Privacy Preserving Neural Network Classification
Beyza Bozdemir¹, Gamze Tillem², Melek Önen¹, Orhan Ermis¹
¹EURECOM, France
²Cyber Security Group, Delft University of Technology, The Netherlands

Privacy goals in PAPAYA

**PAPAYA Objectives**
- **Privacy by Design**
  - PP analytics: Neural Networks (NN), clustering, statistics
- **Different Settings**
  - One data owner vs. multiple data owners
  - One querier vs. multiple queriers

**Use Cases**
- healthcare
  - Arrhythmia Detection
- mobile and phone usage
  - Threat Detection

**Analytics example – Neural Network classification**

### NN Layers and Operations
- **Input Layer**
- **Hidden Layer**
  - Convolutional layer (matrix multiplications)
  - Activation layer (sigmoid, tanh, etc.)
  - Pooling layer
  - Fully connected layer
- **Output Layer** (softmax, etc.)

**Privacy by Design Challenges**
- **Privacy vs. efficiency**
  - Deep NN ⇒ Significant overhead with cryptographic tools
- **Privacy vs. accuracy**
  - Complex operations (sigmoid, softmax, etc.) ⇒ Not suitable to crypto tools

**Privacy by Design**

- **Non-interactive**
- **Only linear operations (eg. AF is approximated to x²)**
- **Expensive in computation cost**
- **No communication cost**

**References**:

**Privacy Preserving Neural Network Classification: A Hybrid Solution**

### Flexible solution: 2 settings

1. **1st Scenario: Client-Server**
   - Paillier for linear operations
   - Optimized computational overhead
   - Less computation time compared to [1]
   - Paillier for x²
   - New interactive protocol to compute x²
   - 2PC for comparison only (ReLU case)
   - Optimized communication overhead
   - Less bandwidth usage compared to [2]
   - Similar level of accuracy as in [1, 2]

2. **2nd Scenario: Two-Server**
   - Paillier for linear operations
   - Optimized computational overhead
   - Less communication overhead
   - Similar level of accuracy as in [1, 2]

**Hybrid Solution**

### Privacy Preserving Neural Network Classification – Existing solutions

**with Homomorphic Encryption [1]**
- Interactive – Client is involved
- Linear operations and comparisons
- Efficient in computation cost
- Expensive in communication cost

**with Secure Two-party Computation [2]**
- Non-interactive
- Only linear operations (eg. AF is approximated to x²)
- Expensive in computation cost
- No communication cost

### Results

<table>
<thead>
<tr>
<th>Technique</th>
<th>Computation Cost (s)</th>
<th>Communication Cost (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE [1]</td>
<td>297</td>
<td>372.2</td>
</tr>
<tr>
<td>2PC [2]</td>
<td>1.2</td>
<td>47.6</td>
</tr>
<tr>
<td>Hybrid Solution</td>
<td>10</td>
<td>1.73</td>
</tr>
</tbody>
</table>

**Technique**
- HE: Homomorphic Encryption
- 2PC: Secure Two-party Computation

**Analytics example – Neural Network classification**

- Computation cost 30-fold better than [1]
- Communication cost 27-fold better than [2]

**References**: