

# Adaptively Robust Resettable Streaming



Edith Cohen<sup>1,3</sup>, Elena Gribelyuk<sup>2,1</sup>, Jelani Nelson<sup>4,1</sup>, Uri Stemmer<sup>3,1</sup>

Google Research<sup>1</sup>, Princeton University<sup>2</sup>, Tel Aviv University<sup>3</sup>, UC Berkeley<sup>4</sup>

## Resettable Streaming Model

- **Input:** updates to an underlying frequency vector  $x \in \mathbb{Z}^n$ , which arrive sequentially one at a time (*worst-case*, fixed in advance).
  - **Insert**( $i, \Delta$ ):  $x_i \leftarrow x_i + \Delta$
  - **Reset**( $i$ ):  $x_i \leftarrow 0$
- **Output:** At the end of the stream,  $A$  outputs an approximation of a given function of the stream.
- **Goal:**  $A$  should use space *sublinear* in the length  $T$  of the update sequence and universe size  $n$ .

## Adversarially Robust Streaming

- **Input:** updates (**Insert** or **Reset**), which arrive sequentially and *adversarially*.
- **Output:** At each time  $t$ ,  $A$  receives an update  $u_t$ , updates its internal state, and returns a *current estimate*  $r_t$ , which is recorded by the *adversary*.

“Future updates may depend on previous estimates”

We consider the  $F_p$  **moment estimation** problem for  $p \in [0,1]$ :

- Given a stream of  $T$  updates to coordinates  $i \in [n]$ , let  $x_i$  denote the frequency of element  $i$ .

$$F_p = \sum_{i \in [n]} |x_i|^p$$

- We let  $F^{(t)}$  denote the value of the relevant statistic on the *prefix* of the stream until time  $t$ . Let  $\epsilon \in (0,1)$  be a fixed accuracy parameter.
- By reduction from set-disjointness, any algorithm which achieves  $(1 + \epsilon)$  relative error (w.p.  $2/3$ ) requires  $\Omega(n)$  space.
- **Goal:** At each step  $t \in T$ , output an estimate  $\hat{F}^{(t)}$  such that  $|\hat{F}^{(t)} - F^{(t)}| \leq \epsilon \cdot \max_{t' \leq t} F^{(t')}$  (*prefix-max* error guarantee)

## Previous Approaches

- **Sketch Switching** [BJWY20]:  $\tilde{O}(T)$  space
- **Differential Privacy Robustification** [HKMMS20]:  $\tilde{O}(\sqrt{T})$  space
- All known algorithms rely on *persistent randomness*, which can be learned by the adversary.

**Question:** can we design adaptively robust sketches for resettable streaming that return a *prefix-max* estimate to fundamental statistics using  $\text{poly}(1/\epsilon, \log 1/\delta, \log T)$  space?

## Main Contribution

**Theorem (Robust resettable cardinality, sum, and Bernstein; informal):** For any  $\epsilon, \delta \in (0,1)$  and a resettable adaptive stream with  $T_{Inc}$  insertions, there exist sketches for cardinality, sum, and Bernstein statistics of size  $k = \text{poly}\left(\frac{1}{\epsilon}, \log\left(\frac{T_{Inc}}{\delta}\right)\right)$  bits. With probability at least  $1 - \delta$ , the sketch maintains an estimate  $\hat{F}^{(t)}$  satisfying  $|\hat{F}^{(t)} - F^{(t)}| \leq \epsilon \cdot \max_{t' \leq t} F^{(t')}$ .

## Overview of our Approach

- We proceed by robustifying the following classical algorithm for distinct elements estimation ( $F_0$ ) [GLH06]:
  - Let  $p \in (0,1)$  be a fixed sampling probability, let  $S = \emptyset$ .
  - For each update in the stream:
    - If **Insert**( $i$ ):
      - $S \leftarrow S \setminus \{i\}$
      - Sample  $b \sim \text{Bernoulli}(p)$ .
      - If  $b = 1$ ,  $S \leftarrow S \cup \{i\}$ .
    - If **Reset**( $i$ ):
      - Set  $S \leftarrow S \setminus \{i\}$ .
  - Return  $\hat{N}_t = \frac{|S|}{p}$
- A standard modification of the above algorithm (by adaptively adjusting the sampling rate) admits a streaming algorithm for resettable  $F_0$  satisfying the *prefix-max* guarantee.
- **Robustification for  $F_0$  (High-Level Idea)**
  - We compose the standard fixed-rate sketch & estimator with the *Binary Tree Mechanism*: the input to the mechanism is a sequence of updates to the *sample-size*, and the output is a sequence of noisy estimates of the prefix-sum ( $\tilde{S}_t$ ). The estimator returns  $\tilde{N}_t = \frac{\tilde{S}_t}{p}$  (protect key membership in sample).
- **Robustification for  $F_1$  (High-Level Idea):**
  - We examine the  $F_1$  (sum) algorithm of [CCD12] and carefully identify the sensitive privacy units. We decompose the estimator into a *protected* (low-sensitivity) component and a *deterministic* component, which can be known to the adversary; we then feed updates to the *protected* part of the estimator to the *Binary Tree Mechanism*.
  - **Technical challenge:** Unlike our robust  $F_0$  algorithm, each private unit contributes to many distinct updates to the private portion of the estimator.

## Conclusions and Future Work

- We further generalize our approach above and design  $\text{poly}(1/\epsilon, \log 1/\delta, \log T)$  space sketches for Bernstein statistics.
- Can we design adaptively robust sketches for super-linear statistics (i.e.  $F_2$ ) using  $\text{poly}\left(\frac{1}{\epsilon}, \log 1/\delta, \log T\right)$  bits of space?
- Can we design adversarially robust sketches under *ReLU* updates?