Table 1: MNIST inference performance comparisons. The network topologies are not identical across previous work, resulting in variations in accuracy.

Framework	Limitation	Accuracy (%)	Latency (s)	Throughput (images/s)
CryptoNets [7]	polynomial activation	98.95	250.00	16.4
Chameleon [17]	3-party	99.00	2.24	1.0
CryptoDL [10]	polynomial activation	99.52	320.00	45.5
GAZELLE [11]	hand-optimized	98.95	0.03	33.3
SecureNN [20]	3-party	99.00	0.08	49.2
XONN [16]	binarized network	98.64	0.16	6.2
nGraph-HE2 [3]	reveals intermediate values	98.62	0.69	2,959.0
CrypTFlow [12]	leaks model architecture	99.31	0.03	33.3
MP2ML (This work)	—	98.60	6.79	33.3

```python
import pyhe_client
from mnist_util import *

# Parse command-line arguments
FLAGS, _ = client_argument_parser().parse_known_args()
# Load data
(x_test, y_test) = load_mnist_test_data(
        FLAGS.start_batch, FLAGS.batch_size)
# Perform inference
client = pyhe_client.HESealClient(
    FLAGS.hostname, FLAGS.port, FLAGS.batch_size,
    { FLAGS.tensor_name:
        (FLAGS.encrypt_data_str, x_test.flatten('C')) })
results = client.get_results()
```

Listing 2: Python3 source code for a client to execute a pre-trained TensorFlow model in MP2ML.

Forschungsgemeinschaft (DFG) — SFB 1119 CROSSING/236615297 and GRK 2050 Privacy & Trust/251805230, and by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within ATHENE.

AVAILABILITY

The open source code of MP2ML is freely available under the permissive Apache license at https://ngra.ph/he.

REFERENCES

[1] Nitin Agrawal, Ali Shahin Shamsabadi, Matt J Kusner, and Adrià Gascón. 2019. QUOTIENT: Two-Party Secure Neural Network Training and Prediction. In *CCS'19*.

[2] Fabian Boemer, Rosario Cammarota, Daniel Demmler, Thomas Schneider, and Hossein Yalame. 2020. MP2ML: A Mixed-Protocol Machine Learning Framework for Private Inference. In *ARES'20*.

[3] Fabian Boemer, Anamaria Costache, Rosario Cammarota, and Casimir Wierzynski. 2019. nGraph-HE2: A High-Throughput Framework for Neural Network Inference on Encrypted Data. In *WAHC'19*.

[4] Fabian Boemer, Yixing Lao, Rosario Cammarota, and Casimir Wierzynski. 2019. nGraph-HE: a graph compiler for deep learning on homomorphically encrypted data. In *ACM International Conference on Computing Frontiers*.

[5] Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. 2017. Homomorphic Encryption for Arithmetic of Approximate Numbers. In *ASIACRYPT'17*.

[6] Daniel Demmler, Thomas Schneider, and Michael Zohner. 2015. ABY - A Framework for Efficient Mixed-Protocol Secure Two-Party Computation. In *NDSS'15*.

[7] Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin Lauter, Michael Naehrig, and John Wernsing. 2016. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *ICML'16*.

[8] Oded Goldreich, Silvio Micali, and Avi Wigderson. 1987. How to play any mental game. In *STOC'87*.

[9] Wilko Henecka, Stefan Kögl, Ahmad-Reza Sadeghi, Thomas Schneider, and Immo Wehrenberg. 2010. TASTY: Tool for Automating Secure Two-party Computations. In *CCS'10*.

[10] Ehsan Hesamifard, Hassan Takabi, Mehdi Ghasemi, and Rebecca N. Wright. 2018. Privacy-preserving Machine Learning as a Service. *PETS'18*.

[11] Chiraag Juvekar, Vinod Vaikuntanathan, and Anantha Chandrakasan. 2018. *GAZELLE*: A Low Latency Framework for Secure Neural Network Inference. In *USENIX Security'18*.

[12] Nishant Kumar, Mayank Rathee, Nishanth Chandran, Divya Gupta, Aseem Rastogi, and Rahul Sharma. 2020. CrypTFlow: Secure TensorFlow Inference. In *S&P'20*.

[13] Jian Liu, Mika Juuti, Yao Lu, and Nadarajah Asokan. 2017. Oblivious neural network predictions via MiniONN transformations. In *CCS'17*.

[14] Pratyush Mishra, Ryan Lehmkuhl, Akshayaram Srinivasan, Wenting Zheng, and Raluca Ada Popa. 2020. DELPHI: A Cryptographic Inference Service for Neural Networks. In *USENIX Security*.

[15] Payman Mohassel and Yupeng Zhang. 2017. SecureML: A system for scalable privacy-preserving machine learning. In *S&P'17*.

[16] M Sadegh Riazi, Mohammad Samragh, Hao Chen, Kim Laine, Kristin E Lauter, and Farinaz Koushanfar. 2019. XONN: XNOR-based Oblivious Deep Neural Network Inference. In *USENIX Security'19*.

[17] M Sadegh Riazi, Christian Weinert, Oleksandr Tkachenko, Ebrahim M Songhori, Thomas Schneider, and Farinaz Koushanfar. 2018. Chameleon: A hybrid secure computation framework for machine learning applications. In *ASIACCS'18*.

[18] Ronald L. Rivest, Len Adleman, and Michael L. Dertouzos. 1978. On Data Banks and Privacy Homomorphisms. *Foundations of Secure Computation, Academia Press*.

[19] SEAL 2019. Microsoft SEAL (release 3.4). https://github.com/Microsoft/SEAL. Microsoft Research, Redmond, WA.

[20] Sameer Wagh, Divya Gupta, and Nishanth Chandran. 2019. SecureNN: 3-Party Secure Computation for Neural Network Training. *PETS'19*.