





Enclave-NN

Alexander Schlögl

Introduction

Neural Nets are universal approximators.

Technical Details

- Forward pass is a series of algebraic operations
- Fully defined by their architecture and weights (parameters)

We focus solely on inference phase, so NNs for us are static functions.

Problem Statement

Monetization requires keeping parameters private.

Problem Statement

Monetization requires keeping parameters private.

Current Method

Online Oracles

- parameters never public
- require sharing of data for inference
- require provider's infrastructure

Problem Statement

Monetization requires keeping parameters private.

Current Method

Online Oracles

- parameters never public
- require sharing of data for inference
- require provider's infrastructure

Idea

Use Trusted Execution Environments (TEEs) to hide parameters during inference

block1_conv1	
block1_conv2	
block1_pool	
block2_conv1	
block2_conv2	
block2_pool	
block3_conv1	
block3_conv2	
block3_conv3	
block3_pool	
block4_conv1	
block4_conv2	
block4_conv3	
block4_pool	
block5_conv1	
block5_conv2	
block5_conv3	
block5_pool	
gap2d	
dense	
dropout	
dense_1	
dropout_1	
dense_2	
	ı

block2_conv1 block3_conv1 block3_conv2 block3_conv3 block4_conv1 block4_conv2 block4_conv3 block5_conv2 block5_conv3 block1_conv1 block1_conv2 block2_conv2 block5_conv1 block3_pool block1_pool block2_pool block4_pool block5_pool dropout_1 dropout dense_1 dense_2 gap2d dense



block1_conv1

block1_conv2

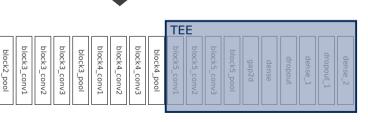
block2

conv2

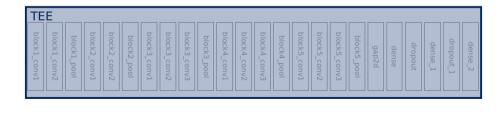
block2_conv1

block1_pool

block1_conv1 block2_conv1 block3_conv1 block3_conv2 block3_conv3 block4_conv1 block4_conv2 block4_conv3 block5_conv2 block5_conv3 block1_conv2 block2_conv2 block5_conv1 block3_pool block1_pool block2_pool block4_pool block5_pool dropout_1 dropout dense_1 dense_2 gap2d dense



block2_conv1 block3_conv1 block3_conv2 block3_conv3 block4_conv1 block4_conv2 block4_conv3 block5_conv2 block5_conv3 block1_conv1 block1_conv2 block2_conv2 block5_conv1 block3_pool block1_pool block2_pool block4_pool block5_pool dropout_1 dropout dense_1 dense_2 gap2d dense



Move last *n* layers into TEE, send protected model to user

Advantages

Less infrastructure required Semi offline usage possible Inference data can stay private

Disadvantages

Performance impact Requires trust in manufacturer Potentially larger attack surface

Evaluation Method

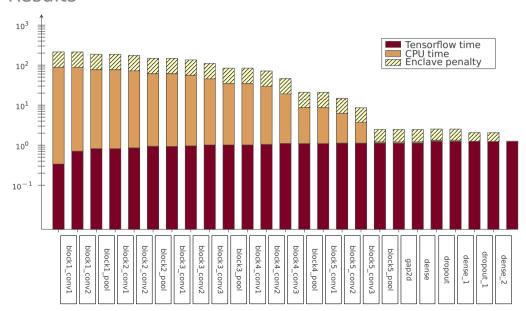
How large is the performance impact?

Procedure

- Split NN
- 2 Compile TEE and native code
- Measure inference time on single input
- Separate CPU impact from TEE impact

Repeat for every possible split in NN

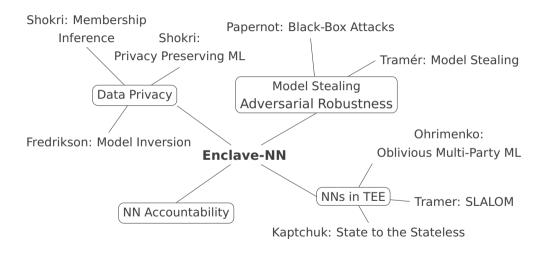
Results



Open Problem: Where to stop

- Generate weights
- Generate code for forward pass
- ✓ Interoparability with TF
- ✔ Run forward pass in TEE
- ✓ Hide weights
- ? Hide architecture
- ? Monetization prototype
- ? Different TEE technologies
- Automate for more code (TEE compiler target)

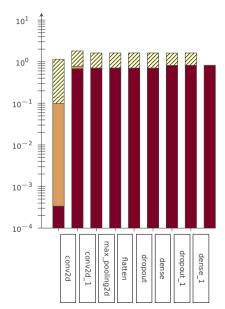
Open Problem: Where to send it



Open Problems

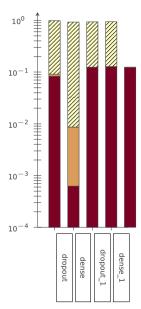
- How to split total work in submissions
- Which community to submit work to

MNIST Results





IMDB Results





Rotten Tomatoes Results

