Relay-Assisted Federated Edge Learning:
Performance Analysis and System Optimization

Lunyuan Chen, Lisheng Fan, Xianfu Lei, Trung Q. Duong, Fellow, IEEE,
Arumugam Nallanathan, Fellow, IEEE, and George K. Karagiannidis, Fellow, IEEE

Abstract

In this paper, we study a relay-assisted federated edge learning (FEEL) network under latency and
bandwidth constraints, where $N$ users collaboratively train a global model assisted by $M$ intermediate
relays and one edge server. For this system, we firstly propose to use the partial aggregation and spectrum
resource multiplexing at the relays for the purpose of enhancing the communication of the relay-assisted
FEEL system. We then perform the system evaluation by deriving analytical and asymptotic expressions
of the system outage probability and convergence rate. For the purpose of improving the system
performance, we further optimize the relay-assisted FEEL network by maximizing the successfully
participated user number in federated learning in each round, through allocating the wireless bandwidth
among users and relays. Specifically, two bandwidth allocation (BA) schemes have been proposed from
either instantaneous channel state information (CSI) or statistical CSI. Simulation results show the
advantages of the proposed BA schemes over other benchmarks in test accuracy and convergence rate
for the considered relay-assisted FEEL network.

Index Terms

Federated learning, edge learning, relay, outage probability, Internet of Things.

L. Chen and L. Fan are both with School of Computer Science, Guangzhou University, Guangzhou 510006, China (e-mail:
2112019037@e.gzhu.edu.cn, lsfan@gzhu.edu.cn).

X. Lei is with the School of Information Science and Technology, Institute of Mobile Communications, Southwest Jiaotong
University, Chengdu 610031, China (e-mail: xflei@home.swjtu.edu.cn).

T. Q. Duong is with the School of Electronics, Electrical Engineering and Computer Science, Queens University, Belfast BT7
1NN, U.K. (e-mail: trung.q.duong@qub.ac.uk).

A. Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London,
London, U.K (e-mail: a.nallanathan@qmul.ac.uk).

G. K. Karagiannidis is with the Aristotle University of Thessaloniki, 54636 Thessaloniki, Greece (e-mail: geokarag@auth.gr).
I. INTRODUCTION

Recently, fast-growing applications of the Internet of Things (IoT) have generated an explosive amount of data to drive artificial intelligence (AI), widely applied in wireless communication, image processing and other fields [1]–[4]. The centralized AI applications need to aggregate distributed data from users into the server for centralized training, which is hard to be achieved due to privacy concerns. To tackle this issue, an intelligent paradigm namely federated learning (FL) was proposed to enable multiple users to train a global model without transmitting the sensitive data [5]–[7]. In this framework, the FL server periodically selects some users as the candidates to join each round’s training. Then, the selected users calculate the training loss, update the weights and transmit the local models to the server. Once the local models are received, the server can aggregate the models and repeat the whole procedure to converge [8], [9].

For the development of wireless networks, mobile edge computing (MEC) has become one of the most advanced technologies for reducing communication latency and energy consumption. For example, MEC could be used for video transmission to suppress jamming [10], where the compression parameter and power control were optimized by a reinforcement learning-based approach. Besides, a deep reinforcement learning was used in making offloading decision against jamming attacks and interference in [11], which could achieve a significant reduction in latency and energy consumption. Therefore, FL is able to apply in the MEC scenarios, where the mobile users perform distributed learning and transmit the trained models to be aggregated at the edge server, called federated edge learning (FEEL) [12]–[14]. The FEEL performance depends on the successfully participated user number in the federated learning, which is however limited by the communication overhead, due to the practical constraints such as latency and wireless bandwidth [15]–[17]. To achieve a reduced communication overhead, a physical-layer quantization scheme was proposed to upload training models, where the compromise between FEEL performance and quantization ratio was revealed [18]. To further cope with the communication overhead, the system resources of FEEL networks can be exploited to support more users to successfully participate into the federated learning [19], [20]. For instance, the trade-off was investigated between the communication overhead and computational capability in [21], by dividing the deep model into several sub-models with different depths, where the authors enabled heterogeneous mobile users to select models of appropriate size to reduce the amount of transmitted data. In
addition, the system resources such as bandwidth can be optimized among users for the FEEL system, in order to meet practical requirements such as latency and energy consumption, by exploiting the channel state information (CSI) [22], [23].

Besides the above techniques, relays can be deployed in FEEL to decrease the communication overhead for enhancing the system communication and learning performance, in which relaying has been shown as an effective technology in wireless communication systems to extend communication coverage and enhance transmission reliability without requiring additional power [24]–[26]. In the relay-assisted FEEL, some intermediate relays can be deployed in the system to help assist the communication between mobile users and the edge server. In this aspect, a FEEL network exploited cooperative relaying with service pricing in [27], where the relays only helped the data communication during the model update between the mobile users and server. In addition, a relay-assisted cooperative FEEL system was investigated in [28], where multiple relays were used to enhance the over-the-air computation performance. Besides assisting the data communication, the relays in the FEEL networks can help perform partial aggregation to reduce the total amount of data required for transmission. In this aspect, a two-tier relay-assisted FL framework was proposed in [29], where the relays assisted the model aggregation for the local gradients to achieve a partially synchronized parallel mechanism. In addition, federated learning aggregation was explored for device-to-device (D2D) communications across the wireless devices, where partial gradient aggregation was performed at the relays to assist the uplink of federated learning [30]. So far, to the best of our knowledge, there has been little work on the relay-assisted FEEL system with limited resources, especially about the framework of performance analysis and system optimization.

In this paper, we study a relay-assisted FEEL system under latency and bandwidth constraints, where $N$ users collaboratively train a global model assisted by one edge server and $M$ intermediate relays. For the relay-assisted FEEL system, we propose a novel framework of performance analysis and system optimization in this paper. Specifically, we begin with the first critical question: “How to design a relay-assisted FEEL system that can make full use of the relays in the edge environment with limited resources?” To answer this question, we propose to use partial aggregation and spectrum resource multiplexing at the relays to enhance the communication of the relay-assisted FEEL system. We then study the second important question: “How to evaluate the system performance of the relay-assisted FEEL?” To answer this question, we provide the analysis of outage probability and perform convergence analysis to reveal the impact of outage
probability on the convergence rate of federated learning. Driven by the system performance analysis of the relay-assisted federated learning, we come to the third important question: “How to optimize the FEEL performance by scheduling the system bandwidth resources?” To answer this question, we provide instantaneous and statistical bandwidth allocation (BA) schemes, which can be applicable depending on specific requirements of communication and computing scenarios. Simulation results are finally provided to illustrate the advantages of the instantaneous and statistical BA schemes.

II. RELAY ASSISTED FEDERATED EDGE LEARNING

In this part, the system model of the relay-assisted federated edge learning (FEEL) network is firstly presented, and then the conventional federated learning is introduced. After that, we present the procedure of the relay-assisted FEEL, assisted by the relay selection and partial aggregation.

A. System Model

A relay-assisted FEEL network is shown in Fig. 1, where $N$ users collaboratively train a global model assisted by one edge server $ES$ and $M$ intermediate decode-and-forward (DF)
relays. Due to severe fading, there is no direct link between the ES and users, i.e., the ES can only communicate with the users via relaying links. Besides assisting the data communication, the relays can also perform the model aggregation, in order to reduce the communication and computing overhead at the server. Let $\mathcal{U} \triangleq \{ U_1, U_2, \ldots, U_N \}$ denote the set of $N$ users, where device $U_k \in \mathcal{U}$ has a local trainable dataset $D_k$ and it can perform local stochastic gradient descent (SGD) on $D_k$. In addition, we use $\mathcal{R} \triangleq \{ R_1, R_2, \ldots, R_M \}$ to denote the relay set, in which each relay can receive the local models from the $N$ IoT users through wireless links. Then, the relays can perform partial aggregation and further transmit the aggregated models to the ES for global aggregation through wireless links. Due to the limitation in the size, each node in the system is assumed to have a single antenna.

B. Conventional Federated Learning

In the conventional FL network, multiple users with distributed data train a global model assisted by one edge server, where the intermediate relays are not involved. For such a system, the FL can be described by

$$\min_w F(w) = \sum_{k=1}^{N} \frac{|D_k|}{|D|} F_k(w),$$

where $w$ represents the global model parameter, $|D_k|$ denotes the training sample amount in user $U_k$, and $|D| = \sum_{k=1}^{N} |D_k|$. Notation $F_k(w)$ is the local loss function of user $U_k$,

$$F_k(w_k) = \sum_{x \in D_k} \frac{1}{|D_k|} \mathcal{L}(w_k, x),$$

where $w_k$ denotes the model parameter of user $U_k$, and $\mathcal{L}(w_k, x)$ is the corresponding loss function. As the data is distributed, it is generally difficult to solve (1) directly. Hence, FL tends to be used by employing an iterative algorithm to train a global model from the users. Specifically, for each round, user $U_k$ calculates the training loss, and then the weights are updated using the gradient descending, given by

$$v_k \leftarrow w_k - \eta \nabla F_k(w_k),$$

where $v_k$ is the updated model parameter of user $U_k$ and $\eta$ denotes the learning rate. After that, the updated local models from multiple users are gathered and aggregated at ES.

\footnote{It is straightforward to adopt the DF relaying protocol to decode and recover the model weights, in order to aggregate models at the relays in this paper. Note that this work can be extended to other relaying protocols, like AF protocol with some minor modification. In particular, we can use the summation property of wireless channels and introduce over-the-air computing technology to aggregate models without decoding and re-modulation, for the AF relaying assisted federated learning systems.}
C. Relay-assisted FEEL

In the considered relay-assisted FEEL, the intermediate relays cooperatively assist the model exchange between users and ES, to extend the coverage and enhance the transmission reliability. Moreover, the relays can perform the operation of aggregation early to cut down the communication cost.

We now detail the procedure of the FEEL assisted by the relays under the paradigm of FedAvg. Specifically, the global model parameter is firstly initialized to \( w_0 \), and then the global model is updated in a number of rounds. At each round, we can divide the model update into the following four steps.

1) **User sampling and model broadcast:** At this step, ES firstly selects a group of users for each round \( t \). It then broadcasts the global model parameters \( w^t \) of the previous round to the selected users with the help of the relays. In particular, ES may uniformly select the user subset \( \mathcal{K} \) out of \( N \) users without replacement, where \( |\mathcal{K}| = K \) is the user number in the user subset \( \mathcal{K} \). Note that the uniform selection can be applied to many scenarios where the importance of users is unknown or identical \([31]–[33]\), and it can guarantee the unbiasedness of the model aggregation with full client participation in each round. For other scenarios where the users have different importance, importance-aware scheduling can be adopted to enhance the federated learning performance.

2) **Local model update:** At this step, user \( U_k \) firstly sets the initial local model parameters as \( w_{k}^{t+1} = w^t \), after receiving the global model parameters \( w^t \) from ES. Then, user \( U_k \) trains its model on its local dataset. Specifically, user \( U_k \) conducts \( E \) epochs of SGD on its local dataset, where there are totally \( e_k = E|D_k|/b \) SGD iterations, and \( b \) is the mini-batch size. Therefore, the local model is updated as

\[
v_{k}^{t+1,e+1} = v_{k}^{t+1,e} - \eta_{t+1} \nabla F_k(w_{k}^{t+1,e}; \xi_k),
\]

where \( \xi_k \) is the data batch uniformly chosen from the local dataset \( D_k \).

3) **Relay selection and partial aggregation:** After finishing the local update, user \( U_k \) needs to transmit its updated weight \( v_{k}^{t+1} \) to a selected intermediate relay \( R_m \). Let \( \mathcal{J}_m \) denote the user subset uploaded to relay \( R_m \), and \( |D_{\mathcal{J}_m}| \) is total the training sample amount in the user subset \( \mathcal{J}_m \). After receiving and decoding the local models, relay \( R_m \) will aggregate the collected models, where some aggregation method can be applied for the considered relay-assisted federated
framework. Without loss of generality, the well-known FedAvg is adopted in this work to aggregate the local trained models, given by

$$w_{t+1}^m = \sum_{U_k \in J_m} \frac{|D_k|}{\sum_{U_k \in J_m} |D_k|} v_{t+1}^k,$$  \hspace{1cm} (5)$$

where the aggregation at the relay $R_m$ is synchronized, which can help reduce the communication overhead and avoid model staleness, by contrast with the asynchronous federated learning. Although the synchronous federated learning has the limitation of waiting for slow learners, i.e., stragglers, such limitation can be alleviated through setting a latency threshold to drop out slow users and using proper resource allocation to avoid a long time waiting.

Note that in the above FedAvg, the problem of “objective inconsistency” may arise due to the heterogeneity in the size of local dataset and local SGD iteration among users [34]. This is because that the aggregated model will be biased towards the users with more SGD iterations, which eventually affects the federated learning performance. To tackle this problem, we can follow the important works in [30], [34] and use the normalized cumulative gradients to replace the FedAvg, given by

$$w_{t+1}^m = w^t + \left( \sum_{U_k \in J_m} \frac{|D_k|}{|D_J|} e_k \right) \sum_{U_k \in J_m} \frac{|D_k|}{|D_J|} \cdot \frac{v_{t+1}^k - w^t}{e_k}.$$  \hspace{1cm} (6)$$

If not specified, FedAvg will be used for aggregating the local trained models in the subsequent sections.

4) Global Aggregation: At this step, each relay needs to send its aggregated model to the edge server via the second-hop relaying link. After gathering all the models from the relays, the ES can perform the aggregation as

$$w_{t+1}^J = \sum_{R_m \in R} \frac{|D_J|}{\sum_{R_m \in R} |D_J|} w_{t+1}^m.$$  \hspace{1cm} (7)$$

D. Problem Formulation

For the considered relay-assisted FEEL system under latency and bandwidth constraints, we can optimize the system performance through minimizing the global loss function, given by

$$\mathbf{P0:} \quad \min \frac{1}{|D|} \sum_{k=1}^{N} \sum_{x \in D_k} L(w_k, x).$$  \hspace{1cm} (8)$$
However, obtaining an exact expression for the global loss function of FEEL is generally hard, which causes much difficulty in solving the optimization in (8). To overcome such difficulty, we turn to perform some analysis on the system performance, as shown in the following section.

III. System Performance Analysis of Relay-assisted FEEL

A. Latency of relay-assisted FEEL

The latency is a critical performance metric in the FEEL, as it determines whether the users can finish the model training and model upload in time or not. When the devices fail to accomplish uploading in time, the effective number of successfully participated users will decrease, causing deterioration in the convergence of federated learning. In the considered relay-assisted FEEL, the latency of each IoT device is related to the computational capability, wireless channel quality, and relay selection. The latency of local training and global aggregation may significantly affect the system training performance. Thus, investigating the latency for the considered relay-assisted FEEL is very important.

The total latency of user $U_k$ is denoted as $T_{\text{total}}^k$, which consists of both the local training latency and the uplink latency. Note that the downlink latency is ignored in this work, as it is generally much smaller than the uplink latency, because the transmit power at the server can be much larger. Specifically, the local training latency $T_{\text{local}}^k$ of user $U_k$ is given by

$$T_{\text{local}}^k = \frac{e_k \cdot b \cdot \rho}{f_k},$$

(9)

where CPU needs $\rho$ cycles to process one sample training, and $f_k$ denotes the computational capability at user $U_k$. Then, the local trained model needs to be uploaded to $ES$ via the uplink relaying links. In this paper, we perform the relay selection based on the instantaneous CSI of the first-hop relaying links

$$m_k^* = \arg\max_{1 \leq m \leq M} |h_{k,m}|^2;$$

(10)

where $h_{k,m}$ is the channel parameter of the link $U_k-R_m$, and it follows Rayleigh fading with $E[|h_{k,m}|^2] = \lambda_{k,m}$. The transmission data rate of the link $U_k-R_{m_k^*}$ is

$$R_{k,m_k^*}^I = B_k^I \log_2 \left(1 + \frac{P_k|h_{k,m_k^*}|^2}{\sigma^2}\right),$$

(11)

\footnote{In order to obtain the instantaneous CSI, each user needs to broadcast the transmission request to all relays, and then the users will send some pilot signals to the relays. After that, the relays can estimate the associated channel parameters and execute the relay selection in (10).}
where $B^I_k$ is the allocated bandwidth of the link $U_k \rightarrow R_{m_k^*}^*$, $P_k$ denotes the transmit power at user $U_k$, and $\sigma^2$ denotes the variance of AWGN.

Note that the transmission in (11) employs orthogonal frequency resources among users. If multiple users employ the same frequency resource to communicate simultaneously, the co-channel interference will arise among the users, and the transmission data rate between user $U_k$ and the selected relay $R_{m_k^*}$ becomes,

$$R^I_{k,m_k^*} = \frac{B^I_k}{m_k^*} \log_2 \left( 1 + \frac{P_k |h_{k,m_k^*}|^2}{\sigma^2 + \sum_{U_i \in J_{m_k^*}, i \neq k} P_i |h_{i,m_k^*}|^2} \right), \quad (12)$$

where $B^I_k$ denotes the bandwidth of the link $R_{m_k^*}^* \rightarrow ES$. From this expression, we can find that the co-channel interference will deteriorate the transmission data rate, and multiple users will have to collaborate or compete in some other domains, such as the power domain in multiuser NOMA systems.

From (11), the transmission latency from user $U_k$ to relay $R_{m_k^*}$ is given by

$$T^I_{k,m_k^*} = \frac{|L|}{R^I_{k,m_k^*}}, \quad (13)$$

where $|L|$ is the size of the uploaded model. After receiving all the model parameters from the user set $J_{m_k^*}$, relay $R_{m_k^*}$ aggregates the local model according to (5). Then, relay $R_{m_k^*}$ needs to transmit the aggregated model to $ES$, where the corresponding transmission data rate from relay $R_{m_k^*}$ to $ES$ is

$$R^II_{m_k^*} = B^II_{m_k^*} \log_2 \left( 1 + \frac{P_{m_k^*} |g_{m_k^*}|^2}{\sigma^2} \right), \quad (14)$$

where $P_{m_k^*}$ denotes the transmit power at relay $R_{m_k^*}$, $g_{m_k^*}$ denotes the instantaneous channel parameter of the link $R_{m_k^*}^* \rightarrow ES$, and it follows Rayleigh fading with $\mathbb{E}[|g_{m_k^*}|^2] = \lambda_{m_k^*}$. In this paper, the relays work in a time-division multiplexing mode, where the dual hops share the same frequency resources, i.e.,

$$B^II_{m_k^*} = \sum_{U_k \in J_{m_k^*}} B^I_k. \quad (15)$$

From $R^II_{m_k^*}$ in (14), the transmission latency from relay $R_{m_k^*}$ and $ES$ is

$$T^II_{m_k^*} = \frac{|L|}{R^II_{m_k^*}}. \quad (16)$$

In summary, the total latency of user $U_k$ is

$$T^\text{total}_k = T^\text{local}_k + T^I_{k,m_k^*} + T^II_{m_k^*}. \quad (17)$$
B. Outage Probability Analysis

From the above $T_k^{\text{total}}$, we begin to analyze the outage probability of user $U_k$. To avoid idle time in the FEEL network, a predetermined latency threshold $\gamma_{th}$ will be set in practice. The user $U_k$ will be dropped from the federated learning, if the associated latency $T_k^{\text{total}}$ is above $\gamma_{th}$. Thus, the effective number of users who can successfully participate in federated learning can be given by

$$K_{\text{eff}} = \sum_{k=1}^{K} \mathbb{I}(T_k^{\text{total}} \leq \gamma_{th}),$$  \hspace{1cm} (18)

where $\mathbb{I}(\cdot)$ denotes the indicator function which returns 1 if the condition is met or 0 otherwise. Accordingly, the expected effective user number is given by

$$\mathbb{E}(K_{\text{eff}}) = \sum_{k=1}^{K} \Pr[T_k^{\text{total}} \leq \gamma_{th}] = K \left( 1 - \frac{1}{K} \sum_{k=1}^{K} \Pr[T_k^{\text{total}} > \gamma_{th}] \right).$$  \hspace{1cm} (19)

From (19), the system outage probability of the FL is given by

$$P_{\text{out}} = \frac{1}{K} \sum_{k=1}^{K} \Pr[T_k^{\text{total}} > \gamma_{th}] = \frac{1}{K} \sum_{k=1}^{K} P_{\text{out},k},$$  \hspace{1cm} (20)

where $P_{\text{out},k}$ is the outage probability of $U_k$ in the process of FL, given by

$$P_{\text{out},k} = \Pr[T_k^{\text{total}} > \gamma_{th}] = \Pr[T_k^{\text{local}} + T_{k,m_k^*}^{I} + T_{m_k^*}^{II} > \gamma_{th}].$$  \hspace{1cm} (21)

For the purpose of analyzing the system outage performance of the relay-assisted FEEL, we need first to derive the outage probability of user $U_k$. In practice, the local training latency of user $U_k$ can be regarded deterministic, as it is not affected by the stochastic channels. Hence, we can re-write $P_{\text{out},k}$ as,

$$P_{\text{out},k} = \Pr \left[ T_{k,m_k^*}^{I} + T_{m_k^*}^{II} > \frac{\gamma_{th} - d_k}{f_k} \right] = \Pr \left[ \frac{|L|}{R_{k,m_k^*}^{I}} + \frac{|L|}{R_{m_k^*}^{II}} > \frac{\gamma_{th} - d_k}{f_k} \right] \hspace{1cm} \text{(22)}$$

where $d_k = e_k b p$ denotes the CPU cycles needed to finish local training for user $U_k$. 
As deriving an exact closed-form solution to \( P_{\text{out},k} \) from (22) is generally hard, we turn to use the inequality of \( \frac{xy}{x + y} < \min(x, y) \) for positive \( x \) and \( y \), and then obtain a tight upper bound for the first form in (22) as,

\[
\frac{R_{k,m_k}^I R_{m_k}^{II}}{|L| (R_{k,m_k}^I + R_{m_k}^{II})} < \frac{1}{|L|} \min(R_{k,m_k}^I, R_{m_k}^{II}).
\]  

(23)

Then, substituting (23) into (22), we can obtain the lower bound on the outage probability of user \( U_k \), which can be analytically solved, as shown in Theorem 1.

**Theorem 1.** A lower bound on the outage probability of user \( U_k \) is

\[
P_{\text{out},k}^{lb} = 1 - \exp \left( \frac{1 - \exp \left( \frac{f_k |L| \ln 2}{\lambda_k |L| (\gamma_k f_k - d_k)} \right)}{\zeta_k m_k^2} \right) \cdot \left( 1 - \prod_{m=1}^M \left( 1 - \exp \left( \frac{1 - \exp \left( \frac{f_k |L| \ln 2}{B_k^* (\gamma_k f_k - d_k)} \right)}{\lambda_k |L| (\gamma_k f_k - d_k)} \right) \right) \right).
\]

(24)

where \( \zeta_k = \frac{P_k}{\sigma^2} \) and \( \zeta_{m_k} = \frac{P_{m_k}^2}{\sigma^2} \) are the transmit SNRs at the user \( U_k \) and relay \( R_{m_k}^* \), respectively, and \( A_k^{II} \) is given by

\[
A_k^{II} = \sum_{i=1}^{K-1} \left( \binom{K-1}{i} B_k^I + \binom{K-2}{i-1} (B_{\text{total}} - B_k^I) \right) \left( \frac{1}{M} \right)^i \left( 1 - \frac{1}{M} \right)^{K-i-1} + B_k^I \left( 1 - \frac{1}{M} \right)^{K-1}.
\]

(25)

*Proof.* See Appendix A.

From (20) and Theorem 1, the lower bound on the system outage probability can be obtained in Theorem 2.

**Theorem 2.** A lower bound on the system outage probability can be given by

\[
P_{\text{out}}^{lb} = \frac{1}{K} \sum_{k=1}^K P_{\text{out},k}^{lb} = \frac{1}{K} \sum_{k=1}^K \left[ 1 - \exp \left( \frac{1 - \exp \left( \frac{f_k |L| \ln 2}{\lambda_k |L| (\gamma_k f_k - d_k)} \right)}{\zeta_k m_k^2} \right) \right] \cdot \left( 1 - \prod_{m=1}^M \left( 1 - \exp \left( \frac{1 - \exp \left( \frac{f_k |L| \ln 2}{B_k^* (\gamma_k f_k - d_k)} \right)}{\lambda_k |L| (\gamma_k f_k - d_k)} \right) \right) \right).
\]

(26)

Note that the above low bound \( P_{\text{out}}^{lb} \) contains elementary functions only, which can be easily computed. Therefore, the system outage probability can be easily evaluated in the whole range of SNR.

\(^3\)Note that in this inequality, the approximation error is large when \( x \) is equal to \( y \), and the approximation accuracy improves when \( x \) differs from \( y \). In general, \( x \) is often different from \( y \) due to random wireless channels, resulting in a fine approximation accuracy on average. Due to these reasons, the inequality of \( \frac{xy}{x + y} < \min(x, y) \) is widely used in the existing works such as [35]–[37].
To obtain more insights on the system design of the relay-assisted FEEL, we use (26) to provide an asymptotic expression for $P_{\text{out}}^{lb}$ for the high SNR region as

$$P_{\text{out}}^{lb} \approx \frac{1}{K} \sum_{k=1}^{K} \left(1 - \left(1 - \prod_{m=1}^{M} \exp \left(\frac{f_k|L| \ln 2}{B_k^m (\gamma_{th} f_k - d_k)} - 1\right)\right) \left(1 - \frac{\exp \left(\frac{f_k|L| \ln 2}{A_k^m (\gamma_{th} f_k - d_k)} - 1\right)}{\lambda_{m_k}^* \zeta_{m_k}^*}\right)\right),$$

(27)

where the Taylor’s series approximation of $\lim_{x \to 0} e^{-x} \approx 1 - x$ is applied [38]. We further use the approximation of $\lim_{x \to 0} (1 - (1 - x)(1 - y)) \approx x + y$ and get the asymptotic expression of $P_{\text{out}}^{lb}$ for high SNR as

$$P_{\text{out}}^{lb} \approx \frac{1}{K} \sum_{k=1}^{K} \left(\prod_{m=1}^{M} \left(\exp \left(\frac{f_k|L| \ln 2}{B_k^m (\gamma_{th} f_k - d_k)} - 1\right)\right) \lambda_{m_k}^* \zeta_{m_k}^* + \left(\frac{\exp \left(\frac{f_k|L| \ln 2}{A_k^m (\gamma_{th} f_k - d_k)} - 1\right)}{\lambda_{m_k}^* \zeta_{m_k}^*}\right)\right) = P_{\text{out}}^{\text{asy}}.$$

(28)

Note that the above asymptotic expression contains two parts, where the first part $O_1$ depends on the transmission between users and relays, while the second part $O_2$ depends on the transmission between relays and edge server. From $P_{\text{out}}^{\text{asy}}$, several insights on the FL system can be obtained,

- The first part $O_1$ decays exponentially with factor $M$, which indicates that the $M$ intermediate relays can be fully exploited.
- When relay number $M$ is large, the first part $O_1$ approaches to 0, and the second part $O_2$ will dominate in the system outage probability, indicating that the transmission between the relays and edge server becomes the system bottleneck.
- The outage performance of the relay-assisted FEEL system improves with a larger $\lambda_{m_k}^*$ and $\lambda_{m_k}^*$, revealing that a better transmission channel can enhance FL transmission.
- Both $O_1$ and $O_2$ are decreasing with respect to $B_k^m$ and $A_k^m$, indicating that a larger bandwidth of user $U_k$ and intermediate relays $m_k^*$ will improve the system outage performance.

C. Convergence Analysis of Relay-assisted FEEL

The convergence of the relay-assisted FEEL is now analyzed, which is of vital importance for the FL training. For this purpose, we first introduce the following assumptions,

Assumption 1: For any user $U_k$, $F_k(\cdot)$ is $\mu$-strongly convex, i.e., for any $w_0$ and $w_1$,

$$F_k(w_1) \geq F_k(w_0) + (w_1 - w_0)^T \nabla F_k(w_0) + \frac{\mu}{2} ||w_1 - w_0||^2.$$

(29)
Assumption 2: For any user $U_k$, $F_k(\cdot)$ is $L$-smooth, i.e., for any $w_0$ and $w_1$,

$$F_k(w_1) \leq F_k(w_0) + (w_1 - w_0)^T \nabla F_k(w_0) + \frac{L}{2} \|w_1 - w_0\|^2.$$  \hspace{1cm} (30)

Assumption 3: For $\xi_k$ uniformly and randomly sampled from the local dataset $D_k$, the variance of user $U_k$ is bounded for all $k$ by

$$\mathbb{E} \left[ \| \nabla F_k(w; \xi_k) - \nabla F_k(w) \|^2 \right] \leq \delta_k^2.$$  \hspace{1cm} (31)

Assumption 4: For all users, the expected second-order moment of the norm of the stochastic gradient is uniformly bounded by $\mathbb{E} [\| \nabla F_k(w; \xi_k) \|^2] \leq G^2$.

In addition to the above assumptions, we use the term $\Gamma = F^* - \sum_{k=1}^{N} p_k F_k^*$ to quantify the degree of non-i.i.d, where $F^*$ and $F_k^*$ are the minimum values of $F$ and $F_k$, respectively. We can find from $\Gamma$’s definition that the data distribution is i.i.d if $\Gamma = 0$, or non-i.i.d otherwise.

From the above assumptions, the convergence performance of the relay-assisted FEEL can be analyzed, which is presented in Theorem 3.

**Theorem 3.** Under Assumption 1-4, with $\psi = \max \left\{ \frac{8 L}{\mu}, e \right\}$, and $\eta_{t} = \frac{2}{\mu(\psi+t)}$, the convergence analysis of the relay-assisted FEEL satisfies

$$\mathbb{E}[F(w^T) - F^*] \leq \frac{L}{\mu(\psi + T)} \left[ \frac{2}{\mu} \left( \sum_{k=1}^{N} p_k \delta_k^2 + 6 L \Gamma + 8(e - 1)^2 G^2 + 4 e^2 G^2 H \right) + \frac{H \psi}{2} \|w^0 - w^*\|^2 \right],$$

in which $H = \sum_{k=1}^{N} p_k \frac{N - K(1 - P_{\text{out}})}{K(1 - P_{\text{out}})}$, and $w^0$ is the initial value of the global model weights.

**Proof.** See Appendix B. \hfill \blacksquare

From Theorem 3, we can find that for the relay-assisted FEEL with partial user participation and user dropout, the terms of $\sum_{k=1}^{N} p_k^2 \delta_k^2$, $6 L \Gamma$, $8(e - 1)^2 G^2$, and $4 e^2 G^2 H$ dominate the convergence performance. Specifically, the term $\sum_{k=1}^{N} p_k^2 \delta_k^2$ is related to the mini-batch SGD used in the local training, and the term $6 L \Gamma$ is related to non-i.i.d data distribution of user data. In particular, the convergence upper bound decreases monotonically with $\Gamma$, and when $\Gamma$ becomes zero, i.e., i.i.d. dataset, the term $6 L \Gamma$ can be removed. Moreover, the terms $8(e - 1)^2 G^2$ and $4 e^2 G^2 H$ are both related to the distributed SGD algorithm and the model aggregation, where the term $4 e^2 G^2 H$ also shows that the effective number of participated users directly affects the convergence upper bound, revealing that a larger outage probability will deteriorate the convergence rate seriously. Thus, it is critical to enhance the convergence performance through reducing the number of users.
dropped from the FEEL training, by designing a bandwidth allocation scheme for the considered system.

IV. BANDWIDTH ALLOCATION FOR RELAY-ASSISTED FEEL

Inspired by the above convergence results that more users successfully participating in each round’s learning process can improve the convergence in Theorem 3, problem $\textbf{P0}$ is reformulated as maximizing the successfully participated user number in each round’s FL by allocating the wireless bandwidth among users and intermediate relays, given by

$$\textbf{P1:} \max_{\{I_k, I_m | U_k \in \mathcal{U}, R_m \in \mathcal{R}\}} K_{eff} = \sum_{k=1}^{K} \mathbb{I}(T_k^{total} \leq \gamma_{th})$$

subject to

$$\sum_{R_m \in \mathcal{R}} I_m \leq B_{total},$$

$$\sum_{U_k \in \mathcal{J}_m} I_k = I_m$$

where (33b) and (33c) are the bandwidth constraints at the relays and users, respectively. These two bandwidth constraints also indicate that multiple users will collaborate or compete with each other in the frequency domain, which can be found in many application scenarios where the users employ some orthogonal frequency resources to communicate, such as OFDMA systems. On the other hand, if the users employ the same frequency resource to communicate simultaneously, co-channel interference will arise, and multiple users have to collaborate or compete in some other domains, such as the power domain in multiuser NOMA systems. In this case, the proposed framework of performance analysis and system optimization in this paper is still applicable, and the results in this work can serve as a useful benchmark for the federated learning with multiuser interference, which can help obtain some insights on the system design.

In the following, the optimization problem is solved by exploiting the instantaneous or statistical CSI, where flexible choices can be provided for the system optimization.

A. Instantaneous Bandwidth Allocation

For the instantaneous bandwidth allocation method, the edge server needs to make bandwidth allocation decision at each time slot, so that the instantaneous bandwidth allocation tends to be used in the system which is sensitive to the performance of communication and training. Due to the indicator function and the coupling of constraints (33b) and (33c), the problem $\textbf{P1}$ is hard to
be directly solved. Thus, we propose to solve this problem by dividing it into two sub-problems: minimizing the total bandwidth required for the selected users and choosing some users to be dropped out from the FEEL process. Specifically, for the first sub-problem, we relax the problem $P1$ by removing the bandwidth constraint (33b), so that all the relays can be allocated by the required bandwidth, in order to support the selected users to successfully participate in FEEL process. The first sub-problem can be given by

$$P2: \min \{B_{I_k}^I, \alpha_{k,m} | U_k \in \mathcal{U}, R_m \in \mathcal{R}\} \sum_{R_m \in \mathcal{R}} B_{I_m}^{II}$$

subject to

$$T_{k}^{loc} + \frac{|L|}{\alpha_{k,m} B_{I_m}^{II} r_{k,m}^{I}} + \frac{|L|}{B_{II}^{II} r_{II}^{II}_m} \leq \gamma_{th}, \forall U_k \in \mathcal{U}, \quad \sum_{U_k \in \mathcal{J}_m} \alpha_{k,m} = 1, \quad 0 \leq \alpha_{k,m} \leq 1,$$

where $r_{k,m}^{I} = \log_2 \left(1 + \frac{P_{k}|h_{k,m}|^2}{\sigma^2}\right)$, $r_{II}^{II} = \log_2 \left(1 + \frac{P_{m}|g_m|^2}{\sigma^2}\right)$, and $\alpha_{k,m}$ is the bandwidth allocation ratio from relay $R_m$ to user $U_k$, which satisfies $0 \leq \alpha_{k,m} \leq 1$ and $\sum_{U_k \in \mathcal{J}_m} \alpha_{k,m} = 1$. Constraint (34b) guarantees that all users can successfully participate in the training process. Constraints (34c) and (34d) are the reformulation of (33c) using $B_{I_k}^I = \alpha_{k,m} B_{II}^{II}$ as the bandwidth allocated to user $U_k$ from relay $R_m$. We can find that the optimal solution of $P2$ should satisfy the conditions given in Theorem 4.

**Theorem 4.** For relay $R_m$, the optimal $B_{II}^{II*}$ and $\alpha_{k,m}^*$ to solve problem $P2$ should satisfy

$$\begin{cases}
T_{k}^{loc} + \frac{|L|}{\alpha_{k,m}^* B_{II}^{II*} r_{k,m}^{I}} + \frac{|L|}{B_{II}^{II*} r_{II}^{II}_m} = \gamma_{th}, \\
\sum_{U_k \in \mathcal{J}_m} \alpha_{k,m}^* = 1, \\
0 \leq \alpha_{k,m}^* \leq 1, \\
B_{II}^{II*} \geq 0.
\end{cases}$$

**Proof.** See Appendix C.

From Theorem 4, we can observe that there is one and only one solution to (35) because of the monotonicity and non-trivial value of $\alpha_{k,m}^*$ and $B_{II}^{II*}$. Moreover, with a given $B_{II}^{II*}$, we can
Algorithm 1: Bisection search of $B_{II}^{*}$ and $\alpha_{k,m}^{*}$

1. **Input** $B_{total}$, $\mathcal{J}_m$
2. $B_{lower} = 0$, $B_{upper} = B_{total}$
3. **while** $B_{lower} < B_{upper}$ **do**
4. \[ B_{mid} = (B_{lower} + B_{upper})/2; \]
5. For $U_k \in \mathcal{J}_m$, calculate the bandwidth ratio $\alpha_{k,m}$ according to (36) with $B_{mid}$;
6. **if** $\sum_{U_k \in \mathcal{J}_m} \alpha_{k,m} < 1$ **then**
7. \[ B_{lower} = B_{mid}; \]
8. **else if** $\sum_{U_k \in \mathcal{J}_m} \alpha_{k,m} > 1$ **then**
9. \[ B_{upper} = B_{mid}; \]
10. **else if** $\sum_{U_k \in \mathcal{J}_m} \alpha_{k,m} = 1$ **then**
11. \[ B_{II}^{*} = B_{mid}, \alpha_{k,m}^{*} = \alpha_{k,m}, U_k \in \mathcal{J}_m; \]
12. break;
13. **end**
14. **end**
15. **Output** $B_{II}^{*}$, \{ $\alpha_{k,m}^{*}$ | $U_k \in \mathcal{J}_m$ \}

get the optimal value of $\alpha_{k,m}$ as

\[
\alpha_{k,m}^{*} = \frac{r_{II}^{*}|L|}{B_{II}^{*} (\gamma_{th} - T_{local}^{L}) r_{k,m}^I r_{II}^{L} - r_{k,m}^I |L|}.
\] (36)

With (36), we can obtain a numerical value of $B_{II}^{*}$ by using an efficient searching algorithm based on the bisection method, as shown in Algorithm 1. In particular, we start the search with the middle point $B_{mid}$ of an initial range $[B_{lower}, B_{upper}]$. With $B_{mid}$ as the bandwidth allocated to relay $R_m$, we then calculate $\alpha_{k,m}$ for each user $U_k \in \mathcal{J}_m$ and sum up all $\alpha_{k,m}$. By comparing $\sum_{U_k \in \mathcal{J}_m} \alpha_{k,m}^{*}$ with 1, we can halve the search region with $B_{upper} = B_{mid}$ if $\sum_{U_k \in \mathcal{J}_m} \alpha_{k,m} > 1$, or halve the search region with $B_{lower} = B_{mid}$ otherwise. The search process will continue until the constraint (35b) is satisfied, which finally outputs the optimal $\alpha_{k,m}$ for $U_k \in \mathcal{J}_m$ and $B_{II}^{*}$.

We proceed to solve the second sub-problem when the total bandwidth needed exceeds the system total bandwidth, i.e., $\sum_{R_m \in \mathcal{R}} B_{II}^{*} > B_{total}$. In this case, the participating users should be adjusted and certain users have to be dropped out to satisfy the bandwidth constraint (33b). Here, a greedy algorithm is utilized to solve the second sub-problem. Specifically, the user with the
Algorithm 2: Greedy based bandwidth allocation algorithm

1 **Input** \( U, R, J_m, B_{\text{total}} \);
2 For \( R_m \in R \), Solve \( B_{II}^*, \alpha_{k,m}^* \) using Algorithm 1;
3 while \( \sum_{R_m \in R} B_{II}^* > B_{\text{total}} \) do
4 \( U_k', R_m' = \arg \max_{U_k \in U, R_m \in R} \alpha_{k,m}^* B_{II}^* \);
5 \( B_k' = 0 \);
6 \( U = U \setminus U_k', J_m \setminus U_k' \);
7 Solve \( B_{II}^*, \alpha_{k,m}^* \) using Algorithm 1 with \( U \) and \( J_m, R_m \in R \);
8 end
9 \( B_{II}^* = B_{II}^*, B_k^i = \alpha_{k,m}^* B_{II}^*, U_k \in J_m, R_m \in R \);
10 **Output** \( \{ B_k^i, B_{II}^* | U_k \in U, R_m \in R \} \)

largest \( \alpha_{k,m}^* B_{II}^* \), i.e., the user occupies the largest bandwidth, will be dropped out from the FEEL process. After removing the firstly dropped out user, we continue to solve problem \( P_2 \) until the constraint \( \sum_{R_m \in R} B_{II}^* \leq B_{\text{total}} \) is satisfied. In this way, we finally solve problem \( P_1 \) with the instantaneous CSI. The greedy based bandwidth allocation algorithm with the instantaneous CSI is summarized in Algorithm 2.

**B. Statistical Bandwidth Allocation**

Besides the above instantaneous BA method, we also provide a statistical bandwidth allocation, which is performed once for many time slots and applicable to the system that is sensitive to the computational complexity of bandwidth allocation at the price of some performance deterioration compared with the instantaneous bandwidth allocation. In this case, we turn problem \( P_1 \) into optimizing the statistical expectation of the successfully participated user number in each round’s FL, given by

\[
P_3: \max_{ \{ B_k^i, B_{II}^* | U_k \in U, R_m \in R \} } \mathbb{E}(K_{\text{eff}}) = K (1 - P_{\text{out}}) \quad (37a)
\]

s.t.
\[
\sum_{R_m \in R} B_{II}^* \leq B_{\text{total}}, \quad (37b)
\]
\[
\sum_{U_k \in J_m} B_k^i = B_{II}^*. \quad (37c)
\]
As obtaining an exact analytical expression for $P_{\text{out}}$ is hard, we turn to employ the derived lower bound $P_{\text{out}}^{lb}$ to help approximate the expectation of the number of users successfully participating in FEEL. Thus, we can reformulate $\textbf{P3}$ into $\textbf{P4}$, given by

$$\textbf{P4}: \max_{\{B_k^I, B_m^I | U_k \in \mathcal{U}, R_m \in \mathcal{R}\}} K(1 - P_{\text{out}}^{\text{lb}})$$

s.t. \hspace{1cm} \sum_{R_m \in \mathcal{R}} B_m^I \leq B_{\text{total}}; \hspace{1cm} (38a)

\sum_{U_k \in \mathcal{J}_m} B_k^I = B_m^I. \hspace{1cm} (38c)

As problem $\textbf{P4}$ is hard to be directly solved, we use the particle swarm optimization (PSO) to solve problem $\textbf{P4}$, which is an intelligent algorithm using a set of “particles” to search for an approximate solution. In PSO, there are $I$ particles, and each particle $i$ has three associated vectors: the velocity $v_i$, the position $p_i$, and the best position $p_{\text{best}i}$. Specifically, $p_i$ is a $K$-dimension vector denoting a feasible solution of bandwidth allocation, where $p_i = \{B_k^I | U_k \in \mathcal{K}\}$, $v_i$ is a $K$-dimension vector of bandwidth variation, where $v_i = \{\Delta B_k^I | U_k \in \mathcal{K}\}$, and $p_{\text{best}i}$ is a $K$-dimension vector of the best solution to the optimization problem for particle $i$. Moreover, there is a global vector $g_{\text{best}}$ used to denote the best solution among all the particles. All the position vectors are potential solutions of the optimization problem evaluated by the fitness function $F_{\text{fitness}}(\cdot)$, measured by $K(1 - P_{\text{out}}^{\text{lb}})$.

For particle $i$ at iteration $t$, its velocity is updated as

$$v_i^t = \omega v_i^{t-1} + \varphi_1 \rho_1 (p_{\text{best}i}^{t-1} - p_i^{t-1}) + \varphi_2 \rho_2 (g_{\text{best}}^{t-1} - p_i^{t-1}), \hspace{1cm} (39)$$

where $\omega$ denotes the inertia weight of the previous velocity, $\varphi_1$ and $\varphi_2$ are two acceleration coefficients, and $\rho_1$ and $\rho_2$ are two random variables uniformly distributed in $[0,1]$. The position of particle $i$ is updated as

$$p_i^t = p_i^{t-1} + v_i^t. \hspace{1cm} (40)$$

After $E$ times of iteration of velocity and position updates, the $g_{\text{best}}$ obtained from $I$ particles can be regarded as a feasible solution to problem $\textbf{P4}$. The PSO based bandwidth allocation algorithm with the statistical CSI is summarized in Algorithm 3.

V. SIMULATION RESULTS

In this part, some analytical and simulation results are presented to validate the proposed studies in this paper. In particular, the basic setting of these simulations is introduced, along
Algorithm 3: PSO based bandwidth allocation algorithm

1. **Input** $U$, $R$, $J_m$, $B_{total}$, $I$, $T$, $\omega$, $\phi_1$, $\phi_2$;

2. **Initialize** Create $I$ particles randomly;

3. **for** $t = 1$ to $T$ **do**

4. **for** $i = 1$ to $I$ **do**

5. Update $v^t_i$ by (19), and update $p^t_i$ by (20);

6. **if** $F_{fitness}(p^t_i) \leq F_{fitness}(p_{best}^t)$ **then**

7. $p_{best}^t = p^t_i$;

8. **end**

9. **if** $F_{fitness}(p^t_i) \leq F_{fitness}(g_{best}^t)$ **then**

10. $g_{best}^t = p^t_i$;

11. **end**

12. **end**

13. **end**

14. $B_{II}^m = \sum_{U_k \in J_m} B_k^I, R_m \in R$;

15. **Output** $\{B_k^I, B_{II}^m | U_k \in U, R_m \in R\}$

with some baselines methods used for comparison. We then present some simulations with the purpose of verifying the derived analysis on the system outage performance. Further, we conduct some more simulations to validate instantaneous and statistical bandwidth allocation schemes.

A. Simulation Settings

The simulations are performed in the considered relay-assisted FEEL system with a total of 200 users. If not specified, for all simulations, there are 500 communication rounds in total, and there are 10 selected users for each communication round. The channels follow Rayleigh flat fading, where the average channel gain of the link $U_k-R_m$ is set to $\lambda_{k,m} = (100 + k)/200$, and the average channel gain between the relays and $ES$ is set to 2. The transmit power at each user and each relay are set to 0.1W and 0.5W, respectively. The computational capability of each user is $1.5 \times 10^7$ cycle/second. In addition, for the PSO based bandwidth allocation algorithm, we use 30 particles and 50 iterations to search for a feasible solution, where the inertia weight of the previous velocity $\omega$ is 0.5 and the two acceleration coefficients $\phi_1$ and $\phi_2$ are both 0.4.
In practice, for the FL task, the Fashion-MNIST dataset is used to perform a classification task, where 60000 training and 10000 test samples are utilized. There are 10 classes of fashion pictures in the training samples, and the number of training samples allocated for each user is uniformly distributed as $|D_k| \in \mathcal{U}(200, 400)$. For the non-i.i.d setting of the Fashion-MNIST dataset, each user is assigned with 2 labels in its local training samples. As to the learning network, we use a CNN composed of two $3 \times 3$ convolution layers, each followed by a batch normalization layer and a $2 \times 2$ max pooling layer, two fully connected layers, a drop out layer between the two fully connected layers, and a soft output layer. For the training of the CNN network, we use the CrossEntropyLoss as the loss function with $\eta = 0.001$, $b = 30$, and $E = 3$.

To verify the effectiveness of the proposed instantaneous and statistical bandwidth allocation schemes, we compare with some baseline methods abbreviated as follows,

- **Ideal FEEL**: There is no bandwidth or latency constraint so that all the selected users can successfully take part in the learning process.
- **Uniform allocation (UA)**: $ES$ performs the uniform bandwidth allocation for all users selected in each communication round.
- **Uniform allocation without partial aggregation (UA-wo-PA)**: $ES$ performs the uniform bandwidth allocation for all users selected in each communication round, and the users upload the model via the selected relay without partial aggregation.
B. Outage Performance Simulation

Fig. 2 depicts the simulated, analytical and asymptotic outage probabilities for the relay-assisted FEEL under UA method versus the transmit SNR, where the transmit SNR of each user ranges from 15dB to 35dB, the transmit SNR of each relay is ten times that of the user, and the total bandwidth of the system is 50MHz. Observing from Fig. 2, we can find that the analytical lower bound fits well with the simulated one, and the asymptotic lower bound converges to the analytical one with high SNR, which shows the correctness of the derived analytical and asymptotic expressions of the system outage probability. Moreover, all the system outage results get improved when SNR becomes larger, as a larger transmit power at users and relays can achieve a reduced latency in the model upload, thus improving the system outage performance. Further to this, it is found that the system outage probability improves with a larger $M$, as more relays can help increase the spatial diversity of the wireless links between users and relays.

C. Federated Learning Performance Simulation

Fig. 3(a) and Fig. 3(b) illustrate the test accuracy and training loss of the aforementioned BA schemes, where $B_{\text{total}} = 60$MHz, and $\gamma_{\text{th}} = 1.2$s. We can observe from Fig. 3(a) and Fig. 3(b) that both the test accuracy and training loss of all BA schemes converge with the increasing communication round. Moreover, the UA-wo-PA performs the worst, because without partial aggregation, more models need to be uploaded through the second hop. Further, the
proposed instantaneous and statistical bandwidth allocation schemes outperform UA, showing the effectiveness of the two bandwidth allocation schemes. Furthermore, the instantaneous bandwidth allocation scheme can achieve a better near-optimal convergence rate and test accuracy than the statistical bandwidth allocation scheme, indicating that the instantaneous CSI can help maximize the number of users who can successfully participate in FL at each round more effectively.

Fig. 4(a) and Fig. 4(b) show the test accuracy and training loss of the aforementioned BA schemes versus the communication round through aggregating the normalized cumulative gradients, where $B_{\text{total}} = 60$MHz, and $\gamma_{\text{th}} = 1.2$s. We can observe that the proposed instantaneous and statistical BA schemes outperform UA, proving the effectiveness of the two bandwidth allocation schemes when aggregating the normalized cumulative gradients. Moreover, aggregating the normalized cumulative gradients can provide a better performance with an improved test accuracy of 1%-1.5% than simply aggregating the trained models in the FedAvg, which demonstrates that the problem of “objective inconsistency” caused by different SGD iterations would deteriorate the federated learning performance, and using normalized cumulative gradients in the aggregation can help solve the inconsistency problem and enhance the system performance.

Fig. 5 is provided to show the test accuracy of the several BA schemes versus $\gamma_{\text{th}}$, where $M \in \{1, 2\}$ and the system latency threshold varies from 0.8s to 1.8s. We can observe that for all the aforementioned schemes except the ideal FEEL one, the test accuracy gets improved with a larger system threshold, as a larger threshold can allow more users successfully to participate
in FEEL. Moreover, for all the aforementioned schemes, the performances with two relays are better than those with only one relay, since more relays can help improve the model transmission rate. In further, the UA and UA-wo-PA schemes have a lower test accuracy than the instantaneous and statistical BA schemes. In particular, when the latency threshold is low, the relay-assisted FEEL system using the UA and UA-wo-PA schemes can not even train an effective model. This is because that only very few users can successfully participate in FEEL under those schemes. However, the proposed instantaneous and statistical BA schemes can achieve sufficiently good performance for various latency thresholds, which proves that instantaneous and statistical BA schemes can provide a feasible bandwidth allocation strategy for the relay-assisted FEEL.

Fig. 6 shows the impact of $B_{\text{total}}$ on the test accuracy of the several bandwidth allocation schemes, where the relay number $M \in \{1, 2\}$, $\gamma_{th} = 1.2s$, and $B_{\text{total}}$ varies from 50MHz to 100MHz. The test accuracy improvements are observed for all the aforementioned schemes except the ideal FEEL one, as $B_{\text{total}}$ increases, indicating that a larger bandwidth can help increase the transmission rate of the models. Moreover, we can see that with the number of relays increasing from 1 to 2, all the bandwidth allocation schemes get improved because more relays can help enhance the outage performance and allow more users successfully participate in FEEL. In further, the proposed instantaneous and statistical BA schemes outperform the other bandwidth allocation schemes for a wide range of $B_{\text{total}}$, and they can achieve almost the same accuracy as the ideal FEEL. These results further verify the proposed bandwidth allocation...
Fig. 6. Test accuracy of the several BA schemes versus $B_{\text{total}}$.

Fig. 7. Test accuracy of the several BA schemes versus $M$.

schemes.

In Fig. 7, the influence of the relay number on the test accuracy of the several BA schemes is studied, where the relay number varies from 1 to 5, $B_{\text{total}} = 60\text{MHz}$, and $\gamma_{th} = 1.2\text{s}$. We can observe from this figure that all the bandwidth allocation schemes are improved with a larger $M$, as spatial diversity and better transmission connections can be provided for model uploading. Moreover, the proposed instantaneous and statistical BA schemes are superior to the other bandwidth allocation schemes, including UA and the UA-wo-PA schemes. In particular,
when there are four relays in the network, the proposed instantaneous and statistical BA schemes can achieve a better test accuracy, at least 5.1% and 3.8% higher than that of the UA and UA-wo-PA schemes. These results indicate that the proposed instantaneous and statistical BA schemes can efficiently exploit multiple relays and improve the performance of the relay-assisted FEEL.

VI. CONCLUSION

In this article, a relay-assisted FEEL system was studied under latency and bandwidth constraints, where we evaluated the system performance by deriving analytical and asymptotic expressions of the system outage probability and the convergence analysis. In order to improve the system performance, we optimized the relay-assisted FEEL network through allocating the wireless bandwidth among users and relays. Specifically, we proposed two bandwidth allocation schemes to maximize the successfully participated user number in each round’s federated learning. Finally, some simulations were demonstrated to verify the instantaneous and statistical bandwidth allocation schemes. The simulation results showed that the proposed instantaneous and statistical BA schemes could outperform the conventional UA and UA-wo-PA schemes, and achieve almost the same performance as the conventional federated learning without latency and bandwidth constraints. In future works, we will study the federated learning with multiuser interference for the considered system, where the proposed framework of performance analysis and system optimization in this paper will be applied.

REFERENCES


PROOF OF THEOREM 1

To prove Theorem 1, we substitute (23) into (22), and then the lower bound of $P_{\text{out},k}$ is,

$$P_{\text{out},k} \geq \Pr \left( \frac{1}{|L|} \min(R^{I}_{k,m_k}, R^{II}_{m_k}) < \frac{1}{\gamma_{th} - T_{k}^\text{local}} \right) = 1 - \Pr \left( \frac{1}{|L|} \min(R^{I}_{k,m_k}, R^{II}_{m_k}) \geq \frac{1}{\gamma_{th} - T_{k}^\text{local}} \right)$$

$$= 1 - \left( 1 - \Pr \left( R^{I}_{k,m_k} < \frac{|L|}{\gamma_{th} - T_{k}^\text{local}} \right) \right) \left( 1 - \Pr \left( R^{II}_{m_k} < \frac{|L|}{\gamma_{th} - T_{k}^\text{local}} \right) \right) . \quad (A.1)$$

After some manipulations, we can further have,

$$P_{\text{out},k} \geq 1 - \left( 1 - \Pr \left( |h_{k,m_k}|^2 < \frac{2 B_k^I (\gamma_{th} - T_{k}^\text{local})}{\zeta_k} - 1 \right) \right) \left( 1 - \Pr \left( |g_{m_k}|^2 < \frac{2 B_k^{II} (\gamma_{th} - T_{k}^\text{local})}{\zeta_{m_k}} - 1 \right) \right)$$

$$= 1 - \left( 1 - \prod_{m=1}^{M} \Pr \left( |h_{k,m}|^2 < \frac{2 B_k^I (\gamma_{th} - T_{k}^\text{local})}{\zeta_k} - 1 \right) \right) \left( 1 - \Pr \left( |g_{m_k}|^2 < \frac{2 B_k^{II} (\gamma_{th} - T_{k}^\text{local})}{\zeta_{m_k}} - 1 \right) \right) . \quad (A.2)$$
As $|h_{k,m}|^2$ and $|g_{m_k}|^2$ are exponentially distributed with $\mathbb{E}[|h_{k,m}|^2] = \lambda_{k,m}$ and $\mathbb{E}[|g_{m_k}|^2] = \lambda_{m_k}$, the analytical lower bound on $P_{out,k}$ is written as,

$$P_{out,k} \geq 1 - \left( 1 - \prod_{m=1}^{M} \left( 1 - \exp \left( \frac{|L| \ln 2}{B_k (\gamma_{th} - T_{local})} \right) \right) \right) \mathbb{E} \left[ \exp \left( \frac{1 - \exp \left( \frac{|L|}{B_{m_k} (\gamma_{th} - T_{local})} \right) \lambda_{k,m} \zeta_k}{\lambda_{m_k}^{*} \zeta_{m_k}^{*}} \right) \right].$$

(A.3)

We can observe from (A.3) that $\exp \left( \frac{1 - \exp \left( \frac{|L|}{B_{m_k} (\gamma_{th} - T_{local})} \right) \lambda_{k,m} \zeta_k}{\lambda_{m_k}^{*} \zeta_{m_k}^{*}} \right)$ is concave for a positive $B_{m_k}^{II}$. By using Jensen’s inequality, we have

$$P_{out,k} \geq 1 - \left( 1 - \prod_{m=1}^{M} \left( 1 - \exp \left( \frac{|L| \ln 2}{B_{m_k}^{II} (\gamma_{th} - T_{local})} \right) \right) \right) \exp \left( \frac{1 - \exp \left( \frac{|L|}{B_{m_k}^{II} (\gamma_{th} - T_{local})} \right) \lambda_{k,m} \zeta_k}{\lambda_{m_k}^{*} \zeta_{m_k}^{*}} \right),$$

(A.4)

where $A_{k}^{II}$ is the expected bandwidth of the relay selected by user $U_k$, given in (25). In this way, we have proven Theorem 1.

**APPENDIX B**

**PROOF OF THEOREM 3**

Before proving Theorem 3, we first present some notations and lemmas to facilitate the proof. Specifically, we define $\bar{w}^t = \sum_{k=1}^{N} p_k w_k^t$ and $\bar{v}^t = \sum_{k=1}^{N} p_k v_k^t$, where $p_k = \frac{|D_k|}{|D|}$. Moreover, in order to simplify the analysis, we change the timeline to SGD iterations and assume that all users have the same $e$ SGD iterations in the convergence analysis. Then, we present two preliminary lemmas used for the proof.

**Lemma 1.** Using (7) and taking the system outage into account, we can write the aggregated global model at time $t$ as

$$w^{t+1} = w^{t+1-e} + \sum_{k=1}^{N} \sum_{k \in K} \mathbb{E}_{T_k^{total}} \left[ \mathbb{I}(T_k^{total} \leq \gamma_{th}) | D_k | \right] (v_k^{t+1} - w^{t+1-e}).$$

(B.1)

**Lemma 2.** The two-stage aggregation in the relay-assisted FEEL is unbiased, i.e., $\mathbb{E}[\bar{w}^t] = \bar{v}^t$. 
\textbf{Proof.} Using (B.1), we have
\[
\mathbb{E}[\bar{w}^t] = \mathbb{E} \left[ w^{t-e} + \sum_{k=1}^{N} \frac{\mathbb{I}(k \in \mathcal{K}, T_k^{\text{total}} \leq \gamma_{th}) |D_k|}{\sum_{k \in \mathcal{K}} \mathbb{I}(T_k^{\text{total}} \leq \gamma_{th}) |D_k|} (v_k^t - w^{t-e}) \right]
\]
\[
= w^{t-e} + \sum_{k=1}^{N} \frac{|D_k|}{N(1 - P_{\text{out}})|D|} \cdot \mathbb{E} \left[ \mathbb{I}(k \in \mathcal{K}, T_k^{\text{total}} \leq \gamma_{th}) \right] (v_k^t - w^{t-e})
\]
\[
= w^{t-e} + \sum_{k=1}^{N} p_k (v_k^t - w^{t-e}) = \sum_{k=1}^{N} p_k v_k^t = \bar{v}^t.
\] (B.2)

Then, we proceed with the proof of Theorem 3 by looking into assumption 2 of Sec.III-C, where we have
\[
\mathbb{E} [F(w^T) - F^*] \leq \frac{L}{2} \mathbb{E} \left[ ||\bar{w}^{t+1} - w^*||^2 \right].
\] (B.3)

Thus, we only need to bound \( \mathbb{E} \left[ ||\bar{w}^{t+1} - w^*||^2 \right] \) for the proof, which can be further written as
\[
\mathbb{E} \left[ ||\bar{w}^{t+1} - w^*||^2 \right] = \mathbb{E} \left[ ||\bar{w}^{t+1} - \bar{v}^{t+1} + \bar{v}^{t+1} - w^*||^2 \right]
\]
\[
= \mathbb{E} \left[ ||\bar{w}^{t+1} - \bar{v}^{t+1}||^2 \right] + \mathbb{E} \left[ ||\bar{v}^{t+1} - w^*||^2 \right] + 2\mathbb{E} \left[ (\bar{w}^{t+1} - \bar{v}^{t+1})^T (\bar{v}^{t+1} - w^*) \right].
\] (B.4)

Next, we bound \( \mathbb{E} \left[ ||\bar{w}^{t+1} - w^*||^2 \right] \) part by part. Specifically, for the first part \( Q_1 \), we have
\[
Q_1 = \mathbb{E} \left[ ||\bar{w}^{t+1} - \bar{v}^{t+1}||^2 \right]
\]
\[
= \mathbb{E} \left[ w^{t-e} + \sum_{k=1}^{N} \frac{\mathbb{I}(k \in \mathcal{K}, T_k^{\text{total}} \leq \gamma_{th}) |D_k|}{\sum_{k \in \mathcal{K}} \mathbb{I}(T_k^{\text{total}} \leq \gamma_{th}) |D_k|} \cdot (v_k^{t+1} - w^{t+1-e}) - \sum_{k=1}^{N} p_k v_k^{t+1} \right]^2
\]
\[
= \mathbb{E} \left[ w^{t-e} + \sum_{k=1}^{N} p_k \frac{\mathbb{I}(k \in \mathcal{K}, T_k^{\text{total}} \leq \gamma_{th})}{K(1 - P_{\text{out}})} (v_k^{t+1} - w^{t+1-e}) - \sum_{k=1}^{N} p_k (v_k^{t+1} - w^{t+1-e}) \right]^2
\]
\[
= \mathbb{E} \left[ \sum_{k=1}^{N} p_k \left( \frac{N \cdot \mathbb{I}(k \in \mathcal{K}, T_k^{\text{total}} \leq \gamma_{th})}{K(1 - P_{\text{out}})} - 1 \right) \cdot (v_k^{t+1} - w^{t+1-e}) \right]^2.
\] (B.5)

Using the convexity of the second-order norm, we further have
\[
Q_1 \leq \mathbb{E} \left[ \sum_{k=1}^{N} p_k \left( \frac{N \cdot \mathbb{I}(k \in \mathcal{K}, T_k^{\text{total}} \leq \gamma_{th})}{K(1 - P_{\text{out}})} - 1 \right) \cdot (v_k^{t+1} - w^{t+1-e}) \right]^2
\]
\[
= \sum_{k=1}^{N} p_k \frac{N - K(1 - P_{\text{out}})}{K(1 - P_{\text{out}})} \cdot \mathbb{E} \left[ (v_k^{t+1} - w^{t+1-e})^2 \right].
\] (B.6)
We can write $\mathbb{E} \left[ \left\| v_k^{t+1} - w^{t+1} - e \right\|^2 \right]$ in (B.6) as

$$
\mathbb{E} \left[ \left\| v_k^{t+1} - w^{t+1} - e \right\|^2 \right] = \mathbb{E} \left[ \left\| \sum_{i=t+1}^{t} \eta_t \nabla F_k(w_k^i; \xi_k^i) \right\|^2 \right] \leq \mathbb{E} \left[ \eta_t \cdot \sum_{i=t+1}^{t} \left\| \nabla F_k(w_k^i; \xi_k^i) \right\|^2 \right]
$$

$$
\leq \mathbb{E} \left[ \eta_t^{2} e \cdot \sum_{t} \left\| \nabla F_k(w_k^i; \xi_k^i) \right\|^2 \right] \leq \eta_t^{2} e^{2} G^{2} \leq 4\eta_t^{2} e^{2} G^{2}, \quad (B.7)
$$

where we use the Cauchy-Schwarz inequality for the first inequality, and we assume that $\eta_t$ is non-increasing with respect to $t$ and $\eta_t \leq 2\eta_t + E$ to derive other inequalities. Then, we can bound the first part $Q_1$ as

$$
Q_1 = \mathbb{E} \left[ \left\| \bar{w}^{t+1} - w^* \right\|^2 \right] \leq \sum_{k=1}^{N} \frac{N - K(1 - P_{out})}{K(1 - P_{out})} \cdot 4\eta_t^{2} e^{2} G^{2}. \quad (B.8)
$$

For the second part $Q_2$, its bound can be found in [39], which can still hold for this paper. Thus, according to [39], we have that for any round $t$, if we choose $\psi = \max \left\{ \frac{8L^2}{\mu}, e \right\}$ and $\eta_t = \frac{2}{\mu(\psi + t)}$, the second part $Q_2$ is bounded as

$$
Q_2 = \mathbb{E} \left[ \left\| \bar{w}^{t+1} - w^* \right\|^2 \right] \leq (1 - \mu \eta_t)\mathbb{E} \left[ \left\| \bar{w}^t - w^* \right\|^2 \right] + \eta_t^2 \left( \sum_{k=1}^{N} p_k^2 \delta_k^2 + 8L \Gamma + 8(e - 1)^2 G^2 \right).
$$

$$
(B.9)
$$

For the third part $Q_3$, we can derive from Lemma 2 that $Q_3$ equals to 0 due to the unbiasedness of $\bar{w}^{t+1}$. By summarizing the above three parts, we have that, for any round $t$, $\mathbb{E} \left[ \left\| \bar{w}^{t+1} - w^* \right\|^2 \right]$ is bounded as

$$
\mathbb{E} \left[ \left\| \bar{w}^{t+1} - w^* \right\|^2 \right] \leq (1 - \mu \eta_t)\mathbb{E} \left[ \left\| \bar{w}^t - w^* \right\|^2 \right] + \eta_t^2 \left( \sum_{k=1}^{N} p_k^2 \delta_k^2 + 6L \Gamma + 8(e - 1)^2 G^2 + 4e^2 G^2 H \right),
$$

$$
(B.10)
$$

in which $H = \sum_{k=1}^{N} p_k \frac{N - K(1 - P_{out})}{K(1 - P_{out})}$. For brevity, we rewrite (B.10) as

$$
\Delta_{t+1} \leq (1 - \mu \eta_t)\Delta_t + C,
$$

$$
(B.11)
$$

where $\Delta_{t+1} = \mathbb{E} \left[ \left\| \bar{w}^{t+1} - w^* \right\|^2 \right]$ and

$$
C = \sum_{k=1}^{N} p_k^2 \delta_k^2 + 6L \Gamma + 8(e - 1)^2 G^2 + 4e^2 G^2 H.
$$

$$
(B.12)
$$
Then, we use the recurrence method to prove that $\Delta_t \leq \frac{v}{\psi + t}$, where $v = \max \left\{ \psi \Delta_0, \frac{\beta^2 C}{\mu \beta - 1} \right\}$.

First, for $t = 0$, we have $\Delta_0 \leq \frac{v}{\psi + 0} \leq \Delta_0$. For $t > 0$, we have

$$\Delta_{t+1} \leq (1 - \mu \eta_t) \Delta_t + \eta_t^2 C = \frac{t + \psi - 1}{(t + \psi)^2} v + \left[ \frac{\beta^2 C}{(t + \psi)^2} - \frac{\mu \beta - 1}{(t + \psi)^2} \right] \leq \frac{1}{t + \psi + 1}. \quad (B.13)$$

Therefore, we have

$$\mathbb{E}[F(w^T) - F^*] \leq \frac{L}{2} \mathbb{E} \left[ \left\| w^{t+1} - w^* \right\|^2 \right] \leq \frac{L}{\mu(\psi + T)} \left[ \frac{2}{\mu} \left( \sum_{k=1}^{N} p_k^2 \delta_k^2 + 6 \Gamma + 8(e - 1)^2 G^2 + 4 e^2 G^2 H \right) + \frac{\mu \psi}{2} \left\| w^0 - w^* \right\|^2 \right]. \quad (B.14)$$

In this way, we have proven Theorem 3.

**APPENDIX C**

**PROOF OF THEOREM 4**

To prove Theorem 4, we start from $\alpha_{k,m} > 0$ and $B_{m}^{II} > 0$ to have

$$\frac{dR_{k,m}^{I}}{d\alpha_{k,m}} = \frac{d}{d\alpha_{k,m}} \left( \alpha_{k,m} B_{m}^{II} \log_2 \left( 1 + \frac{P_k |h_{k,m}|^2}{\sigma^2} \right) \right) = B_{m}^{II} \log_2 \left( 1 + \frac{P_k |h_{k,m}|^2}{\sigma^2} \right) > 0, \quad (C.1)$$

and

$$\frac{dR_{k,m}^{I}}{dB_{m}^{II}} = \frac{d}{dB_{m}^{II}} \left( \alpha_{k,m} B_{m}^{II} \log_2 \left( 1 + \frac{P_k |h_{k,m}|^2}{\sigma^2} \right) \right) = \alpha_{k,m} \log_2 \left( 1 + \frac{P_k |h_{k,m}|^2}{\sigma^2} \right) > 0, \quad (C.2)$$

we can see from (C.1) and (C.2) that $R_{k,m}^{I}$ monotonically increases with $\alpha_{k,m}$ and $B_{m}^{II}$. Therefore, for $\alpha_{k,m} > 0$, and $B_{m}^{II} > 0$, we have that $T_{k,m}^{I}$ and $T_{m}^{II}$ monotonically decrease with $\alpha_{k,m}$ and $B_{m}^{II}$. Moreover, the system latency is determined by the slowest user. Therefore, we can achieve the optimal solution of $P2$ if and only if: 1) all of the bandwidth $B_{m}^{II}$ is allocated (i.e., $\sum_{k \in J_m} \alpha_{k,m} = 1$) and 2) all selected users have the same total latency of $\gamma_{th}$ (i.e., $T_{k}^{total} = T_{k}^{local} + T_{k,m}^{I} + T_{m}^{II} = \gamma_{th}$). So the optimal solution can be given by

$$\begin{aligned}
T_{k}^{total} + \frac{|L|}{\alpha_{k,m}^{*} B_{m}^{II} r_{k,m}^{*}} + \frac{|L|}{B_{m}^{II} r_{m}^{*}} &= \gamma_{th} \\
\sum_{u \in J_m} \alpha_{k,m}^{*} &= 1, \\
0 &\leq \alpha_{k,m}^{*} \leq 1, \\
B_{m}^{II} &\geq 0.
\end{aligned} \quad (C.3)$$

Because of the monotonicity and non-trivial value of $\alpha_{k,m}^{*}$ and $B_{m}^{II*}$, there is one and only one solution to (C.3), which finishes the proof of Theorem 4.