Aprendizado Federado em Redes IoT sem Fio: Novo Algoritmo para a Seleção de Dispositivos e Alocação dos Recursos de Comunicação

Renan R. de Oliveira[1,2], **Rogério S. e Silva**[1,2], **Leandro A. Freitas**[2] **e Antonio Oliveira-Jr**[1,3]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG), GO, Brasil

[2]Instituto Federal de Goiás (IFG), GO, Brasil

[3]Fraunhofer Portugal AICOS, Porto, Portugal

{renan.rodrigues,rogerio.sousa,leandro.freitas}@ifg.edu.br, antoniojr@ufg.br

Abstract. *Federated Learning (FL) allows devices to train a global machine learning model without sharing data. In the context of wireless networks, limited resources and the inherent unreliable nature of the transmission medium introduce delays and errors that compromise the regularity of updating the global model. Therefore, this work proposes a new FL algorithm called DFed-w_{Opt} that considers both the requirements of federated training and a wireless network within the scope of the Internet of Things. To minimize the loss function, DFed-w_{Opt} selects a subset of devices with the largest amount of data for training local models. Then, DFed-w_{Opt} maximizes the probability of successful transmission of models meeting a communication latency and energy consumption policy. The simulation results show that DFed-w_{Opt} increases the number of transmissions and the accuracy of the global model compared to other strategies in the literature.*

Resumo. *O Aprendizado Federado (Federated Learning - FL) permite que dispositivos treinem um modelo global de aprendizado de máquina sem compartilhar dados. No contexto das redes sem fio, os recursos limitados e a natureza não confiável inerente ao meio de transmissão introduzem atrasos e erros que comprometem a regularidade da atualização do modelo global. Dessa forma, este trabalho propõe um novo algoritmo de FL denominado DFed-w_{Opt} que considera tanto os requisitos do treinamento federado quanto de uma rede sem fio no âmbito da Internet das Coisas. Para minimizar a função de perda, DFed-w_{Opt} seleciona um subconjunto de dispositivos com a maior quantidade de dados para o treinamento dos modelos locais. Em seguida, DFed-w_{Opt} maximiza a probabilidade de sucesso da transmissão dos modelos atendendo uma política de latência de comunicação e consumo energético. Os resultados da simulação mostram que DFed-w_{Opt} aumenta a quantidade de transmissões e a acurácia do modelo global em comparação com outras estratégias da literatura.*

1. 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, quando se trata do treinamento, uma suposição é que estes modelos devem ser treinados centralmente na nuvem, usando dados de treinamento de vários dispositivos [Beutel et al. 2020].

O treinamento de modelos de ML em dispositivos de borda despertou um interesse crescente nos últimos anos devido à necessidade do processamento de uma grande quantidade de dados de natureza privada, gerados continuamente por dispositivos como *smartphones*, dispositivos vestíveis, veículos autônomos e outros dispositivos da Internet das Coisas (*Internet of Things* - IoT). Neste contexto, o FL [McMahan et al. 2016] 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.

O FL em redes sem fio apresenta vantagens com relação ao ML centralizado, pois a transmissão dos parâmetros do modelo de ML 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 [Yang et al. 2022]. Além do mais, o FL contribui com a preservação da privacidade dos dados, pois os dados de treinamento permanecem nos dispositivos. Dessa forma, a computação de borda integrada à infraestrutura da rede, conforme proposto na abordagem *Multi-Access Edge Computing* (MEC), contribuiu para a implantação de servidores de parâmetros de FL na borda da rede, ou seja, mais próximo dos dispositivos onde os modelos são treinados.

No entanto, as abordagens de FL em redes sem fio devem considerar a ocorrência de falhas no processo de treinamento distribuído devido a restrições dos recursos de comunicação e a inerente falta de confiabilidade do meio sem fio [Chen et al. 2021]. Além do mais, de acordo com [Zhu et al. 2020], uma estratégia típica para alcançar um desempenho satisfatório no FL é agendar o maior número possível de dispositivos em cada rodada de comunicação. Porém, não é desejável que todos os dispositivos enviem seus modelos locais para o servidor de parâmetros, especialmente quando as atualizações são transmitidas por meio de uma conexão sem fio com recursos limitados [Chen et al. 2022].

Neste contexto, este trabalho propõe um novo algoritmo de FL denominado DFed-w$_{Opt}$[1] (*Data FL Wireless Optimizer*), cujo objetivo é minimizar a função de perda $f(w_{global})$ levando em consideração os requisitos do treinamento federado e da rede IoT sem fio. Dessa forma, para minimizar $f(w_{global})$, propõe-se a sua divisão em dois subproblemas. O primeiro, denominado problema da seleção de dispositivos, tem como objetivo selecionar (a partir de um subconjunto de dispositivos) quais serão os participantes que maximizam a quantidade de dados que é utilizada durante o treinamento dos modelos locais para a próxima rodada de comunicação. O segundo, denominado problema do escalonamento de recursos de comunicação, é tratado como um problema de designação para a alocação de Blocos de Recursos (*Resource Blocks* - RB) de *uplink*, cujo objetivo é maximizar a probabilidade de sucesso da transmissão dos modelos locais dos dispositivos para a BS, atendendo uma política que determina requisitos de latência de comunicação e consumo energético.

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

As contribuições centrais deste artigo são elencadas a seguir: (*i*) apresentam-se as formulações matemáticas que definem o modelo de rede, o modelo de ML, modelo de consumo de energia e o modelo de comunicação utilizado para a simulação da rede IoT sem fio para tarefas de FL; (*ii*) discute-se a formulação do algoritmo DFed-w$_{Opt}$ que seleciona os dispositivos utilizando o método de ordenação seguido por busca linear e maximiza a probabilidade de sucesso da transmissão dos modelos utilizando a técnica de Programação Linear Inteira Mista (*Mixed-Integer Linear Programming* - MILP); (*iii*) descreve-se o algoritmo de DFed-w$_{Opt}$ com base na discussão do problema; (*iv*) apresenta-se a configuração das partições de dados Não-IID dos dispositivos com base nos conjuntos de dados de *benchmark* de FL denominados MNIST e Fashion-MNIST; (*v*) discute-se os resultados de DFed-w$_{Opt}$ em comparação com FedAvg [McMahan et al. 2016] e POC [Cho et al. 2020]. Adicionalmente, com o objetivo de garantir equidade na comparação com DFed-w$_{Opt}$, implementou-se variações de FedAvg e POC, denominadas neste trabalho como FedAvg-w$_{Opt}$ e POC-w$_{Opt}$, incorporando o conhecimento da rede sem fio e utilizando o algoritmo MILP de DFed-w$_{Opt}$ para a alocação dos recursos de comunicação.

Para além desta seção introdutória, o restante deste artigo está organizado em seções, conforme descrito a seguir. A Seção 2 discute os trabalhos relacionados. A Seção 3 apresenta os modelos matemáticos que foram utilizados na simulação da rede IoT sem fio para tarefas de FL. A Seção 4 descreve a formulação de DFed-w$_{Opt}$. A Seção 5 discute os resultados e a análise da simulação. Por fim, a Seção 6 apresenta as considerações finais e indica as orientações para os trabalhos futuros.

2. Trabalhos Relacionados

Desde a proposição original do FL, surgiram um número crescente de estudos dedicados ao aprimoramento deste paradigma. No entanto, muitos destes trabalhos focaram em minimizar os custos de computação ou otimização de estratégias de atualização de parâmetros e não abordaram os problemas relacionados a implementação do FL considerando as incertezas inerentes aos canais sem fio.

Na formulação inicial do FL proposta por [McMahan et al. 2016], os autores introduziram o conceito de FL e apresentaram o desempenho do algoritmo *Federated Averaging* (FedAvg) com diferentes configurações, conjuntos de dados e modelos de ML. No entanto, o trabalho não realizou uma discussão abrangente no contexto das redes sem fio. Em [Cho et al. 2020], os autores apresentam o algoritmo *Power-Of-Choice* (POC), que seleciona os dispositivos para a próxima rodada de comunicação com base nos maiores valores da função de perda local. Como alternativa, os autores sugerem que o servidor utilize informações da perda média acumulada previamente enviada pelos dispositivos junto com os modelos locais. Entretanto, os autores não apresentaram resultados que pudessem destacar a eficácia da proposta em um cenário realista de redes sem fio.

O artigo de [Cao et al. 2023] apresenta uma visão geral das metodologias atuais para o aprendizado distribuído em termos de comunicação. Os autores apontam que o FL apresenta desafios significativos devido às restrições de comunicação e a natureza dinâmica das redes sem fio. Além disso, é destacada a importância da seleção de dispositivos e da otimização dos recursos de comunicação para melhorar a eficiência do FL em redes sem fio. Em [Tran et al. 2019], os autores desenvolveram um dos primeiros traba-

lhos que consideraram tanto as métricas do FL quanto os fatores inerentes às redes sem fio. O artigo apresenta um modelo de FL em redes sem fio considerando características de computação e comunicação. Em seguida, é formulado um problema de minimização para tarefas de FL com o objetivo de otimizar o tempo de treinamento dos modelo locais dos dispositivos levando em consideração o consumo de energia. No entanto, o modelo proposto não avalia os efeitos da implementação de uma política de seleção de dispositivos para cada rodada de comunicação.

O trabalho de [Chen et al. 2021] apresenta um modelo de FL em redes sem fio e uma proposta de seleção de dispositivos e alocação de recursos de forma conjunta por meio da formulação de um problema de otimização que visa minimizar a perda de treinamento do modelo global. De forma semelhante, o estudo de [Chen et al. 2022] investigou a otimização de recursos de comunicação para o FL, onde o problema foi dividido nos subproblemas de escalonamento de dispositivos e alocação de recursos. A política de escalonamento considerou a reutilização de parâmetros obsoletos dos modelos locais para reduzir os custos de comunicação. Contudo, as pesquisas de [Chen et al. 2021] e [Chen et al. 2022] não abordam de maneira abrangente a escalabilidade para ambientes de redes IoT sem fio, uma vez que os resultados foram obtidos com base na iteração de poucos dispositivos, bem como, utilizaram exclusivamente a versão padrão do MNIST como conjunto de dados de referência.

Motivado pelas lacunas identificadas nos trabalhos relacionados, propõe-se o DFed-w$_{Opt}$ como um novo algoritmo de FL que considera tanto os fatores inerentes ao ambiente sem fio como fatores que influenciam o processo de aprendizado do modelo global. A avaliação de DFed-w$_{Opt}$ envolve uma densidade de dispositivos ligeiramente superior em comparação com os trabalhos citados no contexto de redes sem fio e utiliza variações do MNIST e Fashion-MNIST que tornam os dados mais heterogêneos como uma característica intrínseca dos dispositivos das redes IoT sem fio.

3. Modelo do Sistema

Esta seção apresenta os modelos matemáticos utilizados na implementação da simulação da rede IoT sem fio e das tarefas de FL.

3.1. Modelo de Rede

Considere uma rede IoT sem fio com uma BS conectada diretamente a um servidor agregador de FL conforme proposto na abordagem MEC, com um conjunto $S = \{k_1, k_2, ..., k_K\}$ de K dispositivos IoT antes do início de uma rodada de treinamento. Cada cliente possui um conjunto de dados $\mathcal{P}_k$ com $n_k = |\mathcal{P}_k|$ amostras armazenadas em seus respectivos dispositivos locais. Estes dispositivos IoT estão conectados a BS por meio de uma conexão sem fio e possuem capacidade para coletar dados e treinar um modelo local para uma determinada tarefa de FL.

3.2. Modelo de ML

A ideia do FL é realizar a agregação de um modelo global ao longo de várias rodadas de comunicação com base nos parâmetros dos modelos treinados localmente em cada dispositivo minimizando a função de perda $f(w_{global})$ usando os dados $\mathcal{P} = \cup_{k=1}^{K} \mathcal{P}_k$. Dessa forma, o processo de treinamento do FL pode ser expresso como

$$\min_{w_1,\ldots,w_K} f(w_{global}) \triangleq \min_{w_1,\ldots,w_K} \frac{1}{|\mathcal{P}|} \sum_{k=1}^{K} \sum_{n=1}^{n_k} f(w_k, x_{kn}) = \min_{w_1,\ldots,w_K} \sum_{k=1}^{K} \frac{n_k}{m} f_k(w_k) \tag{1}$$

$$\text{s.a.} \quad w_1 = w_2 = \ldots = w_K = w_{global}, \tag{1a}$$

onde $f(w_k, x_{kn})$ é a função de perda que avalia o desempenho do modelo local w_k por meio da observação da saída produzida pelo treinamento com x_{kn} amostras de dados e $m = \sum_{k \in S} n_k$. Por exemplo, para tarefas supervisionadas, a Equação (1) visa encontrar os parâmetros que resultam em previsões que se aproximam ao máximo possível dos rótulos reais em x_{kn} de acordo com uma métrica definida pela função de perda. A restrição (1a) garante que os modelos compartilhados entre os dispositivos e a BS devem ser idênticos para uma determinada tarefa de FL.

Para realizar o treinamento da Equação (1) com base no algoritmo FedAvg, a BS deve transmitir um modelo global w_0 por meio de um enlace sem fio, cujo modelo deve ser criado ou carregado a partir de um modelo pré-treinado. No início de cada rodada de comunicação t, a BS deve selecionar uma fração aleatória dos K dispositivos, gerando um conjunto $\mathcal{S}_t$ de m dispositivos. Em seguida, a BS utiliza os canais de *downlink* para transmitir o estado atual do modelo w_t para cada dispositivo local. Cada participante treina o modelo local utilizando um algoritmo de otimização baseado na Descida do Gradiente Estocástico (*Stochastic Gradient Descent* - SGD) para cada $b \in B$ mini-lotes de $\mathcal{P}_k$ durante E épocas locais, onde $\nabla \ell$ representa o gradiente de ℓ em b. Após a última atualização $w \leftarrow w - \eta \nabla \ell(w; b)$, o dispositivo utiliza um canal de *uplink* para enviar os parâmetros do modelo local w para a BS conectada a um servidor agregador de parâmetros.

Dessa forma, os modelos w_{t+1}^k são recebidos e agregados pelo servidor, gerando o estado atual w_{t+1} do modelo global aplicando a atualização $w_{t+1} \leftarrow \sum_{k \in \mathcal{S}_t} \frac{n_k}{m_t} w_{t+1}^k$, onde $m_t = \sum_{k \in \mathcal{S}_t} n_k$. A agregação trata-se da média ponderada dos parâmetros dos dispositivos, cujos pesos são definidos com base na quantidade de dados utilizada no treinamento local. O processo de treinamento distribuído pode ser repetido por várias rodadas de comunicação até que um determinado valor de convergência seja alcançado ou quando outro critério de parada seja satisfeito [Li et al. 2021].

3.3. Modelo de Consumo de Energia

O modelo de consumo de energia incorpora a energia necessária para a transmissão e o treinamento do modelo local. A demanda energética da BS não é considerada, uma vez que esta normalmente possui um fornecimento contínuo de energia. De acordo com [Chen et al. 2021], consumo de energia de cada dispositivo é dado por

$$e_k(r_k, P_k) = \zeta \omega_k \vartheta^2 Z(w_k) + P_k l_k^U(r_k, P_k), \tag{2}$$

onde ϑ, ω_k e ζ referem-se, respectivamente, a frequência do *clock*, o número de ciclos da unidade central de processamento e o coeficiente de consumo de energia de cada dispositivo. Na Equação (2), a primeira parcela representa o consumo de energia para o treinamento do modelo local no dispositivo e a segunda parcela representa o consumo de energia para a transmissão do modelo local de um determinado dispositivo para a BS.

3.4. 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 [Chen et al. 2021], a taxa de *uplink* do usuário *k* transmitindo os parâmetros do modelo w_k para a BS pode ser formulada como

$$c_k^U(r_k, P_k) = \sum_{n=1}^{R} r_{k,n} B_n^U \mathbb{E}\left(log_2\left(1 + \frac{P_k h_k}{I_n + B_n^U N_0}\right)\right), \tag{3}$$

onde $r_k = [r_{k,1}, ..., r_{k,R}]$ é o vetor de alocação de RBs, R é o número de RBs, $r_{k,n} \in 0, 1$ e $\sum_{n=1}^{R} r_{k,n} = 1$, com $r_{k,n} = 1$ indicando que o RB n está alocado para o dispositivo k. A a largura de banda de cada RB é denotada por B_n^U e P_k é a potência de transmissão do dispositivo k. O ganho do canal entre o dispositivo k e a BS é dado por $h_k = o_k d_k^{-\alpha}$, onde d_k é a distância entre o dispositivo k e a BS, o_k é o parâmetro de desvanecimento de Rayleigh e α é um expoente que afeta como o ganho do canal varia com a distância. $\mathbb{E}(.)$ é a expectativa da taxa de dados em relação a h_k, 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 é geralmente muito maior do que a potência dos dispositivos. Portanto, toda a largura de banda do *downlink* pode ser utilizada para transmitir o modelo global. Dessa forma, a taxa de dados de *downlink* alcançada pela BS ao transmitir os parâmetros do modelo global para cada dispositivo é dada por

$$c_k^D = B^D \mathbb{E}\left(log_2\left(1 + \frac{P_B h_k}{I^D + B^D N_0}\right)\right), \tag{4}$$

onde B^D é a largura de banda utilizada pela BS para transmitir o modelo global para cada dispositivo, P_B é a potência de transmissão da BS e I^D é a interferência causada por outras BSs que não participam da tarefa de FL.

Para formular o atraso de transmissão, assume-se que os modelos de FL são transmitidos por meio de um único pacote. Assim, o atraso de transmissão entre um dispositivo k e a BS no *uplink* e *downlink* podem ser respectivamente formulados como

$$l_k^U(r_k, P_k) = \frac{S_{pkt}^U}{c_k^U(r_k, P_k)}, \tag{5}$$

$$l_k^D = \frac{S_{pkt}^D}{c_k^D}, \tag{6}$$

onde S_{pkt}^U é o tamanho do pacote de *uplink*, ou seja, o número de bits que os dispositivos necessitam transmitir para a BS através do enlace sem fio, considerando o tamanho dos parâmetros dos modelos locais acrescido de outras informações, tais como, a acurácia do treinamento no dispositivo e valor da função de perda local. Da mesma forma, denota-se S_{pkt}^D como sendo o tamanho do pacote de *downlink*, ou seja, o número de bits que a BS necessita transmitir para os dispositivos através do enlace sem fio, considerando o tamanho dos parâmetros do modelo global acrescido de outras informações que podem indicar alguma instrução do servidor para os dispositivos.

Para simular o efeito de erros de transmissão de pacotes, considera-se que a BS não solicitará aos dispositivos o reenvio de modelos quando estes forem recebidos com

erros. Portanto, sempre que um pacote recebido contiver erros, a BS não utilizará o respectivo modelo para a atualização do modelo global. Neste caso, conforme apresentado por [Chen et al. 2021], a taxa de erro de transmissão de pacote do *uplink* é dado por

$$q_k^U(r_k, P_k) = \sum_{n=1}^{R} r_{k,n} \mathbb{E}\left(1 - \exp\left(-\frac{m(I_n + B^U N_0)}{P_k h_k}\right)\right), \tag{7}$$

onde $\mathbb{E}(.)$ é a expectativa da taxa de erro de pacote considerando h_k em r_n, com m sendo um limiar (*waterfall threshold*) que define a qualidade da transmissão.

Considerando que o êxito da transmissão de um pacote dos dispositivos para a BS sem erros é o complemento da taxa de erro de pacote do *uplink*, pode-se formular a probabilidade de sucesso da transmissão do modelo local para a BS como

$$p_k^U(r_k, P_k) = 1 - q_k^U(r_k, P_k). \tag{8}$$

A estimativa de q_k^U ou p_k^U na transmissão de pacotes é útil para a avaliação da qualidade dos canais de comunicação r_n que podem ser alocados para cada dispositivo participante de uma tarefa de FL. Por exemplo, se a probabilidade de sucesso for significativa, isso pode indicar que a transmissão de um modelo local terá maior chance de ser recebido pela BS e agregado ao modelo global pelo servidor de parâmetros.

Por fim, denota-se a probabilidade de sucesso da transmissão do modelo local para a BS considerando o atendimento de uma política da seguinte forma

$$p\gamma_k^U(r_k, P_k) = \begin{cases} p_k^U(r_k, P_k) & \text{se } l_k^U \leq \gamma_T \text{ e } e_k \leq \gamma_E, \\ 0 & \text{caso contrário,} \end{cases} \tag{9}$$

onde γ_T define o requisito de atraso e γ_E define requisito de consumo energético.

4. Formulação do DFed-w$_{Opt}$

Para resolver o problema geral da Equação (1), DFed-w$_{Opt}$ propõe a sua divisão em dois subproblemas apresentados nas Seções 4.1 e 4.2.

4.1. Seleção de Dispositivos

Em uma tarefa de FL, cada dispositivo contribui com um modelo local treinado com dados abrangendo diferentes características, contextos e variações do problema. Dessa forma, direcionando a estratégia para um treinamento local abrangendo mais dados, espera-se DFed-w$_{Opt}$ melhore a representatividade dos modelos treinados localmente. Portanto, ao denotar S_t como uma fração f_t de K dispositivos e b como um vetor binário indicador da seleção de n_p dispositivos, o problema da maximização da quantidade de dados de $n_p \leq |S_t|$ dispositivos pode ser expresso como

$$\max \sum_{k \in S_t} |\mathcal{P}_k| b_k, \tag{10}$$

$$\text{s.a.} \sum_{k=1}^{|S_t|} b_k = n_p, \tag{10a}$$

$$b_k \in \{0, 1\}, \tag{10b}$$

onde $|\mathcal{P}_k|$ é a quantidade de amostras de dados do dispositivo $k \in S_t$ e n_p é a quantidade parcial de dispositivos selecionados para a etapa de escalonamento dos recursos

de comunicação. A restrição (10a) garante que o número de dispositivos selecionados é igual a n_p e a restrição (10b) indica que b é um vetor binário, onde $b_k = 1$ indica que o dispositivo foi selecionado do subconjunto S_t e $b_k = 0$ indica o contrário.

4.2. Escalonamento dos Recursos de Comunicação

Ao denotar uma matriz binária c como um indicador de atribuição de um RB i ao dispositivo k, cujo objetivo é maximizar a probabilidade de sucesso da transmissão dos modelos locais dos dispositivos para a BS, o problema do escalonamento dos recursos de comunicação de DFed-w$_{Opt}$ pode ser formulado como

$$\max \sum_{i=1}^{R} \sum_{k=1}^{n_p} p\gamma_k^U c_{ik}, \tag{11}$$

$$\text{s.a.} \quad \sum_{k=1}^{n_p} c_{ik} = 1 \quad \forall i \in R, \tag{11a}$$

$$\sum_{i=1}^{R} c_{ik} = 1 \quad \forall k \in n_p, \tag{11b}$$

$$\sum c_{ik} = n_f \quad \forall i \in R \text{ e } \forall k \in n_p, \tag{11c}$$

$$c_{ik} \in \{0, 1\}, \tag{11d}$$

onde $n_f \leq n_p$ define o número potencial de dispositivo que devem ser selecionados para a próxima rodada de comunicação. A restrição (11a) garante que cada RB é alocado para um único dispositivo, a restrição (11b) garante que cada dispositivo é atribuído a exatamente um RB, a restrição (11c) garante que o número de dispositivos selecionados é igual a n_f e a restrição (11d) define que c é uma matriz binária, onde $c_{ik} = 1$ indica que o RB i foi alocado para o dispositivo k e $c_{ik} = 0$ indica o contrário.

Por fim, DFed-w$_{Opt}$ denota um vetor binário d que define os dispositivos selecionados para a próxima rodada de comunicação da seguinte forma

$$d_k = \begin{cases} 1 & \text{se } p\gamma_k^U(r_k, P_k) > 0, \\ 0 & \text{caso contrário}, \end{cases} \tag{12}$$

onde $d_k = 1$ indica que o dispositivo foi selecionado para a próxima rodada de comunicação e $d_k = 0$ indica o contrário.

4.3. Algoritmo DFed-w$_{Opt}$

O Algoritmo 1 apresenta a estratégia DFed-w$_{Opt}$ em redes IoT sem fio com base nas Equações (10), (11) e (12). Na linha 6, observa-se que o servidor agregador seleciona n_p dispositivos maximizando a quantidade de dados locais. Em seguida, na linha 8, é realizado o escalonamento dos RBs de *uplink* para n_f dispositivos maximizando a probabilidade de sucesso da transmissão dos modelos. Por fim, na linha 10, é definido o vetor binário d que define os dispositivos selecionados para a próxima rodada de comunicação atendendo ao requisito de atraso γ_T e ao requisito de consumo energético γ_E.

Algoritmo 1: *DFed-w$_{Opt}$ para FL em Redes IoT sem Fio*

Servidor:
  inicialização de w_0
  para cada rodada $t = 1, 2, ...$ **faça**
    $\mathcal{S}_t \leftarrow$ (conjunto aleatório de $max(f_t \cdot \mathcal{K}, 1)$ clientes)
    ▷*Seleção de n_p dispositivos maximizando a qtde local de dados*
    Defina b com $n_p \leq |S_t|$ dispositivos utilizando a Equação (10)
    ▷*Escalonamento dos RBs de uplink para n_f dispositivos considerando γ_T e γ_E*
    Defina c com $n_f \leq n_p$ dispositivos utilizando a Equação (11)
    ▷*Seleção dos dispositivos para a próxima rodada de comunicação*
    Defina d utilizando a Equação (12)
    para cada dispositivo k, onde $d_k = 1$ **em paralelo faça**
      $w_{t+1}^k \leftarrow$ DispositivoSelecionado(k, w_t)
    $m_t \leftarrow \sum_{k \in \mathcal{S}_t} n_k \qquad w_{t+1} \leftarrow \sum_{k \in \mathcal{S}_t} \frac{n_k}{m_t} w_{t+1}^k$
DispositivoSelecionado(k, w): ▷ *Para cada dispositivo k*
  $\mathcal{B} \leftarrow$ (divisão de $\mathcal{P}_k$ em lotes de tamanho $\mathcal{B}$)
  para cada época local i 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

5. Avaliação e Análise de Desempenho

Considere uma rede IoT sem fio atendendo uma área circular com raio r de 500 metros com uma BS no centro. A BS está diretamente associada a um servidor agregador, onde existem $K = 100$ dispositivos conectados para uma tarefa de FL. Conforme a Figura 1, os dispositivos são distribuídos de maneira uniforme e aleatória com distâncias entre 100 e 500 metros da BS. A largura de banda de *uplink* de cada RB é de 1 MHz com a atribuição de uma interferência distinta e incremental. Cada dispositivo é configurado para transmitir os modelos locais para a BS com a potência de 0.01 W. A largura de banda do *downlink* é de 20 MHz, onde a BS transmite os modelos globais para os dispositivos com uma potência de 1 W. A modelagem de h_k incorpora um efeito de desvanecimento que indica que o ganho do canal diminui conforme aumenta a distância dos dispositivos em relação a BS. Na Tabela 1 são apresentados outros parâmetros da simulação.

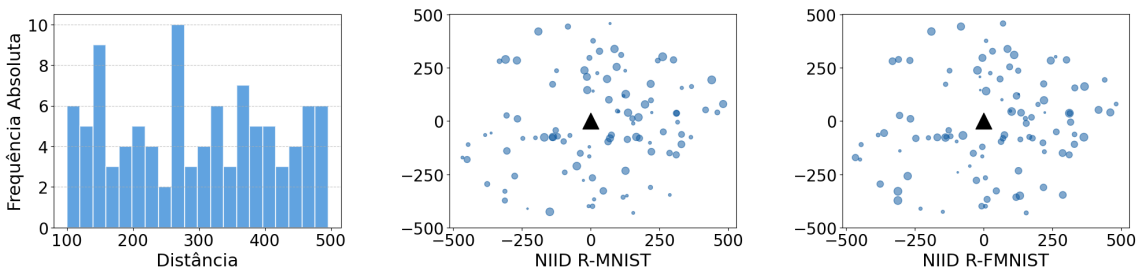


Figura 1. Distância dos dispositivos com relação a BS em metros

As tarefas de FL consideram problemas de classificação de imagens uti-

lizando a base dos conjuntos de dados de *benchmark* utilizados em pesquisas de FL denominados MNIST e Fashion-MNIST, conforme os trabalhos de [McMahan et al. 2016], [Amannejad 2020], [Beutel et al. 2020], [Zhu et al. 2020], [Chen et al. 2021], [Zhao et al. 2022] e [Chen et al. 2022].

Tabela 1. Parâmetros da simulação [Chen et al. 2021]

Parâmetro	Valor	Parâmetro	Valor
α	2	ϑ	10^9
N_0	-174 dBm/Hz	ζ	10^{-27}
m	0.023 dB	ω_k	40

Diferentemente das referências citadas, este trabalho buscou tornar os dados dos dispositivos um pouco mais heterogêneos como uma característica inerente das redes IoT sem fio. Ao utilizar variações do MNIST e Fashion-MNIST para a distribuição de dados de forma Não-IID, introduziu-se um desafio para a otimização das acurácias dos modelos locais e a agregação do modelo global. Isto evitou a convergência trivial para valores próximos à totalidade e permitiu uma observação mais notável sobre desempenho dos algoritmos analisados.

Dessa forma, os conjuntos de dados do MNIST e Fashion-MNIST foram divididos em 10 subconjuntos com amostras do mesmo rótulo, utilizando 75% das amostras para o conjunto de treinamento e 25% para o conjunto de teste. Em seguida, cada dispositivo recebeu uma partição de treinamento e teste, onde: 90% das amostras pertencem ao mesmo rótulo e os 10% restantes pertencem igualmente aos demais rótulos; cada imagem é rotacionada em até 45° em sentido horário ou anti-horário; a quantidade final de dados é dada por um fator entre $[0.25, 1]$ da partição inicial. Neste trabalho, as partições de MNIST e Fashion-MNIST foram nomeadas, respectivamente, como NIID R-MNIST e NIID R-FMNIST. Na Figura 1, cada dispositivo é representado por um círculo, onde o tamanho do círculo é proporcional à quantidade de dados.

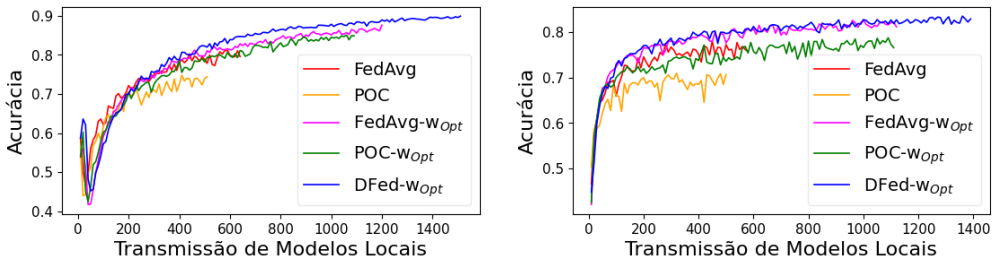
Para a tarefa de classificação do conjunto de dados NIID R-MNIST, foi utilizada uma arquitetura de rede neural do tipo *Multi-Layer Perceptron* (MLP) com 101.770 parâmetros. Por outro lado, para a tarefa de classificação de NIID R-FMNIST adotou-se uma arquitetura de rede neural do tipo *Convolutional Neural Network* (CNN) com 222.516 parâmetros.

5.1. Resultados da Simulação

Esta seção discute os resultados de DFed-w$_{Opt}$ em comparação com os algoritmos FedAvg e POC previamente discutidos nos trabalhos relacionados. O algoritmo FedAvg determina aleatoriamente a seleção de dispositivos e a utilização dos recursos de comunicação. Por outro lado, POC seleciona os dispositivos com base na perda média acumulada, que pode ser previamente enviada pelos dispositivos juntamente com os modelos locais, sem nenhum conhecimento da rede sem fio. Além do mais, com o objetivo de garantir equidade na comparação com DFed-w$_{Opt}$, implementou-se variações de FedAvg e POC, denominadas neste trabalho como FedAvg-w$_{Opt}$ e POC-w$_{Opt}$, incorporando o conhecimento da rede sem fio e utilizando o algoritmo MILP de DFed-w$_{Opt}$ para a alocação dos recursos de comunicação.

Em razão da natureza estocástica inerente à simulação, cada estratégia de FL foi executada por 15 vezes para cada conjunto de dados com o objetivo de obter estatísticas relacionadas ao desempenho médio. As estratégias foram executadas por 200 rodadas de comunicação, utilizando 15 RBs de *uplink* e uma fração f_t de $K = 100$ dispositivos com uma época de treinamento local. Neste caso, FedAvg e FedAvg-w$_{Opt}$ utilizaram o valor de $f_t = 10$. As demais estratégias utilizaram o valor de $f_t = 20$ para definir o subconjunto inicial de dispositivos para a aplicação das estratégias de seleção de dispositivos. Em todas as estratégias, no máximo 10 dispositivos foram selecionados em cada rodada de comunicação após a alocação dos RBs de *uplink* em conformidade com a política que estabelece os requisitos de atraso e consumo energético.

Considerando que o MLP e a CNN possuem um número distinto de parâmetros, foram definidas duas políticas para a avaliação de DFed-w$_{Opt}$, assegurando que qualquer dispositivo pudesse transmitir seu modelo em pelo menos um RB de *uplink*. Para NIID R-MNIST com MLP, utilizou-se $\gamma_T = 200$ ms como requisito de atraso e $\gamma_E = 0.0025$ J como requisito de consumo energético. Para NIID R-FMNIST com CNN, utilizou-se $\gamma_T = 400$ ms e $\gamma_E = 0.005$ J. As Figuras 2a e 2b apresentam a evolução da acurácia dos algoritmos de FL considerando a quantidade de transmissões de atualizações de modelos locais utilizando os referidos conjuntos de dados.



(a) NIID R-MNIST com MLP **(b) NIID R-FMNIST com CNN**

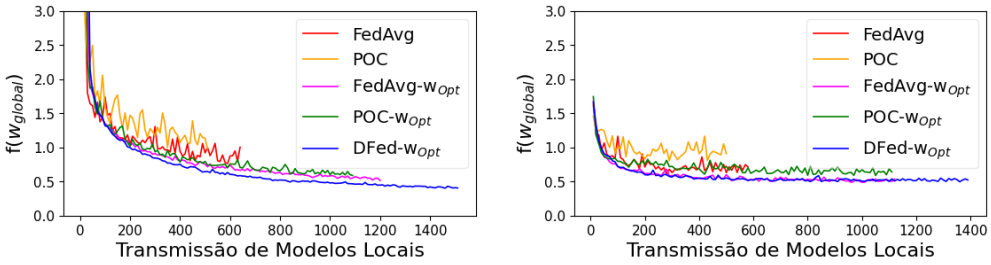
Figura 2. Evolução da acurácia do FL

Observa-se que as acurácias das estratégias de FL aumentam na medida do crescimento do número de transmissões de modelos locais. Neste caso, o compartilhamento frequente de modelos locais durante as rodadas de treinamento permitiu que o modelo global se beneficiasse das informações adquiridas nos padrões de dados dos dispositivos. Após a finalização das 200 rodadas de comunicação, o melhor desempenho dos algoritmos FedAvg, POC, FedAvg-w$_{Opt}$, POC-w$_{Opt}$ e DFed-w$_{Opt}$ apresentaram, respectivamente, uma acurácia no valor de 81.14, 76.11, 87.59, 85.25 e 89.90 para NIID R-MNIST com MLP e de 78.16, 73.50, 82.62, 78.71 e 83.62 para NIID R-FMNIST com CNN.

Destaca-se que DFed-w$_{Opt}$ obteve uma melhor acurácia ao considerar tanto os requisitos do treinamento federado como da rede IoT sem fio. O desempenho de DFed-w$_{Opt}$ é caracterizado pela otimização da alocação dos RBs de *uplink* que permitiu um aumento no número de transmissões de modelos locais. Ao mesmo tempo, a escolha direcionada dos dispositivos para um treinamento local abrangendo um conjunto mais amplo de dados aprimorou a representatividade dos modelos treinados, ampliando a capacidade de DFed-w$_{Opt}$ na generalização do modelo global para a realização de previsões corretas.

De forma análoga, as Figuras 3a e 3b apresentam a evolução da $f(w_{global})$.

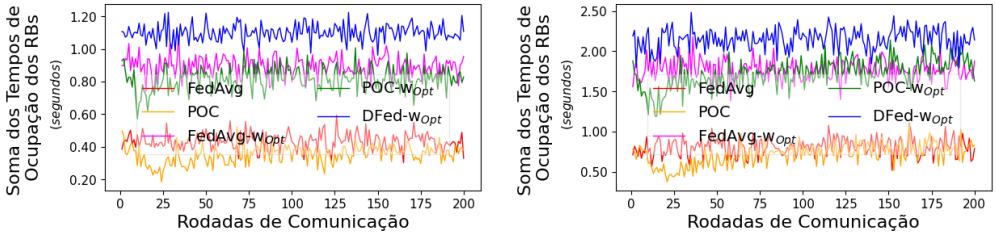
Observa-se que na medida em que o número de transmissões aumenta, $f(w_{global})$ das estratégias de FL diminuem. Além do mais, destaca-se que a $f(w_{global})$ do DFed-w$_{Opt}$ exerce um domínio como um limite inferior com relação aos demais algoritmos.



(a) NIID R-MNIST com MLP **(b) NIID R-FMNIST com CNN**

Figura 3. Evolução da $f(w_{global})$ do FL

As Figuras 4a e 4b apresentam a soma dos tempos de ocupação dos RBs. Neste caso, observa-se que DFed-w$_{Opt}$ apresenta uma utilização mais eficaz dos recursos e indica que os dispositivos mais distantes da BS, com um ganho de canal mais baixo e que exigem o tempo limite estabelecido pela política, conseguem treinar e transmitir seus modelos. Por outro lado, observa-se que FedAvg e POC não conseguem uma alocação apropriada, acarretando na ociosidade do uso dos recursos de comunicação.



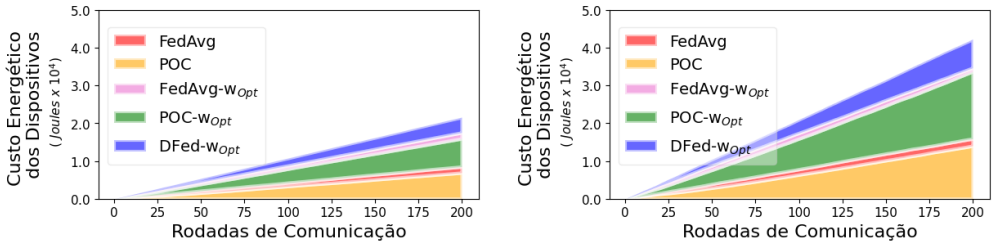
(a) NIID R-MNIST com $\gamma_T = 200$ ms **(b) NIID R-FMNIST com $\gamma_T = 400$ ms**

Figura 4. Tempo de ocupação dos RBs

As Figuras 5a e 5b apresentam a evolução do custo energético dos dispositivos. Observa-se que, para todas as estratégias de FL, o treinamento de NIID R-FMNIST com CNN ocasionou em um impacto significativo no consumo energético quando comparado com o treinamento de NIID R-MNIST com MLP. Isto se deve ao fato da arquitetura CNN possuir 2.21 vezes mais parâmetros que o MLP. Como resultado, foram requeridos valores mais elevados para γ_T e γ_E para assegurar que cada dispositivo pudesse transmitir seus modelos em pelo menos um canal de *uplink*.

Além do mais, é possível constatar que a energia consumida acumulada ao longo das rodadas de comunicação é maior para as estratégias que incorporaram o conhecimento da rede sem fio para a alocação de recursos de comunicação. Isto se deve ao fato do consumo energético dos dispositivos estar diretamente relacionado ao tempo de treinamento e transmissão dos modelos. Além disso, a técnica de MILP, ao maximizar a probabilidade de sucesso da transmissão dos modelos, pode incorporar mais dispositivos que estão distantes da BS, resultando em um aumento no tempo e no consumo energético necessário para a transmissão dos modelos. Neste caso, o consumo energético de DFed-w$_{Opt}$ é justificável pela melhoria da acurácia global do FL em comparação com os demais algoritmos,

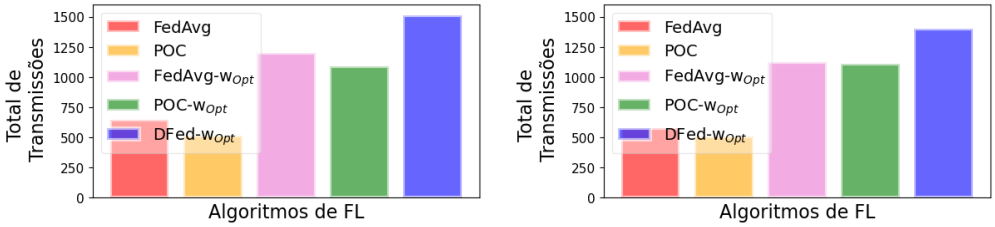
especialmente para os casos de uso onde a precisão é uma prioridade, como por exemplo, em aplicações médicas, financeiras ou de segurança.



(a) NIID R-MNIST com $\gamma_E = 0.0025$ J **(b) NIID R-FMNIST com $\gamma_E = 0.005$ J**

Figura 5. Custo energético dos dispositivos

As Figuras 6a e 6b apresentam o total de transmissões de modelos locais de cada algoritmo de FL. Ao analisar o desempenho de FedAvg e POC, as limitações da rede sem fio ficam ainda mais evidentes quando não há uma alocação eficaz de recursos de comunicação. Como estes algoritmos desconhecem o meio de transmissão, a alocação dos RBs é realizada de forma aleatória. Como consequência, muitos dispositivos são impedidos de transmitir seus modelos.



(a) NIID R-MNIST **(b) NIID R-FMNIST**

Figura 6. Quantidade total de transmissões de modelos locais

Por outro lado, as versões FedAvg-w$_{Opt}$ e POC-w$_{Opt}$ que incorporaram o conhecimento da rede sem fio de DFed-w$_{Opt}$ para a alocação dos recursos de comunicação obtiveram um volume significativamente maior de transmissões de modelos e alcançaram um melhor desempenho do modelo global. Como FedAvg-w$_{Opt}$, POC-w$_{Opt}$ e DFed-w$_{Opt}$ utilizaram a mesma estratégia de alocação dos recursos de comunicação, a diferença na quantidade de transmissões de modelos é atribuída às diferentes estratégias de seleção de dispositivos. Neste caso, desta-se que DFed-w$_{Opt}$ adaptou-se melhor às características da rede sem IoT fio em um cenário de densificação de dispositivos com dados heterogêneos e com quantidades distintas de dados.

6. Considerações Finais

Este trabalho investigou o problema da seleção de dispositivos e da alocação dos recursos de comunicação em redes sem fio para tarefas de FL. Dessa forma, formulou-se o DFed-w$_{Opt}$ como um problema de otimização que considera fatores inerentes ao ambiente sem fio em conjunto com o processo de aprendizado dos modelos. A otimização na alocação dos RBs permitiu um aumento no número de transmissões de modelos enquanto a seleção de dispositivos abrangendo mais dados ampliou a capacidade de generalização

do modelo global. Os resultados mostraram que DFed-w$_{Opt}$ melhorou a acurácia do modelo global em comparação com os algoritmos FedAvg e POC que desconhecem o meio de transmissão, bem como, das suas respectivas versões, denominadas neste trabalho como FedAvg-w$_{Opt}$ e POC-w$_{Opt}$, que incorporaram o conhecimento da rede sem fio para a alocação dos recursos de comunicação. Além do mais, ressalta-se que o código-fonte de DFed-w$_{Opt}$ está disponível para permitir a reprodução e validação dos resultados. Para trabalhos futuros, pretende-se utilizar simuladores de rede que implementam protocolos reais de comunicação para a experimentação de soluções atuais de FL e a proposição de novos algoritmos para a otimização de arquiteturas de FL em redes sem fio visando a minimização do consumo energético e considerando a mobilidade dos dispositivos.

Referências

Amannejad, Y. (2020). Building and Evaluating Federated Models for Edge Computing. In *2020 16th International Conference on Network and Service Management (CNSM)*, pages 1–5.

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., and Lane, N. D. (2020). Flower: A Friendly Federated Learning Research Framework. *CoRR*, abs/2007.14390.

Cao, X., Başar, T., Diggavi, S., Eldar, Y. C., Letaief, K. B., Poor, H. V., and Zhang, J. (2023). Communication-Efficient Distributed Learning: An Overview. *IEEE Journal on Selected Areas in Communications*, 41(4):851–873.

Chen, H., Huang, S., Zhang, D., Xiao, M., Skoglund, M., and Poor, H. V. (2022). Federated Learning Over Wireless IoT Networks With Optimized Communication and Resources. *IEEE Internet of Things Journal*, 9(17):16592–16605.

Chen, M., Yang, Z., Saad, W., Yin, C., Poor, H. V., and Cui, S. (2021). A Joint Learning and Communications Framework for Federated Learning Over Wireless Networks. *IEEE Transactions on Wireless Communications*, 20(1):269–283.

Cho, Y. J., Wang, J., and Joshi, G. (2020). Client Selection in Federated Learning: Convergence Analysis and Power-of-Choice Selection Strategies. *CoRR*, abs/2010.01243.

Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X., and He, B. (2021). A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection. *IEEE Transactions on Knowledge and Data Engineering*, PP:1–1.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2016). Communication-Efficient Learning of Deep Networks from Decentralized Data. *arXiv*.

Tran, N. H., Bao, W., Zomaya, A., Nguyen, M. N. H., and Hong, C. S. (2019). Federated Learning over Wireless Networks: Optimization Model Design and Analysis. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 1387–1395.

Yang, Z., Chen, M., Wong, K.-K., Poor, H. V., and Cui, S. (2022). Federated Learning for 6G: Applications, Challenges, and Opportunities. *Engineering*, 8:33–41.

Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., and Chandra, V. (2022). Federated Learning with Non-IID Data. *arXiv*.

Zhu, G., Wang, Y., and Huang, K. (2020). Broadband Analog Aggregation for Low-Latency Federated Edge Learning. *IEEE Transactions on Wireless Communications*, 19(1):491–506.