Poster: Nebula: an Industrial-purpose Privacy-preserving Machine Learning System

Bingzhe Wu
Peking University
wubingzhe@pku.edu.ccn

Cen Chen
Ant Financial Services Group
chencen.cc@antfin.com

Li Wang
Ant Financial Services Group
raymond.wangl@antfin.com

Lei Wang
Ant Financial Services Group
shensi.wl@antfin.com

Jin Tan
Ant Financial Services Group
tanjin.tj@antfin.com

Chaochao Chen
Ant Financial Services Group
chaochao.ccc@antfin.com

Jun Zhou
Ant Financial Services Group
jun.zhoujun@antfin.com

Guangyu Sun
Peking University
gsun@pku.edu.cn

***Abstract*—In this paper, we present a general system, Nebula, providing holistic solutions for secure multi party collaborative learning and inference. In contrast to previous works, our system can support a wide range of machine learning algorithms and secure computation protocols. Specifically, we encapsulate a number of common computation operations and lower-layer protocols into our system. Thus, our system has significantly reduced the engineering efforts to apply secure learning algorithms in the research paper to real-world scenarios. Moreover, according to different properties of various machine learning algorithms, we adaptively chose secure data representations (i.e., shares) to optimize the computation and communication efficiency of the system.**

I. Introduction

Recently, a number of solutions are presented to leverage advanced cryptographic techniques for building efficient secure collaborative training/inference systems for different machine learning algorithms. The major horse used in these works is the secure multiparty computation (MPC) technique. For example, SecureML has developed a privacy-preserving system for jointly modeling logistic regression and neural networks(NN) in the two-party setting [4]. ABY3 is presented as a general framework containing key building blocks (e.g, different secret sharing schemes) of secure ML model training and inference [3].

Despite these progresses, deploying previous solutions into industrial scenarios still remains a big challenge. On the one hand, implementing these solutions requires cryptographic domain expertise to ensure the security and efficiency of the protocol. On the other hand, it also requires huge engineering efforts to correctly and efficiently implement these solutions in a real industrial setting. Making this more precise, a key ingredient of a collaborative learning system is *secret* data communication (i.e., transmitting secret shares) between different organizations. In practice, this is supported by modifying existing communication protocols accompanying huge engineering-labors. Moreover, the operation heterogeneity of ML models can also be problematic since it comes with frequent switches between different ciphertext domains. Although, this problem has been partly resolved by the previous work named ABY^3. A developer with cryptographic expertise is still needed to make these switches explicitly. In summary, we propose that an ideal collaborative learning system should offer a series of ML operations which are transparent to ML developers, i.e., one can develop their own algorithms without caring about the behind cryptographic and engineering details.

To this end, we introduce our system, Nebula, comprising multiple abstraction layers as shown in Figure 1. At a high level, our system can be seen as a programming framework that facilitates ML developers to develop and deployment their ML models under the data-isolated setting, i.e., building a model using data from different organizations while protecting data privacy. The behind design philosophy is to decouple the high-level APIs of various ML operations and the ML unrelated parts, i.e., different secure computation protocols (e.g., SPDZ protocol) and lower-level infrastructures (e.g., modified MPI library).

With the help of Nebula, ML developers in Ant Financial can build ML models in the data-isolated setting with negligible engineering efforts. Despite the usability, our parameter server based distributed system together with the novel sharing scheme and other optimization strategies can drastically improve the efficiency of the collaborative modeling process of a number of common ML models.

II. System Design

Nebula is a system for secure collaborative learning among different organizations. As most of scenarios in Ant Financial involve only two training parties [1], [2], we focus on a two-server architecture.

A. System Overview

The overall architecture is presented in Figure 1. It contains a number of building blocks to support a wide variety of businesses in Ant Financial. The front-end Data Lab interacts with the backend whenever a party sends a service request. The core of the system is the MPC engine comprising different abstract layers that cover various low-level protocols and high-level operations. Among the MPC engine, we can further build different secure ML models such as logistic regression and

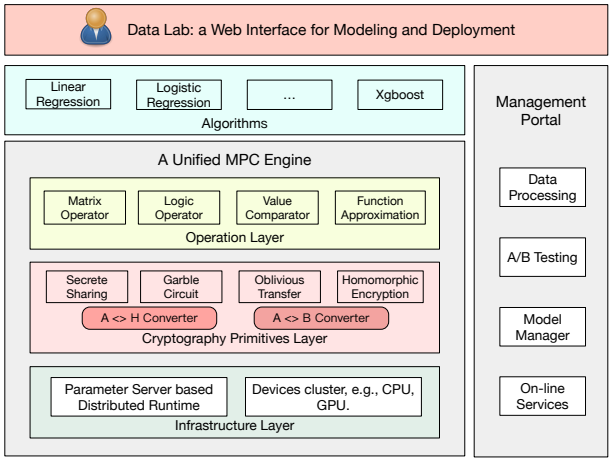

Fig. 1. Overview of Nebula

XGBoost. For further business deployment, the management portal supports a variety of management tasks.

B. MPC Engine

The heart of our system is the MPC engine which consists of three layers, namely, the operation layer, the cryptography primitives layer, and the infrastructure layer. Here we give some design details of each layer:

- **Infrastructure layer** contains the underlying blocks, which are used for building data communication channel and enabling distributing computations on secret shares. Specifically, we present a modify version of MPI library for supporting different secret data types in our system (e.g, additive shares).
- **Cryptography primitives layer** mainly contains three parts, underlying secure protocols (e.g, oblivious transfer, homomorphic encryption), data structures for representing different share types (e.g., additive sharing) and the conversion tools for switches between different sharing types (e.g, additive sharing to Yao's sharing). We also introduce a novel sharing scheme named HE share, which can drastically reduce the communication costs compared with the share schemes used in previous system (e.g, ABY). In practice, this can help us to build more efficient ML models securely and collaboratively.
- **Operation layer** is built upon the lower-level layers, which can be seen as a collection of encapsulates of different secure operations. For example, we have employ the data structures and protocols in the cryptography primitives layer to build secure matrix multiplication for the scenario where inputs and weights reside in different organizations.

C. Runtime Efficiency

An important aspect of a practical system is runtime efficiency, which is evaluated by throughput/latency. In practice, by conducting experiments on previous solutions (e.g., ABY^3), we found that for a number of ML models, the communication overhead introduced by MPC protocols has become the bottleneck of the whole system, especially when the communication between different parties is under the WAN setting. To improve the system efficiency, we introduce a novel sharing scheme in addition to ABY and the associated protocols for switches between this scheme and others. The proposed *HE share* is an asymmetric sharing scheme based on homomorphic encryption. In this scheme, one party holds ciphertexts and the other holds secret key, and they both hold meta information of the data. HE share can convert to arithmetic share and vice versa. Such flexibility to easily convert between Arithmetic and HE shares can significantly speedup the operation computation, especially for multiplication.

TABLE I
EVALUATION OF THE END-TO-END LATENCY OF NEBULA.

Bandwidth(Mbits)	TFE-ABY3	Nebula-HE
10	1394	102
30	465	76
50	277	71

D. Results

Here, we demonstrate some preliminary evaluations of our system. We compare it with an open-sourced framework named Tf-encrypted [1]. We implement a training protocol for secure logistic regression using these two systems and set different network bandwidth (between different parties) to study the running time of the whole system. The running time refers to the end-to-end latency of training 1024 samples and each sample has 100,000 features. As Table I shows, our system can drastically improve the training efficiency of the whole system. We infer the performance boost comes from the using of HE share, which is introduced above.

III. CONCLUSION AND FUTURE WORK

In this work, we present Nebula, a scalable privacy preserving machine learning system in Ant Financial. We describe the system overview and demonstrate how it works. In the future, we will integrate more privacy-preserving algorithms into the system for industry usage and explore GPU/FPGA acceleration for further speed optimization.

REFERENCES

[1] Chaochao Chen, Liang Li, Bingzhe Wu, Cheng Hong, Li Wang, and Jun Zhou. Secure social recommendation based on secret sharing. *arXiv preprint arXiv:2002.02088*, 2020.
[2] Chaochao Chen, Bingzhe Wu, Wenjin Fang, Jun Zhou, Li Wang, Yuan Qi, and Xiaolin Zheng. Practical privacy preserving poi recommendation, 2020.
[3] Payman Mohassel and Peter Rindal. Aby3: A mixed protocol framework for machine learning. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 35–52, 2018.
[4] Payman Mohassel and Yupeng Zhang. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 19–38. IEEE, 2017.

[1] https://github.com/tf-encrypted/tf-encrypted