ICRI-CARS
Secure Systems Group, Aalto University

Max Reuter, Samuel Marchal, Andrew Paverd, N. Asokan

Privacy Preserving Deep Neural Network Prediction using Trusted Hardware

Machine Learning relies on Sensitive Data

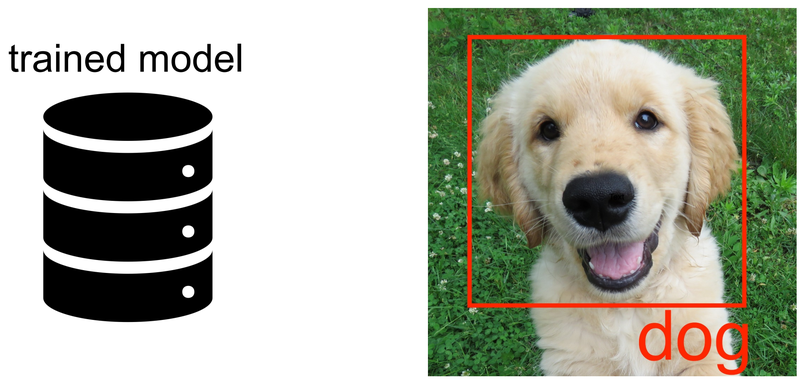
Training machine learning models requires know-how, private data, computational power, etc.

- Service providers want to protect their business advantage (+ intellectual property)

Typical solutions deploy the models to the cloud and allow users to query them

- Users want to protect the privacy of requests

Sensitive data of each party must be protected without compromising functionality



Problems with existing solutions

Existing solutions for protecting user privacy rely on cryptography and oblivious execution:

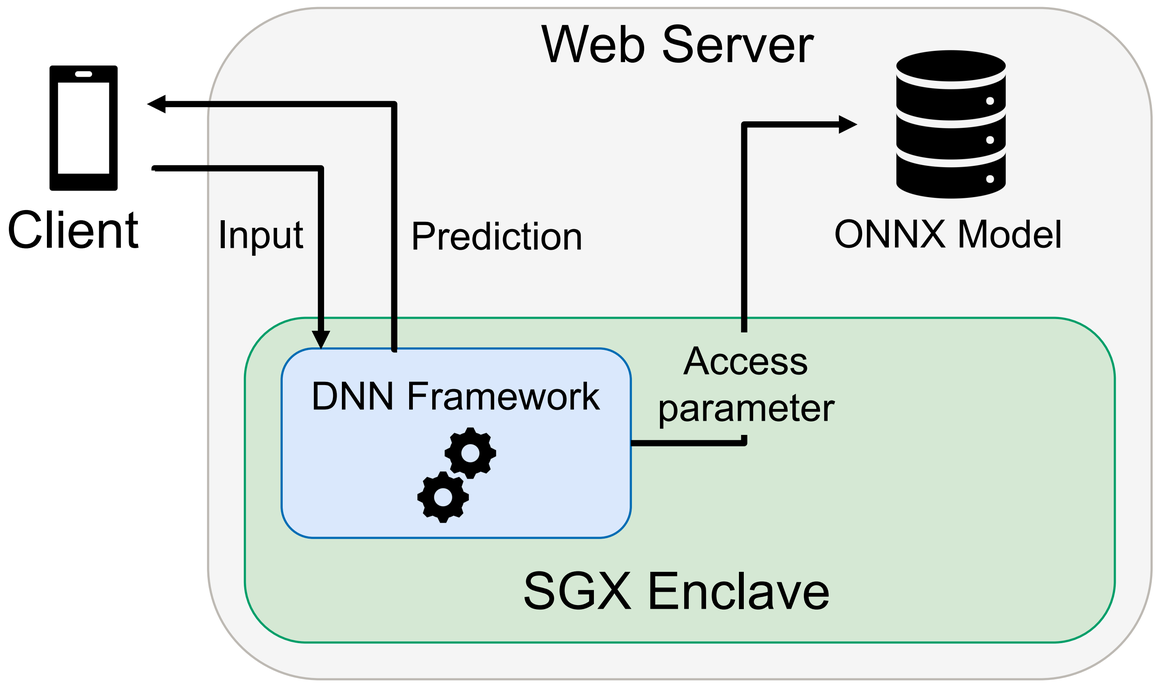
- Introduce large performance overheads
- Only support limited set of operations
- Require changes to existing models
- No input analysis possible

Solution

- Compute prediction on clear data inside trusted execution environment (TEE)
- Prove confidentiality to client using attestation
- Analyse input without compromising user privacy

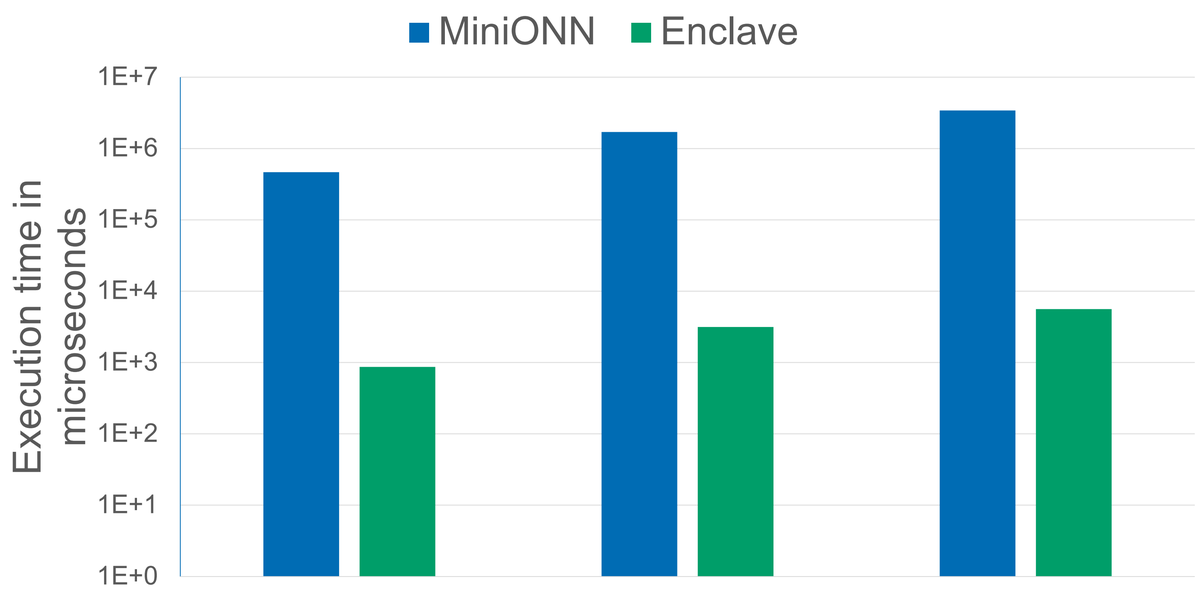
Implementation Details

- Use Intel SGX as TEE
- Adopt ONNX standard to support wide range of models
- Store model weights outside the enclave to address memory limitations



Initial Evaluation

- For general matrix multiplication (GEMM), more than 500 times faster than MiniONN[1]



[1] J. Liu, M. Juuti, Y. Lu, and N. Asokan, “Oblivious Neural Network Predictions via MiniONN Transformations,” Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 619–631, 2017.

Aalto University
Research Institute for Collaborative Autonomous and Resilient Systems