Counting Distinct Elements in the Turnstile Model with Differential Privacy under Continual Observation

Palak Jain, Iden Kalemaj, Sofya Raskhodnikova, Satchit Sivakumar, Adam Smith

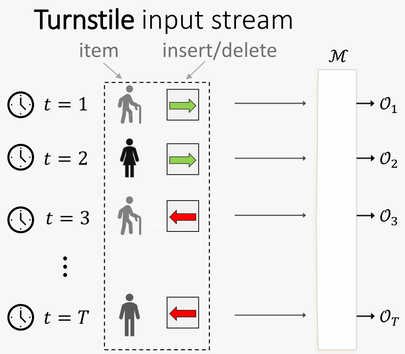
BOSTON UNIVERSITY

Privacy in Streaming Settings

Continual Observation Model of Differential Privacy:

- Introduced by [Dwork Naor Pitassi Rothblum '10] & [Chan Shi Song '10].
- Formalizes privacy in streaming settings where statistics change over time and need to be monitored continuously.

A mechanism in this setting receives inputs continuously over time and at each time produces an output.



Additive error of mechanism $\mathcal{M}$ for CountDistinct:

$$\max_{t\in[T]} |O_t - \text{CountDistinct}(t)| \le \alpha \qquad w.p. \ge 0.99$$

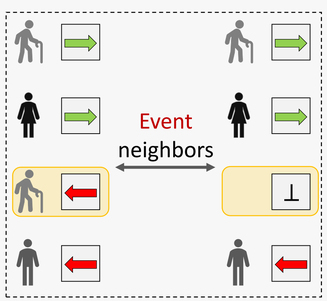
Privacy of mechanism $\mathcal{M}$ for CountDistinct:

Let $\mathcal{M}(x)$ be the **entire list of outputs** of $\mathcal{M}$ on input stream x. A mechanism $\mathcal{M}$ is (ε, δ)-**differentially private** if for all pairs x, x' of **neighboring streams** and all events S in the output space of $\mathcal{M}$

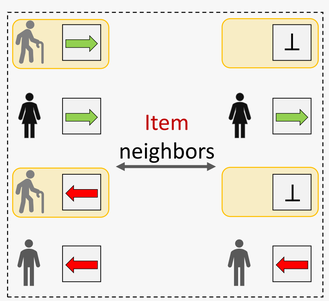
$$\Pr[\mathcal{M}(x) \in S] \le e^{\varepsilon} \cdot \Pr[\mathcal{M}(x') \in S] + \delta.$$

Two common definitions of neighboring streams yield two different levels of privacy protection:

Event-Level Differential Privacy



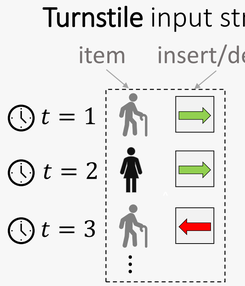
Item-Level Differential Privacy



We study the achievable accuracy of differentially private mechanisms for counting distinct elements in turnstile streams

- Privacy is a central challenge for systems that learn from sensitive data
- Even more challenging when the system's outputs are continuously updated
- Counting the number of distinct elements is a fundamental task
  - e.g., counting the number of distinct accounts logged into a streaming service

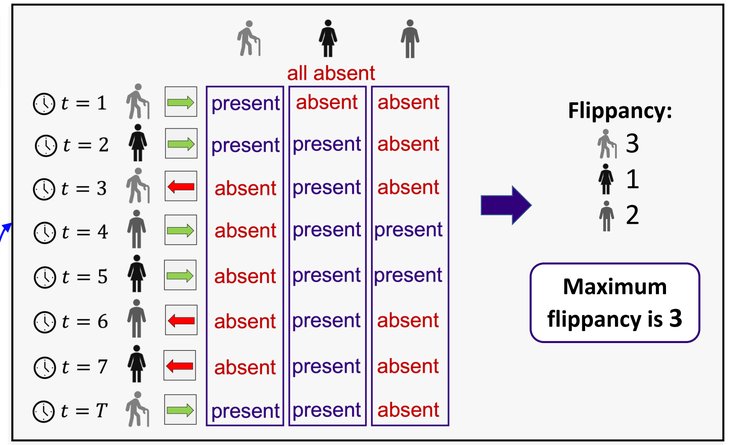
Problem Definition: Counting Distinct Elements



$$\text{CountDistinct}(t) = \sum_{\text{all items } i} \mathbb{1}[i \text{ inserted more than deleted up to step } t]$$

Maximum Flippancy of a Stream

Maximum number of times that an item switches between being present and being absent in the stream



OUR CONTRIBUTIONS

- Design an item-level private mechanism for counting distinct elements in the turnstile model, under continual observation.
- Identify a stream parameter called **maximum flippancy** that is low for many natural streams and analyze the accuracy of the mechanism in terms of it.
- Prove nearly matching lower bounds in terms of the maximum flippancy:
  - Use the sequential embedding technique of **[Jain Raskhodnikova Sivakumar Smith '23]**
  - Rely on deletions to embed multiple instances of base problems into a stream.

Table 1: Bounds on the additive error of differentially private mechanisms for CountDistinct over streams with max flippancy w.

Gap for streams with $w \in (T^{1/2}, T^{2/3})$

$\left(1, o\left(\frac{1}{T}\right)\right)$-DP

	Event-Level Privacy	Item-Level Privacy
Insertion-only Streams ($w = 1$)	$\Theta(\text{polylog } T)$ [Bolot Fawaz Muthukrishnan Nikolov Taft '13] [Folklore]	
Turnstile Streams	$\widetilde{\Omega}(\min(\sqrt{w}\ \text{polylog}\ T, T^{1/4})$ $\widetilde{O}(\min(\sqrt{w}\ \text{polylog}\ T, T^{1/3})$ [Our Work]	$\widetilde{\Theta}(\min(\sqrt{w}\ \text{polylog}\ T, T^{1/3})$ [Our Work]

Our Mechanism for Counting Distinct Elements

High-level structure of our mechanism

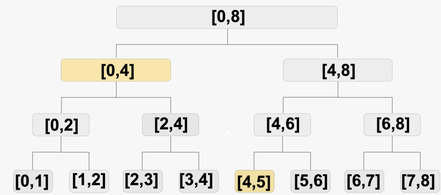
1. Mechanism for known flippancy w (below)

$$\left(\begin{array}{c}\text{Tree mechanism}\\ \text{of [BFMNT'13]}\end{array}\right) + \left(\begin{array}{c}\text{novel analysis for}\\ \textit{deletions with flippancy } w\end{array}\right)$$

2. Extension to dynamically choose w (in paper)

[BFMNT'13] Mechanism for insertion-only streams

1) Create a binary tree with labels as shown below:



2) In node $[s,t]$, store $\text{CountDistinct}(t) - \text{CountDistinct}(s) + (noise)$

3) At time t, sum the values of nodes in dyadic decomposition of $[0,t]$

(Dyadic decomposition of $[0,5]$ is highlighted above.)

[Our insights] Tree mechanism can be modified for deletions

- If flippancy of all items is $\le w$, then
  - Changing one item affects up to ***w* nodes at each level**
  - Each **node sum changes by at most 1**
  - Suffices to add noise $\approx \sqrt{w \log T}$ at each node
- Flippancy bound can be enforced via **stable transformation**

References:

[Chan Shi Song '10] A Private and Continual Release of Statistics. ICALP 2010.
[Dwork Naor Pitassi Rothblum '10] Differential Privacy under Continual Observation. STOC 2010.
[Bolot Fawaz Muthukrishnan Nikolov Taft '13] Private Decayed Sum Estimation under Continual Observation. ICDT 2013.
[Jain Raskhodnikova Sivakumar Smith '23] The Price of Differential Privacy under Continual Observation. ICML 2023.