



WallStreetFeds: Client-Specific Tokens as Investment Vehicles in Federated Learning

Arno Geimer*
SnT, University of Luxembourg
Luxembourg
arno.geimer@uni.lu

Beltran Fiz
SnT, University of Luxembourg
Luxembourg
beltran.fiz@uni.lu

Radu State
SnT, University of Luxembourg
Luxembourg
radu.state@uni.lu

Abstract

Federated Learning (FL) is a collaborative machine learning paradigm which allows participants to collectively train a model while training data remains private. This paradigm is especially beneficial for sectors like finance, where data privacy, security and model performance are paramount. FL has been extensively studied in the years following its introduction, leading to, among others, better performing collaboration techniques, ways to defend against other clients trying to attack the model, and contribution assessment methods.

An important element in for-profit Federated Learning is the development of incentive methods to determine the allocation and distribution of rewards for participants. While numerous methods for allocation have been proposed and thoroughly explored, distribution frameworks remain relatively understudied.

In this paper, we propose a novel framework which introduces client-specific tokens as investment vehicles within the FL ecosystem. Our framework aims to address the limitations of existing incentive schemes by leveraging a decentralized finance (DeFi) platform and automated market makers (AMMs) to create a more flexible and scalable reward distribution system for participants, and a mechanism for third parties to invest in the federation learning process.

CCS Concepts

• **Computing methodologies** → **Cooperation and coordination**.

Keywords

Federated Learning, Decentralized Finance, Automated Market Makers, Incentives, Assets

ACM Reference Format:

Arno Geimer, Beltran Fiz, and Radu State. 2024. WallStreetFeds: Client-Specific Tokens as Investment Vehicles in Federated Learning. In *5th ACM International Conference on AI in Finance (ICAIF '24)*, November 14–17, 2024, Brooklyn, NY, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3677052.3698653>

*Corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.

ICAIF '24, November 14–17, 2024, Brooklyn, NY, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1081-0/24/11
<https://doi.org/10.1145/3677052.3698653>

1 Introduction

Data has emerged as a critical asset in today's digital economy. It drives innovation and enables the rapid advancement of machine learning algorithms. As organizations increasingly rely on data to gain competitive advantages, the ability to collect, analyze, and create collaborative environments has become a significant challenge. Federated Learning (FL) addresses this issue by allowing multiple participants to collaboratively train a machine learning model without the need to share their raw data. Originally conceived by [15] at Google Research, in a classic FL process involves multiple rounds of learning, where participants train a local model using their own data and then share model updates with a central server, where they are aggregated. One of the key advantages of FL is that by keeping data on the devices, sensitive information remains localised, mitigating the risks associated with the new attack vectors a centralised solution could expose all participants to, such as data breaches and unauthorised access. Additionally, FL models often exhibit superior generalizability compared to their centralised counterparts [8].

The financial sector, with its stringent data privacy regulations and sensitive information, stands to benefit immensely from FL. Applications such as fraud detection [35], credit risk assessment [11], and compliance with regulatory frameworks [21] can be significantly enhanced through secure and collaborative model training. However, the adoption and sustained success of Federated Learning in finance will depend on overcoming unique challenges related to privacy, security, and participation incentives.

Privacy and security have been a core focus in FL research from its inception. Early work focused primarily on deploying differential privacy techniques to protect individual data contributions within the aggregated model updates [27], later evolving to encompass a larger scope and attempting to safeguard the entire training process, such as secure multi-party computation and homomorphic encryption [3].

From the economic perspective, the incentives for being in a federation are crucial to motivate client participation. While traditionally the benefits of the global model are shared amongst the participants, the improvements from the global model might not be enough to justify the capital investment required to partake in the federation. In fact recent studies have shown that one of the primary challenges to lasting federated learning partnerships is the initial upfront costs and incentives to remain in a federation for long durations [2].

Blockchain technology and decentralised finance (DeFi) have revolutionised revenue generation models for new companies. By leveraging decentralised ledgers, new businesses can create and manage digital tokens while smart contracts automate and secure transactions, reducing operational costs. These can be offered in

the form of Initial Coin Offerings (ICOs), providing early-stage funding without traditional venture capital constraints. The transparency and security inherent in blockchain technologies also serve to attract a wider range of investors.

Our research aims to address the gap in participation incentives by leveraging on the infrastructure in place provided by DeFi platforms. We propose a federated learning framework that allows participating clients to emit a series of tokens to represent their contributions to a federation's learning process. In doing so, we are able to reduce the barrier to entry for data producers by providing them with alternate means to raise capital.

In this work, we introduce a novel reward payout scheme that enables federated learning participants to monetize their future contributions to a collaboratively trained model. Additionally, we propose an innovative investment vehicle for external actors that emphasizes the learning effectiveness of participants' data.

This paper is structured as follows: Section 2 provides a background on Federated Learning and DeFi. Section 3 goes over related work relevant to our proposed framework. In Section 4, we discuss the limitations of current incentive schemes. Section 5 introduces our tokenized reward-sharing framework, followed by an evaluation and discussion. Finally, Section 6 concludes the paper with some final thoughts and future research directions.

2 Background

This section provides an explanation of Federated Learning as well as its application in different areas. We take a closer look at payment types and various reward distribution and contribution techniques.

2.1 Federated Learning

Federated Learning is a collective machine learning technique allowing privacy-preserving training of a joint model. A federation consists of a central authority, or server, and a collection of clients. After deciding on a model type and parameters, the federation framework and the common task, the central server initiates a model and distributes its weights to the participants. These train the model on their respective dataset and set the updated weights back to the server. The server aggregates the weights, based on a previously decided aggregation strategy, and returns the new model to the clients. This process is iterated until convergence is reached.

Algorithm 1 showcases a standard Federated Learning process. ClientUpdate represents the specific training process used by the clients, which involves minimizing the loss function. ServerUpdate is the aggregation function employed by the central server. It is heavily influenced on the needs and type of federation. Aggregations designed to reduce communication costs by only polling model updates from a few clients have been proposed [5]. Other techniques speed up convergence by giving more weight to specific clients, or by using superior optimization algorithms [20][7]. More stable convergence can be obtained by changing the aggregation function type [31].

2.1.1 Federation types. While the FL process stays the same across the board, due to differences in setups and variables such as the number of clients and the amount of data, two different types of federations have emerged:

Algorithm 1 A Federated Learning framework with K clients.

```

On server: Initialize  $\omega_1$ 
1: for each round  $t = 1, 2, \dots, T - 1$ : do
2:   On server:
     Distribute  $\omega_t$  to the clients
3:   On client  $k$ :
      $\omega_{t+1}^k \leftarrow \text{ClientUpdate}(k, \omega_t)$ 
     Send  $\omega_{t+1}^k$  to the server.
4:   On server:
      $\omega_{t+1} \leftarrow \text{ServerUpdate}(\omega_{t+1}^1, \dots, \omega_{t+1}^K)$ 
5: end for
6: The final model  $\omega_T$  is kept as the federated model.

```

Cross-device federations contain a vast number of clients with sparse datasets. Computational capacities on participants are usually limited. Examples might include IoT devices [17], satellites [14] and electrical meters [24]. The primary challenge is motivating the large number of participants to join and remain in the federation despite the limited computational resources and potential costs related to energy consumption and network usage [10].

Cross-silo federations consist of few participants, typically companies or institutions with extensive datasets, such as hospitals, banks and similar organizations. In these federations, communication costs and computational capabilities are of minor importance, since the participating entities are usually well equipped to handle these problems. However, as participants may be competing entities, privacy preservation as well as security from malicious attacks are of utmost importance. It has been shown [6] that in federated settings without adequate security measures, nosy participants can gain information about other clients' datasets and even singular data samples. Even though a federated model is a net benefit to most actors, privacy concerns may lead to companies refraining from joining a federation. Thus, besides privacy assurances, incentives play a critical role in convincing participants to join a cross-silo federation[9]. Although they can play a role in cross-device federations [10], incentive mechanisms are more commonly applied in cross-silo settings.

2.1.2 Model types and application domains. Federated Learning was introduced as a method of collaboratively training convolutional neural networks for image recognition and LSTMs for language model. Since then, Federated Learning has seen usage in a wide are of domains. Those include tabular data modeling using federated versions of XGBoost [29], used by WeBank among others [12]. Privacy issues have been addressed with the inclusion of homomorphic encryption into Federated Learning[34].

2.2 Smart Contracts

Blockchain based smart contracts became popular through their deployment in the Ethereum blockchain [4]. They are executed by validator nodes with the assurance of the underlying consensus protocol, verifying and enforcing the terms of the smart contract directly in a trustless and transparent manner. The currency used is Ether, which can be utilized to transfer wealth when sent to other accounts or to execute a smart contract. These smart contracts are usually developed in a higher-level language like Solidity, which

is then converted to bytecode and stored on the blockchain. This bytecode is interpreted by the Ethereum Virtual Machine (EVM), which executes the smart contract whenever a transaction interacts with the contract account.

Smart contracts have paved the way for a wide range of decentralized applications that leverage the immutable and decentralized nature of the blockchain. For instance, in the financial sector, smart contracts are used to create DeFi platforms that offer services like lending, borrowing, and trading without traditional financial intermediaries.

3 Related Work

We now explore the work related to our proposed framework. We investigate incentive types in Federated Learning, how contributions may be calculated, and we introduce Automated Market Makers, Decentralized Finance-based exchange platforms.

3.1 Federated Learning Incentive frameworks

3.1.1 Reward Types. While usually not mentioned as a specific aspect of FL in literature, different types of reward have emerged depending on the type of setting and federation. Most common are value-based rewards, either monetary or cryptocurrency-based[18], arising in federations where the central server owns the model and wants to use data owners to improve its performance. These rewards allow participants to be paid before, during and/or at the end of the federation run. Incentive mechanisms form the core dynamics of reward distribution between the server and clients.

3.1.2 Client-pricing incentive mechanism. Client-pricing mechanism allow clients to sell model updates to the server. Since clients are independent entities who make their own decisions, game theory can be used to either maximise the overall payoff of the federation or that of individual clients. Depending on the nature of the federation, clients may use different strategies:

During a **coalitional game**, clients cooperate to maximise the global model performance, entering into contracts between each other and coordinating strategies.

In a **Stackelberg game**, specific clients, the followers, wait for other clients, the leaders, to make their move before deciding on a pricing strategy. The leaders in a Stackelberg game choose their strategy to maximize their profit, while followers aim to gain the maximum utility given the leaders' chosen strategy.

Within a **Non-cooperative game**, players do not care about the outcome of the federation, they are only interested in maximizing their own profit.

An **auction-based** incentive scheme [25] [33] comprises clients determining the cost of being involved in training, and making bids for their participation to the central server. The central server then determines winners, includes them in training, and pays them when finished. Auctions can take on different forms[26], including *open-cry auctions* where bids are public and *sealed-bid auctions* where bids are private.

A **performance-based** incentive mechanism was introduced in [13]. The authors' protocol encourages participants to exchange model updates for tokens, which they can use to buy other model updates for aggregation. A participant who regularly sells their

model update up for trade will collect more tokens and attain a better performing final model.

3.1.3 Server-based incentive mechanisms. Opposed to these client-pricing incentive mechanisms are server-based incentive mechanisms, in which the central server distributes a value-based reward to participants. Here, clients do not price and sell their involvement in the federation, they get paid to be part of the federation. How the reward is distributed is to be decided during the setup of the federation:

A **linear** reward model pays out based on a participant-specific quantity, such as the dataset size of participants. This value might already be known by the central server, as it is used in aggregation strategies as a means to weighting model updates when aggregating.

An **equal** reward model hands out equal rewards to each participant. While this settles disagreements about the fairness of contribution calculations between participants, it may discourage potential substantial clients from joining, as the federation might not be able to cover their costs.

A **contribution-calculation based** incentive model has the server calculate the contribution of a client to the federation and distributes rewards based on this assessment. A basic method of contribution assessment would be the performance of clients' model updates. However, more extensive contribution methods have been proposed, of which the most popular are Shapley values [22]. Given a federation with a set of clients C , the Shapley value ϕ_i of client i can be calculated as

$$\phi_i = \sum_{S \subseteq C \setminus i} \frac{|S|!(C - |S| - 1)!}{C!} (f(\omega_{S \cup i}) - f(\omega_S)),$$

where f is the utility of the model, usually the accuracy or loss on a test set, and ω_S is the combined model of the subset of clients S [23]. Shapley values do necessitate a local test set on the server, and additionally introduce a substantial computational effort, as the amount of per-value computations increases exponentially with the number of clients.

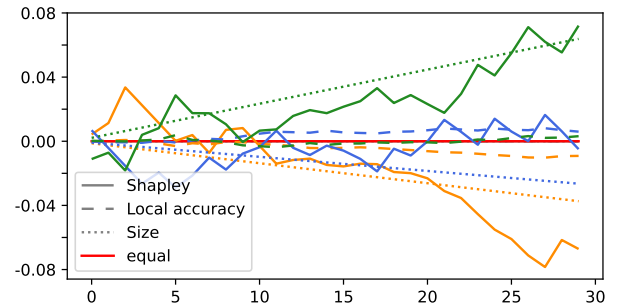


Figure 1: Per-client percentage of reward received as compared to equal payouts, with various contribution calculation methods, on CIFAR-10. Colors indicate different clients.

Figure 1 showcases the cumulative per-round percentages of rewards using different contribution calculations, compared to an

equal payout which distributes the payment evenly between all clients each round. As can be seen, client rewards are more volatile when using Shapley values, allowing clients to collect more rewards depending on the quality of their model updates.

However a contribution technique like Shapley values would accurately reflect utility differences in non-independent and identically (Non-IID) distributed data sources.

3.2 Automated Market Makers

On a traditional exchange market, buyers and sellers can place orders to trade assets. The order book is public information, allowing third parties to gain an insight into the valuation and demand of assets. To match orders, market makers manage the exchange, profiting off the spread. However, in markets with low liquidity, poor availability of assets leads to wide spreads, preventing market makers to properly match orders, resulting in no trade. Thus, in Decentralized Finance (DeFi) applications, automated market makers (AMM) were proposed. Instead of requiring traders to set buy and sell orders and managing the spread, AMMs determine prices algorithmically, allowing for faster and smoother execution of trades.

The most basic type of AMM allows for the exchange of two currencies [16]. Liquidity providers offer a supply of tokens to the AMM, which are pooled and used in any future trades. In such a two-currency AMM, let the currencies be X and Y , with x and y the amount of each currency held by the market maker. The AMM's pricing algorithm determines the value of an asset based on the trading function $f(x, y) = k$ for an arbitrary predetermined value k . Some examples of such trading functions are the multiplicative trading function $f(x, y) = x * y$ or the additive function $f(x, y) = x + y$. When a trader wants to exchange X for Y , the trading function determines how many tokens of Y the trader receives. Meaning, if a trader wants to exchange m amount of X for n amount of Y , the trade is accepted if $f(x, y) = f(x + m, y - n) = k$. It follows that demand drives prices, as an increase in buyers leads to an increase in availability of a token, which results in a higher exchange rate.

Constant mean market makers (CMMM) allow for trading of multiple assets on one single exchange [1]. Their trading function f is the geometric mean of the amount of their n held currencies, i.e. $f(x_1, \dots, x_n) = \prod_{i=1}^n x_i^{\omega_i}$, with $\sum_i \omega_i = 1$. Currency n , the baseline currency, is called the numeraire. It is usually denoted in US Dollars (\$) or Euros (€ or EUR), and its value is always 1. Initial values are reflected in the trading formula, and determined at the creation of the AMM. Ultimately, the AMM scalability is limited by the underlying blockchain protocol [28].

In summary, an automated market maker owns a pool of currencies, and facilitates trading by using a trading function to determine exchange rates. This effectively reduces friction and enhances the overall liquidity of the platform.

4 Limitations of current incentive schemes

Although current Federated Learning incentive frameworks introduce state of the art valuation mechanisms to FL, the implementation and execution of payouts is understudied. Incentivizing participants to join and stay in the federation is still challenging due to, among others, the inflexibility of current payout mechanisms.

This section goes into more details on which problems may arise and their consequences.

4.1 Closed flow of capital

All of the previously mentioned value-based incentive mechanisms are closed systems, as shown in Figure 2: on the basis of a chosen incentive mechanism, monetary transactions are executed directly between the central server and clients. External backers need to enter into a funding agreement with a client or the server directly, which leads to regulatory challenges, necessitating agreement talks and setting up contracts. Such closed incentive schemes severely restrict the inflow of capital to the federation.

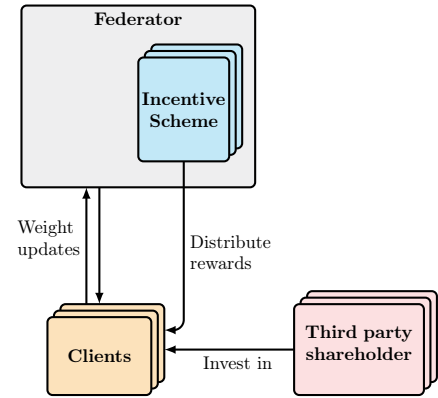


Figure 2: An illustration of incentive schemes, closed systems where third party investments have to be negotiated directly with the client.

Lack of investment can have severe consequences, not only for the participants, but for the federation as a whole: Being underfunded, clients may drop out of the federation, leading to the loss of valuable information gained from their data. The shared model decreases in performance, and loses in value, which diminishes rewards. This can result in a chain reaction: Other clients are not satisfied with rewards anymore and drop out themselves, leading to a further performance loss. In the end, the federation either collapses, or only includes a few clients with a badly-performing shared model. On the other hand, outside investment allows clients to acquire more data and better computing resources in the hopes of receiving more rewards, which improves global model performance.

4.2 Upfront cost

A well-known problem in Federated Learning are the upfront costs encountered when joining the federation: Participants may not own the computational equipment required to carry out the calculations necessary in a Federated Learning environment. To deal with this problem, incentive schemes in which the server pays out a portion of the total rewards in the beginning have been proposed[32]. However, in such a case, the server needs to have the financial means to pay clients upfront, which may not be the case. Additionally, it is not sure whether a client will generate enough income during the federation to justify the initial investment. While a more flexible reward scheme is in the interest of the federation, we argue

that opening up the flow of capital by allowing third parties to get involved takes the burden of financing clients off the server, and leads to a more efficient federation in general.

4.3 Consequences

The closed structure of the federation disallows direct outside investment. A third party interested in the use case, who does not own any training data on their own, has no prospect of profiting off the federation without directly investing in clients. This means entering into direct negotiations with investors, a long and tedious process, in which both parties need to agree on the amount to invest, the specific usage of the money, as well as the return. Such a tedious process can discourage investors.

Due to strict privacy regulations, attracting clients to a federation is already a problem in Federated Learning [2], and potentially useful datasets stay unused. Additionally, current reward distribution systems do not allow for enough versatility to incentivize some types of participants.

As an example, tech startups amass vast amounts of private data. As a result of their exceptionally high failure rate, many startups have to cease operations while sitting on a big pile of unused data. This data could still be monetized by participating in federations and collecting rewards. However, steep upfront costs, coupled with payment inflexibility, make it extremely difficult to join the federation. Due to data privacy issues, there is no possibility for the startup to sell out its data to the federation upfront. The data is not lost because the company does not want to participate; the data is lost because the company is not able to monetize its data upfront.

In summary, two types of problems emerge in Federated Learning incentive mechanisms: On the one hand, clients have difficulties in raising capital, either to enter the federation, or to acquire new data sources. On the other hand, outside investing in the federation is not encouraged, and is only possible by directly funding participants or the server.

A solution would be for participants to go public: By issuing shares in the company, they raise funds, and third party investors can directly invest. However, there are several reasons for staying private: Companies keep more control over their operations, do not have to adhere to compulsory regulations and avoid the risks associated with an IPO.

For these reasons, a mechanic to monetize participation and expected future rewards at any point in time is sensible. Enabling participants to trade expected reward payments to raise capital from the start, helps them fund operations and extend their business. On the other hand, this allows outside investors to exchange capital for the opportunity to profit off potential growth in rewards.

In the following sections, we will introduce and discuss a novel reward distribution platform for Federated Learning and demonstrate how it can mitigate the issues presented here.

5 Tokenized reward sharing framework

As showcased in 4, a standard continuous Federated Learning incentive mechanism consists of a determining of the value of a participant's model update and a subsequent exchange of this update for a monetary reward between the client and the server at each round.

The determination of the price is either left to the participant or the central server. This comes with several challenges, including fair pricing and the inflexibility of reward distribution leading to financing problems. We propose a tokenized reward distribution mechanism for Federated Learning, allowing participants to sell stakes in their involvement in the federation to outside actors.

5.1 Architecture

In a federated learning environment, the only flow of capital are the payouts of rewards from the central server. The revenue of the federation is contained in a closed system, which does not allow outside parties to profit directly. That is what we propose as the basis of our framework: In exchange for investing in a client's operations, third parties receive parts of the rewards received by said client. This is a very similar situation to public companies, where investors can receive dividends from a company's profits as a return on their investment.

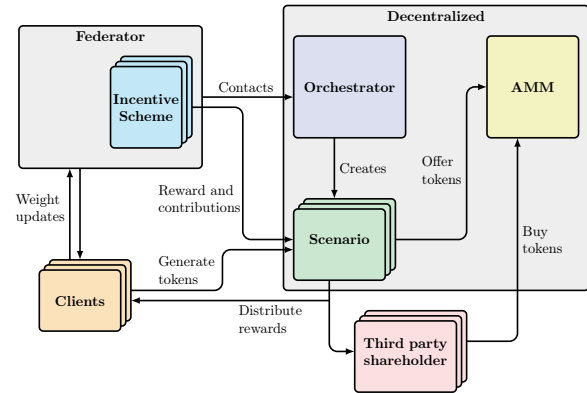


Figure 3: Our share-based reward allocation platform allows payments to go from the server directly to shareholders. Here, the client still holds some shares, as they obtain a part of the reward.

However, the implementation of such a framework in Federated Learning raises some difficulties. A third party investor would still need to invest money directly in a client, which ties them to all of the client's operations. Determining the return for an investor involves negotiations with the client. Additionally, investors can only make fair value assessments of different clients if reward payouts are public information. We use decentralized blockchain technologies, namely smart contracts and tokens, to open up the reward sharing framework: All privacy-sensitive operations are handled by the main server, but everyone can view the distribution of rewards.

More specifically, as shown in Figure 3, our platform introduces three new parties as a decentralized addition to the federation: An **Orchestrator**, a **Scenario** and an **Automated market maker (AMM)**. The Orchestrator is a smart contract, which can be contacted by a interested Federator to create a Scenario. The Scenario, a smart contract created by the Orchestrator for the specific task, manages the process of onboarding clients and acts as the central hub for reward distribution. While the Orchestrator and AMM can be created and owned by the Federator, they may be independent entities, and as such, would rely on themselves for funding.

As decentralization is a desirable trait in Federated Learning we treat both, and by extension also the Scenario, as decentralized standalone entities.

Paying rewards to holders of tokens opens up the incentive scheme: If a third party wants to get involved in a client's operations in the federation, they can purchase tokens and collect a return on their investment via the rewards paid out to the token holders. However, if tokens are directly sold to investors, the system is still closed, as there is no open market for buying and selling.

Thus, we introduce an automated market maker (AMM) to the decentralized part of the federation. As a marketplace for the selling and buying of tokens, the AMM facilitates the exchange and enabling speculation on token prices due to potential future rewards. More specifically, we propose to use a constant mean market maker (CMMM), enabling the exchange of multiple tokens as well as trading for a base currency. This base currency allows any party, be it outside investors or participants, to purchase tokens, and thus involve themselves in a specific client's operations in the federation. The AMM ensures liquidity at any moment, facilitating trading. It is not tied to a specific federation: Indeed, an AMM allows for any token to be listed. Thus, the AMM makes it possible for third parties to invest in different federations, in different clients, as well as in the same client across different federations. Transactions are public: Anyone has insight into token prices and ownership. This transparency enables investors to estimate values, allowing them to determine potentially over- or undervalued clients, and to invest their money accordingly.

Our framework does not depend on a specific incentive scheme: A client-pricing incentive mechanism can just as easily be applied as a server-based incentive mechanism. As long as a flow of capital from the federator to individual clients takes place, the framework can be employed to enable third-party investment.

5.2 User journey

We describe the user journey, from the creation and joining of a federation to the payment of rewards, in more detail.

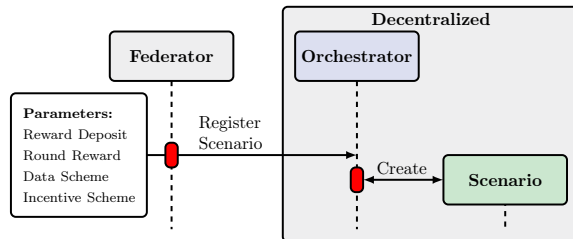


Figure 4: The creation of a specific Scenario.

The Federator, or central server, registers the federation on the Orchestrator, a smart contract (Figure 4). It transmits information on the use case, reward type and amounts, as well as federation specifics such as data scheme, epochs, aggregation strategy and incentive method. The Orchestrator then creates a smart contract for the specific Scenario. The Orchestrator acts as a logbook, keeping track of every federation it has created, enabling prospective clients to choose a Scenario they might want to participate in.

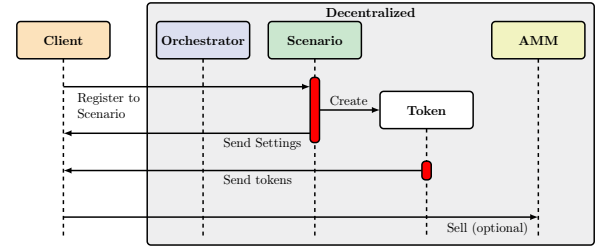


Figure 5: A client joins the Scenario to participate in the federation.

To join a federation (Figure 5), a client contacts the specific Scenario to set up involvement in the federation. During this exchange, the client can choose to create a fixed amount of tokens in their name. The token is tied to the specific client and Scenario, a federation with n clients will contain n tokens at most. Holders of tokens will later receive shares of the specific client's reward, proportional to the amount of tokens they own, just as a shareholder of a company would receive dividends proportional to their held shares. The client receives specific instructions for the setup of the federation, as well as all of their created tokens. They can choose to put their tokens up for sale on the automated market place, or to keep the tokens and collect the full reward.

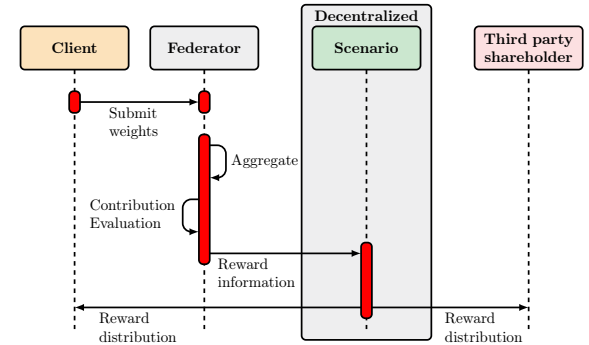


Figure 6: The Federated Learning process and subsequent reward payout.

Each round, the Federator receives model updates from clients (Figure 6), aggregates them and sends back the new global model. Then, the Federator calculates each client's contribution and sends them, along with the round's reward, to the Scenario. Holding ownership information of all tokens, the Scenario splits up the rewards and sends each token holder their respective share of the total reward. This process repeats at each round: The server determines rewards for all clients and the Scenario sends proportional payments to shareholders.

The Orchestrator and the AMM are not tied to a single federation. On the contrary, tokens from different Scenarios with differing use cases, client types and incentive schemes may be listed on an AMM. Thus, a single company participating in several federations can have multiple types of shares listed on a single exchange.

This framework is independent of the incentive mechanism employed, since in all incentive schemes contain a transfer of value from the server to clients. In this instance, the monetary value is instead transmitted to shareholders. This simple change in dynamics does however clear up several problems of a closed incentive scheme, as we will show.

5.3 Framework Advantages

In section 4, we showcased complications which arise when employing a basic reward payout scheme in a federation, namely the closed flow of capital and potentially high upfront costs. Both of these issues are mitigated by our token-based framework.

Our incentive framework opens up the flow of capital as can be seen in the comparison between Figures 2 and 3. Third-party investing is made possible by the sale of tokens on the AMM and subsequent distribution of rewards to token holders by the Scenario. Thus, investors can easily profit off the federation without having to directly finance clients. However, clients collect the proceeds of the sales of their tokens, enabling them to raise capital off of expected future rewards. Reinvesting this capital with the goal of increasing their contribution to the federation increases token-holder value.

Additionally, our framework allows the sale of tokens at any moment. Joiners to a Scenario have the possibility of selling their created tokens at the start, raising capital to purchase computational equipment and set up participation. We can thus mitigate the impact of upfront cost a federation has on prospective clients. As opposed to a company going public, tokens do not come with voting rights and control over the company, clients only waive their right to future rewards by selling tokens.

The possibility of directly investing in clients introduces a few additional benefits. An investor which sees value in a specific use case can directly profit off the federation by buying tokens in all involved clients. If an investor is confident in the quality of the data held by a specific client, they have the possibility of buying their tokens across all Scenarios the client is a part of, increasing their exposure to the client. Similarly, clients have the option of buying other participants' tokens. Thus, if a client is convinced by a use case, they can expose their own exposure by buying competitors' tokens, and collecting a bigger portion of the rewards.

Additionally, investors can choose their risk levels: Directly investing in a client which participates in high-risk use cases, an investor runs the risk of their capital being tied to those use cases. By only buying the client's tokens in a specific federation, the investor shields their investments from these federations, while still being able to fund the client.

Since participants have the option of partaking in the learning of process of several Scenarios, they can boost their overall reputation by showing reliable participation and consistent contribution earnings. This can be an excellent tool to assess a participants' ability to provide valuable data, leading to higher investment interest and increased token prices.

5.4 Pricing

Determining asset prices is an important task, as it determines the profitability of investments. As supply and demand drive token

prices in an automated market maker, we define the trading function such that each asset is, initially, distributed equally. Hence, each client creates an equal amount of tokens when joining the Scenario. This allows a free market to determine fair token values by themselves via third party investors, driving up prices through demand. In literature, different types of reward distributions have been studied: A federator has the possibility of establishing a fixed reward upfront, similarly to a bounty [19], but rewards may also be changing in value and paid out continuously[30].

If a **pre-established reward** amount R is set by the Federator, clients' expected per-round reward can be determined mathematically. With T the total amount of rounds until the reward is fully paid out, and n is the amount of clients in the federation, a client's expected per-round reward E_R^t is $\frac{R}{T \times n}$, with a total expected per-client reward of $\frac{R}{n}$. In such a scenario, each round, tokens lose value proportionally to T , since the total reward left diminishes each round, and they become worthless after the last round.

On the other hand, rewards may change between rounds, presumably due to profit generated by Federator through selling the model. In this case, expected reward calculations include an additional uncertainty parameter, namely the expected profit generated. Additionally, federation which do not include a predetermined total reward can be open-ended, meaning that there is no maximum number of rounds. An expected total per-client reward can theoretically be defined as $E_R = \frac{1}{n} \sum_{t=0} R_t dt$, where R_t is the total reward at round t . However, as per-round rewards are an unknown quantity, other methods, such as federation-adapted price-to-earnings ratios may be better suited to assess token prices.

Of course, a client's reward may be substantially bigger than the expected per-round reward, as actual reward payouts depend additionally on contribution assessments. As showcased in Figure 1, rewards fluctuate considerably between rounds when employing a non-linear and non-equal pricing method. This gives investors the possibility of finding undervalued tokens, enabling speculation. Events such as a client announcing the acquisition of new data sources, a client's unexpected substantial contributions, or the dropping out of a similar competitor, have the potential as serving as catalysts for a token's value.

6 Conclusion and Future work

This work proposes a novel incentive distribution method using client-specific tokens as a means of investment in federated learning. Our framework addresses the limitations of current incentive mechanisms by introducing a more open financial system, allowing third-party investment and mitigating upfront costs for participants in the learning process by providing them with a novel data monetisation approach. Our approach provides a more inclusive environment, potentially attracting new actors and a wider range of data sources, ultimately enhancing the effectiveness of the overall federated learning process. Additionally, enabling speculation on the value of client tokens could provide valuable insights into the participants' internal data quality.

Thanks to the integration of the proposed framework to a decentralized finance platform, it offers a flexible and adaptive reward system that can dynamically respond to the evolving needs and contributions of participants. Participants can choose to raise more

capital through the sale of their tokens or to monetise their data through future contribution earnings. This flexibility is crucial for maintaining participant engagement and incentivizing continuity in their participation.

Future work will involve a more detailed analysis of the economic viability of the proposed framework, as well as in-depth exploration of potential regulatory considerations surrounding the issuance and trading of these client-specific tokens. Additionally, implementing the framework in a specific blockchain environment such as Ethereum and deploying a proof of concept to explore the effect of different pricing strategies would provide further insights into our framework. Finally, further research into the potential integration of advanced cryptographic techniques to enhance security and trust within the framework could further solidify the framework's applicability in privacy-sensitive sectors.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments. This paper was funded in part by the Luxembourg National Research Fund (FNR) under grant number 18047633.

References

- [1] Guillermo Angeris, Akshay Agrawal, Alex Evans, Tarun Chitra, and Stephen Boyd. 2022. Constant function market makers: Multi-asset trades via convex optimization. In *Handbook on Blockchain*. Springer, 415–444.
- [2] Xuan Bi, Alok Gupta, and Mochen Yang. 2023. Understanding partnership formation and repeated contributions in federated learning: An analytical investigation. *Management Science* (2023).
- [3] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. 2016. Practical Secure Aggregation for Federated Learning on User-Held Data. arXiv:1611.04482 [cs.CR] <https://arxiv.org/abs/1611.04482>
- [4] Vitalik Buterin et al. 2013. Ethereum white paper. *GitHub repository* 1 (2013), 22–23.
- [5] Wenlin Chen, Samuel Horváth, and Peter Richtárik. 2020. Optimal Client Sampling for Federated Learning. *CoRR abs/2010.13723* (2020). arXiv:2010.13723 <https://arxiv.org/abs/2010.13723>
- [6] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. 2020. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in neural information processing systems* 33 (2020), 16937–16947.
- [7] Tzu-Ming Harry Hsu, Hang Qi, and Matthew Brown. 2019. Measuring the effects of non-identical data distribution for federated visual classification. *arXiv preprint arXiv:1909.06335* (2019).
- [8] Wenke Huang, Mang Ye, Zekun Shi, Guancheng Wan, He Li, Bo Du, and Qiang Yang. 2023. Federated Learning for Generalization, Robustness, Fairness: A Survey and Benchmark. arXiv:2311.06750 [cs.LG] <https://arxiv.org/abs/2311.06750>
- [9] Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning* 14, 1–2 (2021), 1–210.
- [10] Jiawen Kang, Zehui Xiong, Dusit Niyato, Shengli Xie, and Junshan Zhang. 2019. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. *IEEE Internet of Things Journal* 6, 6 (2019), 10700–10714.
- [11] Chul Min Lee, Joaquín Delgado Fernández, Sergio Potenciano Menci, Alexander Rieger, and Gilbert Fridgen. 2023. Federated Learning for Credit Risk Assessment. In *HICSS*. 386–395.
- [12] Yang Liu, Tao Fan, Tianjian Chen, Qian Xu, and Qiang Yang. 2021. Fate: An industrial grade platform for collaborative learning with data protection. *Journal of Machine Learning Research* 22, 226 (2021), 1–6.
- [13] Lingjuan Lyu, Jiangshan Yu, Karthik Nandakumar, Yitong Li, Xingjun Ma, and Jiong Jin. 2019. Towards Fair and Decentralized Privacy-Preserving Deep Learning with Blockchain. *CoRR abs/1906.01167* (2019). arXiv:1906.01167 <http://arxiv.org/abs/1906.01167>
- [14] Bho Matthiesen, Nasrin Razmi, Israel Leyva-Mayorga, Armin Dekorsy, and Petar Popovski. 2023. Federated learning in satellite constellations. *IEEE Network* (2023).
- [15] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.
- [16] Vijay Mohan. 2022. Automated market makers and decentralized exchanges: a DeFi primer. *Financial Innovation* 8, 1 (2022), 20.
- [17] Dinh C Nguyen, Ming Ding, Pubudu N Pathirana, Aruna Seneviratne, Jun Li, and H Vincent Poor. 2021. Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials* 23, 3 (2021), 1622–1658.
- [18] Shashi Raj Pandey, Lam Duc Nguyen, and Petar Popovski. 2022. Fedtoken: Tokenized incentives for data contribution in federated learning. *arXiv preprint arXiv:2209.09775* (2022).
- [19] Konstantin D Pandl, Florian Leiser, Scott Thiebes, and Ali Sunyaev. 2022. Reward systems for trustworthy medical federated learning. *arXiv preprint arXiv:2205.00470* (2022).
- [20] Sashank Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečný, Sanjiv Kumar, and H. Brendan McMahan. 2020. Adaptive federated optimization. *arXiv preprint arXiv:2003.00295* (2020).
- [21] Mary ROSZEL, Beltran FIZ PONTIVEROS, and Radu STATE. 18 September 2023. FLAIRS: Federated Learning AI Regulatory Sandbox. In *ECML PKDD 2023 Workshops*. Springer Nature.
- [22] Lloyd S Shapley et al. 1953. A value for n-person games. (1953).
- [23] Tianshu Song, Yongxin Tong, and Shuyue Wei. 2019. Profit allocation for federated learning. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2577–2586.
- [24] Afaf Taik and Soumaya Cherkaoui. 2020. Electrical load forecasting using edge computing and federated learning. In *ICC 2020-2020 IEEE international conference on communications (ICC)*. IEEE, 1–6.
- [25] Tra Huong Thi Le, Nguyen H. Tran, Yan Kyaw Tun, Minh N. H. Nguyen, Shashi Raj Pandey, Zhu Han, and Choong Seon Hong. 2021. An Incentive Mechanism for Federated Learning in Wireless Cellular Networks: An Auction Approach. *IEEE Transactions on Wireless Communications* 20, 8 (2021), 4874–4887. <https://doi.org/10.1109/TWC.2021.3062708>
- [26] Xuezhen Tu, Kun Zhu, Nguyen Cong Luong, Dusit Niyato, Yang Zhang, and Juan Li. 2022. Incentive mechanisms for federated learning: From economic and game theoretic perspective. *IEEE transactions on cognitive communications and networking* 8, 3 (2022), 1566–1593.
- [27] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H. Yang, Farhad Farokhi, Shi Jin, Tony Q. S. Quek, and H. Vincent Poor. 2019. Federated Learning with Differential Privacy: Algorithms and Performance Analysis. *CoRR abs/1911.00222* (2019). arXiv:1911.00222 <http://arxiv.org/abs/1911.00222>
- [28] Jiahua Xu, Krzysztof Paruch, Simon Cousaert, and Yebo Feng. 2023. Sok: Decentralized exchanges (dex) with automated market maker (amm) protocols. *Comput. Surveys* 55, 11 (2023), 1–50.
- [29] Mengwei Yang, Linqi Song, Jie Xu, Congduan Li, and Guozhen Tan. 2019. The tradeoff between privacy and accuracy in anomaly detection using federated xgboost. *arXiv preprint arXiv:1907.07157* (2019).
- [30] Shuo Yang, Fan Wu, Shaojie Tang, Xiaofeng Gao, Bo Yang, and Guihai Chen. 2017. On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing. *IEEE Journal on Selected Areas in Communications* 35, 4 (2017), 832–847.
- [31] Dong Yin, Yudong Chen, Ramchandran Kannan, and Peter Bartlett. 2018. Byzantine-robust distributed learning: Towards optimal statistical rates. In *International conference on machine learning*. Pmlr, 5650–5659.
- [32] Han Yu, Zelei Liu, Yang Liu, Tianjian Chen, Mingshu Cong, Xi Weng, Dusit Niyato, and Qiang Yang. 2020. A sustainable incentive scheme for federated learning. *IEEE Intelligent Systems* 35, 4 (2020), 58–69.
- [33] Rongfei Zeng, Shixun Zhang, Jiaqi Wang, and Xiaowen Chu. 2020. Fmore: An incentive scheme of multi-dimensional auction for federated learning in mec. In *2020 IEEE 40th international conference on distributed computing systems (ICDCS)*. IEEE, 278–288.
- [34] Chengliang Zhang, Suyi Li, Junzhe Xia, Wei Wang, Feng Yan, and Yang Liu. 2020. {BatchCrypt}: Efficient homomorphic encryption for {Cross-Silo} federated learning. In *2020 USENIX annual technical conference (USENIX ATC 20)*. 493–506.
- [35] Wenbo Zheng, Lan Yan, Chao Gou, and Fei-Yue Wang. 2020. Federated Meta-Learning for Fraudulent Credit Card Detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, Christian Bessière (Ed.). International Joint Conferences on Artificial Intelligence Organization, 4654–4660. <https://doi.org/10.24963/ijcai.2020/642> Special Track on AI in FinTech.