Al-Quraan, M., Centeno, A. , Zoha, A. , Imran, M. A. and Mohjazi, L. (2023) Federated Learning for Reliable mmWave Systems: Vision-Aided Dynamic Blockages Prediction. In: IEEE Wireless Communications and Networking Conference (WCNC2023), Glasgow, Scotland, UK, 26-29 March 2023, ISBN 9781665491228 (doi: 10.1109/WCNC55385.2023.10118675)

There may be differences between this version and the published version. You are advised to consult the published version if you wish to cite from it.

https://eprints.gla.ac.uk/287650/

Deposited on 12 December 2022

Enlighten – Research publications by members of the University of Glasgow

http://eprints.gla.ac.uk

Federated Learning for Reliable mmWave Systems: Vision-Aided Dynamic Blockages Prediction

Mohammad Al-Quraan, Anthony Centeno, Ahmed Zoha, Muhammad Ali Imran, and Lina Mohjazi
James Watt School of Engineering, University of Glasgow, Glasgow, UK
Email: {m.alquraan.1}@research.gla.ac.uk,
{Anthony.Centeno, Ahmed.Zoha, Muhammad.Imran, Lina.Mohjazi}@glasgow.ac.uk

***Abstract*—Line of sight (LoS) links that use high frequencies are sensitive to blockages, making it challenging to scale future ultra-dense networks (UDN) that capitalise on millimetre wave (mmWave) and potentially terahertz (THz) networks. This paper embraces two novelties; Firstly, it combines machine learning (ML) and computer vision (CV) to enhance the reliability and latency of next-generation wireless networks through proactive identification of blockage scenarios and triggering early handover (HO). Secondly, this study adopts federated learning (FL) to perform decentralised model training so that data privacy is protected, and channel resources are conserved. Our vision-aided proactive HO (PHO) framework localises users using object detection and localisation (ODL) algorithm that feeds a multiple-output neural network (NN) model to predict possible blockages. This involves analysing images captured from the video cameras co-located with the base stations (BSs) in conjunction with wireless parameters to predict future blockages and subsequently trigger PHO. Simulation results show that our approach performs remarkably well in highly dynamic multi-user environments where vehicles move at different speeds, and achieves 93.6% successful PHO. Furthermore, the proposed framework outperforms the reactive-HO methods by a factor of 3.3 in terms of latency while maintaining a high quality of experience (QoE) for the users.**

***Index Terms*—Federated Learning, computer vision, blockage prediction, ultra-dense networks, network latency.**

I. INTRODUCTION

Next-generation wireless networks undergo a substantial design change when operating in high-frequency bands [1]. Obtaining high data rate services from millimeter wave (mmWave) and terahertz (THz) technologies demands downscaling the communication system, resulting in a new network paradigm termed ultra-dense networks (UDN) [2]. Moreover, the use of beamforming enhances the received signal strength (RSS) by forming line-of-sight (LoS) beams. Nevertheless, UDNs face critical challenges due to the sensitivity of high-frequency beams to blockages. These signals suffer high penetration loss and attenuation, leading to a high RSS drop each time an obstacle intercepts the LoS communication link. This problem is aggravated in highly dynamic environments where many dynamic objects can cause frequent blockages.

In literature, several techniques are adopted to overcome the connectivity issue. For example, mmWave channel geometry and signal diffraction characteristics have been studied compared with sub-6GHz to predict whether a mmWave LoS connection will be blocked [3], [4]. Other solutions rely on machine learning (ML) and the dual connectivity (DC) to maintain wireless connectivity and meet the required quality of experience (QoE) for users [5], [6]. However, such solutions' limitations vary in practicality, wasting network resources, and most importantly, they do not solve the link blockage problem since switching between links is still reactive.

To best solve this problem, UDNs require a sense of the surrounding environment to move from reactive to proactive blockage measures. The direct view is essential for UDN communications and is equally important to computer vision (CV), where visual information captures only direct visible objects in the scene, helping to proactively detect blocking objects. Therefore, leveraging vision information collected from the served environment is envisioned to aid the operation of the network rather than relying on wireless information alone, which fails to address this dilemma [3]–[6]. Images are rich in detail that can help solve the blockage problem in UDNs. However, this hinges on two main questions; first, is it possible to detect objects in the environment and identify their mobility information? Second, how can wireless users be distinguished from other passive objects in still images?

In [7], depth images and a deep learning (DL) model are used to predict a user's RSS in the next few hundred milliseconds to assist in handover (HO) decisions. Further, [8] exploits red-green-blue (RGB) images to train a ResNet-18 model and then classify the images based on the blockage status. However, the approaches in [7], [8] do not account for the associated latency until completing a successful HO and, therefore, cannot avoid link blockages, which is critical in highly dynamic UDNs. In our previous work [9] we go further by providing a simple scenario study that intelligently detects blockages and performs optimal proactive HO (PHO) considering the latency required to ensure user HO. The multi-user environment is considered in [10]; this work adopts CV and DL to predict whether the beam will be blocked in the next instance. However, the limited-time prediction will most likely not avoid beam blockages.

In light of the above discussion, this paper extends our study in [9] by providing a CV-aided latency-aware PHO solution that targets practical multi-user UDNs. The proposed framework utilises the object detection and localisation (ODL) algorithm in addition to neural network (NN) to accurately predict blockages and the time when the blockage will occur. Moreover, we noticed that the studies adopting ML approaches follow the centralised training mechanism, which raises data privacy concerns and consumes bandwidth resources. There-

fore, we also adopt federated learning (FL) approach to collaboratively train the NN model, protect the privacy of the visual information, and alleviate pressure on radio channels [11]. The following points summarise the main contributions of this paper:

- We formulate the CV-aided blockage prediction problem for multi-user/object UDNs and develop an end-to-end latency-aware framework that takes advantage of the RGB images to proactively predict blockages and perform PHO so that the QoE of users remains as high as possible.
- We consider FL as a distributed learning approach rather than the conventional centralised learning method to train the model locally in each small base station (SBS) where the visual information resides and secure data privacy while relieving the communication overhead.
- Finally, we validate the accuracy of the proposed framework using modern simulation tools. The simulation results underpin the importance of our solution in maintaining seamless connectivity for highly dynamic UDNs.

II. Network Model

This study targets UDNs which are prevalent in smart cities, where the environment is challenging due to numerous mobile users and obstacles. We consider an outdoor mmWave system consisting of one macro base station (BS) and many SBSs, as depicted in Fig. 1. For clarity, this figure shows only two SBSs and part of a street as a small portion of a UDN. Orthogonal frequency division multiplexing (OFDM) with K subcarriers is adopted as the modulation scheme based on 28GHz. Each SBS has a camera that monitors objects within its field of view. Moreover, it has an M-element antenna array that enables beamforming to serve single-antenna mobile users with beams selected from a predefined beam steering codebook $\mathcal{F} = \{\mathbf{f}_i\}_{i=1}^{B}$, where $\mathbf{f}_i \in \mathbb{C}^{M\times 1}$ and B denotes the total number of beams.

The network focus is to determine the optimal beam that achieves the highest RSS at the user side ($\mathbf{f}^\star$). Given this, we define an area of interest (AoI) as the coverage area that achieves the optimal RSS when the users are connected to the corresponding SBS. QoE is the key performance indicator that this study aims to keep as high as possible; therefore, we assume that the network hands over the users from one SBS to another every time they cross the boundaries of AoIs. Moreover, the geometric mmWave channel model is adopted since it captures the geometrical distribution of the environment and is commonly used in practical mmWave systems [12]. Therefore, the downlink received signal at subcarrier k is

$$y_k = \mathbf{h}_k^T \mathbf{f}^\star s_k + n_k, \tag{1}$$

where $\mathbf{h}^T$ is the transpose of the downlink channel, s is the transmitted symbol, and n represents the additive white Gaussian noise (AWGN). In addition, the RSS at the user side can be determined as follows:

$$RSS = \frac{1}{K}\sum_{k=1}^{K} \left|\mathbf{h}_k^T \mathbf{f}^\star\right|^2. \tag{2}$$

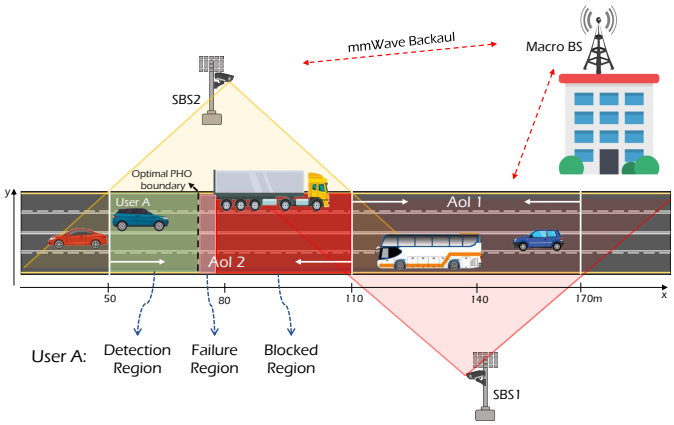


Figure 1: The proposed system model: portion of an UDN including one macro BS and two SBSs each equipped with a vision sensor.

III. Problem Formulation

The beam blockage problem can be formally defined as follows. The camera captures frames of RGB images, and image processing is applied to produce flat[1] RGB (F-RGB) images focused on the AoI. Each F-RGB image is assumed to contain O objects, and every object $o \in O$ will be monitored until it leaves the SBS's AoI. The F-RGBs are fed to an ODL algorithm to obtain boundary-boxes information about every object. This information is then converted to a 6-dimensional metric vector $[x_1, y_1, x_m, y_m, x_2, y_2]$, where the subscripts 1, m, and 2 indicate the upper left, middle, and lower right coordinates of the boundary boxes, respectively. The complete mobility vector ($\mathcal{L}$) of any object o is shaped by adding the it's movement direction (d) and speed (v) as follows $\mathcal{L}_o = [x_{1_o}, y_{1_o}, x_{m_o}, y_{m_o}, x_{2_o}, y_{2_o}, d_o, v_o]_{o=1}^{O}$.

Assuming the number of wireless users in any F-RGB is U, $U \subseteq O$, and a user $u \in U$ is identified from one of all objects (as will be discussed in Section IV-B), then, the $\mathcal{L}$ vector for that user is represented as $\mathcal{L}_u = [x_{m_u}, y_{m_u}, d_u, v_u]$, given that the UE is located in the middle of the object and other objects are possible blockages. Let $S_{u,c}$ represents the combination of the wireless user and a single blocking object, denoted as an obstacle (c), $S_{u,c} = \{\mathcal{L}_u, \mathcal{L}_c\}, u \in U, c \in O\backslash\{u\}$. Therefore, the goal is to classify whether this sample leads to a possible future blockage $b \in \{0, 1\}$, where 0,1 indicate beam non-blockage or blockage, respectively. Moreover, the study predicts the remaining time until the user gets blocked if a link blockage is expected, denoted as T_{BLK}, which could be defined as:

$$T_{BLK} = \begin{cases} i\ , b = 1, \quad \forall i \in \mathbb{R}^+ \\ \text{-1}, b = 0 \end{cases} \tag{3}$$

where -1 means not applicable when the sample $S_{u,c}$ does not lead to a future blockage. Thus, $s_{u,c} = \{b_{u,c}, T_{BLK_{u,c}}\}$ is defined as the labels associated with each data sample $S_{u,c}$.

The objective of this study is achieved by using an ML

[1]Flat term is used to indicate a 2D image that has the same metric width anywhere.

model $f_\Theta(S)$ that can perform classification and regression in parallel. It takes in the user-obstacle vectors S and produces predictions $\hat{s}$. The model predictions are governed by a set of parameters Θ adapted based on dataset of labelled samples $D = \{S_u, s_u\}_{u=1}^{U}$. This dataset trains the ML model to reach high-fidelity for blockage status and time prediction. The following mathematical formulas represent the purpose of the model, which aims to maximise the probability of link status prediction and reduce the blockage time prediction error.

$$\max_{f_\Theta(S)} \prod_{u=1}^{U} \mathbb{P}\left(\hat{b}_{u,c} = b_{u,c} \mid S_{u,c}\right), \quad \forall c \in O \backslash \{u\} \tag{4}$$

$$\min_{f_\Theta(S)} \sum_{u=1}^{U} (|\hat{T}_{BLK_{u,c}} - T_{BLK_{u,c}}|), \quad \forall c \in O \backslash \{u\} \tag{5}$$

IV. CV-Assisted Dynamic Blockage Prediction and PHO

A. Key Idea and Schematic Diagram

This study focuses on a practical scenario that considers multiple dynamic users and objects and extends our previous work [9] that considers a single user and a stationary blocking object. The framework is divided into several subtasks, as illustrated in the schematic diagram in Fig. 2. Initially, the camera at each SBS captures sequences of time-tagged RGB images that are processed to zoom in on the respective AoI. Then, one of the leading-edge ODL algorithms is used to recognise objects and extract the required augmented information. Next, it is necessary to differentiate the wireless users from other obstacles to form $S_{u,c}$ data samples. At this point, the data samples are labelled[2] by blockage status and time. The complete dataset is then stored to train the multi-output model using FL, and when the model is ready, the unlabelled data samples will be fed directly to the model for inference. If the predicted T_{BLK} is greater than the time required by the proposed framework (T_F), it is highly possible to avoid such blockages by requesting a PHO. The following formulas illustrate the main time parameters of the proposed solution:

$$T_F = T_{ODL} + T_{inf} + T_{PHO}, \tag{6}$$

$$T_D \leq T_{BLK} - T_F, \tag{7}$$

where T_{ODL} is the time associated with using the ODL algorithm on two successive F-RGB images. T_{inf} is the model's inference time. T_{PHO} is the time required for performing PHO, and T_D is the time defined to delay triggering the PHO to the point that yields the best QoE.

B. Objects Detection and Users/Obstacles Discrimination

ODL algorithms have recently undergone many advancements allowing for super-fast, real-time, and accurate detection. In this study, a state-of-the-art you only look once (YOLO) version 3 algorithm is adopted to detect various objects in the F-RGB images and produce boundary boxes indicating the positions of the objects in pixel scale [13]. The boundary boxes are then converted to metric scale using the conversion ratio $W_m : W_p$, where W_m and W_p refer to the width of F-RGB images in meters and pixels, respectively. This process is followed by extracting objects' speed and direction to build the $\mathcal{L}$ vector for every object. Performing ODL on two successive F-RGB images is necessary to determine the speed and direction. The direction is determined by noting the displacement in x location, whether to the left or the right. This offset distance is divided by the difference of the corresponding time stamps to get the object's speed. The study in [13] shows that performing ODL on two F-RGB images requires 102ms, i.e. $T_{ODL} = 102ms$. This time will be less if edge computing resources are employed in SBSs.

Identifying wireless users: Moving from a single-user [9] to a multi-user environment necessitates distinguishing each particular user from other objects in the F-RGB image. This study uses a mapping technique in which the exact location of the wireless user in the environment is reflected on the F-RGB images and compared with all boundary boxes. The object with a boundary box centre closest to the user's location will be considered the wireless user in the F-RGB. Several techniques are followed to obtain the user's position in the wireless environment, such as GPS and RSS triangulation, but they fail to provide an accurate location. The shift to higher operating frequency is foreseen to improve the positioning based on the cellular networks [14]. Moreover, several studies have considered this research direction and proposed novel techniques that provide very accurate user localisation [15], [16]. Based on these developments, this work assumes that the radio access network adopts one of these highly accurate methods to provide the location and track the users. Therefore, this study proposes the dynamic positioning table (DPT) in each SBS to keep track of the user's location, which is also converted and reflected on the pixel scale. With DPT tables, it is now possible to differentiate wireless users from other objects and to build the user-obstacle data samples, $S_{u,c}$.

C. Model Training and Inference: FL Approach

The nature of the defined problem is best solved using a model that can do both classification and regression simultaneously. Hence this work develops a multi-output two-hidden layer NN model fed by user-obstacle samples to predict the blockage status and time. In addition, this study adopts FL rather than centralised learning to protect the privacy of the data and relieve the pressure on the communication channels.

Training Phase: During the FL process, the NN model is used as the base model to be trained by the SBSs. The number of clients is set to three. However, the framework can be easily extended to include many SBSs. A parameter server (PS) in the macro BS orchestrates the training process by selecting the number of SBSs participating in each round and sends them the model to start the training. Each SBS exploits its dataset to train the model locally and then sends the model's parameters to the PS for aggregation. Furthermore, an early

[2]Labels of data samples can be obtained analytically in the absence of prior information or by observation, which means monitoring and recording the users blocking status and time.

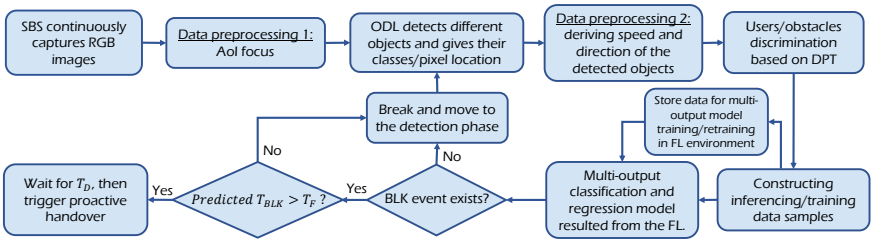


Figure 2: Schematic diagram of the proposed framework.

Table I: Hyperparameters of the NN model.

Parameter	Value
Layers information	Input layer: 10 features 1st Hidden layer: 128 neurons, ReLU 2nd Hidden layer: 64 neurons, ReLU 1st Output (Classifier): 2 neurons, Softmax 2nd Output (Regressor): 1 neuron, Linear
Loss functions	MAE, sparse_categorical_crossentropy
Optimiser	Adam
Learning rate	0.001
Batch size	1000
Data split	70% training - 30% testing

stopping patience technique is developed to avoid suboptimal performance or excessive rounds of unnecessary training.

Inference Phase: After the model has been trained in the FL environment, it is ready to be used for inference. The generated user-obstacle data samples can be fed to the trained model for blockage status and time prediction. Furthermore, the time associated with model inferencing requires approximately 1 ms, i.e., T_{inf} =1ms.

D. Optimal PHO Trigger Point

The proposed framework informs the network of the need for PHO if a blockage is expected. The principal question is when and at what distance the PHO should be triggered. Therefore, this study identifies two regions: detection and failure. The detection region is defined as the region in which the proposed framework monitors and detects the objects within it; in this system model, this region is the same as the AoI. The failure region is where there is no chance of avoiding link interruption due to insufficient time remaining to complete the PHO. The failure region is located just before the blocked region; its width and location vary because of the environment's dynamicity and the user's speed. Fig. 1 illustrates these regions.

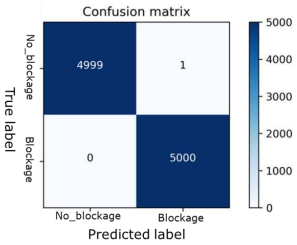
Finding the optimal distance to trigger PHO is linked with maintaining higher values of the user's QoE. If a BLK event is detected, the best scenario is to wait for the maximum delay (T_D^{max}) obtained from equation (7) and then perform PHO. Translating this time to distance gives the optimal distance (D_{opt}) defined as follows:

$$D_{opt} = v \times T_D^{max}. \tag{8}$$

In other words, the optimal PHO must be triggered just before the failure region in the boundary between the detection and failure regions. Triggering the PHO anywhere within the detection region and before the optimal PHO boundary may also avoid the link blockage but at the cost of affecting the perceived QoE. In addition, doing early PHO may impact the balance and the allocation of network resources. Hence, the objective of the proposed framework is to always trigger the PHO within the vicinity of the optimal PHO boundary.

E. PHO Latency

The final required parameter for the proposed framework is T_{PHO}. In a conventional network employing beamforming, if a user loses the connection with the SBS, it will undergo



T_{BLK} categories

Category	Samples	Blockage?
1) Actual $T_{BLK} < T_F$	230	✓
2) Predicted T_{BLK} <= Actual T_{BLK}	2695	×
3) Predicted T_{BLK} > Actual T_{BLK}	2075	✓

Figure 3: Classification and regression model performance.

certain steps to reconnect again. The complete steps are beam failure detection, beam failure recovery, cell search, and contention-based/free random access [10]. According to the 3GPP specifications, each step is associated with a time duration until completion and the total delay time is ∼312.2 ms, indicating the time associated with performing reactive HO when contention-based random access is assumed. Using proactive blockage prediction significantly minimises this time, and the T_{PHO} will be equal to 80ms [9], assuming contention-based radio access since this study targets urban areas with dynamic wireless environments. Consequently, the proposed framework's T_F is now determined and equals 183ms. If the predicted T_{BLK} is greater than 183ms, the framework has a high chance of avoiding link interruption. Otherwise, a link interruption will happen.

V. Performance Evaluation and Results

This section first discusses how the NN model is developed and then delves into describing the simulation setup and discussing the simulation results.

A. FL-based Multi-Output Model Development

The NN model is trained under the FL setup using the federated averaging algorithm. The complete information about the model structure and the selected hyperparameters are shown in Table I. Model performance is tested using ten thousand samples forming 50% blocking and 50% nonblocking. Fig. 3 displays the testing results in which the confusion matrix demonstrates the near-optimal classification accuracy of the model, while the table divides the T_{BLK} into three categories and gives the blockage status for each category.

To give a better understanding, the PHO success rate is defined as $SR_{PHO} = N_s/N_T$, where N_s denotes the number of samples with successful PHO and N_T indicates the total

Table II: PHO success rate versus percent shift.

P_{shift} %	0	1	3	5	7	9	11	13	15
SR_{PHO} %	54	77.1	91	93.4	93.6	93.3	92.8	92.4	92

number of blocking samples. Therefore, we can conclude from the table in Fig. 3 that the success rate is unsatisfactory with 54%. However, this result can be improved by making a trade-off between the SR_{PHO} and the QoE. Therefore, we introduce a new parameter called the percent shift (P_{shift}) defined to reduce the predicted T_{BLK} by this percent. This parameter aims to move as many samples as possible from the third to the second category to enhance the SR_{PHO} at the cost of a slight drop in the QoE. Table II shows how changing the values of the P_{shift} affects the SR_{PHO}, and the best value is 7%. Accordingly, Fig. 4 illustrates the cumulative distribution function (CDF) of the samples with successful PHO versus T_D offset, which indicates how far is the predicted T_D ($\hat{T}_D = \hat{T}_{BLK} - T_F$) from the actual one and defined as:

$$T_{D_{offset}} = \frac{T_D^{max} - \hat{T}_D}{T_D^{max}} \times 100\%, \quad \forall \hat{T}_D \leq T_D^{max}. \tag{9}$$

The optimal PHO point is when the $T_{D_{offset}}$ is zero, and the closer the samples are to this point, the better the performance. Moreover, moving away from this point means an earlier PHO which may affect the QoE. However, performing earlier PHO with some QoE drop is better than losing the connection and establishing it again, which will incur much overhead and increase the network's latency. Finally, the framework is now ready to be used under the considered scenario that will be discussed subsequently.

B. Simulation Setup

The overall performance of the proposed framework is evaluated by considering a practical outdoor environment. The scenario considered in this study is inspired by the vision wireless (ViWi) ASU downtown scenario (ASUDT1) [17], which has a very similar system model to the one adopted for this study. ASUDT1 comprises two mmWave SBSs operating at 28GHz and located 60 m apart on opposite sides of the street. Each has an antenna array that forms LoS beams to serve 60 users moving in straight trajectories. Users are UEs placed in the center of vehicles of different sizes, such as cars, buses, and trucks, moving at different speeds and directions, and each can be seen as a potential obstacle for other users. At each time instance (also known as a scene), the ASUDT1 provides raw data for every user u consisting of a 4-tuple of concurrent information (user location, RGB images, mmWave channel, link status) from each SBS, which helps in evaluating the performance of the proposed solution. The simulation experiments are based on Python programs, and the key performance metrics are the PHO success rate, the network latency, and the perceived QoE.

C. Simulation Results

Several aspects are considered to examine the efficacy of the proposed framework. First, given the dynamicity of the considered environment, the impact of vehicles' speed on performing a successful PHO is studied. The speed is set in a range of 1.5 to 20 mph, and a new parameter called *relative speed* is introduced. For every blocking sample in $S_{u,c}$, the relative speed parameter is defined as the sum of the user's and obstacle's speeds if they are moving towards each other or the difference in their speeds if they are moving in the same direction. This parameter measures how fast a blockage occurs and will be used to investigate its impact on performing successful PHO. Consequently, the relative speed from all blocking samples of the testing dataset is divided into three categories, slow, medium, and fast. Fig. 5(a) demonstrates the results of this study. It can be observed that the PHO success rate is high when the relative speed is low and medium. At the same time, it decreases as the relative speed increases, which is expected due to reducing the T_{BLK} for a blockage, thus reducing the probability of a successful PHO.

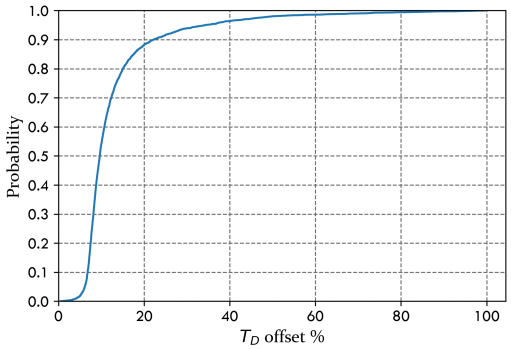


Figure 4: The distribution of the TD offset of the samples with successful PHO.

Then, the latency associated with performing HO on both reactive- and proactive-based approaches is studied. Section IV-E shows that the latency associated with reactive HO is about 312.2 ms. However, the PHO requires only 80 ms, assuming contention-based random access. Following a similar approach from [10], the average HO latency for 5000 users is calculated as:

$$\zeta = \frac{\{\rho \times U\} \times 80 + \{(1-\rho) \times U\} \times 312.2}{U}, \tag{10}$$

where U signifies the total number of users and $\rho \in [0, 1]$ is the percentage of users who successfully perform a PHO. Fig. 5(b) shows the average latency improvement of the proposed PHO framework compared to the reactive mechanism. The average latency is 94.8 ms for our CV-aided PHO solution, which outperforms the reactive HO approach by a factor of 3.3, vital to maintaining connectivity for real-time applications.

The final study adopts a similar approach from [9] of considering moving users running an RTP-based application and measures the average QoE/mean opinion score (MOS) of a group of users with prospect blockages. This study takes advantage of the mmWave channel information provided by the ASUDT1 scenario. Since ASUDT1 provides plentiful information represented as scenes for every location point, this study spotlights on the portion of the street between

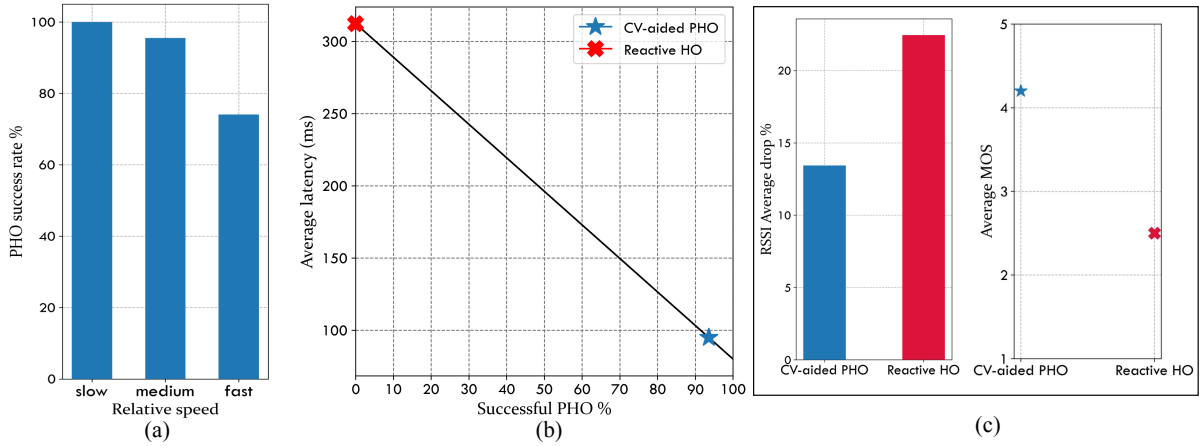


Figure 5: Simulation results: (a) The impact of different relative speeds on the PHO success rate, (b) Comparison of the average latency between the reactive HO and the proposed CV-aided PHO, (c) The RSS percentage drop due to performing reactive and PHO, and how much this drop affects the QoE measured through MOS.

the two SBSs and only focuses on the blockages within the scene interval from 680 to 980. For every blocking, the RSS percentage drop when a user is handed over to another SBS is recorded and is done for all users who encounter blockages in between the two SBSs. Then, the average percentage drop of the RSS is calculated and mapped to the corresponding value of the MOS. However, for the reactive-HO approach, no proactive measures are taken and the percentage drop in RSS is also measured to find the average percentage drop in RSS. Fig. 5(c) illustrates the outcome of this study in which the proposed framework can keep the MOS at a high level despite the small drop in the average RSS. Reactive HO failed to keep users at high MOS during interruption time. This result confirms the potential of the proposed framework for improving the reliability of high-frequency wireless networks and making them suitable for latency-sensitive applications.

VI. Conclusions

This study explored the potential of leveraging vision information to improve the reliability of high-frequency networks by predicting dynamic blockages in advance and taking measures to perform PHO. A NN multi-output model is developed that, combined with CV technology, propose a novel framework capable of accurately predicting blockages and the time needed before the user reaches the blocked region. Moreover, the model is trained using FL to protect data privacy and conserve bandwidth resources. Simulation results indicated that our framework achieves a high PHO success rate of 93.6%, outperforms the reactive-HO approaches by a factor of 3.3 in terms of latency, and maintains the QoE at higher levels. These results highlight a promising solution for beam blockages in multi-user mmWave/THz networks.

References

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surv. Tuts.*, vol. 18, no. 3, pp. 1617–1655, Feb. 2016.

[2] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surv. Tuts.*, vol. 18, pp. 2522–2545, May 2016.

[3] J. Bao, T. Shu, and H. Li, "Handover prediction based on geometry method in mmWave communications-a sensing approach," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops), Kansas City, MO, USA*, May 2018, pp. 1–6.

[4] L. Yu *et al.*, "Long-range blockage prediction based on diffraction fringe characteristics for mmWave communications," *IEEE Commun. Lett.*, Apr. 2022.

[5] K. Qi *et al.*, "Dual connectivity-aided proactive handover and resource reservation for mobile users," *IEEE Access*, vol. 9, pp. 36 100–36 113, Feb. 2021.

[6] C. Wang *et al.*, "Deep learning-based intelligent dual connectivity for mobility management in dense network," in *Proc. IEEE 88th Vehic. Technol. Conf. (VTC-Fall), Chicago, IL, USA*, Aug. 2018, pp. 1–5.

[7] T. Nishio *et al.*, "Proactive received power prediction using machine learning and depth images for mmWave networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2413–2427, Aug. 2019.

[8] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *Proc. IEEE 91st Vehic. Technol. Conf. (VTC2020-Spring), Antwerp, Belgium*, May 2020, pp. 1–5.

[9] M. Al-Quraan *et al.*, "Intelligent blockage prediction and proactive handover for seamless connectivity in vision-aided 5G/6G UDNs," *arXiv preprint arXiv:2203.16419*, Feb. 2022. [Online]. Available: http://arxiv.org/abs/2203.16419

[10] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided 6G wireless communications: Blockage prediction and proactive handoff," *IEEE Trans. Vehic. Technol.*, vol. 70, no. 10, pp. 10 193–10 208, Aug. 2021.

[11] M. Al-Quraan *et al.*, "Edge-native intelligence for 6G communications driven by federated learning: A survey of trends and challenges," *arXiv preprint arXiv:2111.07392*, Nov. 2021. [Online]. Available: http://arxiv.org/abs/2111.07392

[12] Q. C. Li, G. Wu, and T. S. Rappaport, "Channel model for millimeter-wave communications based on geometry statistics," in *Proc. IEEE Globecom Workshops (GC Wkshps), Austin, TX*, 2014, pp. 427–432.

[13] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, Apr. 2018. [Online]. Available: http://arxiv.org/abs/1804.02767

[14] D. Dardari, P. Closas, and P. M. Djurić, "Indoor tracking: Theory, methods, and technologies," *IEEE Trans. Vehic. Technol.*, vol. 64, no. 4, pp. 1263–1278, Feb. 2015.

[15] Rastorgueva-Foi *et al.*, "User positioning in mmW 5G networks using beam-RSRP measurements and Kalman filtering," in *Proc. IEEE 21st Int. Conf. Inf. Fusion (FUSION), Cambridge, UK*, July 2018, pp. 1–7.

[16] M. Koivisto *et al.*, "Joint device positioning and clock synchronization in 5G ultra-dense networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2866–2881, Mar. 2017.

[17] M. Alrabeiah *et al.*, "ViWi: A deep learning dataset framework for vision-aided wireless communications," in *Proc. IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium*, May 2020, pp. 1–5.