6.889 New Developments in Cryptography May 10, 2011

Differential Privacy

Instructors: Shafi Goldwasser, Yael Kalai, Leo Reyzin, Boaz Barak, and Salil Vadhan

Lecturer: Salil P. Vadhan *Scribe: Thomas Steinke*

1 Definition and Recap

Recall the definition of differential privacy.

Definition 1 *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a random mechanism. We say that M is* (ε, δ)-differentially private *if, for every $x, x' \in \mathcal{X}^n$ differing on only one row,*

$$\forall T \subset \mathcal{Y} \quad \mathbb{P}\left[M(x) \in T\right] \leq e^{\varepsilon}\mathbb{P}\left[M(x') \in T\right] + \delta.$$

We say that M is ε-differentially private if it is $(\varepsilon, 0)$-differentially private; equivalently, for every $x, x' \in \mathcal{X}^n$ differing on only one row, $D_\infty(M(x)||M(x')) \leq \varepsilon$, where, for random variables X and Y with range $\mathcal{Y}$, we define

$$D_\infty(X||Y) = \sup_{T \subset \mathcal{Y}} \log\left(\frac{\mathbb{P}\left[X \in T\right]}{\mathbb{P}\left[Y \in T\right]}\right).$$

The typical range of parameters is $\varepsilon \in [1/n, 1]$ being a constant such as 0.01 and $\delta = \text{negl}$ being cryptographically small.

Observe that, for discrete random variables X and Y,

$$D_\infty(X||Y) = \max_{t \in \mathcal{Y}} \log\left(\frac{\mathbb{P}\left[X = t\right]}{\mathbb{P}\left[Y = t\right]}\right).$$

There is also an equivalent simulator-based definition. This says that the amount one learns from $M(x)$ about any row in x is within ε of what one can learn about it from the rest of the database.

Proposition 2 *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a random function. Then M is ε-differentially private if and only if there exists a simulator S such that, for every $i \in [n]$,*

$$D_\infty(M(x)||S(x_{-i})) \leq \varepsilon \quad \wedge \quad D_\infty(S(x_{-i})||M(x)) \leq \varepsilon.$$

Note that we define x_{-i} to be the database with the i^{th} entry hidden.

Note that the equivalence in Proposition 2 can be extended to (ε, δ)-differential privacy.

2 A Bayesian Forumulation of Differential Privacy

Bayes' theorem gives us an equivalent and intuitive definition of differential privacy. This further justifies it being the "right" definition.

Roughly, the next proposition states that the output of a differentially private mechanism does not significantly change an adversary's beliefs about an individual. Note that the output of a (useful) differentially private mechanism will give the adversary information about the distribution. So we must assume that the adversary knows the distribution as part of his prior; otherwise this is false. Indeed, we assume the worst case—the adversary has access to all of the database except one entry.

Proposition 3 *If M is an ε-differentially private, then, for every $x \in \mathcal{X}^n$, $i \in [n]$, distribution X_i on $\mathcal{X}$ (the adversary's prior on x_i), and output y of $M(x)$,*

$$D_\infty(X_i||(X_i|M(X_i, x_{-i}) = y)) \leq \varepsilon \quad \wedge \quad D_\infty((X_i|M(X_i, x_{-i}) = y)||X_i) \leq \varepsilon,$$

where $X|M(X_i, x_{-i}) = y$ denotes the distribution of X_i conditioned on the output of M on X_i and the rest of the database x_{-i} being y.

Proof: By Bayes' theorem

$$\begin{aligned} \mathbb{P}\left[X_i = x_i' | M(X_i, x_{-i}) = y\right] &= \frac{\mathbb{P}\left[M(x_i', x_{-i}) = y\right]}{\mathbb{P}\left[M(X_i, x_{-i}) = y\right]} \mathbb{P}\left[X_i = x_i'\right] \\ &\in e^{\pm\varepsilon} \mathbb{P}\left[X_i = x_i'\right] \quad \text{(by } \varepsilon\text{-differential privacy).} \end{aligned} \tag{1}$$

So the prior and posterior only differ by a multiplicative factor in the range $[e^{-\varepsilon}, e^{+\varepsilon}]$, which gives the result. □

Proposition 4 is a converse to Proposition 3. Moreover, we see a different metric being used—statistical distance.

Proposition 4 *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a randomized mechanism. Suppose that, for every $x \in \mathcal{X}^n$, $i \in [n]$, distributions X_i on $\mathcal{X}$, and $y \in \mathcal{Y}$,*

$$\Delta(X_i, (X_i|M(X_i, x_{-i}) = y)) \leq \varepsilon.$$

Then M is $O(\varepsilon)$-differentially private.

Proof: Let $x, x' \in \mathcal{X}^n$ differ on one row $i \in [n]$; let $x = (x_i, x_{-i})$ and $x' = (x_i', x_{-i})$. Define a distribution X_i by

$$\mathbb{P}\left[X_i = x_i\right] = \mathbb{P}\left[X_i = x_i'\right] = 1/2.$$

Choose γ such that

$$\mathbb{P}\left[X_i = x_i | M(X_i, x_{-i}) = y\right] = (1+\gamma)/2 \quad \wedge \quad \mathbb{P}\left[X_i = x_i' | M(X_i, x_{-i}) = y\right] = (1-\gamma)/2.$$

Then X_i and $X_i|M(X_i, x_{-i}) = y$ are Bernoulli random variables, whence

$$\gamma = \Delta(X_i, (X_i|M(X_i, x_{-i}) = y)) \leq \varepsilon.$$

By Bayes' theorem,

$$\begin{aligned} \frac{\mathbb{P}\left[M(x') = y\right]}{\mathbb{P}\left[M(x) = y\right]} &= \frac{\mathbb{P}\left[M(x_i', x_{-i}) = y\right]}{\mathbb{P}\left[M(X_i, x_{-i}) = y\right]} \Big/ \frac{\mathbb{P}\left[M(x_i, x_{-i}) = y\right]}{\mathbb{P}\left[M(X_i, x_{-i}) = y\right]} \\ &= \frac{\mathbb{P}\left[X_i = x_i' | M(X_i, x_{-i}) = y\right]}{\mathbb{P}\left[X_i = x_i'\right]} \Big/ \frac{\mathbb{P}\left[X_i = x_i | M(X_i, x_{-i}) = y\right]}{\mathbb{P}\left[X_i = x_i\right]} \quad \text{(cf. (1))} \\ &= \frac{1+\gamma}{1-\gamma} = e^{O(\varepsilon)}. \end{aligned}$$

□

Remarks:

- Note that the proof of Proposition 4 only requires

$$\Delta\left(\text{Bernoulli}\left(\frac{1}{2}\right), \text{Bernoulli}\left(\frac{1+\gamma}{2}\right)\right) = \Omega(\gamma).$$

  Any metric on distributions that satisfies this will do.

- Propositions 3 and 4 can be generalized to (ε, δ)-differential privacy. For negligible δ, the statements hold with $1 -$ negligible probability over the randomness of M.

- We allow an arbitrary prior. So differential privacy is resilient to arbitrary side information. So an individual's data is safe, given that it is localized to one row in the database. There is no guarantee if the data is spread over the whole database.

3 Noninteractive Data Release

Ideally, we want a differentially private mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ such that $M(x)$ encodes a number of important statistics. For example, we want to be able to answer several counting queries $\mathcal{P} = \{\pi : \mathcal{X}^n \to [0,1]\}$ within accuracy $\pm\alpha$ from $M(x)$. Note that a counting query is of the form $q : \mathcal{X} \to \{0,1\}$ and the answer is $q(x) = (1/n)\sum_{i\in[n]} q(x_i)$.

One method of noninteractive data release is synthetic data: Synthetic data has exactly the same format as the original data; so M is of the form $M : \mathcal{X}^n \to \mathcal{X}^{n'}$. We say that M is α-accurate for a set $\mathcal{P}$ of queries if, for every $\pi \in \mathcal{P}$, we have $|\pi(x) - \pi(M(x))| \leq \alpha$. Producing synthetic data turns out to be problematic. As such, other forms of data release are of interest too.

3.1 BLR Mechanism

The Blum-Ligett-Roth (BLR) mechanism [BLR] is a ε-differentially private α-accurate noninteractive data release mechanism for a set of counting queries $\mathcal{P}$ as long as

$$\alpha \geq c\left(\frac{\log|\mathcal{X}|\log|\mathcal{P}|}{\varepsilon n}\right)^{1/3}$$

for some universal constant c. Moreover, the BLR mechanism produces synthetic data.

For example, if $\mathcal{X} = \{0,1\}^d$ and $\mathcal{P}$ is the class of conjunctions, then $|\mathcal{P}| = 3^d$ and we require $\alpha \geq c' d^{2/3}(\varepsilon n)^{-1/3}$. So we can produce ε-differentially private α-accurate synthetic data for any d-attribute database of size $n = \text{poly}(d, 1/\alpha, 1/\varepsilon)$.

The downside of the BLR mechanism is that it requires exponential running time and is thus infeasible in practice. This leads to the following open problem: Can we achieve differentially private noninteractive data release in polynomial time for some *interesting* class of queries?

- We know [UV] that creating differentially private synthetic data that preserves even two-conjunctions is hard (assuming that one-way functions exist).

- For non-synthetic-data release, we know of a close connection to the open problem of traitor tracing, which we discuss next.

4 Traitor Tracing

Suppose that we want to broadcast messages—such as a subscriber-only TV channel—to n subscribers (users), but we wish to ensure that non-subscribers do not gain access. Moreover, if subscribers betray the system and give their keys to non-subscribers, we want to be able to identify and punish the traitors. This goal can be formalized as a traitor tracing scheme as follows.

Definition 5 *A (fully resilient)* traitor tracing *scheme consists of polynomial-time algorithms Setup, Enc, Dec, and Trace such that the following hold.*

- *Setup is randomized and* $Setup(1^k, 1^n) = (bk, k_1, \cdots, k_n)$*, where bk is the* broadcast key *and, for* $i \in [n]$*,* k_i *is the* i^{th} user key*.*
- *Enc is randomized, Dec is deterministic, and for every* $i \in [n]$ *and* $m \in \{0,1\}$*,* $Dec_{k_i}(Enc_{bk}(m)) = m$*.*
- *Suppose that A is a probabilistic polynomial-time adversary that takes as input a subset* $\{k_i : i \in I\}$ *of the user keys and black-box access to* Enc_{bk}*. The goal of A is to produce an* α-pirate decoder Dec^**—that is, a polynomial-time algorithm such that* $\mathbb{P}\left[Dec^*(Enc_{bk}(m)) = m\right] \geq 1/2 + \alpha$ *for all m. However, Trace will identify at least one of the keys* k_i *(*$i \in I$*) that A used. In particular,*

$$\mathbb{P}\left[A^{Enc_{bk}}(k_i : i \in I) = Dec^* \wedge Trace^{Dec^*}(bk) \notin \{k_i : i \in I\}\right] = negligible(k)$$

—that is, the probability that A successfully produces an α*-pirate decoder and Trace fails to identify a key used in its construction is negligible. Moreover, the running time of Trace is* $poly(n, k, 1/\alpha)$*.*

We have the following trivial construction of a traitor tracing scheme.

Example 6 *Setup computes independent keys* $k_1 \cdots k_n$ *for a symmetric encryption scheme* (Enc', Dec') *and* $bk = (k_1, \cdots, k_n)$*. Then* $Enc_{bk}(m) = \left(Enc'_{k_1}(m), \cdots, Enc'_{k_n}(m)\right)$ *and* $Dec_{k_i}(c) = Dec'_{k_i}(c_i)$*. Trace functions as follows.*

Let $D_i = (Enc'_{k_1}(0), \cdots, Enc'_{k_i}(0), Enc'_{k_{i+1}}(1), \cdots, Enc'_{k_n}(1))$*. Then* $D_0 \sim Enc(1)$ *and* $D_n \sim Enc(0)$*. By assumption,* $\mathbb{P}\left[Dec^*(D_0) = 1\right] \geq 1/2 + \alpha$ *and* $\mathbb{P}\left[Dec^*(D_n) = 1\right] \leq 1/2 - \alpha$*. So* Dec^* *can* 2α*-distinguish* D_0 *and* D_n*. In particular, by a hybrid argument, there exists* $i^* \in [n]$ *such that* Dec^* *can* $\alpha/(n+1)$*-distinguish* D_{i^*} *from* D_{i^*-1}*. Moreover, with high probability Trace can find such an* i^* *in polynomial-time.*

By semantic security, for every $i \in [n]$*,* $Enc'_{k_i}(0)$ *and* $Enc'_{k_i}(1)$ *are indistinguishable to* Dec^* *unless A had access to* k_i*. Thus, unless A had access to* k_i*,* D_{i-1} *and* D_i *are indistinguishable to* Dec^**. Since* D_{i^*} *and* D_{i^*-1} *are distinguishable, A had access to* i^**.*

Clearly Example 6 is unsatisfactory as the key and message lengths grow linearly with the number of users.

4.1 Sahai-Waters Traitor Tracing Scheme

Boneh, Sahai, and Waters have constructed a traitor tracing scheme [BSW] that has $O(1)$-sized user keys and $O(\sqrt{n})$-sized ciphertexts (ignoring the security parameter k). It remains open to improve this.

4.2 Connection to Differential Privacy

Theorem 7 *An efficient traitor tracing scheme implies that there is no efficient differentially private data release mechanism for some class $\mathcal{P}$ of efficient queries.*

Theorem 7 implies that we cannot have both differentially private data release and traitor tracing schemes. Currently we have neither.

The intuition behind this result is that differentially private data release summarises the 'usefulness' of the data while hiding individual information. However, a traitor tracing scheme prevents keys being summarised without identifying an individual.

Proof: Suppose that (Setup, Enc, Dec, Trace) is a traitor tracing scheme and M is an ε-differentially private data release mechanism. Assume that $\mathcal{X} = \{0,1\}^d = \mathcal{K} = \{\text{user keys}\}$. Let $\mathcal{P} = \{\pi_c : \mathcal{K} \to \{0,1\}\}$, where c ranges over ciphertexts and $\pi_c(k) = \text{Dec}_k(c)$ is a counting query. Consider the following.

- Let $(bk, k_1, \cdots, k_{n+1}) = \text{Setup}(1^k, 1^{n+1})$.
- Let $x = (k_1, \cdots, k_n)$ be the database.
- Compute $y = M(x)$. Then y can be used to answer any query in $\mathcal{P}$ to within $\pm\alpha$ accuracy. We will use y to produce a pirate decoder.
- If $c = \text{Enc}_{bk}(0)$, then $\pi_c(x) = 0$. Likewise, if $c = \text{Enc}_{bk}(1)$, then $\pi_c(x) = 1$. But, from y, we can efficiently estimate $\pi_c(x)$. If $\alpha < 1/2$, we can decode perfectly. So we have a perfect pirate decoder Dec^*.
- Now $\text{Trace}^{\text{Dec}^*}$ outputs one of $k_1 \cdots k_n$ with high probability. This violates the privacy constraint.
- Let k_i be a most likely output of $\text{Trace}^{\text{Dec}^*}$ and let x' be x with k_i replaced with k_{n+1}. The probability that $\text{Trace}^{\text{Dec}^*}$ outputs k_i changes from at least $1/n$ to negligible. This is a large multiplicative difference, which violates the differential privacy of M.

□

It remains open to prove an analogue of Theorem 7 for a more natural class of queries.

References

[BLR] Avrim Blum, Katrina Ligett, and Aaron Roth. *A Learning Theory Approach to Non-Interactive Database Privacy*. STOC 2008.

[BSW] Dan Boneh, Amit Sahai, and Brent Waters. *Fully collusion resistant traitor tracing with short ciphertexts and private keys*. EUROCRYPT 2006.

[UV] Jonathan Ullman and Salil Vadhan. *PCPs and the Hardness of Generating Synthetic Data*. TCC 2011.