Secure XGBoost: Towards Privacy Preserving Collaborative Gradient Boosted Decision Trees
Chester Leung, Andrew Law, Octavian Sima, with Rishabh Poddar, Wenting Zheng, Raluca Ada Popa
University of California, Berkeley

Introduction

- Multiple parties want to perform joint computation over everyone’s data
- Parties are unable to share their data in plaintext for reasons like privacy regulation or business competition
- Need a solution that enables collaborative computation over sensitive datasets
- Example: hospitals that pool patient data for more accurate diagnoses

Goals

- Enable secure collaboration on top of XGBoost
- Limit information leakage with security guarantees
- Provide acceptable latencies while achieving comparable model accuracies and greater security
- Present a simple interface to allow folks with less security background to easily adopt our system

System Overview

- Two modes: federated and coopetitive
- Tradeoff between performance and security
- Python interface for each mode

```
enclave = xgb.Enclave("xgboost_enclave.signed")
# Remote Attestation
enclave.get_remote_report_with_pubkey()
enclave.verify_remote_report_and_set_pubkey()
dtrain = xgb.DMatrix("train.encrypted")
dtest = xgb.DMatrix("test.encrypted")
booster = xgb.Booster()
booster.set_params(params)
for i in range(n_trees):
    booster.update(dtrain, i)
    (dtrain, "train"),(dtest, "test"), i
    booster.predict(dtest)
```

Federated Mode

- Tracker creates “star” topology of parties
- Requires a trusted aggregator that aggregates intermediate summaries
- Leaks intermediate summaries and intermediate models

Coopetitive Mode

- Leverages hardware enclaves to perform training and prediction on sensitive data
- Intermediate summaries and intermediate models are not visible
- Eliminates access pattern leakage with novel oblivious algorithms
- Only predictions are leaked

Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Training Time (s)</th>
<th>Accuracy (AUC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanilla XGBoost</td>
<td>6.40563</td>
<td>0.79018</td>
</tr>
<tr>
<td>Federated XGBoost</td>
<td>6.37262</td>
<td>0.78992</td>
</tr>
</tbody>
</table>

Federated

- 7 nodes, 1M rows

Coopetitive

- Higgs Boson dataset

https://github.com/mc2-project/mc2