URLLC Edge Networks With Joint Optimal User Association, Task Offloading and Resource Allocation: A Digital Twin Approach

Dang Van Huynh, *Graduate Student Member, IEEE*, Van-Dinh Nguyen, *Member, IEEE*, Saeed R. Khosravirad, *Member, IEEE*, Vishal Sharma, *Senior Member, IEEE*, Octavia A. Dobre, *Fellow, IEEE*, Hyundong Shin, *Fellow, IEEE*, and Trung Q. Duong, *Fellow, IEEE*

***Abstract*— This paper addresses the problem of minimising latency in computation offloading with digital twin (DT) wireless edge networks for industrial Internet-of-Things (IoT) environment via ultra-reliable and low latency communications (URLLC) links. The considered DT-aided edge networks provide a powerful computing framework to enable computation-intensive services, where the DT is used to model the computing capacity of edge servers and optimise the resource allocation of the entire system. The objective function is comprised of local processing latency, URLLC-based transmission latency and edge processing latency, subject to both communication and computation resources budgets. In this regard, the minimum latency is obtained by jointly optimising the transmit power, user association, offloading portions, the processing rate of users and edge servers. The formulated problem is highly complicated due to complex non-convex constraints and strong coupling variables. To deal with this computationally intractable problem, we propose an iterative algorithm which decomposes the original problem into three sub-problems and resolve this problem in the fashion of alternating optimisation approach combined with an inner convex approximation framework. Simulation results demonstrate the effectiveness of the proposed method in reducing the latency compared with other benchmark schemes.**

***Index Terms*— Alternating optimization, digital twin, industrial Internet-of-Things, mobile edge computing (MEC), ultra-reliable and low latency communications (URLLC).**

Manuscript received 13 February 2022; revised 30 June 2022 and 3 September 2022; accepted 4 September 2022. Date of publication 12 September 2022; date of current version 18 November 2022. This work was supported in part by the U.K. Royal Academy of Engineering (RAEng) under the RAEng Research Chair and Senior Research Fellowship Scheme under Grant RCSRF2021\11\41, in part by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2019R1A2C2007037 and 2022R1A4A3033401), and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) through its Discovery program. This paper was presented in part at the IEEE International Conference on Communications (ICC), Seoul, South Korea, in May 2022 [1]. The associate editor coordinating the review of this article and approving it for publication was J. Lee. *(Corresponding authors: Trung Q. Duong; Hyundong Shin.)*

Dang Van Huynh, Vishal Sharma, and Trung Q. Duong are with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast, U.K. (e-mail: dhuynh01@qub.ac.uk; v.sharma@qub.ac.uk; trung.q.duong@qub.ac.uk).

Van-Dinh Nguyen was with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, 1855 Luxembourg City, Luxembourg. He is now with the College of Engineering and Computer Science, VinUniversity, Vinhomes Ocean Park, Hanoi 100000, Vietnam (e-mail: dinh.nv2@vinuni.edu.vn).

Saeed R. Khosravirad is with the Nokia Bell Laboratories, Murray Hill, NJ 07964 USA (e-mail: saeed.khosravirad@nokia-bell-labs.com).

Octavia A. Dobre is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3X5, Canada (e-mail: odobre@mun.ca).

Hyundong Shin is with the Department of Electronics and Information Convergence Engineering, Kyung Hee University, Giheung-gu, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: hshin@khu.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3205692.

Digital Object Identifier 10.1109/TCOMM.2022.3205692

I. Introduction

ADVANCED development of communication technologies and powerful computing architecture can enable a variety of computation-intensive and time-sensitive services. In this regard, the ultra-reliable and low latency communications (URLLC) in 5G provides powerful ability to implement a wide range of mission-critical applications [2]. According to the 3GPP Release 15, URLLC aims for stringent requirement with the targeted 1 ms latency and block error rate (BLER) of $10^{-9} - 10^{-5}$ depending on the use cases [3]. This technology opens opportunities to empower various applications in different domains, including factory automation, autonomous vehicles, e-healthcare, and immersive applications. Therefore, combining URLLC with other emerging technologies has attracted much interest from active research groups in both academia and industry. However, there are still many open issues and research challenges in the implementation of URLLC-based communication in practical scenarios due to the complex relationship between the low latency and high reliability requirements [4].

In terms of computing, the mobile edge computing (MEC) technology leverages the powerful computation and storage capacity of nearby edge servers to reduce the overall latency of services. In MEC, task offloading is a key technique that allows constrained devices to fully or partially offload computation-intensive tasks to higher layers equipped with more powerful processors to minimise the latency [5]. Task offloading in the edge computing paradigm opens tremendous opportunities for enabling a large number of novel Internet-of-things (IoT) applications with high quality-of-experience (QoE) and quality-of-service (QoS) requirements. Nevertheless, designing an effective solution for task offloading in edge computing is still challenging due to many issues such as joint computations and communications, heterogeneous architecture, task integration [6], [7], [8].

Recently, digital twin (DT) has emerged as a promising technology which is able to create virtual twins of physical

objects and benefit many domains such as real-time management. DT was firstly introduced by Michael Grieves in 2002 based on the concept of product lifecycle management (PLM) [9]. In 2010, the first practical application of DT was developed by NASA with two identical space vehicles for the Apollo project. During the period from 2011 to present, DT has attracted more attention and its applications span in various domains such as manufacturing, healthcare, entertainment, intelligent transport system, and 6G networks. Therefore, integrating DT with other promising technologies, such as MEC and URLLC opens many opportunities as well as research challenges to enable a next generation of future networked systems.

A. Literature Review

Recent studies in MEC task offloading have mainly focused on the development of optimal designs, aiming to reduce the response time delay [10], [11] and improve energy efficiency [12], [13], [14] and resource management [15], [16], [17], [18]. In particular, a game-theoretic model and a distributed best-response solution have been introduced in [10] to deal with a problem of maximising the expected offloading rate in multiple agents MEC networks. Another distributed solution for optimising long-term average of response time delay with optimal computation offloading strategies was investigated in [11]. In [12], an energy-efficient problem in offloading was addressed and solved with distributed optimisation algorithms based on the subgradient and alternating direction method of multipliers. The energy-efficiency issue was also considered in [14], where an NP-hard intractable problem was tackled by a distributed framework based on a parallel processing approach. In [15], the resource procurement and allocation decisions have been investigated with both offline and online optimisation algorithms. An improved hierarchical adaptive search algorithm has been proposed in [16] to minimise the ultra-dense IoT network energy consumption with joint device association, resource allocation and computation offloading. From the economic perspectives, resource allocation in task offloading has been explored to maximise the utility of services subject to their budget constraints [18].

In order to meet the stringent requirements of low latency and ultra-high reliability, URLLC with finite blocklength [19] has attracted much attention recently. Due to the complexity of short packet transmissions in URLLC, many studies have newly focused on the resource allocation problem for URLLC in industrial automation [20], [21], [22], [23]. In particular, an intensive resource allocation and beamforming design for downlink URLLC systems has been investigated with path-following algorithms in [20]. In [21], a decoding error probability minimisation problem subject to transmit power and latency constraints has been addressed, which is also solved by a path-following algorithm. For uplink URLLC transmission, [22] introduced the closed-form expression of lower bound of achievable date rate for massive multiple-input multiple-output (MIMO) systems and then a joint pilot and payload transmission optimisation was provided for both maximum-ratio combining (MRC) and zero-forcing (ZF) designs. The resource optimisation for URLLC was also presented in [23], where URLLC requirements, application scenarios, and mathematical tools for optimising resource allocation were sufficiently suggested. More recently, combining URLLC with other emerging technologies such as MEC [24], digital twin [25] for mission-critical applications, e.g., Industry 4.0 [26], augmented/virtual reality (AR/VR) has attracted much attention. However, this is still in its infancy, and there are many research issues that can be explored to contribute to this research area [27].

As a promising technology for the next generation of industrial automation, the development of DT opens new opportunities for transforming the cyber-physical systems in terms of intelligence, efficiency and flexibility [9]. In [28], an intelligent cooperation of unmanned aerial vehicle (UAV) swarms combining machine learning and the DT framework has been discussed, which provides more efficient decisions for real-time management. In combining with the MEC architecture, many active researchers recently have shown great interest in DT-assisted task offloading [29], [30], [31], [32]. In this regard, a DT edge network (DITEN) was introduced in [29] to minimise the offloading latency subject to consumed service migration cost based on the Lyapunov optimisation method and deep reinforcement learning (DRL). In [30], DT has combined with DRL to find the optimal solution for edge association in reducing system cost and enhancing the convergence rate with respect to the dynamic network states. Another DT solution for task offloading based on edge collaboration was investigated in [31], where a DRL-based solution was explored to reduce the system power overhead and network delay. In [32], the latency minimisation problem of the DT-aided offloading in edge network has recently been addressed with the alternating optimisation approach.

Overall, URLLC, MEC and DT are three promising technologies to enable the next generation of industrial IoT applications. Each technology itself consists of many technical hurdles to deal with, e.g., short packet transmission in URLLC, optimal resource scheduling in MEC, and integrating DT for management. Therefore, combining these challenging technologies in practical scenarios is likely to open a wide range of opportunities as well as challenges for exploring and contributing to the research community.

B. Motivation and Contributions

Recently, the problem of computation offloading in edge computing with the support of DT has been investigated in the literature [29], [30], [31], [32], [33], [34], [35]. The DT concept has been shown to empower edge networks by creating a digital replica of physical systems. DTs can optimise resource allocation and make accurate decisions to improve system performance. To doing so, recent studies are focusing on the model of dynamic computing allocation in DT-assisted edge networks [29], [30], [31]. The deviation between the estimated processing rate and the real-world configurations of devices is carefully considered to practically model the DT-aided edge network. However, these studies mainly focus on the general task offloading in the MEC architecture while the integration with URLLC and short packet transmission has not taken into consideration. For instance, [33] and [34] have adopted DRL-based algorithms to address the problem of task offloading for the scenarios of industrial IoT and vehicular edge networks, respectively. Again, although these studies target industrial IoT applications, short packet transmissions in URLLC are not considered. In addition, entirely joint

user association, offloading decisions, and resource allocation solutions have not been fully addressed in these studies. The combined problem of MEC, URLLC and DT was firstly investigated in [25], which aims to empower hybrid MEC systems, including URLLC and delay-tolerant services. The solution proposed in [25] targeted to minimise the energy consumption by optimising user association, resource allocation and offloading portions under URLLC-based transmission. A deep learning (DL) architecture was developed with the support of DT to find optimal resource allocation and offloading decisions. However, the important computation parameters (e.g. the processing rate of users and edge servers) has not been considered. In addition, the impact of DT on the system performance (i.e. computing capacity) and the certain deviation existing in DT deployment have been not taken into account in [25]. Thus so far, designing effective solutions for task offloading in MEC with DT technology is still an open research issue, especially for mission-critical applications in industrial scenarios.

Moving beyond the above background, this work proposes a joint communication and computation offloading in URLLC-based edge networks with DT that takes into account all the above issues. We aim to minimise the worst-case latency of task offloading by optimising the user association, transmission power and the processing rate of user equipments (UEs), the offloading portions, and the processing rate of edge servers (ESs). The problem is formulated based on edge network architecture under the DT paradigm, while the wireless communications between UEs and ESs are established via URLLC links, which is applicable in the context of industrial automation. The main contributions of this paper are summarised as follows:

- We first formulate the end-to-end (e2e) latency of DT-empowered URLLC edge networks with joint user association and computation task offloading. The considered problem takes into account both computation and communication aspects. Specifically, the generalised e2e latency minimisation problem deals with user association, transmit power, offloading policies and processing rate of UEs and ESs subject to the service delay and energy consumption requirements of UEs, resulting in a mixed-integer non-convex optimisation problem.
- Combining tools from the inner approximation (IA) framework and alternating optimisation (AO) approach, we develop a low-complexity algorithm to solve the formulated problem in an iterative manner. In particular, we first decompose the original problem into three sub-problems comprising of optimal user association, resource optimisation and optimal offloading. Then, the approximate convex functions are developed base on IA framework to convexify the non-convex constraints which can be solved by the AO-based algorithm.
- For comparison, we provide sub-optimal designs to solve the problem for given user association strategies, *namely* the heuristic approach (HEU) based on the channel condition and random association (RAN). In particular, the heuristic approach directly matches UEs with access points (APs) that have the best channel condition while UEs and APs are randomly assigned to the random user associations.
- Finally, extensive simulations are provided to obtain key insights and investigate the impacts of the involved parameters on the e2e latency of the system. The provided numerical results prove that the proposed solution effectively minimises the e2e latency of the DT-aided URLLC-based wireless edge networks.

C. Paper Structure and Notations

The remainder of this paper is organised as follows. The system model and problem formulation for the DT-aided URLLC-based edge networks are described in Section II. Section III presents our proposed solution to deal with the computation offloading problem taking into account various communication and computation variables. For comparison, we propose sub-optimal designs for user association in Section IV. Next, numerical results and discussions are provided in Section V. Finally, concluding remarks are given in Section VI.

Notation: Throughout the paper, numbers are denoted in lowercase while matrices and vectors are written as bold uppercase and lowercase letters, respectively. We use the notation $x \sim \mathcal{CN}(.,.)$ to represent that x is complex circularly symmetric Gaussian distributed. $\|\cdot\|$ stands for the vector's Euclidean norm and $\mathbb{C}$ denotes the set of all complex numbers. Finally, we denote x_{mk} as a variable x that involves the m-th user (UE) and k-th edge server (ES).

II. System Model and Problem Formulation

A. DT-Empowered Edge Networks With URLLC Model

Fig. 1 illustrates a system model of DT-empowered edge network architecture with URLLC for industrial automation. In this model, the physical layer consists of many UEs and ESs. Meanwhile, the DT layer is able to virtually replicate the physical system, optimise resources, and make decisions to control the whole system via the real-time interaction mechanism. To guarantee stringent requirements on low latency and ultra-high reliable communications, UEs connect with ESs via URLLC links, which is highly applicable for the scenarios of industrial automation. Subsections below fully describe the system model for formulating a latency minimisation problem.

1) URLLC-Based MEC Architecture: We assume M UEs in the set of $\mathcal{M} = \{1, 2, \ldots, M\}$ and K ESs in the set of $\mathcal{K} = \{1, 2, \ldots, K\}$. Each ES is associated with an access point (AP) to provide connections of UEs. User associations of UEs and ESs are indicated by binary variables $\boldsymbol{\pi} = \{\pi_{mk}\}_{\forall m,k} = \{0, 1\}$, when $\pi_{mk} = 1$ there is a connection from m-th UE to the k-th ES; otherwise, $\pi_{mk} = 0$. To guarantee performance, we assume that each ES only serves a maximum of $M_{\max}$ UEs, i.e., $\sum_{m \in \mathcal{M}} \pi_{mk} \leq M_{\max}, \forall k \in \mathcal{K}$. The transmissions between UEs and APs are based on URLLC links with short packet communications to meet high demands on reliability and low latency in industrial automation.

2) Offloading Model in Edge Networks: A computational task from the m-th UE is characterised by $J_m = \{\eta_m, T_m^{\max}\}$, where $\eta_m = C_m/D_m$ is the task complexity (cycles/bit), in which D_m is the task size (bits), C_m is the required CPU cycles (cycles) to execute the task, and $T_m^{\max}$ (s) is the maximum latency for task J_m.

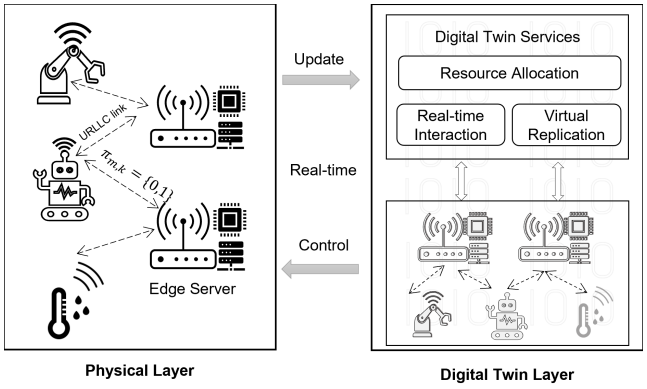


Fig. 1. An exemplary illustration of the DT-empowered URLLC-based edge network model.

In MEC, there are two main approaches for the task offloading model, including binary offloading [36], [37] and partial offloading. Although the binary offloading is considered as a simple solution to implement practical systems, the partial offloading scheme has been widely investigated for delay-constrained applications [11], [25]. In addition, the partial offloading allows the MEC systems to take advantages of parallel processing, which can further reduce the overall latency [38]. Therefore, we consider the partial offloading model in this paper. Let $\boldsymbol{\alpha} = \{\alpha_m\}_{\forall m}$ be a portion of the task processed locally, and $\boldsymbol{\beta} = \{\beta_{mk}\}_{\forall m,k}$ be the offloading factors of the m-th UE to the k-th ES, which satisfy $0 \leq \alpha_m \leq 1$, $0 \leq \beta_{mk} \leq 1$. We assume that the computational task can be planned at a high granularity to provide the partial task offloading ability. By this way, with the J_m task coming from the m-th UE, we have $D_m = \alpha_m D_m + \sum_{k\in K} \pi_{mk}\beta_{mk} D_m$ and $C_m = \alpha_m C_m + \sum_{k\in K} \pi_{mk}\beta_{mk} C_m$, satisfying $\alpha_m + \sum_{k\in K} \pi_{mk}\beta_{mk} = 1$.

3) DT Model: The DT services virtually replicate the physical systems which includes the hardware information, operating applications, real-time states. The DT can interact in real-time with the physical system to control and manage efficiently. According to [9], from the perspective of technical implementation, the main principles of DT are the model of system and data exchanged by the model. DT collects and analyses the state information of objects via functional components, including real-time monitoring and dynamic control services. There are several existing tools and libraries that implement the DT concept in real-world products, such as Modelica, DELMIA, FlexSim, Automod, etc. [9].

The DT of URLLC-based MEC model can be represented as

$$\mathrm{DT} = \{\tilde{\mathcal{M}}, \tilde{\mathcal{K}}\} \tag{1}$$

where $\{\tilde{\mathcal{M}}, \tilde{\mathcal{K}}\}$ are the virtual representation of the physical systems including M UEs and K ESs. Based on real-time updated information from physical objects, digital services in the DT layer provides comprehensive functions to manage and control the system automatically. These services can be listed as data collection, visualisation, analysis, decision-making, and real-time optimisation. The DT services have to promptly provide optimised solutions on tasks offloading, edge selection, estimated processing rate as well as transmit power allocation to guarantee the performance of the entire system.

The DT model for the m-th UE is denoted by $\mathrm{DT}_m^{\mathsf{lo}}$, which is expressed as

$$\mathrm{DT}_m^{\mathsf{lo}} = (f_m^{\mathsf{lo}}, \hat{f}_m^{\mathsf{lo}}) \tag{2}$$

where f_m^{lo} is the estimated processing rate, and $\hat{f}_m^{\mathsf{lo}}$ is the deviation in the processing rate between the physical device and its DT [29], [30], [32]. The DT layer has the estimated processing rate f_m^{lo} to replicate the behaviours of UEs and trigger decisions on optimising physical devices configuration. In this paper, the estimated processing rate is considered as the optimisation variables. The deviation is pre-defined as a percentage of the estimated processing rate for simulations. Similarly, for the k-th ES, its DT ($\mathrm{DT}_k^{\mathsf{es}}$) can be expressed as

$$\mathrm{DT}_k^{\mathsf{es}} = (f_{mk}^{\mathsf{es}}, \hat{f}_{mk}^{\mathsf{es}}) \tag{3}$$

where f_{mk}^{es} is the estimated processing rate of the real ES and $\hat{f}_{mk}^{\mathsf{es}}$ is the deviation value in processing rate estimation between the real ES and its DT. The DT replica of ESs provide the estimated processing rate of ESs to allocate the computing capacity of ESs. This allows the DT services to minimise the processing latency by adjusting offloading portions, user association and computing resource allocation.

B. Communication Model

1) Transmission Model: Each AP is equipped with L antennas and each UE is equipped with single antenna. Let $\mathbf{h}_{mk} \in \mathbb{C}^{L\times 1}$ be the channel vector between the k-th BS and the m-th UE and can be modelled as $\mathbf{h}_{mk} = \sqrt{g_{mk}}\bar{\mathbf{h}}_{mk}$, where g_{mk} denotes the large-scale channel coefficient including the pathloss and shadowing and $\bar{\mathbf{h}}_{mk}$ denotes the small-scale fading following the distribution of $\mathcal{CN}(0, \mathbf{I})$. Let us denote $\mathbf{H}_k \in \mathbb{C}^{L\times M}$ as the channel matrix between M devices and the k-th AP, with $\mathbf{H}_k = [\mathbf{h}_{1k}, \mathbf{h}_{2k}, .., \mathbf{h}_{Mk}]$. Then, the $L \times 1$ received signal vector at the k-th AP is given by $\mathbf{y}_k = \sum_{m=1}^{M} \mathbf{h}_{mk}\sqrt{p_{mk}}s_{mk} + \mathbf{n}_k$, where p_{mk} is the transmission power of the m-th device, s_{mk} is the zero mean and unit variance Gaussian information message from the m-th UE, and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}_L)$ is the additive white Gaussian noise (AWGN) during the data transmission in which N_0 is the noise power.

To improve the wireless transmission performance, we adopt the matched filtering and successive interference cancellation (MF-SIC) at APs. This technique has been widely used for the uplink transmission, which can improve the throughput of the user by removing the strong interference [39]. By applying MF-SIC technique, the decoding order is followed the UEs' index by ordering the channel vector as $\|\mathbf{h}_{1k}\|^2 \geq \|\mathbf{h}_{2k}\|^2 \geq , \cdots, \geq \|\mathbf{h}_{Mk}\|^2, \forall k$. Then, the signal-to-interference-plus-noise (SINR) at the k-th AP of signal transmitted by the m-th UE can be expressed as

$$\gamma_{mk}(\mathbf{p}, \boldsymbol{\pi}) = \frac{\pi_{mk} p_{mk} \|\mathbf{h}_{mk}\|^2}{\mathcal{I}_{mk}(\mathbf{p}, \boldsymbol{\pi}) + N_0} \tag{4}$$

where $\mathcal{I}_{mk}(\mathbf{p}, \boldsymbol{\pi}) = \sum_{n>m}^{M} \pi_{nk} p_{nk} \frac{|\mathbf{h}_{mk}^H \mathbf{h}_{nk}|^2}{\|\mathbf{h}_{mk}\|^2}$ is the interference power produced by UEs $n > m$.

2) URLLC Uplink Transmission Rate: The approximation of transmission rate (bits/s) in URLLC-based connections is given by [22], [23]

$$R_{mk}(\mathbf{p}, \boldsymbol{\pi}) \approx B \log_2 [1 + \gamma_{mk}(\mathbf{p}, \boldsymbol{\pi})]$$

$$-B\sqrt{\frac{V_{mk}\left(\mathbf{p},\boldsymbol{\pi}\right)}{N}}\frac{Q^{-1}\left(\epsilon\right)}{\ln 2} \tag{5}$$

where B is the system bandwidth, N is the blocklength, ϵ is decoding error probability, $\gamma_{mk}\left(\mathbf{p},\boldsymbol{\pi}\right)$ denotes the SINR, $Q^{-1}(.)$ is the inverse function $Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty \exp^{\left(\frac{-t^2}{2}\right)}\,dt$, and V_{mk} is the channel dispersion given by $V_{mk}\left(\mathbf{p},\boldsymbol{\pi}\right)=1-\left[1+\gamma_{mk}\left(\mathbf{p},\boldsymbol{\pi}\right)\right]^{-2}$. When the blocklength N approaches to infinity, the data rate R_{mk} approaches the classic Shannon's equation.

Then, the uplink wireless transmission latency from the m-th UE to the k-th ES for task offloading can be expressed as

$$T_{mk}^{\mathsf{co}}\left(\mathbf{p},\boldsymbol{\pi},\beta_{mk}\right)=\max_{\forall k\in\mathcal{K}}\left\{\frac{\beta_{mk}D_m}{R_{mk}\left(\mathbf{p},\boldsymbol{\pi}\right)}\right\}. \tag{6}$$

C. Computation Model

1) Local Processing: The computational task J_m at m-th UE is partially executed a portion of α_m with the estimated processing rate f_m. As a result, the estimated latency to process the task locally is given by

$$\tilde{T}_m^{\mathsf{lo}}\left(\alpha_m,f_m^{\mathsf{lo}}\right)=\frac{\alpha_m C_m}{f_m^{\mathsf{lo}}}. \tag{7}$$

Assuming the deviation between the physical m-th UE and its DT representation can be acquired in advance and the computing latency gap between real latency and DT estimation can be obtained by

$$\Delta T_m^{\mathsf{lo}}\left(\alpha_m,f_m^{\mathsf{lo}}\right)=\frac{\alpha_m C_m\hat{f}_m^{\mathsf{lo}}}{f_m^{\mathsf{lo}}\left(f_m^{\mathsf{lo}}-\hat{f}_m^{\mathsf{lo}}\right)}. \tag{8}$$

Consequently, the actual time for local processing is given by

$$T_m^{\mathsf{lo}}=\Delta T_m^{\mathsf{lo}}+\tilde{T}_m^{\mathsf{lo}}. \tag{9}$$

2) Edge Processing: At the k-th ES, let us denote the estimated processing rate of the k-th ES as f_{mk}^{es}, the estimated latency of the k-th ES to process the offloaded task J_m is given by

$$\tilde{T}_{mk}^{\mathsf{es}}\left(\pi_{mk},\beta_{mk},f_{mk}^{\mathsf{es}}\right)=\max_{\forall k\in\mathcal{K}}\left\{\frac{\pi_{mk}\beta_{mk}C_m}{f_{mk}^{\mathsf{es}}}\right\}. \tag{10}$$

Then, the gap ΔT_m^{ed} between real latency value and DT estimation latency can be calculated as

$$\Delta T_{mk}^{\mathsf{es}}\left(\pi_{mk},\beta_{mk},f_{mk}^{\mathsf{es}}\right)=\frac{\pi_{mk}\beta_{mk}C_m\hat{f}_{mk}^{\mathsf{es}}}{f_{mk}^{\mathsf{es}}\left(f_{mk}^{\mathsf{es}}-\hat{f}_{mk}^{\mathsf{es}}\right)}. \tag{11}$$

As a result, the actual DT latency to process the offloaded task at ESs can be expressed as

$$T_{mk}^{\mathsf{es}}=\Delta T_{mk}^{\mathsf{es}}+\tilde{T}_{mk}^{\mathsf{es}}. \tag{12}$$

By defining the DT latency of local processing and edge processing in this way, if the deviation increases, the processing time of UEs and ESs rises that results in decreasing the performance of the system. This makes sense in practical scenarios, where the DT tries to estimate as exactly as possible the physical system to perform real-time control and management.

D. Latency Model

The considered end-to-end (e2e) DT latency in the system can be expressed as follows

$$\begin{aligned}T_m^{\mathsf{e2e}}&=T_m^{\mathsf{lo}}+T_{mk}^{\mathsf{co}}+T_m^{\mathsf{es}}=\frac{\alpha_m C_m}{f_m^{\mathsf{lo}}-\hat{f}_m^{\mathsf{lo}}}\\&\quad+\max_{\forall k\in\mathcal{K}}\left\{\frac{\pi_{mk}\beta_{mk}D_m}{R_{mk}\left(\mathbf{p},\boldsymbol{\pi}\right)}\right\}+\max_{\forall k\in\mathcal{K}}\left\{\frac{\pi_{mk}\beta_{mk}C_m}{f_{mk}^{\mathsf{es}}-\hat{f}_{mk}^{\mathsf{es}}}\right\}\end{aligned} \tag{13}$$

which includes the local processing latency at UEs, transmission latency for task offloading, and the edge processing latency at ESs. As mentioned in the task offloading model, the computational task can be offloaded to multiple ESs simultaneously; however, there is a considerable variety of the offloaded portions, transmission time, and ESs' computing capacity to execute the task. To reduce the straggler effect and improve the fairness in the optimisation solution, we apply the maximum operator, i.e., $\max(.)$ in the transmission latency and the ESs' processing latency in (13). Since computation responses from ESs to UEs are typically very small (e.g., control messages) and APs can deliver these signals with very high power, the downlink transmission latency is negligible [5], [11], [29].

E. Energy Consumption Model

Total energy consumption of the m-th UE can be modelled as

$$\begin{aligned}&E_m^{\mathsf{tot}}\left(\alpha_m,\boldsymbol{\beta},\mathbf{p},\boldsymbol{\pi},f_m^{\mathsf{lo}}\right)\\&\quad=E_m^{\mathsf{cp}}+E_m^{\mathsf{cm}}\\&\quad=\alpha_m\frac{\theta}{2}C_m\left(f_m^{\mathsf{lo}}-\hat{f}_m^{\mathsf{lo}}\right)^2+\sum_{k=1}^{K}p_{mk}\frac{\beta_{mk}\pi_{mk}D_m}{R_{mk}\left(\mathbf{p},\boldsymbol{\pi}\right)}\end{aligned} \tag{14}$$

where E_m^{cp}, E_m^{cm} are the energy consumption expression for computation and communication, respectively while the constant $\theta_m/2$ represents the average switched capacitance and the average activity factor of the m-th UE [24].

F. Problem Formulation

In this paper, we aim to minimise the worst-case of e2e DT latency by optimising user association, offloading policies, transmission power and the estimated processing rates of UEs and ESs, subject to constraints of quality-of-service (QoS), energy consumption budget and computation resources budget of UEs and ESs. The problem is mathematically formulated as (15). The constraint (15b) presents maximum latency constraint for every incoming task. The constraint (15c) means that each ES can serve maximum of $M_{\max}$ UEs. The constraints (15e) and (15f) are the requirement of minimum transmission rate and the maximum energy consumption requirement of the UEs, respectively. Finally, the constraint (15g) ensures the computation resources of ESs are not allocated in excess.

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi},\mathbf{p},\mathbf{f}}\max_{\forall m\in\mathcal{M}}\left\{T_m^{\mathsf{e2e}}\left(\boldsymbol{\pi},\alpha_m,\beta_{mk},\mathbf{f},\mathbf{p}\right)\right\} \tag{15a}$$

$$\text{s.t. } T_m^{\mathsf{e2e}}\left(\boldsymbol{\pi},\alpha_m,\beta_{mk},\mathbf{f},\mathbf{p}\right)\le T_m^{\max},\forall m \tag{15b}$$

$$\sum_{m\in\mathcal{M}}\pi_{mk}\le M_{\max},\forall k \tag{15c}$$

$$\alpha_m + \sum_{k\in K} \pi_{mk}\beta_{mk} = 1, \forall m \tag{15d}$$

$$R_{mk}\left(\mathbf{p}, \boldsymbol{\pi}\right) \geq \pi_{mk} R_{\min}, \forall m, k \tag{15e}$$

$$E_m^{\mathsf{tot}}\left(\boldsymbol{\pi}, \alpha_m, \boldsymbol{\beta}, \mathbf{p}, f_m^{\mathsf{lo}}\right) \leq E_m^{\max}, \forall m \tag{15f}$$

$$\sum_{m\in\mathcal{M}} \pi_{mk}\beta_{mk} f_{mk}^{\mathsf{es}} \leq F_{\max}^{\mathsf{es}}, \forall k \tag{15g}$$

$$\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathscr{D}, \boldsymbol{\pi} \in \Pi, \mathbf{p} \in \mathscr{P}, \mathbf{f} \in \mathscr{F} \tag{15h}$$

where $\mathscr{D} \triangleq \{\alpha_m, \beta_{mk}, \forall m, k | 0 \leq \alpha_m \leq 1, 0 \leq \beta_{mk} \leq 1, \forall m, k\}$, $\mathscr{P} \triangleq \{p_{mk}, \forall m, k | 0 \leq \sum_{k\in\mathcal{K}} \pi_{mk} p_{mk} \leq P_m^{\max}, \forall m\}$, $\mathscr{F} \triangleq \{f_m^{\mathsf{lo}}, f_{mk}^{\mathsf{es}}, \forall m, k | 0 \leq f_m^{\mathsf{lo}} \leq F_{\max}^{\mathsf{lo}}, \forall m; 0 \leq f_{mk}^{\mathsf{es}} \leq F_{\max}^{\mathsf{es}}, \forall k\}$, and $\Pi \triangleq \{\pi_{mk}, \forall m, k | \pi_{mk} \in \{0, 1\}, \forall m, k\}$ are the set constraints of offloading decisions, uplink transmission power, processing rates, and association policies, respectively.

III. Proposed Solution

As we can see from problem (15), the objective function (15a) is non-concave and non-smooth while the constraints (15b) to (15h) are highly complicated non-convex constraints. In addition, there are binary variables $\boldsymbol{\pi} \in \Pi$ strongly coupling with other continuous variables, which produces a mixed-integer non-convex optimisation problem. As a result, solving (15) directly with a well-known brute-force search (BFS) approach or branch-and-bound method is computationally challenging and inapplicable, especially in large-scale scenarios.

Therefore, to deal with (15), we first transform the original problem into a computationally tractable problem by replacing the objective function with its upper bound function to provide a smooth linear function. In this regard, with introduced variables $\mathbf{t} \triangleq \{t^{\mathsf{lo}}, t^{\mathsf{co}}, t^{\mathsf{es}}\}$ that satisfy $\tau_m\left(t^{\mathsf{lo}}, t^{\mathsf{co}}, t^{\mathsf{es}}\right) \triangleq t^{\mathsf{lo}} + t^{\mathsf{co}} + t^{\mathsf{es}}$, problem (15) can be equivalently transformed to

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi},\mathbf{p},\mathbf{f},\mathbf{t}} \quad \max_{\forall m\in\mathcal{M}} \{\tau_m(\mathbf{t})\} \tag{16a}$$

$$\text{s.t.} \quad (15c) - (15h) \tag{16b}$$

$$\tau_m(\mathbf{t}) \leq T_m^{\max}, \forall m \tag{16c}$$

$$t^{\mathsf{lo}} \geq \frac{\alpha_m C_m}{f_m^{\mathsf{lo}} - \hat{f}_m^{\mathsf{lo}}}, \forall m \tag{16d}$$

$$t^{\mathsf{co}} \geq \frac{\pi_{mk}\beta_{mk} D_m}{R_{mk}^{ul}\left(\mathbf{p}, \boldsymbol{\pi}\right)}, \forall m, k \tag{16e}$$

$$t^{\mathsf{es}} \geq \frac{\pi_{mk}\beta_{mk} C_m}{f_{mk}^{\mathsf{es}} - \hat{f}_{mk}^{\mathsf{es}}}, \forall m, k. \tag{16f}$$

Lemma 1: Let $(\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star, \boldsymbol{\pi}^\star, \mathbf{p}^\star, \mathbf{f}^\star, \mathbf{t}^\star)$ *be the optimal solution to problem* (16)*, then* $(\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star, \boldsymbol{\pi}^\star, \mathbf{p}^\star, \mathbf{f}^\star)$ *is also the optimal solution to problem* (15) *and vice versa.*

Proof: Please refer to Appendix A. □

Next, to deal with the highly complex non-convex problem (16), we decompose (16) into three sub-problems and solve the problem by applying the AO-IA framework [40], [41]. Three decomposed sub-problems, namely user association optimisation, computation and communication resources optimisation, and offloading policies optimisation which are solved at the i-th iteration. The following subsections clearly demonstrate the development of our proposed solution by approximating non-convex constraints in the three sub-problems.

A. User Association Optimisation

In this sub-problem, we solve problem (16) with fixed values of $(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)}, \mathbf{f}^{(i)}, \mathbf{p}^{(i)})$ to find the next optimal of user association $(\boldsymbol{\pi}^{(i+1)})$. The addressed sub-problem for optimising user association strategies is given as follows

$$\textsf{SP-1:} \quad \min_{\boldsymbol{\pi},\mathbf{t}|\mathbf{p}^{(i)},\mathbf{f}^{(i)},\boldsymbol{\alpha}^{(i)},\boldsymbol{\beta}^{(i)}} \quad \max_{\forall m\in\mathcal{M}} \{\tau_m(\mathbf{t})\} \tag{17a}$$

$$\text{s.t.} \quad (15c) - (15h), (16c), (16e), (16f). \tag{17b}$$

As we can observe from (17), the constraints (15e), (15f), and (16e) are non-convex. Therefore, we process these constraints through the following approximations to generate a convex program of (17).

Convexity of (15e)**:** To address the non-convex constraint (15e), we first rewrite $\gamma_{mk}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right) = \frac{\pi_{mk}}{q_{mk}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right)}$, where $q_{mk}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right)$ is defined as

$$q_{mk}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right) \triangleq \frac{\mathcal{I}_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + N_0}{p_{mk}^{(i)} \|\mathbf{h}_{mk}\|^2}, \forall m, k. \tag{18}$$

We then express the uplink transmission rate in (5) as follows

$$R_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) = \frac{B}{\ln 2}\left[G_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) - \kappa W_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi})\right], \tag{19}$$

where $G_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) = \ln(1 + \gamma_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}))$, $W_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) = \sqrt{1 - \left[1 + \gamma_{mk}\left(\mathbf{p}, \boldsymbol{\pi}\right)\right]^{-2}}$ and $\kappa = Q^{-1}(\epsilon)/\sqrt{N}$.

By following Appendix C, the approximation of (19) can be expressed as

$$\begin{aligned} R_{mk}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right) &\geq R_{mk}^{(i)}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right) \\ &\triangleq \frac{B}{\ln 2}\left[\mathcal{G}_{mk}^{(i)}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right) - \kappa \mathcal{W}_{mk}^{(i)}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right)\right] \end{aligned} \tag{20}$$

under the following trusted regions [20]:

$$q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk} \leq 2\left(q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}_m) + \pi_{mk}^{(i)}\right), \tag{21}$$

$$\frac{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}}{q_m(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)}) + \pi_{mk}^{(i)}} \leq 2\frac{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi})}{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})}, \forall m, k \tag{22}$$

where $\mathcal{G}_{mk}^{(i)}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right)$, and $\mathcal{W}_{mk}^{(i)}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right)$ are defined as (47) and (53) in the Appendix C. Consequently, we innerly approximate constraint (15e) as

$$R_{mk}^{(i)}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}\right) \geq \pi_{mk} R_{\min}, \forall m, k. \tag{23}$$

Convexity of (15f)**:** By introducing new variables $\hat{\mathbf{r}} \triangleq \{\hat{r}_{mk}\}_{\forall m,k}$ that satisfy $\frac{1}{R_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi})} \leq \hat{r}_{mk}$, (15f) is now equivalent to

$$\begin{cases} \alpha_m^{(i)} \dfrac{\theta}{2} C_m \left(f_m^{\mathsf{lo},(i)} - \hat{f}_m^{\mathsf{lo}}\right)^2 \\ \quad + D_m \displaystyle\sum_{k=1}^{K} p_{mk}^{(i)} \beta_{mk}^{(i)} \pi_{mk} \hat{r}_{mk} \leq E_m^{\max}, \forall m \end{cases} \tag{24a}$$

$$\frac{1}{R_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi})} \leq \hat{r}_{mk}, \forall m, k. \tag{24b}$$

The constraint (24b) is convex now, while (24a) is still non-convex; so we follow this inequality

$$xy \leq \frac{1}{2}\left(\frac{\bar{y}}{\bar{x}}x^2 + \frac{\bar{x}}{\bar{y}}y^2\right) \tag{25}$$

with $x = \pi_{mk}$, $\bar{x} = \pi_{mk}^{(i)}$, $y = \hat{r}_{mk}$, $\bar{y} = \hat{r}_{mk}^{(i)}$ to approximate the second part in the left hand-side of (24a). In this regard, the constraint (24a) is innerly approximated as follows

$$\sum_{k=1}^{K} D_m p_{mk}^{(i)} \beta_{mk}^{(i)} \frac{1}{2}\left(\frac{\hat{r}_{mk}^{(i)}}{\pi_{mk}^{(i)}}\pi_{mk}^2 + \frac{\pi_{mk}^{(i)}}{\hat{r}_{mk}^{(i)}}\hat{r}_{mk}^2\right) + \alpha_m^{(i)}\frac{\theta}{2}C_m\left(f_m^{\mathsf{lo},(i)} - \hat{f}_m^{\mathsf{lo}}\right)^2 \leq E_m^{max}, \forall m \tag{26}$$

which is now a convex constraint.

Convexity of (16e)**:** Finally, by using variables $\hat{\mathbf{r}}$ as defined in (24b), we rewrite (16e) as follow

$$t^{\mathsf{co}} \geq \beta_{mk}^{(i)} D_m \pi_{mk} \hat{r}_{mk}, \forall m \in \mathcal{M}, k \in \mathcal{K}. \tag{27}$$

The constraint (27) is still non-convex so we apply (45) with $x = \pi_{mk}$, $\bar{x} = \pi_{mk}^{(i)}$, $y = \hat{r}_{mk}$, $\bar{y} = \hat{r}_{mk}^{(i)}$; then (27) is approximated as follows

$$t^{\mathsf{co}} \geq \beta_{mk}^{(i)} D_m \frac{1}{2}\left(\frac{\hat{r}_{mk}^{(i)}}{\pi_{mk}^{(i)}}\pi_{mk}^2 + \frac{\pi_{mk}^{(i)}}{\hat{r}_{mk}^{(i)}}\hat{r}_{mk}^2\right), \forall m, k \tag{28}$$

which is now the convex constraint.

Based on the above developments, we solve the following approximate convex program of (17) at iteration i:

$$\text{SP-1: Convex} \quad \underset{\substack{\boldsymbol{\pi},\hat{\mathbf{r}},\mathbf{t}|\\ \mathbf{p}^{(i)},\mathbf{f}^{(i)},\boldsymbol{\alpha}^{(i)},\boldsymbol{\beta}^{(i)}}}{\text{minimize}} \quad \max_{\forall m \in \mathcal{M}}\{\tau_m(\mathbf{t})\} \tag{29a}$$

$$\text{s.t.} \quad (15c), (15d), (15g), (15h), (16c), (16f), (21), (22), (23), (24b), (26), (28). \tag{29b}$$

This is a convex problem and can be solved effectively with well-known tools such as CVX [42], and CVXPY [43]. For complexity analysis, the convex problem (29) consists of $2MK + 3M$ scalar decision variables and $7MK + 3M + 2K$ linear and quadratic constraints leading to the per-iteration computational complexity of $\mathcal{O}\left(\sqrt{7MK + 3M + 2K}(2MK + 3M)^2\right)$ [44, Sec. 6].

B. Computation and Communication Resource Optimisation

In this subproblem, we solve (16) with given $(\boldsymbol{\pi}^{(i+1)}, \boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)})$ to find next optimal values of transmit power and processing rate of UEs, ESs $(\mathbf{p}^{(i+1)}, \mathbf{f}^{(i+1)})$. The addressed problem to find optimal computation and transmit power is expressed as follows

$$\text{SP-2:} \quad \underset{\substack{\mathbf{p},\mathbf{f},\mathbf{t}|\\ \boldsymbol{\pi}^{(i+1)},\boldsymbol{\alpha}^{(i)},\boldsymbol{\beta}^{(i)}}}{\text{minimize}} \quad \max_{\forall m \in \mathcal{M}}\{\tau_m(\mathbf{t})\} \tag{30a}$$

$$\text{s.t.} \quad (15e) - (15h), (16c) - (16f). \tag{30b}$$

As can be observed from the sub-problem (30), the constraints (15e), (15f), and (16e) are non-convex. We are now in the position to approximate these constraints and then transform the sub-problem (30) into a convex program.

Convexity of (15e)**:** To address this constraint, we process similarly as in subsection III-A with the inner approximated rate of $R_{mk}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right)$ constructed from $\gamma_{mk}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right) = \frac{p_{mk}}{q_{mk}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right)}$, where $q_{mk}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right)$ is given by

$$q_{mk}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right) \triangleq \frac{\mathcal{I}_{mk}(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}) + N_0}{\pi_{mk}^{(i+1)}\|\mathbf{h}_{mk}\|^2}, \forall m, k. \tag{31}$$

Additionally, by approximating $\mathcal{G}_{mk}^{(i)}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right)$ and $\mathcal{W}_{mk}^{(i)}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right)$ similarly as in the subsection III-A, we obtain the inner approximation of the uplink rate at the i-th iteration as follows

$$\begin{aligned} R_{mk}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right) &\geq R_{mk}^{(i)}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right) \\ &\triangleq \frac{B}{\ln 2}\left[\mathcal{G}_{mk}^{(i)}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right) - \kappa \mathcal{W}_{mk}^{(i)}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right)\right]. \end{aligned} \tag{32}$$

Consequently, the constraint (15e) can be iteratively replaced by

$$R_{mk}^{(i)}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right) \geq \pi_{mk}^{(i+1)} R_{\min}, \forall m, k. \tag{33}$$

Convexity of (15f)**:** By introducing new variables $\check{\mathbf{r}} \triangleq \{\check{r}_{mk}\}_{\forall m,k}$ that satisfy $\frac{1}{R_{mk}} \leq \check{r}_{mk}$, then (15f) is equivalent to

$$\begin{cases} \alpha_m^{(i)}\frac{\theta}{2}C_m\left(f_m^{\mathsf{lo}} - \hat{f}_m^{\mathsf{lo}}\right)^2 + \sum_{k=1}^{K}\beta_{mk}^{(i)}\pi_{mk}^{(i+1)} D_m p_{mk}\check{r}_{mk} \leq E_m^{\max}, \forall m & (34a) \\ \dfrac{1}{R_{mk}^{(i)}\left(\mathbf{p}, \boldsymbol{\pi}^{(i+1)}\right)} \leq \check{r}_{mk}, \forall m, k. & (34b) \end{cases}$$

The constraint (34a) is still non-convex so we follow (45) with $x = p_{mk}$, $y = \check{r}_{mk}$, $\bar{x} = p_{mk}^{(i)}$, $\bar{y} = \check{r}_{mk}^{(i)}$ to iteratively express (34a) as

$$\sum_{k=1}^{K}\beta_{mk}^{(i)}\pi_{mk}^{(i+1)} D_m \frac{1}{2}\left(\frac{\check{r}_{mk}^{(i)}}{p_{mk}^{(i)}}p_{mk}^2 + \frac{p_{mk}^{(i)}}{\check{r}_{mk}^{(i)}}\check{r}_{mk}^2\right) + \alpha_m^{(i)}\frac{\theta_m}{2}C_m\left(f_m^{\mathsf{lo}} - \hat{f}_m^{\mathsf{lo}}\right)^2 \leq E_m^{max}, \forall m \tag{35}$$

which is a convex constraint.

Convexity of (16e)**:** By using $(\check{\mathbf{r}})$ defined in (34b), (16f) can be linearly expressed as

$$t^{\mathsf{co}} \geq D_m \pi_{mk}^{(i+1)} \beta_{mk}^{(i)} \check{r}_{mk}, \forall m, k. \tag{36}$$

Finally, based on the above development, we solve the following approximate convex program of (17) at iteration i:

$$\text{SP-2: Convex} \quad \underset{\substack{\mathbf{p},\mathbf{f},\check{\mathbf{r}},\mathbf{t}|\\ \boldsymbol{\pi}^{(i+1)},\boldsymbol{\alpha}^{(i)},\boldsymbol{\beta}^{(i)}}}{\text{minimize}} \quad \max_{\forall m \in \mathcal{M}}\{\tau_m(\mathbf{t})\} \tag{37a}$$

$$\text{s.t.} \quad (15g), (15h), (16c), (16d), (16f), (33), (34b), (35), (36). \tag{37b}$$

Problem (37) is a convex program, which can be solved efficiently by the CVX package. The convex problem (37) comprises $3MK + 4M$ variables and $6MK + 4M + K$ linear and quadratic constraints. Therefore, solving (37) requires per-iteration complexity of $\mathcal{O}\left(\sqrt{6MK + 4M + K}\ (3MK + 4M)^2\right)$ [44, Sec. 6].

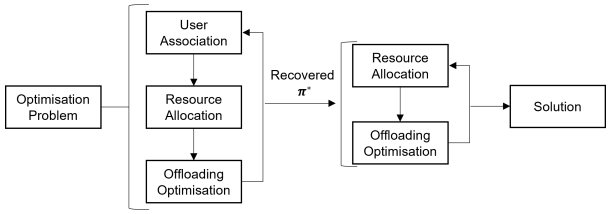


Fig. 2. The flow chart of the optimisation procedure.

C. Offloading Policies Optimisation

In this subsection we solve (16) with fixed $(\mathbf{p}^{(i+1)}, \mathbf{f}^{(i+1)}, \boldsymbol{\pi}^{(i+1)})$ find next optimal values of offloading policies, $(\boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\beta}^{(i+1)})$. The optimisation problem is expressed as follows

$$\text{SP-3: Convex} \quad \underset{\substack{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{t}|\\ \boldsymbol{\pi}^{(i+1)},\mathbf{p}^{(i+1)},\mathbf{f}^{(i+1)}}}{\text{minimize}} \quad \max_{\forall m\in\mathcal{M}}\{\tau_m(\mathbf{t})\} \tag{38a}$$

$$\text{s.t. (15d), (15f), (15g), (15h), (16c)}-\text{(16f).} \tag{38b}$$

The sub-problem (38) is obviously a convex program with all linear constraints, which can be solved by CVX. The per-iteration complexity required for solving (38) is $\mathcal{O}\big(\sqrt{3MK+5M+K}\ (MK+4M)^2\big)$, where the numbers of scalar variables and constraints are $MK+4M$ and $3MK+5M+K$, respectively.

D. Proposed Algorithm

Let us denote $\mathcal{S}_1^{(i)} \triangleq (\boldsymbol{\pi}^{(i)}, \hat{\mathbf{r}}^{(i)})$, $\mathcal{S}_2^{(i)} \triangleq (\mathbf{p}^{(i)}, \mathbf{f}^{(i)}, \check{\mathbf{r}}^{(i)})$, and $\mathcal{S}_3^{(i)} \triangleq (\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)})$ at the i-th iteration, respectively. We now propose Algorithm 1 to solve the problem (16). Due to the relaxation of binary variables of user associations for solving the approximate problems, we implement step 8 and step 9 to recover binary values of $\boldsymbol{\pi}$, repeat the solving procedure for other variables and obtain the final output. The flow chart of the optimisation process is illustrated in Fig. 2.

Complexity and convergence analysis: In the whole algorithm, the major computation complexity comes from solving the joint communication and computation resource allocation subproblem. Given the number of the required iterations I, the worst-case complexity of Algorithm 1 is given by $\mathcal{O}\left(I\sqrt{6MK+4M+K}(3MK+4M)^2\right)$. In terms of convergence, it is important to note that the approximate functions provided in Section III satisfy the properties of IA framework in [41]. According to [40], Algorithm 1 generates sequences of improved points $(\boldsymbol{\pi}^{(i)}, \hat{\mathbf{r}}^{(i)})$, $(\mathbf{p}^{(i)}, \mathbf{f}^{(i)}, \check{\mathbf{r}}^{(i)})$, and $(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)})$ by solving (29), (37) and (38), respectively, with non-increasing e2e latency after each iteration. Additionally, the feasible sets of three problems (29), (37), and (38) are convex and connected. Therefore, the sequences of $(\boldsymbol{\pi}^{(i)})$, $(\mathbf{p}^{(i)}, \mathbf{f}^{(i)})$, and $(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)})$ are guaranteed to achieve at least the local optimal solution of problems (29), (37), and (38), respectively.

Algorithm 1 AO-IA Based Algorithm for Solving (16)

Require: Set $i=0$ and randomly choose initial feasible points $\mathcal{S}_1^{(0)}$, $\mathcal{S}_2^{(0)}$ and $\mathcal{S}_3^{(0)}$ to constraints in (29), (37), (38); Set the tolerance $\varepsilon = 10^{-3}$ and the maximum number of iterations $I^{\max} = 20$.

1: **repeat**
2: Solve problem (29) for given $\mathcal{S}_2^{(i)}, \mathcal{S}_3^{(i)}$ to obtain the next solution of $(\boldsymbol{\pi}^*, \hat{\mathbf{r}}^*)$ and update $\mathcal{S}_1^{(i+1)} := (\boldsymbol{\pi}^*, \hat{\mathbf{r}}^*)$;
3: Solve problem (37) with given $\mathcal{S}_1^{(i+1)}, \mathcal{S}_3^{(i)}$ to obtain the next solution of $(\mathbf{p}^*, \mathbf{f}^*, \check{\mathbf{r}}^*)$ and update $\mathcal{S}_2^{(i+1)} := (\mathbf{p}^*, \mathbf{f}^*, \check{\mathbf{r}}^*)$;
4: Solve problem (38) with given $\mathcal{S}_1^{(i+1)}, \mathcal{S}_2^{(i+1)}$ to obtain the next solution of $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ and update $\mathcal{S}_3^{(i+1)} := (\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$;
5: Set $i := i+1$;
6: **until** Convergence or $i > I^{\max}$.
7: Recover binary values of $\boldsymbol{\pi}^*$ s.t. (15c).
8: Repeat from **Step 1** to **Step 6** with fixed $\boldsymbol{\pi}^*$;
9: **Output**: $\{\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\pi}^*, \mathbf{p}^*, \mathbf{f}^*\}$ and $\max\{\tau_m(\mathbf{t})\}_{\forall m}$.

IV. Sub-Optimal Designs With Given User Associations

In this section, we propose two sub-optimal designs for solving (16) with given user association strategies. In particular, the user association can be obtained by a heuristic approach or random assignments for some practical scenarios to reduce the processing time, i.e., small-scale systems and systems requiring low-complexity.

For the heuristic approach (HEU), the policies of user association $(\boldsymbol{\pi})$ are obtained by directly using channel condition expressed as (39). In this way, at the initial step, UEs are connected with the ESs having the best channel gain, which the sub-optimal user association can be found as:

$$\boldsymbol{\pi}^h = \Big\{\pi_{mk}^h \Big| \pi_{mk}^h = 1 \text{ if } k^* = \arg\max_{k\in\mathcal{K}}\{\|\mathbf{h}_{mk}\|^2\}; \text{ otherwise, } \pi_{mk}^h = 0 \text{ s.t. (15c), (15e)}\Big\}. \tag{39}$$

In this regard, there is no need to solve the sub-problem III-A while the resource optimisation and offloading optimisation in sub-problem III-C and sub-problem III-B are solved with the given sub-optimal user association $(\boldsymbol{\pi}^h)$ to find the minimised e2e latency. More specifically, for the communication and computation optimisation in sub-problem III-B, we now solve (16) for given $(\boldsymbol{\pi}^h, \boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)})$ to obtain the next solution of $(\mathbf{p}^{(i+1)}, \mathbf{f}^{(i+1)})$. For offloading optimisation, the addressed problem is also similar as in the sub-section III-C but solving (38) with given $(\boldsymbol{\pi}^h, \mathbf{p}^{(i+1)}, \mathbf{f}^{(i+1)})$.

For comparison, we additionally provide random user associations (RAN), which randomly assign UEs with ESs to perform task offloading. The expression of the association strategy in this scheme is given by

$$\boldsymbol{\pi}^r = \Big\{\pi_{mk}^r \text{ is randomly generated} \Big| \text{s.t. } \boldsymbol{\pi}\in\Pi, \text{(15c), (15e)}, \forall m,k\Big\}. \tag{40}$$

The algorithm of these sub-optimal designs is given as Algorithm 2. Importantly, we note that Algorithm 2 does not have binary recovering steps compared with Algorithm 1 since user associations are established in advance followed (39) or (40) at the initial step.

V. Numerical Results

In this section, we investigate the effectiveness of the proposed solutions by conducting various simulations. Firstly, the

Algorithm 2 AO-IA Based Algorithm for Solving (16) With Given User Associations

Require: Set $i = 0$ and randomly choose initial feasible points $\mathcal{S}_1^{(0)}$, $\mathcal{S}_2^{(0)}$ and $\mathcal{S}_3^{(0)}$ to constraints in (29), (37), (38); Set the tolerance $\varepsilon = 10^{-3}$ and the maximum number of iterations $I^{\max} = 20$; Obtain sub-optimal solution for user association with (39) or (40).

1: **repeat**
2: Solve problem (37) with given $\mathcal{S}_1^{(i+1)}, \mathcal{S}_2^{(i+1)}$ and $\boldsymbol{\pi}^h/\boldsymbol{\pi}^r$ to obtain the next solution of $(\mathbf{p}^*, \mathbf{f}^*, \check{\mathbf{r}}^*)$ and update $(\mathcal{S}_3^{(i+1)} := (\mathbf{p}^*, \mathbf{f}^*, \check{\mathbf{r}}^*)$;
3: Solve problem (38) with given $\mathcal{S}_1^{(i+1)}, \mathcal{S}_3^{(i)}$ and $\boldsymbol{\pi}^h/\boldsymbol{\pi}^r$ to obtain the next solution of $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ and update $\mathcal{S}_2^{(i+1)} := (\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$;
4: Set $i := i + 1$;
5: **until** Convergence or $i > I^{\max}$.
6: **Output**: $\{\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \mathbf{p}^*, \mathbf{f}^*\}$ and $\max\{\tau_m(\mathbf{t})\}_{\forall m}$.

parameter setting is provided as in the following subsection. Then, the numerical results are discussed in subsection V-B.

For the purpose of performance evaluation, we compare our proposed solution with some benchmark schemes as follows:

- "Fixed Frequency": The addressed problem does not consider optimising the processing rate variables of UEs and ESs. This scheme assumes all UEs and UEs are configured with fixed values of processing rate [25], [31] subject to the energy constraint of UEs (15d) and maximum computation resource budget of ESs (15e).
- "Fixed Power": In this scheme, the transmit power of UEs are fixed and equally set with respective to the maximum power budget, i.e., $\sum_{k\in\mathcal{K}} \pi_{mk} p_{mk} = P_m^{\max}, \ \forall m$. This scheme can be used as a benchmark to compare with existing solutions that have not considered optimising the transmit power variables [45].
- "Fixed Offloading": This scheme is used for demonstrating the impact of offloading optimisation in reducing the e2e latency. Particularly, the offloading portions are fixed and equally set up to all UEs, i.e., $\alpha_m = 0.5 \ \forall m$, while other variables are normally taken into account when executing the proposed algorithm.

A. Simulation Setup and Parameters

In this subsection, we present the parameters setting of our simulations. In particular, we consider a small-scale scenario for factory automation where all ESs and UEs are located within an area of 100 m $\times$ 100 m [24]. The large-scale fading of the channel between the m-th UE to the k-th AP is modelled as $g_{mk} = 10^{\mathbf{PL}(d_{mk})/10}$, where $\mathbf{PL}(d_{mk}) = -35.3 - 37.6 \log_{10} d_{mk}$ [20] denotes the path loss in dB, which is a function of the distance d_{mk}. The noise spectral density is set to -174 dBm/Hz [20] and the URLLC decoding error probability is set to $\epsilon = 10^{-9}$ [22]. Other parameters are summarised in Table I.

B. Numerical Results and Discussions

1) Algorithm Convergence: Fig. 3 clearly illustrates the convergence behaviour of the proposed algorithms. In particular, Algorithm 1 has experienced a slight increase in the e2e latency at the binary recovering point for user association (step 8, Algorithm 1) after sharply declining the latency to obtain the optimal solution. In Algorithm 2, since the solutions of user association are known in advance, the algorithm runs more quickly. In addition, by jointly solving user association and other variables, Algorithm 1 has the best performance in reducing the e2e latency compared with Algorithm 2 for both heuristic (Algorithm 2 - HEU) and random user association (Algorithm 2 - RAN). For instance, under the same system model and transmit power budget of $P_m^{\max} = 23$ dBm, Algorithm 1 has gained optimal latency of approximately 350 ms while these values in Algorithm 2 - HEU, Algorithm 2 - RAN are about 390 ms and 610 ms, respectively. Fig. 3 also compares the results obtained from different level of transmit power budget of UEs. It can be clearly seen that the more power budget the UEs have the lower latency can be achieved. These results evidently prove that the optimisation solution for the transmit power variables has worked effectively.

TABLE I
SIMULATION PARAMETERS

Parameters	Value
Number of antennas	$L = 8$ [46]
Maximum transmit power	$P_m^{\max} = 23$ dBm [46]
Bandwidth	$B = 10$ MHz [47]
Number of UEs	$M = \{8, 9, 10\}$ [47]
Number of ESs	$K = 2$ [25]
Maximum UEs' processing rate	$F_{\max}^{\text{lo}} = 3$ GHz [48]
Total ES processing rate	$F_{\max}^{\text{es}} = 10$ GHz [47]
Maximum UEs each ES serves	$M_{\max} = 7$ UEs
Input task size	$D_m = 100$ KB [32]
Required computation resource	$C_m = 960 \times 10^6$ cycles [24]
Total delay requirement	$T_m^{\max} = 2$ s [47]
Minimum data rate	$R^{\min} = 0.1$ Mbps [32]
Maximum energy consumption	$E_m^{\max} = 1$ Joule [48]
Effective capacitance coefficient	$\theta_m = 10^{-27}$ Watt.s^3/cycle^3 [24]

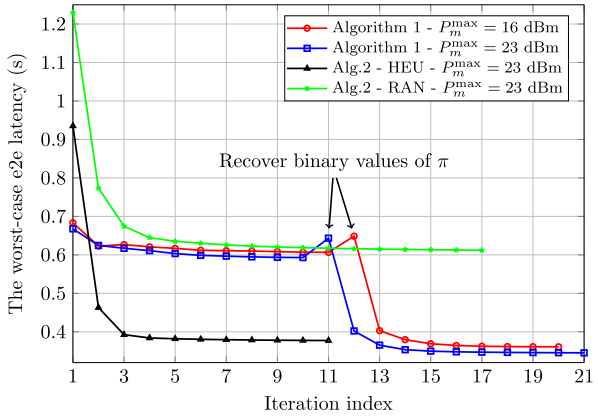


Fig. 3. The convergence of proposed algorithms in the scenario of $M = 8, K = 2$ under two levels of transmit power budget, $P_{\max} = 16$ dBm and $P_{\max} = 23$ dBm, with $E_m^{\max} = 0.6$ J, $F_{\max}^{\text{es}} = 15$GHz.

2) Impact of Maximum Energy Consumption and the Deviation Values: In order to investigate the impacts of the energy consumption budget of UEs ($E_m^{\max}$) and how this parameter affects the system performance, we therefore have conducted simulations with different values of $E_m^{\max}$ under the same system model. Fig. 4 clearly displays the obtained worst-case e2e latency with a range level of $E_m^{\max}$ from 1 J to 1.4 J. Clearly, as observed from the plots, when the energy budget increases, the minimised latency gradually declines in both models of $M = \{8, 9\}$ UEs and $K = 2$ ESs. For instance, the scenario of $M = 9$ UEs, $K = 2$ ESs has witnessed a considerable decrease of 100 ms when $E_m^{\max}$

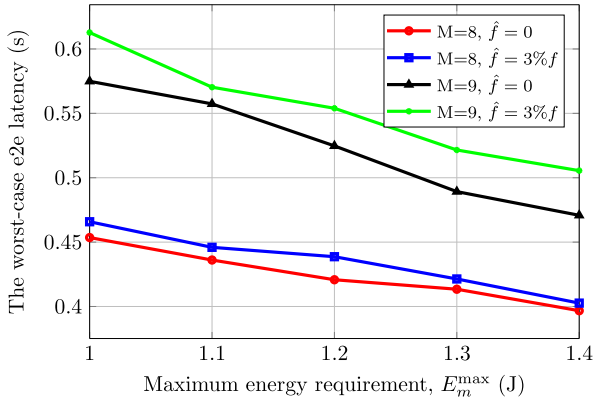


Fig. 4. The worst-case latency obtained by Algorithm 1 with different values UEs energy budget ($E_m^{\max}$) in the scenarios of $M = \{8, 9\}$ and $K = 2$.

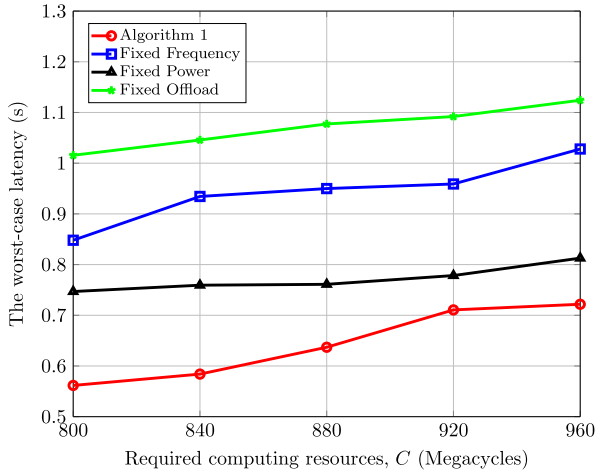


Fig. 5. The worst-case latency versus the required CPU cycles ($C_m \triangleq C, \forall m$) in the scenarios of $M = 10$ and $K = 2$, with $E_m^{\max} = 1$ J.

reaches 1.4 J. In addition, Fig. 4 further investigates the impact of the deviation level in processing rate estimation of DT services ($\hat{f}_m^{\mathsf{lo}}, \hat{f}_{mk}^{\mathsf{es}}$) on the optimal e2e latency. Unsurprisingly, when the deviations increase, the e2e latency increases and correctly follows the expression in (13). For instance, under the same system model and energy budget, the simulations of $\hat{f} = 3\%f, \forall m, k$ response show slightly higher e2e latency compared with the best estimation scheme ($\hat{f} = 0, \forall m, k$).

3) Impact of Required Computation Resource: To demonstrate how required computation resource parameters (C_m) affect the e2e latency, we have conducted simulations among a range values of (C_m). Fig. 5 clearly displays the numerical results of the observed simulations. As we can observe from Fig. 5, increasing the required computation resource under the same algorithm results in rising the e2e latency. For instance, the line of Algorithm 1 experiences a considerable increase in the worst-case e2e latency by approximately 150 ms when the required computation resource of each task climbs to 960 megacycles. In addition, we also compare the worst-case latency of the proposed solution with other benchmarks in Fig. 5. Clearly, the proposed algorithm always outperforms other schemes. In this regard, at the points of $C_m = 800$ megacycles, Algorithm 1 obtains much lower latency than that in the fixed offloading scheme with nearly 650 ms, while comparing with the fixed transmit power and the fixed frequency schemes, the gains are around 200 ms and 300 ms, respectively.

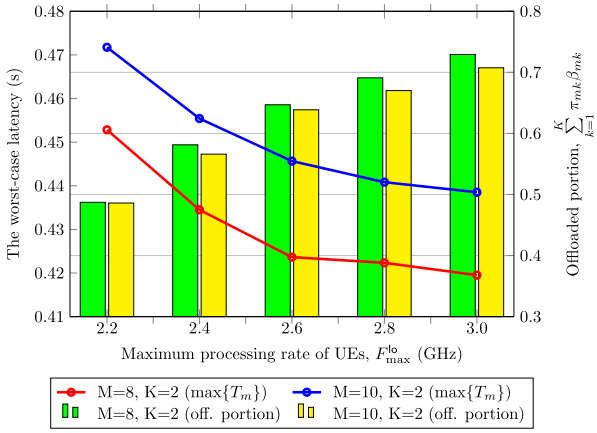


Fig. 6. Impact of local processing rate in reducing the worst-case latency in the scenarios of $M = \{8, 10\}$ and $K = 2$, with $E_{\max} = 1$ J under Algorithm 1.

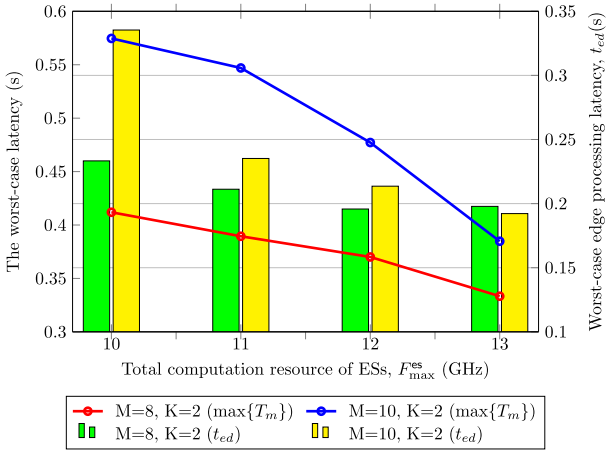


Fig. 7. The worst-case latency versus the total computation resource of ESs in the scenarios of $M = 8, K = 2$ and $M = 10, K = 2$, with $E_{\max} = 1$ J.

4) Impact of Local Processing Rate: Fig. 6 investigates the impact of the maximum processing rate of UEs ($F_{\max}^{\mathsf{lo}}$) in reducing the latency as well as offloading behaviour in different scenarios. As we can observe from the figure, when the UEs' processing rate increases, the worst-case latency gradually decreases in both scenarios of $M = 8, K = 2$ and $M = 10, K = 2$. According to the figure, in the model of $M = 10$ UEs, $K = 2$ ESs, the worst-case latency reduces to around 0.44 s when the maximum processing rate climbs to 3 GHz. Additionally, Fig. 6 also illustrates that the offloaded portion of computational tasks continuously increases when the UEs become more powerful. This is due to the constraint of maximum energy consumption of UEs presented in (14) and (15f). More specifically, when the processing rate of UEs increases, the tasks are quickly executed locally; however, UEs have to consume much energy for computation. As a result, in order to satisfy the energy budget, there is a large portion of the tasks offloaded to the ESs, which is clearly demonstrated through the bars of Fig. 6.

5) Impact of ES Processing Rate: In the MEC architecture, the computation resource of ESs has a strong impact on the system performance. To verify this, Fig. 7 shows the worst-case latency among different values of maximum processing rate of ESs. The figure clearly indicates that when the computation resource of ESs increases, both overall

worst-case latency and edge server processing latency gradually reduce. For example, with the scenarios of $M = 10, K = 2$, the worst-case latency reduces by approximately 200 ms when the total computation resource of ESs climbs to 26 GHz. These results validate that our intelligent tasks offloading solution works effectively. In addition, Fig. 7 also displays the edge processing latency among the range values of $F_{\max}^{\mathsf{es}}$ through the bars of the graph. As we can observe from the figure, increasing the total computation resource of ESs obviously decreases the processing latency of ESs. For instance, in the scenario of $M = 10, K = 2$, the worst-case edge processing latency decreases from nearly 340 ms to 190 ms when the computation resource budget of ESs reaches to 13 GHz.

VI. Conclusion

In this paper, we have investigated the computation offloading problem under the digital twin paradigm in URLLC-based wireless edge networks. The addressed problem has taken into account various factors in both communication and computation variables of practical systems. In this regard, the minimised latency is obtained by optimising the offloading policies, user association strategies, transmit power, and processing rate of UEs and ESs. To solve this challenging problem, we have proposed the AO-IA based algorithm dealing with three decomposed sub-problems, namely user association, offloading policies optimisation, and resource optimisation. For comparison, we have additionally introduced two sub-optimal designs for user association, namely the heuristic approach based on the channel condition and the random association. Various simulations have been conducted to investigate the impact of involved parameters. Extensive numerical results have successfully validated the effectiveness of the proposed solution. Promising future directions could be addressing a problem of heterogeneous computation-intensive tasks in large-scale solutions.

Appendix A
Proof of Lemma 1

In order to prove Lemma 1, we indicate that the constraints (16d)-(16f) must hold with equality at optimal solution. We prove for constraint (16d) and others follow similarly. Firstly, we assume that the equality of (16d) does not hold at the optimum for some m, i.e., existing $t^{\mathsf{lo}} > \frac{\alpha_m^\star C_m}{f_m^{\mathsf{lo}\star} - \hat{f}_m^{\mathsf{lo}}}$. There exists a positive constant $\Delta t^{\mathsf{lo}} > 0$ which is defined as $t^{\mathsf{lo}} - \Delta t^{\mathsf{lo}} = \frac{\alpha_m^\star C_m}{f_m^{\mathsf{lo}\star} - \hat{f}_m^{\mathsf{lo}}}$. As a result, $t^{\mathsf{lo}} - \Delta t^{\mathsf{lo}}$ is also feasible solution for problem (16), but this leads to a strictly lower obtained e2e latency. This disproves the original assumption that the set $(\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star, \boldsymbol{\pi}^\star, \mathbf{p}^\star, \mathbf{f}^\star)$ is the optimal solution to problem (16).

Appendix B
Fundamental Inequalities

We now provide some fundamental inequalities studied in [20] and [46] based on the IA properties [41], which are used to approximate non-convex parts.

1) For all $x > 0, y > 0$, $\bar{x} > 0$ and $\bar{y} > 0$, the function $\ln\big(1 + \frac{x}{y}\big)$ is innerly approximated around the point $(\bar{x} > 0, \bar{y})$ as [20]

$$\ln\big(1 + \frac{x}{y}\big) \geq a - \frac{b}{x} - cy \tag{41}$$

where $a \triangleq \ln\big(1 + \frac{\bar{x}}{\bar{y}}\big) + 2\frac{\bar{x}}{\bar{x}+\bar{y}} > 0, b \triangleq \frac{\bar{x}^2}{\bar{x}+\bar{y}} > 0$, and $c \triangleq \frac{\bar{x}}{(\bar{x}+\bar{y})\bar{y}} > 0$.

2) For the convex function $f(x) = 1/x$ on the domain $x > 0$, its lower bounding concave function around the point $\bar{x}$ is

$$\begin{aligned} f(x) &\geq f(\bar{x}) + \frac{\partial f(x)}{\partial x}\Big|_{x=\bar{x}} (x - \bar{x}) \\ &= \frac{1}{\bar{x}} - \frac{1}{\bar{x}^2}(x - \bar{x}) = \frac{2}{\bar{x}} - \frac{x}{\bar{x}^2}. \end{aligned} \tag{42}$$

3) For the square-over-linear function $f(x, y) = x^2/y$ that is convex on $x \in \mathbb{R}, y > 0$, its lower bounding concave function around the point $(\bar{x}, \bar{y})$ is given as

$$\begin{aligned} f(x,y) \geq f(\bar{x},\bar{y}) &+ \frac{\partial f(x,y)}{\partial x}\Big|_{(x,y)=(\bar{x},\bar{y})} (x - \bar{x}) \\ &+ \frac{\partial f(x,y)}{\partial y}\Big|_{(x,y)=(\bar{x},\bar{y})} (y - \bar{y}) = \frac{2\bar{x}}{\bar{y}}x - \frac{\bar{x}^2}{\bar{y}^2}y. \end{aligned} \tag{43}$$

4) The convex function $g(x) = 1/x^2$ with $x \in \mathbb{R}$ is innerly approximated around the point $\bar{x} \in \mathbb{R}$ as

$$\begin{aligned} g(x) &\geq g(\bar{x}) + \frac{\partial g(x)}{\partial x}\Big|_{x=\bar{x}} (x - \bar{x}) \\ &= \frac{1}{\bar{x}^2} - \frac{2}{\bar{x}^3}(x - \bar{x}) = \frac{3}{\bar{x}^2} - \frac{2x}{\bar{x}^3}. \end{aligned} \tag{44}$$

5) The upper bounding convex function of the product $f(x, y) = xy$ with $x > 0$ and $y > 0$ around the point $(\bar{x}, \bar{y})$ is given by [49, Eq. (B1)]:

$$f(x,y) = xy \leq \frac{1}{2}\left(\frac{\bar{y}}{\bar{x}}x^2 + \frac{\bar{x}}{\bar{y}}y^2\right). \tag{45}$$

6) Finally, for the concave function $h(x) = \sqrt{x}$ over $x > 0$, its upper bounding convex function at the point $\bar{x}$ is

$$h(x) \leq h(\bar{x}) + \frac{\partial h(x)}{\partial x}\Big|_{x=\bar{x}} (x - \bar{x}) = \frac{\sqrt{\bar{x}}}{2} + \frac{x}{2\sqrt{\bar{x}}}. \tag{46}$$

Appendix C
Transmission Rate Approximation

We first rewrite the SINR of UE m as $\gamma_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) = \pi_{mk}/q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi})$. By applying the inequality (41) for $x = \pi_{mk}^{(i)}$, $y = q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}_m)$, $\bar{x} = \pi_{mk}^{(i)}$, and $\bar{y} = q_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})$, we have

$$\begin{aligned} G_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) &\geq a_{mk}^{(i)} - \frac{b_{mk}^{(i)}}{\pi_{mk}} - c_{mk}^{(i)} q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) \\ &\triangleq \mathcal{G}_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) \end{aligned} \tag{47}$$

where

$$\begin{aligned} a_{mk}^{(i)} &= \ln\Big(1 + \frac{\pi_{mk}^{(i)}}{q_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi})}\Big) + \frac{2\pi_{mk}^{(i)}}{\pi_{mk}^{(i)} + q_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi})}, \\ b_{mk}^{(i)} &= \frac{(\pi_{mk}^{(i)})^2}{\pi_{mk}^{(i)} + q_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})}, \\ c_{mk}^{(i)} &= \frac{\pi_{mk}^{(i)}}{\Big(q_{mk}^{(i)}\left(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)}\right) + \pi_{mk}^{(i)}\Big) q_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})}. \end{aligned} \tag{48}$$

To find an upper bounding convex function approximation of $W_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi})$, we apply the inequality (46) for $x = 1 - 1/\left(1+\gamma_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi})\right)^2$ and $\bar{x} = 1 - 1/\left(1+\gamma_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})\right)^2$, yielding

$$\begin{aligned} W_{mk}(\mathbf{p}, \boldsymbol{\pi}_m^{(i)}) &\leq d_{mk}^{(i)} - \frac{e_{mk}^{(i)}}{\gamma_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi})} \\ &= d_{mk}^{(i)} - e_{mk}^{(i)} \frac{q_{mk}^2(\mathbf{p}^{(i)}, \boldsymbol{\pi})}{\left(q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}\right)^2} \end{aligned} \tag{49}$$

where

$$d_{mk}^{(i)} = \frac{\sqrt{V_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})}}{2} + \frac{1}{2\sqrt{V_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})}}, \tag{50}$$

$$e_{mk}^{(i)} = \frac{1}{2\sqrt{V_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})}}. \tag{51}$$

The function $\frac{q_{mk}^2(\mathbf{p}^{(i)}, \boldsymbol{\pi})}{\left(q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}\right)^2}$ in (49) is still not convex [20], which can be further approximated by using inequalities (42) and (43) as

$$\begin{aligned} &\frac{q_{mk}^2(\mathbf{p}^{(i)}, \boldsymbol{\pi}_m)}{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + p_{mk}^{(i)}} \frac{1}{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}} \\ &\geq \frac{2}{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)}) + \pi_{mk}^{(i)}} \left(2f_{mk}^{(i)} q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) \right. \\ &\quad \left. - \frac{q_{mk}^2 p_{mk}^{(i)} \left(q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}\right)}{(q_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}^{(i)})^2}\right) \\ &\quad - \frac{q_{mk}^2(\mathbf{p}^{(i)}, \boldsymbol{\pi})}{(q_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}^{(i)})^2} \end{aligned} \tag{52}$$

over the trusted regions defined in (21) and (22). By substituting (52) to (49), yields

$$\begin{aligned} W_{mk}^{(i)} &\leq d_{mk}^{(i)} - \frac{2e_{mk}^{(i)}}{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)}) + \pi_{mk}^{(i)}} \\ &\quad \left[2f_{mk}^{(i)} q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) - (f_{mk}^{(i)})^2 \left(q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) + \pi_{mk}\right)\right] \\ &\quad + \frac{(f_{mk}^{(i)})^2}{q_{mk}^2(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})} q_{mk}^2(\mathbf{p}^{(i)}, \boldsymbol{\pi}) \triangleq \mathcal{W}_{mk}^{(i)}(\mathbf{p}^{(i)}, \boldsymbol{\pi}) \end{aligned} \tag{53}$$

where $f_{mk}^{(i)} \triangleq \dfrac{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)})}{q_{mk}(\mathbf{p}^{(i)}, \boldsymbol{\pi}^{(i)}) + \pi_{mk}^{(i)}}$.

REFERENCES

[1] D. Van Huynh, V.-D. Nguyen, V. Sharma, O. A. Dobre, and T. Q. Duong, "Digital twin empowered ultra-reliable and low-latency communications-based edge networks in industrial IoT environment," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seoul, South Korea, May 2022, pp. 5651–5656.

[2] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[3] 3GPP, *Study on Scenarios and Requirements for Next Generation Access Technologies*, 3rd Generation Partnership Project (3GPP), document (TR) 38.913, 2018, version 15.0.0.

[4] D. Feng *et al.*, "Toward ultrareliable low-latency communications: Typical scenarios, possible solutions, and open issues," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 94–102, Jun. 2019.

[5] Q. Liu, T. Han, and N. Ansari, "Joint radio and computation resource management for low latency mobile edge computing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–7.

[6] Q. Luo, S. Hu, C. Li, G. Li, and W. Shi, "Resource scheduling in edge computing: A survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2131–2165, 4th Quart., 2021.

[7] D.-B. Ha, V.-T. Truong, and Y. Lee, "Performance analysis for RF energy harvesting mobile edge computing networks with SIMO/MISO-NOMA schemes," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 8, no. 27, Jun. 2021, Art. no. 169425.

[8] J. Xia *et al.*, "Opportunistic access point selection for mobile edge computing networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 695–709, Jan. 2021.

[9] Y. Wu, K. Zhang, and Y. Zhang, "Digital twin networks: A survey," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13789–13804, Sep. 2021.

[10] J. Zhou, D. Tian, Z. Sheng, X. Duan, and X. Shen, "Distributed task offloading optimization with queueing dynamics in multiagent mobile-edge computing networks," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12311–12328, Aug. 2021.

[11] R. Lin *et al.*, "Distributed optimization for computation offloading in edge computing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8179–8194, Dec. 2020.

[12] Y. Xiao and M. Krunz, "Distributed optimization for energy-efficient fog computing in the tactile internet," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2390–2400, Nov. 2018.

[13] T. Bai, J. Wang, Y. Ren, and L. Hanzo, "Energy-efficient computation offloading for secure UAV-edge-computing systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6074–6087, Jun. 2019.

[14] T. T. Vu, D. N. Nguyen, D. T. Hoang, E. Dutkiewicz, and T. V. Nguyen, "Optimal energy efficiency with delay constraints for multi-layer cooperative fog computing networks," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3911–3929, Jun. 2021.

[15] T. Q. Dinh, B. Liang, T. Q. S. Quek, and H. Shin, "Online resource procurement and allocation in a hybrid edge-cloud computing system," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2137–2149, Mar. 2020.

[16] T. Zhou, Y. Yue, D. Qin, X. Nie, X. Li, and C. Li, "Joint device association, resource allocation and computation offloading in ultra-dense multi-device and multi-task IoT networks," *IEEE Internet Things J.*, early access, Mar. 23, 2022, doi: 10.1109/JIOT.2022.3161670.

[17] T. Bai, C. Pan, H. Ren, Y. Deng, M. Elkashlan, and A. Nallanathan, "Resource allocation for intelligent reflecting surface aided wireless powered mobile edge computing in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5389–5407, Aug. 2021.

[18] D. T. Nguyen, L. B. Le, and V. Bhargava, "Price-based resource allocation for edge computing: A market equilibrium approach," *IEEE Trans. Cloud Comput.*, vol. 9, no. 1, pp. 302–317, Jan. 2021.

[19] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, Apr. 2010.

[20] A. A. Nasir, H. D. Tuan, H. H. Nguyen, M. Debbah, and H. V. Poor, "Resource allocation and beamforming design in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1321–1335, Feb. 2021.

[21] A. A. Nasir, "Min-max decoding-error probability-based resource allocation for a URLLC system," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2864–2867, Dec. 2020.

[22] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020.

[23] C. She, C. Yang, and T. Q. S. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 72–78, Jun. 2017.

[24] C.-F. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4132–4150, Jun. 2019.

[25] R. Dong, C. She, W. Hardjawana, Y. Li, and B. Vucetic, "Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4692–4707, Oct. 2019.

[26] D. Van Huynh, S. R. Khosravirad, L. D. Nguyen, and T. Q. Duong, "Multiple relay robots-assisted URLLC for industrial automation with deep neural networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–5.
[27] M. S. Elbamby *et al.*, "Wireless edge computing with latency and reliability guarantees," *Proc. IEEE*, vol. 107, no. 8, pp. 1717–1737, Aug. 2019.
[28] L. Lei, G. Shen, L. Zhang, and Z. Li, "Toward intelligent cooperation of UAV swarms: When machine learning meets digital twin," *IEEE Netw.*, vol. 35, no. 1, pp. 386–392, Jan. 2021.
[29] W. Sun, H. Zhang, R. Wang, and Y. Zhang, "Reducing offloading latency for digital twin edge networks in 6G," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12240–12251, Oct. 2020.
[30] Y. Lu, S. Maharjan, and Y. Zhang, "Adaptive edge association for wireless digital twin networks in 6G," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16219–16230, Nov. 2021.
[31] T. Liu, L. Tang, W. Wang, Q. Chen, and X. Zeng, "Digital-twin-assisted task offloading based on edge collaboration in the digital twin edge network," *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1427–1444, Jan. 2022.
[32] T. Do-Duy, D. Van Huynh, O. A. Dobre, B. Canberk, and T. Q. Duong, "Digital twin-aided intelligent offloading with edge selection in mobile edge computing," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 806–810, Apr. 2022.
[33] Y. Dai, K. Zhang, S. Maharjan, and Y. Zhang, "Deep reinforcement learning for stochastic computation offloading in digital twin networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4968–4977, Jul. 2021.
[34] K. Zhang, J. Cao, and Y. Zhang, "Adaptive digital twin and multi-agent deep reinforcement learning for vehicular edge computing and networks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1405–1413, Feb. 2022.
[35] C. Hu *et al.*, "Digital twin-assisted real-time traffic data prediction method for 5G-enabled internet of vehicles," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2811–2819, Apr. 2022.
[36] M. H. Chen, M. Dong, and B. Liang, "Resource sharing of a computing access point for multi-user mobile cloud offloading with delay constraints," *IEEE Trans. Mobile Comput.*, vol. 17, no. 12, pp. 2868–2881, Dec. 2018.
[37] K. Guo, M. Sheng, J. Tang, T. Q. S. Quek, and Z. Qiu, "Hierarchical offloading for delay-constrained applications in fog RAN," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4257–4270, Apr. 2020.
[38] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4268–4282, Aug. 2016.
[39] L. Fang and L. B. Milstein, "Performance of successive interference cancellation in convolutionally coded multicarrier DS/CDMA systems," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2062–2067, Dec. 2001.
[40] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, pp. 681–683, Jul. 1978.
[41] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.
[42] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*. [Online]. Available: http://cvxr.com/cvx
[43] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2909–2913, Jan. 2016.
[44] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization* (MPS-SIAM Series on Optimization). Philadelphia, PA, USA: SIAM, 2001.
[45] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. S. Quek, "Offloading in mobile edge computing: Task allocation and computational frequency scaling," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3571–3584, Aug. 2017.
[46] A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5861–5871, Sep. 2021.
[47] J. Wang, D. Feng, S. Zhang, A. Liu, and X.-G. Xia, "Joint computation offloading and resource allocation for MEC-enabled IoT systems with imperfect CSI," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3462–3475, Mar. 2021.
[48] V.-D. Nguyen, S. Chatzinotas, B. Ottersten, and T. Q. Duong, "FedFog: Network-aware optimization of federated learning over wireless fog-cloud systems," *IEEE Trans. Wireless Commun.*, early access, Apr. 20, 2022, doi: 10.1109/TWC.2022.3167263.
[49] V.-D. Nguyen, H. V. Nguyen, O. A. Dobre, and O.-S. Shin, "A new design paradigm for secure full-duplex multiuser systems," *IEEE J. Select. Areas Commun.*, vol. 36, no. 7, pp. 1480–1498, Jul. 2018.

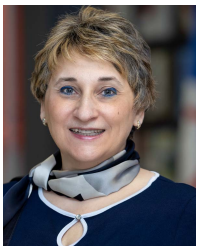
Dang Van Huynh (Graduate Student Member, IEEE) received the B.Eng. degree (Hons.) in information technology and the M.S. degree in computer science from Vietnam National University—Ho Chi Minh City, University of Information Technology (VNUHCM-UIT) in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with the School of Electronics, Electrical Engineering and Computer Science (EEECS), Queen's University Belfast, U.K. From August 2017 to December 2020, he worked as a Research Assistant with the Faculty of Computer Networks & Communications, VNUHCM-UIT. His research interests include wireless communications, industrial Internet of Things, convex optimization, and machine learning for wireless communications.

Van-Dinh Nguyen (Member, IEEE) received the B.E. degree in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2012, and the M.E. and Ph.D. degrees in electronic engineering from Soongsil University, Seoul, South Korea, in 2015 and 2018, respectively.

Since 2022, he has been an Assistant Professor with VinUniversity, Vietnam. He was a Research Associate with the SnT, University of Luxembourg, a Post-Doctoral Researcher and a Lecturer with Soongsil University, a Post-Doctoral Visiting Scholar with the University of Technology Sydney, and a Ph.D. Visiting Scholar with Queen's University Belfast, U.K. His current research activities are focused on the mathematical modeling of 5G/6G cellular networks, edge/fog computing, and AI/ML solutions for wireless communications. He received several best conference paper awards, the Exemplary Editor Award of IEEE COMMUNICATIONS LETTERS 2019, an Exemplary Reviewer Award of IEEE TRANSACTIONS ON COMMUNICATIONS 2018, and IEEE GLOBECOM Student Travel Grant Award 2017. He has authored or coauthored in some 60 papers published in international journals and conference proceedings. He has served as a reviewer for many top-tier international journals on wireless communications, and has also been a technical program committee member for several flag-ship international conferences in the related fields. He is an Editor of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY and IEEE COMMUNICATIONS LETTERS.

Saeed R. Khosravirad (Member, IEEE) received the B.Sc. degree from the Department of Electrical and Computer Engineering, University of Tehran, Iran, the M.Sc. degree from the Department of Electrical Engineering, Sharif University of Technology, Iran, and the Ph.D. degree in telecommunications from McGill University, Canada, in 2015. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the Guest Editor of the *IEEE Wireless Communications* magazine. His research interests include wireless communications theory, ultra-reliable communication for industrial automation, and radio resource management for 5G. He is currently a member of Technical Staff at the Nokia Bell Laboratories.

Vishal Sharma (Senior Member, IEEE) received the B.Tech. degree in computer science and engineering from Punjab Technical University, in 2012, and the Ph.D. degree in computer science and engineering from Thapar University, India, in 2016. From November 2016 to March 2019, he was with the Information Security Engineering Department, Soonchunhyang University, South Korea, in multiple positions (from November 2016 to December 2017, he was a Post-Doctoral Researcher and from January 2018 to March 2019, he was a Research Assistant Professor). He was also a Joint Post-Doctoral Researcher with Soongsil University, South Korea. Before this, he worked as a Lecturer with the Department of Computer Science and Engineering, Thapar University. He is currently working as a Lecturer (an Assistant Professor) with the School of Electronics, Electrical Engineering and Computer Science (EEECS), Queen's University Belfast (QUB), Belfast, U.K. At QUB, he is also a Racial Equity Network School Co-Champion, a BCS Accreditation Liaison, and a Mental Health Ambassador (MHFA). Before coming to QUB, he was a Research Fellow at the Information Systems Technology and Design (ISTD) Pillar, Singapore University of Technology and Design (SUTD), Singapore, where he worked on the future-proof blockchain systems. He has authored/coauthored more than 100 journals/conference papers and book chapters and co-edited two books with Springer. His research interests include autonomous systems, UAV communications, network behavior modeling, 5G and beyond, blockchain, and CPS security. He is a Professional Member of ACM and the Past Chair of ACM Student Chapter-TIET Patiala. He was a recipient of four best paper awards. He serves on the Editorial Board for *IEEE Communications Magazine*, *CAAI Transactions on Intelligence Technology* (IET), *Wireless Communications and Mobile Computing*, *IET Networks*, and *ICT Express*, and the Section Editor-in-Chief of *Drones* journal. He is also the Interim Co-Chair of the IEEE U.K. and Ireland Diversity, Equity, and Inclusion Committee.



Octavia A. Dobre (Fellow, IEEE) received the Dipl.-Ing. and Ph.D. degrees from the Polytechnic Institute of Bucharest, Romania, in 1991 and 2000, respectively.

From 2002 to 2005, she was with the New Jersey Institute of Technology, USA. In 2005, she joined Memorial University, Canada, where she is currently a Professor and the Research Chair. She was a Visiting Professor with the Massachusetts Institute of Technology, USA, and Université de Bretagne Occidentale, France. She has (co)authored over 400 refereed papers in these areas. Her research interests encompass wireless communication and networking technologies, and optical and underwater communications. She serves as the Director of Journals of the Communications Society. She was the Inaugural Editor-in-Chief (EiC) of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY and the EiC of the IEEE COMMUNICATIONS LETTERS. She also served as the general chair, the technical program co-chair, the tutorial co-chair, and the technical co-chair of symposia at numerous conferences. She was a Fulbright Scholar, a Royal Society Scholar, and a Distinguished Lecturer of the IEEE Communications Society. She received Best Paper Awards at various conferences, including IEEE ICC, IEEE Globecom, IEEE WCNC, and IEEE PIMRC. She is an Elected Member of the European Academy of Sciences and Arts, a fellow of the Engineering Institute of Canada, and a fellow of the Canadian Academy of Engineering.

Hyundong Shin (Fellow, IEEE) received the B.S. degree in electronics engineering from Kyung Hee University (KHU), Yongin-si, South Korea, in 1999, and the M.S. and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2001 and 2004, respectively.

During his Post-Doctoral Research at the Massachusetts Institute of Technology (MIT) from 2004 to 2006, he was with the Wireless Communication and Network Sciences Laboratory and the Laboratory for Information Decision Systems (LIDS). In 2006, he joined the KHU, where he is currently a Professor at the Department of Electronic Engineering and the Department of Electronics and Information Convergence Engineering. His research interests include quantum information science, wireless communication, and machine intelligence.

Dr. Shin was honored with the Knowledge Creation Award in the field of Computer Science from Korean Ministry of Education, Science and Technology in 2010. He received the IEEE Communications Society's Guglielmo Marconi Prize Paper Award in 2008 and William R. Bennett Prize Paper Award in 2012. He served as the Publicity Co-Chair for the IEEE PIMRC in 2018 and the Technical Program Co-Chair for the IEEE WCNC (PHY Track 2009) and the IEEE Globecom (Communication Theory Symposium 2012 and Cognitive Radio and Networks Symposium 2016). He was an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2012) and IEEE COMMUNICATIONS LETTERS (2013–2015).

Trung Q. Duong (Fellow, IEEE) received the B.Eng. degree in electrical and electronics engineering from Bach Khoa Sai Gon, Vietnam, in 2002, the M.Sc. degree in computer science from Kyung Hee University, South Korea, in 2005, and the Ph.D. degree in telecommunications systems from the Blekinge Institute of Technology, Sweden, in 2012.

In 2013, he joined Queen's University Belfast, U.K., where he is currently a Chair Professor in telecommunications. He also holds a prestigious Research Chair of Royal Academy of Engineering. His current research interests include quantum communications, wireless communications, signal processing, machine learning, realtime optimization, and data analytic. He currently serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE WIRELESS COMMUNICATIONS LETTERS, and an Executive Editor for IEEE COMMUNICATIONS LETTERS. He has served as an Editor/the Guest Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS, *IEEE Communications Magazine*, IEEE COMMUNICATIONS LETTERS, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He received the Best Paper Award at the IEEE VTC-Spring 2013, IEEE ICC 2014, IEEE GLOBECOM 2016, 2019, IEEE DSP 2017, and IWCMC 2019. He was a recipient of Prestigious Royal Academy of Engineering Research Fellowship (2015–2020) and has won a prestigious Newton Prize 2017.