

Experience-Centric Resource Management in ISAC Networks: A Digital Agent-Assisted Approach

Xinyu Huang[✉], *Member, IEEE*, Yixiao Zhang[✉], *Graduate Student Member, IEEE*, Yingying Pei[✉], *Member, IEEE*, Jianzhe Xue[✉], *Graduate Student Member, IEEE*, and Xuemin Shen[✉], *Fellow, IEEE*

Abstract—In this paper, we propose a digital agent (DA)-assisted resource management scheme for enhanced user quality of experience (QoE) in integrated sensing and communication (ISAC) networks. Particularly, user QoE is a comprehensive metric that integrates quality of service (QoS), user behavioral dynamics, and environmental complexity. The novel DA module includes a user status prediction model, a QoS factor selection model, and a QoE fitting model, which analyzes historical user status data to construct and update user-specific QoE models. Users are clustered into different groups based on their QoE models. A Cramér-Rao bound (CRB) model is utilized to quantify the impact of allocated communication resources on sensing accuracy. A joint optimization problem of communication and computing resource management is formulated to maximize long-term user QoE while satisfying CRB and resource constraints. A two-layer data-model-driven algorithm is developed to solve the formulated problem, where the top layer utilizes an advanced deep reinforcement learning algorithm to make group-level decisions, and the bottom layer uses convex optimization techniques to make user-level decisions. Simulation results based on a real-world dataset demonstrate that the proposed DA-assisted resource management scheme outperforms benchmark schemes in terms of user QoE.

Index Terms—Cramér-Rao bound (CRB), digital agent, integrated sensing and communication (ISAC), quality of experience (QoE), resource management.

I. INTRODUCTION

INTEGRATED sensing and communication (ISAC) is expected to become a cornerstone of sixth-generation (6G) networks, which can seamlessly integrate communication and sensing functionalities within a single communication network infrastructure [1], [2]. The ISAC networks typically leverage waveform design, joint signal processing, and resource sharing to optimize the performance of dual functionalities [3]. By reusing radio spectrum, transmit power, and hardware resources, the ISAC networks aim at enhanced spectral efficiency and

reduced system costs. The ISAC networks are pivotal in scenarios such as autonomous driving, where vehicles rely on seamless communications and high-precision environmental sensing for real-time vehicle control decisions [4], and mobile video streaming, where the integrated system ensures accurate spatial awareness for immersive user experiences [5].

Within ISAC networks, achieving a good trade-off between communication and sensing functionalities necessitates innovative resource management strategies due to their competition for limited resources, such as radio spectrum, transmit power, and computing resources [6]. Conventional service-centric approaches often fail to satisfy the diverse and dynamic requirements of ISAC applications, as user priorities between communication and sensing can vary significantly [7]. Experience-centric resource management prioritizes user-level experience, i.e., quality of experience (QoE), rather than quality of service (QoS) metrics such as throughput or latency. To achieve experience-centric resource management, advanced data processing and resource scheduling algorithms have been developed. For instance, data-driven predictive models are utilized to forecast user resource demands based on real-time user status and behavioral patterns [8]. Adaptive priority-based strategies dynamically allocate resources by constantly re-evaluating each user requirement and status to ensure an effective balance between communication and sensing performance [9]. However, achieving experience-centric resource management requires efficient and accurate user data management.

To efficiently and accurately manage user data, the concept of digital agent (DA) has emerged as a key enabler. Building upon the foundation of digital twins (DTs), a DA represents an evolutionary step that integrates advanced data management functionalities tailored for dynamic and experience-centric resource management [10]. DAs are expected not only to emulate user status in real-time but also to extract fine-grained user characteristics, such as behavioral patterns, to enable precise modeling of user requirements in complex and dynamic environments. Compared with the DA, the DA emphasizes adaptive intelligence rather than mere replication. DA focuses on real-time mapping and simulation validation, whereas DA extracts user- and group-level behavioral patterns to enable customized and automated management. DA relies on high-frequency state synchronization, while DA abstracts representative features such as QoE models, thus reducing synchronization costs. More importantly, DA mainly supports decisions through virtual tests, while DA forms a closed feedback loop that learns user

Received 3 October 2025; revised 23 December 2025; accepted 4 March 2026. Date of publication 10 March 2026; date of current version 6 July 2026. Recommended for acceptance by Y. Shu. (*Corresponding author: Yingying Pei.*)

Xinyu Huang, Yixiao Zhang, Yingying Pei, and Xuemin Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: x357huan@uwaterloo.ca; y3549zha@uwaterloo.ca; y32pei@uwaterloo.ca; sshen@uwaterloo.ca).

Jianzhe Xue is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: jianzhexue@smail.nju.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TMC.2026.3672678>, provided by the authors.

Digital Object Identifier 10.1109/TMC.2026.3672678

behaviors, makes real-time decisions, and continuously evolves with changing network conditions. For example, DAs can leverage multi-modal data sources, such as communication logs, sensor feedback, and application-level interactions, to generate comprehensive user profiles consisting of static attributes (e.g., user preferences), and dynamic states (e.g., mobility trajectories and network traffic variations). Furthermore, DAs possess advanced decision-making capabilities through machine learning and optimization techniques, which can dynamically optimize resource allocation based on user requirements and predicted resource demand fluctuations [8].

Achieving experience-centric resource management in ISAC networks faces significant challenges. First, user-specific QoE models not only vary across users but also evolve over time due to dynamic user behaviors and changing environmental conditions, which complicates model establishment and update. Second, since resource optimization variables are typically coupled with complex and interdependent relationships, how to construct an efficient communication and sensing optimization framework to enhance user QoE is an open research problem. Third, the joint optimization over a large user scale and numerous decision variables introduces high computational complexity, causing traditional optimization methods to converge slowly to an optimal or near-optimal solution. In addition, purely data-driven methods are highly susceptible to data drift.

In this paper, we propose a novel DA-assisted resource management scheme in ISAC networks, which can efficiently manage user-specific QoE models and make tailored resource scheduling strategies to enhance long-term user QoE. The main contributions are summarized as follows:

- First, we propose a DA module to manage user-specific QoE models. Within the DA module, a distance correlation coefficient (DCC) model is utilized to select key QoS factors influencing each user QoE. To capture the dynamics of user behaviors and environmental complexity, a lightweight prediction model is leveraged. Based on the selected QoS factors, along with the predicted user behavioral dynamics and environmental complexity, a personalized QoE model is constructed by data fitting and continuously refined through user feedback. Users with the same QoE model structure are clustered into a user group.
- Second, we utilize Cramér-Rao bound (CRB) to model the impact of allocated communication resources on sensing accuracy. Particularly, each ISAC user has three CRB constraints in terms of distance, velocity, and azimuth angle to achieve sensing accuracy. A joint optimization problem of communication and computing resource management to maximize long-term user QoE is formulated by incorporating both CRB and resource constraints.
- Third, we develop a two-layer data-model-driven algorithm to solve the joint communication and computing resource optimization problem. In the top layer, an advanced deep reinforcement learning (DRL) algorithm is utilized to make group-level network resource allocation decisions. In the bottom layer, the user-level resource allocation problem in each group is transformed into a convex problem and solved by an optimization solver. Experiments based

on a real-world dataset demonstrate the effectiveness of the proposed algorithm across different network resource availability and sensing requirements.

The remainder of this paper is organized as follows. Related works are reviewed in Section II. The system model, comprising a sensing model and a DA-assisted QoE model, is established, and a joint optimization problem is formulated in Section III. A two-layer data-model-driven algorithm is presented in Section IV, followed by simulation results in Section V. Finally, Section VI concludes this study.

II. RELATED WORKS

Resource management in ISAC networks has been extensively studied in recent years. Early research on communication resource management for ISAC mainly focused on non-overlapping approaches, i.e., orthogonal frequency division multiple access (OFDMA) and time division multiple access (TDMA). In the OFDMA approaches, the frequency domain is partitioned into multiple subcarriers, with each subcarrier allocated to either sensing or communication [11], [12], [13]. In the TDMA approaches, a time slot is divided into multiple mini-slots, with all subcarriers in each mini-slot allocated to either sensing or communication [14], [15], [16]. Generally, sensing performance is commonly measured using metrics such as target detection probability, sensing coverage, and sensing continuity [9], [17], while communication performance is evaluated using QoS metrics (e.g., delay and throughput [18]) and QoE metrics (e.g., content quality and rebuffering time [19]). Recently, a unified ISAC waveform design has attracted much attention. A single signal waveform is expected to simultaneously facilitate the sensing and communication functionalities to achieve high spectrum efficiency [9], [20], [21]. Intuitively, allocating more spectrum resources can improve sensing and communication performance. However, spectrum resources are usually limited, and how to efficiently allocate them to achieve satisfactory performance is essential. To address this issue, researchers proposed various CRB models to depict the mathematical relationship between allocated communication resources and sensing performance [9], [22]. To guarantee sensing accuracy, the sensing CRB usually needs to remain below a prescribed threshold. Additionally, some existing works integrate communication resource management with computing or caching resource management to further enhance QoS or QoE [23], [24]. Since such integration introduces additional optimization variables in complex network environments, advanced DRL algorithms have been adopted to make resource management decisions by learning the network state transition probabilities.

Recently, DAs have attracted widespread attention for improving resource management performance. A DA can be regarded as an intelligent proxy of a physical entity on the network side to perform efficient system behavior emulation and data analytics for customized resource management [10]. The emulated behaviors mainly consist of user behaviors, including mobility trajectories and application interactions, and network operation behaviors, including topology changes and network configurations [25]. These behaviors can be emulated by

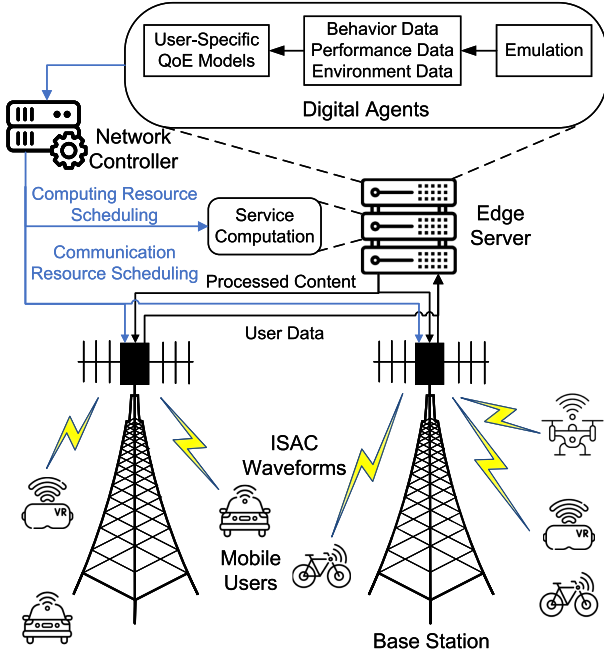


Fig. 1. An illustration of DA-assisted resource management in the ISAC scenario.

prediction-based algorithms in network simulators, e.g., NS-3 and OPENT. Additionally, DAs can abstract useful network characteristics, such as spatiotemporal traffic variations [26], swipe probability distributions [8], and collaborative sensing demands [27], from emulated behaviors and realistic environments. The information maintained in DAs reduces the reliance on handcrafted prior knowledge and enables adaptive resource management under dynamic network conditions.

In summary, most existing works on resource management in ISAC networks focus on performance enhancement based on universal QoE models and purely data-driven or model-based resource scheduling algorithms. Different from the existing studies, our DA-assisted resource management scheme constructs and continuously updates user-specific QoE models that integrate the impact of behavioral dynamics and environmental complexity on selected QoS factors. Moreover, a hybrid data-model-driven resource scheduling algorithm is developed to decouple the joint optimization problem for dimension reduction and utilizes convex optimization to improve the robustness of DRL algorithm.

III. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, a sensing model and a DA-assisted QoE model are established in an ISAC scenario. Based on these models, an experience-centric optimization problem is formulated.

A. ISAC Scenario

As shown in Fig. 1, we consider a DA-assisted resource management framework for the ISAC scenario. The communication network mainly consists of two base stations, one edge server, multiple mobile users and their DAs, and one network controller.

TABLE I
MAIN NOTATIONS AND DEFINITIONS

Notation	Definition
$\mathbf{r}_n(t)$	The reflected signal echoes at time instance t in time slot n
B^{NN}	Null-to-null beam width of a receive antenna
$B_{n,k}^{\text{RMS}}$	User k 's effective bandwidth in time slot n
$H_{n,k}$	User k 's behavioral dynamics in time slot n (Sec. III-C5)
$E_{n,k}$	User k 's environmental complexity in time slot n (Sec. III-C5)
$\mathcal{E}_{n,k}$	User k 's QoE in time slot n (Sec. III-C5)
$\mathcal{S}_{n,k}$	User k 's QoS in time slot n (Sec. III-C5)
$\Upsilon_{n,k}$	User k 's sensing CRB in time slot n
$B_{n,k}$	The allocated bandwidth to user k in time slot n
$P_{n,k}$	The allocated transmit power to user k in time slot n
$C_{n,k}$	The allocated computing resources to user k in time slot n

The base stations transmit unified ISAC waveforms to mobile users and analyze the reflected signal echoes to sense mobile user locations and movement speeds. The ISAC waveforms can support simultaneous data transmission and target sensing. Radio spectrum is partitioned to support the transmission of ISAC waveforms.¹ User data, including behavior, performance, and environment data, are collected periodically for DA update. Each mobile user has a unique DA in the edge server, which predicts user status and abstracts a user-specific QoE model. The QoE model mainly consists of user behavioral dynamics, environmental complexity, and QoS metrics such as service latency and content quality. Mobile users are indexed by $k \in \{1, \dots, K\}$, with each user having a random but distinct velocity and trajectory. The edge server with substantial caching² and computing resources hosts DAs and processes user requests, such as video rendering and data analysis. The network operates in a time-slotted manner, in which time slots are indexed by $n \in \{1, \dots, N\}$. Assume that the variables are constant in each time slot and the number of carriers exceeds the number of mobile users to avoid transmission interference. The network controller is responsible for communication and computing resource management based on real-time user status and their QoE models.

The main notations and their definitions are listed in Table I.

B. Sensing Model

Each base station is equipped with N_t transmit antennas and N_r receive antennas to perform sensing and communication simultaneously. The baseband signals transmitted by the base stations are denoted by $\hat{\mathbf{s}}_n(t) = [\hat{s}_{1,n}(t), \dots, \hat{s}_{K,n}(t)]^T \in \mathbb{C}^{K \times 1}$, where $t \in \mathbb{N}^+$ is the time instance index and each time slot consists of multiple time instances. Superscript T denotes matrix transpose. The transmitted signals need to be calibrated on amplitudes and phases across antennas through

¹In this work, we consider downlink transmission optimization for ISAC waveforms.

²In this work, we consider a fixed caching strategy, i.e., least recently used (LRU), to update DA data and models for data freshness and model accuracy, and focus on communication and computing resource management.

beamforming technology to enhance transmission efficiency, which can be expressed as $\mathbf{s}_n(t) = \mathbf{W}_n \hat{\mathbf{s}}_n(t) \in \mathbb{C}^{N_t \times 1}$. Here, $\mathbf{W}_n = [\mathbf{w}_{1,n}, \dots, \mathbf{w}_{K,n}] \in \mathbb{C}^{N_t \times K}$ is the downlink beamforming matrix in time slot n . The reflected signal echoes from mobile users at time instance t in time slot n can be expressed as [28]

$$\mathbf{r}_n(t) = G \sum_{k=1}^K \beta_{n,k} e^{j2\pi\mu_{n,k}t} \mathbf{b}(\theta_{n,k}) \mathbf{a}^H(\theta_{n,k}) \mathbf{s}_n(t - v_{n,k}) + \mathbf{z}_n(t), \quad (1)$$

where $G = \sqrt{N_t N_r}$ is the total antenna array gain and $\beta_{n,k}$ represents the reflection coefficient that is related to propagation loss. In (1), $v_{n,k}$ is the propagation delay, $\mu_{n,k}$ denotes the Doppler frequency, $\theta_{n,k}$ is the azimuth angle between mobile user k and the base station, and $\mathbf{z}_n(t) \in \mathbb{C}^{N_r \times 1}$ represents the circularly symmetric complex Gaussian noise vector. In (1), vectors $\mathbf{a}$ and $\mathbf{b}$ represent transmit and receive steering vectors, respectively. Superscript H is Hermitian transpose.

To estimate the sensing accuracy of user k on distance $c_{n,k}$, velocity $v_{n,k}$, and azimuth angle $\theta_{n,k}$ in time slot n , we select CRB, which provides a lower bound on the variance of an unbiased estimator. The estimation CRBs for sensing mobile user k in time slot n can be expressed as [9]

$$\begin{cases} \Upsilon_{n,k}^{(1)} = \frac{\lambda_1}{P_{n,k} |\varsigma_{n,k}|^2 |B_{n,k}^{\text{RMS}}|^2}, \\ \Upsilon_{n,k}^{(2)} = \frac{\lambda_2}{P_{n,k} |\varsigma_{n,k}|^2 |T_{n,k}^{\text{Eff}}|^2}, \\ \Upsilon_{n,k}^{(3)} = \frac{\lambda_3}{P_{n,k} |\varsigma_{n,k}|^2 / B^{\text{NN}}}, \end{cases} \quad (2)$$

where $P_{n,k}$ and $\varsigma_{n,k}$ represent the transmit power and downlink sensing channel gain for mobile user k in time slot n , respectively. In (2), $\Upsilon_{n,k}^{(1)}$, $\Upsilon_{n,k}^{(2)}$, $\Upsilon_{n,k}^{(3)}$ represent the sensing performance of distance, velocity, and azimuth angle for the mobile user. A lower CRB value corresponds to a higher sensing accuracy.³ Parameters $\lambda_1, \lambda_2, \lambda_3$ represent different sensing requirements for the distance, velocity, and azimuth angle, respectively; B^{NN} and $B_{n,k}^{\text{RMS}}$ represent a null-to-null beam width and an effective bandwidth for mobile user k , respectively. Effective bandwidth $B_{n,k}^{\text{RMS}}$ can be analyzed on the integral of a frequency-domain signal over its bandwidth [30], given by

$$|B_{n,k}^{\text{RMS}}|^2 = \frac{\int_{B_{n,k}} f^2 |S(f_{n,k})|^2 df}{\int_{B_{n,k}} |S(f_{n,k})|^2 df}, \quad (3)$$

where $f_{n,k}$ is the carrier frequency for mobile user k and $S(f_{n,k})$ is the Fourier transform of time-domain waveform $\hat{\mathbf{s}}_{n,k}(\tilde{t})$. Note that $\tilde{t}$ is the continuous form of t , and $S(f_{n,k})$ can be expressed as $S(f_{n,k}) = \frac{1}{\pi f_{n,k}} \sin(\pi f_{n,k} T^{\text{P}})$ [9], where T^{P} is the pulse width. Therefore, the effective bandwidth $B_{n,k}^{\text{RMS}}$ can be estimated as [9]

$$|B_{n,k}^{\text{RMS}}|^2 \approx B_{n,k} / (2\pi^2 T^{\text{P}}). \quad (4)$$

³In the derivation of the CRB, we consider a noise-limited scenario. We assume that inter-BS interference is effectively mitigated through orthogonal resource allocation and that static clutter is removed via clutter suppression techniques [29]. This assumption allows us to focus on the fundamental performance limits imposed by the signal-to-noise ratio (SNR) and geometric configurations.

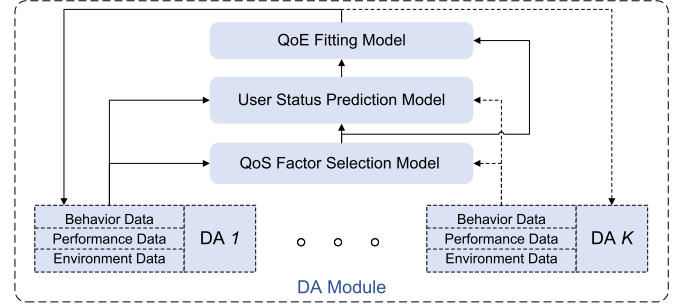


Fig. 2. The operational mechanism in the DA module.

In (2), $T_{n,k}^{\text{Eff}}$ represents the effective time duration for mobile user k , which is used to quantify the temporal spread of the transmit signal's energy [30]. It can be expressed as

$$|T_{n,k}^{\text{Eff}}|^2 = \frac{\int_n \tilde{t}^2 |S_{n,k}(\tilde{t})|^2 d\tilde{t}}{\int_n |S_{n,k}(\tilde{t})|^2 d\tilde{t}}, \quad (5)$$

where $S_{n,k}(\tilde{t})$ is the normalized complex envelope of the transmit signal for mobile user k , given by $S_{n,k}(\tilde{t}) = \hat{s}_{n,k}(\tilde{t}) / \sqrt{\int_n |\hat{s}_{n,k}(\tilde{t})|^2 d\tilde{t}}$. Since the allocation of dwell time (the process of assigning a specific time interval during sensing stages) is not considered in this work, the effective time duration $T_{n,k}^{\text{Eff}}$ is set as a system parameter. It is worth noting that although the BS is aware of the total transmitted waveform, the sensing performance for a specific target k depends on the effective illumination. Due to the use of directional beamforming, the energy intended for other users is spatially separated from target k . Thus, the CRB is determined by the specific signal parameters ($B_{n,k}^{\text{RMS}}$ and $T_{n,k}^{\text{Eff}}$) contained in the beam directed toward target k .

C. DA-Assisted QoE Model

In this subsection, we discuss how to utilize DAs to construct QoE models for individual users.

1) *DA Module Overview:* As shown in Fig. 2, a DA module consists of multiple DAs, each of which stores the corresponding user behavior data (e.g., velocity and engagement time), performance data (e.g., service delay and content quality), and environment data (e.g., location and surrounding objects). The DA module is assumed to be a trusted data storage and processing module in communication networks. The DA data processing functionalities include a QoS factor selection model, a user status prediction model, and a QoE fitting model. In the QoS factor selection model, the input is service performance data from DAs, which are typically multi-dimensional and time-series. Directly using them to update a user QoE model can cause fitting distortion. Therefore, an efficient QoS factor selection model is essential, as it can identify which QoS factors can affect the current QoE model by using a distance correlation coefficient model. The output of QoS factor selection model is selected QoS factors for each user. In the user status prediction model, the input and output are historical DA data and predicted one, respectively.

By using a lightweight prediction model, the future user status can be predicted with low computational overhead. In the QoE fitting model, the input includes the selected QoS factors and emulated contextual information from DA data. Considering the spatiotemporal dynamics in user behaviors and environments, it is important to incorporate their impact into the QoE model because users under different contexts usually have diversified service demands. With the determined QoS factors and emulated contextual information, a comprehensive QoE model can be established for each individual user. The DA data are also used to trigger the QoE model update when the user experience falls below a predefined threshold.

2) *DA Data Collection*: DA data are classified into three categories, i.e., user behavior data, service performance data, and environment data, which are collected from physical networks and the user status prediction model in the DA module. DA data usually exhibit user variability and spatiotemporal discrepancies during data generation. Specifically, different users usually engage with various applications (e.g., social media, streaming, and online gaming), in varying ways, which leads to distinct behavioral patterns. Some users have long and continuous sessions, while others interact in short bursts. This affects the data generation rate and distribution. For example, in short video streaming services, some users may have frequent swipe behaviors while others may watch complete short videos in a sequential manner. This discrepancy can lead to different swipe-behavior uploads. State synchronization is defined as the process of updating the DA database with the latest user behavior, service performance, and environment data through a predictive model and an adaptive data collection mechanism. Specifically, the user status prediction model in the DA module can help generate DA data to guarantee that user status can always be aware on the network side, but it also brings some prediction bias due to data drift. More data needs to be collected from physical networks for the user status prediction model update through incremental learning algorithms when the average prediction accuracy is below a prescribed threshold, as specified in [31]. With the help of the user status prediction model, DA data collection can be adaptive to user status variation patterns. When the user status variation pattern is relatively static, the user status prediction model can have high accuracy and DA data collection can fully rely on the user status prediction model. Otherwise, DA data need to be collected from physical networks to guarantee user status is available in real time and update the user status prediction model. Based on the preceding analysis, we propose an adaptive DA data collection frequency equation, given by

$$l_{k,i} = \bar{l}_i e^{-\nu_i A_{k,i}}, \quad (6)$$

where $l_{k,i}$ denotes the DA data collection frequency for data attribute i from user k . Parameter $\bar{l}_i$ is the standard data collection frequency for data attribute i , which can be retrieved from existing data collection standards, such as [32], [33]. For example, in short video streaming services, user swipe behaviors are usually collected every 5 seconds; in map navigation services, user locations are usually collected every 1 second. Particularly, $e^{-\nu_i A_{k,i}}$ is a scaling factor to adaptively adjust data collection frequency, where ν_i and $A_{k,i}$ denote the attenuation rate of data

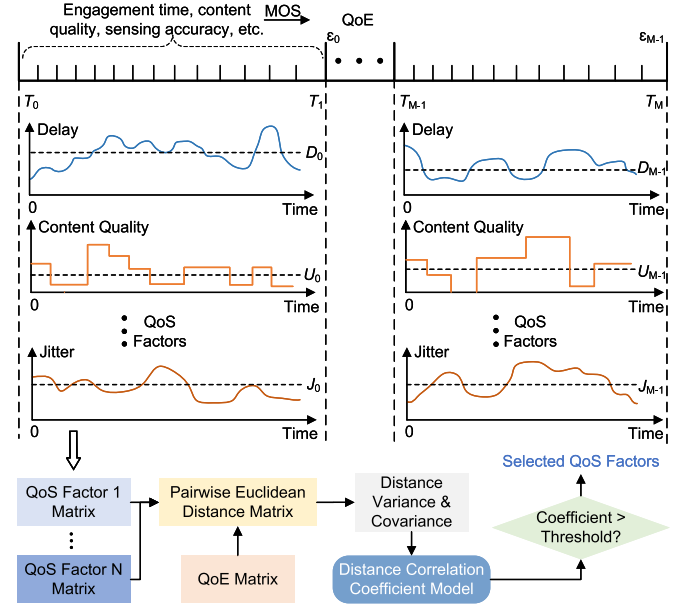


Fig. 3. The QoS factor selection procedure in the DA module.

collection frequency for data attribute i and average prediction accuracy for DA data attribute i of user k , respectively. With the increasing prediction accuracy, DA data collection frequency gradually decreases.

3) *QoS Factor Selection in the DA Module*: As shown in Fig. 3, we present how to select QoS factors for each individual user in the DA module. Since the QoE reflects a user's overall satisfaction over large time durations, the user behavior data and service performance data, such as engagement time, content quality, and sensing accuracy, are collected in each large time duration, e.g., over $[T_0, T_1]$, to estimate the corresponding mean opinion score (MOS). Content quality denotes the quality of transmitted data, which is usually measured when a complete data file is received in the user terminal. Therefore, it is usually a piecewise constant in the time domain. For example, in video streaming services, content quality can refer to structural similarity index measure (SSIM) of video segments or frames. The MOS usually ranges from 1 ~ 5, which has a prescribed mapping relationship with DA data [34]. A higher MOS corresponds to a higher user experience. In this work, we use MOS to measure user QoE. The content quality is equally split into several parts, with a higher content quality corresponding to a higher MOS. Then, the average QoS factors and QoE during each large time duration are used to obtain different matrices. The matrices are input to the DCC model to select QoS factors for user-specific QoE models. Let $\mathbf{D} = [D_0, \dots, D_{M-1}]$ and $\mathbf{Q} = [\mathcal{E}_0, \dots, \mathcal{E}_{M-1}]$ denote the service delay matrix and QoE matrix, respectively, which are used as an example to introduce the operational procedure of the distance correlation coefficient model. In matrices $\mathbf{D}$ and $\mathbf{Q}$, M is the number of selected large time durations, and D_m and $\mathcal{E}_m$ are scalars. The specific procedure is as follows:

First, we need to calculate the pairwise Euclidean distance matrices, i.e., $\mathbf{X}$ for $\mathbf{D}$ and $\mathbf{Y}$ for $\mathbf{Q}$. The scalar element X_{ij}

in matrix $\mathbf{X}$ is defined as the absolute difference $|D_i - D_j|$ between scalar elements D_i and D_j in matrix $\mathbf{D}$, and similarly for Y_{ij} in matrix $\mathbf{Y}$. These matrices are then double-centered using the formula:

$$X_{ij}^* = X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..}, \quad (7)$$

where $\bar{X}_{i.}$ is the mean of the i -th row, $\bar{X}_{.j}$ is the mean of the j -th column, and $\bar{X}_{..}$ is the overall mean of matrix $\mathbf{X}$. The same principle is applied to matrix $\mathbf{Y}$.

The distance covariance $\mathcal{V}^2(\mathbf{D}, \mathbf{Q})$ is then calculated as

$$\mathcal{V}^2(\mathbf{D}, \mathbf{Q}) = \frac{1}{M^2} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} X_{ij}^* Y_{ij}^*. \quad (8)$$

Similarly, the distance variances of matrices $\mathbf{D}$ and $\mathbf{Q}$ are computed as $\mathcal{V}^2(\mathbf{D}, \mathbf{D}) = \frac{1}{M^2} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} X_{ij}^* X_{ij}^*$ and $\mathcal{V}^2(\mathbf{Q}, \mathbf{Q}) = \frac{1}{M^2} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} Y_{ij}^* Y_{ij}^*$, respectively.

Finally, the DCC of matrices $\mathbf{D}$ and $\mathbf{Q}$ is given by [35]

$$\mathcal{F}(\mathbf{D}, \mathbf{Q}) = \frac{\mathcal{V}(\mathbf{D}, \mathbf{Q})}{\sqrt{\mathcal{V}(\mathbf{D}, \mathbf{D}) \times \mathcal{V}(\mathbf{Q}, \mathbf{Q})}}, \quad (9)$$

where the coefficient $\mathcal{F}(\mathbf{D}, \mathbf{Q})$, ranging from 0 to 1, quantifies the association strength between service delay and QoE. A high coefficient indicates a stronger association strength.

Based on the distance correlation coefficient model, the DA module can select which QoS factors are related to the user-specific QoE models, thereby improving the efficiency of QoE model establishment.

4) *User Status Prediction in the DA Module:* The QoE model fitting is performed at the end of each large time duration, which requires time-averaged user status information. Therefore, we construct a large-timescale prediction model in the DA module, which can alleviate the impact of uncertain behaviors and environment variations at the small timescale. To reduce computational consumption, we select a lightweight autoregressive integrated moving average (ARIMA) method to predict the future user status. Taking environmental complexity sequence $\hat{\mathbf{E}} = [E_1, \dots, E_M]$ as an example, the detailed steps for predicting E_{M+1} is given as follows:

First, we need to check the stationarity of the time series. If series $\hat{\mathbf{E}}$ is non-stationary, we can transform it into a stationary state by differencing it w times, and obtain the differenced series, given by

$$y_m = \Delta^w E_m = (1 - L)^w E_m, \quad (10)$$

where L denotes the lag operator, and Δ represents the differencing operation.

Then, we need to determine orders p and q of the ARIMA model. By analyzing the autocorrelation function (ACF) and partial ACF of differenced series y_m , we select appropriate autoregressive order p and moving average order q to establish the ARIMA(p, w, q) model. Specifically, when selecting p and q , the ACF and partial ACF plots need to be observed, where a rapid cutoff in partial ACF suggests a suitable autoregressive order p , and a rapid cutoff in ACF indicates an appropriate moving average order q . By comparing candidate models using information criteria and residual analysis, the optimal parameter combination

can be selected to achieve a balance between prediction accuracy and computational efficiency.

Next, we need to estimate the model parameters. By using the maximum likelihood estimation method, we estimate the autoregressive coefficient ϕ_i and moving average coefficient θ_j . The ARIMA model is formulated as

$$\Phi(L)y_m = \Theta(L)\varepsilon_m, \quad (11)$$

where $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$. In (11), ε_m is the white noise error term.

After building the ARIMA model, differenced series $\hat{y}_{M+1}$ can be predicted as

$$\hat{y}_{M+1} = \sum_{i=1}^p \phi_i y_{M+1-i} + \sum_{j=1}^q \theta_j \varepsilon_{M+1-j}. \quad (12)$$

Since future error terms are unknown, it is typically assumed that $\varepsilon_m = 0$ for $m > M$.

Finally, we can perform inverse differencing to recover the predicted value of the original series E_{M+1} . Depending on the differencing order w , the inverse differencing formula is given by

$$E_{M+1} = \begin{cases} \hat{y}_{M+1} + E_M, & \text{if } w = 1, \\ \hat{y}_{M+1} + 2E_M - E_{M-1}, & \text{if } w = 2. \end{cases} \quad (13)$$

where we only consider differencing order $w \in \{1, 2\}$, as they are the most common cases encountered in practice and can provide an accurate estimation.

Through these steps, an ARIMA model can be built at the DA module to make accurate and lightweight user status prediction.

5) *QoE Model Fitting & Update in the DA Module:* To accurately capture realistic user satisfaction with network services, a comprehensive QoE model is crucial. We consider a QoE model that integrates the impact of behavioral dynamics and environmental complexity on QoS. The behavioral dynamics are analyzed based on the variations in user behavior data, while the environmental complexity is estimated based on the number of surrounding buildings and obstacles in a specific location. When users remain relatively stationary in simpler environments, their QoS requirements can be relaxed. In contrast, in more dynamic or complex settings, stricter QoS requirements are necessary to maintain a satisfactory user experience. This adaptability allows the QoE model to remain responsive to contextual factors, which can enable the network operator to dynamically adjust service quality in response to changing environmental conditions and user behaviors. Based on the preceding analysis, the comprehensive QoE model is constructed as

$$\mathcal{E}_{n,k} = \mathcal{S}_{n,k} \times I(H_{n,k}, E_{n,k}), \quad (14)$$

where $\mathcal{E}_{n,k}$ and $\mathcal{S}_{n,k}$ represent mobile user k 's QoE and QoS in time slot n , respectively. Function $I(\cdot)$ is an impact function and $H_{n,k}$ is mobile user k 's behavioral dynamics. Two QoS factors are considered in the DA-assisted QoS factor selection model, i.e., service delay and content quality, which have linear combination structure and nonlinear one [36], [37], [38], represented by L_1 and L_2 . In each time slot, the QoS of mobile user k can be

expressed as

$$S_{n,k} = \begin{cases} \underbrace{\frac{\omega_{n,k}^{(1)} \xi_k F_{n,k}}{1 + \xi_k F_{n,k}}}_{\text{content quality}} - \underbrace{\omega_{n,k}^{(2)} \left(\frac{\mu F_{n,k}}{C_{n,k}} + \frac{F_{n,k}}{R_{n,k}} \right)}_{\text{service latency}}, & \text{if } M_{n,k} = L_1, \\ \omega_{n,k}^{(3)} \frac{\xi_k F_{n,k}}{1 + \xi_k F_{n,k}} / \left(\frac{\mu F_{n,k}}{C_{n,k}} + \frac{F_{n,k}}{R_{n,k}} \right), & \text{if } M_{n,k} = L_2, \end{cases} \quad (15)$$

where the QoS model structure of user k is denoted by $M_{n,k}$, which can be determined by data fitting methods. The first component is content quality, which has a positive correlation with file size $F_{n,k}$ [31], [39]. In (15), ξ_k is the weighting parameter for content quality, whose value is typically set to 2 [31]. The second component is service latency, where parameter μ is computing density and $R_{n,k}$ is the transmission capability of mobile user k in time slot n . The transmission capability $R_{n,k}$ can be estimated by allocating bandwidth resources, given by $R_{n,k} = B_{n,k} \log(1 + P_{n,k} |h_{n,k}|^2 / (B_{n,k} \sigma_r^2))$, where $h_{n,k}$ is the communication channel gain of mobile user k and σ_r^2 is the noise power spectral density. To balance the content quality and service latency, we introduce three weighting factors $\omega_{n,k}^{(1)}$, $\omega_{n,k}^{(2)}$ and $\omega_{n,k}^{(3)}$, which can be fitted by the mapping between historical QoE and QoS.

As analyzed earlier, impact function $I(\cdot)$ is regarded as an elastic factor influencing the QoS function, where high user dynamics and complex environments correspond to strict QoS requirements, and vice versa. Since behavioral dynamics $H_{n,k}$ and environmental complexity $E_{n,k}$ have different units and ranges, typically larger than 0, they are normalized as $\tilde{H}_{n,k} = H_{n,k} / H_{\max}$ and $\tilde{E}_{n,k} = E_{n,k} / E_{\max}$. Additionally, the impact function should have the following characteristics: when user dynamics and environmental complexity are zero, the user QoS requirement is a predetermined fixed value; as both increase, the value of the function decreases, indicating stricter QoS requirements. Based on the preceding analysis, a parameterized Sigmoid function is utilized to characterize the impact function, given by

$$I(H_{n,k}, E_{n,k}) = 1 - \frac{0.3}{1 + e^{-5(\tilde{H}_{n,k} + \tilde{E}_{n,k} - 1)}}, \quad (16)$$

where $I(\cdot)$ take values in the interval (0.7,1) and is negatively correlated with $H_{n,k}$ and $E_{n,k}$.

Since user behavioral patterns are not always static, the dynamics require updating user QoE models, including model structures (L_1, L_2) and parameters ($\omega_{n,k}^{(1)}, \omega_{n,k}^{(2)}, \omega_{n,k}^{(3)}, \forall k \in \mathcal{K}$). Specifically, the relevance of various factors is first evaluated through the DCC analysis to identify which factors are currently dominant. This ensures that the QoE model remains sensitive to the most pertinent aspects of user experience. Then, the model structures and parameters are updated using data fitting techniques to guarantee that the updated QoE has a low mean square score (MSE) compared with the one estimated by MOS.

We select two representative QoE structures (linear weighted and nonlinear fractional) to formulate the optimization problem.

These two structures capture distinct convex/concave characteristics and allow us to validate the proposed two-layer framework. In practice, users may exhibit more diverse QoE models depending on service types (e.g., entertainment, news, games). The data-model-driven algorithm in Section IV can well handle such heterogeneity: users can be clustered according to their QoE structures at the data-driven layer, while the model-driven layer can incorporate customized optimization solvers for each QoE structure. This ensures that the framework is general and extensible beyond the two QoE examples considered in simulations.

6) *Data and Model Interaction Within the DA Module:* As shown in Fig. 2, there exists a closed loop between DA data and models. Specifically, DA data are input to the QoS factor selection model to identify which kinds of user status are essential for experience-centric network management. Based on the determined user status, the user status prediction model can sample them from historical DA data and predict their future values. The predicted user status is further input to the QoE fitting model to guarantee timeliness and improve accuracy. The fitted QoE models are finally stored in the corresponding DAs, which can be used for experience-centric network management.

The accuracy of the DA-assisted QoE model is ensured by three mechanisms. First, the DA module collects heterogeneous and trusted data sources (behavior, service performance, and environment) that ground the QoE model in real-world user information. Second, we adopt a DCC model to statistically select the most relevant QoS factors, thereby avoiding fitting distortion and improving robustness. Third, the model incorporates behavioral dynamics and environmental complexity via the impact function $I(\cdot)$, and continuously re-fits model structures and parameters when prediction errors exceed a threshold. These mechanisms together guarantee that the DA-assisted QoE model remains adaptive, context-aware, and aligned with real user satisfaction.

Discussion: DA state synchronization offers high fidelity but depends on network conditions and sampling rates. DA emulation provides continuous data flow via prediction. However, as a predictive method, it may experience slight deviations from the ground truth over time, which requires occasional recalibration. The proposed DA emulation module utilizes a lightweight ARIMA model, which is computationally efficient. Compared to DA synchronization, which consumes significant bandwidth to transmit raw data, our approach reduces network usage. Furthermore, due to the lightweight nature of the ARIMA model, the computational overhead is negligible compared with complex AI models.

D. Experience-Centric Optimization Problem Formulation

Since user behaviors and environmental complexity change over time, network resource allocation needs to adapt to the dynamics to guarantee communication and sensing performance. Given the limited communication and computing resources in the ISAC networks, our objective is to maximize long-term user QoE with guaranteed sensing performance. To achieve this goal, the network controller needs to accurately

allocate bandwidths, transmit power, and computing resources to each user. Therefore, the formulated optimization problem is given by

$$\max_{\{\mathbf{B}_n, \mathbf{P}_n, \mathbf{C}_n\}_{n \in \mathcal{N}}} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \frac{1}{K} \sum_{k=1}^K \mathcal{E}_{n,k} \quad (17a)$$

$$\text{s.t. } \Upsilon_{n,k}^{(1)} \leq \alpha_1, \forall k \in \mathcal{K}, n \in \mathcal{N}, \quad (17b)$$

$$\Upsilon_{n,k}^{(2)} \leq \alpha_2, \forall k \in \mathcal{K}, n \in \mathcal{N}, \quad (17c)$$

$$\Upsilon_{n,k}^{(3)} \leq \alpha_3, \forall k \in \mathcal{K}, n \in \mathcal{N}, \quad (17d)$$

$$\sum_{k=1}^K B_{n,k} \leq B_{\text{Total}}, \forall n \in \mathcal{N}, \quad (17e)$$

$$\sum_{k=1}^K P_{n,k} \leq P_{\text{Total}}, \forall n \in \mathcal{N}, \quad (17f)$$

$$\sum_{k=1}^K C_{n,k} \leq C_{\text{Total}}, \forall n \in \mathcal{N}, \quad (17g)$$

where $\mathbf{B}_n, \mathbf{P}_n, \mathbf{C}_n$ represent the bandwidth, transmit power, and computing resource allocation matrices for all mobile users, respectively. Parameters $\alpha_1, \alpha_2, \alpha_3$ denote the sensing CRB thresholds for distance, velocity, and azimuth angle, respectively. Constraints (17e)–(17g) are the resource constraints that guarantee the allocated bandwidths, transmit power, and computing resources do not exceed the system capacity.

IV. TWO-LAYER DATA-MODEL-DRIVEN SOLUTION

In this section, we propose a two-layer data-model-driven algorithm to solve the formulated optimization problem.

A. Overview of Proposed Solution

Since users may have various QoE model structures with unique convexity or concavity properties, their combination can complicate the transformation of the objective function into a purely convex or concave form. Furthermore, due to the lack of future information, finding a globally optimal resource allocation strategy via optimization methods is difficult. To obtain a near-optimal solution, one way is to use DRL algorithms. However, directly inputting all user statuses into a DRL algorithm and making resource allocation decisions for each user can cause the curse of dimensionality [40].

Therefore, we propose a two-layer data-model-driven algorithm, where the DRL algorithm is responsible for making group-level resource allocation decisions, and the convex optimization method handles user-level resource allocation problems, as shown in Fig. 4. The algorithm operates as follows.

First, mobile users are clustered into different groups based on their QoE model structures. Then, the average state of each user group is input to the Taylor temporal difference (TD) algorithm [41] to output group-level bandwidth, transmit power, and computing resource allocation, which can reduce the dimensionality of state and action. Next, the user-level resource allocation problem of each group is transformed into a convex

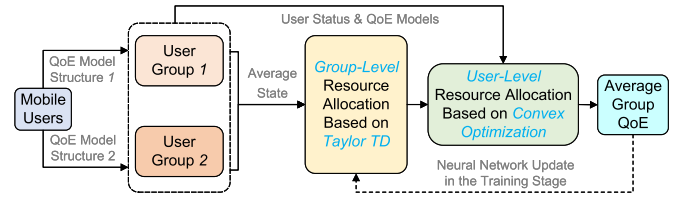


Fig. 4. The proposed two-layer data-model-driven algorithm.

problem solved by convex optimization to achieve an optimal result. Finally, average group QoE is utilized to update the Taylor TD algorithm in the training stage.

B. DRL-Based Group-Level Resource Allocation

In the proposed two-layer data-model-driven algorithm, we adopt the Taylor TD algorithm [41] to determine group-level resource allocation. The Taylor TD algorithm enhances learning stability and accuracy through first-order Taylor corrections. Its ability to model subtle trade-offs and capture fine-grained changes can make it particularly suited for group-level resource allocation decisions, where dynamic and long-term optimization problems can be effectively solved. In the following, we will present the detailed definitions of state, action, reward, and neural network training mechanism in the Taylor TD algorithm.

1) *State*: The state consists of six kinds of elements, i.e., user group i 's average behavioral dynamics $\bar{H}_n^i$, average environmental complexity $\bar{E}_n^i$, average sensing channel gain $\bar{\zeta}_n^i$, average communication channel gain $\bar{h}_n^i$, average file size $\bar{F}_n^i$, and average QoE model parameters $\bar{\omega}_n^i$ at step n . The state can be expressed as

$$\mathbf{s}_n = \{\bar{H}_n^i, \bar{E}_n^i, \bar{\zeta}_n^i, \bar{h}_n^i, \bar{F}_n^i, \bar{\omega}_n^i\}_{i=1,2}. \quad (18)$$

2) *Action*: The action includes bandwidth, transmit power, and computing resource allocation decisions for each group, which can be expressed as

$$\mathbf{a}_n = \{B_n^i, P_n^i, C_n^i\}_{i=1,2}. \quad (19)$$

The action value ranges from 0 to 1. To guarantee that the action can satisfy resource constraints, a normalization operation is needed. For example, the normalized bandwidth resource allocation decision of group i is given by $\hat{B}_n^i = B_{\text{Total}} \times B_n^i / \sum_i B_n^i$. A similar normalization operation is also applied to transmit power allocation and computing resource allocation.

3) *Reward*: The reward is defined as average user QoE, i.e., $r_n = \frac{1}{K} \sum_k \mathcal{E}_{n,k}$, which can be obtained once user-level resource allocation is determined.

4) *Neural Network Training Mechanism*: Traditional TD learning algorithms update value function $Q(\mathbf{s}, \mathbf{a})$ based on the TD error, given by

$$\delta = r + \gamma \max_{\mathbf{a}'} Q(\mathbf{s}', \mathbf{a}') - Q(\mathbf{s}, \mathbf{a}), \quad (20)$$

where δ represents the TD error, γ is the discount factor, and $Q(\mathbf{s}, \mathbf{a})$ and $Q(\mathbf{s}', \mathbf{a}')$ are the current and next state-action value functions, respectively.

In Taylor TD learning, the update is refined with a first-order Taylor correction term, given by

$$Q(s, a + \Delta a) \approx Q(s, a) + \frac{\partial Q(s, a)}{\partial a} \cdot \Delta a, \quad (21)$$

where $\frac{\partial Q(s, a)}{\partial a}$ represents the gradient of $Q(s, a)$ with respect to action a , and Δa is the adjustment to the action. The updated value function is given by

$$Q(s, a) \leftarrow Q(s, a) + \eta \left(\delta + \frac{\partial Q(s, a)}{\partial a} \cdot \Delta a \right), \quad (22)$$

where η is the learning rate. The Taylor correction enables the algorithm to capture local trends in the value function, thereby improving update precision. The Taylor correction is a local refinement term and does not affect the Bellman target [42]. The SCA layer belongs to the environment mapping and thus preserves the Markov property.

Although the state space can be high-dimensional, the Taylor correction term operates on the normalized low-dimensional action space. To avoid instability, we apply gradient clipping to bound the correction term and normalize the action updates within resource constraints. These mechanisms ensure that the Taylor TD algorithm maintains stable updates without gradient explosion. Taylor TD is well-suited for our MDP since the state space is high-dimensional while the action space is low-dimensional. The first-order correction exploits local gradients of the Q-function to improve stability and reduce variance compared with vanilla TD updates. Moreover, Taylor TD preserves the convergence guarantees of classical TD learning and has been proven to converge to the correct Q-value function [41].

C. Model-Based User-Level Resource Allocation

At step n , based on the diversity of QoE model structures, the user-level optimization problem can be decomposed into two subproblems $\mathbf{SP}_1$ and $\mathbf{SP}_2$, as follows:

$$\begin{aligned} \mathbf{SP}_1 : \quad & \max_{\{\mathbf{B}_n, \mathbf{P}_n, \mathbf{C}_n\}} \frac{1}{K_1} \sum_{k=1}^{K_1} \mathcal{E}_{n,k}^{(1)} \\ \text{s.t.} \quad & (17b)-(17g), \end{aligned} \quad (23)$$

$$\begin{aligned} \mathbf{SP}_2 : \quad & \max_{\{\mathbf{B}_n, \mathbf{P}_n, \mathbf{C}_n\}} \frac{1}{K_2} \sum_{k'=1}^{K_2} \mathcal{E}_{n,k'}^{(2)} \\ \text{s.t.} \quad & (17b)-(17g), \end{aligned} \quad (24)$$

where $k \in \{1, \dots, K_1\}$ and $k' \in \{1, \dots, K_2\}$ represent user indexes in user groups 1 and 2, respectively. In (23), (24), $\mathcal{E}_{n,k}^{(1)}$ and $\mathcal{E}_{n,k'}^{(2)}$ are distinguished based on QoS model structures in (15). The total group bandwidth, transmit power, and computing resources in $\mathbf{SP}_1$ and $\mathbf{SP}_2$ are obtained via the Taylor TD algorithm.

Theorem 1: Both subproblems $\mathbf{SP}_1$ and $\mathbf{SP}_2$ are non-convex optimization problems.

Proof: The proof for Theorem 1 is presented in the Appendix in the supplemental material. ■

Based on the above proof, the non-convexity in subproblems $\mathbf{SP}_1$ and $\mathbf{SP}_2$ primarily arises from three factors. First, the QoE functions contain non-convex terms involving $\frac{1}{R_{n,k}}$. Second, some CRB constraints introduce bilinear terms, e.g., $P_{n,k} B_{n,k}$, making the feasible region non-convex. Third, the objective function in subproblem $\mathbf{SP}_2$ includes a fractional term $\frac{\Gamma'_{n,k'}}{\hat{B}_{n,k'}}$, where $\hat{B}_{n,k'}$ depends on $\frac{1}{R_{n,k'}}$, further complicating convexity.

To handle non-convex term $\frac{1}{R_{n,k}}$, we adopt a successive convex optimization (SCA) framework [43]. At the t -th iteration, we obtain a known solution $(B_{n,k}^{(t)}, P_{n,k}^{(t)}, C_{n,k}^{(t)})$ and consequently a known $R_{n,k}^{(t)}$. Let consider a function $f(R) = \frac{1}{R}$, where $R = R_{n,k}$. A first-order Taylor expansion around $R_{n,k}^{(t)}$ is given by

$$\frac{1}{R} \approx \frac{1}{R_{n,k}^{(t)}} - \frac{R - R_{n,k}^{(t)}}{(R_{n,k}^{(t)})^2}. \quad (25)$$

Since $R_{n,k}$ involves $\log_2(1+x)$ with $x = \frac{P_{n,k}|h_{n,k}|^2}{B_{n,k}\sigma_r^2}$, we linearize $\log_2(1+x)$ at $x^{(t)}$, given by

$$\log_2(1+x) = \frac{\ln(1+x)}{\ln(2)}. \quad (26)$$

By approximating $\ln(1+x)$ at $x^{(t)}$, we have

$$\ln(1+x) \approx \ln(1+x^{(t)}) + \frac{x - x^{(t)}}{1+x^{(t)}}. \quad (27)$$

Therefore, the logarithmic function can be estimated as

$$\log_2(1+x) \approx \log_2(1+x^{(t)}) + \frac{x - x^{(t)}}{(1+x^{(t)})\ln(2)}. \quad (28)$$

Substituting (28) into $R_{n,k} = B_{n,k} \log_2(1+x)$, we obtain an affine approximation of $R_{n,k}$ around the current iterate, which ensures that the approximation of $\frac{1}{R_{n,k}}$ is also affine. This step is crucial for maintaining convexity at each iteration.

The CRB constraint involves the product of power and bandwidth, i.e., $P_{n,k} B_{n,k}$, which is a bilinear term and leads to a non-convex feasible set. To convexify this constraint, we introduce an auxiliary variable $Y_{n,k}$ to represent this product. However, the equality constraint $Y_{n,k} = P_{n,k} B_{n,k}$ is still non-convex. Therefore, we apply the McCormick envelope technique [44] to relax this equality, given by

$$\begin{cases} Y_{n,k} \geq P_{n,k} \underline{B} + B_{n,k} \underline{P} - \underline{P} \underline{B}, \\ Y_{n,k} \leq P_{n,k} \bar{B}, \\ Y_{n,k} \leq B_{n,k} \bar{P}, \\ Y_{n,k} \geq 0, \end{cases} \quad (29)$$

where $\underline{P}, \bar{P}, \underline{B}, \bar{B}$ are lower and upper bounds derived from the previous iteration. Specifically, the non-convex bilinear relationship is replaced by its convex hull, which is defined by a set of linear inequalities derived from the lower and upper bounds of the variables. As shown in (29), these linear inequalities form a convex polyhedron that approximates the original feasible region, which makes the optimization problem tractable.

In subproblem $\mathbf{SP}_2$, we face a fractional objective $\frac{\Gamma'_{n,k'}}{\hat{B}_{n,k'}}$, where $\hat{B}_{n,k'} = \frac{\mu F_{n,k'}}{C_{n,k'}} + \frac{F_{n,k'}}{R_{n,k'}}$. To transform it into a convex function, we introduce an auxiliary variable $\phi_{n,k'}$ and use the quadratic transform, given by

$$\frac{\Gamma'_{n,k'}}{\hat{B}_{n,k'}} = \max_{\phi_{n,k'}} \{2\phi_{n,k'}\Gamma'_{n,k'} - \phi_{n,k'}^2\hat{B}_{n,k'}\}. \quad (30)$$

Given $\phi_{n,k'}^{(t)}$ from the previous iteration, updating $\phi_{n,k'}$ and linearizing $\hat{B}_{n,k'}$ ensure that the fractional objective is replaced by a form that can be modeled using convex second-order cone constraints. Combined with the preceding linearization steps, the entire problem at each iteration becomes a convex optimization problem.

Under the SCA framework, after applying first-order Taylor expansions to linearize $\frac{1}{R_{n,k}}$ and $\log_2(1+x)$, employing McCormick envelopes for bilinear terms, and using the quadratic transform for fractional objectives, the transformed subproblems $\mathbf{SP}_1$ and $\mathbf{SP}_2$ at each iteration are convex.

Remark: The SCA step in Algorithm 1 uses surrogates that satisfy the standard first-order consistency conditions: (i) the concave term $\log_2(1+x)$ is replaced by its global affine upper bound obtained by first-order Taylor expansion at the current iterate; (ii) the convex term $\frac{1}{R}$ is locally linearized and augmented with a proximal regularizer $\frac{\tau}{2}\|\mathbf{z} - \mathbf{z}^{(t)}\|_2^2$ to ensure strong convexity and a unique minimizer; (iii) bilinear terms $P_{n,k}B_{n,k}$ are convexified via McCormick envelopes; and (iv) the fractional objective is handled by the quadratic transform, yielding a convex subproblem for fixed auxiliary variables. We initialize from a feasible point and impose box constraints on (B, P, C) , which keep $R_{n,k}^{(t)} > 0$ and the iterates in a compact set. Each convex subproblem is solved exactly using the SLSQP solver. Under these assumptions, the sequence generated by our SCA procedure produces a monotone decrease of the surrogate and any limit point is a stationary solution (i.e., a KKT point) of the original problem.

D. The Procedure and Computational Complexity of Proposed Algorithm

As shown in Algorithm IV-D, we present a two-layer data-model-driven algorithm, named SCA-based Taylor TD (STTD). In the algorithm, group-level and user-level resource allocation decisions are sequentially made. Specifically, in Lines 5 – 7, the Taylor TD neural network makes group-level resource allocation decisions. In Lines 8 – 19, the SCA-based optimization method makes user-level resource allocation for subproblems $\mathbf{SP}_1$ and $\mathbf{SP}_2$. Then, in Lines 20 – 27, the reward is calculated based on average user QoE, and the Taylor TD neural network is updated based on Taylor correction and gradient descent operations.

Remark: In Algorithm 1 (Line 7), the group-level action $\mathbf{a}_n$ is selected either randomly (with probability ς) or according to the Taylor TD policy guided by the online network parameterized by θ . Hence, the Q-network directly participates in the action selection process. The SCA-based optimization layer does not replace the DRL policy. Instead, it refines group-level allocations

Algorithm 1: SCA-based Taylor TD (STTD).

```

1 Input: Users' behavioral dynamics, environmental
   complexity, sensing and communication channel
   gains, file sizes, and QoE model parameters;
2 Output: User-level resource allocation decisions
    $\{B_{n,k}, P_{n,k}, C_{n,k}\}_{k \in \mathcal{K}}$ ;
3 Initialize: Taylor TD parameters, replay memory  $\mathcal{D}$ ,
   policy weight  $\theta$ , target weight  $\theta'$ , initial feasible
   solution  $\{B_k^{(0)}, P_k^{(0)}, C_k^{(0)}, \phi_k^{(0)}\}_{k \in \mathcal{K}}$ ;
4 for each episode  $e \in \{1, \dots, E\}$  do
5   Reset initial state  $s_1$ ;
6   for each step  $n \in \{1, \dots, N\}$  do
7     With probability  $\varsigma$  select a random group-level
       action  $\mathbf{a}_n$ , otherwise, select action based on
       the Taylor TD policy;
8     for each user group  $i \in \{1, 2\}$  do
9       if  $i == 1$  ( $\mathbf{SP}_1$ ) then
10        Linearize non-convex components based
          on Eqs. (25)(28);
11        Conduct bilinear constraints relaxation
          based on Eq. (29);
12        Solve the convex subproblem  $\mathbf{SP}_1$  via
          SLSQP solver to obtain user-level
          decision  $\{B_{n,k}, P_{n,k}, C_{n,k}\}_{k \in \mathcal{K}_1}$ ;
13      end
14      else if  $i == 2$  ( $\mathbf{SP}_2$ ) then
15        Make quadratic transformation for
          fractional objective based on Eq. (30);
16        Linearize non-convex components based
          on Eqs. (25)(28);
17        Solve the convex subproblem  $\mathbf{SP}_2$  via
          SLSQP solver to obtain user-level
          decision  $\{B_{n,k}, P_{n,k}, C_{n,k}\}_{k \in \mathcal{K}_2}$ ;
18      end
19    end
20    Observe reward  $r_t$  based on average user QoE
       and transition to new state  $s_{n+1}$ ;
21    Store transition  $(s_n, \mathbf{a}_n, r_n, s_{n+1})$  in replay
       memory  $\mathcal{D}$ ;
22    Prioritize replay to obtain a transition
        $(s_j, \mathbf{a}_j, r_j, s_{j+1})$  from  $\mathcal{D}$ ;
23    Calculate the TD target based on Eq. (20);
24    Update action-value function using Taylor
       correction in Eq. (22);
25    Perform the gradient descent operation;
26    Every  $\Lambda$  steps, update the target network
       weight with  $\theta' = \theta$ ;
27  end
28 end

```

into user-level decisions and serves as a part of the environment dynamics.

The computational complexity of the STTD algorithm is primarily dominated by the convex optimization for solving subproblems $\mathbf{SP}_1$ and $\mathbf{SP}_2$. Each convex subproblem is solved

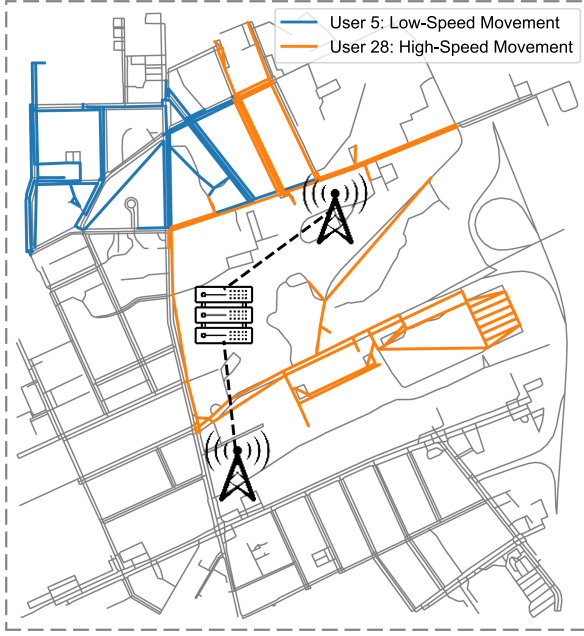


Fig. 5. The simulation region in the Waterloo City, ON, Canada.

by an interior-point method with a worst-case complexity of $O(V^3)$, where V denotes the number of decision variables. This process is repeated for two user groups at each step. Additionally, the Taylor TD updates involve policy network inference and gradient descent, which scale as $O(L_{\text{policy}})$, where L_{policy} represents the operations in the neural network. Replay memory operations, such as prioritized sampling, scale logarithmically with the memory size, which can be expressed as $O(\log |\mathcal{D}|)$. Overall, the total complexity per episode can be expressed as $O(N \cdot (2V^3 + L_{\text{policy}} + \log |\mathcal{D}|))$. Considering the number of episodes, E , the total computational complexity of the algorithm is given by $O(E \cdot N \cdot (2V^3 + L_{\text{policy}} + \log |\mathcal{D}|))$.

V. SIMULATION RESULTS

In this section, we conduct extensive simulations by using the real-world datasets to evaluate the performance of the proposed DA-assisted experience-centric resource management scheme.

A. Simulation Setup

We consider a simulation region in the Waterloo City, ON, Ontario, as shown in Fig. 5. Two base stations equipped with 32 transmit and receive antennas are deployed at the locations $(43.4661^\circ\text{N}, 80.4789^\circ\text{W})$ and $(43.4601^\circ\text{N}, 80.4820^\circ\text{W})$, respectively. The frequency of carriers of two base stations are 28 GHz (n257) and 39 GHz (n260) with bandwidth 400 MHz. The noise power spectral density is set as -174 dBm/Hz [9]. An edge server with computing capacity 15G Cycles/s is connected to two base stations. The computing density for processing user tasks is set as 10^7 Cycles/MB [31]. The total number of ISAC users is 30, where each user moves with different velocities, ranging from 30 ~ 80 Km/h. The variance of pathloss for sensing is calculated based on the standard radar equation, i.e.,

 TABLE II
COMMUNICATION PARAMETERS

Parameter	Description	Value
T^{Eff}	Effective time duration	1 ms [30]
T^{P}	Pulse width	2.5 ns
B_{NN}	Null-to-null beam width	0.076°
P	Total transmit power	40 W
β_{RCS}	Radar cross section	5 m^2
$\hat{e}$	Beamforming gain factor of transmit antenna	1 [9]
$\tilde{e}$	Beamforming gain factor of receive antenna	1 [9]
κ	Array gain	0.8
$\alpha_1, \alpha_2, \alpha_3$	Sensing CRB thresholds	0.01
λ_1	CRB weighting factor of distance	1
λ_2	CRB weighting factor of velocity	10^{-20}
λ_3	CRB weighting factor of azimuth angle	10^{-13}

$\hat{\sigma}_{\alpha_k}^2 = \beta_{\text{RCS}} \lambda_c^2 / (4\pi)^3 d_k^4$, where β_{RCS} and λ_c are the radar cross section of sensing target and the wavelength of carrier, respectively [9]. The sensing channel gain of user k is estimated by $\varsigma_k = \hat{\sigma}_{\alpha_k}^2 \kappa^4 \tilde{e}_k^2 \hat{e}_k^2 / N_r \sigma_r^2$ [9], [45]. The pathloss for communication is calculated based on $32.4 + 20 \log(d_k) + 20 \log(f_c)$ dB [9]. The detailed communication parameter settings are presented in Table II.

We consider a short video streaming service in the simulation region, where 10,000 video sequences are sampled from the YouTube 8M dataset.⁴ The sampled videos include 8 categories, i.e., Entertainment, Games, Food, Sports, Science, Dance, Travel, and News. The time duration of each video sequence is 30 seconds, consisting of 15 equal-length segments. In the DA module, user behavioral dynamics mainly consider user swipe behaviors, which are generated based on watching probability distribution.⁵ To characterize the impact of dynamic channel conditions on ISAC performance, the environmental complexity mainly considers user mobility velocity because high mobility introduces significant Doppler shifts and rapid channel variations. The environmental complexity of user k is estimated as $E_k = \exp((v_k^2 - v_{\text{max}}^2) / 2\rho^2)$ based on the similar principle in [46]. Parameter ρ equals to $v_{\text{max}} / 1.96$, where v_{max} is set as 80 Km/h. Users' MOS are generated based on three principles [10], latency-based probability distribution $\mathcal{N}(5 - 0.4\mathcal{R}, 8; 1, 5)$, quality-based one $\mathcal{N}(1 + 4\mathcal{Q}, 1; 1, 5)$, and combined one $\mathcal{N}(1 + 4\mathcal{Q} - 0.4\mathcal{R}, 0.8; 1, 5)$, where $\mathcal{R}$ and $\mathcal{Q}$ represent service latency and video quality, respectively. The standard DA data collection frequencies for behavior data, service performance data, and environment data are set as 2, 1, and 0.5 per second. The DA data collection attenuation rates of different data attributes are presented in Table IV.

To train an effective STTD algorithm, we set 240 training episodes, each consisting of 60 steps. In each step, the Taylor TD neural network is responsible for making group-level resource allocation decisions. User-level optimization problems are transformed into convex problems, solved by the SLSQP solver. The reward is then analyzed based on the average user QoE. The Taylor TD neural network is updated based on sampled batches. The detailed STTD algorithm parameters are presented

⁴YouTube 8M dataset: <https://research.google.com/youtube8m/index.html>

⁵ACM MM Grand Challenges: <https://github.com/AItransCompetition/Short-Video-Streaming-Challenge/tree/main/data>

TABLE III
STTD ALGORITHM PARAMETERS

Parameter	Value	Parameter	Value
Hidden layer structure	$256 \times 128 \times 64$	Batch size	128
Layer connection	Fully connected	Initial epsilon	0.2
Optimizer	Adam	Final epsilon	0.01
Activation function	ReLU	Memory size	6,000
Maxiter of SLSQP	150	Discount factor	0.9
Step size of SLSQP	10^{-4}	Learning rate	10^{-4}

in Table III. For the STTD algorithm, hyperparameters were selected following common practice in DRL and convex optimization literature. A three-layer fully connected network with structure $256 \times 128 \times 64$ and ReLU activation was adopted to provide sufficient representation capacity while maintaining training stability. The replay memory size (6,000) and batch size (128) follow typical settings in DQN-like algorithms, while Adam optimizer with learning rate 10^{-4} is widely used for stable convergence. The exploration rate ϵ was linearly decayed from 0.2 to 0.01 to balance exploration and exploitation. For the SCA subproblems, the maximum iteration (150) and step size (10^{-4}) were chosen to ensure convergence within reasonable computation time. The discount factor was set to 0.9, which is a conventional choice balancing short-term and long-term QoE optimization.

We compare the proposed DA-assisted resource management scheme with the following benchmark schemes:

- *Without DA (w/o DA)*: A general QoE model consisting of content quality and service quality is utilized. The resource scheduling is based on round robin.
- *Heuristic*: The same DA-assisted QoE model as the proposed scheme, but the resource scheduling uses a greedy algorithm.
- *Purely DRL (PDRL)*: The same DA-assisted QoE model as the proposed scheme, but the resource scheduling relies on a five-layer branch dueling Q network (BDQN) [47], with layer sizes {512, 256, 128, 64, 32}, ReLU activation, Adam optimizer, batch size 128, replay memory 6000, learning rate 10^{-4} , and discount factor 0.9. The exploration rate ϵ decays linearly from 0.2 to 0.01.
- *Hierarchical Reward-based DRL (HDRL)* [48]: The same DA-assisted QoE model, but the resource scheduling relies on a twin delayed deep deterministic policy gradient (TD3) network, where constraint violations and QoE are used to construct a two-layer reward. The TD3 network consists of two critic networks and one actor network, each with hidden layers 400, 300 and ReLU activation. The learning rates for actor and critic are 10^{-3} , target smoothing coefficient $\tau = 0.005$, batch size 100, replay memory 10^5 , and discount factor 0.99.

B. Convergence Performance

As shown in Fig. 6, the convergence performance of the proposed STTD algorithm is evaluated. The envelope area is formulated based on the average reward calculated over every four steps and the corresponding variance. The results show that the reward gradually increases, undergoes minor fluctuations, and

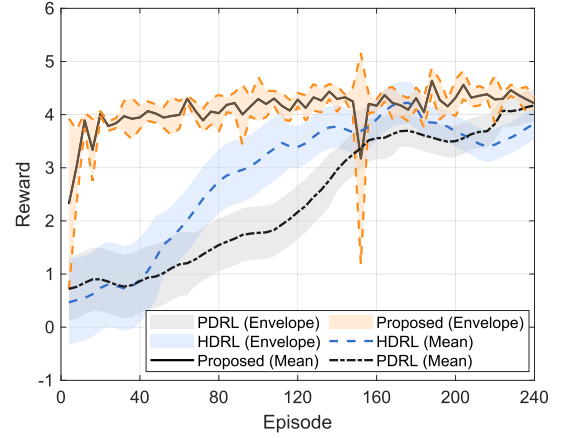


Fig. 6. The convergence performance comparison.

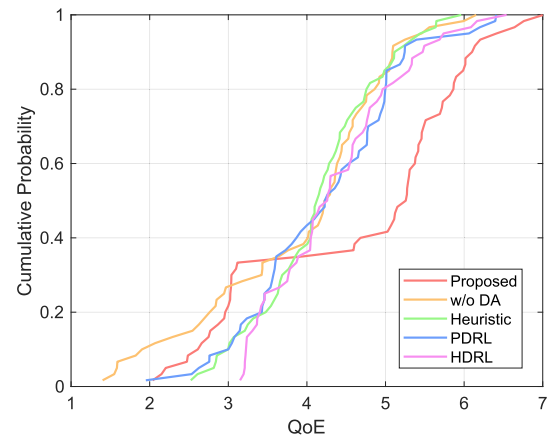


Fig. 7. The cumulative probability of QoE.

ultimately stabilizes around 4.3. Notably, a temporary decrease in the average reward to 3.2 is observed at approximately episode 150, primarily due to extensive explorations. These explorations result in several users' sensing and resource constraints being violated, thereby leading to degraded QoE performance. Overall, the proposed STTD algorithm demonstrates the ability to converge to a high and stable reward, which demonstrates its effectiveness. To further demonstrate the superiority of STTD, we compare its convergence with two DRL baselines, i.e., PDRL and HDRL algorithms. The results show that PDRL algorithm converges more slowly but with smaller variance, while HDRL algorithm converges relatively faster and stabilizes at a higher reward. In contrast, STTD algorithm not only converges quickly but also maintains a higher and more stable reward, which shows its advantages in both training efficiency and stability.

C. QoE Probability Distribution

Fig. 7 depicts the cumulative probability distribution of QoE in one episode under different schemes, including the proposed, w/o DA, Heuristic, PDRL, and HDRL schemes. In the early stage, i.e., low QoE range, the proposed scheme does not exhibit superior performance compared to other schemes such as

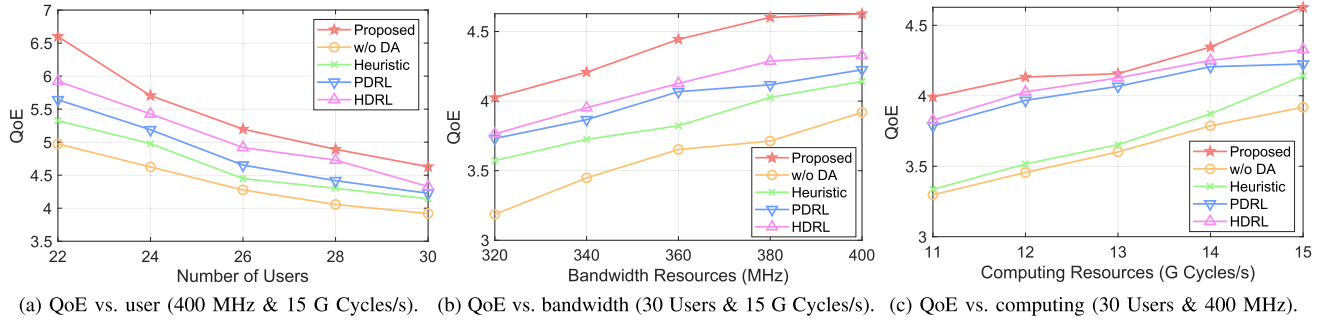


Fig. 8. The QoE performance comparison under different user numbers, bandwidths, and computing resources.

Heuristic and PDRL. This is primarily because the objective of the proposed scheme is to maximize the overall QoE of all users, rather than ensuring fairness across users. Consequently, the proposed scheme sacrifices the QoE performance of some users under poor channel conditions to enable a larger proportion of users to achieve higher QoE. As the QoE increases, this optimization strategy becomes apparent, with the proposed scheme outperforming other methods, as evidenced by its sharper rise and higher cumulative probability at higher QoE values. In contrast, schemes emphasizing fairness, such as HDRL, perform more consistently across all users but fail to achieve the same overall QoE as the proposed scheme. Overall, our proposed scheme can achieve a satisfactory QoE performance compared to other schemes.

D. QoE Performance Under Different Network Resources

We also evaluate the QoE performance under varying network resources, as shown in Fig. 8. As shown in Fig. 8(a), the QoE performance is compared across different numbers of users, with bandwidth and computing resources fixed at 400 MHz and 15 G Cycles/s, respectively. It is observed that as the number of users increases, the QoE of all schemes gradually declines. The proposed scheme exhibits a decrease from an initial value of 6.60 to 4.43, which degraded QoE up to 32.88%. However, it consistently outperforms other schemes, demonstrating the proposed scheme's ability to adapt effectively to varying user numbers and make appropriate network resource allocation decisions. Furthermore, although the QoE gap between the proposed scheme and other schemes narrows with increasing user numbers, the proposed scheme still achieves a QoE that is 17.86% higher than the w/o DA scheme when the user number is 30, which further demonstrates the effectiveness of the proposed DA module.

As shown in Fig. 8(b), the QoE performance is evaluated under varying bandwidth resources, with the number of users and computing resources fixed at 30 and 15 G Cycles/s, respectively. It is observed that as bandwidth resources increase, the QoE of all schemes improves correspondingly. Notably, the QoE performance of the HDRL scheme slightly surpasses that of the PDRL scheme, as the hierarchical reward enables DRL algorithms to explore more efficient network decision-making strategies. However, when the bandwidth resources increase from 380 to 400 MHz, the QoE gain of most schemes becomes

negligible. This is because the current bandwidth resources are sufficient to provide satisfactory networking services, and further increases in bandwidth do not lead to noticeable improvements in service quality. Consequently, the QoE performance eventually stabilizes at a high value.

As depicted in Fig. 8(c), the QoE performance of all schemes is compared under varying computing resources, with the number of users and bandwidth resources fixed at 30 and 400 MHz, respectively. It is observed that as computing resources increase, the QoE gap between the proposed scheme and the baseline schemes initially narrows and then widens beyond 13 G Cycles/s. This phenomenon is primarily attributed to the interplay between computing delay and content quality on the overall QoE. At 13 G Cycles/s, the sensitivity of QoE to computing resources is lower compared to other points, allowing PDRL and HDRL to achieve similar resource allocation decisions as the proposed scheme. However, when computing resources decrease to 11 G Cycles/s, the proposed scheme still achieves a QoE that is 4.18% higher than the HDRL scheme. This demonstrates the ability of the proposed STTD algorithm to effectively adapt to varying computing resources and make satisfactory resource allocation decisions.

E. QoE Performance Under Different DA Data Collection Attenuation Rates

We evaluate how DA data collection attenuation rates affect QoE performance under the proposed DA-assisted resource management scheme, as shown in Table IV. Attenuation rates ν_1, ν_2, ν_3 represent different DA data collection frequency attenuation rates for behavior data, performance data, and environment data. The number of users, bandwidths, and computing resources are set to 30, 400MHz, and 15G Cycles/s, respectively. With the increasing attenuation rate, DA data collection frequency is more sensitive to prediction accuracy, indicating that a low prediction accuracy corresponds to a higher DA data collection frequency and vice versa. It can be observed that when attenuation rates increase, QoE usually increases and gradually increases to a stable state. For instance, when attenuation rate ν_1 increases from 2.00 to 4.00, QoE first increases to a top point 4.63 and then gradually decreases to a stable point 4.52. The reason for this phenomenon is the trade-off between DA data collection cost and ISAC transmission cost. Specifically, with a higher DA data collection frequency, the DA prediction model can have

TABLE IV
THE QoE PERFORMANCE UNDER DIFFERENT DA DATA COLLECTION ATTENUATION RATES

Metric	Attenuation Rate ν_1						Attenuation Rate ν_2						Attenuation Rate ν_3					
	2.00	2.40	2.80	3.20	3.60	4.00	1.50	1.80	2.10	2.40	2.70	3.00	0.50	0.70	0.90	1.10	1.30	1.50
QoE	4.55	4.63	4.57	4.52	4.52	4.52	4.54	4.59	4.63	4.57	4.53	4.50	4.48	4.52	4.57	4.63	4.63	4.63

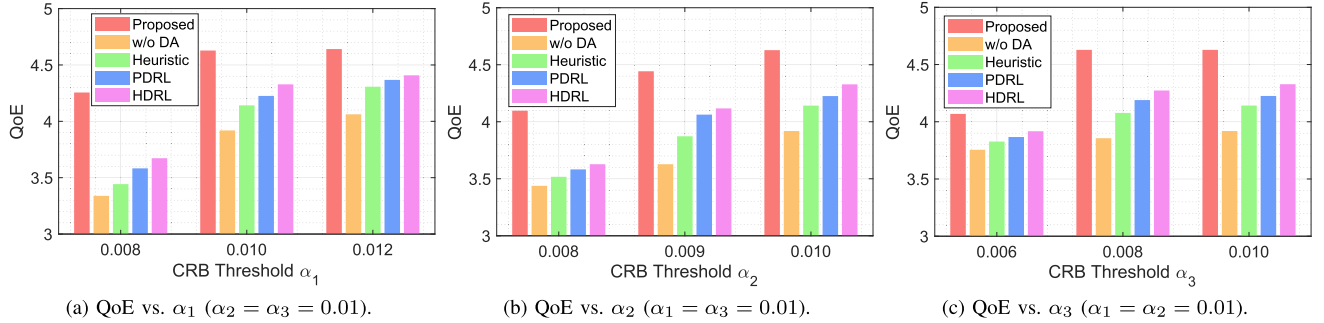


Fig. 9. The QoE performance comparison under different sensing thresholds.

more realistic data to fine-tune its parameters and achieve high accuracy. The more accurate QoE model can guide more tailored resource management, thus increasing QoE. However, higher DA data collection frequency also occupies more communication resources, which decreases data transmission of ISAC services and thus degrades QoE. The trade-off can usually lead to an optimal attenuation rate to maximize QoE. Additionally, different DA data attributes have various sensitive degrees to attenuation rates. Therefore, selecting appropriate attenuation rates for different DA data attributes to maximize QoE is critical.

F. QoE Performance Under Different Sensing Requirements

We further evaluate the QoE performance of different sensing thresholds, as shown in Fig. 9. Specifically, Fig. 9(a) presents the QoE performance comparison of all schemes under different CRB thresholds α_1 , with CRB thresholds α_2 and α_3 fixed at 0.01. It can be observed that as the CRB threshold α_1 gradually increases, the QoE also improves and eventually stabilizes. This behavior can be attributed to looser sensing requirements, which reduce communication resource competition, thereby enhancing QoE. Notably, when the CRB threshold α_1 reaches 0.012, the proposed scheme achieves a QoE that is 14.29% higher than the w/o DA scheme, demonstrating that the proposed DA module effectively enhances user QoE under varying sensing thresholds α_1 .

As shown in Fig. 9(b), we further compare QoE performance under different CRB thresholds α_2 , with CRB thresholds α_1 and α_3 fixed at 0.01. It is observed that as the CRB threshold α_2 decreases, the QoE of all schemes gradually declines. The QoE gap between PDRL and HDRL narrows, decreasing from 0.10 to 0.05, which indicates that the complexity of the network optimization problem limits both DRL algorithms from achieving superior QoE. The proposed scheme consistently achieves the highest QoE among all schemes. Notably, when the CRB threshold α_2 reaches 0.008, the QoE of the proposed scheme remains at 3.84% higher than the HDRL scheme, demonstrating

the effectiveness of the proposed STTD algorithm under varying sensing requirements.

As depicted in Fig. 9(c), the QoE performance of all schemes is further compared under varying CRB thresholds α_3 , when CRB thresholds α_1 and α_2 are fixed at 0.01. It is observed that as the CRB threshold α_3 increases, the QoE performance of all schemes gradually converges to a stable value. For instance, the proposed scheme achieves a QoE of 4.63, while the HDRL scheme stabilizes at 4.33. Under stringent sensing requirements, e.g., $\alpha_3 = 0.006$, the QoE of all schemes experiences significant degradation, as additional communication resources are allocated to sensing tasks, reducing the resources available for QoE enhancement. Nonetheless, the proposed scheme consistently outperforms other schemes, achieving QoE values 3.84% and 8.27% higher than those of the HDRL and w/o DA schemes, respectively. These results validate the effectiveness of the proposed STTD algorithm and DA module in handling varying sensing requirements.

VI. CONCLUSION

In this paper, we have proposed a DA-assisted resource management scheme to maximize long-term user QoE in ISAC networks. A DA module has been proposed to manage user-specific QoE models, and the mathematical relationship between resource allocation and sensing accuracy has been modeled via CRB. The proposed DA-assisted resource management scheme can adaptively update user QoE models based on dynamic user behaviors and environments for tailored resource management. For the future work, we will investigate a DA-assisted service-oriented resource management scheme in ISAC networks.

REFERENCES

- [1] J. A. Zhang et al., "Enabling joint communication and radar sensing in mobile networks—A survey," *IEEE Commun. Surv. Tut.*, vol. 24, no. 1, pp. 306–345, Firstquarter 2022.
- [2] J. Xu et al., "Integrated sensing and communications," *ISAC with Emerging Communication Technologies*. New, York, USA: Springer2023, pp. 589–619. doi: [10.1007/978-981-99-2501-8](https://doi.org/10.1007/978-981-99-2501-8).

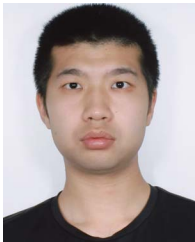
- [3] F. Liu et al., "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.
- [4] H. Kong, C. Huang, J. Yu, and X. Shen, "A survey of mmWave radar-based sensing in autonomous vehicles, smart homes and industry," *IEEE Commun. Surv. Tut.*, vol. 27, no. 1, pp. 463–508, Feb. 2025.
- [5] T. Ma, Y. Xiao, X. Lei, and M. Xiao, "Integrated sensing and communication for wireless extended reality (XR) with reconfigurable intelligent surface," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 5, pp. 980–994, Sep. 2023.
- [6] Z. He, W. Xu, H. Shen, D. W. K. Ng, Y. C. Eldar, and X. You, "Full-duplex communication for ISAC: Joint beamforming and power optimization," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2920–2936, Sep. 2023.
- [7] Y. Gong, Y. Wei, Z. Feng, F. R. Yu, and Y. Zhang, "Resource allocation for integrated sensing and communication in digital twin enabled internet of vehicles," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4510–4524, Apr. 2023.
- [8] X. Huang, W. Wu, S. Hu, M. Li, C. Zhou, and X. Shen, "Digital twin based user-centric resource management for multicast short video streaming," *IEEE J. Sel. Top. Signal Process.*, vol. 18, no. 1, pp. 50–65, Jan. 2024.
- [9] F. Dong, F. Liu, Y. Cui, W. Wang, K. Han, and Z. Wang, "Sensing as a service in 6G perceptive networks: A unified framework for ISAC resource allocation," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3522–3536, Jan. 2022.
- [10] X. Shen, X. Huang, J. Xue, C. Zhou, X. Shi, and W. Zhuang, "Revolutionizing QoE-driven network management with digital agents in 6G," *IEEE Commun. Mag.*, vol. 64, no. 2, pp. 66–73, Feb. 2026.
- [11] Y. Liu, G. Liao, J. Xu, Z. Yang, and Y. Zhang, "Adaptive OFDM integrated radar and communications waveform design based on information theory," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2174–2177, Oct. 2017.
- [12] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, Apr. 2018.
- [13] B. K. Chalise, M. G. Amin, and B. Himed, "Performance tradeoff in a unified passive radar and communications system," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1275–1279, Sep. 2017.
- [14] Z. Xie, R. Li, Z. Jiang, J. Zhu, X. She, and P. Chen, "Optimal scheduling policy for time-division joint radar and communication systems: Cross-layer design and sensing for free," *IEEE Internet Things J.*, vol. 10, no. 23, pp. 20746–20760, Dec. 2023.
- [15] W. Xu, Y. Zhang, J. Wang, L. Zhang, and Q. Li, "Dynamic resource allocation for ISAC enabled internet of vehicles," in *Proc. 30th Annu. Int. Conf. Mobile Comput. Network.*, 2024, pp. 2216–2221.
- [16] B. Zhao, M. Wang, Z. Xing, G. Ren, and J. Su, "Integrated sensing and communication aided dynamic resource allocation for random access in satellite terrestrial relay networks," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 661–665, Feb. 2023.
- [17] X. Gan et al., "Coverage and rate analysis for integrated sensing and communication networks," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 9, pp. 2213–2227, Sep. 2024.
- [18] G. Liu et al., "SensCAP: A systematic sensing capability performance metric for 6G ISAC," *IEEE Internet Things J.*, vol. 11, no. 18, pp. 29438–29454, Sep. 2024.
- [19] Y. He, J. Liu, M. Li, G. Yu, J. Han, and K. Ren, "SenCom: Integrated Sensing and Communication With Practical WiFi," in *Proc. Annual Int. Conf. Mobile Comput. Network.*, Madrid, Spain: Association for Computing Machinery, 2023, vol. 60, pp. 1–16.
- [20] K. Cui, Q. Yang, L. Shen, Y. Zheng, F. Xiao, and J. Han, "Towards ISAC-empowered mmWave radars by capturing modulated vibrations," *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 13787–13803, Dec. 2024.
- [21] J. An, H. Li, D. W. K. Ng, and C. Yuen, "Fundamental detection probability vs. achievable rate tradeoff in integrated sensing and communication systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9835–9853, Dec. 2023.
- [22] H. Hua, T. X. Han, and J. Xu, "MIMO integrated sensing and communication: CRB-rate tradeoff," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 2839–2854, Apr. 2024.
- [23] D. Wen, X. Li, Y. Zhou, Y. Shi, S. Wu, and C. Jiang, "Integrated sensing-communication-computation for edge artificial intelligence," *IEEE Internet Things Mag.*, vol. 7, no. 4, pp. 14–20, Jul. 2024.
- [24] Y. He, G. Yu, Y. Cai, and H. Luo, "Integrated sensing, computation, and communication: System framework and performance optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1114–1128, Feb. 2023.
- [25] 3rd Generation Partnership Project (3GPP), "Study on management aspect of network digital twin," 3GPP, Sophia Antipolis, France, Tech. Rep. TS 28.915, Sep. 2024. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/28_series/28.915/
- [26] C. Zhou et al., "AI-assisted slicing-based resource management for two-tier radio access networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 6, pp. 1691–1706, Dec. 2023.
- [27] M. Li, J. Gao, C. Zhou, L. Zhao, and X. Shen, "Digital-twin-empowered resource allocation for on-demand collaborative sensing," *IEEE Internet Things J.*, vol. 11, no. 23, pp. 37942–37958, Dec. 2024.
- [28] F. Liu, W. Yuan, C. Masouros, and J. Yuan, "Radar-assisted predictive beamforming for vehicular links: Communication served by sensing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7704–7719, Nov. 2020.
- [29] X.-Y. Wang, S. Yang, J. Zhang, C. Masouros, and P. Zhang, "Clutter suppression, time-frequency synchronization, and sensing parameter association in asynchronous perceptive vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 10, pp. 2719–2736, Oct. 2024.
- [30] J. Yan, H. Liu, B. Jiu, B. Chen, Z. Liu, and Z. Bao, "Simultaneous multibeam resource allocation scheme for multiple target tracking," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3110–3122, Jun. 2015.
- [31] X. Huang, X. Qin, M. Li, C. Huang, and X. Shen, "Adaptive digital twin-assisted 3C management for QoE-driven MSVS: A GAI-based DRL approach," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 858–872, Apr. 2025.
- [32] 3rd Generation Partnership Project (3GPP), "Study on RAN-centric data collection and utilization for LTE and NR," 3GPP, Sophia Antipolis, France, Tech. Rep. TS 37.817, Mar. 2022. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/37_series/37.817/
- [33] 3rd Generation Partnership Project (3GPP), "Study on enhancement for data collection for NR and ENDC," 3GPP, Sophia Antipolis, France, Tech. Rep. TS 37.816, Jul. 2019. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/37_series/37.816/
- [34] S. Pezzulli, M. G. Martini, and N. Barman, "Estimation of quality scores from subjective tests-beyond subjects' MOS," *IEEE Trans. Multimedia*, vol. 23, pp. 2505–2519, 2021.
- [35] J. Dueck, D. Edelmann, and D. Richards, "Distance correlation coefficients for lancaster distributions," *J. Multivar. Anal.*, vol. 154, pp. 19–39, 2017.
- [36] H. Zheng, W. Zhao, J. Yang, and A. Bouguettaya, "QoS analysis for web service compositions with complex structures," *IEEE Trans. Serv. Comput.*, vol. 6, no. 3, pp. 373–386, Jul.–Sep. 2013.
- [37] G. Zou, Q. Lu, Y. Chen, R. Huang, Y. Xu, and Y. Xiang, "QoS-aware dynamic composition of web services using numerical temporal planning," *IEEE Trans. Serv. Comput.*, vol. 7, no. 1, pp. 18–31, Jan.–Mar. 2012.
- [38] T. Kimura et al., "QUVE: QoE maximizing framework for video-streaming," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 138–153, Feb. 2017.
- [39] G. Huang, W. Gong, B. Zhang, C. Li, and C. Li, "An online buffer-aware resource allocation algorithm for multiuser mobile video streaming," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3357–3369, Mar. 2020.
- [40] Y. Cai, P. Cheng, Z. Chen, M. Ding, B. Vucetic, and Y. Li, "Deep reinforcement learning for online resource allocation in network slicing," *IEEE Trans. Mobile Comput.*, vol. 23, no. 6, pp. 7099–7116, Jun. 2024.
- [41] M. Garibbo, M. Robeyns, and L. Aitchison, "Taylor TD-learning," in *Proc. Adv. Neural Inf. Process. Syst.*, New Orleans, LA, USA, 2023, vol. 36, pp. 1061–1081.
- [42] J. Buckman, D. Hafner, G. Tucker, E. Brevdo, and H. Lee, "Sample-efficient reinforcement learning with stochastic ensemble value expansion," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, vol. 31, 2018, pp. 8234–8244.
- [43] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, Feb. 2014.
- [44] G. P. McCormick, "Computability of global solutions to factorable non-convex programs: Part I—convex underestimating problems," *Math. Prog.*, vol. 10, no. 1, pp. 147–175, 1976.
- [45] E. Fishler, A. Haimovich, R. Blum, L. Cimini, D. Chizhik, and R. Valenzuela, "Spatial diversity in radars—models and detection performance," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 823–838, Mar. 2006.
- [46] X. Huang, L. He, X. Chen, L. Wang, and F. Li, "Revenue and energy efficiency-driven delay-constrained computing task offloading and resource allocation in a vehicular edge computing network: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8852–8868, Jun. 2022.

- [47] A. Tavakoli, F. Pardo, and P. Kormushev, "Action branching architectures for deep reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, 2018, vol. 32, no. 1, pp. 4131–4138.
- [48] T. Zhao, F. Li, and L. He, "Secure video offloading in MEC-enabled IIoT networks: A multicell federated deep reinforcement learning approach," *IEEE Trans. Ind. Inform.*, vol. 20, no. 2, pp. 1618–1629, Feb. 2024.



Xinyu Huang (Member, IEEE) received the PhD degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2025, the BE degree in Qian Xuesen Experimental Class from Xidian University, Xi'an, China, in 2018, and the MS degree in information and communications engineering from Xi'an Jiaotong University, Xi'an, China, in 2021. He is currently a postdoctoral fellow with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include digital twins, multimedia

communications, ISAC, and network management.



Yixiao Zhang (Graduate Student Member, IEEE) received the BEng degree in communication engineering from the Harbin Institute of Technology (HIT), Shenzhen, China, in 2020, and the MS degree in machine learning from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2023. He is currently working toward the PhD degree with the Department of Electrical and Computer Engineering, the University of Waterloo, Waterloo, ON, Canada. His research interests include network resource management, digital twins, and collaborative robot navigation and control.



Yingying Pei (Member, IEEE) received the PhD degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2026, the BE degree from the Huazhong University of Science and Technology, Wuhan, China, in 2014, and the MSc degree from the University of Macau, Macau, China, in 2019. Her research interests include network resource optimization, mobile-edge computing, and immersive communications.



Jianzhe Xue (Graduate Student Member, IEEE) received the BS degree in communication engineering from Xidian University, Xi'an, China in 2021. He is currently working toward the PhD degree with the School of Electronic Science and Engineering, Nanjing University, China. His current research interests include internet of vehicles, orthogonal time frequency space (OTFS) modulation, and machine learning for wireless communications.



Xuemin (Sherman) Shen (Fellow, IEEE) received the PhD degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently 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, AI for networks, 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, an International fellow of the Engineering Academy of Japan, and a distinguished lecturer of the IEEE Vehicular Technology Society and Communications Society. He was the recipient of the "West Lake Friendship Award" from Zhejiang Province in 2023, President's Excellence in Research from the University of Waterloo in 2022, Canadian Award for Telecommunications Research from the Canadian Society of Information Theory (CSIT) in 2021, 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), Excellent Graduate Supervision Award in 2006 from the University of Waterloo and the Premier's Research Excellence Award (PREA) in 2003 from Province of Ontario, Canada. He is the Past President of the IEEE Communications Society. He was 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 was the Editor-in-Chief of the *IEEE Internet of Things Journal*, *IEEE Network*, and *IET Communications*.