Energy Minimization for Participatory Federated Learning in IoT Analyzed via Game Theory

Alessandro Buratto*, Elia Guerra†, Marco Miozzo†, Paolo Dini†, and Leonardo Badia*
*Dept. of Information Engineering (DEI), University of Padova, Italy
email: {alessandro.buratto.1@phd. , leonardo.badia@}unipd.it
†Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels, Spain
email: {eguerra, marco.miozzo, paolo.dini}@cttc.es

***Abstract*—The Internet of Things requires intelligent decision making in many scenarios. To this end, resources available at the individual nodes for sensing or computing, or both, can be leveraged. This results in approaches known as participatory sensing and federated learning, respectively. We investigate the simultaneous implementation of both, through a distributed approach based on empowering local nodes with game theoretic decision making. A global objective of energy minimization is combined with the individual node's optimization of local expenditure for sensing and transmitting data over multiple learning rounds. We present extensive evaluations of this technique, based on both a theoretical framework and experiments in a simulated network scenario with real data. Such a distributed approach can reach a desired level of accuracy for federated learning without a centralized supervision of the data collector. However, depending on the weight attributed to the local costs of the single node, it may also result in a significantly high Price of Anarchy (from 1.28 onwards). Thus, we argue for the need of incentive mechanisms, possibly based on Age of Information of the single nodes.**

***Index Terms*—Federated learning; Energy consumption; Participatory sensing; Internet of things; Game theory.**

I. Introduction

Participatory sensing, facilitated by the proliferation of mobile devices and the Internet of Things (IoT), has emerged as a promising paradigm for collecting large-scale data from end nodes and exploiting them for many applications [1], [2]. It implies a voluntary contribution of individuals to data provision, forming a distributed sensing network that harnesses user-generated data for diverse applications, such as environmental monitoring, urban planning, and healthcare [3]–[5].

Federated learning (FL) is a distributed approach to artificial intelligence, where devices collaboratively train a shared model with private data, so as to split the computational burden [6]. In context-aware applications, FL can be superimposed to sensing as a collaborative approach for efficient extraction of ambient information from data, leveraging the collective intelligence and computational power of a diverse set of devices to train machine learning models without a costly centralization [7].

Leveraging the distribution of training tasks and shifting the computation from cloud servers to local edge devices reduces the energy consumption associated with transmitting large amounts of data to a central server for processing [8]. As the learning process takes place directly on or close to the participating devices, energy requirements are decreased [9].

Thus, we investigate a combined paradigm of data management, where both sensing and extrapolation of meaning from sensed data are performed by the end nodes, and the role of the network coordinator is to merge the local decision process. Such an integration of participatory sensing and FL can be analyzed through the lens of game theory [10], [11], whose main purpose is to study the strategic interactions among self-interested entities. This can shed light on the resulting system dynamics and highlight relevant trade-offs.

We consider a scenario, where N end nodes are connected to a central data sink that coordinates their federation. Each end node has access to a *private* dataset, either sensed directly from the environment, or assigned by the data sink as a partition of a global dataset, and partake in the learning process on a *voluntary* basis [12], [13]. Their decision about whether to participate or not is iterated through multiple rounds with a probabilistic approach [14]. We assume all nodes to behave identically and not exchange information with each other, thus they have identical and independently distributed (i.i.d.) probability p_i to participate in each round, computed locally. This can be extended to correlated/communicating nodes along the lines of [15].

The common approach is to handle the participation in a centralized fashion at the sink [16], which also assigns the data to the end nodes for local training and handles the federation by merging the resulting parameters. Even in this context, it would be convenient to alternate the participation of nodes over multiple rounds to obtain faster convergence of the FL, avoiding overfitting or entrapment, which would unnecessarily increase the duration of the training as well as cause overconsumption of energy [17], [18]. Thus, an optimal probability of participation for the single node can be computed, which minimizes the time to reach a target accuracy of the FL.

In a fully distributed and participatory context, which is the approach of choice if nodes perform their measurements locally [2], we can consider that nodes act based on their selfish objective, individually computed, minimizing a linear combination of duration and cost. This implementation resounds

This work was supported by the Italian PRIN project 2022PNRR "DIGIT4CIRCLE," by the Spanish project PID2020-113832RB-C22(ORIGIN)/MCIN/AEI/10.13039/50110001103, European Union Horizon 2020 research and innovation programme under Grant Agreement No. 953775 (GREENEDGE) and the grant CHIST-ERA-20-SICT-004 (SONATA) by PCI2021-122043-2A/AEI/10.13039/501100011033

as a *Tragedy of the Commons* scenario [19], since in large IoT systems the end nodes are expected to realize that their individual contribution gives little benefit to the overall duration of the task, and opt for reduced action. This implies an overall suboptimal participation rate throughout the entire network, and a high Price of Anarchy (PoA) [20].

In this spirit, game theory may allow us to exploit participatory sensing and FL to their fullest. Strategic interactions and incentives of participants can ensure effectiveness and energy sustainability of collaborative systems [21] and our study paves the way for designing less power-hungry data-driven applications that benefit from the collective intelligence of diverse participants.

The rest of this paper is organized as follows. In Section II, we discuss related work. Section III describes the system model and discusses our proposed approach to combine federated learning with participatory sensing. We show numerical results in Section IV and we finally conclude in Section V.

II. Related Work

Development of federated learning strategies and minimization of energy consumption are tight-knit goals for future networks exploiting the intelligence available at the end nodes [8]. For example, [17] proposes an online energy-aware dynamic worker scheduling policy, which maximizes the average number of workers scheduled for gradient update at each iteration under a long-term energy constraint. In [22], the reasoning about the energy consumption being related to the use computation resources is also extended to include the wireless communication exchanges. Such proposals compare well with what we consider to be the optimal centralized allocation in the following.

At the same time, participatory sensing and federated learning are scenarios where the actions of individual agents mutually influence the outcome, and it makes sense to invoke an application of game theory [12], [20]. In this context, the shared goal of the activity is at odds with the individual cost paid by the users, which can be declined in multiple ways, e.g., to security and privacy [23] of data owners.

We argue that the quintessential motivation behind distributed approaches to sensing and learning lies in optimizing the resource usage [16], simultaneously aiming to save communication overhead and energy [24].

In [11], the problem of recruiting participants for a federated learning is studied from a game theoretic perspective. The starting point is similar to what we argue in this paper, that is, rational data owners may be unwilling to participate in a collaborative learning process due to excessive resource consumption. Therefore, the authors propose some incentives for their participation designed through game theoretic mechanisms. However, the paper takes a high level perspective, where the motivation behind the individual nodes requiring incentives lies more in their selfishness towards an economic advantage. Instead, we consider this to be more directly related to energy consumption, which in our opinion gives a stronger justification even to well-meaning data owners.

The investigation of incentive design is also traditionally approached through auctions, capturing the supply and demand interaction [13]. Similarly, [10] explores the Nash equilibria (NEs) of a participatory intervention of the nodes minimizing age of information (AoI).

Differently from these contributions, we argue that the problems of participation in federated learning are not just related to the interaction of data exchange and resource utilization, but more fundamentally rooted in the energy consumption [9]. To obtain a factual energy saving from a participatory paradigm, the objective must be shared by all nodes, and the efficiency of strategic choices must be evaluated from this standpoint, i.e., to see whether uncoordinated nodes can harmonize their operation, without unnecessary energy inefficiencies.

Despite the importance of participatory federated learning, this particular issue was scarcely addressed in the literature, which is why our contribution fills a gap and validate such an approach for IoT scenarios.

III. System Model

We consider an FL scenario with a set $\mathcal{N} = \{1 \dots N\}$ of nodes that need to collaborate to the same learning task, communicating their model weights to a central receiver that acts as a sink. The receiver is in charge of collecting the weights from the participating nodes and to combine them, to obtain a global training model. In particular, we consider the FedAvg algorithm [25] where each node $i \in \mathcal{N}$ trains for E local epochs with a loss function ℓ_i on its local dataset the global model and then shares the model updates with the central server via an IEEE 802.11ax wireless link. In the algorithm formulation the receiver, i.e, a central server, is in charge of selecting the subset of clients that participate in the current round.

We give a special twist on this scenario, giving to each node the ability to control its own probability p_i to participate in the task. Differently from approaches where a globally optimal involvement is sought for the end nodes [14], [16], here the participation probability is computed by each node individually, based on local evaluations about the time for task completion and energy consumption. This probability is set a priori and cannot be changed later on. The task is successful when a target validation accuracy is reached at the receiver's side. At every round, each node decides whether to participate in the learning task according to the local probability. Thus, only a subset of nodes $\mathcal{P}^t$ participates in learning round t. Starting from the model in [24], we evaluate the energy consumption of each node depending on the decision to participate in FL round t or not. If client i participates, the energy spent for training is:

$$\mathcal{E}^t_{\mathrm{train},i} = P^t_{\mathrm{hw},i} T^t_{\mathrm{train},i}, \tag{1}$$

where $P^t_{\mathrm{hw,i}}$ is the average power drained by the hardware, i.e., CPU, GPU, and DRAM, and $T^t_{\mathrm{train},i}$ is the training process duration. The energy spent for communication is

$$\mathcal{E}_{\mathrm{tx}} = P_{\mathrm{tx}} T_{\mathrm{tx}}, \tag{2}$$

where P_{tx} and T_{tx} are transmission power and time, respectively. Note that $\mathcal{E}_{\mathrm{tx}}$ is the same for each client in each FL round since the size of the model updates is constant during the whole learning process. Finally, all participating nodes wait for the conclusion of the current FL round. Let T_{round} be the maximum round duration defined by the sink. All participating devices have to start the upload process within T_{round}, otherwise their contribution will be discarded. The idle energy is computed as

$$\mathcal{E}^t_{\text{idle},i} = P_{\text{idle},i}(T_{\text{round}} - T^t_{\text{train},i}), \tag{3}$$

with $P_{\text{idle},i}$ the power consumed while idling and $T_{\text{round}} - T^t_{\text{train},i}$ the idling time. The total energy is given by

$$\mathcal{E}^t_i = \mathcal{E}^t_{\text{train},i} + \mathcal{E}_{\text{tx}} + \mathcal{E}^t_{\text{idle},i}. \tag{4}$$

If a node j does not join the current FL round its energy consumption is given by the idle energy

$$\mathcal{E}_j = P_{\text{idle},j} T_{\text{round}} \tag{5}$$

since the node is waiting for the conclusion of the round. The total energy consumption is

$$\mathcal{E}^t = \sum_{i\in\mathcal{P}^t} \mathcal{E}^t_i + \sum_{j\in\mathcal{N}\setminus\mathcal{P}^t} \mathcal{E}_j. \tag{6}$$

The energy consumed for d rounds is

$$\mathcal{E} = \sum_{t=1}^{d} \mathcal{E}^t. \tag{7}$$

The energy consumption is minimized for the lowest duration d of the FL task, which is concluded when a target value for the validation accuracy is consistently met [22]. This is dependent on the mean number of nodes participating in a given round, as we will later argue in Section IV. The probability that a given number of nodes are active at any given time follows a Poisson-Binomial distribution, because participation of each node is independent of the others. The learning task duration follows a Poisson-Binomial distribution $\mathcal{D} \sim PoiBin(p_i)$, $i{\in}\mathcal{N}$ and the duration value $d(k)$ is a function of the number of participating nodes k. It follows from probability theory that the first moment of this distribution is calculated as

$$\mathbb{E}[\mathcal{D}] = \sum_{i=0}^{N} d(i)\cdot P[m=i], \tag{8}$$

where $P[m = i]$ is the probability that exactly i nodes participate, computed in closed form as [26]

$$P[m] = \frac{\sum_{n=0}^{N}\left\{e^{\frac{-j2\pi nm}{N+1}} \prod_{k=1}^{N}[p_k(e^{\frac{j2\pi n}{N+1}}-1)+1]\right\}}{N+1}. \tag{9}$$

We formalize this distributed optimization task as a static game of complete information $\mathcal{G} = \{\mathcal{N}, \mathcal{A}, \mathcal{U}\}$, where $\mathcal{N}$ is the set of players, i.e., the nodes in the network, $\mathcal{A}$ is the set of actions, namely the participation probability $p_i \in [0, 1]$ for each player i, and $\mathcal{U}$ is the set of the utilities for each player [21]. In order to promote client participation, we implement an incentive mechanism based on the expected AoI of the single node, computed as the ratio of the second order moment and two times the first order moment of the inter-participation time Y as a function of the node's participation probability.

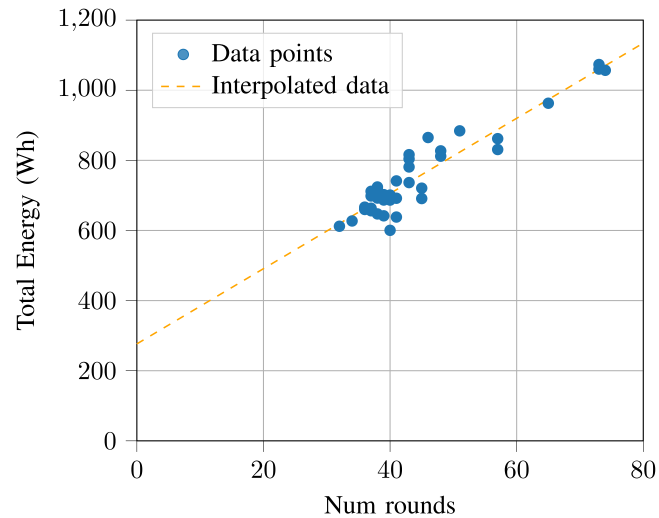

Fig. 1. Total energy spent ($\mathcal{E}$) vs. number of rounds to converge (d).

$$\mathbb{E}[\delta_i] = \frac{\mathbb{E}[Y^2]}{2\mathbb{E}[Y]} = \frac{1}{p_i} - \frac{1}{2}. \tag{10}$$

We define the utility of each player i as

$$u_i = -\mathbb{E}[\mathcal{D}] - \gamma \log(\mathbb{E}[\delta_i]) - cp_i, \tag{11}$$

where $\mathbb{E}[\mathcal{D}]$ follows (8), $\log(\mathbb{E}[\delta_i])$ is the natural logarithm of (10), γ is the weight assigned to the incentive and c is a cost factor taking into account the energy consumption for the node when it participates. This equation combines the two goals of the clients: minimizing the number of rounds to reach convergence and the associated participation cost [3]. It furthermore rewards the clients when they decide to participate actively in the federated learning task.

The NE is found through a one-sided optimization of the utility, which implies that each player takes a *best response* to the unchanged actions of the others [10]. Thus, we solve the system of differential equations

$$\frac{\partial u_i}{\partial p_i} = 0, \quad i = 1, \ldots, N \tag{12}$$

which can be obtained in closed form from (8), (9) and (10). Due to symmetry, this will result in the same value p for all nodes, i.e., $p = p_1 = \ldots = p_N$.

We further calculate the Price of Anarchy (PoA) to evaluate how much a decentralized solution deteriorates the optimal centralized one's performance. The PoA is calculated as [19]

$$PoA = \frac{\max_{s\in\text{NE}} u(s)}{\min_{s\in\mathcal{S}} u(s)} \tag{13}$$

where, at the numerator, we take strategy s from the set of all NEs with the highest cost, and at the denominator, the optimal centralized strategy that minimizes the cost.

	Parameter	Description	Value
Fed. Learning	$\|w\|$	Number of model parameters	11 181 642
	S_w	ResNet-18 model parameters size	44.73 MB
	η	Learning rate	0.01
	N	Number of total clients	50
	E	Local epochs number	5
	T_{round}	Maximum training time	10s
	ℓ_i	Local loss function	Sparse Cat. Crossentropy
	T_{acc}	Target accuracy on CIFAR-10	0.73
	P_{idle}	Idle power consumption	96.85 W
Communication (IEEE 802.11ax)	P_{tx}	Tx power for edge devices	9 dBm
	σ_{leg}	Legacy OFDM symbol duration	4 µs
	N_{sc}	Number of subcarriers (20 MHz)	234
	N_{ss}	Number of spatial streams	1
	T_{e}	Empty slot duration	9 µs
	T_{SIFS}	SIFS duration	16 µs
	T_{DIFS}	DIFS duration	34 µs
	T_{PHY}	Preamble duration	20 µs
	$T_{\text{HE-SU}}$	HE single-user field duration	100 µs
	L_s	Size OFDM symbol	24 bits
	L_{RTS}	Length of an RTS packet	160 bits
	L_{CTS}	Length of a CTS packet	112 bits
	L_{ACK}	Length of an ACK packet	240 bits
	L_{SF}	Length of service field	16 bits
	L_{MAC}	Length of MAC header	320 bits
	CW	Contention window (fixed)	15

TABLE I
SIMULATION PARAMETERS.

IV. PERFORMANCE EVALUATION

A. Experimental Setup

We consider a scenario with $N = 50$ nodes that collaboratively learn ResNet-18 [27], to correctly classify the images in the CIFAR-10 [28] dataset, containing 50 000 training samples that are randomly but fairly divided across all nodes. The input features are 32×32 colored images divided into 10 classes. To evaluate the global model, we extract 7000 samples as a validation set from the test partition of CIFAR-10. ResNet-18 ($|w|$) has 11 181 642 trainable parameters, which require 44.73MB (S_w) if stored as *float32*. In every round, each client executes $E = 5$ local epochs of Stochastic Gradient Descent using sparse categorical accuracy as local loss function (ℓ_i). Convergence is reached when the global model validation accuracy (T_{acc}) is ≥ 0.73 for 3 consecutive rounds to avoid performance spikes. When the training is concluded, each node shares the model update with the central server using transmission power (P_{tx})=9 dBm. Simulation parameters are reported in Table I and a comprehensive description of the communication model can be found in [24]. Experiments run on a server equipped with two Intel Xeon 6230, 188 GB of RAM, and an RTX 2080 Ti.

B. Numerical Results

We now report the results of experiments and simulations. To obtain a realistic model for FL from the perspective of game theory, we executed a wide array of simulations. First, we measured the energy consumption ($\mathcal{E}$) and the number of rounds to reach convergence for several participation probabilities ($p_i \in [0.1, 0.7]$) on the scenario described before. Table II(a) shows the number of rounds to reach convergence (d) and the total energy consumption ($\mathcal{E}$). To assess the training energy ($\mathcal{E}_{\text{train}}$), we use Codecarbon [29] a Python library that estimates

p_i	$\mathcal{E}$	d
0.100	1056.81	74
0.125	1060.25	73
0.130	830.90	57
0.150	1073.33	73
0.160	962.90	65
0.175	600.42	40
0.200	861.87	57
0.225	691.04	45
0.250	638.27	41
0.300	720.66	45
0.350	641.78	39
0.400	691.90	41
0.410	811.87	48
0.420	647.21	38
0.430	736.57	43
0.440	686.69	40
0.450	827.07	48
0.460	884.16	51
0.470	698.03	40
0.480	700.97	40
0.490	686.84	39
0.500	689.25	39
0.510	656.18	37
0.520	660.68	37
0.530	663.44	37
0.540	702.24	39
0.550	741.38	41
0.560	781.14	43
0.570	692.42	38
0.580	659.89	36
0.590	662.56	36
0.600	627.10	34
0.610	666.57	36
0.620	707.24	38
0.630	804.00	43
0.640	865.10	46
0.650	716.03	38
0.660	698.39	37
0.670	816.24	43
0.680	724.07	38
0.690	612.04	32
0.700	711.64	37

(a) One single seed

p_i	$\overline{d}$	$\sigma(d)$	$\overline{\mathcal{E}}$	$\sigma(\mathcal{E})$
0.100	74.50	11.47	1072.14	123.43
0.125	68.00	13.09	1005.97	140.49
0.130	56.00	5.29	862.84	60.19
0.150	62.50	8.81	950.26	100.14
0.160	57.25	6.13	887.80	61.31
0.175	51.00	9.42	797.18	145.67
0.200	51.00	4.55	816.96	37.86
0.225	45.50	3.70	747.44	54.52
0.250	51.00	9.56	803.96	132.64
0.300	46.75	2.75	768.25	41.50
0.350	43.00	5.23	724.40	73.21
0.400	43.25	2.22	734.25	33.22
0.410	44.50	5.32	758.88	62.29
0.420	42.75	4.11	725.76	59.45
0.430	42.75	3.30	734.69	35.41
0.440	43.00	4.08	732.95	49.07
0.450	43.50	4.43	751.96	61.11
0.460	42.75	5.56	750.14	89.77
0.470	39.50	3.11	698.25	33.15
0.480	39.25	6.70	696.30	71.74
0.490	40.67	2.89	709.99	33.48
0.500	40.00	0.82	704.10	11.11
0.510	41.75	3.30	719.96	43.71
0.520	42.50	7.33	729.13	81.90
0.530	40.00	3.16	703.01	37.23
0.540	41.75	4.27	726.11	44.34
0.550	39.50	2.65	706.41	35.12
0.560	40.25	2.99	719.03	48.51
0.570	40.50	4.43	712.93	46.15
0.580	46.25	14.15	771.83	152.41
0.590	39.00	2.58	694.74	27.70
0.600	39.00	4.24	691.24	51.19
0.610	37.75	2.87	682.34	30.05
0.620	39.75	5.56	708.59	58.31
0.630	37.75	3.50	697.93	70.71
0.640	39.75	5.91	726.61	102.68
0.650	39.00	2.16	702.75	23.75
0.660	40.75	4.99	719.79	48.48
0.670	40.00	4.69	725.12	75.90
0.680	41.25	4.03	728.89	36.60
0.690	37.50	3.87	676.75	45.17
0.700	38.25	5.50	696.29	59.19

(b) Average results

TABLE II
NUMBER OF ROUNDS (d) AND ENERGY ($\mathcal{E}$) IN WH SPENT TO CONVERGE WITH DIFFERENT PARTICIPATION PROBABILITIES.

the hardware power consumption as per (1). The results show that low participation probabilities, i.e., $p_i \in [0.15, 0.3]$ do not lead to good performance. The best performance is achieved by $p_i = 0.69$ that converges in 32 rounds, with 612.04 Wh of energy spent. Fig. 1 shows an approximately linear trend between d and $\mathcal{E}$, justifying the assumption of Section III.

To increase the generality of our results, we repeated the simulation with different random seeds. This influences both the decision of the nodes to join the current round (or not), and the initialization of the global model. Simulation runs are executed on the same hardware described before using four RTX2080 Ti in parallel. Due to Codecarbon limitations, it is not possible to distinguish the energy consumption of different processes running in parallel on the same machine, so the total energy is estimated from the linear interpolation shown in Fig. 1. Table II(b) reports the average and standard deviation for the number of rounds and energy to converge.

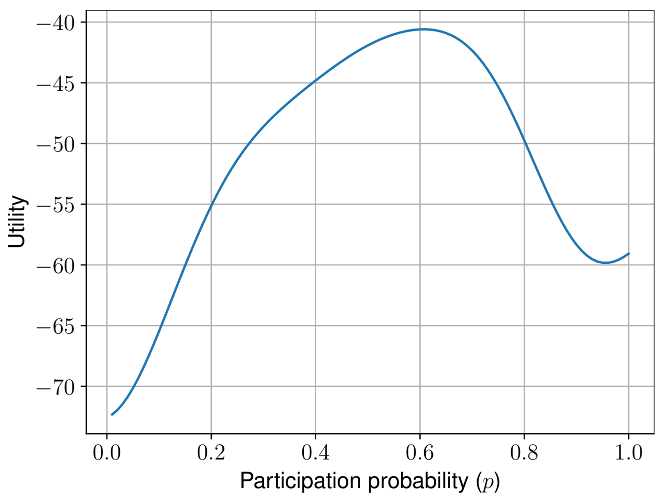

Fig. 2. Utility from a fit of the FL simulation, applying (11) with $c = 0$.

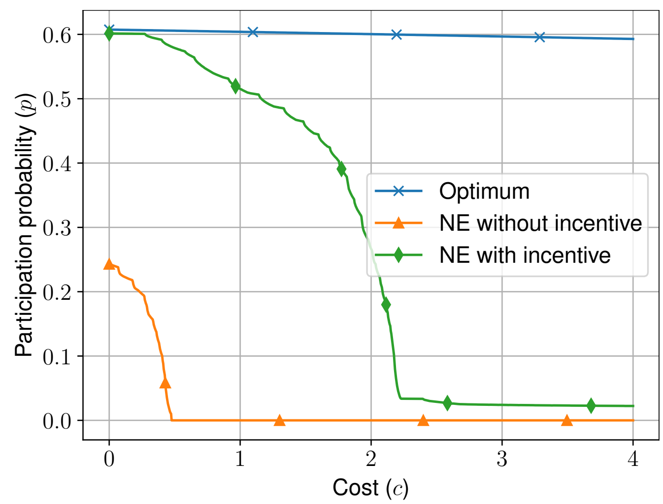

Fig. 4. Nodes' participation probability in the optimal centralized solution and at the NE with and without incentive.

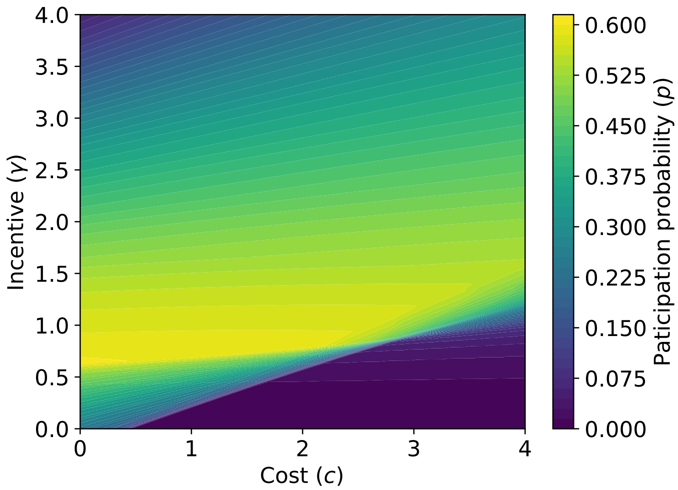

Fig. 3. NE solution of the participation probability for various cost factors c and incentive weights γ.

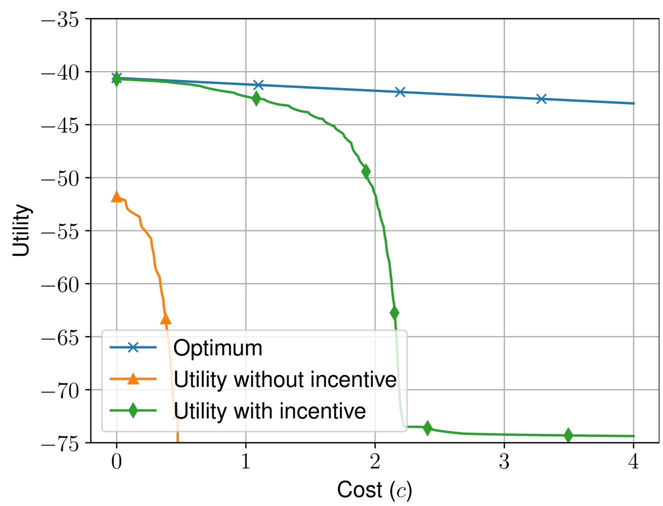

Fig. 5. Utility obtained by the optimal centralized nodes' participation probability and at the NE for various values of the c parameter.

To obtain results for the theoretical setup, we modeled the duration of the FL task using the data in Table II(b). Specifically, we used a polynomial regression model to fit random points with a normal distribution as per mean and standard deviation of d reported in the table. With said model, we calculated the utility as in Fig. 2 for c=0 and γ=0.

Fig. 3 is a contour plot that shows the NE solution for p with different values for the cost factor c and the incentive weight γ. As expected, the nodes start to collaborate more by giving more weight to the incentive offsetting the cost they need to pay for the participation at the FL task. However for higher values of γ it is evident that there is a tradeoff between the cost and the incentive. This is probably due to the shape of our utility function that penalizes the nodes if all of them participate in the task. For the following plots we chose to fix $\gamma \approx 0.6$ as it is the value that obtains the highest participation probability.

Fig. 4 shows the participation probability chosen by the nodes by global maximization of the utility and the NE as obtained by solving (12). For $c = 0$ the optimal participation probability is approximately $p = 0.61$, while the NE without incentive obtains $p = 0.24$. For increasing cost values this solution quickly falls to $p = 0$. This is an evident *Tragedy of the Commons* [19], since selfish users share a common resource, yet their individual participation has little influence on the payoff, so they do not cooperate to a full extent, damaging both the collective and ultimately their own interest. The incentive mechanism based on AoI proves to be effective in improving nodes collaboration as the highest probability is $p = 0.6$. This solution then quickly falls to lower participation as the incentive is not enough to offset the cost but it never reaches $p = 0$.

Fig. 5 shows the evolution of the utility obtained by the solutions. Here it is evident the sudden drop in the NE solution without incentive as soon as the participation probability approaches 0. Conversely, the NE with incentive has a significant drop from the optimum, but then becomes stable. By considering the PoA reported in Fig. 6 we can appreciate

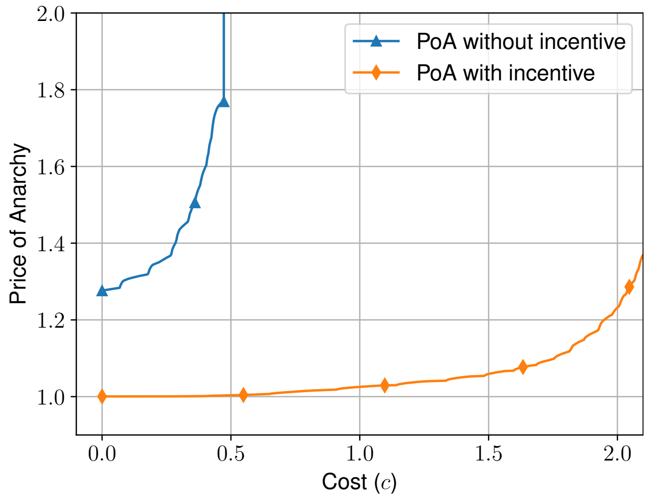

Fig. 6. PoA as in (13) for increasing values of the parameter c.

the importance of the incentive on the NE solution. Not having any kind of incentive obtains $PoA \simeq 1.28$ for $c = 0$, this means that a distributed solution with these constraints has a 28% loss over the performance of a centralized participation schedule. This becomes even worse for higher cost factors and eventually explodes to infinity. The NE with incentive based on AoI manages to obtain $PoA \approx 1$ and does not grow as rapidly as the previous case meaning that this solution allows transmissions even when they become more costly.

V. Conclusions

We explored a collaborative federated learning scenario, where individual nodes are in control of their participation to the learning task. The overall objective is to minimize the task duration and consequently the energy consumption. We tackled this challenge through the instruments of game theory [12] deriving the NE of the resulting allocation, and ultimately computing its energetic performance.

A fully distributed optimization is not feasible without an incentive mechanism in place, as in the best case there would be a 28% performance drop with respect to a centralized solution.

Thus, we introduced an incentive mechanism for the nodes based on AoI that is able to offset the participation cost that they need to sustain [11]. A future extension of this work may consider other more sophisticated incentive mechanisms.

References

[1] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," in *Proc. IEEE PERCOM Wkshps*, 2014, pp. 593–598.

[2] P. Bellavista, L. Foschini, and A. Mora, "Decentralised learning in federated deployment environments: A system-level survey," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–38, 2021.

[3] N. Michelusi, K. Stamatiou, L. Badia, and M. Zorzi, "Operation policies for energy harvesting devices with imperfect state-of-charge knowledge," in *Proc. IEEE ICC*, 2012, pp. 5782–5787.

[4] C. Lanza, E. Angelats, M. Miozzo, and P. Dini, "Urban traffic forecasting using federated and continual learning," in *Proc. IEEE CIoT*, 2023.

[5] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Joint compression of eeg and emg signals for wireless biometrics," in *Proc. IEEE GLOBECOM*, 2018.

[6] N. H. Tran, W. Bao, A. Zomaya, M. N. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. IEEE Infocom*, 2019, pp. 1387–1395.

[7] B. D. Cruz, J. Cheng, Z. Song, and E. Tilevich, "Understanding the potential of edge-based participatory sensing: an experimental study," in *Proc. IEEE VTC-Spring*, 2020.

[8] E. Gindullina, L. Badia, and X. Vilajosana, "Energy modeling and adaptive sampling algorithms for energy-harvesting powered nodes with sampling rate limitations," *Trans. Emerging Telecommun. Techn.*, vol. 31, no. 3, p. e3754, 2020.

[9] Z. Song, J. Cheng, A. Chauhan, and E. Tilevich, "Pushing participatory sensing further to the edge," in *Proc. IEEE EDGE*, 2019, pp. 24–26.

[10] A. Buratto, A. Mora, A. Bujari, and L. Badia, "Game theoretic analysis of AoI efficiency for participatory and federated data ecosystems," in *Proc. ICC Wkshps*, 2023.

[11] X. Tu, K. Zhu, N. C. Luong, D. Niyato, Y. Zhang, and J. Li, "Incentive mechanisms for federated learning: From economic and game theoretic perspective," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 3, pp. 1566–1593, 2022.

[12] V. S. Dasari, B. Kantarci, M. Pouryazdan, L. Foschini, and M. Girolami, "Game theory in mobile crowdsensing: A comprehensive survey," *Sensors*, vol. 20, no. 7, p. 2055, 2020.

[13] M. Cong, H. Yu, X. Weng, and S. M. Yiu, *A game-theoretic framework for incentive mechanism design in federated learning*. Springer, 2020, pp. 205–222.

[14] M. Zhang, G. Zhu, S. Wang, J. Jiang, Q. Liao, C. Zhong, and S. Cui, "Communication-efficient federated edge learning via optimal probabilistic device scheduling," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8536–8551, 2022.

[15] A. Buratto, H. Tuwei, and L. Badia, "Optimizing sensor data transmission in collaborative multi-sensor environments," in *Proc. IEEE COMNETSAT*, 2023.

[16] B. Luo, X. Li, S. Wang, J. Huang, and L. Tassiulas, "Cost-effective federated learning design," in *Proc. IEEE Infocom*, 2021.

[17] Y. Sun, S. Zhou, and D. Gündüz, "Energy-aware analog aggregation for federated learning with redundant data," in *Proc. IEEE ICC*, 2020.

[18] Z. Wang, J. Ma, X. Wang, J. Hu, Z. Qin, and K. Ren, "Threats to training: A survey of poisoning attacks and defenses on machine learning systems," *ACM Comput. Surv.*, vol. 55, no. 7, pp. 1–36, 2022.

[19] L. Prospero, R. Costa, and L. Badia, "Resource sharing in the Internet of Things and selfish behaviors of the agents," *IEEE Trans. Circuits Syst. II*, vol. 68, no. 12, pp. 3488–3492, 2021.

[20] K. Donahue and J. Kleinberg, "Optimality and stability in federated learning: A game-theoretic approach," *Adv. Neur. Inf. Proc. Syst.*, vol. 34, pp. 1287–1298, 2021.

[21] M. Cappelletti, N. Cibin, and L. Badia, "Game-theoretic economic models of duopolies applied to green ICT design," in *Proc. IEEE EPEC*, 2021, pp. 267–272.

[22] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935–1949, 2020.

[23] L. Giaretta, I. Savvidis, T. Marchioro, Š. Girdzijauskas, G. Pallis, M. D. Dikaiakos, and E. Markatos, "PDS 2: A user-centered decentralized marketplace for privacy preserving data processing," in *Proc. IEEE ICDEW*, 2021, pp. 92–99.

[24] E. Guerra, F. Wilhelmi, M. Miozzo, and P. Dini, "The cost of training machine learning models over distributed data sources," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1111–1126, 2023.

[25] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artif. Intell. Stat.* PMLR, 2017, pp. 1273–1282.

[26] M. Fernandez and S. Williams, "Closed-form expression for the Poisson-binomial probability density function," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 2, pp. 803–817, 2010.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[28] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. of Toronto, Toronto, ON, Canada, Tech. Rep., 2009.

[29] V. Schmidt, K. Goyal, A. Joshi, B. Feld, L. Conell, N. Laskaris, D. Blank, J. Wilson, S. Friedler, and S. Luccioni, "CodeCarbon: Estimate and Track Carbon Emissions from Machine Learning Computing," https://github.com/mlco2/codecarbon, 2021.