ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

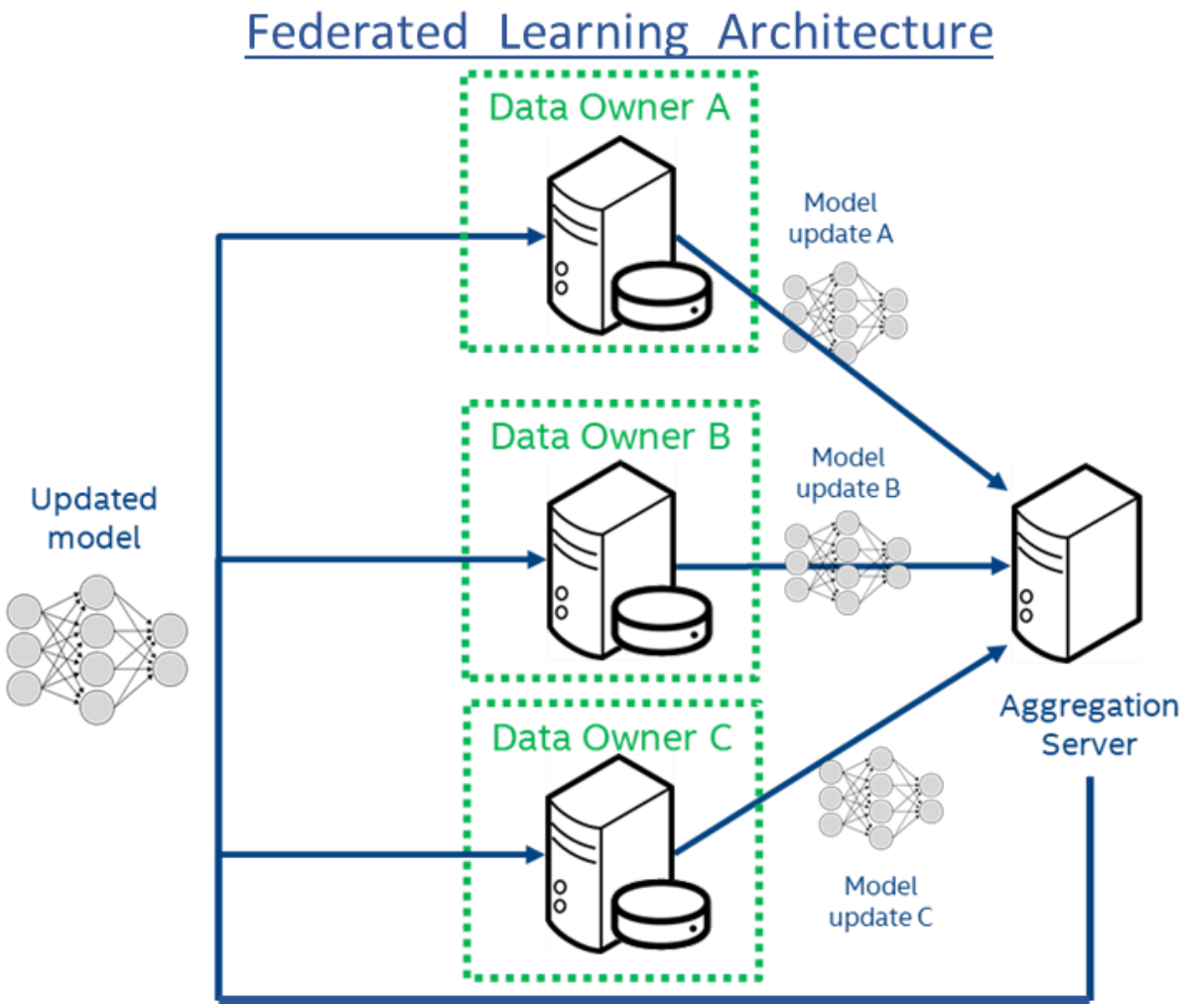
Federated Machine Learning:

Concept and Applications

QIANG YANG, Hong Kong University of Science and Technology, Hong Kong
YANG LIU and TIANJIAN CHEN, Webank, China
YONGXIN TONG, Beihang University, China

ACM Transactions on Intelligent Systems and Technology 2019

Introduction



N data owners $\{\mathcal{F}_1, \ldots \mathcal{F}_N\}$

their respective data $\{\mathcal{D}_1, \ldots \mathcal{D}_N\}$

Traditional way:

Use $\mathcal{D} = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_N$ to train a model $\mathcal{M}_{SUM}$

Fedetated learning:

All data owners train a common model $\mathcal{M}_{FED}$ collaboratively, without exposing their data to the others.

If $|\mathcal{V}_{FED} - \mathcal{V}_{SUM}| < \delta$,

algorithm has δ-accuracy loss.

- **Secure Multiparty Computation**

  Design a computation protocols so that each party knows nothing except its input and output. (Usually computational expensive)

- **Differential Privacy**

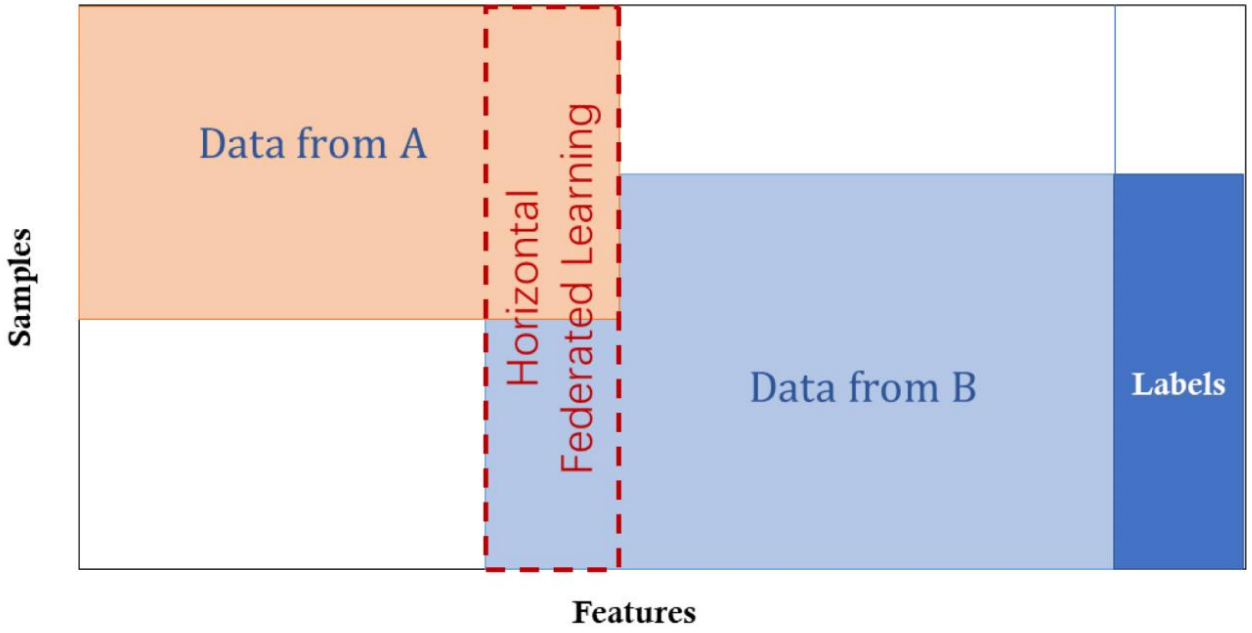
  Involve adding noise to the data, or using generalization methods to obscure certain sensitive attributes until the third party cannot distinguish the individual, thereby making the data impossible to be restore to protect user privacy

- **Homomorphic Encryption**

  Additive homomorphic: $Enc(A) + Enc(B) = Enc(A + B)$

Types of methods

- **Horizontal Federated Learning**

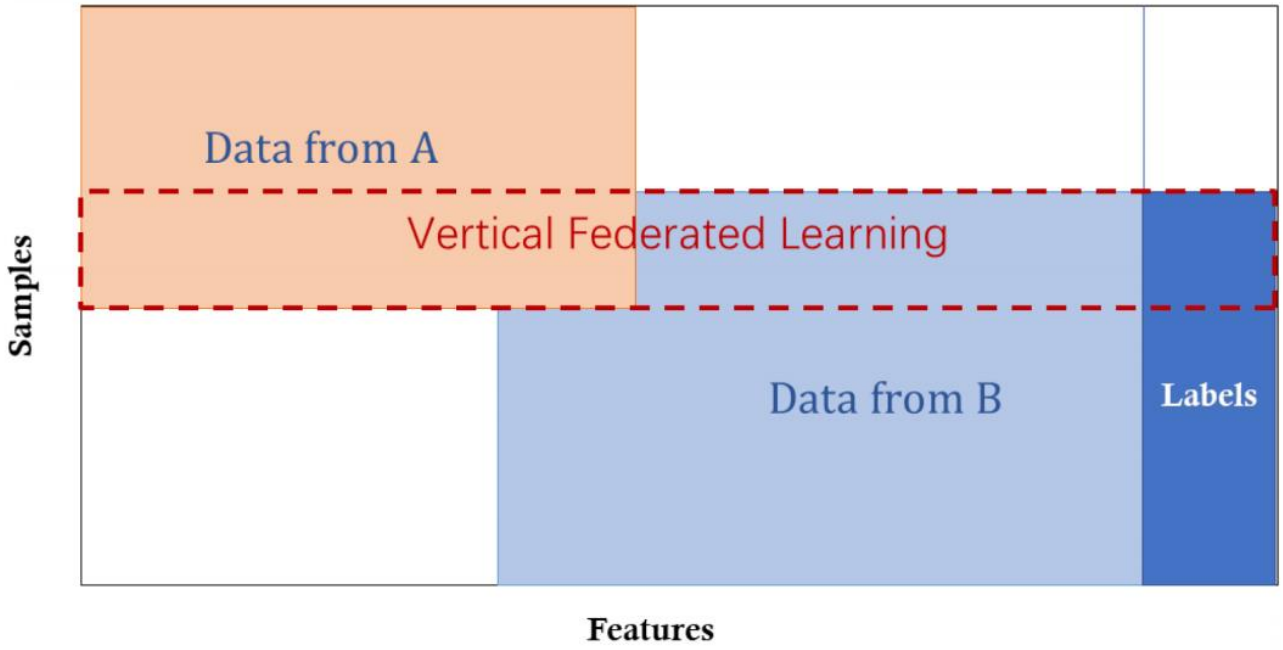


$$\mathcal{X}_i = \mathcal{X}_j, \quad \mathcal{Y}_i = \mathcal{Y}_j, \quad I_i \neq I_j, \quad \forall \mathcal{D}_i, \mathcal{D}_j, i \neq j.$$

Example: Two regional banks may have very different user groups from their respective regions, and the intersection set of their users is very small. However, their business is very similar, so the feature spaces are the same.

Types of methods

- **Vertical Federated Learning**

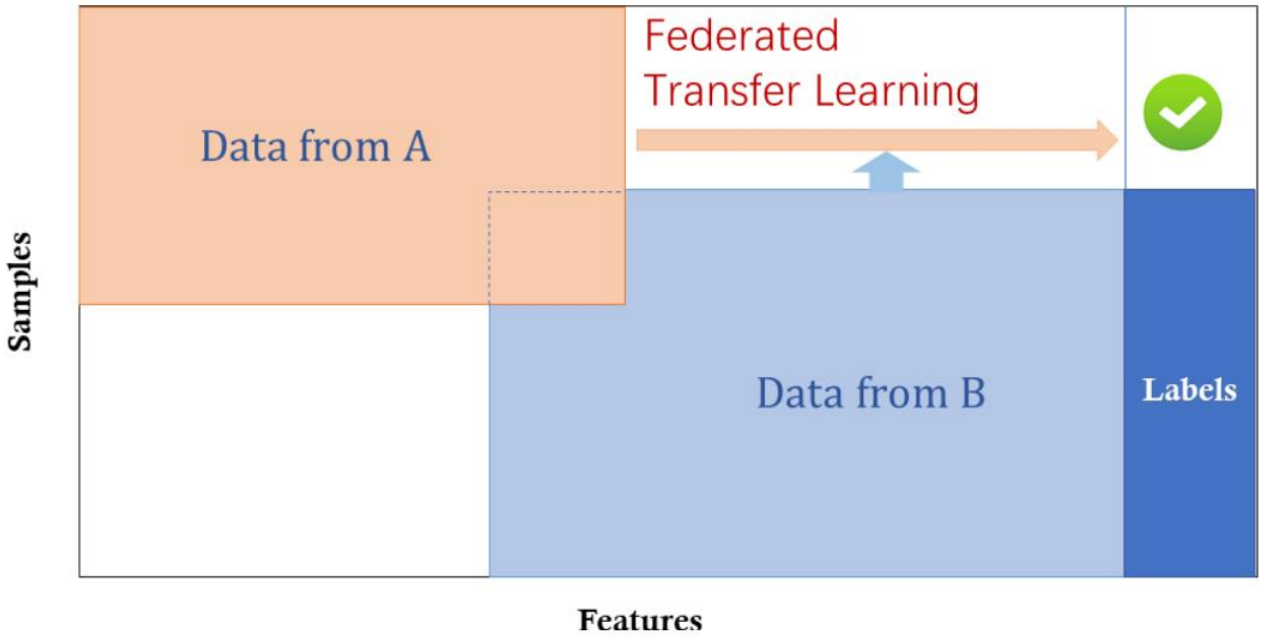


$$\mathcal{X}_i \neq \mathcal{X}_j, \quad \mathcal{Y}_i \neq \mathcal{Y}_j, \quad I_i = I_j \quad \forall \mathcal{D}_i, \mathcal{D}_j, i \neq j.$$

Example: Consider two different companies in the same city: one is a bank and the other is an e-commerce company. Their user sets are likely to contain most of the residents of the area; thus, the intersection of their user space is large.

Types of methods

- **Federated Transfer Learning**

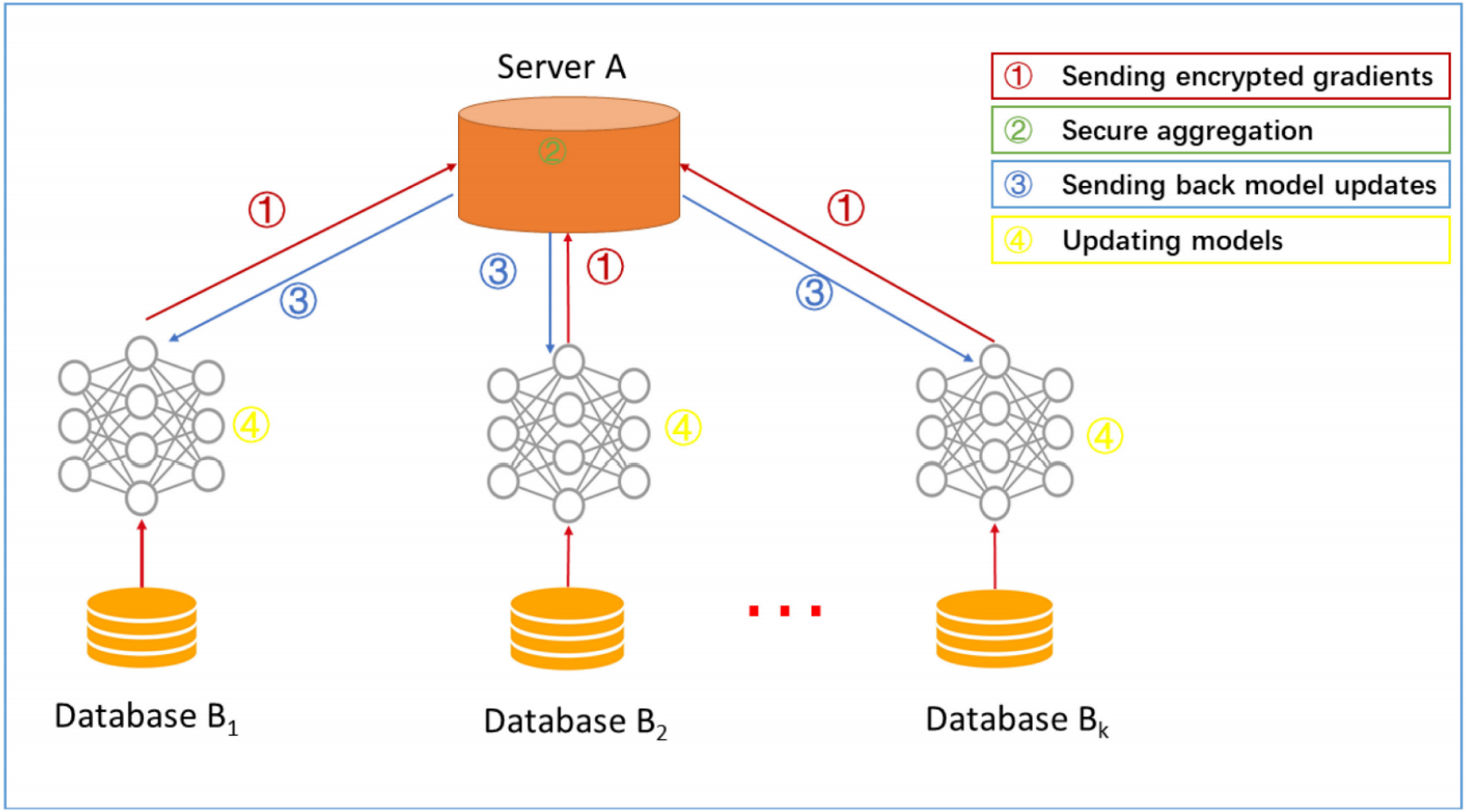


$$\mathcal{X}_i \neq \mathcal{X}_j, \quad \mathcal{Y}_i \neq \mathcal{Y}_j, \quad I_i \neq I_j \quad \forall \mathcal{D}_i, \mathcal{D}_j, i \neq j.$$

Example: Consider two institutions: one is a bank located in China and the other is an e-commerce company located in the United States. Owing to geographical restrictions, the user groups of the two institutions have a small intersection.

Common architecture

- **Horizontal Federated Learning**



All participants will share the final model parameters

◆ Divide existing dataset into shards

- MNIST
- CIFAR
- …

◆ Using the datasets collected by multiple individuals

- **Human Activity Recognition**[5]: Mobile phone accelerometer and gyroscope data collected from 30 individuals, performing one of six activities: {*walking, walking-upstairs, walking-downstairs, sitting, standing, lying-down*}. We use the provided 561-length feature vectors of time and frequency domain variables generated for each instance [3]. We model each individual as a separate task and predict between sitting and the other activities.
- **Vehicle Sensor**[6]: Acoustic, seismic, and infrared sensor data collected from a distributed network of 23 sensors, deployed with the aim of classifying vehicles driving by a segment of road [13]. Each instance is described by 50 acoustic and 50 seismic features. We model each sensor as a separate task and predict between AAV-type and DW-type vehicles.

Method: model selection

- **Vertical Federated Learning**
  (requires a third party collaborator)

1. Identify the common entities

Confirm the common users of both parties without A and B exposing their respective data. The system does not expose users that do not overlap with each other.

2. Encrypted model training

$$\min_{\Theta_A,\Theta_B} \sum_i ||\Theta_A x_i^A + \Theta_B x_i^B - y_i||^2 + \frac{\lambda}{2}(||\Theta_A||^2 + ||\Theta_B||^2).$$

At the end of learning, each party holds only those model parameters associated to its own features. Therefore, at inference time, the two parties also need to collaborate to generate output.

Method: model selection

- **Vertical Federated Learning**

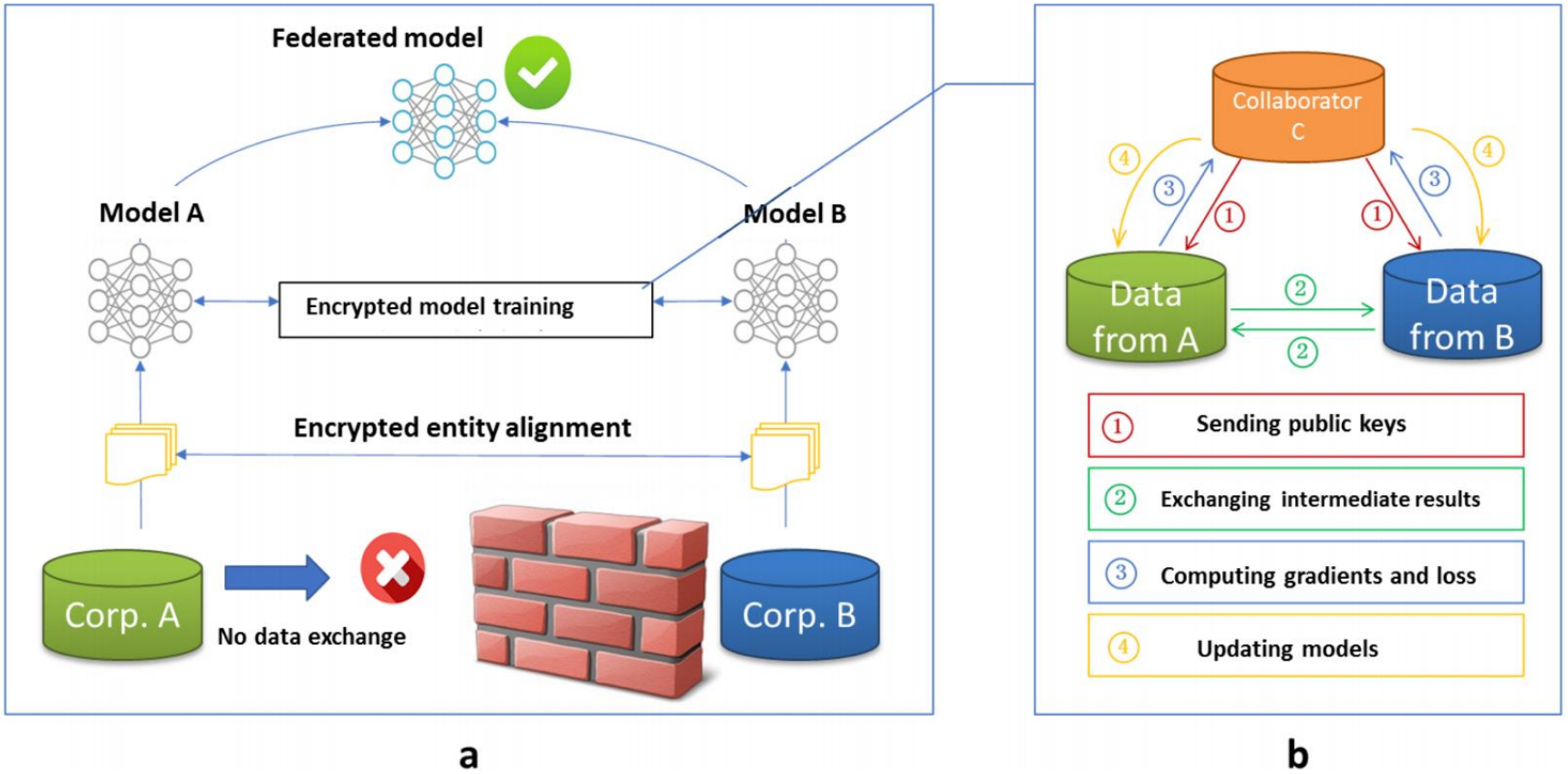


Fig. 4. Architecture for a vertical federated-learning system.

- **Federated Transfer Learning**

1. Share the architecture with vertical federated learning

2. Try to learn a common representation and minimize the errors in predicting the labels for the target-domain party by leveraging the labels in the source-domain party (B in this case).

3. At inference time, it still requires both parties to compute the prediction results

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

THANKS