

# User-Centric Communication Service Provision for Edge-Assisted Mobile Augmented Reality

Conghao Zhou , *Member, IEEE*, Jie Gao , *Senior Member, IEEE*, Shisheng Hu , *Member, IEEE*, Nan Cheng , *Senior Member, IEEE*, Weihua Zhuang , *Fellow, IEEE*, and Xuemin Shen , *Fellow, IEEE*

**Abstract**—Future 6G networks are envisioned to facilitate edge-assisted mobile augmented reality (MAR) via strengthening the collaboration between MAR devices and edge servers. In order to provide immersive user experiences, MAR devices must timely upload camera frames to an edge server for simultaneous localization and mapping (SLAM)-based device pose tracking. In this paper, to cope with user-specific and non-stationary uplink data traffic, we develop a digital twin (DT)-based approach for user-centric communication service provision for MAR. Specifically, to establish DTs for individual MAR devices, we first construct a data model customized for MAR that captures the intricate impact of the SLAM-based frame uploading mechanism on the user-specific data traffic pattern. We then define two DT operation functions that cooperatively enable adaptive switching between different data-driven models for capturing non-stationary data traffic. Leveraging the user-oriented data management introduced by DTs, we propose an algorithm for network resource management that ensures the timeliness of frame uploading and the robustness against inherent inaccuracies in data traffic modeling for individual MAR devices. Trace-driven simulation results demonstrate that the user-centric communication service provision achieves a 14.2% increase in meeting the camera frame uploading delay requirement in comparison with the slicing-based communication service provision widely used for 5G.

**Index Terms**—6G, mobile augmented reality, user-centric service provision, digital twin, graph neural network.

## I. INTRODUCTION

IMMERSIVE communications are expected to extend the enhanced Mobile Broadband (eMBB) communications of 5G networks for providing users with rich immersive experiences in future 6G networks [1]. Augmented reality (AR), a typical form of immersive communications, enables seamless integration of virtual objects into the physical environment surrounding human

users. Driven by the increasing demand for immersive experiences, AR accessible on portable devices such as smart glasses, referred to as mobile AR (MAR), is gaining widespread attention as one of the emerging applications in the 6G era. All MAR applications need device pose tracking, a resource-intensive procedure fundamental to the 3D alignment of virtual objects with the physical environment [2], [3]. Enabling device pose tracking is a key challenge for current MAR devices due to their resource limitations such as limited battery power. To overcome this challenge, leveraging the resources of edge servers via edge-assisted MAR is a promising paradigm [4].

Although communication service provision for edge-assisted applications has been investigated in the context of 5G networks [5], [6], supporting edge-assisted device pose tracking for MAR applications in future 6G networks requires novel solutions for the following two reasons [7]. First, human movements such as changing gaze points are incessant in MAR applications, and differences in the movements of different users result in distinctive network resource demands even if they are using the same MAR application [8]. Most existing communication service provision approaches rely on *service-oriented* demand modeling, which cannot capture the variances in service demands across different users to support personalized MAR user experiences. For example, slicing-based communication service provision widely adopted in 5G networks for aggregated video-type data traffic transmission [9] lacks the capability to characterize the video traffic patterns of individual users. Second, to deal with the uncertainties in human movements, current MAR has incorporated a complex localization-based operational mechanism [4], i.e., simultaneous localization and mapping (SLAM)-based device pose tracking [10] with multiple interacting functionality modules, from the perspective of the application layer. Current communication service provision approaches developed from the networking perspective mostly overlook the impact of this intricate operational mechanism for MAR applications [11], [12], thereby unable to natively cope with uncertainties in human movements [13]. As a result, *user-centric* communication service provision that can properly take the impact of MAR operational mechanism into account and characterize user-specific service demands is essential for future 6G networks to enhance edge-assisted device pose tracking in MAR.

The digital twin (DT) technique can support user-centric communication service provision for device pose tracking in MAR via user device virtualization [7]. The use of DT in communications and networking is still in its infancy. We explore

Received 20 December 2024; revised 15 April 2025; accepted 27 September 2025. Date of publication 6 October 2025; date of current version 4 February 2026. Recommended for acceptance by T. X. Han. (*Corresponding author: Shisheng Hu.*)

Conghao Zhou was with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. He is now with the School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi 710071, China (e-mail: zhouconghao@xidian.edu.cn).

Jie Gao is with the School of Information Technology, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: jie.gao6@carleton.ca).

Shisheng Hu, Weihua Zhuang, and Xuemin Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: s97hu@uwaterloo.ca; sshen@uwaterloo.ca; wzhuang@uwaterloo.ca).

Nan Cheng is with the School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi 710071, China (e-mail: nancheng@xidian.edu.cn).

Digital Object Identifier 10.1109/TMC.2025.3618147

establishing DTs as digital representations of individual MAR devices. By doing so, we leverage DTs to obtain and process high-quantity, high-quality data related to MAR devices [14]. Unlike mathematical methods that explicitly model certain characteristics, such as the statistical frame uploading pattern of an MAR device, DTs can capture implicit yet crucial information about the characteristics using data-driven techniques [15]. Additionally, DTs can be customized to provide the flexibility required to capture the unique characteristics of individual entities, e.g., the service demand pattern of each MAR device [16].

Despite these advantages, research on DTs in communication and networking is still in its infancy. Three issues arise when establishing a user DT supporting user-centric communication service provision for edge-assisted MAR. First, establishing a proper data model that effectively captures the impact of various factors in the intricate SLAM-based device pose tracking on the user-specific service demands is challenging [17]. The selection of key factors for proper demand modeling and the organization of the data corresponding to these factors in a structured way are not explored yet for MAR applications. Second, temporal variations in user movements may lead to non-stationary service demand in MAR. For example, the data traffic load for uploading camera frames may surge intermittently following an event of device pose tracking loss [8]. Non-stationary data traffic presents a challenge to effectively characterizing user-specific service demands. Third, since no modeling technique can perfectly capture user-specific service demands, DT-based communication service provision decisions should account for modeling inaccuracies [9], [18]. However, obtaining a proper representation of modeling inaccuracies, which is crucial for optimizing DT-based communication service provision, is also a challenge.

In this paper, we investigate communication service provision to facilitate timely camera frame uploading, which is essential for supporting SLAM-based device pose tracking in edge-assisted MAR. To address the aforementioned challenges, we establish an augmented reality user digital twin (MARLIN), building on our general DT framework [7], and propose a DT-based user-centric communication service provision approach to support edge-assisted device pose tracking for MAR in 6G networks. MARLIN consists of a customized data model to characterize the service demand of each individual MAR device and various DT operation functions to update the data model. The contributions of this paper are summarized as follows:

- We establish a data model customized for edge-assisted MAR, including well-structured data attributes and differentiated data update mechanisms, which supports capturing the unique service demands of individual MAR devices flexibly.
- We define two novel DT operation functions, which jointly capture the implicit impact of the operational mechanism of SLAM-based device pose tracking on the non-stationary uplink data traffic from each MAR device. One of them implements two data-driven traffic modeling methods with different complexities, and the other adaptively switches between these two methods in response to data traffic variations.

- We propose a user-centric communication service provision algorithm based on MARLIN, which ensures robustness against the inherent inaccuracies resulting from our user-level data traffic modeling.
- We demonstrate the potential of MARLIN in facilitating user-centric communication service provision, by flexibly adapting data granularity to various data-driven modeling methods, through trace-driven simulations.

The remainder of this paper is organized as follows. Section II summarizes related works. Section III presents the considered scenario, system models, and problem formulation. Sections IV and V introduce MARLIN and the proposed DT-based user-centric communication service provision, respectively. Section VI presents the simulation results, and Section VII concludes this paper.

## II. RELATED WORKS

In this section, we first review the related literature on device pose tracking techniques used in MAR. Then, we summarize existing research works on communication service provision from the networking perspective.

### A. Device Pose Tracking in MAR

To align virtual objects with the physical environment within the field of view of each MAR device, tracking the MAR device's pose, including its position and orientation, is fundamental [19]. The SLAM-based device pose tracking techniques and their variants are widely applied in MAR [17]. Such techniques enable accurate 3D alignment by leveraging the 3D location information of physical objects. A trending research direction is using various visual features, e.g., binary robust independent elementary features (BRIEF) [10] in the design of SLAM or MAR systems to enhance the tracking accuracy [20], [21]. Many researchers focus on combining traditional visual features and information collected from additional sensors, e.g., LiDAR [22] and inertial measurement units [23], to improve the tracking accuracy. As a result, the enhanced tracking accuracy comes at the cost of increased operational complexity and resource demands, thereby posing a significant challenge to resource-constrained mobile devices [15].

To overcome the resource constraints of MAR devices, many recent studies have adopted cloud or mobile edge computing paradigms [24]. Specifically, the SLAM system for device pose tracking can be divided into two modules: a computing-intensive mapping module and a lightweight tracking module, which are handled by a cloud/edge server and a mobile device, respectively [4]. Ben Ali et al. develop a practical edge-assisted SLAM system for the mobile device, demonstrating the feasibility and potential of edge-assisted device pose tracking [25]. The authors in [26] and [15] design adaptive edge-assisted device pose tracking schemes aim to overcome constrained uplink data rates and time-varying uplink data rates, respectively, between the edge server and an MAR device. In addition, to support accurate virtual object rendering in multi-user MAR applications, existing studies also investigate coordinate synchronization for the device pose tracking of different MAR devices [8], [27], [28].

Different from aforementioned studies that advance device pose tracking from the application perspective, our work emphasizes service provision from the communication and networking perspective to support camera frame uploading, which is crucial for ensuring the timeliness of device pose tracking in edge-assisted MAR.

### B. Communication Service Provision

Communication service provision in existing works concentrates on configuring networks to manage network resources for meeting service demands. Although communication service provision is a classic research topic in networking, its significance persists due to the increasingly diversified and demanding services [7].

In the early phase of wireless communication network research, mobile communication service provision primarily focuses on the on-demand allocation of communication resources such as the radio spectrum and power at each base station. By modeling communication service demands, e.g., data packet arrival patterns and device location distributions, communication resources can be allocated according to the service demand modeling while considering constraints on resources [29], interference management [30], etc. With the advent of 5G, mobile edge computing (MEC) results in a plethora of computing services that require communication, computing, and data storage resources. Some works investigate multi-resource management problems for MEC in network scenarios such as drone-assisted networks [31], [32], [33] and terrestrial networks [34]. Meanwhile, network slicing, as an innovation in 5G, enables the establishment of isolated virtual networks, i.e., slices, over the same physical network infrastructure and differentiated resource management for different services to satisfy their respective service requirements [35]. As a result, slicing-based communication service provision approaches are developed for various network scenarios [11], [12], [35], [36], [37], [38]. These works focus on minimizing the network resource consumption while satisfying service requirements, including the timely completion of dynamically requested computing tasks [36], the reliability of resource management [11], [37], and end-to-end delay requirements [35], [38].

Despite these advancements, few studies have investigated communication service provision for MAR, especially for device pose tracking. Since the MAR operational mechanism has a significant impact on the service demands of MAR devices, existing communication service provision approaches that overlook the influence of the intrinsic MAR operational mechanism may fail to accurately capture MAR service demands [4]. In addition, the slicing-based communication service provision primarily targets meeting the aggregated service demands of users running the same application, rather than user-specific service demands. As a result, they are not suitable for MAR applications, which involve uncertain user-specific movements [11], [12].

Different from existing works, we aim to achieve user-centric communication service provision for MAR applications. Specifically, we establish a DT for each individual MAR device, which supports a data-driven method to characterize the impact of the

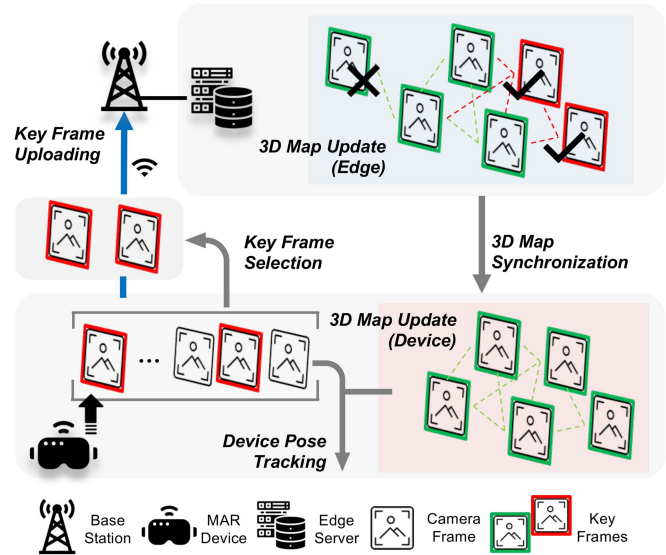


Fig. 1. An illustration of edge-assisted device pose tracking for one MAR device.

complex MAR operational mechanism on user-specific service demands.

## III. EDGE-ASSISTED DEVICE POSE TRACKING IN MAR

### A. Considered Scenario

When a user runs an MAR application, the position and orientation (jointly referred to as the *3D pose*) of the MAR device change over time due to the user movements. The MAR device captures camera frames periodically with a fixed frame rate and tracks its 3D pose based on the captured camera frames, which is crucial for rendering virtual objects at correct locations within the user's field of view [10].

An edge server deployed at a base station (BS) assists the device pose tracking of a set of MAR devices, denoted by  $\mathcal{N}$ , within the communication coverage of the BS. Currently, SLAM techniques are widely adopted to support device pose tracking in MAR [4]. An example paradigm of edge-assisted device pose tracking for one MAR device is illustrated in Fig. 1, in which the edge server and the device collaboratively track its 3D pose. Specifically, the edge server includes a resource-intensive mapping module to construct and update a 3D representation (i.e., a 3D map) of the physical environment surrounding each MAR device for enabling device pose tracking, while the MAR device includes a lightweight tracking module for calculating its real-time 3D pose based on the 3D map [26]. For each MAR device, both the device and the edge server maintain their own 3D maps during device pose tracking. For clarity, we refer to them as the “3D map (device)” and “3D map (edge)” in this paper, respectively.

As shown in Fig. 1, edge-assisted device pose tracking can be divided into four key steps [25]: (i) each MAR device selects a subset of recently captured camera frames as the key frames and uploads these key frames to the edge server over a wireless communication link; (ii) the mapping module equipped at

the edge server updates the 3D map (edge) by incorporating newly uploaded key frames along with previously stored ones; (iii) the edge server sends the updated 3D map (edge) to this MAR device for 3D map synchronization; and (iv) the tracking module at each MAR device updates the 3D map (device) based on the information on the received 3D map (edge) and leverages the updated 3D map (device) to locally calculate the device pose for every camera frame. These four steps are performed iteratively during device pose tracking, as detailed in the following subsections.

### B. Key Frame Uploading

To cope with uncertain user movements, each MAR device with a fixed frame rate periodically selects a number of key frames from recently captured camera frames and uploads them to the edge server for updating the 3D map (edge). Let  $\mathcal{F}_n$  denote the set of all camera frames captured by MAR device  $n \in \mathcal{N}$ . We refer to the duration of  $F$  consecutive camera frames as a time slot, i.e., the period of key frame uploading, and denote the set of all time slots by  $\mathcal{T}$ . Let  $\mathcal{F}_{n,t} \subseteq \mathcal{F}_n$  denote the set of camera frames captured during time slot  $t \in \mathcal{T}$ .

For camera frame  $f \in \mathcal{F}_n$  of MAR device  $n$ , we use  $a_{n,f} \in \{0, 1\}$  to indicate whether it is a key frame or not. Let  $a_{n,f} = 1$  if camera frame  $f$  is selected as a key frame, and  $a_{n,f} = 0$  otherwise. The detail about key frame selection will be presented in Section III-C. Denote the set of key frames selected in time slot  $t$  by  $\mathcal{K}_{n,t} = \{f \in \mathcal{F}_{n,t} | a_{n,f} = 1\}$ . Considering the time-varying number of key frames for uploading [25], we define the number of key frames in each time slot as a random variable, given by:

$$\tilde{k}_{n,t} = \sum_{f \in \mathcal{F}_{n,t}} a_{n,f}, \quad \forall t \in \mathcal{T}. \quad (1)$$

At the end of time slot  $t$ , MAR device  $n$  uploads the set of selected key frames, i.e.,  $\mathcal{K}_{n,t}$ , to the edge server.

Proactive radio spectrum resource management is necessary to ensure timely key frame uploading for facilitating real-time device pose tracking in MAR [4]. Denote the radio spectrum bandwidth reserved for MAR device  $n$  in time slot  $t$  for key frame uploading by  $b_{n,t}$ . For proactive radio spectrum resource management, the value of  $b_{n,t}$  should be determined at the beginning of each time slot. Given  $b_{n,t}$ , the uplink data rate of MAR device  $n$  within time slot  $t$ , denoted by  $r_{n,t}$ , is as follows:

$$r_{n,t} = b_{n,t} \log(1 + \gamma_{n,t}), \quad \forall t \in \mathcal{T}, n \in \mathcal{N}, \quad (2)$$

where  $\gamma_{n,t}$  represents the estimated signal-to-noise ratio (SNR) at MAR device  $n$  in time slot  $t$  [39]. We denote the volume of data (in bits) for uploading each camera frame by  $\alpha$ , the same for all camera frames.

To achieve user-centric communication service provision for MAR, we target ensuring the timely key frame uploading of each individual MAR device, rather than focusing on the performance averaged over multiple devices as in slicing-based communication service provision [12]. Uploading the set of key frames selected by MAR device  $n$  in time slot  $t$ , i.e.,  $\tilde{k}_{n,t}$ , the following

constraint should be satisfied [40]:

$$P(\alpha \tilde{k}_{n,t} \leq T^r r_{n,t}) \geq \epsilon, \quad \forall t \in \mathcal{T}, n \in \mathcal{N}, \quad (3)$$

where  $\epsilon \in [0, 1]$  represents the required reliability of communication service provision, and  $T^r$  represents the maximum tolerable total transmission duration for uploading the selected key frames from an MAR device.

### C. 3D Map Update & Synchronization

The key frames uploaded by each MAR device are used to form or update the 3D map (edge) at the edge server. The 3D map consists of a set of key frames uploaded by the MAR device over time as well as the feature points (FPs), e.g., a wall corner, detected from each key frame [10].<sup>1</sup> The updates of the two 3D maps are different but correlated.

The set of key frames stored in the 3D map (edge) of MAR device  $n \in \mathcal{N}$  at time slot  $t$ , denoted by  $\mathcal{M}_{n,t}^E \subseteq \mathcal{F}$ , changes as shown in the following equation [15], [26]:

$$\mathcal{M}_{n,t}^E = \{\mathcal{M}_{n,t-1}^E \cup \mathcal{K}_{n,t-1}\} \setminus \mathcal{C}_{n,t-1}, \quad (4)$$

where  $\mathcal{C}_{n,t-1} \subseteq \mathcal{M}_{n,t-1}^E$  represents the set of key frames removed from the 3D map (edge)  $\mathcal{M}_{n,t-1}^E$  of MAR device  $n$  in time slot  $t$ . We assume that the update of  $\mathcal{M}_{n,t}^E$  can be accomplished at the end of each time slot. After the 3D map (edge) update, the set  $\mathcal{M}_{n,t}^E$  is downloaded by MAR device  $n$  for updating the 3D map (device) [25].

In MAR applications, key frame selection and uploading is conducted based on an implicit policy, denoted by  $\Pi$ , used by the MAR operational mechanism of device pose tracking [10]. Generally, in a policy, a frame is selected as a key frame if it differs sufficiently from its preceding key frames and thus can provide adequate new environment information, e.g., FPs, while having sufficient overlap with preceding for easily implementing device pose tracking [10]. While factors influencing key frame selection and uploading may differ in existing policies [15], [25], [26], a general policy  $\Pi$  can be formulated as follows:

$$a_{n,f} = \Pi(\mathcal{M}_{n,f}^D, \mathcal{X}_{n,f}, f), \quad \forall f \in \mathcal{F}_t, t \in \mathcal{T}, \quad (5)$$

where  $\mathcal{M}_{n,f}^D$  denotes the set of key frames in the 3D map (device) of MAR device  $n$ , and  $\mathcal{X}_{n,f} = \{f' \in \mathcal{F} | f - X \leq f' < f\}$  represents the set of  $X$  preceding camera frames when camera frame  $f$  is captured by MAR device  $n$ . Given the value of  $a_{n,f}$ , the 3D map (device) is updated as follows:

$$\mathcal{M}_{n,f}^D = \begin{cases} \mathcal{M}_{n,f-1}^D \cup \{f\} & \text{if } a_{n,f-1} = 1; \\ \mathcal{M}_{n,f-1}^D & \text{otherwise.} \end{cases} \quad (6)$$

Given (5) and (6), key frame selection and uploading, depending on policy  $\Pi$ , can be abstracted as a sequential decision making process [15].

<sup>1</sup>A 3D map consists of a set of key frames and the set of FPs identified in each key frame [10]. Since the update of FPs is determined by the update of key frames, we use the adding and deleting of key frames to represent the update of a 3D map for simplicity.

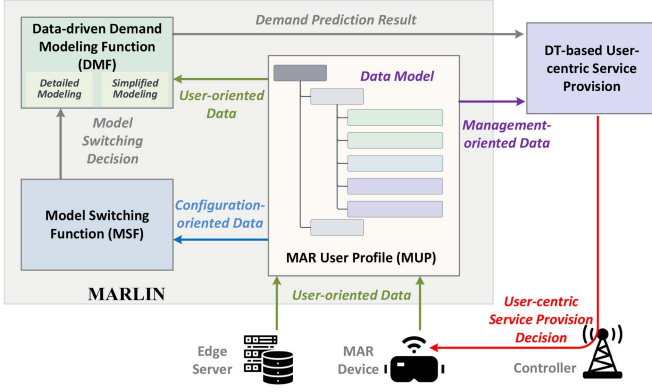


Fig. 2. The workflow of the proposed DT-based user-centric communication service provision approach.

#### D. Problem Formulation

To efficiently support edge-assisted device pose tracking in MAR, we formulate a user-centric communication service provision problem with the objective of minimizing the radio spectrum resource reserved for the key frame uploading of all MAR devices, as follows:

$$\text{P1: } \min_{\{b_{n,t}\}_{n \in \mathcal{N}, t \in \mathcal{T}}} \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} b_{n,t} \quad (7a)$$

$$\text{s.t. } P(\alpha \tilde{k}_{n,t} \leq T^r r_{n,t}) \geq \epsilon, \quad \forall n \in \mathcal{N}, t \in \mathcal{T}, \quad (7b)$$

where constraint (7b) ensures that the key frame uploading of each individual MAR device satisfies the delay requirement with a sufficiently high probability.<sup>2</sup> In Problem P1, random variables  $\tilde{k}_{n,t}$  are unknown a priori, and temporal variations in data traffic of each MAR device may be non-stationary. Existing communication service provision approaches optimize radio spectrum resource reservation based on either mathematical modeling or data-driven prediction [9]. However, mathematically modeling  $\tilde{k}_{n,t}$  is intractable due to the implicit policy  $\Pi$ . While data-driven approaches, e.g., deep neural networks (DNNs), potentially outperform mathematical modeling in the accuracy of approximating a complex policy  $\Pi$ , existing approaches overlook the impact of the MAR operational mechanism of device pose tracking on the data traffic for key frame uploading [25], [26], and thereby struggle to adapt to non-stationary data traffic.

To solve Problem P1, we develop a novel digital twin (DT)-based approach, as shown in Fig. 2. First, we establish a user DT that can characterize the impact of the MAR operational mechanism on the data traffic from individual MAR devices. Second, we propose a user-centric communication service provision method with the consideration of potential modeling inaccuracies. The two modules are presented in Sections IV and V, respectively.

<sup>2</sup>We focus on the camera frame uploading delay that is affected by decisions for the data plane, e.g., network resource provisioning decisions, rather than delay stemming from the control plane, e.g., control signaling delay.

#### IV. MAR USER DIGITAL TWIN ESTABLISHMENT

In this section, we customize an augmented reality user digital twin, called MARLIN, for each individual MAR device. Our designs for MARLIN evolves from the framework presented in [14], [15]. As illustrated by the light grey block in Fig. 2, MARLIN, comprising an *MAR user profile* (MUP) and two *user DT operation functions* (UDTOFs), is deployed at the BS and maintained by the controller to facilitate communication service provision for MAR. In the MUP, we define a *data model* customized for MAR, consisting of a set of well-structured data attributes chosen to characterize the service demand of the corresponding MAR device and to support our user-centric communication service provision. The two UDTOFs are leveraged to process the collected raw data in various ways for updating the data within the MUP accordingly. We introduce the two UDTOFs first, followed by the MUP.

##### A. Data-Driven Demand Modeling Function (DMF)

User-centric communication service provision for MAR requires an accurate model to capture the uplink data traffic of key frame uploading at each individual MAR device. To this end, we use a Markov decision process to model the sequential decision making of each MAR device on key frame selection and uploading.

For MAR device  $n$ , we denote a state by  $\mathbf{s}_{n,f} \in \mathcal{S}_n$  and define  $a_{n,f}$  as an action, i.e., determining whether camera frame  $f$  is a key frame or not, based on the state. Define state transition probability function  $P(\mathbf{s}_{n,f+1} | \mathbf{s}_{n,f}, a_{n,f}) := \mathcal{S}_n \times \mathcal{A}_n \rightarrow \mathcal{S}_n$ . The choice of the action  $a_{n,f}$  depends on the policy  $\Pi$  mentioned in (5).

Given actions  $\mathbf{a}_{n,t} = [a_{n,f}]_{\forall f \in \mathcal{F}_{n,t}} \in \mathcal{A}_n$  in time slot  $t$ , the data traffic of MAR device  $n$  for key frame uploading in time slot  $t$  can be obtained from (1).

The goal of our demand modeling function (DMF) is to approximate policy  $\Pi$  for predicting the data traffic of key frame uploading. A foundation for approximating  $\Pi$  is to define the states. Considering model accuracy and model complexity, we introduce two types of states for a *detailed* and a *simplified* data traffic modeling in the DMF, respectively. In addition to the approximation of policy  $\Pi$ , the DMF should approximate the state transition probability function  $P(\mathbf{s}_{n,f+1} | \mathbf{s}_{n,f}, a_{n,f})$  to support data traffic modeling.

1) *Detailed Modeling*: In SLAM-based device pose tracking, the relations among the sets of FPs corresponding to camera frames are essential for estimating the device poses at the instants of capturing these camera frames, thereby significantly affecting key frame selection and uploading in MAR [20]. In our detailed modeling, we use a graph to model the relations among the sets of FPs corresponding to camera frames and capture the impact of such a graph on the data traffic load of each MAR device [26].

Denote the set of FPs identified in camera frame  $f$  by  $\mathcal{U}_f$ . We model the 3D map (device) of MAR device  $n$  and the camera frame  $f$ , i.e.  $\mathcal{M}_{n,f}^D \cup \{f\}$ , as a weighted undirected graph, denoted by  $\mathcal{G}_{n,f}^D = (\mathcal{M}_{n,f}^D \cup \{f\}, \mathcal{E}_{n,f}^D)$ , where  $\mathcal{E}_{n,f}^D$  represents the set of edges between every pair of camera frames in  $\mathcal{M}_{n,f}^D \cup \{f\}$ . For edge  $e = (f, f') \in \mathcal{E}_{n,f}^D$  connecting camera

frames  $f, f' \in \mathcal{M}_{n,f}^D$ , its weight is defined as the Jaccard coefficient, denoted by  $\varepsilon_{f,f'} \in [0, 1]$ , which quantifies the similarity of sets  $\mathcal{U}_f$  and  $\mathcal{U}_{f'}$  as follows:

$$\varepsilon_{f,f'} = \frac{|\mathcal{U}_f \cap \mathcal{U}_{f'}|}{|\mathcal{U}_f \cup \mathcal{U}_{f'}|}, \quad \forall \mathcal{U}_f \cup \mathcal{U}_{f'} \neq \emptyset, \quad (8)$$

where  $|\cdot|$  represents the cardinality of a set. When many observations are shared by camera frames  $f$  and  $f'$ , the two sets of FPs  $\mathcal{U}_f$  and  $\mathcal{U}_{f'}$  are similar, resulting in a large  $\varepsilon_{f,f'}$ . Similarly, we model the set of camera frames  $\mathcal{X}_{n,f}$  defined right after (5) as another weighted undirected graph, denoted by  $\mathcal{G}_{n,f} = (\mathcal{X}_{n,f}, \mathcal{E}_{n,f})$ . Based on these graphs, we define the state used in the detailed modeling for MAR device  $n$  as follows:

$$\mathbf{s}_{n,f}^D = [\mathcal{G}_{n,f}^D, \mathcal{G}_{n,f}], \quad \forall n \in \mathcal{N}, f \in \mathcal{F}_n, \quad (9)$$

which is abbreviated as the *D-state*.

We adopt a graph convolutional network (GCN)-based technique to approximate policy  $\Pi$  by using the D-state  $\vartheta^D$  as the input and action  $a_{n,f}$  as the output. Let  $\vartheta^D$  denote the parameters of the GCN, and define  $\Pi^D$  as the policy generating action by using this GCN. To make policy  $\Pi^D$  approach policy  $\Pi$ , we optimize the parameters  $\vartheta^D$  by minimizing the following loss function:

$$L(\vartheta^D) = \frac{1}{|\Xi|} \sum_{(a_{n,f}, \mathbf{s}_{n,f}^D) \in \Xi} (a_{n,f} - \Pi^D(\mathbf{s}_{n,f}^D; \vartheta^D))^2, \quad (10)$$

where  $\Xi$  represents a set containing historical information on actions and D-states, stored in the MUP for MAR device  $n$ .

For graphs  $\mathcal{G}_{n,f}^D$  and  $\mathcal{G}_{n,f}$ , we use two GCNs for feature extraction. The extracted features are concatenated, followed by fully-connected layers. The computational complexity for predicting whether a camera frame will be uploaded using the D-model is given by [41], [42]:

$$O(Z^D) = O(|\mathcal{E}_{n,f}|N_{I,1}^D N_{H,1}^D N_{O,1}^D + |\mathcal{E}_{n,f}|N_{I,2}^D N_{H,2}^D N_{O,2}^D + N_{FL}^D), \quad (11)$$

where  $N_{I,1}^D$  and  $N_{I,2}^D$  denote the input feature dimensions of the nodes,  $N_{H,1}^D$  and  $N_{H,2}^D$  denote the hidden layer dimensions of the two GCNs,  $N_{O,1}^D$  and  $N_{O,2}^D$  denote the dimensions of features extracted by the GCNs, and  $N_{FL}^D$  represents the number of parameters in the fully-connected layers. Given  $F$  camera frames to be uploaded within the subsequent time slot, the computational complexity becomes  $O(Z^D F)$ .

2) *D-State Transition Modeling*: In the detailed modeling, the state transition probability in D-state  $\mathbf{s}_{n,f}^D$  is characterized as follows:

$$P(\mathbf{s}_{n,f+1}^D | \mathbf{s}_{n,f}^D, a_{n,f}) = \frac{P(\mathcal{G}_{n,f+1}^D, \mathcal{G}_{n,f+1}, \mathcal{G}_{n,f}^D, \mathcal{G}_{n,f}, a_{n,f})}{P(\mathcal{G}_{n,f}^D, \mathcal{G}_{n,f}, a_{n,f})} \quad (12a)$$

$$= \frac{P(\mathcal{G}_{n,f+1}^D, \mathcal{G}_{n,f}^D, a_{n,f})P(\mathcal{G}_{n,f+1}, \mathcal{G}_{n,f} | \mathcal{G}_{n,f+1}^D, \mathcal{G}_{n,f}^D, a_{n,f})}{P(\mathcal{G}_{n,f+1}^D, a_{n,f})P(\mathcal{G}_{n,f} | \mathcal{G}_{n,f+1}^D, a_{n,f})} \quad (12b)$$

$$= P(\mathcal{G}_{n,f+1}^D | \mathcal{G}_{n,f}^D, a_{n,f}) \frac{P(\mathcal{G}_{n,f+1}, \mathcal{G}_{n,f})}{P(\mathcal{G}_{n,f})} \quad (12c)$$

$$= P(\mathcal{G}_{n,f+1}^D | \mathcal{G}_{n,f}^D, a_{n,f})P(\mathcal{G}_{n,f+1} | \mathcal{G}_{n,f}), \quad (12d)$$

where (12c) holds due to the fact that newly arrived camera frames in  $\mathcal{X}_{n,f+1}$  in graph  $\mathcal{G}_{n,f+1}$  do not depend on graph of the 3D map (device), i.e.,  $\mathcal{G}_{n,f}^D$ , or action  $a_{n,f}$ . Since  $P(\mathcal{G}_{n,f+1}^D | \mathcal{G}_{n,f}^D, a_{n,f})$  can be calculated according to (6), we need to approximate only  $P(\mathcal{G}_{n,f+1} | \mathcal{G}_{n,f})$  for modeling D-state transitions. We build a GCN  $\phi(\mathcal{G}_{n,f}; \theta)$  with parameters  $\theta$  to approximate the D-state transitions. Note that this GCN is used only for graph link prediction, i.e., output only the weights of edges between camera frames, instead of the whole camera frames [43].

3) *Simplified Modeling*: The detailed modeling involves shared observations, e.g., the shared set of FPs, among historical camera frames. Excessive input data may introduce redundancy and decrease service demand modeling accuracy. For example, the procedure of key frame selection and uploading in the MAR operational mechanism is simple for device pose tracking when the variation in device pose is insignificant [10], [25]. The accuracy of detailed modeling, thus, may be degraded since such redundant input data may complicate the GCN training for approximating policy  $\Pi$  in the detailed modeling.

As an alternative, we propose a simplified data-driven data traffic modeling and define the corresponding state used in the simplified modeling for MAR device  $n$ , abbreviated as *S-state*, as follows:

$$\mathbf{s}_{n,f}^S = [a_{n,i}]_{\forall f-T^w \leq i < f}, \quad \forall n \in \mathcal{N}, f \in \mathcal{F}_n, \quad (13)$$

which consists of the actions conducted for the preceding  $T^w$  camera frames. We build a recurrent neural network with parameters  $\vartheta^S$  to approximate policy  $\Pi$  by using the S-state  $\mathbf{s}_{n,f}^S$  as the input. In this case, the approximation of policy  $\Pi$  can be simplified as a temporal sequence prediction [44]. We optimize parameters  $\vartheta^S$  by minimizing the following loss function:

$$L(\vartheta^S) = \frac{1}{|\Xi|} \sum_{(a_{n,f}, \mathbf{s}_{n,f}^S) \in \Xi} (a_{n,f} - \Pi^S(\mathbf{s}_{n,f}^S; \vartheta^S))^2. \quad (14)$$

Since S-state  $\mathbf{s}_{n,f}^S$  consists of only previous actions, S-state transitions are straightforward and do not require additional modeling.

Let  $N^S$  represent the number of cells of the adopted recurrent neural networks, e.g., LSTM [45]. Given that the input size and the output size of the recurrent neural network are  $T^w$  and 1, respectively, the computational complexity for predicting whether a camera frame will be uploaded using the S-model is  $O(Z^S)$ , where  $Z^S$  is the total number of parameters in a standard recurrent neural network with one cell in each memory block and is given by [45]:

$$Z^S = 4 \times N^S \times N^S + 4 \times N^S \times (T^w + 1). \quad (15)$$

Given  $F$  camera frames to be uploaded within the subsequent time slot, the computational complexity becomes  $O(Z^S F)$ .

**Algorithm 1: Model Switching Method.**

```

1 Input:  $M_n, \delta_n^u, \delta_n^b$ ;
2 Initialization:  $h_{n,1} = 1, m_{n,1} = 0$ ;
3 for  $t \in \mathcal{T}$  do
4   Calculate  $\Delta_{n,t}$  by (16);
5   if  $\Delta_{n,t} > \delta_n^u$  then
6      $h_{n,t} \leftarrow 1; m_{n,t} \leftarrow 0$ ;
7   else
8     if  $\Delta_{n,t} < \delta_n^b$  then
9        $m_{n,t} \leftarrow m_{n,t-1} + 1$ ;
10      if  $m_{n,t} \geq M_n$  then
11         $h_{n,t} \leftarrow 0; m_{n,t} \leftarrow 0$ ;
12      end
13    end
14  end
15 end
16 Output:  $h_{n,t}$ 
    
```

### B. Model Switching Function (MSF)

The model switching function (MSF) is designed to adapt the two data-driven models in the DMF to non-stationary uplink data traffic. We propose an easy-to-use model switching method stemmed from our observations from experiments conducted with an open-source SLAM-based device pose tracking platform [10], [25] on real camera frame sequences [46].

In MAR applications, when variations in the physical environment and device pose are insignificant, the procedure of key frame selection and uploading in the MAR operational mechanism is simple, leading to relatively stable key frame uploading; Conversely, a significant variation, e.g., tracking loss, generally complicates the procedure of key frame selection and uploading, potentially resulting in bursts of key frame uploading. We define the variation in the number of uploaded key frames across time slots, denoted by  $\Delta_{n,t}$ , to guide the model switching for MAR device  $n$ :

$$\Delta_{n,t} = |\tilde{k}_{n,t-1} - \tilde{k}_{n,t-2}|, \forall t \in \mathcal{T}, n \in \mathcal{N}. \quad (16)$$

We refer to  $\Delta_{n,t}$  as the *model switching status* of MAR device  $n$ . Define  $h_{n,t} \in \{0, 1\}$  as an indicator representing the resulting choice of our MSF for MAR device  $n$ , i.e., its currently used modeling method. If  $h_{n,t} = 1$ , the detailed modeling is used in time slot  $t$ ; Otherwise, the simplified modeling is used.

The basic idea of model switching is as follows. A larger value of  $\Delta_{n,t}$  indicates a significant variation in device pose, and the detailed modeling is needed to capture the burst of key frame uploading in this case. Otherwise, the simplified modeling is sufficient. We present the model switching method in Algorithm 1. We introduce parameters  $\delta_n^u, \delta_n^b$ , and  $M_n$  to jointly configure the switching condition for MAR device  $n$ , which avoid frequent switching between the two modeling methods. In lines 5–7, parameter  $\delta_n^u$  determines the threshold indicating that a significant variation occurs, and controls the switching to the detailed modeling, i.e.,  $h_{n,t} = 1$ , when the value of  $\Delta_{n,t}$  exceeds the threshold. In lines 8–14, parameters  $\delta_n^b$  and  $M_n$  jointly control the switching to the simplified modeling, wherein  $\delta_n^b$  represents the threshold for an insignificant variation. We

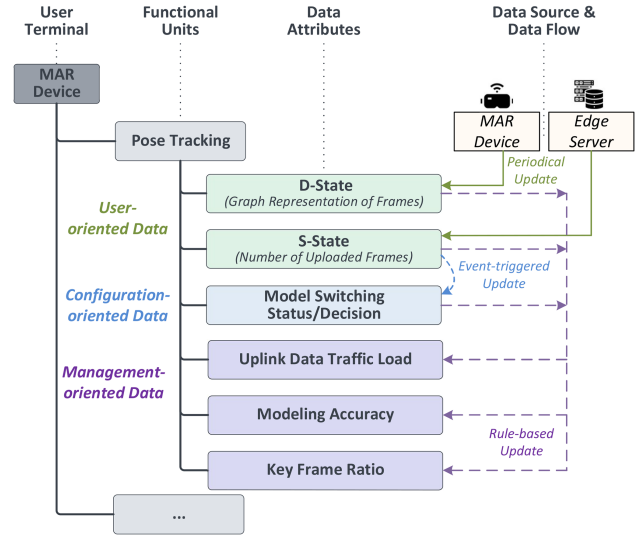


Fig. 3. The hierarchical data model in the MUP.

introduce parameter  $M_n$  to adjust the trigger point for switching from the detailed modeling to the simplified modeling. A larger value of  $M_n$  represents a preference for using detailed modeling in the DMF. The three parameters offer flexibility for user-centric configuration, as they can be configured according to user movements and user-specific psychical environment.

### C. MAR User Profile (MUP)

The MUP includes a structured *data model* consisting of a number of data elements, organized according to a defined schema [14], [47]. As shown in Fig. 3, our defined hierarchical data model is tailored to MAR, which can support user-centric communication service provision. For our data model, we introduce data structure, data attribute, and data update method as follows:

- **Data Structure:** As shown in Fig. 3, at the top level of this hierarchy, there is a “user terminal” representing an individual MAR device such as a pair of smart glasses or a smartphone. We define a number of “functional units” belonging to the MAR device, each relating to a unique functionality, e.g., device pose tracking or 3D rendering, in the MAR application. For each functional unit, e.g., device pose tracking, a set of “data attributes” can be flexibly defined according to its corresponding operation mechanism, thereby enabling MAR user-specific service demand modeling from a data-oriented perspective [14].
- **Data Attribute:** We classify the defined data attributes into three categories. (i) *User-oriented data* support the characterization of service demand for the corresponding MAR device. Illustrated as the green blocks in Fig. 3, key parameters  $s_{n,f}^D, s_{n,f}^S$ , and  $a_{n,f}$  are used by the DMF to model the uplink data traffic. (ii) *Configuration-oriented data* support the effective configuration and execution of UDTOFs, e.g., DMF and MSF, thereby enabling the implementation of the DT itself. Model switching status  $\Delta_{n,t}$  and the choice of switching  $h_{n,t}$  are configured to

determine the switching condition, shown as the blue block in Fig. 3. (iii) *Management-oriented data*, shown as the purple blocks in Fig. 3, facilitate the implementation of user-centric network resource management strategies. In particular, parameters regarding demand modeling inaccuracies, such as  $p_t$ ,  $q_t$ , and  $\lambda_t$ , support the decision making on user-centric network resource allocation, which will be introduced in Section V.

- *Data Update*: The data elements in our data model can be updated. The data sources and data update methods can vary across data attributes. We design three different data update methods for the aforementioned three categories of data attributes. As shown in Fig. 3, the user-oriented data are periodically collected from the MAR device and the edge server. The update of configuration-oriented data is triggered by the update events of user-oriented data. Specifically, the update of model switching status  $\Delta_{n,t}$  in Algorithm 1 is based on the temporal variation in  $a_{n,f}$ , and the change of  $h_{n,t}$  is triggered by the value change of  $\Delta_{n,t}$ . The update of the management-oriented data depends on an update rule tailored to the adopted network management algorithms, which will be presented in Section V.

In contrast to conventional mathematical modeling for capturing service demand, our data model has two potential advantages. First, it has the flexibility to differentiate the data structure and the data attributes for different MAR devices, thereby supporting customized service demand modeling and network resource management. In addition, our data model has a high extensibility to support various functionalities in MAR applications and network management objectives.<sup>3</sup> Due to these advantages, the implicit impact of the MAR operational mechanism on the user-specific service demands can be characterized by using the data model and UDTOFs [7], [48].

## V. DT-BASED USER-CENTRIC COMMUNICATION SERVICE PROVISION

It is well known that neither mathematical models nor data-driven models can guarantee accurate modeling of service demand [7], [9]. Considering the potential modeling inaccuracies resulting from MARLIN, we need to design a robust communication service provision method tailored to data-driven traffic modeling.

Define  $\hat{a}_{n,f}$  as the predicted value of  $a_{n,f}$  by MARLIN. Given the prediction results  $\hat{\mathbf{a}}_{n,t} = [\hat{a}_{n,f}]_{\forall f \in \mathcal{F}_{n,t}}$ , Problem P1 can be transformed as follows:

$$\text{P2: } \min_{\{b_{n,t}\}_{n \in \mathcal{N}, t \in \mathcal{T}}} \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} b_{n,t} \quad (17a)$$

$$\text{s.t. } P(\alpha \tilde{k}_{n,t} \leq T^r r_{n,t} | \hat{\mathbf{a}}_{n,t}) \geq \epsilon, \quad \forall n \in \mathcal{N}, t \in \mathcal{T}, \quad (17b)$$

where constraint (17b) bounds the minimum value of the conditional probability given the predicted results. We solve Problem P2 to achieve robust user-centric communication service provision in both single-user and multi-user scenarios.

<sup>3</sup>This paper only focuses on the communication service provision for a single functional unit, i.e., device pose tracking.

### A. Single-User Communication Service Provision

For the single-user scenario, i.e.,  $|\mathcal{N}| = 1$ , in this subsection, we remove the subscript  $n$  for simplicity. To represent the amount of radio spectrum resource required for timely key frame uploading, we introduce an auxiliary variable  $k_t$ , defined as the number of key frames if the reserved radio spectrum resource  $b_t$  can support timely key frame uploading. As a result, the optimal solution to Problem P2, i.e.,  $b_t^*$ , is given by:

$$b_t^* = \frac{\alpha T^r}{\log(1 + \gamma_t)} k_t^*, \quad \forall t \in \mathcal{T}, \quad (18)$$

where  $k_t^*$  denotes the minimum value of  $k_t$ , given by:

$$k_t^* = \arg \min_{k_t} P \left( \sum_{f \in \mathcal{F}_t} a_f \leq k_t | \hat{\mathbf{a}}_t \right) \geq \epsilon. \quad (19)$$

To determine  $k_t^*$ , we need to first obtain the conditional probability  $P(\mathbf{a}_t = \mathbf{a} | \hat{\mathbf{a}}_t = \hat{\mathbf{a}})$ , i.e., the posterior probability of  $\mathbf{a}_t = \mathbf{a}$ ,  $\forall \mathbf{a} \in \mathcal{A}$  given predicted results  $\hat{\mathbf{a}}_t = \hat{\mathbf{a}}$ ,  $\forall \hat{\mathbf{a}} \in \mathcal{A}$  according to Bayes' theorem:

$$P(\mathbf{a}_t = \mathbf{a} | \hat{\mathbf{a}}_t = \hat{\mathbf{a}}) = \frac{P(\mathbf{a}_t = \mathbf{a})P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})}{\sum_{\mathbf{a} \in \mathcal{A}} P(\mathbf{a}_t = \mathbf{a})P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})}, \quad (20)$$

where  $P(\mathbf{a}_t = \mathbf{a})$  is the prior probability of  $\mathbf{a}_t$ , and  $P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})$  is the likelihood of  $\mathbf{a}_t$  given prediction result  $\hat{\mathbf{a}}_t$ . By updating the value of  $P(\mathbf{a}_t = \mathbf{a} | \hat{\mathbf{a}}_t = \hat{\mathbf{a}})$  based on the update of parameters  $P(\mathbf{a}_t = \mathbf{a})$  and  $P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})$ , we can determine the value of  $k_t^*$ .

The workflow of DT-based user-centric communication service provision for the MAR device is summarized in Algorithm 2. In lines 4–5, at the beginning of time slot  $t$ , the MSF selects a proper modeling method, and the DMF provides the predicted results of the service demand of the MAR device in time slot  $t$ . In lines 6–12, parameters  $P(\mathbf{a}_t = \mathbf{a})$  and  $P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})$  stored in the MUP are used to calculate  $P(\mathbf{a}_t = \mathbf{a} | \hat{\mathbf{a}}_t = \hat{\mathbf{a}})$ , which can help the controller to determine the value of  $k_t^*$ . To indicate the parameters for the detailed and the simplified modeling, we use superscripts “D” and “S”, respectively. In line 13, information on D-states  $s_f^D$  and actions  $a_f$ , collected from the MAR device and the edge server, respectively, is used to update the user-oriented data in the MUP at the end of time slot  $t$ . In lines 14–17, parameters  $P(\mathbf{a}_t = \mathbf{a})$  and  $P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})$  are updated based on the user-oriented data collected in the past  $\tau$  time slots. Parameters  $P(\mathbf{a}_t = \mathbf{a})$  and  $P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})$  in Algorithm 2 represent the management-oriented data defined in Section IV-C, which enhance the robustness of decision making for user-centric communication service provision by accounting for modeling inaccuracies.

Considering the complexity in updating a large number of parameters  $P(\mathbf{a}_t = \mathbf{a})$  and  $P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})$  in the MUP, we analyze the performance of user-centric communication service provision and decrease the number of the used parameters for easily applying Algorithm 2. Without loss of generality, we assume that, given  $\hat{a}_f$ ,  $\forall f \in \mathcal{F}_t$ , random variables  $a_f$ ,  $\forall f \in \mathcal{F}_t$  are independent and identically distributed (i.i.d.), and  $P(a_f | \hat{\mathbf{a}}_t) = P(a_f | \hat{a}_f)$ ,  $\forall f \in \mathcal{F}_t$ .

---

**Algorithm 2: DT-Based User-Centric Communication Service Provision.**


---

```

1 Input:  $\tau$ ;
2 Initialization:  $P^D(\mathbf{a}_t = \mathbf{a}), P^S(\mathbf{a}_t = \mathbf{a}), P^D(\hat{\mathbf{a}}_t = \hat{\mathbf{a}}|\mathbf{a}_t = \mathbf{a}), P^S(\hat{\mathbf{a}}_t = \hat{\mathbf{a}}|\mathbf{a}_t = \mathbf{a}), \forall \mathbf{a} \in \mathcal{A}$ ;
3 for  $t \in \mathcal{T}$  do
4     The MSF obtains  $h_t$  by using Algorithm 1 and
       selects a method from the detailed and simplified
       modeling methods based on  $h_t$ ;
5     The DMF obtains prediction result  $\hat{\mathbf{a}}_t = \hat{\mathbf{a}}$ ;
6     if  $h_t = 1$  then
7         The MUP provides  $P^D(\mathbf{a}_t = \mathbf{a}|\hat{\mathbf{a}}_t)$  by (20);
8         The controller calculates  $k_t^*$  by (19) based
           on  $P^D(\mathbf{a}_t = \mathbf{a}|\hat{\mathbf{a}}_t)$ ;
9     else
10        The MUP provides  $P^S(\mathbf{a}_t = \mathbf{a}|\hat{\mathbf{a}}_t)$  by (20);
11        The controller calculates  $k_t^*$  by (19) based
           on  $P^S(\mathbf{a}_t = \mathbf{a}|\hat{\mathbf{a}}_t)$ ;
12    end
13     $s_f^D, a_f \forall f \in \mathcal{F}_t$  are used to update the
       user-oriented data in the MUP;
14    if  $h_t = 1$  then
15         $P^D(\hat{\mathbf{a}}_{t+1} = \hat{\mathbf{a}}|\mathbf{a}_{t+1} = \mathbf{a}), P^D(\mathbf{a}_{t+1} = \mathbf{a}) \leftarrow$ 
           Estimate based on the user-oriented data
           regarding the detailed modeling during past  $\tau$ 
           time slots;
16    else
17         $P^S(\hat{\mathbf{a}}_{t+1} = \hat{\mathbf{a}}|\mathbf{a}_{t+1} = \mathbf{a}), P^S(\mathbf{a}_{t+1} = \mathbf{a}) \leftarrow$ 
           Estimate based on the user-oriented data
           regarding the simplified modeling during
           past  $\tau$  time slots;
18    end
19 end
20 Output:  $k_t^*$ 

```

---

*Theorem 1:* The probability  $P(\sum_{f \in \mathcal{F}_t} a_f \leq k_t | \hat{\mathbf{a}}_t)$  can be derived in (21), shown at the bottom of this page, which is a non-decreasing function given parameters  $p_t, q_t$ , and  $\lambda_t$ , where  $\hat{A}_t = \sum_{f \in \mathcal{F}_t} \hat{a}_f, F_t = |\mathcal{F}_t|$ ,

$$p_t^{\text{TPR}} = \frac{p_t \lambda_t}{p_t \lambda_t + (1 - q_t)(1 - \lambda_t)}, \quad (22)$$

and

$$p_t^{\text{TNR}} = \frac{q_t(1 - \lambda_t)}{q_t(1 - \lambda_t) + (1 - p_t)\lambda_t}. \quad (23)$$

*Proof:* See Appendix A, available online.  $\square$

Theorem 1 allows us to find the optimal value, i.e.,  $k_t^*$ , based on the closed-form probability  $P(\sum_{f \in \mathcal{F}_t} a_f \leq k_t | \hat{\mathbf{a}}_t)$ , given only three parameters  $p_t = P(a_f = 1 | a_f = 1), q_t = P(\hat{a}_f =$

$0 | a_f = 0)$ , and  $\lambda_t = P(a_f = 1)$  in time slot  $t$ . Thus, parameters  $p_t, q_t$ , and  $\lambda_t$  can replace parameters  $P(\mathbf{a}_t = \mathbf{a})$  and  $P(\hat{\mathbf{a}}_t = \hat{\mathbf{a}} | \mathbf{a}_t = \mathbf{a})$  in Algorithm 2 as management-oriented data maintained by the MUP. By updating the three parameters, MARLIN supports robust user-centric communication service provision for the MAR device.

### B. Multi-User Communication Service Provision

The proposed Algorithm 2 can be extended to multi-user scenarios, i.e.,  $|\mathcal{N}| > 1$ , since the communication service provision for different MAR devices in Problem P2 is independent except for a potential constraint on the total available radio spectrum resource for concurrent uplink transmissions at the BS. In this subsection, we focus on the comparison of user-centric communication service provision and slicing-based communication service provision.

Given the radio spectrum resource reserved for MAR device  $n$  in time slot  $t$ , i.e.,  $b_{n,t}^*$ , in (18), the overall radio spectrum reserved for all MAR devices in time slot  $t$  is given by:

$$B_t^* = \sum_{n \in \mathcal{N}} \frac{\alpha T^r}{\log(1 + \gamma_{n,t})} k_{n,t}^*, \quad \forall t \in \mathcal{T}. \quad (24)$$

In user-centric communication service provision, the reserved radio spectrum  $B_t^*$  can ensure that the probability of timely uploading key frame for each MAR device is larger than the value of  $\epsilon$ .

Since slicing-based communication service provision does not have or use the service demand information of individual MAR devices [7], the service demands of different devices running the same MAR applications are usually assumed to be i.i.d. [35]. Considering a simple scenario where random variables  $\tilde{k}_{n,t}, \forall n \in \mathcal{N}$  are i.i.d. with mean  $\bar{k}_t$  and variance  $\sigma_t^2$ , we can formulate a slicing-based communication service provision problem as follows [11], [12]:

$$\text{P3: } \min_{\{B_t^s\}_{t \in \mathcal{T}}} \sum_{t \in \mathcal{T}} B_t^s \quad (25a)$$

$$\text{s.t. } P\left(\frac{\alpha}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} \tilde{k}_{n,t} \leq T^r r \left(\frac{B_t^s}{|\mathcal{N}|}\right)\right) \geq \epsilon, \quad \forall t \in \mathcal{T}, \quad (25b)$$

where  $B_t^s$  represents the radio spectrum resource reserved for all MAR devices, and (25b) ensures that the probability of timely uploading the number of key frames averaged over all MAR devices is larger than the value of  $\epsilon$ .

*Theorem 2:* The optimal value of  $B_t^s$  in Problem P3 is given by:

$$B_t^{s,*} = \frac{\alpha T^r |\mathcal{N}|}{\log(1 + \bar{\gamma}_t)} \left[ \bar{k}_t + \frac{\Phi^{-1}(\epsilon) \sigma_t^2}{|\mathcal{N}|} \right], \quad (26)$$

---


$$g(k_t; p_t, q_t, \lambda_t) = \sum_{k=0}^{k_t} \sum_{j=\max(0, k - (F_t - \hat{A}_t))}^{\min(\hat{A}_t, k)} \binom{\hat{A}_t}{j} (p_t^{\text{TPR}})^j (1 - p_t^{\text{TPR}})^{\hat{A}_t - j} \binom{F_t - \hat{A}_t}{k - j} (1 - p_t^{\text{TNR}})^{k - j} (p_t^{\text{TNR}})^{F_t - \hat{A}_t - k + j}. \quad (21)$$

TABLE I  
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
$F$	10 frames	$T^r$	0.02 second
$\alpha$	5 Mbits	$\gamma_{n,t}$	15 dB

where  $\Phi^{-1}(\cdot)$  denotes the quantile function of the standard normal distribution,  $\bar{\gamma}_t$  represents the average SNR over all MAR devices, and  $\lceil \cdot \rceil$  is the ceiling function.

*Proof:* See Appendix B, available online.  $\square$

Theorem 2 allows us to obtain the overall radio spectrum reserved for MAR devices in time slot  $t$ , i.e.,  $B_t^{s,*}$ , in slicing-based communication service provision. Comparing (24) and (26), we have two observations. First, the amount of radio spectrum resource reserved for an individual MAR device in the slicing-based communication service provision, i.e.,  $B_t^{s,*}/|\mathcal{N}|$ , approaches  $\alpha T^r \bar{k}_t / \log(1 + \bar{\gamma}_t)$ , which is less than that in user-centric communication service provision when  $\sum_{n \in \mathcal{N}} k_{n,t}^* < \bar{k}_t |\mathcal{N}|$ . Second, slicing-based communication service provision is less sensitive to the variances of individual traffic demands of MAR devices. As shown in Theorem 2, the value of  $\Phi^{-1}(\epsilon) \sigma_t^2 / |\mathcal{N}|$  decreases with the increase of  $|\mathcal{N}|$ . In contrast, the distribution of  $\bar{k}_{n,t}$  significantly influences user-centric communication service provision, and, thus, the user-centric communication service provision is more robust in adapting to user-specific service demands. Consequently, the user-centric communication service provision for MAR can trade off resource consumption for the ability in adapting to user-specific service demand variations.

## VI. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our DT-based, user-centric approach to communication service provision in edge-assisted MAR. Our focus is on the impact of radio spectrum resource management on the timeliness of key frame uploading from the communication and networking perspective.

### A. Simulation Settings

1) *Settings of Considered Scenario and DT:* In our simulation, we use 218 camera frame sequences, corresponding to different user movements in various environments, from the InteriorNet dataset [46] and conduct device pose tracking for the MAR device using the open-source ORB-SLAM3 platform [10]. We use a resource block (RB) as the base unit for radio spectrum resource, each of which is 180 kHz wide (12 subcarriers) in radio spectrum bandwidth and 0.5 ms in time. Other important parameter settings are listed in Table I.

To establish a DT for each MAR device, we first design the DMF in the DT using deep learning techniques. Specifically, for detailed modeling within the DMF, we employ a DNN with parameters  $\vartheta^D$ . This network consists of one graph convolution (GC) layer for graph feature extraction, followed by three fully connected (FC) layers with output dimensions of 416, 64, and 8, respectively. To approximate D-state transitions, we use another

DNN with parameters  $\theta$ , comprising three GC layers followed by three FC layers with output dimensions of 256, 64, and 32, respectively. The input node features of graphs  $\mathcal{G}_{n,f}$  and  $\mathcal{G}_{n,f}^D$  are derived from camera frames processed by the ResNet-50 DNN. Each GC or FC layer is followed by a batch normalization layer and a ReLU activation function. For simplified modeling in the DMF, we adopt a DNN with parameters  $\vartheta^S$ , which includes one LSTM layer with 50 neurons, followed by two FC layers with output dimensions of 32 and 16, respectively.

Second, we design the MSF in the DT by concentrating on three key parameters that enable the adaptation of the DMF to the non-stationary uplink data traffic of each MAR device. We set the trigger point  $M_n$  and the lower variation threshold  $\delta_n^b$  to 3 and 2, respectively, to jointly determine the conditions for switching from detailed modeling to simplified modeling in the DMF. We set the upper variation threshold  $\delta_n^u$  to 4 to determine the condition for switching from simplified modeling back to detailed modeling.

2) *Benchmark Approaches:* We compare the proposed DT-based user-centric communication service provision with the state-of-the-art slicing-based communication service provision, a key innovation for service-oriented network management in 5G [35]. In addition, we adopt the following prevalent service demand modeling approaches, including mathematical model-based and data-driven approaches, as benchmark [9], [49], [50], [51]:

- *Poisson regression:* The number of key frames for uploading in each time slot is assumed to follow a Poisson distribution, a mathematical model widely applied in network resource management. The parameters of the Poisson distribution are estimated based on historical information.
- *LSTM-based neural network:* Following the simplified modeling in the DMF, an LSTM-based recurrent neural network is pre-trained and employed to predict the number of key frames to be uploaded in each time slot.
- *N-BEATS:* Without relying on recurrent neural networks or attention-based mechanisms, N-BEATS uses a stack of fully-connected blocks to capture long-term components for time-series prediction.
- *Transformer-based neural network:* Transformer-based neural networks leverage self-attention mechanisms to enable parallel processing and capture direct relationships across all inputs, making them highly effective for time-series analysis and prediction.
- *Informer:* Transformer-based neural networks designed for long sequence time-series forecasting, featuring a compute-efficient self-attention mechanism and a scalable architecture to handle long outputs.

### B. Performance of Service Demand Modeling

We first compare the service demand modeling performance of MARLIN with that of a mathematical model-based approach. As shown in Fig. 4, compared to the ‘‘Predicted (Poisson Model)’’ approach, we observe that the predicted values by MARLIN more closely match the actual non-stationary uplink data traffic, particularly during bursts in uplink data traffic that

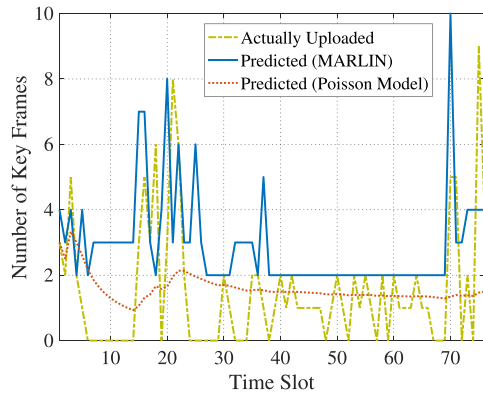
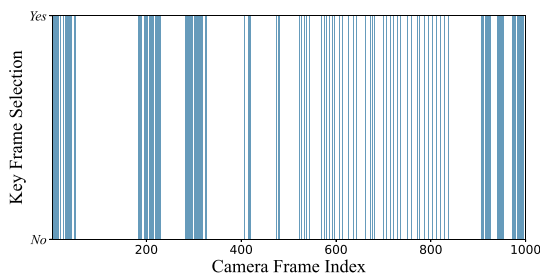
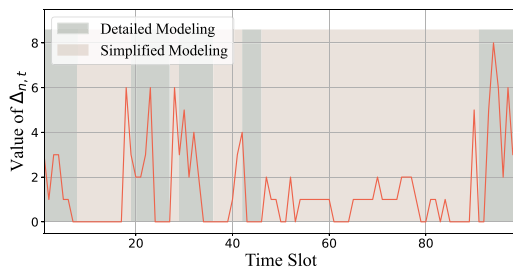


Fig. 4. Performance comparison between MARLIN and Poisson model-based approach.



(a) Non-stationary uplink data traffic due to the MAR operational mechanism of device pose tracking.



(b) The change of model switching status.

Fig. 5. Performance of model switching in adapting to data traffic variations.

may result from device tracking loss or significant changes in the physical environment. In contrast, the widely-used Poisson model captures only large time-scale service demand patterns, such as average values, but falls short in characterizing the detailed temporal patterns in service demand, e.g., fluctuations caused by user movements or the SLAM-specific camera frame uploading mechanism. Therefore, MARLIN outperforms the Poisson model-based approach in terms of representation power. In addition, MARLIN can switch between the detailed and the simplified data-driven modeling according to variations in the number of uploaded key frames, thereby reducing input data redundancy in the detailed modeling.

In Fig. 5, we show the adaptivity of MARLIN to the non-stationary uplink data traffic for key frame uploading. For a camera sequence, the status whether each camera frame is selected as a key frame for uploading or not is shown in Fig. 5(a). It is hard to identify a regular and stationary pattern at first glance since

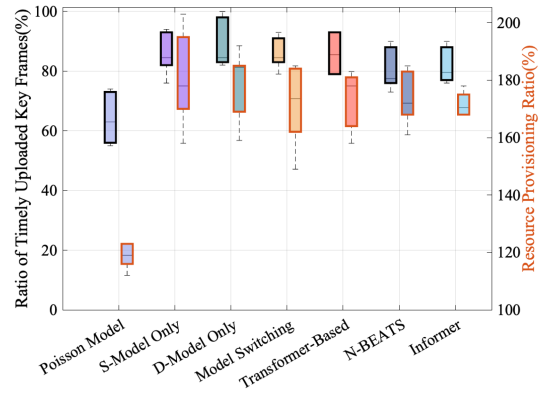


Fig. 6. Radio spectrum resource utilization of the designed model switching method in MARLIN.

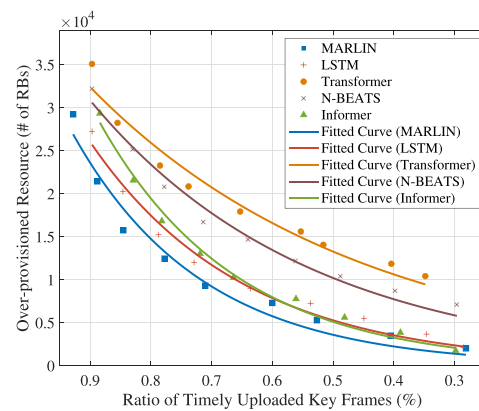


Fig. 7. Performance comparison between the proposed DT-based approach and benchmark data-driven approaches.

the complex operational mechanism of device pose tracking is adopted to address uncertain user movements. We plot the variations in model switching status, i.e.  $\Delta_{n,t}$ , along with the corresponding choice of the detailed and the simplified modeling methods made by our MSF, in Fig. 5(b). We can observe that the detailed modeling method is used during bursts in uplink data traffic, while the simplified modeling method is used in scenarios with less variations in uplink data traffic. This is because MARLIN targets accurately approximating the non-stationary MAR service demand by using contextual knowledge, e.g., the MAR operational mechanism of device pose tracking in dealing with uncertain user movements.

### C. Performance of DT-Based User-Centric Communication Service Provision

In this subsection, we compare the performance of our proposed approach with that of both mathematical model-based and data-driven approaches across 15 camera sequences. Since our objective is to minimize the communication resource consumption while ensuring timely camera frame uploading, we evaluate all approaches from two perspectives: 1) the ratio of timely uploaded key frames (TUKF) and 2) the radio spectrum resource provisioning (RP) ratio, defined as the amount of provisioned radio spectrum resource over the amount of required radio spectrum resource.

TABLE II  
COMPARISON RESULTS ON CAMERA FRAME SEQUENCES WITH DIFFERENT USER MOVEMENT PATTERNS

Approach	3FO4K7I2Q0PG original_1_1		3FO4K8S8DIKV original_3_3		3FO4JWLHP0TI original_7_7		3FO4JU3UAANP random_3_3		3FO4JTFDJQ38 random_7_7	
	TUKF(%)	RP(%)	TUKF(%)	RP(%)	TUKF(%)	RP(%)	TUKF(%)	RP(%)	TUKF(%)	RP(%)
<b>Our Proposed</b>	91	171	93	166	83	162	86	182	83	188
<b>Poisson Model-based</b>	56	116	59	116	73	138	55	112	74	122
<b>LSTM-based</b>	94	182	91	195	86	170	82	176	83	203
<b>N-BEATS</b>	91	185	91	179	76	161	82	173	79	168
<b>Transformer-based</b>	92	183	91	174	79	176	85	184	81	187
<b>Informer</b>	88	175	90	178	76	160	79	169	77	176

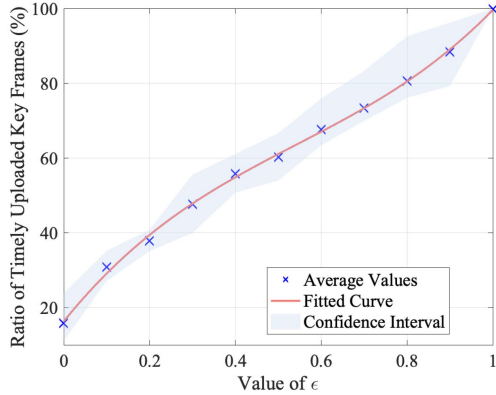


Fig. 8. The value of  $\epsilon$  versus the actual ratio of timely uploaded key frames.

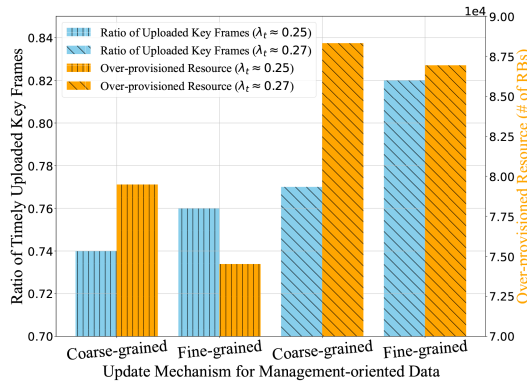


Fig. 9. Performance of the fine-grained and the coarse-grained data update methods in MARLIN.

In Fig. 6, we show the performance when only employing the simplified model (labeled as “S-Model Only”), the performance when only employing the detailed model (labeled as “D-Model Only”), and the performance when switching between two modeling methods (labeled as “Model Switching”), respectively. We can make two important observations. First, “Poisson Model” is inferior to “S-Model Only”, “D-Model Only”, “Transformer-based”, “N-BEATS”, and “Informer” in both aspects since the data-driven modeling methods based on DNNs offer greater representation power. Second, compared to other data-driven benchmark methods, our approach achieves lower RP ratio without sacrificing the performance of TUKF significantly. This demonstrates the advantage of our approach in communication

resource efficiency since the S- and D-models can be adaptively selected in different scenarios via the MSF. Compared to using only a data-driven modeling method, a hidden benefit of “Model Switching” lies in its flexibility to balance the modeling accuracy and the inference costs associated with establishing and updating the models in the MUP, by judiciously selecting from the lightweight S-model and the compute-intensive D-model.

In Table II, we show the performance of these communication service provision approaches on different sub-datasets, i.e., camera frame sequences. In the InteriorNet dataset [46], camera frame sequences are broadly classified into “original” and “random” categories. The former is generated based on human user movement patterns, while the latter is a randomly generated continuous trajectory, where “1–1”, “3–3”, and “7–7” represent different user movement patterns, including velocity and acceleration. In terms of TUKF, we observe that some benchmark approaches may outperform our proposed approach in certain scenarios. For example, the LSTM-based and Transformer-based approaches achieve 3% and 1% higher TUKF, respectively, on the “original\_1\_1” sub-dataset, enabling more timely uploading of key frames. However, this improvement comes at the cost of approximately 12% higher radio spectrum resource consumption. This observation indicates that our approach achieves a more efficient balance between TUKF and RP by dynamically switching between different data-driven models, demonstrating superior adaptability to non-stationary uplink data traffic. Another observation is that the performance gain of our proposed approach on the “original” sub-datasets is greater than that on the “random” sub-datasets. Our hypothesis is that the “original” sub-datasets demonstrate phased fluctuations in human user movement patterns, as shown in Fig. 5(a), in which the proposed “Model Switching” excels. This demonstrates the potential of MARLIN to capture user-specific service demand in future 6G networks.

In Fig. 7, we compare the communication service provision performance of the DT-based approach with that of benchmark data-driven approaches (labeled as “LSTM”, “Transformer”, “N-BEATS”, and “Informer”) in terms of the amount of over-provisioned radio spectrum resource (in RBs) and TUKF. We can observe that our DT-based approach not only reduces over-provisioning but also ensures the timeliness of key frame uploading of the MAR device, for two reasons. First, switching between different data-driven models can effectively capture the joint impact of the complex SLAM-based camera frame uploading mechanism and uncertain user movements on user-specific service demands. Second, our user-centric

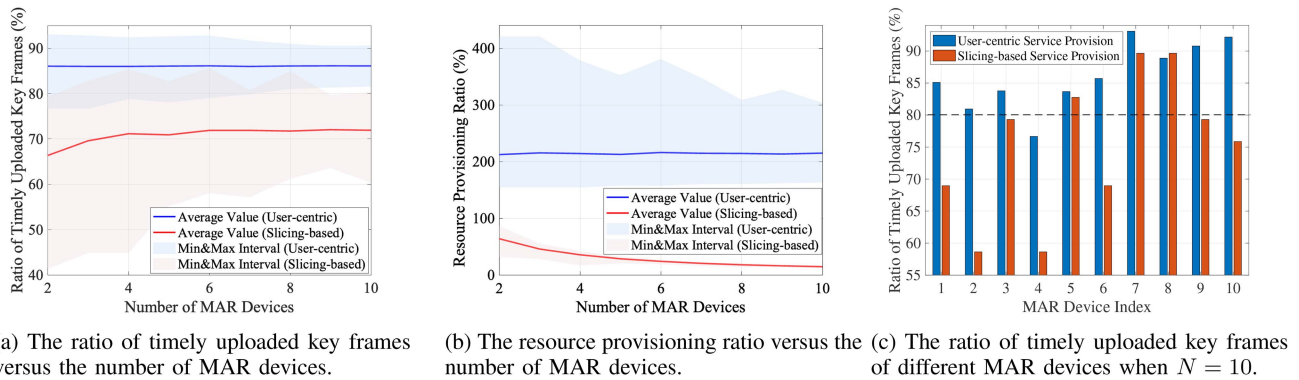


Fig. 10. Performance comparison between user-centric and slicing-based communication service provision.

communication service provision, guided by theoretical analysis benefit from advanced data-driven techniques while accounting for data-driven modeling inaccuracies, thereby effectively mitigating the over-provisioning of radio spectrum resource.

In Fig. 8, given different values of  $\epsilon$ , we plot the actual ratios of timely uploaded key frames with blue markers, each of which is averaged over 6 different camera sequences. The red curve represents a fitted curve, and the light blue shaded area represents the confidence interval between the minimum and maximum values. Given any  $\epsilon$ , the radio spectrum resource is reserved for the MAR device based on (21). We can observe that the actual ratio of timely uploaded key frames closely approaches the value of  $\epsilon$ , despite a slight difference. This is because the result in (21) is based on the assumption that, given  $\hat{a}_f, \forall f \in \mathcal{F}_t$ , random variables  $a_f, \forall f \in \mathcal{F}_t$  are i.i.d., and  $P(a_f|\hat{a}_t) = P(a_f|\hat{a}_f), \forall f \in \mathcal{F}_t$ . However, when the assumption holds, the cost of updating management-oriented data can be reduced. MARLIN can use only three parameters, i.e.,  $p_t$ ,  $q_t$ , and  $\lambda_t$ , as management-oriented data to conduct robust communication service provision with a slight difference in the performance of communication service provision.

In Fig. 9, we compare the performance of fine-grained and coarse-grained data management in MARLIN. As mentioned in Algorithm 2, we can separately update the management-oriented data (labeled as “fine-grained”), e.g.  $p_t$ ,  $q_t$ , and  $\lambda_t$ , for the S-model and D-model. In contrast, we do not differentiate the collection of management-oriented data for the S-model and D-model for the benchmark (labeled as “coarse-grained”). We observe that, given two different values of  $\lambda_t$ , the “fine-grained” method outperforms the “coarse-grained” method in terms of the ratio of timely transmitted key frames and the amount of over-provisioned radio spectrum resource. This is because the “fine-grained” data update method empowers user-centric communication service provision by accurately characterizing the potential modeling inaccuracies of the detailed and the simplified modeling methods, especially when the two modeling methods suit different user movement patterns.

In Fig. 10, we compare the user-centric and the slicing-based communication service provision in terms of the TUKF and the RP. In Fig. 10(a) and (b), we plot the ratios as blue and red lines and illustrate the range between the minimum and maximum values using shaded blue and red areas, respectively,

where each value is averaged over multiple MAR devices and 1,000 simulation runs. In Fig. 10(a), we observe that, with the setting of  $\epsilon = 0.8$ , the user-centric communication service provision significantly outperforms the slicing-based communication service provision in terms of the TUKF. This is because the proposed approach makes communication service provision decisions to meet the delay requirement of each MAR device based on user-specific service demands estimated by using MARLIN, rather than on averaged service demands across multiple MAR devices. In the case of  $N = 10$  and  $\epsilon = 0.8$ , the detailed performance differences across MAR devices are shown in Fig. 10(c). In contrast to slicing-based communication service provision used for 5G, user-centric communication service provision ensures that nearly all MAR devices timely upload their key frames by adapting to user-specific service demands.

In Fig. 10(b), the average amount of radio spectrum resource reserved for an MAR device in the user-centric communication service provision is less than that in slicing-based communication service provision. This demonstrates that user-centric communication service provision can adapt to user-specific service demand variances. Meanwhile, we observe that slicing-based communication service provision is less sensitive to the variances of individual service demands of MAR devices since the value of  $\Phi^{-1}(\epsilon)\sigma_t^2/|\mathcal{N}|$  decreases as the number of devices increases. Both observations align with Theorem 2 and our theoretical analysis in Section V.

## VII. CONCLUSION AND FUTURE WORK

In this paper, we have developed a novel user DT, i.e., MARLIN, and a DT-based user-centric communication service provision approach to support edge-assisted device pose tracking in MAR. In MARLIN, the well-structured data model enables a flexible configuration of user-specific data attributes and personalized data update. The two designed DT operation functions jointly enable an adaptive switching between two data-driven modeling methods for accurately modeling non-stationary data traffic. Trace-driven simulation results have demonstrated the effectiveness and robustness of our user-centric communication service provision approach in coping with modeling inaccuracies. MARLIN provides an extensible prototype of user DT for capturing the impact of user-specific service demands and

reveals the promising potential of fine-grained data management in facilitating user-centric communication service provision in the 6G. In the future, we plan to investigate a hybrid user-centric and slicing-based communication service provision scheme to enhance the quality of experience for MAR users while reducing the cost of data management incurred by user DTs.

## REFERENCES

- [1] X. Shen et al., "Toward immersive communications in 6G," *Front. Comput. Sci.*, vol. 4, 2023, Art. no. 1068478.
- [2] Y. Jin, J. Liu, F. Wang, and S. Cui, "Eubiblio: Edge assisted multi-user 360-degree video streaming," *IEEE Internet Things J.*, vol. 10, no. 17, pp. 15408–15419, Sep. 2023.
- [3] M. Huzafa et al., "ILLIXR: Enabling end-to-end extended reality research," in *Proc. IEEE Int. Symp. Workload Characterization*, 2021, pp. 24–38.
- [4] J. Chen, K. Ramakrishnan, A. Dhakazl, and X. Ran, "Networked architectures for localization-based multi-user augmented reality," *IEEE Commun. Mag.*, vol. 61, no. 12, pp. 104–110, Dec. 2023.
- [5] H. Wang, B. Kim, J. Xie, and Z. Han, "LEAF+AIO: Edge-assisted energy-aware object detection for mobile augmented reality," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5933–5948, Oct. 2023.
- [6] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1784–1797, Mar. 2018.
- [7] X. Shen, J. Gao, W. Wu, M. Li, C. Zhou, and W. Zhuang, "Holistic network virtualization and pervasive network intelligence for 6G," *IEEE Commun. Surv. Tuts.*, vol. 24, no. 1, pp. 1–30, First Quarter 2022.
- [8] X. Ran, C. Slocum, Y.-Z. Tsai, K. Apicharttrisorn, M. Gorlatova, and J. Chen, "Multi-user augmented reality with communication efficient and spatially consistent virtual objects," in *Proc. 16th Int. Conf. Emerg. New. Experiments Technol.*, 2020, pp. 386–398.
- [9] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surv. Tuts.*, vol. 22, no. 2, pp. 905–929, Second Quarter 2020.
- [10] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.
- [11] X. Li, R. Xiao, M. Pan, and N. Zhao, "Risk-averse investment strategy for MEC service provisioning: A data-driven distributionally robust solution," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 24148–24160, Dec. 2022.
- [12] L. Li et al., "Data-driven optimization for cooperative edge service provisioning with demand uncertainty," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4317–4328, Mar. 2021.
- [13] F. Wu et al., "Characterizing Internet card user portraits for efficient churn prediction model design," *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1735–1752, Feb. 2024.
- [14] C. Zhou, S. Hu, J. Gao, X. Huang, W. Zhuang, and X. Shen, "User-centric immersive communications in 6G: A data-oriented framework via digital twin," *IEEE Wireless Commun.*, vol. 32, no. 3, pp. 122–129, Jun. 2025.
- [15] C. Zhou, J. Gao, M. Li, N. Cheng, X. Shen, and W. Zhuang, "Digital twin-based 3-D map management for edge-assisted device pose tracking in mobile AR," *IEEE Internet Things J.*, vol. 11, no. 10, pp. 17812–17826, May 2024.
- [16] R. Sun, N. Cheng, C. Li, F. Chen, and W. Chen, "Knowledge-driven deep learning paradigms for wireless network optimization in 6G," *IEEE Netw.*, vol. 38, no. 2, pp. 70–78, Mar. 2024, doi: [10.1109/MNET.2024.3352257](https://doi.org/10.1109/MNET.2024.3352257).
- [17] J. Linowes and K. Babilinski, *Augmented Reality for Developers: Build Practical Augmented Reality Applications With Unity, ARCore, ARKit, and Vuforia*. Birmingham, U.K.: Packt Publishing Ltd, 2017.
- [18] S. Hu, M. Li, J. Gao, C. Zhou, and X. Shen, "Adaptive device-edge collaboration on DNN inference in AIoT: A digital twin-assisted approach," *IEEE Internet Things J.*, vol. 11, no. 7, pp. 12893–12908, Apr. 2024.
- [19] B. Han, P. Pathak, S. Chen, and L.-F. C. Yu, "CoMIC: A collaborative mobile immersive computing infrastructure for conducting multi-user XR research," *IEEE Netw.*, vol. 37, no. 6, pp. 124–131, Nov. 2023.
- [20] K. Khosoussi, M. Giamou, G. S. Sukhatme, S. Huang, G. Dissanayake, and J. P. How, "Reliable graphs for SLAM," *Int. J. Robot. Res.*, vol. 38, no. 2/3, pp. 260–298, 2019.
- [21] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.
- [22] C. Chou and C. Chou, "Efficient and accurate tightly-coupled visual-lidar SLAM," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 14509–14523, Sep. 2022.
- [23] L. Carlone and S. Karaman, "Attention and anticipation in fast visual-inertial navigation," *IEEE Trans. Robot.*, vol. 35, no. 1, pp. 1–20, Feb. 2019.
- [24] K. Chen, T. Li, H.-S. Kim, D. E. Culler, and R. H. Katz, "MARVEL: Enabling mobile augmented reality with low energy and low latency," in *Proc. 16th ACM Conf. Embedded Netw. Sensor Syst.*, 2018, pp. 292–304.
- [25] A. J. Ben Ali, M. Kouroushli, S. Semenova, Z. S. Hashemifar, S. Y. Ko, and K. Dantu, "Edge-SLAM: Edge-assisted visual simultaneous localization and mapping," *ACM Trans. Embedded Comput. Syst.*, vol. 22, no. 1, pp. 1–31, 2022.
- [26] Y. Chen, H. Inaltekin, and M. Gorlatova, "AdaptSLAM: Edge-assisted adaptive SLAM with resource constraints via uncertainty minimization," in *Proc. IEEE Conf. Comput. Commun.*, 2023, pp. 1–10.
- [27] A. Dhakal, X. Ran, Y. Wang, J. Chen, and K. Ramakrishnan, "SLAM-share: Visual simultaneous localization and mapping for real-time multi-user augmented reality," in *Proc. 18th Int. Conf. Emerg. Netw. Experiments Technol.*, 2022, pp. 293–306.
- [28] P. Ren et al., "Edge AR X5: An edge-assisted multi-user collaborative framework for mobile web augmented reality in 5G and beyond," *IEEE Trans. Cloud Comput.*, vol. 10, no. 4, pp. 2521–2537, Fourth Quarter 2022.
- [29] S.-C. Lin, P. Wang, I. F. Akyildiz, and M. Luo, "Towards optimal network planning for software-defined networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 12, pp. 2953–2967, Dec. 2018.
- [30] S.-H. Wu, C.-H. Ko, and H.-L. Chao, "On-demand coordinated spectrum and resource provisioning under an open C-RAN architecture for dense small cell networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 1, pp. 673–688, Jan. 2024.
- [31] L. Zhang and N. Ansari, "Latency-aware IoT service provisioning in UAV-aided mobile-edge computing networks," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10573–10580, Oct. 2020.
- [32] Y. Qu et al., "Service provisioning for UAV-enabled mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3287–3305, Nov. 2021.
- [33] X. Zhang, J. Wang, and H. V. Poor, "AoI-driven statistical delay and error-rate bounded QoS provisioning for mURLLC over UAV-multimedia 6G mobile networks using FBC," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3425–3443, Nov. 2021.
- [34] Y. Qu et al., "Resilient service provisioning for edge computing," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2255–2271, Feb. 2023.
- [35] Q. Ye, W. Zhuang, S. Zhang, A.-L. Jin, X. Shen, and X. Li, "Dynamic radio resource slicing for a two-tier heterogeneous wireless network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9896–9910, Oct. 2018.
- [36] Z. Jia, M. Sheng, J. Li, D. Zhou, and Z. Han, "VNF-based service provision in software defined LEO satellite networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6139–6153, Sep. 2021.
- [37] Y. Qiu, J. Liang, V. C. Leung, X. Wu, and X. Deng, "Online reliability-enhanced virtual network services provisioning in fault-prone mobile edge cloud," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7299–7313, Sep. 2022.
- [38] G. Wang, S. Zhou, S. Zhang, Z. Niu, and X. Shen, "SFC-based service provisioning for reconfigurable space-air-ground integrated networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 7, pp. 1478–1489, Jul. 2020.
- [39] M. Belgiovine, K. Sankhe, C. Bocanegra, D. Roy, and K. R. Chowdhury, "Deep learning at the edge for channel estimation in beyond-5G massive MIMO," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 19–25, Apr. 2021.
- [40] R. Atawia, H. Abou-Zeid, H. S. Hassanein, and A. Noureldin, "Joint chance-constrained predictive resource allocation for energy-efficient video streaming," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1389–1404, May 2016.
- [41] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017, [arXiv:1609.02907](https://arxiv.org/abs/1609.02907).
- [42] Y. Dai et al., "A survey of graph-based resource management in wireless networks - Part II: Learning approaches," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 4, pp. 2101–2122, Aug. 2025.
- [43] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: A comprehensive review," *Comput. Social Netw.*, vol. 6, no. 1, pp. 1–23, 2019.
- [44] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [45] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," 2014, *arXiv:1402.1128*.
- [46] W. Li et al., "InteriorNet: Mega-scale multi-sensor photo-realistic indoor scenes dataset," in *Proc. Brit. Mach. Vis. Conf.*, 2018, pp. 1–13.
- [47] X. Ma, Q. Zeng, H. Chi, and L. Luo, "No more companion Apps hacking but one dongle: Hub-based blackbox fuzzing of IoT firmware," in *Proc. 21st Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2023, pp. 205–218.
- [48] X. Ma, L. Luo, and Q. Zeng, "From one thousand pages of specification to unveiling hidden bugs: Large language model assisted fuzzing of matter IoT devices," in *Proc. 33rd USENIX Conf. Secur. Symp.*, 2024, Art. no. 268.
- [49] Z. Wang, X. Su, and Z. Ding, "Long-term traffic prediction based on LSTM encoder-decoder architecture," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6561–6571, Oct. 2021.
- [50] B. N. Oreshkin, D. Carпов, N. Chapados, and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," in *Proc. Int. Conf. Learn. Representations*, 2020, pp. 1–31.
- [51] H. Zhou et al., "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 11106–11115.



**Conghao Zhou** (Member, IEEE) received the BEng degree from Northeastern University, Shenyang, China, the MSc degree from the University of Illinois Chicago, Chicago, IL, USA, and the PhD degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada. He was a postdoctoral fellow with the University of Waterloo. He is currently a professor with the School of Telecommunications Engineering, Xidian University, China. His current research interests include immersive communications, AI for networking, and space-air-ground integrated networks.



**Jie Gao** (Senior Member, IEEE) received the PhD degree in electrical engineering from the University of Alberta, Edmonton, AB, Canada, in 2014. He has worked with Toronto Metropolitan University, Toronto, ON, the University of Waterloo, Waterloo, ON, and Marquette University, Milwaukee, WI, USA, as a postdoctoral fellow, research associate, and assistant professor, respectively. He is currently an assistant professor with the School of Information Technology, Carleton University, Ottawa, ON. His research interests include machine learning for

communications and networking, IoT and industrial IoT solutions, and B5G/6G networks. He is serving or has served as an editor of the *IEEE Transactions on Cognitive Communications and Networking* and *IEEE Open Journal of the Communications Society*, a publicity co-chair for IEEE VTC2025-Fall and IEEE MetaCom 2023, a publication co-chair for ACM/IEEE IoTDI 2022, and a co-chair for symposia, tracks, or workshops at various IEEE conferences, including IEEE GLOBECOM, VTC, ICC, and INFOCOM. He received the IEEE VTS OJVT Best Paper Award 2025, the IEEE Best Land Transportation Paper Award 2024, and the Wisconsin Space Grant Consortium Early-Stage Investigator Grant Award 2021.



**Shisheng Hu** (Member, IEEE) received the BEng and MSc degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2018 and 2021, respectively, and the PhD degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2021. He is currently a postdoctoral fellow with the University of Waterloo. His research interests include AI for wireless networks and networking for AI.



**Nan Cheng** (Senior Member, IEEE) received the BE and MS degrees from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively, and the PhD degree from the Department of Electrical and Computer Engineering, University of Waterloo, in 2016. From 2017 to 2019, he was a post-doctoral fellow with the Department of Electrical and Computer Engineering, University of Toronto. He is currently a professor with the State Key Laboratory of ISN and School of Telecommunication Engineering, Xidian University, Shaanxi, China. His current research focuses on B5G/6G, space-air-ground integrated network, Big Data in vehicular networks, and self-driving systems. His research interests also include performance analysis, MAC, opportunistic communication, and application of AI for vehicular networks.



**Weihua Zhuang** (Fellow, IEEE) received the BSc and MSc degrees in electrical engineering from Dalian Maritime University, China, and the PhD degree in electrical engineering from the University of New Brunswick, Canada. Since 1993, she has been a faculty member with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where she is currently a University professor and a University research chair in wireless communication networks. Her current research interests include network architecture, algorithms and protocols, and service provisioning in future communication systems. She was a recipient of the Distinguished Career Award in 2021 from the IEEE Vehicular Technology Society (VTS), the R.A. Fessenden Award in 2021 from IEEE Canada, and the Technical Recognition Award in Ad Hoc and Sensor Networks in 2017 from the IEEE Communications Society. She is a fellow of the Royal Society of Canada (RSC), the Canadian Academy of Engineering (CAE), and the Engineering Institute of Canada (EIC). She is an elected member of the VTS Board of Governors. She was the VTS president (2023–2024) and editor-in-chief of the *IEEE Transactions on Vehicular Technology* (2007–2013), an editor of the *IEEE Transactions on Wireless Communications* (2005–2009), and an IEEE Communications Society distinguished lecturer (2008–2011).



**Xuemin (Sherman) Shen** (Fellow, IEEE) received the PhD degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is a University professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular networks. He is a registered professional engineer of Ontario, Canada, an Engineering Institute of Canada fellow, a Canadian Academy of Engineering fellow, a Royal Society of Canada fellow, a Chinese Academy of Engineering foreign member, and an international fellow of the Engineering Academy of Japan. He received "West Lake Friendship Award" from Zhejiang Province in 2023, President's Excellence in Research from the University of Waterloo in 2022, the Canadian Award for Telecommunications Research from the Canadian Society of Information Theory (CSIT) in 2021, the R.A. Fessenden Award in 2019 from IEEE, Canada, Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society (ComSoc), and Technical Recognition Award from Wireless Communications Technical Committee (2019) and AHSN Technical Committee (2013). He has also received the Excellent Graduate Supervision Award in 2006 from the University of Waterloo and the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He serves/served as the general chair for the 6G Global Conference'23, and ACM Mobihoc'15, Technical Program Committee chair/co-chair for IEEE Globecom'24, 16 and 07, IEEE Infocom'14, IEEE VTC'10 Fall, and the chair for the IEEE ComSoc Technical Committee on Wireless Communications. He is the past president of the IEEE ComSoc, the vice president for Technical & Educational Activities, vice president for publications, member-at-large on the Board of Governors, chair of the Distinguished Lecturer Selection Committee, and member of IEEE Fellow Selection Committee of the ComSoc. He served as the editor-in-chief of the *IEEE Internet of Things Journal*, *IEEE Network*, and *IET Communications*.