Curs 11
Differential Privacy (DP)

Course schedule

1. Why?
2. Cauzalitate
3. Măsurare
4. Modelare și eșantionare
5. Tehnici de analiză
    - Analiza factorială
    - Analiza cluster
    - Analiza de regresie
    - Analiza de rețea
    - Serii de timp
6. Predicție
7. Programare și ML
8. Why Privacy?
9. Privacy Enhancing Techniques
10. Homomorphic Encryption. PIR
11. Differential Privacy
12. Membership Inference Attacks
13. Federated Architecture. Multi-party computation
14. Explainable AI
15. Zero knowledge proof. Blockchain architecture

Contents

Contents

Context

Context

- Companies have to respect GDPR / CCPA / LGDP and also analyze data
- Data needs protection against attackers and security breaches
- Sometimes data has to be publicly shared between businesses for collaboration
- How can these be done?
  - **Differential Privacy**

Context (2)

- https://www.nist.gov/video/what-differential-privacy

What is Differential Privacy?

Differential Privacy (DP)

- Mathematical framework to ensure the privacy of individuals' data
- Goal:
  - To assure an individual that his data remain confidential even if it is used in a statistical study
- Offers privacy guarantees regardless of what an adversary knows
- It is future-proof
- Used on statistical analysis tasks and lately in ML models

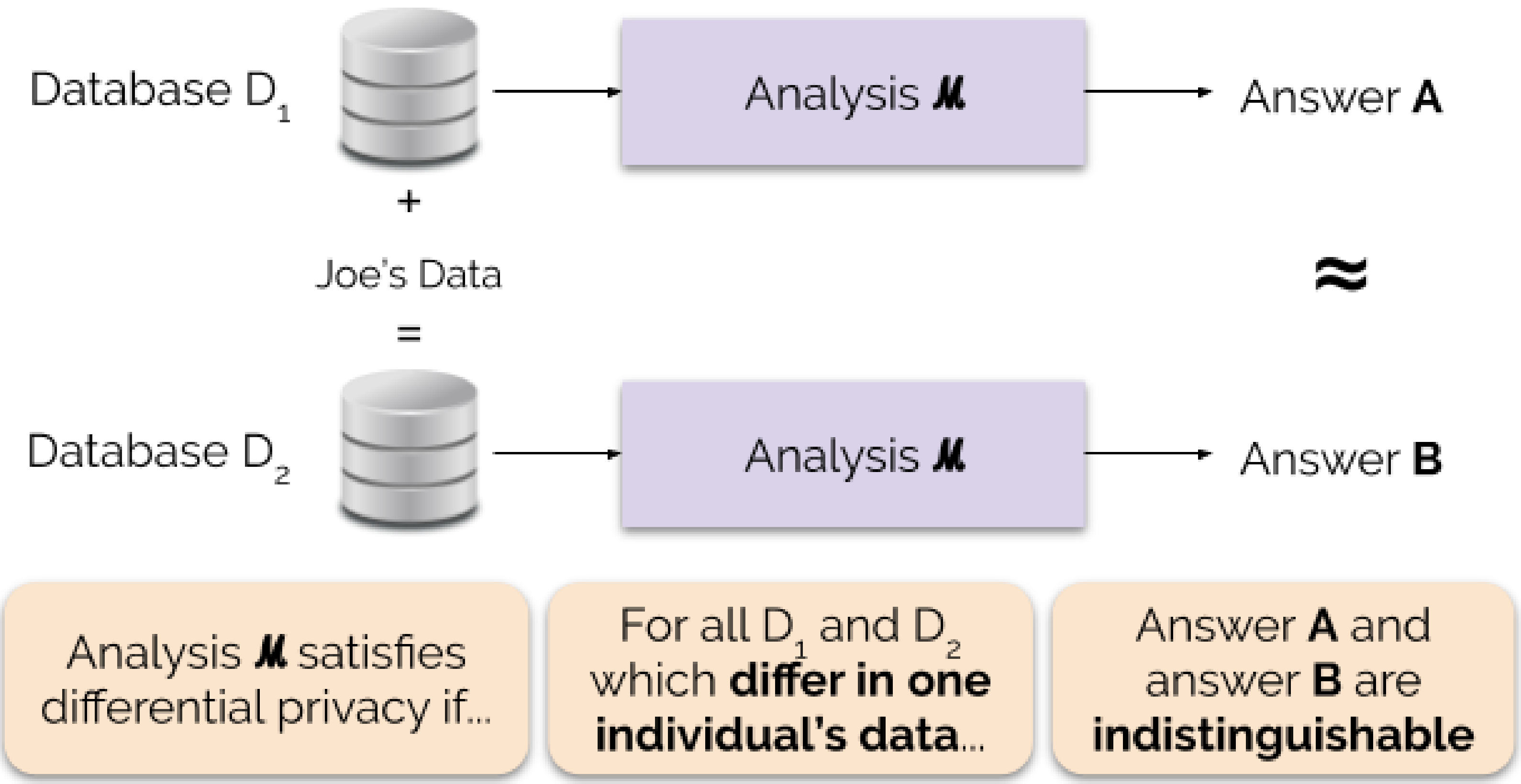
Differential Privacy (DP) (2)

- Offers protection against privacy attacks such as:
  - Background Knowledge Attack
  - Reidentification Attack
  - Membership Inference Attack

- Based upon the idea that the risk of a person' data to be exposed increases every time its data is processed

- Used by big tech companies like Google, Uber, Apple to publicly release studies about sensitive datasets

DP Principle



Source: https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our

Formal Definition of DP

Probability of seeing output O on input D_1

Probability of seeing output O on input D_2

$$\frac{\mathbf{Pr}[\mathcal{M}(D_1) \in O]}{\mathbf{Pr}[\mathcal{M}(D_2) \in O]} \leq e^{\varepsilon}$$

Indistinguishability: bounded ratio of probabilities

Source: https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our

What is this M?

- Any computation that can be done on data
- In DP systems M is a randomized mechanism:
  - The output of M changes probabilistically based upon its input data

How does DP work?

How does DP work?

- Addition of calibrated noise to the output of a statistical query:
  - Done to mask the contribution of an individual data to the final output
  - Chosen in order to preserve a similar accuracy for the analysis

- Multiple mechanisms to add noise

DP mechanisms

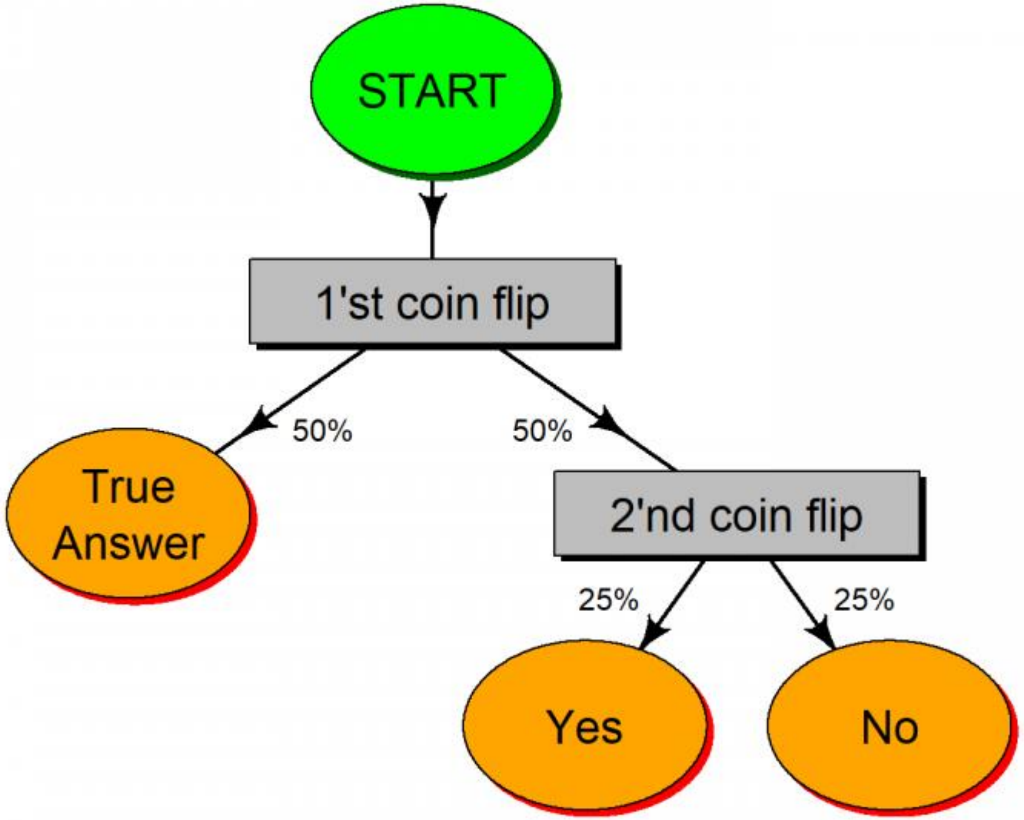
DP mechanisms

- Two main types of techniques used in DP systems to add noise:
  - Laplace mechanism
  - Randomized response and perturbations

Primordial example of DP mechanism

- To protect the privacy of individuals' data
  - Ask each individual a "yes" or "no" question
  - Based upon their response flip a coin
  - If it is head → Add the true answer to the database
  - If it is tails → Flip again the coin
    - If it is head → Add "yes" as the answer to the database
    - If it is tails → Add "no" as the answer to the database

Primordial example of DP mechanism (2)

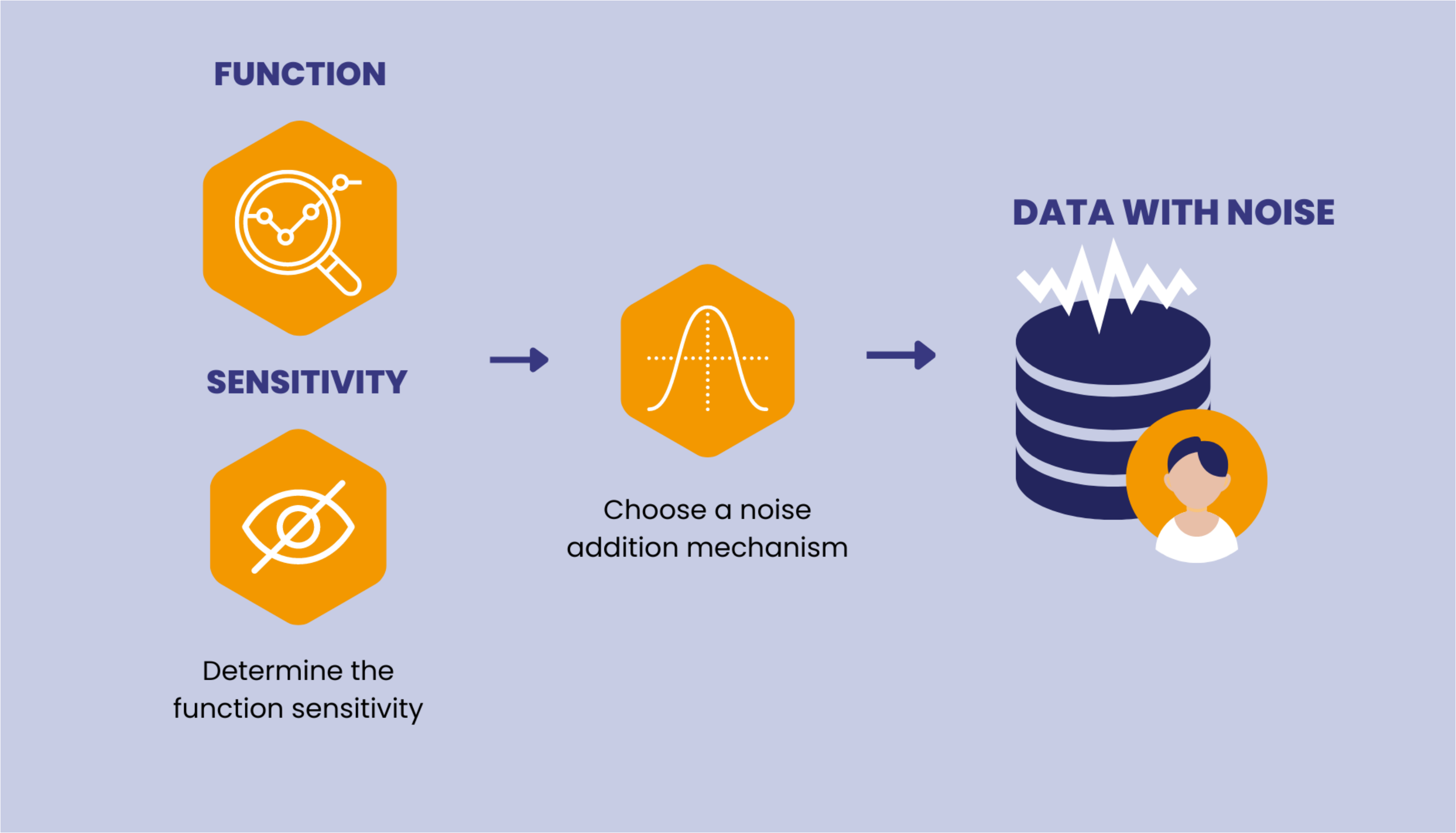


Source: https://www.r-bloggers.com/2020/07/local-differential-privacy-getting-honest-answers-on-embarrassing-questions/

The Laplace mechanism

- Noise is added to the output of a function
- The noise added depends upon the sensitivity of the function
- A function with more sensitivity → More noise to add

The Laplace mechanism (2)



Source: https://www.statice.ai/post/what-is-differential-privacy-definition-mechanisms-examples

The Laplace Mechanism – formal definition

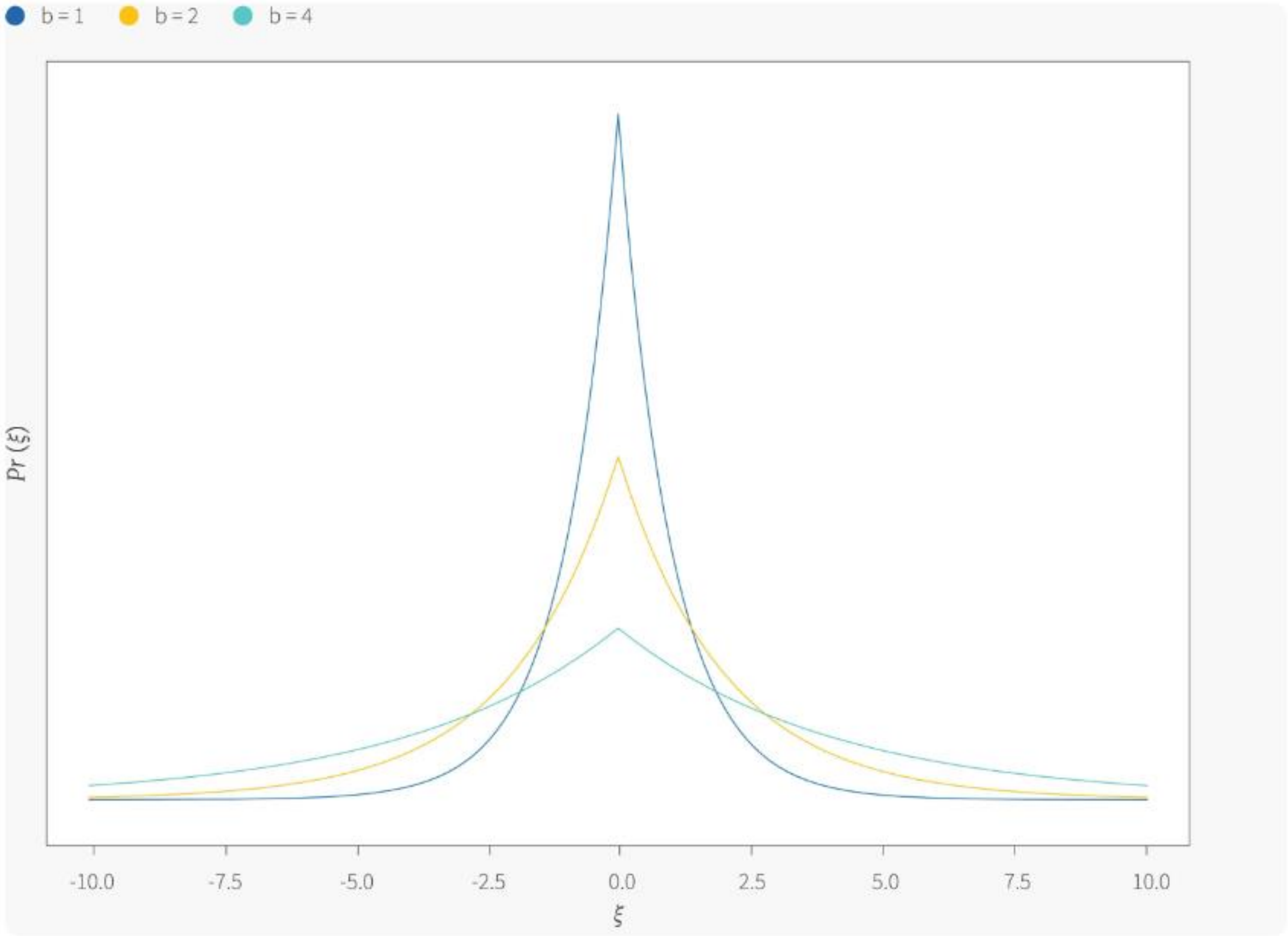
Let $\mathrm{f}[\bullet]$ be a deterministic function of a database $\mathcal{D}$ which returns a scalar value. For instance, it might count the number of entries that satisfy a condition. The Laplace mechanism works by adding noise to $\mathrm{f}[\bullet]$:

$$\mathrm{M}[\mathcal{D}] = \mathrm{f}[\mathcal{D}] + \xi, \tag{3}$$

where $\xi \sim \mathrm{Lap}_{\xi}[b]$ is a sample from a Laplace distribution (figure 2) with scale b. The Laplace mechanism is ϵ-differentially private with $\epsilon = \Delta\mathrm{f}/b$. The term $\Delta\mathrm{f}$ is a constant called the sensitivity which depends on the function $\mathrm{f}[\bullet]$.

Source: https://www.borealisai.com/research-blogs/tutorial-12-differential-privacy-i-introduction/

Laplace Distribution

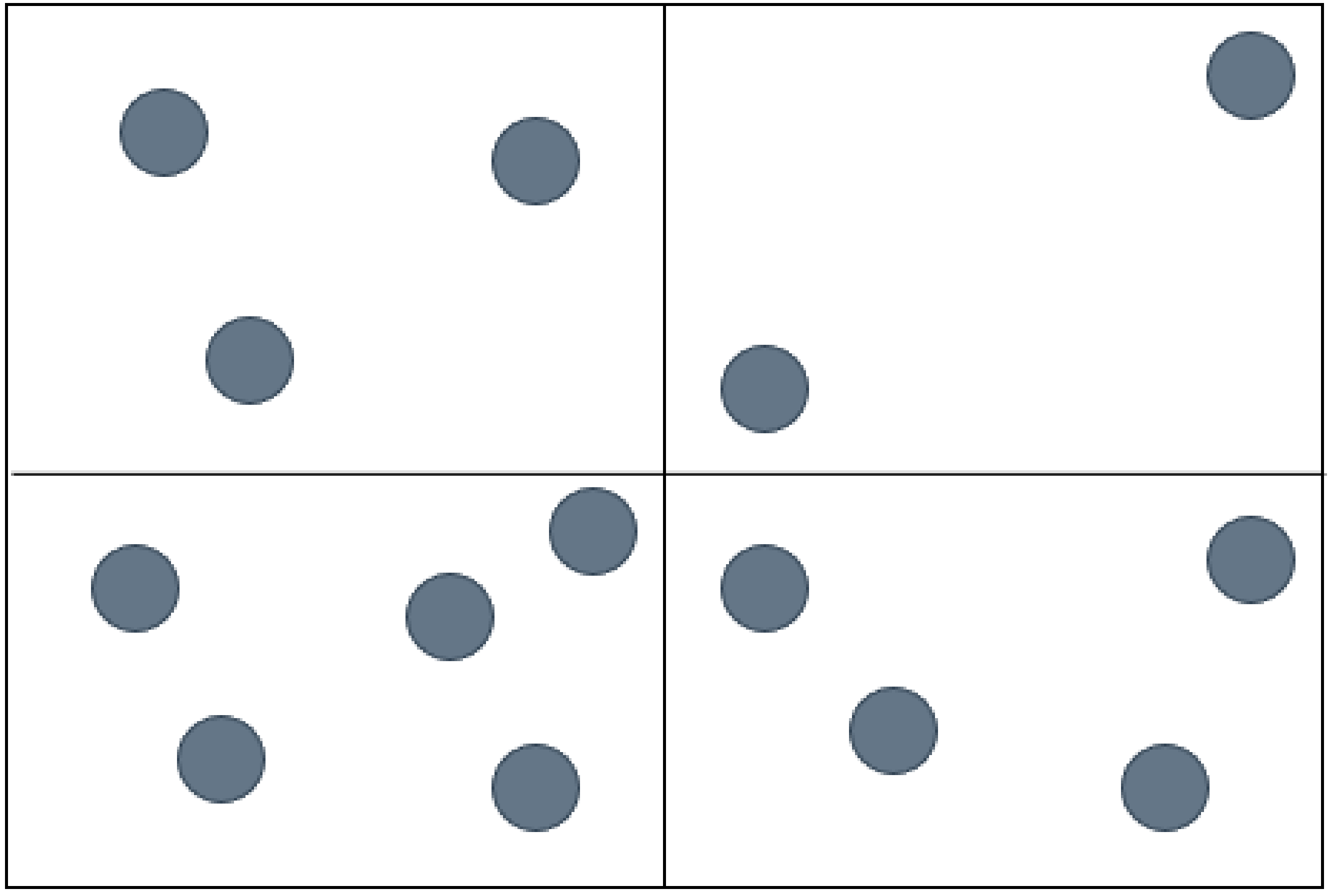


Source: https://www.borealisai.com/research-blogs/tutorial-12-differential-privacy-i-introduction/

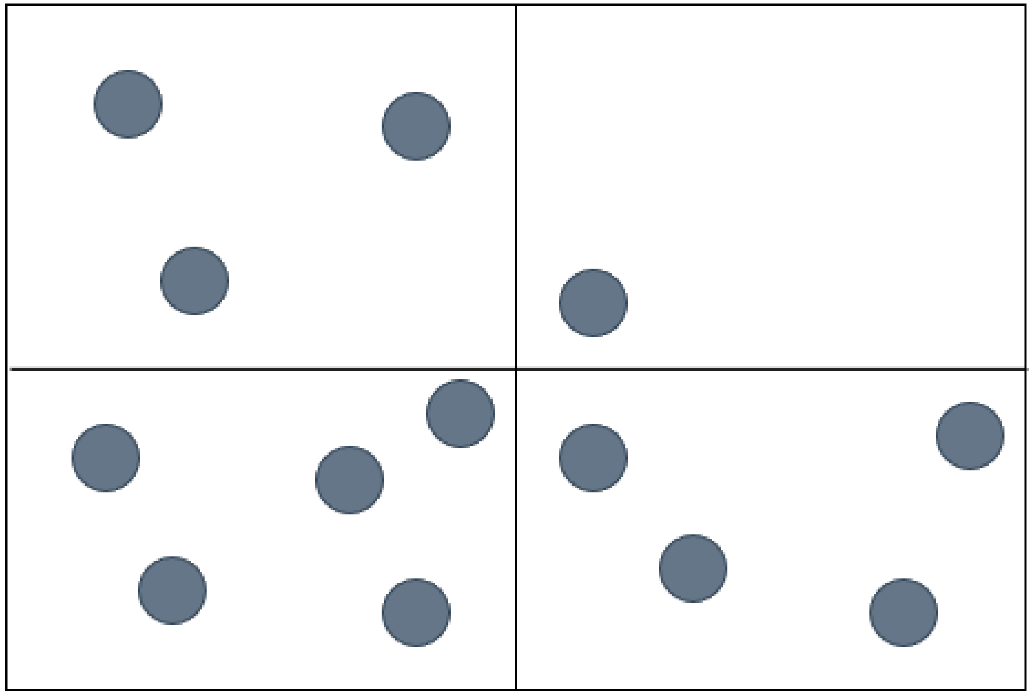
Function sensitivity - Δf

- Describes how much the output of the function can change with the addition or removal of a single element

Examples of function sensitivity

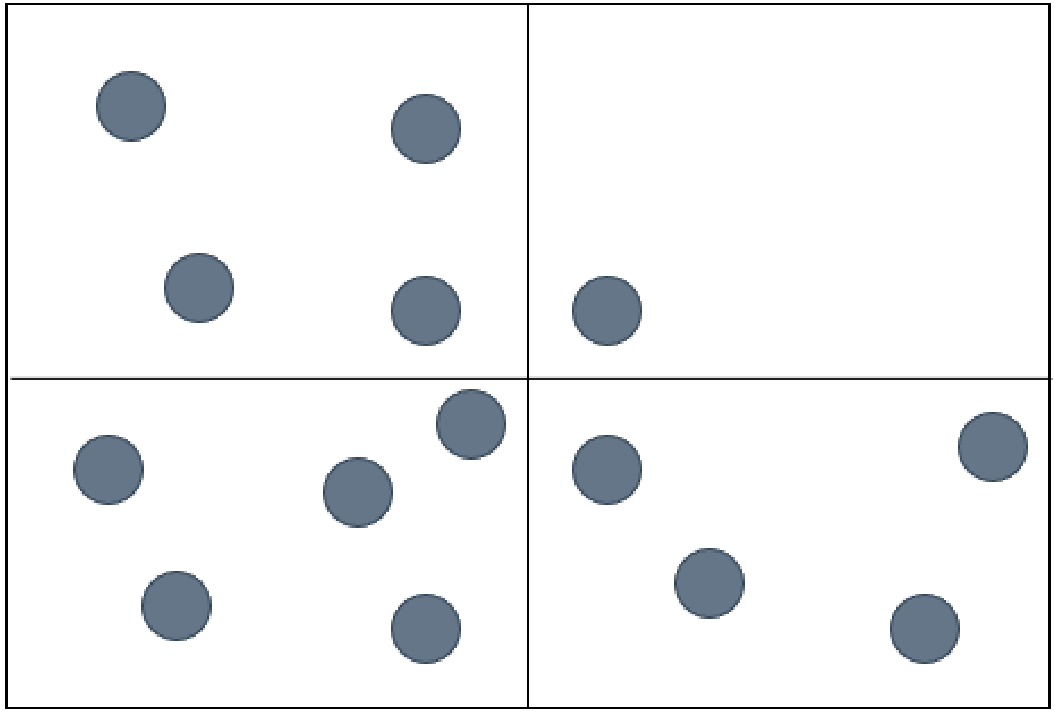


Examples of function sensitivity (2)



Δf = 1

Examples of function sensitivity (3)



$\Delta f = 2$

Randomized response and perturbations

- Ask individuals to answer to a "yes" or "no" question in a randomized manner
- ½ probability to give a truthful answer and ½ probability to give a random response
- Introduce plausible deniability: the mechanism forced respondents to lie
- Ensures that the individuals' answers can be claimed to be the product of chance rather than their true response

Randomized response and perturbations (2)

- Limitations:
  - Can introduce bias when the probability of a truthful answer is too low

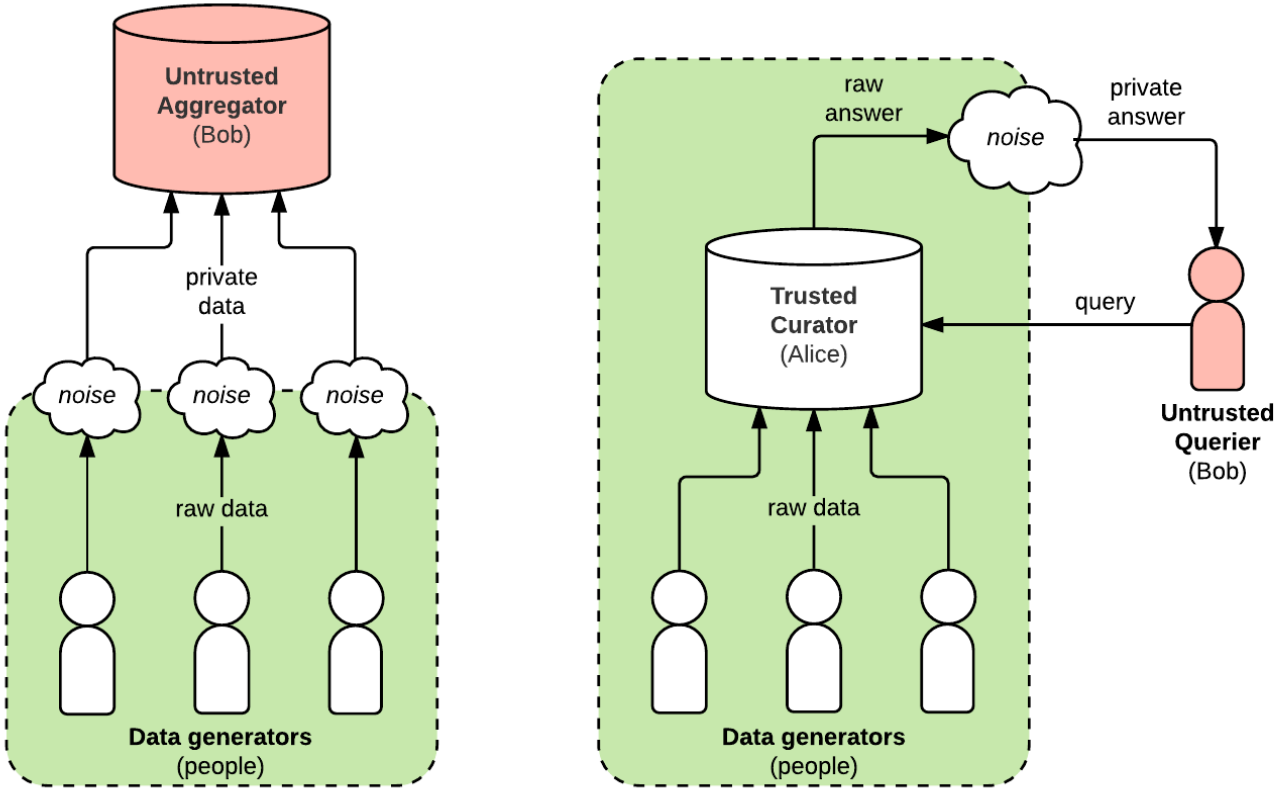
- Solution:
  - Ask multiple questions to better understand the statistical population

Types of DP

Types of DP

- Depending upon where the noise is added there are two types:
  - Local Differential Privacy
  - Global Differential Privacy

Types of DP (2)



Local privacy

Global privacy

Source: https://quantalabs.github.io/Differential-Privacy/

Privacy Preserving Machine Learning

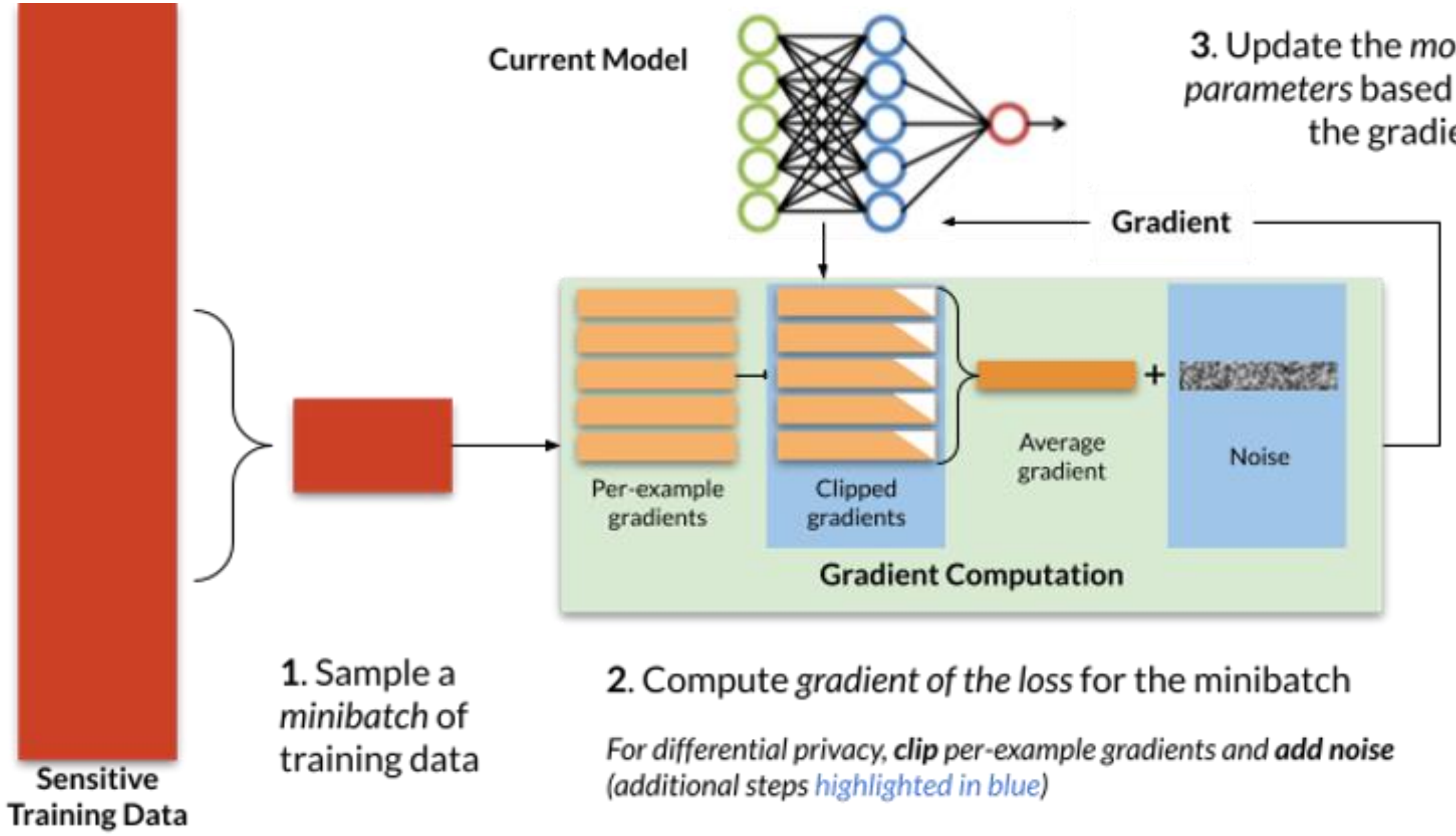
Privacy Preserving Machine Learning with DP

- Two main solutions proposed to achieve privacy in ML models using:
  - **DP-SGD** → Differential Private Stochastic Gradient Descent
  - DPL → Differential Private Logits

DP-SGD

- Noise is added to the gradients during model training
- Uses Gaussian or Laplacian noise
- Needs to clip the gradients when updating the model parameters
  - Done to control the sensitivity of the gradients
- The magnitude of the added noise is proportional to the number of steps of training
- Implementations in Tensorflow Privacy and PyTorch Opacus

DP-SGD (2)



Source: https://www.nist.gov/blogs/cybersecurity-insights/how-deploy-machine-learning-differential-privacy

DP-SGD (3)

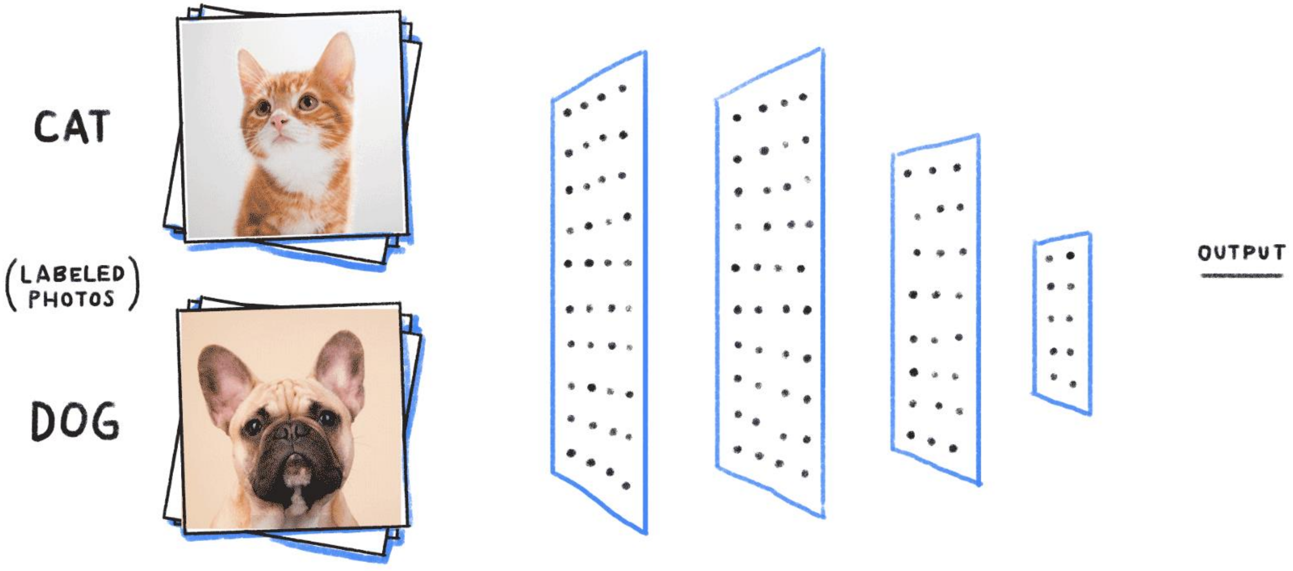
- Limitations:
  - It slows-down the training process
  - Addition of the noise during training can hurt the model accuracy

DP-Logits

- Approach that perturbs only the model outputs
- Can be applied directly on a trained ML model
- Noise is added at the prediction time, not in the training phase

Logits

- Unnormalised predictions (or outputs) of a model



Privacy budget

Disadvantage of DP without constraints

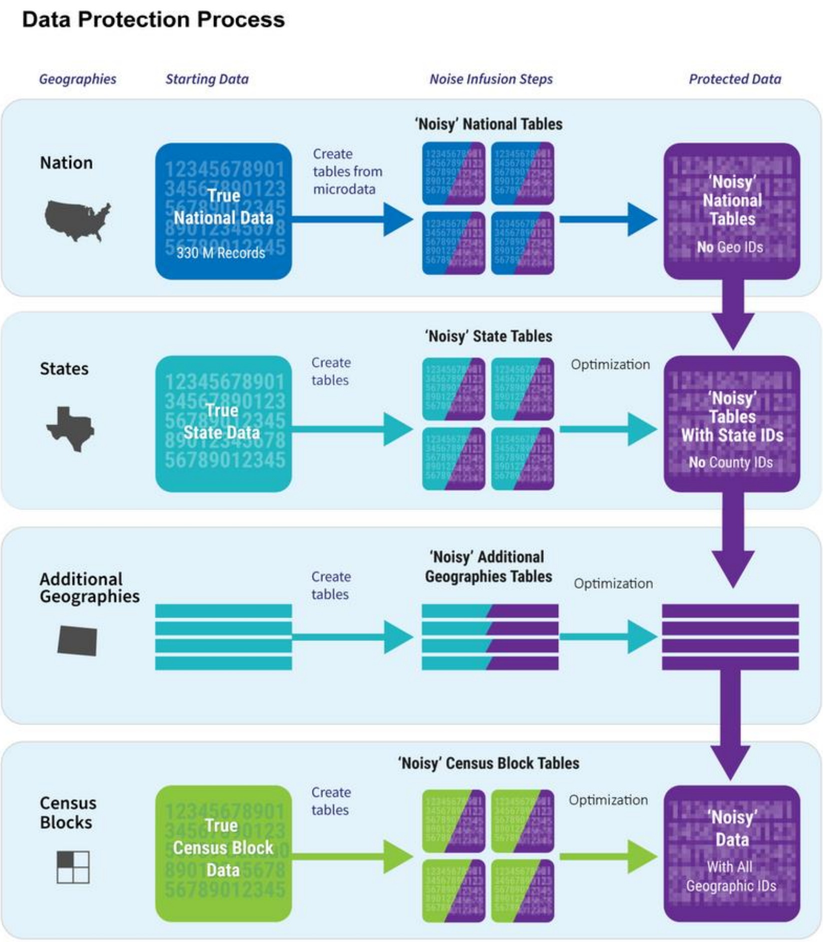
- At each query on the data privacy loss occurs
- Different queries → different results because of the randomness mechanism
- Because of that the level of anonymity of the data decreases (an attacker can filter out the noise)
- Solution: Introduction of the privacy budget

What is privacy budget?

- An upper limit established by a data curator
- Indicates the value of eps from where the data loses its anonymity
- Curators blocks the queries if the cost of the queries done on data is greater than the privacy budget
- 2017: Apple used a privacy budget of 14 per day

Implementations of DP in real world

US Census Bureau



Source: https://www.statice.ai/post/what-is-differential-privacy-definition-mechanisms-examples

Google RAPPOR

- Used to collect security metrics
- Local Differential Privacy
- Allows the analysis of the forest of the client data without the possibility to look at individual trees

Apple

- Used to improve user experience starting from what users do

- Local Differential Privacy

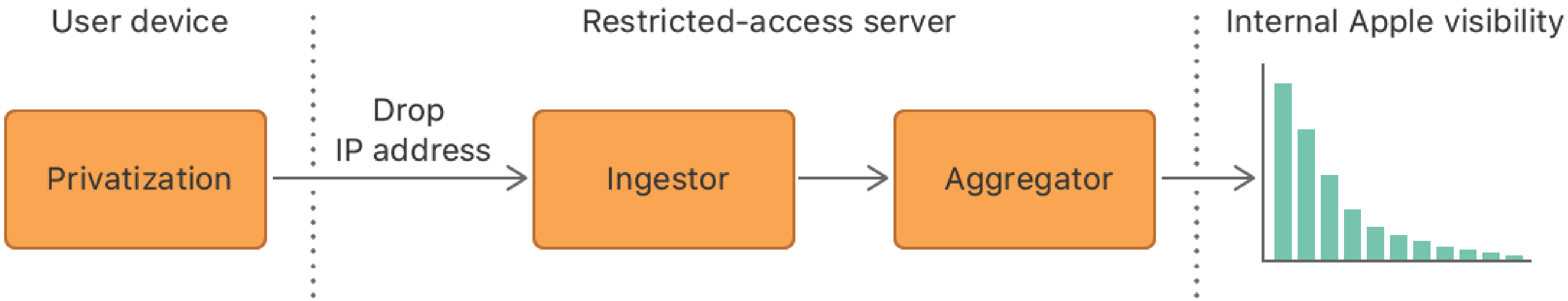
- Privacy budget per-donation → Limit for user' contributions

- Emoji suggestions, QuickType suggestions, Health Type Usage

Apple (2)

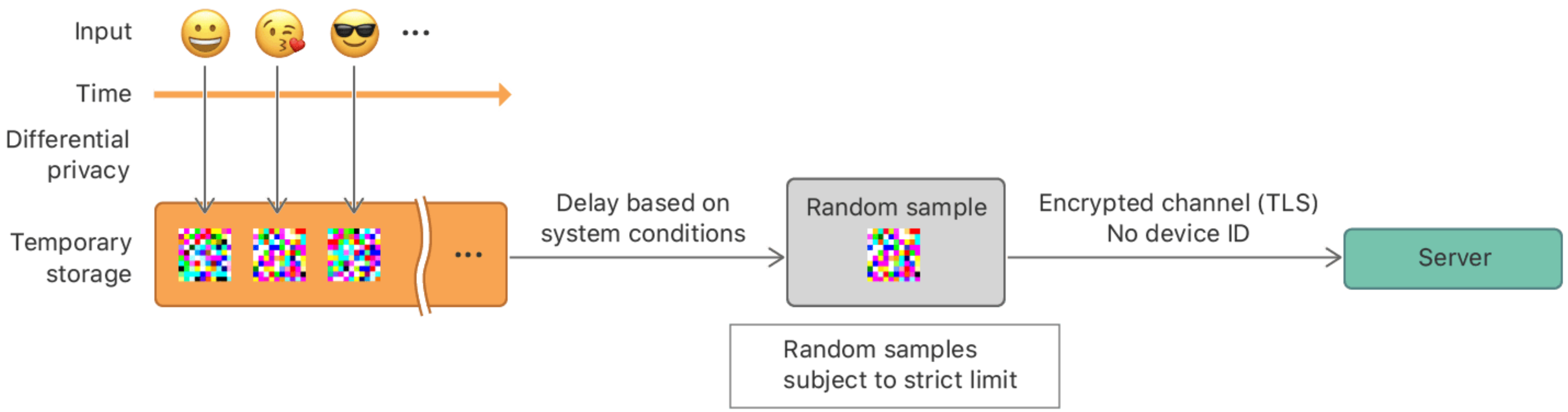
- The collected data is retained for a maximum of three months
- Different privacy budgets per feature

Apple (3)



Source: https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Apple (4)



Source: https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Apple (5)

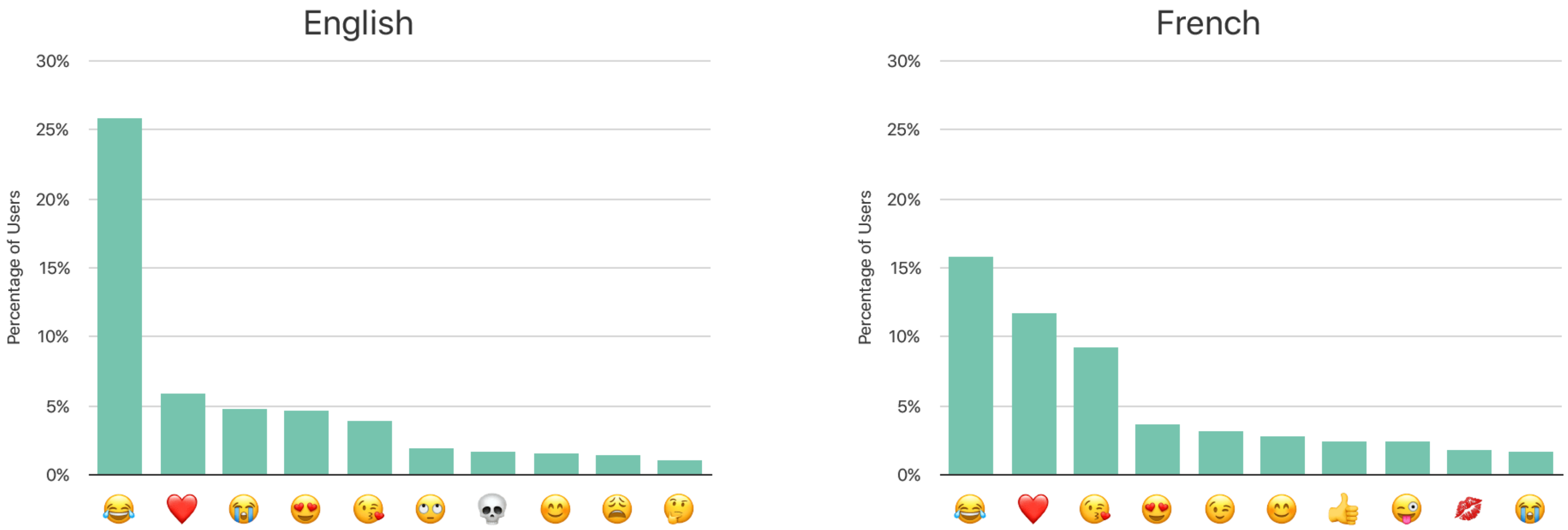
"key": "com.apple.keyboard.Emoji.en_US.EmojiKeyboard",

"parameters": {"epsilon":4,"k":65536,"m":1024},

"records": ["11688,00008200000000000000000200000004..."]

Source: https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Apple (6)

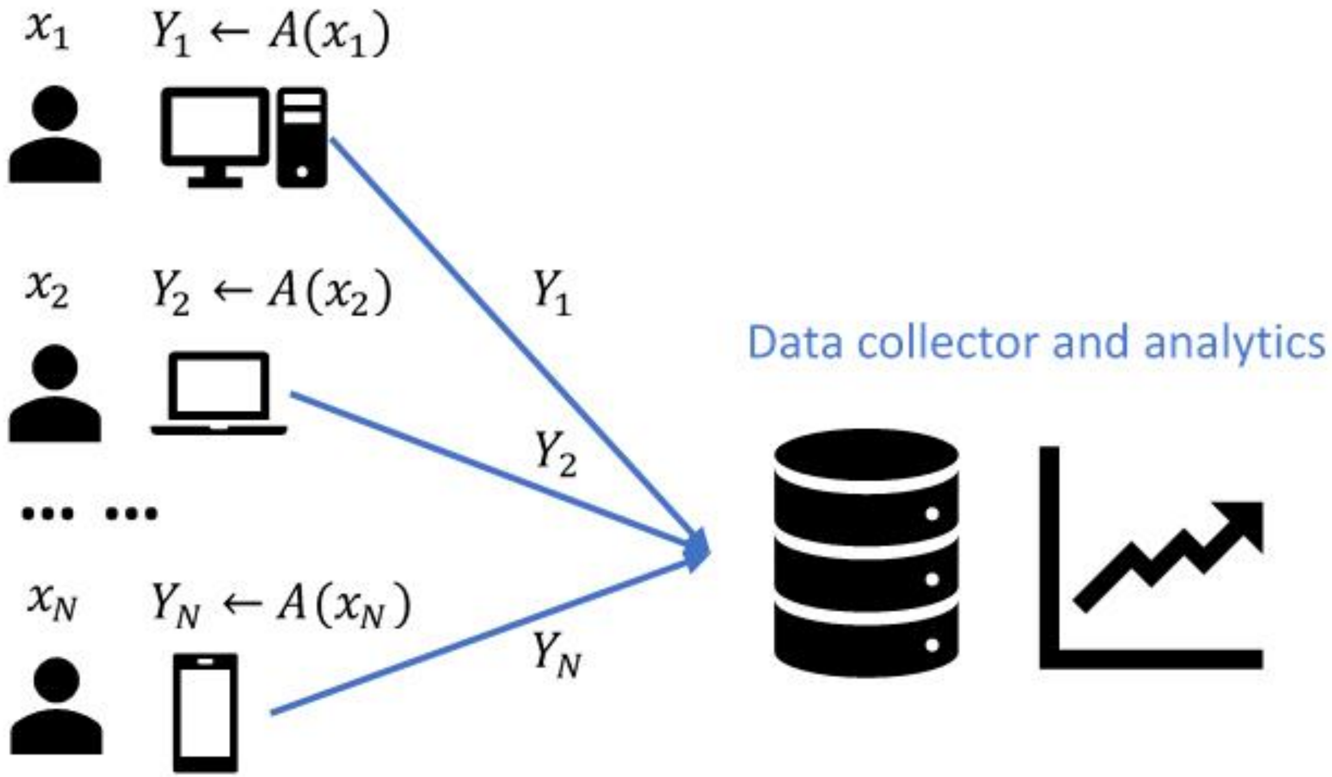


Source: https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Microsoft

- Targets the application telemetry:
  - Application usage statistics in Microsoft Windows
- Local Differential Privacy

Microsoft (2)



Source: https://www.microsoft.com/en-us/research/blog/collecting-telemetry-data-privately/

Limitations of DP

Limitations of DP

- Finding the perfect amount of noise to add is hard
  - Privacy vs Model Utility
- Computationally expensive in some cases
- Needs large databases to work on
- Without privacy budgets slow leaks can lead to full privacy loss

Privacy vs Model Utility

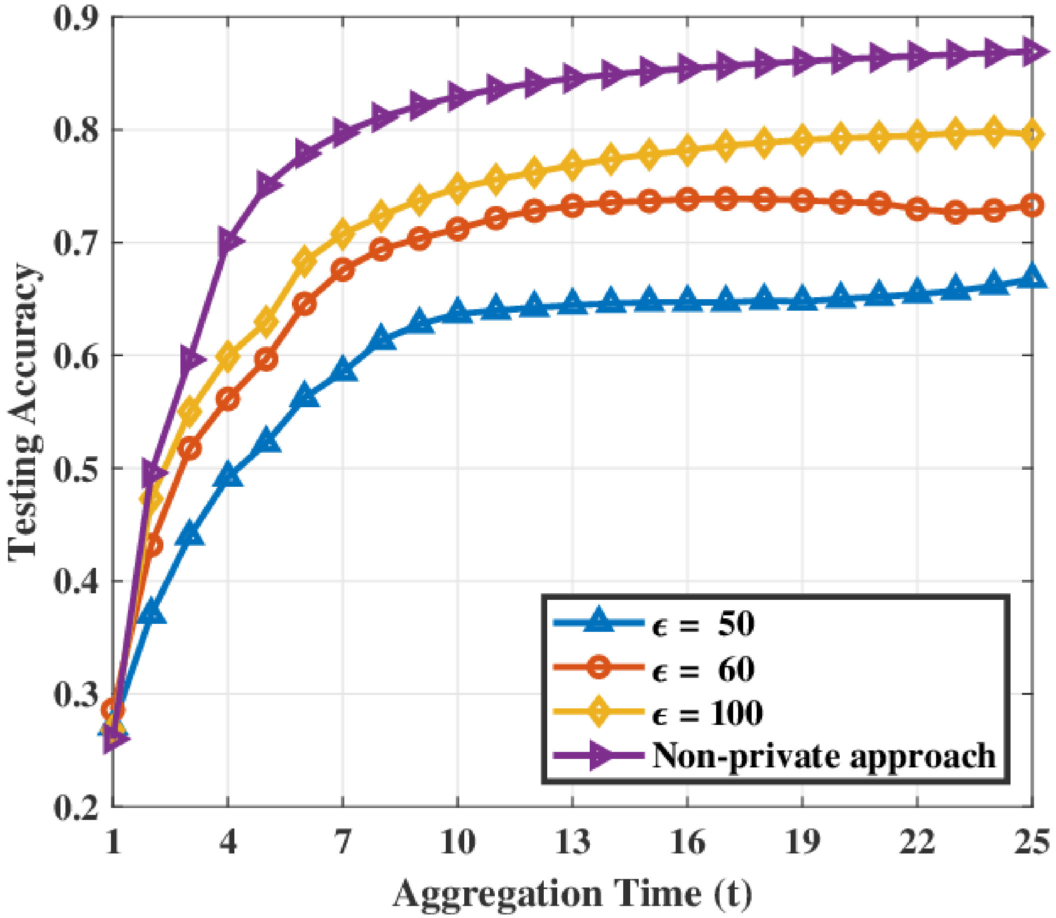


Figure 3: The comparison of training accuracy with various protection levels for

Source: https://www.semanticscholar.org/paper/Federated-Learning-With-Differential-Privacy%3A-and-Wei-Li/afa778ba0ba6333e25671cfb691a4bdda13b2868

Conclusions

Take away

- DP is based upon noise addition
  - On data
  - On predictions

- Not only the ML model accuracy is relevant
  - Also the **data privacy**
    - Privacy vs Model Utility

- To prevent full information leak
  - Privacy budget

References

1. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211-407.
2. https://www.statice.ai/post/what-is-differential-privacy-definition-mechanisms-examples
3. https://www.nist.gov/video/what-differential-privacy
4. https://privacytools.seas.harvard.edu/differential-privacy
5. https://www.educative.io/answers/what-is-differentially-private-sgd
6. Rahimian, S., Orekondy, T., & Fritz, M. (2020). Sampling attacks: Amplification of membership inference attacks by repeated queries. arXiv preprint arXiv:2009.00395.
7. https://www.nist.gov/blogs/cybersecurity-insights/how-deploy-machine-learning-differential-privacy
8. https://opacus.ai/tutorials/
9. https://blog.tensorflow.org/2019/03/introducing-tensorflow-privacy-learning.html

References (2)

10. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

11. https://machinelearning.apple.com/research/learning-with-privacy-at-scale

12. https://www.microsoft.com/en-us/research/blog/collecting-telemetry-data-privately/

13. https://www.ncsl.org/technology-and-communication/differential-privacy-for-census-data-explained

14. Erlingsson, Ú., Pihur, V., & Korolova, A. (2014, November). Rappor: Randomized aggregatable privacy-preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (pp. 1054-1067).