Towards Diverse Device Heterogeneous Federated Learning via Task Arithmetic Knowledge Integration

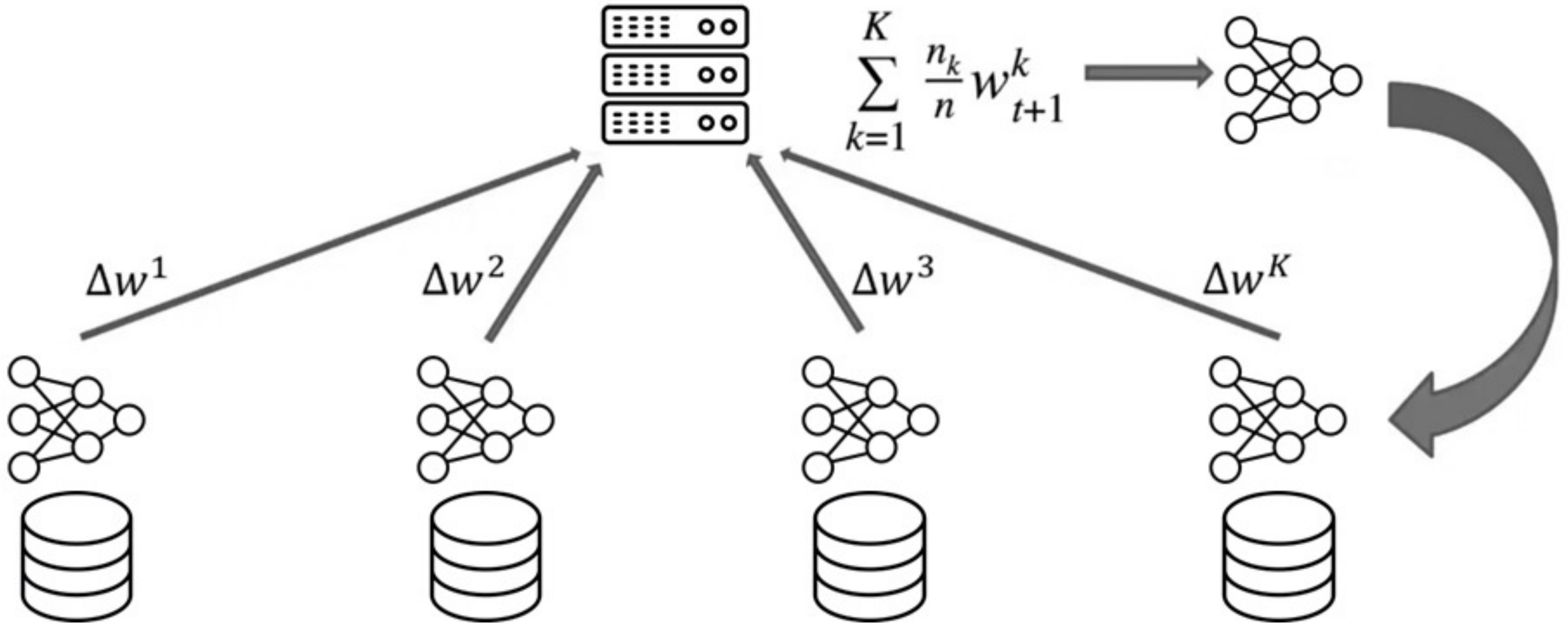
Mahdi Morafah[1], Vyacheslav Kungurtsev[2], Hojin Chang[1], Chen Chen[3], Bill Lin[1]

[1]University of California, San Diego
[2]Czech Technical University in Prague
[3]University of Central Florida

NeurIPS 2024

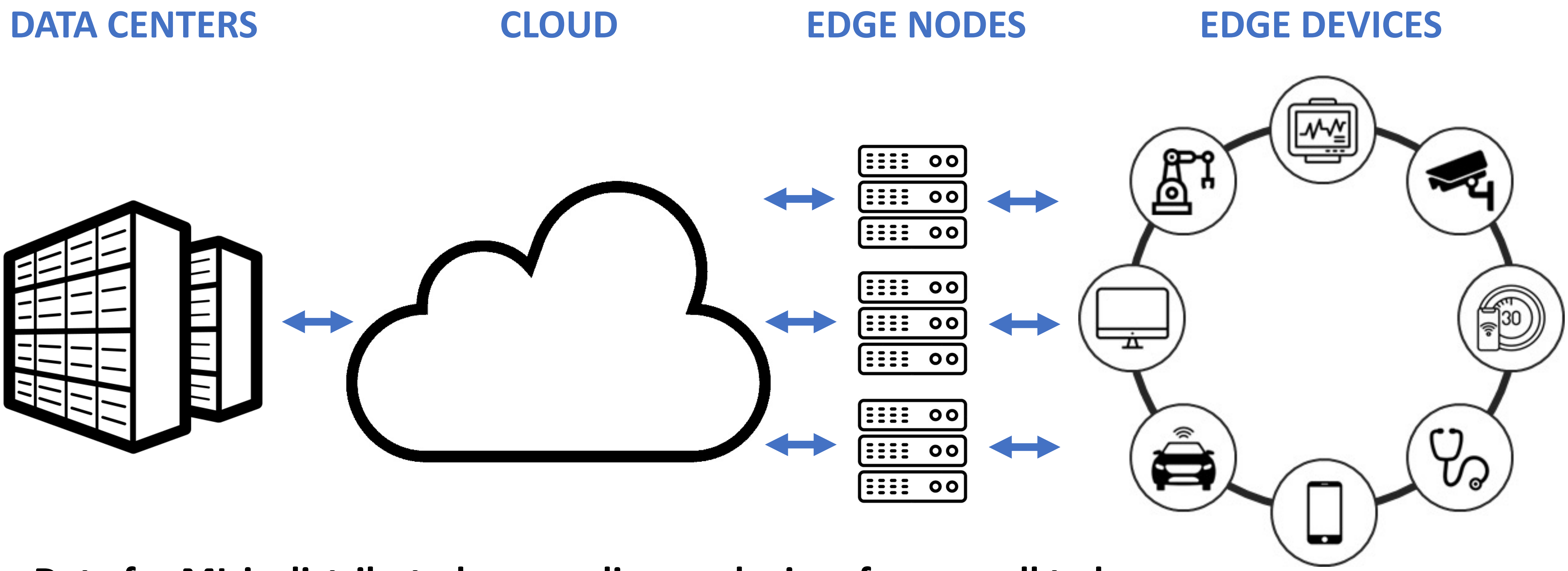
NEURAL INFORMATION PROCESSING SYSTEMS

Standard Federated Learning (FedAvg)



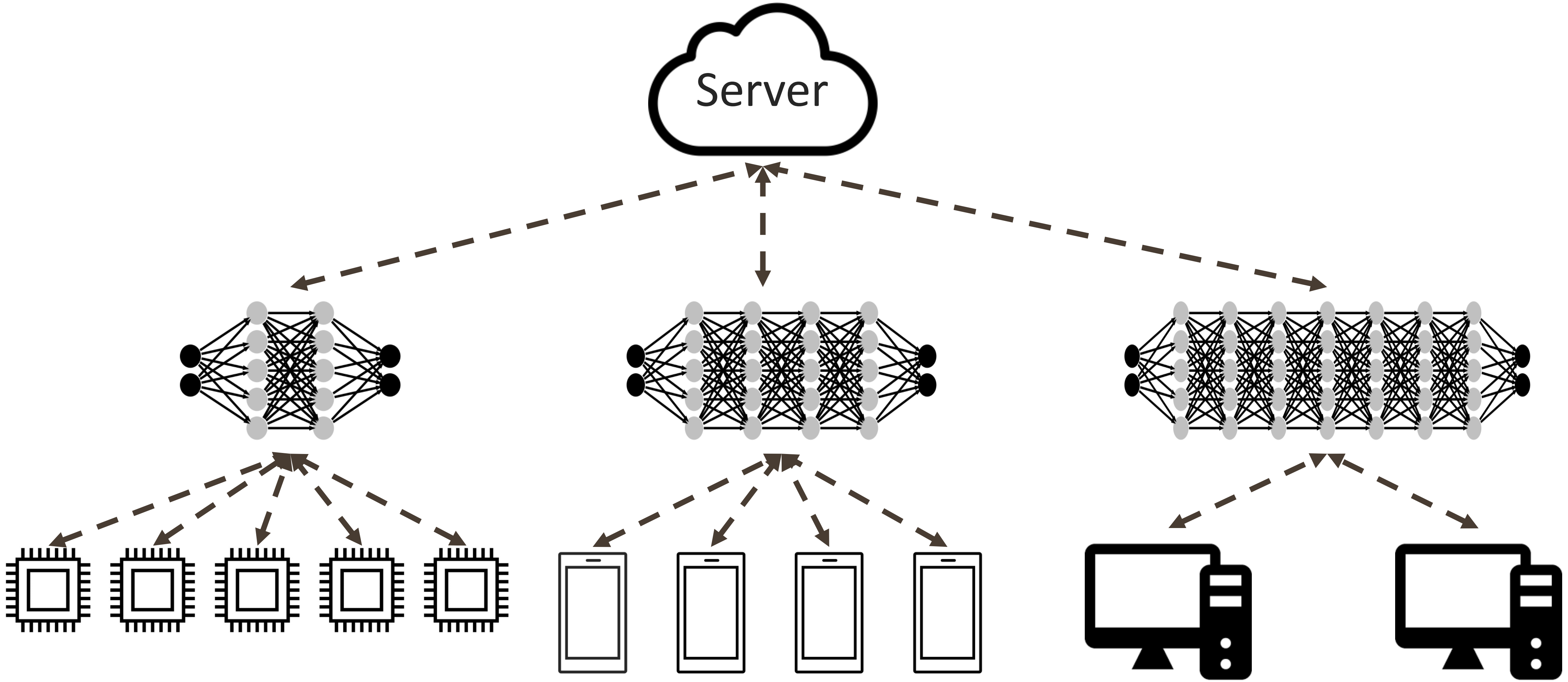
- **Assumes clients can train an identical model**
- **Clients train on local data, weights averaged at the server**
- **Global shared model benefits from all client data**

Heterogeneous Device Prototypes

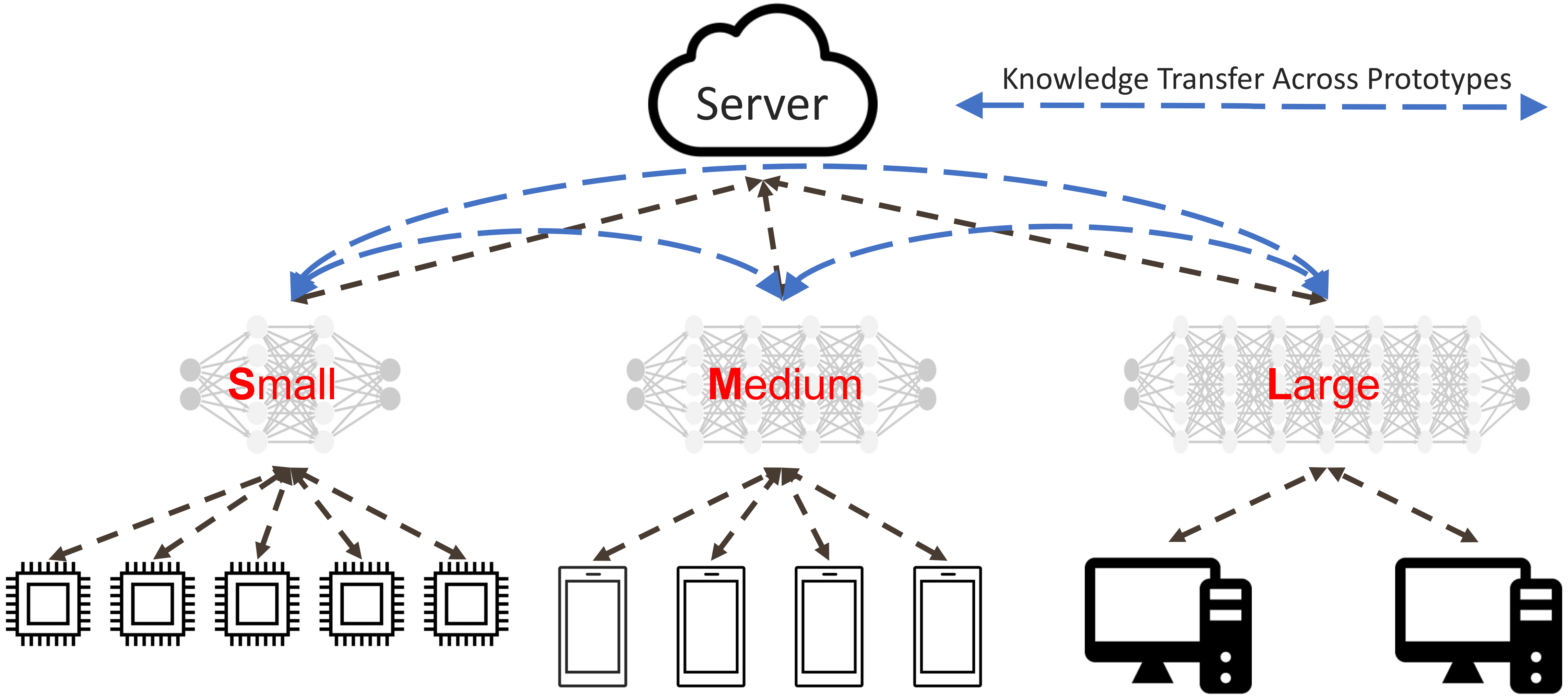


- Data for ML is distributed across diverse devices from small to large
- Learn ML models from diverse devices while maintaining data privacy

FL with Heterogeneous Device Prototypes



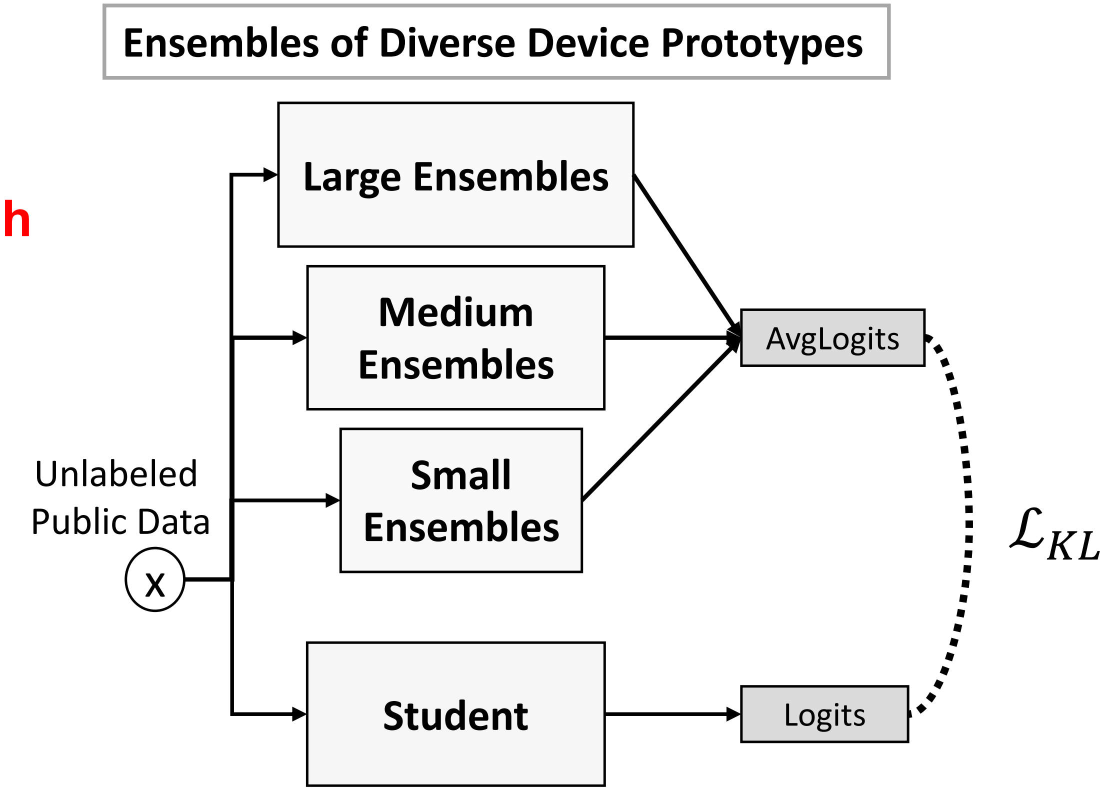
FL with Heterogeneous Device Prototypes



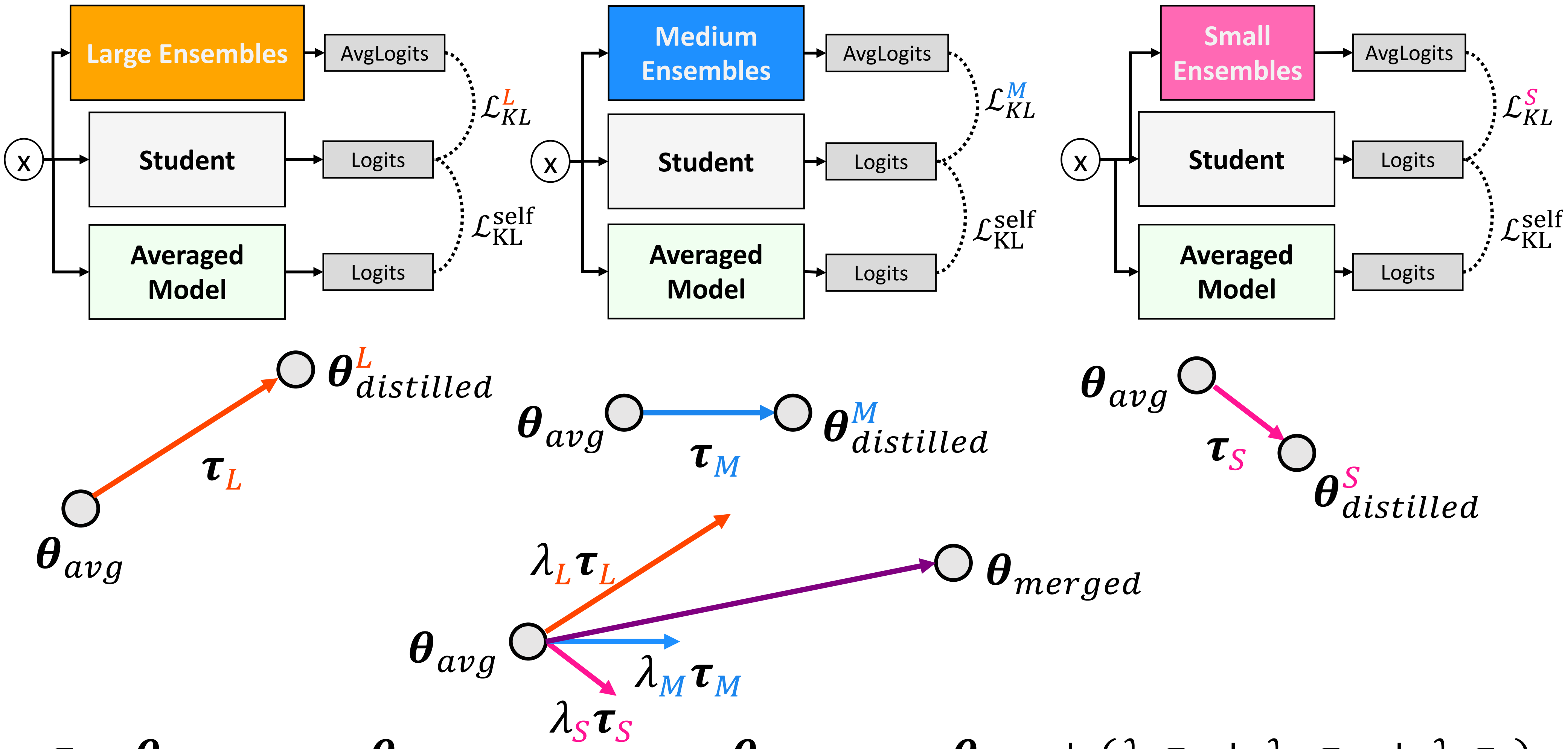
Limitations of Existing KD-based Methods

Information Dilution

one-size-fits-all approach



Proposed Method: TAKFL



$$\boldsymbol{\tau} = \boldsymbol{\theta}_{distilled} - \boldsymbol{\theta}_{avg}$$

$$\boldsymbol{\theta}_{merged} = \boldsymbol{\theta}_{avg} + (\lambda_L\boldsymbol{\tau}_L + \lambda_M\boldsymbol{\tau}_M + \lambda_S\boldsymbol{\tau}_S)$$

Experiment Results (CV Task)

Table 4: **Performance Results for CV task on CIFAR-10 and CIFAR-100.** Training data is distributed among S, M, and L device prototypes in a 1:3:6 ratio, subdivided among clients using Dirichlet distribution. Public datasets are CIFAR-100 for CIFAR-10 and ImageNet-100 for CIFAR-100. Client configurations include 100, 20, and 4 clients for S, M, and L, with sampling rates of 0.1, 0.2, and 0.5. In homo-family settings, architectures are ResNet8, ResNet14, and ResNet18; in hetero-family settings, they are ViT-S, ResNet14, and VGG-16. All models are trained from scratch for 60 rounds. See Appendix F.1 for more details.

		Homo-family Architecture Setting							
Dataset	Baseline	Low Data Heterogeneity				High Data Heterogeneity			
		S	M	L	Average	S	M	L	Average
CIFAR-10	FedAvg	$36.21_{\pm 2.24}$	$46.41_{\pm 2.33}$	$59.46_{\pm 6.17}$	47.36	$22.01_{\pm 0.78}$	$25.26_{\pm 3.89}$	$51.51_{\pm 3.52}$	32.93
	FedDF	$49.31_{\pm 0.15}$	$50.63_{\pm 0.73}$	$49.82_{\pm 0.98}$	49.92	$34.71_{\pm 1.48}$	$35.27_{\pm 4.74}$	$51.08_{\pm 4.04}$	40.35
	FedET	$49.21_{\pm 0.72}$	$55.01_{\pm 1.81}$	$53.60_{\pm 6.47}$	52.61	$29.58_{\pm 3.00}$	$30.96_{\pm 4.70}$	$45.53_{\pm 6.46}$	35.36
	TAKFL	$55.90_{\pm 1.70}$	$57.93_{\pm 3.49}$	$60.58_{\pm 2.35}$	58.14	$37.40_{\pm 1.68}$	$38.96_{\pm 0.17}$	$51.49_{\pm 6.15}$	42.61
	TAKFL+Reg	$\mathbf{56.37}_{\pm \mathbf{0.46}}$	$\mathbf{58.60}_{\pm \mathbf{0.43}}$	$\mathbf{65.69}_{\pm \mathbf{1.28}}$	**60.22**	$\mathbf{40.51}_{\pm \mathbf{1.05}}$	$\mathbf{40.12}_{\pm \mathbf{1.24}}$	$\mathbf{53.24}_{\pm \mathbf{2.51}}$	**44.62**
CIFAR-100	FedAvg	$13.22_{\pm 0.14}$	$21.39_{\pm 1.11}$	$29.47_{\pm 0.86}$	21.36	$11.86_{\pm 0.08}$	$14.63_{\pm 0.65}$	$26.25_{\pm 1.64}$	17.58
	FedDF	$19.54_{\pm 0.20}$	$24.32_{\pm 0.45}$	$29.29_{\pm 1.45}$	24.38	$16.09_{\pm 0.32}$	$19.80_{\pm 0.17}$	$26.59_{\pm 0.25}$	20.83
	FedET	$19.67_{\pm 0.35}$	$25.27_{\pm 0.66}$	$31.10_{\pm 1.53}$	25.35	$11.18_{\pm 1.68}$	$18.22_{\pm 0.35}$	$26.40_{\pm 0.65}$	18.60
	TAKFL	$24.48_{\pm 0.42}$	$27.60_{\pm 0.25}$	$29.84_{\pm 0.94}$	27.31	$\mathbf{22.90}_{\pm \mathbf{0.18}}$	$23.63_{\pm 0.72}$	$26.98_{\pm 0.13}$	24.50
	TAKFL+Reg	$\mathbf{27.18}_{\pm 0.27}$	$\mathbf{29.14}_{\pm 0.20}$	$\mathbf{31.15}_{\pm \mathbf{0.97}}$	**29.16**	$22.88_{\pm 0.37}$	$\mathbf{23.92}_{\pm 0.57}$	$\mathbf{28.01}_{\pm 0.34}$	**24.94**
		Hetero-family Architecture Setting							
Dataset	Baseline	Low Data Heterogeneity				High Data Heterogeneity			
		S	M	L	Average	S	M	L	Average
CIFAR-10	FedAvg	$27.53_{\pm 0.83}$	$47.30_{\pm 3.17}$	$55.10_{\pm 8.60}$	43.31	$20.93_{\pm 1.54}$	$25.62_{\pm 6.04}$	$36.80_{\pm 5.47}$	27.78
	FedDF	$34.15_{\pm 0.87}$	$54.06_{\pm 1.06}$	$69.07_{\pm 4.99}$	52.43	$24.20_{\pm 0.74}$	$34.07_{\pm 3.08}$	$39.81_{\pm 5.45}$	32.69
	FedET	$33.24_{\pm 1.27}$	$58.86_{\pm 0.94}$	$65.56_{\pm 3.49}$	52.55	$24.37_{\pm 1.26}$	$37.77_{\pm 4.71}$	$43.64_{\pm 3.36}$	35.26
	TAKFL	$33.29_{\pm 0.15}$	$57.64_{\pm 0.19}$	$68.44_{\pm 0.66}$	53.12	$24.92_{\pm 1.32}$	$38.07_{\pm 3.19}$	$48.01_{\pm 3.99}$	37.00
	TAKFL+Reg	$\mathbf{33.34}_{\pm \mathbf{3.36}}$	$\mathbf{59.01}_{\pm \mathbf{3.12}}$	$\mathbf{70.22}_{\pm \mathbf{4.40}}$	**54.19**	$\mathbf{25.10}_{\pm \mathbf{1.87}}$	$\mathbf{38.81}_{\pm \mathbf{5.36}}$	$\mathbf{50.26}_{\pm \mathbf{6.42}}$	**38.06**
CIFAR-100	FedAvg	$8.51_{\pm 0.37}$	$22.11_{\pm 0.58}$	$37.91_{\pm 2.60}$	22.84	$7.01_{\pm 0.47}$	$14.94_{\pm 0.96}$	$28.51_{\pm 1.46}$	16.82
	FedDF	$10.46_{\pm 0.17}$	$23.46_{\pm 0.65}$	$36.81_{\pm 0.82}$	23.58	$7.76_{\pm 0.40}$	$18.92_{\pm 0.39}$	$29.81_{\pm 1.09}$	18.83
	FedET	$11.16_{\pm 0.18}$	$25.40_{\pm 0.30}$	$37.38_{\pm 0.60}$	24.65	$8.20_{\pm 0.54}$	$20.66_{\pm 0.50}$	$28.95_{\pm 1.79}$	19.27
	TAKFL	$10.29_{\pm 0.11}$	$27.14_{\pm 0.89}$	$\mathbf{39.15}_{\pm \mathbf{0.88}}$	25.53	$7.88_{\pm 0.68}$	$21.41_{\pm 0.37}$	$31.31_{\pm 0.66}$	20.20
	TAKFL+Reg	$\mathbf{11.25}_{\pm \mathbf{0.37}}$	$\mathbf{27.86}_{\pm \mathbf{0.86}}$	$38.68_{\pm 0.45}$	**25.93**	$\mathbf{8.45}_{\pm \mathbf{0.20}}$	$\mathbf{22.16}_{\pm \mathbf{0.87}}$	$\mathbf{31.95}_{\pm \mathbf{1.13}}$	**20.85**

Experiment Results (NLP Task)

Table 6: **Performance Results for NLP Task on 4 Datasets.**. Training data is distributed among S, M, and L device prototypes in a 1:3:6 ratio, subdivided among clients using Dir(0.5). Client configurations are 8, 4, and 2 clients for S, M, and L, with sample rates of 0.3, 0.5, and 1.0, respectively. Architectures include Bert-Tiny, Bert-Mini, and Bert-Small for S, M, and L, initialized from pre-trained parameters and fine-tuned for 20 communication rounds. See Appendix F.2 for more details.

Private	Public	Baseline	S	M	L	Average
MNLI	SNLI	FedAvg	$36.15_{\pm 0.46}$	$54.47_{\pm 2.48}$	$57.51_{\pm 2.79}$	49.37
		FedDF	$54.21_{\pm 0.15}$	$60.44_{\pm 1.91}$	$66.71_{\pm 1.09}$	60.45
		FedET	$48.03_{\pm 6.32}$	$50.33_{\pm 7.87}$	$53.80_{\pm 6.18}$	50.72
		TAKFL	$57.43_{\pm 0.21}$	$63.58_{\pm 0.31}$	$68.74_{\pm 0.12}$	63.25
		TAKFL+Reg	$\mathbf{57.61_{\pm 0.89}}$	$\mathbf{63.91_{\pm 1.05}}$	$\mathbf{68.96_{\pm 1.10}}$	**63.49**
SST2	Sent140	FedAvg	$54.98_{\pm 1.81}$	$74.71_{\pm 8.22}$	$86.69_{\pm 0.06}$	72.13
		FedDF	$74.41_{\pm 2.62}$	$80.71_{\pm 1.63}$	$84.35_{\pm 1.66}$	79.82
		FedET	$66.63_{\pm 9.14}$	$65.89_{\pm 16.35}$	$70.05_{\pm 15.83}$	67.52
		TAKFL	$74.73_{\pm 0.55}$	$82.17_{\pm 0.31}$	$86.93_{\pm 0.42}$	81.28
		TAKFL+Reg	$\mathbf{74.88_{\pm 0.43}}$	$\mathbf{82.40_{\pm 0.83}}$	$\mathbf{87.33_{\pm 0.63}}$	**81.54**
MARC	Yelp	FedAvg	$33.76_{\pm 1.13}$	$49.08_{\pm 1.28}$	$59.26_{\pm 1.43}$	47.36
		FedDF	$53.01_{\pm 1.24}$	$55.37_{\pm 0.87}$	$56.81_{\pm 0.99}$	55.06
		FedET	$52.63_{\pm 2.29}$	$54.28_{\pm 2.31}$	$56.11_{\pm 2.61}$	54.34
		TAKFL	$55.70_{\pm 2.08}$	$58.64_{\pm 1.75}$	$59.39_{\pm 1.16}$	57.91
		TAKFL+Reg	$\mathbf{55.96_{\pm 1.66}}$	$\mathbf{59.18_{\pm 1.13}}$	$\mathbf{59.61_{\pm 1.89}}$	**58.25**
AG-News	DBPedia	FedAvg	$83.64_{\pm 3.51}$	$83.47_{\pm 2.35}$	$91.48_{\pm 2.22}$	86.20
		FedDF	$85.97_{\pm 2.45}$	$89.10_{\pm 1.85}$	$91.37_{\pm 1.10}$	88.81
		FedET	$75.27_{\pm 3.85}$	$81.13_{\pm 3.21}$	$83.19_{\pm 4.58}$	79.86
		TAKFL	$87.37_{\pm 1.31}$	$90.11_{\pm 1.56}$	$92.48_{\pm 1.12}$	89.99
		TAKFL+Reg	$\mathbf{87.66_{\pm 1.83}}$	$\mathbf{90.30_{\pm 2.05}}$	$\mathbf{92.61_{\pm 1.72}}$	**90.19**

Experiment Results (Scalability)

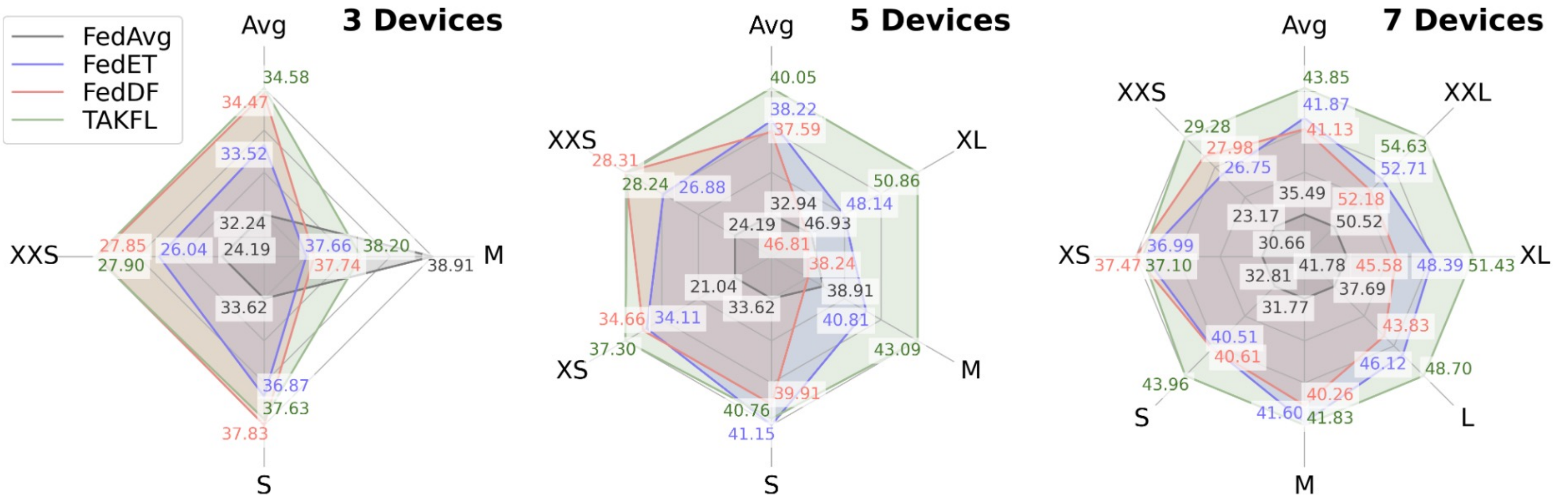


Figure 3: **Scalability Evaluation of TAKFL.** Image classification on CINIC-10 [9] dataset is used to evaluate TAKFL's scalability across device prototypes ranging from XXS to XXL. Training data is distributed among prototypes in a 1:2:3:4:5:6:7 ratio, further subdivided using Dir(0.5). Client configurations range from 35 for XXS to 5 for XXL. Architectures span from ResNet10-XXS for XXS to ResNet50 for XXL prototype, all initialized from scratch and trained over 30 communication rounds. The public dataset is CIFAR-100 [24]. See Appendix D.4 for more details.

Thank you for listening!

Feel free to reach out if you have any questions!