NEURAL INFORMATION PROCESSING SYSTEMS

PubSub-VFL: Towards Efficient Two-Party Split Learning in Heterogeneous Environments via Publisher/Subscriber Architecture

Yi Liu[1], Yang Liu[2], Leqian Zheng[1], Jue Hong[2], Junjie Shi[2],
Qingyou Yang[2], Ye Wu[2], Cong Wang[1]
[1]City University of Hong Kong, [2]ByteDance
yiliu247-c@my.cityu.edu.hk; WeChat: Sea_AAo

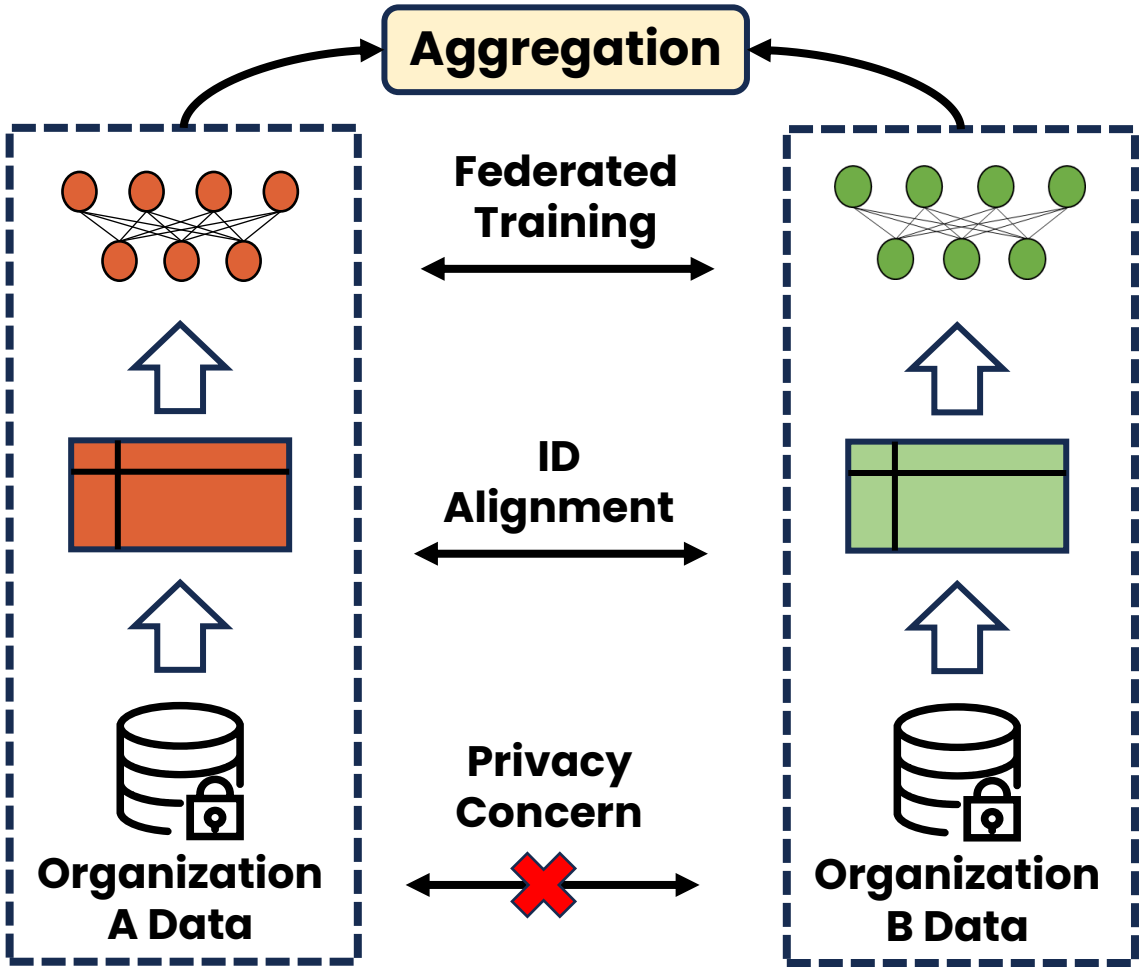
香港城市大學
City University of Hong Kong
Innovating into the Future

ByteDance

BACKGROUND

College of Computing
香港城市大學
City University of Hong Kong

Vertical Federated Learning (VFL) is applicable to the cases that two data sets share the same sample ID space but differ in feature space.



Organization A Data

ID	x1	x2	x3	y
0	value	value	value	0
1	value	value	value	1
2	value	value	value	0
3	value	value	value	1

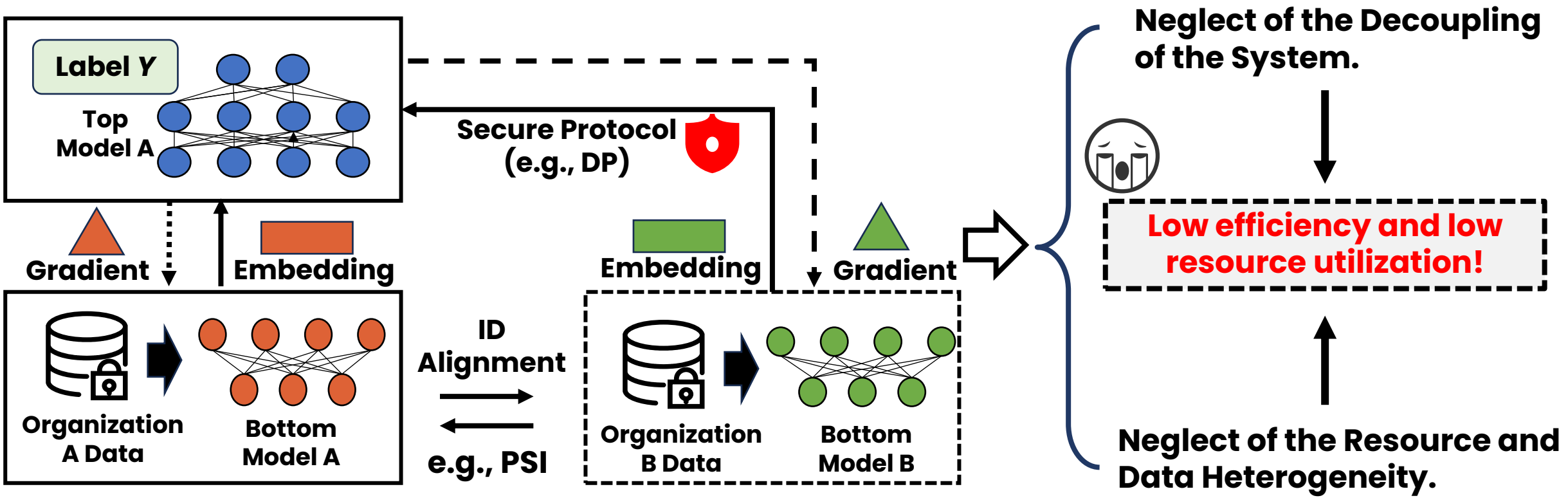
Organization B Data

ID	x4	x5	x6	x7
0	value	value	value	value
2	value	value	value	value
3	value	value	value	value
4	value	value	value	value

Split Learning-based VFL

College of Computing
香港城市大學
City University of Hong Kong

Two-Party Split Learning-based VFL has emerged as a promising solution for secure collaborative learning.

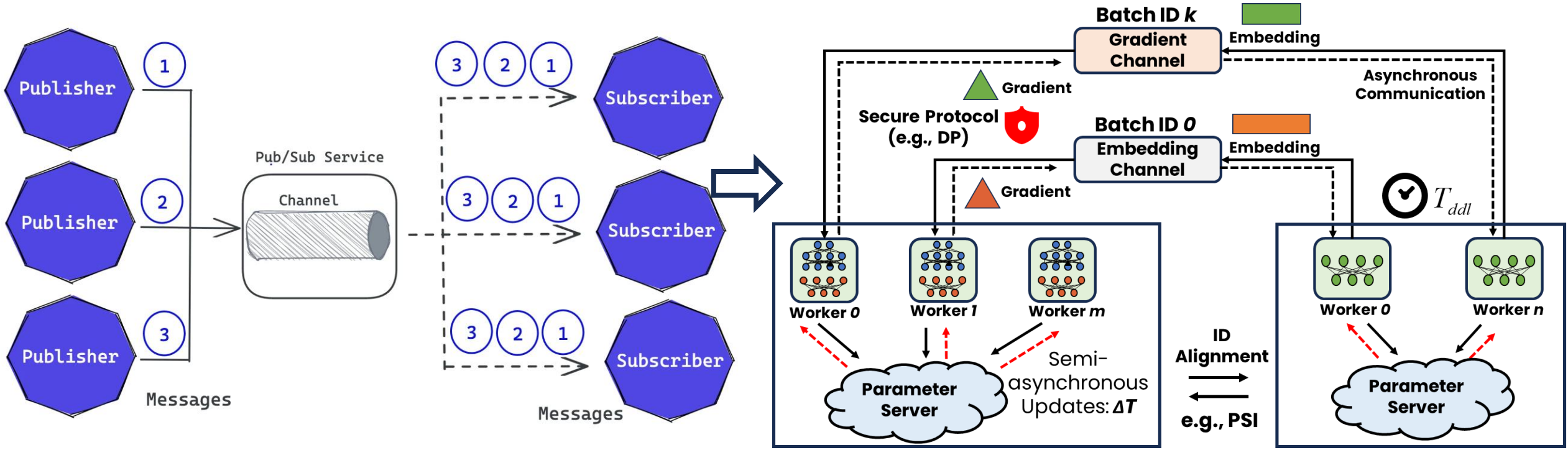


Overview of the split learning-based VFL framework.

PubSub-VFL

College of Computing
香港城市大學
City University of Hong Kong

Core Design 1: We first leverage the ***loose coupling of the PubSub*** model to enable asynchronous exchange of gradients and embeddings in VFL, thereby improving efficiency.

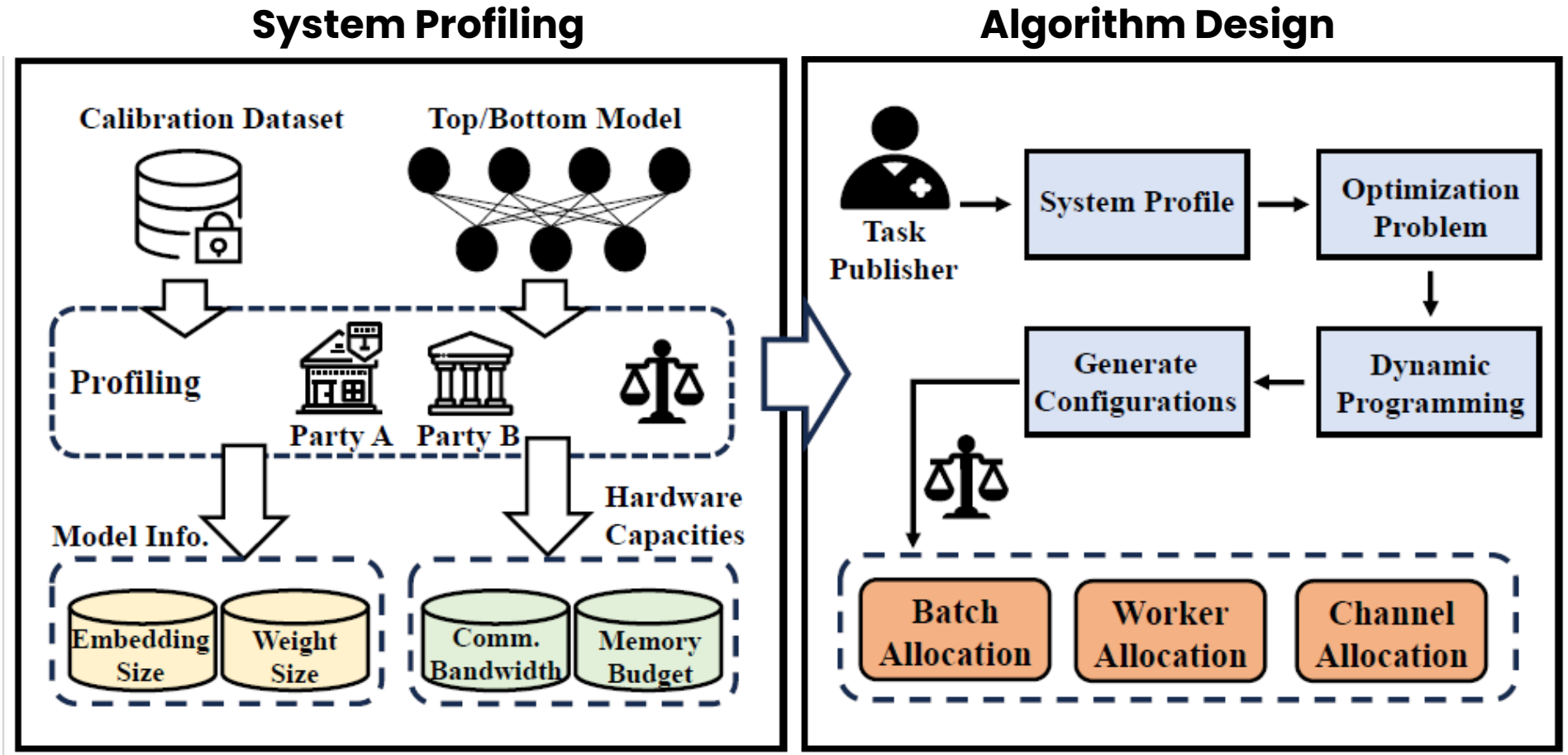


PubSub Model **PubSub-VFL Model**

PubSub-VFL

College of Computing
香港城市大學
City University of Hong Kong

Core Design 2: We ***formalize an optimization problem*** based on participants' system profiles, enabling the selection of optimal hyperparameters while preserving privacy to mitigate the training imbalance caused by resource and data heterogeneity.



RESULTS

College of Computing
香港城市大學
City University of Hong Kong

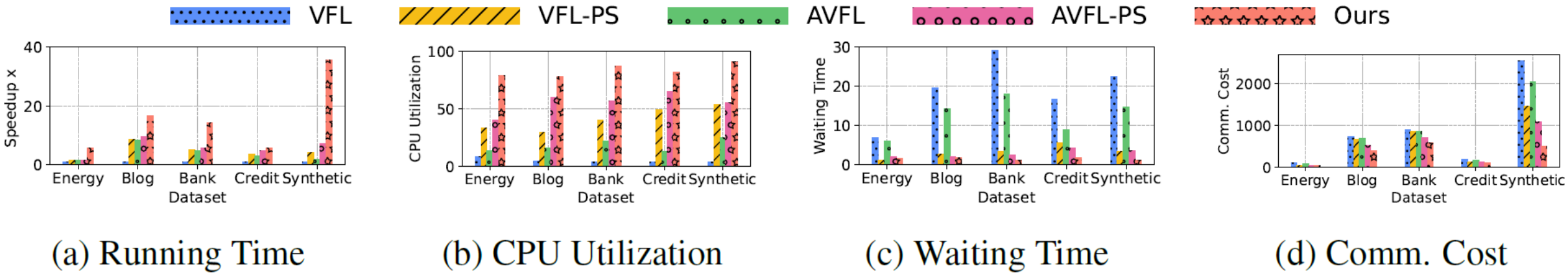


Fig. 1. Comparison with existing baselines in computation and communication efficiency.

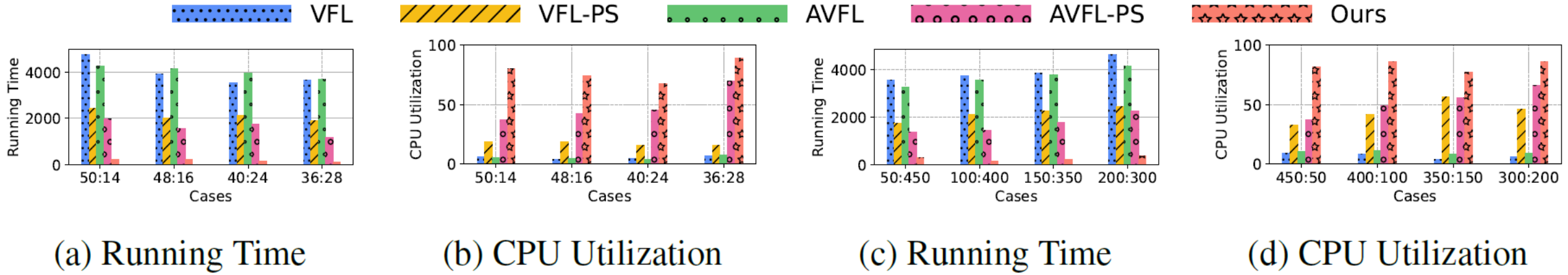


Fig. 2. Comparison with existing baselines on computation efficiency in resource and data heterogeneous scenarios.

RESULTS

College of Computing
香港城市大學
City University of Hong Kong

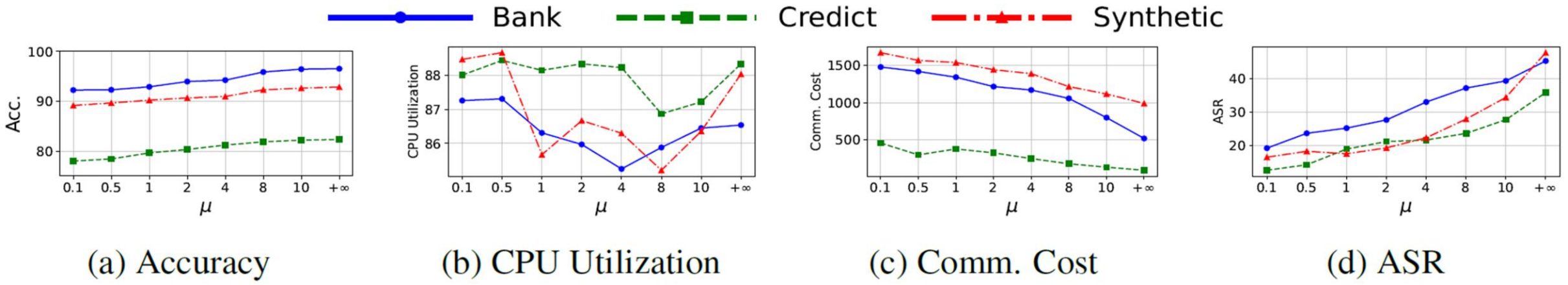


Fig. 3. The impact of privacy budget on the performance, efficiency, and security of PubSub-VFL.

Table 1: Effect of the number of workers.

# of Workers	4	5	**8***	10	20	30	50
Acc.(%)	92.13	92.05	**92.06**	92.28	92.00	92.36	92.21
Time (s)	712.78	805.90	**668.11**	885.01	1420.32	1067.57	1661.74
CPU (%)	67.52	63.30	**88.04**	76.18	42.77	40.78	45.12
Waiting (s)	1.4686	1.9273	**1.5288**	3.461	8.088	9.687	19.843
Comm. (MB)	878.91	1098.63	**888.77**	1318.36	1867.68	1538.09	2197.27

Table 2: Effect of the different batch size.

Batch Size	16	32	64	128	**256**	512	1024
Acc.(%)	91.70	92.06	91.75	92.63	**92.67**	92.36	92.21
Time (s)	987.64	668.11	344.76	124.01	**92.54**	578.69	865.74
CPU (%)	48.64	88.04	90.12	89.97	**91.07**	84.47	52.67
Waiting (s)	1.087	1.5288	1.688	1.263	**1.1389**	1.324	1.789
Comm. (MB)	1298.32	888.77	329.59	439.45	**439.45**	736.89	1070.36

Thanks for your Listening!