HYBRID PRIVACY SCHEME

Authors: Yavor Litchev, Abigail Thomas | Mentor: Yu Xia

Motivation

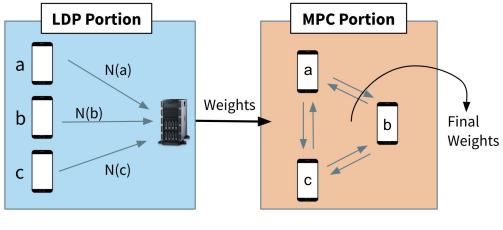
- Our project aims to combine multiple existing algorithms and create a hybrid model to utilize the strengths from various models.

Existing Algorithms

- Two privacy schemes considered: Local Differential Privacy (LDP) and Multi-Party Computation (MPC)
- LDP seeks to protect user privacy by adding random noise to the inputs of the users in order to give people plausible deniability as to what the true values are.
- MPC is a class of algorithms where parties compute a function jointly whilst preserving the privacy of the parties and their inputs to the function.

Privacy Algorithm	Pros	Cons
LDP	fast, high standard of security	accuracy decreases as privacy increases
MPC	high standard of security, maximum accuracy	very slow

Model:

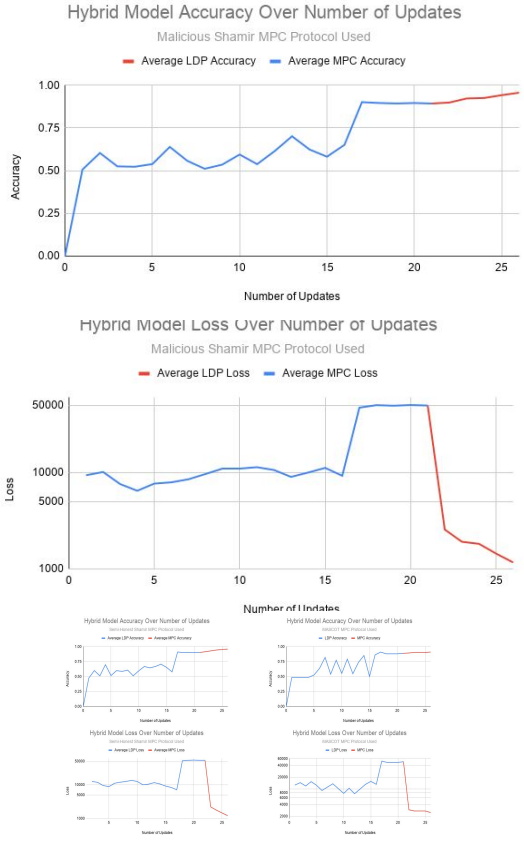


- Train an SVM to stratify the MNIST dataset into 0 or not 0 (modified so zeros are half the dataset)
- Model quickly trains dataset using LDP and then switches to MPC to further fine tune
- Switch occurs when the LDP model is fully trained (predetermined empirically)

Implementation Details:

- There are 4 parties (1-4), with a 5th party (party 0) that is in charge of data distribution
- During the LDP portion, each party adds random noise to their data and sends it all to party 4 (central server) to do the training
- After LDP training, the party 0 (who has the weights) directly submits them to the MPC. Therefore, the weights are not leaked until the end of the program. No other information is leaked and thus the switchpoint is secure
- 4 parties each submits 1 image to the MPC protocol without noise, this is repeated 5 times so that the MPC trains on 20 images.

Results



Hybrid Scheme Testing on SVM, different MPC protocols used

SVM Training	Baseline	Hybrid Model
LDP Portion		
Cost	108,398	49630.0
Accuracy	0.8975	0.8932
Runtime	31.310 secs	23.498 secs
MPC Portion		
Cost	1007.9	1169.1
Accuracy	0.9721	0.9567
Runtime	5.509 hrs	33.226 min

Table of Hybrid Results (With Malicious Shamir MPC Protocol Used) vs Baseline

Conclusion

- Faster than pure MPC
- More accurate than pure LDP
- Security not compromised (proven secure)

Further Research

- Use other ML algorithms (ex. CNN)
- Use other MPC libraries
- Use different datasets (ex. CIFAR)