Energy Efficient Federated Learning over Cooperative Relay-Assisted Wireless Networks

Xinyue Zhang[1], Rui Chen[2], Jingyi Wang[3], Huaqing Zhang[2], Miao Pan[2]
[1]Department of Computer Science, Kennesaw State University, Marietta, GA 30060
[2]Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204
[3]Department of Computer Science, San Francisco State University, San Francisco, CA 94132

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10000664

***Abstract*—Federated learning (FL) is a promising distributed learning paradigm, which can effectively avoid the privacy leakage and communication issues compared with the centralized learning. Specifically, in each training iteration, FL nodes only upload the local training results to the centralized server without disclosure of their raw training dataset and the centralized server will aggregate the local results of all FL nodes and update the global model. To this end, the performance of the global model is highly dependent on the nodes' cooperation. However, it is challenging to motivate mobile edge devices to involve themselves in the FL process without a desired incentive. Another significant concern of the mobile edge devices is the communication and computational energy cost of participation. Therefore, considering the high cost and weak communication channel with the centralized server specially for the distant nodes, in this paper, we propose a relay-assisted energy efficient scheme for federated learning, where each FL computational node is not only motivated by monetary awards based on their local dataset, but also further motivated to function as a relay node to assist distant nodes on local results uploading due to its locality advantage. To achieve a stable pairing solution between FL computational nodes and assisted relays in a distributive fashion, a many-to-one matching algorithm is applied, where each the computational node and relay is unable to deviate with current pairing unilaterally for higher revenue. Extensive simulations are conducted to illustrate the correctness and effectiveness of our proposed scheme.**

***Index Terms*—federated learning, optimization, matching, relay**

I. Introduction

From autonomous vehicles to smart cities, from virtual reality to modern factories, the Internet of Things (IoT) has opened its curtains for an imaginative golden age with the advent of the 5G adoption. Thanks to 5G's ultra high-speed network, mobile edge devices are capable to transmit data at a faster rate, which enables a potential breakthrough of the machine learning technologies [1]. The past few decades have witnessed a soaring number of devices connected to the network, resulting in a significant growth in the amount of data generated as well as the network complexity. Machine learning models can be trained, deployed, as well as interact with each other at multiple levels, such as edge devices, computational machines and cloud servers [2]. With the proliferation of machine learning based applications, there are increasing concerns about privacy leakage from the data owners, since in order to enjoy benefits from the application, the raw data will leave the local devices and be used for training machine learning model on the server's side [3].

This work was supported in part by the U.S. National Science Foundation under grants CNS-2107057, CNS-1801925.

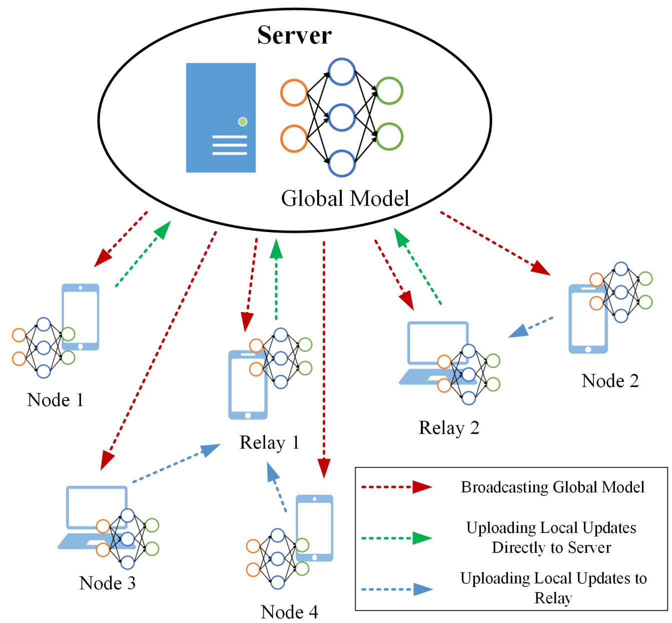

Fig. 1: System overview.

With the advancement of technology, it is critical to balance the privacy risk versus the communication efficiency. Federated learning (FL) [4] was proposed as an emerging distributed learning framework, which is one of the best solutions to the exacerbated privacy issue. As a game breaker, FL comes into play as it parts with the criticized privacy concerns while maintaining all the benefits from a distributed learning system. In this novel framework of distributed learning, the mobile edge devices jointly train a shared model locally in a cooperative manner. To be specific, the centralized server first broadcasts the global model to all the FL edge nodes. Then, the participating nodes can engage in the FL process by training the local model with their own local dataset. The raw data is kept on the edge devices while only the intermediate training results are shared with the centralized server.

Despite the advantages of FL, there is still need for more efforts on motivating nodes participation in order to achieve better global model performance. Since the edge nodes and the centralized server are supposed to exchange the trained model in each training epoch, the communication and energy cost are the main concerns of the potential participating edge mobile devices. Therefore, it is necessary to have a novel incentive mechanism to compensate for the cost of each participant and motivate individuals to engage in cooperation during the FL process. In [5], the authors proposed an auction-based incentive mechanism for FL with the consideration of the learning quality of each participating node. In [6], the authors designed a data trading mechanism in FL to motivate participants to share their private data. By leveraging the game theory, the server can select nodes that can provide a reliable dataset with a high probability. However, in these proposed incentive mechanisms, the authors didn't consider the energy consumption of each FL node during training and local updates.

Moreover, in FL, direct communication channels between nodes and centralized server are always inefficient, or even unavailable, when the nodes are distant, since the transmission cost is expensive. The cooperative relay-assisted wireless networks [7] are an effective solution to enhance the communication reliability. In [8], the authors proposed a synchronization scheme which can transmit models and updates simultaneously via a relay base station by utilizing the relay-assisted wireless network. In [9], the authors leveraged over-the-air computation technique on relays to assist local updates transmission and further reduced the communication cost in FL. However, these papers involved one or several relay base stations and they consider that the FL nodes can only upload local updates via relays. Additionally, it takes extra efforts to establish relays in pre-determined locations. In the scenarios where FL nodes are flexibly distributed, we expect each node to be self-motivated to function as a relay in a distributive way, and the computational node also prefers to upload to the closest optimal relay nearby with higher revenues.

To tackle the issues discussed above, in this paper, we proposed a relay-assisted energy efficient scheme for FL, where each computational node or relay is self-motivated to join FL based on its own utilities. Our salient contributions are listed as follows.

- The proposed relay-assisted energy efficient scheme motivates large volume of edge devices to function as computational nodes and relays, resulting in high revenues for FL in a distributive fashion.
- A many-to-one matching algorithm is applied in the proposed scheme to achieve stable pairing results between computational nodes and relays, where no computational nodes or relays are able to unilaterally deviate from current pairing with higher revenues.
- Extensive simulations are conducted with high performance of the proposed scheme, compared with scenarios without relays.

The rest of paper is organized as follows. In Section II, we describe the overview of our system, and discuss the energy consumption of each FL node. In Section III, we formulate the utility function and optimization problem for both computational nodes and relays with the consideration of the participation reward and the energy consumption of training and transmission. We provide further analysis on the formulated problem and propose a matching algorithm for the stable pairing mechanism between computational nodes and relays in the distributive fashion in Section IV. In Section V, we provide extensive simulation results and evaluate the performance of our proposed scheme. Finally, we draw conclusions in Section VI.

II. System Model

The overview of our proposed method is shown in Figure 1. We consider that there is one centralized server with a global learning model and multiple mobile edge devices as potential participating FL nodes. Since edge devices located far from the centralized server may exist, we expect that some of the FL nodes can also work as relays to help other distant nodes with model data transmission based on local updates so as to improve the data transmission reliability. Therefore, the FL nodes are expected to optionally function as two roles in the proposed scheme. The first one is working as a normal computational node that can do the model training based on the local dataset and send the local updates directly to the centralized server or via the nearby relay node. The second kind is to function as a relay so as to assist distant computational nodes for data transmission. In order to motivate mobile edge devices to cooperate in the training process, the centralized server provides incentive to each participating FL node based on the size of the local dataset applied in local model training. If a FL node decides to select another FL node to assist with the local update's transmission, this node is supposed to provide partial incentive as a reward to the relay. Each relay can optionally assist with the data communication according to the evaluation of energy consumption. The objective of each FL node is to maximize its own utility based on the incentive and energy consumption.

In our system scenario, we consider a set of FL nodes $\mathcal{K} = \{1, \cdots, k, \cdots, N\}$. The energy consumption of each node k consists of three parts, including the computational energy from local training E_k^{comp}, data transmission energy due to local update E_k^{node}, and corresponding transmission energy E_k^{relay} if the node functions as a relay. Given the local training dataset size D_k, and the energy consumed for one training data sample on the edge size ρ_k, the energy consumption of each node k for local training is,

$$E_k^{comp} = \rho_k D_k. \tag{1}$$

Note that ρ_k is dependent on the computing architecture and hardware of each edge node k. We consider the allocated bandwidth of each node k for transmitting local update is B, the variance of white Gaussian channel noise is N_0, and the channel state of from each node k to relay i is h_{ki}. We also assume that M (in bit) denotes the size of the applied machine learning model in federated learning and the total time for uploading the machine learning model is no more than $\bar{\tau}$. Hence, the transmission energy consumption of each node k for local updates can be denoted as [10], which is shown as follows

$$E_k^{node} = \sum_i a_{ki} E_{ki}^{node}, \tag{2}$$

$$E_{ki}^{node} = \frac{\bar{\tau} N_0 B}{(h_{ki})^2} \left(2^{\frac{M}{\bar{\tau} B}} - 1 \right), \tag{3}$$

where a_{ki} is a binary variable indicating if the computational node k communicates with the relay i, satisfying $\sum_i a_{ki}^t \leq 1, k \in \mathcal{K}, i \in \mathcal{K} \bigcup \{0\}$. When $i \in \mathcal{K}$, a relay node i helps the computational node k for the local update transmission. When $i = 0$, we assume the node communicates with the centralized server directly. If one FL node works as a relay, there will be additional energy consumption for assisting other computational nodes to transmit their local updates. The additional energy consumption of edge node i as a relay can be denoted as

$$E_i^{relay} = \frac{\bar{\tau} N_0 B}{(h_i)^2} \left(2^{\frac{L}{\bar{\tau} B}} - 1 \right). \tag{4}$$

Based on the energy consumption evaluation, each FL node tries to maximize its own revenue and minimize its energy consumption cost in the FL process. The utility function of each FL node k is formulated and analyzed in Section III.

III. Problem Formulation

Under the system scenario, each node behaves distributively and aims to maximize its own utility. In this section, we formulate the utility function of each FL node according to its serving role as computational node and relay, respectively. We assume each FL node is able to serve as computational node and as a relay for other computational nodes simultaneously, and it optimizes its own utilities as a computational node and as a relay based on the behavior of others in the FL process.

A. Utility Analysis of Computational Node

With the updated model based on computation and training with local preserved data set, each computational node seeks the centralized server or a nearby relay for the communication on model parameters. For each computational node k, we consider the total utility for each computational node as the revenue obtained from the local preserved data D_k minus energy cost due to the transmission and computation.

Accordingly, if the computational node k communicates with the centralized server directly, the utility for the computational node sending to the centralized server directly is

$$U_k^{node-server} = \max\{\phi D_k - \beta E_{k0}^{node} - \gamma E_k^{comp}, 0\}, \tag{5}$$

which is the revenue gained from the local dataset owned by computational node D_k minus the communication cost from the computational node k to centralized server E_{k0}^{node}. The computation cost of the node k E_k^{comp}. ϕ is the weight of reward on private dataset in monetary value. β and γ are the weights of each type of cost to evaluate in monetary values. If the utility is less than 0, we assume the node is unwilling to join the federated learning network due to high costs.

On the other hand, we suppose the computational node k can also select one relay for the data transmission. In order to motivate node to relay the data communication, we suppose each computational node rewards the relay α ratio of its own revenue, $\alpha \in (0, 1)$. Therefore, the utility for computational node k on each relay $i (\forall i \in \{1, 2, \ldots, N\}, i \neq k)$, is

$$U_{ki}^{node-relay} = \max\{(1-\alpha)\phi D_k - \beta E_{ki}^{node} - \gamma E_k^{comp}, 0\} \tag{6}$$

which is revenue gained from the local dataset owned by computational node D_k minus the motivation cost to the relay $\alpha\phi D_k$, together with the subtraction of the communication cost from the computational node k to relay i E_{ki}^{node} and the computation cost E_k^{comp}.

Based on the above options, the optimization problem for computational node k, $\forall k \in \{1, 2, \ldots, N\}$, can be shown as

$$\begin{aligned} &\max_{a_{ki}, \forall i \in \{1,2,\ldots,N\}} \quad \{U_k^{node-server}, \sum_{i=1}^{N} a_{ki} U_{ki}^{node-relay}\}, \\ &s.t. \begin{cases} \sum_{i=1}^{N} a_{ki} \leq 1, \\ a_{ki} \in \{0, 1\}, \quad \forall i \in \{1, 2, \ldots, N\}, \quad i \neq k, \end{cases} \end{aligned} \tag{7}$$

where each computational node selects relay optimally, with observation of the behaviors of other nodes. With the optimized potential solution, the computational node further compares the utilities of communicating directly with the centralized server and sending the data via the most optimal relay chosen as the option with highest utility. We assume each computational node can pair with, at most, one relay for data transmission.

B. Utility Analysis of Relay Node

Motivated by the reward from other computational nodes, each node evaluates its own capacity and communication cost to the centralized server and optionally functions as relay for other nodes to gain revenue. We suppose each node i can relay data transmission for, at most, C_i other nodes. The utility of relay i when assisting data transmission of computational node k is denoted as

$$U_{ki}^{relay} = \max\{\alpha\phi D_k - \beta E_i^{relay}, 0\}, \tag{8}$$

which is the reward from computational node k $\alpha\phi D_k$ minus the communication cost of relay E_i^{relay}. Similar as the computational node, if the utility of relay is less than 0, we suppose the relay is unwilling to help on data transmission due to high cost.

Accordingly, the optimization problem for relay i, $\forall i \in \{1, 2, \ldots, N\}$ is

$$\begin{aligned} &\max_{a_{ki}, \forall k \in \{1,2,\ldots,N\}} \quad \sum_{k=1}^{N} a_{ki} U_{ki}^{relay}, \\ &s.t. \begin{cases} \sum_{k=1}^{N} a_{ki} \leq C_i, \\ a_{ki} \in \{0, 1\}, \quad \forall k \in \{1, 2, \ldots, N\}, \quad k \neq i, \end{cases} \end{aligned} \tag{9}$$

where each relay optimizes its selections on the computational nodes based on its own capacity C_j and the behavior of other relays and computational nodes. If there are more than C_k computational nodes requesting to be served, the relay k can only choose the top C_k ones and reject the rest to achieve highest utility.

IV. System Analysis

According to the formulated problem, each node functions as a computational node k, $\forall k \in \{1, 2, \ldots, N\}$ and each node functions as a relay i, $\forall i \in \{1, 2, \ldots, N\}$ seeks the pairing a_{ki}, so as to optimize its own utility based on the behavior of others. In order to achieve an optimal and stable result in a distributive fashion, matching theory is introduced as a powerful tool [11], [12], where computational nodes and relays try to be matched with each other and achieve stable pairing solutions.

Following the formulated optimization problems, in the distributive relay-assisted federated learning network, computational nodes and relays try to pair with each other to gain a high and stable utility. Based on the utilities, the preference list of each computational node k can be denoted as follows

$$\begin{aligned} &P_k^{comp} = \{U_{ki}^{node-relay}, \quad \forall i \in \{1, 2, \ldots, N\}\}, \\ &s.t.\ U_{ki}^{node-relay} > U_k^{node-server}, \end{aligned} \tag{10}$$

which is composed of the list of relays that can bring higher utility for the computational node, compared with communicating with the centralized server directly.

Similarly, we consider the preference list of each relay i as

$$\begin{aligned} &P_i^{relay} = \{U_{ki}^{relay}, \quad \forall k \in \{1, 2, \ldots, N\}\}, \\ &s.t.\ U_{ki}^{relay} > 0, \end{aligned} \tag{11}$$

which is composed of the list of computational nodes candidates that can bring revenues.

Based on preference lists, a many-to-one matching algorithm is proposed between computational nodes and relays, with details illustrated in Algorithm 1 [13]. At the beginning stage, each computational node k keeps a preference list for all potential relays nearby, and is initially set as active status with $s_k = 1$. Each relay i also keeps a preference list for all computational nodes and initializes a priority queue q_i to record the top most promising computational nodes based on its capacity C_k. During the matching process, for each round R, each active computational node first proposes to the most favourable relays according to its pointer r_k in the preference list. If there are no more relays to propose to, the computational node is set as deactivated and regarded as no longer available for matching. Receiving the proposal from computational nodes, each relay i' temporally added the computational node into its queue $q_{i'}$. Next, each relay i evaluates all of its potential candidates in the queue q_i, and select the top C_i computational nodes based on its preference list, due to the capacity. The computational node k gets popped out and will keep active for next round of proposals with $s_k = 1$, and computational node k is kept in the queue q_i of relay i, $\forall i \in \{1, 2, \ldots, N\}$ updates its status as deactivated with $s_k = 0$. The round of proposals and selections continues until all computational nodes are deactivated, meaning the computational nodes are either matched with a preferred relay or there are no available relays to match with. Meanwhile, the relay k with empty queue q_k is not to function as a relay for other computational nodes, while the others are matched with all computational nodes in its queue and serve as a relay for the data transmission.

Algorithm 1 Matching Algorithm for Relay-Assisted Federated Learning Network.

for Computational node k **do**
  Sort the preference list P_k^{comp} in (10) from high to low;
  Initialize pointer r_k at the first relay in P_k^{comp} with highest utility value;
  Initialize status $s_k = 1$ as actively seeking relays;
end for
for Relay i **do**
  Sort the preference list P_i^{relay} in (11) from high to low;
  Initialize priority queue q_i for matched computational nodes.
end for
Initialize Round $R = 0$;
while $\sum_k s_k > 0$ **do**
  $R++$;
  Section: Active computational nodes propose to relays;
  for Computational node k **do**
    if $s_k = 1$ **then**
      if Pointer r_k already points to last one in the preference list **then**
        Set $s_k = 0$;
        continue;
      end if
      Pointer r_k++;
      $q_{i'}$ appends k;
      Set $s_k = 0$;
    end if
  end for
  Section: Relays determine which computational node to choose based on its capacity;
  for Relay i **do**
    while Size of $q_i >$ Capacity C_i **do**
      q_i pops k' with lowest value;
      Set $s_{k'} = 1$;
    end while
  end for
end while

Lemma 1. *For each computational node in the algorithm, the pointer r_k of the computational node k in its sorted preference list moves in one direction.*

Proof. As depicted in the proposed matching algorithm, the relay i, $\forall i \in \{1, 2, \ldots, N\}$, maintains the priority queue q_i

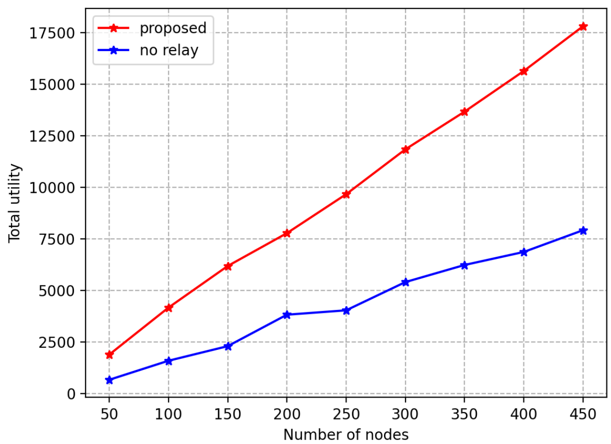

Fig. 2: Total utility with different number of nodes

and keep the highest C_k proposed computational nodes if it is out of the capacity. Therefore, for any two rounds $R = x$ and $R = y$ in the proposal and selection loops, where $x < y$, we get

$$\left[\sum_{k=1}^{N} a_{ki} U_{ki}^{relay}\right](x) \leq \left[\sum_{k=1}^{N} a_{ki} U_{ki}^{relay}\right](y), \tag{12}$$

where the utility of the relay i at round y should be no less than the utility at round x.

Accordingly, if we assume the relay i rejects the proposal of the computational node k at round x, at round y there should be C_i computational nodes in the priority queue q_i and each of the matching computational nodes brings higher revenues for the relay i than computational node k. Therefore, each computational node k should propose in one direction based on its sorted preference list, and cannot propose to the relay previously rejected and get accepted in a later round. □

Lemma 2. *All computational nodes and relays will achieve a stable matching result.*

Proof. According to Lemma 1, after each computational node proposes based on its preference list and gets deactivated with $s_k = 0$, each computational node cannot propose to other relays unilaterally to get matched and higher utility. Therefore, following the same proof as in [14], all computational nodes and relays achieve a stable matching result. □

V. Performance Evaluation

In this section, simulation results are presented to evaluate our proposed framework. In the simulated scenario of relay-assistant federated learning network, without specific explanation, there are 50 nodes allocated randomly in a square urban area with an edge of 100 meters, and the centralized server is located at the left bottom corner of area. The neural network's model size is $M = 3.4 \times 10^5$ bits. The local dataset size D is uniformly distributed for all the computational nodes from 10 Gigabytes to 100 Gigabytes. In the simulations, we assume that all of the nodes will participate in the training. The proposed scheme can be easily extended in the case of FL with partial participation where only a subset of computational nodes need to transmit their models. For the wireless communication network, we consider an OFDMA system with bandwidth B = 10 MHz. The variance of the complex white Gaussian channel noise is $N_0 = 10^{-12}$. We suppose that the uploading time for each round of model training will be $\bar{\tau} = 300$ms. We set 2 as the capacity of each relay. The ratio of motivated award is $\alpha = 0.3$. The weight factors β and γ are set as 0.1 and 0.1, respectively.

In Fig.2, we evaluate the total utility based on the number of nodes. In comparison with the proposed relay-assistant scenario, we also consider the scenario where each computational node communicates with the centralized server directly. As shown in the figure, generally, the total utility for both scenarios increases as the number of nodes increase, and the proposed relay-assistant scenario keeps in higher total utility than the no relay scenario. Thus, in the proposed scenario, each computational node is able to communicate with the a nearby relay to reduce the energy cost from data transmission, and the relays are also motivated to help on the data transmission to the centralized server. Furthermore, we also notice as the number of nodes increases, the difference gap between the proposed relay-assistant scenario and the no-relay scenario also increases. With more relay candidates located nearby, the computational nodes are able to propose and be matched for higher revenue.

To evaluate how many relays are actually motivated to serve in the proposed scenario, we evaluate the number of active relays based on the number of nodes. We suppose the capacity of each relay as 1, 2, 3, 5, 8, respectively. With the number of nodes increasing, as shown in Fig. 3, more computational nodes are able to be matched with nearby relays and the number of active relays increases. When the number of nodes is small, each relay will not achieve its maximum capacity in the stable matching status. Thus, for the scenarios with different capacities on the relay, the total number of active ones are similar. However, when the number of nodes is large, more relays will be motivated and active when the capacity of relays is small, since the computational node would have to seek and pair with other relays due to the low capacity of each of them. Accordingly, we can conclude the relays are able to help increase the general performance of the federated learning network and when increasing the capacity of each relay, some relays will be more popular and can bring high benefits for the network due to each relay's advantage.

Furthermore, we evaluate the average utility of each node based on the value of the motivation reward ratio α in Fig. 4. With 10, 50 and 500 nodes distributed in the simulated scenarios, we observe the average utility of each node behaves differently with different α values. When the total volume of nodes is small, the computational nodes are still able to motivate and be matched with relays for the data transmission

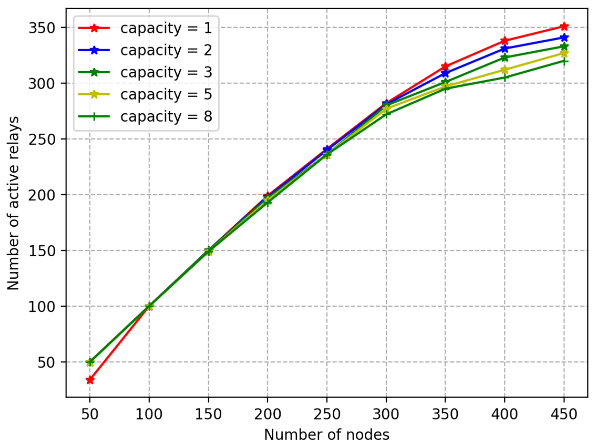

Fig. 3: Number of active relays with different number of nodes

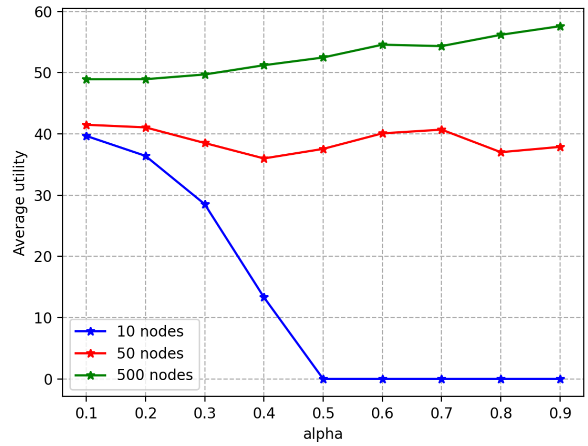

Fig. 4: Average utility versus motivation reward ratio α

with small α values. However, with the ratio of α increasing, considering the high cost to motivate relays, the revenue gained for the computational node is less than the energy cost itself. Thus, the computational nodes are less likely to join as the edge node in federated learning. In the scenario of 10 nodes, when $\alpha > 0.5$, we notice the average utility achieves 0, which means there are no computational nodes willing to join in the federated learning network. From the 0 utility, we also discover that the 10 nodes are unwilling to join in federated learning and communicate with centralized server directly. However, with a small ratio of α values, they are able to join with the assistant of matched relays.

As the total volume of nodes increases, in Fig. 4, we notice the average utility increases when α is large since each computational node has better chance to match and communicate with relays nearby with lower transmission power, even though high ratio of revenues are applied for the motivation. In the scenario with 50 nodes, we notice the average utility for each node is similar with different α values. In the scenario with 500 nodes, as the α increases, the average utility increases since more nodes are motivated to serve as a relay and help by reducing the transmission cost for more computational nodes.

VI. Conclusion

In this paper, a relay-assisted energy efficient scheme for federated learning has been proposed, where FL nodes are self-motivated to join into the network as computational nodes and relays to achieve high monetary awards. With the application of the many-to-one matching algorithm, the computational nodes and relays achieve stable matching results, where no computational node or relay is able to deviate from the current matching results unilaterally for higher revenue. Extensive simulations are conducted to show the correctness and effectiveness of our proposed scheme.

References

[1] D. Wang, D. Chen, B. Song, N. Guizani, X. Yu, and X. Du, "From iot to 5g i-iot: The next generation iot-based intelligent algorithms and 5g technologies," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 114–120, October 2018.

[2] I. Alvarado, "Ai-enabled iot, network complexity and 5g," in *2019 IEEE Green Energy and Smart Systems Conference (IGESSC)*, Long Beach, CA, March 2019.

[3] X. Zhang, J. Ding, S. M. Errapotu, X. Huang, P. Li, and M. Pan, "Differentially private functional mechanism for generative adversarial networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, December 2019.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, Ft. Lauderdale, FL, April 2017.

[5] Y. Deng, F. Lyu, J. Ren, Y.-C. Chen, P. Yang, Y. Zhou, and Y. Zhang, "Fair: Quality-aware federated learning with precise user incentive and model aggregation," in *2021 IEEE Conference on Computer Communications (INFOCOM)*, Virtual, May 2021.

[6] R. Hu and Y. Gong, "Trading data for learning: Incentive mechanism for on-device federated learning," in *2020 IEEE Global Communications Conference (GLOBECOM)*, Taipei, Taiwan, December 2020.

[7] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information theory*, vol. 50, no. 12, pp. 3062–3080, November 2004.

[8] Z. Qu, S. Guo, H. Wang, B. Ye, Y. Wang, A. Zomaya, and B. Tang, "Partial synchronization to accelerate federated learning over relay-assisted edge networks," *IEEE Transactions on Mobile Computing*, May 2021.

[9] Z. Lin, H. Liu, and Y.-J. A. Zhang, "Relay-assisted cooperative federated learning," *IEEE Transactions on Wireless Communications (Early Access)*, March 2022.

[10] J. Xu and H. Wang, "Client selection and bandwidth allocation in wireless federated learning networks: A long-term perspective," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1188–1200, October 2020.

[11] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: fundamentals and applications," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 52–59, 2015.

[12] P. Harrenstein, D. Manlove, and M. Wooldridge, "The joy of matching," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 81–85, 2013.

[13] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.

[14] H. Zhang, Y. Xiao, S. Bu, D. Niyato, F. R. Yu, and Z. Han, "Computing resource allocation in three-tier iot fog networks: A joint optimization approach combining stackelberg game and matching," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1204–1215, 2017.