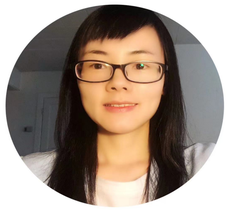
TP-NET: Training Privacy-Preserving Deep Neural Networks under Side-Channel Power Attacks

Summer 2022

Team Members



Hui Hu

Research area:
Machine Learning
Data Mining
Privacy
Security

Jessa Gegax

Research area:
Machine Learning
Privacy
Security

Clay Carper

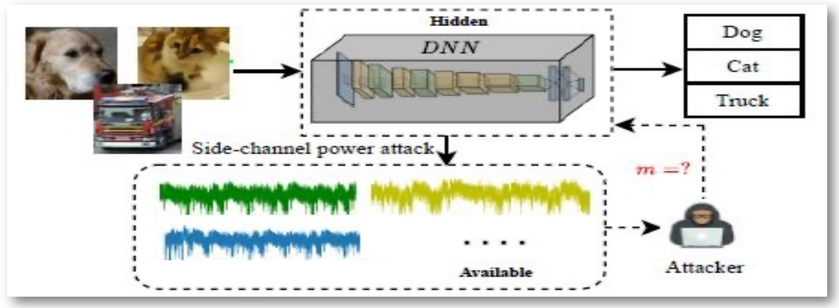
Research area:
Machine Learning
Hardware Security

Introduction

Deep learning has been widely used in various fields such as medical systems, recommendation systems, and computer vision[1]. While it achieves remarkable performance, privacy in deep learning is becoming increasingly prominent with the emergence of numerous attacks. In particular, recent studies have shown that existing deep neural networks (DNNs) are extremely vulnerable to side-channel attacks[2]. For example, the internal structure of a DNN is easily inferred via side-channel power attacks. Further, the leakage of model internal information may lead to users' extremely sensitive predictions being leaked, such as whether or not a user is an HIV carrier. Therefore, it is critically important to protect the model's internal information for avoiding users' privacy leakage under side-channel power attacks. However, to date, few efficient solutions have been proposed for training privacy-preserving DNNs under powerful side-channel power attacks.

Problem Statement

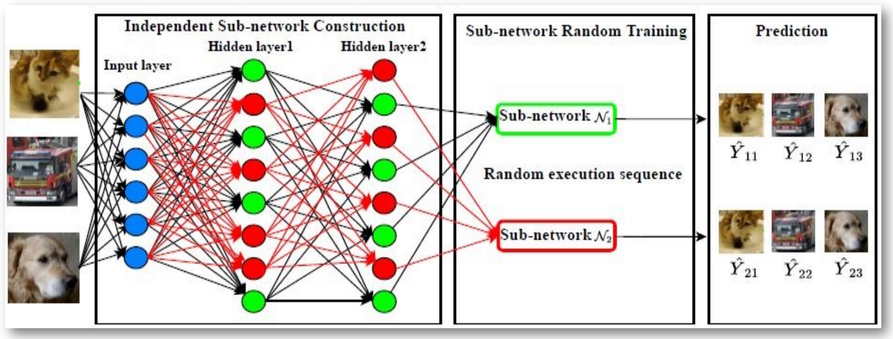
The number of hidden nodes in a DNN is essential internal structural information for inferring model parameter set. However, this information is easily inferred under side-channel power attacks. Therefore, in this work, we focus on preventing a side-channel power attacker from inferring the number of hidden nodes (m), as the figure shows.



Methods

We propose a novel solution for training privacy-preserving DNNs under side-channel power attacks, called **TP-NET**. It includes three steps:

- *Independent Sub-network Construction*, which generates multiple independent sub-networks via randomly selecting nodes in each hidden layer.
- *Sub-network Random Training*, which randomly trains multiple sub-networks such that power traces keep random in the temporal domain.
- *Prediction*, which outputs the predictions made by the most accurate sub-network.



Experiment Results

Table 1 and Table 2 present the classification accuracy and privacy-preserving performance of **TP-NET**.

Table 1. Node classification accuracy

# of hidden nodes	Models	DIABETES	COMPAS
8 nodes	Traditional DNN	0.6257	0.7644
	TP-NET	**0.6268**$^{\uparrow 0.17\%}$	**0.6938**$^{\downarrow 9.24\%}$
10 nodes	Traditional DNN	0.6415	0.7852
	TP-NET	**0.6461**$^{\uparrow 0.72\%}$	**0.7712**$^{\downarrow 1.78\%}$
14 nodes	Traditional DNN	0.6233	0.7720
	TP-NET	**0.6172**$^{\downarrow 0.98\%}$	**0.7616**$^{\downarrow 1.35\%}$

Table 2. Inference accuracy on the Diabetes dataset by using k-NN

Models	Structures	k=3	k=6
Traditional DNN	(8,8)(14,14)	**1.0000**	**1.0000**
TP-NET	Structure 6	0.6193$^{\downarrow 38.07\%}$	0.6163$^{\downarrow 38.37\%}$
	Structure 7	**0.5970**$^{\downarrow 40.30\%}$	**0.6007**$^{\downarrow 39.93\%}$

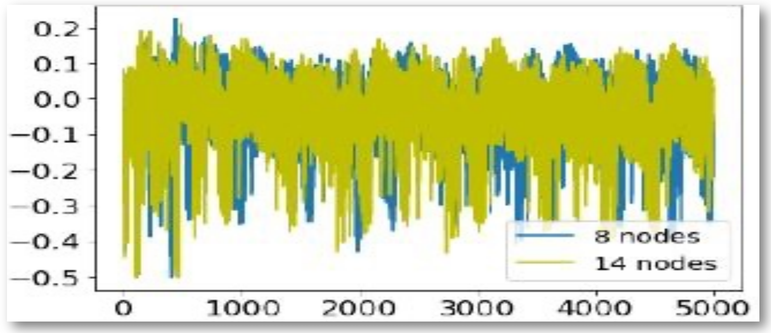


Figure 1. Power traces of neural networks with 8 nodes and 14 nodes in each hidden layer.

(a) As Table 1 shows, the classification accuracy of **TP-NET** is competitive compared with the traditional DNN. Take 14 nodes as an example, **TP-NET** has slightly lower classification accuracy on two datasets, which only decreases the classification accuracy by 0.98% on Diabetes dataset and 1.35% on COMPAS dataset respectively.

(b) As Table 2 shows, **TP-NET** decreases inference accuracy on the number of hidden nodes significantly compared with the traditional DNN.

Advisor: Dr. Mike Borowczak

Group Members:

- Hui Hu (hhu1@uwyo.edu)
- Jessa Gegax (jessagegaxrandazzo@gmail.com)
- Clay Carper (ccarper2 @uwyo.edu)

References:

[1] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," Annual review of biomedical engineering, vol. 19, pp. 221–248, 2017.

[2] S. Wolf, H. Hu, R. Cooley, and M. Borowczak, "Stealing machine learning parameters via side channel power attacks," in 2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI). IEEE, 2021, pp. 242–247.