UC RIVERSIDE

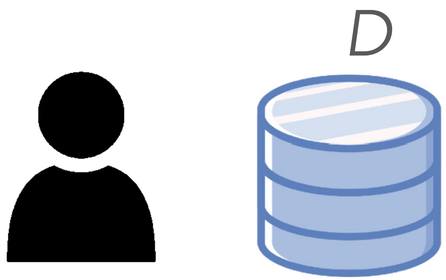
To Shuffle or not to Shuffle: Auditing DP-SGD with Shuffling

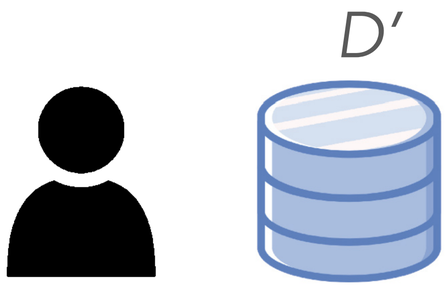
Meenatchi Sundaram Muthu Selva Annamalai, Borja Balle,
Jamie Hayes, and Emiliano De Cristofaro

Differential Privacy

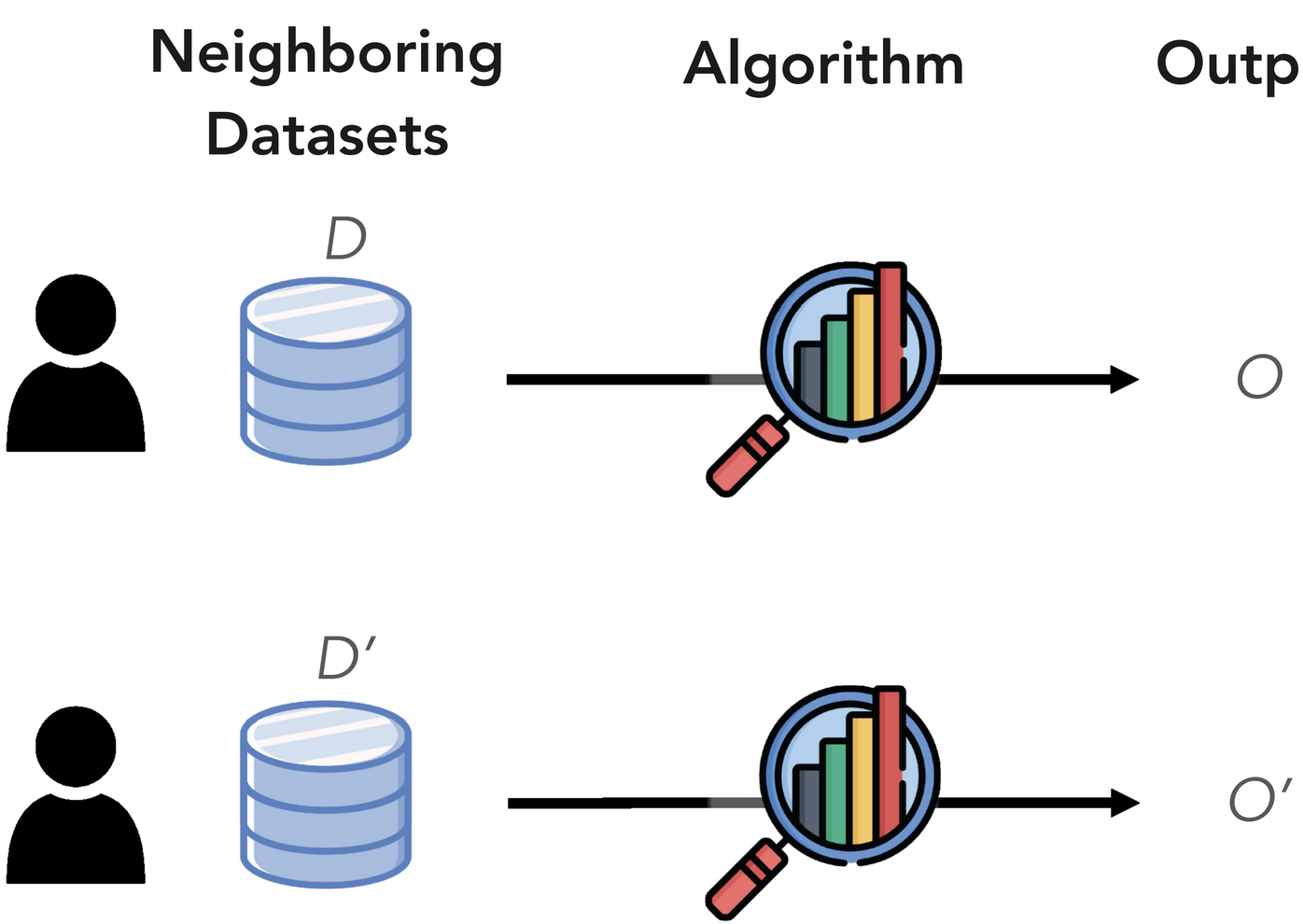
Differential Privacy

Neighboring Datasets



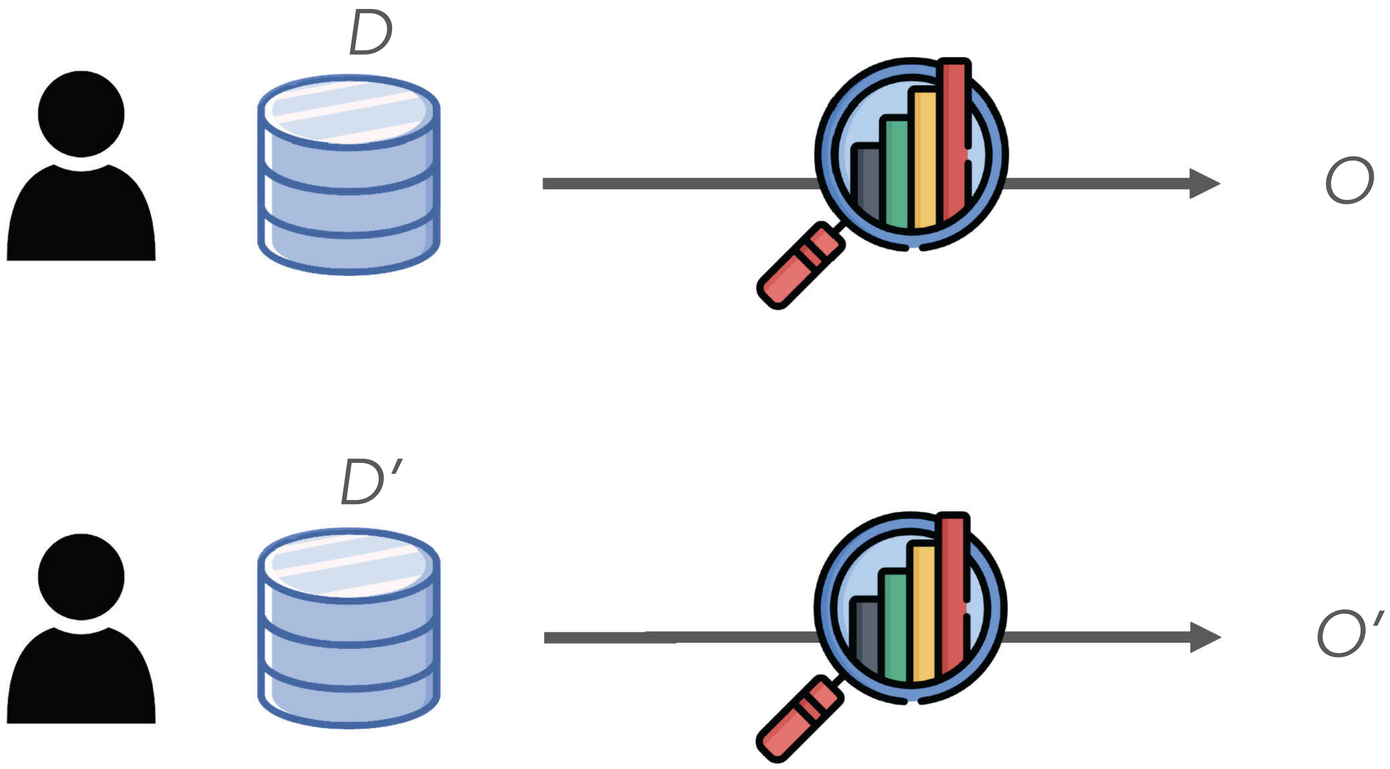


Differential Privacy



Differential Privacy

Neighboring Datasets | **Algorithm** | **Output**



O and O′ are roughly similar
(can't distinguish between them)
for any D, D′

With probability bounded by privacy parameter ε

DP-SGD

1 iteration

Dataset

x_1

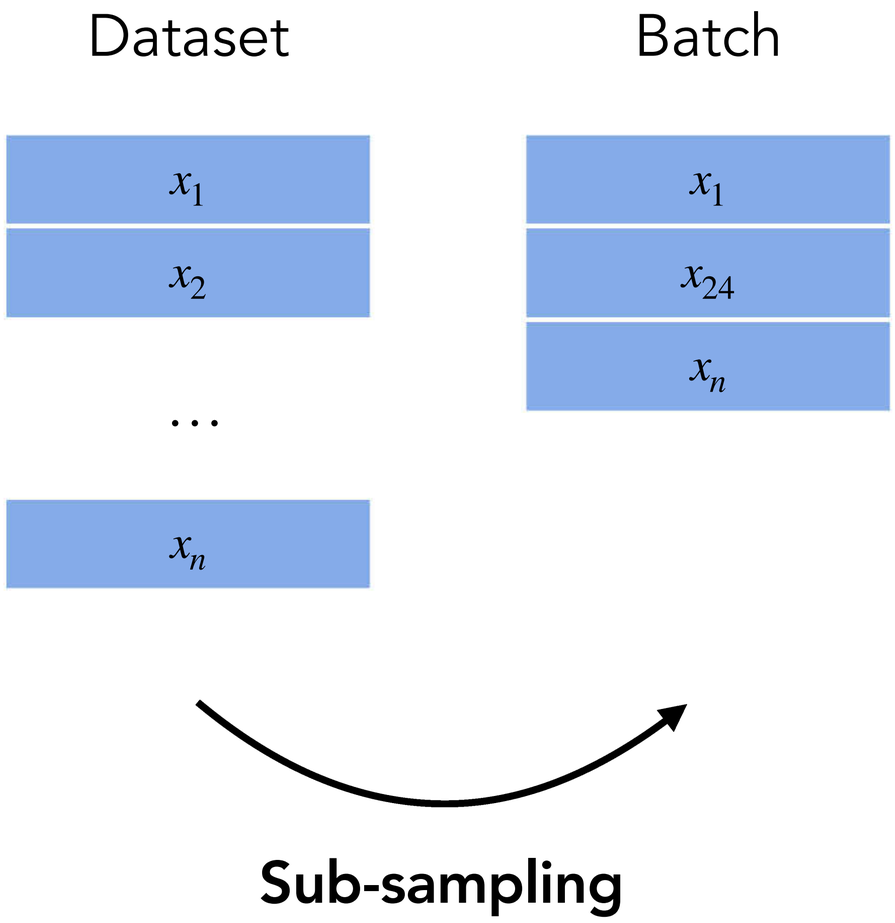
x_2

...

x_n

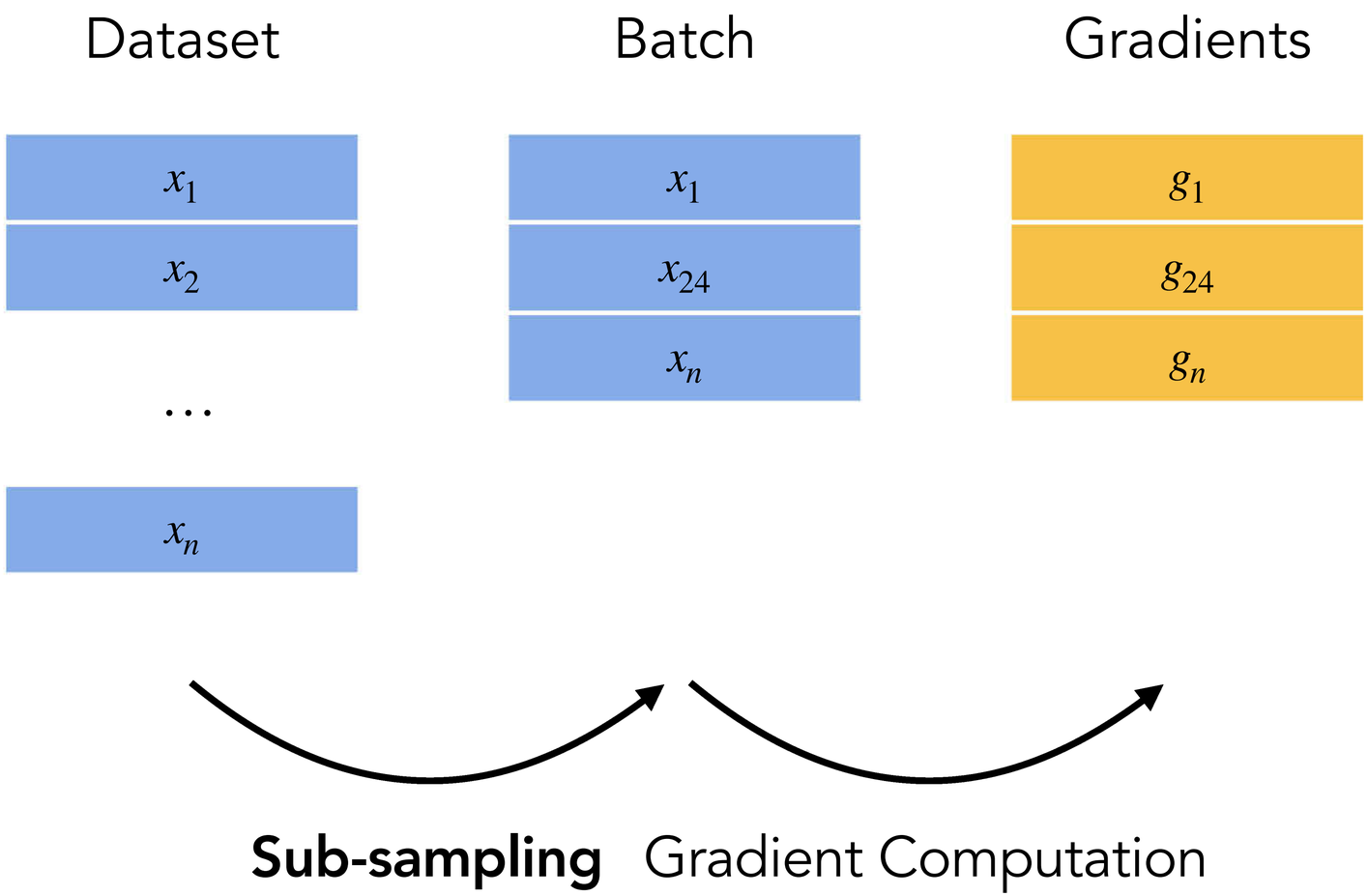
DP-SGD

1 iteration



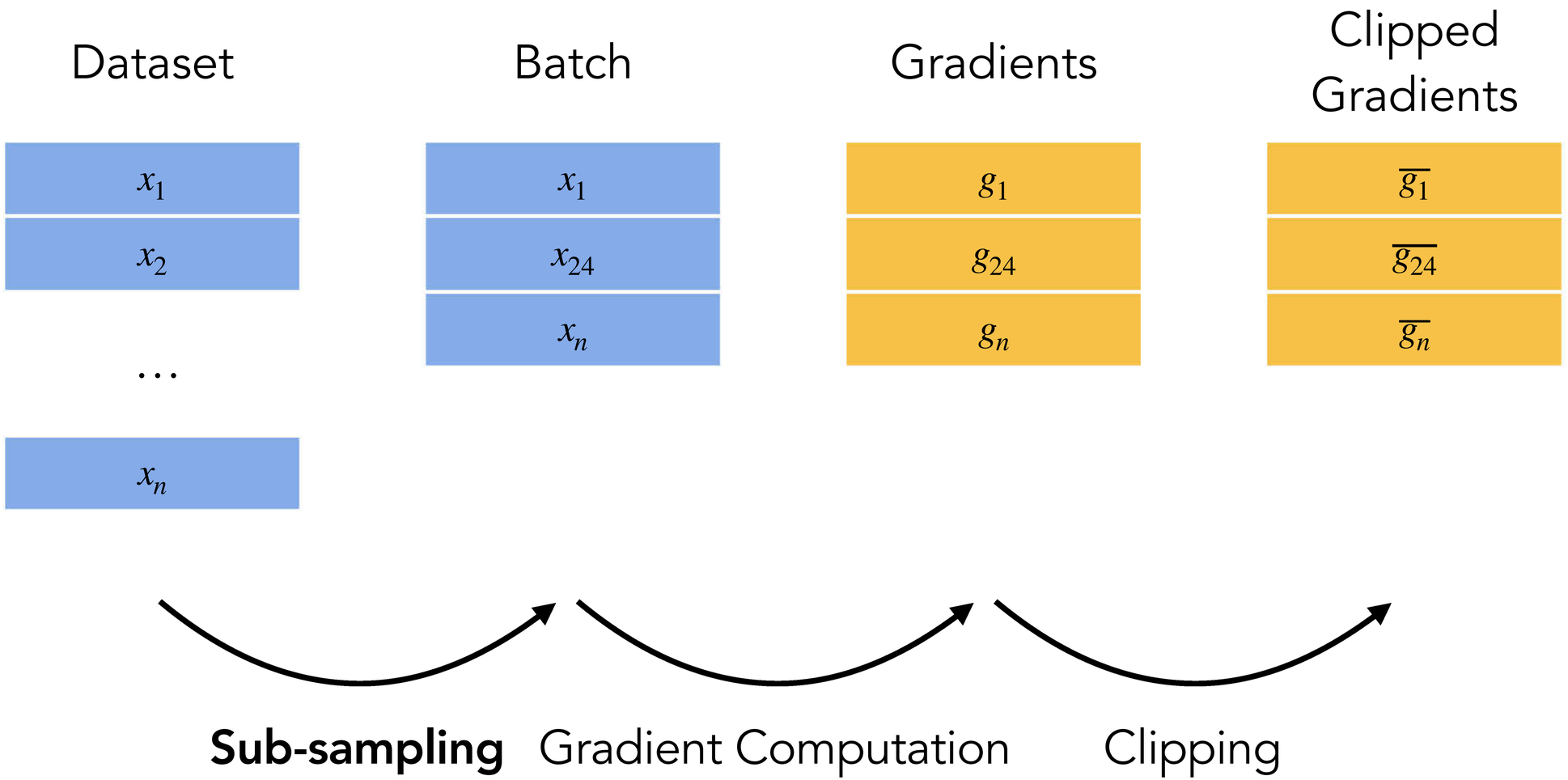
DP-SGD

1 iteration



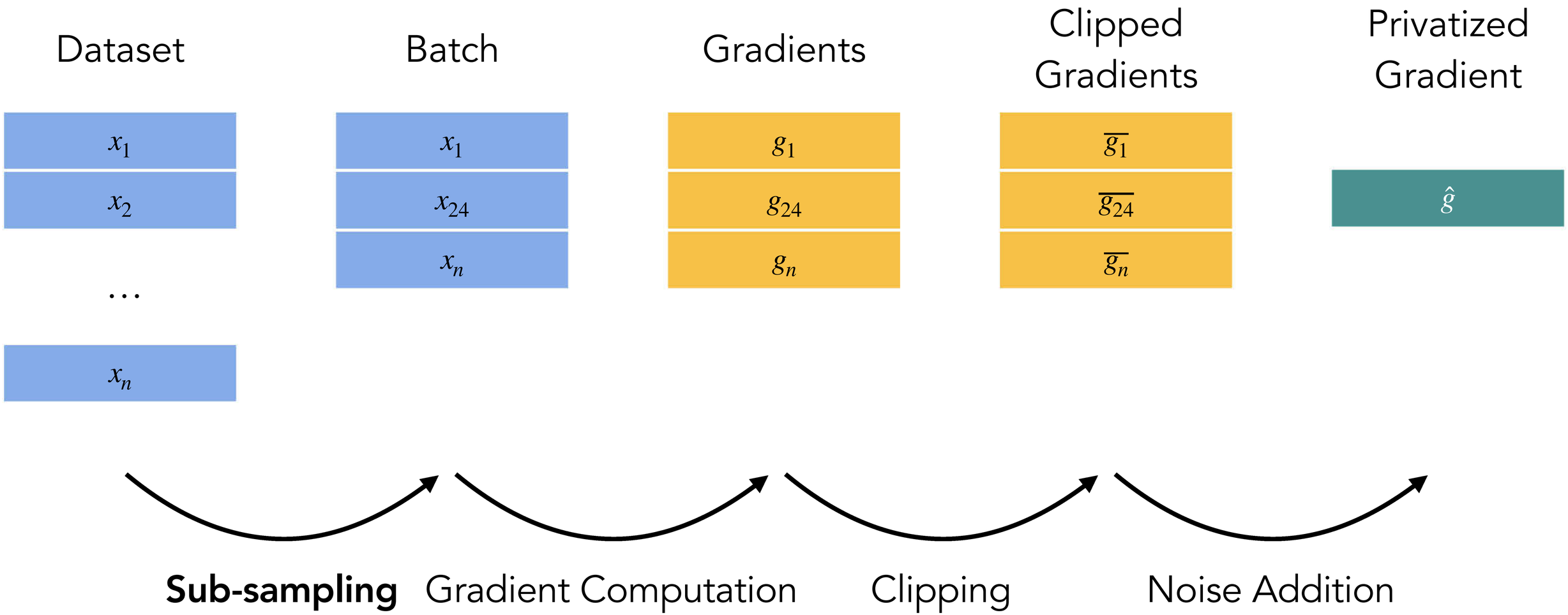
DP-SGD

1 iteration



DP-SGD

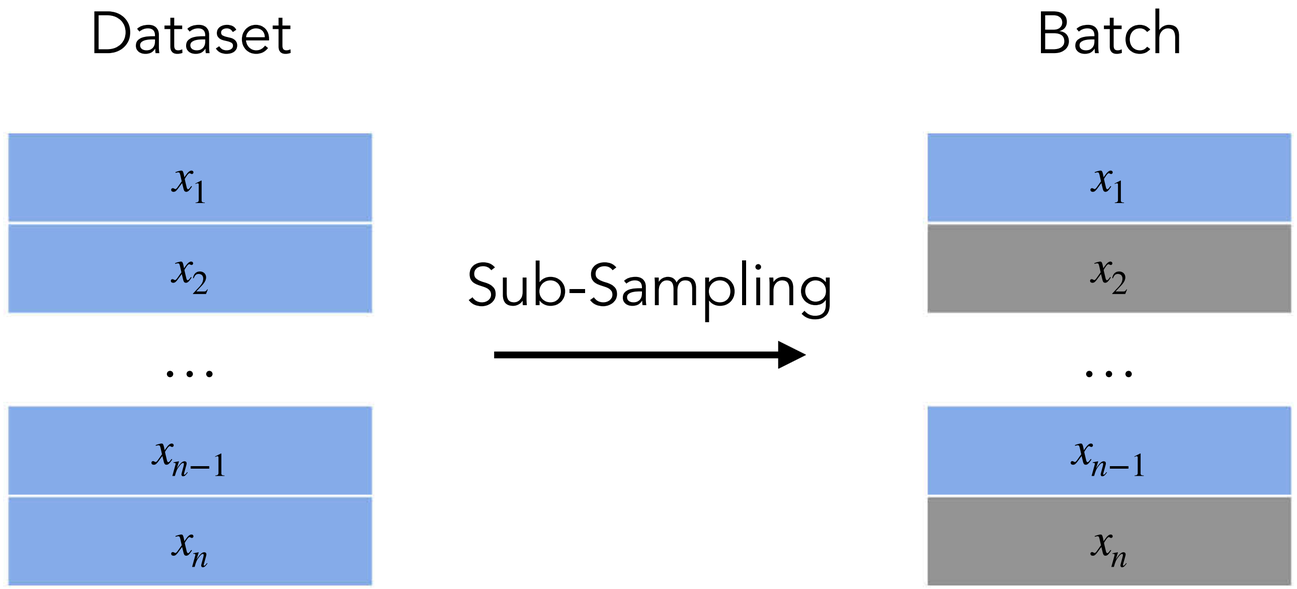
1 iteration



Sub-Sampling

Why Sub-Sample?

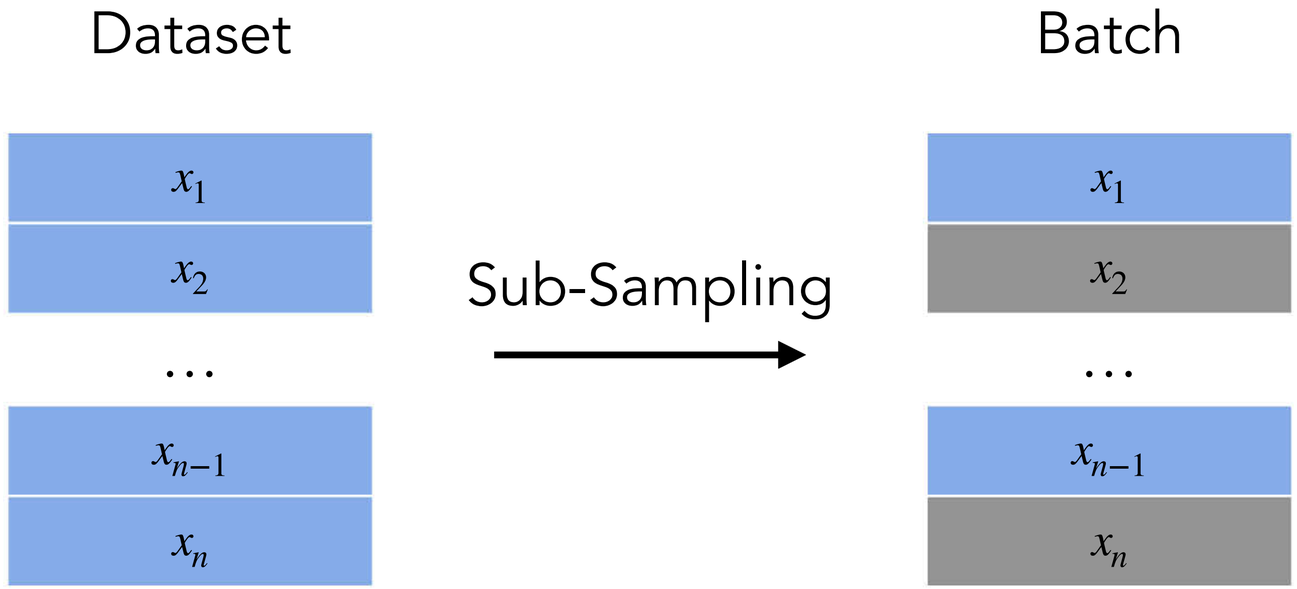
- Sub-sampling: choosing batches through a random process



Sub-Sampling

Why Sub-Sample?

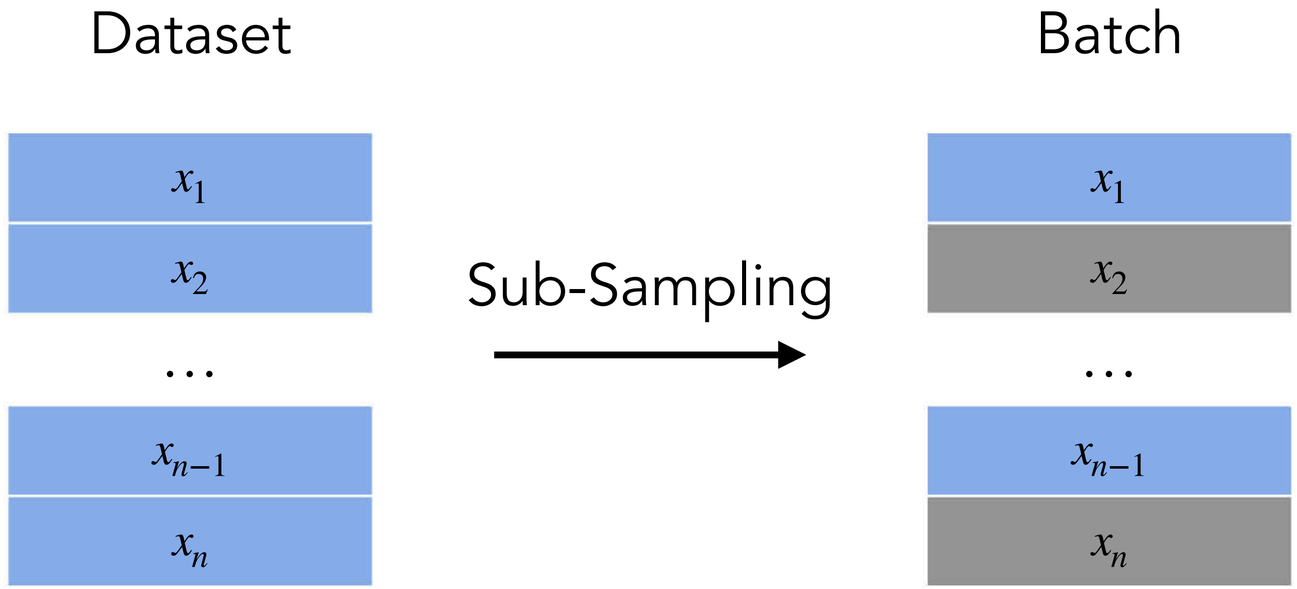
- Sub-sampling: choosing batches through a random process
- SGD naturally processes dataset in batches (for efficiency and utility)



Sub-Sampling

Why Sub-Sample?

- Sub-sampling: choosing batches through a random process
- SGD naturally processes dataset in batches (for efficiency and utility)
- Sub-sampling amplifies DP guarantees substantially[1]



[1] B. Balle, G. Barthe, M. Gaboardi. Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences. NeurIPS, 2018.

Sub-Sampling

Sub-Sampling

Poisson	Shuffling
Sample each record **independently** with prob. q	Shuffle dataset and **sequentially select** B records

Sub-Sampling

	Poisson	Shuffling
	Sample each record **independently** with prob. q	Shuffle dataset and **sequentially select** B records
DP Guarantees	**Known privacy amplification** theorem	**Unknown** guarantees

Sub-Sampling

	Poisson	Shuffling
	Sample each record **independently** with prob. q	Shuffle dataset and **sequentially select** B records
DP Guarantees	**Known privacy amplification** theorem	**Unknown** guarantees
Speed	**Slow**, requires random access to entire dataset	**Fast**, iterate over B records sequentially

Sub-Sampling

	Poisson	Shuffling
	Sample each record **independently** with prob. q	Shuffle dataset and **sequentially select** B records
DP Guarantees	**Known privacy amplification** theorem	**Unknown** guarantees
Speed	**Slow**, requires random access to entire dataset	**Fast**, iterate over B records sequentially
Hardware Optimizations	**None**, due to random batch sizes	**XLA** optimization from fixed-batch size

What's the Problem?

What's the Problem?

- Prior work[1,2] **calibrate noise added to Poisson sub-sampling**, even though they **implement shuffling in practice** (for efficiency)

[1] S. De, L. Berrada, J. Hayes, S. L. Smith, and B. Balle. Unlocking High-Accuracy Differentially Private Image Classification through Scale. arXiv:2204.13650, 2022.
[2] X. Li, F. Tramer, P. Liang, and T. Hashimoto. Large Language Models Can Be Strong Differentially Private Learners. In ICLR, 2022.

What's the Problem?

- Prior work[1,2] **calibrate noise added to Poisson sub-sampling**, even though they **implement shuffling in practice** (for efficiency)
- Related work[3,4] **analyze lower bound** privacy leakage of shuffling **theoretically** using only simplified version of DP-SGD

[1] S. De, L. Berrada, J. Hayes, S. L. Smith, and B. Balle. Unlocking High-Accuracy Differentially Private Image Classification through Scale. arXiv:2204.13650, 2022.
[2] X. Li, F. Tramer, P. Liang, and T. Hashimoto. Large Language Models Can Be Strong Differentially Private Learners. In ICLR, 2022.
[3] L. Chua, B. Ghazi, P. Kamath, R. Kumar, P. Maurangsi, A. Sinha, and C. Zhang. How Private are DP-SGD Implementations? In ICML, 2024.
[4] L. Chua, B. Ghazi, P. Kamath, R. Kumar, P. Maurangsi, A. Sinha, and C. Zhang. Scalable DP-SGD: Shuffling vs Poisson Subsampling. In NeurIPS, 2024.

What's the Problem?

- Prior work[1,2] **calibrate noise added to Poisson sub-sampling**, even though they **implement shuffling in practice** (for efficiency)
- Related work[3,4] **analyze lower bound** privacy leakage of shuffling **theoretically** using only simplified version of DP-SGD
- We don't know how to derive tight theoretical guarantees for shuffling

[1] S. De, L. Berrada, J. Hayes, S. L. Smith, and B. Balle. Unlocking High-Accuracy Differentially Private Image Classification through Scale. arXiv:2204.13650, 2022.
[2] X. Li, F. Tramer, P. Liang, and T. Hashimoto. Large Language Models Can Be Strong Differentially Private Learners. In ICLR, 2022.
[3] L. Chua, B. Ghazi, P. Kamath, R. Kumar, P. Maurangsi, A. Sinha, and C. Zhang. How Private are DP-SGD Implementations? In ICML, 2024.
[4] L. Chua, B. Ghazi, P. Kamath, R. Kumar, P. Maurangsi, A. Sinha, and C. Zhang. Scalable DP-SGD: Shuffling vs Poisson Subsampling. In NeurIPS, 2024.

What's the Problem?

- Prior work[1,2] **calibrate noise added to Poisson sub-sampling**, even though they **implement shuffling in practice** (for efficiency)
- Related work[3,4] **analyze lower bound** privacy leakage of shuffling **theoretically** using only simplified version of DP-SGD
- We don't know how to derive tight theoretical guarantees for shuffling

Actual privacy guarantees of SOTA private models are unknown!

[1] S. De, L. Berrada, J. Hayes, S. L. Smith, and B. Balle. Unlocking High-Accuracy Differentially Private Image Classification through Scale. arXiv:2204.13650, 2022.
[2] X. Li, F. Tramer, P. Liang, and T. Hashimoto. Large Language Models Can Be Strong Differentially Private Learners. In ICLR, 2022.
[3] L. Chua, B. Ghazi, P. Kamath, R. Kumar, P. Maurangsi, A. Sinha, and C. Zhang. How Private are DP-SGD Implementations? In ICML, 2024.
[4] L. Chua, B. Ghazi, P. Kamath, R. Kumar, P. Maurangsi, A. Sinha, and C. Zhang. Scalable DP-SGD: Shuffling vs Poisson Subsampling. In NeurIPS, 2024.

DP Auditing

DP Auditing

Choose neighboring datasets

D

$D' = D +$ **Alice**

DP Auditing

Choose neighboring datasets

Run DP-SGD (Shuffle) repeatedly

D

...

$D' = D +$ **Alice**

...

DP Auditing

Choose neighboring datasets

Run DP-SGD (Shuffle) repeatedly

Run Membership Inference Attack (MIA)

D

$D' = D +$ **Alice**

DP Auditing

Choose neighboring datasets

Run DP-SGD (Shuffle) repeatedly

Run Membership Inference Attack (MIA)

D

$D' = D +$ **Alice**

Convert FPR and FNR to empirical ε_{emp}

Methodology

Methodology

- Follow DP auditing framework of Nasr et al.[1] **adapted to shuffling**

[1]M. Nasr, J. Hayes, T. Steinke, B. Balle, F. Tramer, M. Jagielski, N. Carlini, and A. Terzis. Tight Auditing of Differentially Private Machine Learning. In USENIX Security, 2023

Methodology

- Follow DP auditing framework of Nasr et al.[1] **adapted to shuffling**
  - **Use likelihood ratio for MIA**: Optimized for shuffling

[1]M. Nasr, J. Hayes, T. Steinke, B. Balle, F. Tramer, M. Jagielski, N. Carlini, and A. Terzis. Tight Auditing of Differentially Private Machine Learning. In USENIX Security, 2023

Methodology

- Follow DP auditing framework of Nasr et al.[1] **adapted to shuffling**
  - **Use likelihood ratio for MIA**: Optimized for shuffling
  - **Increasing adversarial power**: Inject gradient of target canary, final record in each batch, or all records

[1]M. Nasr, J. Hayes, T. Steinke, B. Balle, F. Tramer, M. Jagielski, N. Carlini, and A. Terzis. Tight Auditing of Differentially Private Machine Learning. In USENIX Security, 2023

Methodology

- Follow DP auditing framework of Nasr et al.[1] **adapted to shuffling**
  - **Use likelihood ratio for MIA**: Optimized for shuffling
  - **Increasing adversarial power**: Inject gradient of target canary, final record in each batch, or all records
  - (DP guarantees should hold against ALL , **even pathologically strong** adversaries)

[1]M. Nasr, J. Hayes, T. Steinke, B. Balle, F. Tramer, M. Jagielski, N. Carlini, and A. Terzis. Tight Auditing of Differentially Private Machine Learning. In USENIX Security, 2023

Methodology

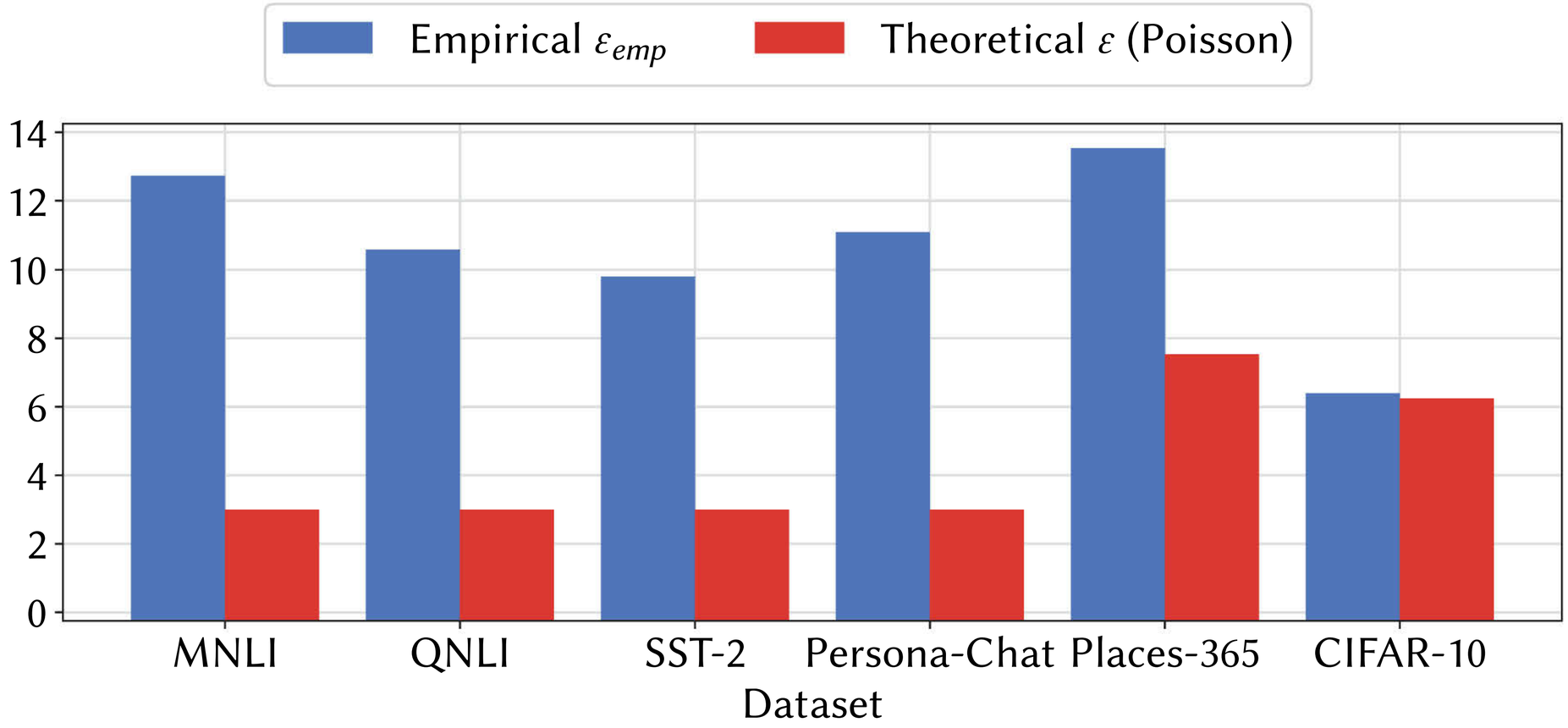
- Follow DP auditing framework of Nasr et al.[1] **adapted to shuffling**
  - **Use likelihood ratio for MIA**: Optimized for shuffling
  - **Increasing adversarial power**: Inject gradient of target canary, final record in each batch, or all records
  - (DP guarantees should hold against ALL , **even pathologically strong** adversaries)
- Compare empirical ε_{emp} obtained from auditing DP-SGD (Shuffle) and compare to theoretical ε intended by DP-SGD (Poisson)

[1]M. Nasr, J. Hayes, T. Steinke, B. Balle, F. Tramer, M. Jagielski, N. Carlini, and A. Terzis. Tight Auditing of Differentially Private Machine Learning. In USENIX Security, 2023

Experiments

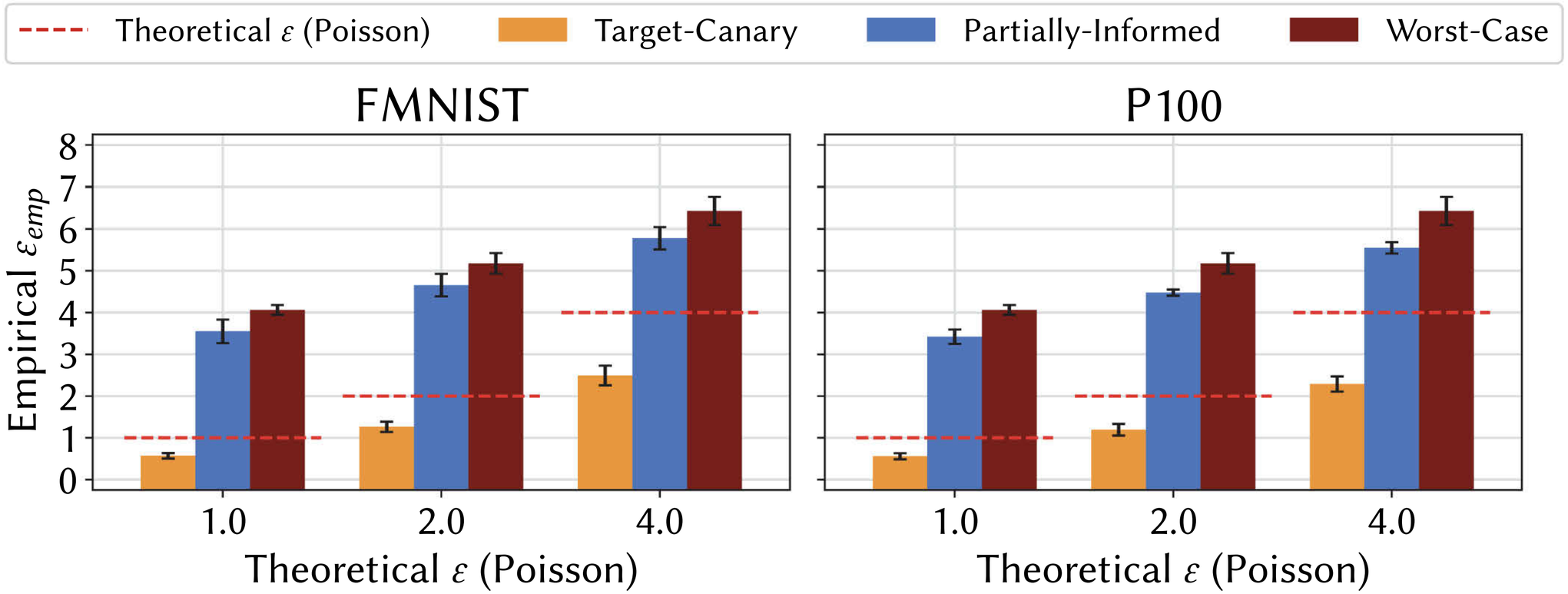
- **Datasets:** FMNIST, Purchase-100, CIFAR-10
- **Models:** LeNet, MLP, CNN
- Train 1M models for auditing

SOTA models* using Shuffling



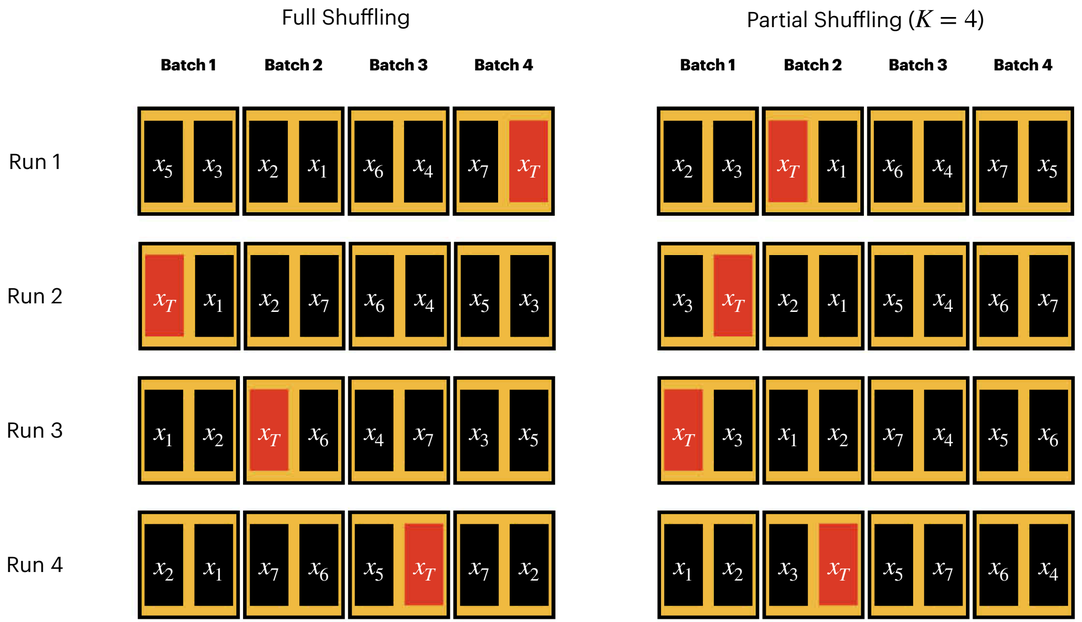
*training on datasets is simulated under Worst-Case threat model

Impact of Adversarial Power

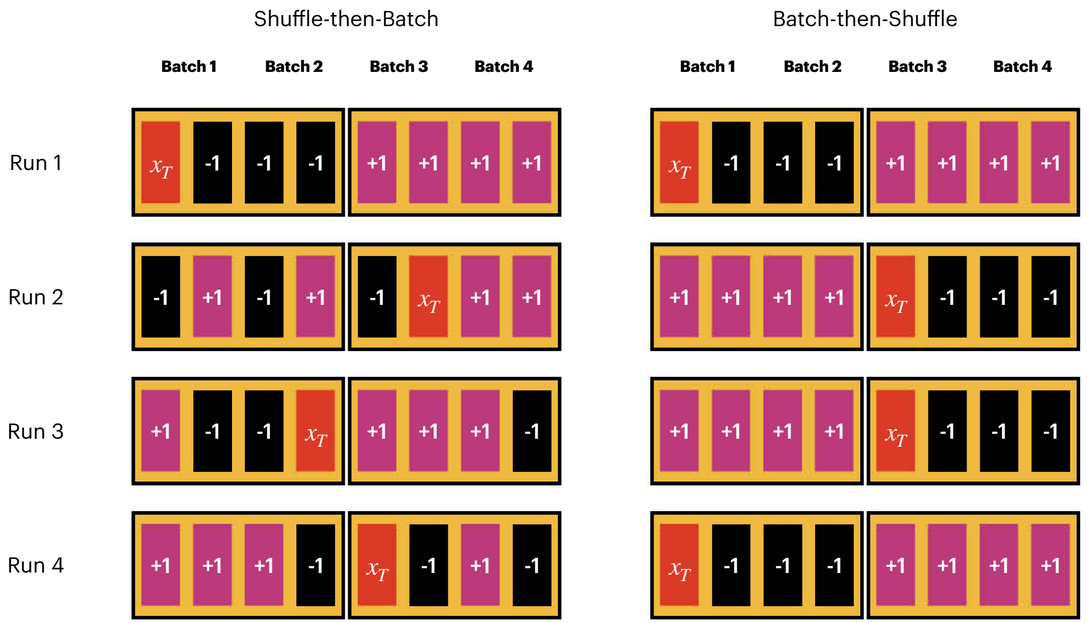


Bugs in Shuffling

Partial Shuffling

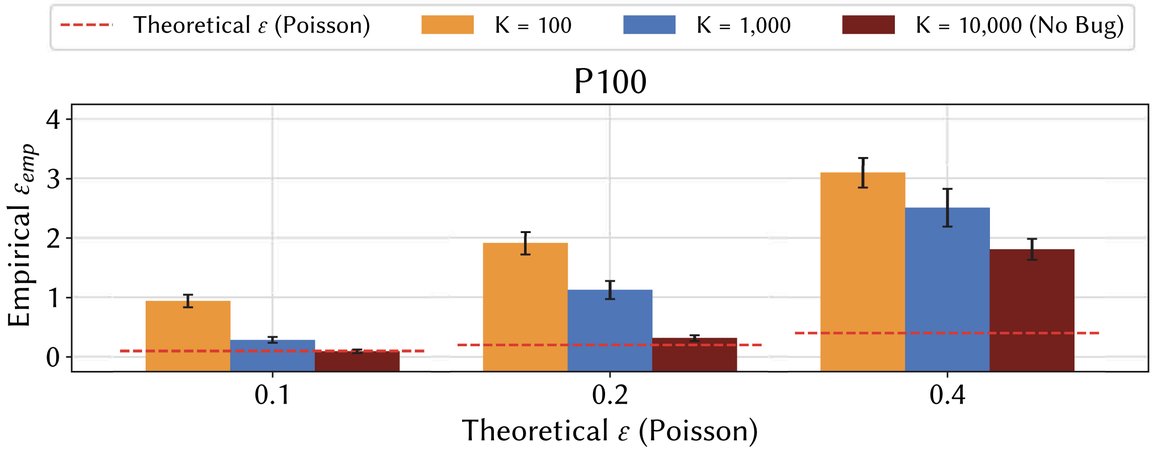


Batch-then-Shuffle

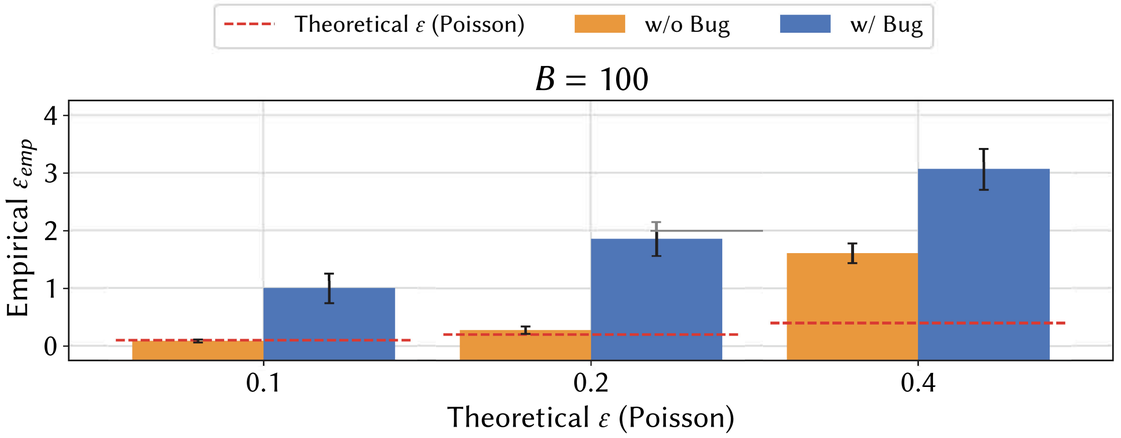


Bugs in Shuffling

Partial Shuffling

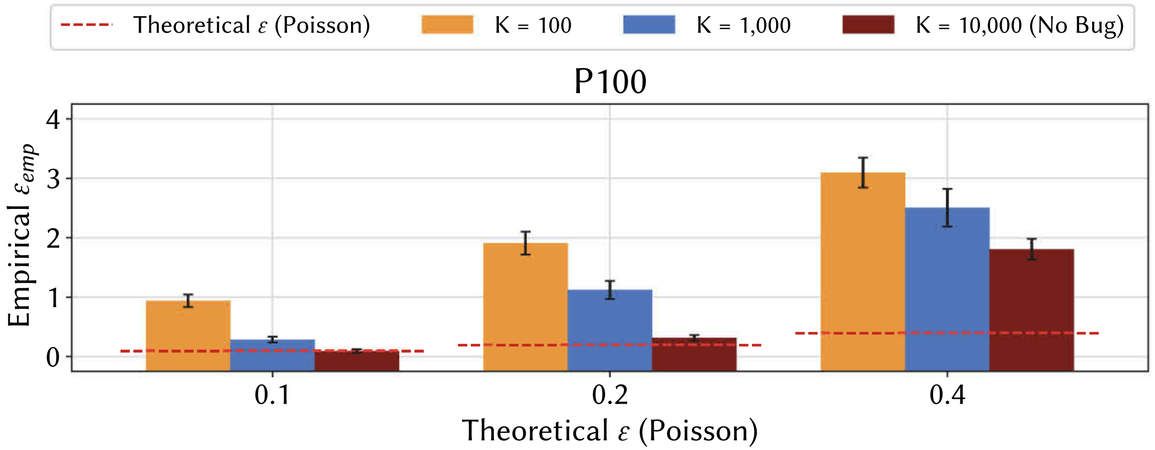


Batch-then-Shuffle

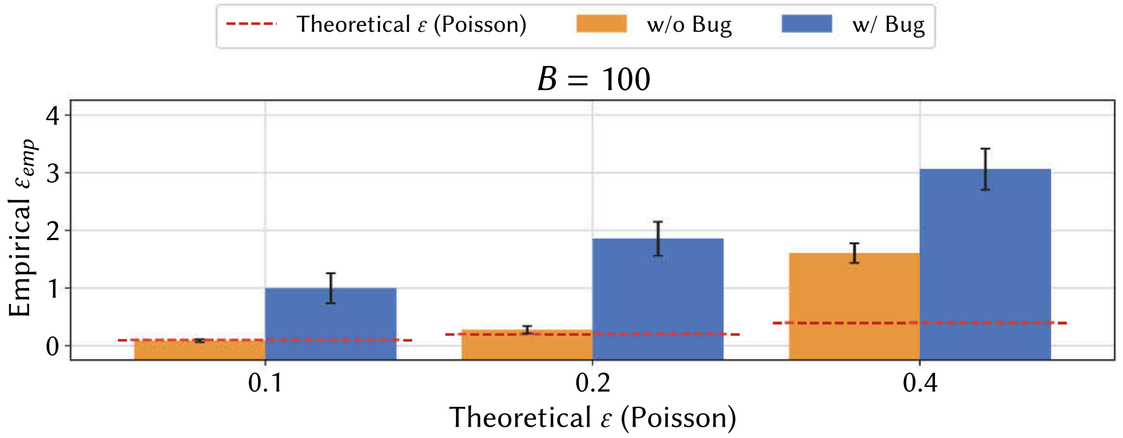


Bugs in Shuffling

Partial Shuffling



Batch-then-Shuffle



Bugs in shuffling can exacerbate gap between theoretical Poisson analysis and actual privacy leakage

Conclusion

Conclusion

- **First audit** of DP-SGD (Shuffle)

Conclusion

- **First audit** of DP-SGD (Shuffle)
- **Novel audit procedure** for shuffling

Conclusion

- **First audit** of DP-SGD (Shuffle)
- **Novel audit procedure** for shuffling
- Theoretical guarantees of SOTA models heavily overestimated!

Conclusion

- **First audit** of DP-SGD (Shuffle)
- **Novel audit procedure** for shuffling
- Theoretical guarantees of SOTA models heavily overestimated!

- Recommendations

Conclusion

- **First audit** of DP-SGD (Shuffle)
- **Novel audit procedure** for shuffling
- Theoretical guarantees of SOTA models heavily overestimated!

- Recommendations
  - **Avoid shuffling** in real-world deployments of DP-SGD

Conclusion

- **First audit** of DP-SGD (Shuffle)
- **Novel audit procedure** for shuffling
- Theoretical guarantees of SOTA models heavily overestimated!

- Recommendations
  - **Avoid shuffling** in real-world deployments of DP-SGD
  - Use alternative sampling schemes (e.g., **Balls-in-Bins**[1,2])

Conclusion

- **First audit** of DP-SGD (Shuffle)
- **Novel audit procedure** for shuffling
- Theoretical guarantees of SOTA models heavily overestimated!

- Recommendations
  - **Avoid shuffling** in real-world deployments of DP-SGD
  - Use alternative sampling schemes (e.g., **Balls-in-Bins**[1,2])

[1] C. A. Choquette-Choo, A. Ganesh, S. Haque, T. Steinke, and A. G. Thakurta. Near-Exact Privacy Amplification for Matrix Mechanisms. In ICLR, 2025.
[2] L. Chua, B. Ghazi, C. Harrison, E. Leeman, P. Kamath, R. Kumar, P. Manurangsi, A. Sinha, and C. Zhang. Balls-and-bins sampling for DP-SGD. In AISTATS, 2025

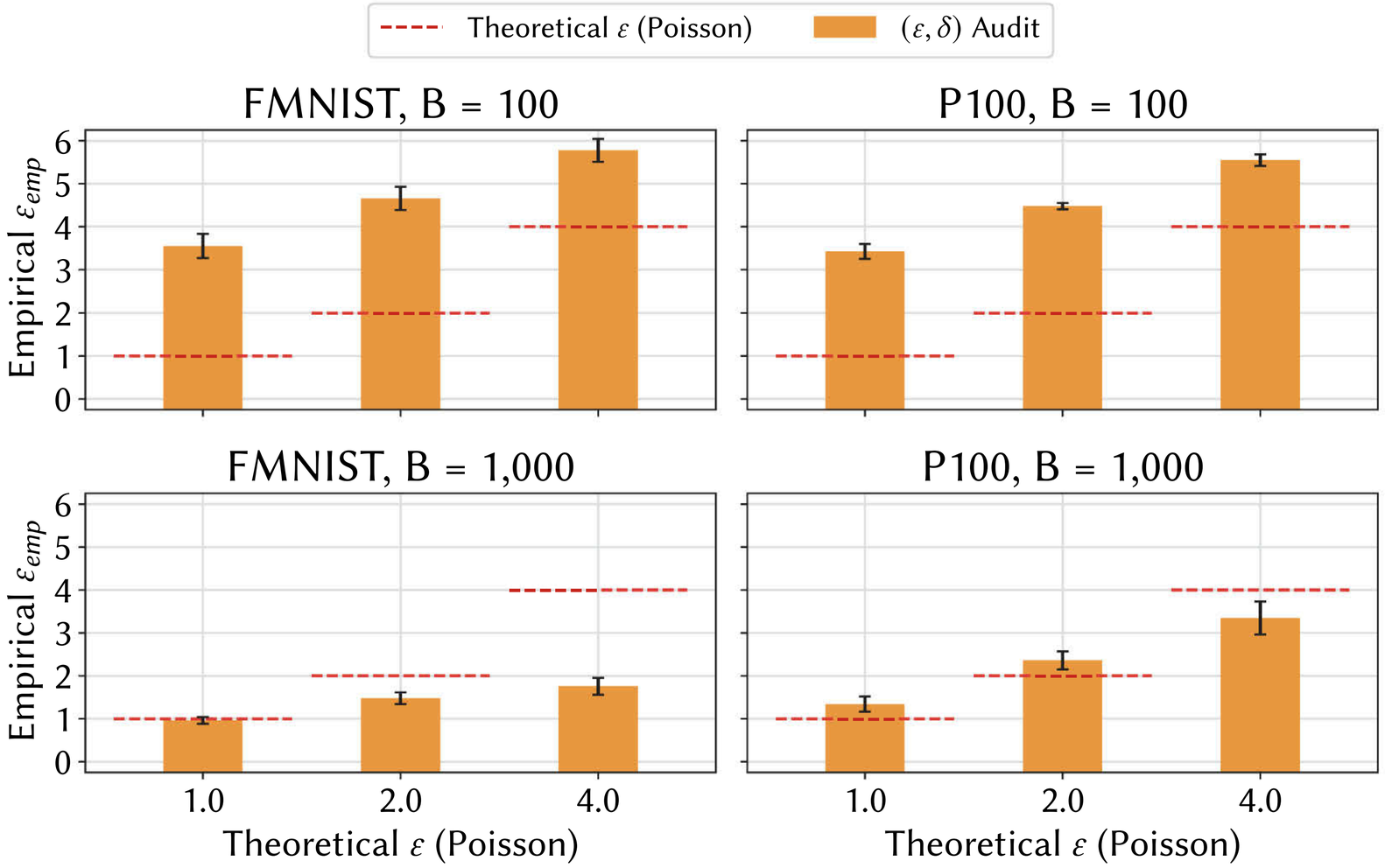
Conclusion

- **First audit** of DP-SGD (Shuffle)
- **Novel audit procedure** for shuffling
- Theoretical guarantees of SOTA models heavily overestimated!

- Recommendations
  - **Avoid shuffling** in real-world deployments of DP-SGD
  - Use alternative sampling schemes (e.g., **Balls-in-Bins**[1,2])

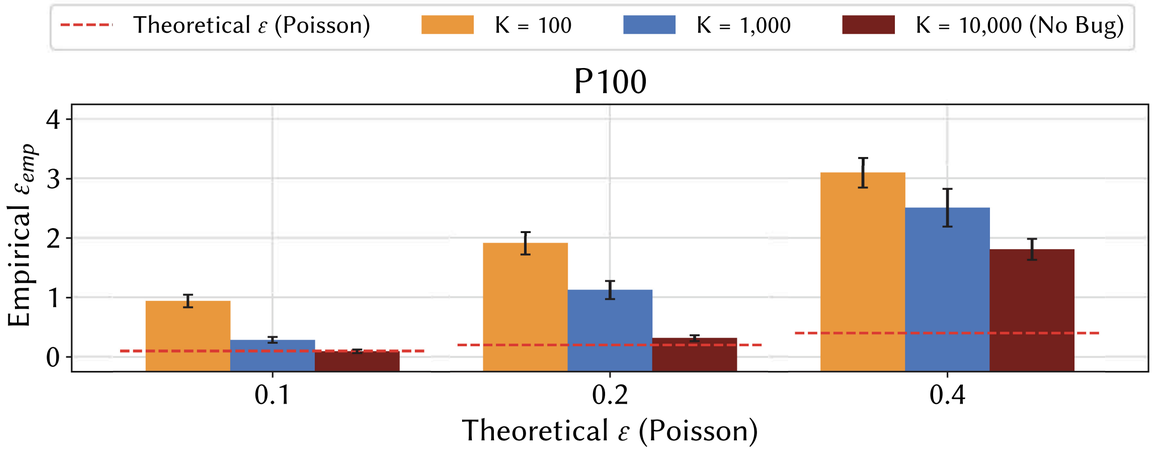
[1] C. A. Choquette-Choo, A. Ganesh, S. Haque, T. Steinke, and A. G. Thakurta. Near-Exact Privacy Amplification for Matrix Mechanisms. In ICLR, 2025.
[2] L. Chua, B. Ghazi, C. Harrison, E. Leeman, P. Kamath, R. Kumar, P. Manurangsi, A. Sinha, and C. Zhang. Balls-and-bins sampling for DP-SGD. In AISTATS, 2025

Impact of Batch Size

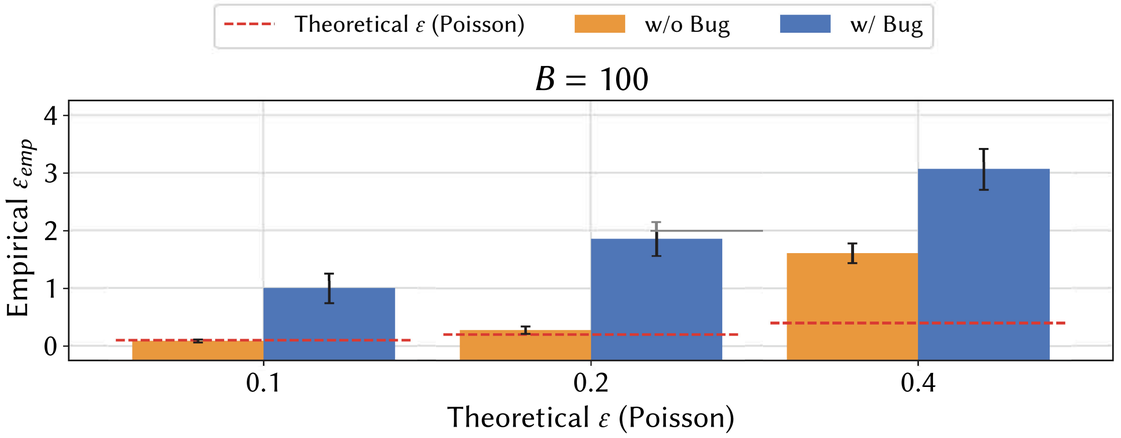


Bugs in Shuffling

Partial Shuffling

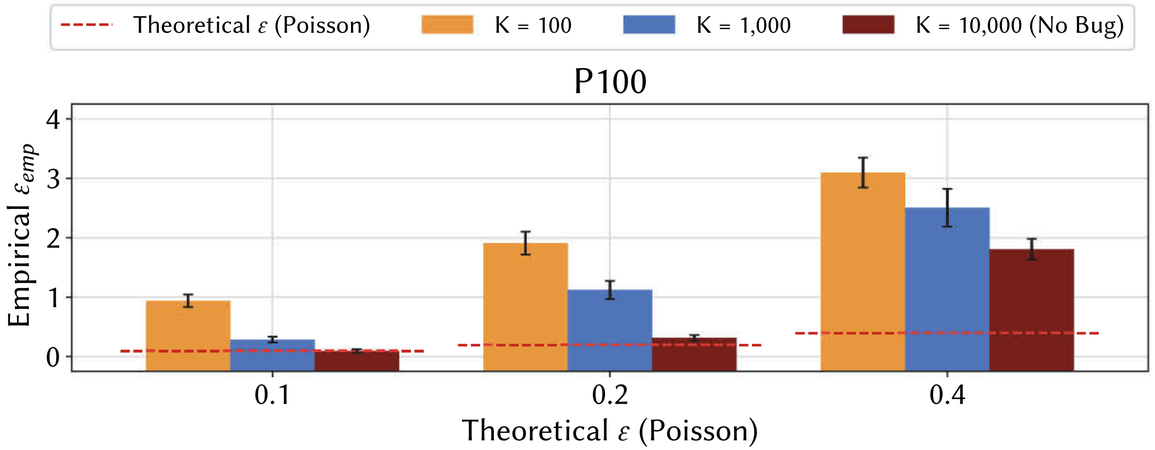


Batch-then-Shuffle

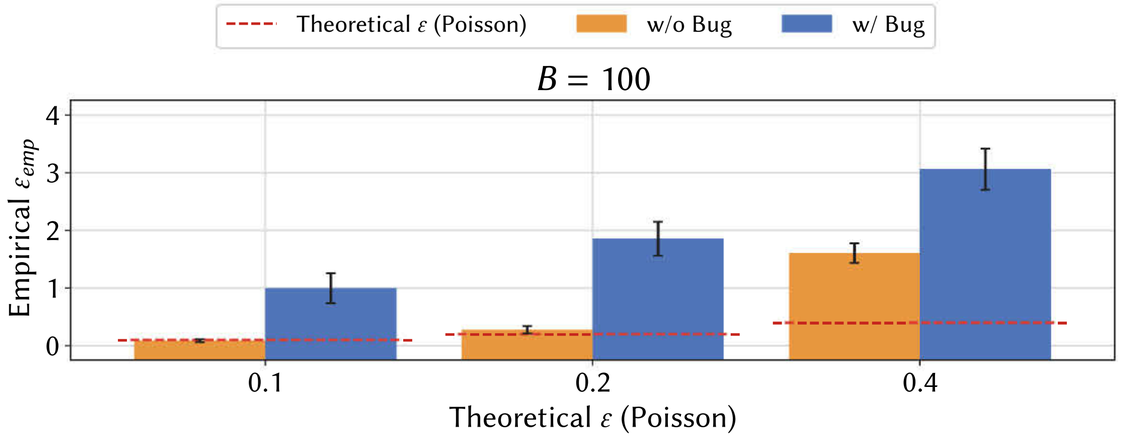


Bugs in Shuffling

Partial Shuffling



Batch-then-Shuffle



Bugs in shuffling can exacerbate gap between theoretical Poisson analysis and actual privacy leakage