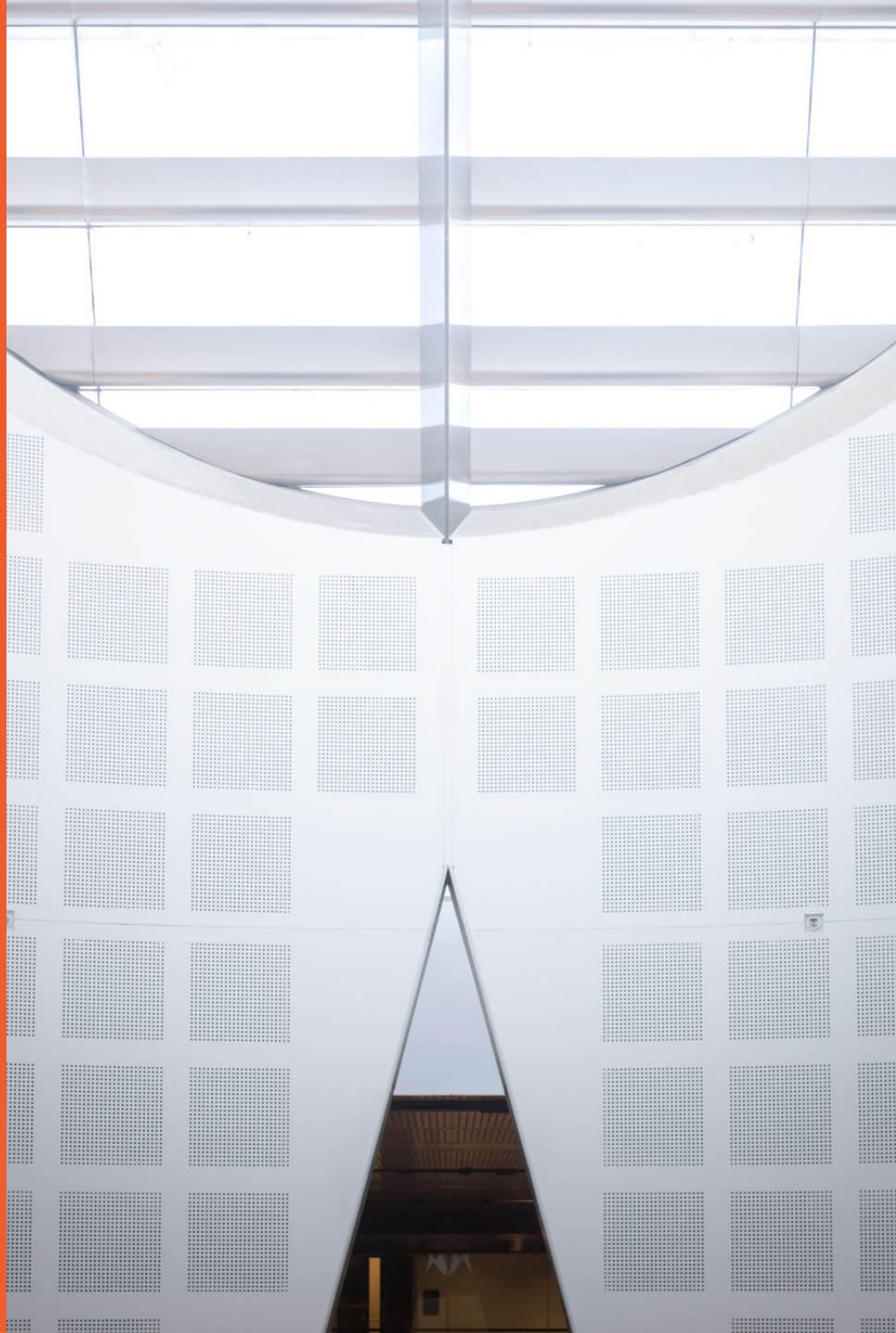
Data Privacy in Healthcare

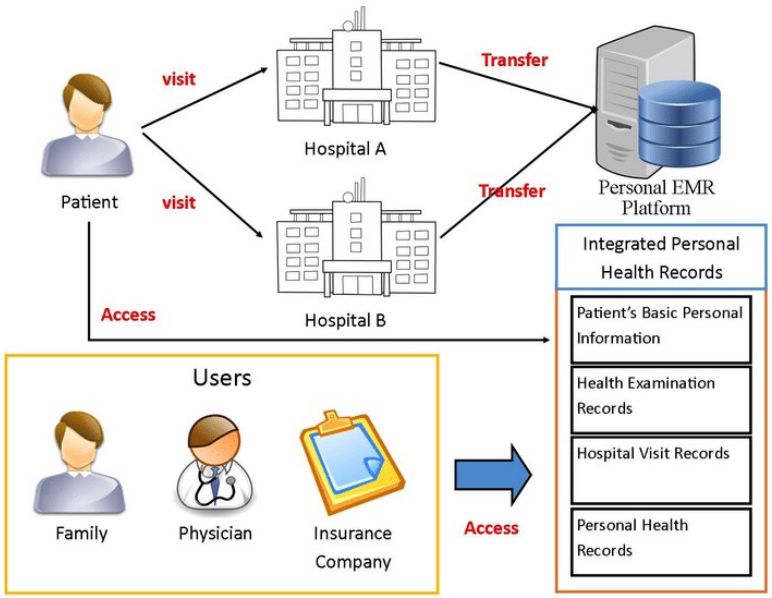
Reference: Healthcare Data Analytics, Chapter 15

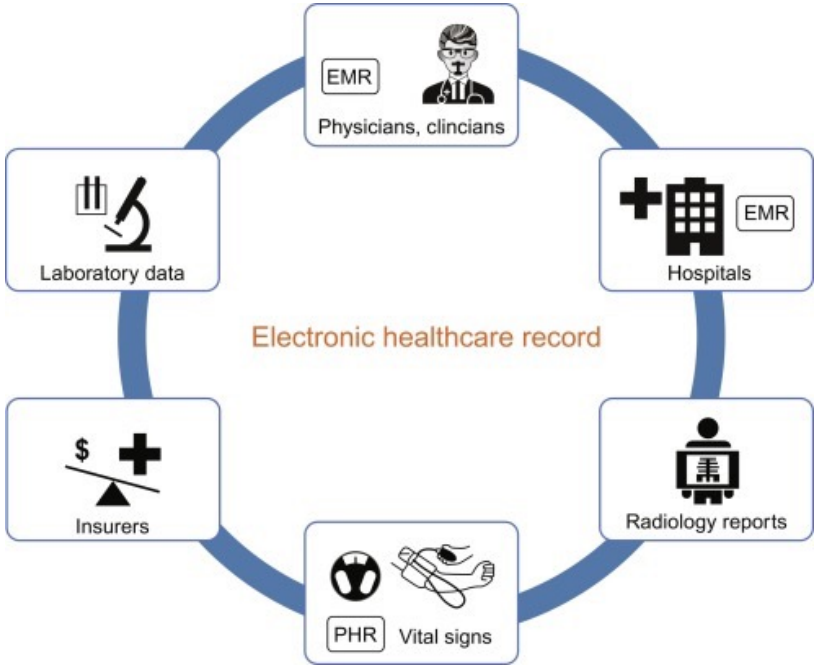


THE UNIVERSITY OF SYDNEY

Electronic Health Records

Electronic Health Records (EHRs) is a digital version of a patient's paper chart. **EHRs** are real-time, patient-centered records that make information available instantly and securely to authorized users. **EHRs** contain a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results.[1]





[1]. https://www.healthit.gov/faq/what-electronic-health-record-ehr

Usage of EHRs

Meaningful use of EHRs can help improve care coordination, reduce disparities, engage patients and their families and improve population and public health.

Such Meaningful use can only be achieved through **carefully controlled sharing and exchanging of personal health information and complying** with existing regulations such as the Health Insurance Portability and Accountability Act (HIPAA), otherwise the **privacy** of patients may be severely damaged.

In the United States, 75% of patients have expressed concerns about uninformed sharing of their health information.[1]

[1]. A. Raman. Enforcing privacy through security in remote patient monitoring ecosystems. In 6th International Special Topic Conference on Information Technology Applications in Biomedicine, pages 298–301, 2007.

Concerns of EHRs

Research indicates EHRs attract some of the highest prices on the dark web. The estimated mean value of EHR on criminal markets is $250 (up to $1000). EHRs are seen as good source of **Personally Identifiable Information (PII)**, as so many attributes are stored, and the data is more likely to be accurate. [1]

Medical Identity Theft

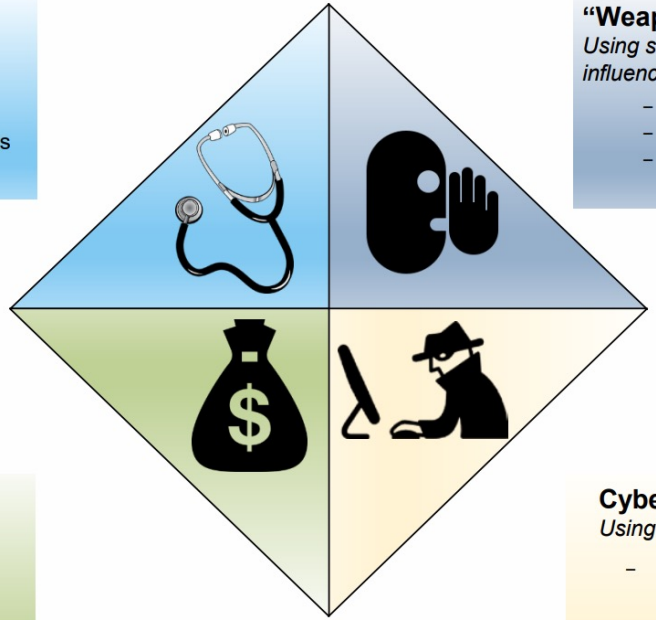
Utilizing someone else's personal medical information to obtain medical services

- Prescriptions for drugs
- Surgeries and medical procedures
- False medical insurance claims

"Weaponizing" healthcare data

Using sensitive healthcare data to threaten, extort, or influence individuals

- leverage for ransom payments
- Can be real or doctored data
- Public figures and VIPs highly susceptible



Financial Fraud

Using PII in medical records to establish credit/banking profiles for financial gain.

- Healthcare organizations often carry financial data of individuals
- Loans and lines of credit often require data found in medical records
- False tax return claims

Cyber Campaigns

Using healthcare data to support hacking campaigns

- Contact information can be targeted for phishing and scams
- Credential/authentication information can be used for access/privilege escalation

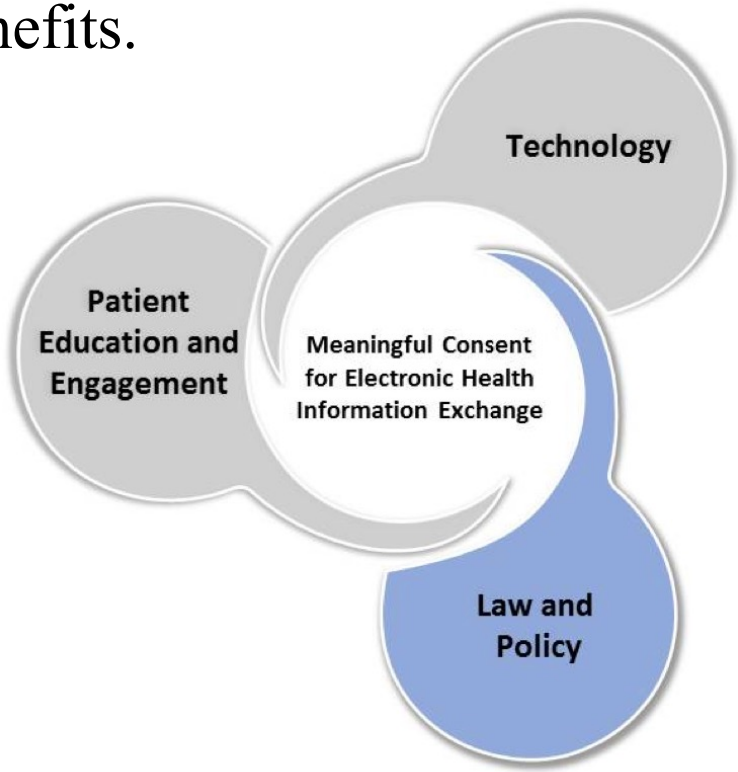
[1].https://content.govdelivery.com/attachments/USDHSFACIR/2019/04/25/file_attachments/1199378/Dark%20Web%20primer.pdf

EHRs Privacy & Utility

Privacy vs Utility:

Regulations on data privacy may create an elaborate set of cross-subsidies that reduces the total level of social wealth as it transfers wealth between parties. However, in a complex healthcare system, the negative consequences for open access of health information overwhelm the idealistic economical benefits.

For example, insurance companies and employers can maliciously utilize such data to increase their revenues, discriminating out unhealthy subpopulations. Thus, there exists a delicate **equilibrium point between utility and privacy**, and an extreme point cannot be a solution.



Introduction

Privacy is a subjective and contextual concept, and it conveys different connotations and interpretations in different fields. In the healthcare sector, the definition of privacy is commonly accepted as "a person's right and desire to control the **disclosure of their personal health information**", where the type of health information ranges from a person's identity to disease/medication history. We primarily focus on applying **privacy-preserving algorithms** to healthcare data for secondary-use data publishing.

Introduction

In practice, multiple privacy-preserving algorithms can be applied to EHRs in different stages. And there exist various privacy metrics, such as **k-anonymity** and ***ε*-differential privacy**.

We first introduce data pre-processing techniques, such as **variable reduction/coarsening**. Then we introduce some privacy-preserving techniques, such as **generalization/suppression**, **differential privacy**, **data imputation,** and **federated learning**.

EHRs Data Overview and Pre-processing

EHRs Variable Reduction

Our objective is *to publish a dataset for specified research objective while protecting patients' privacy*. Taking Texas Inpatient Public Use Data File from Texas Department of State Health Services as an example:

Variable Reduction: Given patients' records, Let us assume that we need to model the relationship between demographic factors (sex, address), insurance, and hospital charges. We first **remove irrelevant variables** except for five research objective: sex (of an infant), zip code, payment source (primary insurance), length of stay, and total charges.

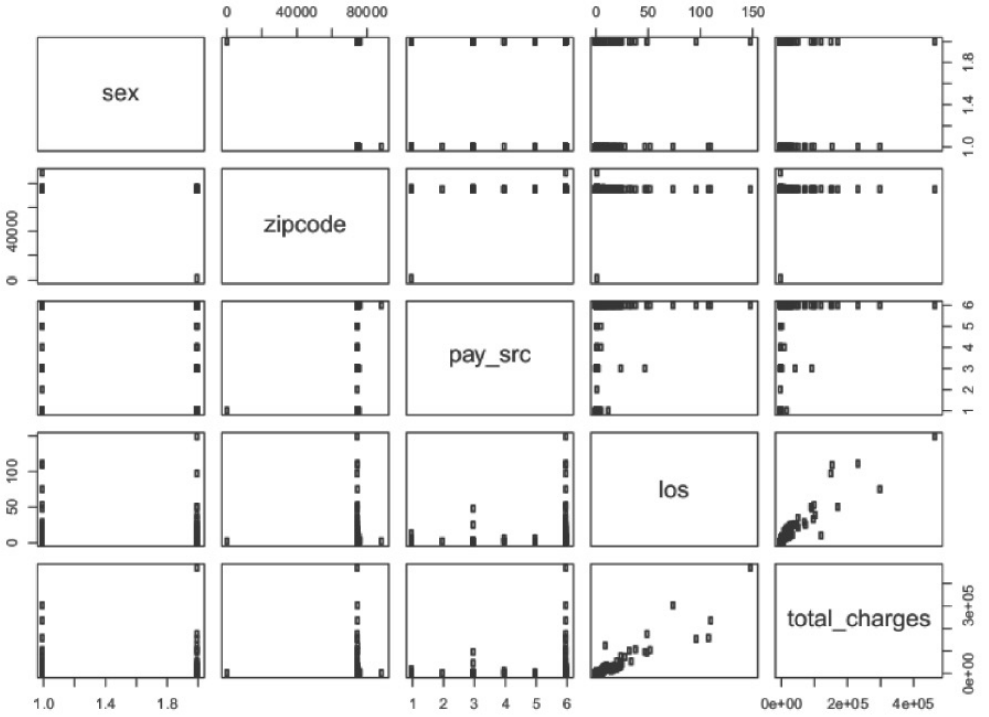
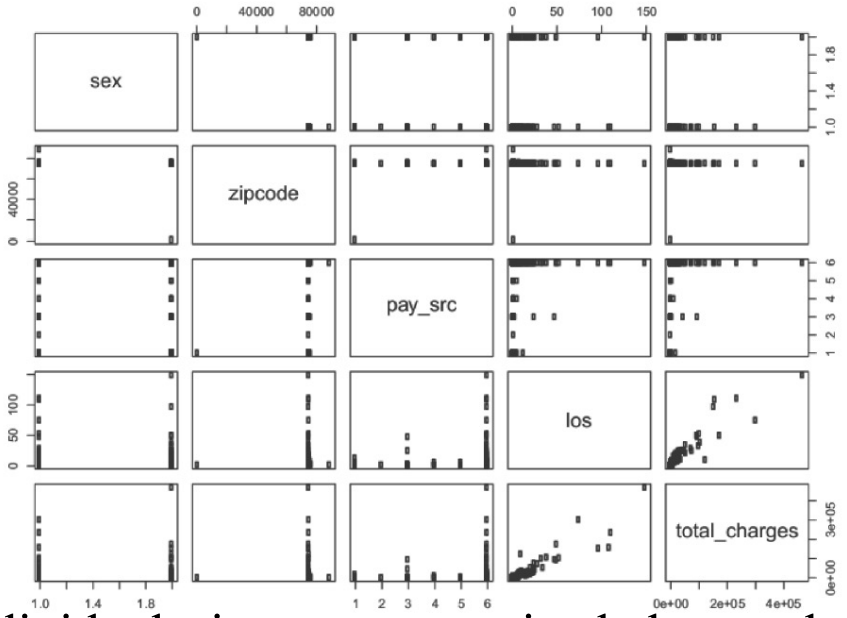


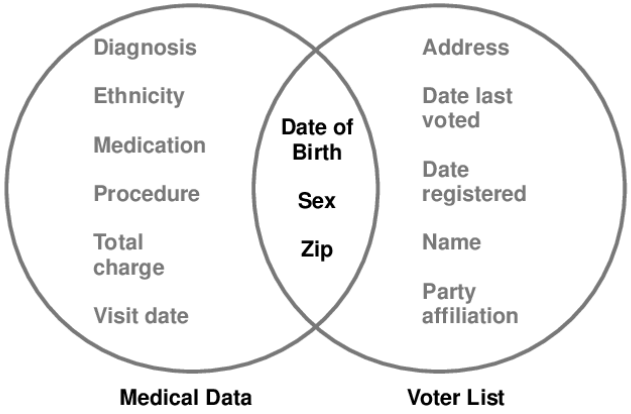
FIGURE 15.1: Cross-scatter plots of the original Texas inpatient data.

Linking Attack

After variable reduction, we need to check the characteristics of data. For example, the numeric variables tend to have many **unique entries** in the figure, which can be easily utilized by a **linking attack**.



Linking attack: An attempt to re-identify individuals in an anonymized dataset by combining the data with background information. The linking may use unique entries.



Private data

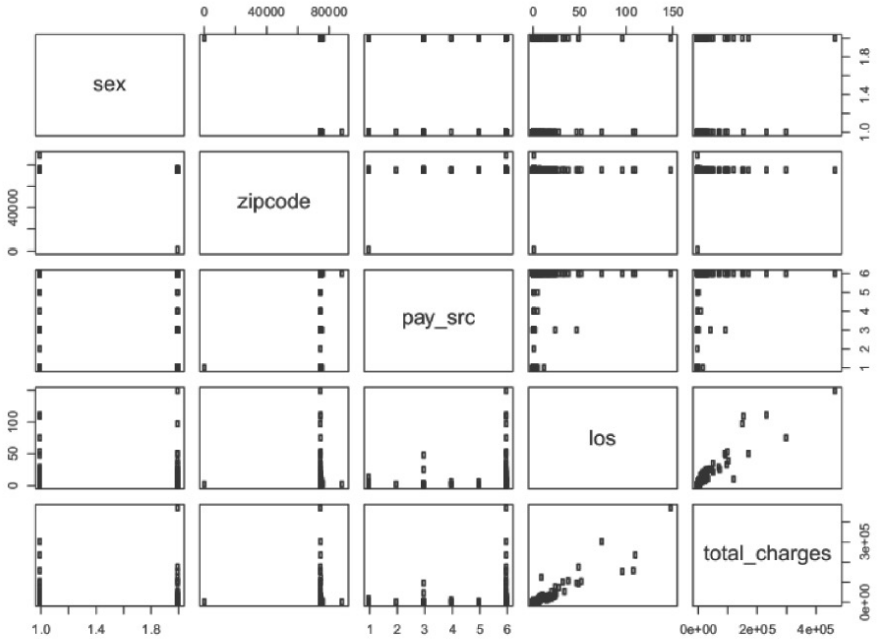
DOB	Sex	ZIP	Salary
1/21/76	M	53715	50,000
4/13/86	F	53715	55,000
2/28/76	M	53703	60,000
1/21/76	M	53703	65,000
4/13/86	F	53706	70,000
2/28/76	F	53706	75,000

Public data

SSN	DOB	Sex	ZIP
11-1-111	1/21/76	M	53715
33-3-333	2/28/76	M	53703

EHRs Data Overview and Pre-processing

Coarsening: We can bin the original numeric values of total charges into 20 ranges: [0, 500),{500, 1000), …[9500, 10000) to remove identifiable data points.



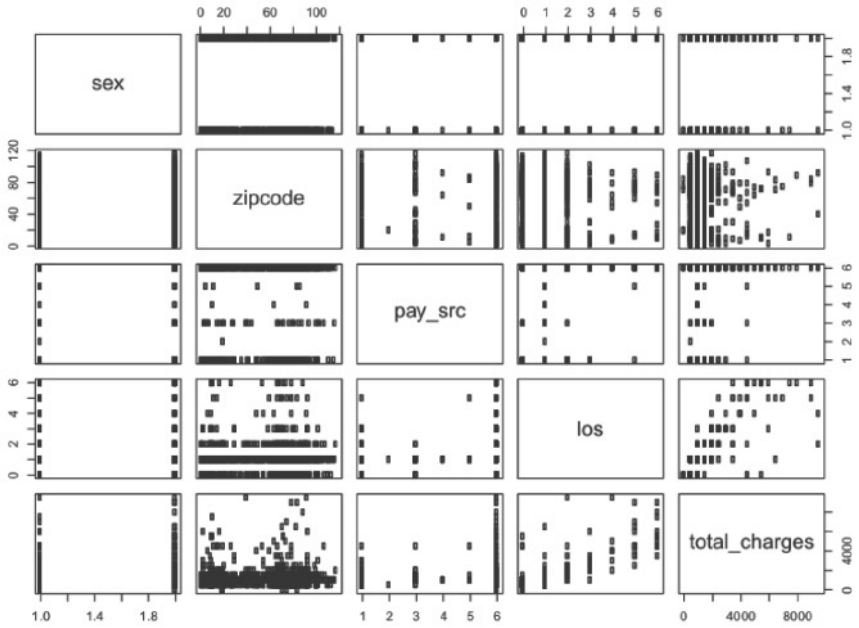


FIGURE 15.1: Cross-scatter plots of the original Texas inpatient data.

FIGURE 15.2: Scatter plots of the preprocessed Texas inpatient data.

We can see that **uniqueness** is a key concept in privacy-preserving techniques. These procedures are simple and effective, however, not sufficient for comprehensive privacy projection.

Generalization/Suppression

K-anonymity

Definition (**quasi-identifier**): A quasi-identifier is a set of variables within a dataset that may be **empirically unique**. Therefore, in principle, such a quasi-identifier can be used to uniquely identify a population unit.

Definition (**k-anonymity**): A Table is said to satisfy k-anonymity if and only if the quasi-identifier of the table appears with **at least k occurrences**.

We can use **k-anonymity** to formalize and quantify the disclosure risk of unique populations. In other words, to adhere the **k-anonymity** principle, each row in a dataset should be indistinguishable with at least k-1 other rows.

A Noanonymized database example of patient records: k=1 with quasi-identifier of name, age or religion.

Name	Age	Gender	Religion	Disease
Ramsha	30	Female	Hindu	Cancer
John	19	Male	Christian	Viral infection
Rambha	19	Male	Hindu	Cancer
Yadu	24	Female	Hindu	Viral infection
Salima	28	Female	Muslim	TB
Bahukasna	23	Male	Buddhist	TB

Generalization/Suppression

Generalization: Individual values of attributes are replaced with a broader category. For example, the value of "23" of the attribute Age can be replaced by "20 < Age < 30".

Suppression: Certain values of the attributes are replaced by an asterisk '*'. All or some values of a column may be replaced by '*'. In the anonymized table below, we have replaced all the values in the 'Name' attribute and all the values in the 'Religion' attribute with a '*'.

Through the combination of generalization and suppression, we can achieve k-anonymity for same value of k. Taking a Noanonymized table as an example:

Name	Age	Gender	Religion	Disease
Ramsha	30	Female	Hindu	Cancer
John	19	Male	Christian	Viral infection
Rambha	19	Male	Hindu	Cancer
Yadu	24	Female	Hindu	Viral infection
Salima	28	Female	Muslim	TB
Bahukasna	23	Male	Buddhist	TB

Generalization/Suppression

Name	Age	Gender	Religion	Disease
Ramsha	30	Female	Hindu	Cancer
John	19	Male	Christian	Viral infection
Rambha	19	Male	Hindu	Cancer
Yadu	24	Female	Hindu	Viral infection
Salima	28	Female	Muslim	TB
Bahukasna	23	Male	Buddhist	TB

Name	Age	Gender	Religion	Disease
*	(20,30]	Female	*	Cancer
*	(10,20]	Male	*	Viral infection
*	(10,20]	Male	*	Cancer
*	(20,30]	Female	*	Viral infection
*	(20,30]	Female	*	TB
*	(20,30]	Male	*	TB

Age Generalization

Name & Religion Suppression

k-anonymity (k=2), at least 2 rows with same attributes.

Differential Privacy

Differential Privacy

With generalization and suppression, it is still difficult to prevent unique identification of individuals.

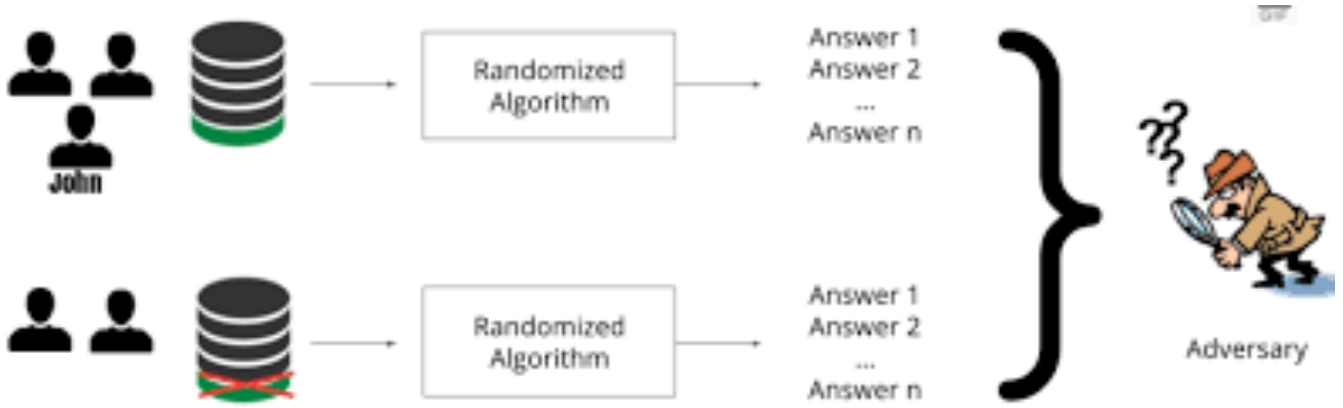
Name	Age	Gender	Zip Code	Smoker	Diagnosis
*	60–70	Male	191**	Y	Heart disease
*	60–70	Female	191**	N	Arthritis
*	60–70	Male	191**	Y	Lung cancer
*	60–70	Female	191**	N	Crohn's disease
*	60–70	Male	191**	Y	Lung cancer
*	***50–60***	***Female***	191**	N	HIV
*	50–60	Male	191**	Y	Lyme disease
*	50–60	Male	191**	Y	Seasonal allergies
*	***50–60***	***Female***	191**	N	Ulcerative colitis

From this fictional hospital database, if we know Rebecca is 55 years old and in this database, then we know she has 1 of 2 diseases.

How to guarantee? Differential Privacy.

Differential Privacy

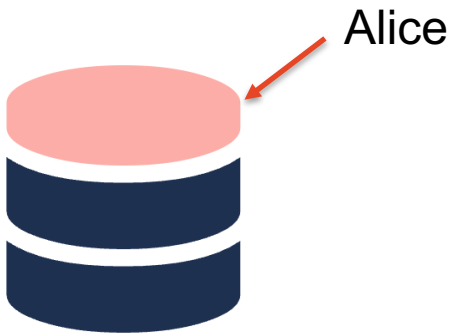
- Proposed by Cynthia Dwork in 2006
- Intuition: Perturb the result (e.g., by adding noise) such that the chance that the perturbed result will be C is nearly the same, whether or nor you submit your info.
- Differential privacy builds conceptually on a prior method known as **randomized response**. Here, the key idea is to introduce a randomization mechanism that provides plausible deniability.



Differential Privacy: A Conceptual Demo

Hypothetical Dataset

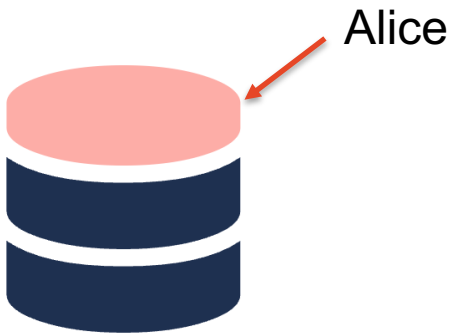
- A small country's health database with 1m people.
- We want to keep Alice's data private.



Differential Privacy: A Conceptual Demo

Without protection

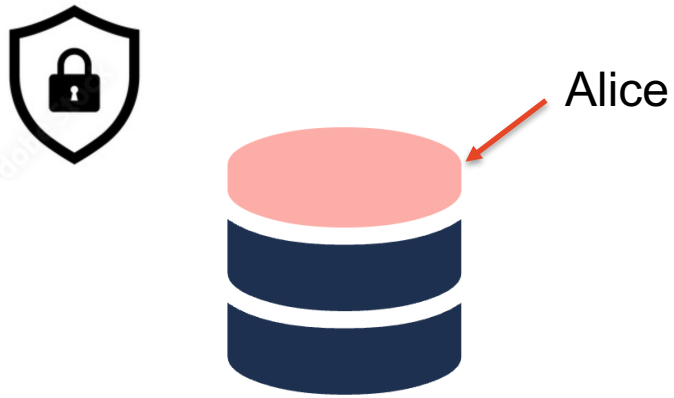
- Query: How many have diabetes?
- Result: 2,000 (Exact number returned).
- Remove Alice's data and query again.
- Result: 1,999 (Exact number returned).
- Adversary: Gotcha!



Differential Privacy: A Conceptual Demo

Solution: Differential Privacy

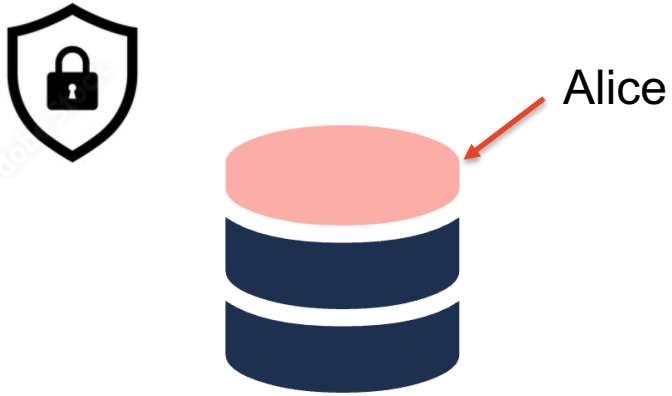
- A technique to protect individual's data in shared databases.
- Adds 'noise' to the result to obscure exact data.



Differential Privacy: A Conceptual Demo

With Differential Privacy

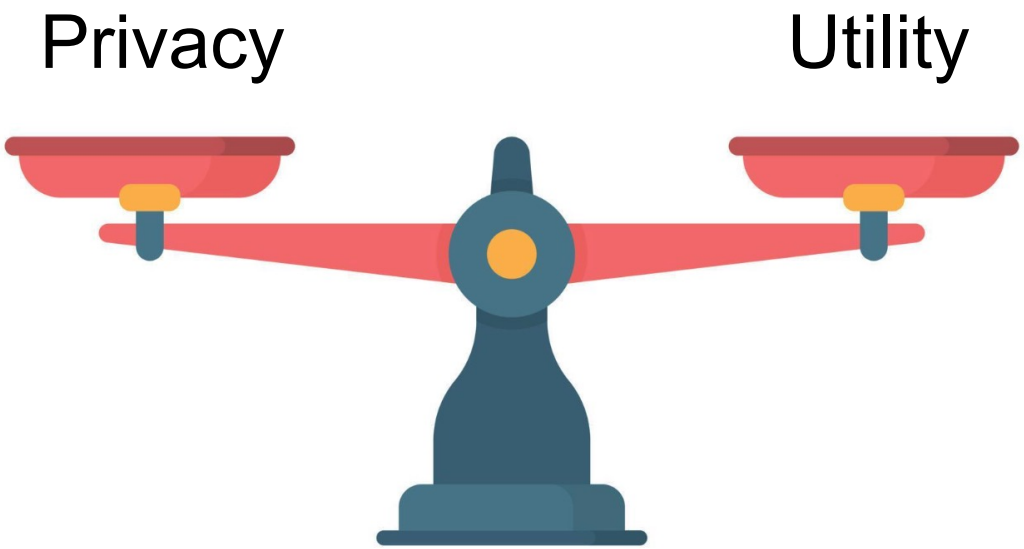
- Query: How many have diabetes?
- Result: 2,176 (Exact number protected).
- Remove Alice's data and query again.
- Result: 2,371 (Exact number protected).
- Adversary: ???



Differential Privacy: A Conceptual Demo

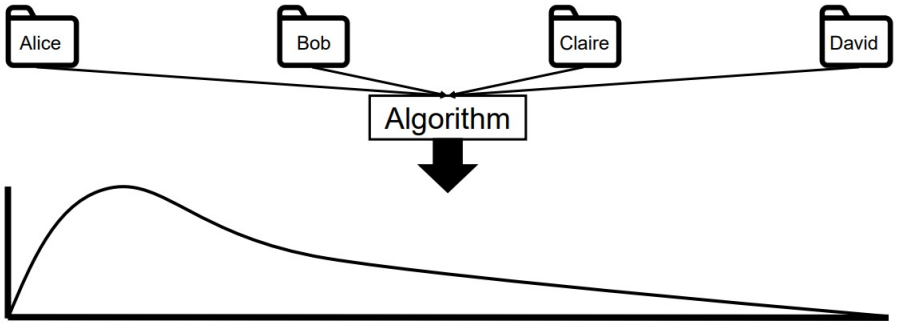
Privacy-Utility Trade-Off

- Adding noise increases privacy.
- But too much noise can reduce data's utility.

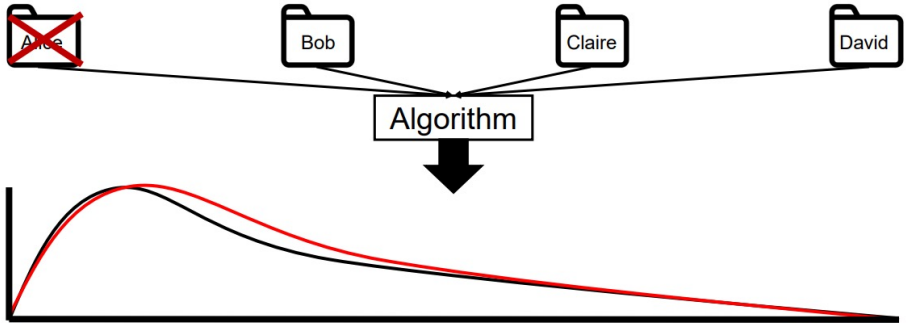


Differential Privacy

Assume we have two datasets which agree except for a single entry "Alice". If the probability density function of the output distributions of any algorithms are nearly identically for both datasets, the attacker cannot tell datasets apart and cannot learn much about the individual. Thus, this property provides plausible deniability.



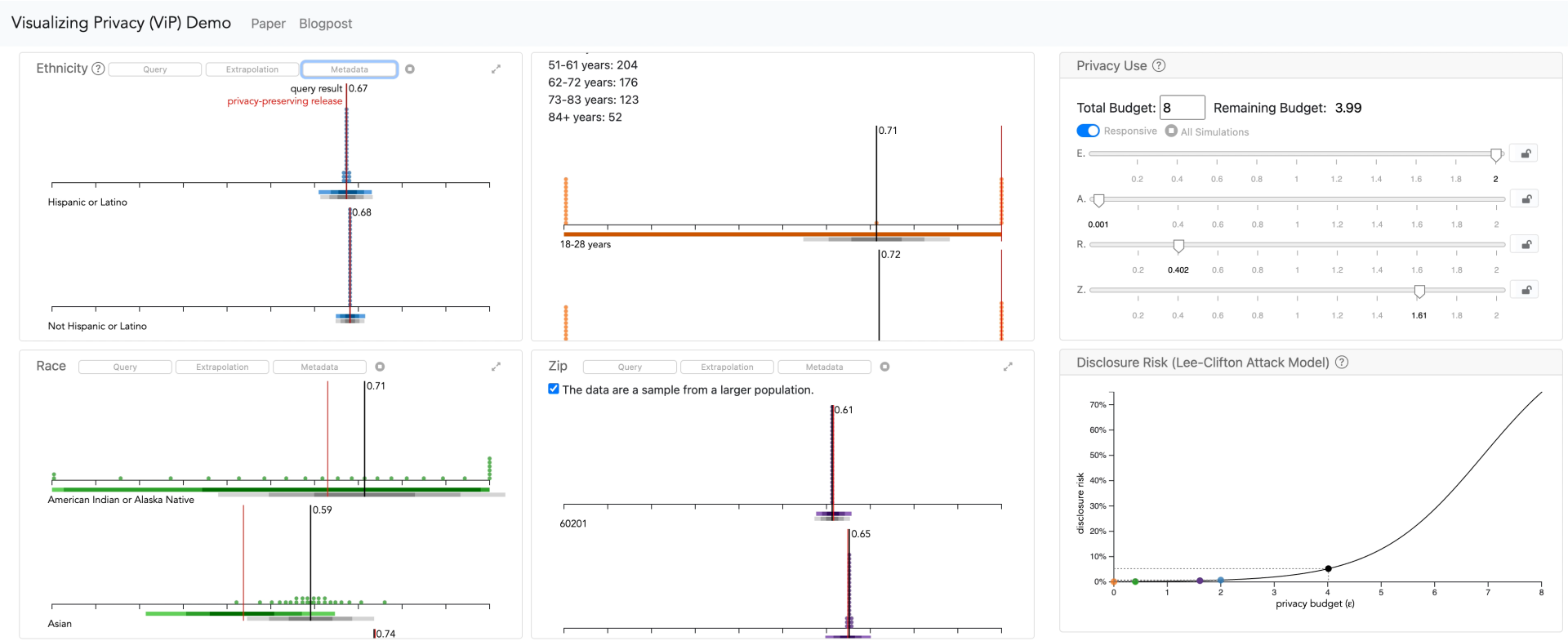
PDF of output distribution



Can't tell if used Alice's data, let alone what her data was!

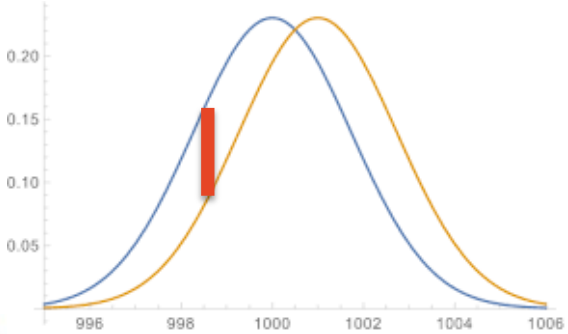
Differential Privacy

Demo: an interactive visualization interface for differential privacy



[1]. https://priyakalot.github.io/ViP-demo/

Differential Privacy



Definition of Differential Privacy (ϵ-DP) :

A mechanism $\mathcal{M}$ is ε-differentially private if for any two neighbouring databases X and X', and any set S of possible responses

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} Pr[\mathcal{M}(x') \in S]$$

Two databases X and X' are neighbouring if they agree except for a single entry.

Note: for small ε, $\exp(\varepsilon) \approx 1 + \varepsilon$.

The term ϵ controls how much the output of the mechanism can differ between the two adjacent databases and captures how much privacy is lost when the mechanism is run on the database. Large values of ϵ correspond to only weak assurances of privacy while values close to zero ensure that less privacy is lost.

Differential Privacy

Formally, differential privacy includes δ in the definition (ϵ, δ)-DP:

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

Think $0 \leq \delta \ll \frac{1}{n}$ (often, cryptographically small)

δ: Probability of information accidentally being leaked.

When δ is negligible or equal to zero: It is called ϵ-differential privacy. (ϵ, δ)-differential privacy says that **the absolute value of the privacy loss will be bounded by ε with probability at least 1−δ**.

Differential Privacy

How is differential privacy achieved?

Addition of carefully crafted random noise.

Laplace Mechanism:

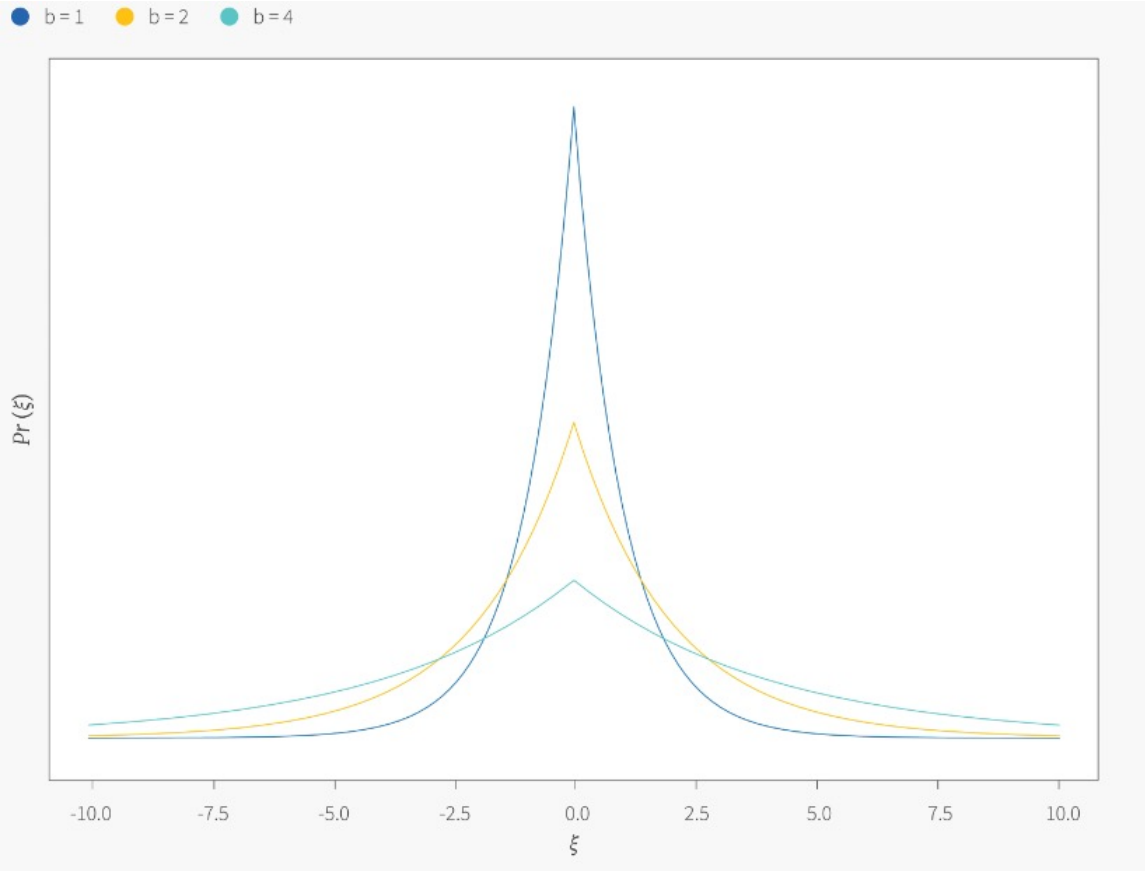
Add Laplace noise (i.e., noise from the Laplace distribution). Given function f which is the original real valued query/function we planned to execute on the dataset, the Laplace mechanism works by adding noise ξ to f

Noisy answer: return $f(x)+ \xi$,

where $\xi \sim Lap(b)$ is a sample from a Laplace distribution with scale $b=1/\epsilon$. The Laplace mechanism is ϵ-differentially private.

Differential Privacy

$$p(y;\mu,b) = \frac{1}{2b}\exp\left(-\frac{|y-\mu|}{b}\right)$$



Laplace distributions

Differential Privacy

Other Mechanisms?

Gaussian Mechanism:

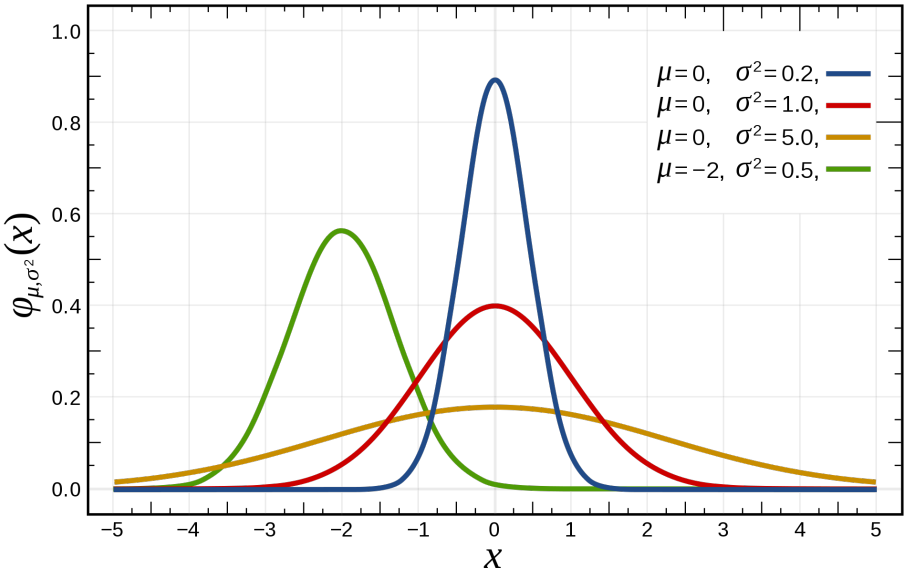
Add Gaussian noise (i.e., noise from the Gaussian distribution).

$$\mathcal{M}_{\text{Gauss}}(x, f, \epsilon, \delta) = f(x) + \mathcal{N}\left(\mu = 0, \sigma^2 = \frac{2\ln(1.25/\delta) \cdot (\Delta f)^2}{\epsilon^2}\right)$$

provides (ϵ, δ)-DP.

Gaussian distributions:

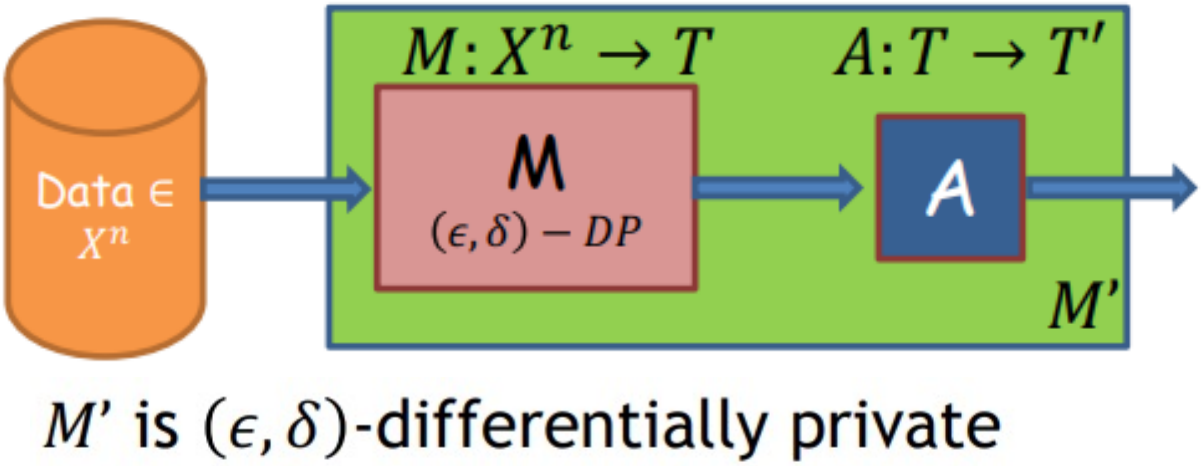
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Differential Privacy

Properties of differential privacy:

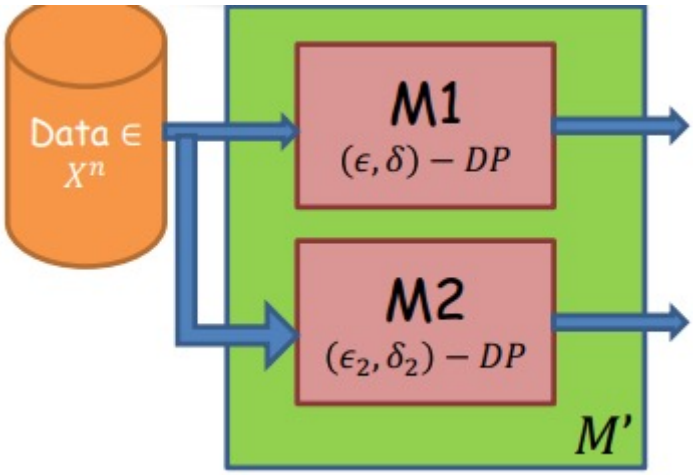
Post processing: Arbitrary data-independent transformations to the results of DP outputs cannot affect their privacy guarantees.



Differential Privacy

Properties of differential privacy:

Composition: If f is (ϵ_1, δ_1)-DP and g is (ϵ_2, δ_2)-DP, then f(x), g(x) is $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-DP.



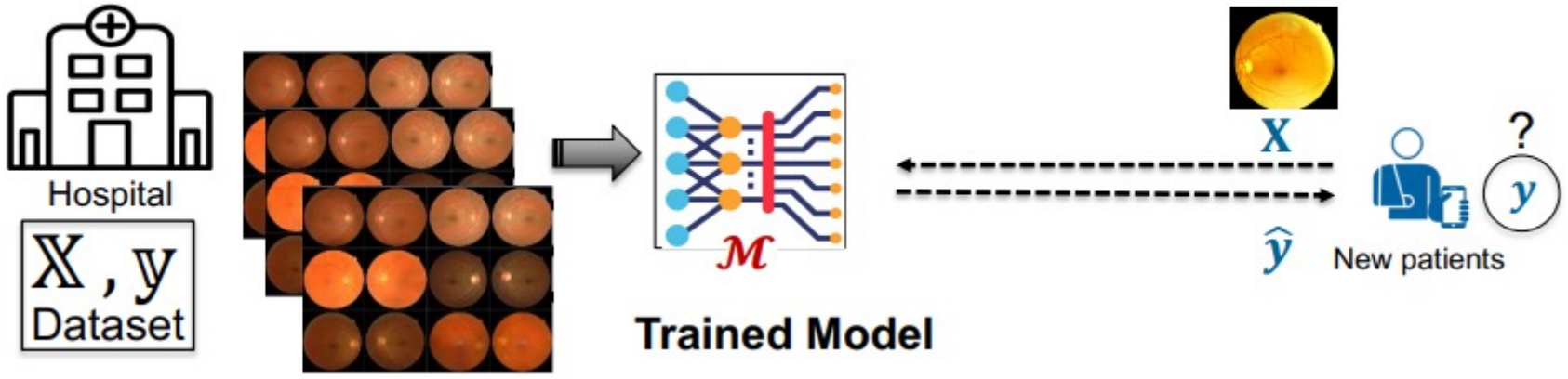
M' is $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$- differentially private

Privacy in Machine Learning of Healthcare

Privacy in Machine Learning of Healthcare

Privacy in Deep Neural Networks (DNNs):

The optimization of DNNs requires large amount of data entries. The trained model M must not reveal the presence (or absence) of any patient in the training set. [1]

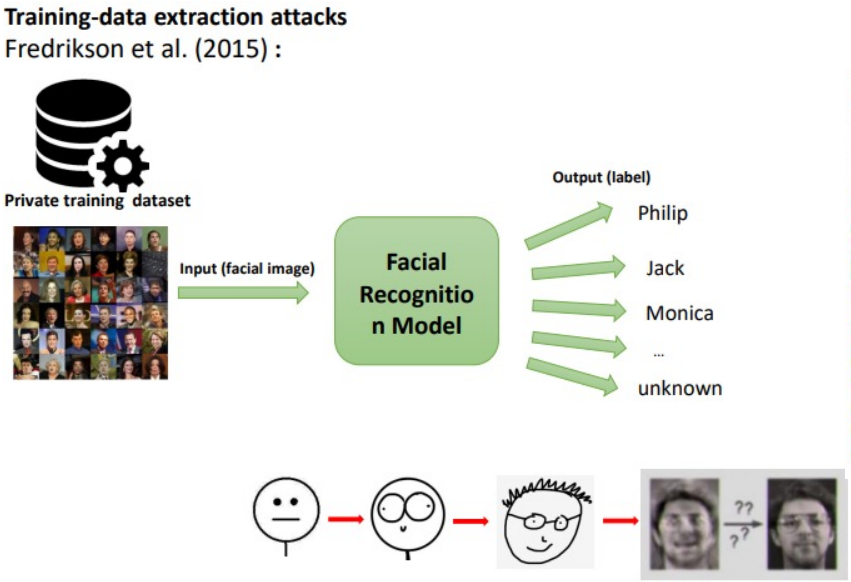


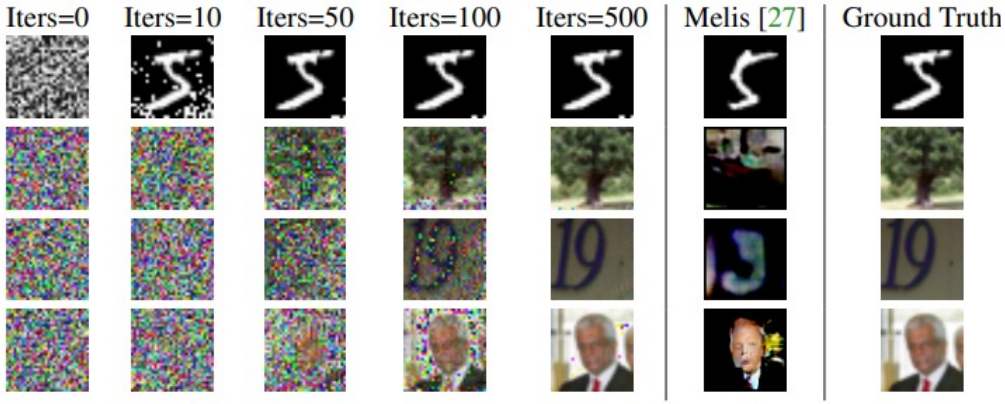
[1]. Abadi, Martin, et al. "Deep learning with differential privacy." Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. 2016.

Privacy in Machine Learning of Healthcare

Privacy in Deep Neural Networks (DNNs):

It is possible to recover the private training data with a trained model or the gradients. [1,2] It could be dangerous for some researchers to release their trained models on private patients' data.





[1]. Carlini, Nicholas, et al. "Extracting training data from large language models." 30th USENIX Security Symposium (USENIX Security 21). 2021.
[2]. Zhao, Bo, Konda Reddy Mopuri, and Hakan Bilen. "idlg: Improved deep leakage from gradients." arXiv preprint arXiv:2001.02610 (2020).

Privacy in Machine Learning of Healthcare

Privacy in Deep Neural Networks (DNNs):

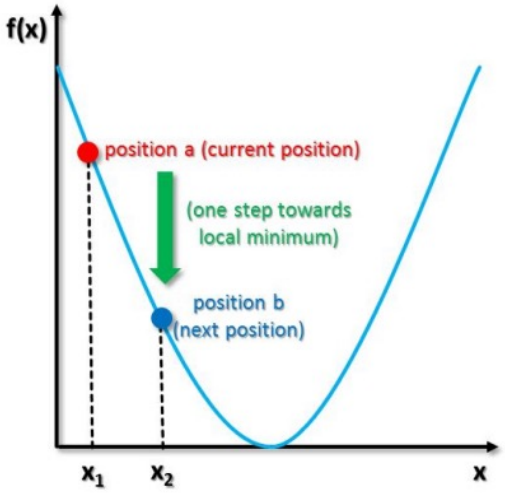
Since there exist privacy risks of DNNs, how can we guarantee the privacy of DNNs? Again, Differential privacy! Incorporate DP in the optimization of DNNs.

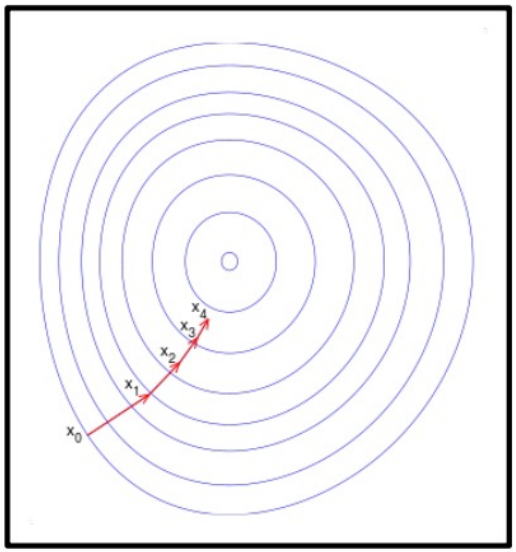
Gradient Descent Algorithm

$$W_{t+1} = W_t - \eta \nabla \mathrm{f}(\mathrm{W_t})$$

W: model parameters

η : learning rate





Privacy in Machine Learning of Healthcare

Optimization of DNNs:

- Gradient Descent:
  - The cost gradient is based on the complete training set, can be costly and longer to converge to minimum
- Stochastic Gradient Descent:
  - Update the weight after each training sample
  - Converges faster but the path towards minimum may zig-zag
- Mini-Batch Gradient Descent:
  - Update the weights based on small group of training samples
  - For example, randomly sampling training data with batch size of B. The gradient of this batch can be computed as $g_B = 1/B \sum_{x \in B} \nabla_\theta \mathcal{L}(\theta, x)$

Differentially Private SGD

Modify the mini-batch gradient descent via clipping and noise.

Algorithm 1 Differentially private SGD (Outline)

Input: Examples $\{x_1, \ldots, x_N\}$, loss function $\mathcal{L}(\theta) = \frac{1}{N}\sum_i \mathcal{L}(\theta, x_i)$. Parameters: learning rate η_t, noise scale σ, group size L, gradient norm bound C.

Initialize θ_0 randomly

for $t \in [T]$ **do**

Take a random sample L_t with sampling probability L/N

Compute gradient

For each $i \in L_t$, compute $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t}\mathcal{L}(\theta_t, x_i)$

Clip gradient

$\bar{\mathbf{g}}_t(x_i) \leftarrow \mathbf{g}_t(x_i) / \max\left(1, \frac{\|\mathbf{g}_t(x_i)\|_2}{C}\right)$

Add noise

$\tilde{\mathbf{g}}_t \leftarrow \frac{1}{L}\left(\sum_i \bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I})\right)$

Descent

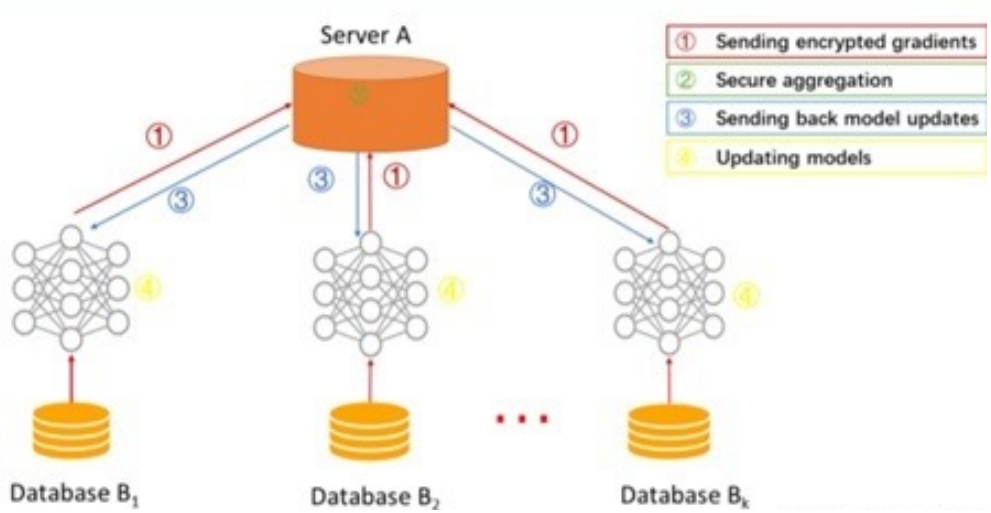
$\theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{\mathbf{g}}_t$

Output θ_T and compute the overall privacy cost (ε, δ) using a privacy accounting method.

Federated Learning

Federated Learning

Federated learning (also known as collaborative learning) is a machine learning technique that **trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging them**. This approach stands in contrast to traditional centralized machine learning techniques where all the local datasets are uploaded to one server, as well as to more classical decentralized approaches which often assume that local data samples are identically distributed.



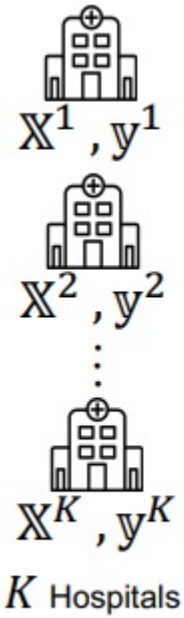
Federated Learning in Healthcare

The patients' EHRs are kept privately in each hospital. Federated learning enables the optimization of DNNs without exchanging these private data.

Medical Dataset are **Distributed** and Kept **Private**.

Patients' **Privacy** is as important as Patient's **Health**

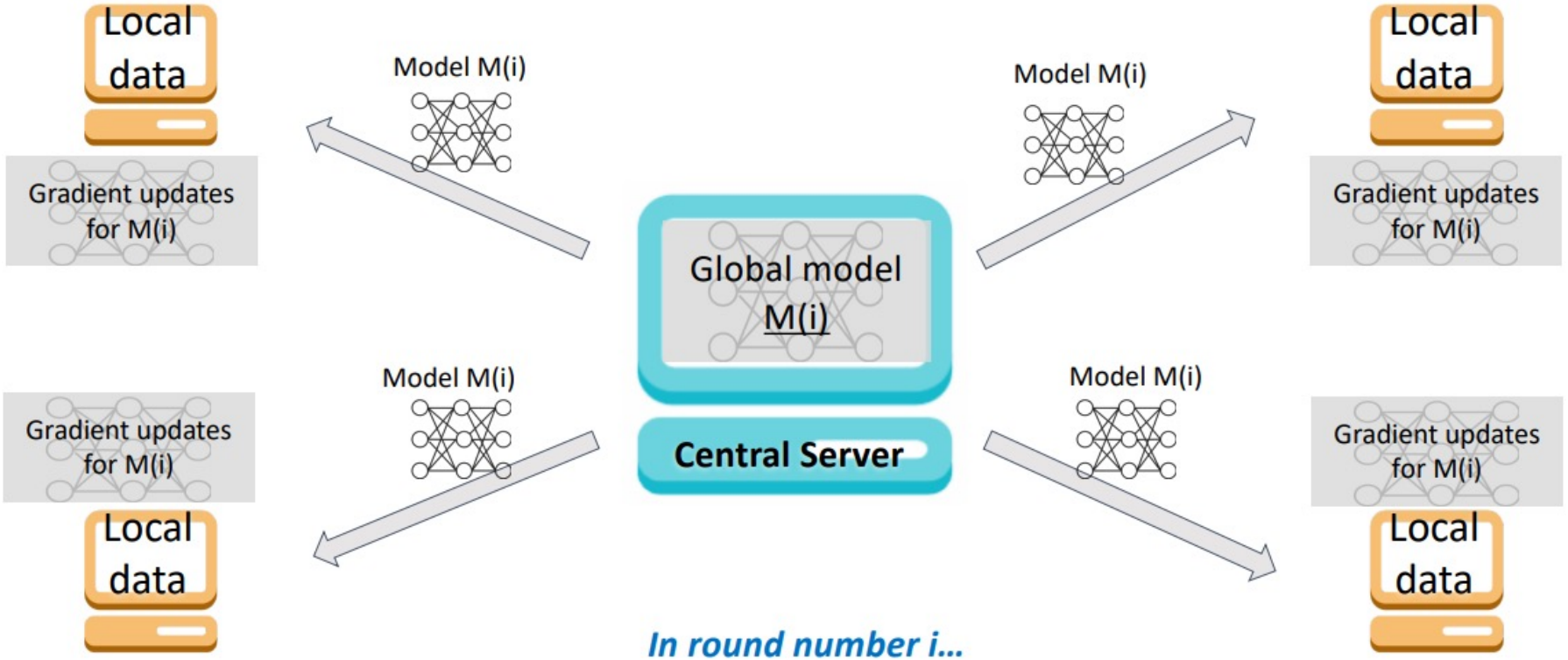
Protected by the **law**!



Federated Learning in Healthcare

Overview of Federated Learning steps:

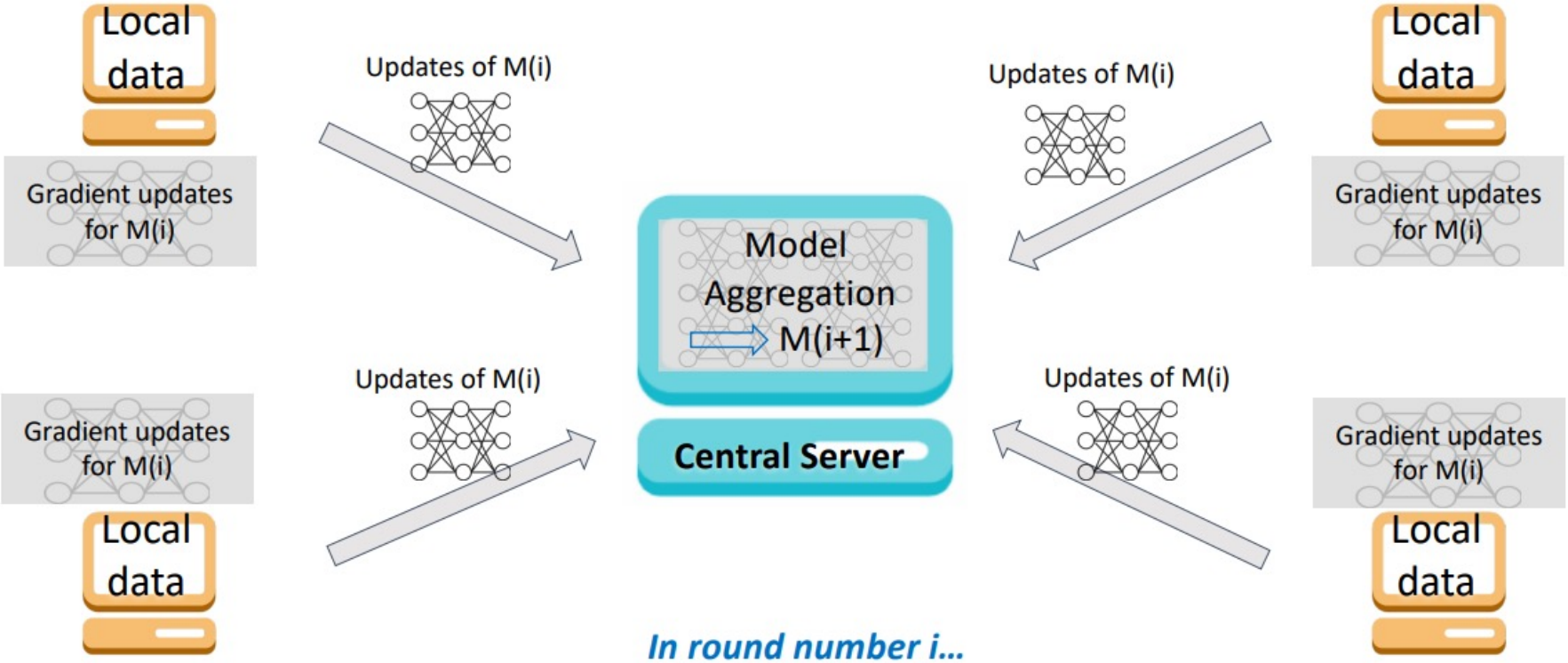
1. Initialize DNN on the central server and send model to clients



Federated Learning in Healthcare

Overview of Federated Learning steps:

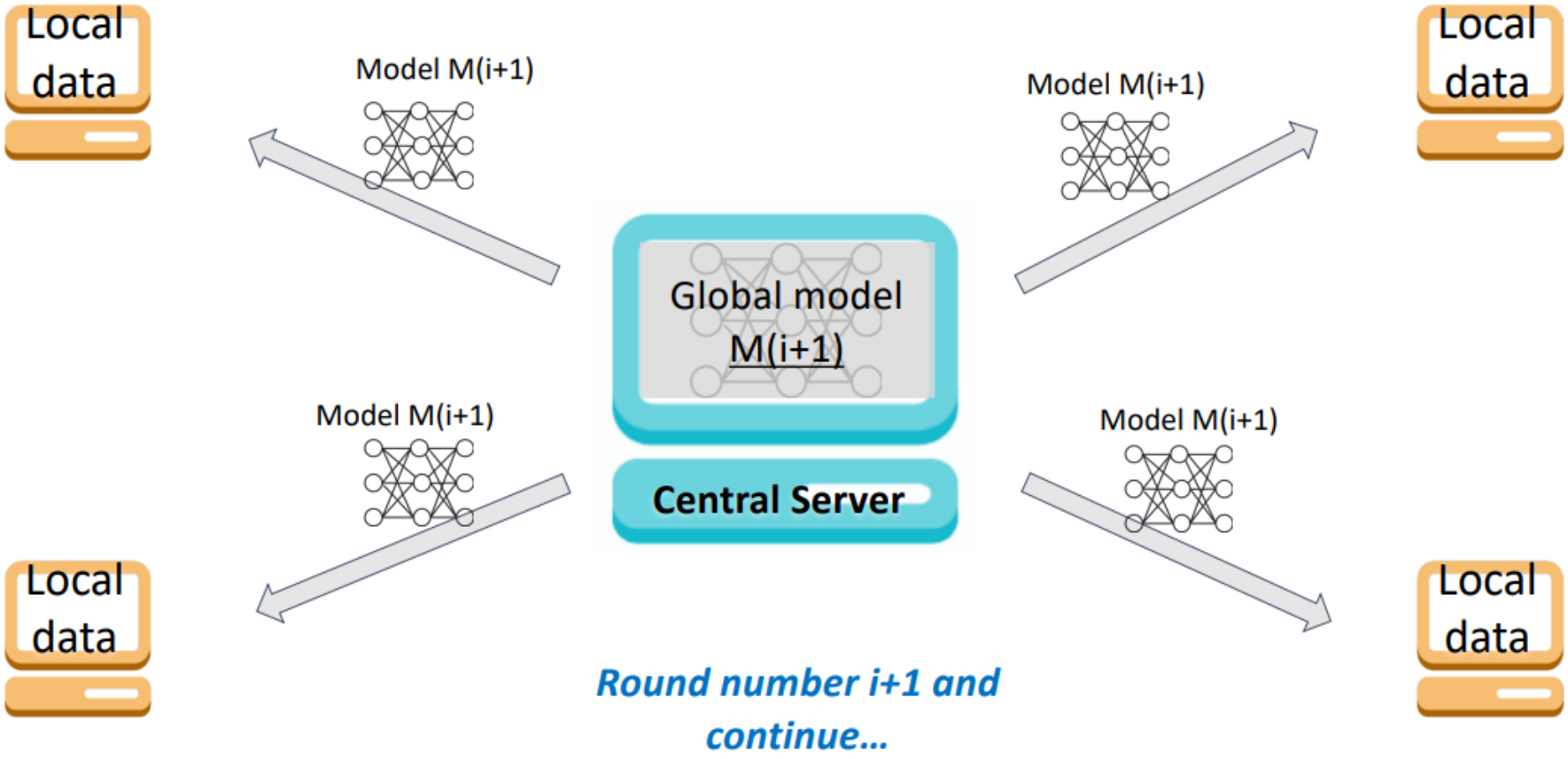
2. Each client performs local gradients decent steps



Federated Learning in Healthcare

Overview of Federated Learning steps:

3. Aggregate the gradients to form new model and send model to clients



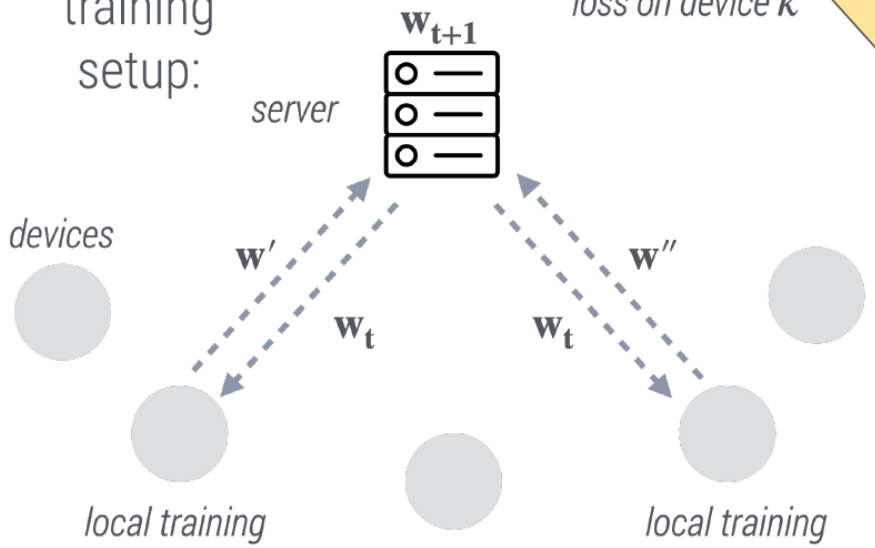
Federated Learning in Healthcare

Federated Optimization:

objective: $\min_w f(w) = \sum_{k=1}^{m} p_k F_k(w)$

loss on device k

training setup:



Typically consider solving an ERM objective, which is a (possibly) weighted average of losses across the m devices and their local data, i.e.,

$$\min_w \sum_{k=1}^{m} p_k \sum_{i=1}^{n_k} \ell(h(x_k^{(i)}; w), y_k^{(i)})$$

However, challenges discussed translate to other common ML objectives as well

m	number of devices
N	total number of data points
n_k	number of data points on device k
$(x_k^{(i)}, y_k^{(i)})$	i-th data point on device k
w	model parameters

Federated Learning in Healthcare

Federated Challenge:

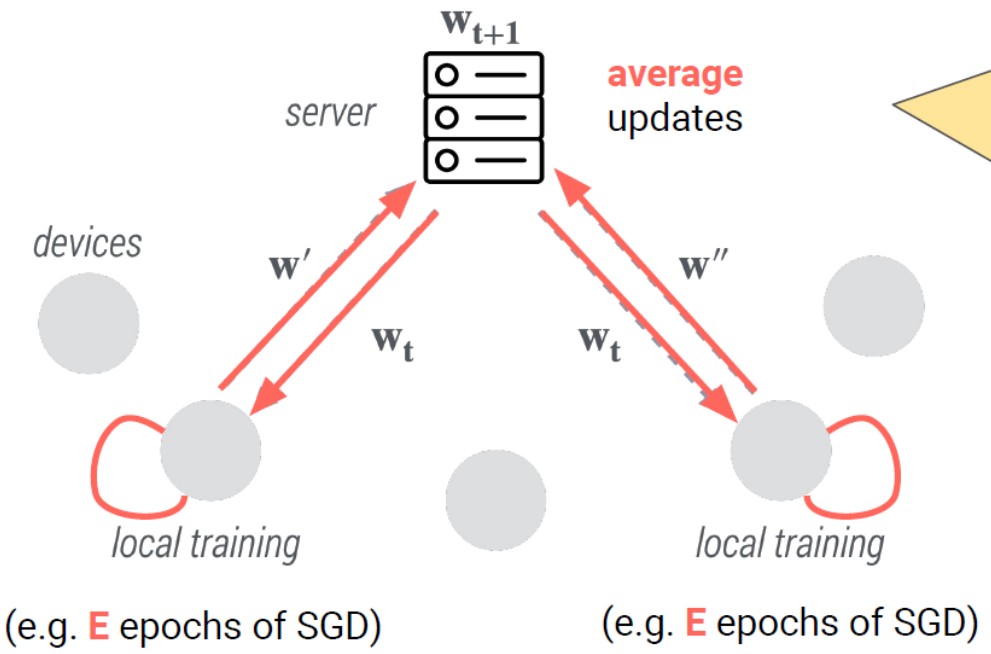
expensive communication
- massive, slow networks

privacy concerns
- user privacy constraints

- Expensive communication:
  1. Limiting number of devices
  2. Reducing number of communication round
  3. Reducing size of messages
- Privacy concerns:
  1. DP-SGD

Federated Learning in Healthcare

Federated Averaging (FedAvg):



- At each communication round:
  (i) run SGD locally, then
  (ii) average the model updates
- Can add privacy mechanisms to procedure (more later ...)
- Reduces communication by:
  (i) performing local updating,
  (ii) communicating with a subset of devices

Federated Learning in Healthcare

Federated Averaging (FedAvg):

Distributed SGD: computation on device k

for $i \in$ *mini-batch* B
  $\Delta \mathbf{w} \leftarrow \Delta \mathbf{w} - \alpha \nabla f_i(\mathbf{w})$
end
$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$

FedAvg: computation on device k

for $t = 1, 2, \ldots,$ *local iterations* T
  $\Delta \mathbf{w} \leftarrow \Delta \mathbf{w} - \alpha \nabla f_{i_t}(\mathbf{w})$
  $\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$
end

Why is it useful to perform `local-updating`?

1. Can perform more local computation (i.e., more than just one mini-batch)
2. Incorporate updates more quickly (immediately apply gradient information)

✓ **Can lead to method converging in many fewer communication rounds**

Data Imputation

Data Imputation

Sometimes we wish to share data with each other without compromising privacy. For example, pooling medical records between hospitals might be particularly useful when the data pertains to rare diseases.

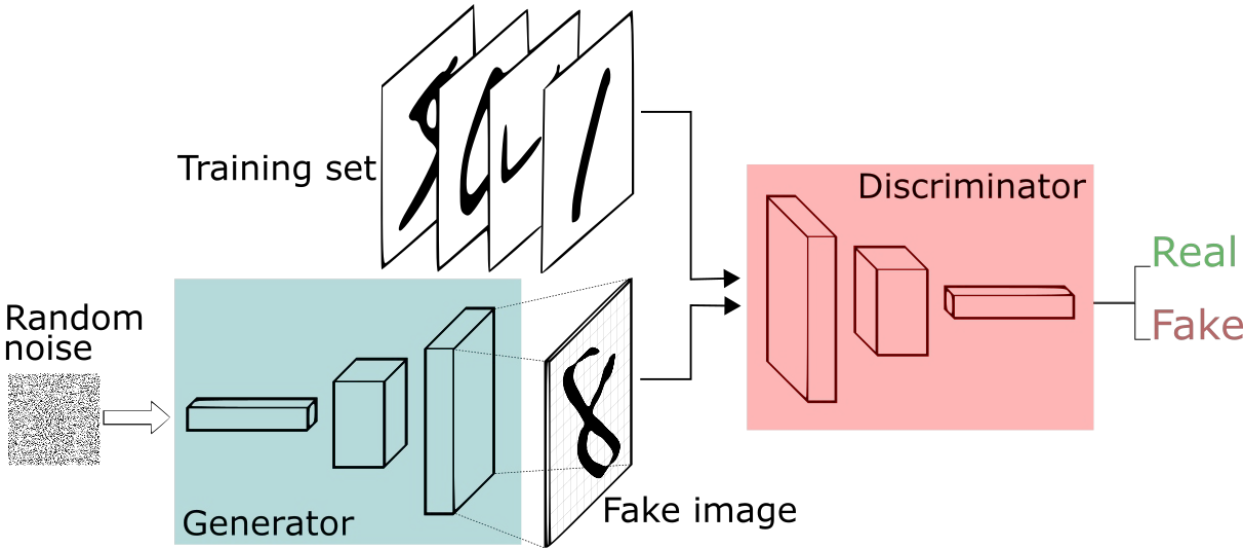
In such situations, organizations wish to **protect the privacy of the data without preventing their researchers from doing their jobs**.

The answer to this conundrum is not to release the data at all. Instead, we can **release synthetic data** which looks and acts like the real data, but which has been **generated in a differentially private manner**.

Data Imputation

Generative Adversarial Networks (GANs):

- This consists of two networks: a **generator**, $G(z;\theta_G)$ and a **discriminator** $D(x;\theta_D)$.
- The generator is a network with parameters θ_G that takes a sample of random noise z, and transforms this into a data point x.
- The discriminator is a second network with parameters D that takes a data point x and seeks to classify that data point as being either real data or fake (i.e., created by the generator).



Data Imputation

Training a GAN can be thought of as a two-player game in which the discriminator tries to identify which data points are fake, and the generator tries to generate data points which fool the discriminator. The players take turns updating their parameters by taking steps of stochastic gradient descent based on the current state of the other player. The generator takes a few steps of SGD trying to minimize

$$\max_{\theta_D}\left[\mathbb{E}_x[\log D[x;\theta_D]] - \mathbb{E}_z[\log D[G(z;\theta_G);\theta_D]]\right]$$

followed by the discriminator taking steps of SGD trying to maximize

$$\min_{\theta_G}\left[-\mathbb{E}_z[\log D[G(z;\theta_G);\theta_D]]\right]$$

Data Imputation

DP-GAN: Incorporate differential privacy in Generative Discriminator Network.

Algorithm 1 Differentially Private Generative Adversarial Nets

Require: α_d, learning rate of discriminator. α_g, learning rate of generator. c_p, parameter clip constant. m, batch size. M, total number of training data points in each discriminator iteration. n_d, number of discriminator iterations per generator iteration. n_g, generator iteration. σ_n, noise scale. c_g, bound on the gradient of Wasserstein distance with respect to weights

Ensure: Differentially private generator θ.

1: Initialize discriminator parameters w_0, generator parameters θ_0.
2: **for** $t_1 = 1, \ldots, n_g$ **do**
3: **for** $t_2 = 1, \ldots, n_d$ **do**
4: Sample $\{\mathbf{z}^{(i)}\}_{i=1}^{m} \sim p(\mathbf{z})$ a batch of prior samples.
5: Sample $\{\mathbf{x}^{(i)}\}_{i=1}^{m} \sim p_{data}(\mathbf{x})$ a batch of real data points.
6: For each i, $g_w(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) \leftarrow \nabla_w \left[f_w(\mathbf{x}^{(i)}) - f_w(g_\theta(\mathbf{z}^{(i)})) \right]$
7: $\bar{g}_w \leftarrow \frac{1}{m}(\sum_{i=1}^{m} g_w(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) + N(0, \sigma_n^2 c_g^2 I))$.
8: $w^{(t_2+1)} \leftarrow w^{(t_2)} + \alpha_d \cdot RMSProp(w^{(t_2)}, \bar{g}_w)$
9: $w^{(t_2+1)} \leftarrow clip(w^{(t_2+1)}, -c_p, c_p)$
10: **end for**
11: Sample $\{\mathbf{z}^{(i)}\}_{i=1}^{m} \sim p(\mathbf{z})$, another batch of prior samples.
12: $g_\theta \leftarrow -\nabla_\theta \frac{1}{m} \sum_{i=1}^{m} f_w(g_\theta(\mathbf{z}^{(i)}))$
13: $\theta^{(t_1+1)} \leftarrow \theta^{(t_1)} - \alpha_g \cdot RMSProp(\theta^{(t_1)}, g_\theta)$
14: **end for**
15: **return** θ.