
GradSec : a TEE-based Scheme Against Federated Learning Inference Attacks

ResilientFL, 2021

Workshop on Systems Challenges in Reliable and Secure Federated Learning

October 25th, 2021

***Aghiles AIT MESSAOUD¹, Sonia BEN MOKHTAR²,
Vlad NITU², Valerio SCHIAVONI³***

¹ESI, Algiers

²LIRIS-CNRS, France

³University of Neuchâtel, Switzerland

Summary

Introduction

State-of-the-art

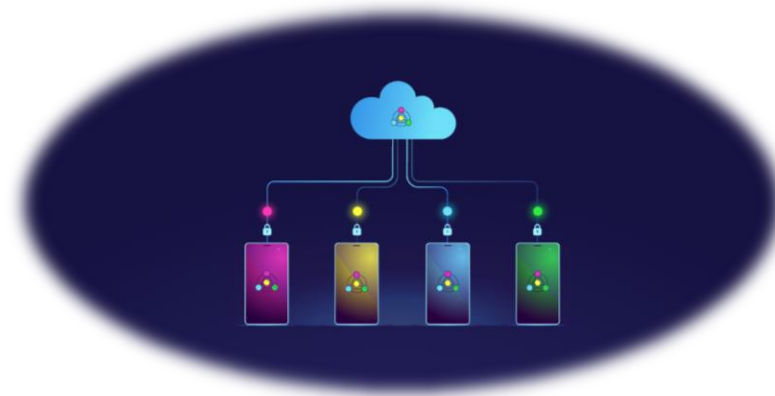
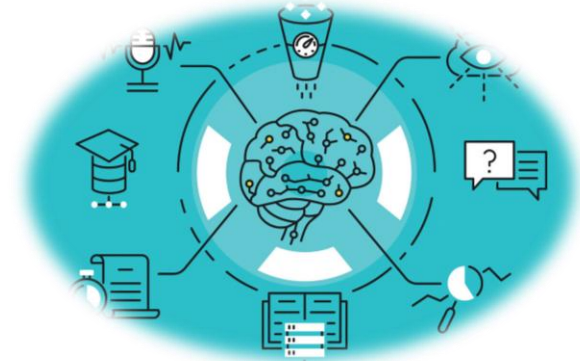
Contributions

Context

Customers are increasingly aware about their privacy

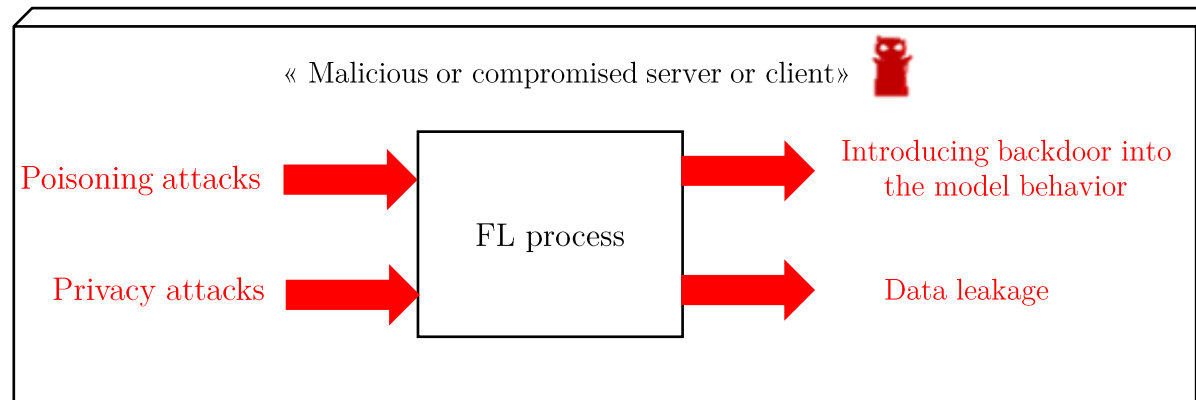


Learning models are increasingly fed with private data

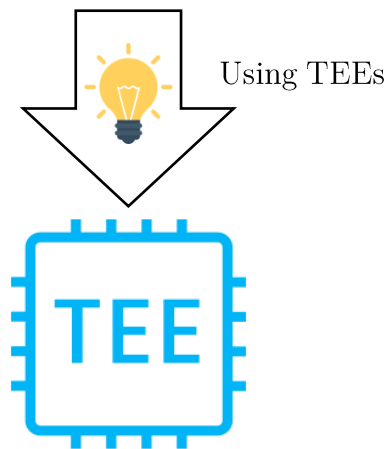


Advent of Federated Learning (FL) to ensure privacy-preserving training

Problem



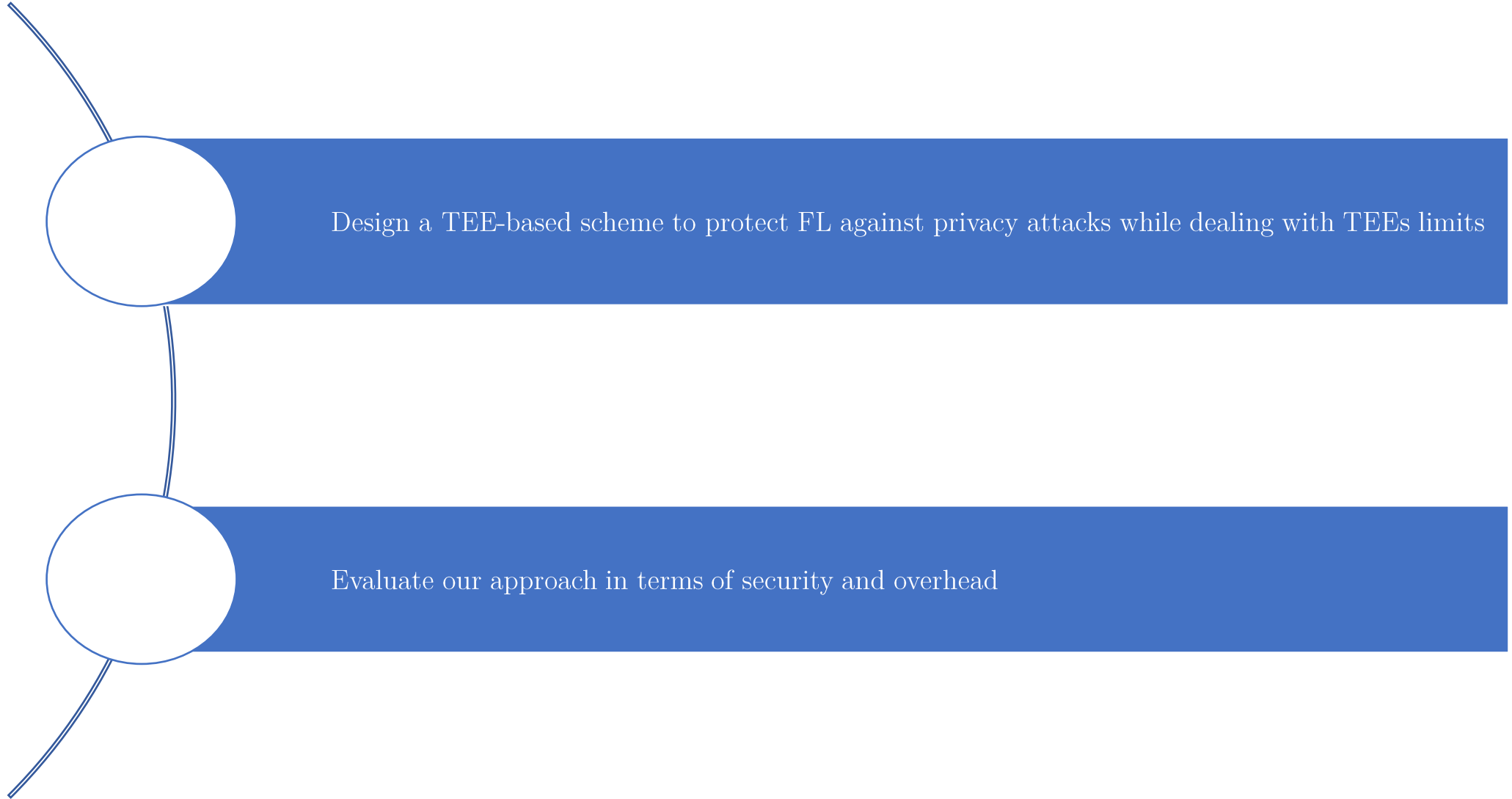
Problem 1 : FL is vulnerable to many attacks



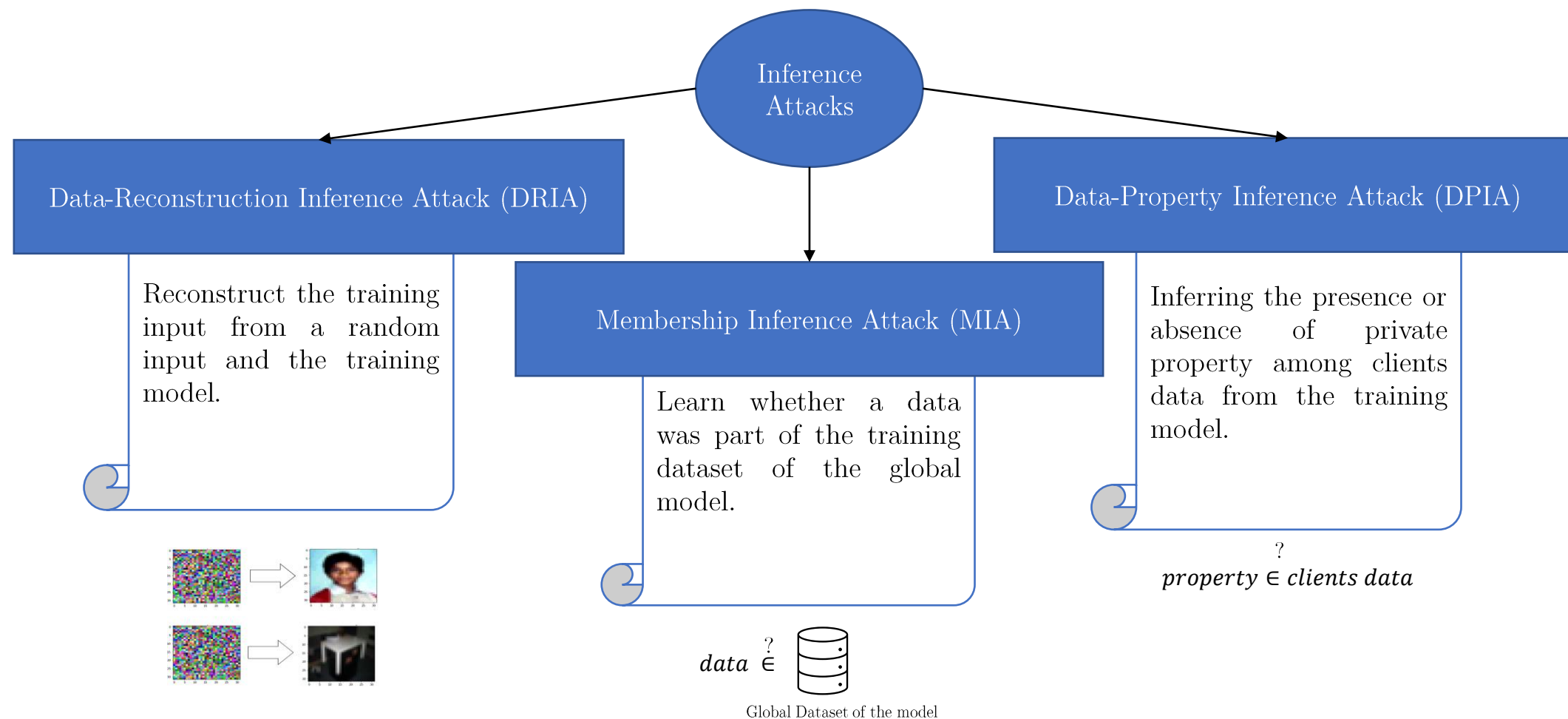
Problem 2 : TEEs offer limited secure memory (spatial constraint) and high latency (temporal constraint)



Objectives

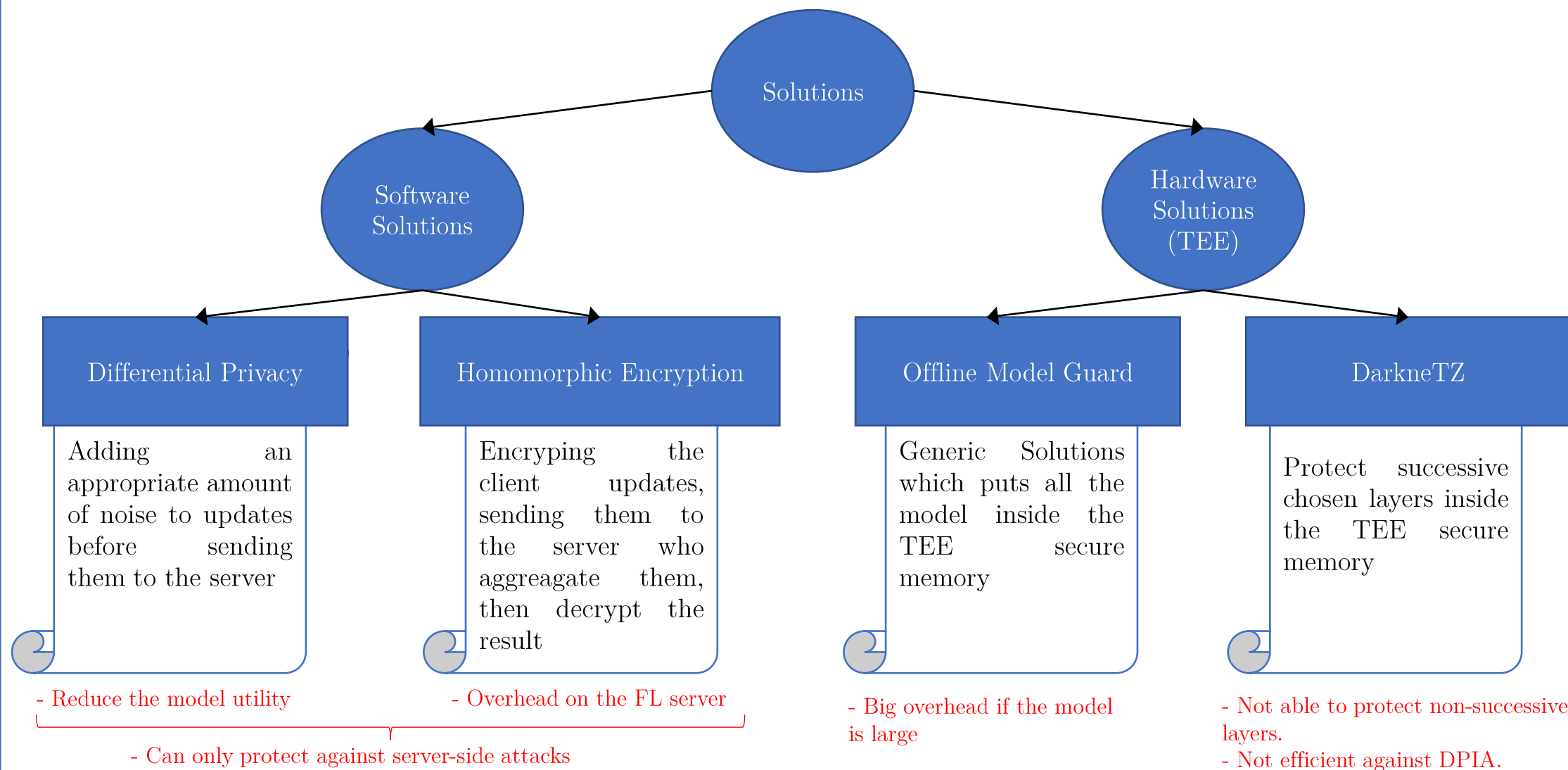


Inference Attacks in FL



Common point: The use of gradients emitted by the model to work

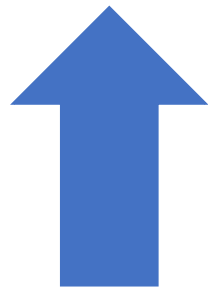
State-of-the-art approaches



Assumptions

- Securing FL against the most cited Inference Attacks (DRIA, MIA, DPIA).
- Considering the previous attacks carried out by the clients (the FL server uses Secure Aggregation).
- The FL models used are exclusively Feed-forward Neural Networks (Fully-connected or Convolutional, no Recurrent).
- The FL models use Stochastic Gradient Descent Algorithm [8] to update their weights.

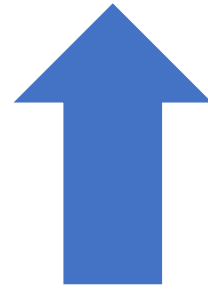
Securing models per layer



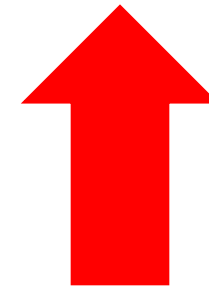
Number of protected layers in TEE



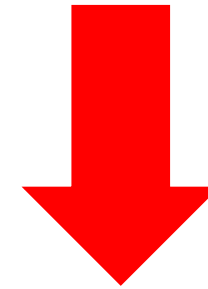
Number of hidden Gradients to the attacker



Improved security against Inference Attacks



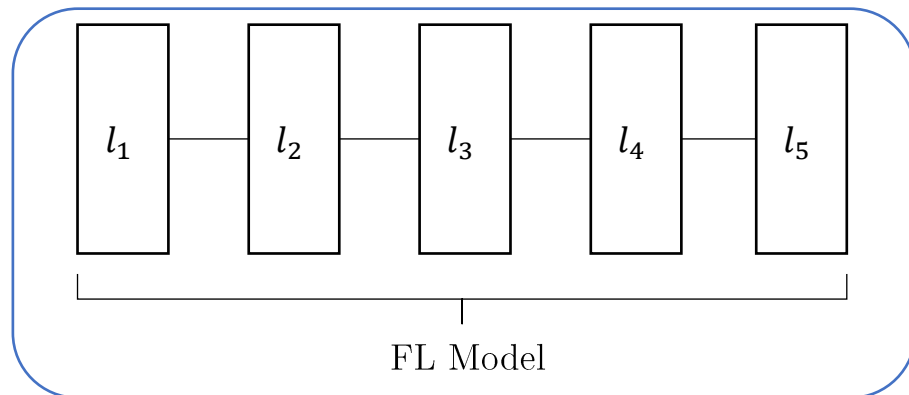
High Latency



Decreasing of available secure memory

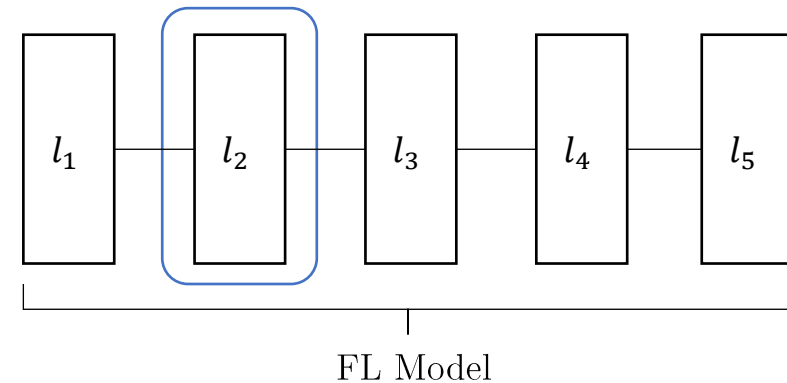
Ideal solution

TEE Enclave



Realistic solution

TEE Enclave



Sources of Gradients leakage of layer l

Source 1 : Compute the difference between two consecutive snapshots of the model

Formula to update weights model :

$$W_l^{(t+1)} \leftarrow W_l^{(t)} - \lambda \textcolor{red}{dW_l}$$

Consecutive weights of
the model



Gradients deduction

$$\textcolor{red}{dW_l} = \frac{W_l^{(t)} - W_l^{(t+1)}}{\lambda}$$

Solution : Put W_l in TEE secure memory

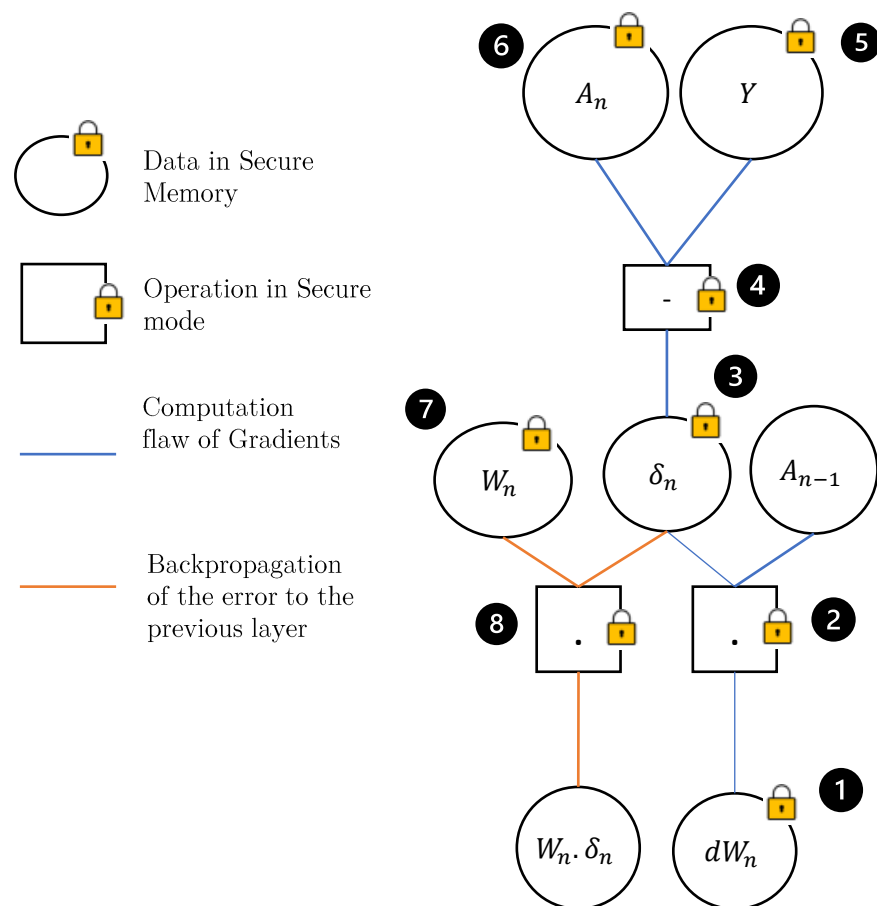
Source 2 : Backpropagation computation flaw

Operation	Designation
.	Regular dot product
\otimes	Convolutional dot product

Solution : Secure most important parts of backpropagation computation in the TEE Secure Memory

Securing Backpropagation using TEEs

Securing last layer ($l=n$) Gradients

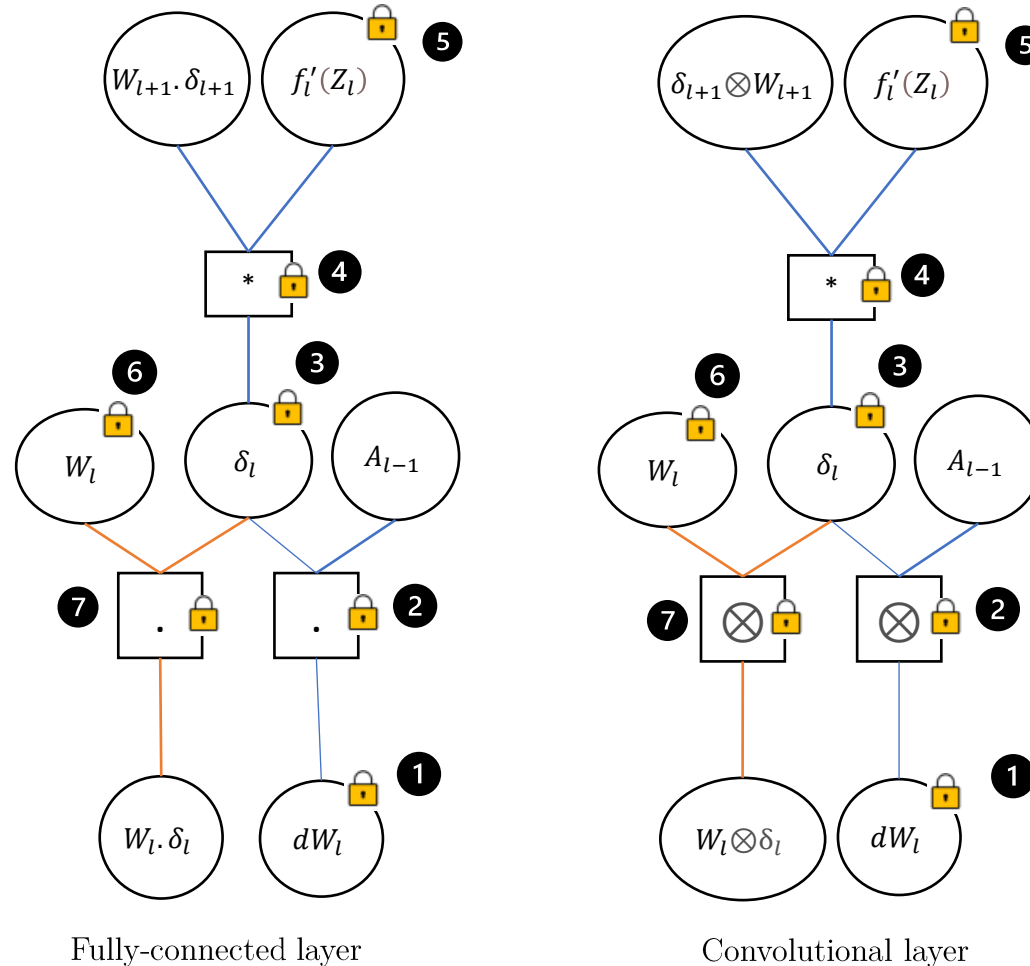


Secured Data/Operation	Justification
1	Represents the Gradients we want to secure
2	Avoid disclosing dW_n in the cache memory
3	Avoid the calculation of dW_n thanks to securing δ_n operand
4	Avoid disclosing δ_n in the cache memory
5	Avoid the computation of δ_n thanks to securing Y operand
6	Additional safety measure to avoid the computation of δ_n if Y is known by the attacker
7	Avoid source 1 of Gradients leakage to compute dW_n
8	Avoid disclosing δ_n or W_n in the cache memory

Operation	Designation
.	Regular dot product
*	Hadamard dot product
\otimes	Convolutional dot product

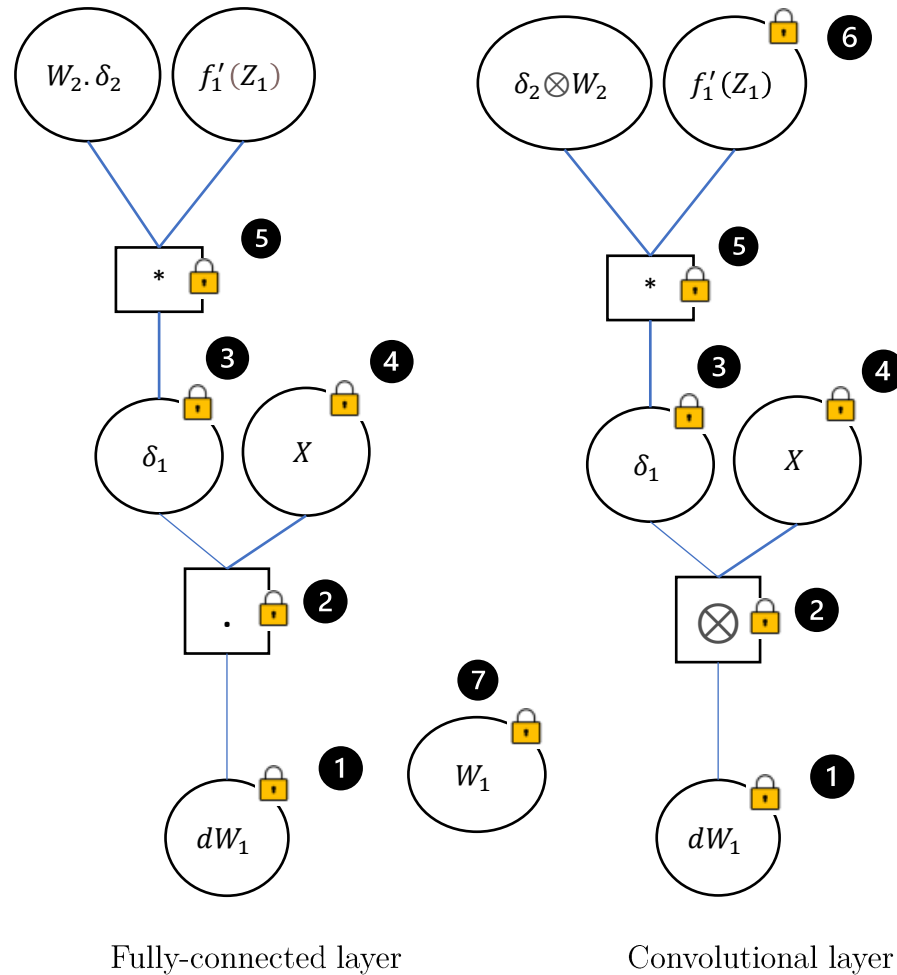
Securing Backpropagation using TEEs

Securing layer $1 < l < n$ Gradients

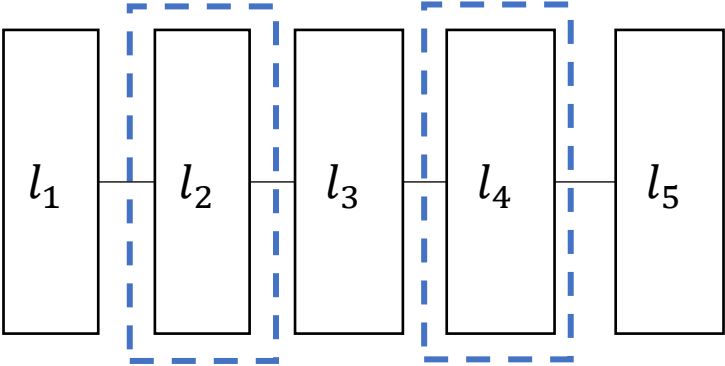
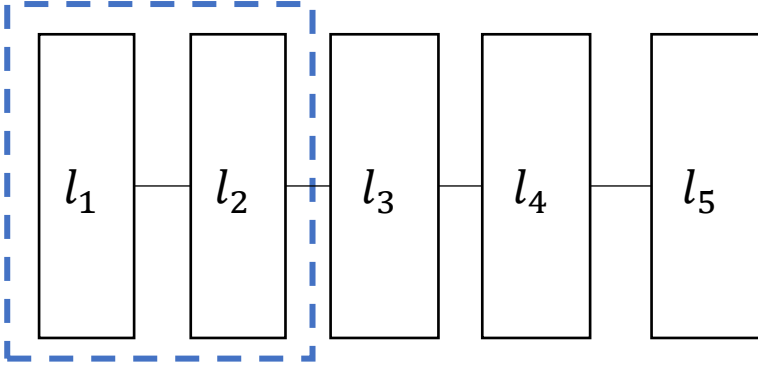


Securing Backpropagation using TEEs

Securing first layer $l = 1$ Gradients



Static GradSec and Dynamic GradSec

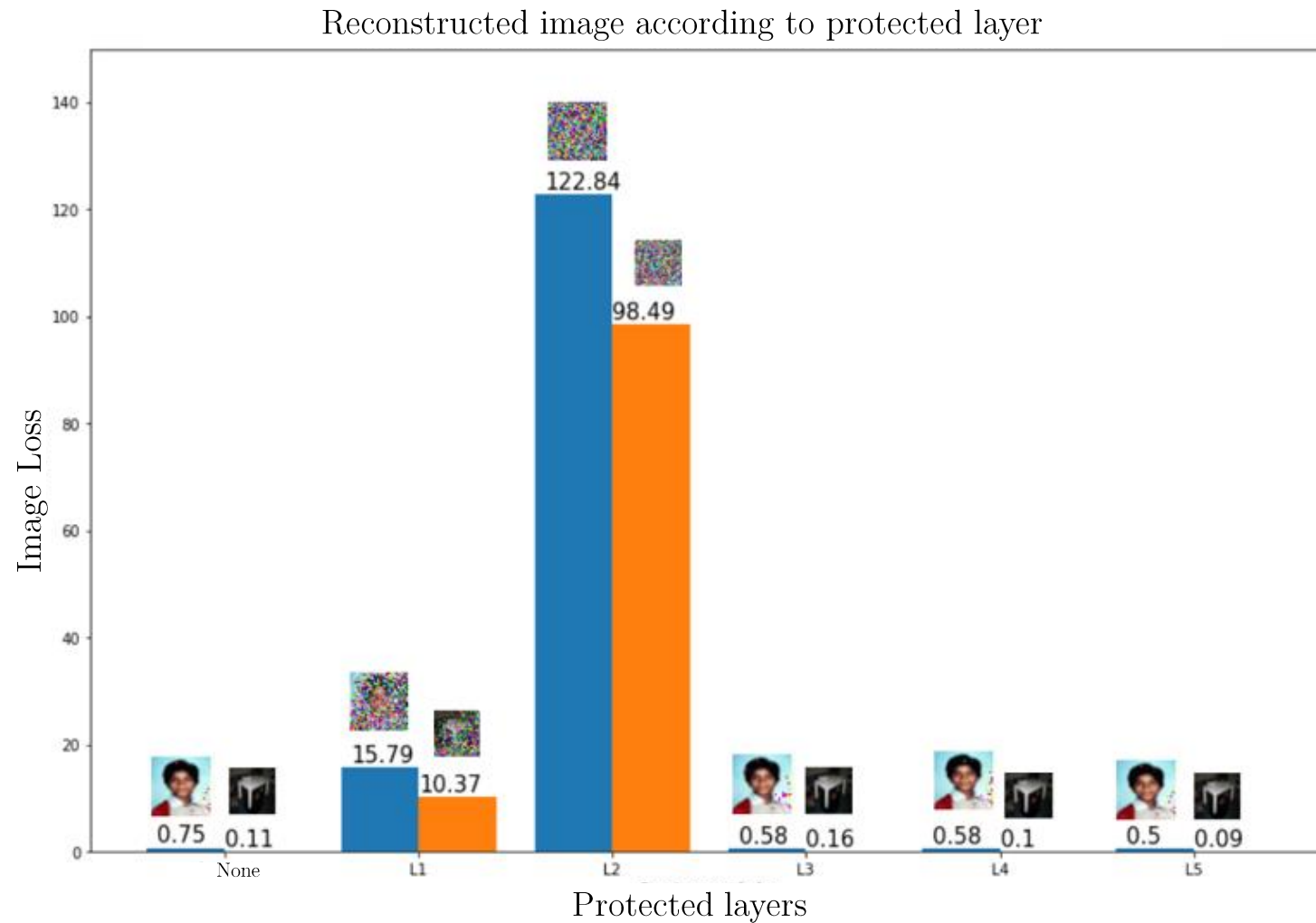
	Static GradSec	Dynamic GradSec
Principle	Protecting the same layers during all FL cycles	Changing the protected layers as the FL cycles through a moving window (MW)
Parameters	<p><i>protected_layers</i> : List of protected layers</p>	<p>➤ $size_{MW}$: Number of protected layers during each FL cycle</p> <p>➤ V_{MW} : Vector of distribution of probability protection</p>
Overview	<div> <div> <div></div> TEE Enclave </div> <div> Protected_layers = {l_2, l_4} </div>  </div>	<div> <div> <div></div> TEE Enclave </div> <div> $V_{MW} = [0.3 \ 0.2 \ 0.1 \ 0.4]$ </div> <div> Size_{MW} = 2 </div>  </div>

Evaluation : Experimental setup



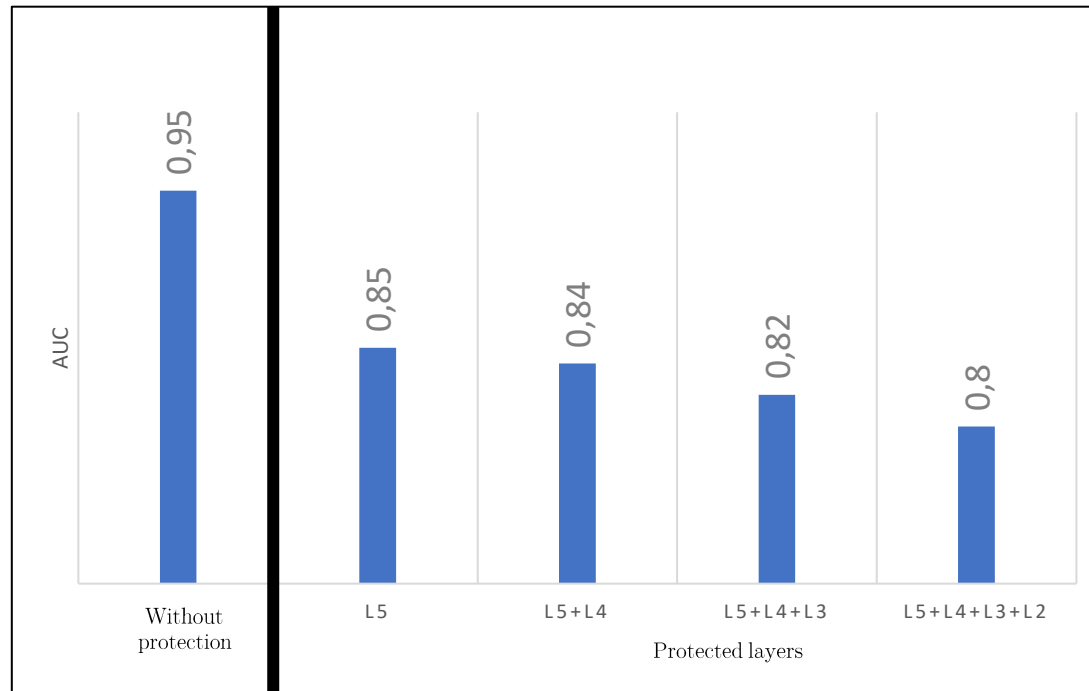
	Metric	Dataset	Model to attack	Protection method
DRIA	Image Loss	CIFAR-100	LeNet-V1 (4 conv2D+ 1 Dense)	Static
MIA				
DPIA	AUC	LFW	LeNet-V2 (3 Conv2D + 2 Dense)	Dynamic

Evaluation : GradSec against DRIA



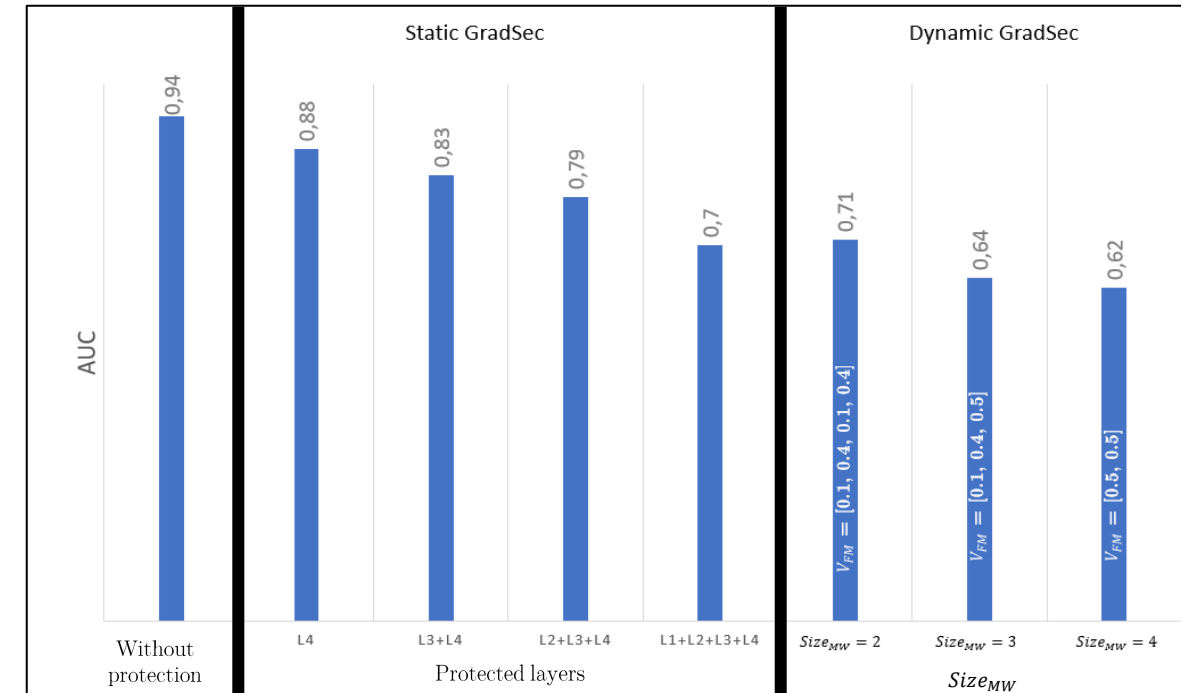
- Conclusion :
- First layers are the most sensitive → layers which contains the most important informations for model inversion [12]
 - We should protect the L2 layer

Evaluation : GradSec against MIA and DPIA



Static GradSec against MIA

- The last layer is the most sensitive → layers that contain latent informations necessary to get membership informations [12]
- Limited interest to protect many layers
- We should protect only L5 layer



Static GradSec and Dynamic GradSec against DPIA

- **Protection offered by securing statically 4 layers is equivalent to the protection offered by securing dynamically 2 layers.**
- **Dynamic GradSec is more efficient than Static GradSec against DPIA**

Evaluation : Comparison with DarkneTZ



		GradSec	DarkneTZ
Protection granularity		Per layer	
Cost of Individual protection against attacks	DRIA	Protecting early layers (2 nd)	
	MIA	Protecting last layers (5 th)	
	DPIA	Dynamic GradSec ($size_{MW} = 2$)	Protecting 4 layers permanently
Cost for grouped protection		DRIA and MIA (2 nd and 5 th)	DRIA and MIA (2 nd , 3 rd , 4 th and 5 th layer)

16% more efficient
against DPIA

8% more efficient in
grouped protection

Thanks for your attention

Aghiles AIT MESSAOUD
Email: ga_aitmessaoud@esi.dz