Cooperative Interference Estimation using LSTM-based Federated Learning for In-X Subnetworks

Pramesh Gautam, MohammadAmin Vakilifard, Carsten Bockelmann, Armin Dekorsy
Department of Communications Engineering, University of Bremen, Germany,
Email: {gautam, vakilifard, bockelmann, dekorsy}@ant.uni-bremen.de

***Abstract*—"Network of subnetworks" is envisioned to be a key enabler in a 6G network with extremely low (100 μs) latency and high-reliability (99.9999%-99.99999%) in demanding applications. However, to achieve this level of performance, it is necessary to introduce a proactive and robust interference estimation considering the random mobility of subnetworks in an ultra-dense environment. We propose long-short term memory (LSTM) to learn the non-linear behavior of interference power time series for robust estimation and prediction in in-X subnetworks. This proposed method empowers to prediction/estimation of interference on the subnetwork itself. The achieved estimation result is compared with the moving average-based estimator. Furthermore, we introduce federated learning (FL) in-X subnetworks' interference estimation, which learns cooperatively from the interference power vector of subnetworks participating in training. The results indicate that the proposed FL-based estimator achieves a higher convergence speed and lower estimation error.**

***Index Terms*—In-X, Smart Industry, 6G, Interference Estimation, Interference Prediction, Subnetworks, LSTM, FL.**

I. Introduction

While the current and still evolving 5G communication system will significantly outperform older standards, in as little as ten years it won't be able to keep up with the growing needs of intelligent and automated systems [1]. According to the European vision for the 6G white paper [2], there is increasing demand for vertical applications to be integrated into 6G. Such specialized networks may then exist as a subnetwork within a surrounding 6G network. Some of the applications of in-X subnetworks are in-robots, in-vehicle, in-body, in-house, etc., with different service demands. These subnetworks have their own extreme requirements in terms of low latency, high reliability, life-criticality, and data rate, etc [3]. These subnetworks can be either connected to the existing cellular network directly or via local interactive devices or may remain unconnected to the 6G network.

These use cases manifest that the mobility model of each subnetwork is different and ranges from rapid to slow movement as well as deterministic or random path, which results in highly dynamic and diverse interference conditions. It is anticipated that industrial wireless networks replace the traditional wired industrial network infrastructures, including EtherCAT (Ethernet for Control Automation Technology), Profinet, and time-sensitive network (TSN) solutions to achieve a similar level of reliability, latency, and increased data rate as a more flexible solution [4]. Hence, in this research, we focus on the in-robot/factory scenario, i.e., an industrial scenario of an in-X subnetwork that demands extremely low latencies. Due to this requirement, the resource allocation needs to be carried out instantly. This necessitates proactive interference estimation/prediction.

This work is supported by the German Ministry of Education and Research (BMBF) under the grants of 16KISK109 (6G-ANNA) and 16KISK016 (Open6GHub).

The prediction of interference power can be classified into classical statistical and data-driven methods like artificial intelligence (AI) [5]. Classical statistical-based interference prediction is based on extracting the statistical properties of the interference like mean and distributions by considering tail or extreme value theory to predict the next time steps of interference power values [6]. However, statistics related to randomness due to the mobility of subnetworks and wireless channel statistics lead to rather complicated mathematical derivations which usually only work for some limited cases. One instance of these methods is [7] in which the conventional weighted moving average-based is used as an interference estimator. Due to its simplicity and low computational requirement, we have chosen the same moving average estimator as the baseline technique.

Artificial intelligence provides an increasing ability to deal with complex data, thus becoming a key enabler in 6G [1]. Machine learning, a branch of artificial intelligence, helps in the estimation of interference from complex unknown and potentially non-linear real-time data efficiently. In [8], regression-based up-link interference identification is proposed, in which a comparison is shown between linear and non-linear least-squares regression algorithms along with multi-layer perception. In [5], the authors used AI-powered interference estimation using an non-linear Autoregressive Neural Network (NARNN). There, faster convergence is achieved by using Levenberg-Marquadt backpropagation (LMBP). However, backpropagation requires a lot of computations, especially for deep neural networks, which can make training of the networks computationally infeasible [9]. Furthermore, it is necessary to consider existing ML problems like vanishing and exploding gradient problems while selecting existing ML based approaches as well. In this research, we have considered the LSTM method, which consists of a set of gates to control when information enters the memory that also solves the aforementioned problems, i.e., vanishing and exploding gradients [10].

A. Contributions

The contributions of this research work are summarized here:

Contribution 1: The LSTM model is introduced for the estimation of interference in individual subnetworks, which can capture complex statistical dependencies that may exist in complex unknown and potentially non-linear dynamic interference patterns collected at the subnetworks. The results illustrate that the use of LSTM helps achieve a low estimation error by addressing problems like vanishing and exploding gradients.

Contribution 2: The proposed LSTM-based interference estimation achieves low estimation error at the cost of higher convergence time. We introduce FL based (FedAVG-In-X) interference estimation algorithm which helps to achieve fast convergence speed and low estimation error as compared to the LSTM-based estimation technique deployed in a subnetwork.

B. Outline

The paper is structured as follows: the system model is introduced in Section II. The base and proposed interference estimation model is explained in Section III. In Section IV, the results are discussed. The conclusion of the proposal is presented in Section V.

Notation: Bold small and capital letters denote vectors and matrices, respectively. $|.|$ denotes the cardinality of a set. Sets are calligraphic.

II. System Model

A. Subnetwork Deployment

Assume a set of $\mathcal{N} = \{n_1, \ldots, n_N\}$ with $|\mathcal{N}| = N$ subnetworks have been deployed in a factory within an area $A \subset \mathbb{R}^2$ and moving with velocity v m/s in a random direction β. One gNB/local interactive device, e.g., an access point (AP) or a drone in the outdoors, provides coverage and data exchange for all N subnetworks. Each subnetwork consists of a set of $\mathcal{M}_{n_q} = \{m_1, \ldots, m_M\}$ with $|\mathcal{M}_{n_q}| = M$, $\forall\ n_q \in \mathcal{N}$ sensor-actuator pairs. Fig. 1 shows the deployment and the working principle of the subnetwork. For interference modeling, it is assumed that downlink (DL) from the controller to the actuator and uplink (UL) from the sensor to the control unit. The control unit of the subnetwork is placed in the center as an example. The positions of the subnetworks are given as $\mathbf{P} = [\mathbf{p}_{n_1}, \ldots, \mathbf{p}_{n_N}]$. Assume that $b\left(\mathbf{p}_{n_q}, r\right)$ is a disc with radius r centered at position $\mathbf{p}_{n_q}$, illustrating the area occupied by the n_qth subnetwork. To place a new subnetwork, the admissible area for the n_q th subnetwork consists of the original area A excluding the occupied area of all $(q-1)$ subnetworks, that is, $\tilde{A} \subset A \setminus \bigcup_{k=1}^{q-1} b(\mathbf{p}_{n_k}, r)$ [11]. Furthermore, we also assume that $\tilde{A}$ provides extra space to avoid overlapping of subnetworks. Then, the conditional pdf for the center of the n_qth subnetwork can be given as [11]

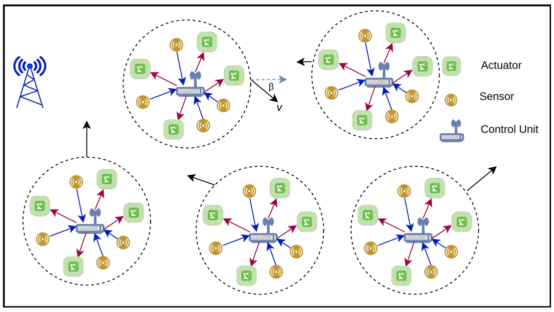

Fig. 1. Subnetwork with sensor-actuator pair (not to scale)

$$f(\mathbf{p}_{n_q}|\mathbf{p}_{n_1}, \ldots, \mathbf{p}_{n_{q-1}}) = \begin{cases} \dfrac{1}{|\tilde{A}|} & for\ \mathbf{p}_{n_q} \in \tilde{A} \\ 0 & else \end{cases} \tag{1}$$

where $|\tilde{A}|$ is Lebesgue measure of $\tilde{A}$. In the factory, the number of sensor-actuator pairs and control units can be assumed to be fixed in area A. The position of the sensor / actuator m_a of the n_q subnetwork is therefore $\boldsymbol{\phi}_{n_q m_a} = \mathbf{p}_{n_q} + \boldsymbol{\xi}_{n_q m_a}$. So, the location of all sensor-actuator pairs in subnetwork n_q can be given as $\boldsymbol{\Xi}_{n_q} = \{\boldsymbol{\xi}_{n_q m_1}, \boldsymbol{\xi}_{n_q m_2}, \ldots, \boldsymbol{\xi}_{n_q m_M}\}$ which is distributed as binomial point process over a disc $b\ (0, r)$ for $|\mathcal{M}_{n_q}|$ sensor-actuator pairs as the uniform pdf of $\frac{1}{\pi r^2}$.

B. Interference Modeling

We focus on the interference experienced by subnetwork n_q during uplink in a high-density subnetwork scenario by all other active links using the same frequency band in the area A. To ensure extreme communication requirements *within* a subnetwork, orthogonality of communication links should be guaranteed [4]. We consider a network with N subnetworks and N_{ch} frequency bands partitioned from a total bandwidth, B. Furthermore, we assume time division duplexing (TDD) for intra-subnetwork communication. Each subband is split into M orthogonal OFDM subcarriers, where each subcarrier is occupied by only one actuator and one sensor for downlink (DL) and uplink (UL) transmission. This ensures there is no intra-subnetwork interference. Each subnetwork uses a single frequency band out of N_{ch}. We assume the number of subbands $N_{ch} < N$. Therefore, multiple subnetworks will interfere in a single subband. In the following, interference analysis is carried out for the n_qth subnetwork among $\mathcal{N}$ subnetworks during the uplink. Interference is caused by subnetworks $\mathcal{K} \setminus \{n_q\}$, where $\mathcal{K} \subset \mathcal{N}$ is a set of subnetworks utilizing the same subband. As a result, there exists dynamic inter-subnetwork interference due to the subband sharing by more than one moving subnetworks.

To describe the interference at the n_qth subnetwork we first define a set of communication links except those belonging to itself as:

$$\mathcal{C}_{n_q} = \underbrace{\left\{ \bigcup_{\substack{j=1 \\ j \neq q}}^{N-1} \mathcal{M}_{n_j} \right\}}_{I} \bigcup \underbrace{\left\{ \mathcal{N} \setminus \{n_q\} \right\}}_{II} \tag{2}$$

where the I is for all sensor-actuator pairs, the II is for all control units.

With the set $\mathcal{C}_{n_q}$ we can describe the interference using an appropriate channel model. Eq 3 sums over all contributing links, where $h_{c_e}(t)\ \forall\ c_e \in \mathcal{C}_{n_q}$ captures small-scale fading. The path loss is $l_{c_e}(t) = min(1,|\boldsymbol{\phi}_{c_e}(t) - \mathbf{p}_{n_q}(t)|^{-\sigma})$ with the path loss exponent σ[12] and $\boldsymbol{\phi}_{c_e}$ represents the position of each communication link from $\boldsymbol{\phi}_{n_q m_a}$. Small-scale fading is modeled as a circular symmetric Gaussian distributed with $\mathbb{E}[|h_{c_e}(t)|^2] = 1$ [13]. The symbol $\zeta_{c_e}(t)$ denotes shadowing based on spatially correlated Gaussian random fields model with a decorrelation distance of δ, which is [14]. If we assume unit transmit power for any time t, we obtain the interference power at control unit $\boldsymbol{p}_{n_q}$ as

$$I_{\boldsymbol{p}_{n_q}}(t) = \sum_{c_e \in \mathcal{C}_{n_q}} |h_{c_e}(t)|^2 \cdot l_{c_e}(t) \cdot \gamma_{c_e}(t) \cdot \zeta_{c_e}(t)\,, \tag{3}$$

The symbol $\gamma_{c_e}(t)$ is a Bernoulli random variable indicating 1 when the link is active and 0 when there is no transmission. For the n_qth subnetwork, we consider S interference powers, in successive time instants collected in a vector $\mathbf{I}_{\mathbf{p}_{n_q}} = [I_{\mathbf{p}_{n_q}}(0), \ldots, I_{\mathbf{p}_{n_q}}(S)]$. This interference power vector $\mathbf{I}_{\mathbf{p}_{n_q}}$ can be seen as a stationary time series from which we can capture the long-term non-linear relationships.

III. Interference Estimation

In this section, we discuss the baseline interference estimation technique, i.e., a moving average estimator, and propose centralized and decentralized machine learning-based estimation techniques.

A. Moving Average Estimation

The baseline method is the weighted average-based interference estimator, which is used in link adaptation for traditional enhanced mobile broadband (eMBB) services. The interference measured at time t is filtered with a low-pass first-order infinite impulse response (IIR) filter to obtain the interference [15]:

$$\hat{I}_{\mathbf{p}_{n_q}}(t+1) = \alpha I_{\mathbf{p}_{n_q}}(t) + (1-\alpha)\hat{I}_{\mathbf{p}_{n_q}}(t) \tag{4}$$

where $I_{\mathbf{p}_{n_q}}$ is the true interference power measured at previous time step, and $\hat{I}_{\mathbf{p}_{n_q}}$ is the estimated interference of the n_qth subnetwork and $0 < \alpha < 1$ is the forgetting factor of the filter.

B. LSTM-based Interference Estimation

An LSTM is a type of recurrent neural network that is able to capture long-term dependencies in sequential data [16]. The interference power vector $\mathbf{I}_{p_{n_q}}$ is considered as the input for the LSTM model. LSTM exploits the complex and potentially non-linear relationship contained in the interference power vector. Furthermore, it can learn to store and retrieve information from long-term dependencies, allowing to make better predictions than other types of recurrent neural networks. In interference prediction, we are utilizing it to predict the interference power at time $t+1$ based on the series of interference power vectors till time t, i.e. $[I_{\mathbf{p}_{n_q}}(0), I_{\mathbf{p}_{n_q}}(1), \ldots, I_{\mathbf{p}_{n_q}}(t)]$. LSTMs are composed of cells that can remember or forget information, as well as input, output, and forget gates that decide which information should be retained and which should be discarded over time. It allows for better handling of long-term dependencies and also solves the vanishing or exploding gradient problem [10]. A detailed explanation of LSTM can be found in [16]. The data consisting of interference power of size S collected at the n_qth subnetwork is split into training (90%) and test data (10%). The training data is transformed and restructured with a moving window of length S_w and a moving step size s_m such that it can feed into LSTM. After training, the model is tested and analyzed in terms of root mean square error and Mean absolute percentage error. The evaluation matrices are presented and analyzed in Section IV. This interference estimation technique can be deployed in subnetworks without connection to the surrounding 6G networks.

C. Cooperative Interference Estimation Using LSTM Based Federated Learning

Up to this point, we have considered interference in a single subnetwork. However, it is clear that the interference power of different subnetworks might be correlated and therefore might improve estimation performance. Let us consider two subnetworks like n_q , $n_u \in \mathcal{K}$ which are facing interference from $\mathcal{K} \setminus \{n_q\}$ and $\mathcal{K} \setminus \{n_u\}$ respectively. There exists a cross-correlation between interference power vectors as both subnetworks account for the same interferers $\mathcal{K} \setminus \{n_q, n_u\}$. We can infer that the subnetworks in $\mathcal{K}$ have higher importance than the other subnetworks $\mathcal{N} \setminus \mathcal{K}$ as they are likely to cross-correlated with the result of small-scale fading. As the subchannels on different bands are likely to be uncorrelated, i.e., accounting subnetwork $\mathcal{N} \setminus \mathcal{K}$ in cooperative training may result in marginal gain only. As a result, we are interested to analyze the performance by considering $n_q \in \mathcal{K}$ subnetworks. Moreover, knowledge of the spatiotemporal distribution in the Gaussian random field of the area A helps to learn about the shadowing effect. Constraining data just to $\mathcal{K}$, though, might still limit the observation of global effects like shadowing. Therefore, it is necessary to analyze considering all subnetworks $n_q \in \mathcal{N}$ in training and examine the performance of the proposed estimator. The analysis of these two cases helps to decide on a computationally efficient approach, e.g. reducing the number of participating subnetworks in training, but also gives intuition for the selection of participants among all subnetworks, which is numerically analyzed in Section IV-C. Hence, cooperative estimation trained with local data and empowered by global knowledge from other subnetworks' learning may prove to be useful for estimation with lower error and higher convergence speed.

Federated learning is a distributed machine learning method that allows multiple devices, such as smartphones or IoT devices, to train a machine learning model collaboratively [17]. In federated learning, each device trains a local model on its own local data, and the global model is then updated by aggregating the updates from the local models. One of the main benefits of federated learning is that it allows the training of machine learning models on large amounts of decentralized data without the need to centralize the data on a single server. We introduce the LSTM-based FL estimator (FedAVG-in-X) with all $\mathcal{N}$ and interfering $\mathcal{K}$ clients in this research work, which is explained in Algorithm 1, and the following.

Algorithm 1 FedAVG-in-X Interference Estimation Algorithm

Collect interference power vector for all $n_k \in \mathcal{K}$ as $\mathbf{I}_{\mathbf{p}_{n_k}}$ // replace $\mathcal{K}$ by $\mathcal{N}$ if all subnetworks participates in training.

gNB/Server Executes:
Initialize model $\theta_0^{n_k}$ with Xavier Initialization
for each round $p = 1,2,\dots$ **do**
  for each subnetwork $n_k \in \mathcal{K}$ in parallel **do**
    $\theta_{p+1}^{n_k} \leftarrow$ SubnetworkUpdate(n_k, θ_p, $\mathbf{I}_{\mathbf{p}_{n_k}}$)
    $\theta_{p+1}^{n_k} \leftarrow \sum_{n_k \in \mathcal{K}} \frac{|\mathbf{I}_{\mathbf{p}_{n_k}}|}{\sum_{n_k \in \mathcal{K}} |\mathbf{I}_{\mathbf{p}_{n_k}}|} \theta_{p+1}^{n_k}$
  end for
end for

Subnetwork/Client Executes:
Interference power vector $\mathbf{I}_{\mathbf{p}_{n_k}}$
SubnetworkUpdate(n_k, θ_p, $\mathbf{I}_{\mathbf{p}_{n_k}}$)
for each local epoch i from 1 to E **do**
  $\theta_p^{n_k} \leftarrow$ trained model of LSTM model with input $\mathbf{I}_{\mathbf{p}_{n_k}}$
end for
Predicted interference $\hat{I}_{\mathbf{p}_{n_k}}[S_w + 1] \leftarrow$ SubnetworkUpdate(n_k, θ_p, $\mathbf{I}_{\mathbf{p}_{n_k}[1:S_w]}$)
return $\theta_p^{n_k}$

In our work, we have considered the interference power vector $\mathbf{I}_{\mathbf{p}_{n_k}}$ $\forall\ n_k \in \mathcal{N}$ as a local data set, which is disjoint, balanced, and non-i.i.d. and trained the LSTM model as explained in Section III-B. The updated model of each subnetwork $n_k \in \mathcal{K}$ is uploaded to the cellular infrastructure that aggregates the models and sends the aggregated model back to the subnetworks. This process is carried out until convergence as shown in Fig. 2.

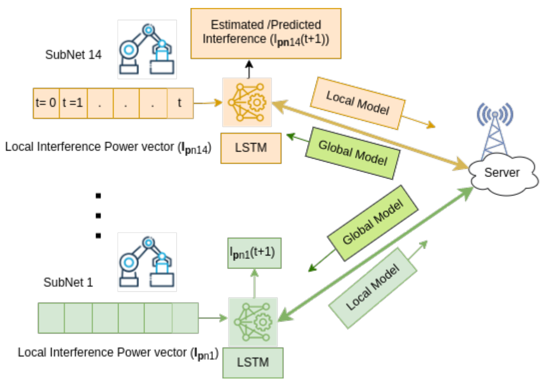

Fig. 2. Interference Estimation using FL for subnetwork. The index of subnetwork is considered based on using same channel during simulation

The machine learning model is trained by the subnetworks collaboratively to solve the following optimization problem [17]:

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^a} \sum_{n_k \in \mathcal{K}} \frac{|\mathbf{I}_{\mathbf{p}_{n_k}}|}{\sum_{n_k \in \mathcal{K}} |\mathbf{I}_{\mathbf{p}_{n_k}}|} \sum_{i=1}^{|\mathbf{I}_{\mathbf{p}_{n_k}}|} \frac{1}{|\mathbf{I}_{\mathbf{p}_{n_k}}|} f(\mathbf{I}_{\mathbf{p}_{n_k}}[i]; \boldsymbol{\theta}) \tag{5}$$

where $\mathcal{K} \subset \mathcal{N}$. The $|.|$ denotes the length of vector. The objective $\frac{1}{|\mathbf{I}_{\mathbf{p}_{n_k}}|} f(\mathbf{I}_{\mathbf{p}_{n_k}}[i]; \boldsymbol{\theta})$ is an empirical loss function defined by the training task, and $f(\mathbf{I}_{\mathbf{p}_{n_k}}[i]; \boldsymbol{\theta})$ is the training loss for the data point $i \in \{1, \dots, |\mathbf{I}_{\mathbf{p}_{n_k}}|\}$ and the model parameter $\boldsymbol{\theta}$ with dimension a.

In this research, we have considered federated averaging (FedAVG) which incorporate E local epochs of training in each subnetwork. However, this Fed-in-X approach will only be applicable for subnetworks connected to existing cellular networks.

TABLE I
SIMULATION SETTING FOR INTERFERENCE ANALYSIS AND ESTIMATION

Parameter	Value
Deployment Parameter	
Number of subnetworks , $\lvert\mathcal{N}\rvert$	16
Number of sensor-actuator pair, $\lvert\mathcal{M}\rvert$	1
Interfering subnetworks, $\lvert\mathcal{K}\rvert$	4
Deployment density~(subnetwork/km^2)	40000
Mobility Parameter	
Mobility Model	RDMM
Cell Radius, r	2 [m]
Velocity, v	2 [m/s]
Minimum distance, d	3 [m]
Channel Parameter	
Path Loss Exponent, σ	2.55
Shadow fading standard deviation	5.7 [dB]
Decorrelation distance, δ	5 [m]
Carrier frequency, f_c	6 [GHz]
Number of available Channel, N_{ch}	4
LSTM Parameters	
Activation Function	tanh
Loss Function	Mean Square Error (MSE)
Learning Rate	0.01
Number of layer and hidden units per layer	1 and 64
Optimizer	Adaptive moments (Adam)
Weight Initialization	Xavier

IV. Simulation Setup and Analysis

A. Simulation Setup

We have assumed that 16 subnetworks are deployed in a region of $20 \times 20\ m^2$ and each subnetwork is moving with a speed of $2\ m/s$ in a random direction with $\beta \sim U(0,2\pi)$. The subnetwork follows the random direction mobility model (RDMM) [18]. Each subnetwork consists of one active sensor-actuator pair randomly located within radius r from the center of the subnetwork. The next position of subnetwork at time t moving with speed v can be given as $[x(t-1) + vt\cos(\beta),\ y(t-1) + vt\sin(\beta)]$. It has been assumed that if two subnetworks are about to collide, i.e., the minimum distance between the center of two subnetworks is less than d, the subnetwork changes the direction. The interference power at time t is calculated by equation (3) using the channel parameters mentioned in Table I. The channel parameters are considered based on the technical report produced by 3GPP [19]. The time series data associated with position at time $t = 0, 1, \dots, S$ is stored as interference vector $\mathbf{I}_{\mathbf{p}_{n_q}} = [I_{n_q}(0), I_{n_q}(1), \dots, I_{n_q}(S)]\ \forall\ n \in 1, \dots, N$. Based on this local dataset, interference estimation is performed.

B. Evaluation Metrics

1) Estimation Error: To analyze the performance of interference estimation, we have considered two performance metrics, namely root mean square error (RMSE) and mean absolute percentage error (MAPE). Root mean square error (RMSE) is given by

$$RMSE = \sqrt{\frac{1}{T_{test}} \sum_{t=1}^{T_{test}} (\hat{I}_{n_q}(t) - I_{n_q}(t))^2}. \quad (6)$$

MAPE is defined as

$$MAPE = \frac{100\%}{T_{test}} \sum_{t=1}^{T_{test}} \left| \frac{I_{n_q}(t) - \hat{I}_{n_q}(t)}{I_{n_q}(t)} \right| \quad (7)$$

where $\hat{I}_{n_q}(t)$ is the predicted interference power and $I_{n_q}(t)$ is the true interference value for the t^{th} time instance of n_qth subnetwork with T_{test} samples.

2) Convergence Analysis: Convergence analysis is the process of evaluating whether a machine learning model has reached a stable state, where the model's performance on the training data is no longer improving with additional training. There are several ways for convergence analysis. The convergence is analyzed in a similar way in terms of training loss as mentioned in [17], [20].

C. Results and Analysis

The moving average estimator predicts the interference value at time $t+1$ based on (4) with the forgetting factor $\alpha = 0.01$. For comparison with LSTM, we have stored the predicted interference power value with a moving average-based estimator on the estimated interference power vector. Performance is analyzed using performance metrics like RMSE and MAPE from true and estimated interference power vectors.

In the case of LSTM based method with a univariate interference power vector, it is necessary to restructure the data into multiple inputs and outputs to create a local dataset in the subnetwork. The interference power vector of size $S = 10000$ is divided into 90% train and 10% test data. Those train and test data are restructured with a window size of $S_w = 20$ and a moving step of size $s_m = 1$. Utilizing these sequences of interference power values from train data, LSTM extracts the existing complex and potentially non-linear relationships. The details of the LSTM parameters are mentioned in Table I. The LSTM model is trained for 150 epochs in the nth subnetwork and the performance is compared with the estimated interference power vector using a moving average estimator.

Performance metrics are calculated for the entire test data. Fig. 3 illustrates the comparison of RMSE in the n_qth subnetwork trained only with the interference it captures, and another one with data of all the interference captured by all the N subnetworks. It shows that the performance is better when the n_qth subnetwork is trained by only its captured interference compared to the case it is trained by also all other subnetworks' captured interference. It takes into account only the correlated time series observed by itself. Although both converge, convergence with all subnetworks' interference power vectors together using a comparable size of LSTM (n-to-n LSTM) for training is much slower. This encourages us to empower training in a decentralized way without sharing the

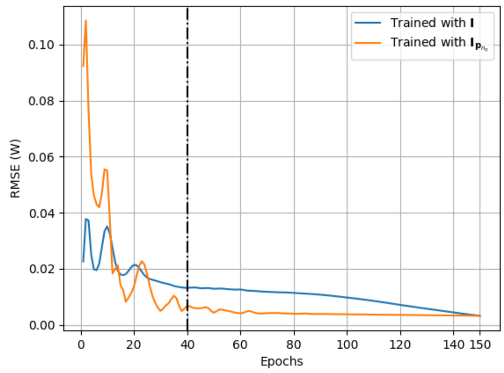

Fig. 3. RMSE of the interference prediction for the n_qth subnetwork LSTM model trained with $\mathbf{I}_{\mathbf{p}_{n_q}}$ on interfered by $\mathcal{K} \setminus \{n_q\}$ compared to $\mathbf{I} = [\mathbf{I}_{\mathbf{p}_{n_1}}, \dots, \mathbf{I}_{\mathbf{p}_{n_N}}]$.

interference power vector itself. This reduces the effort used for collecting and sharing interference data in the n_qth subnetwork. The RMSE and MAPE are approximately $0.0048\ W$ and 10%

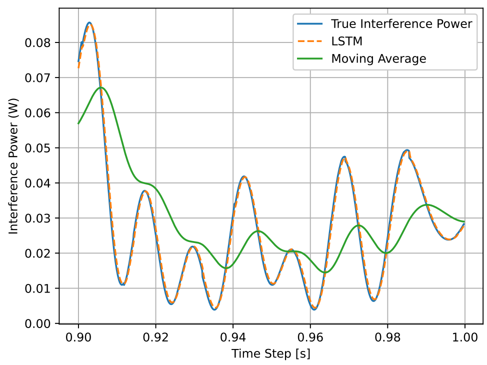

Fig. 4. Prediction of interference for a fixed time step (test data) after training 150 epochs.

for the n_qth subnetwork trained with local data after training 40 epochs. After training 150 epochs, the RMSE and MAPE are $0.0032\ W$ and 5.804%. For 1000 time steps (test data), the RMSE and MAPE for the moving average-based estimator are $0.016\ W$ and 69.174%.

Fig. 4 shows the prediction based on the proposed LSTM-based estimator and the moving average-based estimator. It can be observed that the LSTM model performs better than the moving average-based estimator. We are not considering a moving average-based estimator for further analysis since an LSTM-based predictor is superior.

We consider two cases with $\mathcal{K}$ and $\mathcal{N}$ are considered participants in FL training. The local model of the participating subnetwork is trained by local data. These subnetworks send the model to the server. The server then aggregates them and sends the aggregated model which will be utilized by the subnetworks. The FL model is repeated for 150 communication rounds. Based on simulation results, it took 40 epochs to reach the convergence region for the LSTM model, while FedAVG took less than 10 and 15 local epochs (i.e., approx. $p = 2$ and $p = 3$ communication rounds with $E = 5$ as mentioned in Algorithm 1) for $\mathcal{K} = 4$ and $\mathcal{N} = 16$ participating clients, respectively. It shows that cooperative interference estimation helps to speed up convergence more than 2 times. Furthermore, it is necessary to analyze the error performance as well.

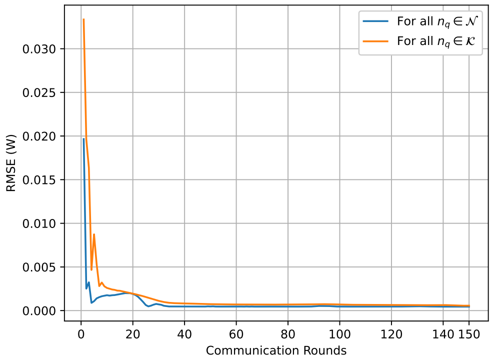

Fig. 5. RMSE comparison of test data for all subnetworks participating in training $\mathcal{N}$ and only subnetworks operating in the same channel $\mathcal{K}$ for FedAVG.

The RMSE analysis shown in Fig. 5 illustrates the RMSE of FedAVG with participating $\mathcal{N}$ and $\mathcal{K}$ subnetworks. The results show that considering $\mathcal{N}$ subnetworks instead of all $\mathcal{K}$ achieves the performance with marginal difference (approx. $0.7 - 1\%$) in terms of error metrics. This result illustrates the influence of the participating $\mathcal{K}$ instead of $\mathcal{N}$ subnetworks on the estimation error is marginal, which is computationally efficient. The RMSE and MAPE achieved at 150th epoch for LSTM-based estimator are $0.0032\ W$ and 5.86% approximately comparable to 7th communication round (35 local epochs) with $0.0031\ W$ and 4.86% in case of FL-based estimator with participating $\mathcal{K}$ clients. This result illustrates the intuitive performance comparison of the FL-based estimator and LSTM-based estimator at individual subnetworks. Furthermore, it shows that the FL-based estimator outperforms in terms of both convergence and prediction error. The reason behind the better performance of FedAVG is that it trains a few epochs at the subnetwork which helps to learn more about a local dataset and the global knowledge from the aggregated model helps learn the full statistics of interfering subnetworks as well. From the results, we have observed that at the 30th round, a MAPE of $\approx 2\%$ and an RMSE of $0.000566\ W$ have been achieved. The MAPE in the 150th communication round is 0.9%.

V. Conclusion

In this paper, we have used an LSTM-based interference estimation technique using only the interference power vector available at each subnetwork. ML-based interference estimation outperforms the conventional moving average estimation method with a high margin in terms of RMSE and MAPE. Already a purely local LSTM model results in low MAPE and RMSE with slow convergence speed. Furthermore, we have introduced the concept of cooperative interference estimation for subnetworks operating on the same channel without losing much compared to including all participating subnetworks. This model with federated learning (FedAVG-in-X) shows a high convergence speed compared to the model without federated learning. It shows that such a proactive cooperative interference estimation technique can be a potential proactive approach for estimation in an application like an in-X subnetwork which needs to work in a fraction of ms latency and extremely high reliability.

References

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, “Toward 6G networks: Use cases and technologies,” *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[2] C. Bernardos and et al., “European vision for the 6G network ecosystem,” Tech. Rep., 06 2021.

[3] G. Berardinelli, P. Baracca, R. O. Adeogun, S. R. Khosravirad, F. Schaich, K. Upadhya, D. Li, T. Tao, H. Viswanathan, and P. Mogensen, “Extreme communication in 6G: Vision and challenges for ‘in-X’subnetworks,” *IEEE Open Journal of the Communications Society*, vol. 2, pp. 2516–2535, 2021.

[4] X. Du, T. Wang, Q. Feng, C. Ye, T. Tao, L. Wang, Y. Shi, and M. Chen, “Multi-agent reinforcement learning for dynamic resource management in 6G in-X subnetworks,” *IEEE Transactions on Wireless Communications*, 2022.

[5] C. Padilla, R. Hashemi, N. H. Mahmood, and M. Latva-Aho, “A nonlinear autoregressive neural network for interference prediction and resource allocation in URLLC scenarios,” in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2021, pp. 184–189.

[6] M. Bennis, M. Debbah, and H. V. Poor, “Ultrareliable and low-latency wireless communication: Tail, risk, and scale,” *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[7] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, “MAC layer enhancements for ultra-reliable low-latency communications in cellular networks,” in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 1005–1010.

[8] Y. Guo, C. Hu, T. Peng, H. Wang, and X. Guo, “Regression-based uplink interference identification and SINR prediction for 5G ultra-dense network,” in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[9] A. Suliman and Y. Zhang, “A review on back-propagation neural networks in the application of remote sensing image classification,” *Journal of Earth Science and Engineering*, vol. 5, no. 1, 2015.

[10] N. K. Manaswi, *RNN and LSTM*. Berkeley, CA: Apress, 2018.

[11] L. Menholt and L. V. Friis, “Mathematical modelling and prediction of interference power in in-robot subnetworks,” Master’s thesis, Mathematical Engineering-Aalborg University, 2022.

[12] J. F. Schmidt, U. Schilcher, M. K. Atiq, and C. Bettstetter, “Interference prediction in wireless networks: Stochastic geometry meets recursive filtering,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2783–2793, 2021.

[13] P. Dent, G. E. Bottomley, and T. Croft, “Jakes fading model revisited,” *Electronics letters*, vol. 13, no. 29, pp. 1162–1163, 1993.

[14] S. Lu, J. May, and R. J. Haines, “Effects of correlated shadowing modeling on performance evaluation of wireless sensor networks,” in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*. IEEE, 2015, pp. 1–5.

[15] N. H. Mahmood, O. A. Lopez, H. Alves, and M. Latva-Aho, “A predictive interference management algorithm for URLLC in beyond 5G networks,” *IEEE Communications Letters*, vol. 25, no. 3, pp. 995–999, 2020.

[16] H. Sak, A. Senior, and F. Beaufays, “Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition,” *arXiv preprint arXiv:1402.1128*, 2014.

[17] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[18] M. Liu, Y. Wan, and F. L. Lewis, “Analysis of the random direction mobility model with a sense-and-avoid protocol,” in *2017 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2017, pp. 1–6.

[19] 3GPP, “5G: Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 16.1.0 release 16),” 2020.

[20] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, “On the convergence of fedavg on non-iid data,” *arXiv preprint arXiv:1907.02189*, 2019.