Eficiência Energética no Aprendizado Federado em Redes IoT sem Fio

Renan R. de Oliveira, Luan G. S. Oliveira, Carlos E. da S. Santos, Kleber V. Cardoso e Antonio Oliveira-Jr

***Resumo*— Este artigo propõe um algoritmo de Aprendizado Federado em redes IoT sem fio que seleciona um subconjunto de dispositivos em cada rodada com base na qualidade dos dados e da comunicação. Em seguida, é realizada a alocação dos recursos de comunicação e a definição da potência de transmissão, onde a métrica EMD é utilizada como um fator de ponderação na agregação dos modelos. O algoritmo é implementado por uma abordagem exata baseada em MILP e por uma metaheurística baseada em Algoritmos Genéticos. Em comparação com outras abordagens, as implementações do algoritmo proposto preservam a precisão do modelo global e apresentam maior eficiência energética.**

***Palavras-Chave*— Aprendizado Federado, Redes IoT sem Fio, Eficiência Energética, MILP, Algoritmos Genéticos.**

***Abstract*— This article proposes a Federated Learning algorithm for wireless IoT networks that selects a subset of devices in each round based on data and communication quality. Communication resources and transmission power are then allocated, with the EMD metric employed as a weighting factor during model aggregation. The algorithm is implemented through an exact approach based on MILP and a metaheuristic based on Genetic Algorithms. Compared to other approaches, the implementations of the proposed algorithm preserve the global model's accuracy and achieve greater energy efficiency.**

***Keywords*— Federated Learning, Wireless IoT Networks, Energy Efficiency, MILP, Genetic Algorithm.**

I. Introdução

As projeções para as redes B5G/6G indicam um cenário caracterizado por aplicações inteligentes emergentes, possibilitando que modelos de Aprendizado de Máquina (*Machine Learning* – ML) sejam executados em dispositivos de borda heterogêneos e com recursos limitados. No entanto, esses modelos são geralmente treinados de forma centralizada na nuvem, o que pode comprometer a privacidade, aumentar o tráfego e sobrecarregar as redes de comunicação [9].

O Aprendizado Federado (*Federated Learning* – FL) [4] foi introduzido como uma abordagem de ML descentralizada que permite que dispositivos treinem de forma colaborativa um modelo compartilhado mantendo os dados privados nos dispositivos. Nesta abordagem, somente os parâmetros dos modelos treinados localmente são compartilhados com o servidor agregador. No contexto das redes sem fio, a transmissão dos parâmetros de modelos em vez dos dados de treinamento entre os dispositivos e a Estação Base (*Base Station* – BS) pode economizar energia, recursos de rede e latência da comunicação [16].

Neste contexto, este trabalho propõe o algoritmo FL-w_{OPT}[1], formulado como um problema de otimização para a seleção de dispositivos e o escalonamento de recursos de comunicação. Em cada rodada, FL-w_{OPT} seleciona um subconjunto de dispositivos com base na qualidade da distribuição dos dados, utilizando a métrica EMD (*Earth Mover's Distance*) [15] e a qualidade da conexão com a BS, utilizando o valor médio do SINR (*Signal to Interference plus Noise Ratio*).

A métrica EMD é calculada localmente por cada dispositivo, considerando a distância entre a distribuição de seus rótulos e uma distribuição uniforme. Quando essa diferença é significativa, os parâmetros do modelo do dispositivo pode comprometer a eficiência da agregação e a acurácia do modelo global. Além do mais, FL-w_{OPT} utiliza uma estratégia de escalonamento que otimiza a alocação de Blocos de Recursos (*Resource Blocks* – RBs) de *uplink* e a potência de transmissão dos dispositivos. Por fim, FL-w_{OPT} utiliza a métrica EMD como um fator de ponderação na agregação do modelo global.

O algoritmo FL-w_{OPT} é avaliado por meio de duas estratégias. A primeira, denominada FL-w_{EMD}^{MILP}, é uma abordagem exata baseada na técnica de Programação Linear Inteira Mista (*Mixed-Integer Linear Programming* – MILP), resolvida com a biblioteca PuLP [11]. A segunda, denominada FL-w_{EMD}^{AG}, utiliza uma metaheurística baseada em Algoritmos Genéticos (*Genetic Algorithm* – GA), em busca de soluções viáveis em vez de limitar-se na busca de soluções ótimas.

Além do mais, o desempenho de FL-w_{EMD}^{MILP} e FL-w_{EMD}^{AG} é comparado com dois algoritmos da literatura. O método $FedProx_{SINR}$ é baseado no FedProx [12], que utiliza regularização proximal para limitar o desvio do modelo local em relação ao global. Neste estudo, $FedProx_{SINR}$ incorpora uma heurística que aloca os canais de *uplink* com maior SINR aos dispositivos mais distantes da BS. Além do mais, é atribuída uma potência de transmissão proporcional à distância, com base na potência mínima e máxima que satisfaz as restrições do problema. O algoritmo FLoWN, baseado em [8], otimiza a potência de transmissão e o escalonamento de RBs para reduzir os erros de transmissão. A potência $P^*(r_i)$ é escolhida como o menor valor que satisfaz as restrições operacionais e o escalonamento de RBs é modelado como um problema de emparelhamento bipartido, resolvido pelo algoritmo Húngaro.

Os resultados das simulações demostram que FL-w_{EMD}^{MILP} e FL-w_{EMD}^{AG} preservam a precisão do modelo global e apresentam maior eficiência energética. Para além desta seção introdutória, o restante deste artigo está organizado conforme descrito

Renan R. de Oliveira, Universidade Federal de Goiás, Goiânia-GO e Instituto Federal de Goiás, Goiânia-GO, e-mail: renan.rodrigues@ifg.edu.br. Luan G. S. Oliveira, Universidade Federal de Goiás, Goiânia-GO, e-mail: luangabriel@egresso.ufg.br. Carlos E. da S. Santos, Instituto Federal do Tocantis, Palmas-GO, e-mail: carlosedu@ifto.edu.br. Kleber V. Cardoso, Universidade Federal de Goiás, Goiânia-GO, e-mail: kleber@ufg.br. Antonio Oliveira-Jr, Universidade Federal de Goiás, Goiânia-GO e Fraunhofer Portugal AICOS, Porto-Portugal, e-mail: antoniojr@ufg.br.

[1]Disponível em https://github.com/LABORA-INF-UFG/FL-wOpt

a seguir. A Seção II discute os trabalhos relacionados. A Seção III apresenta os modelos matemáticos utilizados na simulação. A Seção IV descreve a formulação do problema de otimização e a proposta do algoritmo FL-w$_{\text{OPT}}$. A Seção V discute os resultados e a análise da simulação. Por fim, a Seção VI apresenta as considerações finais e indica as orientações para os trabalhos futuros.

II. Trabalhos Relacionados

Na formulação do FL proposta por [4], os autores introduziram o conceito de FL e apresentaram o desempenho do algoritmo *Federated Averaging* (FedAvg). Entretanto, os autores negligenciam os desafios do FL em ambientes com restrições energéticas e enlaces de comunicação não confiáveis.

No trabalho de [13], os autores apontam os desafios do FL devido a natureza dinâmica das redes sem fio. O estudo em [8] propõe um algoritmo de FL que otimiza a alocação de recursos de comunicação visando minimizar a perda do modelo global. De forma semelhante, o artigo de [5] investiga o escalonamento de recursos no FL, considerando a reutilização de parâmetros para reduzir os custos de transmissão. No entanto, as estratégias de [8] e [5] não abordam de forma abrangente a escalabilidade para redes IoT sem fio, utilizando poucos dispositivos e o conjunto MNIST padrão.

Além disso, os estudos de [6] e [7] abordam a eficiência energética no FL por meio do agendamento de recursos de comunicação. Em [10], os autores investigaram o problema da alocação de recursos no FL utilizando MILP e *Deep Q-Network* (DQN) como técnica de otimização. No entanto, os estudos de [6], [7] e [10] não incorporam estratégias de seleção de dispositivos.

III. Modelo do Sistema

A. Modelo de Rede para FL

Considere uma rede IoT sem fio com uma BS diretamente conectada a um servidor agregador de FL, conforme proposto na abordagem MEC (*Multi-Access Edge Computing*) [2], com um conjunto $S = \{i_1, i_2, ..., i_N\}$ de N dispositivos IoT. Cada dispositivo possui um conjunto de dados $\mathcal{P}_i$ com $n_i = |\mathcal{P}_i|$ amostras armazenadas localmente. Esses dispositivos estão conectados à BS por meio de uma conexão sem fio e possuem a capacidade de coletar dados e treinar um modelo local para uma determinada tarefa de FL.

B. Modelo de Comunicação

Considere a técnica de acesso múltiplo por divisão de frequência ortogonal (OFDMA) para o *uplink*, onde cada dispositivo ocupa um RB. De acordo com [8], a taxa de *uplink* do dispositivo *i* transmitindo os parâmetros do modelo w_i para a BS pode ser formulada como $c_i^U(r_i, p_i) = \sum_{n=1}^{R} r_{i,n} B_n^U \mathbb{E}\left(log_2\left(1 + \frac{P_i h_i}{I_n + B_n^U N_0}\right)\right)$, onde $r_{i,n} = [r_{i,1}, ..., r_{i,R}]$ é o vetor de alocação de RBs, $r_{i,n} \in \{0, 1\}$ e $\sum_{n=1}^{R} r_{i,n} = 1$, com $r_{i,n} = 1$ indicando que o RB n está alocado para o dispositivo i, e $r_{i,n} = 0$ indicando o contrário. A potência de transmissão do dispositivo é definida por P_i e o ganho do canal é dado por $h_i = o_i d_i^{-\alpha}$, onde d_i é a distância entre o dispositivo i e a BS, o_i é o parâmetro de desvanecimento de Rayleigh e α é um expoente que afeta como o ganho do canal varia com a distância. Além do mais, $\mathbb{E}(.)$ é a expectativa da taxa de dados em relação a h_i, N_0 é a densidade espectral da potência do ruído e I_n é a interferência em r_n causada por outros dispositivos.

A potência de transmissão da BS geralmente é muito superior à dos dispositivos. Dessa forma, toda a largura de banda do *downlink* pode ser utilizada para a transmissão do modelo global. Assim, a taxa de dados no *downlink* pode ser definida como $c_i^D = B^D \mathbb{E}\left(log_2\left(1 + \frac{P_B h_i}{I^D + B^D N_0}\right)\right)$, onde B^D é a largura de banda e P_B é a potência de transmissão da BS. Além do mais, I^D é a interferência causada por outras BSs que não participam da tarefa de FL.

Assume-se que os modelos de FL são transmitidos por meio de um único pacote. Portanto, o atraso de transmissão entre um dispositivo i e a BS no *uplink* e *downlink* podem ser, respectivamente, formulados como $l_i^U(r_i, P_i) = \frac{S_{pkt}^U}{c_i^U(r_i, P_i)}$ e $l_i^D = \frac{S_{pkt}^D}{c_i^D}$, onde S_{pkt}^U é o tamanho do pacote de *uplink* e S_{pkt}^D é o tamanho do pacote de *downlink*.

Considera-se que a BS não solicitará aos dispositivos o reenvio de modelos quando estes forem recebidos com erros. Neste caso, conforme apresentado por [8], a taxa de erro de transmissão de pacote do *uplink* é dado por $q_i^U(r_i, p_i) = \sum_{n=1}^{R} r_{i,n} \mathbb{E}\left(1 - \exp\left(-\frac{m(I_n + B^U N_0)}{P_i h_i}\right)\right)$, onde $\mathbb{E}(.)$ é a expectativa da taxa de erro de pacote considerando h_i em r_n, com m sendo um limiar (*waterfall threshold*) que define a qualidade da transmissão.

Dessa forma, pode-se formular a probabilidade de sucesso da transmissão do modelo local para a BS como $p_i^U(r_i, p_i) = (1 - q_i^U)$ `if` $(l_i^U \leq \gamma_T$ & $e_i^U \leq \gamma_E$ & $q_i^U \leq \gamma_Q)$ `else` 0, onde p_i define a potência do dispositivo, e γ_T, γ_E e γ_Q definem, respectivamente, o requisito de latência, consumo energético e da taxa de erros de pacotes.

C. Modelo de Consumo Energético

Baseado em [8], o modelo de consumo energético dos dispositivos para o treinamento e transmissão do modelo w_i podem ser formulados como $e_i^T = \zeta \omega_i \vartheta^2 Z(w_i)$ e $e_i^U = P_i l_i^U$, onde ϑ, ω_i e ζ referem-se, respectivamente, a frequência do *clock*, o número de ciclos da unidade central de processamento e o coeficiente de consumo energético de cada dispositivo. Dessa forma, a computação da energia consumida pelo dispositivo i em uma rodada é dada por $e_i^R = e_i^T + e_i^U$. A demanda energética da BS não é considerada, uma vez que esta normalmente possui um fornecimento contínuo de energia.

IV. Formulação do Algoritmo

A. Seleção de Dispositivos

Considere S_t como uma fração f_t de N dispositivos e b como um vetor binário indicando a seleção de n_p dispositivos. Dessa forma, a etapa de seleção dos dispositivos participantes em cada rodada de comunicação pode ser formulada como

$$\text{minimizar} \sum_{i \in S_t} \left(\text{EMD}_i + \frac{1}{\text{SINR}_i^{avg}}\right) b_i, \tag{1}$$

onde EMD_i é o valor da métrica EMD do dispositivo, calculada com base na distância da distribuição de seus rótulos

com uma distribuição uniforme. Além do mais, $\text{SINR}_i^{avg} = \frac{1}{R}\sum_{j=1}^{R} \text{SINR}_i$ é a média dos valores de SINR para o dispositivo i ao longo de todos os RBs e n_p é o número parcial de dispositivos selecionados para a etapa de escalonamento dos recursos de comunicação. A normalização Min-Max foi utilizada para equalizar as distribuições dos termos da Equação (1), garantindo que estejam na mesma escala e contribuam de forma equilibrada. A restrição $\sum_{i=1}^{|S_t|} b_i = n_p$ garante que o número de dispositivos selecionados seja igual a n_p e $b_i \in \{0,1\}$ indica que b é um vetor binário, onde $b_i = 1$ indica que o dispositivo i foi selecionado do subconjunto S_t.

B. Escalonamento dos Recursos de Comunicação

Considere uma matriz binária c como um indicador de atribuição de um dispositivo i no RB j. Dessa forma, o problema do escalonamento dos RBs de *uplink* para maximizar a probabilidade de sucesso da transmissão de modelos pode ser formulado como

$$\text{maximizar} \sum_{i=1}^{n_p}\sum_{j=1}^{R} p_i^U(r_j, p_i^{\max}) \times c_{ij}, \tag{2}$$

onde $p_i^{\max}$ é a potência máxima do dispositivo. A restrição $\sum_{i=1}^{n_p} c_{ij} = 1 \ \forall j \in R$ garante que cada RB é alocado para um único dispositivo e a restrição $\sum_{j=1}^{R} c_{ij} = 1 \ \forall i \in n_p$ garante que cada dispositivo é atribuído a exatamente um RB. A restrição $l_i^U \leq \gamma_T \ \& \ e_i^U \leq \gamma_E \ \& \ q_i^U \leq \gamma_Q$ estabelece os requisitos de latência, consumo energético e da taxa de erro de pacote. A restrição $\sum c_{ij} \leq n_f$ garante que o número final de dispositivos é no máximo n_f e $c_{ij} \in \{0,1\}$ define que c é uma matriz binária, onde $c_{ij} = 1$ indica que o RB j foi alocado para o dispositivo i.

C. Ajuste da Potência de Transmissão

Após definir a alocação dos RBs de *uplink* com $p_i^{\max}$, a potência de transmissão de cada dispositivo pode ser reduzida visando a minimização do consumo energético, atendendo as restrições γ_T, γ_E e γ_Q. Dessa forma, a potência de transmissão ajustada, denotada por p_i^+, pode ser definida da seguinte forma

$$p_i^+ = \lambda \times [p_i^\gamma, \ldots, p_i^{\max}], \tag{3}$$

onde p_i^γ é o valor da menor potência que atende as restrições operacionais e λ é um fator que escalona o valor da potência entre p_i^γ e $p_i^{\max}$, à partir de um vetor de potências discretizado. Neste caso, quando $\lambda = 0$ é selecionado o menor valor de potência que atende as restrições operacionais, e quando $\lambda = 1$ é selecionada a maior potência do vetor discretizado.

D. Proposta do Algoritmo FL-w$_{opt}$

O algoritmo 1 apresenta um visão geral de FL-w$_{\text{OPT}}$ em redes IoT sem fio. Inicialmente, o algoritmo inicializa um modelo w_0. Em cada rodada de comunicação, o algoritmo seleciona n_p dispositivos utilizando a Equação (1). Em seguida, é definido o vetor c com a atribuição de RBs de *uplink* para os dispositivos com $p_i^{\max}$, resolvendo a Equação (2). A potência de transmissão p_i^+ de cada dispositivo é ajustada, resolvendo a Equação (3). Em seguida, o servidor transmite o modelo global para os dispositivos pelos canais de *downlink*.

Algoritmo 1: FL-w$_{\text{OPT}}$ em Redes IoT sem Fio

Servidor:
  Inicialize w_0
  para cada rodada $t = 1, 2, \ldots$ **faça**
    $\mathcal{S}_t \leftarrow$ (selecione m dispositivos)
    Defina b com $n_p \leq |S_t|$ resolvendo a Eq. (1)
    Defina c com $n_f \leq n_p$ com $p_i^{\max}$ com a Eq. (2)
    Defina p_i^+ resolvendo a Eq. (3)
    para cada $c_{ij} = 1$ **em paralelo faça**
      $w_{t+1}^i \leftarrow$ ClienteUpdate(i, w_t)
      $n_i \leftarrow n_i \times \frac{1}{EMD_i}$
    $m_t \leftarrow \sum_{i \in \mathcal{S}_t : c_{ij}=1} n_i$
    $w_{t+1} \leftarrow \sum_{i \in \mathcal{S}_t : c_{ij}=1} \frac{n_i}{m_t} w_{t+1}^i$
ClienteUpdate(i, w): ▷ *Para cada dispositivo i*
  $\mathcal{B} \leftarrow$ (divisão de $\mathcal{P}_i$ em lotes de tamanho B)
  para cada época local j de 1 até E **faça**
    para cada $b \in \mathcal{B}$ **faça**
      $w \leftarrow w - \eta \nabla \ell(w; b)$
  retorne w para o servidor

Cada participante treina um modelo local durante E épocas locais. Após o treinamento, cada dispositivo utiliza o canal de *uplink* definido em c para enviar o modelo atualizado para o servidor. Em seguida, os modelos locais são recebidos e agregados pelo servidor, utilizado a métrica EMD de cada dispositivo com um fator de ponderação de n_i, gerando o estado atual w_{t+1} do modelo w_{global}.

E. Implementação de FL-w$_{OPT}$ baseada em MILP e GA

O Algoritmo 1 é implementado por meio das estratégias FL-w$_{\text{EMD}}^{\text{MILP}}$ e FL-w$_{\text{EMD}}^{\text{AG}}$. A primeira utiliza uma abordagem exata baseada na técnica MILP resolvido pela biblioteca PuLP e a segunda utiliza GA como técnica de otimização.

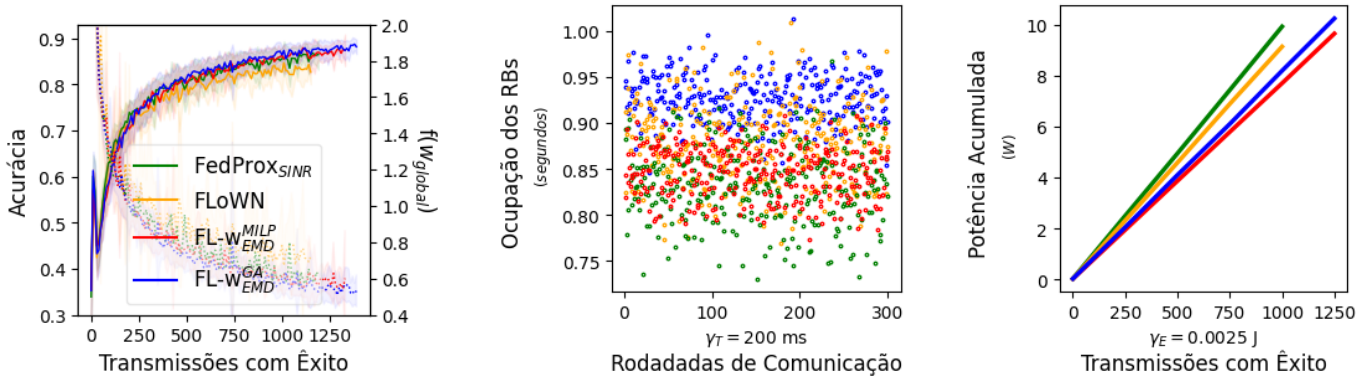
O algoritmo FL-w$_{\text{EMD}}^{\text{AG}}$ codifica cada indivíduo do GA como uma permutação que representa uma possível alocação dos RBs dos dispositivos. A função de avaliação estima a qualidade da solução com base em $p_i^U(r_j, p_i^{\max})$. O operador de *crossover* combina os genes dos pais como uma permutação e o operador de mutação troca genes no cromossomo para manter a diversidade das soluções.

O algoritmo baseia-se em uma população inicial gerada aleatoriamente e evolui por várias gerações. Após o término das gerações, FL-w$_{\text{EMD}}^{\text{AG}}$ retorna o melhor indivíduo, que indica uma solução viável para a alocação dos RBs de *uplink*. Em vez de buscar uma solução estritamente ótima, FL-w$_{\text{EMD}}^{\text{AG}}$ privilegia a obtenção de soluções viáveis que satisfazem todas as restrições operacionais.

V. Configurações e Análise de Desempenho

A. Configuração da Simulação

Considere uma rede IoT sem fio em uma área circular de raio $r = 500$ metros, com uma BS central com um servidor agregador e $N = 100$ dispositivos distribuídos uniforme e aleatoriamente entre 100 e 500 metros. Cada RB de *uplink* possui largura de banda de *uplink* de 1 MHz, com interferência



(a) Evolução da acurácia e $f(w_{global})$. (b) Ocupação dos RBs de *uplink*. (c) Potência acumulada.

Fig. 1: Acurácia, $f(w_{global})$, tempo de ocupação dos RBs de *uplink* e potência acumulada.

distinta e incremental. A potência dos dispositivos é limitada a valores discretos definidos por uma progressão aritmética no intervalo $[0.005, 0.01]$ W, com incremento de 2.5×10^{-4} W e mediana de 0.007 W. A largura de banda de *downlink* é de 20 MHz, sendo a potência de transmissão da BS de 1 W para transmissão dos modelos globais. O ganho do canal h_k considera desvanecimento, refletindo a degradação do sinal com o aumento da distância em relação à BS. Conforme [8], os demais parâmetros são: $\alpha = 2$, $N_0 = -174$ dBm/Hz, $m = 0.023$ dB, $\vartheta = 10^9$, $\zeta = 10^{-27}$ e $\omega_i = 40$.

As tarefas de FL envolvem problemas de classificação de imagens utilizando o conjunto de dados MNIST, amplamente adotado em pesquisas de FL, conforme [1], [3], [4], [5], [8], [14] e [15]. Diferentemente dessas abordagens, este trabalho utiliza uma variação do MNIST com heterogeneidade intrínseca, refletindo características típicas de redes IoT sem fio. Cada dispositivo recebeu uma partição com 90% das amostras de um rótulo dominante e 10% distribuídas uniformemente entre os demais, com imagens rotacionadas em até $45°$ (sentido horário ou anti-horário). O total de dados por dispositivo foi reduzido por um fator entre $[0.25, 1]$ da partição original.

Neste estudo, cada algoritmo foi executado por 300 rodadas de comunicação. As partições do MNIST foram nomeadas como NIID R-MNIST. Para a tarefa de classificação do conjunto de dados, utilizou-se uma arquitetura de rede neural do tipo *Multi-Layer Perceptron* (MLP) com 101.770 parâmetros treináveis. Cada modelo foi treinado por uma época de treinamento local, definindo $\gamma_T = 200$ ms como requisito de atraso, $\gamma_E = 0.0025$ J como requisito de consumo energético e $\gamma_Q = 0.3$ como taxa de erros de pacotes. Os algoritmos FL-w$^{\text{MILP}}_{\text{EMD}}$ e FL-w$^{\text{AG}}_{\text{EMD}}$ utilizaram o valor de $\lambda = 0.2$ como fator de escalonamento da potência dos dispositivos. Esses parâmetros foram definidos empiricamente de modo a garantir que qualquer dispositivo fosse capaz de transmitir seu modelo em pelo menos um RB de *uplink*.

B. Resultados da Simulação

A Figura 1a apresenta a acurácia e $f(w_{global})$ dos algoritmos de FL considerando o número de transmissões com êxito, com $n_p = 2 \times n_f$ onde no máximo $n_f = 6$ dispositivos são selecionados em cada rodada de comunicação. Esse limite considera tanto as restrições de largura de banda quanto a complexidade computacional da solução, que limitam a quantidade de dispositivos que podem ser atendidos simultaneamente. As curvas contínuas mostram o desempenho médio, e as regiões sombreadas representam o desvio padrão ao longo de 15 execuções.

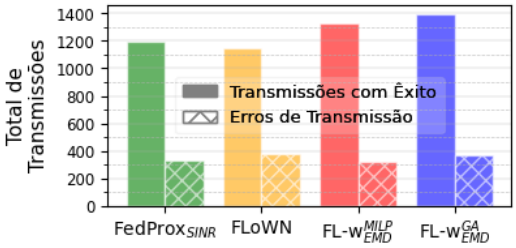
As acurácias dos algoritmos melhoram conforme aumenta o número de transmissões com êxito, pois o modelo global passa a incorporar mais informações dos modelos locais dos dispositivos. Consequentemente, o valor da função de perda $f(w_{global})$ diminui, conforme aumenta o número de transmissões. Os resultados mostram que FL-w$^{\text{MILP}}_{\text{EMD}}$ e FL-w$^{\text{AG}}_{\text{EMD}}$ alcançam uma melhor acurácia e maior número de transmissões com êxito. Este desempenho é devido à estratégia de seleção de dispositivos, que considera tanto a qualidade da comunicação, representada pelo SINR médio, quanto a qualidade dos dados, avaliada pela métrica EMD.

Em contraste, FedProx$_{\text{SINR}}$ e FLoWN adotam a seleção aleatória de dispositivos, o que pode resultar na seleção de participantes com dados mais heterogêneos ou com condições de enlace desfavoráveis. Adicionalmente, FedProx$_{\text{SINR}}$ utiliza uma heurística que aloca canais de *uplink* com maior SINR a dispositivos mais distantes, o que pode priorizar dispositivos mais propensos a erros de transmissão. Por outro lado, FLoWN seleciona a menor potência viável segundo as restrições operacionais, operando frequentemente no limite mínimo, o que também pode aumentar a taxa de erros de transmissão.

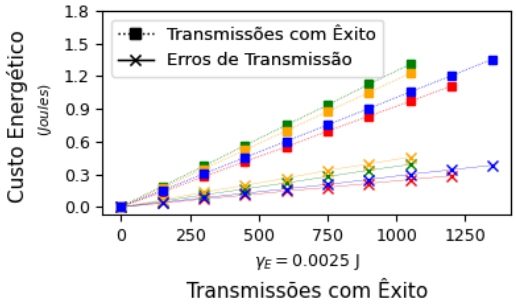
A Figura 1b apresenta a soma dos tempos de ocupação dos RBs de *uplink* em cada rodada. FL-w$^{\text{MILP}}_{\text{EMD}}$ apresenta um tempo de ocupação intermediário, refletindo sua alocação ótima via MILP, que otimiza a taxa de dados e reduz o tempo de transmissão. Em contraste, o FL-w$^{\text{AG}}_{\text{EMD}}$ exibe maior tempo de ocupação, pois prioriza soluções viáveis por meio do GA, permitindo que mais dispositivos transmitam com êxito, ainda que com maiores tempos de ocupação dos RBs. FLoWN apresenta variabilidade, com picos do tempo de ocupação em várias rodadas. Essa característica decorre da estratégia de operar com potência mínima, que reduz a taxa de dados e aumenta o tempo necessário para concluir as transmissões. Além disso, FedProx$_{\text{SINR}}$ apresenta o menor tempo de ocupação ao priorizar canais com maior SINR para dispositivos distantes

e ajustar a potência conforme a distância, reduzindo o tempo das transmissões com êxito.

A Figura 1c apresenta a potência acumulada, evidenciando o impacto das estratégias de seleção de dispositivos, escalonamento de recursos e ajuste de potência de cada algoritmo. FedProx$_{\text{SINR}}$ apresenta a maior potência acumulada, pois utiliza uma compensação de potência proporcional à distância dos dispositivos. Por outro lado, FLoWN opera com a menor potência, ainda que suas falhas de transmissão também contribuam para o valor da potência acumulada. Além disso, FL-w$^{\text{MILP}}_{\text{EMD}}$ e FL-w$^{\text{AG}}_{\text{EMD}}$ adotam estratégias de seleção que priorizam dispositivos com melhores condições de comunicação, o que reduz a necessidade de atribuição de potências elevadas. Dessa forma, mesmo operando com potências de transmissões reduzidas, a Figura 2a indica que FL-w$^{\text{MILP}}_{\text{EMD}}$ e FL-w$^{\text{AG}}_{\text{EMD}}$ aumentam o número de transmissões com êxito.



(a) Total de Transmissões.



(b) Consumo Energético.

Fig. 2: Total de transmissões e consumo energético.

A Figura 2b apresenta o custo energético das transmissões com êxito e dos erros de transmissão. FedProx$_{\text{SINR}}$ apresenta o maior consumo energético, devido sua estratégia com maior potência acumulada. Além disso, embora o FLoWN opere com a menor potência possível, essa estratégia eleva os custos energéticos acumulados devido ao aumento do tempo de transmissão e da taxa de erros. Em contraste, FL-w$^{\text{MILP}}_{\text{EMD}}$ e FL-w$^{\text{AG}}_{\text{EMD}}$ apresentam o menores custos energéticos, pois priorizam dispositivos com melhores condições de comunicação e otimizam a alocação de potência ao considerar o equilíbrio entre o consumo energético e o número de transmissões com êxito. Ao evitar o uso estrito da menor potência de transmissão, FL-w$^{\text{MILP}}_{\text{EMD}}$ e FL-w$^{\text{AG}}_{\text{EMD}}$ minimizam tanto o desperdício energético causado por erros de transmissão quanto o consumo excessivo devido ao uso de potências elevadas. Como resultado, o custo energético acumulado por transmissões com êxito é menor em comparação às demais estratégias.

VI. Conclusões

Este trabalho apresentou um algoritmo para aprimorar a seleção de dispositivos e o escalonamento dos recursos de comunicação no FL em redes IoT sem fio. As versões baseadas em MILP e GA demonstraram melhor eficiência energética e mantiveram a acurácia do modelo global em relação a outras estratégias. A implementação com GA mostrou-se viável, pois obteve um desempenho comparável à abordagem baseada em MILP. Como trabalhos futuros, pretende-se investigar outras soluções baseadas em metaheurísticas e aprendizado por reforço, com o objetivo de otimizar a alocação de recursos em cenários de FL escaláveis e com diferentes topologias de rede.

VII. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

[1] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, T. Parcollet, and N. D. Lane, "Flower: A Friendly Federated Learning Research Framework", *CoRR*, arXiv:2007.14390, 2020.

[2] ETSI, "Multi-access Edge Computing (MEC). Framework and Reference Architecture, ETSI GS MEC 003 V3.1.1," Mar. 2022.

[3] G. Zhu, Y. Wang, and K. Huang, "Broadband Analog Aggregation for Low-Latency Federated Edge Learning", *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 491-506, 2020.

[4] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data", *arXiv*, arXiv:1602.05629, 2016.

[5] H. Chen, S. Huang, D. Zhang, M. Xiao, M. Skoglund, and H. V. Poor, "Federated Learning Over Wireless IoT Networks With Optimized Communication and Resources", *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 16592-16605, 2022.

[6] J. Perazzone, S. Wang, M. Ji, and K. S. Chan, "Communication-Efficient Device Scheduling for Federated Learning Using Stochastic Optimization," in *Proceedings of the IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 2022, pp. 1449–1458.

[7] M. Beitollahi and N. Lu, "Multi-frame Scheduling for Federated Learning over Energy-Efficient 6G Wireless Networks," in *Proceedings of the IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2022, pp. 1–6.

[8] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A Joint Learning and Communications Framework for Federated Learning Over Wireless Networks", *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269-283, 2021.

[9] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated Learning over Wireless Networks: Optimization Model Design and Analysis", in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 1387-1395, 2019.

[10] Oliveira, R. R., Silva, R. S., Freitas, L. A. e Oliveira-Jr, A. "Deep Q-Network para a Alocação dos Recursos de Comunicação do Aprendizado Federado em Redes IoT sem Fio", XLII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT), Brasil, 2024.

[11] Stuart Mitchell. *PuLP: A Linear Programming Toolkit for Python*. Disponível em: `https://coin-or.github.io/pulp/`, 2009.

[12] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar e Virginia Smith. *Federated Optimization in Heterogeneous Networks*. In: *Conference on ML and Systems (MLSys)*, 2020.

[13] X. Cao, T. Başar, S. Diggavi, Y. C. Eldar, K. B. Letaief, H. V. Poor, and J. Zhang, "Communication-Efficient Distributed Learning: An Overview", *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 4, pp. 851-873, 2023.

[14] Y. Amannejad, "Building and Evaluating Federated Models for Edge Computing", in *2020 16th International Conference on Network and Service Management (CNSM)*, pp. 1-5, 2020.

[15] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data", *arXiv*, arXiv:1806.00582, 2022.

[16] Z. Yang, M. Chen, K. Wong, H. V. Poor, and S. Cui, "Federated Learning for 6G: Applications, Challenges, and Opportunities", *Engineering*, vol. 8, pp. 33-41, 2022.