Proposal: Light-weighted Neural Networks Encryption Accelerator

Group name: ANN_encryption_byLLE

Group members: Enze Chen (ec3576), Liqin Zhang (lz2809), Lanxiang Hu (lh3116)

Feb 24, 2022

Abstract:

In this work, we presented the Integer Vector Homomorphic Encryption scheme on embedded devices as a prototype for encrypting neural networks. We will use the FPGA as the hosting computational core that reads arbitrary encrypted functions from the client with integer vectors stored in memory as inputs, carries out vectorized calculations in the custom accelerator over the encrypted domain, and returns encrypted computational results to the client.

Background:

In the past few decades, deep learning (DL) and DL models have revolutionized many classification tasks. However, many existing DL models are computationally expensive and memory consuming. To deploy those models on smartphones and embedded devices, light-weight architecture designs are studied and put in use for better speed-accuracy trade-off.

While we are in the dawn of the DL explosion for smartphones and embedded devices, many DL-based models are not well-protected. For example, A research in 2019 showed that out of 218 DL-based Android apps, only less than 20% of them use encryption, another 20% of them use obfuscation as the protection mechanism and the rest were left utterly unprotected [1]. One major reason for this lack of security is that neural networks as ciphertext need to be malleable so that once an encrypted trained model is sent to the cloud where big data is stored, the model can be used effectively.

To resolve this problem, a Homomorphic Encryption (HE) scheme is introduced and employed in this project for encrypting plaintext, so that operations on the ciphertext translate into operations on the plaintext. Specifically, Integer Vector Homomorphic Encryption is adopted to support and accelerate vectorized computations ubiquitous to neural networks [2].

Design:

Hardware:

1. Hardware serves as the cloud host and accelerator in this project, it's responsible for carrying out addition, linear transformation, weighted inner product for integer vectors that comply with the

encryption scheme. Based on these three operations, arbitrary polynomials on integer vectors can be computed efficiently and this functionality needs to be supported as well.

2. Memory modules are needed to load and store encrypted data, weight matrices and key-switching matrices.

Software:

1. Software serves as the client in this project, and it's responsible for homomorphic encryption and decryption processes and key switching operation that generates public keys (including data pre-encryption and data upload).
2. Data as inputs will be pre-configured, encrypted and transferred to the hardware [3].
3. Software will generate inputs as neural networks for HE algorithm to encrypt. All computations can be decomposed into a combination of vector additions, linear transformations and weighted inner products. For nonlinear operations in the model, such as the sigmoid activation function, we rely on approximation to make the model agree with our HE computation [4].

Milestones:

1. Hardware implementation of key modules in the computational core to support addition, linear transformation and weighted inner product over the encrypted domain.
2. Software implementation of encryption, description and key-switching.
3. Hardware-software interface implementation (to preload encrypted data in static time, and to load weight matrices and key-switching matrices in runtime).

References:

1. Mengwei Xu, Jiawei Liu, Yuanqiang Liu, Felix Xiaozhu Lin, Yunxin Liu, and Xuanzhe Liu. 2019. A First Look at Deep Learning Apps on Smartphones. In The World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 2125–2136.
2. Zhou, Hongchao and Gregory W. Wornell. "Efficient homomorphic encryption on integer vectors and its applications." 2014 Information Theory and Applications Workshop (ITA) (2014): 1-9.
3. Yu, A. et al. "Efficient Integer Vector Homomorphic Encryption." (2015).
4. Yang, Zhaoxiong et al. "FPGA-Based Hardware Accelerator of Homomorphic Encryption for Efficient Federated Learning." ArXiv abs/2007.10560 (2020): n. pag.