Differential Privacy in New Settings

Dheeraj Kumar Singh

Abstract

This paper discusses the concepts of pan-privacy and continual observation under the scope of differential privacy and proposes algorithms to realize these objectives. Pan-privacy is focused on providing privacy against external as well as internal agents while continual observation focuses on creating algorithms that can maintain privacy while continuously processing data streams in real time.

Continual Observation, Differential Privacy, Pan-Privacy, Data Privacy

I. Introduction

Differential privacy is a new and novel approach to provide statistical disclosure without undermining the privacy of any of the individual(s) involved. In other words, differential privacy can be summarised as achieving the results of the analysis of a database in such a fashion that the probability distribution of the analysis is the same whether any individual opts in or out of the data set i.e. to make sure that anything that happens is equally probable if the individual joins or does not join the dataset. This paper discusses the concepts of continual observation and pan-privacy in the domain of differential privacy.

Pan-privacy is aimed at providing protection beyond the notion of traditional attacks (external attacks such as a hacker trying to gain access to the data of individuals etc.) i.e. protecting the privacy of individuals against intrusions into the internal state of the system as well (such as subpoena etc). Thus, the objective is to maintain a differentially private internal state so that even in case of an intrusion into the internal state of the algorithm, the privacy of individuals is not compromised. This encourages more people to participate thereby having a positive impact on the output.

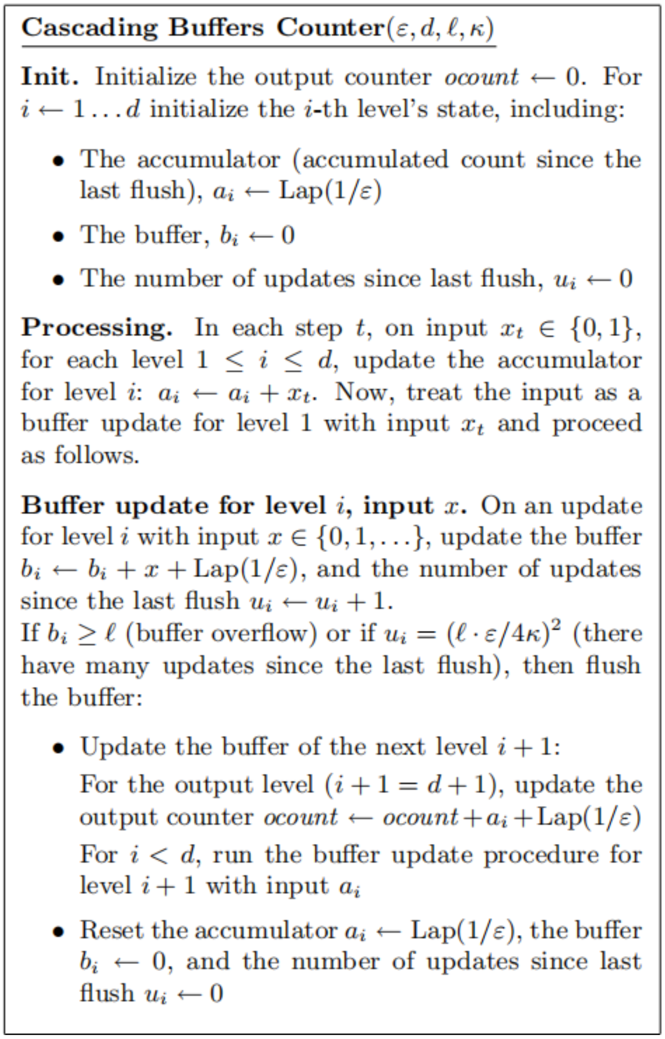
Continual observation aims at providing privacy to individuals while dealing with algorithms that continuously observe and process data streams over time with a reasonable accuracy in the result. Unlike static datasets, data streams pose challenges due to their dynamic nature. In continual observation the algorithm must adapt to changing data distributions, process and provide results in real time all while maintaining privacy of individuals.

In this paper, we present solutions for achieving continual observation and pan-privacy within the context of differential privacy.

II. Your Procedure or Your Method

The paper outlines new settings for achieving differential privacy. Here's an overview of these settings:

1. **The Counter Primitive**:



Cascading Buffers Counter$(\varepsilon, d, \ell, \kappa)$

Init. Initialize the output counter *ocount* $\leftarrow 0$. For $i \leftarrow 1 \dots d$ initialize the i-th level's state, including:

- The accumulator (accumulated count since the last flush), $a_i \leftarrow \mathrm{Lap}(1/\varepsilon)$
- The buffer, $b_i \leftarrow 0$
- The number of updates since last flush, $u_i \leftarrow 0$

Processing. In each step t, on input $x_t \in \{0,1\}$, for each level $1 \leq i \leq d$, update the accumulator for level i: $a_i \leftarrow a_i + x_t$. Now, treat the input as a buffer update for level 1 with input x_t and proceed as follows.

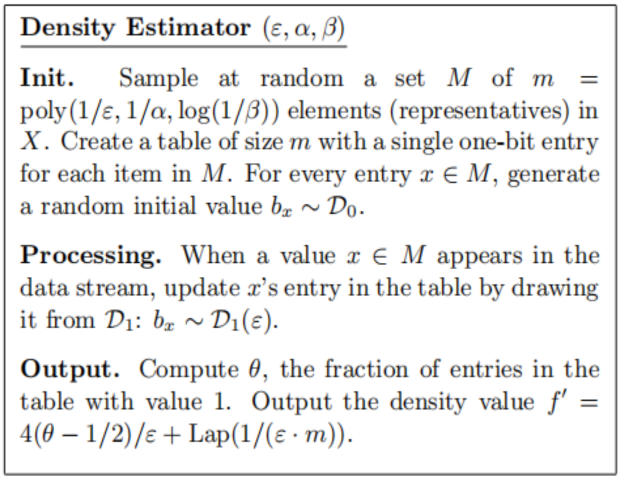
Buffer update for level i, input x. On an update for level i with input $x \in \{0,1,\dots\}$, update the buffer $b_i \leftarrow b_i + x + \mathrm{Lap}(1/\varepsilon)$, and the number of updates since the last flush $u_i \leftarrow u_i + 1$.
If $b_i \geq \ell$ (buffer overflow) or if $u_i = (\ell \cdot \varepsilon / 4\kappa)^2$ (there have many updates since the last flush), then flush the buffer:

- Update the buffer of the next level $i+1$:
  For the output level $(i+1 = d+1)$, update the output counter *ocount* $\leftarrow$ *ocount* $+ a_i + \mathrm{Lap}(1/\varepsilon)$
  For $i < d$, run the buffer update procedure for level $i+1$ with input a_i
- Reset the accumulator $a_i \leftarrow \mathrm{Lap}(1/\varepsilon)$, the buffer $b_i \leftarrow 0$, and the number of updates since last flush $u_i \leftarrow 0$

(This image is taken from C. Dwork's work:Differential Privacy in New Settings[1])

This approach implements a mechanism to count updates in a stream. At the same time it maintains differential privacy by generating Laplace noise and doing buffer updates while flushing them whenever it reaches a threshold.

2. **Density Estimator**:



Density Estimator $(\varepsilon, \alpha, \beta)$

Init. Sample at random a set M of $m = \mathrm{poly}(1/\varepsilon, 1/\alpha, \log(1/\beta))$ elements (representatives) in X. Create a table of size m with a single one-bit entry for each item in M. For every entry $x \in M$, generate a random initial value $b_x \sim \mathcal{D}_0$.

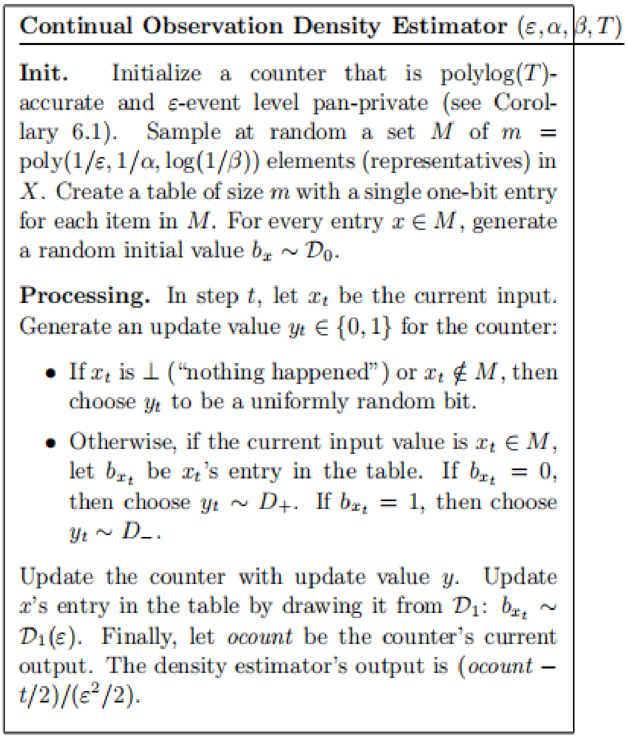
Processing. When a value $x \in M$ appears in the data stream, update x's entry in the table by drawing it from $\mathcal{D}_1$: $b_x \sim \mathcal{D}_1(\varepsilon)$.

Output. Compute θ, the fraction of entries in the table with value 1. Output the density value $f' = 4(\theta - 1/2)/\varepsilon + \mathrm{Lap}(1/(\varepsilon \cdot m))$.

(This image is taken from C. Dwork's work:Differential Privacy in New Settings[1])

It is a mechanism for estimating densities of observed values while preserving privacy. It is a simple approach that is based on updating the estimator with each new observation.

3. **Continuous Observation Density Estimator**:



Continual Observation Density Estimator $(\varepsilon, \alpha, \beta, T)$

Init. Initialize a counter that is $\mathrm{polylog}(T)$-accurate and ε-event level pan-private (see Corollary 6.1). Sample at random a set M of $m = \mathrm{poly}(1/\varepsilon, 1/\alpha, \log(1/\beta))$ elements (representatives) in X. Create a table of size m with a single one-bit entry for each item in M. For every entry $x \in M$, generate a random initial value $b_x \sim \mathcal{D}_0$.

Processing. In step t, let x_t be the current input. Generate an update value $y_t \in \{0, 1\}$ for the counter:

- If x_t is $\perp$ ("nothing happened") or $x_t \notin M$, then choose y_t to be a uniformly random bit.
- Otherwise, if the current input value is $x_t \in M$, let b_{x_t} be x_t's entry in the table. If $b_{x_t} = 0$, then choose $y_t \sim D_+$. If $b_{x_t} = 1$, then choose $y_t \sim D_-$.

Update the counter with update value y. Update x's entry in the table by drawing it from $\mathcal{D}_1$: $b_{x_t} \sim \mathcal{D}_1(\varepsilon)$. Finally, let $ocount$ be the counter's current output. The density estimator's output is $(ocount - t/2)/(\varepsilon^2/2)$.

(This image is taken from C. Dwork's work:Differential Privacy in New Settings[1])

It is in simple terms a complex version of density estimator. The idea is to update the estimator with each new observation hence generating a binary update value based on the input and updating an internal counter accordingly. It also maintains a table to keep track of the observed values and corresponding bits.

The code implementation for these are available and can be found at https://github.com/ardor03/AI-Research-Paper.git .

III. Results

In this paper, we implemented the approaches proposed in the research paper "Differential Privacy in New Settings, Cynthia Dwork" that focused on Continual Observation and Pan-Privacy. The implementation includes the various aspects such as data stream processing, cascading buffers, counters and density estimation techniques mentioned in the paper.

IV. Conclusion

The paper highlights that the approaches mentioned in the research paper "Differential Privacy in New Settings, Cynthia Dwork" are practical and can be applied to real world systems (by taking into consideration scalability issues etc.). These approaches can act as a foundation for further research into the topic.

V. References

[1] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: The SuLQ framework. In *Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, June 2005.

[2] I. Dinur and K. Nissim. Revealing information while preserving privacy. In *Proceedings of the Twenty Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 202–210, 2003.

[3] C. Dwork. A firm foundation for private data analysis. *Communications of the ACM (to appear)*.

[4] C. Dwork. An ad omnia approach to defining and achieving private data analysis. In F. Bonchi, E. Ferrari, B. Malin, and Y. Saygin, editors, *Privacy, Security, and Trust in KDD, First ACM SIGKDD International (PinKDD), Revised Selected Papers*, volume 4890 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 2007.

[5] C. Dwork. The differential privacy frontier. In *Proceedings of the 6th Theory of Cryptography Conference (TCC)*, 2009.

[6] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: privacy via distributed noise generation. In *Advances in Cryptology: Proceedings of EUROCRYPT*, pages 486–503, 2006.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Theory of Cryptography Conference*, pages 265–284, 2006.

[8] C. Dwork, F. McSherry, and K. Talwar. The price of privacy and the limits of lp decoding. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages pp. 85–94, 2007.

[9] C. Dwork, M. Naor, T. Pitassi, and G. Rothblum. Differential privacy under continual observation. Manuscript in preparation, 2009.

[10] C. Dwork, M. Naor, T. Pitassi, G. Rothblum, and S. Yekhanin. Pan-private streaming algorithms. Manuscript submitted for publication, 2009.

[11] C. Dwork and K. Nissim. Privacy-preserving datamining on vertically partitioned databases. In *Proceedings of CRYPTO 2004*, volume 3152, pages 528–544, 2004.

[12] C. Dwork and S. Yekhanin. New efficient attacks on statistical disclosure control mechanisms. In *Proceedings of CRYPTO 2008*, pages 468–480, 2008.

[13] M. Hardt and K. Talwar. On the geometry of differential privacy. arXiv:0907.3754v2, 2009.

[14] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual Symposium on Foundations of Computer Science*, 2007.

[15] S. Warner. Randomized response: a survey technique for eliminating evasive answer bias. *JASA*, pages 63–69, 1965.

[16] Cynthia Dwork, “Differential Privacy in New Settings”.