An Additional Definition

Jie Zhang, Zhiwu Li

Differential privacy is a privacy-preserving method with a rigorous mathematical definition, which offers a mechanism (or a function) that publishes aggregate information about a statistical database, where the private information in it is protected or restricted. In other words, if a dataset is considered as an input of a differential privacy mechanism, the addition or deletion of any one record (or element) in it does not affect the query result, i.e., an intruder cannot capture the private information with the slight modification of the dataset.

Definition 1 (Differential Privacy) Let ϵ be a positive real number, and $\mathcal{F}$ be a randomized mechanism (or function) that takes a dataset as input. Let $Im(\mathcal{F})$ denote the image of $\mathcal{F}$. The mechanism $\mathcal{F}$ provides ϵ-differential privacy if for any two datasets O_1 and O_2 that differ on a single element, for all $O_3 \subseteq Im(\mathcal{F})$, the following inequality holds:

$$\mathbb{P}\left(\mathcal{F}(O_1) \in O_3\right) \leq \exp(\epsilon) \cdot \mathbb{P}\left(\mathcal{F}(O_2) \in O_3\right), \tag{1}$$

where the value of $\mathcal{F}$ at a dataset O_1 or O_2 is contained in the sample space, i.e., $Im(\mathcal{F})$, with a probability decided by the randomness used in the mechanism. The notation $\mathbb{P}\left(\mathcal{F}(O_i) \in O_3\right)$ with $i \in \{1, 2\}$ represents the probability that the output of $\mathcal{F}$ at O_i belongs to O_3. $\diamondsuit$

In Definition 1, the value ϵ evaluates the performance of differential privacy. Namely, a smaller value of ϵ implies a finer difference between the probabilities of $\mathcal{F}(O_1) \in O_3$ and $\mathcal{F}(O_2) \in O_3$, i.e., the intruder is less likely to distinguish the two datasets. On the contrary, a larger ϵ means a lower degree of users' private information protection.

Remark 1 *By Definition 1, the two datasets O_1 and O_2, are required to be different on a single element. However, this work introduces event string differential privacy in the field of DESs modeled by DFAs $\mathcal{G} = (X, \Sigma, f, x_0)$, where the event set Σ can be partitioned into two disjoint parts: observable event set Σ_o and unobservable event set Σ_{uo}, and defines $\mathcal{F}_\psi : \Sigma_o^k \to \Sigma_o^k$ as the mechanism specifically defined on a set of event strings. In the modified definition of differential privacy catering for a DES, the input and output of $\mathcal{F}_\psi$ are two event strings with the same length, whose similarity is measured by the trajectory distance, which is a metric defined in Section 4.*