FLChain: A Blockchain for Auditable Federated Learning with Trust and Incentive

Xianglin Bao
School of Computer Science
USTC
Hefei, China
e-mail:baoxl@mail.ustc.edu.cn

Cheng Su
School of Computer Science
USTC
Hefei, China
e-mail:gotzeus@mail.ustc.edu.cn

Yan Xiong
School of Computer Science
USTC
Hefei, China
e-mail:yxiong@ustc.edu.cn

Wenchao Huang
School of Computer Science
USTC
Hefei, China
e-mail:huangwc@ustc.edu.cn

Yifei Hu
School of Computer Science
USTC
Hefei, China
e-mail:huyifei@mail.ustc.edu.cn

***Abstract*—Federated learning (shorted as FL) recently proposed by Google is a privacy-preserving method to integrate distributed data trainers. FL is extremely useful due to its ensuring privacy, lower latency, less power consumption and smarter models, but it could fail if multiple trainers abort training or send malformed messages to its partners. Such misbehavior are not auditable and parameter server may compute incorrectly due to single point failure. Furthermore, FL has no incentive to attract sufficient distributed training data and computation power. In this paper, we propose FLChain to build a decentralized, public auditable and healthy FL ecosystem with trust and incentive. FLChain replace traditional FL parameter server whose computation result must be consensual on-chain. Our work is not trivial when it is vital and hard to provide enough incentive and deterrence to distributed trainers. We achieve model commercialization by providing a healthy marketplace for collaborative-training models. Honest trainer can gain fairly partitioned profit from well-trained model according to its contribution and the malicious can be timely detected and heavily punished. To reduce the time cost of misbehavior detecting and model query, we design DDCBF for accelerating the query of blockchain-documented information. Finally, we implement a prototype of our work and measure the cost of various operations.**

***Keywords*—blockchain; federated learning; incentive; decentralize; trust;**

I. Introduction

Machine learning models bring significant improvement across industries [1][2][3][4] but numerous sensitive real-world data are required for model training. Nowadays, data privacy becomes a worldwide major issue and strict data protection regulations have been issued. Huge amount of data are scattered across different organization and data combination is much more difficult. Federated learning (shorted as FL) is a distributed and privacy-preserving machine learning method proposed by google [5][6][7], where massive amount of decentralized data sets can be trained and complementary knowledge can be transferred among distributed model trainers. It preserves privacy of distributed training data and learns a global model by aggregating locally computed updates via a semi-honest parameter server. Trainers without enough data samples are allowed to build more powerful models via FL.

Despite these benefits, there are some notable but lacking of attention problems in Federated learning. Firstly, FL lacks auditability of malicious trainers and dishonest cooperation can disrupt model training. Incorrect masked gradients and unmasked shares can be uploaded to parameter server by dishonest local trainers. In real world, centralized parameter server is vulnerable under single-point-failure and may bring loss to all trainers. Furthermore, FL assume that distributed training data is sufficient and model training can be voluntarily and honestly done by trainers. In fact, real world trainers are reluctant to train model without incentive and unwilling to cooperation due to the worry of dishonest partners[8], [9]. So, it is vital to build a public auditable and decentralized mechanism for federated learning with trust and incentive.

To solve current sufferings, we put forward FLChain, a federated learning blockchain providing incentive and misbehavior deterrence for collaborative-modeling. FLChain document information of trainers and FL models and act as traditional parameter server of FL and its computation result must be consensual on-chain. Incorrect computation result is dropped by FLChain and its generator will be punished. We motivate misbehavior detection by providing reward for misbehavior detector and compensation for affected trainers.

Note that our work is not trivial since we make contributions as follows: 1) We propose FLChain to distribute trust and incentive among trainers. 2) We evaluate reliability and contribution of trainers for fair profit partition. We propose federation establishment algorithm for maximizing total profit of trainers. 3) By the way, FLChain provide a public marketplace of well-trained model for profit gaining. We propose a method to reduce the time cost of blockchain query for timely misbehavior detection and model purchase.

978-1-7281-4024-7/19/$31.00 ©2019 IEEE
DOI 10.1109/BIGCOM.2019.00030

II. Related Work

A. Blockchain

Blockchain is an emerging technology with its decentralized, immutable, sharing and time-order ledger. Transactions are stored into blocks containing timestamps and references and growth as a chain. Its pseudonymous FLChain nodes competitively collect transactions and generate new block for block reward so that the chain is continuously lengthened. Blockchain inspires us to build incentive, transparency, and fairness for our work, with its incentive feature [10] and smart contracts supporting Turing-complete programmability [11].

B. Federated Learning

Federated learning is also called as collaborative learning or distributed learning, combining machine learning and distributed computation, which maintains a global model commonly generated by the parameter server and distributed trainers. In federated learning, training data is partitioned and trained by trainers locally and individually. Local gradients are sent to the parameter server who aggregates those gradients and updates global model parameters accordingly. Then trainers download the updated parameters from the server and step into the next training iteration. The distributed training process repeats until the model accuracy are higher than prespecified thresholds [4].

Federated learning achieves input privacy for honest users, which make it harder to additionally guarantee correctness and availability [12][13][14]. It assumes users' inputs are always in correct aggregation once the parameter server receives local training result and cannot prevent attacker-controlled clients from sending malformed messages to others, which leads to abortion of federated learning protocol. The actively adversarial trainers can chose arbitrary values sending to parameter server and the output of federated learning protocol will be distorted [15][16].

Therefore, our work is vital by making corrupt clients be identifiable with verifiable training process and public auditing and evaluating each trainer's reliability to distribute trust.

C. Blockchain-based Machine Learning

Ref [17] describes an approach to link the crowd workforce with AI trainers and companies. The crowd label and validate data and are paid via cryptocurrencies. OpenMined [18] is an open source community implementing a federated learning architecture based on smart contracts. Ref [19] proposes DeepChain which provides secure deep learning training and blockchain-based incentive. The incentive of existing works only consider the machine Learning related attributes such as model accuracy and ignore each trainer's contribution and reliability while our work can provide a more stable environment for collaborative training.

Our work not only partition model profit fairly according to each trainer's contribution and reliability, but also consider the simultaneous situation of multiple taining tasks with our federation establishment algorithm, considering trainer collaborative intentions and reliability with the purpose of maximizing the total profit of training tasks.

III. System model

A. Overview

Our work replace traditional parameter server of Federated Learning with FLChain, but provides comparable security, by combining federated learning with cryptographic technique and blockchain to achieve traceable, auditable and privacy-preserving collaborative machine learning. The overview of FLChain is presented at Fig.1.

In FLChain, trainers of a federation are selected by FLChain according to their collaboration intention and reliability. They commonly train a federated learning model in an iterative manner. Firstly, trainers gossip for cryptographic settings and parameters for model initialization. The cryptographic settings and model parameters are uploaded to FLChain. When gradient generation and gradients aggregation are finished, trainers collaborative decrypt the aggregation result and update their local models accordingly. Each masked gradient and decrypted shares can be checked by its proof and the dishonest will be reported and punished. After that the federation will step into the next training iteration and when federated learning is finished, the reliability of each trainer will increase according to its partners evaluation and their model can be purchased via FLChain, whose price is set according to each trainer's expectation.

Parties who need a well-trained model should pay for model usage and the profit of the model is partitioned according to each trainer's reliability and contribution. Model users who purchased model can upload its model evaluation to FLChain and the reliability of trainer will be updated. The problematic training will be halted and its model are forbidden to sale. Misbehavior detectors will be rewarded and affected trainer will be compensated by FLChain.

B. Concepts

To guarantee not only confidentiality of training data and gradients, but also auditability of federated learning processes, FLChain trainer mask local gradients for privacy-preserving aggregation calculation and employ secret sharing mechanism in [20] to collaboratively unmask aggregation result. Furthermore, each masked gradient and unmasking part requires auditable proof to deter dishonest behaviors for secure federated learning. **Gradients generation** is a process in FL where trainers train their models locally and independently. At the end of each local machine learning iteration, verifiable and masked local gradients are uploaded to FLChain by each trainer. **Gradients aggregation** is a process where FLChain nodes aggregate FLChain-documented local gradients and upload aggregation results to FLChain. FLChain receive aggregation results from FLChain nodes and select the leader of nodes whose aggregation result will be documented by FLChain. **Collaborative unmasking** is a process where aggregated gradients are downloaded from FLChain and collaborative decrypted by trainers. After that

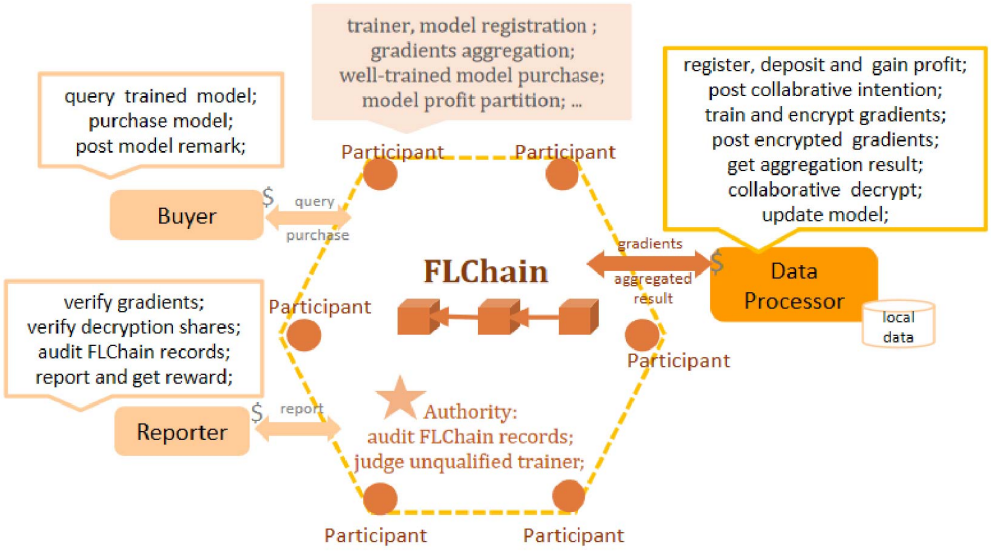

Fig. 1. FLChain Overview

process, trainers update model parameter and begin the next iteration of local model training. **Forced halting** is a process to halt the problematic collaborative modeling and only happens when honest reporter find that there does exist misbehavior in FL.

In particular, a certain amount of deposit should be frozen during federated learning and the dishonest one should pay penalty to FLChain, which used as the reward for misbehavior reporter and the compensation for affected trainers. The latest work for FLChain growth will generate some reward as the leader rewards. Trainers will gain profit for their federated learning contributions from FLChain, and pay for the usage of well-trained federated learning model. Recent works have shown the model-based pricing is reasonable for machine learning [21]. For deterrence, each trainer in FLChain should has a certain amount of deposit before federated learning. Next, we will explain the related terms used in FLChain.

FLChain node is the entities in FLChain and they are motivated to compute gradients aggregation result for collaborative model update, can be any trainer registered in FLChain. **Leader** is selected from FLChain nodes via FLChain consensus protocol to document the correct result of masked gradients aggregation into FLChain, and the work will be rewarded. Reward can be consumed for paying well-trained models usage fee in the future. **Authority** in FLChain is responsible for judging the unqualified trainers in the collaborative modeling processes and auditing the records documented by FLChain. **Trainer** is the FLChain node who need to achieve higher performance model with its insufficient computational power and training data, is the same entity as local trainer defined in traditional federated learning. **Federation** is a set of DPs who cooperate in the same FL modeling task and commonly own the FLM. **Federated Learning Model**, shorted as FLM, is a machine learning model collaborative trained by multiple trainers.

C. Threats and Goals

In federated learning, malicious trainer may generate incorrect gradients to mislead the collaborative training process. The federation will get erroneous results if incorrect gradients are uploaded. In collaborative unmasking, dishonest FLChain nodes may give a problematic unmasking share leading to federated learning failure. We should achieve the auditability of gradient collecting and collaborative unmasking. Also, the malicious may send incorrect incorrect aggregation result to trainers. We should achieve deterrence of incorrect gradients aggregation result and incentive of correct aggregation computation.

Smart contracts function may be executed overtime or mistakenly by malicious parties. Dishonest FLChain nodes may delay trading or terminate a contract for her own benefit and they may be selfish to save their cost for training by aborting local training process early, which makes the honest ones suffer losses. We should achieve deterrence of overtime and incorrect smart contract execution. Malicious party may illegally collect data to be qualified of collaborative modeling. In order to be selected into collaborative training federation and get high-performance machine-learning models, unqualified party lacking training data may collect data through illegal channel, which violates its privacy policies or corresponding privacy regulations. We ought to achieve deterrence of illegal training data collection.

IV. Design Details

A. Collaborative Modeling

The whole collaborative modeling in FLChain consists of several processes, including trainer registration, federation

TABLE I
DP (TRAINER)

Fields	Descriptions
pk^{PSU}	pseudonymous public key of trainer
dr	its data resource description, consists of data features, samples amount, labels
$rely$	the reliability of j^{th} trainer, decided by its partners and model buyers remarks
$FLMs$	federated learning models trained by j^{th} trainer
$status$	qualified, unqualified, suspending
$TorF$	a boolean, present the availability of j^{th} trainer
$height$	block height of last version of this DP

TABLE II
FLM (FEDERATED LEARNING MODEL)

Fields	Descriptions
fid	federated learning model id
DPs	trainers' pseudonymous public keys of a federated learning model's federation
$price$	model price
$TorF$	T means training federation is establishing, otherwise not establishing
$status$	unavailable, onsale, problematic
$height$	block height of last version

recursion preparation, federation establishment, collaborative training, FLM purchase and misbehavior detecting.

Trainer register its identical information and data resource to get its pseudonymous public key and corresponding secret key in FLChain. FLChain will document its identical information, data resources description, history collaborative training activities and reliability into FLChain block. Trainer (also called trainer, here shorted as DP) can update its FLChain-documented registration information via FLChain and the new one will include the block height which documented the old one. The fields of DP and FLM are present at Table I and Table II.

Federation recursion preparation consists of FLM description publishing and collaboration intention publishing in FLChain. Trainer publish the description of FLM they wanted to be collaboratively trained in FLChain. Each trainer receives all FLM descriptions whose collaborative training federation recursion not yet begun. Trainers select FLMs they interested in and publish collaboration intentions of corresponding FLMs. Collaboration intention consists of pseudonymous public key of trainer dp,federated learning model FLM, expected price ep, expected income $einc$, cost estimation $cost$. Then FLChain will allocate trainers to federations according to their FLM intentions. The Federations Establishment algorithm is presented at Algorithm1.

After a federation is established, the cryptographic system implemented by trust component like SGX, steps into collaborative training preparation of cryptographic setting and secret sharing for federation trainers. Collaborative training preparation protocol is presented at Algorithm2.

Algorithm 1 Federation Establishment Function

Input:
federated learning models to be trained $FLMs$;
Collaboration Intention of trainers CIs;
Output:
allocation result $A_{i,j}$, $A_{i,j}$=1 means DP_j is allocated in FLM_i, $A_{i,j}$=0 means not DP_j is not allocated in FLM_i

Allocate trainers into collaboratively training federations
select trainers who has published CIs
DPs is the set of selected trainers
for each $i \in FLMs$ **do**
  for each $j \in DPs$ **do**
    Set $a_{i,j} = 0$;
set maximum profit $PFT_{max} = 0$;
for each $FLM_i \in FLMs$ **do**
  for each $DP_j \in DPs$ **do**
    if $RELY_j * (EINC_{i,j} - COST_{i,j}) \geq p_{max}$ **then**
      PFT_{max}=$RELY_j * (EINC_{i,j} - COST_{i,j})$;
      FLM_{max}=FLM_i,DP_{max}=DP_j;
while $FLMs \neq \emptyset$ and $DPs \neq \emptyset$ **do**
  if $PFT_{max} \geq 0$ **then**
    $A_{FLM_{max},DP_{max}}$ = 1
    Remove DP_{max} from DPs;
    if federation of $FLMs_{max}$ training is full **then**
      Remove FLM_{max} from $FLMs$;
return A;

After collaborative training preparation is done, trainers in each federation step into iterations of collaborative training. The collaborative training protocol is demonstrated in Algorithm3. Data processes train their models locally and gradients are generated. Then trainers mask their local gradients via their cryptographic secrets and upload the encrpted gradients with proofs to FLChain. FLChain document all verified masked gradients and FLChain nodes download all masked gradients to calculate aggregation result. FLChain nodes publish aggregation result and FLChain will select a FLChain node with the highest reliability as the FLChain leader to document its result into the new generated FLChain block.

Trainers download result and collaboratively decrypt aggregation result. Each trainer should send its decrypted shares of result with proofs to the cryptographic system and fully decrypted result will be returned. Then, an iteration of FLM is done and data processes will evaluate each others. Each trainer's reliability is updated according to its partners evaluaion and the next local traning iteration begins.

After FLM is well-trained, and its price is set acorrding to trainers intention and the profit shares of data processes are decided by not only the intention, but also the reliability and training data resources amount. Then the model description and commodity information of FLM will be documented to FLChain.

Trainer who want to use well-trained FLM check the FLM information and its trainers information via FLChain, and

Algorithm 2 Collaborative Training Preparation

Trainer DP_j**:**

- Upload its information into FLChain and FLChain register a new DP.
- Send some amount of deposits to FLChain public account for secure computation.
- Send collaboration intention to FLChain.

FLChain:

- Register and document DP into FLChain.
- Register and document a new FLM FLM into FLChain.
- Allocate trainers into the federation of training FLM according to their collaboration intention.

CryptoSystem (a trusted component like SGX):

Suppose there are N trainers $[DP_1, DP_2, ..., DP_n]$ allocated into a federation, whose pseudonymous public keys are $[P_1, P_2, ..., P_n]$ and private keys are $[S_1, S_2, ..., S_n]$. Current block height is h and a secrect sharing threshold t to decrypt a cipher is set. f is a function of secret sharing protocol. Store Lagrange interpolation coefficient with respect to each P_j.

- Generate the product of two selected safe primes n_{FLM}, $g_{FLM} \in Z^*_{n^2_{FLM}}$,$\theta = as, a, s \in Z^*_{n_{FLM}}$.
- Generate cryptographic parameters $[r_0, r_1, ..., r_n], r_i \in Z^*_{n_{FLM}}$
- Randomly divide a trainer's FLM secret s into N shares $s_j, j \in [1, 2, ..., N], s = f(s_1 + ... + s_N)$.
- Generate public parameters to verify unmasked shares $v, \{v_j\} \in Z^*_{n_{FLM}}, j \in [1, ..., N]$.
- Public key of FLM is $PK_{FLM} = (n_{FLM}, g_{FLM}, \theta, v, \{v_j\})$.
- Send PK_{FLM}, s_i to other trainers $DP_j \in [DP_1, DP_2, ..., DP_n]$
- Upload public verification parameter $v, \{v_j\}$ of FLM to FLChain.

Trainer DP_j**:**

- Upload commitment of secret shares to FLChain $Enc(s_j, h, Sign(s_j||h, S_j))$
- Gossip the initial weight W_0 of FLM, upload $C(W_0) = g^{W_0}_{FLM} * r_0^{n_{FLM}}, r_0 \in Z^*_{n_{FLM}}$
- Provides local models initial weights $C(W_{0j}) = g^{W_{0j}}_{FLM} * r_j^{n_{FLM}}, r_j \in Z^*_{n_{FLM}}$,

FLChain update FLM **information.**

Algorithm 3 Collaborative Training

Suppose the current iteration of collaborative training is i, pseudonymous public keys of trainers $[DP_1, DP_2, ..., DP_n]$ are $[P_1, P_2, ..., P_n]$, corresponding with $[r_1, r_2, ..., r_n], r_j \in Z^*_{n_{FLM}}$

Trainer DP_j**:**

- Local training model, generate gradient G_{ij}.
- Mask gradient of local model $C(G_{ij})$.
- Generate its correctness proof $PFC_{ij} = proof_C(C(G_{ij}); G_{ij}, r_j; PK_j)$
- Upload masked training results $C(G_{ij})$ along with proofs PFC_{ij}.
- Receive $\{C(G_{ij})\}, \{PFC_{ij}\}$, verify masked gradient is the generated from G_{ij} with r_j via $verf_C(C(G_{ij}), PFC_{ij}, PK_j$.
- Report incorrect $C(G_{ij})$ along with PFC_{ij}, PK_j to FLChain, get reward and compensation.

FLChain:

- Get correct $\{C(G_{ij})\}$ along with current iteration i.
- FLChain nodes aggregate verified $\{C(G_{ij})\}$ and upload aggregation results $C(G_i)$.
- Leader selection.

Trainer DP_j

- Download aggregation result $C(G_i)$ documented in FLChain.
- Unmask aggregation result with s_j and get unmasked share CD_{ij}, generate proof $PFD_{ij} = proof_D(C(G_i), CD_{ij}, v, v_j); PK_j)$.
- Gossip unmask share CD_{ij} along with its proof PFD_{ij}.
- Receive unmasked shares $\{CD_{ij}\}$ and proofs $\{PFD_{ij}\}$.
- Verify each unmasked shares via $verf_{CD}(C_{ij}, CD_{ij}, v, v_{ij}, PFD_{ij}, PK_j)$.
- Report incorrect CD_{ij} with $v, v_{ij}, PFD_{ij}, PK_j$.
- Calculate G_i and update local model weights.
- Step into next iteration, update current iteration number i.

If training is finfished or halted, FLChain will update the information of FLM.

purchase FLM documented in FLChain whose trainers are all honest and audited by decentralized detectors. The buyers evaluate the performance FLM they purchased and FLChain update each trainer reliability once in a certain interval. The profit of each FLM divided and distributed automatically via FLChain.

Any one detect misbehavior can report it to FLChain and the reports will be handled transparently via FLChain. FLChain force halts collaborative training of misbehaved trainer and forbidden the sale of FLM trained by the dishonest. A certain deposit of the dishonest will be confiscated and used as the compensation to the honest and the reward to reporter. Anyone detecting incorrect encrpted gradients, decrypted aggregation result shares can send incorrect information with its proofs to FLChain. The registration information and collaborative modeling history of trainers are documented by FLChain permanently and transparently which decentralized detectors can pulic audit. Unqualified trainer who legally collect training data can be reported by pulic and transparently judged by authority. The honest reporter will gain rewards from FLChain and the dishonest will lose some report fee.

The collaborative training process and commercialization details of FLM should observe the policies of FLChain to achieve fairness, incentives and deterrence. The FLChain policies consists of misbehavior detecting reward, limitation of a FLM trainer's amount and computation method of FLM price, profit shares and trainer reliability.

B. FLChain Architecture

We use four-layer blockchain architecture including data layer, network layer, extension layer and application layer. The FLChain architecture are demonstrated at Fig.2.

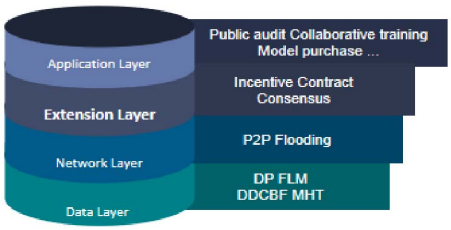

Fig. 2. FLChain Architecture

The data layer of FLChain includes DP, FLM, DDCBF and MHT. DP and FLM is introduced above and we propose DDCBF to accelerate FLChain query. In FLChain, FLM buyer need to query the legalty of FLM and purchase FLM allowed for sale, including unproblematic FLM and problematic FLM but judged legal. Detector need to query FLChain records and listen the event and judgement from FLChain. Bloom filter (BF) is a space-efficient probabilistic data structure used to check if a set contains an element but can not remove an element while counting bloom filter (CBF) support it [22]. CBF replace every bit to a counter, which will add one if an element hash is mapped to it and decrease one if remove an elements. Our DDCBF demonstrated at Fig.3, doubles dual CBF to quickly sift problematic FLM still in judging and a dual count bloom filter to quickly sift FLM allowed for sale with high possibility, which consists of DCBF1 and DCBF2.

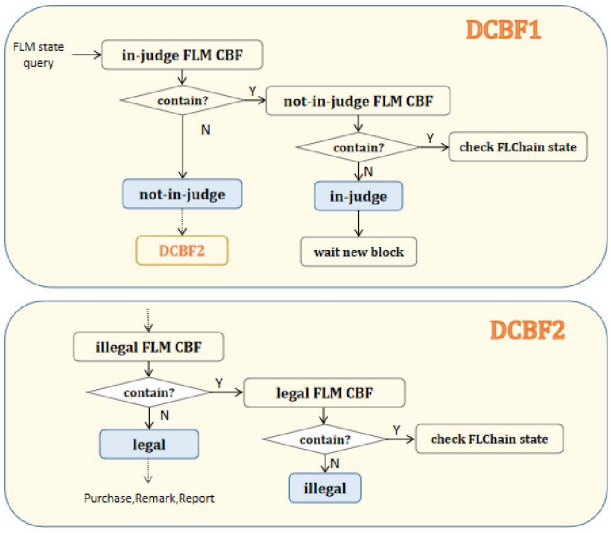

Fig. 3. DDCBF

The FLChain node need to generate MHT containing all DPs and FLMs. The latest statuses of DP and FLM are updated in DDCBF and DPs, FLMs, DDCBF and MHT are stored into FLChain block, MHT is demonstrated at Fig.4. Our network layer are well compatible with existing P2P network protocol and flooding transmission mechanism, due to the approximate transaction size.

Our application layer provides the FLM marketplace for trainers and decentralized supervision platform for misbehavior detectors. The FLM buyer query FLM allowed for sale, send FLM purchase request, get well-trained FLM and upload evaluation to FLChian. Collaborative trainers deposit for collaborative training, set cryptographic parameters, collaborative train and gain fair profit shares from FLChain. Misbehavior detectors query and audit FLM and DP documented in FLChain, verify masked gradients and decrypted aggregation shares, report misbehavior to FLChain and the dishonest one will be punished. In next subsection, we will introduce extension layer in details.

C. Extension Layer

To spread trust in FLChain, FLChain evaluates each FLChain node reliability and leverage the reliability-based leader selection to achieve FLChain consensus. The trainer's reliability affect profit shares, decided by not only the traning data amount of data FLChain node, but also the history remarks of its partner and FLM buyer. The poor trainers are of low reliability due to their poor training and low remarks, and the misbehaved one is punished with reliability deduction. To build incentive and deterrence mechanism, FLChain provide marketplace for model commercialization and partition profit fairly between collaborative training partners according to their contributions and reliabilities. What's more, FLChain confiscate deposit of the dishonest, and provides reward for decentralize misbehavior reporters and compensation for affected honest trainers. Our secure collaborative modeling is implemented by reliability-based consensus and several FLChain contracts,including incentive contract, federation establishment contract, collaborative training contract and misbehavior report contract.

We achieve safety, correctness and liveness of our FLChain based on our reliability-based consensus, which randomly selects leader from high reliable FLChain nodes. The selected leader will document its aggregation result into the new block after achieving consensus and gain block reward. All honest FLChain nodes agree on a same FLChain and require any information documented in its block are proved correct. FLChain nodes are willing to continuously perform activities in FLChain to win the reward which keeps FLChain alive.

Incentive contract works for deposit management, model commercialization and incentive distribution, which provides service like deposit bonding and deposit withdraw, FLM purchase and profit partition, reporter rewarding and compensation provision. **Federation establishment contract** works for collaborative training preparation and cryptographic setting, which provides service to invoke FLM federation events, such as federation recursion description upload and collaboration intension upload. **Collaborative training contract** works

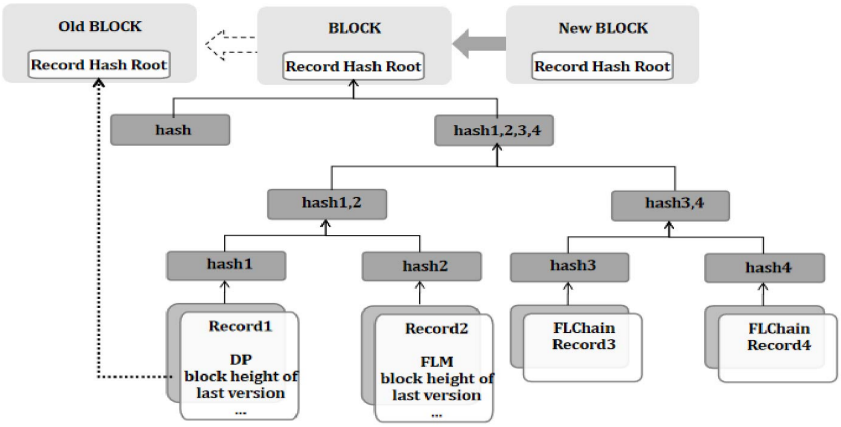

Fig. 4. FLChain MHT

for model training and provide service to post/get masked local gradients, post/get aggregation result and update/upgrade FLM. **Collaborative training contract** not only achieve an auditable and verifiable privacy-preserving federated learning, but also build a fair and reliable cooperation environment for trainers with dynamic reliability and trainer contribution evaluation. **Misbehavior report contract** receive report of incorrect gradient, incorrect aggregation or unqualified FLChain node, to achieve public audit of FL and timely detection of the malicious.

V. Analysis

A. Incentive and Deterrence

In this section, we will analyze the incentive and deterrence in FLChain. In particular, we model the incentives of each entity in FL ecosystem by considering the flow of payments among entities for each FLChain event. Using this model, we demonstrate two important guarantees that hold in FLChain: 1) Incentives for honest collaborative training and misbehavior reporting: uploading incorrect masked gradients or decrypted aggregation shares that violates a FLM and unqualified collaborative trainner violates privacy regulations results in a higher payout than honest actions, while misbehavior report results in FLChain rewards. 2) Misincentives against dishonest reporter: falsely reporting an honest trainer does not result in a profit for a reporter or trainer.

In order to get the properties of compensation of trainers affected by misbehavior, rewards for honest detectors and the deterrence of the malicious, our work requires that: 1) The affected trainer receive a higher net payout after a honest report. 2) A successful misbehavior report results in a higher net payout for the detector than an unsuccessful report or no report at all. 3) The malicious who upload incorrect masked gradient or unmasking shares have a negative net payout.

For each case, we consider the payments made in FLChain related actions and summary it at Table III. We analysis the net profit of each entity by calculate total payments it received and deducting total payments it made at Table IV. We

TABLE III
List of Payments sent for each action.

Event	From	To	Amount
Register DP	DP	PAC	d_{DP}
Upload Collaboration Intention	DP	FED	f_{CI}
Register FLM	FED	PAC	f_{FLM}
Purchase FLM	BUY	PAC	p_{FLM}
-	PAC	FED	p_{FLM}
-	FED	DP_j	p^j_{FLM}
Report misbehavior	RP	PAC	f_1
-	MDP	PAC	d_{DP}
-	PAC	RP	r_1
-	PAC	FED	$d_{DP} - r_1$
Report unqualified DP	RP	PAC	f_2
-	PAC	AUTH	f
-	UQDP	PAC	d_{DP}
-	PAC	RP	r_2
-	PAC	FED	$d_{DP} - f - r_2$
Dishonest misbehavior report	RP	PAC	f_1
Dishonest unqualified report	RP	PAC	f_2

consider a single FLM lifetime, and use the following notation: DP denotes the trainer account, HDP denotes a honest trainer account, MDP denotes a misbehaved trainer account, $UQDP$ denotes a unqualified trainer account, FED denotes a federation account, RP denotes a detector account, PAC denotes public account, $AUTH$ denotes authority account. The model price p_{FLM} can be set as the average price of trainers collaborative intentions, which must be larger than the training cost c, and the minimum deposit of a trainer is $d >> p_{FLM}$. The profit share of a trainer can be set as $shares^j = RELY^j/N$. We calculate the profit share of a trainer via $p^j_{FLM} = p_{FLM} * shares^j * RELY^j$. The comparison of affected trainer can be $cp = p^j_{FLM}$ and the detector reward r is larger than is report fee $f_1 or f_2$.

B. Security Analysis

1) Privacy preserving and auditability of collaborative training: In FLChain each FLChain node individually masks and then uploads gradients obtained from her local model. All

TABLE IV
INCENTIVE FOR EACH ENTITY.

Entity	Well-train	Unrep.	Rep.	Dishonest Rep.
HDP	$p-c$	$-c$	$cp-c$	0
MDP	-	-	$-d$	-
$UQDP$	$p-c$	$-c$	$-d-c$	-
RP	0	0	r_1-f_1	$-f_1$
-	0	0	r_2-f_2	$-f_2$
PAC	0	0	$\Sigma(f_1-r_1)$ $+d-\Sigma cp$	Σf_1
-	0	0	$\Sigma(f_2-r_2)$ $+d-\Sigma cp-f$	Σf_2

gradients are used to update parameters of the collaborative model collaboratively by all FLChain nodes, who then obtain updated parameters via collaborative unmasking in each iteration. Here, collaborative unmasking means that at least t FLChain nodes provide their secret shares to decrypt a cipher. We assume that FLChain nodes do not expose their own data and at least t FLChain nodes are honest. Then each partys local gradients cannot be exposed to anyone else, unless at least t FLChain nodes collude.

During gradient collecting, trainers transactions are consist of masked gradients and correctness proofs, allowing the third party to audit whether a FLChain node gives a correctly masked construction of gradients. Trainers collaboratively decrypt the parameters and provide their unmasking shares with proofs for correctness verification, allowing any third party audit whether the unmasking shares are correct or not.

2) Deterrence of misbehavior: Any trainers in FLChain should register its data resources into FLChain for collaborative modeling and public audit. Trainers can act as FLChain nodes and both of them should have enough deposit in FLChain. FLChain applies the monetary penalty mechanism, revoking the pre-frozen deposit of dishonest trainers and imposing penalty FLChain nodes who do not behaved punctually and correctly. FLChain defines a time point for each smart contracts function and results of the function are verified. If the verification failed, it means that there exist FLChain node not being punctual or incorrectly execute the function.

For deterrence and transparency, anyone can audit training data information documented in FLChain and report dishonest or unqualified FLChain node to FLChain. FLChain check the balance of the deposit of each trainer and pre-frozen it. The problematic collaborative training can be halted by FLChain and the dishonest will lose its pre-frozen deposit after FLChain judgement. Affected trainers will be well compensated when their collaborative training is halted due to the dishonest.

VI. IMPLEMENTATION AND EVALUATION

A. Evaluation

We implement our FLChain prototype based on FLChain and has ten FLChain nodes who document DP and FLM in SHA-512 MHT stored in FLChain. We select dataset with 55,000 training samples, 5,000 verification samples and 10,000 test samples. We separately allocate different amounts trainers

TABLE V
COLLABORATIVE TRAINING WITH 4 DPs

DP	Val.	Acc.	DP	Val.	Acc.
DP_1	0.9683	0.9696	DP_3	0.9721	0.9702
DP_2	0.9712	0.9719	DP_4	0.9698	0.9689
Single Set	0.9579	0.9618	Full Set	0.9767	0.9795

TABLE VI
COLLABORATIVE TRAINING WITH 10 DPs

DP	Val.n	Acc.	DP	Val.	Acc.
DP_1	0.9701	0.9676	DP_6	0.9713	0.9702
DP_2	0.9723	0.9719	DP_7	0.9703	0.9676
DP_3	0.9721	0.9692	DP_8	0.9692	0.9721
DP_4	0.9690	0.9689	DP_9	0.9688	0.9698
DP_5	0.9683	0.9676	DP_{10}	0.9727	0.9673
Single Set	0.9520	0.9561	Full Set	0.9767	0.9795

in a federation and each one trains one subset of total samples. We set a baseline trainer only training the local model on its dataset without collaboration. Training iteration is 1500, epoch is 1, learning rate is 0.5, min batch size is 64 and optimizer function is stochastic gradient descent. We divide dataset into 1) 4 pieces and each DP has 13750 training samples. 2) 10 pieces and each DP has 5500 training samples.

The training validation and test accuracy of single dataset, full dataset and collaborative training with four and ten trainers are compared at Table V and Table VI. We can find that each trainer has higher validation and acurracy than the single set as baseline, and each model performance is approach to the model trained with full dataset. The comparison of two tables shows that model perfomrances of trainers with smaller dataset will improve more via collaborative training. At Table VII, we summary record size and analyze the capacity of each block according to the size of total block and the empty block.

We set FLChain difficulty as 0x160000 and the average block generation interval is 6700 ms. The total training time consists of gradient masking, gradients upload, masked gradients download, aggregation result upload, aggregation result download, collaborative unmasking. At Table VIII, we summary the average time of FLChain event. In traditional

TABLE VII
RECORD AND BLOCK

Component	Parameter	Component	Parameter
DP	2KB	Block Size	2MB
FLM	2KB	Empty Block	2.6KB
$DDCBF$	1MB	Block capacity	500 records

TABLE VIII
AVERAGE TIME COST

Event	Time (ms)	Event	Time (ms)
query MHT	15.68	total query	33.31
query $DDCBF$	13.42	total on-chain time	653500
update record	184.54	total training	401396719

blockchain, block query time cost is logarithmic correlation to total records amount and blockchain should be updated. We find that the average of total query time in FLChain is quite small thanks to DDCBF. Time cost of on-chain action such as masked gradients upload, masked gradients download, aggregation result upload and aggregation result download is small and the most proportions of total time cost are encryption and decryption for privacy-preserving training. It suggests that training tasks with different trainers amount and their total training times will similar to the average.

VII. Conclusion

We propose FLChain for federated learning with trust and incentive. FLChain stores trainer information and verifiable training details for public auditability, and its incentive mechanism motivates trainer be honest and detectors report misbehavior in FLChain. Incorrect mask of local gradient is dropped by FLChain and the dishonest can be detected by public and punished via FLChain. Our work is not trivial when it is vital and hard to provide enough incentive and deterrence. We provide a healthy marketplace for collaborative-training models for public and honest trainer gain fairly partitioned profit from well-trained model according to its contribution. The malicious can be timely detected and heavily punished. To reduce the time cost of blockchian query, we design DDCBF for FLChain and it is no need to go through blockchain for most state query. Finally, we implement a prototype of our work and measure the collaborative model performance and time cost of various operations. In future, we will propose a more elaborate method of reliability evaluaion and incentive strategy for FLChain. We plan to design a time-save method of privacy preserving and training auditing to reduce total time cost.

Acknowledgment

The research is supported by National Natural Science Foundation of China under Grant No.61572453, No.61202404, No.61520106007, No.61170233, No.61572454, and the Fundamental Research Funds for the Central Universities, No. WK2150110009.

References

[1] Lampos V , Miller A C , Crossan S , et al. Advances in nowcasting influenza-like illness rates using search query logs[J]. Scientific Reports, 2015, 5(1):12760.

[2] Paparrizos J , White R W , Horvitz E . Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results[J]. Journal of Oncology Practice, 2016, 12(8).

[3] Hinton G , Deng L , Yu D , et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups[J]. IEEE Signal Processing Magazine, 2012, 29(6):82-97.

[4] Gawehn E , Hiss J A , Schneider G . Deep Learning in Drug Discovery[J]. Molecular Informatics, 2016, 35(1):3-14.

[5] Bonawitz K , Ivanov V , Kreuter B , et al. [ACM Press the 2017 ACM SIGSAC Conference - Dallas, Texas, USA (2017.10.30-2017.11.03)] Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, - CCS Ï7 - Practical Secure Aggregation for Privacy-Preserving Machine Learning[C]// Acm Sigsac Conference on Computer & Communications Security. ACM, 2017:1175-1191.

[6] Konen, Jakub, Mcmahan H B , Ramage D , et al. Federated Optimization: Distributed Machine Learning for On-Device Intelligence[J]. 2016. Konen, Jakub, Mcmahan H B , Yu F X , et al. Federated Learning: Strategies for Improving Communication Efficiency[J]. 2016.

[7] Mcmahan H B , Moore E , Ramage D , et al. Federated Learning of Deep Networks using Model Averaging[J]. 2016. Bagdasaryan E , Veit A , Hua Y , et al. How To Backdoor Federated Learning[J]. 2018.

[8] Shokri R , Shmatikov V . Privacy-Preserving Deep Learning[C]// ACM Conference on Computer and Communications Security (CCS). ACM, 2015.

[9] Heigold G , Vanhoucke V , Senior A , et al. Multilingual acoustic models using distributed deep neural networks[C]// Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.

[10] Eyal I , Gencer A E , Sirer E G , et al. Bitcoin-NG: A Scalable Blockchain Protocol[J]. 2015.

[11] Delmolino K , Arnett M , Kosba A , et al. Step by Step Towards Creating a Safe Smart Contract: Lessons and Insights from a Cryptocurrency Lab[J]. 2016.

[12] Wang G , Wang T , Zhang H , et al. Man vs. machine: practical adversarial detection of malicious crowdsourcing workers[C]// Usenix Conference on Security Symposium. USENIX Association, 2014.

[13] Gu T , Dolan-Gavitt B , Garg S . BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain[J]. 2017.

[14] Rubinstein B I P , Nelson B , Huang L , et al. ANTIDOTE:understanding and defending against poisoning of anomaly detectors[C]// Acm Sigcomm Conference on Internet Measurement. DBLP, 2009.

[15] Steinhardt J , Koh P W , Liang P . Certified Defenses for Data Poisoning Attacks[J]. 2017.

[16] Biggio B , Nelson B , Laskov P . Poisoning Attacks against Support Vector Machines[J]. 2012.

[17] "Dbrain - the blockchain platform to label data and to build ai apps,2018, Version 0.2." [Online]. Available:https://dbrain.io/DbrainWhitepaper.pdf.

[18] "Openmined - building safe artificial intelligence." [Online]. Available: https://openmined.org/.

[19] Weng J , Zhang J , M Li , et al. Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive[J]. Cryptology ePrint Archive, Report 2018/679, 2018.

[20] Phong L T , Aono Y , Hayashi T , et al. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(5):1333-1345.

[21] Chen L , Koutris P , Kumar A . Model-based Pricing for Machine Learning in a Data Marketplace[J]. 2018.

[22] Tarkoma S , Rothenberg C E , Lagerspetz E . Theory and Practice of Bloom Filters for Distributed Systems[J]. IEEE Communications Surveys & Tutorials, 2012, 14(1):131-155.