Secure Systems Group, Aalto University†, University of Waterloo+

Vasisht Duddu†, Oskari Järvinen†, Lachlan J. Gunn†, N. Asokan††

ML Property Attestation using TEEs

• Clients of ML-based services cannot verify that responses come from the right model
• Algorithms, datasets, and training parameters cannot be verified after training
• ML property attestation can prove such properties efficiently and scalably

1 Introduction
- Measured model and dataset metrics used to demonstrate the quality of models & inferences
- Need to link dataset, training parameters to model, model to inference input/output
- New advances (e.g., Intel AMX) allow training/running complex models within TEEs

2 The problem
- Cryptographic proofs inefficient or don’t scale
- ML-based methods are inaccurate
- Current methods focus only on specific properties
- Current certification services require outsourcing both training and inference

3 Our solution
Use remote attestation to prove properties like:
• Which model produced an inference
• How accurate is the model
• How was the model trained
• What data was used to train it
• How representative was the training set

4 Implementation
- SGX enclaves perform ML tasks and attest process/performance claims
- Verifier combines attestations to link output to input, model, training dataset

5 Conclusion
TEE-based ML property attestation is efficient, scalable & versatile