Entropy-minimizing Mechanism for Differential Privacy of Discrete-time Linear Feedback Systems

Yu Wang, Zhenqi Huang, Sayan Mitra and Geir E. Dullerud

Abstract—The concept of differential privacy stems from the study of private query of datasets. In this work, we apply this concept to metric spaces to study a mechanism that randomizes a deterministic query by adding mean-zero noise to keep differential privacy. For one-shot queries, we show that $\epsilon$-differential privacy of an $n$-dimensional input implies a lower bound $n - n \ln(\epsilon/2)$ on the entropy of the randomized output, and this lower bound is achieved by adding Laplacian noise. We then consider the $\epsilon$-differential privacy of a discrete-time linear feedback system in which noise is added to the system output at each time. The adversary estimates the system states from the output history. We show that, to keep the system $\epsilon$-differentially private, the output entropy is bounded below, and this lower bound is achieved by an explicit mechanism.

I. INTRODUCTION

The concept of $\epsilon$-differential privacy comes from the study of privacy-preserving queries of datasets [1]. To keep an entry of a dataset undetectable in general queries, the outputs of the queries should be randomized to blur the difference when the entry changes. A query $Q$ is called $\epsilon$-differentially private if for any two adjacent datasets $D_1, D_2$ that differ in only one entry, we have

$$P(Q(D_1) \in O) \leq e^{\epsilon} P(Q(D_2) \in O),$$

where $O$ is an arbitrary set of possible outputs.

When it comes to dynamic systems, the quantities involved are small sets of continuous number like real vectors, instead of datasets of numerous discrete values. Therefore, norms are employed to define adjacency. The common approach is to first defining a threshold $c > 0$ and then calling the two states $x_1, x_2 \in \mathbb{R}^n$ adjacent if $\|x_1 - x_2\| \leq c$ [2], [3]. But, two weaknesses exist in this idea: the choice of $c$ is arbitrary and the binary nature of adjacency remains. Two states $x_1, x_2$ are not adjacent even if $\|x_1 - x_2\| - c$ is small but positive.

On the other hand, randomizing the outputs undermines the accuracy of the query. In dynamic systems, this hinders efficient communications between system components, especially in multi-agent systems [4], [5]. Therefore, it is always desirable to find a mechanism minimizing the inaccuracy while preserving $\epsilon$-differential privacy.

Previous results show that there is trade-off between privacy and accuracy. For datasets, it has been observed that $\epsilon$-differential privacy imposes a upper bound on the min entropy that measures the information leakage of the queries. [6], [7]. For dynamics systems, it has been shown that, under the usual definition, $\epsilon$-differential privacy imposes a lower bound on the variance of queries and the minimum is achieved by staircase distributions [8], [9], [10].

In this work, we study the relationship between privacy and accuracy in the framework of a discrete-time feedback control system, where accuracy is measured by Shannon entropy and a version of $\epsilon$-differential privacy without adjacency is adopted [11] (see Section II). We first study an $\epsilon$-differentially private noise-adding mechanism for one-shot queries that provides the best output accuracy. Then, we consider the $\epsilon$-differential privacy of a discrete-time linear feedback system, which is characterized by the interplay of differential privacy and system dynamics. At each time, noise is added to the system outputs to keep the system states private against the adversary who has access to the system outputs. Since the system states and outputs at different time are related via the system dynamics, the adversary can filter the noise using the output history.

In the sequel, the preliminaries are presented in Section II. In Section III, we prove that, for a one-shot $n$-dimensional input, there is a lower bound $n - n \ln(\epsilon/2)$ on the entropy of the output for an $\epsilon$-differentially private noise-adding mechanism, and the lower bound is achieved by Laplacian noise with parameter $1/\epsilon$. In Section IV, we introduce a discrete-time linear feedback system where the noise-adding mechanism is incorporated to keep the system states $\epsilon$-differentially private, and give the definition of $\epsilon$-differential privacy and output entropy for the system. In Section V, we show that the $\epsilon$-differential privacy of the system implies a lower bound on the output entropy, and give explicitly a mechanism that achieves the lower bound. Finally, we conclude this work in Section VI.

II. PRELIMINARIES

Denote respectively the set of natural numbers, positive integers, positive real numbers and real numbers by $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{R}_+$ and $\mathbb{R}$. Let $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ be the set of $n$-dimensional real vectors and the set of $m$ by $n$ real matrices. Denote the positive orthant in $\mathbb{R}^n$ by $\mathbb{R}^n_+$. In this work, scalars and vectors are in lower case; matrices are in upper case; random variables are in roman font.

For $n \in \mathbb{N}$, let $[n] = \{0, 1, 2, \ldots, n\}$. For $x \in \mathbb{R}$, denote the absolute value of $x$ by $|x|$. For $x \in \mathbb{R}^n$, denote the $i$-th component of $x$ by $x_i$; i.e., $x = (x_1, x_2, \ldots, x_n)$. For $x, y \in \mathbb{R}^n$, let $x \cdot y = \sum_{i=1}^n x_i y_i$. For a function $f$, denote the image of $f$ by $\text{Im}(f)$. For a real-valued function $f$ on $\mathbb{R}$,
denote the first derivative by $f'$; for a real-valued function on $\mathbb{R}^n$, denote the gradient by $\nabla f$.

A metric space is denoted by $(\mathcal{M}, \rho(\cdot, \cdot))$ where $\rho(\cdot, \cdot)$ is the metric on $\mathcal{M}$. $\mathbb{R}^n$ equipped with a norm $\| \cdot \|$ is a metric space, denoted by $(\mathbb{R}^n, \| \cdot \|)$, where $\rho(x, y) = \|x - y\|$. In particular, the $\ell_1$-norm on $\mathbb{R}^n$, which will be frequently referred to later, is defined as $\|x\|_1 = \sum_{i=1}^{n} |x_i|$.

A random variable $X$ on $\mathbb{R}$ obeys Laplace distribution with parameter $\alpha$, denoted by $X \sim \text{Lap}(\alpha)$, if its probability distribution function is

$$f_X(x) = \frac{1}{2\alpha} \exp\left(\frac{|x|}{\alpha}\right).$$

Similarly, on $\mathbb{R}^n$, $X \sim \text{Lap}_n(\alpha)$ if its probability distribution function is

$$f_X(x) = \left(\frac{1}{2\alpha}\right)^n \exp\left(\frac{|x|}{\alpha}\right).$$

Let $X$ be a random variable on $\mathbb{R}^n$ with probability distribution function $f(x)$.

**Definition 1:** The entropy of $X$ is

$$H(X) = -\int_{\mathbb{R}^n} f(x) \ln(f(x)) \, dx.$$  

Informally speaking, the entropy is a measure of the uncertainty of a random variable, either positive or negative. Conditioning reduces entropy; for two random variables $X$ and $Y$,

$$H(X | Y) \leq H(X).$$

with equality iff $X$ and $Y$ are independent.

Suppose that we have some data $x$ taken from a metric space $(\mathcal{I}, \rho(\cdot, \cdot))$. To keep the query of $x$ private, we pass $x$ through a mechanism $\mathcal{M}$ which generates a randomized output $\mathcal{M}(x)$.

**Definition 2:** The mechanism is called $\epsilon$-differentially private for given $\epsilon > 0$, if the inequality

$$\mathbb{P} [\mathcal{M}(x_1) \subseteq O] \leq \exp(\epsilon \rho(x_1, x_2)) \mathbb{P} [\mathcal{M}(x_2) \subseteq O]$$

holds for any inputs $x_1, x_2$ and a set of possible outputs $O$.

In this work, as shown in Figure 1, we let the metric space be a real normed space $(\mathbb{R}^n, \| \cdot \|)$ and the mechanism $\mathcal{M}$ generate the randomized output by adding noise $N(x)$.

**III. Entropy Analysis for Differentially Private One-shot Queries**

In this section, we show that, for a single query, $\epsilon$-differential privacy implies a bound on the noise of the entropy and the bound is achieved by adding Laplacian noise.

Let $\mathcal{M}$ be a randomizing mechanism with input set $\mathcal{I} = (\mathbb{R}^n, \| \cdot \|)$. For any input $x \in \mathbb{R}^n$, the output of the mechanism $Y = \mathcal{M}(x)$ is a random variable on $\mathbb{R}^n$ with probability density function $f_x(y)$. We call $p(x, y) = f_x(y)$ the output distribution function, which is assumed to be absolutely continuous in $x$ and $y$. In addition, we assume that the noise added is mean-zero; i.e.,

$$\int_{\mathbb{R}^n} yp(x, y) \, dy = x.$$  

**Lemma 1:** A mechanism $\mathcal{M}$ with absolutely continuous output distribution function $p(x, y)$ is $\epsilon$-differentially private iff for any $x \in \mathbb{R}^n$ and unit vector $\hat{n}$ under the norm $\| \cdot \|$,

$$|\hat{n} \cdot \nabla_x p(x, y)| \leq \epsilon p(x, y)$$

holds for almost every $y$, where the gradient $\nabla_x p(x, y)$ is taken for fixed $y \in \mathbb{R}^n$.

**Proof:** Necessity. By Definition 2, for any two inputs $x_1, x_2 \in \mathbb{R}^n$ and an output set $O \subseteq \mathbb{R}^n$, we have

$$\int_{O} p(x_1, y) \, dy \leq \exp(\epsilon \|x_1 - x_2\|) \int_{O} p(x_2, y) \, dy.$$  

Since $p(x, y)$ is absolute continuous, we have $p(x_1, y) \leq \exp(\epsilon \|x_1 - x_2\|) p(x_2, y)$ for any $y \in \mathbb{R}^n$. Thus,

$$\frac{p(x_1, y) - p(x_2, y)}{\|x_1 - x_2\|} \leq \frac{\exp(\epsilon \|x_1 - x_2\|) - 1}{\|x_1 - x_2\|} p(x_2, y).$$

By letting $x_1$ converge to $x_2$, we have $\frac{\|x_1 - x_2\|}{\|x_1 - x_2\|} = \epsilon p(x_2, y)$, for $x$ almost everywhere, abbreviated as a.e. Since $x_1$ can approach $x_2$ in arbitrary direction, we have $|\hat{n} \cdot \nabla_x p(x, y)| \leq \epsilon p(x, y)$ for arbitrary unit vector $\hat{n}$ and $x$ a.e.

Sufficiency. For any two inputs $x_1, x_2 \in \mathbb{R}^n$, define

$$g(z) = p(x_2 + \hat{n} z, y),$$

where $\hat{n} = \frac{x_1 - x_2}{\|x_1 - x_2\|}$. Clearly, $g(0) = p(x_2, y)$ and $g(\|x_1 - x_2\|) = p(x_1, y)$. By (8), $|\hat{n} \cdot \nabla_x p(x, y)| \leq \epsilon g(z)$ for $z$ a.e. Thus,

$$|\ln p(x_1, y) - \ln p(x_2, y)| = |\ln g(\|x_1 - x_2\|) - \ln g(0)| \leq \int_0^{\|x_1 - x_2\|} \frac{|g'(z)|}{g(z)} \, dz \leq \epsilon \|x_1 - x_2\|.$$  

Therefore,

$$p(x_1, y) \leq \exp(\epsilon \|x_1 - x_2\|) p(x_2, y)$$

Finally, we finish the proof by integrating (13) over $y$ on any output set $O \subseteq \mathbb{R}^n$.

Given mechanism $\mathcal{M}$, we use the output entropy

$$H(\mathcal{M}) = \sup_{x \in \mathbb{R}^n} \int_{\mathbb{R}^n} -p(x, y) \ln p(x, y) \, dy.$$  

to evaluate the accuracy of outputs. In the following, we propose several properties of a mechanism $\mathcal{M}$ that minimizes the output entropy $H(\mathcal{M})$.

**Lemma 2:** Given a $\epsilon$-differentially private $\mathcal{M}$ with output distribution function $p(x, y)$, we can construct an $\epsilon$-differentially private mechanism $\mathcal{N}$ with output probability distribution function $q(x, y)$, such that (i) $H(\mathcal{N}) \leq H(\mathcal{M})$,
and (ii) for any \( x \in \mathbb{R}^n \), \( g_x(y) = q(x, y - x) \) is even with respect to each \( y_m (m \in [n]) \).

**Proof:** Without loss of generality, assume \( m = 1 \). Let

\[
H_+ (\mathcal{M}) = \sup_{x \in \mathbb{R}^n} \int_{[x_1, \infty) \times \mathbb{R}^{n-1}} -p(x, y) \ln p(x, y) dy,
\]

\[
H_- (\mathcal{M}) = \sup_{x \in \mathbb{R}^n} \int_{(-\infty, x_1] \times \mathbb{R}^{n-1}} -p(x, y) \ln p(x, y) dy.
\]

(15)

For fixed \( \{x_i : i \neq 1, i \in [n]\} \) and \( \{y_i : i \neq 1, i \in [n]\} \), if \( H_+ (\mathcal{M}) \leq H_- (\mathcal{M}) \), then define

\[
q(x_1, y_1) = \begin{cases} p(2a - x_1, 2a - y_1), & x_1 \in L^+, \\ p(x_1, y_1), & x_1 \in L^- \end{cases}
\]

(23)

otherwise, define

\[
q(x_1, y_1) = \begin{cases} p(2a - x_1, 2a - y_1), & x_1 \in L^+, \\ p(x_1, y_1), & x_1 \in L^- \end{cases}
\]

(24)

Clearly, \( q(x_1, y_1) = q(2a - x_1, 2a - y_1) \). In addition, it is easy to check that \( q(x, y) \geq 0 \), \( \int_{\mathbb{R}^n} q(x, y) dy = 1 \) and \( \int_{\mathbb{R}^n} y q(x, y) dy = x \). By Lemma 2, we have \( p(a, y_i) = p(a, 2a - y_1) \), thus \( q(x_1, y_i) \), i.e. \( q(x, y) \) is absolutely continuous and \( \nabla_x q(x, y) = |\nabla_x p(x, y)|. \) Consequently, \( n|\nabla \cdot \nabla_x(x, y) | \leq \epsilon q(x, y) \) for \( x \) a.e. By Lemma 1, \( N \) is a well-defined \( \epsilon \)-differentially private mechanism and

\[
H(N) = \min \{H_+ (\mathcal{M}), H_- (\mathcal{M})\} \\
\leq \max \{H_+ (\mathcal{M}), H_- (\mathcal{M})\} = H(M),
\]

(25)

where the equality holds iff \( H_+ (\mathcal{M}) = H_- (\mathcal{M}) \).□

By the above two Lemmas, we derive the following result.

**Lemma 4:** There exists an \( \epsilon \)-differentially private mechanism \( \mathcal{M} \), such that (i) \( \mathcal{M} \) minimizes the output entropy \( H(\mathcal{M}) \), and (ii) the noise added is independent of the input \( x \in \mathbb{R}^n \).

**Proof:** Let \( q(x, y) = p(x, y - x) \). Then \( q(x, y) \) is the probability distribution function of the noise \( \mathcal{N}(x) \) for input \( x \). By Lemma 3, for any \( a \in \mathbb{R} \) and \( m \in [n] \), fixing \( \{x_i : i \neq m, i \in [n]\} \) and \( \{y_i : i \neq m, i \in [n]\} \), we can have \( q(x_m, y_m) = q(2a - x_m, -y_m) \). By Lemma 2, we can have \( q(x_m, y_m) = q(2a - x_m, -y_m) = q(2a - x_m, y_m) \). Therefore, \( q(x, y) \) is independent of \( x_m \), hence independent of \( x \).

By Lemma 2 and Lemma 4, if a mechanism \( \mathcal{M} \) minimizes the output entropy \( H(\mathcal{M}) \), the output distribution function will be in the form \( p(x, y) = f(y - x) \), where \( f \) is an absolutely continuous probability distribution function. For each \( i \in [n] \), \( f \) is a even function with respect to \( x_i \). For now, we assume a scalar input, that is \( n = 1 \). Then the problem of minimizing (14) becomes

**Problem 1 (Scalar Case):**

Minimize: \( H(f) = -\int_{[0, \infty)} f(x) \ln f(x) dx \),

subject to: \( f(x) \) is absolutely continuous,

\( f(x) \geq 0 \), \n\( |f'(x)| \leq \epsilon f(x) \) a.e.,

\( \int_{[0, \infty)} f(x) dx = \frac{1}{2} \).
Lemma 5: f(x) is nonincreasing if it solves Problem 1.

Proof: Let f(x) be a function that solves Problem 1. Let g(x) = sup_{y \geq x} f(y). Clearly, g(x) \geq f(x) for x \geq 0. Assume that f(x) is not non-increasing. Then for some x^* > 0, g(x^*) > f(x^*). By continuity of f, there exists a "largest" non-empty interval (a, b) containing x^*, on which g(x) > f(x). Note that b is finite since f(x) > 0 and lim_{x \to \infty} f(x) = 0. In addition, g(b) = f(b). Let d = \frac{1}{f(b)} \int_0^b f(x)dx, then d \in [0, b - a).

For a = 0, define

\[ h(x) = \begin{cases} 
  f(b), & x \in [0, d], \\
  f(x + b - d), & x \in [d, \infty].
\end{cases} \tag{26} \]

Otherwise, we have f(a) = g(a) = f(b) = g(b) and define

\[ h(x) = \begin{cases} 
  f(x), & x \in [0, a], \\
  f(b), & x \in [a, a + d], \\
  f(x + b - a - d), & x \in [a + d, \infty],
\end{cases} \tag{27} \]

In both cases, h(x) satisfies the constraints in Problem 1 and H(h) < H(f). This is in contradiction with the assumption.

Theorem 6: For a \( \epsilon \)-differentially private mechanism \( \mathcal{M} \) with input set \( \mathcal{I} = (\mathbb{R}, \cdot, \cdot) \), the output entropy is lower bounded by \( 1 - \ln(\epsilon/2) \) and the minimum is achieved by \( p(x, y) = \frac{\epsilon}{2} \exp(-\epsilon y - x) \).

Proof: It suffices to show that in the above minimization Problem 1, \( H \geq \frac{1}{2} - \frac{1}{2} \ln(\epsilon/2) \) and the minimum is achieved by \( f(x) = \frac{\epsilon}{2} \exp(-\epsilon x) \).

Let \( F(x) = \int_{-\infty}^{\infty} f(y)dy \). Noting that \( \lim_{x \to \infty} f(x) = 0 \), by the definition of \( \epsilon \)-differential privacy,

\[ F(x) \geq \int_{-\infty}^{x} \frac{|f'(y)|}{\epsilon} dy \geq \frac{1}{\epsilon} \int_{-\infty}^{x} f'(y)dy \]

In particular, \( f(0) \geq \epsilon F(0) = \frac{\epsilon}{2} \). In (28), the first equality holds iff \( |f'(y)| = \epsilon f(x) \) for x a.e. The second equality holds iff \( f'(x) = -\epsilon f(x) \) is monotonous. Considering that \( f'(x) > 0 \) and \( f(\infty) = 0 \), \( f(x) \) is monotonously decreasing. Thus, \( f(x) = \epsilon F(x) \) iff \( f'(x) = -\epsilon f(x) \).

By Lemma 5, \( f'(x) \leq 0 \) a.e. Thus,

\[ H(f) = -\int_{-\infty}^{\infty} f(x) \ln f(x)dx \]

\[ = -\int_{-\infty}^{x} f(x) \left( \ln f(0) + \int_{-\infty}^{x} f'(y)dy \right) dx \]

\[ = \frac{1}{2} \ln f(0) - \int_{-\infty}^{x} f'(y)f(y)dy \]

\[ \geq -\frac{1}{2} \ln f(0) - \int_{-\infty}^{\infty} f'(y)f(y)dy \]

\[ = \frac{f(0)}{\epsilon} - \frac{1}{2} \ln f(0), \]

where the equality holds iff \( f'(x) = -\epsilon f(x) \).

Since \( f(0) \in (0, \frac{\epsilon}{2}) \) and \( \frac{f(0)}{\epsilon} - \frac{1}{2} \ln f(0) \) is decreasing on this interval, we have \( H \geq \frac{1}{2} - \frac{1}{2} \ln(\frac{\epsilon}{2}) \). Again, the equality holds if \( \epsilon f'(x) = -\epsilon f(x) \) a.e.

Finally, by \( f(x) \geq 0 \) and \( \int_{0,\infty} f(x)dx = \frac{1}{2} \), we derive that the minimum is achieved by

\[ f(x) = \frac{\epsilon}{2} \exp(-\epsilon x). \tag{30} \]

In this case, we have \( H(f) = \frac{1}{2} - \frac{1}{2} \ln(\frac{\epsilon}{2}) \).

Remark 7: If we replace the restriction \( \int_{0,\infty} f(x)dx = \frac{1}{2} \) in Problem 1 with \( \int_{0,\infty} f(x)dx = \frac{\epsilon}{2} \) where \( \epsilon > 0 \) is a real constant, then the above proof will still work. In this case, we have \( H \geq \frac{1}{2} - \frac{1}{2} \ln(\frac{\epsilon}{2}) + \frac{1}{2} \ln \epsilon \) and the minimum is achieved by \( f(x) = \frac{\epsilon}{2} \exp(-\epsilon x) \).

For multidimensional inputs, \( n \geq 2 \), noting that we have taken the \( \ell_1 \)-norm on \( \mathbb{R}^n \), the problem of minimizing (14) becomes

Problem 2 (Multi-dimensional Case):

Minimize: \( H(f) = -\int_{\mathbb{R}^n} f(x) \ln f(x)dx \),

subject to: \( f(x) \) is absolutely continuous, \( f(x) \geq 0 \),

\[ \frac{\partial f(x)}{\partial x_i} \leq \epsilon f(x), \forall i \in [n] \text{ a.e.,} \]

\[ \int_{\mathbb{R}^n} f(x)dx = \frac{1}{2^n}. \]

Theorem 8: For a \( \epsilon \)-differentially private mechanism \( \mathcal{M} \) with input set \( \mathcal{I} = (\mathbb{R}^n, \| \cdot \|_1) \), the output entropy is lower bounded by \( H(\mathcal{M}) \geq n - n \ln(\epsilon/2) \) and the minimum is achieved by \( p(x, y) = \left( \frac{\epsilon}{2} \right)^n \exp(-\epsilon x) \).

Proof: It suffices to show that in the minimization Problem 2, \( H \geq \frac{n}{2} - \frac{n}{2} \ln(\epsilon/2) \) and the minimum is achieved by \( f(x_1, x_2, \ldots, x_n) = \left( \frac{\epsilon}{2} \right)^n \exp(-\epsilon(x_1 + x_2 + \ldots + x_n)) \).

For each fixed \( x_2, x_3, \ldots, x_n \), let

\[ g_{x_2, x_3, \ldots, x_n}(x_1) = f(x_1, x_2, \ldots, x_n), \tag{31} \]

then we have \( g_{x_2, x_3, \ldots, x_n}(x_1) \geq 0 \), \( |g'_{x_2, x_3, \ldots, x_n}(x_1)| \leq \epsilon g_{x_2, x_3, \ldots, x_n}(x_1) \) and

\[ H(f) = -\int_{\mathbb{R}^n} \int_{0,\infty} g_{x_2, x_3, \ldots, x_n}(x_1)dx_1 dx_2 dx_3 \ldots dx_n. \tag{32} \]

By Remark 7, to minimize H, we should have

\[ f(x_1, x_2, \ldots, x_n) = g_{x_2, x_3, \ldots, x_n}(x_1) = \exp(-\epsilon x_1) \ln(x_2, x_3, \ldots, x_n) \tag{33} \]

where \( h(x_2, x_3, \ldots, x_n) \) is some function of \( x_2, x_3, \ldots, x_n \).

By repeating the above argument, we derive that the minimum is achieved by

\[ f(x_1, x_2, \ldots, x_n) = k \exp(-\epsilon(x_1 + x_2 + \ldots + x_n)) \tag{34} \]

where \( k \) is some constant. Finally, by \( \int_{\mathbb{R}^n} f(x)dx = \frac{1}{2^n} \), we have \( k = \left( \frac{\epsilon}{2} \right)^n \). In this case, the lower bound is \( H(f) = \frac{n}{2} - \frac{n}{2} \ln(\epsilon/2) \).
The system output $Z(t)$ is accessible to the adversary. As shown in (38), due to the feedback setting, given the past history of outputs $\{Z(t) \mid t \in \mathbb{N}\}$, we can derive the trajectory of system states $\{X(t) \mid t \in \mathbb{N}\}$ from the initial system state $X(0)$. Therefore, protecting the $\epsilon$-differential privacy of the initial system state is equivalent to protecting the $\epsilon$-differential privacy of the whole system state trajectory.

For $t \geq 0$, the adversary $\mathcal{A}$ estimates the initial system state from the past history of output $\{Z(i) \mid i \in [t]\}$. The best estimation of $X(0)$ at time $t$ is a random variable such that

$$
\tilde{X}(t) = \mathbb{E} \left[ X(0) \mid Z(0), Z(1), \ldots, Z(t) \right].
$$

We denote the probability density function of $\tilde{X}(t)$ by $\tilde{h}_t$. Since the adversary gets cumulating outputs, the entropy of $X(t)$ is not increasing.

The mechanism $\mathcal{L}$ is $\epsilon$-differentially private up to time $t \in \mathbb{N}$, if for any pair of initial states $x_1, x_2 \in \mathbb{R}^n$, and output history $\{z(i) \mid i \in [t]\}$,

$$
\mathbb{P} \left[ Z(1) = z_1, \ldots, Z(t) = z_t \mid X(0) = x_1 \right] \\
\mathbb{P} \left[ Z(1) = z_1, \ldots, Z(t) = z_t \mid X(0) = x_2 \right] \\
\leq \exp(\epsilon \|x_1 - x_2\|)
$$

The output entropy $\mathbf{H}(\mathcal{L}, t)$ of the mechanism $\mathcal{L}$ at time $t \in \mathbb{N}$ is defined by

$$
\mathbf{H}(\mathcal{L}, t) = \mathbf{H}(\tilde{X}(t)).
$$

Since we have no a prior knowledge on $X(0)$, by Bayes formula, (42) is equivalent to

$$
\tilde{h}_t(x_1) \leq \exp(\epsilon \|x_1 - x_2\|) \tilde{h}_t(x_2).
$$

Roughly speaking, the shape of distribution of $\tilde{X}(t)$ is flat enough to mask two different guesses $x_1, x_2$ of the initial state.

### V. Entropy Analysis

In this section, we consider the $\epsilon$-differentially private mechanisms and derive the one that gives the most accurate output up to time $t \geq 0$; i.e., a mechanism $\mathcal{L}$ that minimizes the output entropy $\mathbf{H}(\mathcal{L}, i)$ at each time $i \in [t]$ while preserving $\epsilon$-differential privacy.

First we assume that the norm used in this section is the $\ell_1$-norm $\| \cdot \|_1$. By combining (35) (36), we have

$$
X(t + 1) = (A + B)X(t) + BN(t),
$$

which implies that

$$
X(t) = (A + B)^t X(0) + \sum_{i=0}^{t-1} (A + B)^{t-i-1} BN(i).
$$

Thus, $Z(0) = X(0) + N(0)$ and for $t \geq 1$

$$
Z(t) = (A + B)^t X(0) + \sum_{i=0}^{t-1} (A + B)^{t-i-1} BN(i) + N(t)
$$

![Fig. 2. Block Diagram for an $\epsilon$-Differentially Private Discrete-time Linear Feedback System](image-url)
On the other hand, by (38) and (39), we have $N(0) = Z(0) - X(0)$ and for $t \geq 1$, 
\begin{equation}
N(t) = Z(t) - X(t) = Z(t) - A^t X(0) - \sum_{i=0}^{t-1} A^{i-1} B Z(i) \tag{48}
\end{equation}

Lemma 9: Given the initial state $X(0)$ and a history of outputs $\{Z(i) \mid i \in [t-1]\}$, we have 
\begin{equation}
\mathbb{E}[N(t) \mid X(0), Z(0), Z(1), \ldots, Z(t-1)] = 0. \tag{49}
\end{equation}

Proof: By (47) and (48), there is an one-to-one linear map between $\{X(0), (X(0), N(0), N(1), \ldots, N(t-1))\}$ and $\{X(0), Z(0), Z(1), \ldots, Z(t-1)\}$. Then by (40), (49) holds.

Lemma 10: Given an output history $\{z_i \mid i \in [t]\}$, the following equality holds: 
\begin{equation}
A^t \mathbb{E} [X(t)] = w_t, \tag{50}
\end{equation}
where $w_0 = z_0$ and $w_t = z_t - \sum_{i=0}^{t-1} A^{t-i-1} B z_i$ for $t \geq 1$.

Proof: Take conditional expectation on (48) with respect to $\{Z(i) = z_i \mid i \in [t]\}$.

Theorem 11: If a mechanism is $\epsilon$-differentially private up to time $t \geq 0$, then the output entropy obeys 
\begin{equation}
\mathbb{H}(L, i) \geq n - n \ln(\frac{\epsilon}{2}) \tag{51}
\end{equation}
for $i \in [t]$. The equality holds when $N(0) \sim \text{Lap}_n(1/\epsilon)$, and for $t \geq 1$, $N(t) = AN(t - 1)$.

Proof: Denote the joint probability distribution function of $\{N(i) \mid i \in [t]\}$ by $p_t(n_0, n_1, \ldots, n_t)$. For an output history $\{z_i \mid i \in [t]\}$, by (48), 
\begin{align*}
\tilde{h}_0(x) &= c_0 p_0(w_0 - x) \\
\tilde{h}_1(x) &= c_1 p_1(w_0 - x, w_1 - A x) \\
& \ldots \\
\tilde{h}_t(x) &= c_t p_t(w_0 - x, \ldots, w_t - A^t x) \\
& \ldots
\end{align*}
(52)

where $\{c_t \mid t \in \mathbb{N}\}$ are unifying constants. To minimize the entropy of $\tilde{X}(t)$ at each time $t \in \mathbb{N}$ while ensuring $\epsilon$-differential privacy, by Theorem 6, we should have 
\begin{equation}
\tilde{h}_t(x) = \frac{\epsilon}{2} \exp(-\epsilon \|x - \mathbb{E}[\tilde{X}(t)]\|_1). \tag{53}
\end{equation}

Plug (50) into (52). Since the new equalities holds for arbitrary $x$ and $\{w_t \mid t \in \mathbb{N}\}$, by comparing the arguments, we have 
\begin{align*}
p_0(x_0) &= d_1 \exp(-\epsilon \|x_0\|_1) \\
p_1(x_0, A x_1) &= d_2 \exp(-\epsilon \|x_1\|_1) \\
& \ldots \\
p_t(x_0, \ldots, A^t x_t) &= d_t \exp(-\epsilon \|x_t\|_1) \\
& \ldots
\end{align*}
(54)

where $\{x_t \mid t \in \mathbb{N}\} \subseteq \mathbb{R}$ and $\{d_t \mid t \in \mathbb{N}\}$ are unifying constants. A mechanism that satisfies (54) is $N(0) \sim \text{Lap}(1/\epsilon)$ and for $t \geq 1$, $N(t) = AN(t - 1)$. The noise $N(1)$ is determined by $N(0)$; $N(2)$ is determined by $N(0), N(1)$; and so on.

In this case, the output entropy is independent of time $t$, 
\begin{equation}
\mathbb{H}(L, 1) = \mathbb{H}(L, 2) = \ldots = \mathbb{H}(L, t) = n - n \ln(\frac{\epsilon}{2}). \tag{55}
\end{equation}
This is because all the randomness is brought to the system at the initial time.

VI. CONCLUSION

In this work, we introduced the entropy-minimizing mechanism for differential privacy. In particular, we studied a query mechanism which randomizes a deterministic quantity by adding mean-zero noise to keep differential privacy. For one-shot queries, we showed that for $n$-dimensional input, $\epsilon$-differential privacy implies a lower bound $n - n \ln(\epsilon/2)$ on the entropy of the randomized output of the mechanism, and this lower bound is achievable by adding Laplacian noise with parameter $\epsilon$. Then we studied a discrete-time linear feedback system that simplifies the system in [12], where the adversary is modeled as a filter on the system output. We demonstrate that there is a lower bound on the entropy if the mechanism ensures the $\epsilon$-differential privacy of the system. This lower bound is achievable by $N(0) \sim \text{Lap}_n(1/\epsilon)$, and for $t \geq 1$, $N(t) = AN(t - 1)$.

REFERENCES