3L3-OS-06b-4

差分プライバシについての暗号理論的な考察

Some Observations on Differential Privacy from a Cryptographic Viewpoint

松田 隆宏
Takahiro Matsuda

産業技術総合研究所 セキュアシステム研究部門
Research Institute for Secure Systems (RISEC), National Institute of Advanced Industrial Science and Technology (AIST)

Differential privacy (DP), introduced by Dwork in 2006, is one of formal definitions of privacy for indivisuals in statistical databases, and has been intensively studied in a wide range of research areas, such as databases, data mining, machine learning, cryptography, and more generally theoretical computer science. In this paper, we briefly review the definition of DP and several recent results that are related to cryptography and/or have some cryptographic flavor, and make some observations on them.

1. はじめに

個人情報からなるデータベースに関する統計情報の公開では、何をもって個人のプライバシが守られるとするかの定義が非常に難しい。差分プライバシ (Differential Privacy, DP) [4] は、2006 年に Dwork が提唱した (クエリ応答型の統計データベースでの) プライバシの定義であり、近年、データベース、データマイニング、機械学習、暗号理論を始めとする理論計算機科学の様々な分野で急速に普及している*1。本稿では、差分プライバシの定義について簡単に振り返り、その後著者が暗号理論的に興味深いと考える点についてそれに関連する結果を紹介しつつ考察を行う。

なお、差分プライバシについての詳細は、五十嵐らによる解説 [17] や、2011 年春に MIT/ボストン大学で開かれた講義の講義録 [10] (英語) などを、差分プライバシについての初期の主要な結果については Dwork のサーベイ [5] などを参照されたい。また、データの中の秘密情報を守りつつ有用な情報を取り出し利活用する技術は一般にプライバシ保護データマイニングと呼ばれるが、それらについては佐久間らによる解説 [18, 19] などを参照されたい。

本稿の構成 §2. では様々な表記法の導入や基礎的な定義を行う。§3. では本稿で考える差分プライバシの設定及び定義、効用などについて説明する。§4. では、差分プライバシの定義について、いくつかの観点から考察を行う。§5. は本稿のまとめである。

2. 準備

“$\mathbb{N}$”、“$\mathbb{R}$”、及び “$\mathbb{R}_{\geq 0}$” はそれぞれ、自然数全体の集合、実数全体の集合、及び非負実数全体の集合を表す。“e” は自然対数の底を表す。S が有限集合の場合、“$x \leftarrow S$” で S から要素を一様ランダムに選び、x に代入する操作を表す。$\mathcal{M}$ が確率分布の場合、“$x \leftarrow \mathcal{M}$” で、分布 $\mathcal{M}$ に従って x を選ぶ操作を表す。

ベクトル $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ の L_p-ノルムを “$||x||_p$” で表す。すなわち、$||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$ である。

本稿では、“セキュリティパラメータ” を “k” で表す。関数 $f : \mathbb{N} \to [0, 1]$ が、全ての正の多項式 p と十分大きな全ての $k \in \mathbb{N}$ に対し $f(k) < 1/p(k)$ を満たす場合、f を “無視できる”(negligible) という。“negl” と書いて、非特定の無視できる関数を表すことにする。以下では頻繁に “多項式時間”、“指数的に大きい”、“無視できる” の様な言葉を使って議論を行うが、特に断りが無い場合は、セキュリティパラメータに対して多項式時間、指数的に大きい、無視できる、などの意味とする。暗号理論では慣例的に、要素技術 (あるいは、“正当なエンティティ”) のアルゴリズムを確率多項式時間チューリング機械で、攻撃者 (安全性を破ろうとしたり、分布を識別しようとしたりするエンティティ) を確率多項式時間チューリング機械や多項式サイズの回路族でモデル化することが多い*2。以下では表記の簡単化のため、要素技術のアルゴリズムが確率的多項式時間アルゴリズムであることも攻撃者が多項式サイズの回路族であることも、単に “PPTA” (Probabilistic Polynomial Time Algorithm の略) と書くことにする。

連絡先: t-matsuda@aist.go.jp

*1 [4] 以前に、Dwork ら [7] が差分プライバシと本質的に同等の定義を “(ϵ-)Indistinguishability” という名で提案しており、[7] が差分プライバシの提案論文だとされることも多い。

3. 差分プライバシ

本節では、差分プライバシ、メカニズム、効用などの定義を振り返る。

3.1 モデル

本稿では、“データベース” を、データ空間 $\mathcal{D}$ の要素を d 個持つベクトルとして表すことにし、データベースの要素数 d は既知の値とする。従って、データベース全体の集合は $\mathcal{D}^d$ となる。以下では d がセキュリティパラメータの関数となっている場合も考える。データベース $D \in \mathcal{D}^d$ と $D' \in \mathcal{D}^d$ の要素が一つだけ異なる場合、“D と D' は隣接している” と言う。

本稿では簡単のため、差分プライバシが提唱された際 [7, 4] に考えられた最も基礎的な設定である対話型データベースの設定のみを考えることにする。この設定では、データベース D は信頼できるサーバに保管されており、データベースに関する統計情報を利用したいユーザ (クライアント) は、サーバへ統計情報を計算する関数を記載した “クエリ”$q : \mathcal{D}^d \to \mathcal{R}$ を発行する。このクエリ q は、予め決められたクエリ集合 $\mathcal{Q}$ から選ばれることにする。(例えば、(人名, 身長) からなるデータベース D で、平均身長を問い合わせできる対話型データベースであれば、$\mathcal{Q}$ は “部分集合 $S \subseteq \mathcal{D}^d$ を選び、S 内の人の平均身長を返す” 様なクエリ全体と考えればよい。$\mathcal{Q}$ が大きければ大きいほど、多くの種類の統計情報が計算できることになる。) サーバは、統計値 $q(D)$ を、ノイズを乗せるなどして加工し、ユーザに返答 a を返す。この状況で、a から D に含まれる (または含まれない) 個人の情報が必要以上に漏れてしまうことを防ぎたい。差分プライバシの研究では、この様な動作を行う

*2 暗号理論や計算の複雑さの理論では、単に “多項式時間アルゴリズム” と言うときは、入力のサイズに対し多項式時間であることを指すが、特に暗号理論では、セキュリティパラメータ k の多項式時間で動作することを明示するために、1 が k 個並んだ文字列 “1^k” を各アルゴリズムの入力として明示的に書くことも多い。本稿では表記の簡単化のために 1^k などは省略する。

(確率的) アルゴリズムは慣例的にメカニズム (Mechanism) と呼ばれるため、本稿でもその呼び名を用いる[*3]。

クエリ集合 $\mathcal{Q}$ についてのメカニズム $\mathsf{M} : \mathcal{D}^d \to \mathcal{R}$ の動作だけに限ってその入出力を確認すると、M は、データベース $D \in \mathcal{D}^d$ とクエリ $q \in \mathcal{Q}$ を受け取り[*4]、内部で計算を行い、返答 $a \in \mathcal{R}$ を返す。

3.2 差分プライバシの定義

差分プライバシ [4] は、非常に大雑把に言えば、データベース中のどの個人のデータも、その個人のデータがあろうと無かろうと、メカニズムの出力 (加工された統計情報) の分布に大きな変化が無い、ということを保証する。どの個人のデータも、出力の分布に大きな変化を与えないならば、データベース中の個人のプライバシが守られている、という考えに基づいている。

ここでは形式的な定義を振り返る。差分プライバシは、その "強度" を表す実数のパラメータ ϵ を持つ。本稿では、Dwork ら [6] の定義を元に、加算的な誤差 δ を考慮した定義を用いる。次節の議論で、ϵ や δ(及びその他のパラメータ) がセキュリティパラメータ k のどの様な関数となっているかを取り扱うため、ここでも ϵ、δ は k の関数として考える。

定義 1. $\epsilon, \delta : \mathbb{N} \to \mathbb{R}_{\geq 0}$、$d : \mathbb{N} \to \mathbb{N}$ とする。メカニズム $\mathsf{M} : \mathcal{Q} \times \mathcal{D}^d \to \mathcal{R}$ が以下を満たす場合、M はクエリ集合 $\mathcal{Q}$ についての (ϵ, δ)-DP を満たすと言う*:* 全ての隣接するデータベースの組 $D, D' \in \mathcal{D}^d$、全てのクエリ $q \in \mathcal{Q}$、及び全ての出力空間の部分集合 $S \subseteq \mathcal{R}$ について以下が成り立つ。

$$\Pr[\mathsf{M}(q, D) \in S] \leq e^{\epsilon(k)} \cdot \Pr[\mathsf{M}(q, D') \in S] + \delta(k) \qquad (1)$$

(ただし確率は M *の使用する乱数を一様ランダムに選ぶ場合を考える。)*

不等式 *(1)* において δ が無視できる関数の場合、M は $(\epsilon, \mathsf{negl})$-DP を満たすと言う。さらに、不等式 *(1)* において常に $\delta(k) = 0$ の場合、M は ϵ-DP を満たすと言う。 ♢

出力が実数 (のベクトル) である様なクエリの場合、そのクエリの L_1-敏感度 (L_1-Sensitivity) に従って、適切に調整されたパラメータのラプラスノイズを加えることで、上記の定義を満たすことができることが知られている [7, 4]。

3.3 メカニズムの効用

差分プライバシの定義は、それ単独ではあまり意味をなさない。例えば、入力に限らず常にゼロ (あるいは任意の固定値や単なる乱数) を返すメカニズムは 0-DP を満たす。しかし、この様なメカニズムの出力に全く統計情報としての価値 (あるいは "効用"、Utility) が無いことは明らかである。従って差分プライバシを考える際は、メカニズムの出力の効用を同時に考えることが必要となる。

効用は、データベースや計算される統計情報の種類によって様々な設定の仕方が有り得るが、議論の対象が統計情報の場合は、メカニズムによって加工された値が真の統計値とどれほど近いか (遠いか) を表す "誤差" が設定されることが多い。(誤差が小さければ、効用が高いと考える。)

ここでは、実数のベクトルを出力するメカニズムについての効用の一つの例として、§4.2 節で紹介する Groce ら [11] の研究で使用された、二つのパラメータ p と v を持つ誤差の評価関数である "平均 (p, v)-誤差" (Average (p, v)-Error) の定義を振り返る。

定義 2. $p > 0$、$v \geq 1$、及びデータベース $D \in \mathcal{D}^d$ に対し、メカニズム M のクエリ $q : \mathcal{D}^d \to \mathbb{R}^n$ に対する "平均 (p, v)-誤差" を、$(||\mathsf{M}(D) - q(D)||_p)^v$ の *(*M の乱数を一様ランダムに選ぶ場合の*)* 期待値と定義する。 ♢

p と v の値によって、様々な "誤差" の定義を捉えることができる。例えば、平均二乗誤差など、統計でよく用いられる誤差の評価関数も平均 (p, v)-誤差で記述できる。

4. 考察

本節では、差分プライバシの定義について、いくつか暗号理論的な視点から考察を行う。

暗号理論では多くの場合、要素技術によって細かな差異はあるが、要素技術 (例えば公開鍵暗号) を構成するアルゴリズムは全て PPTA である。また、暗号理論では (慣例的に)、"現実的に実行可能かどうか" を、アルゴリズムが多項式時間で動作するかどうかで切り分けて考える。従って、ほぼ全ての暗号技術の安全性の定義では、攻撃者として多項式アルゴリズム (あるいは多項式サイズの回路) を考え、攻撃者が方式の安全性を破る "自明な方法に対する優位性" が無視できることを要求する[*5]。例えば、最も基礎的な暗号要素技術である "一方向関数" の定義は以下の様になる。

定義 3. 以下の二つの性質を満たす場合、*(*確定的な*)* 関数 $f : \{0,1\}^* \to \{0,1\}^*$ を一方向関数であると言う*: (1)* f は多項式時間で計算可能であり、*(2)* 全ての *PPTA* 攻撃者 $\mathcal{A}$ に対し、$\Pr_{x \leftarrow \{0,1\}^k}[\mathcal{A}(f(x)) \in f^{-1}(f(x))]$ が無視できる。 ♢

上記の様な暗号理論での典型的な暗号要素技術の安全性の定義と対比して、本稿では、差分プライバシの定義 (定義 1) の、以下の点に着目したい。

(1) 差分プライバシの定義の中には、"攻撃者" が明示的に出てこない。従って、例えば想定される攻撃者の能力が (少なくともそのままでは) 直感的にはそれほど明らかではない。

(2) クエリ q が多項式時間で計算可能かどうかや、メカニズムが多項式時間アルゴリズムかどうか、などが規定されていない。

(3) データベースのサイズ d が多項式かどうかが規定されていない。

(1) について、差分プライバシでは、無制限の計算能力を持つ攻撃者を考えることで、より (暗号理論での) 安全性定義らしい (と思われる) 定義と等価であることを容易に示すことができる。すなわち、差分プライバシは、"1 回のクエリを行う計算能力無制限の攻撃者に対するメカニズムの出力の識別不能性" を保証している。この点について、§4.1 においてもう少し詳細に議論する。

上記に関連して興味深いと思われる問題は、現在の差分プライバシの定義を、計算量的な (つまり PPTA の) 攻撃者のみを考えることで緩和できるのか、ということである。すなわち、計算能力無制限のアルゴリズムではなく、PPTA の攻撃者のみを考え、それに対してのみ識別不可能性を満たすことを要求する様なプライバシの定義を考えた場合に、"プライバシ" の定義を達成しやすくなるのか、ということである。(例えば、同じ計算時間のメカニズムでも、よりよい効用を持つ、同じパラメータ (ϵ, δ) を達成するために、計算量的な差分プライバシの方が効率よく達成できるのか、など。) これについては、§4.2 において、Mironov ら [14] による計算量的な差分プライバシの定義、及び Groce ら [11] による上記の問いに対する否定的な結果を紹介する。

---

*3 差分プライバシの文脈では、データにノイズを乗せるなどして元の統計情報を "隠す" ため、サニタイザ (Sanitizer) と呼ばれることもある。

*4 統計値 $q(D)$ は D と q から計算できるため、ここではメカニズムの入力として考えていない。

*5 情報理論的な安全性を持つ暗号技術や、"多項式時間" や "無視できる" などの漸近的な表現を用いない "Exact Security" と呼ばれる型の安全性定義も存在するが、ここでは考えていない。

最後に、上記の (2) と (3) は、暗号理論における “実行可能かどうか” に関連する、重要な点である。例えばあるメカニズムが、「差分プライバシを満たすが、その動作時間がデータベースの要素数 d に対して指数的になる」場合、実運用では比較的小さなサイズのデータベースの場合にしか使用できず、柔軟性や拡張性に乏しい。また、多くの場合、現実的に興味があるのは、クエリやメカニズムが多項式時間で計算できる場合である。実際、データベース $D \in \mathcal{D}^d$ から、合成データ (Synthetic Data) ([1] など) と呼ばれるデータを出力する様なメカニズムは、既存の最も効率が良い方式でも要素数 d に対し指数時間かかる方式しか知られていない。

これらの点に関連する興味深い結果として、§4.3 において、Ullman [16] による差分プライバシ、計算時間、効用とデータベースへのクエリ回数に関するある種の限界 (トレードオフ) を示した結果を紹介する。

4.1 攻撃者を用いた差分プライバシの定義

以下の様なプライバシの定義 (便宜上 DP′ と呼ぶ) を考える。

定義 4. $\epsilon, \delta : \mathbb{N} \to \mathbb{R}_{\geq 0}$、$d : \mathbb{N} \to \mathbb{N}$ とする。メカニズム $\mathsf{M} : \mathcal{Q} \times \mathcal{D}^d \to \mathcal{R}$ が以下を満たす場合、M はクエリ集合 $\mathcal{Q}$ についての (ϵ, δ)-DP′ を満たすと言う*:* 全ての隣接するデータベースの組 $D, D' \in \mathcal{D}^d$、全てのクエリ $q \in \mathcal{Q}$、及び全ての計算能力無制限の攻撃者 $\mathcal{A}$ について以下が成り立つ。

$$\Pr[\mathcal{A}(\mathsf{M}(q, D)) = 1] \leq e^{\epsilon(k)} \cdot \Pr[\mathcal{A}(\mathsf{M}(q, D')) = 1] + \delta(k) \quad (2)$$

*(*ただし確率は M 及び $\mathcal{A}$ の使用する乱数を一様ランダムに選ぶ場合を考える。*)* ♢

以下に、定義 1 と定義 4 が等価であることを示す。このことから、DP は、計算能力無制限の攻撃者に対するある種の識別不能性を保証しているということが分かる。

定理 1. $\epsilon, \delta : \mathbb{N} \to \mathbb{R}_{\geq 0}$、$d : \mathbb{N} \to \mathbb{N}$、そして $\mathsf{M} : \mathcal{D}^d \to \mathcal{R}$ をクエリ集合 $\mathcal{Q}$ についてのメカニズムとする。M が $\mathcal{Q}$ についての (ϵ, δ)-DP を満たすならば、M は $\mathcal{Q}$ についての (ϵ, δ)-DP′ も満たす。また、その逆も成り立つ。

証明のスケッチ (DP ⇒ DP′ の方向) 隣接するデータベースの組 $D, D' \in \mathcal{D}^d$、クエリ $q \in \mathcal{Q}$、及び M の (ϵ, δ)-DP′ に対する任意の無制限の計算能力を持つ攻撃者 $\mathcal{A}$ を任意に固定する。一般性を失わず、$\mathcal{A}$ は決定的なアルゴリズムとしてよい*6。M の出力空間 $\mathcal{R}$ の部分集合 $S_{\mathcal{A}}$ を、$S_{\mathcal{A}} = \{x \in \mathcal{R} | \mathcal{A}(x) = 1\}$ と定義する。M は (ϵ, δ)-DP を満たす $\mathcal{Q}$ についてのメカニズムなので、当然特に最初に固定したデータベース組 D, D' とクエリ q、及びこの部分集合 $S_{\mathcal{A}}$ について不等式 (1) が成り立つ。また、$S_{\mathcal{A}}$ の定義より、全ての $x \in \mathcal{R}$ について $x \in S_{\mathcal{A}} \Leftrightarrow \mathcal{A}(x) = 1$ となることを用いて、$\Pr[\mathsf{M}(q, D) \in S_{\mathcal{A}}] = \Pr[\mathcal{A}(\mathsf{M}(q, D)) = 1]$ 及び $\Pr[\mathsf{M}(q, D') \in S_{\mathcal{A}}] = \Pr[\mathcal{A}(\mathsf{M}(q, D')) = 1]$ を示せる。これらを不等式 (1) の両辺に代入すれば、不等式 (2) が得られる。データベースの組 $D, D' \in \mathcal{D}^d$ とクエリ $q \in \mathcal{Q}$ の選び方、そして攻撃者 $\mathcal{A}$ の選び方は任意だったので、あらゆるデータベースの組、クエリ、計算能力無制限の攻撃者について不等式 (2) が成り立つ。つまり、M は (ϵ, δ)-DP′ を満たす。

(DP′ ⇒ DP の方向) 上記と逆のことをすればよい。つまり、出力空間 $\mathcal{R}$ の任意の部分集合 $S \subseteq \mathcal{R}$ を固定し、以下の様な DP′ に対する攻撃者 $\mathcal{A}_S$ を考える: $\mathcal{A}_S$ は値 $x \in \mathcal{R}$ を入力として受け取り、(必要ならば自身の持つ無制限の計算能力を使って)$x \in S$ かどうかを調べ、そうならば 1 を、$x \notin S$ ならば 0 を出力する。後は同様である。 □

*6 $\mathcal{A}$ が仮に確率的アルゴリズムであっても自身の無制限の計算能力によって、“最高の乱数” 自分自身で調べることができる。

4.2 計算量的な差分プライバシとその限界

Mironov ら [14] は、攻撃者として PPTA のみ想定する場合の差分プライバシを 2 種類導入した。[14] ではシミュレータを使った定義と識別不能性に基づく定義が導入されたが、ここでは “識別不能性” に基づく計算量的な差分プライバシ (IND-CDP) のみを振り返る。

定義 5. $\epsilon : \mathbb{N} \to \mathbb{R}_{\geq 0}$ とする。メカニズム $\mathsf{M} : \mathcal{Q} \times \mathcal{D}^* \to \mathcal{R}$ が以下を満たす時、M はクエリ集合 $\mathcal{Q}$ についての ϵ-IND-CDP を満たすと言う*:* 全ての *PPTA* 攻撃者 $\mathcal{A}$、全てのクエリ $q \in \mathcal{Q}$、全ての多項式 d、全ての十分大きな $k \in \mathbb{N}$、及び全ての隣接するデータベースの組 $D, D' \in \mathcal{D}^d$ について以下が成り立つ。

$$\Pr[\mathcal{A}(\mathsf{M}(q, D)) = 1] \leq e^{\epsilon(k)} \cdot \Pr[\mathcal{A}(\mathsf{M}(q, D')) = 1] + \mathsf{negl}(k)$$

*(*ただし、確率は M と $\mathcal{A}$ の使用する乱数を一様ランダムに選ぶ場合を考える。*)* ♢

上記の定義において、“PPTA $\mathcal{A}$ を “計算能力無制限の攻撃者 $\mathcal{A}$” と置き換えると、データベースの要素数 d が任意の多項式の場合の $(\epsilon, \mathsf{negl})$-DP′ の定義と一致する。従って、定理 1 から、少なくとも $(\epsilon, \mathsf{negl})$-DP よりも弱いプライバシの定義となっていることが分かる。

では、この “弱められた”(様に見える) 安全性は、元来の DP よりも達成し易いのだろうか? Groce ら [11] は、その問いに対し 2 つの否定的な結果を与えた。

- $\mathcal{Q}$ を効率的に計算可能な (多項式サイズ回路で記述可能な) クエリの集合とし、ϵ-IND-CDP を満たす効率的なメカニズムの、一方向関数*7を部品としたブラックボックス構成 M が存在すると仮定する。(大雑把に言えば、M は関数 (サブルーチン)F を部品として持ち、F をブラックボックスとして取り扱い、「F が一方向関数ならば、M は ϵ-IND-CDP である」という命題を証明できる、ということである。) このとき、全ての効用関数、全ての多項式 d、及び全てのデータベース $D \in \mathcal{D}^d$ に対し M と無視できる差を除いて同じ効用を持ち、M とほぼ同じ実行時間で動作する $(\epsilon, \mathsf{negl})$-DP を満たすメカニズム $\widehat{\mathsf{M}}$ が存在する。(この結果は、M の出力空間 $\mathcal{R}$ に依らずに成り立つ。)
- $p > 0$、$v \geq 1$、そして $\mathsf{M} : \mathcal{Q} \times \mathcal{D}^n \to \mathbb{R}^n$ (n は定数) を ϵ-IND-CDP を満たし、しかもその平均 (p, v)-誤差が多項式 ℓ 以下である様な効率的なメカニズムであるとする。このとき、M とほぼ同一の実行時間を持ち、しかもその平均 (p, v)-誤差が $\ell + \mathsf{negl}$ 以下である様な ϵ-DP を満たす別の効率的なメカニズム $\widehat{\mathsf{M}} : \mathcal{Q} \times \mathcal{D}^n \to \mathbb{R}^c$ が存在する。

暗号理論におけるブラックボックス構成とは、暗号要素技術 (疑似乱数生成器など) を別の暗号要素技術 (一方向関数など) を部品として構成する場合の構成の種類のことであり、ある暗号技術を部品として別の暗号技術が構成できる場合、自然な構成方法は全てブラックボックス構成に当てはまる。(より詳細なブラックボックス構成の定義などは、[15] などを参照されたい。) 従って、上記の両結果はいずれも、計算量的な差分プライバシを達成するメカニズムが構成できる場合、それとほぼ同等の実行時間、効用を持つ通常の差分プライバシを満たすメカニズムが構成できる、ということを示している。これはすなわち、計算量的な差分プライバシを達成するのは、通常の差分プライバシを達成するのと比べて容易になっていない、ということを意味している。

なお、計算量的差分プライバシと通常の差分プライバシには (達成し易さという意味で) ほとんど差がない、という [11]

*7 [11] によれば、一方向関数ではなく、公開鍵暗号や落とし戸付き関数など、暗号要素技術として定式化できる大部分の要素技術からのブラックボックス構成を考えても、本結果と同様の結果を示せる。

の結果は、本稿で考えているクライアント/サーバ型の設定でサーバ側の出力するデータに対し差分プライバシを考える場合の話である。McGregor ら [13] は、二者以上が分割してデータベースを保持しており、二者間秘匿計算プロトコルで両者のデータベースの組み合わせから (各エンティティが相手に対して差分プライバシを満たすように) 統計情報を計算する、という設定では、計算量的な差分プライバシと通常の差分プライバシの間には、達成に必要なノイズの大きさなどに本質的な差があることを示した。

4.3 効率的なメカニズムの複数回クエリへの応答の限界

統計情報において “ある性質 (述語) を満たすレコードは、データベース中に何個あるか (あるいは、全体に対する割合はどのくらいか)?” という “数え上げ型” のクエリ (Counting Query) は最も基礎的なものの一つである。ここでは、効率的に計算可能な (多項式サイズの回路で述語を計算するクエリを記述可能な) 数え上げクエリ全体からなる集合を $\mathcal{Q}_{\mathtt{eff}}$ と書く。

本節では、§3.1 で説明したクライアント/サーバ型の設定において、クライアントが複数個のクエリを同時に行う状況を考える。この様な状況において、ϵ-DP を満たすメカニズムに n 個のクエリを行った場合、$n\epsilon$-DP を満たすことが知られている [7, 4]。しかし、n があまり大きくなりすぎると、$n\epsilon$ が大きくなりすぎてしまい、実質的に何も守られないことになる。

Ullman [16] は、数え上げクエリについてのメカニズムについてのクエリ回数、データベースの大きさ、差分プライバシ、効用についてのトレードオフを表す以下の様な成果を示した:

- d を k の多項式とする。一方向関数が存在する場合、各データの空間 $\mathcal{D}$ が $\{0,1\}^k$ である様なデータベース (つまり、データベース全体の空間は $(\{0,1\}^k)^d$)、出力空間が実数 $\mathbb{R}$ である様な $\mathcal{Q}_{\mathtt{eff}}$ についての効率的なメカニズム M は、$\widetilde{\Theta}(d^2)$ 回のクエリ $q_1, q_2, \cdots \in \mathcal{Q}_{\mathtt{eff}}$ に応答する場合、$(O(1), o(1/d))$-DP を達成しながら同時に全てのクエリに対する誤差を (1/2 以下の) 定数とはできない。

$\epsilon = O(1)$、$\delta = o(1/d)$ は差分プライバシとして非常に弱いパラメータであるため、この結果は、差分プライバシを保ちつつ、$\mathcal{Q}_{\mathtt{eff}}$ ついての効率的なメカニズムが本質的に持つ、複数回のクエリへ応答する差異の限界を示している。

なお、上記と対比的な結果として、Blum ら [2] は、d^2 回を超える数え上げクエリに対して差分プライバシを (意味のあるパラメータに対して) 達成するメカニズムを示したが、このメカニズムは d について指数時間で動作する。また、上記の Ullman [16] の結果は、全ての効率的な数え上げクエリ $\mathcal{Q}_{\mathtt{eff}}$ を受け付けることができるメカニズムについての限界を示しているだけであり、例えば $\mathcal{Q}_{\mathtt{eff}}$ の部分集合 (でかつ統計的に興味深いもの) のクラスに対して $O(d^2)$ 回以上答えることができるメカニズムも知られている ([12] など)。(これらの関連研究については、[16] を参照されたい。)

また、[16] の結果は、不正者を追跡可能な放送型暗号 (Traitor Tracing)[3] を利用して、効率的な数え上げ型クエリの部分集合を構成することで行う。このクエリに DP を満たしつつかつ誤差を定数以下とできてしまうと、不正者追跡型の放送型暗号の安全性を破れてしまうことを示す。暗号要素技術が差分プライバシにおいて否定的な結果を導出するために使用されるというその手法自体も、暗号理論的に非常に興味深い。

5. おわりに

本稿では、近年盛んに研究されている差分プライバシの定義について、暗号理論の視点から、最近の興味深い結果を紹介するとともに、いくつかの簡単な考察を行った。

Dwork [5] や五十嵐ら [17] も指摘しているように、差分プライバシでは、概念の提案からの歴史がまだ浅く、“パラメータの ϵ や δ がどの様な値であれば、実際にどの程度個人のプライバシが現実問題として守られるのか” や、“そもそもその定義自体がプライバシの定義として妥当なのか”、などの非常に基礎的な問いに対して、研究コミュニティの中でさえ意見の一致が得られていない。(例えば最近でも、ゼロ知識プライバシ [9] や Crowd-Blending プライバシ [8] など、未だに新しいプライバシの定義が導入されている。) これは、暗号方式や電子署名などの、25 年以上の歴史を持つデファクト標準の安全性定義があり、既に要素技術が多く実運用されている基礎的な暗号技術とは大きく状況が異なる。

差分プライバシを満たすメカニズムが実運用されるまでには解決されるべき数多くの興味深い研究課題が残されており、魅力的な研究分野であると言えるだろう。

謝辞 本稿の執筆にあたり、有益なコメントをいただいた産業技術総合研究所の花岡悟一郎氏、縫田光司氏、照屋唯紀氏、及び東京大学の大畑幸矢氏、そして多大なるご協力をいただいた筑波大学の矢内直人氏に深く感謝いたします。

参考文献

[1] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *PODS 2007*, pp.273-282, ACM, 2007.

[2] A. Blum, K. Ligett, and A.Roth. A learning theory approach to non-interactive database privacy. In *STOC 2008*, pp. 609-618, ACM, 2008.

[3] B. Chor, A. Fiat, and M. Naor. Tracing traitors. In *CRYPTO 1994*, LNCS 839, pp.257–270, Springer, 1994.

[4] C. Dwork. Differential privacy. In *ICALP 2006 Part 2*, LNCS 4052, pp.1–12, Springer, 2006.

[5] C. Dwork. Differential privacy: A survey of results. In *Proc. of TAMC 2008*, LNCS 4978, pp.1–19, Springer, 2008.

[6] C. Dwork, K. Kenthapadi, F. McSHerry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Proc. of EUROCRYPT 2006*, LNCS 4004, pp.486–503, Springer, 2006.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proc. of TCC 2006*, LNCS 3876, pp.265–284, Springer, 2006.

[8] J. Gehrke, M. Hay, E. Lui, and R. Pass. Crowd-blending privacy. In *Proc. of CRYPTO 2012*, LNCS 7417, pp.479–496, Springer, 2012.

[9] J. Gehrke, E. Lui, and R. Pass. Twords privacy for social networks: A zero-knowledge based definition of privacy. In *Proc. of TCC 2011*, LNCS 6597, pp.432–449, Springer, 2011.

[10] S. Goldwasser, B. Barak, L. Reyzin, Y. Kalai, and S. Vadhan. 2011 年春に開かれた MIT とボストン大学での講義 “New Developments in Cryptography” の講義録の Class 11, Class 12, Class 13. http://www.cs.bu.edu/ reyzin/teaching/s11cs937/

[11] A. Groce, J. Katz, and A. Yerukhimovich. Limits of computational differential privacy in the client/server setting. In *Proc. of TCC 2011*, LNCS 6597, pp.417–431, Springer, 2011.

[12] M. Hardt, G.N. Rothblum, and R.A.Servedio. Private data release via learning thresholds. In *Proc. of SODA 2012*, pp. 168–187, SIAM, 2012.

[13] A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S.P. Vadhan. The limits of two party differential privacy. In *Proc. of FOCS 2010*, pp.81–90, IEEE Society, 2010.

[14] I. Mironov, O. Pandey, O. Reingold, and S.P. Vadhan. Computational differential privacy. In *CRYPTO 2009*, LNCS 5677, pp.126–142, Springer, 2009.

[15] O. Reingold, L. Trevisan, and S. Vadhan. Notions of reducibility between cryptographic primitives. In *Proc. of TCC 2004*, LNCS 2951, pp.1–20, Springer, 2004.

[16] J. Ullman. Answering $n^{2+o(1)}$ counting queries with differential privacy is hard, 2012. To appear in STOC 2013. Available at http://arxiv.org/pdf/1207.6945v2.pdf.

[17] 五十嵐大, 高橋克己. “注目のプライバシー Differential Privacy” コンピュータソフトウェア Vol.29, No.4 (2012), pp. 40-49.

[18] 佐久間淳, 小林重信. “プライバシ保護データマイニング” 人工知能学会誌 Vol.24, No.2 (2009), pp. 283-294.

[19] 佐久間淳, 高橋克己. “クラウドストレージにおける個人情報の利活用とプライバシー保護” 情報処理 Vol.52, No.6 (2011), pp. 706-715.