# A Game-Theoretic Framework for Incentive Mechanism Design in Federated Learning

Mingshu Cong[1,2(✉)], Han Yu[3], Xi Weng[4], and Siu Ming Yiu[2]

[1] LogiOcean Technologies, Shenzhen, China
miranda.cong@logiocean.com
[2] The FinTech and Blockchain Lab, The University of Hong Kong, Pok Fu Lam, Hong Kong
smyiu@cs.hku.hk
[3] School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore
han.yu@ntu.edu.sg
[4] Guanghua School of Management, Peking University, Beijing, China
wengxi125@gsm.pku.edu.cn
http://www.logiocean.com

**Abstract.** Federated learning (FL) has great potential for coalescing isolated data islands. It enables privacy-preserving collaborative model training and addresses security and privacy concerns. Besides booming technological breakthroughs in this field, for better commercialization of FL in the business world, we also need to provide sufficient monetary incentives to data providers. The problem of FL incentive mechanism design is therefore proposed to find out the optimal organization and payment structure for the federation. This problem can be tackled by game theory.

In this chapter, we set up a research framework for reasoning about FL incentive mechanism design. We introduce key concepts and their mathematical notations specified under the FML environment, hereby proposing a precise definition of the FML incentive mechanism design problem. Then, we break down the big problem into a demand-side problem and a supply-side problem. Based on different settings and objectives, we provide a checklist for FL practitioners to choose the appropriate FL incentive mechanism without deep knowledge in game theory.

As examples, we introduce the Crémer-McLean mechanism to solve the demand-side problem and present a VCG-based mechanism, PVCG, to solve the demand-side problem. These mechanisms both guarantee truthfulness, i.e., they encourage participants to truthfully report their private information and offer all their data to the federation. Crémer-McLean mechanism, together with PVCG, attains allocative efficiency, individual rationality, and weak budget balancedness at the same time, easing the well-known tension between these objectives in the mechanism design literature.

Supported by LogiOcean Co., Ltd.


Q. Yang et al. (Eds.): Federated Learning, LNAI 12500, pp. 205–222, 2020.
https://doi.org/10.1007/978-3-030-63076-8_15

**Keywords:** Federated learning · Mechanism design · Game theory

## 1 Introduction

In most industries, data are segregated into isolated data islands, among which direct data sharing is restricted by laws and regulations such as the General Data Protection Regulation (GDPR) [6]. *Federated learning* (FL) [12] has emerged in recent years as an alternative solution to train AI models based on distributedly stored data while preserving data privacy. Commercial FL platforms have been developed, e.g., *TensorFlow Federated* (TFF) from Google and *FATE* from WeBank. Industries such as finance, insurance, telecommunications, healthcare, education, and urban computing have great potential to benefit from FL technologies.

In real application scenarios of FL, where data providers are profit-seeking business entities, FL may not be economically viable because of the *free rider problem*, i.e., a rational data provider may hold back its data while expecting others to contribute all their data to the federation. Without proper incentives, it is hard to prevent such free-riding activities because the FL model, as a virtual product, has characteristics of *club goods*, i.e., it is non-rivalrous in consumption.

In order to incentivize data providers to offer their best datasets to federated learning, we need to pay data providers enough monetary reward to cover their costs. The marginal monetary reward for contributing more data should be no less than the marginal cost hence incurred. Also, we aim to maintain a balanced budget and optimize for social welfare. At least three sources of *information asymmetry* intertwined in this problem: 1) the datasets owned by each data provider, 2) costs incurred to each data provider, and 3) model users' valuations on the trained FL model. An *FL incentive mechanism*, formulated as a function that calculates payments to participants, is designed to overcome these information asymmetries and to obtain the above-mentioned objectives. The problem of *FL incentive mechanism design* is to find the optimal FL incentive mechanism.

In this chapter, we first propose a game-theoretic model for analyzing the FL incentive mechanism design problem. We provide a checklist to specify heterogenous game settings and mechanism design objectives, together with four benchmark theorems that help FL practitioners to choose the appropriate FL incentive mechanism. Then, under our research framework, we provide two examples of FL incentive mechanisms, one on the demand side and the other on the supply side. The proposed Crémer-McLean mechanism and Procurement-VCG (PVCG) mechanism encourage FL participants to truthfully report their type parameters and offer their best datasets to the federation. These mechanisms also provide theoretical guarantees for incentive compatibility, allocative efficiency, individual rationality, and weak budget balancedness.

## 2 Problem Setup

In this section, we set up a game-theoretic environment for our following discussions. For readers unfamiliar with game theory and mechanism design, this section also provides necessary background knowledge.

### 2.1 The Game-Theoretic Environment

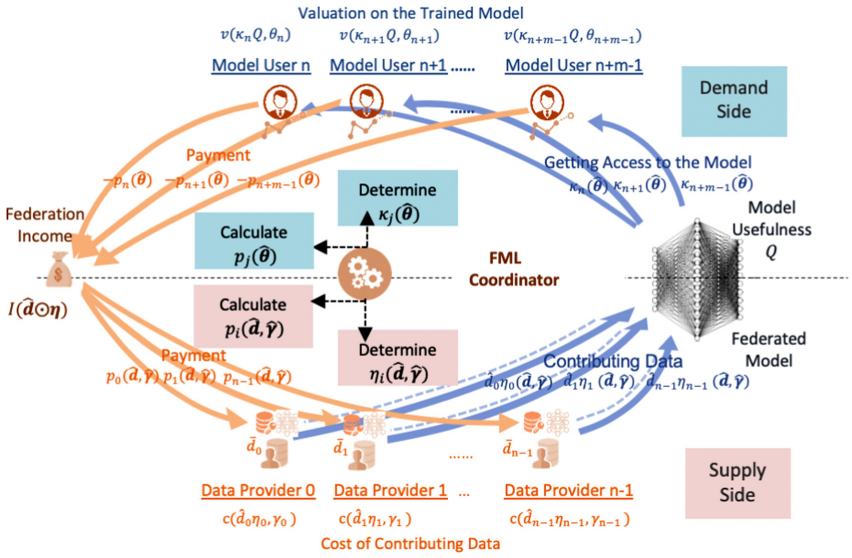


**Fig. 1.** The circular flow diagram of federated learning.

The environment of FL incentive mechanism design is set up as follows:

- There exists a set of $n$ *data providers*, denoted by $N = (0, \ldots, n-1)$, and another set of $m$ *model users*, denoted by $M = (n, \ldots, n+m-1)$;
- Each data provider $i \in N$ owns a dataset $\bar{d}_i$. It claims it owns a dataset $\hat{d}_i$. The federation accepts a dataset $d_i \leq \hat{d}_i$ from this data provider. We call $\eta_i = d_i \oslash \hat{d}_i$ the *acceptance ratio*, where $\oslash$ denotes element-wise division.
- Trained on datasets $\boldsymbol{d} = (d_0, \ldots, d_{n-1})$ from all data providers, the usefulness of the federated model is $Q(\boldsymbol{d})$. Model users may be granted limited access to the federated model such that the usefulness of the federated model to model user $j$ is $\kappa_j Q(\boldsymbol{d})$, where $\kappa_j$ is called the *access permission*.
- Each data provider $i \in N$ has a *cost type* $\gamma_i \in \Gamma_i$. Its cost of contributing data $d_i$ is $c(d_i, \gamma_i)$. The collection of cost types of all data providers forms the *cost type profile* $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_n)$. Data provider $i$ may report a different cost type $\hat{\gamma}_i$.

- Each model user $j \in M$ has a *valuation type* $\theta_j \in \Theta_j$. Its valuation on the trained federated model is $w(\kappa_j Q(\boldsymbol{d}), \theta_j) = v(\boldsymbol{d}, \kappa_j, \theta_j)$. The collection of valuation types of all model users forms the *valuation type profile* $\boldsymbol{\theta} = (\theta_n, \ldots, \theta_{n+m-1})$. Model user $j$ may report a different valuation type $\hat{\theta}_j$.
- The payment to data provider $i \in N$ is $p_i \geq 0$. The payment to model user $j \in M$ is $p_j \leq 0$. We denote $\boldsymbol{p}^s = (p_0, \ldots, p_{n-1})$ and $\boldsymbol{p}^d = (p_n, \ldots, p_{n+m-1})$. The *federation income* is $I = -\sum_{j=n}^{n+m-1} p_j$; the *federation expenditure* is $E = \sum_{i=0}^{n-1} p_i$; the *federation profit* is $P = \sum_{l=0}^{n+m-1} p_l$.
- Participants' preferences are represented by *quasi-linear utility functions* $u_i(\cdot) = p_i(\cdot) - c_i(\cdot), i \in N$ and $u_j(\cdot) = p_j(\cdot) + v_j(\cdot), j \in M$.
- The social effect of federated learning is measured by *social surplus*, defined as $S(\cdot) = \sum_{j=n}^{n+m-1} v_j(\cdot) - \sum_{i=0}^{n-1} c_i(\cdot)$, which includes consumer surplus $S^d = \sum_{j=n}^{n+m-1} v_j(\cdot)$ and producer surplus $S^d = -\sum_{i=0}^{n-1} c_i(\cdot)$.
- There is user-defined *unfairness functions* $\varpi^s(\boldsymbol{p}^s, \boldsymbol{c})$ and $\varpi^d(\boldsymbol{p}^s, \boldsymbol{v})$ that measures the unfairness among data providers and model users.

Figure 1 illustrates the flows of economic resources in this federated learning game. Table 1 lists the mathematical symbols.

**Table 1.** List of mathematical symbols

| Symbol | Meaning |
|---|---|
| i | Index of data provider |
| j | Index of model user |
| $\bar{d}_i$, $\hat{d}_i$, or $d_i$ | Owned/claimed/accepted dataset |
| $Q(\boldsymbol{d})$ | Usefulness of federated model |
| $\gamma_i$ or $\hat{\gamma}_i$ | True/reported cost type |
| $\theta_j$ or $\hat{\theta}_j$ | True/reported valuation type |
| $p_i, p_j$ | Payment to participants |
| $\eta_i(\cdot)$ | Acceptance ratio of datasets |
| $\kappa_j(\cdot)$ | Access permission to the federated model |
| $c(d_i, \gamma_i)$ | Individual cost function |
| $v(\boldsymbol{d}, \kappa_j, \theta_j)$ | Individual valuation function |
| $u(\cdot)$ | Utility function |
| $I(\cdot), E(\cdot), P(\cdot)$ | Federation income/expenditure/profit |
| $S(\cdot), S^d(\cdot), S^s(\cdot)$ | Social surplus/consumer surplus/producer surplus |
| $\varpi^s(\cdot), \varpi^d(\cdot)$ | Unfairness functions |

### 2.2 Definition of the FL Incentive Mechanism Design Problem

With these concepts and notations introduced so far, we present a formal definition for the problem of *FML incentive mechanism design*.

**Definition 1 (FL Incentive Mechanism Design).** *FL incentive mechanism design is to design the optimal* $\boldsymbol{p}^s(\cdot)$, $\boldsymbol{\eta}(\cdot)$, $\boldsymbol{p}^d(\cdot)$, $\boldsymbol{\kappa}(\cdot)$, *as functions of claimed* $\hat{\boldsymbol{d}}$ *and reported* $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\theta}}$, *in order to achieve a set of objectives in Sect. 2.3.*

There are three sources of intertwined information asymmetry, $\hat{\boldsymbol{d}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$, in the FL incentive mechanism design problem. When all variables are considered simultaneously, this problem becomes extremely complicated. As a tradition in the economic literature, we separate this big problem into a demand-side subproblem and a supply-side sub-problem. Formally, we introduce the following assumption:

**Assumption 1 (Separation between Data Supply and Model Demand).** *The data supply market and the model demand market are separated. When an FL participant is both a data provider and a model user, its decision as a data provider does not affect his decision as a model user, or vice versa.*

With Assumption 1, we can define the two subproblems as follows.

**Definition 2 (Supply-Side FL Incentive Mechanism Design).** *Given that the federation Income* $I(Q)$ *and the model quality* $Q(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta})$ *are exogenous functions, the supply-side FL incentive mechanism design is to design the optimal* $p_i(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$ *and* $\eta_i(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$, $i = 1, \ldots, n$, *as functions of claimed datasets* $\hat{d}_i$, $i = 0, \ldots, n-1$ *and reported cost types* $\gamma_i$, $i = 1, \ldots, n$, *in order to achieve some desirable objectives in Sect. 2.3.*

**Definition 3 (Demand-Side FL Incentive Mechanism Design).** *Given that the model quality* $Q$ *is an exogenous constant, the demand-side FL incentive mechanism design is to design the optimal* $p_j(\hat{\boldsymbol{\theta}})$ *and* $\kappa_j(\hat{\boldsymbol{\theta}})$, $j = 1, \ldots, m$, *as functions of reported benefit types* $\hat{\theta}$, $j = 1, \ldots, m$, *in order to achieve some desirable objectives in Sect. 2.3.*

### 2.3 Objectives of FL Incentive Mechanism Design

Below is a list of desirable properties of FL incentive mechanism design. For detailed explanations of these objectives, refer to [10].

- *(Incentive Compatibility, IC)* IC is attained if in equilibrium, all participants report their types truthfully, i.e., $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$. Different types of equilibriums correspond to different IC conditions, which can be one of *Nash Incentive Compatibility (NIC)*, *Dominant Incentive Compatibility (DIC)*, *Baysian Incentive Compatibility (BIC)*, or *Perfect Bayesian Incentive Compatibility (PBIC)*.
- *(Individual Rationality, IR)* A mechanism is *individually rational (IR)* if this mechanism does not make any player worse off than if he quits the federation, i.e.,

$$u_i(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) \geq 0, \forall i \in N \quad \text{and} \quad u_j(\hat{\boldsymbol{\theta}}) \geq 0, \forall j \in M. \tag{1}$$

  In games of incomplete information, IR can be ex-ante IR, interim IR or ex-post IR.

– *(Budget Balance, BB)* A mechanism is *weakly budget balanced (WBB)* if for all feasible outcomes, the sum of payments is less than or equal to zero. i.e.,

$$\sum_{l=1}^{n+m-1} p_l(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}) \leq 0, \forall \hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}. \tag{2}$$

It is *strongly budget balanced (SBB)* if the equity holds. In games of incomplete information, BB can be ex-ante BB, interim BB or ex-post BB.
– *(Social Optimization) social optimization* can be *social surplus maximization (SSM)* when social surplus is maximized or *profit maximization (PM)* if the federation profit is maximized. Social surplus maximization implies *allocative efficiency (AE)*.
– *(Fairness)* We desire to minimize the unfairness function.

## 3 Specifications of FL Incentive Mechanisms

### 3.1 Non-standard Game Settings

Besides game settings in Sect. 2.2, several non-standard game settings also need to be specified when we design FL incentive mechanisms. These non-standard game settings include:

– *(Level of Information Asymmetry)* On the demand side, there may be or may not be *information asymmetry* on valuation types. On the supply side, there may be a) no information asymmetry, b) information asymmetry on datasets only, c) about cost types only, or d) about both.
– *(Mode of System Evolution)* If the FL game is played for only once, it corresponds to a static mechanism. If the FL game is played repeatedly or parameters change over time, it corresponds to a dynamic mechanism.
– *(Belief Updates)* In a dynamic FL game, as time passes by, agents update their beliefs based on *heuristic belief updates* or *Bayesian belief updates*, according to which agents update their information based on some heuristic rules or Bayesian rules, respectively.
– *(Controllable Parameters)* The FL coordinator may determine $\boldsymbol{p}^s(\cdot)$, $\boldsymbol{\eta}(\cdot)$, $\boldsymbol{p}^d(\cdot)$, and $\boldsymbol{\kappa}(\cdot)$,but in some situations, some of these parameters are not controllable. For example, it may not be possible to set up access control on the FML model so that $\boldsymbol{\kappa}(\cdot)$ is not controllable, or it may not be possible to reject datasets offered by data providers so that $\boldsymbol{\eta}(\cdot)$ is not controllable. Also, there are cases where *price discrimination* is not possible so that the unit price of data/model services has to be the same for all data providers/model users.
– *(Functional Forms)* On the supply side, exact forms of the federation income function $I(Q)$, the model quality function $Q(\boldsymbol{d})$, and the individual cost functions $c(d_i, \gamma_i)$ need to be specified. On the demand side, the form of the individual valuation function $w(\kappa_j Q, \theta_j)$ need to be specified.

### 3.2 Measures of Objectives

As a general rule, objectives in Sect. 2.3 cannot be attained simultaneously, so we would like to know how well each objective is achieved by a given FL incentive mechanism. There are also cases where the constraints of some objectives are approximately achieved. The performance of an FL incentive mechanism on achieving these objectives can be evaluated according to the following measures.

- *(Data offering rate, DOR)* DOR is defined as the total data offered by all data providers to the total data owned by all data providers, i.e.,

$$DOR = \frac{\sum_{i=0}^{n-1} \hat{d}_i}{\sum_{i=0}^{n-1} \bar{d}_i}. \tag{3}$$

  The data offering rate varies from 0.0 to 1.0, with 1.0 indicating all data being offered. When a payment scheme is incentive-compatible, the data offering rate is 1.0.
- *(Individual rationality index, IRI)* Rational data providers are not expected to stay in the federation if their costs cannot be covered by payments.
  The individual rationality indicator $IR_i$ for data provider i is defined as $IR_i = 1$ if $p_i - c_i \geq 0$ and $IR_i = 0$ otherwise.
  The ideal case is that the payment scheme satisfies individual rationality for all participants. For general cases, we measure the individual rationality index (IRI), defined as the average of individual rationality indicators, i.e., $IRI = \sum_{i=0}^{n-1} w_i \times IR_i$, where $w_i, i \in N$ are user-defined weights for the relative importance of data owner I, e.g., $w_i = \frac{d_i}{\sum_{l=0}^{n-1} d_l}$.
  The individual rationality index varies from 0.0 to 1.0, with 1.0 indicating individual rationality constraints satisfied for all participants.
- *(Budget surplus margin, BSM)* The budget surplus is the difference between the total income received from model users and the total payments paid to data owners. In practice, the budget surplus is the profit made by the coordinator. Budget surplus margin is the ratio of the budget surplus to total revenue of the federation, i.e.,

$$BSM = \frac{-\sum_{j=n}^{n+m-1} p_j - \sum_{i=0}^{n-1} p_i}{-\sum_{j=n}^{n+m-1} p_j}. \tag{4}$$

  The budget surplus margin varies from $-\infty$ to 1.0, with 0.0 indicating a break-even point, and positive/negative values indicating net profits/losses, respectively.
- *(Efficiency Index, EI)* In federated learning, allocative efficiency is achieved when social surplus is maximized. The efficiency index(EI) is the ratio of realized social surplus to the maximum possible social surplus, i.e.,

$$EI = \frac{S(\hat{\boldsymbol{d}}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\eta}(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}), \boldsymbol{\kappa}(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}))}{\max_{\boldsymbol{\eta}, \boldsymbol{\kappa}} S(\bar{\boldsymbol{d}}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\kappa})} \tag{5}$$

  EI varies from $-\infty$ to 1.0, with 1.0 indicating allocative efficiency.

– *(Fairness Index, FI)* In FL, we want the payment of a unit of contributed data to be the same for all data providers. We set the unfairness function to be the variance of the normalized unit price, i.e., rescaled to $[0.0, 1.0]$, i.e.,

$$\varpi(\boldsymbol{p}^s, \boldsymbol{d}) = Var\{\frac{p_i/d_i}{\sum_{i=0}^{n-1} p_i / \sum_{i=0}^{n-1} d_i}\}. \tag{6}$$

The normalized unit price is invariant with the change of measure.
The fairness index (FI) is the realized unfairness function rescaled to $[0.0, 1.0]$, i.e.,

$$FI = \frac{1}{1 + \varpi(\boldsymbol{p}^s, \boldsymbol{d})}, \tag{7}$$

which varies from 0.0 to 1.0, with 1.0 indicating the absolute fairness.

**Table 2.** Checklist for specifications of FL incentive mechanisms

| | Game settings | Specifications |
|---|---|---|
| Demand-side settings | Information asymmetry on valuation types | Yes/No |
| | Access permission control on FL model | Yes/No/Partial |
| | Price discrimination | Yes/No |
| | Specification of individual valuation functions | Specification |
| Supply-side settings | Information asymmetry on datasets | Yes/No |
| | Information asymmetry on cost types | Yes/No |
| | Ability to reject data owners | Yes/No/Partial |
| | Price discrimination | Yes/No |
| | Specification of individual cost functions | Specification |
| | Specification of the federation income function | Specification |
| | Specification of the model quality function | Specification |
| Other settings | Mode of system evolution | Static/Dynamic |
| | Belief updates | No/Heuristic/Bayesian |
| | Objectives | Measures |
| Objectives | IC | Data offering rate |
| | IR | Individual rationality index |
| | BB | Budget Surplus Margin |
| | Social optimization | Efficiency index |
| | Fairness | Fairness index |

### 3.3 A Checklist for FL Incentive Mechanisms

Designing FL incentive mechanisms often requires deep knowledge in game theory, a field unfamiliar to most FL practitioners. Nevertheless, for FL practitioners to apply an FL incentive mechanism, they only need to make sure the game settings of the targeted mechanism is a good approximation of the real-world scenario. Besides, they would like to know how well the mechanism achieves the objectives listed in Sect. 2.3. We recommend that a checklist of specifications, e.g., Table 2, is provided with every FL incentive mechanism so that FL practitioners can easily choose the right mechanism without understanding the inner workings of these mechanisms.

### 3.4 Benchmarks for Choosing FL Incentive Mechanisms

When choosing FL incentive mechanisms, simpler game settings and fewer objectives are preferred. There is well-known tension between the multiple objectives listed in Sect. 2.3 [8]. We can prove that when game settings become more complicated or more objectives are optimized, the expected social surplus attained by the optimal mechanism is reduced. Formally, we have the following benchmark theorems. For proofs of these theorems, refer to [4].

**Theorem 1 (More controllable parameters, better social optimization).** *The more parameters can be controlled by the FL coordinator, the larger is the expected social surplus attained by the optimal FL incentive mechanism.*

**Theorem 2 (Less information asymmetry, better social optimization).** *When the IC constraint is concerned, the more accurate is the prior belief on $\bar{\boldsymbol{d}}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, the larger is the expected social surplus attained by the optimal FL incentive mechanism.*

**Theorem 3 (More constraints, worse social optimization).** *The more constraints (such as IC, IR, BB), the smaller is the expected social surplus attained by the optimal FL incentive mechanism.*

According to these theorems, it would always be helpful if the FL coordinator can better estimate the datasets and type parameters of FL participants. Also, objectives in Sect. 2.3 compete with each other. If an objective is not a concern for an FL scenario, the FL coordinator should not choose an FL incentive mechanism optimized for that objective.

## 4 A Demand-Side FL Incentive Mechanism - Crémer-McLean Mechanism

In this section and the next section, we provide two examples of FL incentive mechanisms, on the demand side and the supply side, respectively.

### 4.1 Crémer-McLean Theorem

The demand-side mechanism introduced in this section is an application of the famous Crémer-McLean mechanism [5]. In order to apply Crémer-McLean mechanism, we put two assumptions on the prior distribution Prior($\boldsymbol{\theta}$). For more discussions on these assumptions, refer to [2].

**Assumption 2 (Crémer-McLean condition).** *The prior distribution of $\boldsymbol{\theta}$ satisfies the "Crémer-McLean condition" if there are no $j \in M$, $\theta_j \in \Theta_j$ and $\lambda_j : \Theta_j \backslash \{\theta_j\} \mapsto \mathbb{R}_+$ for which*

$$Prior(\boldsymbol{\theta}_{-j}|\theta_j) = \sum_{\theta_j' \in \Theta_j \backslash \{\theta_j\}} \lambda(\theta_j') Prior(\boldsymbol{\theta}_{-j}|\theta_j'), \quad \forall \boldsymbol{\theta}_{-j} \in \boldsymbol{\Theta}_{-j}. \tag{8}$$

The Crémer-McLean condition is often referred to as *correlated types*. To understand this, one can understand agent $j$'s belief about other agents' types $\text{Prior}(\boldsymbol{\theta}_{-j}|\theta_j)$ as a vector with as many entries as $\boldsymbol{\Theta}_{-j}$ has elements. Each $\theta_j$ corresponds to such a vector. The Crémer-McLean condition requires that none of these vectors can be written as a convex combination of other vectors. Note that the Crémer-McLean condition is obviously *violated* when agent $j$'s conditional beliefs are independent of his type, i.e., all these vectors are identical.

The assumption of correlated types is reasonable for the FL scenario. It is highly possible that when the FL model brings high value to one model user, it also brings high value to other model users.

**Assumption 3 (Identifiability condition).** *The prior distribution of* $\boldsymbol{\theta}$ *satisfies the "identifiability condition" if, for all other prior distributions* $Prior'(\boldsymbol{\theta}) \neq Prior(\boldsymbol{\theta})$ *such that* $Prior'(\boldsymbol{\theta}) > 0$ *for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*, there is at least one model user* $j$ *and one valuation type* $\theta_j \in \Theta_j$ *such that for any collection of nonnegative coefficients* $\lambda(\theta_j^{'}), \theta_j^{'} \in \Theta_j$*, we have*

$$Prior'(\boldsymbol{\theta}_{-j}|\theta_j) \neq \sum_{\theta_j^{'} \in \Theta_j} \lambda(\theta_j^{'}) Prior(\boldsymbol{\theta}_{-j}|\theta_j^{'}) \tag{9}$$

*for at least one* $\boldsymbol{\theta}_{-j} \in \boldsymbol{\Theta}_{-j}$.

Intuitively, this condition says that for any alternative prior distribution $\text{Prior}'(\boldsymbol{\theta}) > 0$, there is at least one agent and one type of that agent such that this agent cannot randomize over reports in a way that makes the conditional distribution of all other types under $\text{Prior}'(\boldsymbol{\theta})$ indistinguishable from the conditional distribution of all other types under $\text{Prior}(\boldsymbol{\theta})$. In practice, we do not need to worry about this assumption because identifiability is generic in the topological sense, i.e., for almost all prior distributions, we can assume the identifiability condition holds. We have the following proposition, of which the proof can be found in [9].

**Proposition 1 (Genericity of identifiability).** *Suppose there are at least three agents* ($m \geq 3$)*. Also, if* $m = 3$*, then at least one of the agents has at least three types. Then almost all prior distributions* $Prior(\boldsymbol{\theta})$ *are identifiable.*

Provided Assumption 2 and 3, we can guarantee the existence of an interim truthful and interim individual rational demand-side mechanism that attracts full consumer surplus. Here, interim incentive compatibility means truth-telling is superior to other strategies in expectation under the conditional prior distribution of other agents' types, i.e.,

$$\begin{aligned} &\mathbb{E}_{\text{Prior}(\boldsymbol{\theta}_{-j}|\theta_j)}[w(\kappa_j(\theta_j, \boldsymbol{\theta}_{-j})Q, \theta_j) + p_j(\theta_j, \boldsymbol{\theta}_{-j})] \\ \geq &\mathbb{E}_{\text{Prior}(\boldsymbol{\theta}_{-j}|\theta_j)}[w(\kappa_j(\hat{\theta}_j, \boldsymbol{\theta}_{-j})Q, \theta_j) + p_j(\hat{\theta}_j, \boldsymbol{\theta}_{-j})], \quad \forall j \in M, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \hat{\theta}_j \in \Theta_j; \end{aligned} \tag{10}$$

interim individual rationality means the expected utilities of all agents are nonnegative, under the conditional prior distribution of other agents' types, i.e.,

$$\mathbb{E}_{\text{Prior}(\boldsymbol{\theta}_{-j}|\theta_j)}[w(\kappa_j(\theta_j, \boldsymbol{\theta}_{-j})Q, \theta_j) + p_j(\theta_j, \boldsymbol{\theta}_{-j})] \geq 0, \quad \forall j \in M, \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{11}$$

The *Crémer-McLean Theorem* is:

**Theorem 4 (Crémer-McLean Theorem).** *When the Crémer-McLean condition and the identifiability condition hold for Prior($\boldsymbol{\theta}$), for any decision rule $\boldsymbol{\kappa}(\hat{\boldsymbol{\theta}})$, there exists an interim incentive compatible and interim individually rational payment rule $\boldsymbol{p}(\hat{\boldsymbol{\theta}})$ that extracts full consumer surplus, i.e.,* $-\sum_{j=n}^{n+m-1} p_j(\hat{\boldsymbol{\theta}}) = \sum_{j=n}^{n+m-1} w(\kappa_j(\hat{\boldsymbol{\theta}})Q, \theta_j)$.

As an application of this theorem, we can set $\kappa_j(\hat{\boldsymbol{\theta}}) \equiv 1$, i.e., every model user gets full access permission to the FL model. In this case, $w(\kappa_j(\hat{\boldsymbol{\theta}})Q, \theta_j) = w(Q, \theta_j)$, and we can find an interim incentive compatible and interim individually rational payment rule $\boldsymbol{p}(\hat{\boldsymbol{\theta}})$ such that $-\sum_{j=n}^{n+m-1} p_j(\hat{\boldsymbol{\theta}}) = \sum_{j=n}^{n+m-1} w(Q, \theta_j)$.

As an example, consider the following payment rule:

$$p_j(\hat{\boldsymbol{\theta}}) = -w(Q, \hat{\theta}_j) + \beta[\alpha - ln(\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j)], \tag{12}$$

where $\beta$ and $\alpha$ are two constants. We can prove that when $\beta$ is large enough, $p_j(\hat{\boldsymbol{\theta}})$ is interim incentive compatible. To understand this, noticing that if model user $j$ reports a $\hat{\theta}_j$ lower than his true $\theta_j$.

To see this, noticing that the Lagrange equation for model user $j$ to maximize its utility is

$$\begin{aligned} &\frac{\partial}{\partial \text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j)}\{\mathbb{E}_{\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\theta_j)} p_j(\hat{\boldsymbol{\theta}}) + \lambda[\sum_{\hat{\boldsymbol{\theta}}_{-j}} \text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j) - 1]\} \\ =& -\frac{\partial \mathbb{E}_{\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\theta_j)} w(Q, \hat{\theta}_j)}{\partial \text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j)} - \beta \cdot \frac{\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\theta_j)}{\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j)} + \lambda = 0, \end{aligned} \tag{13}$$

where $\lambda$ is the Lagrange multiplier for the constraint $\sum_{\hat{\boldsymbol{\theta}}_{-j}} \text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j) \equiv 1$. When $\beta$ is large enough compared to $\frac{\partial \mathbb{E}_{\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\theta_j)} w(Q, \hat{\theta}_j)}{\partial \text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j)}$, solving the Lagrange equation in Eq. 13 results in

$$\frac{\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\theta_j)}{\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j)} \simeq \frac{\lambda}{\beta}, \forall \hat{\boldsymbol{\theta}}_{-j}. \tag{14}$$

Therefore, $\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\theta_j)$ has to be equivalent to $\text{Prior}(\hat{\boldsymbol{\theta}}_{-j}|\hat{\theta}_j)$, i.e., $\hat{\boldsymbol{\theta}}_{-j} = \boldsymbol{\theta}_{-j}$.

If we set $\alpha = \mathbb{E}_{\text{Prior}(\boldsymbol{\theta})} ln[\text{Prior}(\boldsymbol{\theta}_{-j}|\theta_j)]$, the payment rule in Eq. 12 is ex-ante individual rational and extracts full consumer surplus ex ante. We can use automated mechanism design to find a *Crémer-McLean mechanism* that is also ex-post individual rational and extracts full consumer surplus ex post, as explained in the following sub-section.

### 4.2 Training Crémer-McLean Mechanism

The Crémer-McLean payments can be calculated by automated mechanism design techniques, e.g., refer to [1]. The method presented in this section is

slightly different from that in [1], compared to which our method extracts full consumer surplus ex post instead of ex ante.

The Crémer-McLean payments $\boldsymbol{p}(\boldsymbol{\theta})$ should simultaneously satisfy the three constraints in the following equation set 15, corresponding to ex-post full consumer surplus extraction, interim incentive compatibility and ex-post individual rationality, respectively.

$$\begin{cases} -\sum_{j=n}^{n+m-1} p_j(\boldsymbol{\theta}) = \sum_{j=n}^{n+m-1} w(Q, \theta_j), \ \forall \boldsymbol{\theta}; \\ \sum_{\boldsymbol{\theta}'_{-j}} [w(Q, \theta_j) + p_j(\theta_j, \boldsymbol{\theta}'_{-j})] \text{Prior}(\boldsymbol{\theta}'_{-j}|\theta_j) \geq 0, \ \forall j \in M, \theta_j \in \Theta_j; \\ \sum_{\boldsymbol{\theta}'_{-j}} [p_j(\theta_j, \boldsymbol{\theta}'_{-j}) - p_j(\hat{\theta}_j, \boldsymbol{\theta}'_{-j})] \text{Prior}(\boldsymbol{\theta}'_{-j}|\theta_j) \geq 0, \ \forall j \in M, \theta_j \in \Theta_j. \end{cases} \tag{15}$$

Crémer-McLean Theorem guarantees that there is a solution $\boldsymbol{p}(\boldsymbol{\theta})$ to Eq. 15. In order to find such a solution, we can minimize the following LOSS in Eq. 16, because it is easy to see that $\boldsymbol{p}(\boldsymbol{\theta})$ is a solution to Eq. 15 i.f.f. it minimizes the LOSS in Eq. 16 to 0. With such a LOSS function, we can easily learn the demand-side Crémer-McLean payments by applying standard backpropagation algorithms.

$$\begin{aligned} LOSS = \{ & \sum_{j=n}^{n+m-1} [w(Q, \theta_j) + p_j(\boldsymbol{\theta})] \}^2 \\ & + \sum_{j=n}^{n+m-1} ReLu\{ - \sum_{\boldsymbol{\theta}'_{-j}} [w(Q, \theta_j) + p_j(\theta_j, \boldsymbol{\theta}'_{-j})] \text{Prior}(\boldsymbol{\theta}'_{-j}|\theta_j) \} \\ & + \sum_{j=n}^{n+m-1} ReLu\{ - \sum_{\boldsymbol{\theta}'_{-j}} [p_j(\theta_j, \boldsymbol{\theta}'_{-j}) - p_j(\hat{\theta}_j, \boldsymbol{\theta}'_{-j})] \text{Prior}(\boldsymbol{\theta}'_{-j}|\theta_j) \}, \end{aligned} \tag{16}$$

where $\boldsymbol{\theta}, \boldsymbol{\theta}', \hat{\boldsymbol{\theta}}$ are drawn randomly from the prior distribution of $\boldsymbol{\theta}$.

## 5 A Supply-Side FL Incentive Mechanism - PVCG

As a counterpart of Crémer-McLean mechanism, which is optimal on the demand side, we introduce an optimal supply-side procurement auction in this section. This proposed procurement auction, accompanied by the demand-side Crémer-McLean mechanism, maximizes producer surplus by incentivizing data providers to offer all their data to the federation and truthfully report their cost types. For more discussions on PVCG, refer to [3].

As explained in Sect. 2.3. When designing the supply-side mechanism, we assume the federation income $I(Q)$ and the model quality $Q(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta})$ are exogenous functions. For example, when Crémer-McLean mechanism is adopted on the demand side, we know the federation income is:

$$I(Q) = -\sum_{j=n}^{n+m-1} p_j(\boldsymbol{\theta}) = \sum_{j=n}^{n+m-1} w(\kappa_j(\boldsymbol{\theta})Q, \theta_j), \tag{17}$$

where $\boldsymbol{\theta}$ is assumed to be an exogenous parameter, so we can ignore it when we focus on the supply side. Because the federation income indirectly depends on $Q(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta})$, we also write $I(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}) = I(Q(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}))$.

### 5.1 The Procurement Auction

One can carry out the proposed procurement auction and compute the payments to data providers by following the following steps.

*Step 1. Data providers claim datasets to offer and bid on cost types*

As the first step, every data provider submits a sealed bid for their respective claimed datasets and cost types. The claimed dataset $\hat{d}_i$ is the best dataset that data provider $i$ claims it can offer to federated learning. It may differ from the dataset $\bar{d}_i$ actually owned by data provider $i$. Similarly, the reported cost type $\hat{\gamma}_i$ may differ from the true cost type $\gamma_i$.

*Step 2. The coordinator chooses the optimal acceptance ratios*

Then, the coordinator decides how many data to accept from each data provider. It chooses $d_i \leq \hat{d}_i, i = 0, \ldots, n-1$ that maximize the social surplus. Equivalently, the coordinator calculates the optimal *acceptance ratio* $\eta_i \in [0,1]^{\dim(d_i)} = d_i \oslash \hat{d}_i$ such that $d_i = \hat{d}_i \odot \eta_i$, where $[0,1]$ denotes the interval between 0 and 1.

The optimal acceptance ratios $(\eta_0^*, \ldots, \eta_{n-1}^*) = \boldsymbol{\eta}^*$ are calculated according to the following formula:

$$\begin{aligned} \boldsymbol{\eta}^* &= \operatorname{argmax}_{\boldsymbol{\eta} \in [0,1]^{\dim(x_i) \times n}} \{S(\hat{\boldsymbol{x}} \odot \boldsymbol{\eta}, \hat{\boldsymbol{\gamma}})\} \\ &= \operatorname*{argmax}_{\boldsymbol{\eta} \in [0,1]^{\dim(d_i) \times n}} I(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}) - \sum_{i=0}^{n-1} c_i(\hat{d}_i \odot \eta_i, \hat{\gamma}_i). \end{aligned} \tag{18}$$

Because different $(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$ results in different $\boldsymbol{\eta}^*$, $\boldsymbol{\eta}^*$ is written as $\boldsymbol{\eta}^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$. Correspondingly, the maximum producer surplus is denoted by $S^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) = I(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})) - \sum_{i=0}^{n-1} c_i(\hat{d}_i \odot \eta_i^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\gamma}_i)$.

It is worth noting that although $S^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$ and $S(\boldsymbol{d}, \boldsymbol{\gamma})$ both represent producer surplus, they are different functions. The first parameter $\boldsymbol{d}$ in $S(\cdot)$ is the accepted dataset, whereas the first parameter $\hat{\boldsymbol{d}}$ in $S^*(\cdot)$ is the claimed dataset. $\boldsymbol{d}$ and $\hat{\boldsymbol{d}}$ are related by $\boldsymbol{d} = \hat{\boldsymbol{d}} \odot \boldsymbol{\eta}^*$.

*Step 3. Data providers contribute accepted datasets to federated learning*

In this step, data providers are required to contribute the accepted dataset $\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}^*$ to federated learning. Since in the first step, data provider $i$ has claimed the ability to offer a dataset no worse than $\hat{d}_i$, if it cannot contribute $\hat{d}_i \odot \eta_i^* \leq \hat{d}_i$, we impose a high punishment on it. With the contributed datasets, data providers collaboratively produce the output virtual product, bringing income $I(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}^*)$ to the federation.

*Step 4. The coordinator makes transfer payments to data providers according to the PVCG sharing rule*

In this final step, the coordinator pays data providers according to the PVCG sharing rule. The PVCG payment

$$p_i(\cdot) = \tau_i(\cdot) + h_i^*(\cdot) \tag{19}$$

is composed of two parts, the *VCG payment* $\tau_i$ and the *optimal adjustment payment* $h_i^*$. The VCG payment is designed to induce truthfulness, i.e., the reported capacity limits $\hat{\boldsymbol{d}}$ and reported cost type $\hat{\boldsymbol{\gamma}}$ are equal to the true capacity limits $\bar{\boldsymbol{d}}$ and true cost type $\boldsymbol{\gamma}$. The adjustment payment is optimized so that ex-post individual rationality and ex-post weak budget balancedness can also be attained.

With $\boldsymbol{\eta}^*$ calculated in Step 2, the VCG payment $\tau_i$ to data provider $i$ is:

$$\begin{aligned}\tau_i &= S^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) - S^*_{-i}(\hat{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}) + c(\hat{d}_i \odot \eta_i^*(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\gamma}})) \\ &= [I(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})) - I(\hat{\boldsymbol{d}}_{-i} \odot \boldsymbol{\eta}^{-i*}(\hat{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}))] \\ &\quad - \sum_{k=0,\neq i}^{n-1} [c(\hat{d}_k \odot \eta_k^*(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}), \hat{\gamma}_k) - c(\hat{d}_k \odot \eta_k^{-i*}(\hat{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}, \hat{\boldsymbol{\theta}}), \hat{\gamma}_k)],\end{aligned} \tag{20}$$

where $(\hat{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ denotes the claimed datasets and the reported cost types excluding data provider $i$. $\boldsymbol{\eta}^{-i*}$ and $S^*_{-i}(\hat{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ are the corresponding optimal acceptance ratios and maximum producer surplus. Note that $\boldsymbol{\eta}^{-i*}$ is different from $\boldsymbol{\eta}^*_{-i}$: the former maximizes $S(\hat{\boldsymbol{d}}_{-i} \odot \boldsymbol{\eta}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$, whereas the latter is the component of $\boldsymbol{\eta}^*$ that maximizes $S(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta}, \hat{\boldsymbol{\gamma}})$. $\boldsymbol{\tau} = (\tau_0, \ldots, \tau_{n-1})$ is a function of $(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$, written as $\boldsymbol{\tau}(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$.

The *adjustment payment* $h_i(\hat{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ is a function of $(\hat{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$. The optimal adjustment payments $(h_0^*(\cdot), \ldots, h_{n-1}^*(\cdot)) = \boldsymbol{h}^*(\cdot)$ are determined by solving the following *functional equation* (a type of equation in which the unknowns are functions instead of variables; refer to [11] for more details):

$$\begin{aligned}&\sum_{i=0}^{n-1} \mathrm{ReLu}[-(S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S^*_{-i}(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i})) - h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i})] \\ &+ \mathrm{ReLu}\{\sum_{i=0}^{n-1} [(S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S^*_{-i}(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i})] - S^*(\boldsymbol{d}, \boldsymbol{\gamma})\} \\ &\quad \equiv 0, \qquad \forall (\bar{\boldsymbol{d}}, \boldsymbol{\gamma}) \in \mathrm{supp}(\mathrm{Prior}(\boldsymbol{d}, \boldsymbol{\gamma})),\end{aligned} \tag{21}$$

where $\mathrm{supp}(\mathrm{Prior}(\boldsymbol{d}, \boldsymbol{\gamma}))$ is the *support* of the *prior distribution* $\mathrm{Prior}(\boldsymbol{d}, \boldsymbol{\gamma})$ of the true parameters $(\boldsymbol{d}, \boldsymbol{\gamma})$. Support is a terminology from *measure theory*, defined as $\mathrm{supp}(\mathrm{Prior}(\boldsymbol{d}, \boldsymbol{\gamma})) = \{(\boldsymbol{d}, \boldsymbol{\gamma}) | \mathrm{Prior}(\boldsymbol{d}, \boldsymbol{\gamma}) > 0\}$. In general, there is no closed-form solution to Eq. 21, so we employ neural network techniques to learn the solution, as is explained in the following sub-section.

Through rigorous mathematical derivation, we can prove that with some reasonable assumptions, the PVCG payment rule thus calculated is dominant incentive compatible, allocative efficient, ex-post individual rational, and ex-post weak budget balanced. For detailed proofs of these properties, refer to [3].

### 5.2 Learning the Optimal Adjustment Payments

We can prove that the solution $\boldsymbol{h}^*(\cdot)$ to Eq. 21, if existing, is also a solution to the following minimization problem:

$$\boldsymbol{h}^*(\cdot) = \text{argmin}_{\boldsymbol{h}(\cdot)} \mathbb{E}_{(\bar{\boldsymbol{x}},\boldsymbol{\gamma},\boldsymbol{\theta})}\{\text{LOSS}\}, \tag{22}$$

where the expectation is over the prior distribution of $(\bar{\boldsymbol{x}}, \boldsymbol{\gamma}, \boldsymbol{\theta})$. Here, we bring back the valuation type $\boldsymbol{\theta}$ because we want the adjustment payments applicable to all possible $\boldsymbol{\theta}$. Note that different $\boldsymbol{\theta}$ results in different federation income function $I(\hat{\boldsymbol{d}} \odot \boldsymbol{\eta})$. Hence the maximum producer surplus also depends on $\boldsymbol{\theta}$.

LOSS is defined as

$$\text{LOSS} = \text{Loss1} + \text{Loss2} = 0, \tag{23}$$

where

$$\begin{aligned} \text{Loss1} = \sum_{i=0}^{n-1} \text{ReLu}[-(S^*(\bar{\boldsymbol{x}}, \boldsymbol{\gamma}, \boldsymbol{\theta}) - S^*_{-i}(\bar{\boldsymbol{x}}_{-i}, \boldsymbol{\gamma}_{-i}, \boldsymbol{\theta})) \\ - h_i(\bar{\boldsymbol{x}}_{-i}, \boldsymbol{\gamma}_{-i}, \boldsymbol{\theta})] \quad \text{and} \end{aligned} \tag{24}$$

$$\begin{aligned} \text{Loss2} = \text{ReLu}[\sum_{i=0}^{n-1} [(S^*(\bar{\boldsymbol{x}}, \boldsymbol{\gamma}, \boldsymbol{\theta}) - S^*_{-i}(\bar{\boldsymbol{x}}_{-i}, \boldsymbol{\gamma}_{-i}, \boldsymbol{\theta})) \\ + h_i(\bar{\boldsymbol{x}}_{-i}, \boldsymbol{\gamma}_{-i}, \boldsymbol{\theta})] - S^*(\bar{\boldsymbol{x}}, \boldsymbol{\gamma}, \boldsymbol{\theta})]. \end{aligned} \tag{25}$$

This fact informs us that we can learn the optimal adjustment payments $\boldsymbol{h}^*(\cdot)$ by minimizing the expected LOSS function. Also, we know neural networks can approximate arbitrary continuous functions to arbitrary precisions [7]. Therefore, we construct $n$ neural networks $\text{NET}_i^h, i \in N$ to approximate $h_i(\cdot), i \in N$. Output nodes of these $n$ networks, denoted by $\text{NET}_i^h.o, i \in N$, are combined into a single *composite neural network* in Fig. 2 with the loss function in Eq. 23–25.

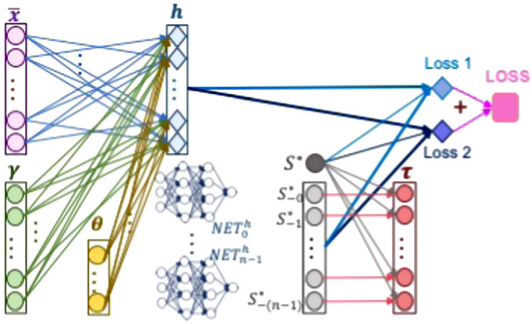


**Fig. 2.** The structure of the composite neural network of PVCG

The training data $(\bar{\boldsymbol{d}}^t, \boldsymbol{\gamma}^t, \boldsymbol{\theta}^t), t = 0, 1, \ldots, T$ are drawn from their prior distribution Prior$(\bar{\boldsymbol{d}}, \boldsymbol{\gamma}, \boldsymbol{\theta})$ and $T$ is the sample size. For the $t$th sample, $\boldsymbol{\tau}^t = \boldsymbol{\tau}(\bar{\boldsymbol{d}}^t, \boldsymbol{\gamma}^t, \boldsymbol{\theta}^t)$, $S^{*t} = S^*(\bar{\boldsymbol{d}}^t, \boldsymbol{\gamma}^t, \boldsymbol{\theta}^t)$, and $S^{*t}_{-i} = S^*(\bar{\boldsymbol{d}}^t_{-i}, \boldsymbol{\gamma}^t_{-i}, \boldsymbol{\theta}^t)$. Since we only need synthetic data to train this network, we can generate as many data as needed. As a result, LOSS can be minimized to its theoretical minimum almost perfectly in experiments.

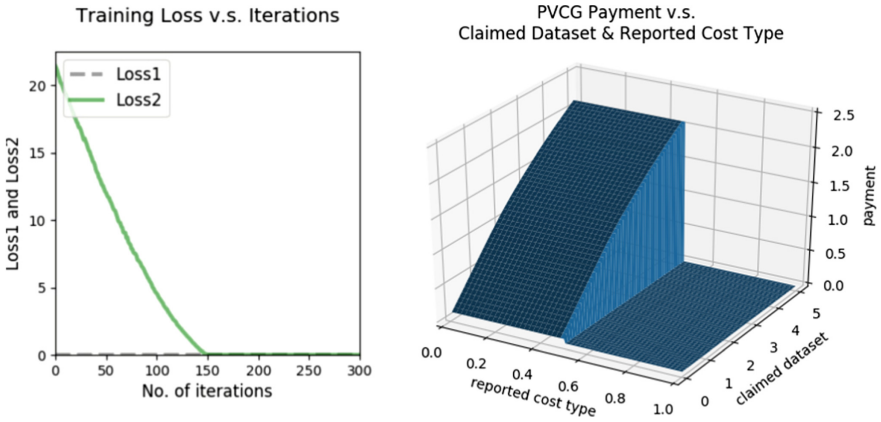


**Fig. 3.** Training loss v.s. iterations (left) and PVCG payment v.s. reported capacity limit & reported cost type (right)

To illustrate the effectiveness of this neural network method, we learned the adjustment payments for a hypothetical scenario. We set the individual valuation functions and individual cost functions as follows:

$$v(\boldsymbol{d}) = \theta_i \sqrt{n(\sum_{k=0}^{n-1} d_k)} \text{ and } c_i(d_i, \gamma_i) = \gamma_i d_i, i \in N. \tag{26}$$

We report the experiment results for $n = 10, m = 2$, Prior$(\bar{x}_i) =$ Uniform$[0, 5], i \in N$, Prior$(\gamma_i) =$ Uniform$[0, 1], i \in N$, and Prior$(\theta_j) = [0, 1], j \in M$. We let $\text{NET}^h_i, i \in N$ each have three 10-dimensional hidden layers.

The loss curve is shown in the left figure of Fig. 3. The training loss fast converges to 0, as expected. After we obtain the trained networks $[\text{NET}^h_i], i \in N$, we can use trained networks to calculate PVCG payments $\boldsymbol{p}(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ for any reported $(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$. For illustration, we draw $p_0$, the payment to data provider 0, with respect to $\hat{d}_0$ and $\hat{\gamma}_0$ in the right figure in Fig. 3, fixing parameters of other participants at $\hat{d}_i \equiv 2.5, \hat{\gamma}_i \equiv 0.5, i \in N, \neq 0, \hat{\theta}_j \equiv 0.5, j \in M$.

We can see that $p_0$ increases with $\hat{d}_0$. This indicates that the more data a data provider claim, the more data are accepted from this data provider; hence, it receives higher payments. Also, $p_0$ remains constant with $\gamma_0$ when $\gamma_0$ is below a threshold and sharply drops to around 0 when $\gamma_0$ passes the threshold.

This implies that the payment to a data provider should only be affected by its contribution to the federated learning process rather than its cost, but if a data provider's cost is too high, the optimal social choice is to exclude this data provider from the federation and thus pay it nothing.

## 6 Summary

In this chapter, we set up a game-theoretic framework for studying FL incentive mechanisms. We introduced the key concepts, mathematical symbols, definitions, and key assumptions that are necessary for readers to understand the FL incentive mechanism design problem and its objectives. Then, we suggest breaking down the original complicated problem into two sub-problems: a demand-side problem and a supply-side problem. We provide a checklist for FL practitioners to quickly understand the specifications and objectives of any given FL incentive mechanism so that real-world FL practitioners can choose the most appropriate mechanism without understanding its internal workings.

As examples, we introduced two FL incentive mechanisms designed under our proposed framework: the Crémer-McLean mechanism on the demand side and a VCG-based procurement auction, PVCG, on the supply side. These mechanisms both guarantee truthfulness, i.e., they encourage participants to truthfully report their private information and offer all their data to the federation. The Crémer-McLean mechanism, together with PVCG, attains allocative efficiency, individual rationality, and weak budget balancedness at the same time, easing the tension between these objectives.

## References

1. Albert, M., Conitzer, V., Lopomo, G.: Assessing the robustness of Cremer-McLean with automated mechanism design. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
2. Börgers, T., Krahmer, D.: An Introduction to the Theory of Mechanism Design. Oxford University Press, Oxford (2015)
3. Cong, M., Weng, X., Yu, H., Qu, J., Yiu, S.M.: Optimal procurement auction for cooperative production of virtual products: Vickrey-Clarke-Groves Meet Crémer-McLean. CoRR. arXiv:2007.14780 (2000)
4. Cong, M., Weng, X., Yu, H., Qu, Z.: FML incentive mechanism design: concepts, basic settings, and taxonomy. In: the 1st International Workshop on Federated Machine Learning for User Privacy and Data Confidentiality (FL-IJCAI 2019) (2019)
5. Crémer, J., McLean, R.P.: Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdepen0 denty. Econometrica **53**, 345–361 (1985)
6. EU: Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. Off. J. Eur. Union (OJ) **59**(1–88), 294 (2016)

7. Funahashi, K.I.: On the approximate realization of continuous mappings by neural networks. Neural Netw. **2**(3), 183–192 (1989)
8. Jackson, M.O.: Mechanism theory. Available at SSRN 2542983 (2014)
9. Kosenok, G., Severinov, S.: Individually rational, budget-balanced mechanisms and allocation of surplus. J. Econ. Theory **140**(1), 126–161 (2008)
10. Narahari, Y.: Game Theory and Mechanism Design, vol. 4. World Scientific, Singapore (2014)
11. Rassias, T.: Functional Equations and Inequalities, vol. 518. Springer, Dordrecht (2012). https://doi.org/10.1007/978-94-011-4341-7
12. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: concept and applications. ACM Trans. Intell. Syst. Technol. (TIST) **10**(2), 12 (2019)