Resource Allocation of NOMA Communication Systems for Federated Learning

Marija Poposka, *Student Member, IEEE*, Borche Jovanovski, Valentin Rakovic, *Senior Member, IEEE*, Daniel Denkovski, and Zoran Hadzi-Velkov, *Senior Member, IEEE*

***Abstract*—Federated learning is a new communication and computing concept that allows naturally distributed data sets (e.g., as in data acquisition sensors) to be used to train global models, and therefore successfully addresses privacy, power and bandwidth limitations in wireless networks. In this paper, we study the communications problem of latency minimization of a multi-user wireless network used to train a decentralized machine learning model. To facilitate low latency, the wireless stations (WSs) employ non-orthogonal multiple access (NOMA) for simultaneous transmission of local model parameters to the base station, subject to the users' maximum CPU frequency, maximum transmit power, and maximum available energy. The proposed resource allocation scheme guarantees fair resource sharing among WSs by enforcing only a single WS to spend the maximum allowable energy or transmit at maximum power, whereas the rest of the WSs transmit at lower power and spend less energy. The closed-form analytical solution for the optimal values of resource allocation parameters is used for efficient online implementation of the proposed scheme with low computational complexity.**

***Index Terms*—wireless federated learning; non-orthogonal multiple access; latency minimization.**

I. Introduction

The next-generation wireless networks are envisioned to integrate communication and computing by employing artificial intelligence [1]. The novel concept of federated learning (FL) is suitable for these networks because the massive amounts of data acquired by the wireless (e.g. sensor) nodes can be exploited without transferring them over the wireless channel into the network cloud but rather processing them locally. Wireless federated learning (WFL) is a novel concept for distributed machine learning that helps wireless stations (WSs) to collaboratively build a shared learning model while locally preserving their own training data [2]. In particular, WSs train their local models using their local datasets over multiple training rounds and then offload the model parameters to the central server (i.e., the base station - BS) over the wireless channel. The BS aggregates the locally updated model parameters into a single set of model parameters, which is used to update the global learning model and transmit it back wirelessly to the WSs. The "federated averaging" scheme combines the local updates by simply averaging them in the central server [3].

The particularities of the wireless channel must be taken into account when designing WFL systems. Due to path loss, WSs at different distances to the BS can achieve different transmission rates, which means that it would take different times for different WS to transmit the same amount of information (i.e., the locally updated model parameters). Therefore, the convergence time of the federated learning should consider both the local processing time in the WSs and the transmission time from WSs to the BS. Papers [4] and [5] propose schemes that minimize the convergence time in a WFL system based on orthogonal frequency division multiple access (OFDMA) and FDMA, respectively. To facilitate low latency, the WSs can employ non-orthogonal multiple access (NOMA) for simultaneous transmission of their local model parameters to the BS. WFL systems employing NOMA have been studied in recent papers [6]-[9]. In [6], WFL uses NOMA for "over the air" aggregation/computation of uncoded analog transmissions from multiple users and minimizes the learning optimality gap. Paper [7] maximizes the weighted-sum rate over the NOMA uplink, subject to the maximum power constraint of the WSs. Paper [8] minimizes the communication latency of the NOMA-based WFL system, but neglects the maximum power constraint and the maximum central processing unit (CPU) frequency of the WS, both of which have a significant effect on the system design. Paper [9] proposes a user selection scheme that ensures selected NOMA users can achieve sufficiently high rates to provide a quality update to the central model.

In this paper, we propose a novel resource allocation scheme for a WFL system based on NOMA, which minimizes the sum of the durations of the communications phase and the local computation phase during the machine learning process. Due to its fairness properties [10], NOMA is a suitable multiple access method for WFL, because, regardless of their distance to the central server, the WSs typically offload the same amount of information (i.e., the model parameters of their identical local learning models). The proposed scheme considers both the power and the energy constraint imposed on each WS, as well as, the CPU frequency of the WS. We present an original analytical method for solving the resource allocation problem, which relies on a novel expression for calculating the transmit powers of the WSs, indexed according to the order of successive interference cancellation (SIC). As outlined in the Numerical results section, compared to the time division multiple access (TDMA), NOMA can more successfully handle the low latency requirements of the considered WFL system in the relevant range of system parameters [11].

II. System model

A. Wireless FL model

The considered WFL system consists of a single BS and K WSs. The BS has a central server that hosts the global learning model, whereas each WS hosts the local learning model and its own dataset. From the communications aspect, the BS is equipped with a single transmitting/receiving antenna, and

Manuscript received March 23, 2023; revised May 8, 2023 and June 9, 2023; accepted June 11, 2023. This work was supported in part by the EU Horizon 2020 program under WideHealth project (No. 952279). The editor coordinating the review of this paper and approving it for publication was Zhaohui Yang. *(Corresponding author: Marija Poposka.)*

The authors are with the Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University, 1000 Skopje, Macedonia (email: {poposkam, borchej, valentin, danield, zoranhv}@feit.ukim.edu.mk).

each WS is also equipped with a single antenna. The base station collects the local model parameters from the WS and applies a FL strategy (e.g. "federated averaging"), for updating the central model [3]. The training process of both global and local models runs over multiple training rounds (TRs), during which the global and the local models converge. Each TR consists of 2 phases: local computation phase, and communication phase. We assume that the kth WS has available local data set $\mathcal{D}_k$, that consists of D_k samples. During the local computation phase, each WS trains its local model using its local dataset and ML updating method (such as the stochastic gradient descent method for numerical optimization). Next, during the communications phase, all WSs simultaneously transmit their local model parameters to the BS by employing NOMA. The FL round is completed by the BS aggregating the local model parameters, updating the central model, and broadcasting these parameters back to the WSs.

B. Local computation phase

In the local computation phase, the kth WS trains its local model over I_k iterations, starting with model parameters' values received from the BS at the end of the previous TR. For the kth WS, its CPU frequency is denoted by f_k (CPU cycles/second), the computation load needed to process a single data sample is denoted by L_k (CPU cycles/sample), and the number of iterations in each local computation phase in any given TR is denoted by I_k. Therefore, $a_k = I_k L_k D_k$ is the total number of CPU cycles needed by the kth WS to realize its local computation in one TR. We assume $a_k = a_0,\ \forall k$, in which case the duration of the local computation of the kth WS is given by $\tau_k = a_0/f_k$. The amount of energy consumed by the kth WS during the local computation phase within a single TR is determined by $E_k^{LC} = \alpha a_k f_k^2$, where α is the energy efficiency coefficient that depends on the CPU architecture of the WS. Since all WSs transmit simultaneously, the duration of the local computation phase, τ_0, is calculated as the maximum of the individual computation phases of all WSs, i.e.,

$$\tau_0 = \max_{1\le k\le K}\left\{\frac{a_0}{f_k}\right\}. \tag{1}$$

C. Transmission phase

The communications phase consists of the uplink transmission phase (from WSs to the BS) and the downlink transmission phase (from BS to the WS)[1]. The uplink transmission has duration t_0 and is realized by simultaneous transmission of all WS to the BS utilizing NOMA. The order of information decoding at the BS corresponds to the order of decreasing gains of the respective channels from the WSs to the BS, $h_k (1 \le k \le K)$. Assuming the WS are indexed according in increasing order of their channel gains, $|h_1| \le |h_2| \le \cdots \le |h_K|$, BS decodes the information from the employment of SIC in decreasing order WS index, thus removing the interference from previously decoded WSs. The BS first decodes the codeword transmitted by WS K, while being exposed to interference from all other WSs, then subtracts the Kth WS's decoded signal from the received signal, then decodes the codeword transmitted by WS $K-1$, as so on. In this case, the achievable rate of kth WS is given by

$$R_k = B\log_2\left(\frac{|h_k|^2 p_k}{\sum_{j=1}^{k-1}|h_j|^2 p_j + N_0}\right),\ 1\le k\le K \tag{2}$$

where B is the communication bandwidth of the uplink channel, N_0 is the thermal noise power at the BS receiver, and p_k is the transmit power of kth WS.

The amount of data comprising the local model parameters of kth WS, b_k, that need to be offloaded to the BS are equal to b_0 (i.e., $b_k = b_0, \forall k$). Given the rate R_k, the time needed by the kth WS to offload the model parameters is $t_k = b_0/R_k$. The energy needed by the kth WS to transmit this much data is given by $E_k^{WT} = p_k t_k$. Due to NOMA, we impose a constraint $t_k = t_0, \forall k$, in which case, the duration t_0 of the uplink transmission phase is calculated by [12, Lemma 1]

$$t_0 = \frac{k b_0}{B\log_2\left(\sum_{i=1}^{k} p_i x_i + 1\right)},\quad 1\le k\le K, \tag{3}$$

where $x_i = |h_i|^2/N_0$. In (3), t_0 must have the same value for any k, yielding a set of K equations that must be satisfied (i.e., the constraint $C1$ in the following section).

III. Latency minimization

We aim to minimize the duration of the system latency, T_0, during a single TR, defined as the sum of the duration of the local computation phase, τ_0, determined by (1), and the duration of the transmission phase, t_0, determined by (3). In doing so, we determine the optimal values of the transmit powers, $p_k, \forall k$, and CPU frequencies of the WSs, $f_k, \forall k$. The optimization problem leading to the proposed resource allocation scheme is given by:

$$\begin{aligned}
&\underset{t_0, p_k, f_k}{\text{Minimize}}\ t_0 + \max_{1\le k\le K}\left\{\frac{a_0}{f_k}\right\}\\
&\text{subject to:}\\
&\quad C1: \tfrac{t_0}{k} B\log_2\left(1+\textstyle\sum_{i=1}^{k} x_i p_i\right) = b_0, \forall k\\
&\quad C2: p_k \le P_{max},\ \forall k\\
&\quad C3: \alpha a_0 f_k^2 + p_k t_0 \le E_{max},\ \forall k\\
&\quad C4: f_k \le f_{max},\ \forall k
\end{aligned} \tag{4}$$

where constraint $C1$ refers to (3), and $C2$ refers to the maximum transmit power constraint of each WS, P_{max}. Constraint $C3$ applies to the maximum available energy constraint of each WS, E_{max}, and $C4$ applies to the maximum CPU frequency of each WS, f_{max}. Note, (4) can be transformed into a convex optimization problem (c.f. (16) Appendix A), and its solution is given by the following Theorem.

Theorem 1. *The optimal value of the transmit power of the kth WS is given by:*

$$p_k^* = \frac{e^{kz^*} - e^{(k-1)z^*}}{x_k}, \tag{5}$$

where z^* is given as

$$z^* = \min\left\{z_{01}, z_{02}, \ldots, z_{0K}, g_{01}, g_{02}, \ldots, g_{0K}\right\}, \tag{6}$$

[1]During the downlink transmission, the BS broadcasts the aggregated model parameters to the WS. However, we neglect the duration of the BS broadcast transmission, because the broadcast channel has a much higher capacity compared to the capacity per user of the multiple access channel. Even if the broadcast phase is not neglected, its duration would appear as a constant additive term in the overall system latency and would not affect the considered design optimization [5], [8], [12].

where z_{0k} is determined as the solution of following equation:

$$e^{kz_{0k}} - e^{(k-1)z_{0k}} = P_{max}x_k, \tag{7}$$

whereas g_{0k} is determined together with the CPU frequency f_{0k}, as the solution of the following set of equations:

$$1 - g_{0k}(k-1) - e^{g_{0k}} + kg_{0k}e^{g_{0k}} = 2\alpha a_0 (f_{0k})^3 x_k e^{-g_{0k}(k-1)} \tag{8a}$$

$$\frac{e^{kg_{0k}} - e^{(k-1)g_{0k}}}{g_{0k}} = \frac{x_k \left(E_{max} - \alpha a_0 (f_{0k})^2\right)}{c_0} \tag{8b}$$

The optimal CPU frequency of the kth WS is determined by:

$$f_k^* = \begin{cases} f_{0k}, & f_{0k} < f_{max} \\ f_{max}, & f_{0k} \geq f_{max} \end{cases} \tag{9}$$

The system latency during a single TR is minimized at

$$T_0^* = \frac{c_0}{z^*} + \max\left\{\frac{a_0}{f_k^*}\right\}, \tag{10}$$

where c_0 is defined by

$$c_0 = \frac{b_0 \log(2)}{B}. \tag{11}$$

Proof: Please refer to Appendix A. ■

A. Complexity analysis

Interior-point methods typically solve convex optimization problems as (16) in a number of iteration steps, n_1, which is almost always in the range between $10 < n_1 < 100$ [13, Section 1.3.1]. Each iteration step requires $\max\left\{S_1^3, S_1^2 S_2\right\}$ operations, where S_1 is the total number of variables and S_2 is the total number of constraints [13]. Since $S_1 = 2K + 2$ and $S_2 = 4K$, the complexity for solving (4) is given by $\mathcal{O}\left(n_1 K^3\right)$. On the other hand, for each k, the set of equations (8a)-(8b) can be efficiently solved by the Newton's method [14]. Actually, replacing f_{0k} from (8b) into (8a), we obtain a single (twice continuously differentiable) nonlinear equation in variable g_{0k}, whose single root can be estimated with quadratic rate of convergence [14, Section 2.2]. For example, Newton's method needs approximately 20 iterations to achieve accuracy of six decimal places, and, therefore, similarly to n_1, we can safely assume that the typical number of steps, n_2, needed to solve (8a)-(8b) is almost always in the range $10 < n_2 < 100$.

IV. Numerical results

For the performance evaluation of the proposed WFL scheme, we assume $K = 20$ WSs distributed uniformly along the distance to the BS as $d_k = 50k$ meters $(1 \leq k \leq K)$. Each WS has its local dataset available for FL training of size $D_k = 1000$ samples. The computational load of each WS needed to process one data sample is $L_k = 1000$ CPU cycles per sample, and the number of local training iterations in each TR is $I_k = 100$. The total number of CPU cycles needed by the kth WS to realize its local computation in one TR is $a_k = D_k L_k I_k = a_0 = 10^8$. The computation efficiency is $\alpha = 10^{-28}$, the thermal noise power before the BS receiver is $N_0 = 10^{-14}$W, and the transmission bandwidth is $B = 10$MHz. We assume that the wireless nodes and the wireless environment are nearly static, and thus the wireless signals are only subject to large-scale fading. Even in presence

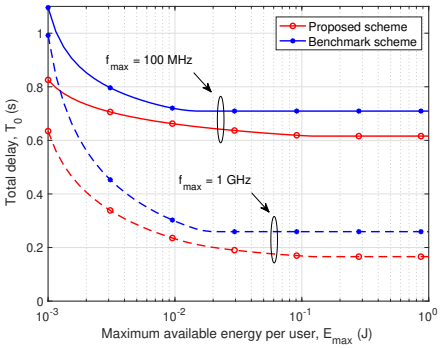

Fig. 1. Impact of maximum available energy E_{max} on latency T_0

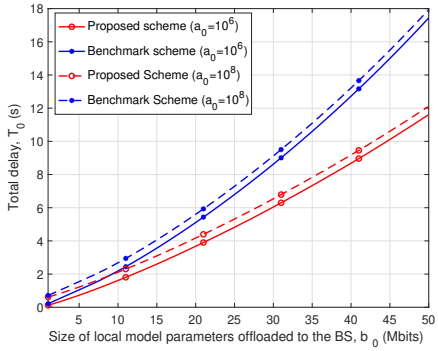

Fig. 2. Impact of the size of local model parameters b_0 on latency T_0

of small-scale (random) fading[2], instead of their instantaneous values, we can use the averaged channel gains dependent on large-scale fading [15]. Specifically, the uplink channel from the kth WS to the BS is exposed only to the deterministic path loss, i.e., $|h_k|^2 = \Omega_k$, where Ω_k is determined by $\Omega_k = 10^{-3} d_k^{-3}$. The model assumes 30dB path loss at a reference distance of 1m, and path loss exponent equal to 3.

To compare the performance of the proposed scheme, we have introduced a TDMA benchmark scheme. The resource allocation of the TDMA-based benchmark scheme is obtained similarly to the NOMA-based scheme, i.e., its parameters are obtained as a solution to an optimization problem analogous to (4), with an objective function given by $\sum_{k=1}^{K} t_k + \max\limits_{1 \leq k \leq K} \{a_0 / f_k\}$, while maintaining constraints $C1$, $C2$, $C3$, and $C4$ as in (4). The duration of the local computation time, τ_0, is unchanged, whereas the communication phase consists of K successive transmissions of duration $t_k (1 \leq k \leq K)$, each allocated to a different WS. This optimization problem is easily transformed into a convex one and then solved numerically, for example, by using the CVX software.[3]

Fig. 1 depicts the system latency T_0 vs. the maximum available energy E_{max}, given $P_{max} = 1$W and $b_0 = 1$Mbit. The two pairs of curves correspond to the proposed and bench-

[2]In the case of slow fading, depending on the parameters associated with the FL model, the values of E_{max} and f_{max} can be adjusted to ensure that the duration of a single TR is less than the channel coherence time (i.e., the channel is constant for the entire TR), c.f. Fig. 1.

[3]It cannot be straightforwardly assumed that NOMA will outperform TDMA, because, unlike TDMA, the concurrent NOMA transmissions are subject to mutual interference, [16].

mark scheme and are plotted for different f_{max} ($f_{max} = 100$ MHz - solid line and $f_{max} = 1$ GHz - dashed line). As the maximum available energy E_{max} increases, each station can transmit at higher transmit power p_k and/or can process at higher CPU frequency f_k (i.e., due to constraint C3). Thus, each WS can deliver the same amount of information for a shorter period of time. However, increasing the maximum available energy E_{max} above a threshold value does not lead to further reduction of T_0, since the latency saturates when E_{max} exceeds a certain energy threshold. The saturation occurs because, although the WSs have sufficient energy at their disposal, the system latency cannot be further reduced due to the peak transmit power P_{max} (c.f. constraint $C2$) and maximum CPU frequency constraint (i.e., due to constraint C4). Compared to the benchmark scheme, the saturation of the proposed NOMA-based scheme occurs for higher values of E_{max}, because when TDMA is used, WSs need much more transmit power in order to deliver the same amount of information, due to the small duration of the successive transmission intervals t_k. Increasing f_{max}, the total duration of the TR is reduced, for both schemes, as the computation phase depends on the corresponding CPU frequency, according to a_0/f_k. Thus, higher f_k reduces latency but increases the WSs' power consumption, and the proposed resource allocation achieves an optimal balance among these two trends.

The impact of the parameters of the learning model on the latency can be illustrated by changing the values of the parameter a_0. Since $a_0 = I_k L_k D_k$, the higher a_0, the higher the training accuracy of the local learning models of each WS. Fig. 2 depicts the system latency T_0 versus the size of the local model parameters offloaded to the BS, b_0, where a_0 appears as a parameter ($a_0 = 10^6$ and $a_0 = 10^8$). The latency increases with the transmit data size for both schemes due to longer times required for wireless transmission. The NOMA-based scheme achieves lower latency than the benchmark scheme, and the gap grows with increasing b_0. This means that the proposed NOMA-based scheme improves its training accuracy performance over the benchmark scheme. If we fix the training duration, our proposed scheme can sustain a higher number of TR, which leads to improved accuracy of the learning model [17]. For widely used FL loss functions, such as linear or logistic loss functions, due to the exponentially decaying dependence between the accuracy of the global WFL model and the number of TRs [17, Eq. (17)], a small increase in the number of TRs leads to very high improvement in accuracy.

V. Conclusion

In this paper, we present a novel resource allocation scheme that minimizes the total latency in NOMA-based wireless FL networks. The proposed resource allocation scheme considers the practical limitations of a wireless network and calculates the relevant system parameters analytically. The analytical solutions facilitate efficient online algorithms when the system operates in a wireless fading environment. The latency of the proposed NOMA-based scheme is lower than the latency of the corresponding TDMA-based benchmark scheme.

Appendix A

Introducing a new optimization variable T_0, the $\min$ operator in the objective function of (4) is converted into an additional constraint $C5$, which transforms (4) as

$$\underset{T_0,t_0,p_k,f_k}{\text{Minimize}}\ T_0$$

$$\begin{aligned}\text{subject to:}\quad & \\ C1:\ & \tfrac{t_0}{k} B \log_2\left(1+\textstyle\sum_{i=1}^{k} x_i p_i\right) = b_0, \forall k \\ C2:\ & p_k \le P_{max},\ \forall k \\ C3:\ & \alpha a_0 f_k^2 + p_k t_0 \le E_{max},\ \forall k \\ C4:\ & f_k \le f_{max},\ \forall k \\ C5:\ & t_0 + \tfrac{a_0}{f_k} \le T_0,\ \forall k\end{aligned} \tag{12}$$

Next, we tackle the constraint $C1$ by rewriting it for two consecutive indices, $k-1$ and k, as follows:

$$1 + \sum_{i=1}^{k-1} p_i x_i = e^{z(k-1)}, \tag{13}$$

$$1 + \sum_{i=1}^{k-1} p_i x_i + p_k x_k = e^{z\,k}, \tag{14}$$

where $z = c_0/t_0$ with c_0 given by (11). Subtracting (14) from (13), we obtain

$$p_k = \frac{e^{zk} - e^{z(k-1)}}{x_k}. \tag{15}$$

Inserting (15) into $C2$ and $C3$, (12) is transformed into the following convex optimization problem:

$$\underset{T_0,z,f_k}{\text{Minimize}}\ T_0$$

$$\begin{aligned}\text{subject to:}\quad & \\ C2':\ & \tfrac{e^{zk}-e^{z(k-1)}}{x_k} \le P_{max}, \quad \forall k \\ C3':\ & \alpha a_0 f_k^2 + \tfrac{c_0}{x_k}\tfrac{e^{zk}-e^{z(k-1)}}{z} \le E_{max}, \quad \forall k \\ C4':\ & f_k \le f_{max}, \qquad \forall k \\ C5':\ & \tfrac{c_0}{z} + \tfrac{a_0}{f_k} \le T_0. \quad \forall k\end{aligned} \tag{16}$$

The functions $\left(e^{zk} - e^{z(k-1)}\right)$ and $\left(e^{zk} - e^{z(k-1)}\right)/z$ in the constraints $C2'$ and $C3'$ are convex functions, which therefore yields the convexity of (16). The solution given by Theorem 1 is obtained by merging the solutions of two special-case optimization problems: an optimization problem with $E_{max} \to \infty$ (special case 1), and an optimization problem with $P_{max} \to \infty$ (special case 2) .

A. Special case 1: $E_{max} \to \infty$

In this case, (16) reduces to:

$$\underset{z}{\text{Minimize}}\left(\frac{c_0}{z}\right)$$

$$\text{subject to:}\quad C2':\ \tfrac{e^{zk}-e^{z(k-1)}}{x_k} \le P_{max},\ \forall k \tag{17}$$

For any k, the left-hand side of $C2'$ is monotonically increasing in z, and, therefore, the constraint $C2'$ is equivalently expressed as $z \le z_{0k}, \forall k$, where z_{0k} is determined as a solution of (7). Thus, (17) is feasible if $z \le z^*$, where its optimal solution, z^*, is determined by:

$$z^* = \min\left\{z_{01}, z_{02}, ..., z_{0K}\right\}. \tag{18}$$

Note, constraint $C2'$ is satisfied with strict equality only for the single WS, whereas the remaining $K-1$ WSs satisfy $C2'$ with strict inequalities. Given (18), the the optimal transmit powers of the WSs are determined by (5).

B. Special case 2: $P_{max} \to \infty$

In this case, introducing the variable change $t_0 = c_0/g$, (16) is transformed as

$$\underset{T_0, g, f_k}{\text{Minimize}}\ T_0$$

subject to:

$$\begin{aligned} C3^{'}:&\quad \alpha a_0 f_k^2 + \tfrac{c_0}{x_k}\tfrac{e^{gk}-e^{g(k-1)}}{g} \le E_{max},\ \forall k \\ C4':&\quad f_k \le f_{max},\quad \forall k \\ C5':&\quad \tfrac{a_0}{f_k} + \tfrac{c_0}{g} \le T_0,\ \forall k \end{aligned} \tag{19}$$

For $k \ge 1$ and $g \ge 0$, the function $\left(e^{gk} - e^{g(k-1)}\right)/g$ exceeds the value of 1, and monotonically increases in both g and k. Note, (19) is feasible if the following condition is satisfied:

$$\frac{x_k E_{max}}{c_0} > 1,\ \forall k. \tag{20}$$

Otherwise, the WSs do not have sufficient energy to transmit the desired amount of information to the BS. Additionally, the solution of $y_0 = \left(e^{gk} - e^{g(k-1)}\right)/g$ with respect to g, denoted by $g_k^*(y_0)$, increases with increasing y_0 ($y_0 > 1$) and decreasing k ($k \ge 1$). Similarly to the special case 1, the constraint $C2'$ is satisfied with strict equality for only a single WS, whereas the rest of $K-1$ WSs satisfy $C2'$ with strict inequalities, in which case, $C2^{'}$ is equivalently expressed as

$$g \le \min_{1 \le k \le K} \left\{ g_k^* \left(\frac{x_k \left(E_{max} - \alpha a_0 f_k^2\right)}{c_0} \right) \right\}. \tag{21}$$

Let us now denote the index of the single WS that satisfies $C2'$ with strict equality by k^*. Clearly, the same WS also satisfies $C4'$ and $C5'$ with strict equality, and the rest of the WSs with inequality. Thus, instead of $3K$ constraints, (19) can be analyzed as a convex optimization problem with only 3 equality constraints. Its Lagrangian is given by

$$\begin{aligned} \mathcal{L} = \quad & T_0 + \lambda \left(\tfrac{e^{gk^*} - e^{g(k^*-1)}}{g} - \tfrac{x_{k^*}\left(E_{max} - \alpha a_0 f_{k^*}^2\right)}{c_0} \right) \\ & + \gamma \left(\tfrac{a_0}{f_{k^*}} + \tfrac{c_0}{g} - T_0 \right) + \mu \left(f_{k^*} - f_{max} \right), \end{aligned} \tag{22}$$

where λ, γ, and μ are non-negative Lagrange multipliers associated with each of the three constraints, respectively. Setting the derivatives of $\mathcal{L}$ with respect to τ_0, g and f_{k^*} to zero, we obtain

$$\begin{aligned} &\tfrac{d\mathcal{L}}{dT_0} = 1 - \gamma = 0 \\ &\tfrac{d\mathcal{L}}{dg} = -\tfrac{\gamma c_0}{g^2} + \lambda \left(\tfrac{k^* e^{k^* g} - (k^*-1)e^{(k^*-1)g}}{g} - \tfrac{e^{k^* g} - e^{(k^*-1)g}}{g^2} \right) = 0 \\ &\tfrac{d\mathcal{L}}{df_{k^*}} = -\tfrac{\gamma a_0}{f_{k^*}^2} + \lambda \tfrac{2\alpha a_0 f_{k^*} x_{k^*}}{c_0} + \mu = 0 \end{aligned} \tag{23}$$

From the first equation of (23), we obtain $\gamma = 1$, which validates that $C3^{'}$ for k^* is satisfied with strict equality. The second equation is transformed as:

$$\lambda = \frac{c_0\ e^{g(1-k^*)}}{1 + g(1-k^*) - e^g + k^* g e^g} \ge 0, \tag{24}$$

which is positive for any $g > 0$ and $k \ge 0$, which validates that $C2^{'}$ for k^* is satisfied with strict equality. Inserting (24) into the third equation of (23), we obtain:

$$\mu = \frac{a_0}{f_{k^*}^2} - \frac{2\alpha a_0 f_{k^*} x_{k^*} e^{g(1-k^*)}}{1 + g(1-k^*) - e^g + k^* g e^g} \ge 0. \tag{25}$$

Scenario 1: Let us assume $f_{k^*} < f_{max}$, yileding $\mu = 0$. Setting (25) to zero, and setting $C2'$ to equality, we obtain the set of equations given in (8). The solution of this set, (g_{0k^*}, f_{0k^*}), is the desired optimal solution if $f_{0k^*} < f_{max}$. Otherwise, $f_{0k^*} = f_{max}$, and g^* is determined by Scenario 2.

Scenario 2: Let us assume that $f_{k^*} = f_{max}$, in which case $\mu > 0$. The optimal value of g is determined directly from the equation (8b). The condition for the validity of this case is obtained from (25) as follows:

$$2\alpha f_{max}^3 x_{k^*} e^{g_{0k^*}(1-k^*)} < 1 + g_{0k^*}(1-k^*) - e^{g_{0k^*}} + k^* g_{0k^*} e^{g_{0k^*}}. \tag{26}$$

If condition (26) is satisfied, the optimal solution is given by (g_{0k^*}, f_{max}). Considering both scenarios, g^* is finally determined by:

$$g^* = \min\left\{ g_{01}, g_{02}, ..., g_{0K} \right\}. \tag{27}$$

Since either $C2'$ or $C3'$ can be active in (4), the general solution presented by Theorem 1 is obtained by merging together (18) and (27), which yields z^* given by (6).

REFERENCES

[1] K. B. Letaief, W. Chen, Y. Shi, J. Zhang and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Comm. Mag.*, vol. 57, no. 8, pp. 84-90, Aug. 2019

[2] Y. Liu *et al.*, "Federated learning for 6G communications: Challenges, methods, and future directions," *China Commun.*, vol. 17, no. 9, pp. 105–118, Sep. 2020

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proc. AISTATS*, vol. 54, 2017, pp. 1273–1282.

[4] M. Chen, H. V. Poor, W. Saad, and S. Cui, "Convergence time minimization of federated learning over wireless networks", *Proc. ICC 2020*, pp. 1-6

[5] Z. Yang *et al.*, "Delay minimization for federated learning over wireless communication networks", *Proc. ICML 2020*, Vienna, Austria, 2020

[6] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE J. Sel. Areas Comm.,* vol. 39, no. 1, pp. 170-185, Jan. 2021

[7] Ma, Xiang, Haijian Sun, and Rose Qingyang Hu. "Scheduling policy and power allocation for federated learning in NOMA based MEC," *2020 IEEE GLOBECOM*, pp. 1-7., 2020.

[8] P. S. Bouzinis, P. D. Diamantoulakis and G. K. Karagiannidis, "Wireless Federated Learning (WFL) for 6G Networks—Part II: The Compute-Then-Transmit NOMA Paradigm," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 8-12, Jan. 2022

[9] H. Sun, X. Ma and R. Q. Hu, "Adaptive federated learning with gradient compression in uplink NOMA," in *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16325-16329, Dec. 2020

[10] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347-2381, Dec. 2017

[11] Z. Chen, Z. Ding, X. Dai and R. Zhang, "An optimization perspective of the superiority of NOMA compared to conventional OMA," *IEEE Trans. Sig. Proc.*, vol. 65, no. 19, pp. 5191-5202, Oct. 2017

[12] F. Fang *et al.*, "Optimal resource allocation for delay minimization in NOMA-MEC networks," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7867-7881, Dec. 2020

[13] S.Boyd, and L.Vandenberghe, "Convex optimization," *Cambridge university press*, 2004.

[14] K. Atkinson, "An introduction to numerical analysis", *John Wiley & Sons*, 2nd Edition, 1991

[15] M. Salehi, H. Tabassum and E. Hossain, "Accuracy of distance-based ranking of users in the analysis of NOMA aystems," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5069-5083, July 2019

[16] Q. Wu, W. Chen, D. W. K. Ng and R. Schober, "Spectral and energy-efficient wireless powered IoT networks: NOMA or TDMA?," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6663-6667, July 2018.

[17] Z. Yang, M. Chen, W. Saad, C. S. Hong and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *IEEE Trans. Wir. Commun.*, vol. 20, no. 3, pp. 1935-1949, March 2021