

Privacy Preserving Machine Learning

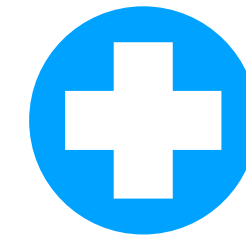
28 March 2023 @ TrainRDM

Tanja Šarčević and Anastasia Pustozero

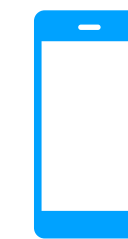
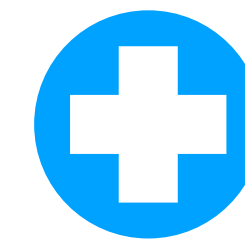
PhD students at TU Wien



Data is used to improve many services



Data is used to improve many services



Support the Guardian
Available for everyone, funded by readers
[Contribute](#) [Subscribe](#)

Search jobs Sign in Search International edition

The Guardian

For 200 years

News Opinion Sport Culture Lifestyle More

The Cambridge Analytica Files
Cambridge Analytica

Patrick Greenfield
@pgreenfielduk
Mon 26 Mar 2018 00:53 BST

The Cambridge Analytica files: the story so far

Google Is Fined \$57 Million Under Europe's Data Privacy Law

By Adam Satariano

Facebook Data Breach -- What To Do Next

Kate O'Flaherty Senior Contributor
Cybersecurity
Straight Talking Cyber

Twitter faces \$250m fine over new data privacy issue

Social media giant accused of personal information misuse for tailored advertising

© Tue, Aug 4, 2020, 11:32

Kate Conger

phone displaying Facebook

ssive data breach affecting
n place three days earlier,

Data collection challenges

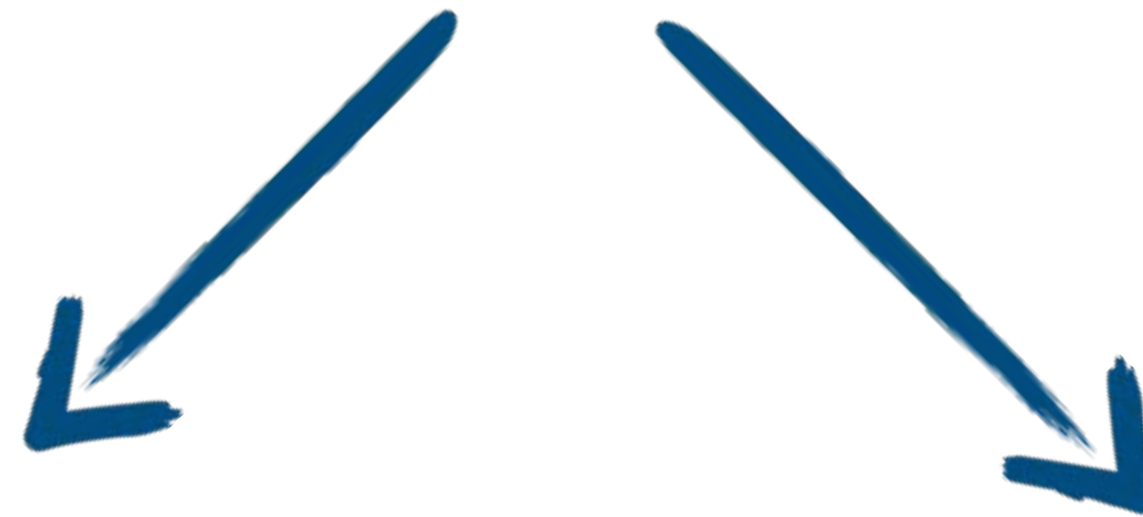
- Data usually contains private or sensitive information
- Collecting private data brings a lot of bureaucracy and legal responsibilities
- Leakage or unauthorised access to the data poses threats to the data owners



Privacy-Preserving Data Analysis

Privacy-preserving **Data Publishing**

k -anonymity
l -diversity
t -closeness
Differential privacy
Synthetic data



Privacy-preserving **Computation**

Homomorphic encryption
Secure multi-party computation
Federated Learning

Pseudonymisation

- A state of disguised identity
- Pseudonym identifies a holder, that is, one or more human beings who possess but do not disclose their true names (legal identities)
- It enables a consolidation of a persons' data without revealing identities
 - Data can also mean books, paintings, etc...

ID	Name	Date of birth	City of residence
1	William Smith	1/2/73	Berkeley, California
2	Anna Williams	23/8/79	Berkeley, CA
ID	Pseudonym	Date of birth	City of residence
1	John Doe	1/2/73	Berkeley, California
2	Jane Doe	23/8/79	Berkeley, CA

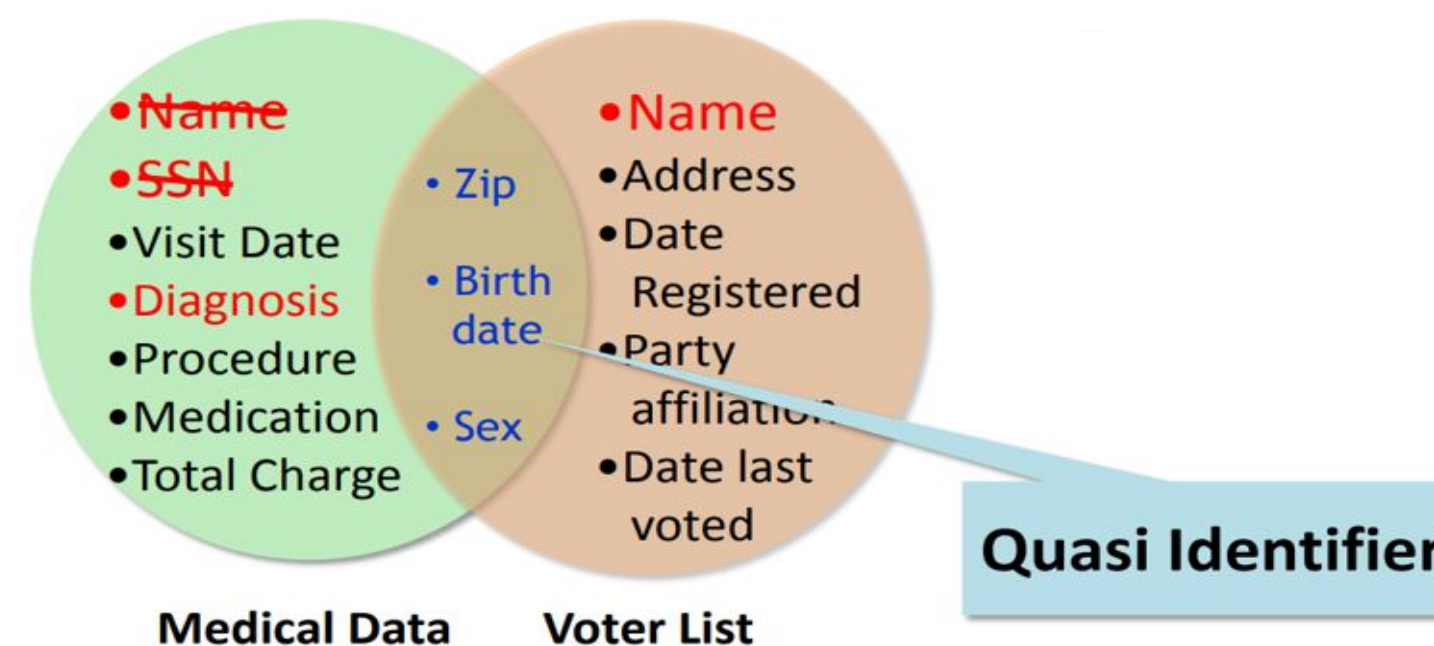
Pseudonymisation

- GDPR:
 - “...*personal data ... that can no longer be attributed to a specific data subject without the use of additional information*”
 - pseudonymized personal data **remain** personal data nonetheless, provided the controller **or** another party has the means to **reverse** the process
- Thus the same principles for storing, processing, sharing, etc. still apply!

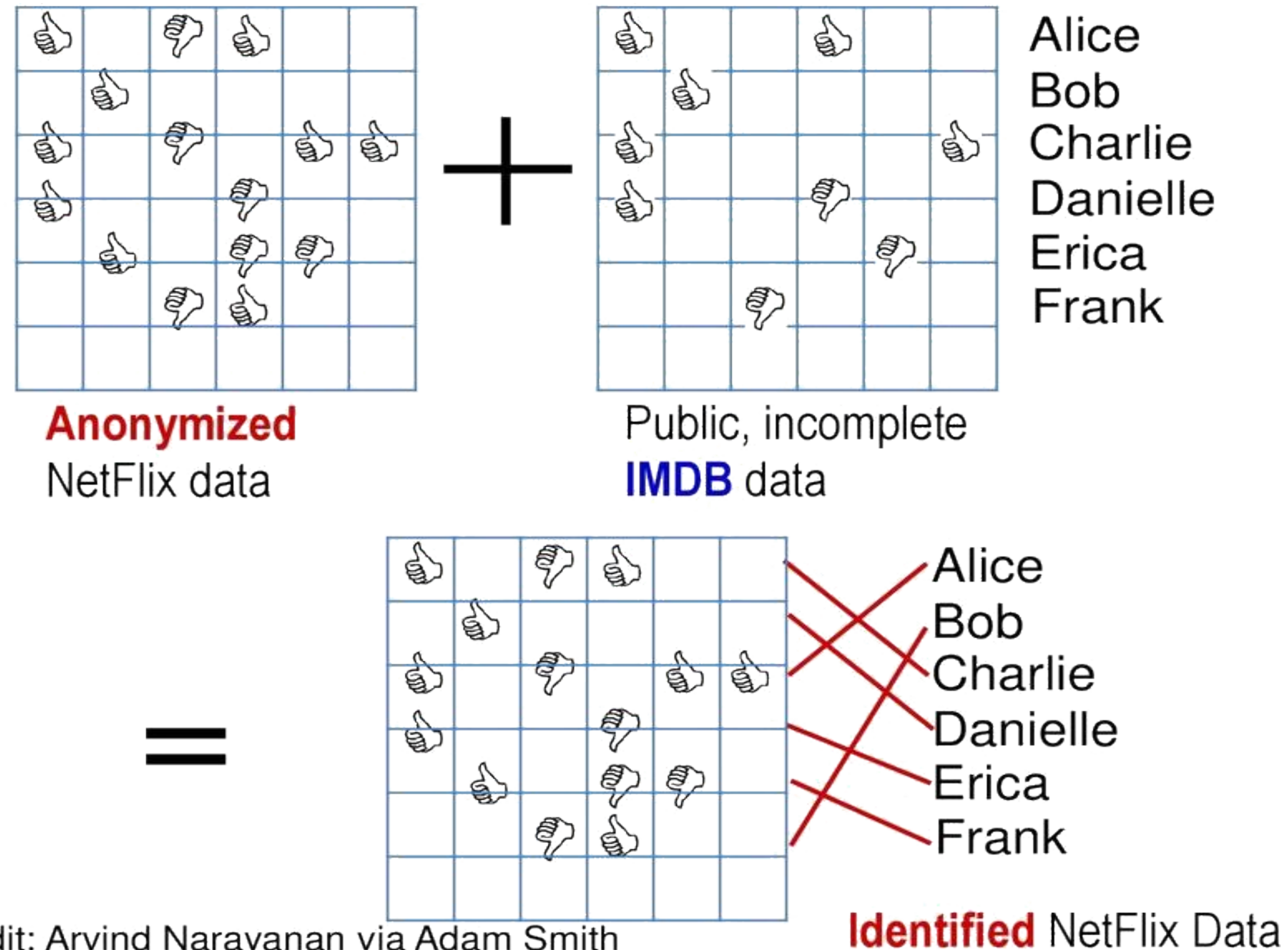


Data Sanitisation

- **Pseudonymisation:** remove directly identifying information
 - *That is often **not** enough!*
- Massachusetts Health records of public employees
 - With the birthdate, ZIP Code, gender: Governor of Massachusetts William Weld uniquely identified
 - Linkage attack with public voting records
 - Other examples:
 - Netflix prize (2006),
 - AOL data,



Record Linkage attacks



Data Sanitisation

- **Anonymisation of microdata:** sanitize also quasi-identifiers (QI)
 - Those attributes that can identify when used in combination
 - Birthdate, ZIP Code, gender, occupation, ...
- *Issues?*
 - List is not complete
 - Case-dependent
 - Adversary's background knowledge!
 - Dependent on the available other data (present **AND** future!)
- Anonymised data is not subject to GDPR regulations anymore!



Data Sanitisation: HIPAA



- The Privacy Rule of the US *Health Insurance Portability and Accountability Act* of 1996 (HIPAA) establishes comprehensive protections for medical privacy (*revised & came into effect 2002*)
- Protected health information (PHI) is “identifiable” health information acquired in the course of serving patients
 - One of the few authoritative sources that lists identifiable attributes
- Sanitisation standard before data sharing in medical domains (research and professional)

Data Sanitisation: 18 HIPAA Identifiers

- Names
- All geographic subdivisions smaller than a State
- All elements of dates (except year)
- Telephone numbers
- Fax numbers
- Electronic mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images and any comparable images
- Any other unique identifying number, characteristic, or code



Data anonymisation

k -anonymity

- Ensures that at least k records have same QI
- Or: cardinality of any query result on released data should be at least k
- Achieved via
 - Generalisation of values (exact age to a range of values, ...)
 - Suppression of values
 - Microaggregation

	QI ₁	QI ₂	S ₁
ID	Age	Zip	Disease
1	5	15	Flu
2	15	25	Fever
3	28	28	Diarrhea
4	25	15	Fever
5	22	28	Flu
6	32	35	Fever
7	38	32	Flu
8	35	25	Diarrhea

	QI ₁	QI ₂	S ₁
ID	Age	Zip	Disease
1	0-20	10-30	Flu
2	0-20	10-30	Fever
3	20-30	10-30	Diarrhea
4	20-30	10-30	Fever
5	20-30	10-30	Flu
6	30-40	20-40	Fever
7	30-40	20-40	Flu
8	30-40	20-40	Diarrhea

Samarati, Pierangela, and Latanya Sweeney. "Protecting privacy when disclosing information: k -anonymity and its enforcement through generalization and suppression." (1998).

Data anonymisation

k -anonymity

- Ensures that at least k records have same QI
- Or: cardinality of any query result on released data should be at least k
- Achieved via
 - Generalisation of values (exact age to a range of values, ...)
 - Suppression of values
 - Microaggregation

Birth day	gender	Zipcode
21/1/79	M	53715
10/1/79	F	55410
1/10/44	F	90210
21/2/83	M	02274
19/4/82	M	02237



	Birth day	gender	Zipcode
group 1	* /1/79	human	5****
	* /1/79	human	5****
suppress	1/10/44	F	90210
group 2	* /*/8*	M	022**
	* /*/8*	M	022**

Samarati, Pierangela, and Latanya Sweeney. "Protecting privacy when disclosing information: k -anonymity and its enforcement through generalization and suppression." (1998).

Data anonymisation

k -anonymity

- Ensures that at least k records have same QI
- Or: cardinality of any query result on released data should be at least k
- Achieved via
 - Generalisation of values (exact age to a range of values, ...)
 - Suppression of values
 - Microaggregation

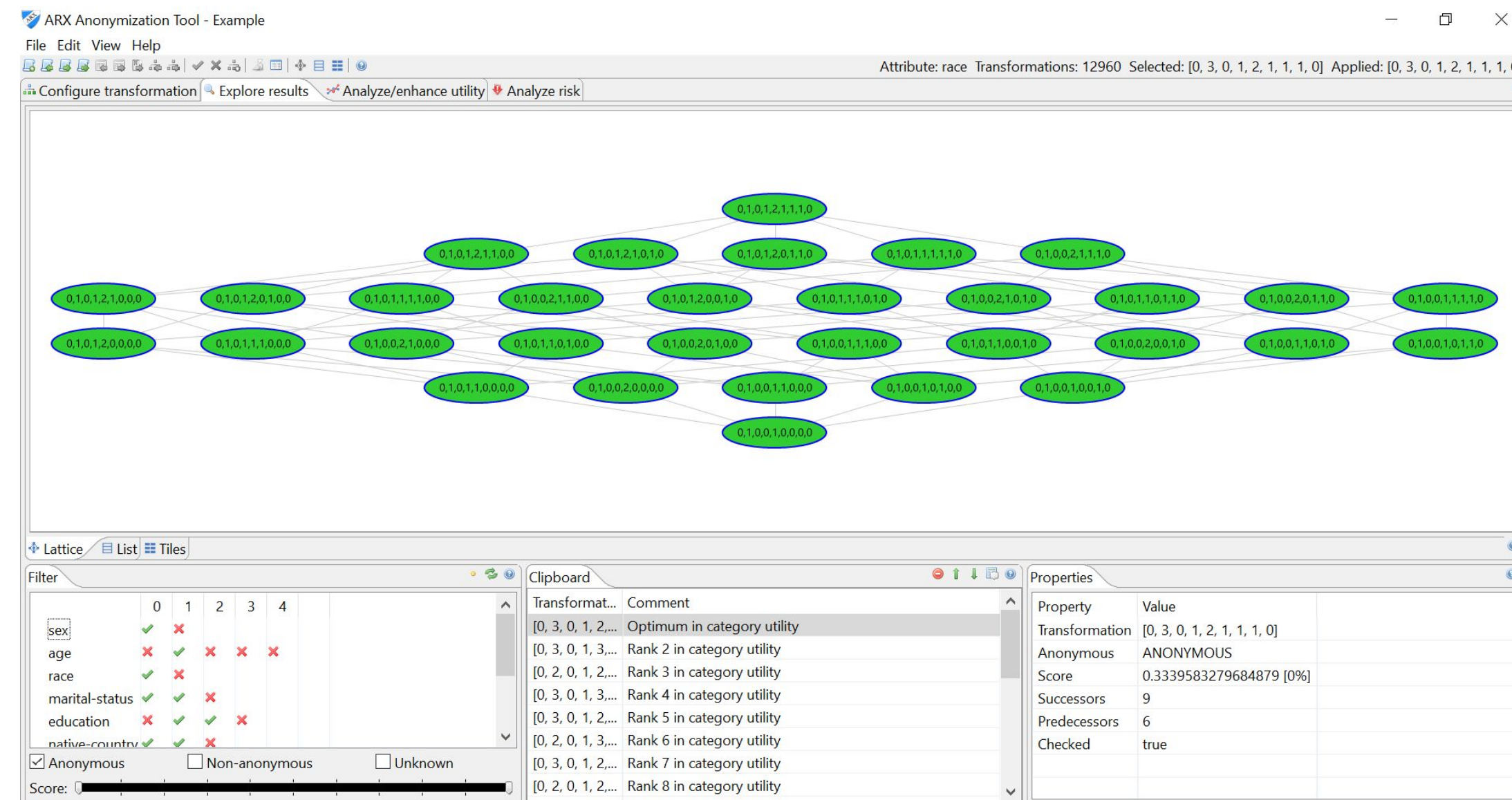


Height	Weight	High Cholesterol
5'4"	158	N
5'3"	162	Y
5'6"	161	N
5'8"	157	N

Height	Weight	High Cholesterol
5'5"	160	N
5'5"	160	Y
5'5"	160	N
6'0"	155	N

Solving k -anonymity: Tools

- ARX:
 - Flash algorithm
 - <https://arx.deidentifier.org/>
- Amnesia:
 - <https://amnesia.openaire.eu/>
- UTD Anonymization Toolbox:
 - Datafly, Incognito, Mondrian
 - <http://www.cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php>
- Microaggregation tool:
 - https://github.com/CrisesUrv/microaggregation-based_anonymization_tool



k -anonymity: multiple solutions

Record	Name	SSN	Age	Location	Sex	Race	Diagnosis	Income
r ₁	Alice	123456789	32	San Diego	M	W	AIDS	17,000
r ₂	Bob	323232323	30	Los Angeles	M	W	Asthma	68,000
r ₃	Charley	232345656	42	Wichita	M	W	Asthma	80,000
r ₄	Dave	333333333	30	Kansas City	M	W	Asthma	55,000
r ₅	Eva	666666666	35	Lincoln	F	W	Diabetes	23,000
r ₆	John	214365879	20	Lincoln	M	B	Asthma	55,000
r ₇	Casey	909090909	25	Wichita	F	B	Diabetes	23,000



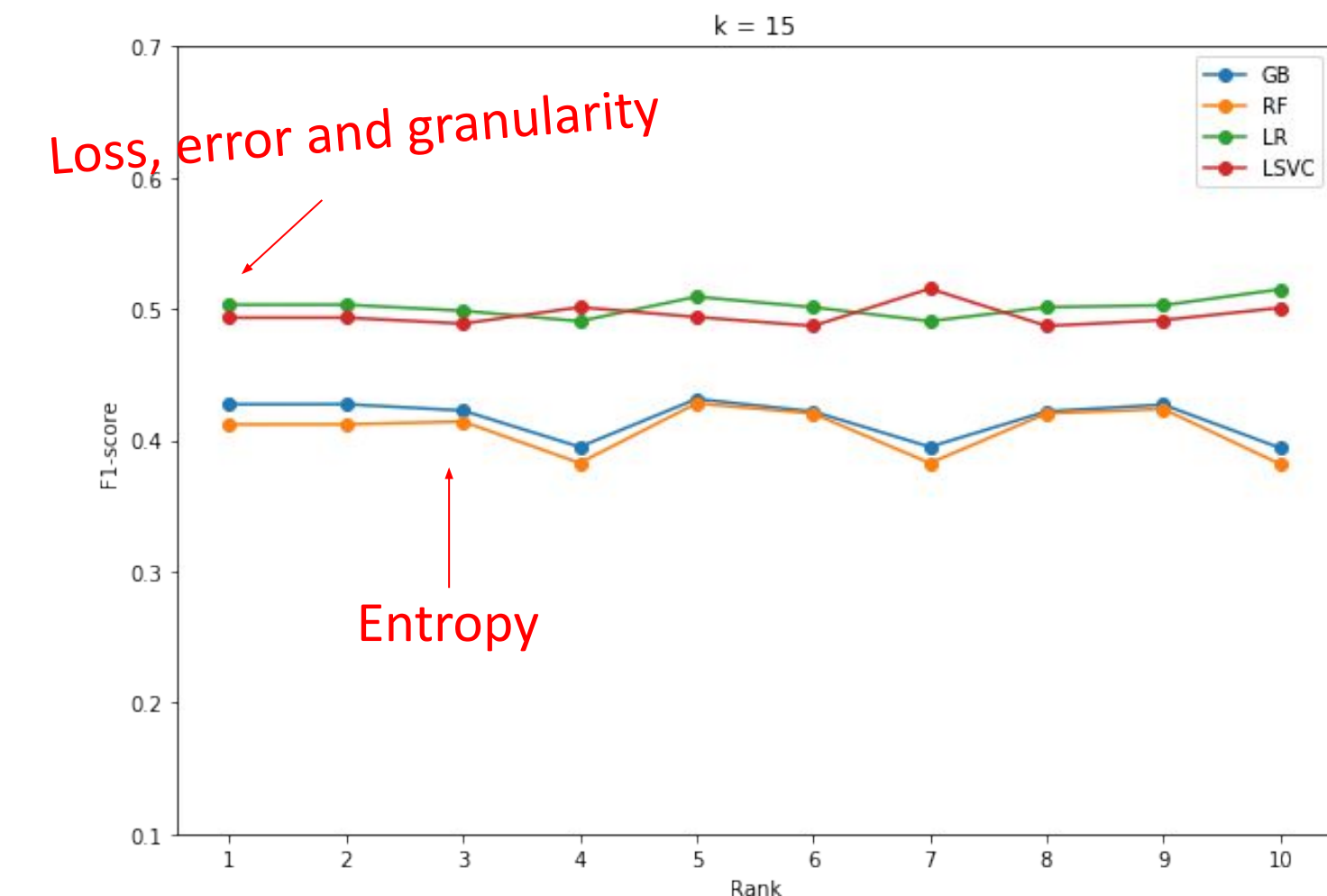
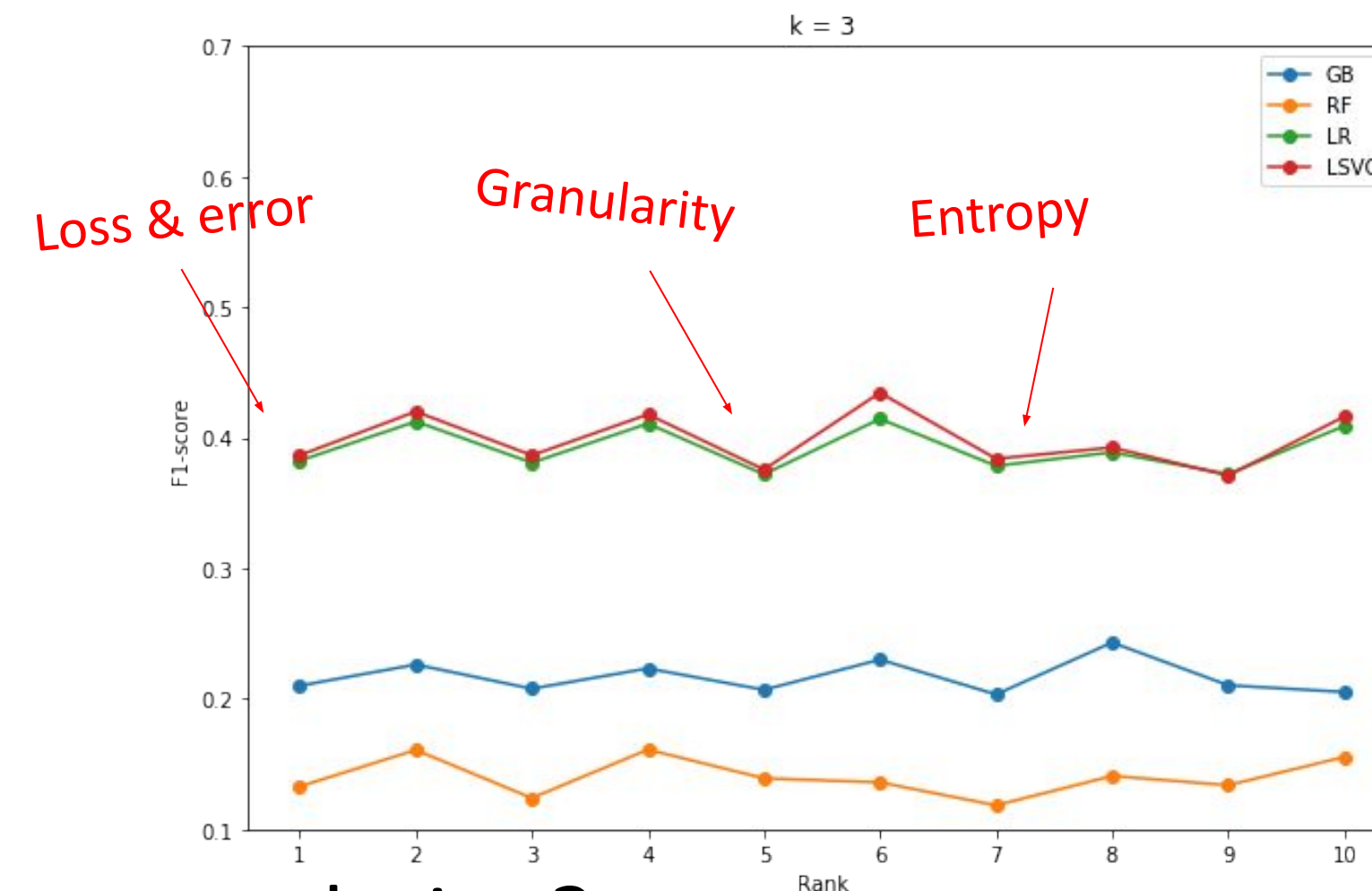
Record	Age	Location	Sex	Race
r ₁	30-32	California	M	W
r ₂	30-32	California	M	W
r ₃	30-42	MidWest	*	W
r ₄	30-42	MidWest	*	W
r ₅	30-42	MidWest	*	W
r ₆	20-25	MidWest	*	B
r ₇	20-25	MidWest	*	B

Record	Age	Location	Sex	Race
r ₁	30-32	California	M	W
r ₂	30-32	California	M	W
r ₃	25-42	Kansas	*	*
r ₄	25-42	Kansas	*	*
r ₇	25-42	Kansas	*	*
r ₅	20-35	Lincoln	*	*
r ₆	20-35	Lincoln	*	*

Data anonymisation

Utility of k -anonymous data

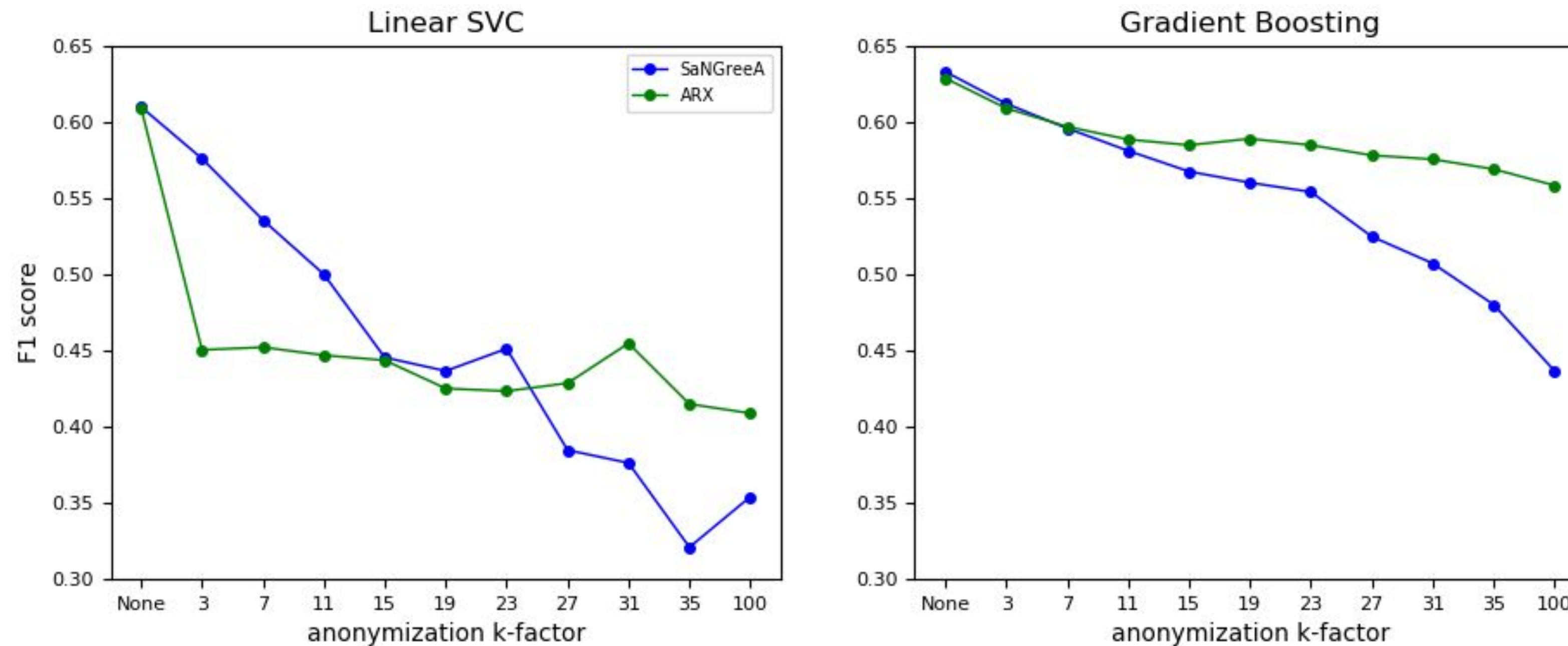
- Should preserve as much *information* as possible
- Notions of utility of anonymised data:
 - Precision
 - Entropy
 - Loss
 - Error
 - ML performance
 -
- What's the best anonymous solution?
 - Stability of solutions
 - Other metrics usually indicate good model performance



Data anonymisation

Utility of k -anonymous data

- Increasing the level of anonymity k , the information loss also increases



- Local (SaNGreeA) vs Global (Flash/ARX) transformation * Experiments on Adult dataset (target: education-num)

Challenges

- Privacy level k – no standard
- Utility of the resulting data analysis decreased
- Does not necessarily ensure privacy
 - Expert inspection of the anonymised data before publishing
 - Additional sanitisation methods need to be applied
 - l -diversity
 - t -closeness

Attacks Against K-Anonymity

- k-Anonymity does not provide **privacy** if
 - Sensitive values in an equivalence class lack diversity
 - The attacker has background knowledge

Homogeneity Attack

Bob

Zipcode	Age
47678	27

A 3-anonymous patient table

Zipcode	Age	Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
4790*	≥40	Flu
4790*	≥40	Heart Disease
4790*	≥40	Bladder Cancer
476**	3*	Heart Disease
476**	3*	Bladder Cancer
476**	3*	Bladder Cancer

L-diversity principles

- Each equivalence class has at least l well-represented sensitive values

Bob

Zipcode	Age
47678	27

?

A 3-anonymous patient table

Zipcode	Age	Disease
476**	20-40	Heart Disease
476**	20-40	Heart Disease
476**	20-40	Bladder Cancer
4790*	≥40	Flu
4790*	≥40	Heart Disease
4790*	≥40	Bladder Cancer
476**	20-40	Heart Disease
476**	20-40	Heart Disease
476**	20-40	Bladder Cancer

Limitations of l-Diversity

- l-diversity is insufficient to prevent attribute disclosure

Bob

Zip

Age

47678

27

Similarity Attack

], which is relatively low

Zipcode	Age	Salary	Disease
476**	2*	3K	Gastric Ulcer
476**	2*	4K	Gastritis
476**	2*	5K	Stomach Cancer
4790*	≥40	6K	Gastritis
4790*	≥40	11K	Flu
4790*	≥40	8K	Bronchitis
476**	3*	7K	Bronchitis
476**	3*	9K	Pneumonia
476**	3*	10K	Stomach Cancer

- Conclusions:
 - Bob's salary is in [3k,5k], which is relatively low
 - Bob has some stomach-related disease
- l-diversity does not consider semantic meanings of sensitive values

t-closeness

Bob	
Zip	Age
47678	27

?

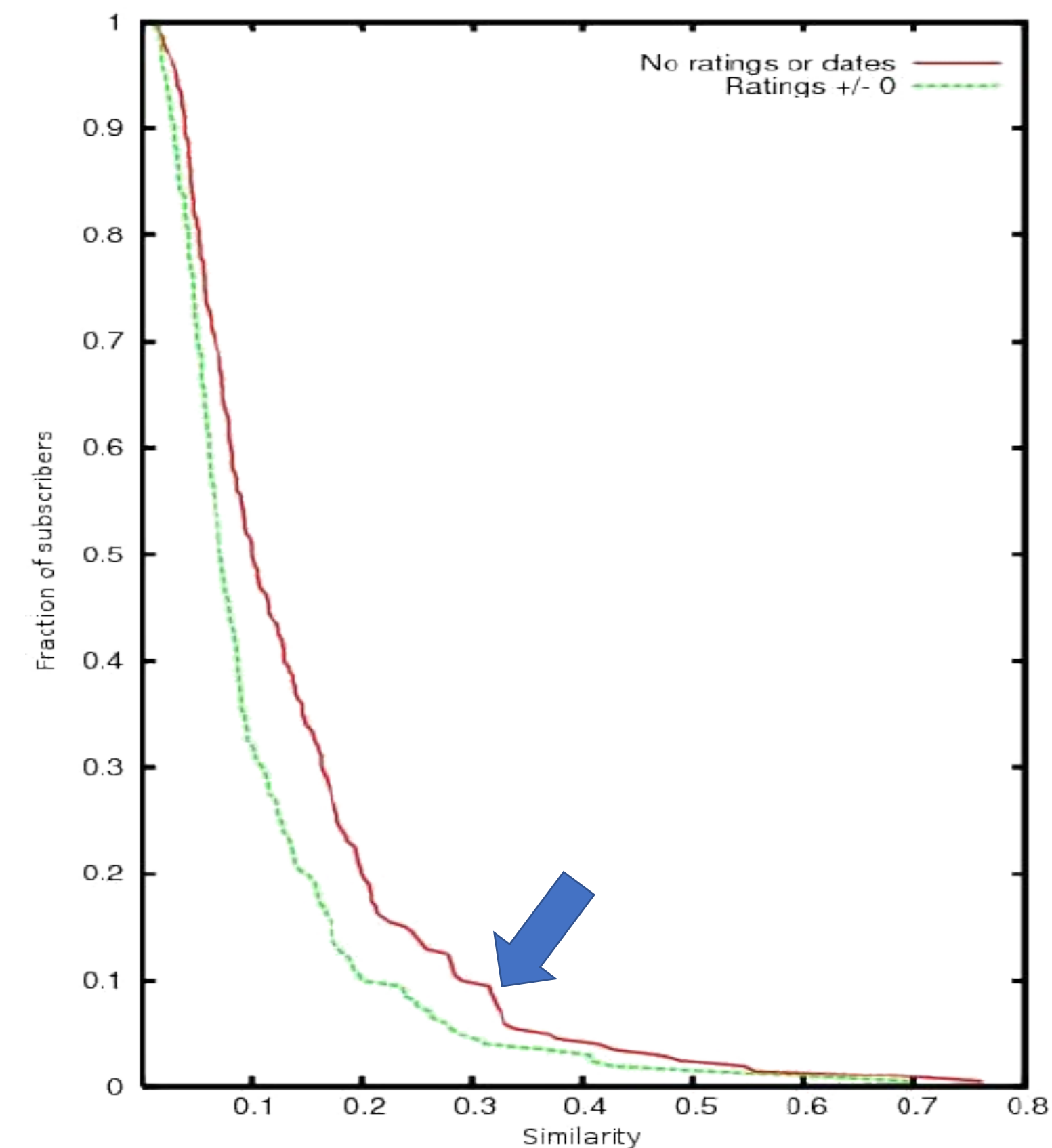
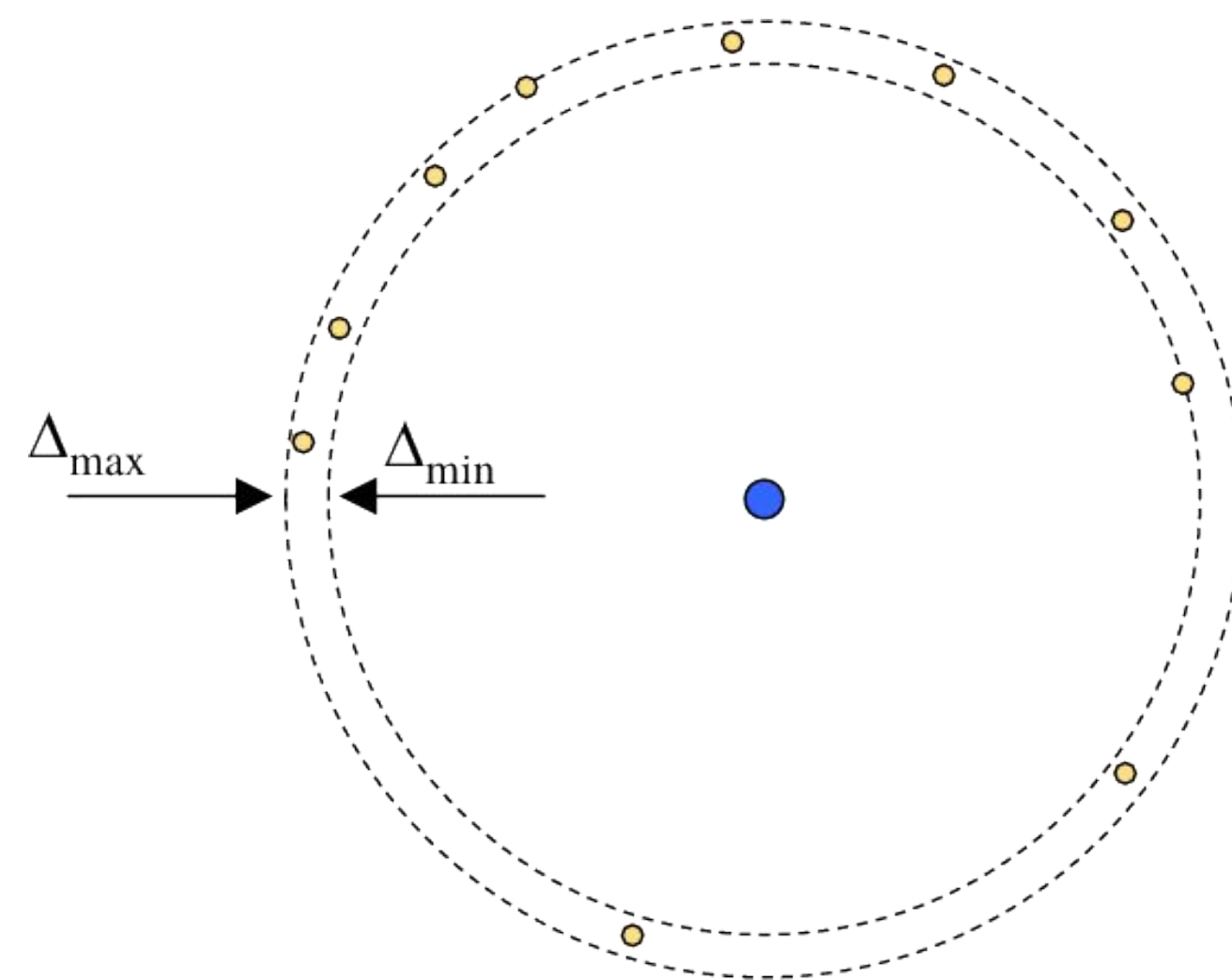
Similarity Attack

Zipcode	Age	Salary	Disease
476**	<40	3K	Gastric Ulcer
476**	<40	9K	Pneumonia
476**	<40	5K	Stomach Cancer
4790*	≥40	6K	Gastritis
4790*	≥40	11K	Flu
4790*	≥40	8K	Bronchitis
476**	<40	7K	Bronchitis
476**	<40	4K	Gastritis
476**	<40	10K	Stomach Cancer

- **Privacy = information gain of an observer**
- Distribution of the sensitive attribute in each equivalence class should be *similar* to distribution of the sensitive attribute in the whole table

Anonymisation: other limitations

- In very high-dimensional spaces data matrices often get very sparse
 - Only a few items are actually similar to each other



Ownership protection in data publishing

Data ownership protection

- Why protecting the data?
 - Data owner used a lot of resources to collect/create the data (money, human experts, time...)
 - Sensitive data (e.g. medical data) needs to be shared with researchers
 - Privacy implications: only the trusted parties get the data and should not share it further
- Threats of sharing the data:
 - Data theft / unauthorised usage
 - Privacy implications for sensitive data (e.g. medical data): only the trusted parties get the data and should not share it further
- The goal: controlled data sharing
 - Share full data
 - Trace the unauthorised data re-distribution



Controlled data sharing

Watermarking & Fingerprinting

- Embedding owner's signature into the data

Perceptible



vs.

Imperceptible



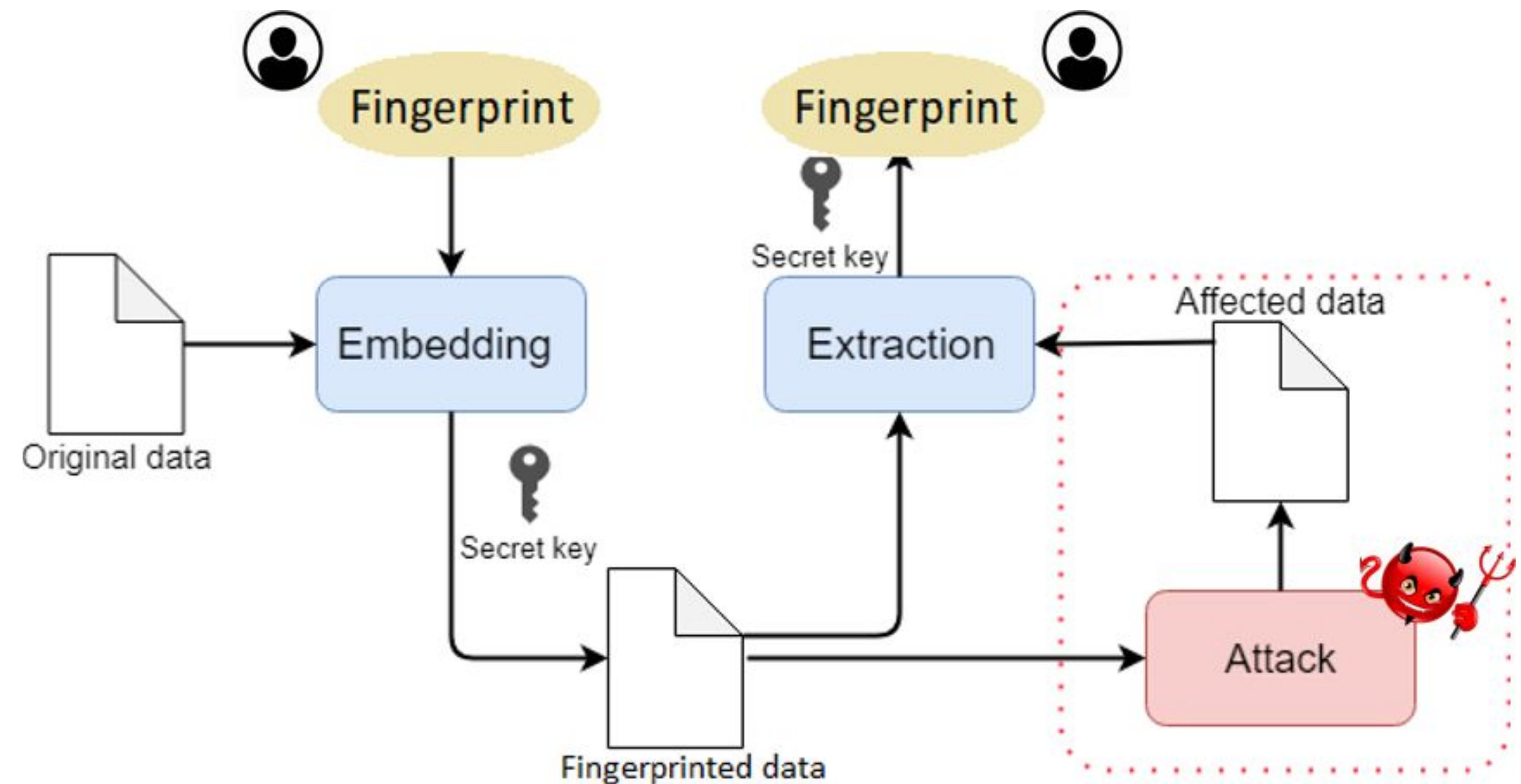
Age	Blood Pressure	Diabetes
32	64	1
31	66	0
50	72	1
48	70	0

Age	Blood Pressure	Diabetes
33	64	1
31	68	0
50	72	1
47	70	0

Requirements & workflow

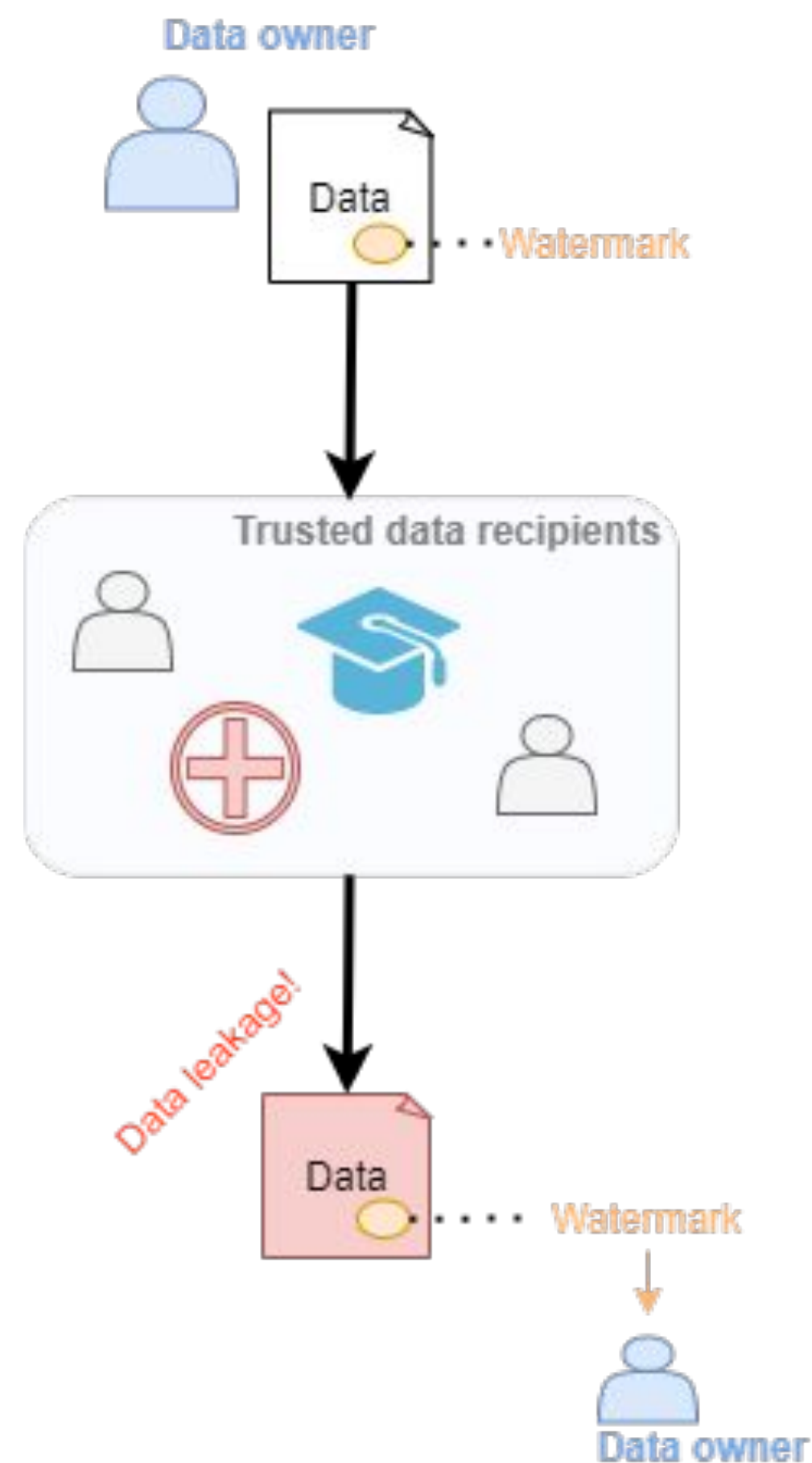
Requirements:

- Recognisable by the owner
- Not detectable and consequently removable by the recipients
- Robust to the attacks
- Does not change the utility of the data too much

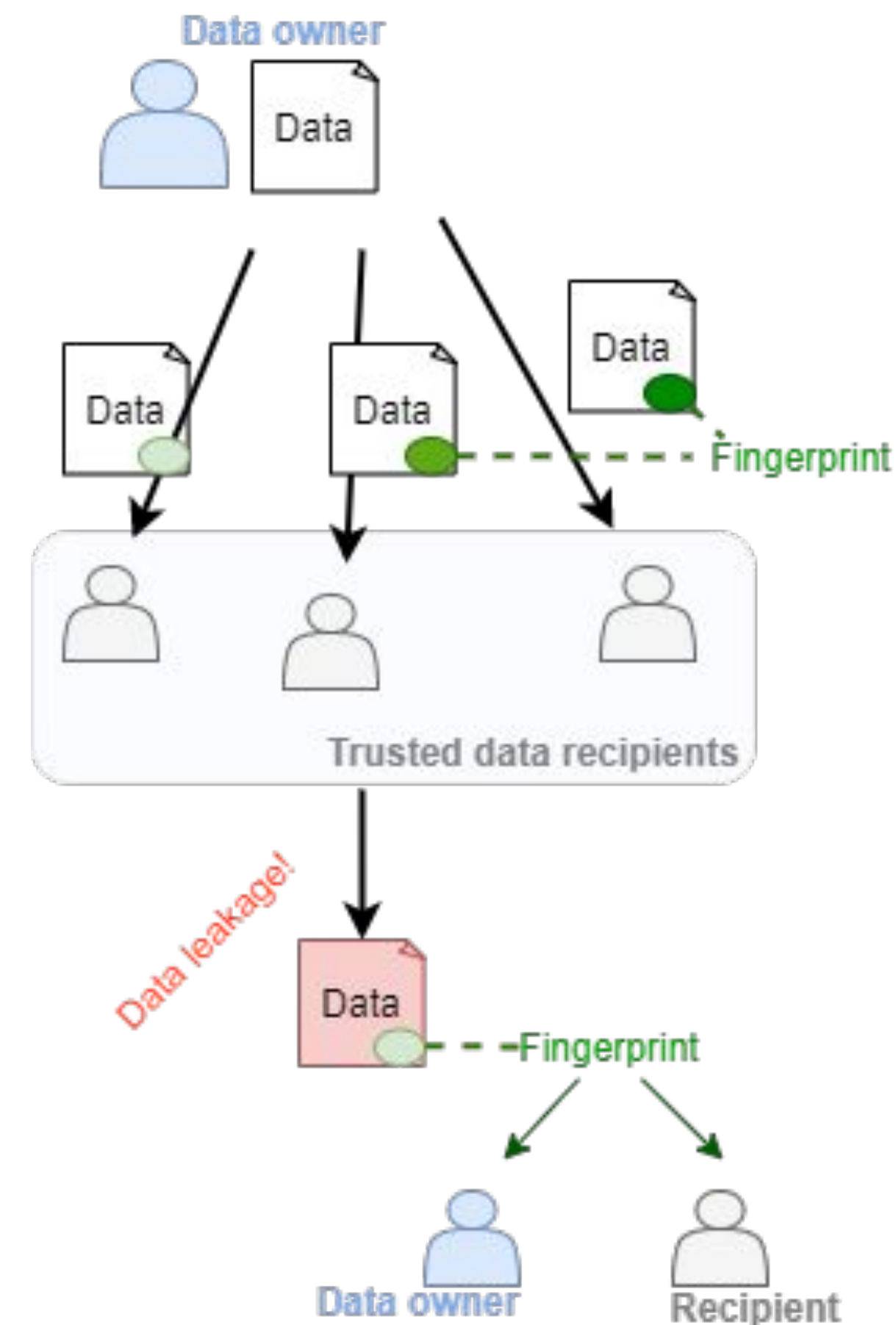


Watermark vs. Fingerprint

- **Watermark**: owner ID

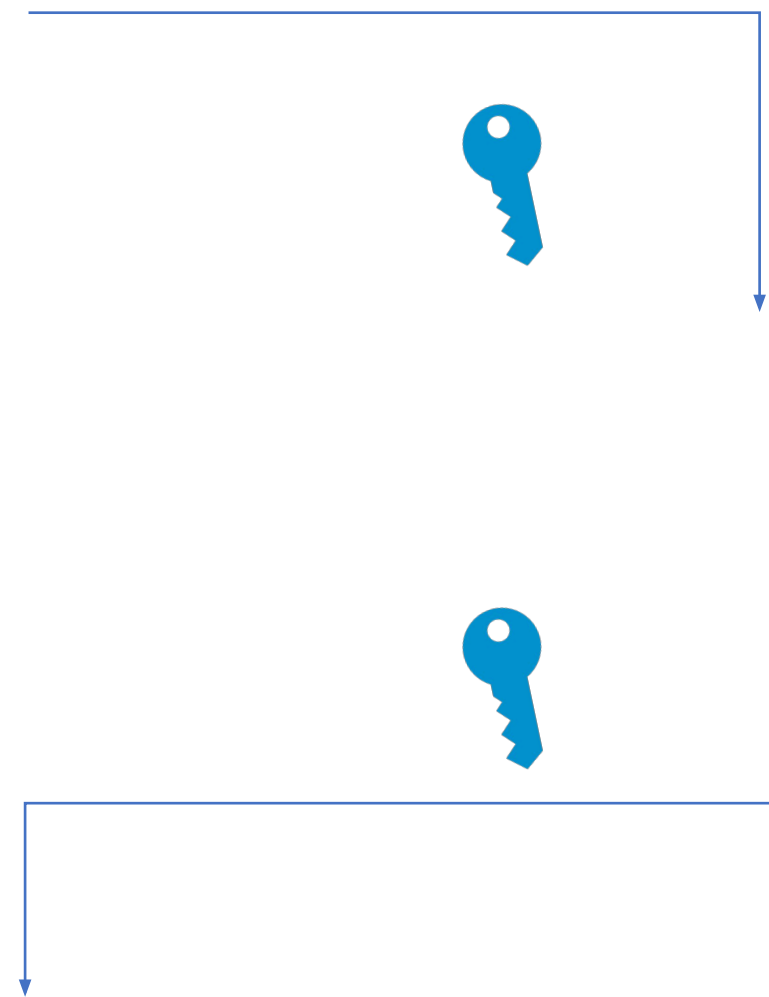


- **Fingerprint**: owner ID + recipient ID



Watermarking & fingerprinting data

1 0 1 0



1 0 1 0

Age	Blood Pressure	Diabetes
32	64	1
31	66	0
50	72	1
48	70	0
31	65	0
75	77	1
67	68	0
53	68	0
60	81	1
51	76	1
39	84	0



Li Y, Swarup V, Jajodia S. Fingerprinting relational databases: Schemes and specialties. IEEE Transactions on Dependable and Secure Computing. 2005 Apr 11;2(1):34-45.

Watermarking & fingerprinting data

1 0 1 0



1 0 1 0

Age	Blood Pressure	Diabetes
30	64	1
31	66	1
50	72	1
48	71	0
31	65	0
75	77	1
67	68	0
52	68	0
60	81	1
51	77	1
38	84	0



Li Y, Swarup V, Jajodia S. Fingerprinting relational databases: Schemes and specialties. IEEE Transactions on Dependable and Secure Computing. 2005 Apr 11;2(1):34-45.

Watermarking & fingerprinting data

1 0 1 0



1 0 1 0

Age	Blood Pressure	Diabetes
30	64	1
31	66	1
50	72	1
48	71	0
31	65	0
75	77	1
67	68	0
52	68	0
60	81	1
51	77	1
38	84	0
44	75	0
51	69	1
72	70	0
65	68	0
81	80	0

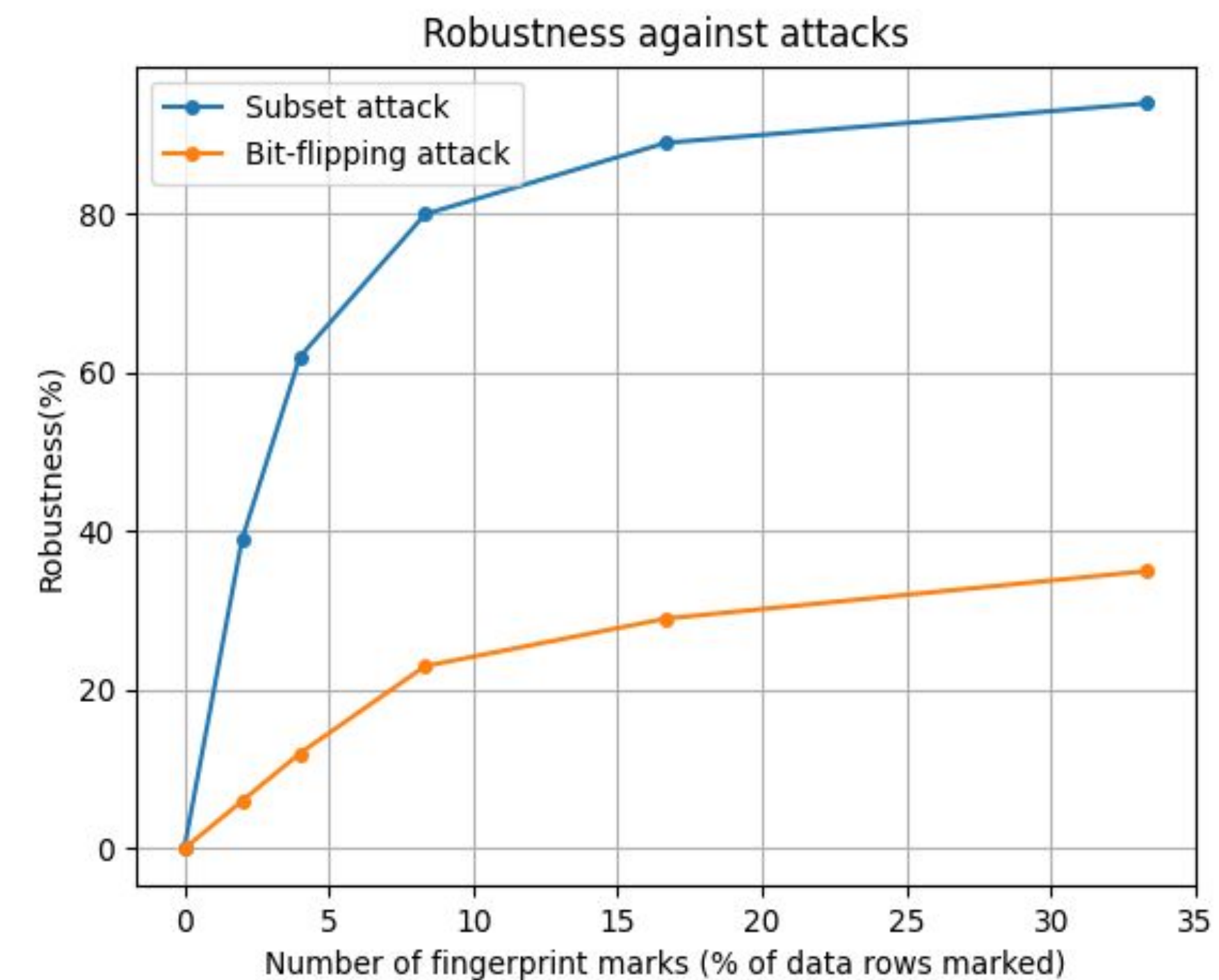
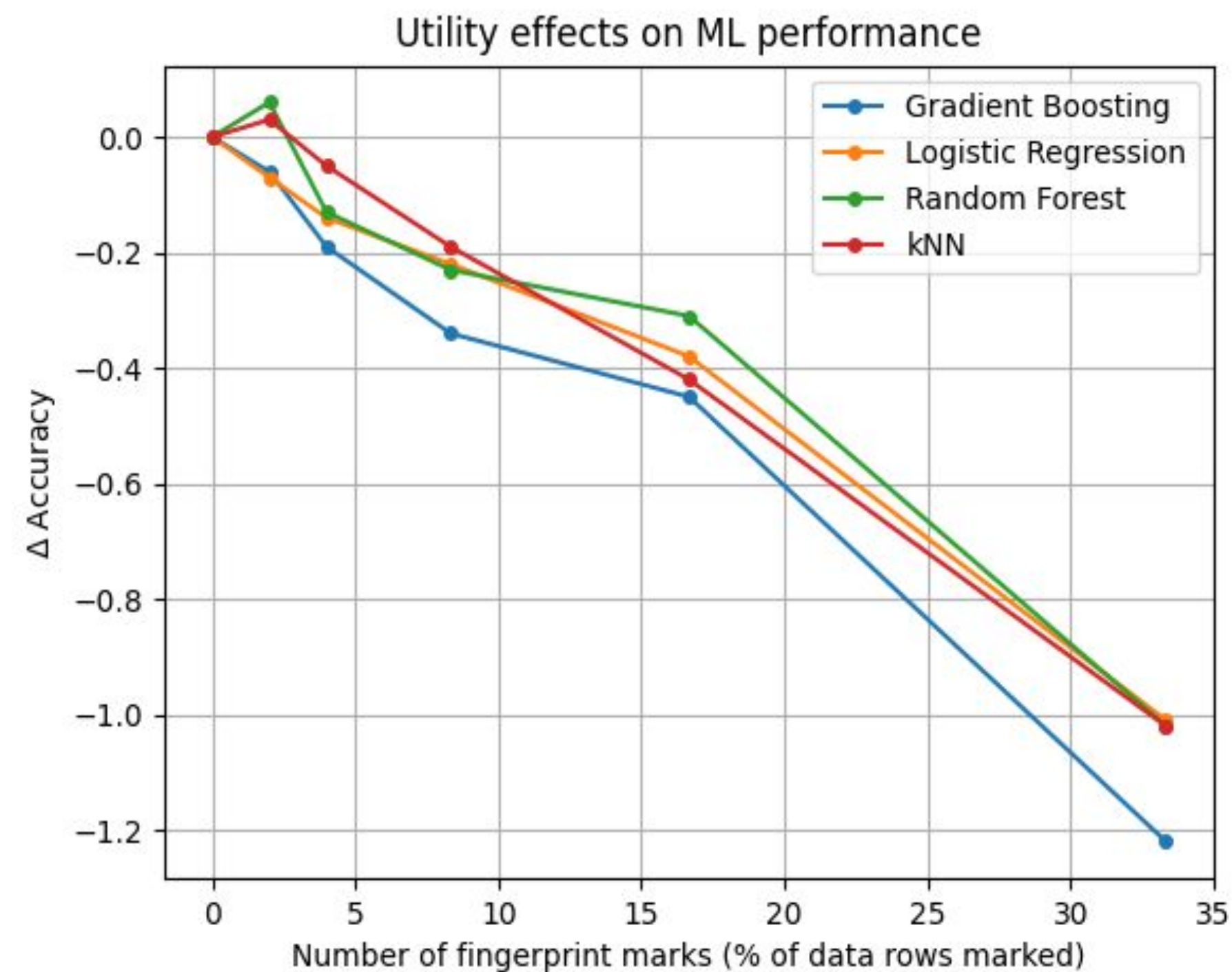
Trade-off:

Robustness vs. data utility

Li Y, Swarup V, Jajodia S. Fingerprinting relational databases: Schemes and specialties. IEEE Transactions on Dependable and Secure Computing. 2005 Apr 11;2(1):34-45.

Trade-off: Robustness vs. utility

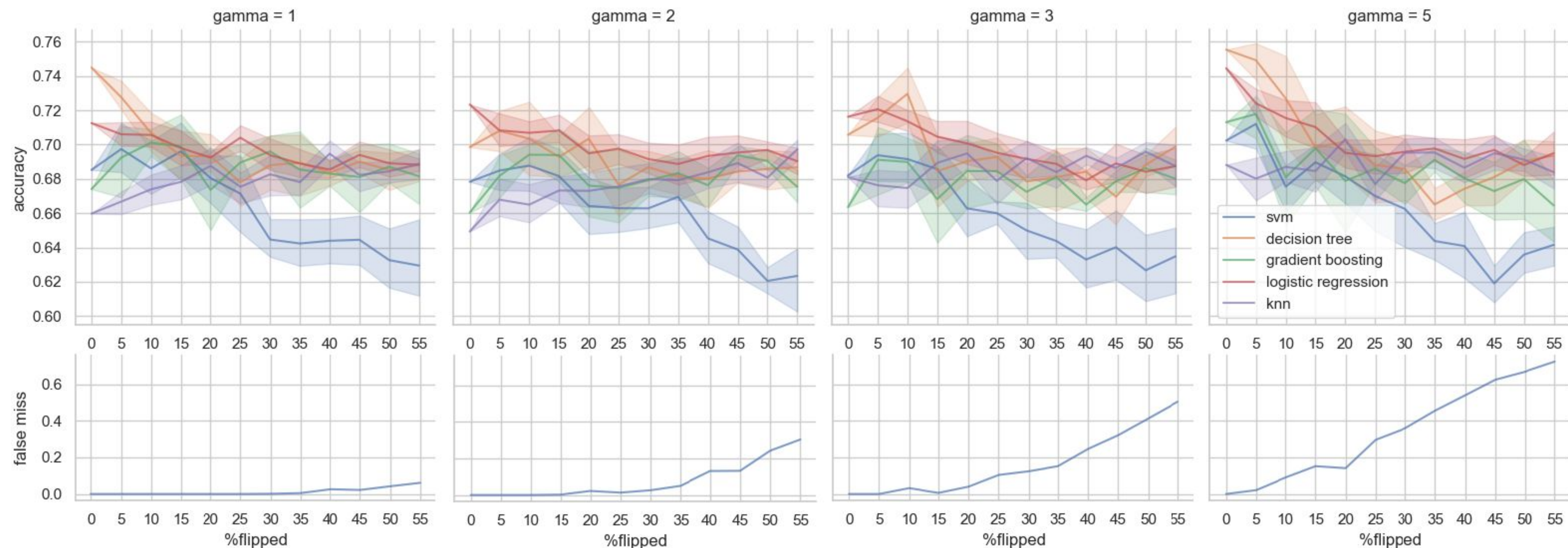
- Robustness against attacks → **maximise** modifications
- Preserve data utility → **minimise** modifications



Fingerprinting relational data

Attacker loss

- Attacking is no free lunch for the attacker



Šarčević, Tanja, Rudolf Mayer, and Andreas Rauber. "Adaptive Attacks and Targeted Fingerprinting of Relational Data." *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022.

Fingerprinting relational data

Challenges

- *Minor alterations in categorical data?*

age	sex	employed	Alzheimers stage
59	Male	No	late
67	Male	No	late
...	...	...	...



age	sex	employed	Alzheimers stage
59	Male	Yes	late
68	Female	No	late
...	...	...	...

- *kNN-based fingerprinting: „modify the value to something that is likely to occur in the original dataset „*

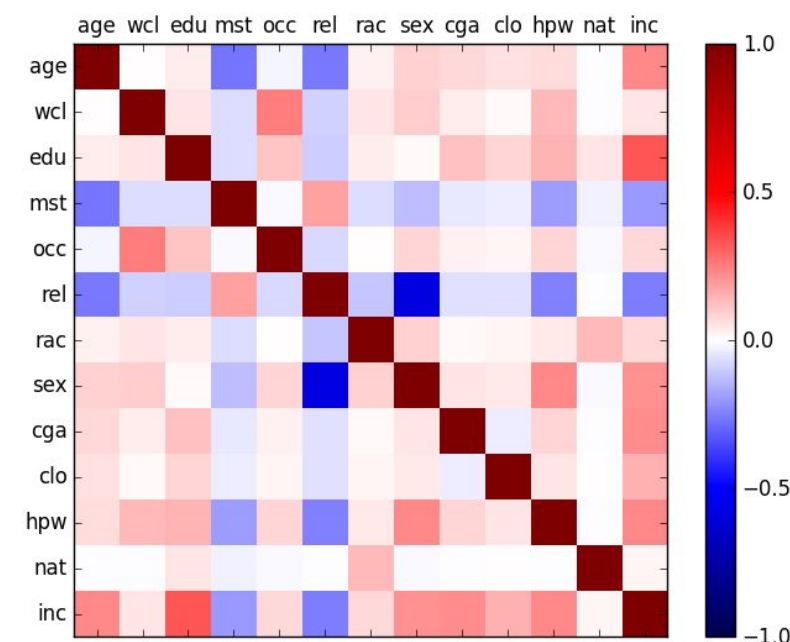
Sarcevic, Tanja, and Rudolf Mayer. "A Correlation-Preserving Fingerprinting Technique for Categorical Data in Relational Databases." *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, Cham, 2020.

<https://github.com/tanjascats/fingerprinting-toolbox>, version 0.1.0, accessed 2022-03-25

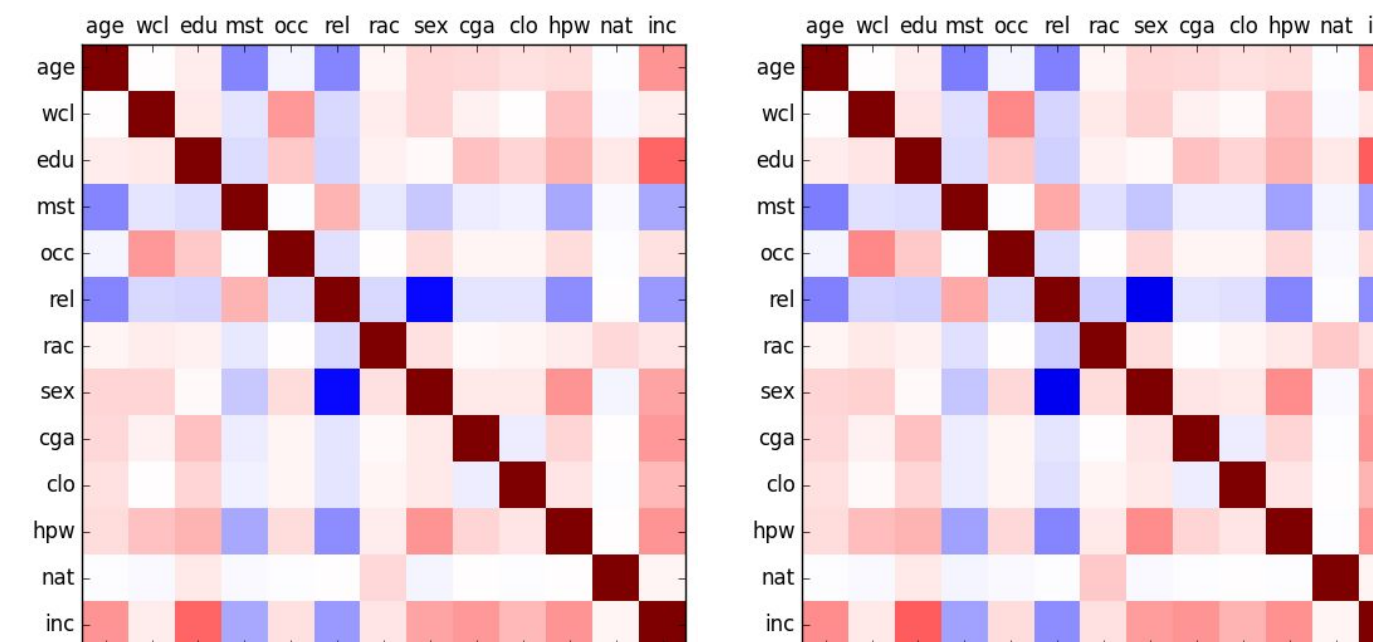
Fingerprinting relational data

Challenges

- Different data types: categorical vs. numerical



a) Original correlations



b) Correlations in fingerprinted datasets

Fig: Effects of the fingerprint on statistical properties (Adult data)

		NB	SVM	KNN	RF	LR
Adult	Original	79.7%	81.0%	81.5%	82.2%	82.3%
	$\gamma = 1$	79.7%	79.9%	80.6%	80.4%	82.0%
	$\gamma = 2$	80.3%	80.3%	81.6%	80.3%	82.1%

Fig: Effects of the fingerprint on the Machine Learning task

Sarcevic, Tanja, and Rudolf Mayer. "A Correlation-Preserving Fingerprinting Technique for Categorical Data in Relational Databases." *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, Cham, 2020.

<https://github.com/tanjascats/fingerprinting-toolbox>, version 0.1.0, accessed 2022-03-25

Data fingerprinting: summary

- Reactive ownership protection method
 - Verification of the owner
 - Trace of the source of unauthorised usage
 - Trade-off: robustness vs. data utility
- 
- Can be applied in many domains
 - Multimedia, tabular data, software, ML models (our recent research 😊)

Privacy Preserving Computation

Privacy-Preserving Data Analysis

Privacy-preserving **Data Publishing**

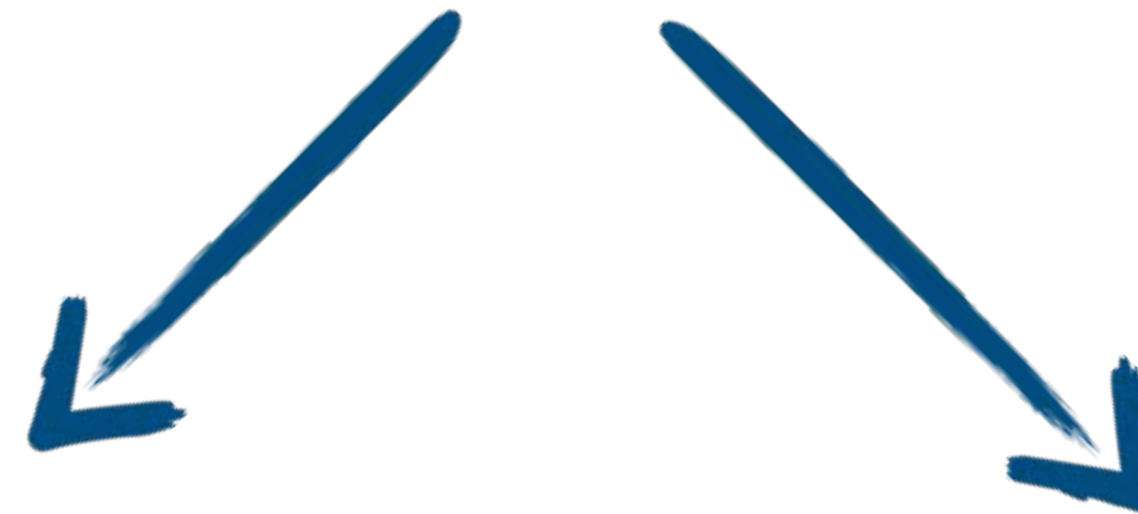
k-anonymity

l-diversity

t-closeness

Differential privacy

Synthetic data



Privacy-preserving **Computation**

Homomorphic encryption

Secure multi-party
computation

Federated Learning

Homomorphic encryption

- is a form of [encryption](#) that permits users to **perform computations on its encrypted data** without first decrypting it.
- Allows to securely perform computation or store the data at the third party (e.g. cloud provider).



Homomorphic encryption

Example

Input		[3, 7, 12]
Encryption	shift input by +2	[5, 9, 14]
Task (performed at the third party)	compute sum	28
Decrypt	shift the output back	$28 - (2 \cdot 3) = 22$

Homomorphic encryption

Issues:

- Supports relatively small number of operations: sum and multiplication
- *Extremely* computational expensive

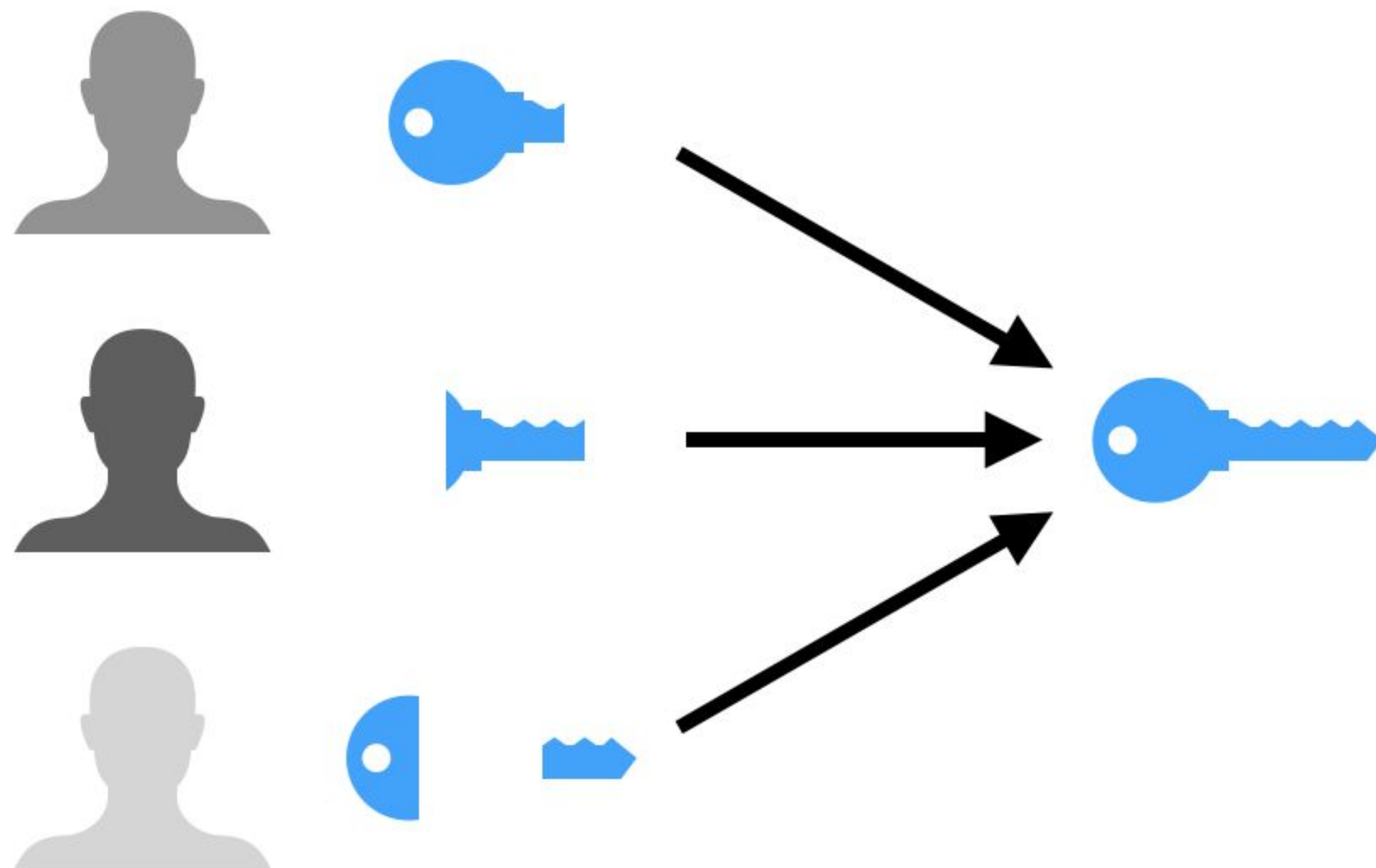
Secure multi-party computation (SMPC)

is a cryptographic protocol allowing several parties (data owners) to jointly compute a public function over their private data, when no individual party can see the other parties' data.

- SMPC is based on **secret sharing**

Secure multi-party computation (SMPC)

Secret sharing allows one to distribute a secret among several parties by distributing shares to each party, only combining all shares together it is possible to compute a final function.



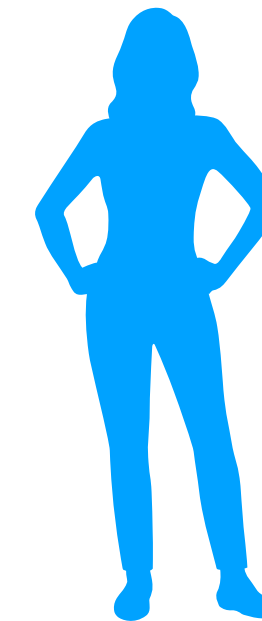
Example:

11 scientists are working on a secret project and lock up the documents in the cabinet which can be open only if at least 7 scientists are present.

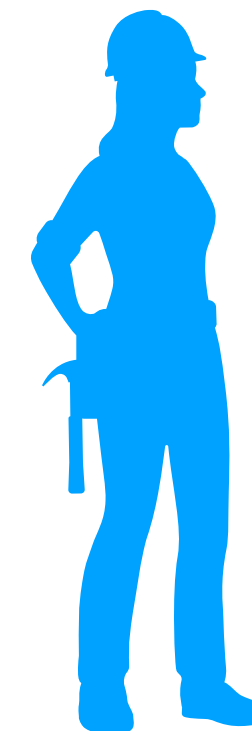
The secret is split to 11 parts in a way that any 7 of this parts can reveal the secret.

Secure multi-party computation (SMPC)

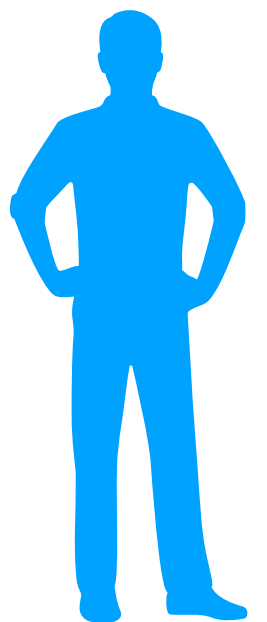
Example: Compute the average salary.



A: 3100



C: 2800



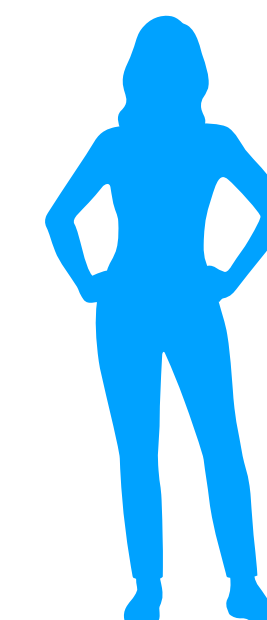
B: 2500

Secure multi-party computation (SMPC)

Example: Compute the average salary.

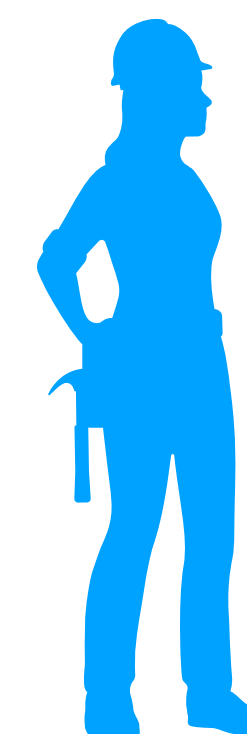
Each participant:

1. Splits their salary to 3 parts, randomly



AA: 1100
AB: 500
AC: 1500

A: 3100



CA: 1000
CB: 400
CC: 1400

C: 2800



BA: 750
BB: 900
BC: 850

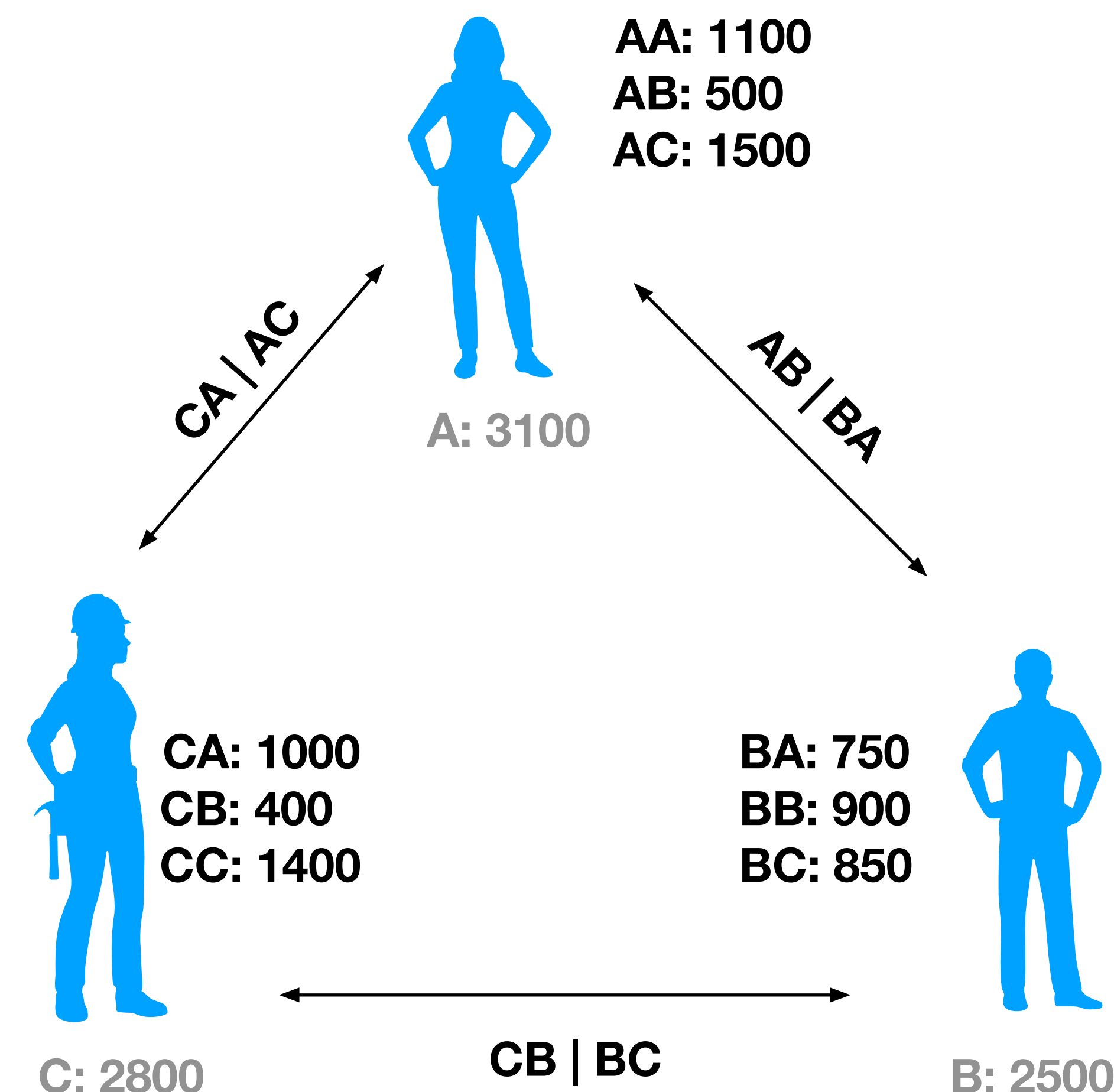
B: 2500

Secure multi-party computation (SMPC)

Example: Compute the average salary.

Each participant:

1. Splits their salary to 3 parts, randomly
2. Shares with each of the other participants one of the parts

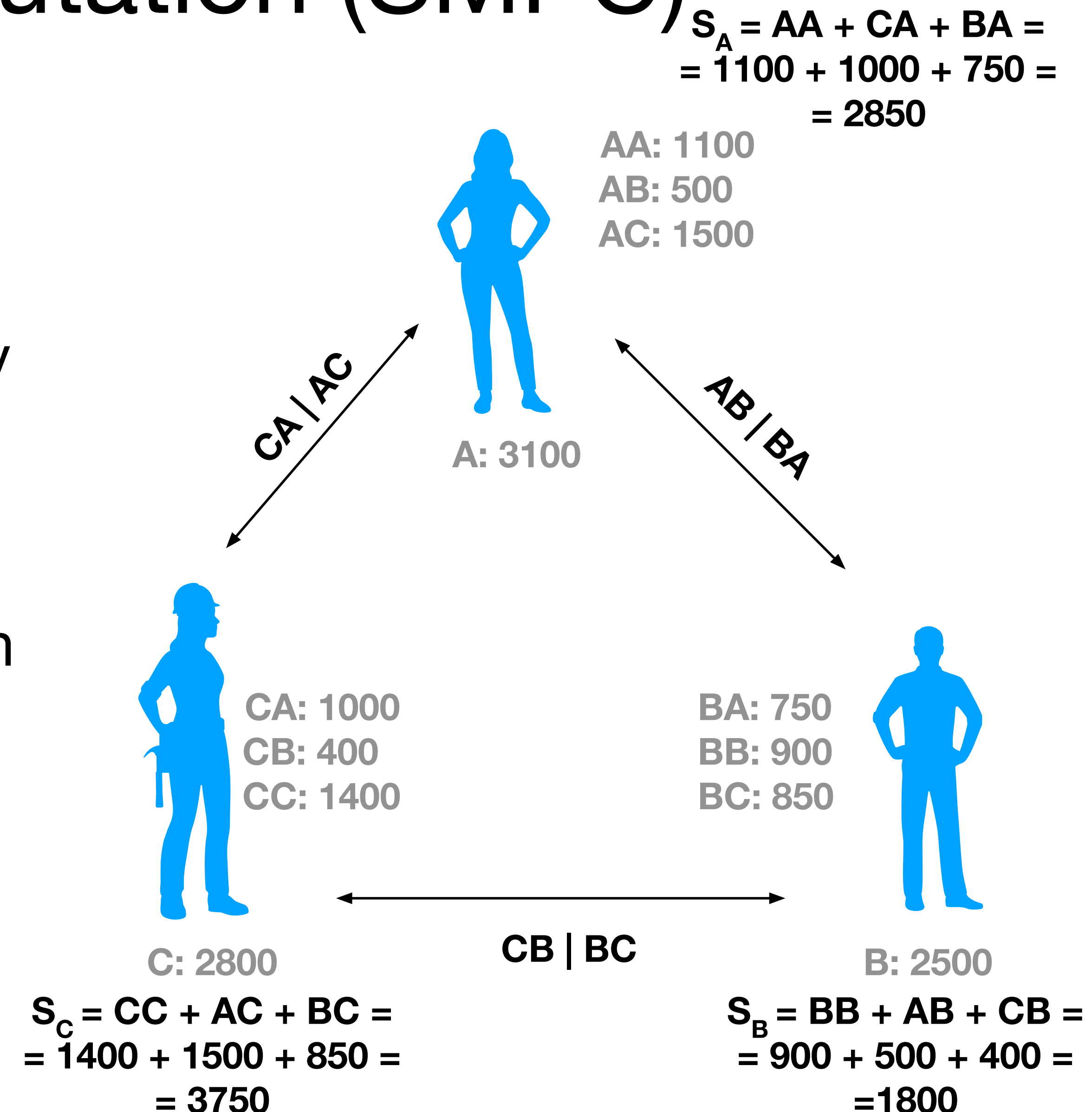


Secure multi-party computation (SMPC)

Example: Compute the average salary.

Each participant:

1. Splits their salary to 3 parts, randomly
2. Shares with each of the other participants one of the parts
3. Sums all received parts with their own left

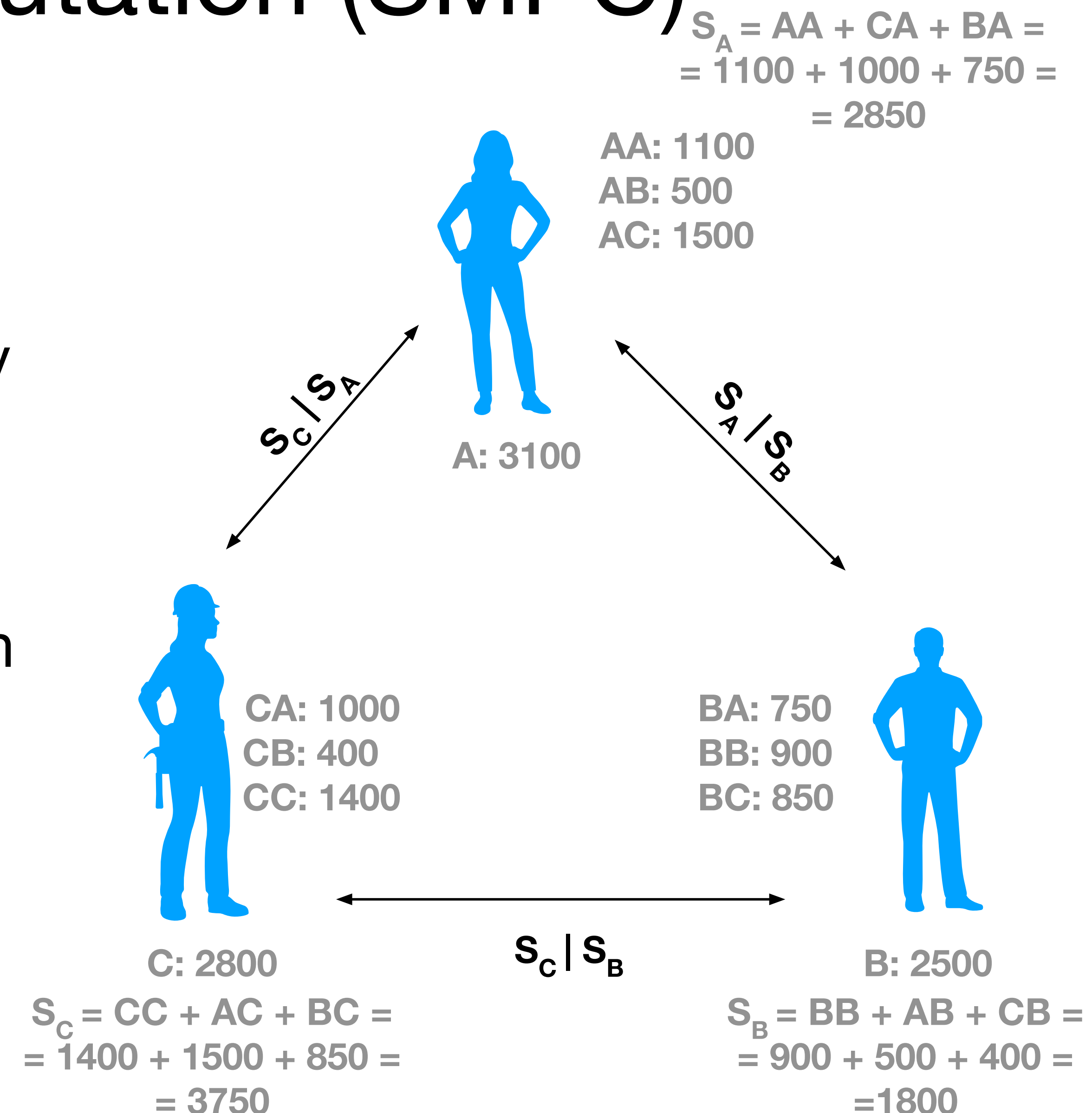


Secure multi-party computation (SMPC)

Example: Compute the average salary.

Each participant:

1. Splits their salary to 3 parts, randomly
2. Shares with each of the other participants one of the parts
3. Sums all received parts with their own left
4. Shares the sum

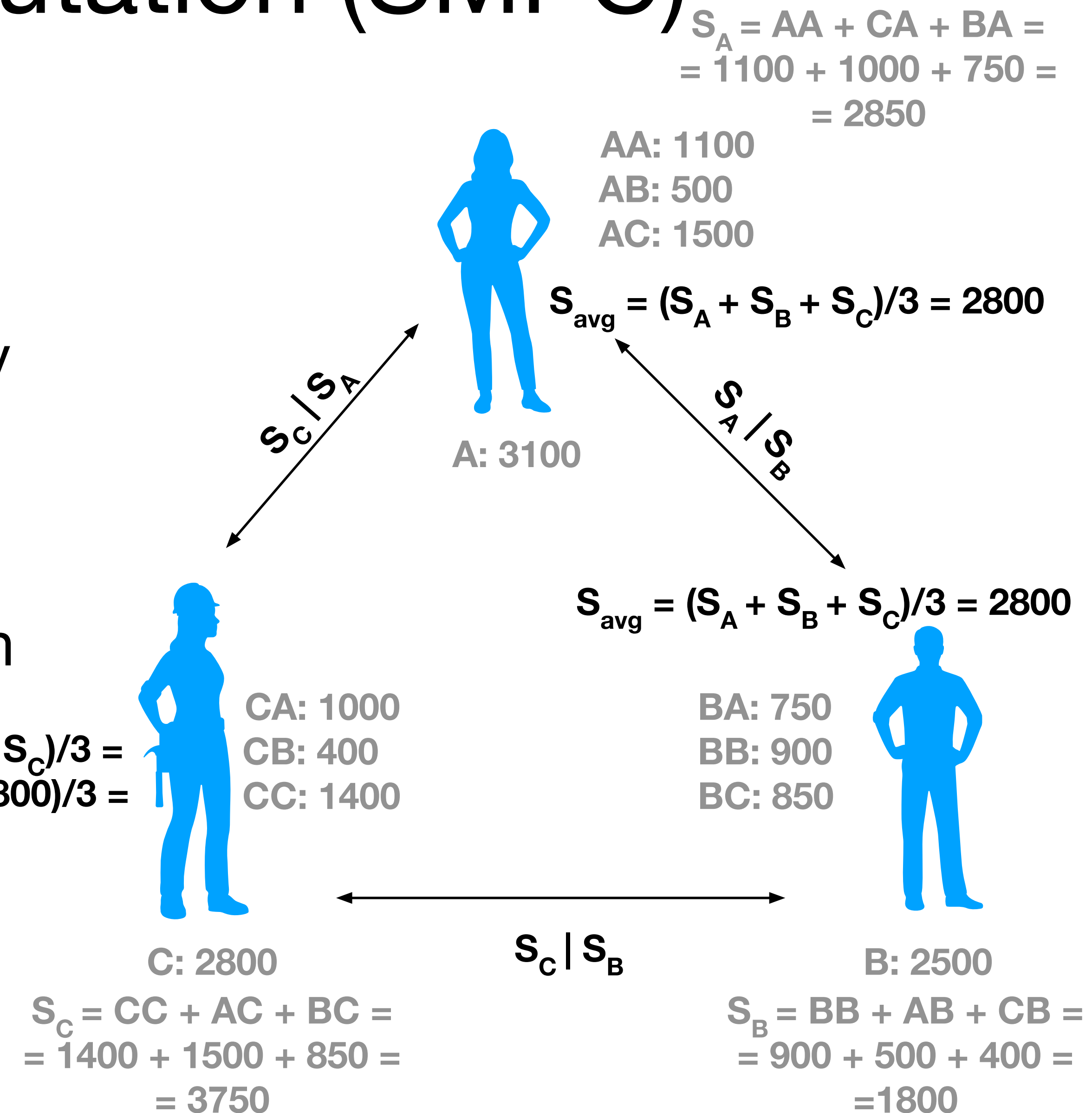


Secure multi-party computation (SMPC)

Example: Compute the average salary.

Each participant:

1. Splits their salary to 3 parts, randomly
2. Shares with each of the other participants one of the parts
3. Sums all received parts with their own left
4. Shares the sum
5. Computes the final sum and divides by 3

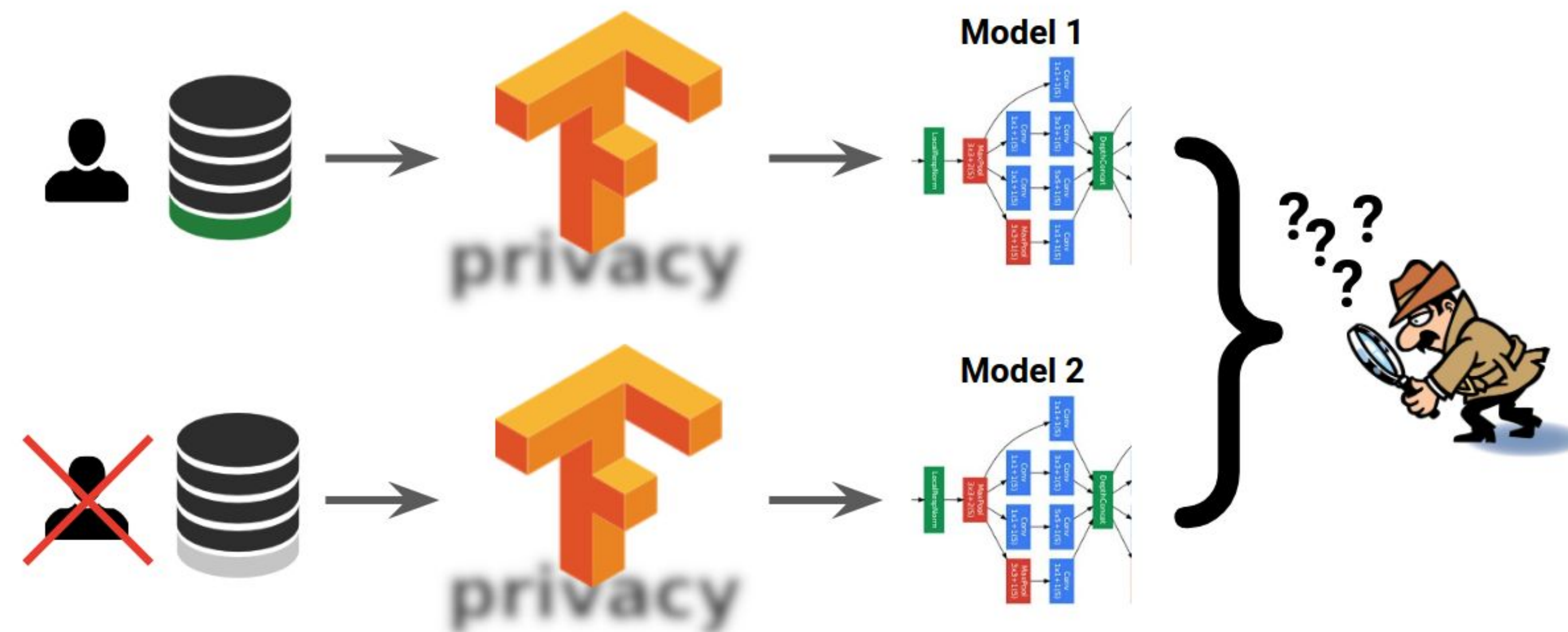


Differential Privacy

Differential Privacy

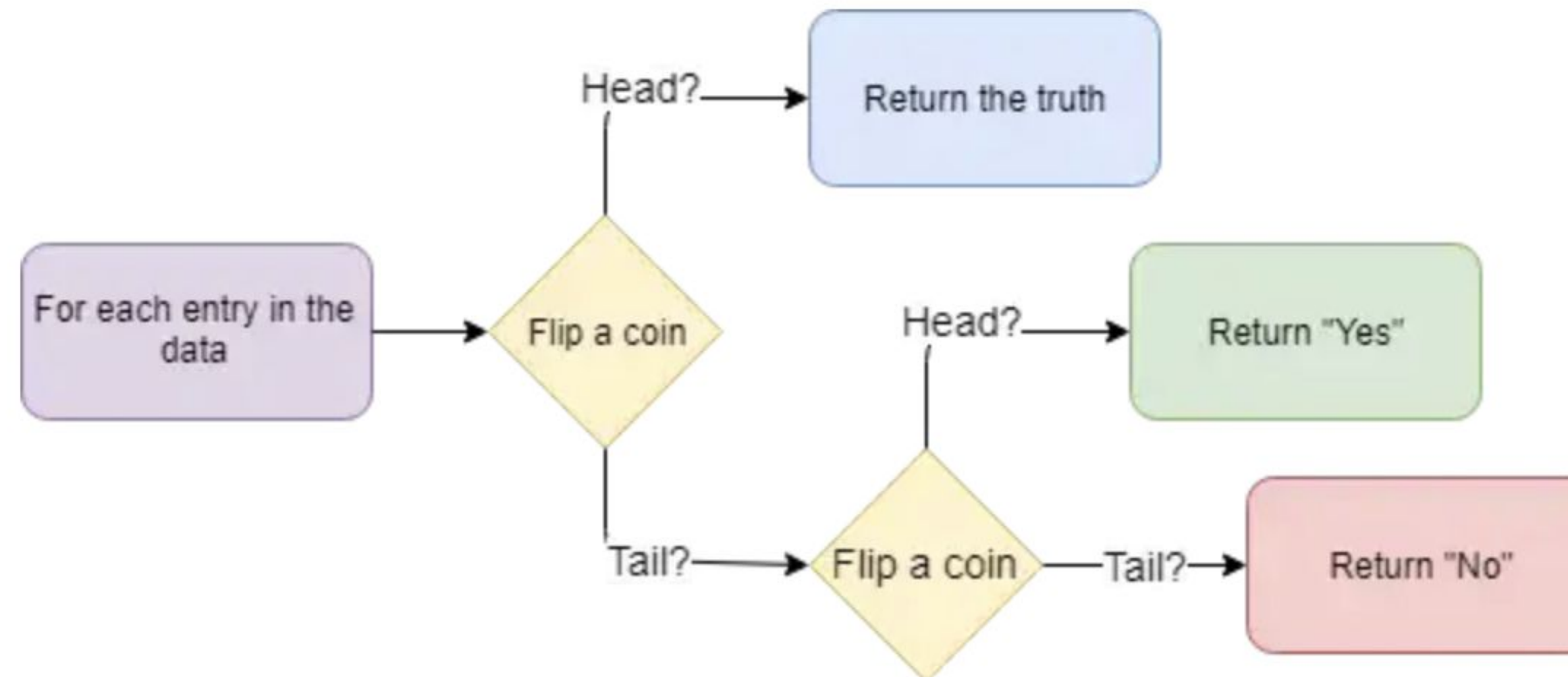
- Allows to add noise to the data, model or output of a model
- Noise should not influence overall statistics

Idea: DP guarantees that removing any sample from the dataset should lead to basically the same output as an output based on the full dataset



Differential Privacy: Randomised response

Imagine that you are a social data scientist, who wants to perform an analysis on a survey data about a very taboo behaviour. Each entry in the data is an answer (the truth) of individuals, “yes” or “no”, in the surveyed population.

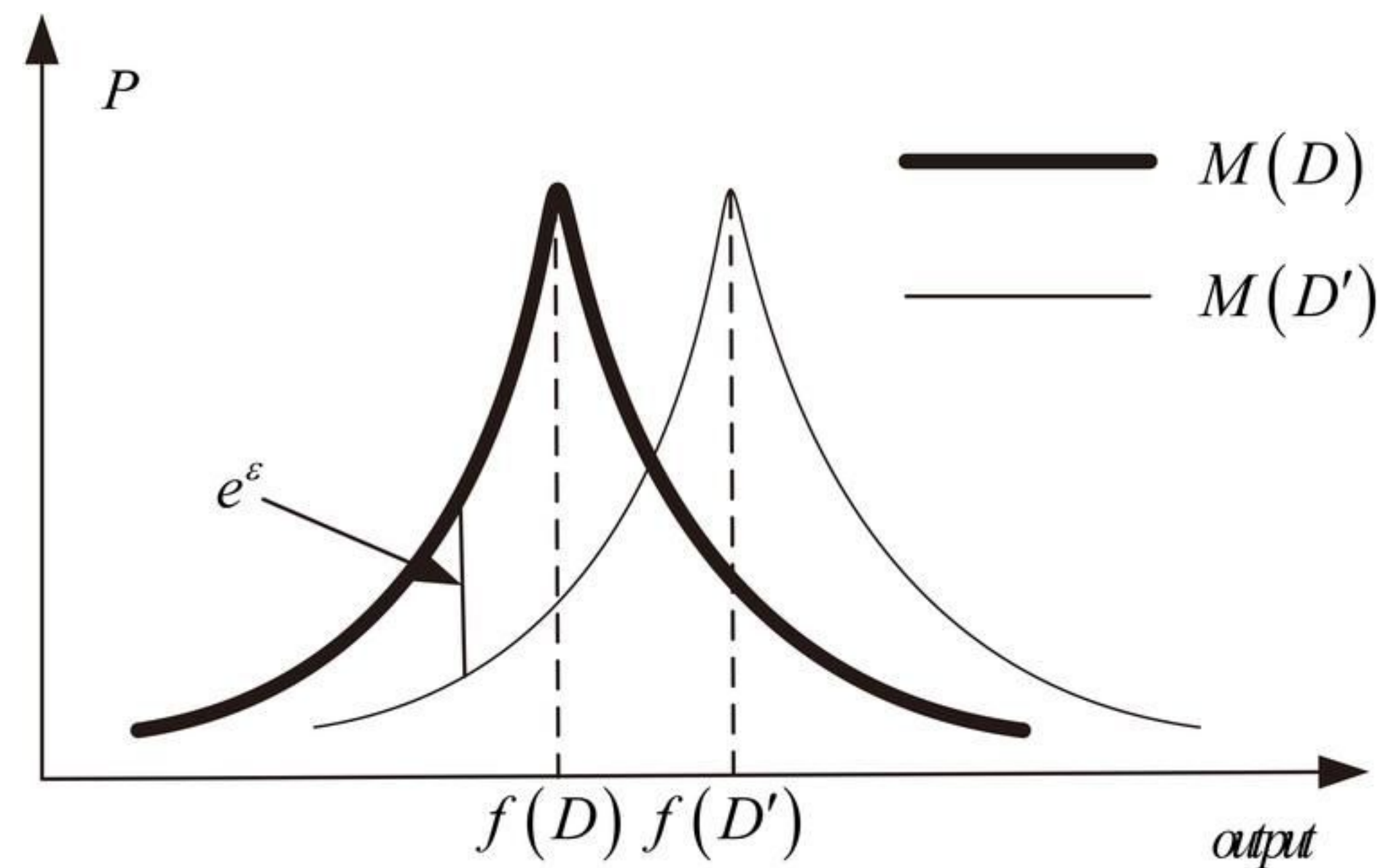


Protect each person with “plausible deniability”, because a person is plausible to deny the answer by the randomness of flipping a coin.

Differential Privacy: definition

A randomized algorithm $\mathcal{M}$ is ϵ -differentially private if for all possible sub-range of $\mathcal{M}$, say $\mathcal{S}$, and for all neighbouring dataset D_1, D_2 , the probability that $\mathcal{M}$ gives the same output based on the two input datasets is very similar.

$$Pr[\mathcal{M}(D_1) \in \mathcal{S}] \leq \exp(\epsilon) Pr[\mathcal{M}(D_2) \in \mathcal{S}]$$



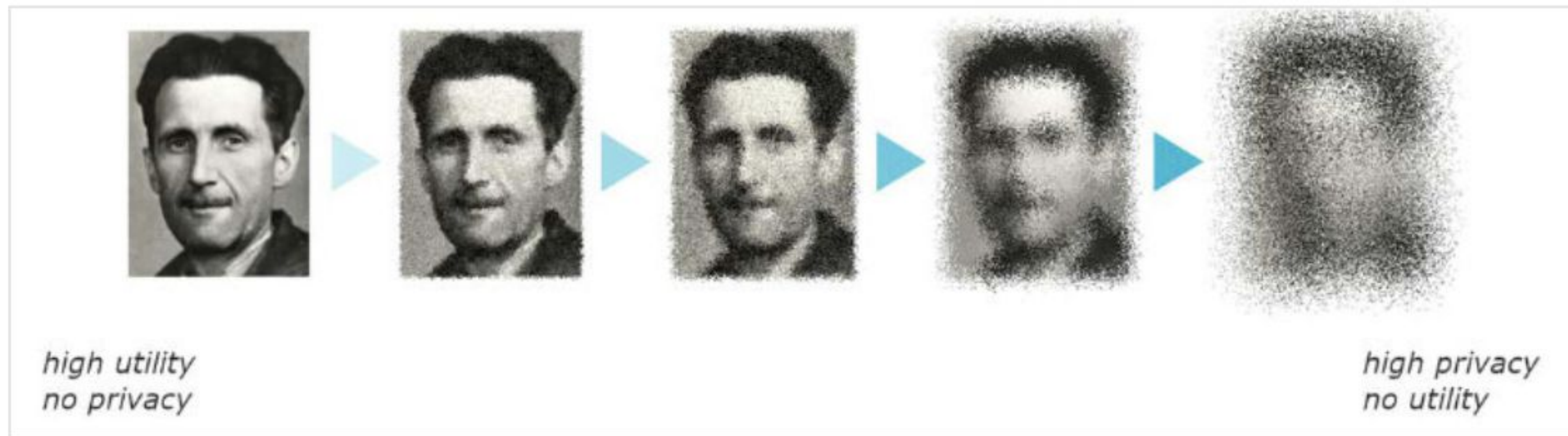
Differential Privacy

In machine learning, to achieve differential privacy of an algorithm one can **add noise** to:

- Training data: **input perturbation**
- Objective function or gradient during model training: **objective or gradient perturbation**
- Final model: **output perturbation**
- Model's predictions: **predictions perturbation**

Differential Privacy

Input perturbation:



$$X'_i = X_i + \text{noise}, \forall X_i \in X_{train}$$

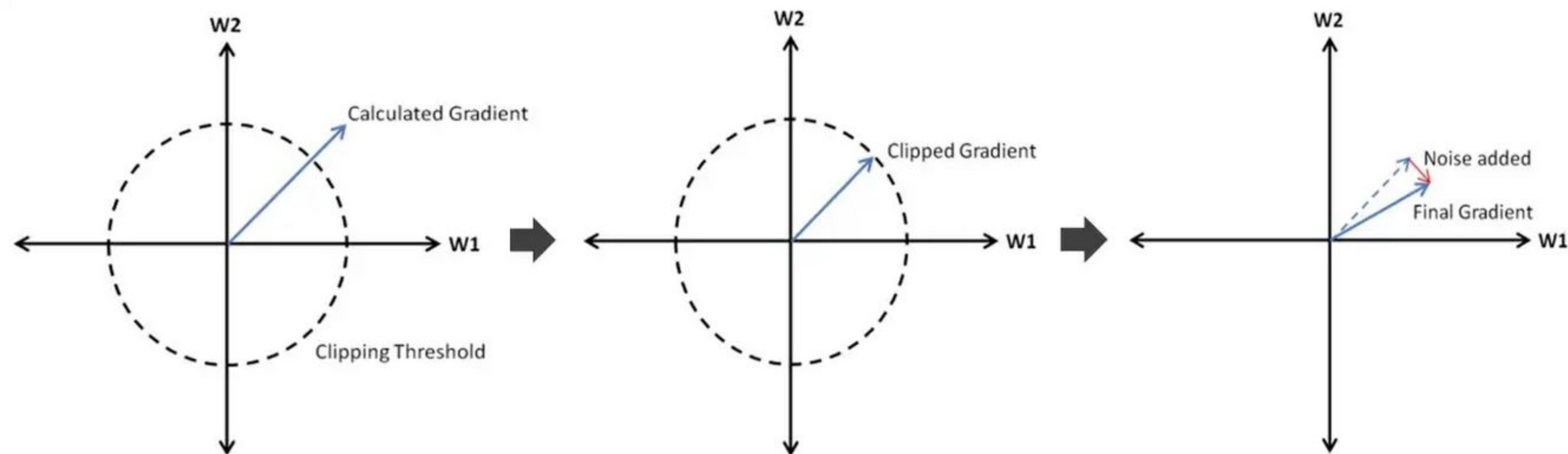
Differential Privacy

Objective perturbation:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n \overset{\text{objective function (loss)}}{l(\theta, X_i, y_i)} + \underbrace{\lambda R(\theta)}_{\text{Regularisation}} + \text{noise}$$

↑
model

Gradient perturbation:



Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. 2011. Differ- entially Private Empirical Risk Minimization. J. Mach. Learn. Res. 12 (2011), 1069–1109.

Martín Abadi, Andy Chu, Ian J. Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep Learning with Differential Privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016. ACM, 308–318.

Differential Privacy

Output perturbation:

$$\theta = \underset{\substack{\uparrow \\ \text{trained model}}}{\theta} + \textit{noise}$$

Prediction perturbation:

$$y = f_{\theta}(X_i, \theta) + \textit{noise}$$

Differential Privacy

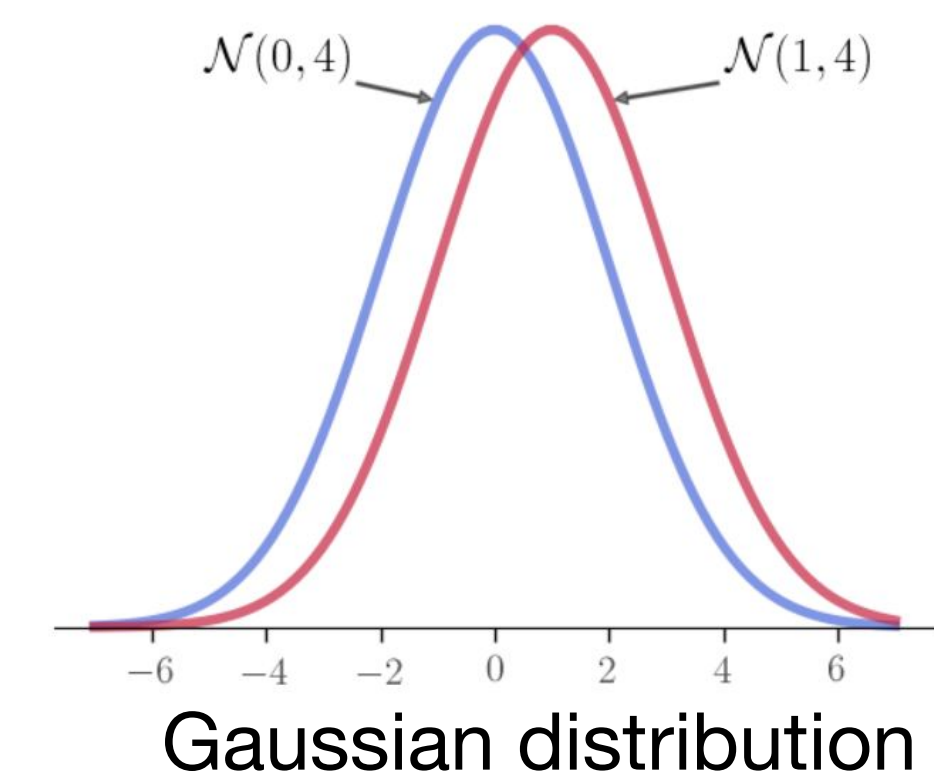
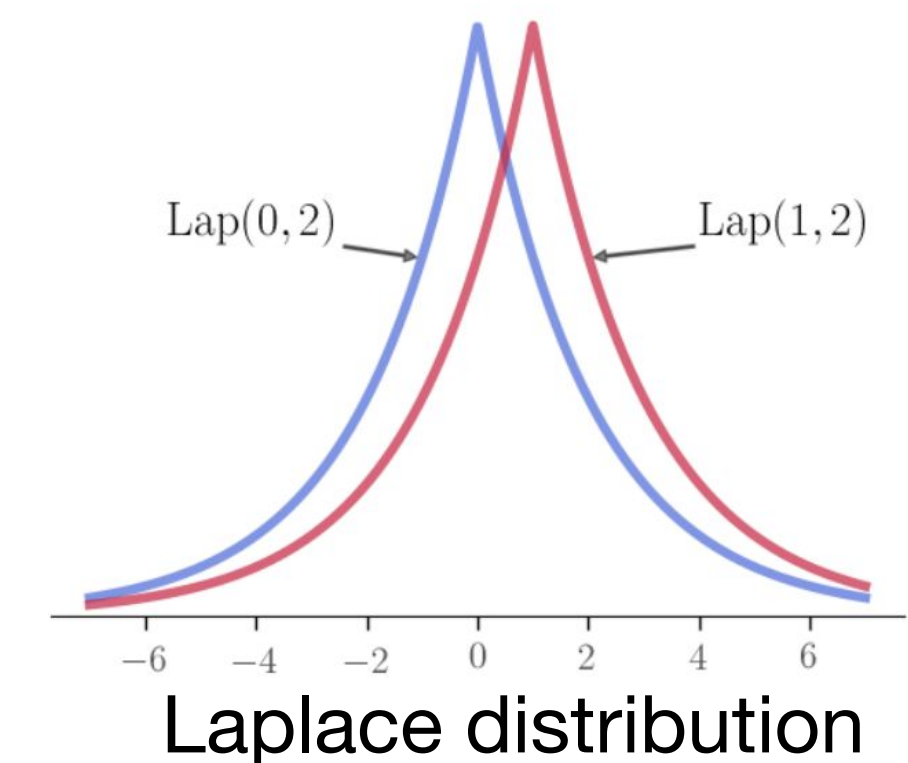
How to calculate the noise?

Example for output perturbation:

$$\theta = \underset{\substack{\uparrow \\ \text{trained model}}}{\theta} + \text{noise}$$

Additive noise mechanisms:

- *Laplace Mechanism* - adds noise drawn from a Laplace distribution
- *Gaussian Mechanism* - adds noise drawn from a Gaussian distribution



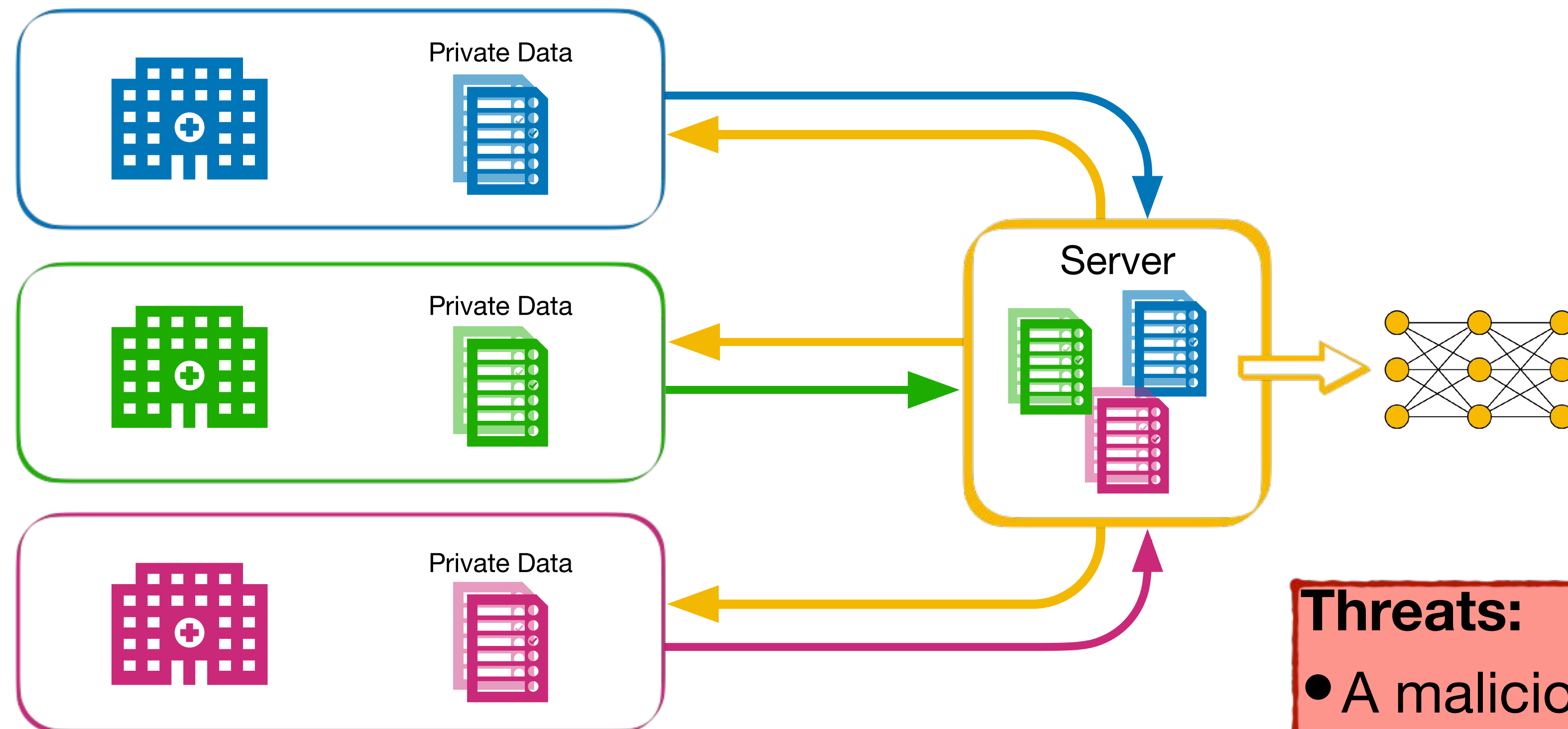
Differential Privacy

- Is not an algorithm but rather a **property** of an algorithm.
- The **main issue** of DP is the **drop in utility of the model**. By adding noise to the data, model, or predictions the accuracy of the model's predictions can drop significantly.
- One can use a parameter **ϵ (epsilon)** to quantify the privacy loss and regulate privacy level and the trade-off between privacy and utility.
- **Composition property:** quantifying loss enables the control and analysis of cumulative privacy losses across multiple computations

Federated Learning

Centralised Machine Learning

A server collects private data to train a machine learning model

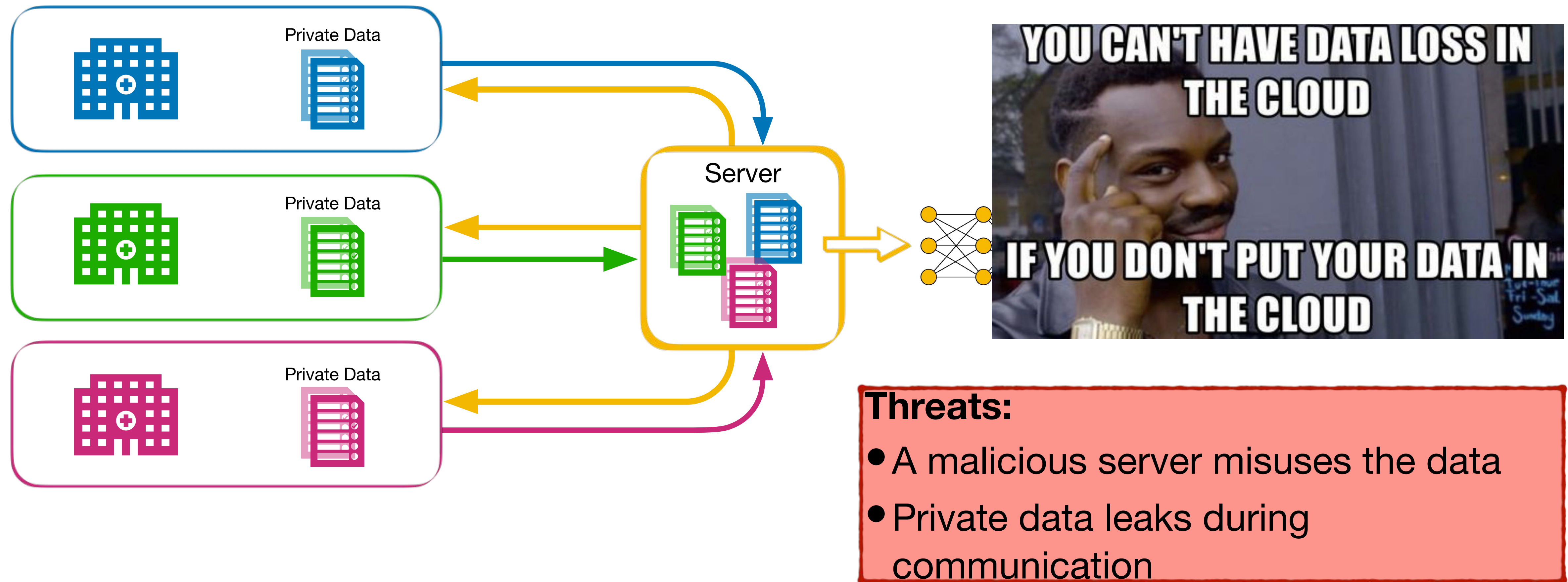


Threats:

- A malicious server misuses the data
- Private data leaks during communication

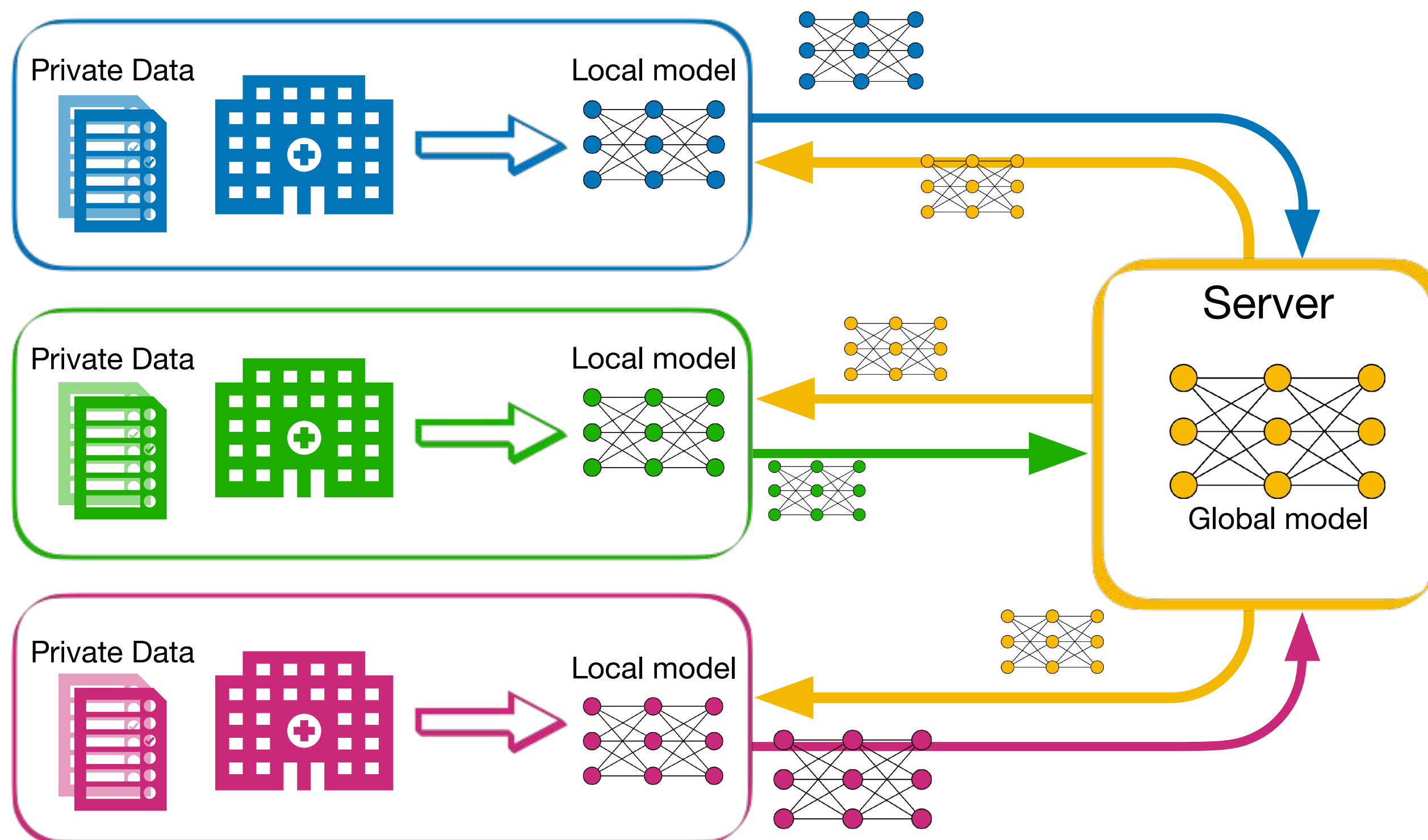
Centralised Machine Learning

A server collects private data to train a machine learning model



Federated Learning

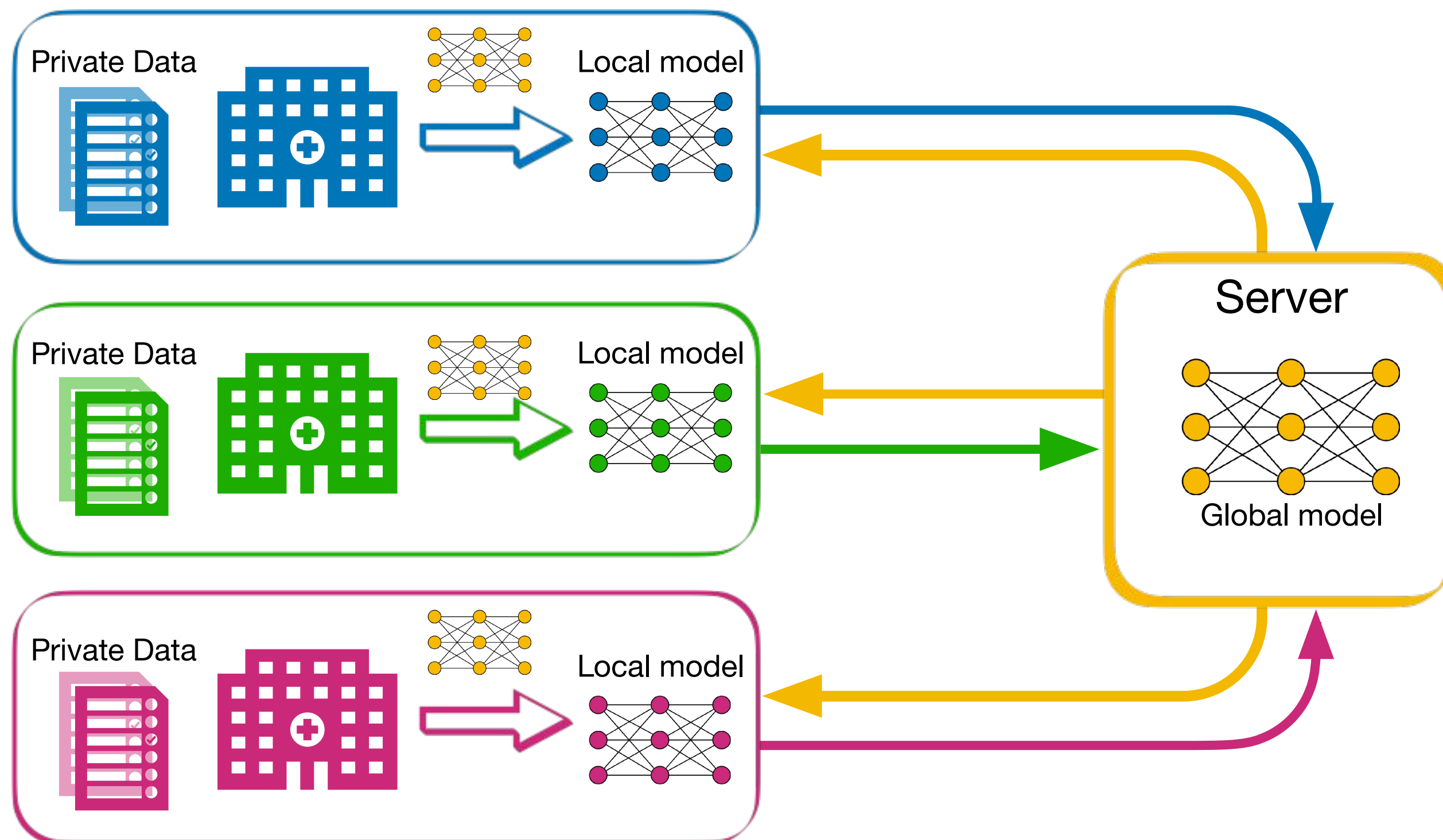
A server collects machine learning models instead of sensitive data



1. Each party (**client**) trains a **local model** on its private training data
2. Each client **shares its local model** with a Server
3. The Server aggregates all locally trained models into a **global model**
4. The Server sends a global model back to the clients for the following training or usage

Federated Learning

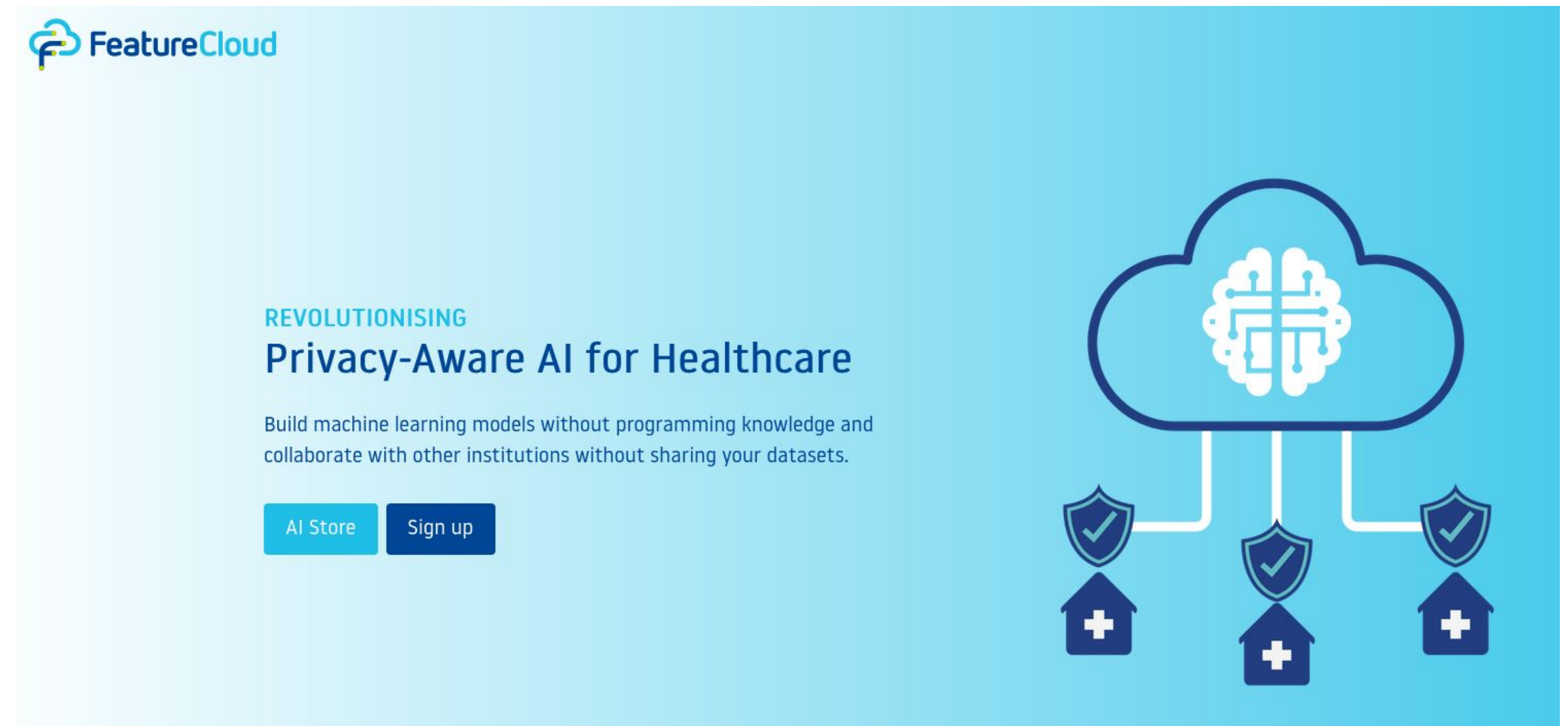
A server collects machine learning models instead of sensitive data



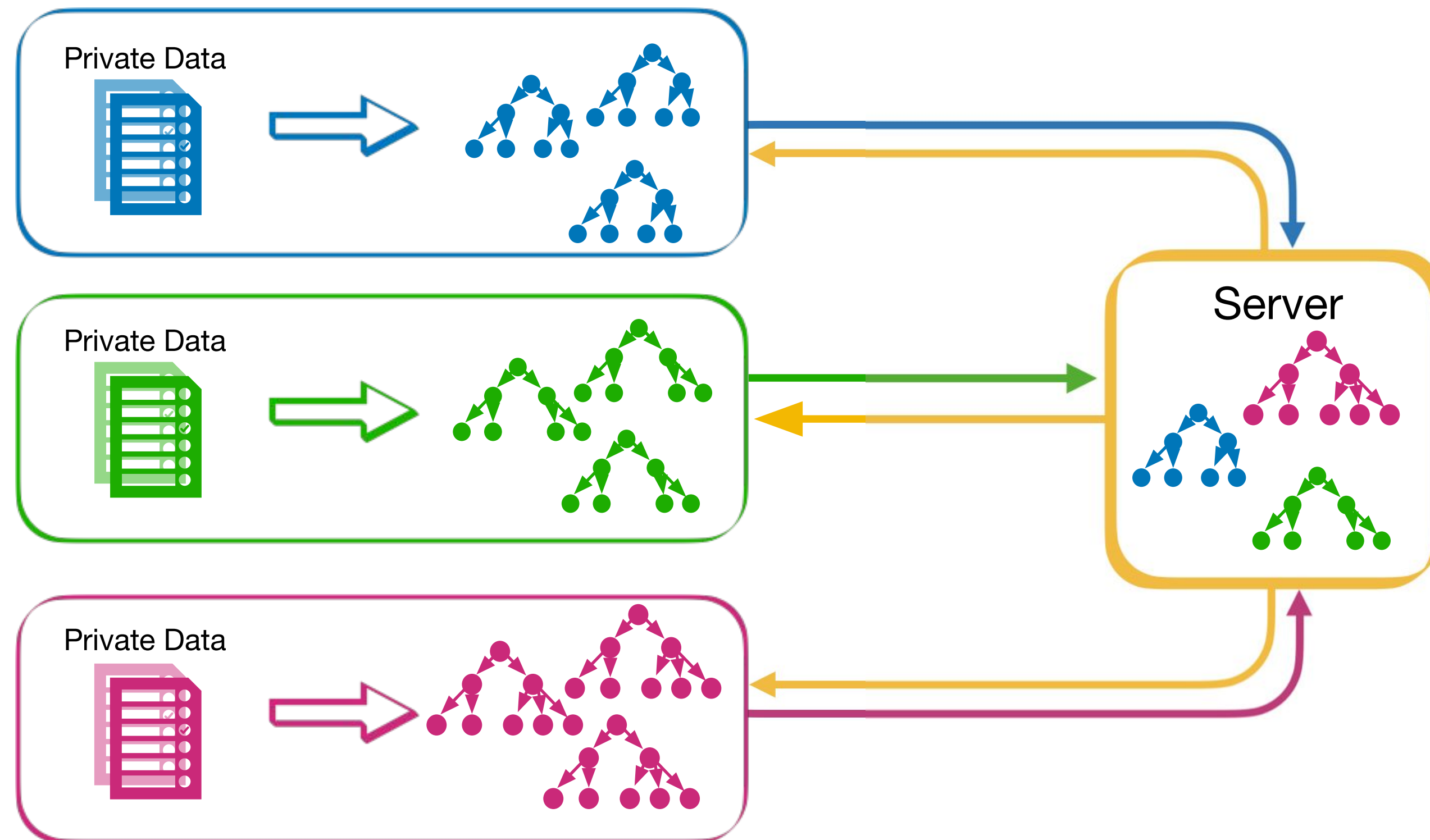
1. Each party (**client**) trains a **local model** on its private training data
2. Each client **shares its local model** with a Server
3. The Server aggregates all locally trained models into a **global model**
4. The Server sends a global model back to the clients for the following training or usage

Federated Learning: Applications

- **Mobile devices**
 - Google, iOS mobile keyboard
 - “Hey Siri”
- **Medical domain**
 - Pharmaceutical discovery
 - Research on health data records
- **IoT**
 - Smart cities
 - Smart transportation
 - Unmanned Aerial Vehicles (UAV)



Federated Forest



Options:

- Each node trains locally a decision tree (or a random forest) and sends it to the server
- Each node trains trees on specific features and sends only them to the server

Parameters:

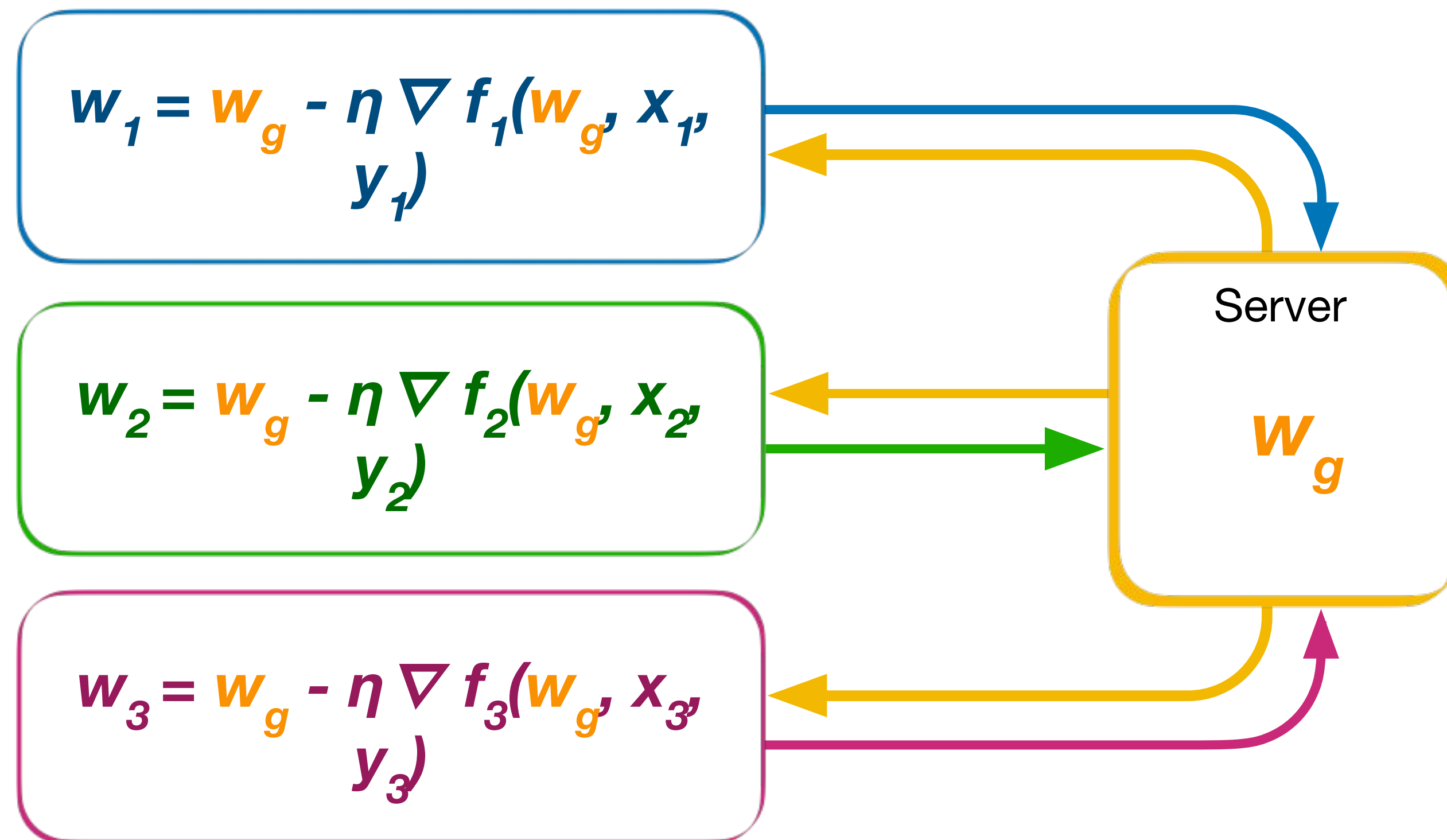
- Choose stopping criteria (max depth of the trees, limiting number of remaining samples in leaf nodes)
- Randomly create extra trees locally (to bring the noise)
- Weights of the trees

L. A. C. de Souza, G. Antonio F. Rebello, G. F. Camilo, L. C. B. Guimarães and O. C. M. B. Duarte, "DFedForest: Decentralized Federated Forest," 2020 IEEE International Conference on Blockchain (Blockchain), 2020, pp. 90-97, doi: 10.1109/Blockchain50366.2020.00019.

Liu, Yang, Yingting Liu, Zhijie Liu, Junbo Zhang, Chuishi Meng and Yu Zheng. "Federated Forest." ArXiv abs/1905.10053 (2019): n. pag.

Liu, Yang, Mingxi Chen, Wenxi Zhang, Junbo Zhang and Yu Zheng. "Federated Extra-Trees with Privacy Preserving." ArXiv abs/2002.07323 (2020): n. pag.

Federated Averaging

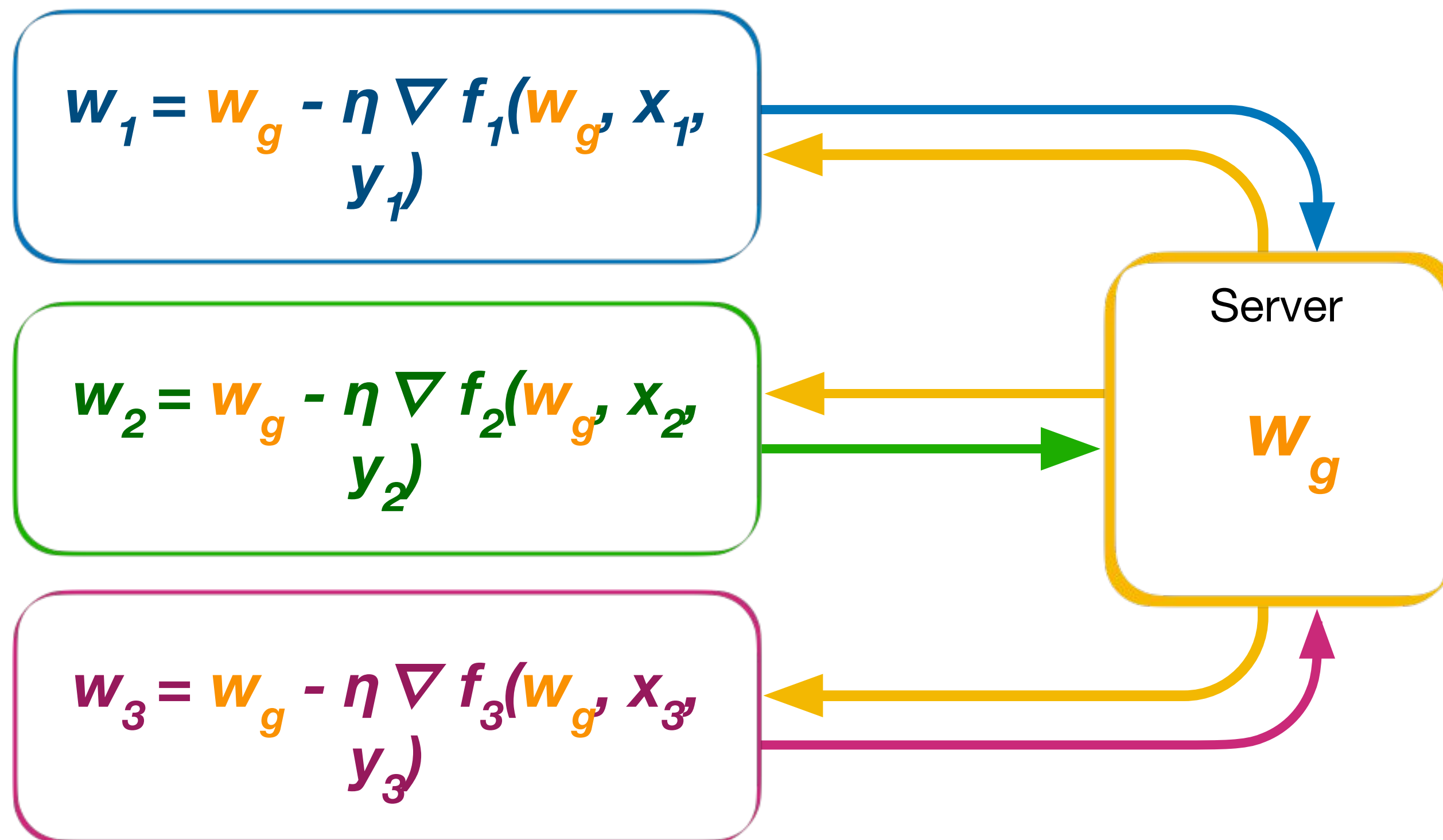


The server collects locally trained models and computes a global models as a (weighted) average of the local models:

$$w_g = \sum_{k=1}^K \frac{n_k}{n} w_k$$

McMahan, H. B., Eider Moore, D. Ramage, S. Hampson and B. A. Y. Arcas. "Communication-Efficient Learning of Deep Networks from Decentralized Data." AISTATS (2017).

Federated Averaging



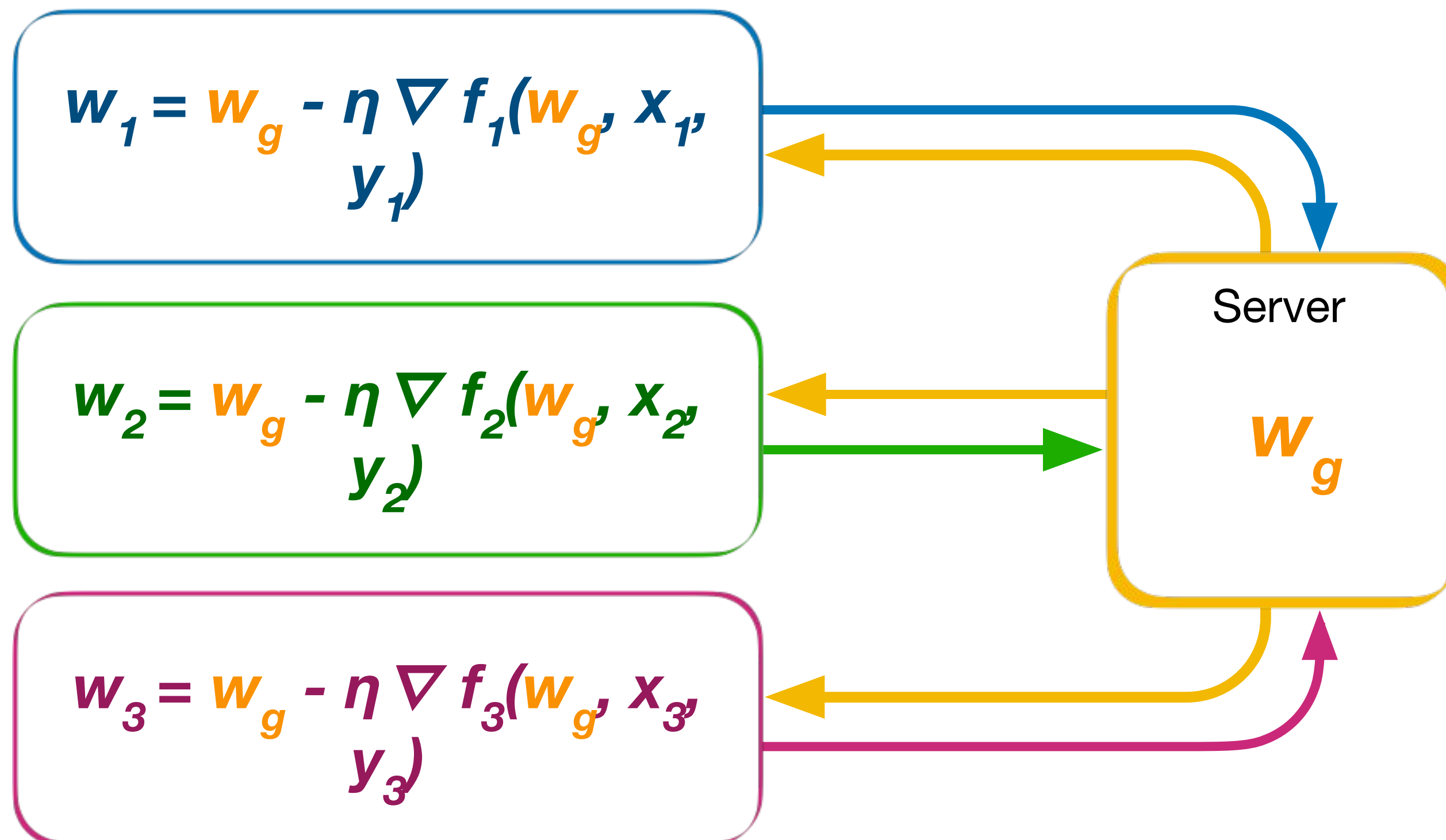
The server collects locally trained models and computes a global models as a (weighted) average of the local models:

$$w_g = \sum_{k=1}^K \frac{n_k}{n} w_k$$

Problem:

- Slow convergence if clients start training from different models

Federated Averaging



The server collects locally trained models and computes a global models as a (weighted) average of the local models:

$$w_g = \sum_{k=1}^K \frac{n_k}{n} w_k$$

Problem:

- Slow convergence if clients start training from different models

Solution:

- Randomly initialise the first global model and send it to the nodes for training

McMahan, H. B., Eider Moore, D. Ramage, S. Hampson and B. A. Y. Arcas. "Communication-Efficient Learning of Deep Networks from Decentralized Data." AISTATS (2017).

Federated Learning: Challenges

- Communication costs
- Data and systems heterogeneity
- Security risks
- **Privacy risks**

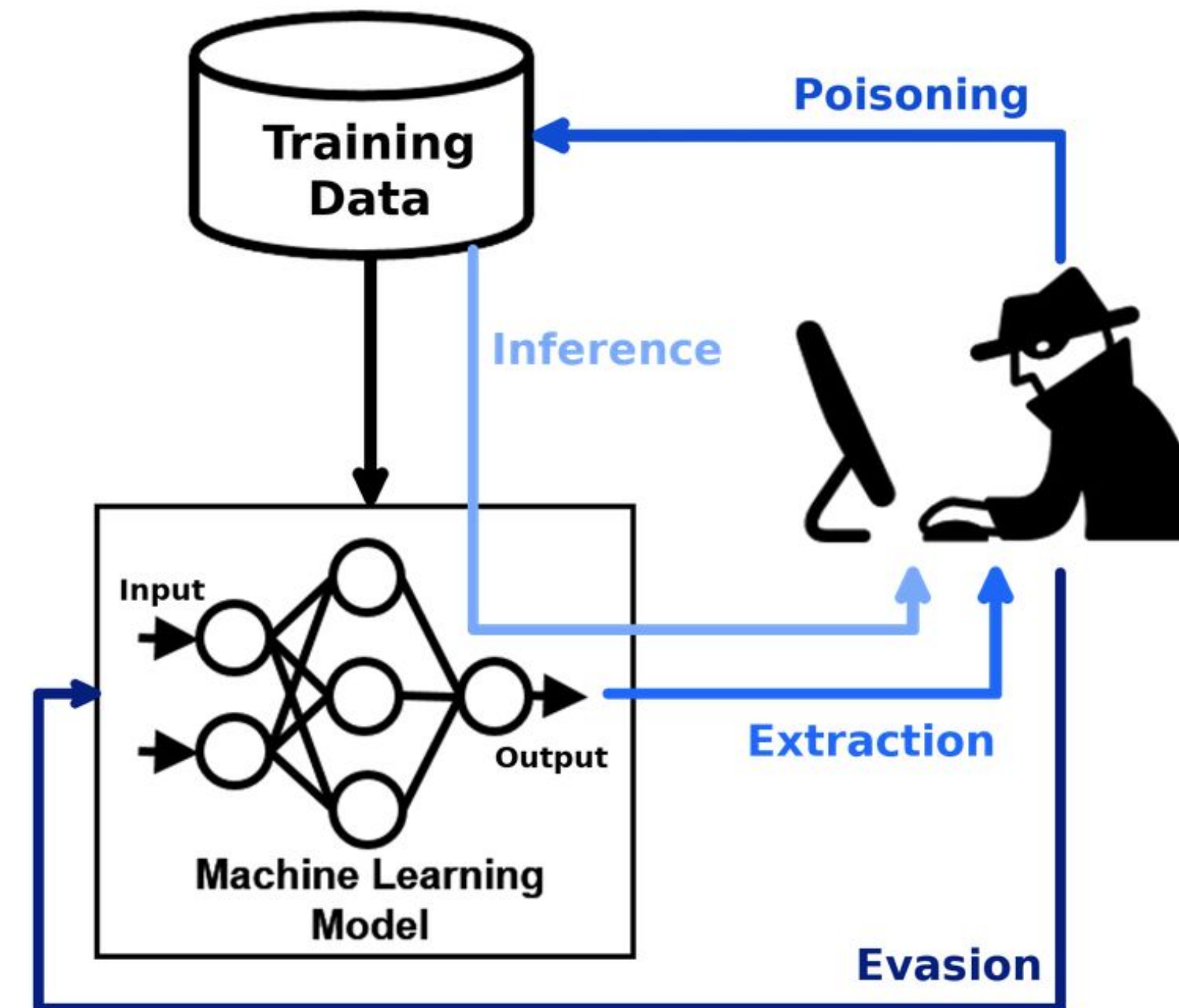
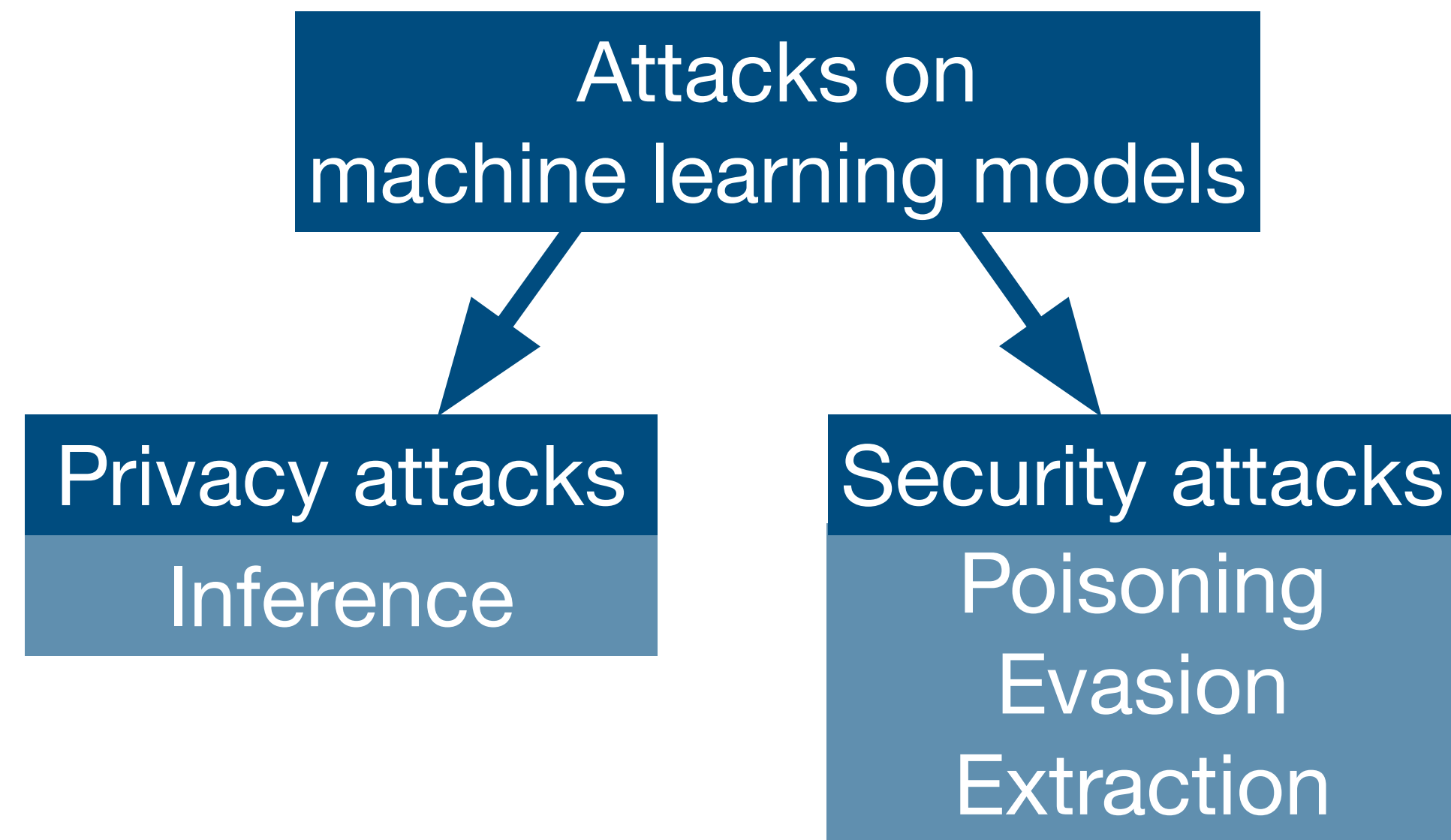
Privacy and Security

threats in ML

ML Security and Privacy risks

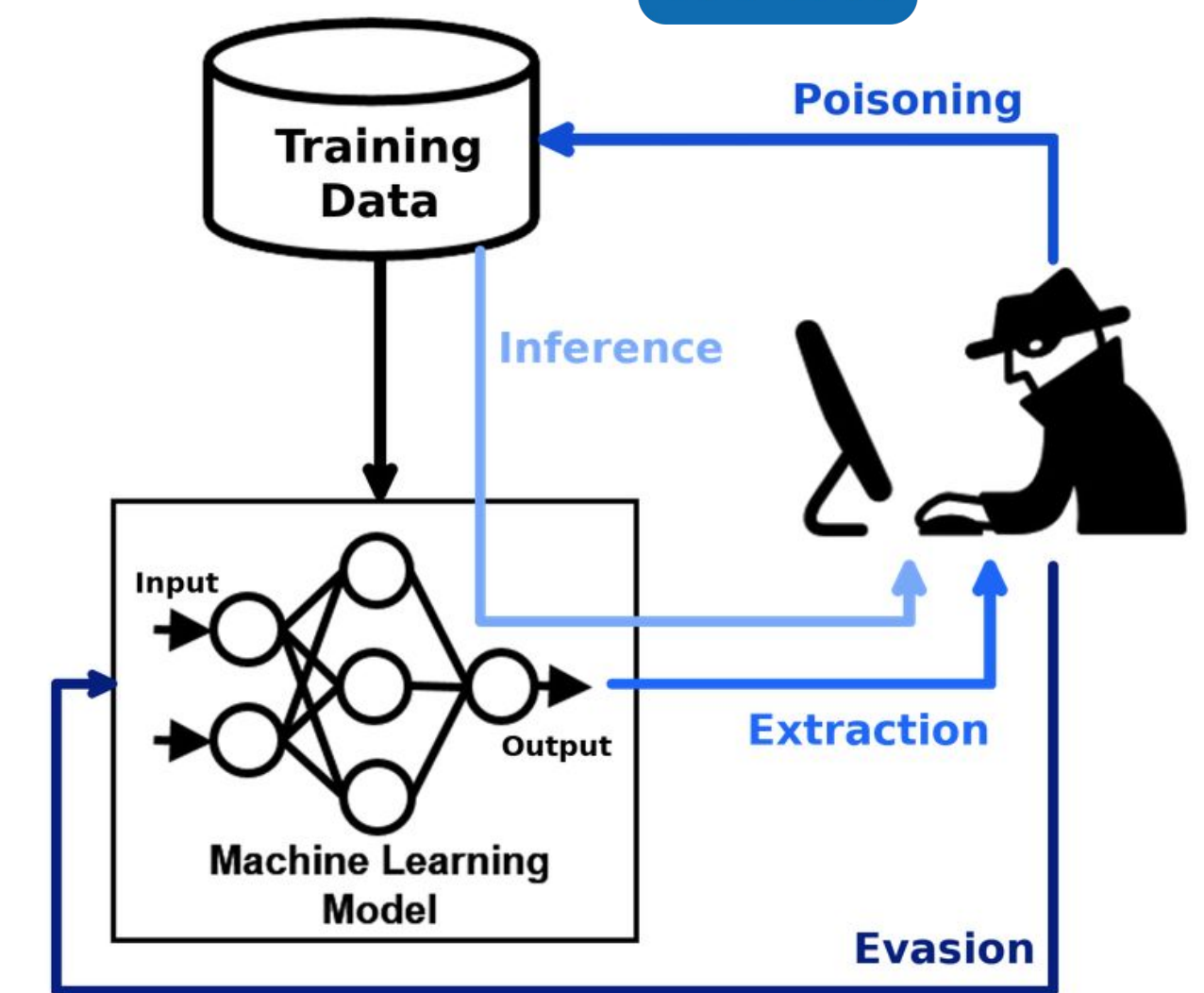
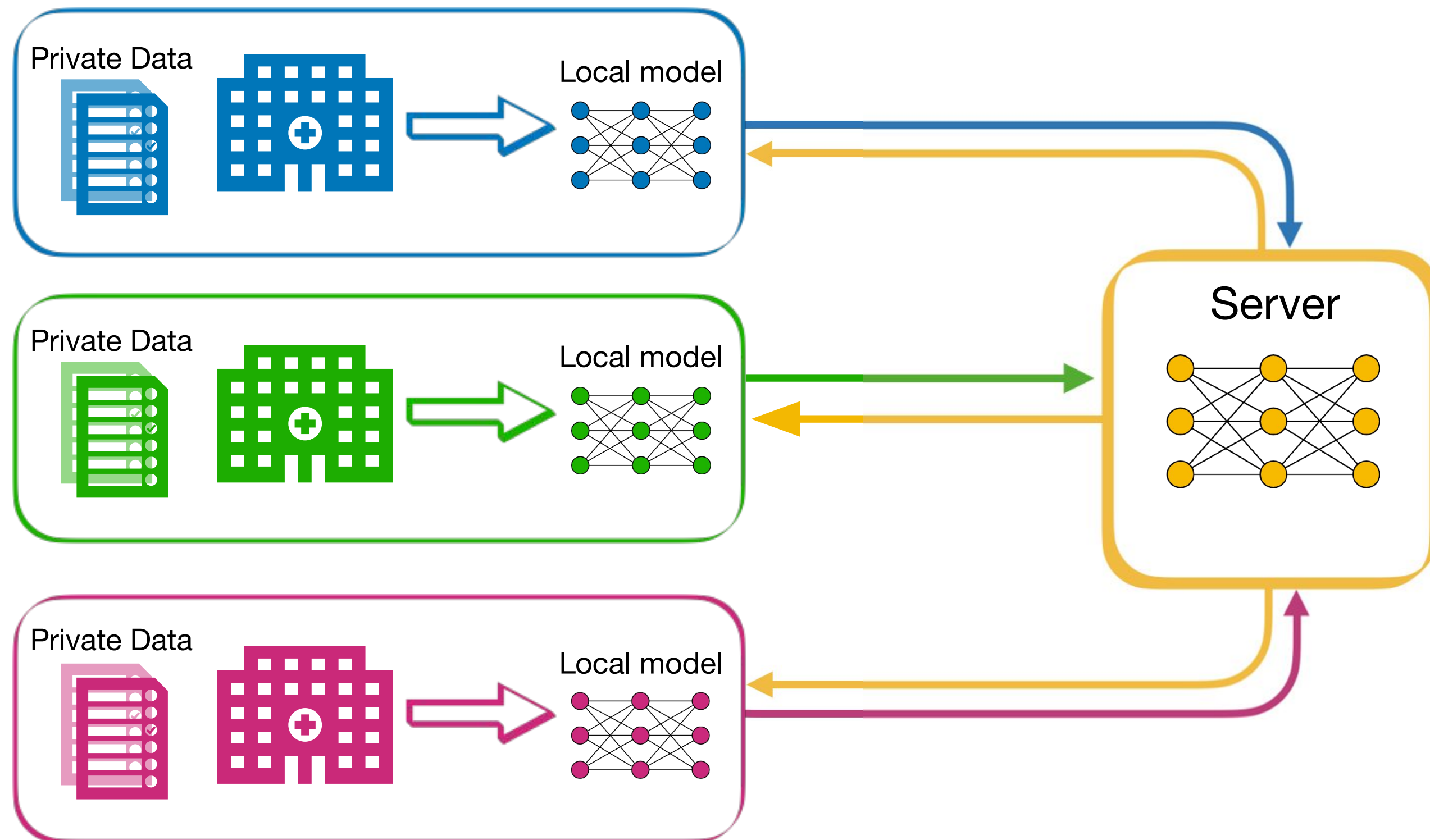


ML Security and Privacy risks



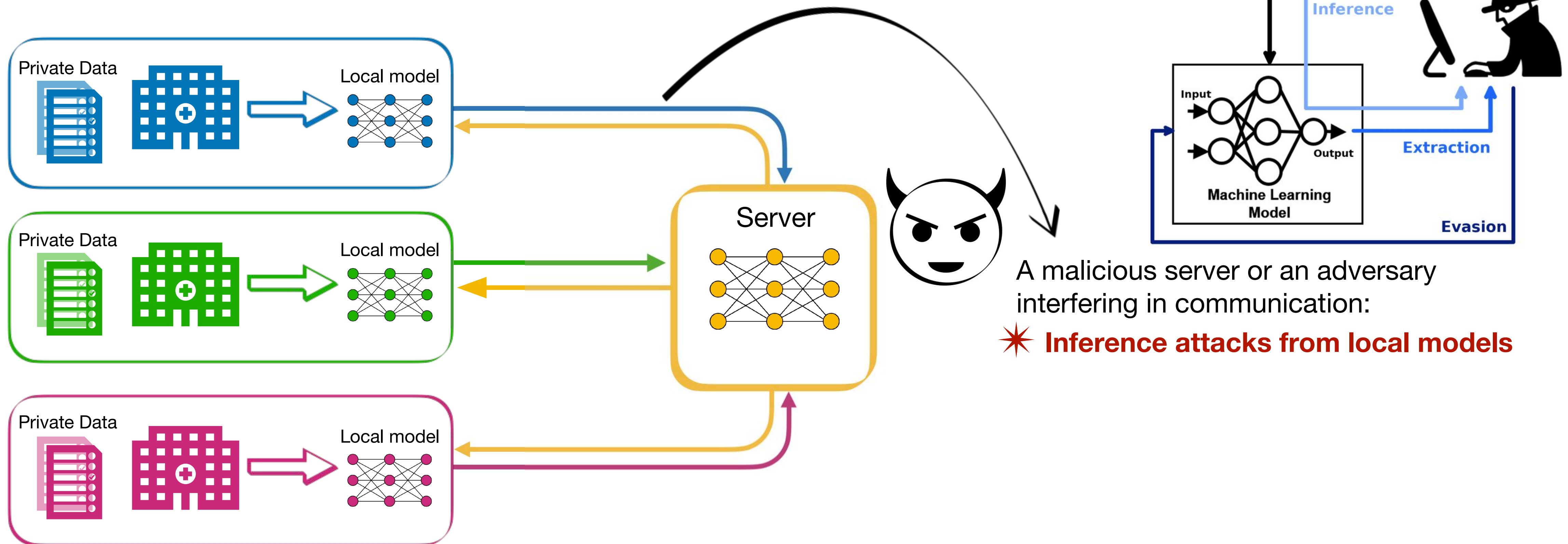
Despite FL allows to communicate machine learning models instead of data, these **models can become a surface for attacks**

Threats in Federated Learning

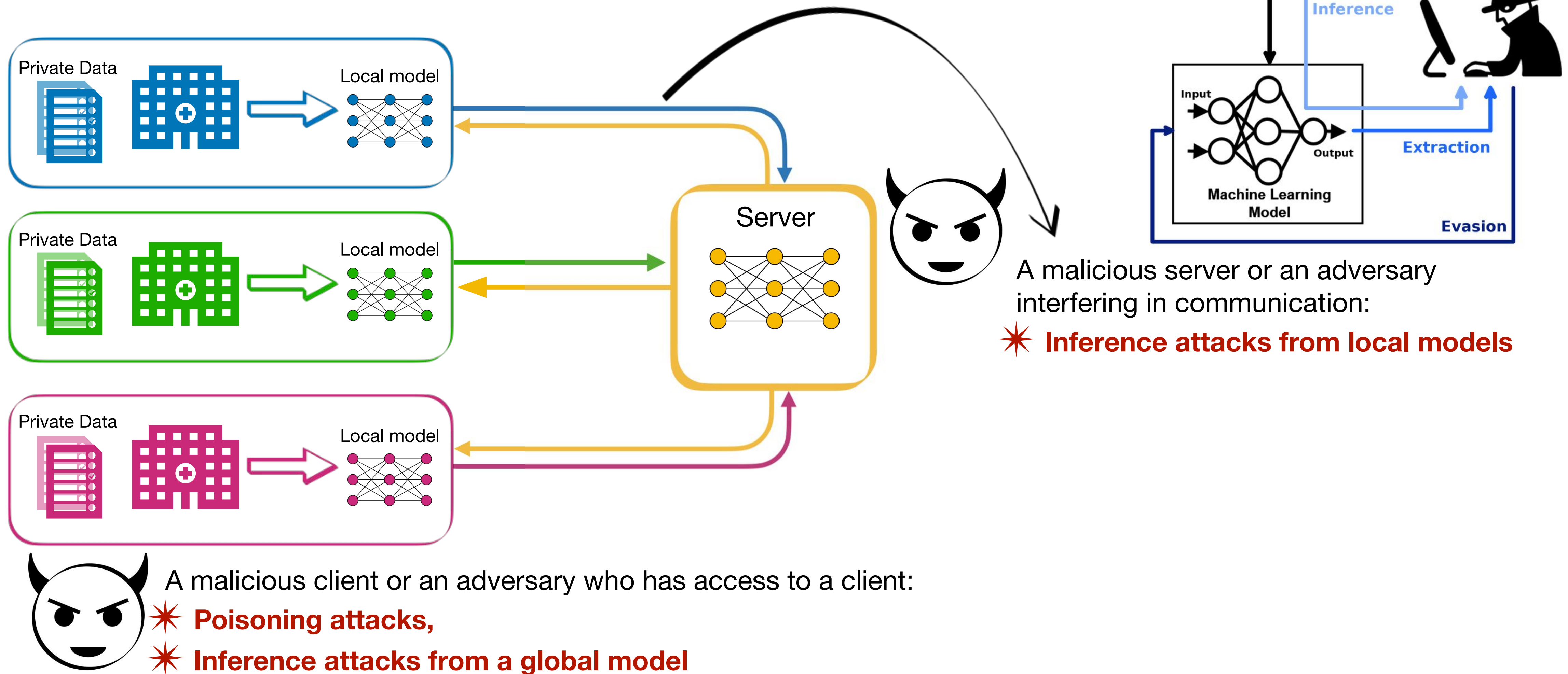


Who can be an
adversary in federated
learning?

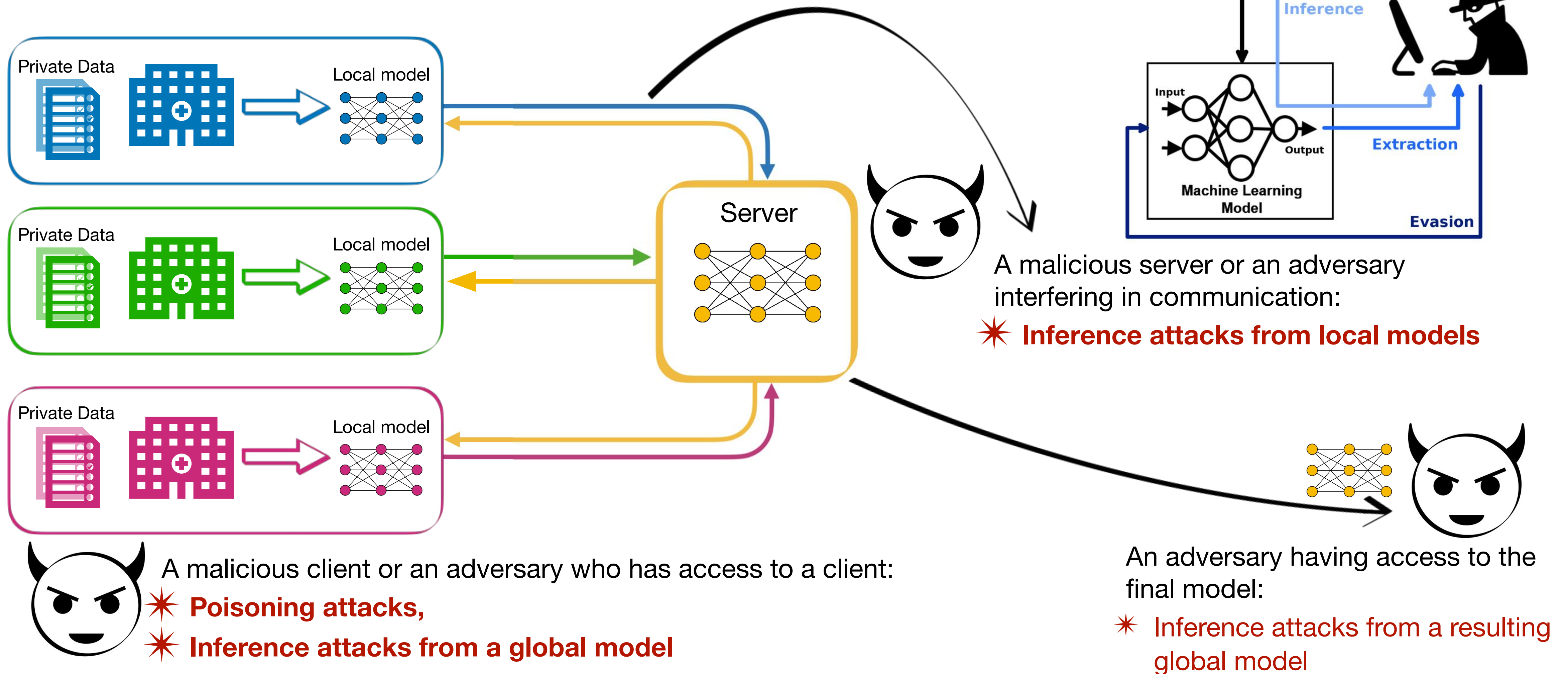
Threats in Federated Learning



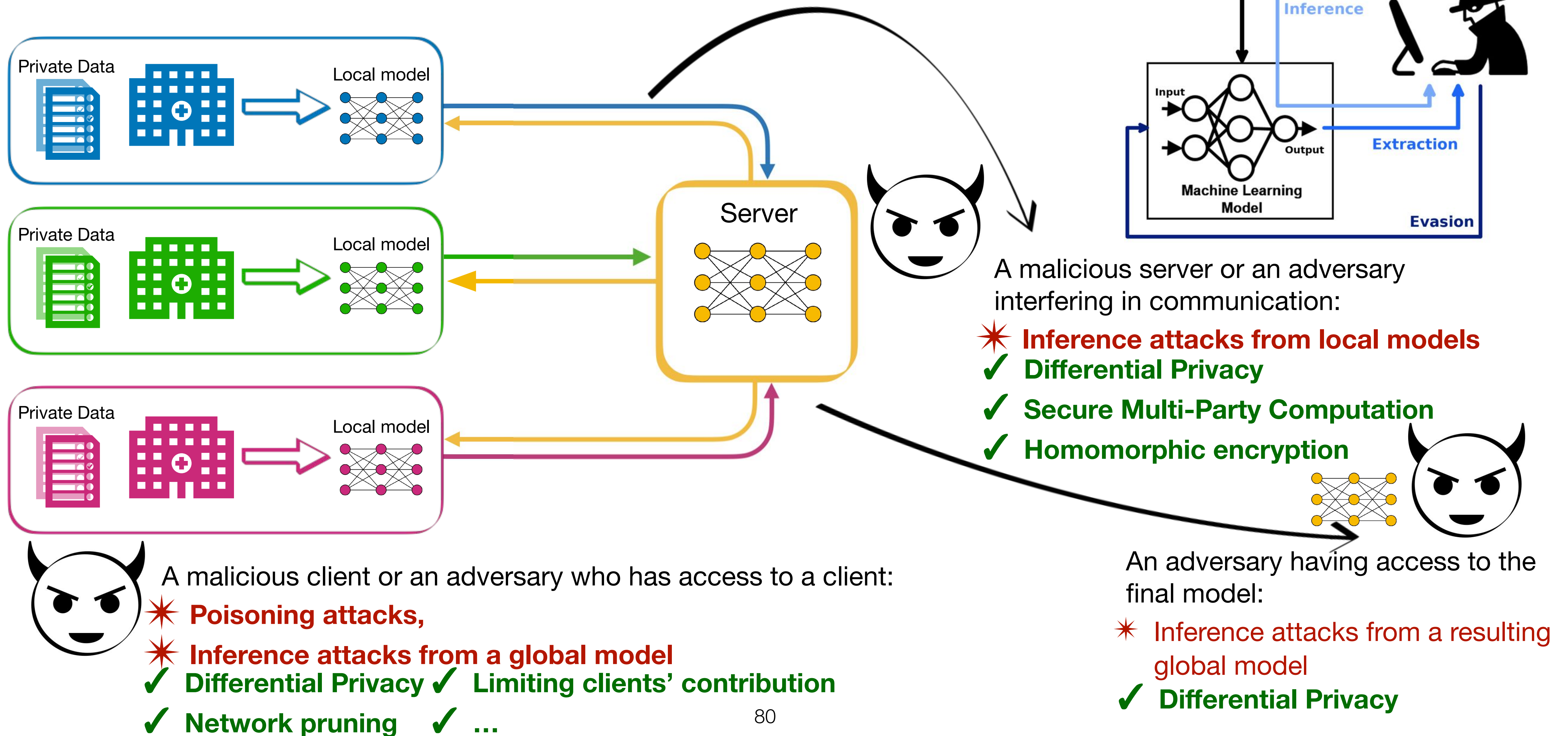
Threats in Federated Learning



Threats in Federated Learning



Threats in Federated Learning



Takeaway

- Privacy preserving data publishing and privacy preserving computation **allow performing data analysis while keeping data private**
- **Privacy comes at a cost** and results either in increased computation or communication costs or reduction of the data quality
- Privacy preserving techniques can be combined together

Thank you!

Anastasia Pustozero**va**, TU Wien & SBA Research
email: apustozero@sba-research.org

Tanja Šarčević, TU Wien & SBA Research
email: tsarcevic@sba-research.org