Diffusing Private Data Over Networks

Fragkiskos Koufogiannis and George J. Pappas, *Fellow, IEEE*

***Abstract*—The emergence of social and technological networks has enabled rapid sharing of data and information. This has resulted in significant privacy concerns, where private information can be either leaked or inferred from public data. The problem is significantly harder for social networks, where we may reveal more information to our friends than to strangers. Nonetheless, our private information can still leak to strangers as our friends are their friends and so on. In order to address this important challenge, in this paper, we present a privacy-preserving mechanism that enables private data to be diffused over a network. In particular, whenever a user wants to access another users' data, the proposed mechanism returns a differentially private response that ensures that the amount of private data leaked depends on the distance between the two users in the network. While allowing global statistics to be inferred by users acting as analysts, our mechanism guarantees that no individual user, or a group of users, can harm the privacy guarantees of any other user. We illustrate our mechanism with two examples: one on synthetic data where the users share their global positioning system coordinates, and the other on a Facebook ego network where a user shares her infection status.**

***Index Terms*—Data privacy, social network services.**

I. Introduction

In THE era of social networks, individuals' profiles include an increasing amount of private information such as their age, location, and shopping habits. Besides users' intention to share this information for social interaction, their private data enable systems such as location-based services and collaborative recommender engines, that is, systems that are not part of their friendship network. Therefore, although users consent to share their private data with their friends, when this is not the case, severe privacy concerns are raised.

Traditionally, these privacy concerns are mitigated by restricting access rights (e.g., on Facebook); more precisely, only users indicated as friends are granted access to each user's personal information. However, such an approach has severe limitations as follows: first, this scheme is inflexible since users cannot be partitioned into exactly two groups, that is, friends and strangers. Instead, privacy concerns gradually increase from family members and friends to acquaintances and, finally, strangers. Second, a scheme based on access rights keeps private information local, which limits the ability of inferring statistics of the entire network. For instance, consider users acting as network analysts who are interested in statistics over the whole population of the social network such as population density maps and epidemic monitoring. This limits the utility of the network. Hence, an alternative mechanism that allows global statistics on the whole population and respecting individuals' privacy is needed. For example, when individuals share their shopping habits, they contribute to the social welfare by allowing for more efficient recommender systems.

Frameworks for providing privacy guarantees are differential privacy [1] and information-theoretic privacy [2]. However, most of the previous approaches in both frameworks do not consider variable privacy levels in a network, where the level of privacy depends on friendship distance. Hereafter, we consider a network where users wish to share their private data under privacy guarantees, where the strength of these guarantees is quantified by the distance on the graph. Within the context of a social network, users wish to communicate accurate information with little privacy guarantees to their close friends, whereas they desire strong privacy guarantees whenever their private data are communicated to distant areas of the network. From the point of view of a user acting as a network analyst, statistics over the whole network need to be possible while ensuring privacy guarantees. In the model proposed here, we assume a trusted network operator and, thus, we do not provide privacy guarantees of the users against the operator. Instead, we protect the users' privacy from all other users of the network. In addition, our work targets networks, in general, that may lack a single network operator. In that case, users have different trust levels against different users and, therefore, wish to share their private data with different privacy guarantees. In that case, the trust levels can either be hand-picked by the users or can be inferred from the network structure. For the scope of this work, we assume that these trust levels are given.

Multiple privacy-preserving frameworks that formalize privacy guarantees have appeared in the literature, for example, [2] and [3]. Commonly, privacy-preserving approaches add artificial noise to the accessed private data. This noise is designed such that the resulting response conveys little information about the private data. Specifically, an information-theoretic approach [2] constrains the mutual information between the private data and the released signal. Similarly, differential privacy [1], [3]

Manuscript received October 9, 2015; revised June 30, 2016, September 29, 2016, and January 17, 2017; accepted February 18, 2017. Date of publication February 23, 2017; date of current version September 17, 2018. 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, a Semiconductor Research Corporation program sponsored by Microelectronics Advanced Research Corp. and the Defense Advanced Research Projects Agency, and in part by the National Science Foundation (NSF) under Grant CNS-1505799 and the Intel-NSF Partnership for Cyber-Physical Systems Security and Privacy. A preliminary version of this work was presented at the 2016 American Control Conference and considered only single-dimensional private data. Recommended by Associate Editor B. Sinopoli.

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: fkouf@seas.upenn.edu; pappasg@seas.upenn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCNS.2017.2673414

requires that the statistics of the noisy response should be almost independent of perturbations of the private data. In this work, we adopt the framework of differential privacy because of its strong privacy guarantees, yet the underlying problem can be formulated under other privacy frameworks.

Within differential privacy, an extensive family of privacy-preserving mechanisms has emerged. The application range of these mechanisms varies from solving linear problems [4], [5], distributed convex optimization [6], Kalman filtering [7], and consensus that protects the network topology [8] to smart metering [9], [10] and traffic flow estimation [11]. In particular, the problems introduced in the aforementioned line of research share a common underlying abstract problem that can be stated as follows: given the private data and a predefined privacy requirement, we need to design a differentially private algorithm, called mechanism, which accurately approximates a desired quantity. Then, a single sample from the mechanism is published and is used as a proxy for the exact response, so a curious user cannot confidently infer the original private data. Instead of considering a single privacy level and assuming that the responses are publicly released, that is, everyone receives the same response, in this paper, we consider the novel problem of assigning different privacy levels for different users. Moreover, contrary to publishing the responses, we assume that they are securely communicated to each user. Therefore, the aforementioned works do not address the problem introduced here. Furthermore, in [12] and [13], multicomponent private data and different privacy levels for different components are considered, that is, in a user's profile, typically, stronger privacy is required for the component representing salary compared to that of age. Contrary to previous works that focus on variable privacy for different components of one's data, our paper focuses on different privacy levels that depend on friendship status. The work closest to ours is [14], where the problem of relaxing the privacy level after, for example, supplementary payments to the owners of the sensitive data was considered. Although some of the tools in [14] are leveraged to provide a solution, here, we consider a different problem, which is the problem of releasing sensitive data to multiple parties with different privacy levels and has not been studied before.

This paper is organized as follows. Section II informally describes the problem of diffusing private data across a network, then, provides a model of the system, reviews differential privacy, and derives a formal statement of the problem. Section III introduces a composite mechanism based on a Markov stochastic process and presents low-complexity algorithmic implementations of this mechanism. We demonstrate our approach with two illustrative examples in Section IV: one on synthetic network where a user releases her global positioning system (GPS) coordinates and one on a Facebook ego network where a user shares her infection status.

II. Problem Formulation

Here, the problem of releasing private information over networks (i.e., social networks) is formulated. First, we provide an informal description of the problem whose formal statements are presented in Problems 1 and 2 in the end of this section. Let a network be represented as a graph $G = (\mathcal{V}, \mathcal{E})$, where each node $i \in \mathcal{V}$ is a user and each edge $(i, j) \in \mathcal{E}$ represents a friendship relation between users i and j. Also, we assume that each user i owns a private data $u_i \in \mathcal{U}$, where $\mathcal{U}$ is the set of possible private data, and wishes to share their private data with the rest of the users under privacy guarantees. Specifically, user i generates an approximation y_{ij} of u_i and securely communicates y_{ij} to user j. More specifically, each user i requires her data u_i to be $\epsilon(d_{ij})$-differential privacy against user j (differential privacy is overviewed in Section II-B), where d_{ij} is a distance function $d_{ij} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$ and $\epsilon : \mathbb{R}_+ \to \mathbb{R}_+$ is a decreasing function that converts distance d to a privacy level $\epsilon(d)$. Therefore, we need to design a mechanism that generates accurate responses $\{y_{ij}\}_{j\in\mathcal{V}}$, while satisfying different privacy constraints for different recipients based on the distance on the network. Specifically, accuracy is meant in the expected mean-squared error sense; the response y_{ij} should an accurate but private proxy of the private data u_i.

In order to formalize these statements as in Problem 1 and, eventually, in Problem 2, we need to revisit some concepts and known results. Subsequently, modeling assumptions are presented in Section II-A, whereas differential privacy is briefly reviewed in Section II-B. We present a conventional approach, that is, a scheme based on access rights in Section II-C, whereas Section II-D formally presents the problem of diffusing private data over networks.

A. System Model

Consider a network represented as a graph G with $|\mathcal{V}| = N$ nodes. For simplicity, we assume that the graph is undirected and unweighted, although this assumption can be removed. Each node $i \in \mathcal{V}$ represents a user and $(i, j) \in \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the friendship relation between users i and j. Each user owns a private data $u_i \in \mathcal{U}$. Typical examples of private data include the following.

1) *Timestamps:* Let $u_i \in \mathbb{R}$ be a real-valued representation of a timestamp such as date of birth, for example, unix time [15] is a popular way of mapping timestamps to integers.
2) *Location:* Let $u_i \in \mathbb{R}^2$ be the GPS coordinates of the residence of an individual i.
3) *Binary states:* Let $u_i \in \{0, 1\}$ indicate user's i status such as infected or healthy, married or single, etc.

Furthermore, we want the severity of the privacy concerns to scale with the distance between two nodes. Typical choices for the distance function $d : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$ are as follows.

1) *Shortest path distance:* Let d_{ij} be the length of the shortest path connecting nodes i and j.
2) *Resistance distance:* Let d_{ij} be the resistance between nodes i and j, where the edges of graph G are associated with unit resistors [16].

Distance functions that are more suitable for social networks have been proposed in the literature; for example, see [17, Fig. 2]. In this paper, we assume that these distances are given and we focus on preserving the privacy of the users' data. Moreover, distances d_{ij} may not depend only on the structure of the

underlying network but also on the attributes of the nodes. For instance, a family relationship between users i and j may lead to a smaller value of d_{ij}. Furthermore, directed edges (e.g., blocked users) can also be allowed in social network scenarios.

Additionally, we assume the existence of a trusted central authority. Users provide their noiseless private data to this authority—already the case with modern social networks—which executes a privacy-preserving mechanism, adds noise, and securely communicates the responses to each user. Then, differential privacy protects user's i private data from inference attacks by an adversarial user j while honest users locally run any postprocessing such as a recommendation system. The assumption of a central authority can be relaxed by considering honest but curious users with only local secure communications.

B. Differential Privacy

Differential privacy is a formal framework that provides rigorous privacy guarantees. Differentially private algorithms add noise in order to make it hard for a curious user to infer whether someone's data has been used in the computation. The dependence of this noisy response on the private data is required to be bounded, as formally stated in Definition 1. The strength of this bound is quantified by the nonnegative parameter $\epsilon \in [0, \infty)$, called privacy level, where smaller values of ϵ imply stronger privacy guarantees. Moreover, an adjacency relation $\mathcal{A}$ is a symmetric binary relation over the set of private data $\mathcal{U}$, which includes the pairs of private data (u, u') that should be rendered almost indistinguishable. Furthermore, a mechanism[1] $Q : \mathcal{U} \to \Delta(\mathcal{Y})$ is a randomized map from the space of private data to the space of responses.

Definition 1 (Differential privacy [1]): Let $\epsilon > 0$, $\mathcal{U}$ be the space of private data, and $\mathcal{A} \subseteq \mathcal{U} \times \mathcal{U}$ be an adjacency relation. The mechanism $Q : \mathcal{U} \to \Delta(\mathcal{Y})$ is ϵ-differentially private if

$$\mathbb{P}(Qu \in \mathcal{S}) \leq e^{\epsilon}\, \mathbb{P}(Qu' \in \mathcal{S}), \quad \text{for all } \mathcal{S} \subseteq \mathcal{Y}$$

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

In this paper, we consider real-valued private data $\mathcal{U} = \mathbb{R}^n$ and the following adjacency relation:

$$(u, u') \in \mathcal{A}_2 \Leftrightarrow \|u - u'\|_2 \leq \alpha \tag{1}$$

where $\alpha \in \mathbb{R}_+$ is a small constant. Practically, adjacency relation $\mathcal{A}_2$ requires that, given the output of mechanism Q, a curious user should not be able to infer the private input u within a radius of α. A popular differentially private mechanism is the Laplace mechanism, which is near optimal [10], [18], is used as a building block for many mechanisms, and is described next.

Theorem 2 (Laplace mechanism [1]): Consider the mechanism $Q : \mathbb{R}^n \to \Delta(\mathbb{R}^n)$ that adds Laplace distributed noise:

$$Qu = u + V, \quad \text{where } V \sim \text{Lap}\left(\frac{\alpha}{\epsilon}\right)$$

where Lap(b) has density $d\mathbb{P}(V = v) = e^{-\frac{\|v\|_2}{b}}$. Then, mechanism Q is ϵ-differential private under adjacency relation $\mathcal{A}_2$.

[1]For a set $\mathcal{T}$ and a rich-enough σ-algebra $\mathcal{T}$ on it, we denote the set of all probability measures on $(T, \mathcal{T})$ with $\Delta(T)$. Specifically, for Euclidean spaces $T = \mathbb{R}^n$, we consider the Borel's σ-algebra.

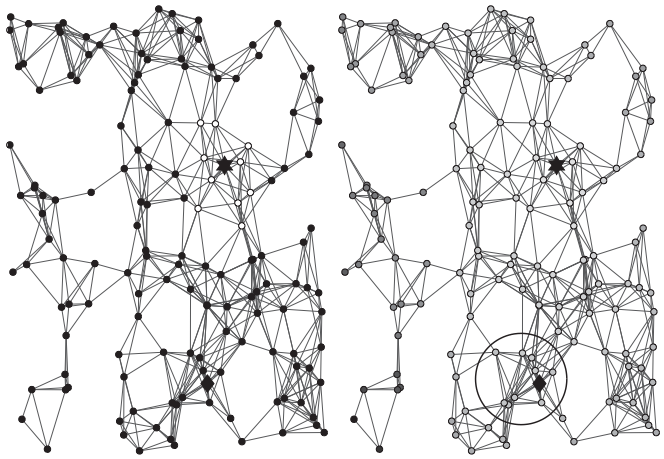

Fig. 1. Synthetic network with 150 nodes and 1256 edges is shown. Each node represents a user of the network and each edge indicates a friendship. The user indicated with the star wishes to share her sensitive information with the rest of the network. Privacy concerns can be addressed by managing access privileges. Under an access right scheme (a), only friends of the starred user (white nodes) are granted access to the exact information, whereas any other member (black nodes) have no access. Such a scheme partitions users to only two groups: friends and strangers. Moreover, each user has access only to local information and cannot estimate the global state of the network. Therefore, any estimator constructed by the diamond user will be independent of the data of the starred user and, thus, biased. On the other hand, (b) proposes an approach where users' privacy concerns scale with the distance from others. Friends (lighter-colored nodes) receive a less noisy version of the private data, whereas strangers (darker-colored nodes) receive only heavily perturbed versions. Despite the increased noise, estimates of aggregate statistics are possible. However, coalitions might be encouraged and initial privacy guarantees can quickly degrade. For example, users within the circle can combine their estimates and infer the private data of the starred user. (a) Access right scheme. (b) Distance-based scheme.

C. Access Rights Scheme

Now, we describe a typical approach for handling privacy concerns in social network while highlighting its limitations and motivating the need for a more sophisticated privacy-aware approach. Fig. 1(a) shows a synthetic network with 150 nodes, where the starred node wishes to share her sensitive information with the rest of the network. Privacy concerns can be handled by regulating access privileges. For example, friends of a user can access her data, whereas every other user cannot. Such a scheme has limitations. On one hand, users are coarsely partitioned to friends and strangers as depicted in Fig. 1(a); friends of the star-labeled user are colored white, whereas strangers are colored black. Instead, the distance between two users can be more finely quantified by a real-valued function. On the other hand, each user has access only to neighboring information. Although restricting access rights settles privacy concerns, computing global statistics on the network is impossible, limiting the utility of the network. Indeed, any estimator of global quantities (mean value, histogram, etc.) will be biased. Therefore, users may choose to collaborate, merge their local information, and damage any privacy guarantees. Fig. 1(b) overcomes these limitations by defining a distance function $d : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$, which quantifies the strength of the privacy concerns. In this case, users share

privacy-aware versions of their profile with every member of the network.

D. Diffusing Sensitive Information Over a Social Network

Under the modeling introduced in Section II-A, we pose the problem of designing a mechanism that diffuses private data over a network in the following.

Problem 1: Design a privacy-aware mechanism $Q : \mathcal{U} \to \Delta\left(\mathcal{U}^N\right)$ that privately releases user's i sensitive data $u_i \in \mathcal{U}$ over a social network. Specifically, design mechanism Q that generates N responses $\{y_j\}_{j=1}^N$, where y_{ij} is the securely communicated response to user j. Furthermore, for the adjacency relation (1) (where, for simplicity, $\alpha = 1$), the mechanism Q needs to satisfy the following properties:

1) *Variable privacy:* The mechanism must generate the response y_{ij} for private data u_i, which is $\epsilon(d_{ij})$-differential private.
2) *Optimal utility:* Response y_{ij} must be an accurate approximation of the sensitive data u_i, that is, for real-valued private data, it should minimize the expected squared-error

$$\mathbb{E}_Q \|y_{ij} - u_i\|_2^2 \quad \forall i, j \in \{1, \dots, n\}.$$

Although the utility part of the problem is stated in a multiobjective sense, it can be stated in a single objective

$$\sum_{i,j \in \{1,\dots,n\}} \mathbb{E}_Q \|y_{ij} - u_i\|_2^2.$$

Specifically, whenever individual i shares her sensitive information to another individual j, she requires $\epsilon(d_{ij})$-differential privacy, where $\epsilon(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is a decreasing function that converts a distance d to a privacy level $\epsilon(d)$. People residing close (w.r.t. a distance) to individual i receive a loose privacy constraint $\epsilon_{ij} \gg 1$, whereas strangers get noisier versions $\epsilon_{ij} \ll 1$.

Problem 1 admits a straightforward but unsatisfying approach. Let $y_{ij} = u_i + V$, where $V \sim \text{Lap}\left(\epsilon(d_{ij})^{-1}\right)$, independently for each user $j \in \mathcal{V}$. Subsequently, a group of users $j \in A \subseteq \mathcal{U}$ have the incentive to collaborate share their estimates $\{y_{ij}\}_{j \in A}$ in order to derive a more accurate estimator y_A of u_i described by

$$y_A = \sum_{j \in A} w_j y_{ij}.$$

Fig. 1(b) depicts a group of users forming such a coalition. The possibly large group A resides far away from the user indicated by the star, $d_{ij} \gg 1, \forall j \in A$. Although each user j in the group A receives a highly noisy estimate of u_i, estimator y_A is more accurate. The composition theorem of differential privacy [1] guarantees only $(\sum_{j \in A} \epsilon(d_{ij}))$-privacy, which can be rather looser than each of the $\epsilon(d_{ij})$-privacy guarantees; larger values of ϵ imply less privacy.

Therefore, Problem 1 is subject to coalition attacks. Thus, we restate Problem 1 by requiring that any group A that exchanges their estimates $\{y_{ij}\}_{j \in A}$ cannot produce a better estimator of u_i than the best estimator among the group y_{ij^*}, where $j^* = \arg\min_{j \in A} d_{ij}$ is the user closest to user i. This problem can be stated as follows.

Problem 2: Design a privacy-aware mechanism $Q : \mathcal{U} \to \Delta\left(\mathcal{U}^N\right)$ that releases an approximation of user's i sensitive data $u_i \in \mathcal{U}$ over a social network. Specifically, mechanism $\mathcal{M}$ generates N responses $\{y_{ij}\}_{j=1}^N$ and securely communicates response y_{ij} to user j. Mechanism Q needs to satisfy the following.

1) *Privacy:* For any group of users $A \subseteq \mathcal{V}$, response $\{y_{ij}\}_{j \in A}$ must be $\max_{j \in A} \epsilon(d_{ij})$-differential private.
2) *Performance:* Response y_{ij} must be an accurate approximation of the sensitive data u_i.

Here, the performance of the mechanism is captured by the need that each response y_{ij} approximated private data u_i as accurately as possible. Specifically, response y_{ij} is not a noiseless copy of data u_i because of the need for 1) ϵ_{ij}-privacy of user i against user j and 2) guarding against coalitions due to the existence of multiple users. The first limitation stems from the privacy–utility tradeoff, whereas, for the second, we will prove that the existence of multiple users apart from i and j does not incur any further performance loss.

III. Main Results

In this section, we approach the problem of diffusing private data over a network. Section III-A derives the needed theoretical results and establishes that the accuracy of each estimate y_{ij} depends only on the distance d_{ij}. Moreover, algorithmic implementations of the composite mechanism Q should scale for vast social networks. Section III-B provides algorithmic implementations of the mechanism Q with complexity $O\left(\ln\left(\frac{\max_{i,j \in V} \epsilon(d_{ij})}{\min_{i,j \in V} \epsilon(d_{ij})}\right)\right)$.

A. Private Stochastic Process

For n-dimensional real-valued private data $u \in \mathbb{R}^n$, we derive a composite mechanism that generates the response y_{ij} that user j receives as an approximation of user's i private data u_i. This mechanism has the following two properties. First, the accuracy of the response y_{ij} depends solely on the distance d_{ij} between nodes i and j. Specifically, the expected squared error $\mathbb{E}\|y_{ij} - u_i\|_2^2$ does not depend on any other parameters of the network (e.g., size, topology) or the rest of the responses $\{y_{ik}\}_{k \in \mathcal{V} \setminus \{j\}}$. Second, any group of users $A \subseteq \mathcal{V}$ that decides to collaborate and share their responses $\{y_{ij}\}_{j \in A}$ cannot infer anything more about user's i private data. Algorithmic aspects of the composite mechanism are deferred until Section III-B.

Definition 3 introduces a continuous-domain stochastic process $\{V_\epsilon\}_{\epsilon > 0}$, which is used in Theorem 4 to define a composite privacy-preserving mechanism. Properties, sampling algorithms, and the derivation of it are deferred for later.

Definition 3: Define the stochastic process $\{V_\epsilon\}_{\epsilon > 0}$ with the following properties.

1) For $\epsilon > 0$, it is $d\mathbb{P}(V_\epsilon = v) \propto e^{-\epsilon \|v\|_2}$.
2) The process is Markov, that is, for any $0 < \epsilon_1 < \epsilon_2 < \epsilon_3$, it holds that $V_{\epsilon_1} \perp V_{\epsilon_3} | V_{\epsilon_2}$.

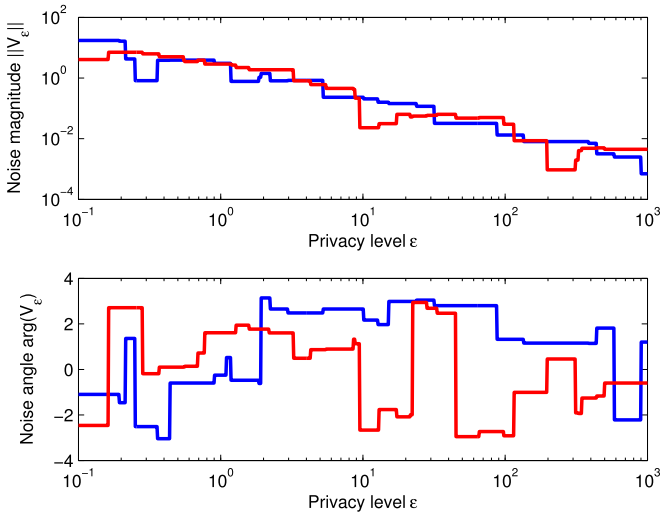

Fig. 2. Two samples of the 2-D process, which is the underlying object for diffusing private GPS coordinates over a network.

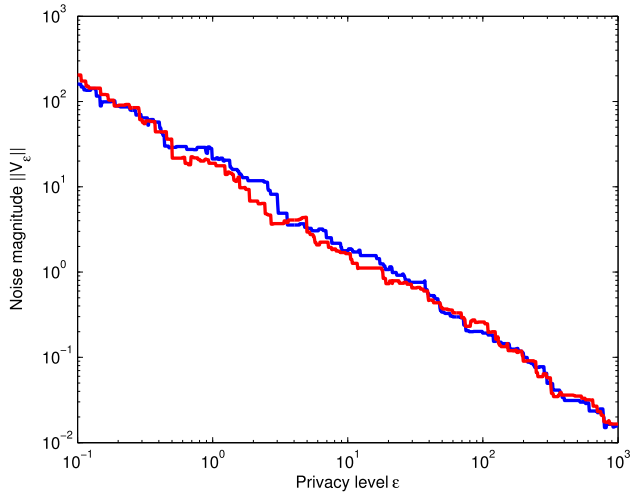

Fig. 3. ℓ_2-norm of two samples of the stochastic process $\{V_\epsilon\}_{\epsilon>0}$ in high dimensions ($n = 20$), which can be used to diffuse private signals over networks, such as power consumption in smart grids.

3) For any $0 < \epsilon_1 < \epsilon_2$, with $\tau = \frac{\epsilon_2}{\epsilon_1} - 1$, it is

$$d\mathbb{P}(V_{\epsilon_1} = v_1 | V_{\epsilon_2} = v_2) \propto \delta(v_1 - v_2) + \frac{(n+1)\epsilon_1^{1+\frac{n}{2}} \|v_1 - v_2\|_2^{1-\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}} K_{\frac{n}{2}-1}(\epsilon_1 \|v_1 - v_2\|_2)\tau + O(\tau^2)$$

where K is the modified Bessel function of the second kind.

Theorem 4: Let $d_{ij} \in \mathbb{R}_+$ denote the distance between users i and j, and $u_i \in \mathbb{R}$ be the private data of user i. Consider the mechanism Q that generates the responses:

$$y_{ij} = u_i + V^{(i)}_{\epsilon(d_{ij})}$$

where $\{V^{(i)}_\epsilon\}_{\epsilon>0}$ is a sample of a Markov stochastic process $\{V_\epsilon\}_{\epsilon>0}$. Then, mechanism Q provides a solution to Problem 2. In particular, it has the following properties.

1) The variance of response y_{ij} is $n\,(n+1)\,\epsilon(d_{ij})^{-2}$ and, thus, depends only on the distance between users i and j.
2) For any subset of users $A \subseteq \mathcal{V}$, the mechanism that releases the responses $\{y_{ij}\}_{j\in A}$ is $\left(\max_{j\in\mathcal{V}} \epsilon(d_{ij})\right)$-differential private.

The proof of Theorem 4 is presented in Appendix A. The main idea is introducing correlation between the responses $\{y_{ij}\}_{j\in\mathcal{V}}$. For $n = 1$, the stochastic process $\{V_\epsilon\}_\epsilon$ has closed-form expressions, whereas, for $n > 1$, closed-form expressions are derived only for the infinitesimal increments $V_{\epsilon+d\epsilon} - V_\epsilon$. Nonetheless, we derive handles that allow for exact (in the sense that we do not use approximations of the process) and efficient (in the algorithmic complexity sense) sampling of the process. Furthermore, our proof techniques are robust and can possibly be applied beyond the Laplace mechanism; for example, the K-norm mechanism [19] that appears in a different setting than the one considered here.

Fig. 2 pictures two samples of the stochastic process $\{V_\epsilon\}_{\epsilon>0}$, for $n = 2$, in polar coordinates and shows that the process is a jump process, that is, with high probability, the process is constant in small intervals. Fig. 3 pictures two samples of the process in high dimensions. The process is again lazy; yet, the jumps are more often.

A major consequence of Theorem 4 is that mechanism Q does not incentivize coalitions. Specifically, consider a group of curious users $A \subseteq \mathcal{V}$ who wish to estimate u_i more accurately and, thus, collaborate and share their knowledge $\{y_{ij}\}_{j\in A}$. In practice, such a group can be fake accounts of a single but distant (in the sense of d) user. Then, given this shared knowledge, the best estimator is

$$\hat{u}_i = y_{ij^*}\big|_{j^*\in\arg\min_{j\in A} d_{ij}}.$$

Therefore, user j^* is not benefited by such a coalition, and thus, she has no incentive to participate in the coalition and share her information y_{ij^*}. In fact, Theorem 4 solves Problem 2 in the best possible way: the existence of multiple users asking for the same private data under different privacy levels does not require additional privacy-preserving noise.

B. Algorithmic Implementation

Sampling from a continuous-domain stochastic process can often be performed only approximately. For example, consider the Brownian motion $\{B_t\}_{t\in[0,1]}$, which, for sampling purposes, requires storing an infimum of real values. Contrary to Brownian motion, the private process $\{V_\epsilon\}_{\epsilon>0}$ rarely changes value and is, thus, lazy. More formally, restricted to a sufficiently small interval $[\epsilon_1, \epsilon_2]$, the stochastic process $\{V_\epsilon\}_{\epsilon\in[\epsilon_1,\epsilon_2]}$ is constant with high probability. Furthermore, assuming the existence of an algorithm for computing the distance d_{ij}, the response y_{ij} can be generated during runtime. This property is crucial, since it circumvents the $O\left(N^2\right)$ memory requirements of a static implementation. Proposition 5 characterizes the distribution of the number of jumps in a bounded interval.

Algorithm 1: Sampling from the stochastic process V_ϵ over a bounded interval $\epsilon \in [\epsilon_1, \epsilon_2]$ can be performed both efficiently (with complexity $O\left(\ln\left(\frac{\epsilon_2}{\epsilon_1}\right)\right)$) and exactly (in the sense that we are not discretizing the interval or approximating the processes).

Require: Dimension n; Privacy levels ϵ_1 and ϵ_2, such that $\epsilon_2 > \epsilon_1 > 0$.
function SAMPLEPRIVATEPROCESSL2 $(n, \epsilon_1, \epsilon_2)$
  $k \leftarrow 1$
  $\epsilon^{(1)} \leftarrow \epsilon_2$
  $r \sim \text{Gamma}\left(n, \frac{1}{\epsilon_2}\right)$
  $v_1^{(1)}, \dots, v_n^{(1)} \overset{\text{i.i.d.}}{\sim} \text{Gaussian}(0,1)$
  $v^{(1)} \leftarrow \frac{r}{\|v^{(1)}\|_2} v^{(1)}$
  while $\epsilon(k) > \epsilon_1$ **do**
    $\delta\epsilon \sim \text{Exponential}(n+1)$
    $\epsilon^{(k+1)} \leftarrow e^{-\delta\epsilon}\, \epsilon^{(k)}$
    $r \sim \text{Bessel}\left(\frac{n}{2} - 1, \frac{1}{\epsilon^{(k+1)}}\right)$
    $\delta v_1, \dots, \delta v_n \overset{\text{i.i.d.}}{\sim} \text{Gaussian}(0,1)$
    $\delta v \leftarrow \frac{r}{\|\delta v\|_2} \delta v$
    $v^{(k+1)} \leftarrow v^{(k)} + \delta v$
    $k \leftarrow k + 1$
  end while
  Return $\{(\epsilon^{(i)}, v^{(i)}\}_{i=1}^k$
end function

Proposition 5: The number of jumps that the process $\{V_\epsilon\}_{\epsilon>0}$ performs in the interval $[\epsilon_1, \epsilon_2]$ is Poisson distributed with mean value $(n+1)\ln\left(\frac{\epsilon_2}{\epsilon_1}\right)$.

$$\mathbb{P}\left(k \text{ jumps in } [\epsilon_1, \epsilon_2]\right) = \frac{x^k}{k!} e^{-x}$$

where $x = (n+1)\ln\left(\frac{\epsilon_2}{\epsilon_1}\right)$.

Corollary 6: Process $\{V_\epsilon\}_{\epsilon>0}$ performs $\mathbb{E}[k] = (n+1)\ln 2$ jumps (in expectation, with variance $\text{Var}[k] = (n+1)\ln 2$) for every doubling of the privacy level, that is, in the interval $[\epsilon, 2\epsilon]$.

This laziness renders samples from the process highly compressible. Indeed, given the locations $\{\epsilon^{(i)}\}_{i=1}^k$ of the jumps and the values[2] $\{V_{\epsilon^{(i)}_-}\}_{i=1}^k$ near those points, a sample can be exactly reconstructed. The number k of jumps over a bounded interval $[\epsilon_1, \epsilon_2]$ is itself a random variable and captures the memory needs of our approach.

Furthermore, Proposition 5 suggests an efficient algorithm for directly sampling from the process $\{V_\epsilon\}_{\epsilon\in[\epsilon_1,\epsilon_2]}$, which we present in Algorithm 1. Algorithm 1 draws a sample $\{v_\epsilon\}_{\epsilon\in[\epsilon_1,\epsilon_2]}$ from the stochastic process V_ϵ over a bounded interval $\epsilon \in [\epsilon_1, \epsilon_2]$. This sample $\{v_\epsilon\}$ is the main object that performs diffusion of private data; whenever a user j requests user's i private data u_i residing d_{ij} away, the estimator $y_{ij} = u_i + v_{\epsilon(d_{ij})}$.

Algorithm 1 draws a sample of the stochastic process by sampling V_{ϵ_2}. Next, the algorithm proceeds toward smaller values of the privacy level ϵ by sampling the dormant time and the size of the jump. Specifically, the algorithm initializes a trace of the process by sampling from the Laplace mechanism. This is done in two steps: using the Gaussian distribution, the direction is drawn uniformly from the $n-1$-sphere, and then, the magnitude is drawn from the Gamma distribution. Then, the algorithm extends this trace backwards in ϵ by sampling for the location of the next jump. The logarithm of the positions where jumps occur define a Poisson process with rate $\lambda = n+1$, and thus, the length $\delta\epsilon = \ln\epsilon^{(i)} - \ln\epsilon^{(i+1)}$ of the interval until the next jump is exponentially distributed with density $\delta\epsilon \sim \lambda e^{-\lambda\,\delta\epsilon}$. Finally, conditioned on the event of a jump at $\epsilon^{(i)}$, the size $\delta v = V_{\epsilon^{(i)}_-} - V_{\epsilon^{(i)}_+}$ of the jump is "Bessel" distributed with parameter $\frac{1}{\epsilon^{(i)}}$. The algorithm recycles until the level ϵ_1 is reached. Additionally, responses y_{ij} are generated upon request, and thus, there is no excessive memory requirement $O(N^2)$ for storing all the responses $\{y_{ij}\}_{i,j\in\mathcal{V}}$. The number of iterations that Algorithm 1 performs is a random variable and is characterized by Proposition 5.

TABLE I
DISTRIBUTIONS THAT ARE USED BY ALGORITHM 1

Distribution	Param.	Supp.	Density
Laplace	$\beta > 0$	$x \in \mathbb{R}$	$\frac{1}{2\beta} e^{-\frac{\lvert x\rvert}{\beta}}$
Exponential	$\lambda > 0$	$x \in \mathbb{R}_+$	$\lambda e^{-\lambda x}$
Gamma	$n \in \mathbb{N}$, $\beta > 0$	$x \in \mathbb{R}_+$	$\frac{1}{\Gamma(n)\beta^n} x^{n-1} e^{-\frac{x}{\beta}}$
Bessel	$n \in \mathbb{N}$,	$x \in \mathbb{R}_+$	$\frac{4}{\Gamma\left(\frac{n}{2}\right)(2\beta)^{\frac{n}{2}+1}}$
		$\beta > 0$	$x^{\frac{n}{2}} K_{\frac{n}{2}-1}\left(\frac{x}{\beta}\right)$

Sampling from these distributions can be performed using a uniform random variable and the quantile function.

Typical single-dimensional ($n = 1$) private data are date of birth, salary, and health status. For $n = 2$, our results are applicable to geoindistinguishability [20] which is differential privacy for GPS locations and is experimentally illustrated in Section IV-A. Finally, the case $n \to \infty$ appeals to private signals that appear in filtering problems and smart grid applications.

For completeness, Table I presents the parameterization of the elementary distributions used by the proposed algorithms. We note that the Bessel distribution decays exponentially and has closed-form expressions for odd n. Nonetheless, it is a single-dimensional distribution, and thus, discretization and sampling through the inverse cumulative function are possible.

IV. ILLUSTRATIVE EXAMPLES

We present two applications that depict diffusion of private data over a network. These examples show that bits of private information can be spread over the whole network, which allows users to estimate global quantities, such as epidemic spreading, while providing strong privacy guarantees.

A. Synthetic Data

We consider the synthetic network in Fig. 5 with $N = |\mathcal{V}| = 150$ nodes and $|\mathcal{E}| = 1256$ edges, where edges are formed based

[2]We use the notation $V_{\epsilon_-} = \lim_{\tau\uparrow\epsilon} V_\tau$ and $V_{\epsilon_+} = \lim_{\tau\downarrow\epsilon} V_\tau$.

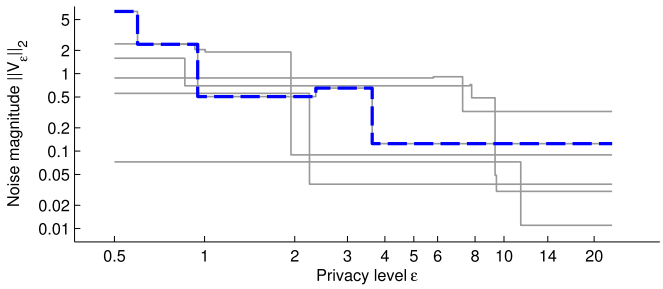

Fig. 4. Agent i uses Algorithm 1 with $n = 2$ and generates a single sample of the stochastic process. For small values of privacy level, high noise values are more likely, whereas, for loose privacy levels ($\epsilon \to \infty$), the noise values decrease in magnitude. Despite the continuity of the domain $\epsilon \in [0, \infty)$, the process performs only a few jumps.

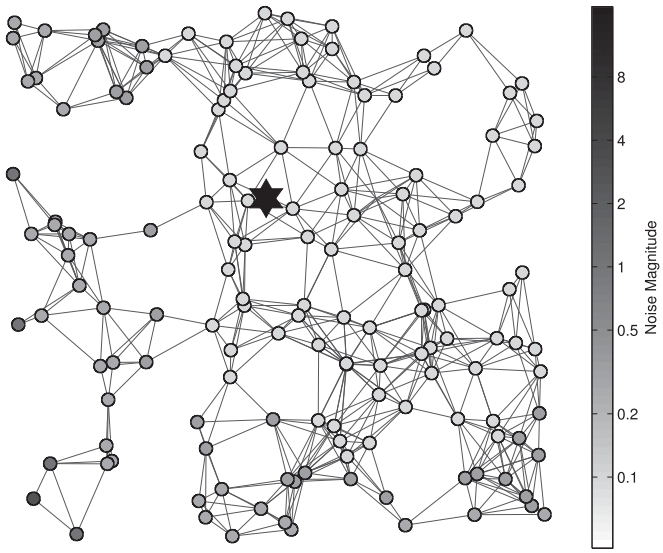

Fig. 5. Each individual j gets the value $u_i + V_{\epsilon(d_{ij})}$, where u_i is the true sensitive data, d_{ij} is the number of hops between users i and j, and V_ϵ is the result of Algorithm 1.

on proximity. Each user $i \in \{1, \ldots, N\}$ wishes to publish her vector-valued private data $u_i \in \mathbb{R}^2$, such as her GPS coordinates. For simplicity, we focus on a single user; our technique can be applied independently for each user. The distance d_{ij} between users i and j is captured by the shortest path length. We choose an exponential function $\epsilon(\cdot)$ that converts distances $d_{ij} \in \{1, \ldots, 9\}$ to privacy levels $\epsilon(d_{ij}) \in [.5, 15]$. The function $\epsilon(\cdot)$ that converts distances d_{ij} to privacy levels $\epsilon(d_{ij})$ can be different for each agent. In fact, user i may require any privacy level against user j. In practice, these privacy levels can be manually chosen by each user or automatically generated by the system operator based on the structure of the network and preferences of the nodes. In any case, all privacy levels $\epsilon_{ij} = \epsilon(d_{ij})$ are assumed public knowledge.

Algorithm 1 is executed by user i for $n = 2$ and the norms of several traces are shown in Fig. 4. For tight values of privacy level ($\epsilon \to 0$), large amounts of noises are added. In Fig. 5, nodes are colored based on the accuracy $\|y_{ij} - u_i\|_2$ of the response y_{ij} they receive.

Although we have assumed the existence of a secure communication channel between any two users of the network and the existence of a central authority which computes the distances d_{ij}, an implementation that relaxes these assumptions is possible. Specifically, assuming only local communications between neighboring users, an honest-but-curious model, and knowledge of the privacy levels—which can be performed also in a decentralized manner—a distributed approach is possible. In such an implementation, user i sends to all her neighbors the signal $\{u_i + V_\epsilon\}_{\epsilon \in (0, \epsilon(1))}$. Then, each user j receives the signal $\{u + V_\epsilon\}_{\epsilon \in [0, \epsilon(d_{ij}))}$, trims it to $\{u + V_\epsilon\}_{\epsilon \in [0, \epsilon(d_{ij}+1))}$, and broadcasts it to her friends. An application of this variation is left for future work.

B. Real Dataset: Facebook

In this section, we present an application of diffusing sensitive data on a real network. Specifically, an "ego network" [21] is a subgraph $G = (\mathcal{V} \cup \{\text{Alice}\}, \mathcal{E})$ of Facebook induced by a single user, Alice, and her friends $\mathcal{V}$. Fig. 7 plots such an ego network, where the bottom-left node is the user whose neighborhood is captured. The rest of the nodes represent Alice's friends, edges represent friendships between her friends, whereas the edges between Alice and her friends are omitted for clarity. We assume that Alice's infection status is captured by a single bit $u \in \{0, 1\}$. Then, Alice wishes to share this information with her friends in a privacy-preserving way.

For each friend $i \in \mathcal{V}$, the distance d_i is calculated by a central authority. Values $\{d_i\}_{i \in \mathcal{V}}$ are independent of the private data u and can be computed without any privacy requirements. The strength of the friendship between Alice and friend i is quantified by the value of d_i and can be computed using methods from the social network literature such as the `score` functions suggested in [17]. Here, distances d_i are evaluated according to

$$d_{ij} = \Gamma_{ii} + \Gamma_{jj} - 2\Gamma_{ij} \tag{2}$$

where $\Gamma \in \mathbb{R}^{n \times n}$ is the pseudoinverse of the Laplace matrix L of the network. Due to space limitations, we use the fact that our technique allows postprocessing of the responses y_{ij} and, thus, is applicable for private bits.

Initially, Alice executes Algorithm 1 in order to generate a single sample $\{w_\epsilon : \epsilon \in [\underline{\epsilon}, \bar{\epsilon}]\}$ of the stochastic process $\{V_\epsilon : \epsilon > 0\}$, where $\underline{\epsilon}$ (respectively, $\bar{\epsilon}$) is a lower (respectively, upper) bound of the quantity $\min_{i \in V} \epsilon(d_i)$ (respectively, $\max_{i \in V} \epsilon(d_i)$). Function $\epsilon(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is a decreasing function, which converts distances d_i to privacy levels $\epsilon_i = \epsilon(d_i)$. In this example, we chose $\epsilon(d) = \exp(-3.3d + 4)$, which leads to privacy levels within $[.5, 15]$. In practice, any decreasing function resulting in any range of privacy levels can be used. The exact expression is considered a designer's choice that balances the users' privacy needs and the accuracy of network-wide statistics. Next, individual responses are generated during runtime. Whenever user i requests access to the sensitive data u, the

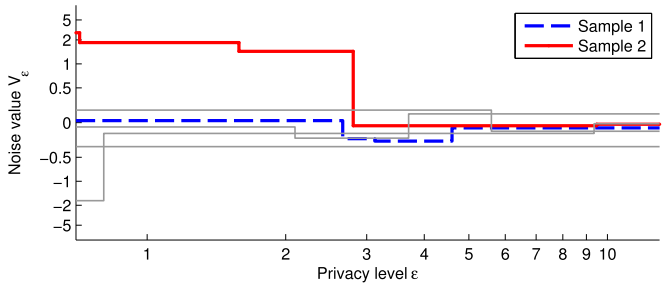

Fig. 6. Two samples of the stochastic process generated by Algorithm 1. The samples are private information; a malicious user i can subtract the noise $w_{\epsilon(d_i)}$ from the received response y_i and exactly infer the private data u.

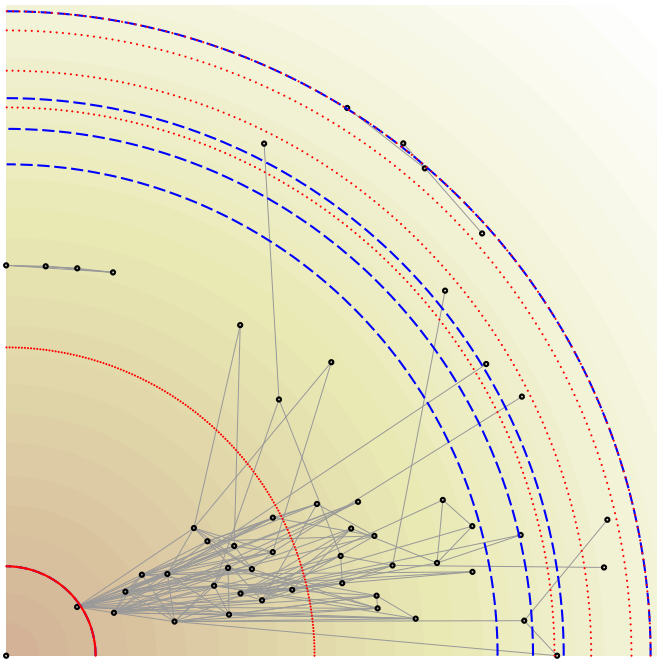

Fig. 7. Ego network is the part of the Facebook network that is visible from a fixed user A (ego), shown in the bottom-left corner of the plot. Each friend i is plotted at distance d_i. The locations of the jumps of the two samples shown in Fig. 6 are depicted by the blue and red circles. Although users residing within consecutive circles receive identical responses y_i, they are assigned different privacy levels $\epsilon(d_i)$ and, thus, have different confidence levels.

response y_i is securely communicated to user i:

$$y_i = \Pi_{\{0,1\}}(u + w_{\epsilon(d_i)})$$

where Π_S is the projection operator on the set S.

Fig. 6 depicts two executions of Algorithm 1 with $n = 1$, whereas Fig. 7 plots the ego network centered around Alice. In particular, Alice is shown on the bottom-left corner and each friend i is plotted at distance d_i from her. The blue and red circles mark the jumps of the stochastic process for the two samples w_ϵ^{blue} and w_ϵ^{red}. Counterintuitively, friends i lying within two consecutive blue circles receive exactly the same response y_i although they are assigned different privacy levels $\epsilon(d_i)$. The paradox is settled by noticing that the boundary circles are random variables themselves. Therefore, users receiving identical responses have different confidence levels.

V. Conclusion

In this work, we considered the case of a network where each user owns a private data $u \in \mathbb{R}^n$ such as her salary or her infection status and wishes to share approximations of this private data with the rest of the network under differential privacy guarantees. Specifically, we assumed that user i requires $\epsilon(d_{ij})$-differential privacy against user j, where $\epsilon(\cdot)$ is a decreasing function and d_{ij} is the distance induced by the underlying network between users i and j. In this context, we derived a composite mechanism that generates the response y_{ij} as user's j approximation of user's i private data. The accuracy of the response y_{ij} depends only on the allocated privacy level $\epsilon(d_{ij})$ and not on the size or other parameters of the network. An important property of our proposed mechanism is the resilience to coalitions, where we considered a group of users combining their received responses for more accurate approximations. Practically, this means that scenarios where an adversarial user creates multiple fake accounts cannot weaken the privacy guarantees. Algorithms for sampling from this composite mechanism were also provided. In particular, the complexity of these algorithms is independent of the size of the network, which renders them scalable, and is dictated only by the extreme privacy levels $\min_{i\in\mathcal{V}} \epsilon(d_{ij})$ and $\max_{i\in\mathcal{V}} \epsilon(d_{ij})$. Finally, we provided two illustrative examples: one on a synthetic network where users communicate their private GPS locations, and the other where a user shares her infection status with her Facebook ego network.

Concluding, reasons that users share their private data can be either for social interaction or for allowing crowdsourced statistics. In the first case, we view the proposed approach as a "soft-access scheme" to users' profiles, which substitutes the existing privacy settings in a social network. On the other hand, further work is required to address crowdsourced statistics. In that case, for accurate statistics, there is a tradeoff between the number of users residing far away and privacy level required against them.

Appendix

A. Proof of Theorem 4

Theorem 4 is established in multiple steps. First, we focus on the discrete-domain process $\{V_{\epsilon_i}\}_{i=1}^m$, where $\epsilon_1 \leq \cdots \leq \epsilon_m$ and, in particular, on the case of $m = 2$, with $\epsilon_1 \leq \epsilon_2 < \sqrt{2}\epsilon_1$, where the second inequality is due to technical reasons. Next, we prove the Markov property, which allows m discrete privacy levels. Finally, we pass to the limit and derive the continuous-domain process $\{V_\epsilon\}_{\epsilon>0}$, as stated in Theorem 4.

Proof for two privacy levels: We consider the stochastic process V_ϵ supported on two privacy levels $\{\epsilon_1, \epsilon_2\}$, where $\epsilon_1 \leq \epsilon_2 < \sqrt{2}\epsilon_1$. Allowing generalized functions, we assume that the joint distribution of V_{ϵ_1} and V_{ϵ_2} has density

$$\begin{aligned}\mathbb{P}(V_{\epsilon_1} = x, V_{\epsilon_2} = y) &= l_{\epsilon_1,\epsilon_2}(x, y) \\ &=: g(x, y), \quad x, y \in \mathbb{R}^n \end{aligned} \tag{3}$$

Density (3) should satisfy the following marginal distributions and privacy constraints:

$$\int_{\mathbb{R}^n} g(x,y)d^n y = \epsilon_1^n C_1 e^{-\epsilon_1 \|x\|_2}$$
$$\int_{\mathbb{R}^n} g(x,y)d^n x = \epsilon_2^n C_1 e^{-\epsilon_2 \|y\|_2}$$
$$\|\nabla_x g(x,y) + \nabla_y g(x,y)\|_2 \leq \epsilon_2\, g(x,y)$$

where $C_1 = \frac{\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}\Gamma(n+1)}$. The first two constraints express that V_{ϵ_1} and V_{ϵ_2} should be Laplace distributed with parameters $\frac{1}{\epsilon_1}$ and $\frac{1}{\epsilon_2}$, respectively. The last constraint enforces that the mechanism that releases $(u+V_{\epsilon_1}, u+V_{\epsilon_2})$ must be ϵ_2-private, and thus, the mechanism's log density needs to be ϵ_2-Lipschitz. We solve for densities g of the form

$$g(x,y) = \epsilon_2^n C_1 \phi(x-y) e^{-\epsilon_2 \|y\|_2}$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ is a (possibly generalized) function satisfying

$$\int_{\mathbb{R}^n} \phi(x-u)\epsilon_2^n e^{-\epsilon_2\|u\|_2} d^n u = \epsilon_1^n e^{-\epsilon_1\|x\|_2}$$
$$\int_{\mathbb{R}^n} \phi(u) d^n u = 1. \tag{4}$$

The first equation in (4) is an n-dimensional convolution with solution

$$\mathcal{F}\phi(s) = \frac{\mathcal{M}(s;\epsilon_1)}{\mathcal{M}(s;\epsilon_2)} \tag{5}$$

where $\mathcal{M}(s;\epsilon) = \mathcal{F}\left\{\epsilon^n e^{-\epsilon\|x\|_2}\right\}(s)$, and $s \in \mathbb{R}^n$ is the frequency. Solution (5) satisfies the second equation in (4) since

$$\int_{\mathbb{R}^n} \phi(u) d^n u = \mathcal{F}\phi(s)|_{s=0} = \frac{\mathcal{M}(0;\epsilon_1)}{\mathcal{M}(0;\epsilon_2)} = 1.$$

Finally, we need to prove that, for ϕ given in (5), density g is well defined, specifically

$$\phi(z) \geq 0 \quad \forall z \in \mathbb{R}^n.$$

This is proven under the assumption that $\epsilon_2 < \sqrt{2}\epsilon_1$; this assumption will eventually be removed. According to Lemma 7, we get

$$\mathcal{F}\phi(s) = \frac{\mathcal{M}(s;\epsilon_1)}{\mathcal{M}(s;\epsilon_2)} = \left(\frac{\epsilon_1}{\epsilon_2}\right)^{n+1} \left(1 + \frac{\epsilon_2^2 - \epsilon_1^2}{\epsilon_1^2 + \rho^2}\right)^{\frac{n+1}{2}}$$
$$= \left(\frac{\epsilon_1}{\epsilon_2}\right)^{n+1} \sum_{k=0}^{\infty} \binom{\frac{n+1}{2}}{k} \left(\frac{\frac{\epsilon_2^2}{\epsilon_1^2} - 1}{1 + \frac{\rho^2}{\epsilon_1^2}}\right)^k$$

where $\rho = \|s\|_2$. The sum on the right-hand side is an infinite series only when n is even, and, for $\epsilon_2 < \sqrt{2}\epsilon_1$, it converges uniformly in s to the left-hand side. Lemma 8 can be used to invert the series:

$$\phi(x) = \left(\frac{\epsilon_1}{\epsilon_2}\right)^{n+1} \sum_{k=0}^{\infty} \binom{\frac{n+1}{2}}{k} *^k \left\{ \left(\frac{\epsilon_2^2}{\epsilon_1^2} - 1\right) \epsilon_1^n (2\pi)^{-\frac{n}{2}} (\epsilon_1 r)^{1-\frac{n}{2}} K_{\frac{n}{2}-1}(\epsilon_1 r) \right\} \tag{6}$$

where $r = \|x\|_2$, $K_k(x)$ is the modified Bessel function of the second kind, and $*$ is the n-dimensional convolution. Since $\frac{\epsilon_2^2}{\epsilon_1^2} - 1 \geq 0$ and $K_{\frac{n}{2}-1}(r) \geq 0$, density g is well defined. ■

Next, we prove that the discrete-domain stochastic process $\{V_{\epsilon_i}\}_{i\in\{1,\ldots,m\}}$ is Markov.

Proof of the Markov property: Consider the discrete-domain process $\{V_{\epsilon_i}\}_{i\in\{1,\ldots,m\}}$ supported on m nondecreasing privacy levels $\{\epsilon_1,\ldots,\epsilon_m\}$, and the joint distribution that satisfies the Markov property:

$$d\mathbb{P}(V_{\epsilon_i} = v_i,\ \forall i) = l_{\epsilon_{1:m}}(v_1,\ldots,v_m)$$
$$= d\mathbb{P}(V_{\epsilon_1} = v_1) d\prod_{i=2}^{m} \mathbb{P}(V_{\epsilon_i} = v_i | V_{\epsilon_{i-1}} = v_{i-1})$$
$$= l_{\epsilon_1}(v_1) \prod_{i=2}^{m} \frac{l_{\epsilon_{i-1:i}}(v_{i-1}, v_i)}{l_{\epsilon_{i-1}}(v_{i-1})} \tag{7}$$

where $l_\epsilon(v) \propto e^{-\epsilon\|v\|_2}$ is the n-dimensional Laplace distribution with parameter ϵ^{-1} and $l_{\epsilon_1,\epsilon_2}(v_1,v_2)$ is the density g from the previous proof. Then, the joint distribution $l_{\epsilon_{1:m}}$ satisfies the following properties.

1) *Accuracy:* Each coordinate V_{ϵ_i} is optimally distributed, that is, Laplace distributed with parameter ϵ_i^{-1}:

$$\mathbb{P}(V_{\epsilon_i} = v_k) = \int_{\mathbb{R}^{n(m-1)}} l_{\epsilon_{1:m}}(v_1,\ldots,v_m) dv_{-i}$$
$$= l_{\epsilon_i}(v_i)$$

where $dv_{-i} = dv_1 \cdots dv_{i-1}\, dv_{i+1} \cdots dv_m$.

2) *Privacy:* The mechanism that releases $\{y_i\}_{i=1}^m$, where $y_i = u + V_{\epsilon_i}$, is ϵ_m-private. Indeed, the mechanism can be expressed as

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{m-1} \\ y_m \end{bmatrix} = \begin{bmatrix} u + V_{\epsilon_1} \\ \vdots \\ u + V_{\epsilon_{m-1}} \\ u + V_{\epsilon_m} \end{bmatrix}$$
$$= (u + V_{\epsilon_m}) + \begin{bmatrix} \sum_{i=2}^m V_{\epsilon_{i-1}} - V_{\epsilon_i} \\ \vdots \\ V_{\epsilon_{m-1}} - V_{\epsilon_m} \\ 0 \end{bmatrix}.$$

Density $l_{\epsilon_{i-1},\epsilon_i}$ defined in (6) can be rewritten in the form

$$l_{\epsilon_{1:m}}(v_1,\dots,v_m) = d\mathbb{P}(V_m = v_m)\prod_{i=1}^{m-1} d\mathbb{P}(V_i = v_i | V_{i+1} = v_{i+1})$$

where

$$d\mathbb{P}(V_i = v_i | V_{i+1} = v_{i+1}) = \frac{\ell_{\epsilon_{i:i+1}}(v_i, v_{i+1})}{\ell_{i+1}(v_{i+1})}$$

depends only on the quantity $v_{\epsilon_i} - v_{\epsilon_{i+1}}$. Therefore, V_m is independent of the differences $V_{\epsilon_i} - V_{\epsilon_{i+1}}$. Thus, the mechanism can be viewed as the composition of the ϵ_m-private mechanism that releases $u + V_{\epsilon_m}$ postprocessed by adding independent noise. Since differential privacy is resilient to postprocessing [1], the overall mechanism is ϵ_m-private. ■

Finally, we derive the continuous-domain process $\{V_\epsilon\}_{\epsilon>0}$ by passing to the limit as the $m \to \infty$, $\epsilon_1 = 0$, and $\epsilon \to \infty$. Specifically, we derive closed-form expressions that lead to efficient algorithms for sampling of the continuous-domain stochastic process.

Proof of the continuous-domain process: In density (6), let $\epsilon_1 = \epsilon$ and $\epsilon_2 = (1+\tau)\epsilon$, where $0 < \tau \ll 1$. Then, we prove that we can safely ignore high-order terms:

$$\begin{aligned}\phi_\epsilon(x) &\propto \delta(x) + \mathcal{F}^{-1}\left\{\frac{(n+1)\tau}{1+\frac{\rho^2}{\epsilon^2}}\right\} + O\left(\tau^2\right)\\ &= \delta(x) + \frac{\epsilon^n(n+1)}{(2\pi)^{\frac{n}{2}}}(\epsilon r)^{1-\frac{n}{2}}K_{\frac{n}{2}-1}(\epsilon r)\tau + O\left(\tau^2\right)\end{aligned} \tag{8}$$

where $r = \|x\|_2$. We discretize a bounded interval $[\underline{\epsilon}, \overline{\epsilon}]$ by considering $K+1$ points $\epsilon^{(i)} = q^i\underline{\epsilon}$, where $q = \left(\frac{\overline{\epsilon}}{\underline{\epsilon}}\right)^{K^{-1}}$, and define the random variable Z as follows:

$$Z := V_{\underline{\epsilon}} - V_{\overline{\epsilon}} = \sum_{i=1}^{K} V_{\epsilon^{(i-1)}} - V_{\epsilon^{(i)}}$$

where the random variables $\{V_{\epsilon^{(i)}}\}_{i=0}^K$ form a discrete-domain stochastic process introduced in (7). For large K, the step $\tau = q - 1$ becomes arbitrarily small, and thus, we use the first-order approximation in (8) for each telescoping term $(V_{\epsilon^{(i-1)}} - V_{\epsilon^{(i)}}) \sim \phi_{\epsilon^{(i)}}$. Finally, the random variable Z is distributed as

$$\begin{aligned}Z &\sim *_{i=1}^N \phi_{\epsilon^{(i)}}(Z)\\ &= *_{i=1}^N \left\{\delta(Z) + \frac{\left(\epsilon^{(i)}\right)^n(n+1)}{(2\pi)^{\frac{n}{2}}}(\epsilon^{(i)}\|Z\|_2)^{1-\frac{n}{2}}\right.\\ &\qquad \left. K_{\frac{n}{2}-1}(\epsilon^{(i)}\|Z\|_2)\tau\right\} + O(\tau)\end{aligned}$$

where we let $\tau \to 0$. This proves that we can approximate the continuous-domain stochastic process by a first-order approximation of the discrete-domain process. ■

Equation (7) characterizes the stochastic process $\{V_\epsilon\}_{\epsilon>0}$. The atom renders the stochastic process lazy; with high probability, the process is constant over sufficiently small intervals. The linear term governs the statistics of the jump.

B. Proof of Proposition 5

We now provide the proof of Proposition 5 that characterizes the jumps of the stochastic process $\{V_\epsilon\}_{\epsilon>0}$ and, thus, captures the complexity of Algorithm 1.

Proof: Consider the first-order approximation of the backwards conditional distribution ϕ_ϵ derived in (8), where $0 < \epsilon$ and $0 < \delta \ll 1$:

$$\begin{aligned}\mathbb{P}(V_\epsilon = x | V_{(1+\delta)\epsilon} = y) &\approx (1+(n+1)\tau)^{-1}\\ &\left(\delta(x) + \frac{\epsilon^n(n+1)}{(2\pi)^{\frac{n}{2}}}(\epsilon r)^{1-\frac{n}{2}}K_{\frac{n}{2}-1}(\epsilon r)\tau\right).\end{aligned} \tag{9}$$

Let $a_n(x)$ denote the probability that the process performs n jumps in the interval $[\epsilon, e^x\epsilon]$. Equation (9) shows that, for sufficiently small intervals $[\epsilon, (1+\tau)\epsilon]$, the process remains constant with probability $(1+(n+1)\tau)^{-1}$; therefore, $a_n(x)$ is invariant of ϵ. Under the discretization introduced earlier, where $\underline{\epsilon} \leftarrow \epsilon$ and $\overline{\epsilon} \leftarrow e^x\epsilon$:

$$a_0(x) = \mathbb{P}(0 \text{ jumps in } [\epsilon, e^x\epsilon]) = e^{-(n+1)x}.$$

A limiting argument is used to compute $a_1(x)$:

$$\begin{aligned}a_1(x) &= \lim_{K\to\infty}\sum_{k=1}^{K}\mathbb{P}\left(0 \text{ jumps in } [\epsilon, \epsilon^{(k-1)}]\right)\\ &\quad \mathbb{P}\left(1 \text{ jump in } [\epsilon^{(k-1)}, \epsilon^{(k)}]\right)\mathbb{P}\left(0 \text{ jumps in } [\epsilon^{(k)}, e^x\epsilon]\right)\\ &= (n+1)xe^{-(n+1)x}.\end{aligned}$$

A similar argument provides a recurrent equation and eventually

$$a_k(x) = \frac{((n+1)x)^k}{k!}e^{-(n+1)x}. \tag{10}$$

Therefore, for a bounded interval $[\underline{\epsilon}, \overline{\epsilon}]$, the number n of jumps is characterized by distribution (10), which is the Poisson distribution with mean value $(n+1)\ln\left(\frac{\overline{\epsilon}}{\underline{\epsilon}}\right)$. ■

C. Fourier Transform Pairs

In this section, we derive two Fourier pairs used in the proof of Theorem 4. By convention, the following definition of Fourier transform $f \overset{\mathcal{F}}{\leftrightarrow} F$ is used:

$$\mathcal{F}\{f\}(s) = \int_{\mathbb{R}^n} f(x)e^{-jx\cdot s}d^n x$$

where $f, F : \mathbb{R}^n \to \mathbb{R}$.

Lemma 7: The n-dimensional Fourier transform $\mathcal{F}$ of $f : \mathbb{R}^n \to \mathbb{R}$:

$$\begin{aligned}f(x) &= e^{-\|x\|_2}\\ \mathcal{F}\{f\}(s) &= \frac{\pi^{\frac{n}{2}}\Gamma(n+1)}{\Gamma\left(\frac{n}{2}+1\right)}\left(1+\|s\|_2^2\right)^{-\frac{n+1}{2}}\end{aligned}$$

where $s \in \mathbb{R}^n$.

Lemma 8: The n-dimensional Fourier transform $\mathcal{F}$ of $f: \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|^{1-\frac{n}{2}} K_{\frac{n}{2}-1}(\|x\|)$, is

$$\mathcal{F}\{f\}(s) = \frac{(2\pi)^{\frac{n}{2}}}{1+\rho^2}$$

where $x \in \mathbb{R}^n$, $\rho = \|s\|_2$, and $K_k(z)$ is the modified Bessel function of the second kind.

The integrals are formulated using spherical coordinates and, then, symbolically evaluated with Mathematica 10.0. For a nonautomated evaluation of the expressions, we refer the reader to MathWorld [22] and references therein, and integral lookup tables [23]. We remark that the Bessel function $K_k(z)$ diverges at $z = 0$; for $0 < z \ll 1$, it is $K_k(z) \approx \frac{\Gamma(k)}{2}\left(\frac{2}{x}\right)^k$. Therefore, its Fourier integral converges as the limit of the Laplace transform. This technicality is circumvented here by using lookup tables.

REFERENCES

[1] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3/4, pp. 211–407, Aug. 2014.

[2] L. Sankar, R. Rajagopalan, S. Mohajer, and V. Poor, "Smart meter privacy: A theoretical framework," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

[3] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Autom., Lang. Program.*, 2006, pp. 1–12.

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

[5] S. Han, U. Topcu, and G. Pappas, "Differentially private convex optimization with piecewise affine objectives," in *Proc. IEEE Conf. Dec. Control*, 2014, pp. 2160–2166.

[6] M. Hale and M. Egerstedty, "Differentially private cloud-based multiagent optimization with constraints," in *Proc. Amer. Control Conf.*, 2015, pp. 1235–1240.

[7] J. Le Ny and G. Pappas, "Differentially private filtering," *IEEE Trans. Autom. Control*, vol. 59, no. 2, pp. 341–354, Feb. 2014.

[8] V. Katewa, A. Chakrabortty, and V. Gupta, "Protecting privacy of topology in consensus networks," in *Proc. Amer. Control Conf.*, 2015, pp. 2476–2481.

[9] G. Acs and C. Castelluccia, "I have a dream!(differentially private smart metering)," in *Proc. 13th Int. Conf. Inf. Hiding*, 2011, pp. 118–132.

[10] F. Koufogiannis, S. Han, and G. Pappas, "Computation of privacy-preserving prices in smart grids," in *Proc. IEEE Conf. Dec. Control*, 2014, pp. 2142–2147.

[11] J. Le Ny, A. Touati, and G. Pappas, "Real-time privacy-preserving model-based estimation of traffic flows," in *Proc. ACM/IEEE 5th Int. Conf. Cyber-Phys. Syst.*, 2014, pp. 92–102.

[12] M. Alaggan, S. Gambs, and A. Kermarrec, "Heterogeneous differential privacy," arXiv preprint arXiv:1504.06998, 2015.

[13] H. Ebadi, D. Sands, and G. Schneider, "Differential privacy: Now it's getting personal," in *Proc. 42nd Annu. ACM SIGPLAN-SIGACT Symp. Principles Program. Lang.*, 2015, pp. 69–81.

[14] F. Koufogiannis, S. Han, and G. Pappas, "Gradual release of sensitive data under differential privacy," arXiv preprint arXiv:1504.00429, 2015.

[15] Wikipedia, Unix time—Wikipedia, the free encyclopedia, 2015. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Unix_time&oldid=662052467

[16] D. Babić, D. Klein, I. Lukovits, S. Nikolić, and N. Trinajstić, "Resistance-distance matrix: A computational algorithm and its application," *Int. J. Quantum Chem.*, vol. 90, no. 1, pp. 166–176, 2002.

[17] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.

[18] Y. Wang, Z. Huang, S. Mitra, and G. Dullerud, "Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems," in *Proc. IEEE Conf. Dec. Control*, 2014, pp. 2130–2135.

[19] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Proc. 42nd ACM Symp. Theory Comput.*, 2010, pp. 705–714.

[20] M. Andrés, N. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proc. ACM SIGSAC Conf. Comput. Commun. Securely*, 2013, pp. 901–914.

[21] J. Leskovec and J. Mcauley, "Learning to discover social circles in ego networks," in *Proc. Adv. Neural Inf. Process. Syst. Conf.*, 2012, pp. 548–556.

[22] E. W. Weisstein, Hypergeometric function. From mathworld—A wolfram web resource, 2015. [Online]. Available: http://mathworld.wolfram.com/HypergeometricFunction.html

[23] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. North Chelmsford, MA, USA: Courier Corp., 1964.

Fragkiskos Koufogiannis received the M.S.E. degree in electrical engineering and the A.M. degree in statistics from the University of Pennsylvania, Philadelphia, PA, USA, in 2016, and the Diploma in electrical and computer engineering from Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2011. He completed his diploma's thesis at École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. He is currently working toward the Ph.D. degree with the Department of Electrical and Systems Engineering, University of Pennsylvania.

His current work focuses on techniques for processing sensitive data.

George J. Pappas (S'90–M'91–SM'04–F'09) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 1998.

He is currently the Joseph Moore Professor of Electrical and Systems Engineering with the University of Pennsylvania, Philadelphia, PA, USA, where he is a member of the General Robotics, Automation, Sensing and Perception Laboratory and serves as the Chair of the Department of Electrical and Systems Engineering. His current research interests include hybrid and embedded systems, hierarchical control systems, distributed control systems, nonlinear control systems, with applications to robotics, unmanned aerial vehicles, biomolecular networks, and green buildings.

Dr. Pappas has received numerous awards, including the National Science Foundation (NSF) CAREER Award in 2002, the NSF Presidential Early Career Award for Scientists and Engineers in 2002, the 2009 George S. Axelby Outstanding Paper Award, and the 2010 Antonio Ruberti Outstanding Young Researcher Prize. He received the Eliahu Jury Award for Excellence in Systems Research at the University of California, Berkeley.