Minimum Exposure Approach for Trustworthy Vertical Federated Learning

Dashan GAO

Supervisor: Prof. Qiang YANG

Co-supervisor: Prof. Xin YAO

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

SUSTech Southern University of Science and Technology

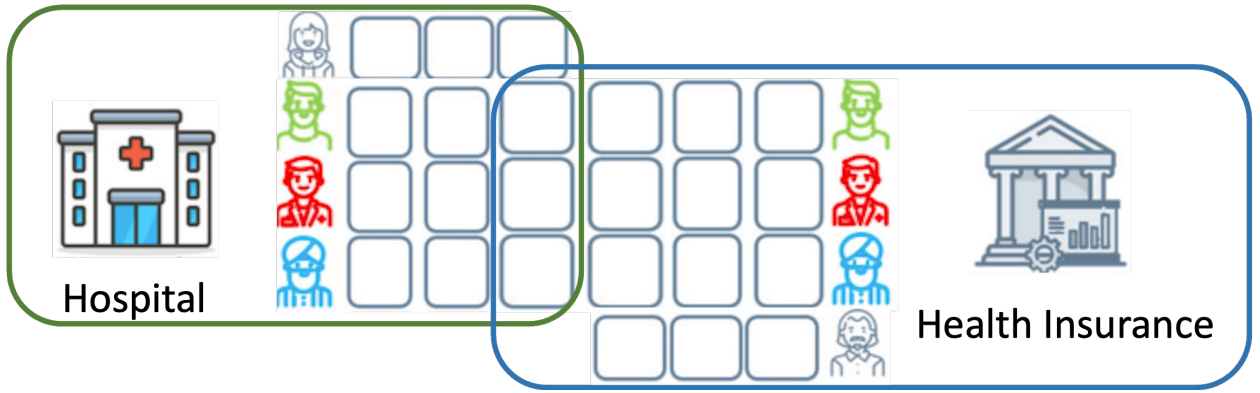
Contents

Background

- Data privacy attracts growing attention.
- Strict regulations such as GDPR are formulated.

Background

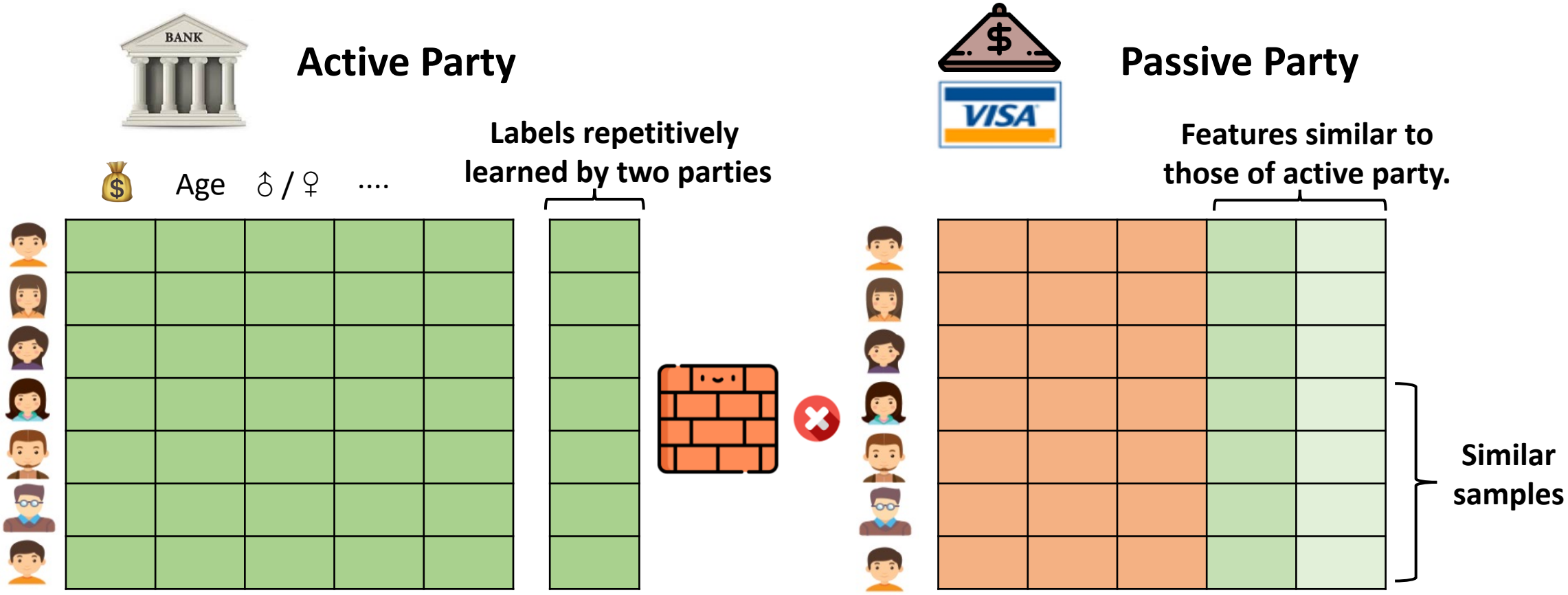
- Federated Learning [McMahan et al., 2016] is a paradigm that enables learning from separated parties in a privacy-preserving manner.
  - Data are distributed and privacy-sensitive.
  - Labeled data are insufficient in each single party.
- Vertical Federated Learning (VFL) was proposed by [Yang et al., 2019].
  - Different feature spaces in each party.
  - Common in cross-enterprise scenarios.



H. Brendan McMahan, Eider Moore, Daniel Ramage, and Blaise Agüeray Arcas. Federated learning of deep networks using model averaging. CoRR,abs/1602.05629, 2016.
Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. ACM Trans. Intell. Syst. Technol., 10(2), Jan 2019.

Example: Information Exposure in Naive VFL

[Visa Research, 2023]

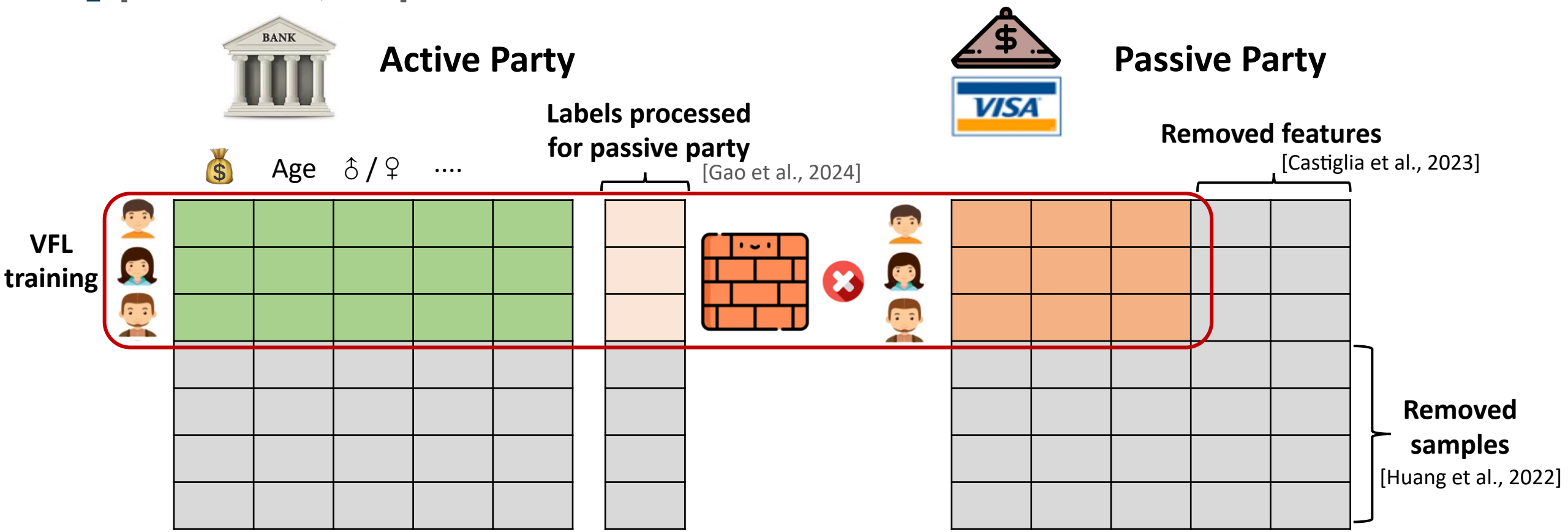


Naive VFL trains on the **entire dataset**, including similar samples and features. The exposure of unnecessary information leads to privacy risks and inefficiencies without extra utility gains.

Visa Research. (2023). Secure collaborative machine learning. https://usa.visa.com/dam/VCOM/regional/na/us/about-visa/research/documents/secure-collaborative-machine-learning.pdf

Example: Information Exposure in Ideal VFL

[Visa Research, 2023]



Ideal VFL **minimizes the unnecessary information exposure** by removing similar samples and features while maintaining utility.

- Castiglia, T., Zhou, Y., Wang, S., Kadhe, S., Baracaldo, N., & Patterson, S. (2023). LESS-VFL: Communication-Efficient Feature Selection for Vertical Federated Learning. Proceedings of the 40th International Conference on Machine Learning, 202:3757-3781.
- Huang,L.,Li,Z.,Sun,J.,Zhao,H.:Coresetsforverticalfederatedlearning:Regularizedlin- ear regression and k-means clustering. Advances in Neural Information Processing Systems 35, 29566–29581 (2022)

Guiding Principle in This Thesis

- **"Everything should be made as simple as possible, but not simpler."**

  — Albert Einstein

Explanation in trustworthy VFL:

- Excessive information exposure does not improve utility but only increases privacy and efficiency risks.
- Conversely, insufficient information exposure can harm utility, leading to inevitable utility-privacy-efficiency trade-offs.
- Therefore, we pursue **Minimum-Necessary Information Exposure (MNIE).**

Contents

Vertical Federated Learning – Definition

extending from [Yang et al., 2019]

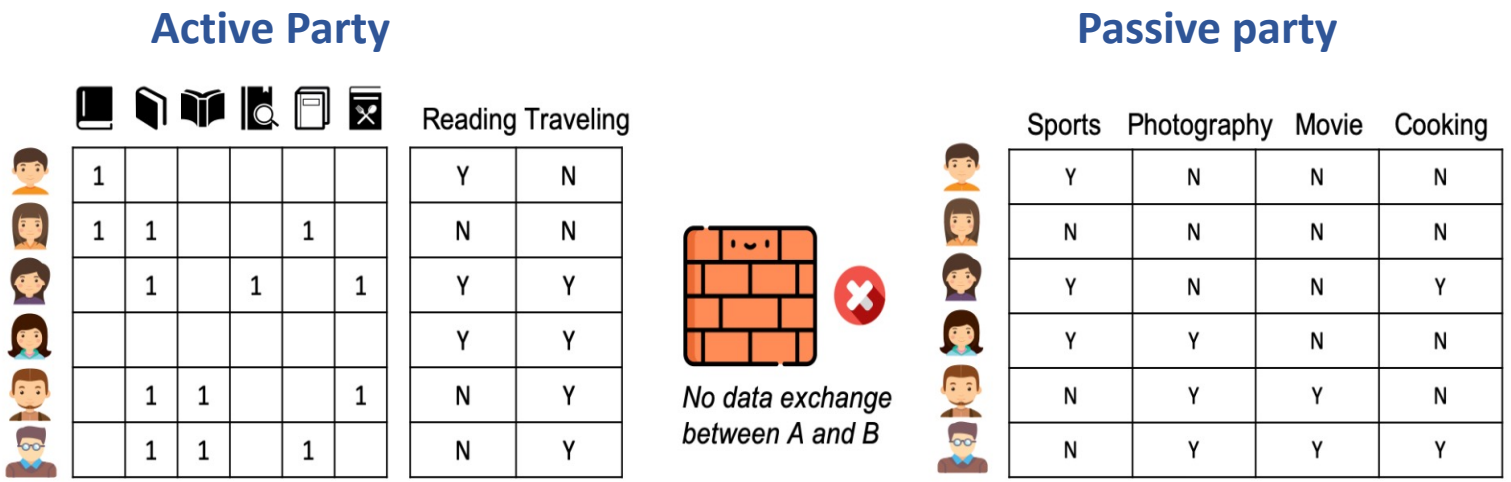
- N parties $\{P_1, \dots, P_N\}$ with isolated datasets $\{\mathcal{D}_1, \dots, \mathcal{D}_N\}$
- A dataset $\mathcal{D}$ consists of three components:
  - Feature space $\mathcal{X}$
  - Label space $\mathcal{Y}$
  - ID space $\mathcal{I}$
- The three components can be different among parties.
  - Active party: $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}, \mathcal{I}\}$ Can do on-site learning.
  - Passive party: $\mathcal{D} = \{\mathcal{X}, \mathcal{I}\}$ Can only provide auxiliary features.
- A model $\mathcal{M}_{fed}$ is learned using a federated learning approach on the isolated datasets.

Multi-objective Trade-offs in VFL

- There are multi-objective trade-offs in trustworthy VFL:

1. **Utility**: improve the model performance.
2. **Efficiency**: efficiently train the VFL model.
3. **Privacy**: protect data asset and user privacy.

This thesis aims to find unified approaches to improve these objectives.



Definition of Minimum-Necessary Information Exposure (MNIE)

- We formally define the concept of **minimum-necessary information exposure in trustworthy VFL**.

Definition 1.5.3 (Minimum-Necessary Information Exposure (MNIE)). *Given the original global dataset* $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_K$ *held by* K *parties, find a transformed dataset* $\mathcal{D}'$ *that minimizes the information exposure* $E(\mathcal{D}')$ *while ensuring that the utility of the model trained on* $\mathcal{D}'$ *is within an acceptable threshold* ε *of the utility on the original dataset* $\mathcal{D}$. *Formally:*

$$\begin{aligned} &\underset{\mathcal{D}'}{\textit{Minimize}} && E(\mathcal{D}') && \textit{(minimum)} \\ &\textit{Subject to} && \min_w \mathcal{L}(w, \mathcal{D}') \leq \min_w \mathcal{L}(w, \mathcal{D}) + \varepsilon && \textit{(necessary),} \end{aligned}$$

where:

- $\mathcal{L}(w, \mathcal{D})$ and $\mathcal{L}(w, \mathcal{D}')$ are the loss functions evaluated on datasets $\mathcal{D}$ and $\mathcal{D}'$.
- $\varepsilon \geq 0$ is a small constant representing an acceptable increase in loss.

Our Works Categorized from Two Perspectives

Our works categorized from **transfer learning perspective.**

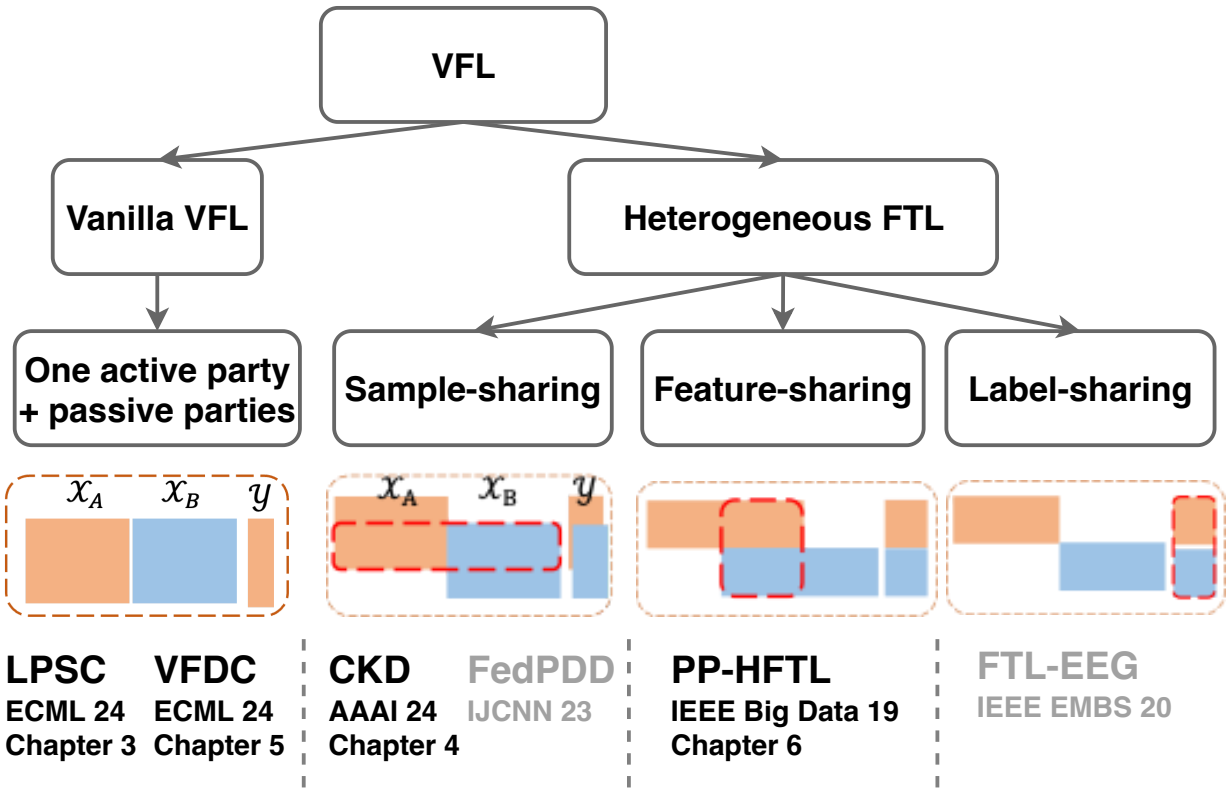


Figure 2.1

Our works categorized based on the **information exposure** perspective. This categorization is followed throughout the presentation.

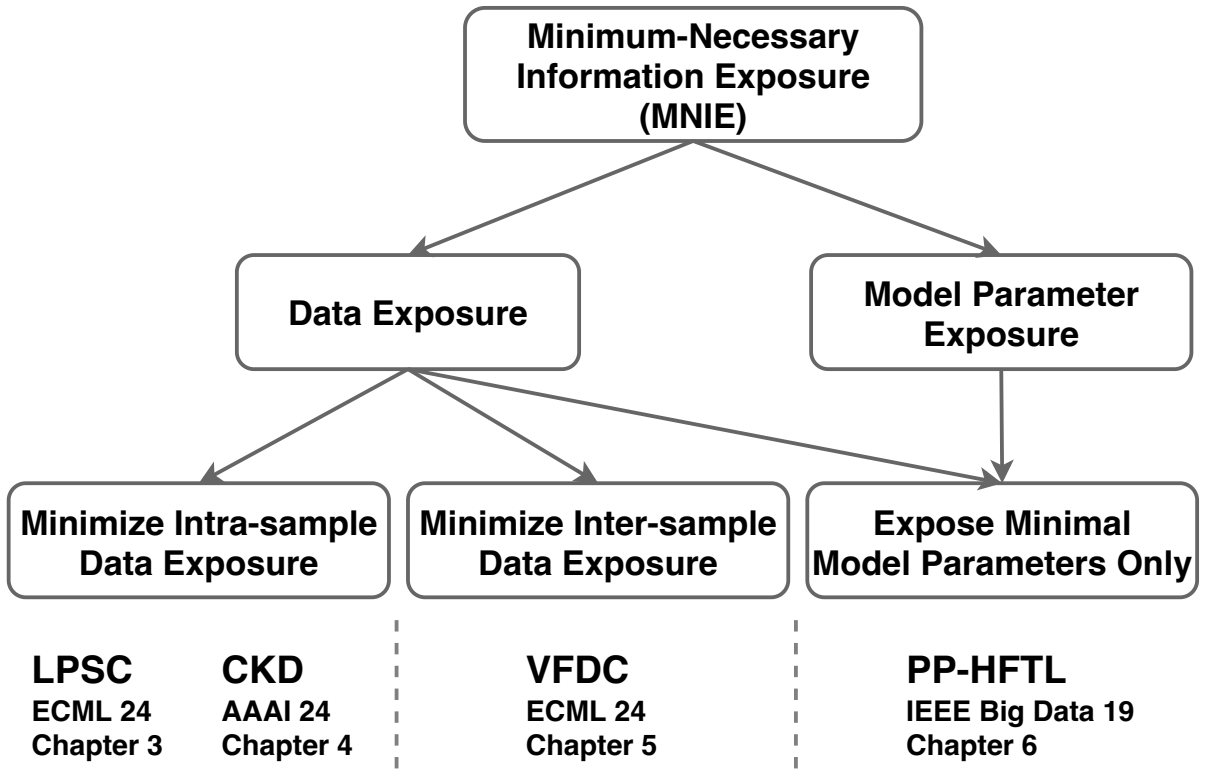


Figure 2.2

Minimize Intra-Sample Label Exposure
(LPSC Chapter 3, CKD Chapter 4)

- Encode original labels to minimize label information exposure for privacy.
- Ensure effective VFL model training to maintain utility.
- Adapted MNIE objective function:

$$\underset{\mathcal{D}',y'}{\text{Minimize}} \quad E(y') \qquad \textit{(minimum label exposure)}$$

$$\text{Subject to} \quad \min_{w} \mathcal{L}(w, \mathcal{D}', y') \leq \min_{w} \mathcal{L}(w, \mathcal{D}, y) + \varepsilon \qquad \textit{(necessary utility)}$$

where:

- y' represents the transformed (encoded) labels to ensure privacy.
- $E(y')$ quantifies the exposure of the transformed labels y', focusing specifically on minimizing intra-sample label information exposure.

Minimize Inter-Sample Data Exposure by Dataset Condensation (VFDC, Chapter 5)

- Generate condensed synthetic dataset for privacy and efficiency improvement while maintaining utility.
- Adapted MNIE objective function:

$$\underset{D'}{\text{Minimize}} \quad E(D') \qquad \textit{(minimum inter-sample exposure)}$$

$$\text{Subject to} \quad \min_{w} \mathcal{L}(w, D') \leq \min_{w} \mathcal{L}(w, D) + \varepsilon \qquad \textit{(necessary utility)}$$

where:

- D' represents the **condensed synthetic dataset** that retains essential information while reducing the number of samples.
- $E(D')$ quantifies the data exposure of the synthetic dataset D', focusing specifically on minimizing exposure by reducing **the number of samples**.

Expose Model Parameters Only via Secure Computation (PP-HFTL, Chapter 6)

- Directly transform original dataset to model parameters via secure computation, to achieve high utility and strong privacy protection.
- Adapted MNIE objective function:

$$\text{Minimize} \quad \underset{w}{} \quad E(w) \qquad \textit{(minimum model parameter exposure)}$$

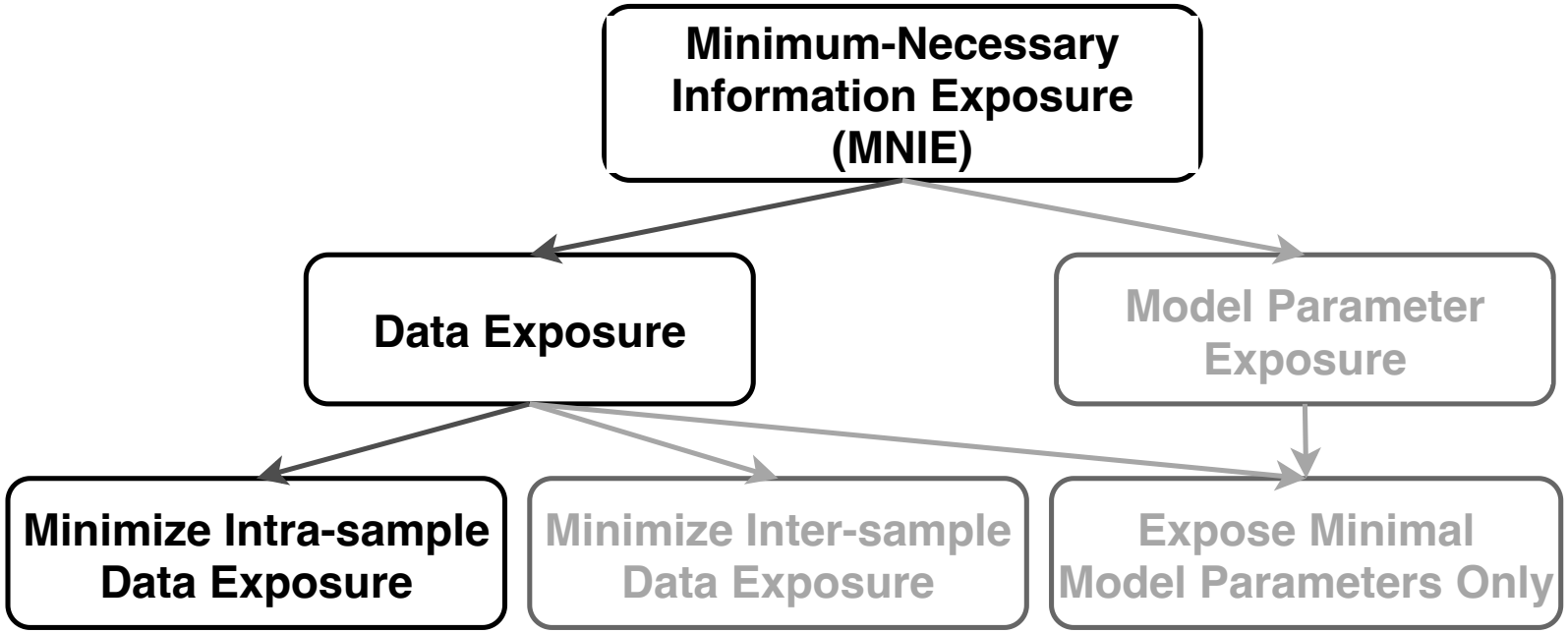
$$\text{Subject to} \quad \min_{w} \mathcal{L}(w) \leq \min_{w} \mathcal{L}(w, \mathcal{D}) + \varepsilon \qquad \textit{(necessary utility)}$$

where:

- $w = \text{Transform}(\mathcal{D})$ represents the model parameters directly derived from the original dataset $\mathcal{D}$, effectively reducing the exposure of intermediate data.
- $E(w)$ quantifies the exposure of the model parameters, ensuring that only the minimum-necessary information from $\mathcal{D}$ is retained in w.

Contents

3. LPSC: Label Privacy Source Coding in VFL (ECML PKDD 2024)



LPSC
ECML 24
Chapter 3

CKD
AAAI 24
Chapter 4

VFDC
ECML 24
Chapter 5

PP-HFTL
IEEE Big Data 19
Chapter 6

Vanilla VFL and Privacy Threats

Existing methods [Li et al., 2022, Fu et al., 2022]:

- Adopt a model-splitting paradigm:
  - Top model protects label.
  - Bottom models protect features.
- Dilemmas: existing perturbation methods leads to **privacy-utility trade-off**.
- Fundamental cause: directly train forward embeddings to fit labels, which contains **unnecessary information exposure**.

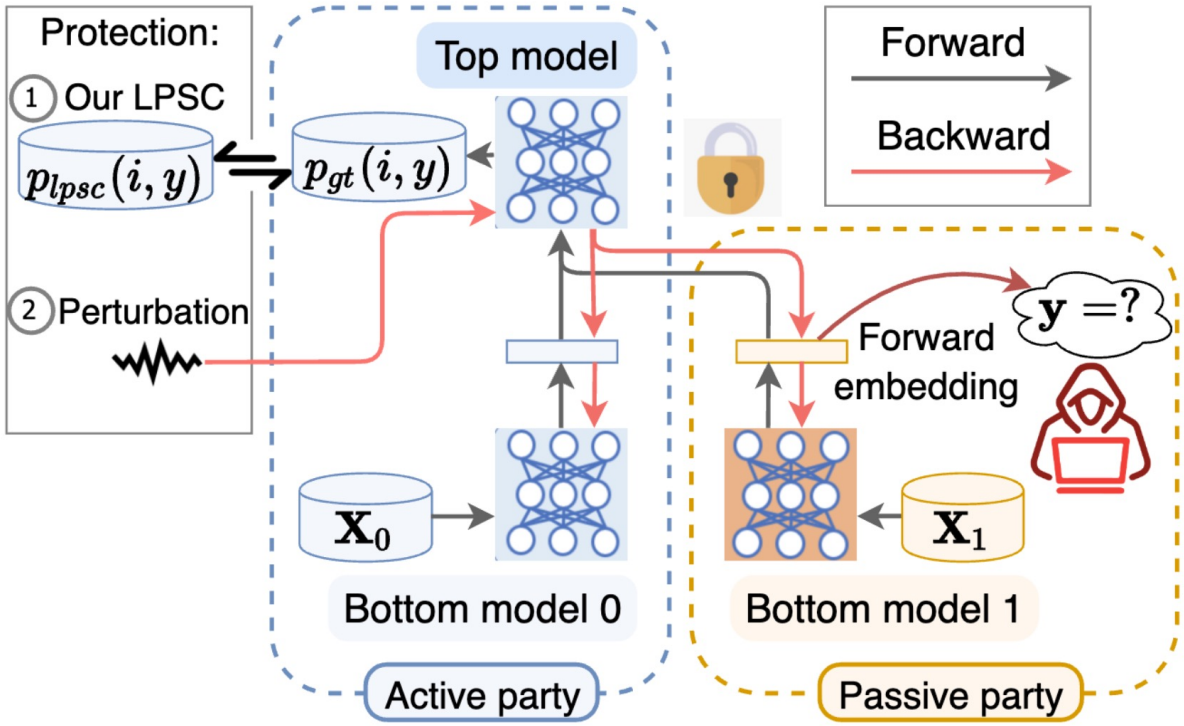


Figure 3.1: Vanilla VFL and label privacy threat

- Oscar Li, Jiankai Sun, Xin Yang, Weihao Gao, Hongyi Zhang, Junyuan Xie, Virginia, Smith, and Chong Wang. Label leakage and protection in two-party split learning. International Conference on Learning Representations (ICLR), 2022.
- Chong Fu, Xuhong Zhang, Shouling Ji, Jinyin Chen, Jingzheng Wu, Shanqing Guo, Jun Zhou, Alex X. Liu, and Ting Wang. Label inference attacks against vertical federated learning. In 31st USENIX Security Symposium (USENIX Security 22), pages 1397–1414, Boston, MA, August 2022. USENIX Association.

Preliminary: Privacy

- **Private label information** is defined as the ID-label joint distribution.

Definition 1 (Private Label Information). *In VFL, the private label information that the active party aims to protect is defined as the dataset's ID-label joint distribution* $p_{gt}(i, y)$.

- **Mutual Information Privacy (MIP)** can be used to measure the label information leaked to the passive party in VFL.

Definition 2 (ϵ-MIP). *A mechanism $\mathcal{M}$ satisfies ϵ-MIP if the mutual information between any input X and the output Y is limited to ϵ bits, formally:*

$$I(X;Y) \leq \epsilon \text{ bits}.$$

In line with this, our goal is to ensure that LPSC satisfies to ϵ-MIP, effectively protecting label privacy in the offline phase.

Insight of Label Privacy Source Coding (LPSC)

- We aim to propose an optimization problem to encode the **minimum-necessary label information** in the original label information.
- Therefore, we propose Label Privacy Source Coding (LPSC).

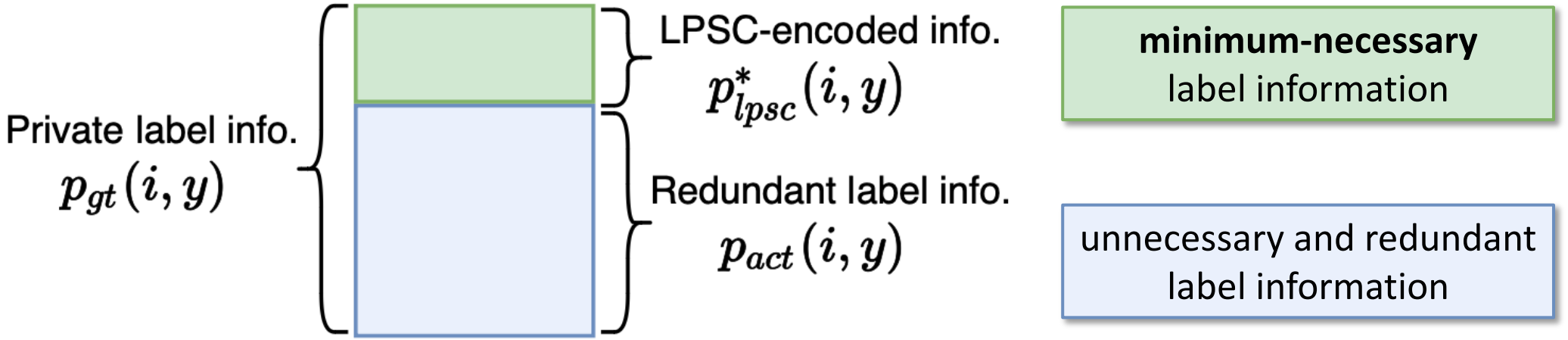


Figure 3.2: The schematic graph of LPSC.

Label Privacy Source Coding Problem

- **We encode minimum-necessary label information as follows:**

Definition 3 (Label Privacy Source Coding). *Given the ground-truth private label information $p_{gt}(i,y)$ and the active party P_0's learned private label information $p_{act}(i,y)$ from its features $\boldsymbol{X}_0$, the label privacy source coding problem is to optimize a new joint distribution $p_{lpsc}(i,y)$ as follows:*

$$\begin{aligned} &\max_{p_{lpsc}(i,y)} \; I(p_{gt}(i,y); p_{lpsc}(i,y)) && \textit{(necessary, utility)} \\ &s.t. \;\; I(p_{act}(i,y); p_{lpsc}(i,y)) = 0 && \textit{(minimum, privacy)} \end{aligned} \tag{2}$$

where $I(\cdot;\cdot)$ denotes mutual information.

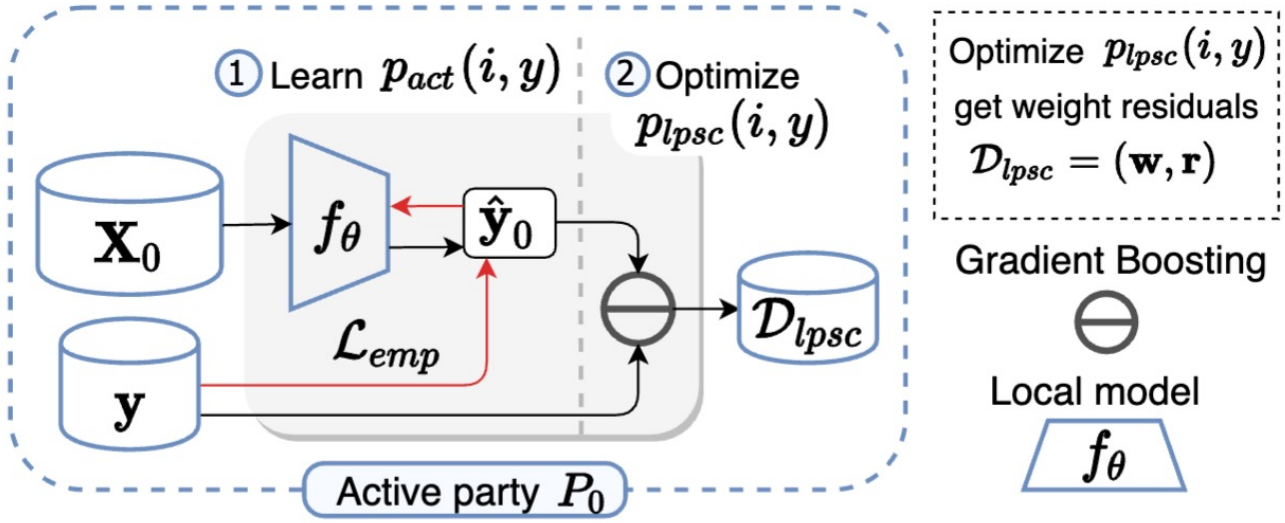
- **Mutual Information Privacy (MIP)-based privacy guarantee**

Theorem 1 (Privacy Guarantee). *LPSC satisfies ϵ-MIP. The privacy leakage is bounded by $\epsilon = H(p_{gt}(i,y)|p_{act}(i,y))$, the conditional entropy of the ground-truth label distribution $p_{gt}(i,y)$ given the active party's label distribution $p_{act}(i,y)$. Formally,*

$$I(p_{gt}(i,y); p^*_{lpsc}(i,y)) \leq \epsilon \; bits,$$

where $p^*_{lpsc}(i,y)$ represents the optimal solution of Eq. 2 in the LPSC problem.

Label Privacy Source Coding



An overview of the label privacy source coding (LPSC).

The active party P_0:

1) trains a local model f_θ on its labeled data to learn $p_{act}(i,y)$.
2) optimizes the $p_{lpsc}(i,y)$ via gradient boosting.

Gradient Boosting Solves LPSC

We prove that gradient boosting can efficiently optimize LPSC.

$$\max_{\mathcal{D}_{lpsc}} \ I(\mathcal{D}_{gt}; \mathcal{D}_{lpsc}) \qquad \textit{(necessary, utility)}$$

$$s.t. \ \ I(\mathcal{D}_{act}; \mathcal{D}_{lpsc}) = 0 \qquad \textit{(minimum, privacy)}$$

1) Learning $\mathcal{D}_{act}$: To learn $\mathcal{D}_{act}$, which is the label privacy contained in local features X_0, the active party P_0 only needs to learn label $\mathcal{D}_{act}(Y|I)$ as the sample weight $\mathcal{D}_{act}(I) \sim U$ is a uniform distribution. To do so, P_0 trains model f_θ on its labeled data $\mathcal{D}^{loc} = \{X_0^{loc}, Y^{loc}\}$ as follows:

$$\theta^* = \arg\min_{\theta} \ \mathbb{E}_{(\boldsymbol{x}_0^{loc}, y^{loc}) \sim \mathcal{D}^{loc}} \left[\mathcal{L}_{emp}(y^{loc}, f_\theta(\boldsymbol{x}_0^{loc})) \right], \quad (4)$$

where $\mathcal{L}_{emp}$ denotes the empirical loss. Thereby, P_0 gets its learned label privacy $\mathcal{D}_{act} = D(I, \hat{Y}_0) = D(I, f_\theta(X_0))$.

2) Optimizing $\mathcal{D}_{lpsc}$: We point out that the *gradient boosting algorithm optimizes the LPSC problem* by taking AdaBoost [12] as an example.

Theorem IV.3. *Assuming fixed conditional distribution $\mathcal{D}_{lpsc}(Y|I) = \mathcal{D}_{gt}(Y|I)$ and let U denote uniform distribution, the LPSC problem IV.1 can be reduced to:*

$$\min_{\mathcal{D}_{lpsc}(I)} \ D_{KL}(\mathcal{D}_{lpsc}(I) \,||\, U) \qquad (5)$$

$$s.t. \ \ \sum_{i=1}^{|I|} \mathcal{D}_{lpsc}(i) y_i f_\theta(\boldsymbol{x}_{0,i}) = 0,$$

where i is the sample index and $|I|$ is the sample number.

Theorem IV.4. *[27] The solution of the convex optimization problem Eq. 5 is equivalent to AdaBoost [12]:*

$$\mathcal{D}_{lpsc}(i) = \frac{e^{-\alpha y_i f_\theta(\boldsymbol{x}_{0,i})}}{\sum_{i=1}^{|I|} e^{-\alpha y_i f_\theta(\boldsymbol{x}_{0,i})}},$$

where $\alpha = \frac{1}{2}\ln(\frac{1-\epsilon}{\epsilon})$ and ϵ is the classification error of f_θ. $\mathcal{D}_{lpsc}(I)$ can be computed in $O(|I|)$ time-complexity.

LPSC+ Framework for Flexible Privacy-Utility Trade-off

- LPSC provides upper-bounded label privacy protection, which may not achieve satisfactory privacy protection.
- Therefore, we introduce **adversarial training** to enable flexible privacy-utility trade-off and provide stronger privacy protection.

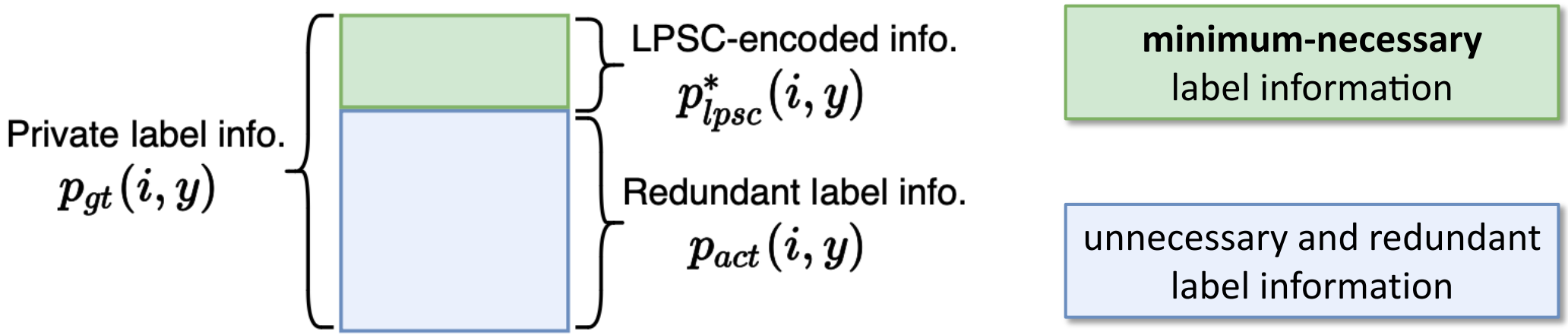


Figure 3.2: The schematic graph of LPSC.

LPSC+ Framework

The framework of LPSC+Adv with two phases:

1. **Offline phase:** LPSC encodes minimum-necessary label information.
2. **Federated training phase:** a federated model is trained via a *utility loss and privacy loss* to enable trading utility for extra privacy.

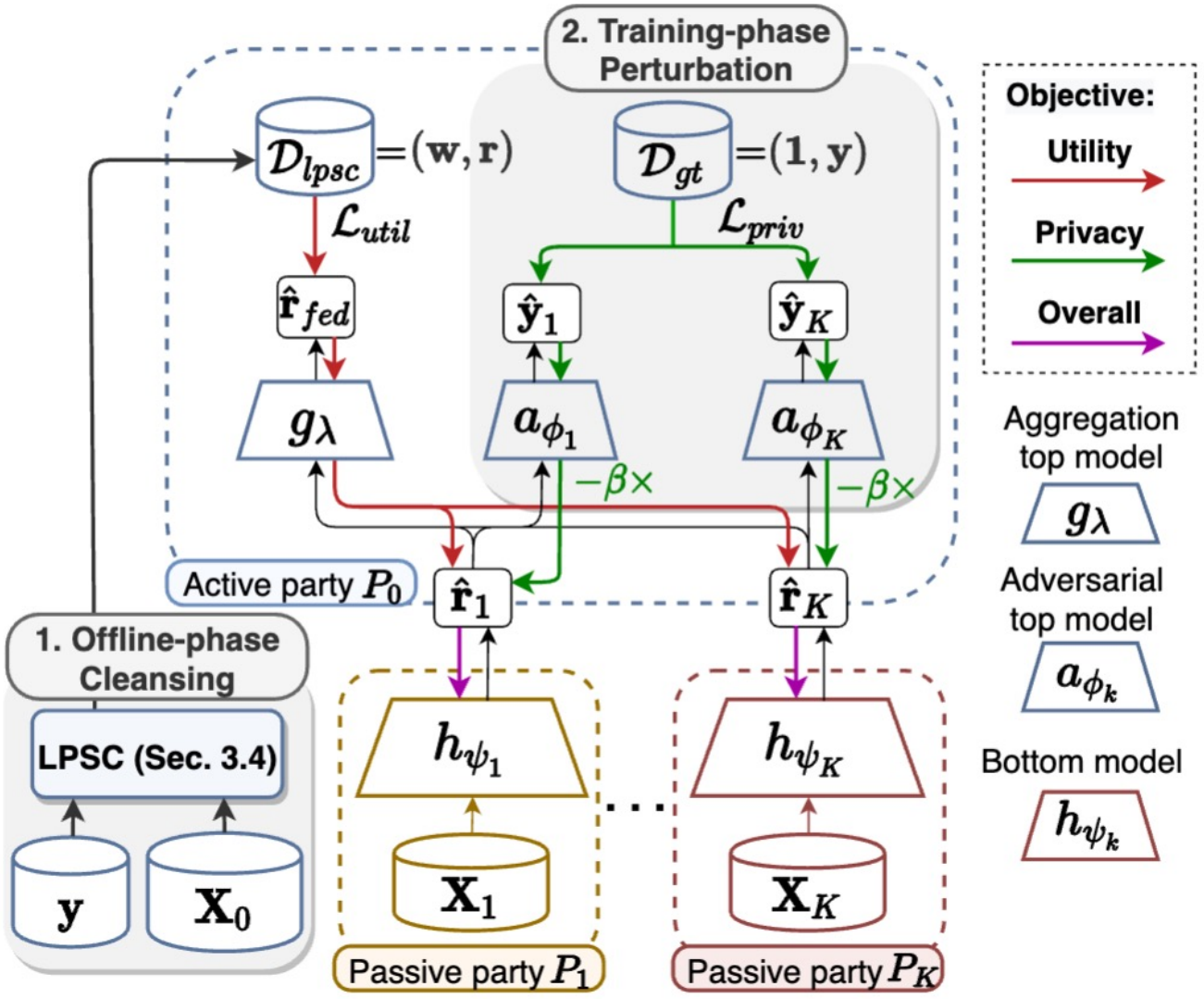


Figure 3.3

LPSC+Adv Framework Objectives

LPSC Utility Objective

$$\min_{\lambda,\{\psi_k\}_{k=1}^{K}} \sum_{i \in \boldsymbol{i}} w_i \cdot \mathcal{L}_{util}\left(r_i, h_{fed}(i)\right)$$

$(w_i, r_i) \in \mathcal{D}_{lpsc}$: LPSC-encoded weight-residuals.

Adversarial Privacy Objective

$$\max_{\psi_k} \min_{\phi_k} \mathbb{E}_{i \sim p_{gt}(i)} \left[\mathcal{L}_{priv(k)}\left(y_i,\ a_{\phi_k} \circ h_{\psi_k}(i)\right)\right]$$

$$s.t.\ \forall k \in [1, \ldots, K],$$

$\{a_{\phi_k}\}_{k=1}^{K}$: adversarial top models

$\{h_{\psi_k}\}_{k=1}^{K}$: bottom models

Overall Objective

$$\min_{\lambda,\{\psi_k\}_{k=1}^{K}} \max_{\{\phi_k\}_{k=1}^{K}} \left\{ \underbrace{\sum_{i \in \boldsymbol{i}} w_i \cdot \mathcal{L}_{util}\left(r_i, h_{fed}(i)\right)}_{\text{LPSC utility objective}} - \beta \cdot \underbrace{\sum_{k=1}^{K} \sum_{i \in \boldsymbol{i}} \frac{1}{|\boldsymbol{i}|} \cdot \mathcal{L}_{priv(k)}\left(y_i,\ a_{\phi_k} \circ h_{\psi_k}(i)\right)}_{\text{Adversarial privacy objective}} \right\}$$

Experiments: Visualization of Passive Party's Model Output

- Figure 3.4 visualizes the distributions of output logits by fitting **original labels v.s. LPSC-encoded labels** for 6 passive parties.
- It is **more challenging** to distinguish the output distributions between classes when the bottom models are trained with **LPSC-encoded labels**.

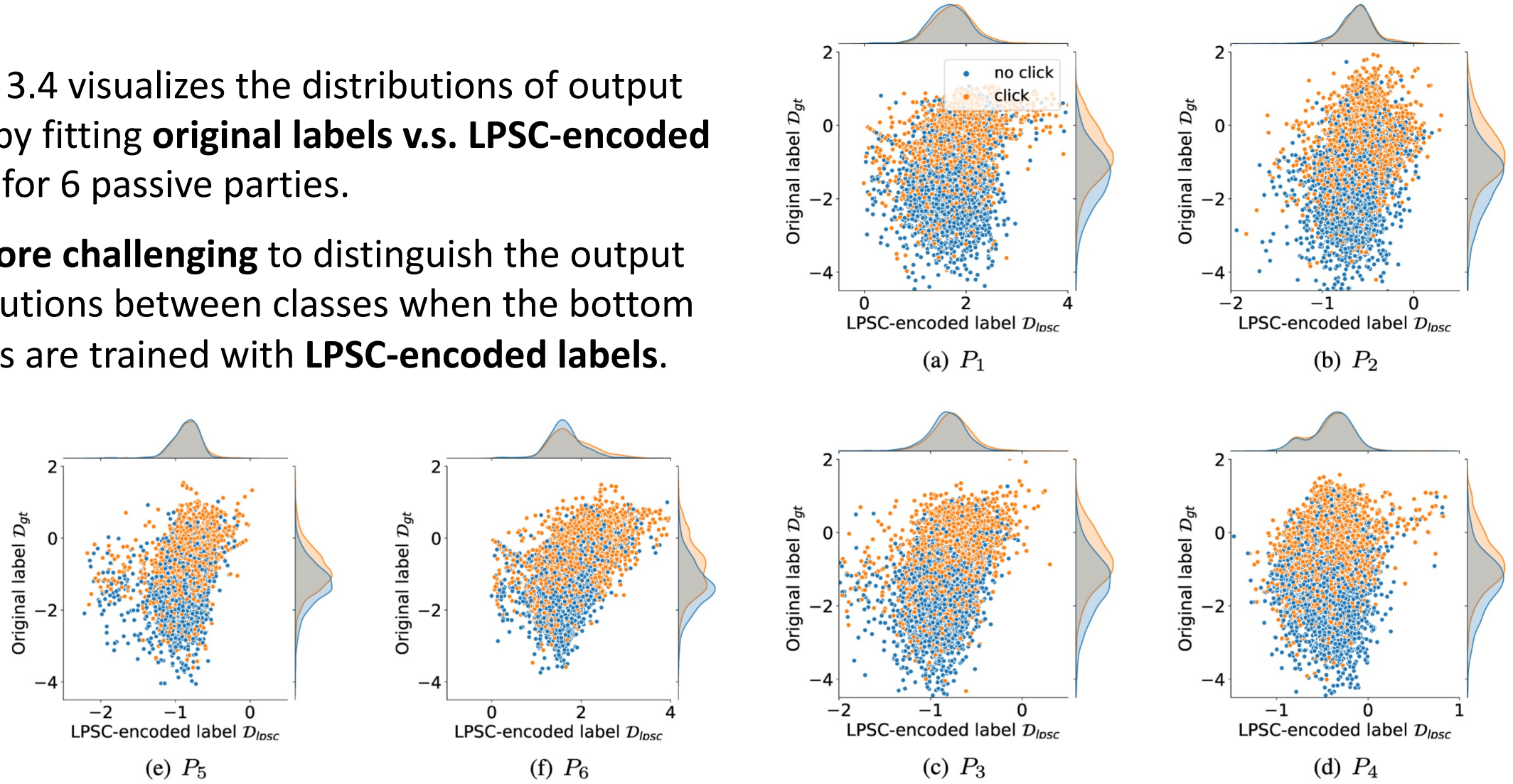


Figure 3.4

Experiments: LPSC Protects Label Privacy for Free

- Table 3.1 presents the **L**abel **L**eakage **LL-AUC** against Norm, Spectral, and PMC attacks and the FL-AUC on four datasets.
- The LL-AUC of LPSC against three attacks is significantly lower than that of the original labels. → **LPSC provides strong label privacy protection**.
- The FL-AUC of LPSC is comparable to that of the original labels. → **LPSC does not sacrifice model utility**.

Table 3.1: Comparative results of privacy and utility of VFL fitting original labels vs. LPSC-encoded labels.

Dataset	Target	Privacy (LL-AUC) ↓			Utility ↑
		Norm	Spectral	PMC	FL-AUC
Criteo	Label	0.673	0.689	0.711	0.768
	LPSC	**0.523**	**0.538**	**0.571**	0.766
Avazu	Label	0.620	0.648	0.705	0.749
	LPSC	**0.531**	**0.555**	**0.577**	0.751
MIMIC-III	Label	0.577	0.593	0.611	0.763
	LPSC	**0.528**	**0.535**	**0.558**	0.763
Cardio	Label	0.582	0.618	0.664	0.722
	LPSC	**0.517**	**0.542**	**0.567**	0.724

Experiments: Privacy-Utility Trade-off Comparison

- Figure 3.5 shows the privacy-utility trade-off curves on four datasets.
- An ideal trade-off should have a large FL-AUC and a small LL-AUC, thus in the **upper-left** corner.
- Our **LPSC** + **adversarial training** is the closest to the ideal trade-off on all four datasets.

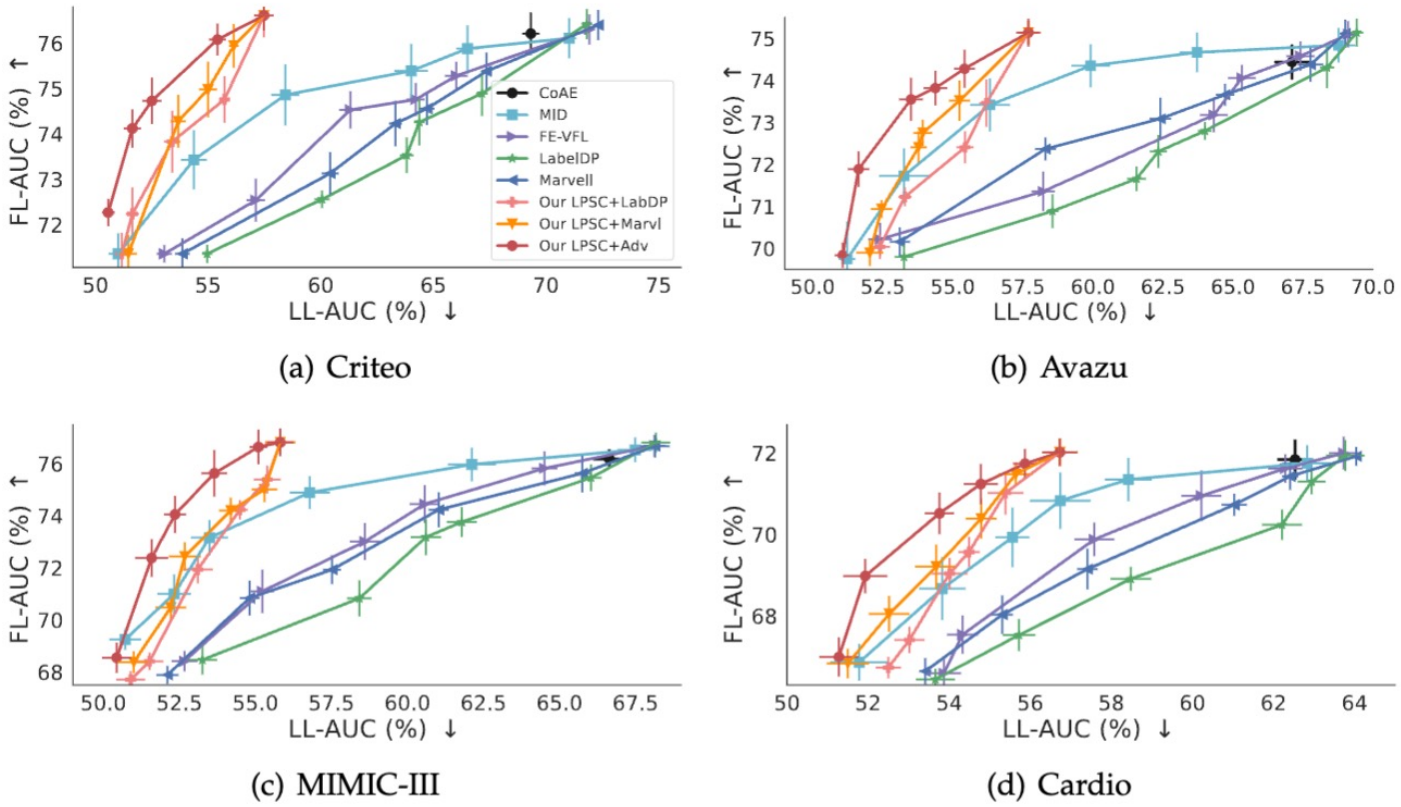


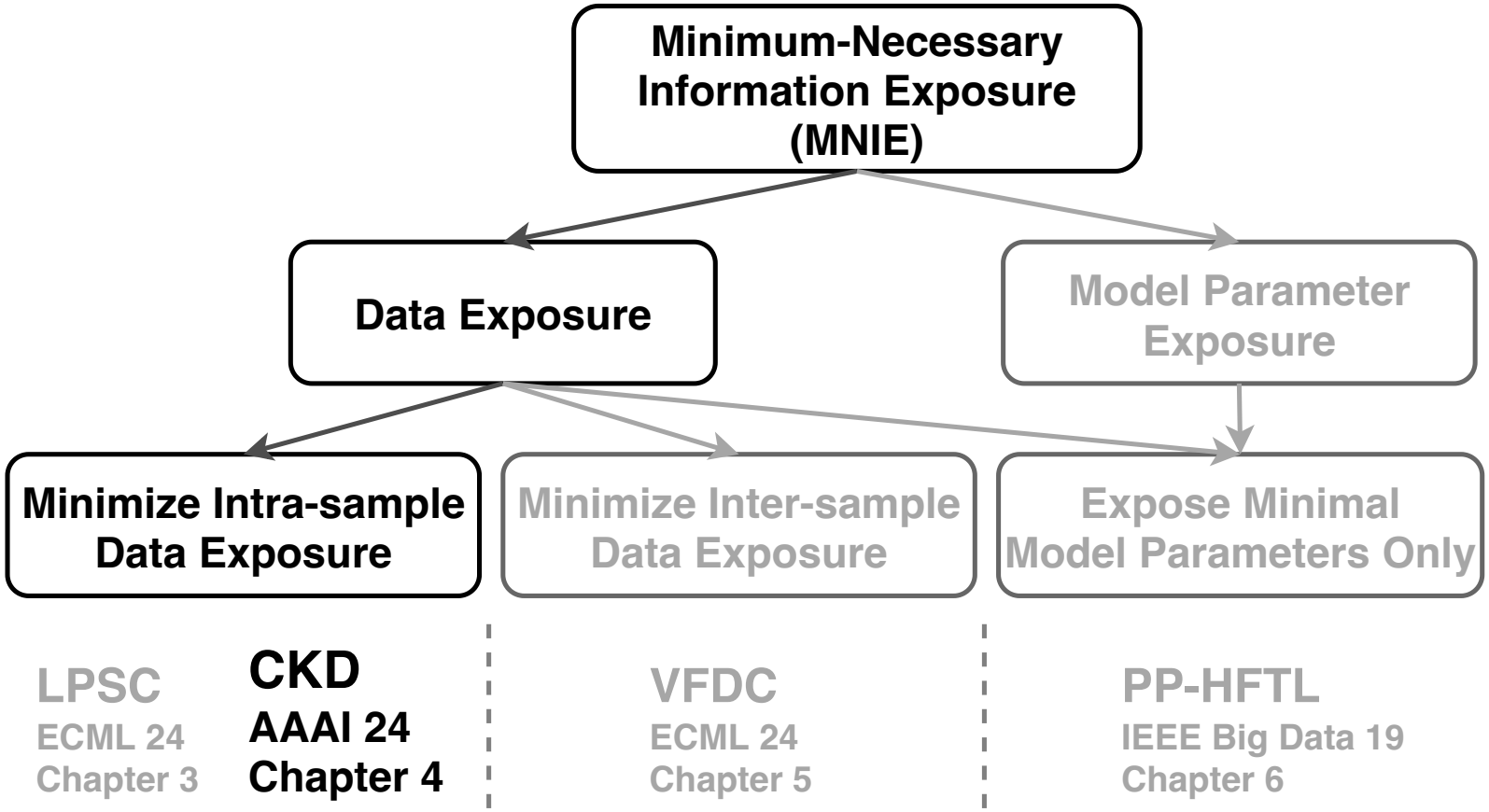
Figure 3.5: Privacy-utility trade-off of different protection methods against the Passive Model Completion (PMC) attack on four datasets.

Conclusion

- We study label privacy protection in VFL by formulating an offline **Label Privacy Source Coding (LPSC)** problem with a ϵ-MIP privacy guarantee**.**
- We prove that **gradient boosting** efficiently solves the LPSC problem.
- We propose a two-phase LPSC+Adv framework which incorporates **adversarial training** in during training and enables a superior privacy-utility trade-off.
- Experimental results on four datasets demonstrate the efficacy of LPSC+Adv framework.

Contents

4. CKD: Complementary Knowledge Distillation for Robust and Privacy-Preserving Model Serving in VFL (AAAI 2024)



Background: Challenges

The VFL model inference faces two major challenges:

- **Robustness:**

  Unavailable features due to arbitrarily aligned samples and straggler parties.

  Note: distinct from Byzantine robustness.

- **Label leakage of test data:**

  Semi-honest passive party attacks test data's label privacy from the bottom models' output embeddings.

  Note: protecting test label is to protect $p(y|x_k)$

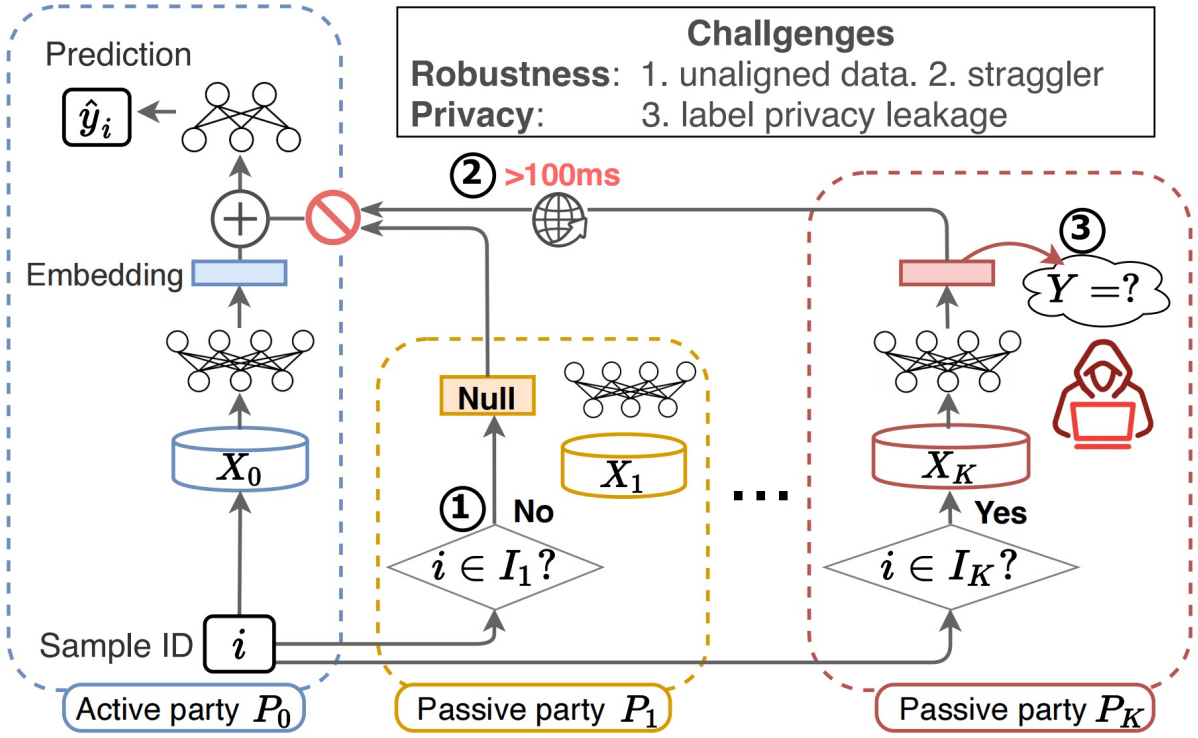


Figure 4.1. VFL inference challenges

Complementary Label Coding (CLC)

We introduce the concept of **Complementary Label Coding (CLC),** to decouple the private label information $p_{gt}(i, y)$ into two distinct components:

1. The **redundant label info** $p_{act}(i, y)$ which is locally learned by the active party's model f_θ.
2. The **complementary label info** $p_{clc}(i, y)$ that the active party has yet to learn.

We formulate the **CLC objective** with a mutual information constraint as follows:

$$\min_{p_{clc}(i,y)} \mathbb{R}_{p_{gt}(i,y)}(f_{CKD}) \quad \textit{(necessary)}$$
$$s.t. \quad I(p_{act}(i, y); p_{clc}(i, y)) = 0 \quad \textit{(minimum)}$$

where $f_{CKD}(i) = Merge(f_\theta(i), h^*_{pas}(i))$,

$$h^*_{pas}(i) = \arg\min_{h_{pas}} \mathbb{R}_{p_{clc}(i,y)}(h_{pas}).$$

Necessary information exposure for **utility**

Minimum information exposure for **privacy**

Complementary Label Coding (CLC)

We demonstrate that the proposed CLC objective can be reduced to the **LogitBoost** [Yoav and Robert, 1997] objective and solved via the Newton Method.

The optimized joint distribution $p_{clc}(i, y)$ provides weights and pseudo-residuals $\mathrm{D}_{clc} = (\mathrm{w}, r)$ of training data to train passive party's bottom model.

Proposition 1. *Given $p_{gt}(i) \sim U$ is a uniform distribution, ground-truth label $y_i = p_{gt}(y|i)$, local model output logit $f_\theta(i)$, and the expected passive model output $h^*_{pas}(i)$. The original CLC objective in Eq. 2 is equivalent to LogitBoost (Freund and Schapire 1997) objective:*

$$\min_{p_{clc}(i,y)} \sum_{i=1}^{n} \frac{1}{n} \mathcal{L}_{CE}(y_i,\ f_\theta(i) + h^*_{pas}(i)), \tag{5}$$

Proposition 2. *(Freund and Schapire 1997) Given $\hat{y}_{0,i} = \sigma(f_\theta(i))$ is the locally predicted probability. Using the Newton method, the optimization result $p_{clc}(i, y)$ of Eq. 5 is:*

$$p_{clc}(i) = \frac{\hat{y}_{0,i}(1-\hat{y}_{0,i})}{\sum_{j=1}^{n} \hat{y}_{0,j}(1-\hat{y}_{0,j})},\ p_{clc}(y|i) = \frac{y_i - \hat{y}_{0,i}}{\hat{y}_{0,i}(1-\hat{y}_{0,i})}. \tag{6}$$

Sample weights Pseudo-residuals

Theorem 1 (Privacy guarantee). *When the standard error of the local model f_θ trained on $p_{gt}(i, y)$ satisfies $\mathbb{R}_{p_{gt}(i,y)}(f_\theta) \to 0$, the privacy leakage from CLC-encoded results $p_{clc}(i, y)$ satisfies $I(p_{gt}(i, y); p_{clc}(i, y)) \to 0$.*

Yoav Freund and Robert E Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1):119–139, 1997.

Protect Label Privacy via CLC

- Active party learns label $D_{gt} = (1, y)$:

$$\mathcal{L}_{loc} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{L}_{CE}(y_i, f_\theta(i))$$

- Passive party learns re-weighted pseudo-residuals $D_{clc} = (w, r)$:

$$\mathcal{L}_{clc} = \sum_{i=1}^{n} w_i \cdot ||r_i - h_{pas}(i)||_2^2$$

- Federated model is the sum of local prediction and passive residuals:

$$f_{CKD}(i) = f_\theta(i) + \alpha \cdot h_{pas}(i)$$

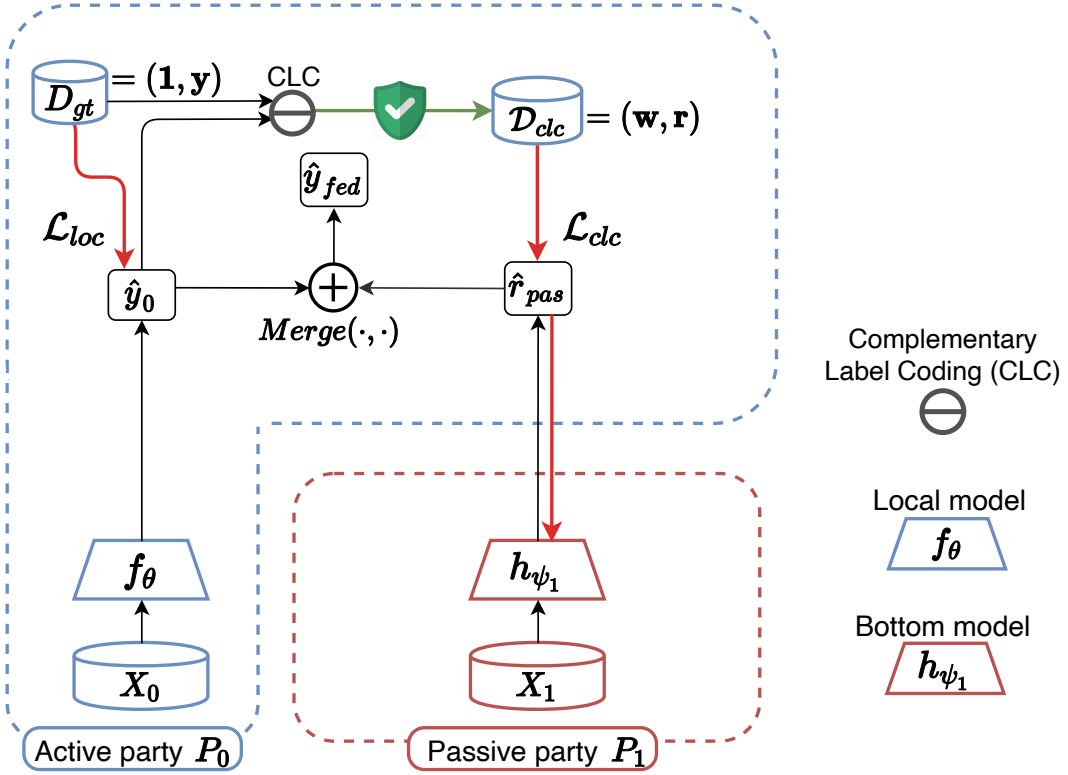


Figure 4.3: CLC-based VFL

CKD for Robust Model Inference

To address robustness against partially-aligned data, we use **online ensemble distillation** to distill the aggregated residual among parties.

Robustly aggregated teacher residuals:

$$h_{pas}(i) = g_\lambda \circ \{h_{\psi_k}\}_{k\in\mathcal{K}}(i) = \frac{\sum_{k\in\mathcal{K}} \lambda_k \cdot h_{\psi_k}(i)}{\sum_{k\in\mathcal{K}} \lambda_k}$$
$$s.t.\ \ \lambda_k \geq 0, \ \ \forall k \in [1, K],$$

Passive-to-active distillation:

$$\mathcal{L}_{p2a} = T^2 \sum_{i=1}^{n} w_i \cdot \mathbf{KL}(\sigma(f_{CKD}(i)/T) || \sigma(f_\theta(i)/T))$$

Passive-to-passive distillation:

$$\mathcal{L}_{p2p(k)} = T^2 \sum_{i=1}^{n} w_i \cdot D_{KL}(\sigma(h_{pas}(i)/T) || \sigma(h_{\psi_k}(i)/T))$$

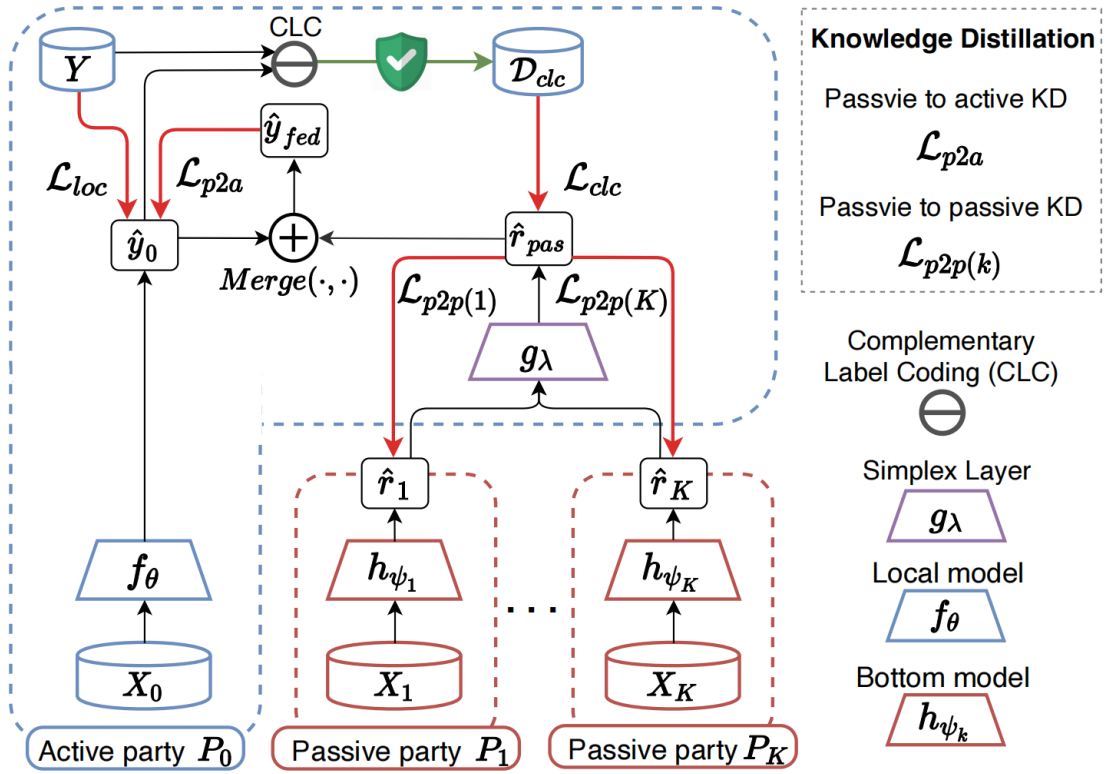


Figure 4.4: Overview of proposed CKD

Experiments: Privacy-Robustness Trade-off

- The comparative results of utility and privacy on four datasets.
- CKD protects label privacy via CLC and improves robustness (utility) via distilling knowledge to the local model.

Method	Criteo			Avazu			HetRec			MIMIC-III		
	L-Util↑	P-Util↑	Priv↓	L-Util↑	P-Util↑	Priv↓	L-Util↑	P-Util↑	Priv↓	L-Util↑	P-Util↑	Priv↓
Local	72.2	-	-	69.8	-	-	68.6	-	-	63.9	-	-
VFEns	72.2	74.1	72.7	69.8	74.0	71.2	68.6	72.9	69.0	63.9	69.0	64.9
PtyDrop	70.8	73.5	62.4	67.9	73.5	60.4	66.7	72.6	61.6	62.4	67.8	57.3
SplitKD	**72.6**	72.6	70.5	70.2	70.2	69.8	**69.0**	69.0	69.2	**64.2**	64.2	65.1
VFMD	72.4	74.3	73.2	70.2	74.2	72.5	68.8	73.1	69.4	64.2	69.2	65.3
(Our) CKD	72.5	**74.5**	**59.7**	70.2	**74.3**	**57.9**	68.9	**73.4**	**60.6**	64.1	**69.3**	**56.7**

Table 1: The comparative results of utility and privacy on four datasets. *L-Util* and *P-Util* indicate the AUC (%) on active party's **l**ocal data $\boldsymbol{x}_0$ and **p**artially-aligned data $\{\boldsymbol{x}_0, \boldsymbol{x}_k\}_{k\in K}$, respectively. *Priv* is the privacy leakage AUC (%) of bottom models.

Experiments: Impact of KD Losses

- Passive-to-active KD:

  Improve local utility and privacy by transferring complementary label knowledge.

- Passive-to-passive KD:

  Improve partially-aligned utility and decrease privacy. Trade privacy for robustness.

Loss		Criteo			Avazu		
$\mathcal{L}_{p2a}$	$\mathcal{L}_{p2p}$	L-Util↑	P-Util↑	Priv↓	L-Util↑	P-Util↑	Priv↓
×	×	72.1	74.2	59.6	69.8	74.0	57.8
✓	×	72.5	74.4	**59.4**	70.1	74.2	**57.6**
✓	✓	72.5	**74.5**	59.7	**70.2**	**74.3**	57.9

Table 2: Impact of $\mathcal{L}_{p2a}$ and $\mathcal{L}_{p2p}$ on CKD. *L-Util* and *P-Util* denote the AUC (%) on **l**ocal data $\boldsymbol{x}_0$ and **p**artially-aligned data $\{\boldsymbol{x}_0, \boldsymbol{x}_k\}_{k \in K}$, respectively. *Priv* is the privacy leakage AUC (%) of bottom models against the PMC attack.

Conclusion

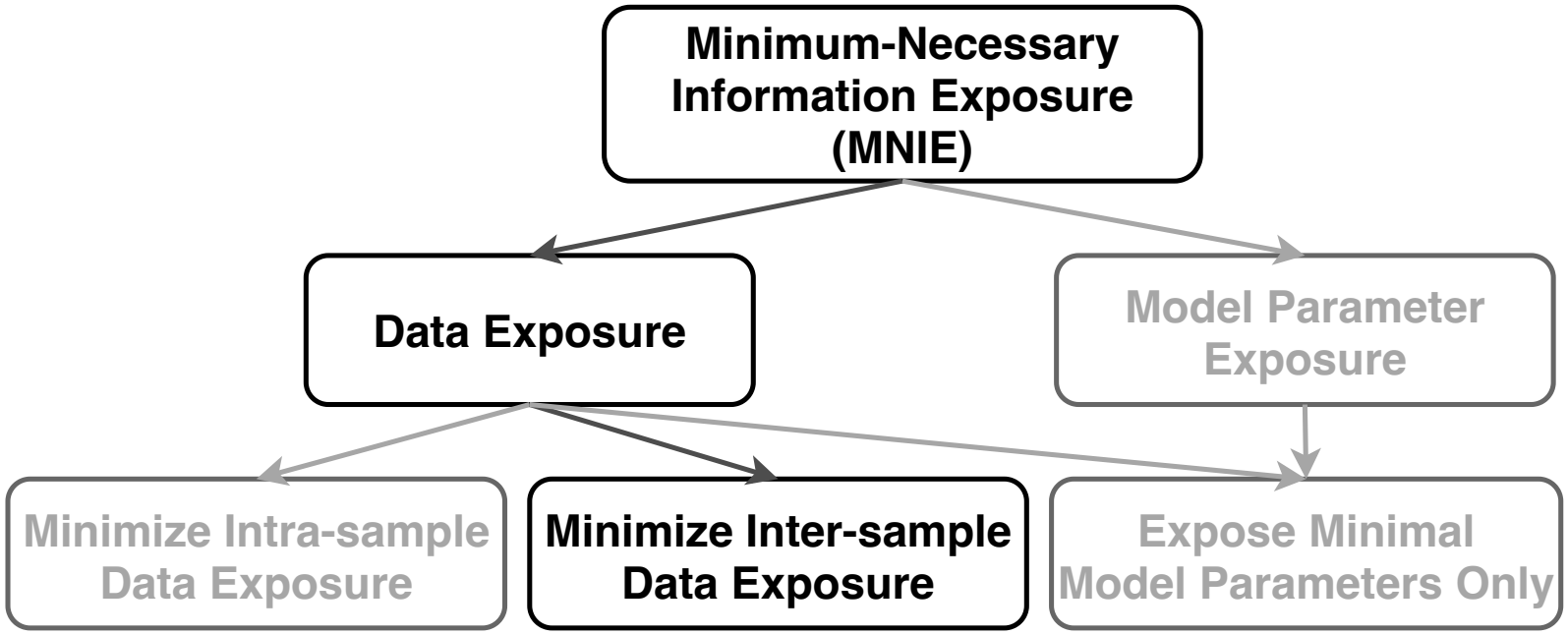
- Key insight: In VFL, passive parties should only learn and transfer the "complementary label information", which is **minimum-necessary** and not learned by the active party's local model.

Contributions:

- We introduce the Complementary Label Coding (CLC) that generates re-weighted pseudo-residuals for label privacy protection.
- We propose Complementary Knowledge Distillation (CKD) to **transfer** complementary label knowledge to improve robustness.

Contents

5. VFDC: Secure Dataset Condensation for Privacy-Preserving and Efficient VFL



Motivation

- Vanilla VFL methods train models on the **entire real** dataset.
- However, they face a dual challenge of
  1. real-data privacy,
  2. training-phase efficiency.
- Therefore, we aim to address this dual challenge by securely generating a **small synthetic** dataset to eliminate **inter-sample data exposure.**

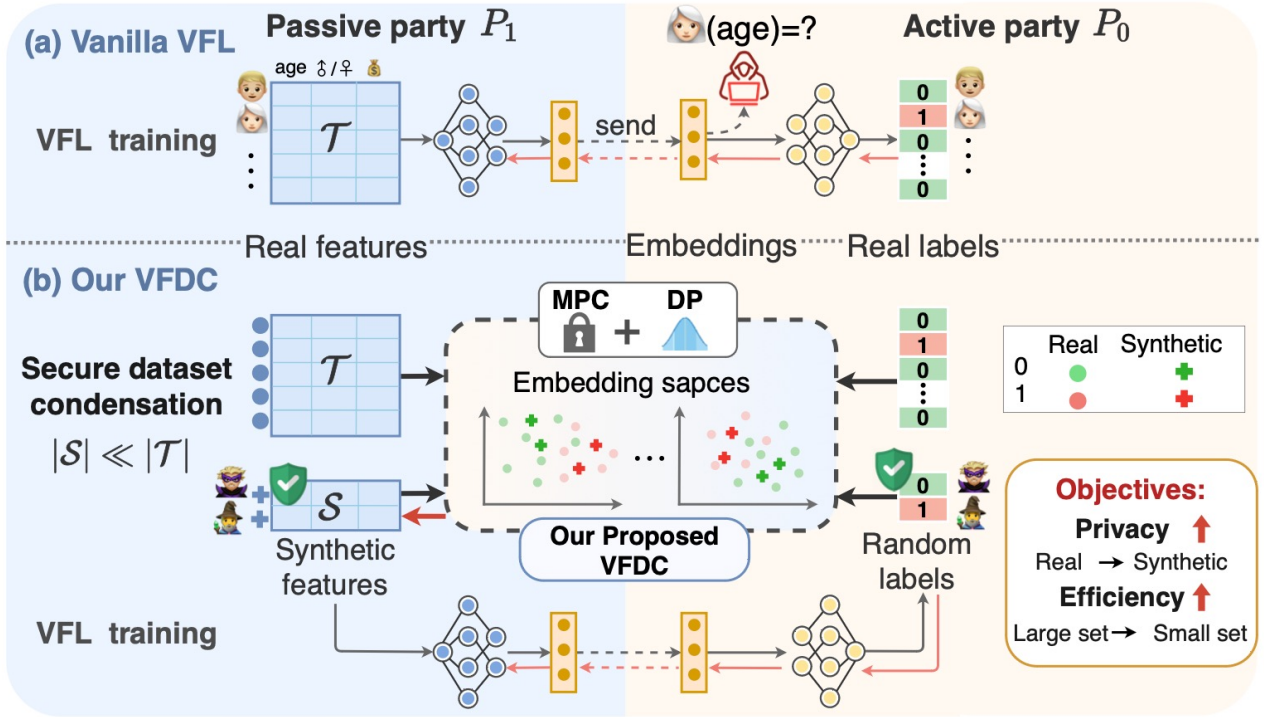


Figure 5.1: Schematic comparison between vanilla VFL and VFDC. (a) Vanilla VFL trains on the entire real dataset. (b) Our VFDC securely generates a small synthetic dataset.

Problem Definition

- **Utility objective**: The federated model trained on synthetic dataset $\mathcal{S}$ has high utility on real data distribution:

$$\min \mathbb{E}_{(\boldsymbol{x},y)\in\mathcal{T}} \left[\mathcal{L}(\psi_\theta(\boldsymbol{x}), y)\right]$$
$$\text{where } \theta = \arg\min \mathbb{E}_{(\boldsymbol{x},y)\in\mathcal{S}} \left[\mathcal{L}(\psi_\theta(\boldsymbol{x}), y)\right].$$

- **Privacy objective**:
  - No feature and label privacy leakage during class-wise secure aggregation.
  - Feature privacy satisfies differential privacy (DP).
- **Efficiency objective:**
  - Minimize the training iterations to converge.

Preliminary

- **Dataset condensation** (DC) aims to generate a significantly smaller condensed dataset S, with $|S| \ll |T|$.
- Objective: The model trained on the **condensed dataset S** achieves a comparable performance to that trained on the **entire dataset T**.

$$\mathbb{E}_{\boldsymbol{x}\sim P_{\mathcal{D}}}\mathcal{L}(f_{\theta^{\mathcal{T}}}(\boldsymbol{x}), y) \approx \mathbb{E}_{\boldsymbol{x}\sim P_{\mathcal{D}}}\mathcal{L}(f_{\theta^{\mathcal{S}}}(\boldsymbol{x}), y)$$

- We minimize the Maximum Mean Discrepancy (MMD) Loss:

$$\sum_{c=0}^{C-1} \mathbb{E}_{\boldsymbol{\vartheta}\sim P_{\vartheta}} \left\| \frac{1}{|\mathcal{T}_c|}\sum_{i=1}^{|\mathcal{T}_c|} \psi_{\boldsymbol{\vartheta}}\left(\boldsymbol{x}_{c,i}\right) - \frac{1}{|\mathcal{S}_c|}\sum_{j=1}^{|\mathcal{S}_c|} \psi_{\vartheta}\left(\boldsymbol{s}_{c,j}\right) \right\|^2$$

Proposed VFDC Method

1. **Left**: The passive party computes the DP-protected embeddings of real features.
2. **Middle**: Two parties engage in class-wise secure aggregation.
3. **Right**: Active party computes MMD loss and sends back gradients.

MMD loss:

$$\left\| \frac{1}{|\mathcal{T}_c|} \sum_{i=1}^{|\mathcal{T}_c|} \psi_{\boldsymbol{\vartheta}}\left(\boldsymbol{x}_{c,i}\right) - \frac{1}{|\mathcal{S}_c|} \sum_{j=1}^{|\mathcal{S}_c|} \psi_{\vartheta}\left(\boldsymbol{s}_{c,j}\right) \right\|^2$$

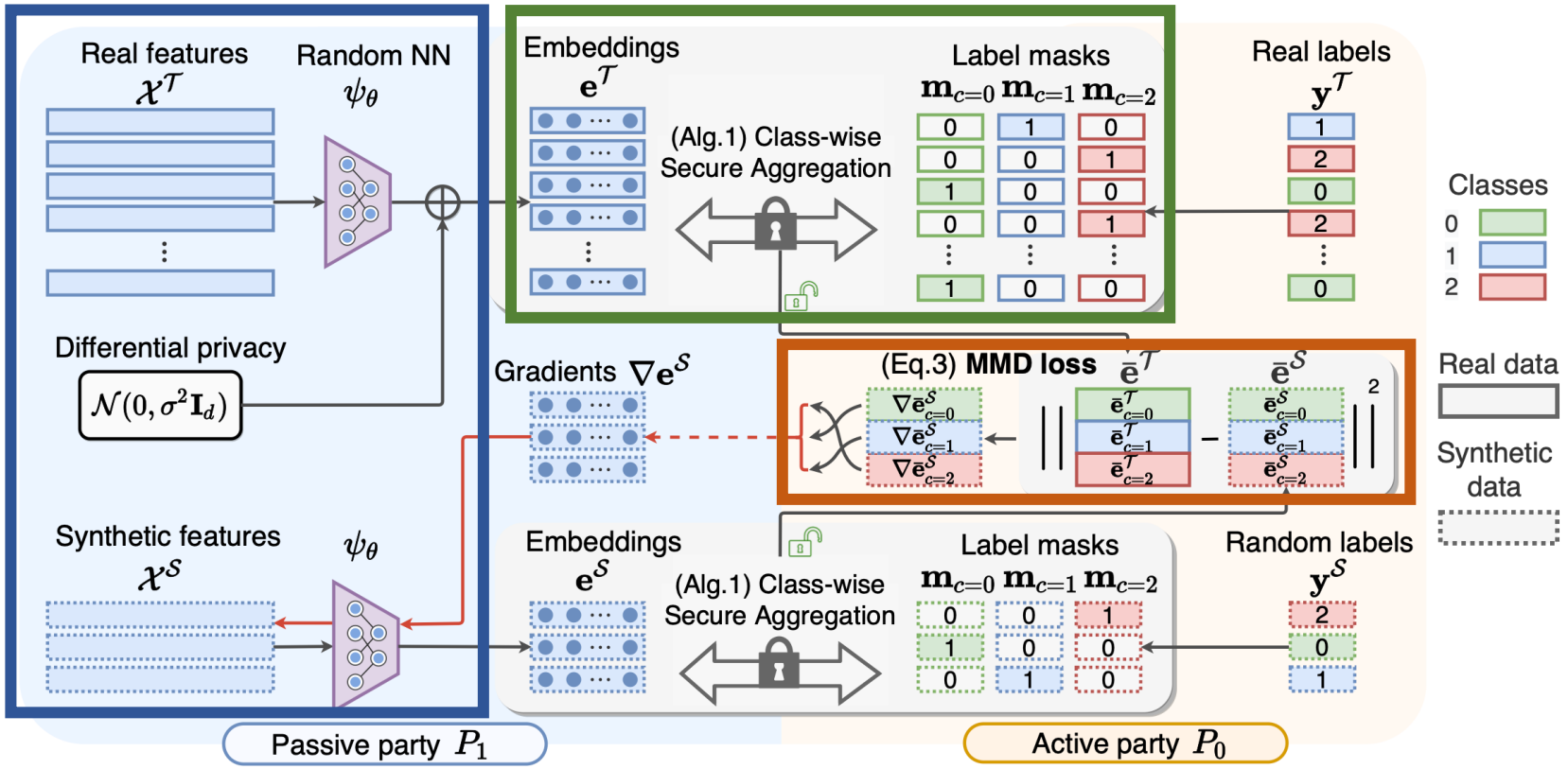


Figure 5.3

Privacy Analysis

- Class-wise secure aggregation protects sample-wise embeddings and labels.

Theorem 2. *The class-wise secure aggregation (Algorithm 1) ensures that the passive party gains no information about real data labels, while the active party learns only the average class embeddings.*

- Feature Privacy protection:

1) Secure aggregation + DP enables smaller noise:

Theorem 3. *Given each epoch of Algorithm 2 satisfies* (ϵ, δ)*-DP, there exists constants* r_1 *and* r_2 *such that given sampling probability* q *and epoch number* T*, and* $\epsilon < r_1 q\sqrt{T}$*, Algorithm 2 satisfies* (ϵ', δ)*-DP, with* $\epsilon' = r_2 q\sqrt{T}\epsilon$ *over* T *epochs.*

2) Repetitive random model initialization:

The continuously re-initialized embedding extraction model makes it challenging to reconstruct the real features from embeddings.

Experimental Results

- RQ1: What is the visual quality of VFDC-generated dataset?

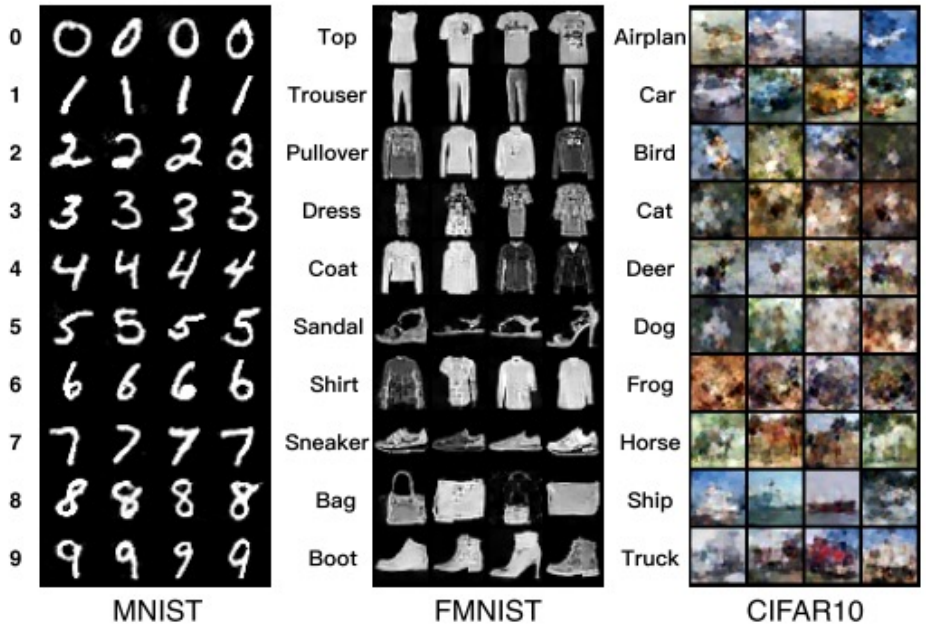


Figure 5.4: Visualization of the VFDC-generated synthetic images.

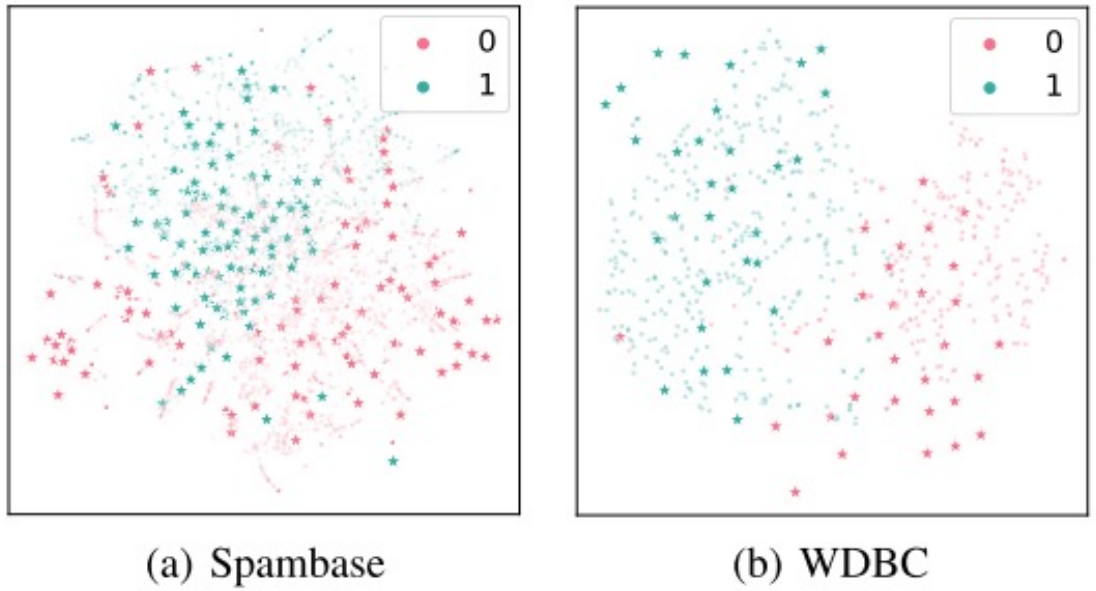


Figure 5.5: Visualization of the distribution of the original dataset and the condensed dataset by VFDC on two datasets using t-SNE. The star-shaped points denote the VFDC-generated samples.

Experimental Results

- RQ2: What is the privacy-utility trade-off of VFDC compared to other methods on different datasets?

Table 5.1: Comparison of accuracy (utility) of models trained by datasets generated via different methods.

	Spc	Rat.%	Coreset-LDP			DC-LDP			(Ours) VFDC			Vanilla-DC	Whole
DP $\epsilon \downarrow$			10	50	100	10	50	100	10	50	100	∞	∞
FMNIST	50	0.83	78.3	78.8	79.8	74.0	82.1	84.4	**88.2**	**88.3**	**88.3**	88.5	93.5
	100	1.7	80.7	81.6	81.9	74.2	84.0	84.8	**88.9**	**89.0**	**89.0**	89.1	
	200	3.3	81.8	82.6	83.1	75.4	84.3	85.0	**89.4**	**89.6**	**89.7**	89.8	
CIFAR	50	1.0	37.2	39.4	41.1	38.1	49.0	59.8	**62.1**	**63.0**	**63.3**	63.4	84.8
	100	2.0	40.3	44.5	50.9	37.6	50.4	60.9	**64.1**	**64.1**	**64.5**	65.3	
	200	4.0	52.3	54.5	56.3	37.0	50.7	61.3	**66.9**	**67.0**	**67.5**	67.9	
MNIST	50	0.83	90.1	92.1	93.4	91.3	95.6	96.7	**98.3**	**98.3**	**98.4**	98.4	99.6
	100	1.7	90.2	92.3	93.5	91.5	95.6	96.9	**98.6**	**98.6**	**98.7**	98.6	
	200	3.3	91.3	93.0	93.3	92.0	95.8	97.0	**98.7**	**98.7**	**98.7**	98.7	
MIMIC	10	0.11	63.6	70.8	70.8	41.2	52.2	71.4	**78.2**	**78.4**	**78.9**	79.5	81.7
	30	0.34	64.2	71.6	70.3	53.6	67.1	70.9	**78.4**	**78.9**	**79.1**	80.3	
	50	0.56	65.6	70.4	71.9	59.1	69.3	72.3	**78.8**	**79.3**	**79.9**	80.4	
Spambase	10	0.48	63.4	72.7	80.2	58.6	63.4	82.8	**89.7**	**90.1**	**90.5**	90.7	94.2
	30	1.45	65.8	74.3	82.6	57.2	65.9	81.8	**90.1**	**90.5**	**91.0**	91.3	
	50	2.42	70.1	75.2	83.8	56.3	69.1	82.7	**90.9**	**91.2**	**91.5**	91.7	
WDBC	10	4	89.4	90.4	92.5	86.1	91.8	92.9	**97.0**	**97.1**	**97.1**	97.2	97.3
	20	8	89.7	90.8	92.1	88.7	91.8	93.2	**97.0**	**97.1**	**97.2**	97.2	
	30	12	90.6	91.0	93.3	89.9	93.2	93.9	**97.1**	**97.2**	**97.3**	97.3	

Experimental Results

- RQ3: How does VFDC improve training efficiency compared to other methods?

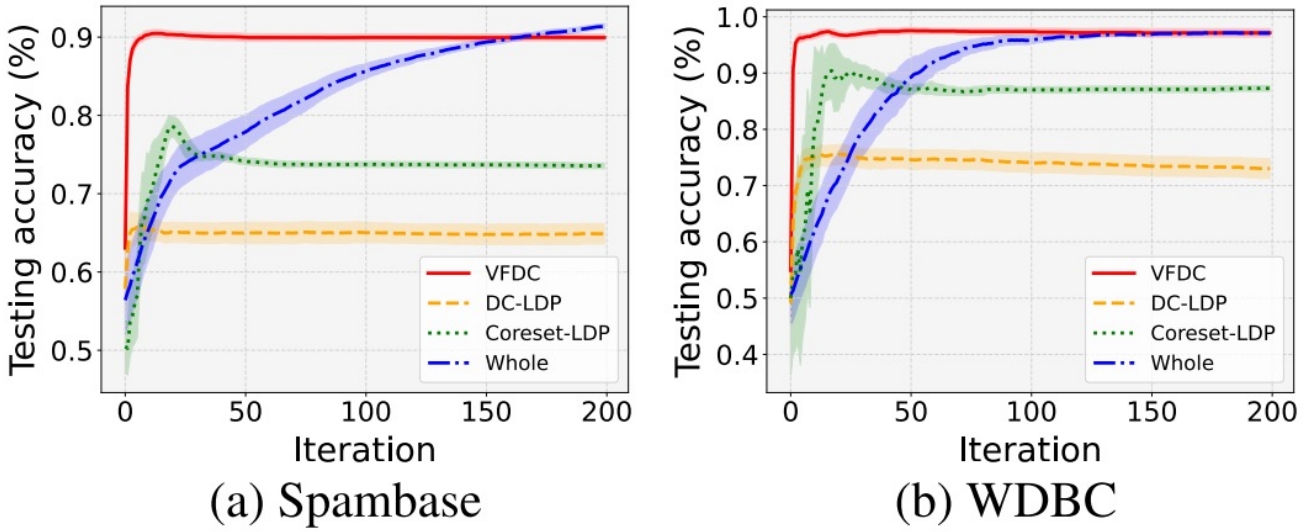


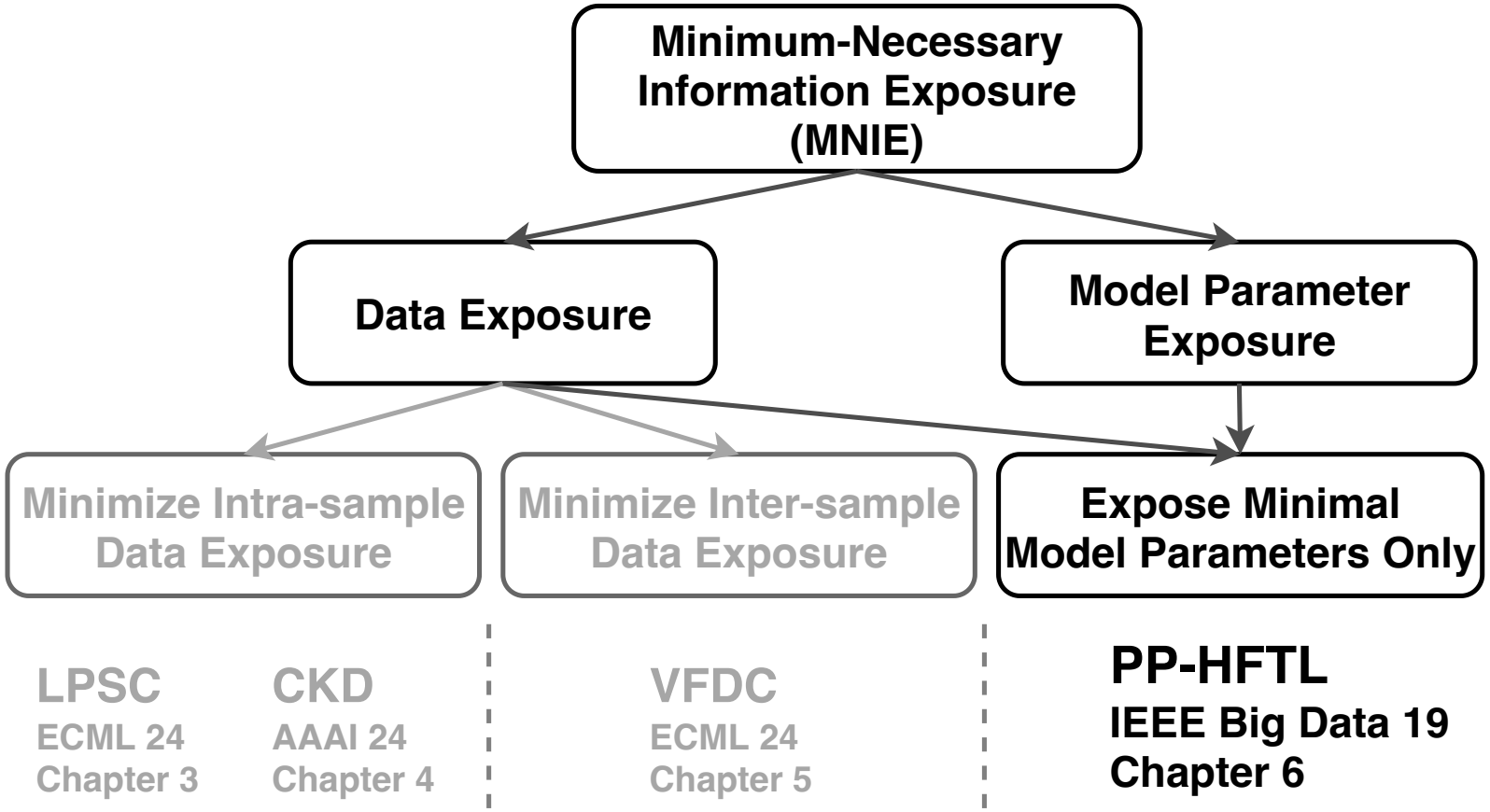
Figure 5.6: Performance variation of the VFL model trained using dataset output by different methods across iterations.

Conclusion

- To address the dual challenges of **privacy** and **efficiency** in VFL, we present VFDC for **small synthetic dataset generation**.
- VFDC has **three-fold mixed protection** mechanism, merging class-wise secure aggregation, DP, and repetitive model initialization.
- Experimental results show that VFDC achieves high training efficiency, sample-level data privacy, and utility.

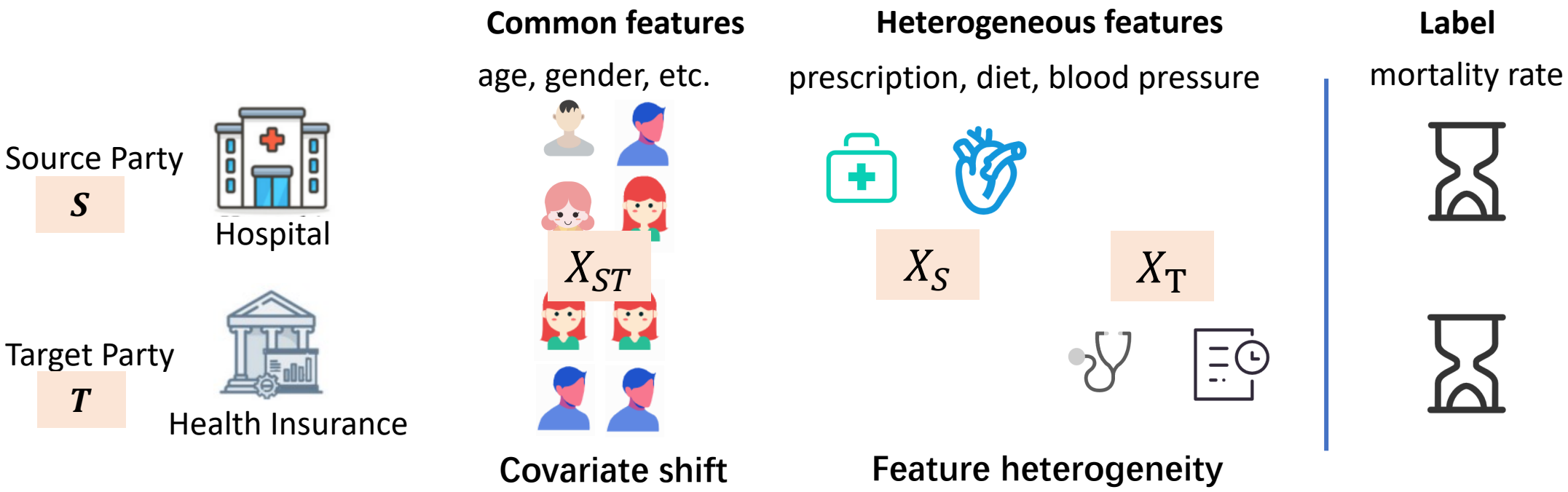
Contents

6. PP-HFTL: Privacy-Preserving Heterogeneous Federated Transfer Learning (IEEE Big Data 2019)



Motivation and Challenges

- An example [Zhang et al., 2022] of feature-sharing HFTL in healthcare scenario:
  - Common features with covariate shift
  - Heterogeneous features in each party
  - No semi-trusted third party



R. Zhang, H. Li, M. Hao, H. Chen and Y. Zhang, "Secure Feature Selection for Vertical Federated Learning in eHealth Systems," ICC 2022 - IEEE International Conference on Communications, Seoul, Korea, Republic of, 2022, pp. 1257-1262, doi: 10.1109/ICC45855.2022.9838917.

PP-HFTL Phase 1: Secure Transfer Learning

1. Secure domain adaptation

Compute instance weight $\langle \alpha \rangle$ by training a domain classifier via SLR.

$$\alpha(x_i) = \frac{P(x_i \in \mathcal{D}_T)}{1 - P(x_i \in \mathcal{D}_T)} = e^{w_k x_i} \qquad e^z \approx 1 + z\left(1 + \frac{z}{2}\left(\ldots(1 + \frac{z}{n-1}(1 + \frac{z}{n}))\right)\right)$$

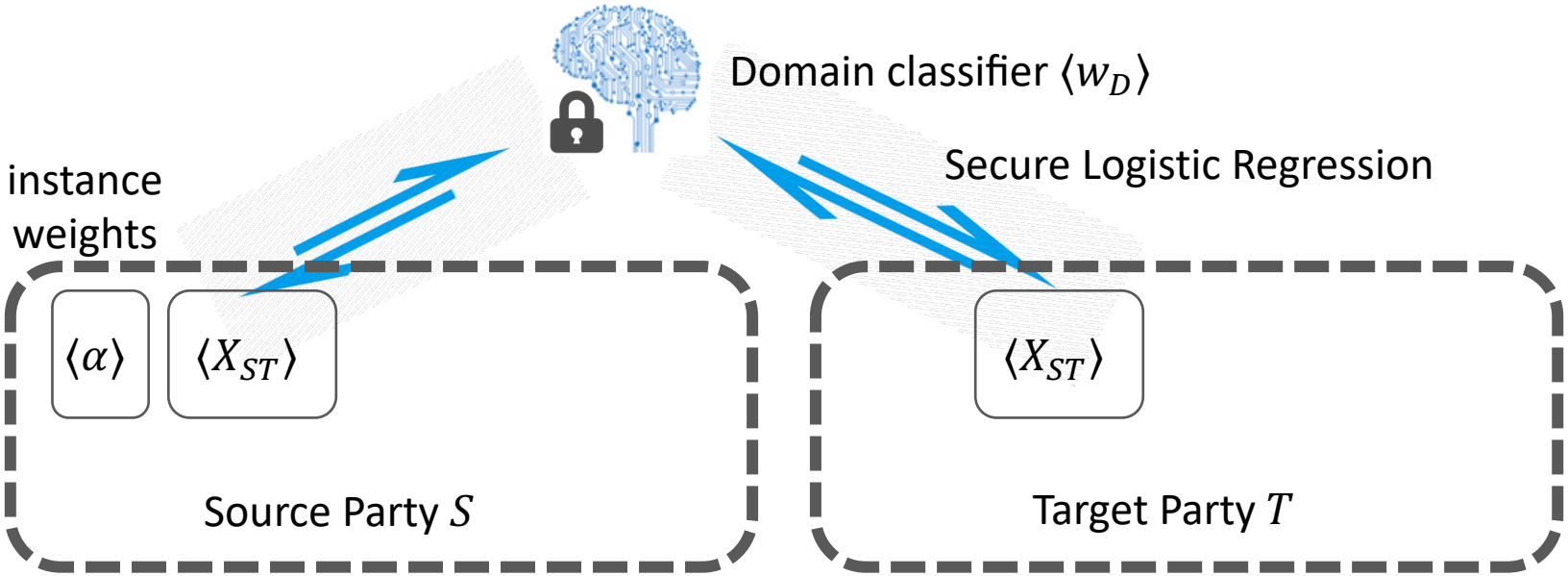


Figure 6.1

PP-HFTL Phase 1: Secure Transfer Learning

2. Secure feature mapping

Each party locally learns a linear regression model for feature mapping.

$$\min_{\theta^{S,T}} ||diag(\alpha) \cdot (X_S^S - X_{ST}^S \theta^S)||_F^2 + \lambda \sum_{n=1}^{d_2} ||\theta_n^S||_F^2$$

$$\min_{\theta^{T,S}} ||X_T^T - X_{ST}^T \theta^T||_F^2 + \lambda \sum_{n=1}^{d_2} ||\theta_n^T||_F^2$$

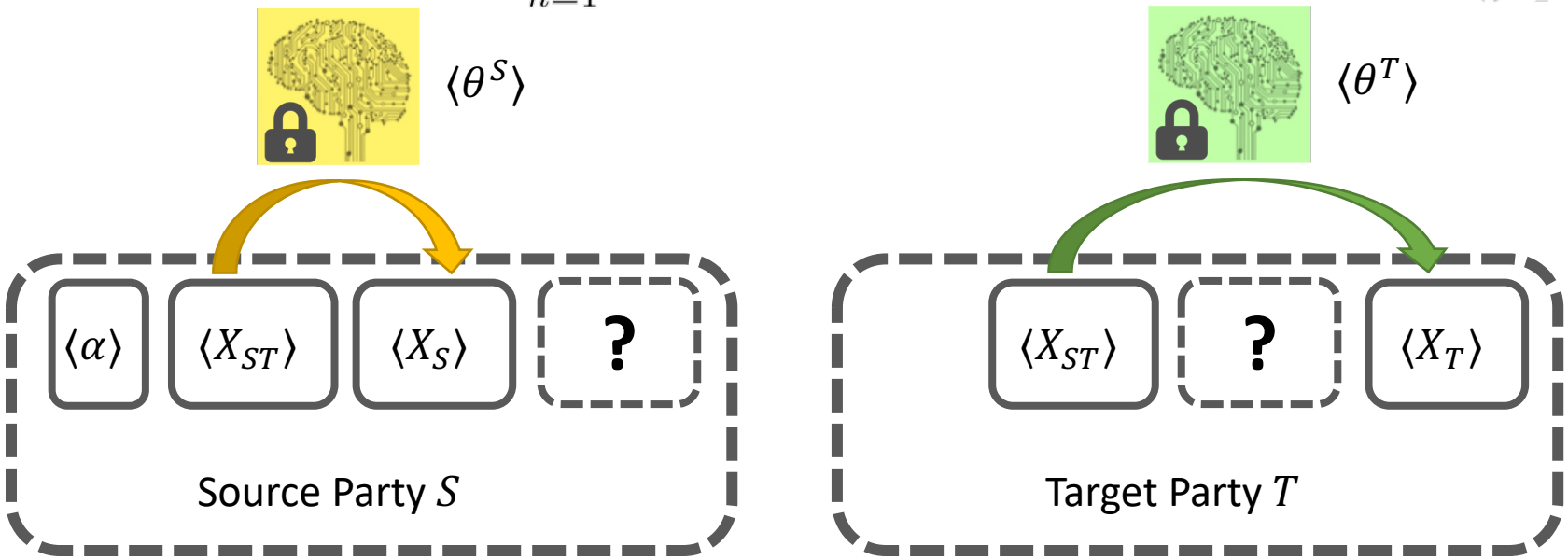


Figure 6.2

PP-HFTL Phase 1: Secure Transfer Learning

2. Secure feature mapping

Each party locally learns a linear regression model for feature mapping.

Do private feature inference.

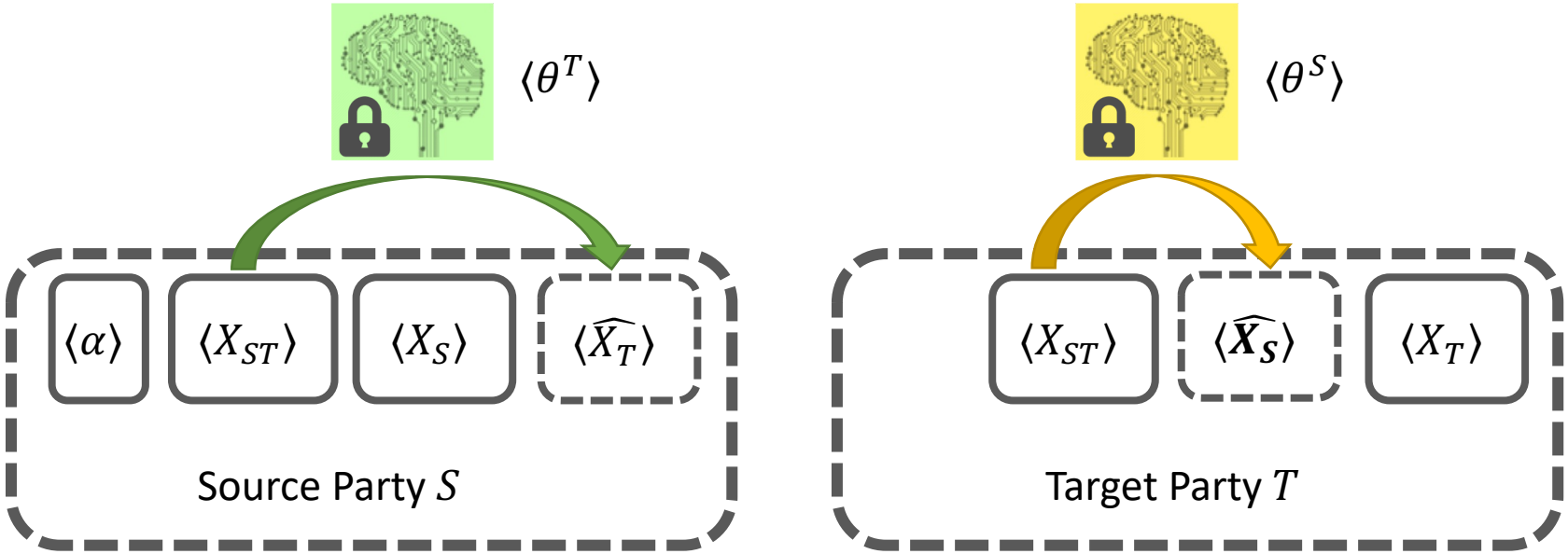


Figure 6.3

PP-HFTL Phase 2: Secure Federated Learning

Parties conduct secure logistic regression to train label prediction model $\langle w \rangle$

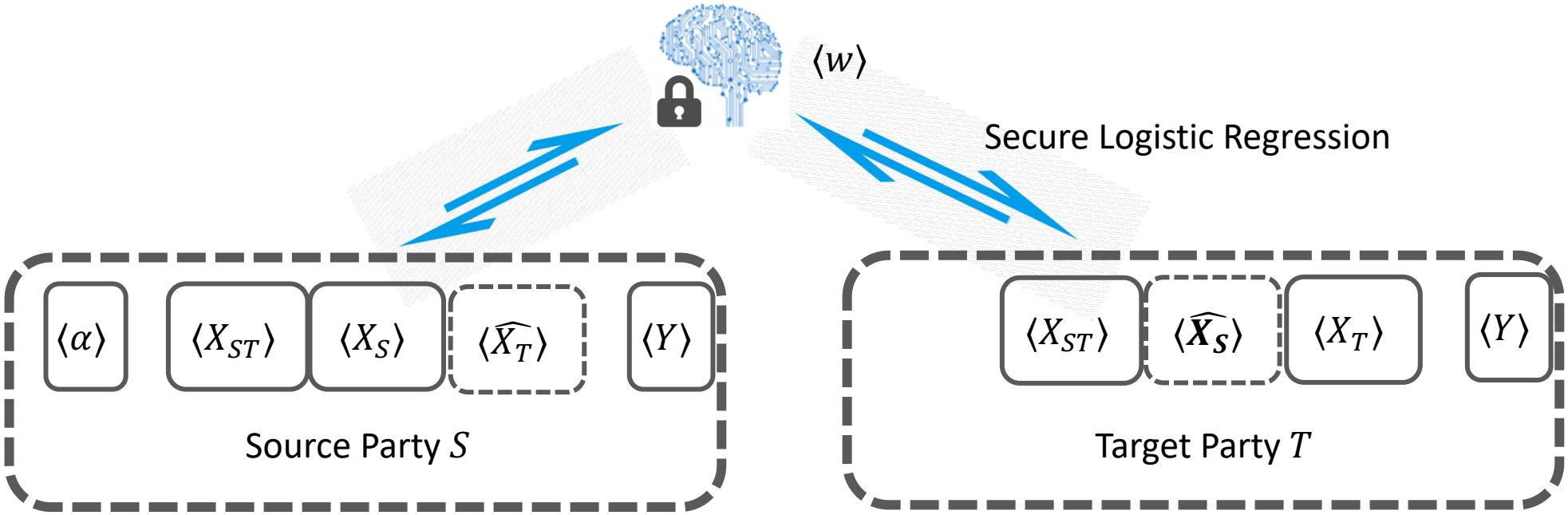


Figure 6.4

PP-HFTL Phase 3: Secure Model Integration

- Label inference requires $\langle \theta^S \rangle$ and $\langle w \rangle$, which is inefficient.

Target party T inference process:

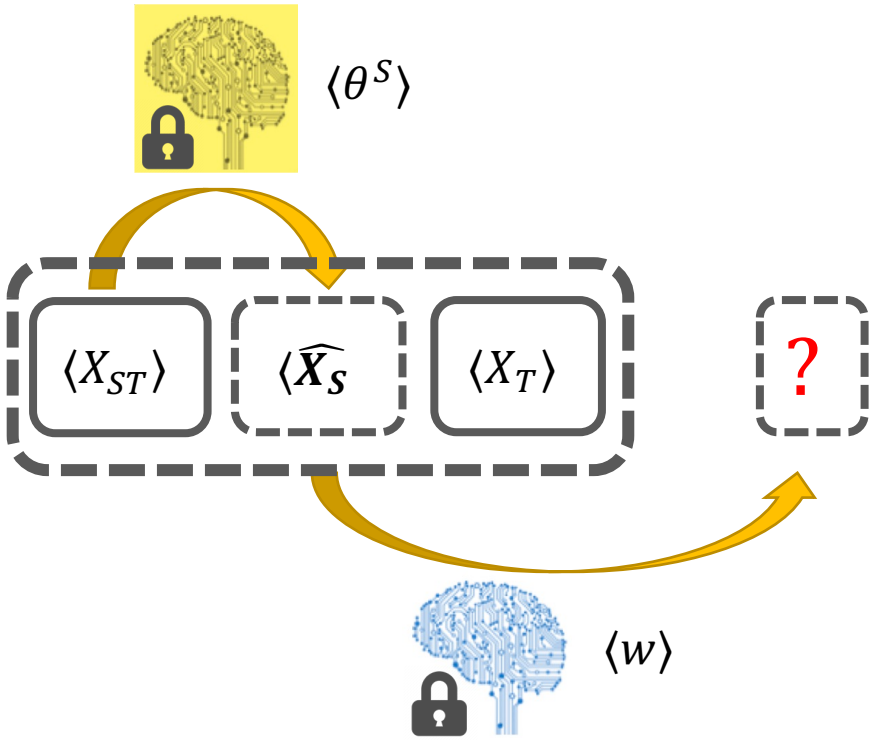


Figure 6.5

PP-HFTL Phase 3: Secure Model Integration

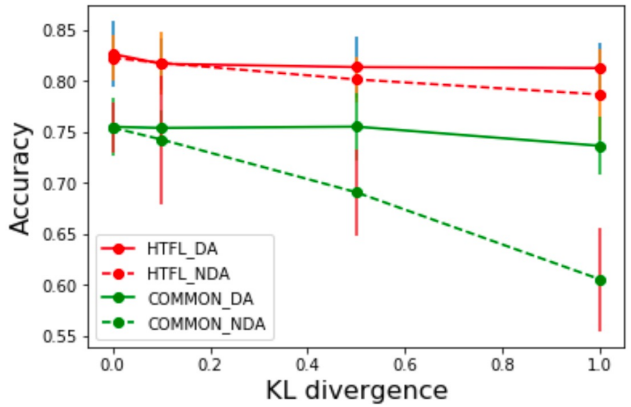
- Label inference requires $\langle \theta^S \rangle$ and $\langle w \rangle$, which is inefficient
- Integrate $\langle \theta^S \rangle$ and $\langle w \rangle$ and reveal w^T to target party T.

$$\langle w^T \rangle = \begin{bmatrix} \langle \theta^S \rangle \cdot \langle w_S \rangle + \langle w_{ST} \rangle \\ \langle w_T \rangle \end{bmatrix}$$

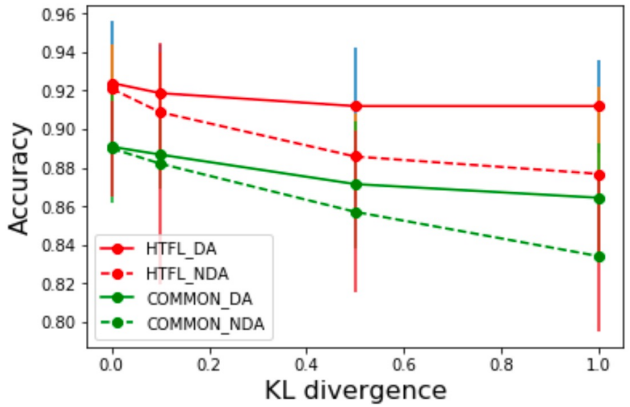
$$\text{where } w = \begin{bmatrix} w_{ST} \\ w_S \\ w_T \end{bmatrix}$$

- Party T can do label inference locally with w^T: $\hat{Y} = Sigmoid(X \cdot w^T)$

Experiments Performance of transfer learning

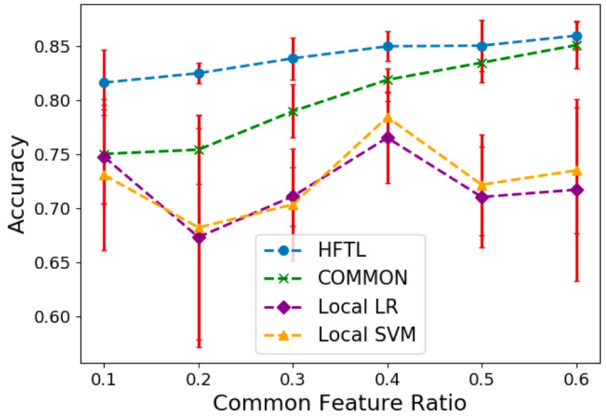


Spambase

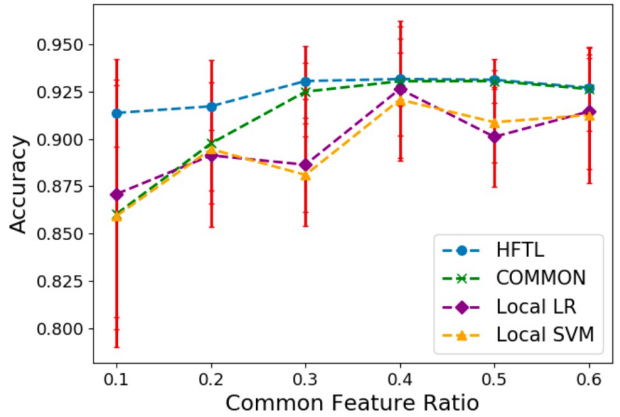


WDBC

PP-HFTL with **secure domain adaptation** is robust to covariate shift



Spambase

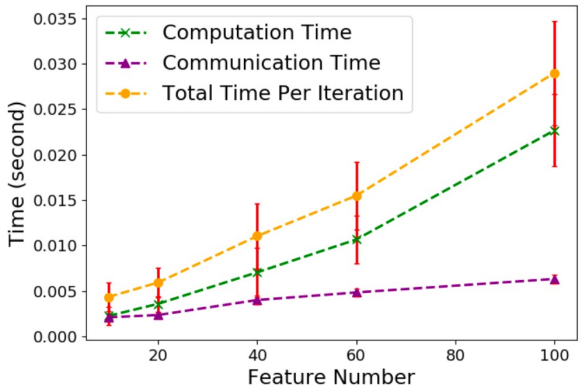


WDBC

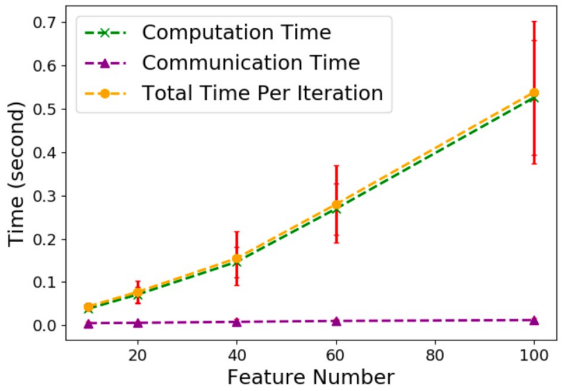
PP-HFTL with **secure feature mapping** is robust to feature heterogeneity

Figure 6.6

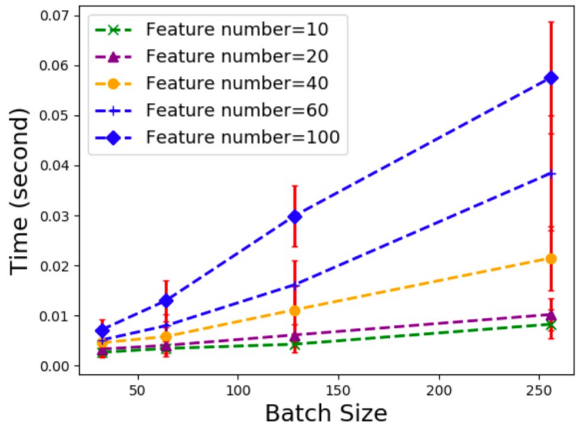
Experiments Efficiency and scalability of HFTL



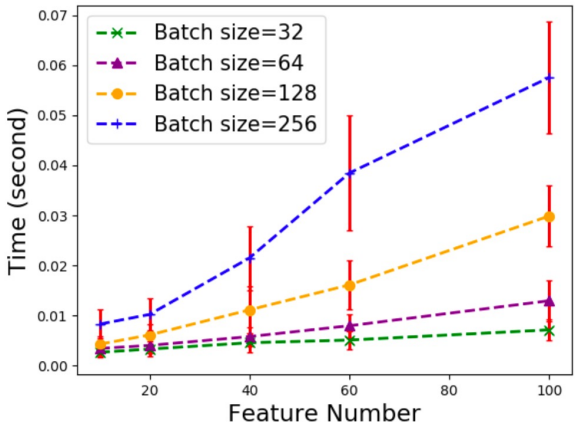
SS-based HFTL



HE-based HFTL



time vs. batch size n



time vs. # features m

Figure 6.7

Secret sharing-based HFTL is 20x more efficient than HE-based PP-HFTL.

$O(nm)$ complexity to batch size n and feature number m.

Conclusion

- We propose heterogeneous federated transfer learning (PP-HFTL) to enable secure VFL to address **covariate shift** and **feature space heterogeneity.**
- Two variants of PP-HFTL are proposed without introducing a collaborator.
- The final model achieves **minimum-necessary model exposure** for efficient inference.
- Experiments on four datasets demonstrate our approach can transfer knowledge from other parties with practical efficiency and scalability.

Contents

Conclusions

- We propose a unique **taxonomy** of information exposure in VFL.
- Our four studies on **minimum-necessary information exposure (MNIE)** explore VFL from three aspects:

1. Various types of information exposure
2. More objectives.
3. More complex settings.

More Objectives →

More Complex Settings ↓

	Privacy & Utility	Privacy & Utility & Robustness	Privacy & Utility & Efficiency
Vanilla VFL	**LPSC** (Intra-Exp) ECML PKDD 24		**VFDC** (Inter-Exp) ECML PKDD 24
Sample-sharing HFTL		**CKD** (Intra-Exp) AAAI 24	
Feature-sharing HFTL	**PP-HFTL** (Model-Exp) IEEE Big Data 19		

Intra-Exp: Intra-sample exposure, **Inter-Exp**: inter-sample exposure, **Model-Exp**: model exposure.

References

- Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. ACM Trans. Intell. Syst. Technol., 10(2), Jan 2019.
- Visa Research. (2023). Secure collaborative machine learning. https://usa.visa.com/dam/VCOM/regional/na/us/about-visa/research/documents/ secure-collaborative-machine-learning.pdf
- Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. Homomorphic encryption for arithmetic of approximate numbers. In International conference on the theory and application of cryptology and information security, pages 409–437. Springer, 2017.
- Fangcheng Fu, Yingxia Shao, Lele Yu, Jiawei Jiang, Huanran Xue, Yangyu Tao, and Bin Cui. Vf2boost: Very fast vertical federated gradient boosting for cross-enterprise learning. In Proceedings of the 2021 International Conference on Management of Data, pages 563–576, 2021.
- Chong Fu, Xuhong Zhang, Shouling Ji, Jinyin Chen, Jingzheng Wu, Shanqing Guo, Jun Zhou, Alex X. Liu, and Ting Wang. Label inference attacks against vertical federated learning. In 31st USENIX Security Symposium (USENIX Security 22), pages 1397–1414, Boston, MA, August 2022. USENIX Association.
- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Proc. of Theory of cryptography conference, 2006.
- Dashan Gao, Sheng Wan, Lixin Fan, Xin Yao, and Qiang Yang. Complementary Knowledge Distillation for Robust and Privacy-Preserving Model Serving in Vertical Federated Learning. AAAI, 2024.
- Dashan Gao, Sheng Wan, Hanlin Gu, Lixin Fan, Xin Yao, and Qiang Yang. Label Privacy Source Coding in Vertical Federated Learning. ECML PKDD 2024.
- Dashan Gao, Canhui Wu, Xiaojin Zhang, Xin Yao, and Qiang Yang. Secure Dataset Condensation for Privacy-Preserving and Efficient Vertical Federated Learning. ECML PKDD 2024.
- Dashan Gao, Yang Liu, Anbu Huang, Ce Ju, Han Yu, and Qiang Yang. Privacy-preserving heterogeneous federated transfer learning. In 2019 IEEE International Conference on Big Data (Big Data), pages 2552–2559, 2019.

References

- Dashan Gao, Ce Ju, X. Wei, Y. Liu, Tianjian Chen, and Q. Yang. HHHFL: Hierarchical heterogeneous horizontal federated learning for electroencephalography. ArXiv, abs/1909.05784, NeurIPS FL Workshop, 2019
- Dashan Gao, Ben Tan, Ce Ju, Vincent Zheng, and Qiang Yang. Federated Factorization Machine for Secure Recommendation with Sparse Data. AAAI RSML Workshop, Virtual, 2021.
- Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. 2017. Practical Secure Aggregation for Privacy-Preserving Machine Learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17).
- Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex J Smola. A kernel method for the two-sample-problem. In Advances in neural information processing systems, pages 513–520, 2007.
- Zhenghang Ren, Liu Yang, & Kai Chen(2022). Improving Availability of Vertical Federated Learning: Relaxing Inference on Non-overlapping Data. ACM Transactions on Intelligent Systems and Technology (TIST).
- Di Chai, Leye Wang, Kai Chen and Qiang Yang, "Secure Federated Matrix Factorization," in IEEE Intelligent Systems, vol. 36, no. 5, pp. 11-20, 1 Sept.-Oct. 2021, doi: 10.1109/MIS.2020.3014880
- Di Chai, Leye Wang, J. Zhang, Liu Yang, S. Cai, Kai Chen, & Qiang Yang. (2022, August). Practical Lossless Federated Singular Vector Decomposition over Billion-Scale Data. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 46-55).
- Kewei Cheng, Tao Fan, Yilun Jin, Yang Liu, Tianjian Chen, Dimitrios Papadopou- los, and Qiang Yang. Secureboost: A lossless federated learning framework. IEEE Intelligent Systems, 36(6):87–98, 2021.
- Oscar Li, Jiankai Sun, Xin Yang, Weihao Gao, Hongyi Zhang, Junyuan Xie, Virginia, Smith, and Chong Wang. Label leakage and protection in two-party split learning. International Conference on Learning Representations (ICLR), 2022.

Thank You

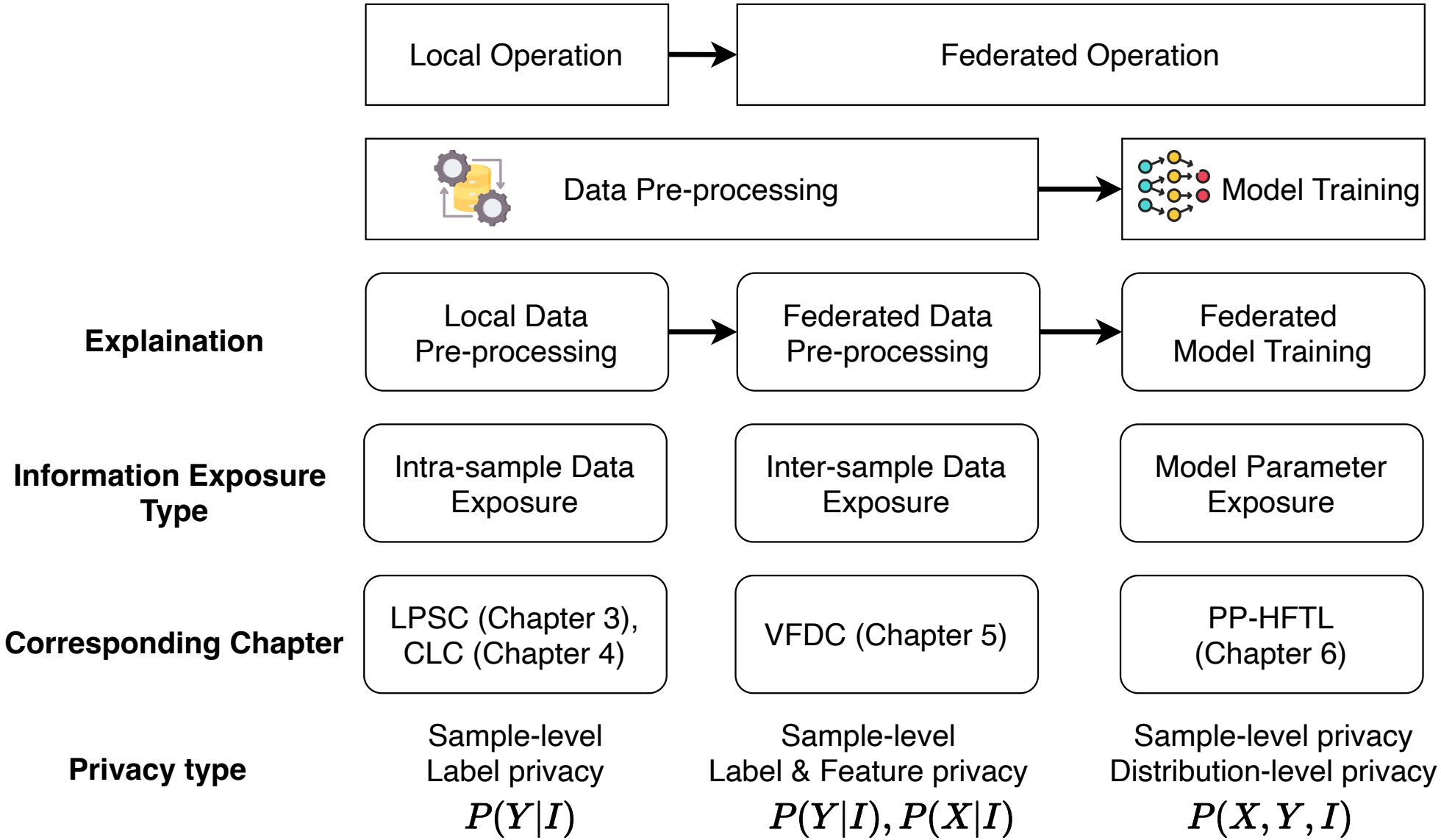
Q & A

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

SUSTech Southern University of Science and Technology

Backup slides

Comparison of Information Exposures



Summary of This Thesis

- One topic:
  - **Minimum Exposure Approach to Trustworthy VFL.**
  - Goal: Minimizing information exposure to improve trade-offs in VFL.
- Three dimensions:
  - **Exposure Types**: Intra-sample exposure, Inter-sample exposure, Model exposure.
  - **Objectives**: Utility, Privacy, Efficiency, Robustness, ...
  - **VFL Settings**: Vanilla VFL, Heterogeneous Federated Transfer Learning.
- Four works:

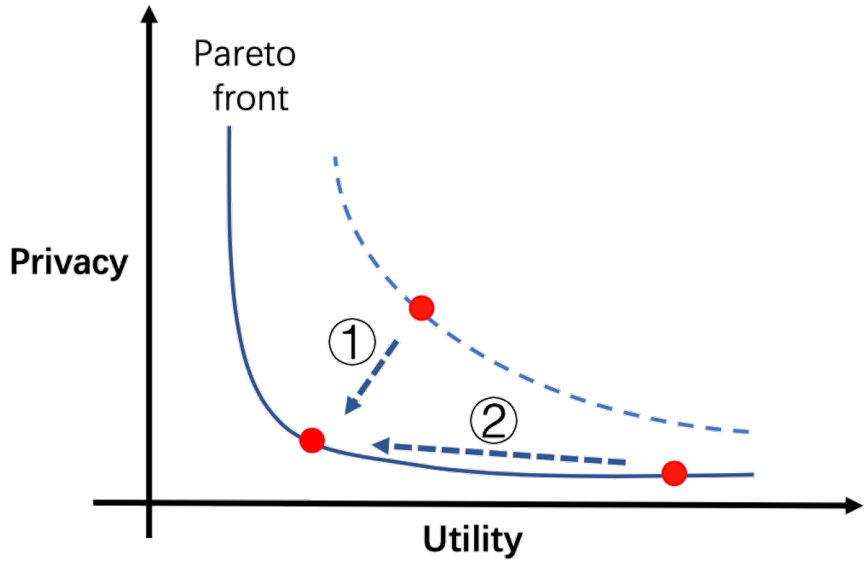
Papers	Conference	Targeted Exposure	Objectives	VFL Setting
LPSC	ECML PKDD	Intra-sample **exposure**	Privacy & Utility	Vanilla VFL
CKD	AAAI	Intra-sample **exposure**	Privacy & Utility & Robustness	Sample-sharing HFTL
PP-HFTL	IEEE Big Data	Data & Model **exposure**	Privacy & Utility	Feature-sharing HFTL
VFDC	ECML PKDD	Inter-sample **exposure**	Privacy & Utility & Efficiency	Vanilla VFL

Strategies for Multi-objective Trade-offs

- The design of VFL algorithms is to search for a Pareto front in the multi-objective trade-off space.
- There are two strategies to optimize the VFL approach:

1. Identify and reduce **unnecessity**, e.g., reduce unnecessary (redundant) communications and cryptographic operations[Fu et al. 2021].
2. Find the **knee point**. Trade one objective with a little decline for another with significant benefit [Cheon et al. 2017].

We focus on the first strategy by proposing MNIE.



- Fangcheng Fu, Yingxia Shao, Lele Yu, Jiawei Jiang, Huanran Xue, Yangyu Tao, and Bin Cui. Vf2boost: Very fast vertical federated gradient boosting for cross-enterprise learning. In Proceedings of the 2021 International Conference on Management of Data, pages 563–576, 2021.
- Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. Homomorphic encryption for arithmetic of approximate numbers. In International conference on the theory and application of cryptology and information security, pages 409–437. Springer, 2017.

Background of CKD

- Existing studies [Li et al., 2019, Ren et al., 2020] tackle the robustness challenge via knowledge distillation.
- However, they train the passive party's bottom model to **fit labels**, leading to label privacy leakage from the output of passive party's bottom model.

- Our idea is to train the passive party's bottom model to **fit residuals** of active party's local model.
- By doing so, the passive party only contributes residuals to the teacher model, thus protects label privacy.

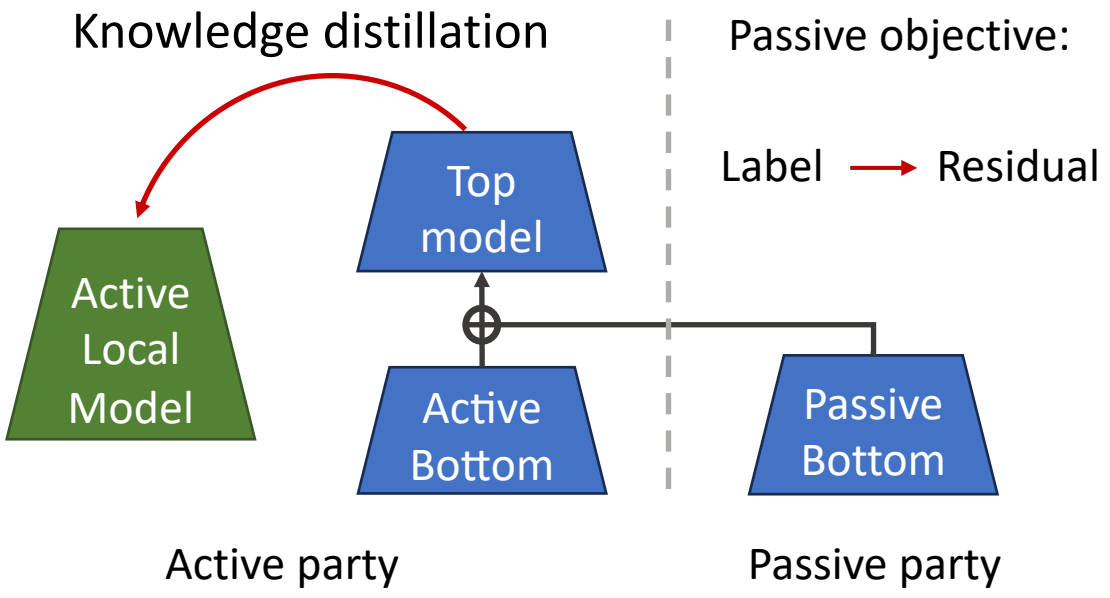
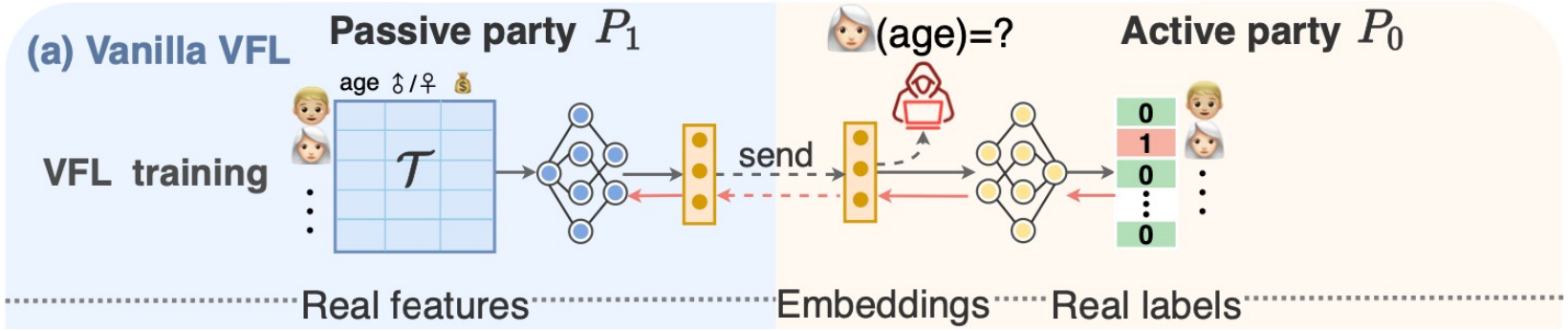


Figure 4.2

- Li, Wenjie, et al. "Semi-supervised cross-silo advertising with partial knowledge transfer." arXiv e-prints (2022): arXiv-2205.
- Ren, Zhenghang, Liu Yang, and Kai Chen. "Improving availability of vertical federated learning: Relaxing inference on non-overlapping data." ACM Transactions on Intelligent Systems and Technology (TIST) 13.4 (2022): 1-20.

Preliminary of VFDC

- Setting:
  - **Two parties**: active party P_0 possesses real labels y, and passive party P_1 holds real features x. The objective in VFL is for these two parties to collaboratively train a model.
- Threat Model:
  - '**Honest-but-curious**' assumption for all parties in VFL.
  - Active party seeks to infer feature privacy from the passive parties, and vice versa, the passive parties aim to infer label privacy from the active party.



Preliminary of VFDC

- **Differential Privacy (DP):**

Definition 1 (Differential Privacy [9]). *A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-differential privacy if, for any two neighboring datasets $\mathcal{D}$ and $\mathcal{D}'$ differing in only one sample, and for any output event E, the following inequality holds:*

$$\Pr[\mathcal{M}(\mathcal{D}) \in E] \leq e^{\epsilon} \Pr[\mathcal{M}(\mathcal{D}') \in E] + \delta,$$

where ϵ is the privacy budget, and δ is the fault-tolerance probability.

- **Secure Aggregation**:

Secure aggregation corresponds to computing the sum of multiple inputs while keeping the individual inputs confidential.

Preliminary of VFDC

- **Dataset condensation** (DC) aims to generate a significantly smaller condensed dataset S, with $|S| \ll |T|$.
- Objective: The model trained on the **condensed dataset S** achieves a comparable performance to that trained on the **entire dataset T**.

$$\mathbb{E}_{\boldsymbol{x}\sim P_{\mathcal{D}}}\mathcal{L}(f_{\theta^{\mathcal{T}}}(\boldsymbol{x}), y) \approx \mathbb{E}_{\boldsymbol{x}\sim P_{\mathcal{D}}}\mathcal{L}(f_{\theta^{\mathcal{S}}}(\boldsymbol{x}), y)$$

- We minimize the Maximum Mean Discrepancy (MMD) Loss:

$$\sum_{c=0}^{C-1} \mathbb{E}_{\boldsymbol{\vartheta}\sim P_{\vartheta}} \left\| \frac{1}{|\mathcal{T}_c|}\sum_{i=1}^{|\mathcal{T}_c|} \psi_{\boldsymbol{\vartheta}}\left(\boldsymbol{x}_{c,i}\right) - \frac{1}{|\mathcal{S}_c|}\sum_{j=1}^{|\mathcal{S}_c|} \psi_{\vartheta}\left(\boldsymbol{s}_{c,j}\right) \right\|^2$$

Class-wise Secure Aggregation in VFDC

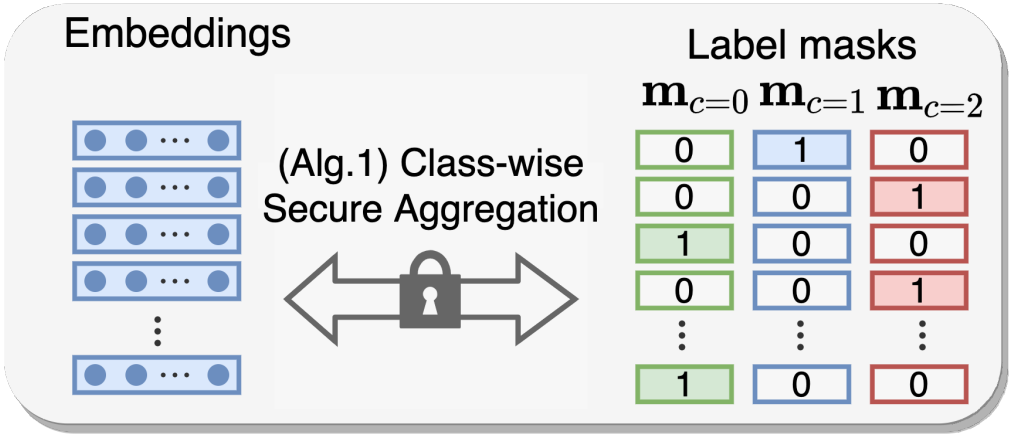


Figure 5.2

Algorithm 1 Class-wise Secure Aggregation

Input: Label vector $\boldsymbol{y}$ in active party P_0, feature embeddings $\boldsymbol{e}$ of real data in passive party P_1, class number C.
Output: Class-wise average embeddings $\bar{\boldsymbol{e}} = \{\bar{\boldsymbol{e}}_c\}_{c=0}^{C-1}$.

procedure CLS_SECAGG($\boldsymbol{y}$, $\boldsymbol{e}$)
  Active party P_0 maps label vector $\boldsymbol{y}$ to mask vectors $\{\boldsymbol{m}_c\}_{c=0}^{C-1}$ for each class c.
  The key distribution server generates Beaver's triples and distributes to P_0 and P_1.
  for each class c in parallel **do**
    P_0 and P_1 secret share $\boldsymbol{m}_c$ and $\boldsymbol{e}$ with each other and get shared $\langle \boldsymbol{m}_c \rangle$ and $\langle \boldsymbol{e} \rangle$.
    P_0 and P_1 compute inner product $\langle \boldsymbol{m}_c \cdot \boldsymbol{e} \rangle$ via Beaver's triples technique.
    P_0 and P_1 reconstruct $\langle \boldsymbol{m}_c \cdot \boldsymbol{e} \rangle$ to P_0 to get $\boldsymbol{m}_c \cdot \boldsymbol{e}$.
    P_0 computes the average embedding $\bar{\boldsymbol{e}}_c = \frac{\boldsymbol{m}_c \cdot \boldsymbol{e}}{\sum \boldsymbol{m}_c}$.
  end for
  Return Average embeddings $\bar{\boldsymbol{e}} = \{\bar{\boldsymbol{e}}_c\}_{c=0}^{C-1}$.
end procedure

Backup Slides on Secure Computation

Security and Privacy-Preservation in FL

- Security Definitions:
  - Honest and incurious security:
    - The adversaries honestly follow the protocol and do not infer knowledge.
  - Semi-honest security:
    - The adversaries strictly follow the prespecified protocol without deviation. However, they will collect all received intermediate data and try to derive knowledge from it.
  - Malicious security:
    - The adversaries can arbitrarily deviate from the protocol in their attempt to cheat.
- Semi-honest security attracts the most attention as
  - Semi-honest security balances security and efficiency.
  - Malicious security can be achieved by adding message commitment to semi-honest protocols.

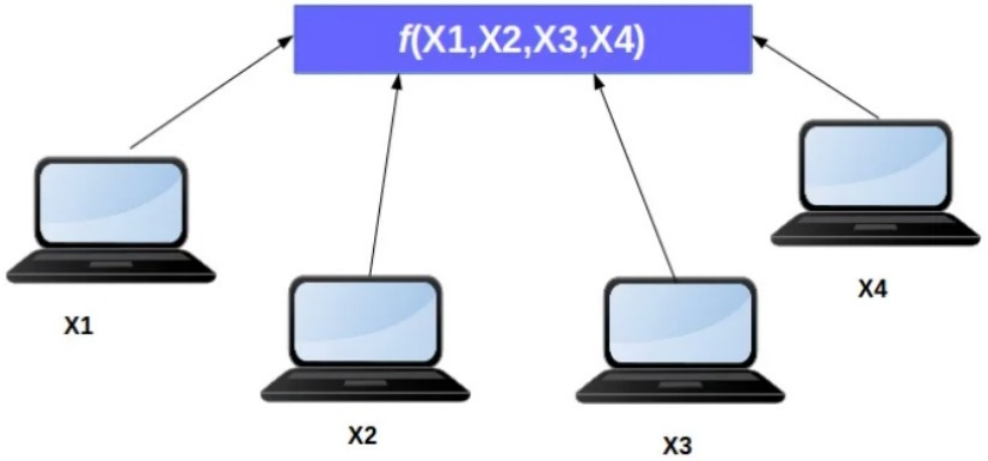
Security and Privacy-Preservation in FL

General Privacy Attacks

- Attack：
  1. Model inversion attack (learn model feature representation of each class)
  2. Data reconstruction attack (infer training data)
  3. Membership inference attack
- Phase:
  - Training phase/ inference phase
- Access:
  - White box access/ black box access

Security and Privacy-Preservation in FL - Key Security Techniques

- Secure Multi-party Computation (MPC) [Yao, 1982]
  - Subfield of cryptography
  - Enables parties to jointly compute a function over their inputs while keeping those inputs private
  - Low efficiency (communication cost) & no utility loss
  - MPC can be achieved by:
    - Secret sharing
    - Oblivious transfer
    - Threshold homomorphic encryption



Security and Privacy-Preservation in FL - Key Security Techniques

- Homomorphic Encryption (HE) [Rivest et al., 1978]
  - Enables algebraic operations on ciphertexts
  - Low efficiency (computation costs) & no utility loss
  - Partial HE vs. Fully HE

- Paillier [Paillier, 1999]
  - additive HE

$$\begin{aligned} E(m_1) * E(m_2) &= (g^{m_1} r_1^n \pmod{n^2}) * (g^{m_2} r_2^n \pmod{n^2}) \\ &= g^{m_1+m_2}(r1 * r2)^n \pmod{n^2} \\ &= E(m_1 + m_2) \end{aligned}$$

- RSA [Rivest et al., 1978]
  - multiplicative HE

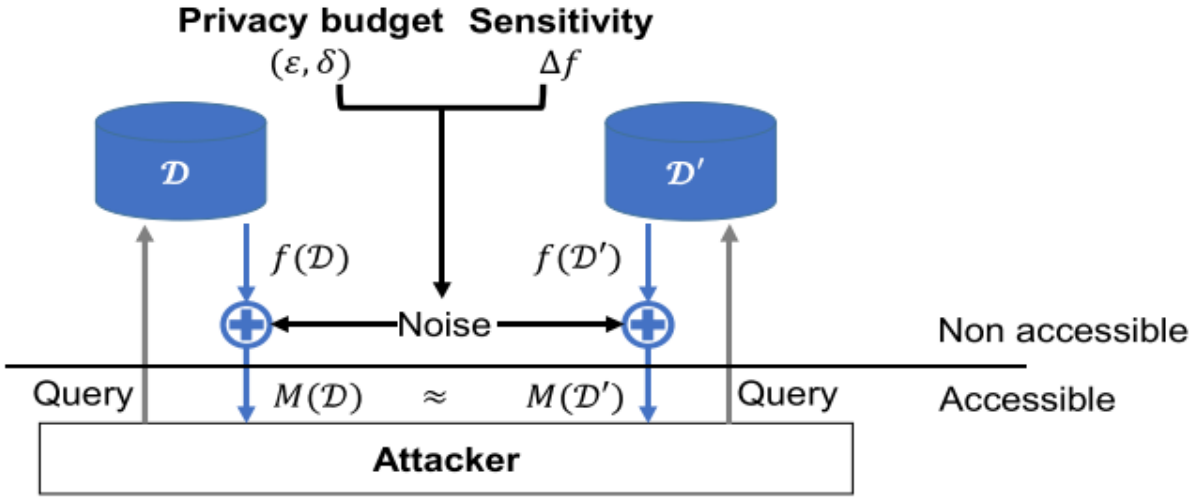
$$\begin{aligned} E(m_1) * E(m_2) &= (m_1^e \pmod{n}) * (m_2^e \pmod{n}) \\ &= (m_1 * m_2)^e \pmod{n} \\ &= E(m_1 * m_2). \end{aligned}$$

Security and Privacy-Preservation in FL - Key Security Techniques

- Differential Privacy (DP) [Dwork, 2006]
  - A randomized mechanism that protects **membership** information.
  - Guarantees an adversary cannot deduce any membership information with high confidence from released datasets, models, or gradients.
  - High efficiency & utility loss

- (ε, δ)-DP:

$$\Pr[M(\mathcal{D}) \in S] \leq \exp(\varepsilon)\Pr[M(\mathcal{D}') \in S] + \delta$$



[Bae, et al., 2018]