response from the differential private algorithm. Importantly, data users may be tempted to collaborate by sharing the responses they received and, thus, harming the owner's privacy.

In this paper, we focus on designing differentially private mechanisms in a multi-owner multi-user scenario. Specifically, we consider multiple data owners holding a piece of private data and multiple users, each interested in a different function of the entire private data. We formulate the problem of multi-owner multi-user privacy and explore some variations of it. Next, we focus on owners possessing real-valued private data and users seeking linear functions of this data. Within approximate differential privacy, we propose a Gaussian-based mechanism and express the privacy constraints as constraints on the covariance matrix. Then, we design the mechanism by relaxing these constraints to semi-definite problem. We illustrate our approach in the setting of n agents, each acting both as an data owners and data users. Our approach is numerically evaluated both in terms of

Authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, PA, USA.

This work was supported in part by the TerraSwarm Research Center, one of six centers supported by the STARnet phase of the Focus Center Research Program (FCRP) a Semiconductor Research Corporation program sponsored by MARCO and DARPA and in part by NSF CNS-1505799 and the Intel-NSF Partnership for Cyber-Physical Systems Security and Privacy.

efficiency (how much noisier the responses are because of the multiple users) and in terms of incentive to collaborate (how much more information users can gain by collaborating).

The paper is structured as follows. Section II informally introduces the problem of multi-owner, multi-user privacy and provides the system model. Next, differential privacy is revisited and a formal statement of the problem is derived. Section III initially provides an approach to the problem based on existing results and, then, uses a semi-definite problem to design a mechanism which approximately de-incentivizes users from colluding and harming privacy guarantees. We demonstrate our approach in Section IV with an example of a synthetic social network where agents act both as data owners and data users and wish to estimate their local average. Finally, Section V concludes our work.

II. Problem Formulation

We now introduce the problem of multi-owner multi-user privacy. Initially, we informally state the general problem and, in the end of this section, we formulate a concrete instance of the problem. We consider two, possibly overlapping, groups of agents: *data owners* and *data users*. Specifically, we consider n data owners with owner $i \in [n] = \{1, \dots, n\}$ possessing private data $u_i \in \mathcal{U}_i$, where $\mathcal{U}_i$ is the set of possible values for the private data of owner i. Let $\mathbf{u} = [u_i]_{i=1}^n$ denote the set of everyone's private data. We also consider m data users, where each user $j \in [m] = \{1, \dots, m\}$ is interested in some function $q_j = q_j(\mathbf{u})$ of the private data. Furthermore, we quantify the severity of the privacy concerns that owner i has against user j with the *privacy level* ϵ_{ij}, where smaller values indicate more severe privacy concerns —an overview of differential privacy is presented in Subsection II-A. Assuming there exists a trusted operator of the system, we wish to design a (randomized) mechanism $\mathcal{M}$ which, given the set of private data $\mathbf{u} \in \times_{i=1}^n \mathcal{U}_i$, computes the set of responses $\mathbf{Y} = [Y_i]_{j=1}^m$ and securely communicates response Y_j to user j as a proxy of $q_j(\mathbf{u})$. From a utility point of view, we wish each response Y_j to be a good approximation of $q_j(\mathbf{u})$. From a privacy point of view, we wish to guarantee that, given the response Y_j, private data u_i remains ϵ_{ij}-differentially private.

However, data users might decide to collude, share their responses, and violate the privacy needs of a data owner. Therefore, mechanism $\mathcal{M}$ should not incentivized such coalitions. In particular, for any group $\mathcal{J} \subseteq \{1, \dots, m\}$ of data users and any data owner i, there should exist a user $j^* \in \mathcal{J}$ that does not gain any more information about u_i by participating in group $\mathcal{J}$ and, thus, leaves the coalition.

In this paper, we focus on the case of real-valued private data u_i, linear queries q_j, and the notion of approximate differential privacy. Before formally stating our problem as in Problem 1 in Subsection II-D, we review differential privacy in Subsection II-A and related results from the literature in Subsection II-E.

A. Differential Privacy

Differential privacy [3] —see [12] for a survey— is a framework that provides strong and formal privacy guarantees. Intuitively, given a private data as input, a differentially private algorithm returns a noisy response such that an adversary who observes this response cannot confidently infer the private data. In particular, as shown in Definition 1, the dependency of the noisy response on the private data should be bounded and is captured by the pair of constants $(\epsilon, \delta) \in \mathbb{R}_+^2$ called *privacy level*; smaller values of the parameters correspond to stronger privacy.

Definition 1 (Differential Privacy [3])**.** *Let (ϵ, δ) be a privacy level, $\mathcal{U}$ be the space of private data, and $\mathcal{A} \subseteq \mathcal{U} \times \mathcal{U}$ be an adjacency relation. Then, a mechanism $\mathcal{M} : \mathcal{U} \to \Delta(\mathcal{Y})$ is (ϵ, δ)-differentially private if*[*]

$$\mathbb{P}(\mathcal{M}u \in \mathcal{S}) \leq e^{\epsilon}\, \mathbb{P}(\mathcal{M}u' \in \mathcal{S}) + \delta, \quad \forall \mathcal{S} \subseteq \mathcal{Y},$$

for all adjacent inputs $(u, u') \in \mathcal{A}$.

For $\delta > 0$, we refer to Definition 1 as approximate differential privacy, whereas for $\delta = 0$ we retrieve pure differential privacy. Here, we will consider real-valued private data $\mathcal{U} = \mathbb{R}$ and an adjacency relation of the form

$$(u, u') \in \mathcal{A}_\alpha \;\Leftrightarrow\; |u - u'| \leq \alpha,$$

for some constant $\alpha > 0$. For simplicity, we will assume that $\alpha = 1$. Such an adjacency relation is commonly used in the literature, e.g. [13], [14], [6].

A popular approximate differential private mechanism is the Gaussian mechanism defined as follows.

Theorem 2 (Gaussian Mechanism [3])**.** *Consider the mechanism $\mathcal{M} : \mathbb{R} \to \Delta(\mathbb{R})$ which adds Gaussian noise*

$$\mathcal{M}(u) = c\,u + V, \text{ where } V \sim \mathcal{N}\left(0, \frac{c^2}{\kappa(\epsilon, \delta)^2}\right),$$

where $c \in \mathbb{R}$, $\kappa(\epsilon, \delta) = \frac{2\epsilon}{K + \sqrt{K^2 + 2\epsilon}}$, $K = Q^{-1}(\delta)$[†]*, and $\mathcal{N}(0, \sigma^2)$ is the normal distribution with variance σ^2. Then $\mathcal{M}$ is (ϵ, δ)-differentially private under adjacency relation $\mathcal{A}_1$.*

In the following, the term *privacy level* will refer to the parameter ϵ in the context of pure differential privacy and to the value $\kappa(\epsilon, \delta)$ in approximate differential privacy.

B. SISO to MIMO Privacy

Treating data owners as inputs and data users as outputs, Figure 1 categorizes some of the literature in differential privacy. Specifically, work in [3] introduced differential privacy in a single-owner, single-user setting. They consider a single, possible high-dimensional, private input u and a single privacy level ϵ, and focused on designing ϵ-differential private mechanisms; i.e. given the response y of the mechanism, the private input u is ϵ-private. The majority of the following

[*]For a set T and a rich-enough σ-algebra $\mathcal{T}$ of it, we denote with $\Delta(T)$ the set of all probability measures on $(T, \mathcal{T})$.

[†]Q is the tail probability of the standard normal distribution

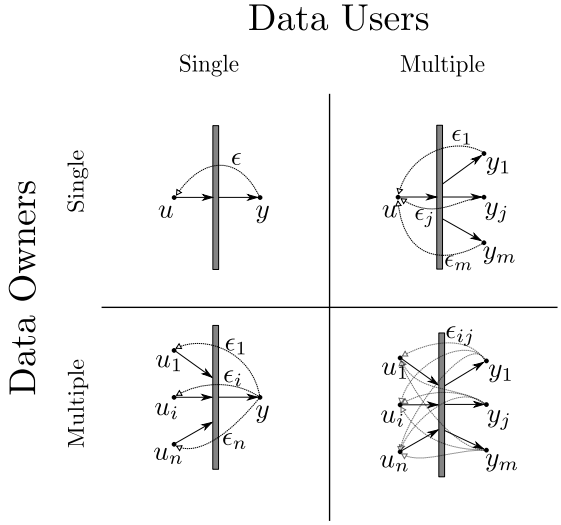

Fig. 1: A categorization of differentially private mechanism based on the number of data owners (inputs) and data users (outputs).

work proposed private mechanisms in a variety of fields ranging from Kalman filtering [6], consensus algorithms [15] ,and optimization problems [16], [7] to smart metering [17] and traffic flow estimation [18]. Importantly, although different components of the private data u belong to different owners, the proposed mechanisms protect the input u as a whole with a single privacy level ϵ.

Next, authors of [9] and those of [10] considered n data owners, each with a private data u_i and a privacy level ϵ_i, and propose a mechanism that computes a single output y which is publicly announced. Such a setting, which was also used in [19], can be thought as multi-owner single-user privacy and, practically, is interpreted as in that, given the response y, each private data u_i remains ϵ_i-private.

Furthermore, authors in [11] considered the single owner, multiple users case. A single data owner shares a private data u with m users under different privacy levels ϵ_j, $j \in [m]$ by responding with y_j to user j. The authors propose a mechanism such that, given y_j, private data u remains ϵ_j-private. Importantly, the proposed mechanism does not incentivize coalitions among the users; i.e. users are not willing to collude and share information in order to damage the owner's privacy.

According to such a categorization, present work considers the multi-owner, multi-user scenario, where each data owner i has a different privacy level ϵ_{ij} against each data user j. Additionally, each data user is interested in a different aspect of the private data. For example, within a sensor network, based on their location, sensors are interested in mostly local information. As in the SIMO case, it is not enough to guarantee the privacy of owner i against user j and, thus, the MIMO case cannot be decomposed to m MISO systems. Instead, we need to model any possible interactions among the data users that can lead to privacy breaches. Since we assume the existence of a trusted system operator, we focus only on users' interactions that occur after the execution of the differentially private mechanism.

C. Effects of Coalitions

As briefly mentioned above, the case of designing a private mechanism $\mathcal{M}$, with $\mathcal{M}(\mathbf{u}) = \begin{bmatrix} Y_1 & \cdots & Y_m \end{bmatrix}$, that serves m data users does not decompose into m independent mechanisms $\mathcal{M}_j$, $j \in [m]$, where $\mathcal{M}_j(\mathbf{u}) = Y_j$. Such a decomposition is not possible due to possible interactions among the data users. In fact, data users may collaborate and exchange information for different reasons. Two models of the possible interaction among data users are the following.

- *Curious coalitions:* Consider a group of data users $\mathcal{J} \subseteq \{1, \ldots, m\}$ and the responses $\mathbf{Y}_{\mathcal{J}} = [Y_j]_{j \in \mathcal{J}}$ that they receive as a proxy to the quantities of interest $[q_j(\mathbf{u})]_{j \in \mathcal{J}}$. If there is a post-processing of the coalition's knowledge $\mathbf{Y}_{\mathcal{J}}$ such that *each* colluding user $j \in \mathcal{J}$ extracts a more accurate proxy Y'_j of the quantity of interest $q_j(\mathbf{u})$, then, coalition $\mathcal{J}$ is stable. We call such a group a *curious coalition* because users focus on simply improving the accuracy of their received responses while ignoring any privacy requirements.
- *Adversarial coalitions:* In this case, a subset $\mathcal{J} \subseteq [m]$ of the users collude and share their information $\mathbf{Y}_{\mathcal{J}}$ in order to infer private data u_i and violate the privacy of a targeted data owner i. In this case, data users are considered adversarial; for example, they might be multiple personae of a single adversary.

Beyond curious and adversarial coalitions, as introduced here, further models exist. For example, in the case that data users also act as data owners, they may or may not care about their own privacy levels whenever they participate in a curious coalition. Specifically, agents participate in a coalition $\mathcal{J}$ only if (i) the accuracy of the received responses is improved and (ii) their privacy level is not compromised, even against other members of the coalition. Another model considers users who take publicly observed actions based on their received responses and, thus, information is exchanged between users.

In this paper, we will mostly focus on adversarial coalitions. As in the SIMO case [11], the main technique against coalitions is introducing correlation among the responses $\{Y_j\}_{j=1}^m$ that the data users receive. Specifically, the main result of the SIMO case designs a mechanism such that, in each possible coalition, there exists a data user who does not benefit by colluding and, thus, leaves the coalition. Another possible, but not exploited here, technique to de-incentivize coalitions is considering mechanisms that return some side information Z_j to data user j. Then, response Y_j is used by honest users, whereas, side information Z_j is *"used to gratify the curiosity"* of dishonest users.

D. Linear Queries of Real-Valued Private Data

Here, we assume that each data owner i possesses a scalar private data $u_i \in \mathbb{R}$ and each data user j is interested in a linear form of all the private data

$$q_j(\mathbf{u}) = \sum_{i=1}^{n} a_{ij} u_i.$$

Moreover, we assume that each data owner i requires $(\epsilon_{ij}, \delta_{ij})$-differential privacy against data user j. To that end, a trusted system operator receives all private data $\mathbf{u}$, computes the desired quantities $\{q_j(\mathbf{u})\}_{j=1}^m$, adds noise $\mathbf{V}$, and returns the response $Y_j = q_j(\mathbf{u}) + V_j$ to user j:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} q_1(\mathbf{u}) + V_1 \\ \vdots \\ q_m(\mathbf{u}) + V_m \end{bmatrix} = A\,\mathbf{u} + \mathbf{V}, \qquad (1)$$

where $A = [a_{ij}]$ is the matrix of the coefficients and $\mathbf{V}$ is a privacy-enforcing noise. Now, the problem of MIMO privacy can be formulated.

Problem 1 (MIMO Privacy)**.** *Consider a set of data users $[n]$ and a set of data users $[m]$. For any user $j \in [m]$, any subset of users $\mathcal{J} \subseteq [m]$, and any data owner $i \in [n]$, consider the mechanism in Equation* (1)*, and let $\mathcal{M}_{ij}$ be the sub-mechanism that releases Y_j and let $\mathcal{M}_{i\mathcal{J}}$ be the sub-mechanism that releases $\{Y_j\}_{j\in\mathcal{J}}$, i.e.*

$$\mathcal{M}_{ij}(u_i) = Y_j \quad \textit{and} \quad \mathcal{M}_{i\mathcal{J}}(u_i) = [Y_j]_{j\in\mathcal{J}}.$$

Design the noise $\mathcal{V}$ such that,

- *the mechanism $\mathcal{M}_{ij}$ is $(\epsilon_{ij}, \delta_{ij})$-private and*
- *for any group $\mathcal{J}$ and any owner i, there exists a user $j^* \in \mathcal{J}$ such that, if $\mathcal{M}_{ij^*}$ is (ϵ, δ)-private, then, $\mathcal{M}_{i\mathcal{J}}$ is also (ϵ, δ)-private.*

The first constraint of Problem 1 protects each owner's private data against each user, whereas, the second constraint provides privacy against adversarial coalitions. Specifically, for any coalition $\mathcal{J}$ and any target owner i, there exists a data user j^* who does not gain any additional knowledge about owner i (in the sense of privacy level) by participating in the coalition $\mathcal{J}$ and, thus, opts out of it.

E. A Basic Approach

A base approach can be established by reducing the MIMO system to n SIMO systems using the results in earlier work [11], where owners independently diffuse their private data. Specifically, owner i samples the Brownian motion $\{B_t^{(i)}\}_{t\geq 0}$ and responds to user j with a proxy Z_{ij} of her private data u_i

$$Z_{ij} = u_i + B^{(i)}_{\kappa(\epsilon_{ij},\delta_{ij})^{-1}}.$$

This ensures $(\epsilon_{ij}, \delta_{ij})$-privacy of owner i from user j, whereas, users cannot break the privacy guarantees by forming adversarial coalitions. Then, each user estimates the quantity of interest as

$$Y_j = \sum_{i\in[n]} a_{ij}\, Z_{ij} = q(\mathbf{u}) + \sum_{i\in[n]} a_{ij}\, B^{(i)}_{\kappa(\epsilon_{ij},\delta_{ij})^{-1}}.$$

Although such an *input-perturbation* approach protects owner's privacy, results in accumulating privacy-preserving noise from every input. Additionally, this approach cannot be extended to handle curious coalitions or de-incentivize only a specific set $\mathbb{J}$ of adversarial coalitions $\mathcal{J} \in \mathbb{J} \subset 2^{[m]}$.

III. MIMO Privacy through Semi-definite Programming

In this section, we employ the Gaussian mechanism and build a solution to a relaxed version of Problem 1. Specifically, we assume that the system operator adds Gaussian noise with zero mean and covariance matrix $\Sigma \in \mathbb{S}^m$, where $\mathbb{S}^m$ is the set of positive-semidefinite matrices of size $m \times m$, i.e.

$$\mathbf{Y} = A\,\mathbf{u} + \mathbf{V}, \quad \text{where} \quad \mathbf{V} \sim \mathcal{N}(\mathbf{0}_{m\times 1}, \Sigma). \qquad (2)$$

A. Analysis of a Coalition

In order to provide an approach to Problem 1, we need to analyze the effect of a coalition. To this end, consider a coalition of data users $\mathcal{J} \subseteq [m]$ and a target data owner $i \in [n]$. For the mechanism defined in Equation (2), the following lemma characterizes the privacy level that owner i receives against the group $\mathcal{J}$.

Lemma 3. *For a coalition $\mathcal{J}$ and a target owner i, the mechanism $\mathcal{M}_{i\mathcal{J}}$,*

$$\mathcal{M}_{i\mathcal{J}}(u_i) = [Y_j]_{j\in\mathcal{J}},$$

is (ϵ, δ)-private if

$$\kappa^2(\epsilon, \delta) \geq a_{i\mathcal{J}}^T\, \Sigma_{\mathcal{J}}^{-1}\, a_{i\mathcal{J}},$$

where $a_{i\mathcal{J}} = [a_{ij}]_{j\in\mathcal{J}} \in \mathbb{R}^{|\mathcal{J}|}$ and $\Sigma_J = [\Sigma_{jk}]_{j,k\in\mathcal{J}} \in \mathbb{S}^{|\mathcal{J}|}$.

Sketch of proof. We focus on the following probability as a function of u_i

$$\mathbb{P}\left(\mathcal{M}_{i\mathcal{J}}(u_i) = [y_j]_{j\in\mathcal{J}}\right). \qquad (3)$$

and we re-write the responses y_i as noisy observations of the private data u_i

$$y_j = \sum_{k\in[n]} a_{kj}\, u_k + V_j \Leftrightarrow$$

$$\frac{y_j}{a_{ij}} - \sum_{k\in[n]} \frac{a_{kj}}{a_{ij}} u_k = u_i + \frac{1}{a_{ij}} V_j \Leftrightarrow$$

$$z_{ij} = u_i + \frac{1}{a_{ij}} V_j, \quad \forall j \in \mathcal{J}$$

where z_{ij} is considered an observation; it does not depend on the noise V_j or the private data u_i. Next, consider the optimal Bayesian linear estimator

$$\hat{u}_i = \sum_{j\in\mathcal{J}} w_j\, z_{ij} = u_i + \sum_{i\in\mathcal{J}} \frac{w_j}{a_{ij}} V_j, \qquad (4)$$

for appropriate weights w_j with $\sum_{j\in\mathcal{J}} w_j = 1$ and let $\{o_1, \ldots, o_{|\mathcal{J}|-1}\}$ be $|\mathcal{J}| - 1$ orthogonal to $\hat{u}_i$ linear combinations of the observations z_{ij}. Then, the mechanism in Equation (3) can be viewed as the mechanism that releases $\hat{u}_i$ as in Equation (4) followed by a post-processing which appends to $\hat{u}_i$ the *independent* responses $\{o_1, \ldots, o_{|\mathcal{J}|-1}\}$.

The statement follows from observing that the mechanism in Equation (4) is a Gaussian mechanism with variance $\left(a_{i\mathcal{J}}^T\, \Sigma_{\mathcal{J}}^{-1}\, a_{i\mathcal{J}}\right)^{-1}$ and the resilience to post-processing theorem [3].

A formal but less intuitive proof follows by massaging Equation (3)

$$\begin{aligned}\mathbb{P}\left(\mathcal{M}_{i\mathcal{J}}(u_i)=[y_j]\right) &= \mathbb{P}\left(V_j = a_{ij}\, z_{ij} - a_{ij}\, u_i\right) \qquad (5)\\ &\propto e^{-\frac{1}{2}[a_{ij}\, z_{ij}-a_{ij}\, u_i]^T \Sigma_J^{-1}[a_{ij}\, z_{ij}-a_{ij}\, u_i]}\\ &= e^{-\frac{1}{2}C_1\, u_i^2 + C_2\, u_i - \frac{1}{2}C_3},\end{aligned}$$

with

$$\begin{aligned}&C_1 = [a_{ij}]^T\, \Sigma_{\mathcal{J}}^{-1}\, [a_{ij}], \quad C_2 = [z_{ij}]^T\, \Sigma_{\mathcal{J}}^{-1}\, [a_{ij}],\\ &C_3 = [z_{ij}]^T\, \Sigma_{\mathcal{J}}^{-1}\, [z_{ij}],\end{aligned}$$

where, for clarity, we have dropped the subscript $j \in \mathcal{J}$ in all stacked vectors, e.g. $[y_j]$. Next, we compare Equation (5) to the probability density of a Gaussian mechanism $\mathcal{W}$ that releases $u_i + W$, where $W \sim \mathcal{N}(0, \sigma_w^2)$,

$$\mathbb{P}(u_i + W = w) \propto e^{-\frac{1}{2}u_i^2\, \sigma_w^{-2} + u_i\, w\, \sigma_w^{-2} - \frac{1}{2}w^2 \sigma_w^{-2}}, \qquad (6)$$

and we identify the terms

$$\sigma_w^{-2} = [a_{ij}]^T \Sigma_J^{-1} [a_{ij}], \quad w = \frac{[a_{ij}]^T \Sigma_{\mathcal{J}}^{-1} [a_{ij} z_{ij}]}{[a_{ij}]^T \Sigma_{\mathcal{J}}^{-1} [a_{ij}]}.$$

Since Equation (6) satisfies the (ϵ, δ)-privacy constraint if $\kappa(\epsilon, \delta) \leq \sigma_w^{-1}$, then, Equation (3) also satisfies the (ϵ, δ)-privacy constraint and, thus, $\mathcal{M}_{i\mathcal{J}}$ is (ϵ, δ)-private.

□

B. Design of Covariance Matrix

We now formulate the problem of designing the covariance matrix Σ which provides a solution to Problem 1. Specifically, we provide a solution to Problem 1 by formulating the optimization problem in Theorem 4.

Theorem 4. *Consider n data owners with private data $\mathbf{u} = [u_i]_{i\in[n]} \in \mathbb{R}^n$ and m data users where $\kappa_{ij} = \kappa(\epsilon_{ij}, \delta_{ij})$ is the privacy level of owner i against user j. Then, consider the mechanism $\mathcal{M}$ that securely returns Y_j to user j*

$$\mathcal{M}\mathbf{u} = A\,\mathbf{u} + \mathbf{V} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix},$$

where $\mathbf{V} \sim \mathcal{N}(\mathbf{0}_{m\times 1}, \Sigma)$ and $\Sigma \in \mathbb{S}^m$ satisfies the constraints

$$\Sigma_{jj} \geq \frac{a_{ij}^2}{\kappa_{ij}^2},\ \forall i \in [n],\ j \in [m] \quad \textit{and}$$

$$\frac{1}{a_{i\mathcal{J}}^T\, \Sigma_{\mathcal{J}}^{-1}\, a_{i\mathcal{J}}} \geq \min_{j\in\mathcal{J}} \frac{\Sigma_{jj}}{a_{ij}^2},\ \forall \mathcal{J} \subseteq [m],\ i \in [n].$$

Then, mechanism $\mathcal{M}$ is multi-input multi-out private. Specifically, $\mathcal{M}$ satisfies the privacy requirements and does not incentivize adversarial coalitions.

Sketch of proof. The first set of constraints follows from the Gaussian mechanism and requires that owner's i data remains private from user j. The second set of constraints refers to the correlation of the responses that different users receive and is interpreted as follows. For any coalition $\mathcal{J}$ and any targeted owner i, according to Lemma 3, the most privacy-violating inference of the adversarial coalition has variance

$$\frac{1}{a_{i\mathcal{J}}^T\, \Sigma_{\mathcal{J}}^{-1}\, a_{i\mathcal{J}}}.$$

We, then, require that there exists a colluding user $j^* \in \mathcal{J}$ that already has inferred user's i data u_i with smaller variance. □

Theorem 4 provides necessary conditions for a mechanism to be MIMO private. On the downside, Theorem 4 has exponentially many in the number of users constraints. Additionally, the covariance matrix may be over-constrained and, thus, the feasibility problem might be infeasible. Lastly, the second set of constraints is non-convex which makes the design of the covariance matrix challenging.

Nonetheless, the expressitivity of Theorem 4 allows focusing only a subset of potential coalitions. For example, the system designer can choose to focus only on coalitions up to fixed size or ignore coalitions across non-cooperative groups of users. In particular, for agents that act both as data owners and data users, we can ignore coalitions where agent i participates and attempt to attack herself.

Next, in Theorem 5 we propose a convex relaxation which provides privacy of each owner from each user and approximately de-incentivizes users' coalitions.

Theorem 5. *In the setting of Theorem 4, let the covariance matrix Σ be the solution of the following optimization problem*

$$\begin{aligned}\underset{\Sigma\in\mathbb{S}^m}{\textit{minimize}} \quad & \|\textit{diag}(\Sigma)\|_p\\ \textit{s.t.} \quad & \Sigma_{jj} \geq \max_i \frac{a_{ij}^2}{\kappa_{ij}^2},\ \forall j\\ & \begin{bmatrix} D_{i\mathcal{J}}^{-1} & a_{i\mathcal{J}}^T \\ a_{i\mathcal{J}} & \Sigma_{\mathcal{J}} \end{bmatrix} \succeq 0,\ \forall \mathcal{J}, i,\end{aligned}$$

where

$$D_{i\mathcal{J}} = \min_{j\in\mathcal{J}} \left[a_{ij}^{-2}\, \max_{l\in[n]} \left(\frac{a_{lj}}{\kappa_{lj}} \right)^2 \right].$$

Then, $\mathcal{M}$ is approximately MIMO private. Specifically, $\mathcal{M}$ provides κ_{ij} privacy of owner i from user j and approximately de-incentives coalitions.

Proof. Let Σ satisfy the constraints of Theorem 4. Then, for any $\mathcal{J} \subseteq [m]$ and $i \in [n]$

$$\frac{1}{a_{i\mathcal{J}}^T\, \Sigma_{\mathcal{J}}^{-1}\, a_{i\mathcal{J}}} \geq \min_{j\in\mathcal{J}} \frac{\Sigma_{jj}}{a_{ij}^2} \geq \min_{j\in\mathcal{J}} \left[\frac{1}{a_{ij}^2} \max_{l\in[n]} \left(\frac{a_{lj}^2}{\kappa_{lj}^2} \right) \right] = D_{i\mathcal{J}}.$$

Since $D_{i\mathcal{J}} > 0$ and using Schur's complement we get

$$\frac{1}{a_{i\mathcal{J}}^T\, \Sigma_{\mathcal{J}}^{-1}\, a_{i\mathcal{J}}} \geq D_{i\mathcal{J}} \ \Leftrightarrow\ \begin{bmatrix} D_{i\mathcal{J}}^{-1} & a_{i\mathcal{J}}^T \\ a_{i\mathcal{J}} & \Sigma_{\mathcal{J}} \end{bmatrix} \succeq 0.$$

Regarding accuracy, user's j response has variance Σ_{jj} and, thus, we choose an objective function that minimizes the diagonal elements of Σ. This formulation approximates the

design of a MIMO private mechanism using a convex semi-definite program. □

IV. Case Study: Smooth Local Averaging

For a case study, we consider n agents who act both as data owners and data users. Specifically, we assume that each user i has a scalar private data $u_i \in \mathbb{R}$. For example, data u_i can capture the health status of agent i or a privately computed exposure to risk (e.g. debt-to-equity ratio). Then, each agent wishes to estimate the smooth local average $q_i(\mathbf{u})$ of its neighborhood. For instance, such an average captures the probability of an agent getting infected by other nodes or the cascade exposure to risk.

Specifically, we consider n agents each placed at location $x_i \in [0,1]^2$ uniformly randomly. Then, each agent wishes to compute a smooth local average q_i of the private data $\mathbf{u}$,

$$q_i(\mathbf{u}) = \sum_{j \neq i} \|x_i - x_j\|^{-1} u_j.$$

In words, agent i weighs input more from nearby agents than from more distant ones. Additionally, we assume that the agents are connected by an undirected graph $G = ([n], E)$, where $E \subseteq [n]^2$ captures the friendships; agents i and j are connected with an edge $(i,j) \in E$ whenever they are friends. Then, let $d : [n]^2 \to \mathbb{R}_+$ be a measure of how far away agents i and j lie in this graph, captures the trust level between the two agents, and quantifies the privacy level required between i and j as follows

$$\epsilon_{ij} = d^{-1}(i,j), \quad \delta_{ij} = .01\,\epsilon_{ij}, \quad \kappa_{ij} = \kappa(\epsilon_{ij}, \delta_{ij}).$$

Here, we choose $d(i,j)$ to be the resistance distance [20], [21]. Next, we apply Theorem 5 in order to design a MIMO private mechanism. Specifically, we consider the following SDP, where we only consider coalitions of size up to $m_{\max}$ and worst variance of the responses that agents receive.

$$\begin{aligned} \underset{\Sigma \in \mathbb{S}^n}{\text{minimize}} \quad & \max_i \Sigma_{ii} \\ \text{s.t.} \quad & \Sigma_{ii} \geq \max_j \left(\frac{a_{ij}}{\kappa_{ij}}\right)^2, \ \forall i \in [n]; \\ & \begin{bmatrix} D_{i\mathcal{J}}^{-1} & a_{i\mathcal{J}}^T \\ a_{i\mathcal{J}} & \Sigma_{\mathcal{J}} \end{bmatrix} \succeq 0, \\ & \forall\, \mathcal{J} \subseteq [n] \text{ s.t. } |\mathcal{J}| \leq m_{\max} \text{ and } \forall\, i \in [n] \setminus \mathcal{J}. \end{aligned}$$

We evaluate our approach by computing the worst-case incentive to form a coalition; given any potential coalition $\mathcal{J}$ and any targeted agent i, we compute how much more information the most-informed agent $j^* \in \mathcal{J}$ can extract about the targeted agent by participating in the coalition. Formally, we define Incentive as

$$\text{Incentive} := \max_{\substack{i \in [n], \\ i \notin \mathcal{J} \subseteq [n] \\ \text{s.t. } |\mathcal{J}| \leq m_{\max}-1}} \ \min_{\substack{j \notin \mathcal{J} \\ j \neq i}} \frac{\kappa \text{ of } i \text{ from } \mathcal{J} \cup \{j\}}{\kappa \text{ of } i \text{ from } j},$$

where the expression κ of i from j is the privacy level that protects agent's i private data from the response that agent j receives; as a reminder, larger values correspond to less privacy. Specifically, $\min$ chooses the most informed agent j^* and the $\max$ chooses the worst-case option over all possible coalitions and possible targets. Figure 2 plots this quantity for different sizes of the network $n \in [2, 16]$ and coalition sizes up to $m_{\max} \in [2, 4]$.

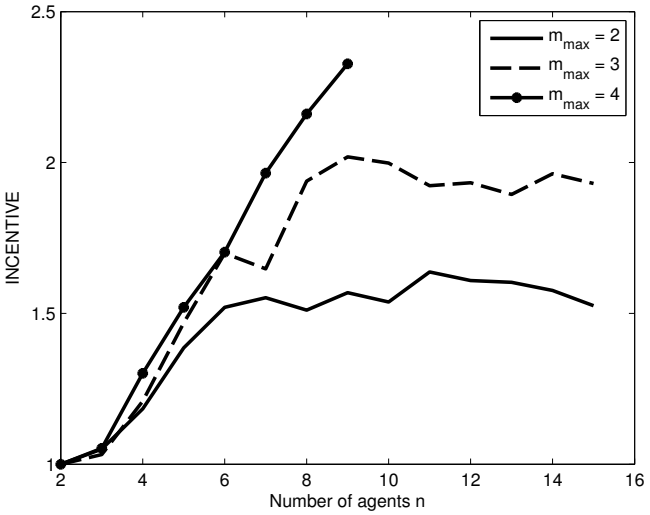

Fig. 2: The semidefinite constraints are not binding and, thus, there exists some incentive for agents to form adverserial coalitions. Allowing coalitions of size up to $m_{\max} = 2, 3, 4$, we compute Incentive which captures this gap. Note that Incentive ≥ 1 and larger values result to stronger incentives for agents to collaborate.

Furthermore, we evaluate the effect of performance of the SDP in the following two ways:

- We compare the variance $\sqrt{\Sigma_{ii}}$ of the response Y_i that agent i receives to that of the baseline introduced in Subsection II-E and is based on the SIMO case [11]. Specifically, Figure 3 plots the ratio Improvement defined as

$$\text{Improvement} := \max_{i \in [n]} \frac{\sqrt{\sum_{\substack{j \in [n], \\ j \neq i}} \left(\frac{a_{ij}}{\kappa_{ij}}\right)^2}}{\sqrt{\Sigma_{ii}}},$$

  where the numerator is the variance of the baseline. As argued in Subsection II-E, the baseline completely de-incentivizes coalitions of any size. Importantly, the baseline assumes that each agent i retains all the information $\{Z_{ji}\}_{j \in [n]}$ as side information, whereas, the proposed approach does not require agents to retain such side information.
- We compare the proposed method to the case where we do not protect the private data against coalitions. This figure of merit captures how performance degrades in order to defeat coalitions. Figure 4 compares the variance $\sqrt{\Sigma_{ii}}$ that agent i observes to the variance $\max_i \frac{a_{ij}}{\kappa_{ij}}$ that agent i would ideally observe in the absence of the rest of the users:

$$\text{Inefficiency} := \max_{i \in [n]} \frac{\sqrt{\Sigma_{ii}}}{\max_{\substack{j \in [n], \\ j \neq i}} \frac{a_{ij}}{\kappa_{ij}}}.$$

In all cases, we averaged the figures of merit over 20 executions.

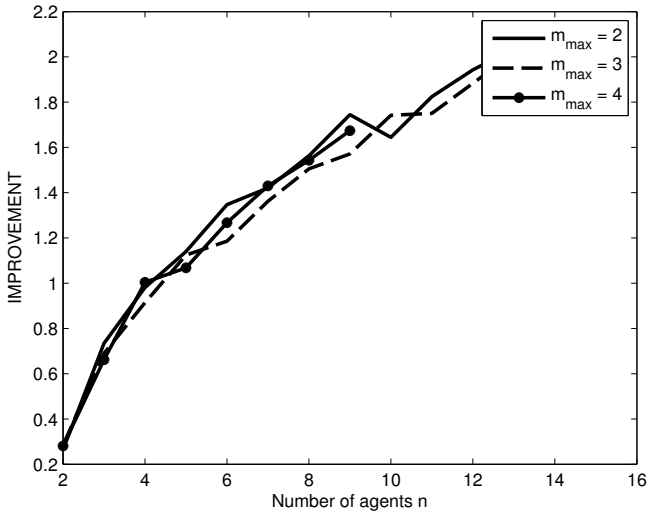

Fig. 3: A baseline approach to MIMO privacy utilizes [11], where agents independently diffuse their private data. The figure of merit IMPROVEMENT catpures the performance of proposed approach to such a baseline. Althouhg, for very small sizes, the baseline performs better, the proposed approach outperforms the baseline for larger networks.

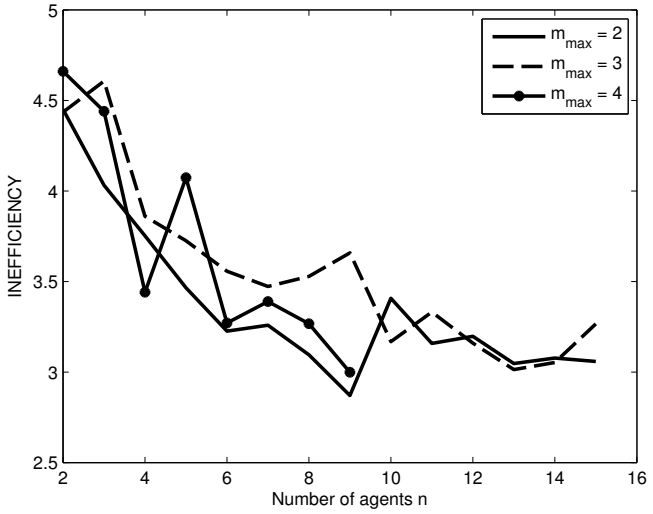

Fig. 4: The existence of multiple users force the privacy-enforcing mechanism to inject more noise. We quantify the toll on the accuracy of the responses by plotting INEFFICIENCY which compares the amount of noise added to that of a mechanism that ignores possible coalitions and adds only privacy-preserving noise.

V. DISCUSSION

Concluding, we introduced a taxonomy of the privacy-aware approaches within differential privacy. We motivated the use of multi-input multi-output private mechanism which return different pieces of private information to different users. Such mechanisms need to account for possible interactions between users. To this end, we employed approximate differential privacy and we introduced a model with real-valued private data and linear queries. After deriving the privacy constraints for such a MIMO private mechanism, we derived a relaxed SDP with *approximately* de-incentivizes coalitions. Finally, we evaluated the proposed approach on synthetic data and quantified the performance loss due to the co-existence of multiple users.

Future work can consider richer models of interactions of users such as *curious coalitions* introduced earlier. Additionally, an approach for pure differential privacy is would be interesting as well as extending the results to non-linear queries. Furthermore, efficiently solving the SDP for large populations of users is problematic and distributed approaches are helpful. Finally, as described earlier, we defer for the future the design of MIMO private mechanisms that return side information as a, probably *exact*, de-incentivizing means.

REFERENCES

[1] L Atzori, A Iera, and G Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[2] G Adomavicius and A Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

[3] C Dwork, F McSherry, K Nissim, and A Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, pages 265–284. Springer, 2006.

[4] F McSherry and I Mironov. Differentially private recommender systems: building privacy into the net. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 627–636. ACM, 2009.

[5] C Dwork, M Naor, T Pitassi, and GN Rothblum. Differential privacy under continual observation. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 715–724. ACM, 2010.

[6] J Le Ny and GJ Pappas. Differentially private filtering. *Automatic Control, IEEE Transactions on*, 2014.

[7] J Hsu, Z Huang, A Roth, and ZS Wu. Jointly private convex programming. *arXiv preprint arXiv:1411.0998*, 2014.

[8] Z Huang, S Mitra, and N Vaidya. Differentially private distributed optimization. In *arXiv preprint arXiv:1401.2596*, 2014.

[9] H Ebadi, D Sands, and G Schneider. Differential privacy: Now it's getting personal. In *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 2015.

[10] M Alaggan, S Gambs, and AM Kermarrec. Heterogeneous differential privacy. *arXiv preprint arXiv:1504.06998*, 2015.

[11] F Koufogiannis and G Pappas. Diffusing private data over networks. *arXiv preprint arXiv:1511.06253*, 2015.

[12] C Dwork and A Roth. The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 2013.

[13] Y Wang, Z Huang, S Mitra, and GE Dullerud. Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems. In *IEEE Conference on Decision and Control*, 2014.

[14] M Hardt and K Talwar. On the geometry of differential privacy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 705–714. ACM, 2010.

[15] Z Huang, S Mitra, and G Dullerud. Differentially private iterative synchronous consensus. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, pages 81–90. ACM, 2012.

[16] S Han, U Topcu, and GJ Pappas. Differentially private convex optimization with piecewise affine objectives. In *IEEE Conference on Decision and Control*, 2014.

[17] F Koufogiannis, S Han, and GJ Pappas. Computation of privacy-preserving prices in smart grids. In *IEEE Conference on Decision and Control*, 2014.

[18] J Le Ny, A Touati, and GJ Pappas. Real-time privacy-preserving model-based estimation of traffic flows. In *ICCPS'14: ACM/IEEE 5th International Conference on Cyber-Physical Systems*, 2014.

[19] M Kearns, A Roth, ZS Wu, and G Yaroslavtsev. Privacy for the protected (only). *arXiv preprint arXiv:1506.00242*, 2015.

[20] DJ Klein and M Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.

[21] W Xiao and I Gutman. Resistance distance and laplacian spectrum. *Theoretical Chemistry Accounts*, 110(4):284–289, 2003.