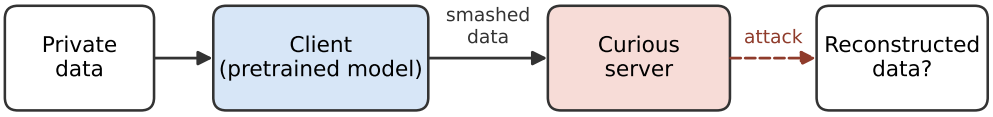
Split Learning Meets Pretrained Models: Revisiting Passive Inference Attacks



Split learning (SL) lets a client train a neural network with a powerful server by sending only intermediate representations (*smashed data*) instead of raw data. The recent SDAR attack [1] showed that an honest-but-curious server can nevertheless reconstruct the client's private images by training a simulator of the client model on public data. However, SDAR and earlier attacks were only evaluated on small CNNs trained from scratch, whereas real clients fine-tune public pretrained backbones such as ResNets (especially ResNet-50), Vision Transformers, or CLIP/DINOv2 encoders. This may strengthen the attack, since the server can start its simulator from the same public checkpoint, or weaken it, since fine-tuned transformer representations may be harder to simulate. Which effect dominates is unknown.

The goal of this thesis is to determine how private SL really is for modern pretrained models. The student is expected to reproduce SDAR using its public code [2] and extend it to pretrained CNNs and ViTs fine-tuned on CIFAR-100 and Tiny ImageNet datasets. Experiments will vary the split point, the attacker's knowledge of the checkpoint and the fine-tuning strategy (full, frozen layers, Low-Rank Adaptation (LoRA)), and measure leakage with perceptual (Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS)) and task-relevant privacy metrics. Finally, the student will design and evaluate a lightweight client-side defence against an adaptive attacker.

[1] X. Zhu et al., "Passive Inference Attacks on Split Learning via Adversarial Regularization," NDSS 2025.

[2] `https://github.com/zhxchd/SDAR_SplitNN`

Practicalities

Promotor:	Bart Preneel		
Daily supervisor:	Alireza Aghabagherloo	Alireza.Aghabagherloo@esat.kuleuven.be	office B01.27
	Roozbeh Sarenche	Roozbeh.Sarenche@esat.kuleuven.be	office B01.13
	Mahdi Rahimi	Mahdi.rahimi@esat.kuleuven.be	office B01.27
Nature of the work:	25% literature, 15% theoretical work, 60% software		
Number of students:	1		

 © 2026 – 2027 COSIC