Federated Deep Learning for Immersive Virtual Reality over Wireless Networks

Mingzhe Chen*,¶,§, Omid Semiari‡, Walid Saad†, Xuanlin Liu*, and Changchuan Yin*

*Beijing Key Laboratory of Network System Architecture and Convergence,
Beijing University of Posts and Telecommunications, Beijing, China,
Emails: {chenmingzhe,xuanlin.liu,ccyin}@bupt.edu.cn.
¶Department of Electrical Engineering, Princeton University, Princeton, NJ, USA.
§The Future Network of Intelligence Institute, The Chinese University of Hong Kong, Shenzhen, China.
‡Department of Electrical and Computer Engineering, University of Colorado Colorado Springs, Colorado Springs, CO, USA,
Email: osemiari@uccs.edu
†Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA, Email: walids@vt.edu.

Abstract*—In this paper, the problem of enhancing the virtual reality (VR) experience for wireless users is investigated by minimizing the occurrence of breaks in presence (BIPs) that can detach the users from their virtual world. To measure the BIPs for wireless VR users, a novel model that jointly considers the VR applications, transmission delay, VR video quality, and users' awareness of the virtual environment is proposed. In the developed model, the base stations (BSs) transmit VR videos to the wireless VR users using directional transmission links so as to increase the data rate of VR users, thus, reducing the number of BIPs for each user. Therefore, the mobility and orientation of VR users must be considered when minimizing BIPs, since the body movements of a VR user may result in blockage of its wireless link. The BIP problem is formulated as an optimization problem which jointly considers the predictions of users' mobility patterns, orientations, and their BS association. To predict the orientation and mobility patterns of VR users, a distributed learning algorithm based on the machine learning framework of deep echo state networks (ESNs) is proposed. The proposed algorithm uses concept from *federated learning* to enable multiple BSs to locally train their deep ESNs using their collected data and cooperatively build a learning model to predict the entire users' mobility patterns and orientations. Using these predictions, the user association policy that minimizes BIPs is derived. Simulation results demonstrate that the developed algorithm reduces the users' BIPs by up to 16% and 26%, respectively, compared to centralized ESN and deep learning algorithms.

I. Introduction

Deploying virtual reality (VR) applications over wireless networks is an essential stepping stone towards flexible deployment of pervasive VR applications [1]. However, to enable a seamless and immersive wireless VR experience, it is necessary to introduce novel wireless networking solutions that can meet stringent quality-of-service (QoS) requirements of VR applications in terms of delivering high data rates and low latency. In wireless VR networks, the sudden data rate reductions or large delay can negatively impact the users' VR experience (e.g., due to interruptions in VR video streams). Due to such an interruption in the virtual world, VR users will experience *breaks in presence (BIPs)* events that can be detrimental to their immersive VR experience. While the fifth-generation (5G) new radio supports operation at high frequency bands (with abundant bandwidth) as well as flexible frame structure to minimize latency, performance of communication links at high frequencies is highly prone to blockage. That is, if an object blocks the wireless link between the BS and a VR user, the data rate can drop significantly and lead to a BIP. In addition to wireless factors such as delay and data rate, behavioral metrics related to each VR user such as the user's *awareness* can also affect BIPs. Here, awareness is defined as each wireless VR user's perceptions and actions in its individual VR environment. For instance, a user might be too sensitive to slight variations in VR video quality changes, while another user might be more tolerant. Therefore, to minimize the BIPs of VR users, it is necessary to jointly consider all of the wireless environment and user-specific metrics that cause BIPs, such as link blockage, user mobility, user orientation, user association, and user awareness.

Recently, several works have studied a number of problems related to wireless VR networks [2]–[5]. In [2], the authors develop a framework for mobile VR delivery to alleviate the traffic burden over wireless networks. The authors in [3] study the problem of supporting visual and haptic perceptions over wireless cellular networks. The work in [4] proposes a concrete measure for the delay perception of VR users. Our previous works in [5] studied the problems of resource allocation and 360° content transmission for wireless VR users. However, most of these existing works do not provide a comprehensive BIP model that accounts for the transmission delay, the quality of VR videos, VR applications, and user awareness. Moreover, the prior art in [2]–[5] does not jointly consider the impact of the users' body movements when using mmWave communications with highly directional links to support high data rates for VR video transmissions.

To address this challenge, machine learning techniques [6] can be used to predict the users' movements and proactively determine the user associations that can minimize BIPs. However, the existing works for user movement predictions focus on scenarios where users are not mobile, and hence, user association does not change with time. Hence, the data for each VR user's movement can be collected by its associated BS. In contrast, in real mobile VR scenarios, users will move and change their association and the data related to the users' movement is dispersed across multiple BSs. The BSs may

This work was supported in part by the National Natural Science Foundation of China under Grants 61671086, 61629101, and 61871041, in part by Beijing Natural Science Foundation and Municipal Education Committee Joint Funding Project under Grant KZ201911232046, in part by the 111 Project under Grant B17007, in part by the U.S. National Science Foundation under Grants CNS-1836802 and CNS-1941348, and in part by grants No. ZDSYS201707251409055, No. 2017ZT07X152, No. 2018B030338001, and No. 2018YFB1800800.

not be able to transmit all of their collected data on the users' movements to each other, due to the high overhead of data transmission. Moreover, sending all the information to a centralized processing server will cause very large latencies that cannot be tolerated by VR applications. Thus, centralized machine learning algorithms will not be useful to predict real-time movements of the VR users. To this end, *a distributed learning framework that can be trained by the collected data at each BS and cooperatively build a learning model that can predict the entire users' mobility and orientations is needed.*

The key contribution of this work is to develop a novel framework for minimizing BIPs within VR applications that operate over wireless networks. To our best knowledge, *this paper is the first to analyzes how a wireless network with distributed learning can minimize BIP for VR users and enhance their virtual world experience*. For wireless VR users, we mathematically model the BIP that jointly considers VR applications, the delay of VR video and tracking information transmission, VR video quality, and the users' awareness. To minimize the BIP of wireless VR users, we develop a federated echo state network (ESN) learning algorithm that enables BSs to locally train their machine learning algorithms using the data collected from the users' locations and orientations. Then, the BSs can cooperatively build a learning model by sharing their trained models to predict the users' mobility patterns and orientations. Based on these predictions, we perform fundamental analysis to find an efficient user association for each VR user that minimizes the BIPs. Simulation results demonstrate that our proposed algorithm can achieve, respectively, 16% and 26% gains in terms of total BIPs compared to the centralized ESN and deep learning algorithm in [7].

II. System Model and Problem Formulation

Consider a cellular network that consists of a set $\mathcal{B}$ of B BSs that service a set $\mathcal{U}$ of U VR users. In this model, BSs act as VR controllers that can collect the tracking information related to the users' movements via VR sensors and use the collected data to generate the VR videos for their associated users. In particular, the uplink is used to transmit tracking information such as users' locations and orientations from the VR devices to the BSs, while the downlink is used to transmit VR videos from BSs to VR users. For user association, the VR users can associate with different BSs for uplink and downlink data transmissions. We consider practical scenarios when the type of VR application can depend on the location of the user. For example, a given user that works in a lab may use certain VR applications for training or research purposes, while using the VR device for entertainment at home. This information will be used by BSs to predict users' locations and orientations and proactively determine efficient user associations.

A. Transmission Model

We consider both uplink and downlink transmission links between BSs and VR users. The VR users can operate at both mmWave and sub-6 GHz frequencies. The VR videos are transmitted from BSs to VR users over the 28 GHz band. Meanwhile, the tracking information is transmitted from VR devices to their associated BSs over a sub-6 GHz frequency band [8]. This is due to the fact that sub-6 GHz frequencies with limited bandwidth cannot support the large data rates required for VR video transmissions. However, it can provide reliable communications for sending small data sized users' tracking information. Next, we first introduce the transmission of the users' tracking information in the uplink. Then, we specify the VR video transmission via downlink mmWave links.

1) Uplink Transmissions of User Tracking Information: Let (x_{it}, y_{it}) be the Cartesian coordinates for the location of user i at time t and S be the data size of each user's tracking information, including location and orientation. S depends on the VR system. The rate for transmitting the tracking information from VR user i to BS j is:

$$c_{ij}^{\mathrm{UL}}\left(x_{it}, y_{it}\right) = \frac{F^{\mathrm{UL}}}{U_j^{\mathrm{UL}}}\log_2\left(1 + \frac{P_u g_{ij} d_{ij}^{-\beta}\left(x_{it}, y_{it}\right)}{\sum\limits_{k\in\mathcal{U}, k\notin\mathcal{U}_j} P_u g_{kj} d_{kj}^{-\beta}\left(x_{kt}, y_{kt}\right) + \rho^2}\right), \tag{1}$$

where F^{UL} is the total uplink bandwidth of each BS j which is assumed to be equal for all BSs, U_j^{UL} represents the number of VR users associated with BS j over uplink, $\mathcal{U}_j$ is the set of VR users associated with BS j, P_u is the transmit power of each VR user (assumed equal for all users), g_{ij} is the Rayleigh fading channel gain, d_{ij} is the distance between VR user i and BS j at time t, and ρ^2 is the noise power.

2) Downlink VR Video Transmission: In downlink, antenna arrays are deployed at BSs to perform directional beamforming over the mmWave frequency band. For simplicity, a sectored antenna model [9] is used to approximate the actual array beam patterns. This simplified antenna model consists of four parameters: the half-power beamwidth ϕ, the boresight direction θ, the antenna gain of the mainlobe Q, and the antenna gain of the sidelobe q. Let φ_{ij} be the phase from BS j to VR user i. The antenna gain of the transmission link from BS j to user i is:

$$G_{ij} = \left\{ \begin{array}{l} Q, \text{ if } \left|\varphi_{ij} - \theta_j\right| \leqslant \frac{\phi}{2}, \\ q, \text{ if } \left|\varphi_{ij} - \theta_j\right| > \frac{\phi}{2}. \end{array} \right. \tag{2}$$

Since the VR device is located in front of the VR user's head, the mmWave link will be blocked, if the user rotates. Let χ_{it} be the orientation of user i at time t and ϑ be the maximum angle using which BS j can directly transmit VR videos to a user without any human body blockage. ϕ_{ij}' denotes the phase from user i to BS j. For user i, the blockage effect caused by its own body can be given by:

$$b_i\left(\chi_{it}\right) = \left\{ \begin{array}{l} 1, \text{ if } \left|\varphi_{ij}' - \chi_{it}\right| \leqslant \vartheta, \\ 0, \text{ if } \left|\varphi_{ij}' - \chi_{it}\right| > \vartheta. \end{array} \right. \tag{3}$$

We assume that each VR user's body constitutes a single blockage area and n_{ijt} represents the number of VR users located between user i and BS j at time t. If there are no users located between user i and BS j that block the mmWave link, $(b_i\left(\chi_{it}\right) + n_{ij} = 0)$, the communication link between user i and BS j is line-of-sight (LoS). If the mmWave link between user i and BS j is blocked by the user i's own body $(b_i\left(\chi_{it}\right) = 1)$ or blocked by other users located between user i and BS j

($n_{ij} > 0$), the communication link between user i and BS j is said to be non-line-of-sight (NLoS).

Considering path loss and shadowing effects, the path loss for a LoS link and a NLoS link between VR user i and BS j in dB will be given by [9]:

$$h_{ij}^{\rm LoS}\left(x_{it}, y_{it}\right) = L_0 + 10\chi_{\rm LoS}\log\left(d_{ij}\left(x_{it}, y_{it}\right)\right) + \mu_{\sigma_{\rm LoS}}, \quad (4)$$

$$h_{ij}^{\rm NLoS}\left(x_{it}, y_{it}\right) = L_0 + 10\chi_{\rm NLoS}\log\left(d_{ij}\left(x_{it}, y_{it}\right)\right) + \mu_{\sigma_{\rm NLoS}}, \quad (5)$$

where $L_0 = 20\log\left(\frac{d_0 f_c 4\pi}{\nu}\right)$ is the free space path loss. Here, d^0 represents the reference distance, f_c is the carrier frequency and ν is the light speed. $\chi_{\rm LoS}$ and $\chi_{\rm NLoS}$ represent the path loss exponents for the LoS and NLoS links, respectively. $\mu_{\sigma_{\rm LoS}}$ and $\mu_{\sigma_{\rm NLoS}}$ represent Gaussian random variables with zero mean, respectively. $\sigma_{\rm LoS}$ and $\sigma_{\rm NLoS}$ represent the standard deviations for LoS and NLoS links in dB, respectively. The downlink data rate of VR video transmission from BS j to user i is:

$$c_{ij}^{\rm DL}\left(x_{it}, y_{it}, b_i\left(\chi_{it}\right), n_{ij}\right) = \begin{cases} F^{\rm DL}\log_2\left(1 + \dfrac{P_{\rm B} G_{ij}}{10^{h_{ij}^{\rm LoS}/10}\rho^2}\right), \text{if } b_i\left(\chi_{it}\right) + n_{ij} = 0, \\ F^{\rm DL}\log_2\left(1 + \dfrac{P_{\rm B} G_{ij}}{10^{h_{ij}^{\rm NLoS}/10}\rho^2}\right), \text{if } b_i\left(\chi_{it}\right) + n_{ij} > 0, \end{cases} \quad (6)$$

where $F^{\rm DL}$ is the bandwidth allocated to each user and $P_{\rm B}$ is the transmit power of each BS j.

B. Break in Presence Model

In a VR application, the notion of a BIP represents an event that leads VR users to realize that they are in a fictitious, virtual environment, thus ruining their immersive experience. In other words, a BIP event transitions a user from the immersive virtual world to the real world [10]. For wired VR, BIP can be caused by various factors such as hitting the walls/ceiling, loss of tracking with the device, or talking to another person from the real world [10]. For wireless VR, BIP can be also caused by the delay of VR video and tracking information transmission, the quality of the VR videos received by the VR users, and the inaccurate tracking information received by BSs.

To model such BIPs, we jointly consider the delay of VR video and tracking information transmission and the quality of VR videos. We first define a vector $\boldsymbol{l}_{i,t}\left(c_{ij}^{\rm DL}\left(x_{it}, y_{it}, b_i\left(\chi_{it}\right), n_{ij}\right)\right) = \left[l_{i1,t}, \ldots, l_{iN_L,t}\right]$ that represents a VR video that user i received at time t with $l_{ik,t} \in \{0,1\}$. $l_{ik,t} = 0$ indicates that pixel k is not successfully received by user i, and $l_{ik,t} = 1$, otherwise. We also define a vector $\boldsymbol{m}_{i,t}\left(G_A\right) = \left[m_{i1,t}, \ldots, m_{iN_L,t}\right]^{\rm T}$ that represents the weight of the importance of each pixel constructing a VR video, where $m_{ik,t} \in [0,1]$ and G_A represents a VR application such as an immersive VR game or a VR video. $m_{ik,t} = 1$ indicates that pixel k is one of the most important elements for the generation of G_A. Here, in each VR application G_A, a number of pixels can be compressed at the BS and recovered by the user. Hence, the pixels that can be compressed by the BSs not important. However, some of the pixels cannot be compressed by the BS and, hence, they need to transmit to the VR users. Therefore, each pixel will have different importance and $m_{ik,t} \in [0,1]$. Then, the BIP of VR user i caused by the wireless transmission will be given by:

$$P_{it}^{\rm W}\left(x_{it}, y_{it}, \chi_{it}, \boldsymbol{a}_{i,t}^{\rm UL}, \boldsymbol{a}_{i,t}^{\rm DL}\right) = \mathbb{1}_{\left\{\frac{A}{a_{ij,t}^{\rm UL} c_{ij}^{\rm UL}\left(x_{it}, y_{it}\right)} + \frac{D\left(\boldsymbol{l}_{i,t}\left(a_{ik,t}^{\rm DL} c_{ik}^{\rm DL}\left(x_{it}, y_{it}, b_i\left(\chi_{it}\right), n_{ik}\right)\right)\right)}{a_{ik,t}^{\rm DL} c_{ik}^{\rm DL}\left(x_{it}, y_{it}, b_i\left(\chi_{it}\right), n_{ik}\right)} \leqslant \gamma_{\rm D}\right\}} \vee \mathbb{1}_{\left\{\boldsymbol{l}_{i,t}\left(a_{ik,t}^{\rm DL} c_{ik}^{\rm DL}\left(x_{it}, y_{it}, b_i\left(\chi_{it}\right), n_{ik}\right)\right)\boldsymbol{m}_{i,t}\left(G_A\right) \geqslant \gamma_{\rm Q}\right\}}, \quad (7)$$

where $\mathbb{1}_{\{x\}} = 1$ as x is true, $\mathbb{1}_{\{x\}} = 0$, otherwise. $\mathbb{1}_{\{x\}} \vee \mathbb{1}_{\{y\}} = 1$ as y or x is true, $\mathbb{1}_{\{x\}} \vee \mathbb{1}_{\{y\}} = 0$, otherwise. $\boldsymbol{a}_{i,t}^{\rm UL} = \left[a_{i1,t}^{\rm UL}, \ldots, a_{iB,t}^{\rm UL}\right]$ is a vector that represents user i's uplink association with $a_{ik,t}^{\rm UL} \in \{0,1\}$ and $\sum\limits_{k \in \mathcal{B}} a_{ik,t}^{\rm UL} = 1$. Similarly, $\boldsymbol{a}_{i,t}^{\rm DL} = \left[a_{i1,t}^{\rm DL}, \ldots, a_{iB,t}^{\rm DL}\right]$ is a vector that represents user i's downlink association with $a_{ik,t}^{\rm DL} \in \{0,1\}$ and $\sum\limits_{k \in \mathcal{B}} a_{ik,t}^{\rm DL} = 1$. γ_D and γ_Q represent the target delay and video quality requirements, respectively. In (7), $\frac{A}{c_{ij}^{\rm UL}\left(x_{it}, y_{it}\right)}$ represents the time used for tracking information transmission from user i to BS j. $\frac{D\left(\boldsymbol{l}_{i,t}\left(c_{ik}^{\rm DL}\left(x_{it}, y_{it}, b_i\left(\chi_{it}\right), n_{ik}\right)\right)\right)}{c_{ik}^{\rm DL}\left(x_{it}, y_{it}, b_i\left(\chi_{it}\right), n_{ik}\right)}$ represents the transmission latency for sending the tracking information from BS k to user i. For simplicity, hereinafter, $P_{it}^{\rm W}$ is referred as $P_{it}^{\rm W}\left(x_{it}, y_{it}, \chi_{it}, \boldsymbol{a}_{i,t}^{\rm UL}, \boldsymbol{a}_{i,t}^{\rm DL}\right)$. (7) shows that if the delay of VR video and tracking information transmission exceeds the target delay threshold allowed by VR systems or the quality of VR video cannot meet the video requirement, users will experience a BIP ($P_{it}^{\rm W}$=1). From (7), we can also see that, the BIP of user i caused by wireless transmission depends on user i's location, orientation, VR applications, and user association. (7) represents the BIP caused by wireless networking factors such as transmission delay and video quality. Next, we show the BIP model that jointly considers wireless transmission, the VR applications, and the users awareness. The BIP of user i can be given by [11]:

$$P_i\left(x_{it}, y_{it}, G_A, \chi_{it}, \boldsymbol{a}_{i,t}^{\rm UL}, \boldsymbol{a}_{i,t}^{\rm DL}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(G_A + P_{it}^{\rm W} + G_A P_{it}^{\rm W} + \epsilon_i + \epsilon_{G_A|i} + \epsilon_B\right), \quad (8)$$

where ϵ_i is the user i's awareness measured by VR users, $\epsilon_{G_A|i}$ is joint effect caused by user i's awareness and VR application G_A, and ϵ_B is a random effect. ϵ_i, $\epsilon_{G_A|i}$, and ϵ_B follow the Gaussian distribution [11] with zero mean and variances σ_i^2, $\sigma_{G_A|i}^2$, and σ_B^2, respectively. In (8), the value of $P_i\left(x_{it}, y_{it}, G_A, \chi_{it}, \boldsymbol{a}_{i,t}^{\rm UL}, \boldsymbol{a}_{i,t}^{\rm DL}\right)$ quantifies the average number of BIPs that user i can identify during a period. From (8), we can see that as the VR application for user i changes, the value of BIP will change. For example, a given user watching VR videos will experience fewer BIPs compared to a user engaged in an immersive first-person shooting game. This is due to the fact that in an immersive game environment, users are fully engaged with the virtual environment, as opposed to some VR applications that require the user to only watch VR videos. In (8), we can also see that the BIPs depend on the users' awareness. This means that different users will have different actions and perceptions when they interact with the virtual environment and, hence, different VR users may experience different levels of BIP.

C. Problem Formulation

From (8), we can see that the BIP of each user depends on the user's locations and orientations as well as its associations. Using an effective learning algorithm to predict users' locations and orientations, BSs can proactively determine the users' association to improve the downlink and uplink data rates and minimize BIP for each VR user. The BIP minimization problem can be given as follows:

$$\min_{\hat{x}_{it},\hat{y}_{it},\hat{\chi}_{it},\boldsymbol{a}_{i,t}^{\mathrm{UL}},\boldsymbol{a}_{i,t}^{\mathrm{DL}}} \sum_{i\in\mathcal{U}} P_i\left(\hat{x}_{it},\hat{y}_{it},G_A,\hat{\chi}_{it},\boldsymbol{a}_{i,t}^{\mathrm{UL}},\boldsymbol{a}_{i,t}^{\mathrm{DL}}\right) \tag{9}$$

$$\text{s. t.}\quad U_j \leqslant V, \quad \forall j\in\mathcal{B}, \tag{9a}$$

$$a_{ij,t}^{\mathrm{UL}} \in \{0,1\}, \quad \forall i\in\mathcal{U}, \forall j\in\mathcal{B}, \tag{9b}$$

$$a_{ij,t}^{\mathrm{DL}} \in \{0,1\}, \quad \forall i\in\mathcal{U}, \forall j\in\mathcal{B}, \tag{9c}$$

$$\sum_{j\in\mathcal{B}} a_{ij,t}^{\mathrm{UL}} = 1, \quad \forall i\in\mathcal{U}, \tag{9d}$$

$$\sum_{j\in\mathcal{B}} a_{ij,t}^{\mathrm{DL}} = 1, \quad \forall i\in\mathcal{U}, \tag{9e}$$

where $\hat{x}_{it}$, $\hat{y}_{it}$, and $\hat{\chi}_{it}$ are the predicted locations and orientation of user i at time t. U_j is the number of VR users associated with BS j over downlink and V is the maximum number of users that can be associated with each BS. (9b) and (9d) show that each user can associate with only one uplink BS while (9c) and (9e) indicate that each user can associate with only one BS at downlink. From (9), we can see that the BIPs of each user will depend on the user association as well as the users' locations and orientations. Meanwhile, the user association depends on the locations and orientations of the VR users. As the users' locations and orientations will continuously change as time elapses, BSs must proactively determine the user association to reduce the BIPs of VR users. Hence, it is necessary to introduce a learning algorithm to predict the users' locations and orientations in order to determine the user association and minimize BIPs of VR users.

III. Federated Echo State Learning for Predictions of the Users' Location and Orientation

Federated learning is a decentralized learning algorithm [12] that can operate by using training datasets that are distributed across multiple devices (e.g., BSs), instead of being centralized at one location or device [13]. For our system, one key advantage of federated learning is that it can allow multiple BSs to locally train their ESNs using their collected data and cooperatively build a learning model by sharing their locally trained models. Compared to existing federated learning algorithms [13] that use matrices to record the users' behavior and cannot analyze the correlation of the users' behavior data, we propose an ESN-based federated learning algorithm that can use an ESN to efficiently analyze the data related to the users' mobility and orientation since an ESN that is a recurrent neural network is good at analyzing time-related data. Moreover, ESNs only need to train an output weight matrix, hence, they reduce the training complexity of the federated learning algorithms. Next, we first introduce the components of the federated ESN learning model. Then, we explain the entire procedure of using our federated ESN learning algorithm to predict the users' mobility patterns and and orientation.

A. Components of Federated ESN Learning Algorithm

A federated ESN learning algorithm consists of four components: a) agents, b) input, c) output, and d) local ESN model, which are specified as follows:

- *Agent*: In our system, each BS j must implement at most U learning algorithms.
- *Input:* The input of the federated ESN learning algorithm that is implemented by BS j for the predictions of each VR user i is defined by a vector $\boldsymbol{x}_{ij} = \left[\boldsymbol{x}_{ij,1},\cdots,\boldsymbol{x}_{ij,T}\right]^{\mathrm{T}}$ that represents the information related to user i's mobility and orientation where $\boldsymbol{x}_{ij,t} = \left[\xi_{ij1,t},\ldots,\xi_{ijN_x,t}\right]$ represents user i's information related to mobility and orientation at time t. This information includes user i's locations, orientations, VR applications, and the time that user i associates with BS j. N_x is the number of properties that constitute a vector $\boldsymbol{x}_{ij,t}$. The input of the proposed algorithm will be combined with the ESN model to predict users' orientation and mobility patterns. BSs will use these predictions to determine user associations.
- *Output:* For each user i, the output of the federated ESN learning algorithm at BS j is a vector $\boldsymbol{y}_{ij,t} = \left[\hat{\boldsymbol{y}}_{ijt+1},\ldots,\hat{\boldsymbol{y}}_{ijt+Y}\right]$ of user i's locations and orientations where $\hat{\boldsymbol{y}}_{ijt+k} = \left[\hat{x}_{it+k},\hat{y}_{it+k},\hat{\chi}_{it+k}\right]$ with $\hat{x}_{it+k}$ and $\hat{y}_{it+k}$ being the predicted location coordinates of user i at time $t+k$ and $\hat{\chi}_{it+k}$ being the estimated orientation of user i at $t+k$. Y is the number of future time slots that a federated ESN learning algorithm can predict. The predictions of the locations and orientations can be used to determine the user's association.
- *Local ESN model:* For each BS j, a local ESN model is used to build the relationship between the input of all BSs and the predictions of the users' mobility and orientation. The local ESN model consists of the input weight matrix $\boldsymbol{W}_j^{\mathrm{in}} \in \mathbb{R}^{N_W\times T}$, recurrent matrix $\boldsymbol{W}_j \in \mathbb{R}^{N_W\times N_W}$, and the output weight matrix $\boldsymbol{W}_j^{\mathrm{out}} \in \mathbb{R}^{Y\times(N_W+T)}$. The values of $\boldsymbol{W}_j^{\mathrm{in}}$ and $\boldsymbol{W}_j$ are generated randomly. However, the output weight matrix $\boldsymbol{W}_j^{\mathrm{out}}$ need to be trained according to the inputs of all BSs. A parallel ESN model in which the ESNs are connected in series is used for the proposed algorithm.

B. ESN Based Federated Learning Algorithm for Users' Location and Orientation Predictions

Next, we introduce the entire procedure of training the proposed ESN-based federated learning algorithm. Our purpose of training ESN is to find an optimal output weight matrix in order to accurately predict the users' mobility patterns and orientations.

To introduce the training process, we first explain the ESN neuron state. The neuron states of the proposed algorithm implemented by BS j for the predictions of user i are:

$$\boldsymbol{\mu}_{j,t} = \boldsymbol{W}_j \boldsymbol{\mu}_{j,t-1} + \boldsymbol{W}_j^{\text{in}} \boldsymbol{x}_{ij,t}. \tag{10}$$

Based on the states of neurons and the inputs, the ESN can estimate the output, which is:

$$\hat{\boldsymbol{y}}_{ij,t} = \boldsymbol{W}_{j,t}^{\text{out}} \begin{bmatrix} \boldsymbol{x}_{ij,t} \\ \boldsymbol{\mu}_{j,t} \end{bmatrix}. \tag{11}$$

From (11), we can see that to enable an ESN to predict the users' mobility patterns and orientations, we only need to adjust the value of the output weight matrix. However, each BS can collect only partial data for each user and, hence, we need to use a distributed learning algorithm to train the ESNs. To introduce the distributed learning algorithm, we first define two matrices which are given by:

$$\boldsymbol{H}_j = \begin{bmatrix} \boldsymbol{x}_{ij,1} & \boldsymbol{\mu}_{j,1} \\ & \vdots & \\ \boldsymbol{x}_{ij,T} & \boldsymbol{\mu}_{j,T} \end{bmatrix} \text{ and } \boldsymbol{E}_j = [\boldsymbol{e}_{ij,1}, \ldots, \boldsymbol{e}_{ij,T}],$$

where $\boldsymbol{e}_{ij,t}$ is the desired locations and orientations of each VR user, given the ESN input $\boldsymbol{x}_{ij,t}$. Then, the training purpose can be given as follows:

$$\min_{\boldsymbol{W}^{\text{out}}} \frac{1}{2} \left(\sum_{j=1}^{B} \left\| \boldsymbol{W}^{\text{out}} \boldsymbol{H}_j^{\text{T}} - \boldsymbol{E}_j \right\|^2 \right) + \frac{\lambda}{2} \|\boldsymbol{W}^{\text{out}}\|. \tag{12}$$

(12) is used to find the optimal output weight matrix $\boldsymbol{W}^{\text{out}}$ according to which the BSs can predict the entire users' locations and orientations without the knowledge of the users' data collected by other BSs. From (12), we can see that, each BS j needs to adjust its output weight matrix $\boldsymbol{W}_j^{\text{out}}$ and find the optimal output weight matrix $\boldsymbol{W}^{\text{out}}$. The update of $\boldsymbol{W}_j^{\text{out}}$ is given by:

$$\begin{aligned} \boldsymbol{W}_{j,t+1}^{\text{out}} = & \varsigma^{-1} \left[\boldsymbol{I} - \boldsymbol{H}_j^{\text{T}} \left(\varsigma \boldsymbol{I} + \boldsymbol{H}_j \boldsymbol{H}_j^{\text{T}} \right) \boldsymbol{H}_j^{\text{T}} \right] \\ & \times \left(\boldsymbol{H}_j^{\text{T}} \boldsymbol{E}_j - \boldsymbol{n}_{j,t} + \varsigma \boldsymbol{W}_t^{\text{out}} \right), \end{aligned} \tag{13}$$

where ς is the learning rate and $\boldsymbol{W}_t^{\text{out}}$ is the optimal output weight matrix that the ESN model of each BS needs to find. From (13), we can see that $\boldsymbol{W}_{j,t+1}^{\text{out}}$ is the output weight matrix that is generated at BS j. $\boldsymbol{W}_{j,t+1}^{\text{out}}$ can only be used to predict partial mobility patterns and orientations given the users' data collected by BS j. $\boldsymbol{W}_{j,t+1}^{\text{out}}$ is different from the output weight matrices of other BSs. The optimal output weight matrix is:

$$\boldsymbol{W}_{t+1}^{\text{out}} = \frac{B \varsigma \hat{\boldsymbol{W}}_{t+1}^{\text{out}} + B \hat{\boldsymbol{n}}_t}{\lambda + \varsigma B}, \tag{14}$$

where $\hat{\boldsymbol{W}}_{t+1}^{\text{out}}$ and $\hat{\boldsymbol{n}}_{t+1}^{\text{out}}$ can be calculated as follows:

$$\hat{\boldsymbol{W}}_{t+1}^{\text{out}} = \frac{1}{B} \sum_{j=1}^{B} \boldsymbol{W}_{j,t+1}^{\text{out}}, \ \hat{\boldsymbol{n}}_t = \frac{1}{B} \sum_{j=1}^{B} \boldsymbol{n}_{j,t}. \tag{15}$$

In (13), $\boldsymbol{n}_{j,t}$ is the deviation between the output weight matrix $\boldsymbol{W}_{j,t+1}^{\text{out}}$ of each BS j and the optimal output weight matrix $\boldsymbol{W}_{t+1}^{\text{out}}$ that the ESN model of each BS needs to converge, which is given by:

$$\boldsymbol{n}_{j,t+1} = \boldsymbol{n}_{j,t} + \gamma \left(\boldsymbol{W}_{j,t+1}^{\text{out}} - \boldsymbol{W}_{t+1}^{\text{out}} \right). \tag{16}$$

Algorithm 1 Federated ESN learning algorithm for mobility and orientation predictions

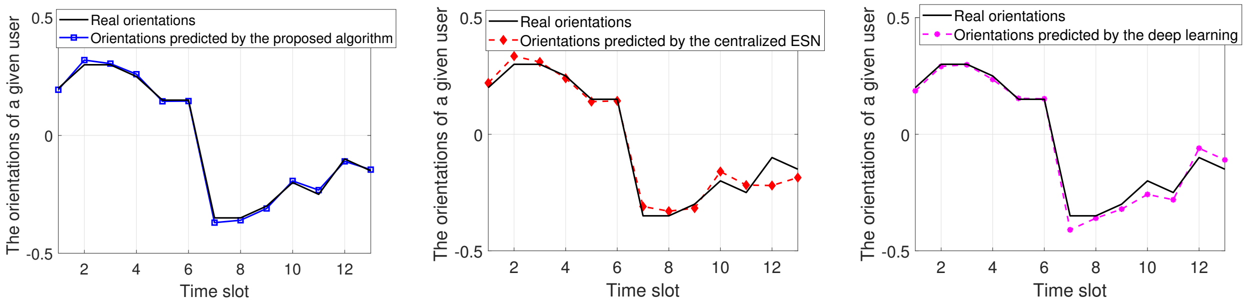
```
Input: Training data set (local), $\boldsymbol{x}_{ij}$.
Initialization: Each BS $j$ generates the ESN model for each user including $\boldsymbol{W}_j^{\text{in}}$ (local), $\boldsymbol{W}_j$ (global), and $\boldsymbol{W}_j^{\text{out}}$ (local).
Obtain the matrices $\boldsymbol{H}_j$ and $\boldsymbol{E}_j$ based on (10).
for time $t$ do
    Compute $\boldsymbol{W}_{j,t+1}^{\text{out}}$ using (13).
    Calculate $\hat{\boldsymbol{W}}_{t+1}^{\text{out}}$ and $\hat{\boldsymbol{n}}_t^{\text{out}}$ based on (15).
    Calculate $\boldsymbol{W}_{t+1}^{\text{out}}$ based on (14).
    Compute $\boldsymbol{n}_{j,t+1}$ based on (16).
    Compute $\|\boldsymbol{r}_{j,t+1}\|$ and $\|\boldsymbol{s}_{j,t}\|$.
    If $\|\boldsymbol{r}_{j,t+1}\| \leqslant \gamma_A$ or $\|\boldsymbol{s}_{j,t}\| \leqslant \gamma_A$, the algorithm converges.
end for
```

$\boldsymbol{W}_{t+1}^{\text{out}}$ is the global optimal output weight matrix that can be used to predict the entire mobility patterns and orientations of a given user. This means that using $\boldsymbol{W}_{t+1}^{\text{out}}$, each BS can predict the entire user's mobility patterns and orientations as the BS only collects partial data related to the user's mobility and orientations. As time elapses, $\boldsymbol{W}_{j,t+1}^{\text{out}}$ will finally converge to $\boldsymbol{W}_{t+1}^{\text{out}}$. In consequence, all of BSs can predict the entire mobility patterns and orientations of each user. To measure the convergence, we define two vectors which can be given by $\boldsymbol{r}_{j,t} = \boldsymbol{W}_{j,t}^{\text{out}} - \boldsymbol{W}_t^{\text{out}}$ and $\boldsymbol{s}_{j,t} = \boldsymbol{W}_t^{\text{out}} - \boldsymbol{W}_{t-1}^{\text{out}}$. As $\|\boldsymbol{r}_{j,t+1}\| \leqslant \gamma_A$ or $\|\boldsymbol{s}_{j,t}\| \leqslant \gamma_A$, the proposed algorithm converges. As the learning algorithm converges, each BS can use its own ESN to predict the entire mobility and orientation of each VR user. According to these predictions, BSs can determine the user association to minimize the BIPs of VR users. Algorithm 1 summarizes the entire process of using ESN based federated learning algorithm for the predictions of the users' mobility patterns and orientations. Based on the predictions of the users' orientations and mobility patterns, we can use a reinforcement learning algorithm given in [14] to find a sub-optimal solution. The reinforcement learning algorithms can learn the VR users state and exploit different actions to adapt the user association according to the the predictions of the users' mobility and orientation. After the learning step, each BS will find a sub-optimal user association.

IV. Simulation Results

For our simulations, we consider a circular area with radius $r = 500$ m, $U = 20$ wireless VR users, and $B = 5$ BSs distributed uniformly. The orientation data is collected from a first-person shooter game at the Youtube Website. In particular, we record the users' orientations from 25 videos of the first-person shooter VR game. For comparison purposes, we consider the deep learning algorithm in [7] and the ESN algorithm in [15], as two baseline schemes. All statistical results are averaged over a large number of independent runs.

Fig. 1 show the predictions of the VR users' orientations as time elapses. To simplify the model training, the collected data related to orientations are mapped to $[-0.5, 0.5]$. From Fig. 1, we observe that the proposed algorithm can predict the users' orientations more accurately than the centralized ESN and deep learning algorithms. Figs. 1(b) and 1(c) also show that the prediction error mainly occur at time slot 8 to 12.



(a)

(b)

(c)

Fig. 1. Predictions of the VR users' orientations as time elapses.

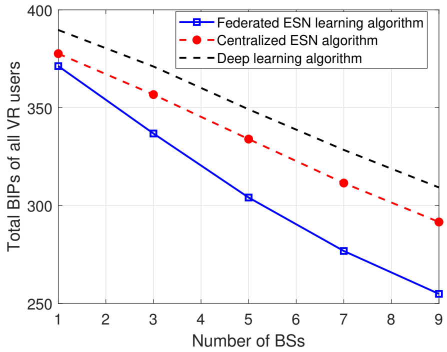


Fig. 2. Total BIPs experienced by VR users as the number of BSs varies.

This is due to the fact that the proposed algorithm can build a learning model that predicts the entire mobility and orientation of each user. In particular, the output weight matrices of all ESN algorithms implemented by each BS will converge to a common matrix. Hence, BSs can predict the entire mobility and orientations of each VR user.

Fig. 2 shows how the total BIP of all VR users changes as the number of BSs varies. From Fig. 2, we can see that, as the number of BSs increases, the total BIP of all VR users decreases. That is because as the number of BSs increases, the VR users have more connection options. Hence, the blockage caused by human bodies will be less severe, thereby improving the data rates of VR users. Fig. 2 also shows that the proposed algorithm can achieve up to 16% and 26% reduction in the number of BIPs, respectively, compared to centralized ESN algorithm and deep learning algorithm for a network with 9 BSs. These gains stem from the fact that the centralized ESN and deep learning algorithms can partially predict the mobility and orientation of each VR user as they rely only on the local data collected by a BS. In contrast, the proposed algorithm facilitates cooperation among BSs to build a learning model that can predict the entire users' mobility and orientations.

V. Conclusion

In this paper, we have developed a novel framework for minimizing BIPs within VR applications that operate over wireless networks. To this end, we have developed a BIP model that jointly considers the VR applications, transmission delay, VR video quality, and the user's awareness. We have then formulated an optimization problem that seeks to minimize the BIP of VR users by predicting users' mobility and orientation, as well as determining the user association. To solve this problem, we have developed a novel federated learning algorithm based on echo state networks. The proposed federated ESN algorithm enables the BSs to train their ESN with their locally collected data and share these models to build a global learning model that can predict the entire mobility pattern and orientations of each VR user. Using these predictions, each BS can determine the user association in both uplink and downlink. Simulation results have shown that, when compared to the centralized ESN and deep learning algorithms, the federated ESN approach achieves significant performance gains in terms of BIPs.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, to appear, 2019.

[2] Y. Sun, Z. Chen, M. Tao, and H. Liu, "Communication, computing and caching for mobile VR delivery: Modeling and trade-off," *arXiv preprint arXiv:1804.10335*, April 2018.

[3] J. Park and M. Bennis, "URLLC-eMBB slicing to support VR multimodal perceptions over wireless cellular systems," *available online: arxiv.org/abs/1805.00142*, May 2018.

[4] A. Taleb Zadeh Kasgari, W. Saad, and M. Debbah, "Human-in-the-loop wireless communications: Machine learning and brain-aware resource management," *IEEE Transactions on Communications*, to appear, 2019.

[5] M. Chen, W. Saad, and C. Yin, "Virtual reality over wireless networks: Quality-of-service model and learning-based resource management," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5621–5635, Nov. 2018.

[6] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Communications Surveys Tutorials*, to appear, 2019.

[7] N. T. Nguyen, Y. Wang, H. Li, X. Liu, and Z. Han, "Extracting typical users' moving patterns using deep learning," in *Proc. of IEEE Global Communications Conference*, Anaheim, CA, USA, Dec 2012.

[8] O. Semiari, W. Saad, M. Bennis, and M. Debbah, "Integrated millimeter wave and sub-6 Ghz wireless networks: A roadmap for joint mobile broadband and ultra-reliable low-latency communications," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 109–115, April 2019.

[9] O. Semiari, W. Saad, M. Bennis, and Z. Dawy, "Inter-operator resource management for millimeter wave multi-hop backhaul networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5258–5272, Aug 2017.

[10] J. Jerald, *The VR book: Human-centered design for virtual reality*, Morgan & Claypool, Sept. 2015.

[11] J. Chung, H. J. Yoon, and H. J. Gardner, "Analysis of break in presence during game play using a linear mixed model," *ETRI journal*, vol. 32, no. 5, pp. 687–694, Oct. 2010.

[12] M. M. Amiri and D. Gunduz, "Computation scheduling for distributed machine learning with straggling workers," *arXiv preprint arXiv:1810.09992*, Oct. 2018.

[13] V. Smith, C. K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," in *Proc. of Advances in Neural Information Processing Systems*, Long beach, CA,USA, Dec. 2017.

[14] M. Bennis and D. Niyato, "A Q-learning based approach to interference avoidance in self-organized femtocell networks," in *Proc. of IEEE Global Communications Conference Workshops*, Miami, FL, USA, Dec 2010.

[15] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE Journal on Selected Areas on Communications (JSAC)*, vol. 35, no. 5, pp. 1046–1061, May 2017.