Private and Reliable Neural Network Inference

Nikola Jovanović Marc Fischer Samuel Steffen Martin Vechev

ETH Zurich, Switzerland

{nikola.jovanovic, marc.fischer, samuel.steffen, martin.vechev}@inf.ethz.ch
Motivation: ML as a Service

privacy-preserving inference

inference with reliability guarantees

FHE
This Work

- Randomized Smoothing
- Argmax Approximation
- FHE Optimizations

private + reliable inference

FHE

Phoenix

Fairness

Robustness
Guarantees of Phoenix

- Client Data Privacy
- Reliability Guarantees

- FHE
- private+reliable inference

- Phoenix
- fairness
- robustness
Background: Randomized Smoothing

- **Logits:** $[0.1 \ 2.4 \ 0.3] \ldots$
- **Predictions:** $[0 \ 1 \ 0] \ldots$
- **Counts:** $[13 \ 81 \ 6]$
- **Prediction (“Cat”)**

**Argmax**

$\sum y_n$

+ **Probabilistic Reliability Guarantee** (“The prediction is robust.”)
Overview of Phoenix

Prior Work (Batched Inference)

Key Challenge!

Soft Counting Heuristic

Rotate + Add + Optimizations

Rewrite of BinPValue

ARGMAX

\[ x_1 \to x_2 \to \ldots \to x_n \]

\[ x_1 \oplus_p \text{noise} \]

\[ z_1 \to z_2 \to \ldots \to z_n \]

\[ y_1 \to y_2 \to \ldots \to y_n \]

\[ \sum y_i \to \text{cnt} \]

1 Ciphertext
Argmax Approximation

Cheon et al. (ASIACRYPT ’20)

Full algorithm in our paper

Algorithm 2 Approximation of Argmax for BNS-CKKS

1. function ArgmaxHE
2. Input: $z = [z_0, \ldots, z_{n-1}]$, $d_0^{(1)}, d_0^{(2)}, d_0^{(3)}$
3. Output: $y = [y_0, \ldots, y_{n-1}]$ as in Eq. (3)

4. $z' = z \cdot d_0^{(1)}(z, i)$

5. $a_0 = 0$

6. for $i = 0$ to $n-1$
7.     $a_i = a_{i-1} + d_0^{(2)}(z, i)$
8.     $y_i = \text{score}(d_0^{(3)}(z, i), a_i)$
9.     $y_i = \text{score}(d_0^{(3)}(z, i), a_i)$
10. end for

11. return $y$

Bound P(violation)

Conditions + Rescale

$[a, 1] \rightarrow [1-2^b, 1]$
Implementation & Evaluation

Available on GitHub: eth-sri/phoenix

Consistency

Efficiency

The results are equivalent to the ones obtained in non-private evaluation

Viable latencies and communication costs
Summary: **Phoenix**

Client Data Privacy

Reliability Guarantees