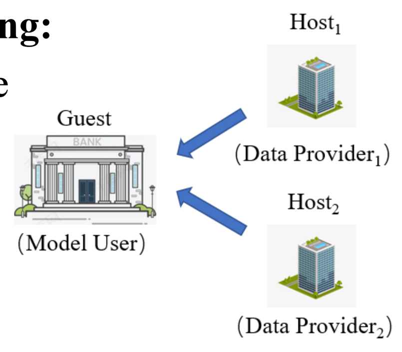
Fed-EINI: An Efficient and Interpretable Inference Framework for Decision Tree Ensembles in Federated Learning
一种高效可解释的联邦集成树推理框架

Xiaolin Chen; Shuai Zhou; Bei Guan; Kai Yang; Yongji Wang
2021 IEEE International Conference on Big Data (Big Data)

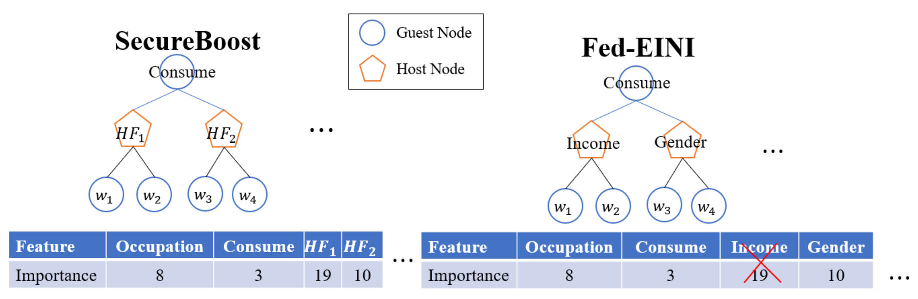
Motivation

☐ **Vertical Federated Learning:**

- Share the same sample space
- Different feature space
- Models used by the Guest
- Data provider by Hosts



☐ **Interpretability**



Reasons for disclosing the meanings of features:

- judge the reasonability of Federated AI models.
- prove the compliance of models to business regulators [1]

☐ **Challenges:** disclosing the meanings of features, data breaches, efficiency

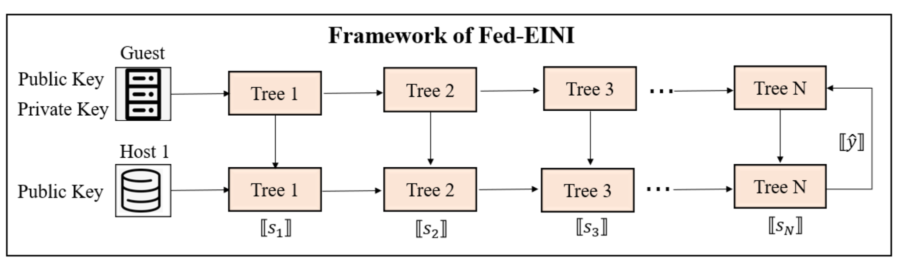
☐ **Solution:** additively homomorphic encryption, confusion items, hidden secure decision paths

Fed-EINI : A Two-Stage Inference Algorithm

Key Observation: The prediction result of a tree can be expressed as the intersection of results of the sub-models held by all parties.

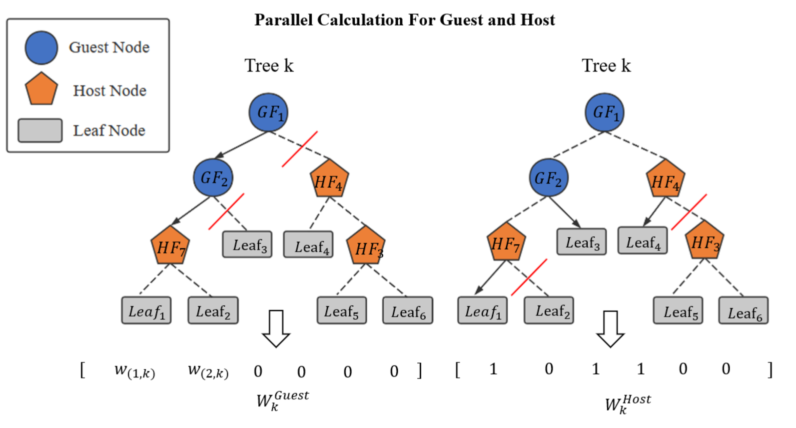
$$f_k(\boldsymbol{x}) = w_{(j,k)},\ where\ j \in \bigcap_{\{m=1\}}^{M} f_k^m(\boldsymbol{x}^m)$$

Fed-EINI: compute the candidate sets of each party $f_k^m(x^m)$ locally, and securely compute the inference results



☐ **Stage1: Parallel Calculation:**

- Each participant generates candidate sets of leaf nodes.



$$\boldsymbol{W}_k^{Host} = \left[s_{(j,k)}: s_{(j,k)} = \begin{cases} 1 & if\ j \in f_k^{Host}(x^{Host}) \\ 0 & otherwise \end{cases}\right]$$

$$\boldsymbol{W}_k^{Guest} = \left[s_{(j,k)}: s_{(j,k)} = \begin{cases} w_{(j,k)} & if\ j \in f_k^{Guest}(x^{Guest}) \\ 0 & otherwise \end{cases}\right]$$

☐ **Stage 2: Synchronization**

- Guest: encrypts and send all decision vectors to Host
- Host: merges decision vectors and takes the sum of them

Algorithm 1 Fed-EINI: an efficient and interpretable inference framework.

Input : $x^{\text{Guest}}, x^{\text{Host}}, \{f_k^{\text{Guest}}\}_{k=1}^K, \{f_k^{\text{Host}}\}_{k=1}^K$
Output : Y
Set $[q] = 0, [S_k] = 0$
for $k = 1, \cdots, K$ **do**
 Stage 1: Parallel Calculation
  Guest&Host:Load parameters of f_k^{Guest} or f_k^{Host};
  Guest&Host:generate W_k^{Guest} or W_k^{Host} for x^{Guest} or x^{Host} during tree search according to equation(8)(9);
 Stage 2: Synchronization
  Guest: Encrypt and push $[\![\mathbf{W}_k^{\text{Guest}}]\!]$ to Host;
  Host: Pull $[\![\mathbf{W}_{k-1}]\!]$ from Guest;
  Host: $[\![S_k]\!] = \langle \mathbf{1}, [\![\mathbf{W}_k^{\text{Guest}}]\!] \circ \mathbf{W}_k^{\text{Host}} \rangle$;
end
Host: Push value $[\![\hat{y}]\!] = \sum_{k=1}^{K} [\![S_k]\!]$ to Guest;
Guest: Decrypt and get the prediction $\hat{y}$;

$$[\![\hat{y}]\!] = \sum_{k=1}^{K} [\![S_k]\!] = \sum_{k=1}^{K} \langle \mathbf{1}, [\![W_k^{Guest}]\!] \circ [\![\boldsymbol{W}_k^{Host}]\!] \rangle$$

Analysis

☐ **Security & Interpretability :**

- disclose the meaning of features while hidden decision path
- achieve the same security as existing framework with semi-honest assumption.

		SecureBoost		Fed-EINI	
		Guest	Host	Guest	Host
Model Information	Model structure	√	√	√	√
	Weights of leaf nodes	√	×	√	×
	Model parameters	√	√	√	√
	Splitting rules	Local nodes	Local nodes	Local nodes	Local Nodes
Data Information	Local data	√	√	√	√
	Number of features	√	√	√	√
	Meaning of features	Local features	Local features	All features	Local features
	Decision path	Complete Decision path	Decision path based on local nodes	Decision path based on local nodes	Decision path based on local nodes

Simulation

☐ **Efficiency:**

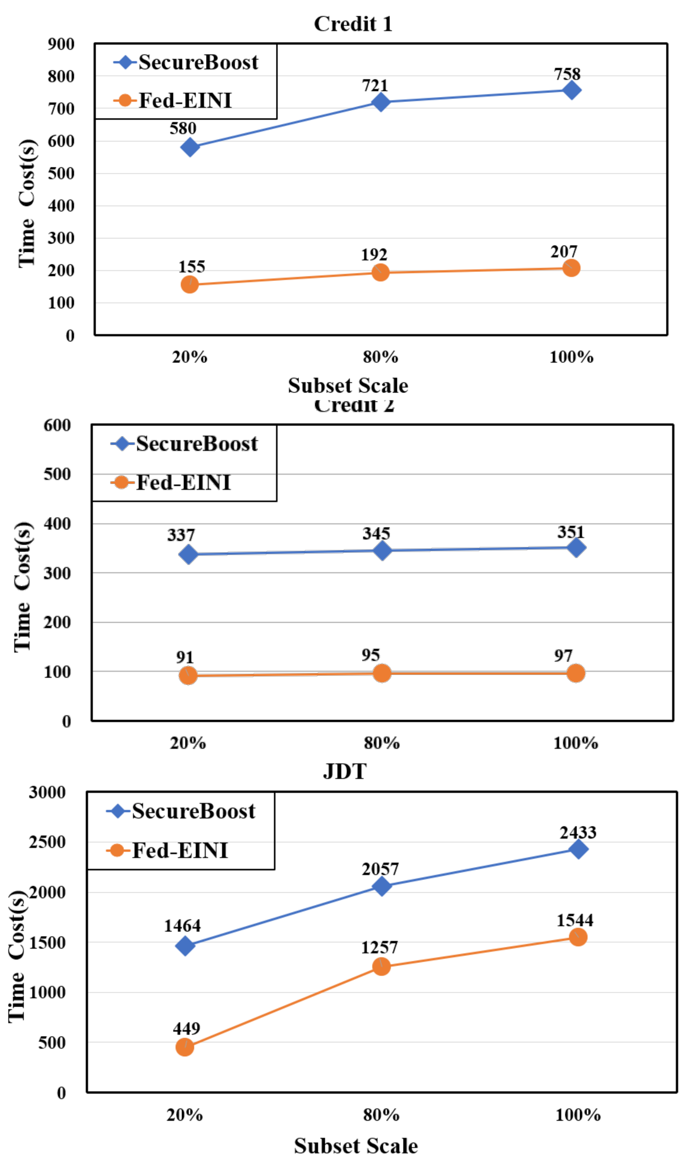
- Parallel Inference: each party generates all the candidates in parallel based on its local splitting condition and local data
- One-Round Communication: The inference of each tree only needs to communicate once at last layer.

☐ **Accuracy Metrics:**

- Conduct numerical experiments with the proposed Fed-EINI and the multi-interactive framework (SecureBoost[2] as representative).

	Sampling Rate	SecureBoost AUC	SecureBoost KS	Fed-EINI AUC	Fed-EINI KS
Credit1	20%	0.880	60.3	0.880	60.3
	80%	0.853	55.0	0.853	55.0
	100%	0.855	54.9	0.855	54.9
Credit2	20%	0.773	43.0	0.773	43.0
	80%	0.771	41.0	0.771	41.0
	100%	0.771	40.9	0.771	40.9
JDT	20%	0.636	19.9	0.636	19.9
	80%	0.639	19.8	0.639	19.8
	100%	0.638	19.7	0.638	19.7

☐ **Efficiency Metrics :**



Conclusions

☐ disclose the meaning of features while hidden decision path
☐ achieve the same security as existing framework with semi-honest assumption
☐ accuracy and efficiency of Fed-EINI

References

- [1]The People’s Bank of China. Evaluation specification of artificial intelligence algorithm in financial application. Financial Industry Standard, Beijing, China, 2021.
- [2] Cheng, K., Fan, T., Jin, Y., Liu, Y., Chen, T., and Yang, Q. SecureBoost: A lossless federated learning framework. arXiv preprint arXiv:1901.08755, 2019.

Authors

- Xiaolin Chen: chenxiaolin18@mails.ucas.ac.cn
- Yongji Wang：ywang@itechs.iscas.ac.cn