Reliable Intelligent Reflecting Surface-Assisted Mobile Edge Computing Systems: A Physical Layer Security and Encryption Design

Ti Ti Nguyen, Vu Nguyen Ha, Thanh-Dung Le, Duc-Dung Tran, Symeon Chatzinotas, Kim-Khoa Nguyen

***Abstract*—Mobile edge computing (MEC) has emerged as a promising technology to extend the functionality of end-users' wireless devices while prolonging their battery life by offloading computationally intensive tasks to remote edge servers. However, the inherent broadcast nature of wireless transmission during offloading introduces notable security challenges. To address this issue, we propose leveraging intelligent reflecting surface (IRS) technology to enhance physical layer security (PLS). Nevertheless, attaining high PLS for all users in dense networks with multiple malicious terminals is challenging. In this paper, we investigate the physical layer encryption (PLE) to complement the PLS in enabling secure wireless transmission. Since such encryption and decryption processes require computation resources, we aim to optimize the encryption decision, offloading decision, as well as wireless and computing resource allocations. Our objective is to minimize the maximum weighted energy consumption while satisfying practical constraints, including limited computing and wireless resources, fulfilling minimum user rate requirements, and complying with IRS conditions. To tackle the non-convex objective and constraints, we explore the utilization of bisection search and successive convex approximation (SCA) methods. Our numerical results confirm the efficiency of the proposed design in terms of energy consumption and network capacity within a secure MEC network.**

***Index Terms*—Intelligent reflecting surface, physical layer encryption, physical layer security, MIMO, MEC, computation offloading.**

I. Introduction

RECENTLY, the rapid evolution of artificial intelligence-driven services and high-speed communications have motivated a variety of computation-intensive and delay-sensitive mobile applications, such as object recognition, real-time social interactions, e-health, natural language processing, and augmented/virtual reality (AR/VR) [1]. However, executing these resource-hungry desktop-level applications on personal mobile platform poses challenges due to the energy limitations and constrained computational capacity on mobile devices. Mobile edge computing (MEC) technology has recently been emerged as a viable design paradigm aimed at enhancing the mobile usability and extending the mobile battery life by offloading intensive computational tasks to remote edge servers [2–5]. This paradigm empowers mobile devices to efficiently handle computation-intensive and energy-hungry applications within MEC systems [6]. Recently, the massive multiple-input multiple-output (MIMO) has become a key component of the MEC system to improve the offloading efficiency [6–9]. In [9], the zero-forcing (ZF) beamforming method is applied to support the transmission of offloading data. On the other approach, the MIMO design is optimized according to the CSI, objective, and resource constraints in MEC systems [6–8]. Additionally, PLS has been extensively studied in the literature; see recent surveys [10, 11] and the references therein.

Despite the benefits brought by MEC, the wireless task offloading introduces severe security concerns. The inherent broadcast nature of wireless transmission necessitates robust security measures to safeguard sensitive information from potential eavesdropping attacks. A promising approach to achieving confidentiality and integrity in wireless communication systems is the physical layer security (PLS) paradigm, which exploits the characteristics of wireless channels to deteriorate the signal quality attained by eavesdroppers [12–17]. In [15], the ergodic secrecy rate is employed to determine the number of tasks offloaded to the edge in intelligently connected vehicle networks. In [16], friendly jamming from both the BS and nearby mobile users is introduced to enhance data confidentiality during uploading in MEC systems. Focusing on a single-antenna, multi-user orthogonal transmission scenario, the study in [17] aims to minimize total energy consumption while considering constraints on secrecy outage and latency. Besides, PLS has been studied extensively in the literature, see recent surveys [10, 11] and the references therein.

To increase the interference at eavesdroppers while improving the signal quality at legitimate users, the intelligent reflecting surface (IRS) technology has attracted significant attention [18, 19]. By leveraging a numerous low-cost passive elements capable of manipulating signal reflections through adjustable phase shifts and/or amplitudes, with the aid of IRSs, the signals from different communication links can be added coherently at the desired receiver to enhance the received signal strength or can be added destructively at undesired receivers to avoid the information leakage [20]. Recently, the integration of IRS into MEC systems to improve the PLS level has gained significant attention [21–26]. In [21, 22], IRSs are employed to enhance the security of the UAV-MEC network, assuming that the offloading rate is less than or equal to the secrecy rate. The study in [24] compares the performance in terms of computation energy efficiency in MEC systems with and without IRS. It shows that energy can be saved

Ti Ti Nguyen, Vu Nguyen Ha, Thanh-Dung Le, Duc-Dung Tran, and Symeon Chatzinotas, are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg (E-mails: {titi.nguyen, vu-nguyen.ha, thanh-dung.le, duc.tran, symeon.chatzinotas}@uni.lu)

Kim-Khoa Nguyen is with †Ecole de Technologie Supérieure, Canada (email: kim-khoa.nguyen@etsmtl.ca)

TABLE I: A Brief Comparison of Recent Works

Ref.	PLS	PLE	IRS-PLS	IRS-PLS-PLE
[12–17]	✓			
[18–26]	✓		✓	
[27–32]		✓		
Our paper	✓	✓	✓	✓

more than two times if both IRS and PLS techniques are utilized to avoid the information leakage to the malicious node. In [23], the IRS is adopted to prevent information leakage and improve legitimate reception quality in the aerial secure offloading system. Considering the partial offloading strategy, Wang et al. in [23] propose to maximize the secrecy energy efficiency by optimizing the CPU frequency, offloading ratio, transmit power, phase shifts, and UAV locations. In [25], Li et al. consider the integration of MEC, IRS, and wireless power transfer (WPT).

These aforementioned works [13–16, 18, 20–26] have primarily focused on enhancing the secrecy rate as a criterion for designing PLS in wireless systems. However, it should be noted that achieving the same secrecy rate does not necessarily ensure equal levels of information leakage, as the signal-to-interference-plus-noise ratio (SINR) can vary among users and eavesdroppers. Therefore, in this paper, we propose a practical secure criterion based on the bit error probability, which directly determines eavesdroppers' successful rate of data decoding.

Furthermore, with the escalating reliance on wireless networks and the exponential surge in the number of connected devices, ensuring high PLS for all users poses a considerable challenge. It is particularly critical in scenarios involving multiple isotropic transmissions from single-antenna users, such as wireless offloading. In such scenarios, employing PLE paradigm becomes essential to protect users from potential eavesdroppers, especially when wireless resources cannot guarantee a bit error probability below an acceptable threshold [27, 28]. In [29, 30], a selective encryption problem is investigated to satisfy the desired multiple confidentiality and data integrity levels. Recent studies, such as [27, 31, 32], have explored the hybrid crypto-PLS protocol to provide secure solutions. In [27], Sadig et al. consider the joint design of encryption and PLS schemes for 4-node Gaussian wiretap channels. The secrecy rate is adopted to model the PLS. In [32], the encrypted data are securely transmitted and satisfy the secrecy rate requirement. Although PLE entails additional computations needed for the execution of the corresponding encryption and decryption algorithms, none of the aforementioned works [27–32] delved deeply into the computational energy associated with the encryption process. Additionally, their proposed solution fails to account for situations where certain users in the MEC system achieve low PLS levels.

Motivated by these gaps in research, we propose a joint PLS and PLE framework for enhancing the security in MEC systems. A brief comparison of our work to recent research is shown in Table I. Our contribution is summarized as follows:

- We introduce a practical secure criterion that relies on the bit error probability (BEP). This criterion directly determines the successful rate of data decoding by eavesdroppers.
- We present a novel joint mechanism of PLE and PLS to enable secure computation offloading in MEC systems. Our approach ensures the protection of user data through two conditions: either the data is encrypted, or the error probability experienced by potential eavesdroppers surpasses an acceptable threshold.
- We formulate the binary computation offloading problem to minimize the maximum weighted consumed energy of all users subject to constraints on the joint PLE and PLS, limited computing resource, limited energy at the edge server, limited transmit power, and reflecting condition.
- We design a bisection search algorithm to tackle the min-max objective. This enables us to transform the optimization problem into a feasibility verification of two subsets. The first subset comprises users who perform their task locally, while the second subset consists of users who offload their tasks to the edge.
- We propose to employ an alternative method to address the feasibility verification of offloaded users, which is a mixed-integer non-linear programming (MINLP) problem. In particular, we decompose this problem into the subproblems of determining the reflecting coefficients and encryption decision, the detection matrix, the transmit power and computing resource, and the resource association at the edge server. We then iteratively solve these subproblems until the feasibility condition is satisfied. The successive convex approximation (SCA) is applied to convexify the underlying non-convex subproblems.
- Extensive experiments have been carried out to evaluate the proposed designs' performance gains compared with the conventional strategies. The results confirm the high performance of the proposed algorithm.

The remainder of this paper is organized as follows. Section II presents the system model, the joint PLE and PLS mechanism, and the computation offloading problem formulation. Section III introduces the proposed algorithm to address the min-max objective. Section IV presents the proposed feasibility verification algorithm. Section V evaluates the performance of the proposed algorithm. Finally, Section VI concludes the paper and presents future directions. In Table II, we summarize important notations in this paper, particularly those related to system configurations, offloading and resource allocation, and security.

II. System model

Considering an IRS-assisted multi-user MEC system including L edge servers co-located with a base station (BS) assisted by I IRSs to serve K single-antenna legitimate users and manage the eavesdropping from E single-antenna eavesdroppers, as shown in Fig. 1. In this system, one assumes the BS is equipped with M_0 antennas, and IRS i has M_i reflecting elements. For convenience, we denote the sets of users, eavesdroppers, IRSs, and edge servers as $\mathcal{K}$, $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{L}$, respectively. Reflecting practical frameworks, one further assumes eavesdroppers operate independently without pre-arranged cooperative strategies.

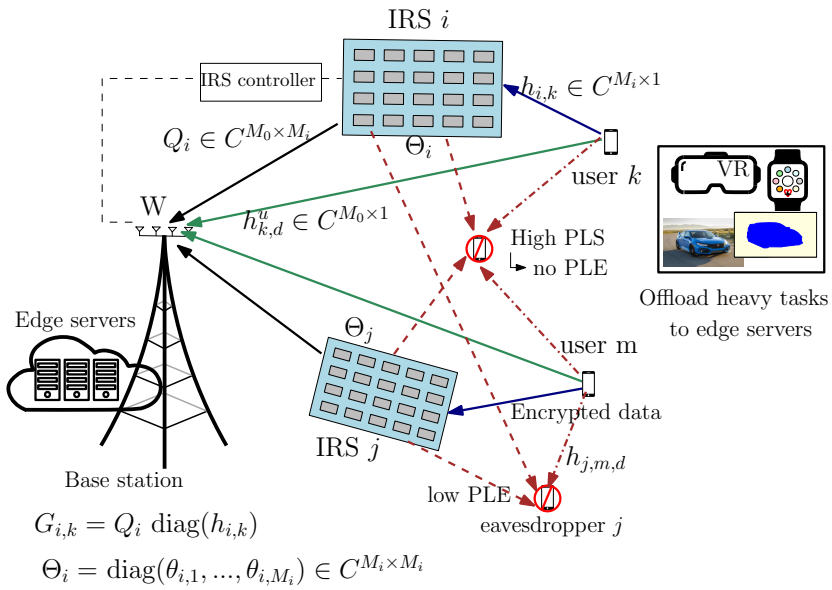

Fig. 1: A Intelligent Reflecting Surface-Assisted Mobile Edge Computing System Architecture.

We assume that user $k \in \mathcal{K}$ needs to execute a computational task from his/her application, and this task can be executed locally at its mobile device or offloaded and executed in an edge server for energy saving with the maximum allowable delay $\eta_k^{\max}$. Moreover, the task of user k requires a number of CPU cycles c_k and a number of data bits b_k (e.g., to transmit the involved programming states to the BS) [2]. We now introduce a binary offloading decision variable s_k for the task of user k as,

$$s_k = \begin{cases} 1, & \text{if user } k\text{'s is executed locally,} \\ 0, & \text{if user } k\text{'s is offloaded to an edge server.} \end{cases} \tag{1}$$

Once users decide to offload their tasks to edge servers, their data are transmitted over air to the BS. In addition to the direct user-to-BS links, reflected links from multiple IRSs enhance signal transmission and forwarding efficiency. Leveraging this over-air transmission, eavesdroppers can intercept multiple versions of the signals through both direct and RIS-related reflecting links between themselves and the users. The following section will describe the achievable and leakage information rates.

A. Wireless Communication Model

1) Achievable Information Rate: To ease the notation, we denote $\boldsymbol{h}_k^{\mathsf{BS}} \in \mathbb{C}^{M_0 \times 1}$, $\boldsymbol{h}_{i,k}^{\mathsf{RIS}} \in \mathbb{C}^{M_i \times 1}$, and $\boldsymbol{Q}_i^{\mathsf{R2B}} \in \mathbb{C}^{M_0 \times M_i}$ as the channel vectors between the BS and user k, between IRS i and user k, and between IRS i and BS, respectively. Let $\boldsymbol{\theta}_i \in \mathbb{C}^{M_i \times 1}$ be the coefficient vector of IRS i, $\theta_{i,m}$ be the m^{th} element of vector $\boldsymbol{\theta}_i$, and $\boldsymbol{\Theta}_i = \mathrm{diag}(\boldsymbol{\theta}_i) \in \mathbb{C}^{M_i \times M_i}$. The overall channel between user k and the BS consists of the direct channel $\boldsymbol{h}_k^{\mathsf{BS}}$ and the reflected channels from user k to the BS via IRSs (i.e., $\boldsymbol{Q}_i^{\mathsf{R2B}} \boldsymbol{\Theta}_i \boldsymbol{h}_{i,k}^{\mathsf{RIS}}$ via IRS i)[1] Thus, the overall channel between user k and the BS is expressed as

$$\boldsymbol{z}_k^{\mathsf{info}} = \sum_{i \in \mathcal{I}} \boldsymbol{Q}_i^{\mathsf{R2B}} \boldsymbol{\Theta}_i \boldsymbol{h}_{i,k}^{\mathsf{RIS}} + \boldsymbol{h}_k^{\mathsf{BS}}. \tag{2}$$

[1]As in [33, 34], multiple reflections between IRSs are neglected.

At the BS, a MIMO detection vector, $\boldsymbol{w}_k \in \mathbb{C}^{M_0 \times 1}$, is designed to efficiently extract the signal from user k. Hence, the instantaneous SINR of user k's signal received at the BS, denoted as γ_k^{info}, can be then given by

$$\gamma_k^{\mathsf{info}} = \frac{p_k |\boldsymbol{w}_k^H \boldsymbol{z}_k^{\mathsf{info}}|^2}{\sum_{j \in \mathcal{K}, j \neq k} p_j |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2}. \tag{3}$$

where p_k is the transmit power of user k and σ_k^2 is the received noise power at user k. Regarding the signal of user k at other receiving terminals, its achievable rate at the BS can be described based on Shannon formula as

$$r_k^{\mathsf{info}} = \log_2 \left(1 + \gamma_k^{\mathsf{info}}\right). \tag{4}$$

2) Information Leakage Rate: Denote $h_{e,k}^{\mathsf{D,leak}}$ the direct channel coefficient between user k and eavesdropper e. Let $\boldsymbol{h}_{i,k,e}^{\mathsf{R,leak}} = \mathrm{diag}(\boldsymbol{h}_{i,k}^{\mathsf{RIS}}) \boldsymbol{h}_{i,e}^{\mathsf{RIS}}$. The overall channel between user k and eavesdropper e, including the direct channel $h_{e,k}^{\mathsf{D,leak}}$ and the reflected channels $\boldsymbol{\theta}_i^H \boldsymbol{h}_{i,k,e}^{\mathsf{R,leak}}, \forall i \in \mathcal{I}$, is expressed as

$$z_{k,e}^{\mathsf{leak}} = \sum_{i \in \mathcal{I}} \boldsymbol{\theta}_i^H \boldsymbol{h}_{i,k,e}^{\mathsf{R,leak}} + h_{e,k}^{\mathsf{D,leak}}. \tag{5}$$

The SINR of the leakage information of user k at eavesdropper $e \in \mathcal{E}$, denoted as $\gamma_{e,k}^{\mathsf{leak}}$, can be described given by

$$\gamma_{e,k}^{\mathsf{leak}} = \frac{p_k |z_{k,e}^{\mathsf{leak}}|^2}{\sum_{j \in \mathcal{K}, j \neq k} p_j |z_{j,e}^{\mathsf{leak}}|^2 + \sigma_e^2}. \tag{6}$$

Then, the rate at the eavesdropper e (i.e., information leakage rate) can be given as

$$r_{e,k}^{\mathsf{leak}} = \log_2 \left(1 + \gamma_{e,k}^{\mathsf{leak}}\right). \tag{7}$$

B. Security Model

1) Information Leakage Ratio Definition: Assuming that the data of user k is transmitted at a rate based on Shannon formula r_k^{info}, and the maximum number of bits that eavesdropper $e \in \mathcal{E}$ can successfully decode, corresponding to the

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3607599

TABLE II: Important Notations

Notations	Description
M_0/M_i	Number of antennas at BS/IRS i
$\boldsymbol{h}_k^{\mathsf{BS}}$	Direct channels between the BS and user k
$\boldsymbol{h}_{i,k}^{\mathsf{RIS}}$	Channel between IRS i and user k
$\boldsymbol{Q}_i^{\mathsf{R2B}}$	Channel between IRS i and the BS
$h_{e,k}^{\mathsf{D,leak}}$	Direct channel between user k and eavesdropper e
$\boldsymbol{W}/\boldsymbol{w}_k$	MIMO detection matrix/vector for user k
$\theta_{i,m}/\boldsymbol{\theta}_i$	Coefficient element/vector of IRS i
$\boldsymbol{z}_k^{\mathsf{info}}$	Overall channel between the BS and user k
$z_{k,e}^{\mathsf{leak}}$	Overall channel between user k and eavesdropper e
$\gamma_k^{\mathsf{info}}/r_k^{\mathsf{info}}$	SINR/ rate per Hz of user k
$\gamma_{e,k}^{\mathsf{leak}}/r_{e,k}^{\mathsf{leak}}$	SINR/rate per Hz received at eavesdropper e when it eavesdrops the signal from user k
$\epsilon_{e,k}$	Information leakage ratio when eavesdropper e decodes the signal from user k
ϵ^{th}	Threshold for security criteria
$\eta_k^{\max}$	Maximum delay of executing the task of user k
c_k	Number of CPU cycles of the task of user k
$f_k/F_k^{\max}$	CPU clock speed of user k and its maximum
$c_k^{\mathsf{en}}/c_k^{\mathsf{de}}$	Number of CPU cycles to encrypt and decrypt the data
c_k^{Ed}	Number of required CPU cycles of user k at edge server
f_k^{Ed}	Allocated CPU clock speed to user k at the edge server
$\xi_k^{\mathsf{Lo}}/t_k^{\mathsf{Lo}}$	Local computational energy and time of user k
t_k^{Tx}	Transmission time of user k
t_k^{Ed}	Execution time of user k at the edge server
ξ_k^{Tx}	Transmission energy of user k
$\alpha_{k,l}$	Binary variable where $\alpha_{k,l}=1$ if the task of user k is executed by edge server l
x_k	Binary encryption decision variable where $x_k=1$ if the data of user k is encrypted
s_k	Binary offloading decision variable where $s_k=1$ if the task is executed locally
p_k	Transmit power of user k
$K/\mathcal{K}$	Number/set of users
$E/\mathcal{E}$	Number/set of eavesdroppers
$L/\mathcal{L}$	Number/set of edge servers
$I/\mathcal{I}$	Number/set of IRSs

transmit powers $p_j, \forall j \in \mathcal{K}$, is $r_{e,k}^{\mathsf{leak}}$ bits per second per Hz [35, 36]. It means that in one second and per Hz of bandwidth, the number of error bits once eavesdropper $e \in \mathcal{E}$ tries to decode the captured data from user k can be represented as $[r_k^{\mathsf{info}} - r_{e,k}^{\mathsf{leak}}]^+$, where $[x]^+$ denotes $\max\{0, x\}$. Therefore, we define the information leakage ratio of user k's data at eavesdropper e as follows,

$$\epsilon_{e,k} = \min\left(1, \frac{r_{e,k}^{\mathsf{leak}}}{r_k^{\mathsf{info}}}\right). \tag{8}$$

We further determine that transmission is considered secure by the PLS criteria if and only if $\epsilon_{e,k}$ is less than an acceptable threshold ϵ^{th} (where $0 \le \epsilon^{\mathsf{th}} \le 1$). It means that with a sufficiently small SINR $\gamma_{e,k}^{\mathsf{leak}}$, the eavesdropper $e \in \mathcal{E}$ cannot successfully decode the transmit message of user k, even if user k does not employ any data encryption method [37]. In addition, we define $1-\epsilon_{e,k}$ as the BEP at the eavesdropper e.

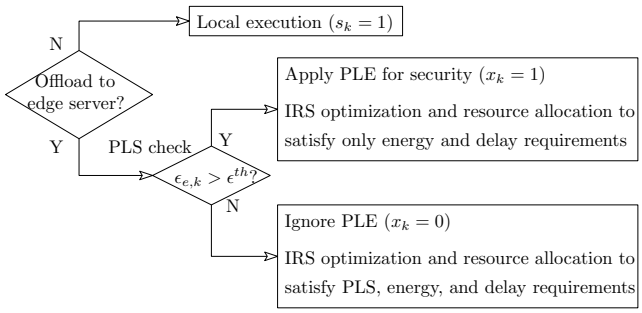

Fig. 2: Flowchart of the proposed joint PLE and PLS in IRS-assised MEC systems.

2) Data Encryption: In an uplink communication system, where single-antenna users isotropically transmit their data to the BS, $\epsilon_{e,k}$ might be greater than an acceptable threshold; thereby, information leakage cannot be avoided [38]. According to [39], a message can be reconstructed even from partial information of the original message. To prevent smart eavesdroppers from reconstructing the message [40], we consider communication systems supporting both PLS and PLE. Specifically, user k does not encrypt its data when there is a potential threat from eavesdroppers if the information leakage ratio at eavesdropper e is less than a threshold ϵ^{th}. Otherwise, user k must encrypt its data[2].

Let x_k be a binary encryption decision variable where $x_k = 1$ if the data of user k is encrypted; otherwise $x_k = 0$. Then, the joint PLS and PLE constraint can be expressed as follows,

$$(1-x_k)\epsilon_{e,k} \le (1-x_k)\epsilon^{\mathsf{th}}, \forall k \in \mathcal{K}. \tag{9}$$

The relation between the encryption decision and offloading decision is then expressed as follows,

$$x_k \le 1 - s_k, \forall k \in \mathcal{K}. \tag{10}$$

The joint PLE, PLS, and resource allocation process is summarized in Fig. 2. Specifically, when a user offloads its task to the edge server and the PLS protection is sufficient to meet the security requirements, PLE is not needed. Otherwise, PLE is applied to ensure the security requirement is satisfied.

C. Energy and Latency Estimation

Let c_k^{en} and c_k^{de} be the number of CPU cycles to encrypt and decrypt the data of user k, respectively. Then, the local computational energy and time[3] due to user k can be expressed, respectively, as follows [1],

$$\begin{aligned} \xi_k^{\mathsf{Lo}} &= \kappa_k f_k^2 \left(s_k c_k + x_k c_k^{\mathsf{en}}\right), \\ t_k^{\mathsf{Lo}} &= \frac{s_k c_k + x_k c_k^{\mathsf{en}}}{f_k}, \end{aligned} \tag{11}$$

where κ_k denotes the energy coefficient specified in the CPU model and f_k denotes the CPU clock speed of user k (CPU

[2]Assume that users and the BS have prior knowledge of key management, with all necessary coordination for encryption, decryption, and security parameters.

[3]Since cryptographic key renewal periods far exceed computation task delays, key management energy and time are neglected [41, 42].

cycles/second or Hz), which is assumed to be smaller than the UE's maximum clock speed $F_k^{\max}$.

We assume that the task of a specific user can not run in parallel. Let $\alpha_{k,l}$ be a binary variable where $\alpha_{k,l} = 1$ if the task of user k is executed by edge server $l \in \mathcal{L}$. Then, we have

$$\sum_{l\in\mathcal{L}} \alpha_{k,l} \leq 1, \forall k \in \mathcal{K}. \tag{12}$$

The number of required CPU cycles of user k at the edge server is then given by

$$c_k^{\mathsf{Ed}} = (1 - s_k)c_k + x_k c_k^{\mathsf{de}} \tag{13}$$

Let f_k^{Ed} be the allocated CPU clock speed to user k at the edge server. The computational time required by user k at the edge server is computed as $t_k^{\mathsf{Ed}} = c_k^{\mathsf{Ed}} \left(f_k^{\mathsf{Ed}}\right)^{-1}$. The total computational energy at co-located edge servers is calculated as follows,

$$\xi^{\mathsf{Ed}} = \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(\sum_k \alpha_{k,l} f_k^{\mathsf{Ed}}\right)^2 \sum_k \alpha_{k,l} c_k^{\mathsf{Ed}}. \tag{14}$$

where κ_l^{Ed} denotes the energy coefficient due to the CPU equipped at edge sever l. The transmission time of user k is given by

$$t_k^{\mathsf{Tx}} = \frac{(1 - s_k)b_k}{Br_k^{\mathsf{info}}}, \tag{15}$$

where B is the bandwidth. The transmission energy is then given by $\xi_k^{\mathsf{Tx}} = p_k t_k^{\mathsf{Tx}}$.

The total latency experienced by an offloaded task has different components, namely the time required for sending transmission bits to the edge server, the computation time in the edge server, and the time required for downloading the results to the UE from the edge server. For many practical applications, the amount of data required to report the computation outcome is typically much smaller than the amount of offloading data [1]. Therefore, we omit the time for sending back the computation outcome from the edge server to mobile users in our design. Consequently, user k needs to either execute the task locally or offload b_k bits to the BS within the maximum delay $\eta_k^{\max}$, which is expressed as follows,

$$t_k^{\mathsf{Lo}} + t_k^{\mathsf{Tx}} + t_k^{\mathsf{Ed}} \leq \eta_k^{\max}, \forall k \in \mathcal{K}. \tag{16}$$

D. Problem Formulation

In addition to energy efficiency, fairness among users has gained increasing attention, especially in heterogeneous networks where devices may have significantly different capabilities [43, 44]. While minimizing total energy consumption may favor devices with better channel conditions or more computation resources, it can lead to unfair resource allocation and potentially exclude weaker devices from effectively offloading their tasks. In this paper, we focus on addressing the fairness issue by minimizing energy consumption from the users' perspective in a secure MEC network. The primary objective is to maximize the service lifetime of users while ensuring fairness among them. To this end, we propose a design that involves optimizing multiple aspects, including encryption decisions, offloading decisions, computation, and radio resource allocation. The design aims to minimize the maximum weighted energy consumption at mobile users considering latency, joint PLE and PLS constraint, and limited computation-radio resource conditions. This problem can be formulated as follows,

$$\begin{aligned}
(\mathcal{P}) \quad & \min_{\boldsymbol{p},\boldsymbol{\Theta},\boldsymbol{W},\boldsymbol{s},\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{f},\boldsymbol{f}^{\mathsf{Ed}}} \quad \max_k \ \beta_k \left(\xi_k^{\mathsf{Lo}} + \xi_k^{\mathsf{Tx}}\right) \\
& \text{s.t. } \xi^{\mathsf{Ed}} \leq \xi^{\mathsf{Ed},\max}, && \text{(17a)} \\
& \quad p_k \leq p_k^{\max}, && \text{(17b)} \\
& \quad 0 \leq f_k \leq F_k^{\max}, && \text{(17c)} \\
& \quad 0 \leq f_k^{\mathsf{Ed}}, && \text{(17d)} \\
& \quad \sum_k \alpha_{k,l} f_k^{\mathsf{Ed}} \leq F^{\max}, \forall l \in \mathcal{L}, && \text{(17e)} \\
& \quad \sum_{l\in\mathcal{L}} \alpha_{k,l} \geq 1 - s_k, && \text{(17f)} \\
& \quad |\theta_{i,m}| = 1, && \text{(17g)} \\
& \quad (9), (10), (12), (16),
\end{aligned}$$

where $\boldsymbol{p} = \{p_k, \forall k \in \mathcal{K}\}$, $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_i, \forall i \in \mathcal{I}\}$, $\boldsymbol{W} = [\boldsymbol{w}_k, \forall k \in \mathcal{K}]$, $\boldsymbol{s} = \{s_k, \forall k \in \mathcal{K}\}$, $\boldsymbol{x} = \{x_k, \forall k \in \mathcal{K}\}$, $\boldsymbol{f} = \{f_k, \forall k \in \mathcal{K}\}$, $\boldsymbol{f}^{\mathsf{Ed}} = \{f_k^{\mathsf{Ed}}, \forall k \in \mathcal{K}\}$, and $\boldsymbol{\alpha} = \{\alpha_{k,l}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}\}$. In problem $(\mathcal{P})$, β_k is the energy weight of user k, $\xi^{\mathsf{Ed},\max}$ is the maximum consumed energy at the edge server, $p_k^{\max}$ is the maximum transmit power of user k, $F_k^{\max}$ is the maximum clock speed of user k, $F^{\max}$ is the edge server's maximum clock speed.

In this problem, constraint (17a) represents the energy consumption requirement at the edge server, while limited computing and wireless resources at the mobile users are captured in constraints (17b) and (17c), respectively. Constraint (17e) represents the sharing of computing resources at the edge server. Specifically, it allows a particular edge server to execute multiple tasks in the considered interval. Constraint (17f) states that if a user k offloads its task to the edge server, at least one assigned edge server should be available to process that task. Meanwhile, constraint (17g) illustrates the reflective condition of the IRS's elements.

III. Algorithm Development

In this paper, we assume a central controller is responsible for globally managing the MEC network. The controller collects data from all users and network components, determines encryption decisions, offloading strategies, computation, and radio resource allocation, and then communicates these decisions back to the users and network elements through secure control channels. The following sections detail our proposed algorithm for obtaining these decisions.

A. Bisection Search Algorithm

To gain insight into its non-smooth min-max objective function, we recast $(\mathcal{P})$ into the following problem,

$$\begin{aligned}
(\mathcal{P}_0) \ & \min_{\Omega_1 \cup \xi} \ \xi \\
& \text{s.t. } \beta_k \left(\xi_k^{\mathsf{Lo}} + \xi_k^{\mathsf{Tx}}\right) \leq \xi, \forall k, && \text{(18a)} \\
& \quad (17a) - (17g), (9), (10), (12), (16),
\end{aligned}$$

Algorithm 1 Joint Offloading, PLE and PLS, IRS coefficient control, and Resource Allocation algorithm

1: **Initialize**: Compute $\xi_k^{\mathsf{Lo}}, \forall k \in \mathcal{K}$ as in (19), choose ζ, assign $\xi^{\min} = 0$, $\xi^{\max} = \max_k(\xi_k^{\mathsf{Lo}})$, and set BOOL = *False*.
2: **while** $(\xi^{\max} - \xi^{\min} > \zeta)$ & (BOOL = *False*) **do**
3: Assign $\xi = (\xi_{\mathsf{max}} + \xi_{\mathsf{min}})/2$, and then define sets $\mathcal{A} = \{k|\xi_k^{\mathsf{Lo}} \leq \xi\}$ and $\mathcal{B} = \mathcal{K}/\mathcal{A}$.
4: Check feasibility of $(\mathcal{P}_\mathcal{B})$ as in Section III-C.
5: **if** $(\mathcal{P}_\mathcal{B})$ *is feasible* **then** $\xi^{\max} = \xi$, BOOL = *True*, **else** $\xi^{\min} = \xi$, BOOL = *False*, **end if**
6: **end while**

where $\mathbf{\Omega}_1$ is the set of variables in problem $(\mathcal{P})$ and ξ is an auxiliary variable. We also denote $\mathbf{\Omega}_{1,k}$ is the set of variables corresponding to user k in $\mathbf{\Omega}_1$. $(\mathcal{P}_0)$ is a MINLP problem which is difficult to solve due to the complex fractional of the transmission time and energy consumption, the logarithmic transmission rate function, and the mix of binary and continuous variables.

To tackle the problem, we adapt the method proposed in [2], where bisection search is employed to exploit the monotonicity of the objective function ξ, and SCA is used to iteratively solve a sequence of convex subproblems that approximate the original non-convex problem. In particular, we first categorize the users into two distinct sets: the *"locally executing user set"* consists of users who execute their applications locally, while the *"offloading user set"* comprises users who offload their applications for processing at the edge server. The proposed algorithm is developed based on the bisection search approach where in each search iteration, we perform: 1) user classification based on the current value of ξ using the results in Theorem 1 below; 2) feasibility verification for sub-problem $(\mathcal{P}_\mathcal{B})$ of $(\mathcal{P}_0)$ corresponding to the offloading user set $\mathcal{B}$; and 3) updates of lower and upper bounds on ξ according to the feasibility verification outcome. The detailed design is presented in the following.

B. User Classification

Let $\mathcal{A}$ and $\mathcal{B}$ be the locally executing and the offloading user sets, respectively. We further define any pair of sets $(\mathcal{A}, \mathcal{B})$ satisfying $\mathcal{B} = \mathcal{K}\backslash\mathcal{A}$ as a user classification. By denoting $\Omega_\mathcal{B} = \cup_{k\in\mathcal{B}}\Omega_{1,k}$, then for a given classification $(\mathcal{A}, \mathcal{B})$, problem $(\mathcal{P}_0)$ can be addressed by solving two sub-problems $(\mathcal{P}_\mathcal{A})$ and $(\mathcal{P}_\mathcal{B})$ for the users in sets $\mathcal{A}$ and $\mathcal{B}$, respectively, as follows,

$$
\begin{aligned}
(\mathcal{P}_\mathcal{A}) \quad & \min_{\{f_k\}_{k\in\mathcal{A}},\xi} \ \xi \\
\text{s.t} \quad & (\text{CA0}): \ \beta_k\kappa_k(f_k)^2 c_k \leq \xi, \forall k \in \mathcal{A}, \\
& (\text{CA1}): \ c_k/\eta_k^{\mathsf{max}} \leq f_k \leq F_k^{\mathsf{max}}, \forall k \in \mathcal{A}, \\
(\mathcal{P}_\mathcal{B}) \quad & \min_{\Omega_\mathcal{B},\xi} \ \xi \\
\text{s.t} \quad & (18a), (17a) - (17g), (9), (10), (12), (16), \forall k \in \mathcal{B},
\end{aligned}
$$

To attain more insight into the user classification, we now study the relationship between $(\mathcal{P}_\mathcal{A})$ and $(\mathcal{P}_\mathcal{B})$ in the following proposition and theorem.

Proposition 1. *The optimal objective value of* $(\mathcal{P}_\mathcal{A})$ *can be expressed as* $\xi_\mathcal{A}^\star = \max_{k\in\mathcal{A}} \xi_k^{\mathsf{Lo}}$, *where* ξ_k^{Lo} *is defined as*

$$\xi_k^{\mathsf{Lo}} = \beta_k\kappa_k \left(\min\left(\frac{c_k}{\eta_k^{\max}}, F_k^{\max}\right)\right)^2 c_k. \tag{19}$$

Proof. Because the optimization variables associated with the locally executing users are independent from the optimization variables of the other users and $\beta_k\kappa_k(f_k)^2 c_k$ is a monotonically increasing function with respect to f_k, the consumed energy is smallest if f_k is smallest. □

Theorem 1. *If* $\xi^\star$ *is the optimum objective value of problem* $(\mathcal{P}_2)$, *then an optimal classification,* $(\mathcal{A}^\star, \mathcal{B}^\star)$, *can be determined as* $\mathcal{A}^\star = \{k|\xi_k^{\mathsf{Lo}} \leq \xi^\star\}$, *and* $\mathcal{B}^\star = \mathcal{K}\backslash\mathcal{A}^\star$.

Proof: The proof is given in Appendix A in [2]. ■

C. General Optimal Algorithm Design

To solve the problem $(\mathcal{P})$, we utilize the results from Theorem 1 to develop an optimal algorithm. The algorithm iteratively solves the sub-problems $(\mathcal{P}_\mathcal{A})$ and $(\mathcal{P}_\mathcal{B})$ while updating the sets $(\mathcal{A}, \mathcal{B})$ until the optimal sets $(\mathcal{A}^\star, \mathcal{B}^\star)$ are obtained. The general optimal algorithm is outlined in Algorithm 1. In this algorithm, we employ the bisection search to find the optimum $\xi^\star$ where upper bound $\xi^{\max}$ and lower bound $\xi^{\min}$ are iteratively updated until the difference between them becomes sufficiently small, and the sets $\mathcal{A}$ and $\mathcal{B}$ no longer change. At convergence, the optimal classification solution can be obtained by merging the solutions of $(\mathcal{P}_\mathcal{A})$ and $(\mathcal{P}_\mathcal{B})$. The optimal solution of $(\mathcal{P}_\mathcal{A})$ can be determined using Proposition 1 and the verification of the feasibility of $(\mathcal{P}_\mathcal{B})$ is addressed in the following. The relationship among the (sub)problems involved in solving $(\mathcal{P})$ is illustrated in Fig. 3. Specifically, an alternating optimization approach is applied to the proposed subproblems $(\mathcal{P}_1) - (\mathcal{P}_7)$ to verify the feasibility of the MINLP problem $(\mathcal{P}_\mathcal{B})$. In this framework, subproblems $(\mathcal{P}_1) - (\mathcal{P}_3)$ determine the IRS's coefficients and encryption decision, $(\mathcal{P}_4) - (\mathcal{P}_6)$ handle the transmit power and computing resource allocation, and $(\mathcal{P}_7)$ addresses the computing resource association at the edge server.

IV. Feasibility verification of $(\mathcal{P}_\mathcal{B})$

For a given ξ, to address the challenge of feasibility verification for $(\mathcal{P}_\mathcal{B})$, which is hindered by the non-convex nature of the constraints, we adopt an alternating optimization approach. This approach involves iteratively optimizing each set of variables while keeping the values of other variables fixed in the corresponding sub-problems. We describe how to solve different subproblems in the following.

A. Determining the IRS's Coefficients and Encryption Decision

For given $\xi, \boldsymbol{s}, \boldsymbol{p}, \boldsymbol{W}, \boldsymbol{f}, \boldsymbol{f}^{\mathsf{Ed}}$, if $x_k = 1$, (16) and (18a) imply the following condition:

$$\gamma_k^{\mathsf{info}} \geq \gamma_{k,1}, \tag{20}$$

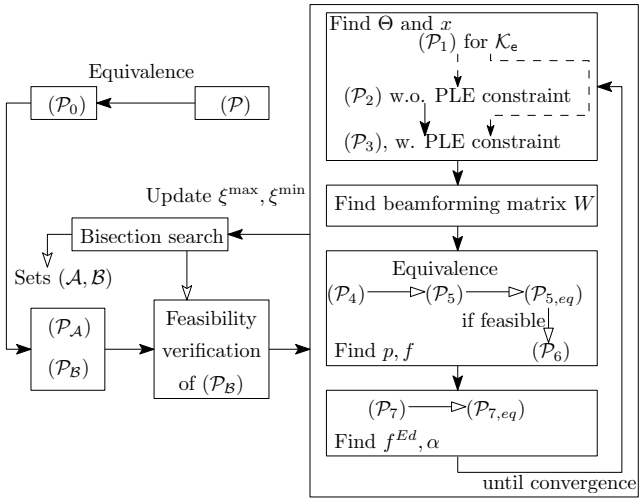

Fig. 3: Relationship between the (sub)problems when solving $(\mathcal{P})$.

where $\gamma_{k,1} = 2^{\frac{b_k}{B\min\left\{\frac{\xi}{\beta_k p_k} - \frac{\kappa_k f_k^2(c_k^{\mathsf{en}})}{p_k}, \eta_k^{\max} - \frac{c_k^{\mathsf{en}}}{f_k} - \frac{c_k + c_k^{\mathsf{de}}}{f_k^{\mathsf{Ed}}}\right\}}} - 1$. Meanwhile, if $x_k = 0$, (9), (16), and (18a) imply the following condition:

$$\begin{aligned} &\gamma_k^{\mathsf{info}} \geq \gamma_{k,2}, \\ &\gamma_{k,e} \leq (1+\gamma_k^{\mathsf{info}})^{\epsilon^{\mathsf{th}}} - 1 \end{aligned} \tag{21}$$

where $\gamma_{k,2} = 2^{\frac{b_k}{B\min\left\{\frac{\xi}{\beta_k p_k}, \eta_k^{\max} - \frac{c_k}{f_k^{\mathsf{Ed}}}\right\}}} - 1$. In (20) and (21), it is observed that $\gamma_{k,1}$ is larger than $\gamma_{k,2}$.

If constraints (20) and (17g) are satisfied when $x_k = 1, \forall k \in \mathcal{B}$, problem $(\mathcal{P}_\mathcal{B})$ is feasible. However, if the constraints are not satisfied when $x_k = 1$ for all $k \in \mathcal{B}$, we can determine the level of violation when solving the problem with the assumption $x_k = 1$. By examining the violated constraints, we can propose an effective method to determine the encryption decision and the IRSs' coefficients through the degree of constraint violation. In particular, we first identify the set of encrypted users and non-encrypted users in set $\mathcal{B}$ as $\mathcal{K}_\mathsf{e}$ and $\mathcal{K}_\mathsf{u}$, respectively. Mathematically, they are expressed as $\mathcal{K}_\mathsf{e} = \{k \in \mathcal{B} | x_k = 1\}$ and $\mathcal{K}_\mathsf{u} = \{k \in \mathcal{B} | x_k = 0\}$. At the beginning, assuming $x_k = 1, \forall k \in \mathcal{B}$. It means $\mathcal{K}_\mathsf{e} \equiv \mathcal{B}$ and $\mathcal{K}_\mathsf{u} = \emptyset$. The feasibility verification is equivalently transformed as follows

$$\begin{aligned} (\mathcal{P}_1)\ &\min_{\Theta,\zeta} \zeta \text{ s.t. } (17g), \\ &p_k |\boldsymbol{w}_k^H \boldsymbol{z}_k^{\mathsf{info}}|^2 \geq \Psi_k \gamma_{k,1} - \zeta, \forall k \in \mathcal{K}_\mathsf{e}, \end{aligned} \tag{22a}$$

where

$$\Psi_k = \sum_{j \in \mathcal{K}, j \neq k} p_j |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2. \tag{23}$$

Noted that if $\zeta \leq 0$, the problem is feasible. However, if $\zeta > 0$, it is not possible to satisfy all constraints of $(\mathcal{P}_\mathcal{B})$ with $x_k = 1$ for all $k \in \mathcal{K}_\mathsf{e}$. In such cases, when $\zeta > 0$ and $(\mathcal{P}_\mathcal{B})$ is feasible, the encryption decision for some users must be changed to $x_k = 0$. We define the violation level as $v_k = 1 - \gamma_{k,1}/\gamma_k^{\mathsf{info}}$. A negative value of v_k indicates a violation of the corresponding constraint for user k. As v_k increases, the violation level reduces. The additional resource required to satisfy (22a) is lower when v_k is getting larger. Therefore, we will change the decision from $x_k = 1$ to $x_k = 0$ for the user with the largest v_k value to optimize resource usage.

To implement this, we sort the list of users in $\mathcal{K}_\mathsf{e}$ in ascending order of v_k. Then, we will change and assign $x_k = 0$ where k is the first element of $\mathcal{K}_\mathsf{e}$. This process is represented as follows: $\mathcal{K}_\mathsf{u} = \mathcal{K}_\mathsf{u} \cup k$ and $\mathcal{K}_\mathsf{e} = \mathcal{K}_\mathsf{e} \backslash k$. We assume the PLE constraint is satisfied when $x_k = 0$ and re-verify this later. The problem is now rewritten as follows,

$$\begin{aligned} (\mathcal{P}_2)\ &\min_{\Theta,\zeta} \zeta \text{ s.t. } (17g), \\ &p_k |\boldsymbol{w}_k^H \boldsymbol{z}_k^{\mathsf{info}}|^2 \geq \Psi_k \gamma_{k,1} - \zeta, \forall k \in \mathcal{K}_\mathsf{e}, \qquad (24a) \\ &p_k |\boldsymbol{w}_k^H \boldsymbol{z}_k^{\mathsf{info}}|^2 \geq \Psi_k \gamma_{k,2} - \zeta, \forall k \in \mathcal{K}_\mathsf{u}, \qquad (24b) \end{aligned}$$

The process is repeated until $\zeta \leq 0$ or $\mathcal{K}_\mathsf{e} = \emptyset$. It can be verified that if $\zeta > 0$, the problem is infeasible. Otherwise, we continue to solve the problem with the PLE constraint as follows,

$$\begin{aligned} (\mathcal{P}_3)\ &\min_{\Theta,\zeta} \zeta \\ &\text{s.t. } (17g), (24a), (24b), \\ &p_k |z_{k,e}^{\mathsf{leak}}|^2 \leq \zeta + \Psi_{k,e} \left((1+\gamma_k^{\mathsf{info}})^{\epsilon^{\mathsf{th}}} - 1 \right), \forall k, e, \end{aligned} \tag{25a}$$

where $\Psi_{k,e} = \left(\sum_{j \in \mathcal{K}, j \neq k} p_j |z_{j,e}^{\mathsf{leak}}|^2 + \sigma_e^2 \right)$.

Constraint (25a) in problem $(\mathcal{P}_3)$ is challenging to address due to the non-integer power $(1+\gamma_k^{\mathsf{info}})^{\epsilon^{\mathsf{th}}}$. To overcome this, we employ the iterative method, where in iteration n, we define

$$\delta_k^{\min} = \left(1 + \gamma_k^{\mathsf{info}(n)}\right)^{\epsilon^{\mathsf{th}}} - 1, \tag{26a}$$

$$\gamma_k^{\mathsf{info}(n)} = \frac{p_k |\boldsymbol{w}_k^H \left(\boldsymbol{z}_k^{\mathsf{info}}\right)^{(n-1)}|^2}{\sum_{j \in \mathcal{K}, j \neq k} p_j |\boldsymbol{w}_k^H \left(\boldsymbol{z}_j^{\mathsf{info}}\right)^{(n-1)}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2}. \tag{26b}$$

Based on that, constraint (25a) can be rewritten as

$$p_k |z_{k,e}^{\mathsf{leak}}|^2 \leq \Psi_{k,e} \delta_k^{\min} + \zeta, \forall k \in \mathcal{K}_\mathsf{u}, \tag{27a}$$

Then, we will determine $\boldsymbol{\theta}$ to satisfy (17g), (24a), (24b), (27a). Specifically, non-convex constraint (27a) can be rewritten as follows,

$$\begin{aligned} &\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} \boldsymbol{\theta}_i^H \mathcal{F}_{i,j,k,e}(\delta_k^{\min}) \boldsymbol{\theta}_j + \mathcal{F}_{0,k,e}(\delta_k^{\min}) \\ &+ \sum_{i \in \mathcal{I}} 2\mathbb{R}\left(\boldsymbol{\theta}_i^H \mathcal{F}_{i,k,e,\mathsf{d}}(\delta_k^{\min}) \right) \leq 0, \forall e \in \mathcal{E}, k \in \mathcal{K}, \end{aligned} \tag{28}$$

where

$$\begin{aligned} \mathcal{F}_{i,j,k,e}(x) &= p_k \boldsymbol{h}_{i,k,e}^{\mathsf{R,leak}} \boldsymbol{h}_{j,k,e}^H - x \sum\nolimits_{k' \neq k} p_{k'} \boldsymbol{h}_{i,k',e} \boldsymbol{h}_{j,k',e}^H, \\ \mathcal{F}_{0,k,e}(x) &= p_k |h_{e,k}^{\mathsf{D,leak}}|^2 - x \sum\nolimits_{k' \neq k} p_{k'} |h_{e,k',\mathsf{d}}|^2 - x\sigma_e^2, \\ \mathcal{F}_{i,k,e,\mathsf{d}}(x) &= p_k \boldsymbol{h}_{i,k,e}^{\mathsf{R,leak}} \left(h_{e,k}^{\mathsf{D,leak}}\right)^* - x \sum_{k' \neq k} p_{k'} \boldsymbol{h}_{i,k',e} \left(h_{e,k',\mathsf{d}}\right)^*, \end{aligned}$$

Non-convex constraint (24a) or (24b) is rewritten as follows,

$$\begin{aligned} &\sum\nolimits_{i \in \mathcal{I}} \sum\nolimits_{j \in \mathcal{I}} \boldsymbol{\theta}_i^H \mathcal{G}_{i,j,k}(\gamma_{k,0}) \boldsymbol{\theta}_j + \mathcal{G}_{0,k}(\gamma_{k,0}) \\ &+ \sum\nolimits_{i \in \mathcal{I}} 2\mathbb{R}\left(\boldsymbol{\theta}_i^H \mathcal{G}_{i,k,\mathsf{d}}(\gamma_{k,0}) \right) \leq 0, \end{aligned} \tag{29}$$

where

$$\mathcal{G}_{i,j,k}(x) = x_k\bigg(\sum_{k'\neq k} p_{k'}\check{\boldsymbol{G}}_{i,k,k'}\boldsymbol{G}_{j,k'}\bigg) - p_k\check{\boldsymbol{G}}_{i,k,k}\boldsymbol{G}_{j,k},$$

$$\mathcal{G}_{0,k}(x) = -p_k|\boldsymbol{h}^{\mathsf{u}}_{k,\mathsf{d}}|^2 + x_k\bigg(\sum_{k'\neq k} p_{k'}|\boldsymbol{h}^{\mathsf{u}}_{k',\mathsf{d}}|^2 + \sigma_k^2|\boldsymbol{w}_k^H|^2\bigg),$$

$$\mathcal{G}_{i,k,\mathsf{d}}(x) = -p_k\check{\boldsymbol{G}}_{i,k,k}\boldsymbol{h}^{\mathsf{u}}_{k,\mathsf{d}} + x_k\sum_{k'\neq k} p_{k'}\check{\boldsymbol{G}}_{i,k,k'}\boldsymbol{h}^{\mathsf{u}}_{k',\mathsf{d}},$$

where $\boldsymbol{G}_{i,k} = \boldsymbol{Q}_i^{\mathsf{R2B}}\,\mathrm{diag}(\boldsymbol{h}^{\mathsf{RIS}}_{i,k})$, $\check{\boldsymbol{G}}_{i,k,k'} = \boldsymbol{G}^H_{i,k'}\boldsymbol{w}_k\boldsymbol{w}_k^H$, $\gamma_{k,0} = \gamma^{(n)}_{k,1}$ if (24a) is considered, and $\gamma_{k,0} = \gamma^{(n)}_{k,2}$ if (24b) is considered.

In (28) and (29), terms $\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\boldsymbol{\theta}_i^H\mathcal{F}_{i,j,k,e}(\delta_k^{\min})\boldsymbol{\theta}_j$ and $\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\boldsymbol{\theta}_i^H\mathcal{G}_{i,j,k}(\gamma_{k,0})\boldsymbol{\theta}_j$ are non-convex. However, according to [45], for given Hermitian matrices $\boldsymbol{L}$ and $\boldsymbol{S}$ such that $\boldsymbol{L}\succeq\boldsymbol{S}$ and for an arbitrarily given $\boldsymbol{\theta}_{(0)}$, we have

$$\boldsymbol{\theta}^H\boldsymbol{S}\boldsymbol{\theta}\leq\boldsymbol{\theta}^H\boldsymbol{L}\boldsymbol{\theta}-2\mathbb{R}\big\{\boldsymbol{\theta}^H(\boldsymbol{L}-\boldsymbol{S})\boldsymbol{\theta}_{(0)}\big\}+\boldsymbol{\theta}^H_{(0)}(\boldsymbol{L}-\boldsymbol{S})\boldsymbol{\theta}_{(0)}. \tag{30}$$

It can be verified that $\boldsymbol{L} = \lambda_{\max}\boldsymbol{I}$, where $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{S}$, we will have $\boldsymbol{L}\succeq\boldsymbol{S}$. Then, we apply (30) to approximate (28) and (29), which are given by (31) and (32), respectively, where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T,...,\boldsymbol{\theta}_I^T]^T$,

$$\begin{aligned}
\mathcal{F}_{k,e,\mathsf{d}}(\delta_k^{\min}) &= [\mathcal{F}_{1,k,e,\mathsf{d}}(\delta_k^{\min})^T,...,\mathcal{F}_{I,k,e,\mathsf{d}}(\delta_k^{\min})]^T,\\
\mathcal{F}_{k,e}(\delta_k^{\min}) &= \begin{bmatrix}\mathcal{F}_{1,1,k,e}(\delta_k^{\min}) & ... & \mathcal{F}_{1,I,k,e}(\delta_k^{\min})\\ \vdots & ... & \vdots\\ \mathcal{F}_{I,1,k,e}(\delta_k^{\min}) & ... & \mathcal{F}_{I,I,k,e}(\delta_k^{\min})\end{bmatrix},\\
\mathcal{G}_{k,\mathsf{d}}(\gamma_{k,0}) &= [\mathcal{G}_{1,k,\mathsf{d}}(\gamma_{k,0})^T,...,\mathcal{G}_{I,k,\mathsf{d}}(\gamma_{k,0})]^T,\\
\mathcal{G}_{k}(\gamma_{k,0}) &= \begin{bmatrix}\mathcal{G}_{1,1,k}(\gamma_{k,0}) & ... & \mathcal{G}_{1,I,k}(\gamma_{k,0})\\ \vdots & ... & \vdots\\ \mathcal{G}_{I,1,k}(\gamma_{k,0}) & ... & \mathcal{G}_{I,I,k}(\gamma_{k,0})\end{bmatrix},
\end{aligned} \tag{33}$$

and $\lambda^{\max}_{k,e}$ and $\lambda^{\max}_{k}$ are the maximum eigenvalue of $\mathcal{F}_{k,e}(t)$ and $\mathcal{G}_k(t)$, respectively. Because $\theta^H_{i,m}\theta_{i,m} = 1$, thus $\boldsymbol{\theta}^H\lambda^{\max}_{k,e}\boldsymbol{\theta} = \lambda^{\max}_{k,e}\sum_{i\in\mathcal{I}}M_i$. Then, the subproblem $\mathcal{P}_3$ in iteration n is approximated as a feasibility verification of constraints (17g), (31), and (32). The Lagrangian function considering these constraints is given by

$$\mathcal{L}_{(\mathcal{P})_{\mathsf{IRS}}} = -2\mathbb{R}\big(\boldsymbol{\theta}^H\boldsymbol{\chi}\big) + \sum_{(k,e)}(1-x_k)\varrho_{k,e}\boldsymbol{\delta}_{k,e,0} + \sum_{k\in\mathcal{K}}\varrho_k\boldsymbol{\delta}_{k,0}, \tag{34}$$

where $\varrho_{k,e}$ and ϱ_k are Lagrangian multipliers, and $\boldsymbol{\chi} = \sum_{k\in\mathcal{K},e\in\mathcal{E}}(1-x_k)\varrho_{k,e}\boldsymbol{\delta}_{k,e} + \sum_{k\in\mathcal{K}}\varrho_k\boldsymbol{\delta}_k$, $\boldsymbol{\delta}_{k,e} = \Big(\lambda^{\max}_{k,e}\boldsymbol{I} - \mathcal{F}_{k,e}(\delta_k^{\min}t)\Big)\boldsymbol{\theta}_{(0)} - \mathcal{F}_{k,e,\mathsf{d}}(\delta_k^{\min})$, $\boldsymbol{\delta}_k = \big(\lambda^{\max}_k\boldsymbol{I} - \mathcal{G}_k(\gamma_{k,0})\big)\boldsymbol{\theta}_{(0)} - \mathcal{G}_{k,\mathsf{d}}(\gamma_{k,0})$, $\boldsymbol{\delta}_{k,e,0} = \lambda^{\max}_{k,e}\sum_{i\in\mathcal{I}}M_i + \mathcal{F}_{0,k,e}(\delta_k^{\min}) + \boldsymbol{\theta}^H_{(0)}\Big(\lambda^{\max}_{k,e}\boldsymbol{I} - \mathcal{F}_{k,e}(\delta_k^{\min})\Big)\boldsymbol{\theta}_{(0)}$, $\boldsymbol{\delta}_{k,0} = \lambda^{\max}_k\sum_{i\in\mathcal{I}}M_i + \mathcal{G}_{0,k}(\gamma_{k,0}) + \boldsymbol{\theta}^H_{(0)}\big(\lambda^{\max}_k\boldsymbol{I} - \mathcal{G}_k(\gamma_{k,0})\big)\boldsymbol{\theta}_{(0)}$.

For given $\varrho_{k,e}$ and ϱ_k, $\forall k\in\mathcal{K}, e\in\mathcal{E}$, the problem becomes

$$(\mathcal{P}_3)^{(n)}\ \max_{\boldsymbol{\theta}}\ 2\mathbb{R}\big(\boldsymbol{\theta}^H\boldsymbol{\chi}\big)\ \text{s.t. (17g)}.$$

We denote the element (i,m) (i.e., the $(m+\sum_{j<i}M_j)^{\mathsf{th}}$ element of vector $\boldsymbol{\chi}$) by $\chi_{i,m}$. Then, it is trivial to obtain the optimal solution to the problem, which is $\theta^{(n+1)}_{i,m} = \boldsymbol{e}^{-j\angle\chi_{i,m}}$. The Lagrangian multipliers are updated as follows, $\varrho^{(n+1)}_{k,e} = \Big[\varrho^{(n)}_{k,e} + \psi\big(\gamma^{\mathsf{leak}}_{e,k} - \delta_k^{\min}\big)\Big]^+$, $\varrho^{(n+1)}_{k} = \Big[\varrho^{(n)}_{k} + \psi\big(\gamma_{k,0} - \gamma^{\mathsf{info}}_k\big)\Big]^+$, where ψ is the step size. Noted that this approach is also utilized to solve $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$.

B. Determining the Detection Matrix $\boldsymbol{W}$

According to [46], the optimal MIMO detection vector $\boldsymbol{w}_k$ is given by

$$\begin{aligned}
\boldsymbol{w}_k &= \underset{w_k}{\mathrm{argmax}}\,\gamma^{\mathsf{info}}_k\\
&= \underset{w_k}{\mathrm{argmax}}\,\frac{\boldsymbol{w}_k^H p_k\boldsymbol{z}^{\mathsf{info}}_k\big(\boldsymbol{z}^{\mathsf{info}}_k\big)^H\boldsymbol{w}_k}{\boldsymbol{w}_k^H\Big(\sum_{j\in\mathcal{K},j\neq k}p_j\boldsymbol{z}^{\mathsf{info}}_j\big(\boldsymbol{z}^{\mathsf{info}}_j\big)^H + \sigma_k^2\boldsymbol{I}\Big)\boldsymbol{w}_k}\\
&= \max\mathrm{eigenvector}\Big(\boldsymbol{U}_k\Big),
\end{aligned} \tag{36}$$

where $\boldsymbol{U}_k = \left(\sum_{j\in\mathcal{K},j\neq k}p_j\boldsymbol{z}^{\mathsf{info}}_j\big(\boldsymbol{z}^{\mathsf{info}}_j\big)^H + \sigma_k^2\boldsymbol{I}\right)^{-1}p_k\boldsymbol{z}^{\mathsf{info}}_k\big(\boldsymbol{z}^{\mathsf{info}}_k\big)^H$.

C. Determining the Transmit Power and the Computing Resources

For given $\xi,\boldsymbol{\theta},\boldsymbol{W},\boldsymbol{s}$, and $\boldsymbol{x}$, constraints (9), (18a), (17a), (17b), (17c), (17d), (17e), (17f), (12), and (16) need to be verified. However, because $\alpha_{k,l}$ is an integer variable, the feasibility verification becomes more challenging. However, there are constraints (12), (17a), (17e), and (17f) related to $\alpha_{k,l}$. In order to address this issue, we first ignore $\alpha_{k,l}$ and constraints (12), (17a), (17e), and (17f) . The feasibility verification becomes

$$\begin{aligned}
(\mathcal{P}_4)\ &\min_{\boldsymbol{p},\boldsymbol{f},\boldsymbol{f}^{\mathsf{Ed}},\zeta}\ \zeta\\
&\text{s.t. }\beta_k\left(\xi^{\mathsf{Lo}}_k + \xi^{\mathsf{Tx}}_k\right)\leq\xi+\zeta,\forall k, && (37a)\\
&\quad t^{\mathsf{Lo}}_k + t^{\mathsf{Tx}}_k + t^{\mathsf{Ed}}_k\leq\eta^{\max}_k+\zeta,\forall k, && (37b)\\
&\quad (9),(17b),(17c),(17d).
\end{aligned}$$

Problem $(\mathcal{P}_4)$ is non-convex due to (9), (37a) and (37b). In order to address this issue, we consider t^{Tx}_k as an auxiliary variable. Then, the problem is rewritten as follows

$$\begin{aligned}
(\mathcal{P}_5)\ &\min_{\boldsymbol{p},\boldsymbol{f},\boldsymbol{f}^{\mathsf{Ed}},\boldsymbol{t}^{\mathsf{Tx}},\zeta}\ \zeta\\
&\text{s.t. }\xi^{\mathsf{Lo}}_k + p_k t^{\mathsf{Tx}}_k\leq\frac{\xi+\zeta}{\beta_k}, && (38a)\\
&\quad B\log_2(1+\gamma^{\mathsf{info}}_k)\geq\frac{b_k}{t^{\mathsf{Tx}}_k},\forall k, && (38b)\\
&\quad (9),(17b),(17c),(17d),(37b),
\end{aligned}$$

By substituting γ^{info}_k and $\gamma_{k,e}$ in (3) and (6), respectively, the problem is rewritten as follows,

$$\begin{aligned}
(\mathcal{P}_{5,\mathsf{eq}})\ &\min_{\boldsymbol{p},\boldsymbol{f},\boldsymbol{f}^{\mathsf{Ed}},\boldsymbol{t}^{\mathsf{Tx}},\zeta}\ \zeta\\
&\text{s.t. (17b),(17c),(17d),(37b),(38a),(40),(41).}
\end{aligned}$$

Problem $(\mathcal{P}_{5,\mathsf{eq}})$ is non-convex due to the objective and con-

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3607599

$$\boldsymbol{\theta}^H \lambda_{k,e}^{\max} \boldsymbol{\theta} - 2\mathbb{R}\left\{\boldsymbol{\theta}^H \left(\lambda_{k,e}^{\max} \boldsymbol{I} - \mathcal{F}_{k,e}(t)\right) \boldsymbol{\theta}_{(0)}\right\} + \boldsymbol{\theta}_{(0)}^H \left(\lambda_{k,e}^{\max} \boldsymbol{I} - \mathcal{F}_{k,e}(t)\right) \boldsymbol{\theta}_{(0)} + 2\mathbb{R}\left(\boldsymbol{\theta}^H \mathcal{F}_{k,e,\mathsf{d}}(t)\right) + \mathcal{F}_{0,k,e}(t) \leq 0, \tag{31}$$

$$\boldsymbol{\theta}^H \lambda_{k}^{\max} \boldsymbol{\theta} - 2\mathbb{R}\left\{\boldsymbol{\theta}^H \left(\lambda_{k}^{\max} \boldsymbol{I} - \mathcal{G}_{k}(t)\right) \boldsymbol{\theta}_{(0)}\right\} + \boldsymbol{\theta}_{(0)}^H \left(\lambda_{k}^{\max} \boldsymbol{I} - \mathcal{G}_{k,e}(t)\right) \boldsymbol{\theta}_{(0)} + 2\mathbb{R}\left(\boldsymbol{\theta}^H \mathcal{G}_{k,\mathsf{d}}(t)\right) + \mathcal{G}_{0,k}(t) \leq 0, \tag{32}$$

$$B\log_2(\sum\nolimits_j p_j |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2) - \frac{b_k}{t_k} \geq B\log_2(\sum\nolimits_{j\neq k} p_j |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2). \tag{40}$$

$$\log(\sum_j p_j |z_{j,e}^{\mathsf{leak}}|^2 + \sigma_e^2) + \epsilon^{\mathsf{th}} \log(\sum_{j\neq k} p_j |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2) \leq \log(\sum_{j\neq k} p_j |z_{j,e}^{\mathsf{leak}}|^2 + \sigma_e^2) + \epsilon^{\mathsf{th}} \log(\sum_j p_j |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2). \tag{41}$$

$$\log\left(\sum_{j\neq k} p_j |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2\right) \leq \log\left(\sum_{j\neq k} p_j^{(n-1)} |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2\right) + \frac{\sum_{j\neq k}(p_j - p_j^{(n-1)}) |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2}{\sum_{j\neq k} p_j^{(n-1)} |\boldsymbol{w}_k^H \boldsymbol{z}_j^{\mathsf{info}}|^2 + \sigma_k^2 |\boldsymbol{w}_k^H|^2}. \tag{42}$$

$$\log\left(\sum\nolimits_j p_j |z_{j,e}^{\mathsf{leak}}|^2 + \sigma_e^2\right) \leq \log\left(\sum\nolimits_j p_j^{(n-1)} |z_{j,e}^{\mathsf{leak}}|^2 + \sigma_e^2\right) + \frac{\sum_j (p_j - p_j^{(n-1)}) |z_{j,e}^{\mathsf{leak}}|^2}{\sum_j p_j^{(n-1)} |z_{j,e}^{\mathsf{leak}}|^2 + \sigma_e^2}. \tag{43}$$

straints (38a), (40), and (41). In order to convexify these non-convex functions, the iterative method is applied. In particular, in iteration n, we convexify these non-convex functions as follows,

$$p_k t_k \leq 0.5\big(p_k^2/\nu_{k,\mathsf{t}}^{(n)} + t_k^2 \nu_{k,\mathsf{t}}^{(n)}\big),), \tag{44}$$

where $\nu_{k,\mathsf{t}}^{(n)} = p_k^{(n-1)}/t_k^{(n-1)}$.

In addition, constraints (40) and (41) have the difference of convex (DC) form. As such, we apply the DC technique to convexify these constraints, which are given by (42) and (43). By substituting (42), (43), and (44) into the (38a), (40), and (41), the problem becomes convex, which can be solved effectively by the CVX solver. Then, problem $(\mathcal{P}_4)$ is feasible if ζ is non-positive.

On the other hand, in the alternating method applying to the feasibility verification problem, if $(\mathcal{P}_4)$ is feasible, besides satisfying constraints, the assignment of variables in this particular subproblem needs to yield a broader feasible region for the remaining subproblems. It means we will determine p_k, f_k, and f_k^{Ed}, for all k such that (12), (17a), (17e), and (17f) are satisfied if there exists a feasible solution of $\{\alpha_{k,l}, \forall k, \forall n\}$.

It can be observed that the biggest challenge of determining $\{\alpha_{k,l}, \forall k, \forall n\}$ is to assign computing resources at the edge server to satisfy the energy consumption constraint, which is reflected in constraint (17a). As shown in (14), the energy consumption is a quadratic function of the computing frequency. Therefore, when determining p_k, f_k, and f_k^{Ed}, for all k to satisfy (16), (17b), (17c), (17d), and (18a), we also minimize the total computing resource at the edge server if $(\mathcal{P}_4)$ is feasible. The optimization problem is given by

$$\begin{aligned}(\mathcal{P}_6) \quad & \min_{\boldsymbol{p},\boldsymbol{f},\boldsymbol{f}^{\mathsf{Ed}},\boldsymbol{t}^{\mathsf{Tx}}} \sum_k f_k^{\mathsf{Ed}} \\ \text{s.t. } & (9),(16),(17b),(17c),(17d),(18a), \\ & f_k^{\mathsf{Ed}} \leq F^{\max}, \forall n, \end{aligned} \tag{45a}$$

The process to convexify non-convex constraint is similar to solving $(\mathcal{P}_4)$.

D. Determining the Computing Resource Association at the Edge Server

In this section, we determine $\alpha_{k,l}, \forall k, n$ satisfying (12), (17a), (17e), and (17f). In addition, to improve the objective in the alternative optimization method, if feasible points exist, we also minimize the energy consumption ξ^{Ed}.

1) Optimization-based algorithm: The feasibility verification in this section is given by

$$\begin{aligned}(\mathcal{P}_7) \quad & \min_{\boldsymbol{\alpha}} \xi^{\mathsf{Ed}} \\ \text{s.t. } & (17e),(17f). \end{aligned}$$

The term $\left(\sum_k \alpha_{k,l} f_k^{\mathsf{Ed}}\right)^2 \sum_k \alpha_{k,l} c_k^{\mathsf{Ed}}$ in the consumed energy ξ^{Ed} is non-convex even with continuous $\alpha_{k,l}$. The following proposition is presented to address this issue.

Proposition 2. *Problem $(\mathcal{P}_7)$ is equivalently transformed as follows,*

$$\begin{aligned}(\mathcal{P}_{7,\mathsf{eq}}) \quad & \min_{\boldsymbol{\alpha},\check{\boldsymbol{\alpha}},\check{\boldsymbol{f}}^{\mathsf{Ed}}} \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(F^{\max}\right)^2 \sum_k \check{\alpha}_{k,n} c_k^{\mathsf{Ed}} \\ \textit{s.t. } & 0 \leq \check{\alpha}_{k,n} \leq \min\{\check{f}_n^{\mathsf{Ed}}, \alpha_{k,l}\}, && (47a) \\ & \check{\alpha}_{k,n} \geq \check{f}_n^{\mathsf{Ed}} + \alpha_{k,l} - 1, && (47b) \\ & \left(F^{\max}\right)^2 \check{f}_n^{\mathsf{Ed}} \geq \left(\sum_k \alpha_{k,l} f_k^{\mathsf{Ed}}\right)^2, && (47c) \\ & (17e),(17f). \end{aligned}$$

Proof. From (47c) and (17e), it can be verified that $0 \leq \check{f}_n^{\mathsf{Ed}} \leq 1$. From (47a) and (47b), we have $\check{\alpha}_{k,n} = \check{f}_n^{\mathsf{Ed}} \alpha_{k,l}$.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3607599

Let $\alpha_{k,l}^*$ and $\xi^{\mathsf{e},*}$ be the optimal solution and the objective of $(\mathcal{P}_6)$, respectively. We have

$$\begin{aligned}\xi^{\mathsf{e},*} \le \xi^{\mathsf{Ed}} &= \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(\sum_k \alpha_{k,l} f_k^{\mathsf{Ed}}\right)^2 \sum_k \alpha_{k,l} c_k^{\mathsf{Ed}} \\ &\le \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(F^{\max}\right)^2 \check{f}_n^{\mathsf{Ed}} \sum_k \alpha_{k,l} c_k^{\mathsf{Ed}} \\ &= \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(F^{\max}\right)^2 \sum_k \check{\alpha}_{k,n} c_k^{\mathsf{Ed}}, \forall \alpha_{k,l}, \check{\alpha}_{k,n}, \check{f}_n^{\mathsf{Ed}}.\end{aligned} \tag{48}$$

Let $\alpha_{k,l}^*$ is the optimal solution of $(\mathcal{P}_7)$. If we set $\check{f}_n^{\mathsf{e},*} = \left(\sum_k \alpha_{k,l} f_k^{\mathsf{Ed}}\right)^2 / \left(F^{\max}\right)^2$ and $\check{\alpha}_{k,n}^* = \check{f}_n^{\mathsf{e},*} \alpha_{k,l}^*$, $\{\alpha_{k,l}^*, \check{\alpha}_{k,n}^*, \check{f}_n^{\mathsf{e},*}\}$ is also a feasible solution of $(\mathcal{P}_{7,\mathsf{eq}})$. Then, we have:

$$\begin{aligned}&\min_{\boldsymbol{\alpha},\check{\boldsymbol{\alpha}},\check{\boldsymbol{f}}^{\mathsf{Ed}}} \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(F^{\max}\right)^2 \sum_k \check{\alpha}_{k,n} c_k^{\mathsf{Ed}} \\ &\le \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(F^{\max}\right)^2 \sum_k \check{\alpha}_{k,n}^* c_k^{\mathsf{Ed}} \\ &= \sum_{l\in\mathcal{L}} \kappa_l^{\mathsf{Ed}} \left(\sum_k \alpha_{k,l}^* f_k^{\mathsf{Ed}}\right)^2 \sum_k \alpha_{k,l}^* c_k^{\mathsf{Ed}} = \xi^{\mathsf{e},*}.\end{aligned} \tag{49}$$

From (48) and (49), it completes the proof. □

It can be verified that $(\mathcal{P}_{7,\mathsf{eq}})$ is an MINLP problem where all constraints and the objective are either linear or convex, which can be solved by the CVX solver. When the network size increases, solving this MINLP problem may be highly complex. In order to obtain a low-complexity solution, we consider the many-to-one matching game with externalities as follows.

2) Matching-based algorithm: Let $\mathcal{S} = \{f_1^{\mathsf{Ed}}, ..., f_K^{\mathsf{Ed}}\}$ and $\mathcal{D} = \{1, ..., L\}$ be the set of computing frequency clock speeds and edge servers, respectively.

Definition 1. *A many-to-one matching is a subset $\mu \in \mathcal{S} \times \mathcal{D}$ such that $|\mu(f)| = 1$ and $|\mu(l)| \le K$, where $\mu(f) = \{l \in \mathcal{D} : (f,l) \in \mu\}$ and $\mu(l) = \{f \in \mathcal{S} : (f,l) \in \mu\}$.*

The reference value for the computing resource f_k^{Ed} executed by edge server $l \in \mathcal{L}$ is $\kappa_l^{\mathsf{Ed}} \left(f_k^{\mathsf{Ed}} + \sum_{k'\neq k} \alpha_{k',l} f_{k'}^{\mathsf{Ed}}\right)^2 \left(c_k^{\mathsf{Ed}} + \sum_{k'\neq k} \alpha_{k',l} c_{k'}^{\mathsf{Ed}}\right)$. According to [47], the proposed matching game has externalities, where the preference value for the computing resource f_k^{Ed} not only depends on the edge server that it is matched with but also on the $f_{k'}^{\mathsf{Ed}}, k' \neq k$ matched to the same edge server.

To find the solution of the matching game with externalities, we define the preference value of edge server $l \in \mathcal{L}$ on the set of computing resource $\mathcal{S}_\mu$ under the matching state μ as the energy consumption in edge server $l \in \mathcal{L}$, i.e.,

$$U(\mathcal{S}_\mu, \mu) = \kappa_l^{\mathsf{Ed}} \left(\sum\nolimits_{k\in\mathcal{S}_\mu} f_k^{\mathsf{Ed}}\right)^2 \left(\sum\nolimits_{k\in\mathcal{S}_\mu} c_k^{\mathsf{Ed}}\right).$$

To define the exchange stability, we first consider the concept of swap matching as follows,

$$\mu_f^{f'} = \mu \backslash \{(f,\mu(f)),(f',\mu(f'))\} \cup \{(f,\mu(f')),(f',\mu(f))\}. \tag{50}$$

Algorithm 2 Task-edge server association using matching theory (TCAM) algorithm

1: **Initialize**:
- Variable $\alpha_{k,l}, \forall k,l$ are randomly assigned subject to constraints (12), (17e), and (17f).

2: **repeat**
3: For each f_k^{Ed}, search for another $f_{k'}^{\mathsf{Ed}}$ or an open spot $\mathcal{O}$ to form a swap-blocking pair.
4: **if** $\{f_k^{\mathsf{Ed}}, f_{k'}^{\mathsf{Ed}}\}$ or $\{f_k^{\mathsf{Ed}}, \mathcal{O}\}$ forms a swap-blocking pair **then**
5: Update $\mu = \mu_{f_k^{\mathsf{Ed}}}^{f_{k'}^{\mathsf{Ed}}}$ or $\mu = \mu_{f_k^{\mathsf{Ed}}}^{\mathcal{O}}$, respectively.
6: **end if**
7: **until** matching μ is two-sided exchange-stable.

In other words, a swap matching enables switching from $(f,\mu(f))$ to $(f,\mu(f'))$ and from $(f',\mu(f'))$ to $(f',\mu(f))$ while keeps all other matchings. Further, the computing resource involved in the swap can be a hole '$\mathcal{O}$' representing an open spot, thus allowing a single computing resource moving to available vacancies.

Definition 2. *For $i = \{f,f'\}, f \in \mathcal{S}, f' \in \mathcal{S} \cup \mathcal{O}$, let define*

$$\mathcal{U}_i(\mu) = U(\mathcal{S}_{\mu(f)}, \mu) + U(\mathcal{S}_{\mu(f')}, \mu) \tag{51}$$

Let $\mathcal{F}_{\mathsf{pair}}$ be the set of $\{f,f'\}$ such that constraints (12)*,* (17e)*, and* (17f) *are satisfied under either matching state μ or $\mu_f^{f'}$.*

Then, the $\{f,f'\} \in \mathcal{F}_{\mathsf{pair}}$ is a swap-blocking pair if and only if

$$\mathcal{U}_i(\mu_f^{f'}) < \mathcal{U}_i(\mu). \tag{52}$$

According to [47], a matching μ is a two-sided exchange-stable if no swap-blocking pairs exists. Therefore, we propose the algorithm to determine the binary decision $\alpha_{k,l}$ that attains the two-sided exchange-stable matching as in Algorithm. 2. Indeed, if the converged matching $\mu^\star$ is not the two-sided exchange-stable matching, a swap-blocking pair exists after the final matching $\mu^\star$. However, as shown in Algorithm. 2, the algorithm does not terminate until all swap-blocking pairs are eliminated. Therefore, the converged $\mu^\star$ must be a two-sided exchange-stable matching.

Furthermore, the decrease of consumed energy at the edge server after each swap operation shows the convergence of Algorithm. 2. Indeed, the energy consumption under matching μ is given by $\xi_\mu^{\mathsf{Ed}} = \sum_{l\in\mathcal{L}} U(\mathcal{S}_\mu, \mu)$. After the swap operation of $i = \{f,f'\} \in \mathcal{F}_{\mathsf{pair}}$, we define $q_{\mu\to\mu_f^{f'}} = \xi_{\mu_f^{f'}}^{\mathsf{Ed}} - \xi_\mu^{\mathsf{Ed}} = \mathcal{U}_i(\mu_f^{f'}) - \mathcal{U}_i(\mu) < 0$. In this case, $q_{\mu\to\mu_f^{f'}}$ is the difference of energy consumption at the edge server under the matching state $\mu_f^{f'}$ and that under the matching state μ.

Finally, the feasibility verification of problem $(\mathcal{P}_\mathcal{B})$ is presented in Algorithm 3. Note that, in each iteration to perform feasibility verification of $(\mathcal{P}_\mathcal{B})$, we will alternatively determine $\{\boldsymbol{\Theta}, \boldsymbol{x}\}$, $\boldsymbol{W}$, $\{\boldsymbol{p}, \boldsymbol{f}, \boldsymbol{f}^{\mathsf{Ed}}\}$, $\boldsymbol{\alpha}$ as in Sections IV-A, IV-B, IV-C, and IV-D, respectively. The convergence proof of our proposed alternating optimization-based algorithm follows a similar approach to that in [48]. It is worth noting that, in addition to verifying feasibility, problems $(\mathcal{P}_1) - (\mathcal{P}_7)$ are also designed to optimize control variables such that, when multiple feasible solutions exist, the algorithm avoids selecting boundary feasible points whenever possible. As a result, in the next iteration of the bisection search, when the

Algorithm 3 Algorithm for the feasibility verification of problem $(\mathcal{P}_{\mathcal{B}})$ corresponding to the value ξ

1: **Initialize**: Apply zero-forcing-based method to get the initial $\boldsymbol{W}$ and block-diagonalization-based method to get the initial $\boldsymbol{\Theta}$ [50]. Assign $p_k = p_k^{\max}$, $f_k = F_k^{\max}$, $f_k^{\mathsf{Ed}} = F^{\max} L/|\mathcal{B}|$, $\mathcal{K}_{\mathsf{e}} \equiv \mathcal{B}$, and $\mathcal{K}_{\mathsf{u}} = \emptyset$.
2: **Repeat**
3: Solve $(\mathcal{P}_1)$ with set $\mathcal{K}_{\mathsf{e}}$ to determine ζ.
4: **if** $\zeta \leq 0$ **then** Go to line 12.
5: **else**
6: **while** $\zeta > 0$ OR $\mathcal{K}_{\mathsf{e}} \neq \emptyset$ **do**
7: Sort users in $\mathcal{K}_{\mathsf{e}}$ in ascending order of $v_k = 1 - \gamma_{k,1}/\gamma_k^{\mathsf{info}}$.
8: Select k is the first element of $\mathcal{K}_{\mathsf{e}}$, and assign $\mathcal{K}_{\mathsf{u}} = \mathcal{K}_{\mathsf{u}} \cup k$ and $\mathcal{K}_{\mathsf{e}} = \mathcal{K}_{\mathsf{e}} \backslash k$.
9: Solve $(\mathcal{P}_2)$ to get ζ.
10: **end while**
11: **if** $\zeta \leq 0$ **then**
12: Solve $(\mathcal{P}_3)$ by iteratively solve $(\mathcal{P}_3)^{(n)}$ until convergence.
13: **end if**
14: **end if**
15: Determining $\boldsymbol{W}$ as in Section IV-B.
16: Determining ζ by solving $(\mathcal{P}_{5,\mathsf{eq}})$ as presented in Section IV-C.
17: **if** $\zeta \leq 0$ **then**
18: Solve $(\mathcal{P}_6)$ to get $\boldsymbol{p}, \boldsymbol{f}, \boldsymbol{f}^{\mathsf{Ed}}$ as presented in Section IV-C.
19: Solve $(\mathcal{P}_7)$ to get $\boldsymbol{\alpha}$ as presented in Section IV-D.
20: **if** $(\mathcal{P}_3)$, $(\mathcal{P}_5)$, $(\mathcal{P}_7)$ are feasible and $\xi^{\mathsf{Ed}} \leq \xi^{\mathsf{Ed,max}}$ **then**
21: Return $(\mathcal{P}_{\mathcal{B}})$ is feasible.
22: **end if**
23: **end if**
24: **Until** convergence or $(\mathcal{P}_{\mathcal{B}})$ is feasible.

bisection value is decreased, the updated problem is more likely to remain feasible, thereby accelerating convergence of the overall procedure.

Remark 1. *The formulated problem accounts for practical constraints in modern MEC systems, including energy efficiency, latency requirements, security, and limited wireless and computing resources. The proposed Algorithm 1, which integrates Algorithm 2 for feasibility checking in Step 5 and Algorithm 3 for task–edge server association in Step 19 of Algorithm 2, is specifically designed to address this realistic setting. Furthermore, its implementation is practically feasible within current MEC system architectures. Specifically, we assume a central controller [49] located at a coordinating edge server, with access to global network information such as channel state and user task parameters. This controller executes the proposed optimization algorithms and then send control decisions (i.e., offloading assignments, IRS configurations, encryption decisions, and resource allocation) to users and other system components. To reduce communication overhead, Algorithm 1 can be implemented as follows: 1) The controller sends a bisection search value to users. Each user then determines whether it belongs to set $\mathcal{A}$ or $\mathcal{B}$. It only transmits its CSI and task parameters to the controller if it belongs to set $\mathcal{B}$. The controller run Algorithm 2 and updates the bisection search value accordingly. Once Algorithm 1 converges, the the controller sends the final control decisions to the users in set $\mathcal{B}$. To reduce computational burden and latency, key algorithmic components such as the bisection search and successive convex approximation (SCA) are chosen for their scalability and convergence guarantees. Additionally, we evaluate the overall algorithmic complexity in Section IV-E, providing insight into the required number of operations and their dependency on system parameters.*

E. Complexity Analysis

We analyze the computational complexity of the proposed Algorithm 1 in terms of the required number of arithmetic operations. The while-loop for the bisection search of ξ requires $\log_2\left(\frac{\xi^{\max}-\xi^{\min}}{2}\right)$ iterations. The worst-case complexity of solving $(\mathcal{P}_{\mathcal{A}})$ is $\mathcal{O}(K)$. In the feasibility verification of problem $(\mathcal{P}_{\mathcal{B}})$, when determining the IRS's coefficients and the encryption decision, the key complexity lies in computing $\mathcal{G}_{i,j,k}, \forall i,j,k$; $\mathcal{F}_{i,j,k,e}, \forall i,j,k,e$; and $\boldsymbol{\chi}$, where each is respectively characterized by $\mathcal{O}\left(I^2K^2M_0^2\max(M_i)\right)$, $\mathcal{O}\left(I^2K^2E\max(M_i)\right)$, and $\mathcal{O}\left(\left(\sum_{i\in\mathcal{I}}M_i\right)^2KE\right)$. Let N_i be the number of iterations when solving $(\mathcal{P})_i$ where i ranges from 1 to 6. Consequently, the complexity for determining the IRS's coefficients and encryption decisions is expressed as $\mathcal{O}_1 = \mathcal{O}(N_2I^2K^3M_0^2\max(M_i) + N_2\left(\sum_{i\in\mathcal{I}}M_i\right)^2K + N_3I^2K^2\max(M_i)\left(M_0^2+E\right) + N_3\left(\sum_{i\in\mathcal{I}}M_i\right)^2KE)$. The determination of the detection matrix $\boldsymbol{W}$ incurs a complexity of $\mathcal{O}(M_0^3)$. The complexity associated with solving $(\mathcal{P}_{5,\mathsf{eq}})$, featuring $4K+1$ variables and $(6+E)K$ constraints, is denoted as $\mathcal{O}_5 = \mathcal{O}(N_5\sqrt{(6+E)K}((10+E)K+1)(4K+1)^2) = \mathcal{O}(N_5E^{3/2}K^{5/2})$. Correspondingly, the complexity of solving $(\mathcal{P}_6)$ is $\mathcal{O}(N_6E^{3/2}K^{5/2})$. The worst-case complexity of the matching algorithm to solve $(\mathcal{P}_7)$ is $\mathcal{O}(LK)$. Thus, the overall complexity of the proposed Algorithm 3 for the feasibility verification is $N_{\mathsf{B}}\left(\mathcal{O}_1 + \mathcal{O}(M_0^3) + (N_5+N_6)\mathcal{O}(E^{3/2}K^{5/2}) + \mathcal{O}(LK)\right)$, where N_{B} is the number of iteration of Algorithm 3. Consequently, the overall complexity of Algorithm 1 is $\log_2\left(\frac{\xi^{\max}-\xi^{\min}}{2}\right)N_{\mathsf{B}}(\mathcal{O}_1 + \mathcal{O}(M_0^3) + (N_5+N_6)\mathcal{O}(E^{3/2}K^{5/2}) + \mathcal{O}(LK))$.

V. Numerical results

In this section, we numerically evaluate the performance of the proposed framework. Consider K users, E eavesdroppers, and I IRSs uniformly distributed at random within a cell coverage area with radius of 800 meters. The BS is at the centered point of this area. The channel between user k and eavesdropper e is modeled as $h_{e,k}^{\mathsf{D,leak}} = u_{k,e}\sqrt{\vartheta_0 q_{k,e}^{-\alpha}}$ where $\alpha = 2.6$ if $q_{k,e} \leq D^{\mathsf{th}}$, otherwise $\alpha = 3.1$, where $\vartheta_0 = 10^{-3.24-2\log_{10}(f_{\mathsf{fre}})}$, $f_{\mathsf{fre}} = 5$ GHz, $u_{k,e} \sim \mathcal{CN}(0,1)$, and $q_{k,e}$ is the distance between user k and eavesdropper e. We also set $D^{\mathsf{th}} = 150$ meters. The channel between BS/IRS i (i.e., BS if $i=0$) and UE/ED k is $\boldsymbol{h}_{i,k}^{\mathsf{RIS}} = \boldsymbol{u}_{i,k}^{\mathsf{MIMO}}\sqrt{\vartheta_0 q_{i,k}^{-\alpha_i}}$, where $\boldsymbol{u}_{i,k}^{\mathsf{MIMO}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{M_i})$, $q_{i,k}$ is the distance between UE/ED k and the BS/IRS i, and $\alpha_i = 2.8$ if $i=0$ (i.e., between the BS and users), otherwise $\alpha_i = 2$ (i.e., between the IRS i and users). Note that $\boldsymbol{h}_{k,\mathsf{d}}^{\mathsf{u}} = \boldsymbol{h}_{0,k}$. The channel between BS and IRS i is expressed as $\boldsymbol{Q}_i^{\mathsf{R2B}} = \sqrt{\vartheta_0 q_i^{-\alpha_{\mathsf{IRS}}}}\boldsymbol{a}(\rho_{i,1}^{\mathsf{Rx}}, \rho_{i,2}^{\mathsf{Rx}})\boldsymbol{a}(\rho_{i,1}^{\mathsf{Tx}}, \rho_{i,2}^{\mathsf{Tx}})^H$, where q_i is the distance between IRS i and the BS, and $(\rho_{i,1}^{\mathsf{Rx}}, \rho_{i,2}^{\mathsf{Rx}})$ and $(\rho_{i,1}^{\mathsf{Tx}}, \rho_{i,2}^{\mathsf{Tx}})$ are the (azimuth, elevation) angles of arrival and departure corresponding, respectively.

In all experiments, we set the default value of the number of users $K = 10$, the number of eavesdroppers $E = 2$,

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3607599

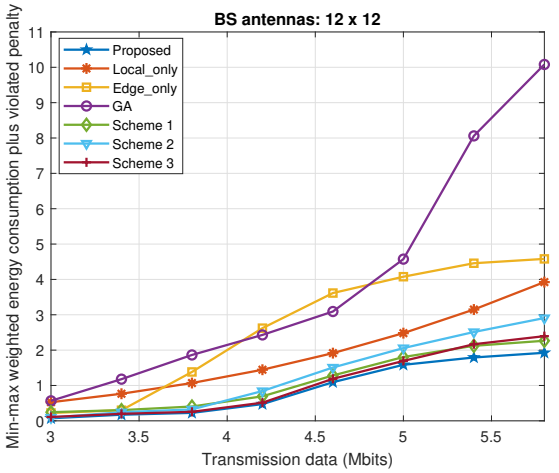

Fig. 4: The impact of the amount of data on the max-min weighted energy consumption.

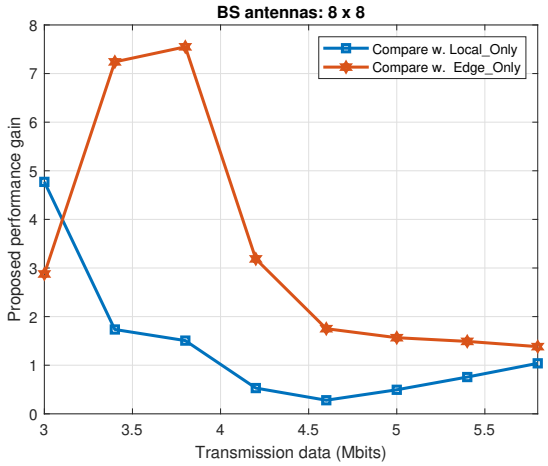

Fig. 5: The performance gain of the proposed algorithm compared to the 'Local_Only' and 'Edge_Only' strategies.

the number of IRSs $I = 3$, threshold $\epsilon^{\text{th}} = 0.1$, objective weight $\beta = 0.8 + 0.2\mathcal{U}(0,1)$ where $\mathcal{U}(0,1)$ is a random value following the uniform distribution, bandwidth $B = 1$ MHz, and the number of edge servers $L = 6$. The data size $b_k =$ band width $\times$ 3.5 bits, $c_k = (300 + 300\mathcal{U}(0,1))b_k$. The Advanced Encryption Standard (AES) algorithm is considered for data encryption. The encryption and decryption CPU cycles are set the same as $c_k^{\text{en}} = c_k^{\text{de}} = 50b_k$ [1, 41]. The user maximum transmit power $p_k^{\max} = 0.24$, and maximum delay $\eta_k^{\max} = 1$ second for all $k \in \mathcal{K}$. The default BS antennas are $M_0 = 10 \times 10$ and IRSs' antennas are $M_i = 10 \times 10, \forall i \in \mathcal{I}$. The noise power is set as noise power $= B \times 3.1812 \times 10^{-20}$.

Fig. 4 shows the system performance versus different amount of transmission data in Mbits. The violated penalty is computed as

$$\text{Pen} = \frac{1}{K}\left(\left[\frac{\xi^{\text{Ed}} - \xi^{\text{Ed,max}}}{\xi^{\text{Ed,max}}}\right]^+ + \sum_{l\in\mathcal{L}}\left[\frac{\sum_k \alpha_{k,l} f_k^{\text{Ed}} - F^{\max}}{F^{\max}}\right]^+ + U\right), \tag{53}$$

where $U = \sum_k \left[\frac{p_k - p_k^{\max}}{p_k^{\max}}\right]^+ + \sum_k \left[\frac{f_k - F_k^{\max}}{F_k^{\max}}\right]^+$. In Fig. 4, we compare the performance of three different strategies: 'Local_Only,' where all users' tasks are executed locally on their mobile devices; 'Edge_Only,' where all tasks are offloaded and executed on the edge server; 'Proposed,' which corresponds to the suggested strategy discussed in Sections III and IV. We also include 'GA' where a genetic algorithm is employed to solve problem $(\mathcal{P})$ [51, 52], 'Scheme 1' where the transmit power is fixed at $p_k^{\max}$, 'Scheme 2' where encryption is not applied, and 'Scheme 3' where the user-server association is selected randomly. It is noted that in 'Local_only', 'Edge_only', 'Scheme 1', 'Scheme 2', and 'Scheme 3', the remaining variables are optimized following the methods described in Sections III and IV. As depicted in Fig. 4, our proposed strategy outperforms both the 'Local_Only' and 'Edge_Only' approaches significantly. In Fig. 5, we examine the performance of 'Local_only' and 'Egde_only' scheme where the proposed performance gain is computed as

$$\text{Gain} = \frac{(\xi + \text{Pen})_{\text{local/edge}} - (\xi + \text{Pen})_{\text{Proposed}}}{(\xi + \text{Pen})_{\text{Proposed}}}. \tag{54}$$

As the transmission data increases, resulting in more extensive computational tasks, the computing resources available on the mobile devices become insufficient to meet the required delay, leading to an increase in violated penalties. The 'Edge_Only' strategy faces limitations due to the lack of wireless resources and favorable transmission conditions for all connections between the users and the BS. Therefore, an optimal offloading decision can yield remarkably high benefits in wireless communications scenarios with limited resources, as observed in Fig. 5. Moreover, compared to the 'GA' strategy for addressing this complex MINLP problem, our proposed algorithm achieves significantly better performance, further demonstrating its effectiveness. When compared to 'Scheme 1', the optimization of transmit power contributes to improved system performance. Similarly, comparing 'Scheme 2' with the proposed algorithm highlights the advantages of incorporating PLE to complement PLS in meeting the security requirements. Lastly, compared to 'Scheme 3', the use of appropriate user-server association further enhances overall system performance. Additionally, a performance gain of up to 7.5 times can be obtained by applying the proposed strategy compared to the 'Edge_Only' strategy.

Remark 2. *It is worth noting that the lowest-complexity scheme is 'Local_only', where the local computing resource is determined directly using a closed-form solution, and no further optimization of the remaining variables is required. However, when the latency-sensitive task becomes heavier and local computing resources are insufficient to meet the latency requirements, more advanced schemes are needed. The complexities of 'Edge_only', 'Scheme 1', 'Scheme 2', and 'Scheme 3' are provided in Section IV-E, where the complexity associated with determining the offloading decision, transmit power, encryption decision, and user-server association, respectively, is excluded. The complexity analysis of genetic algorithm can be found in [52, 53]. The GA parameter settings used in Fig. 4 are as follows: the population size is set to 200, the number of generations is 100, the mutation rate is 0.1, and the crossover rate is 0.6.*

In Fig. 6, we present the violated penalty for strategies

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3607599

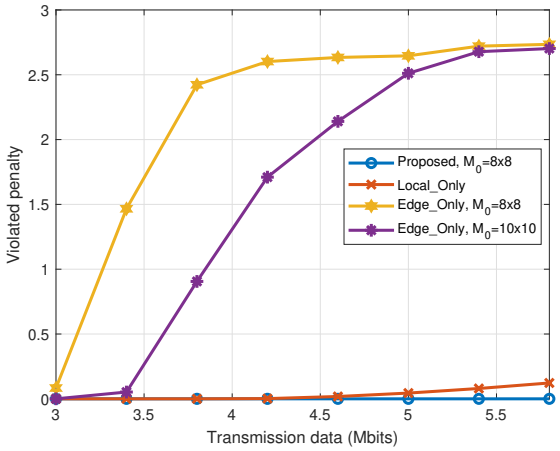

Fig. 6: The comparison of the violated penalty.

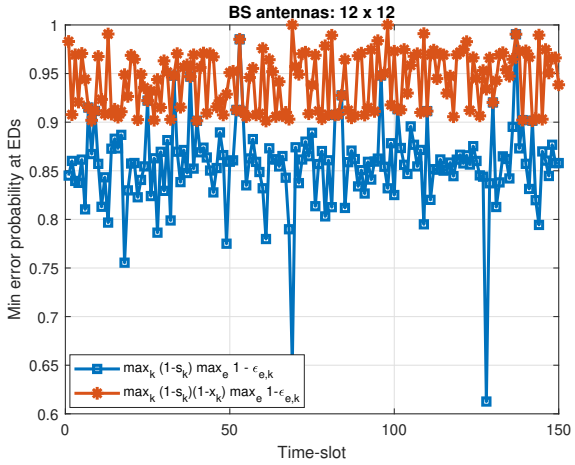

Fig. 8: The impact of the encryption decision on the error probability at eavesdroppers.

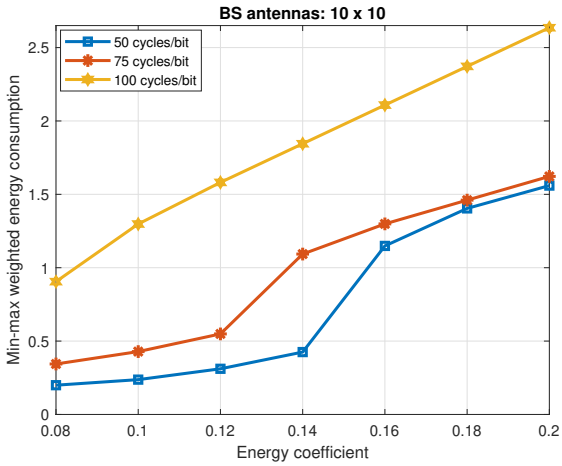

Fig. 7: The impact of the energy coefficient on the max-min weighted energy consumption.

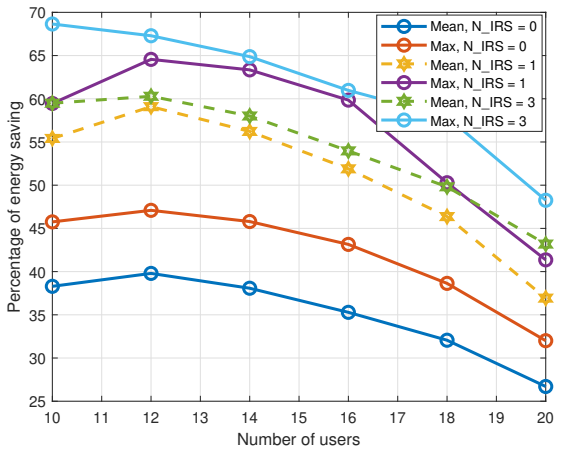

Fig. 9: The energy saving versus number of users.

employing various transmission data values and numbers of antennas (M_0). Our proposed algorithm ensures that all constraints are met by utilizing the available resources, resulting in a violated penalty of 0. However, the violated penalty of the 'Local_Only' strategy remains unaffected by the number of antennas in the BS. When the computational task is not intensive, it can be executed locally without violating the latency constraint. Nevertheless, as depicted in Fig. 4, this approach consumes more energy compared to a suitable offloading strategy. Conversely, the performance of the 'Edge_Only' strategy relies heavily on the transmission capacity to successfully offload all users' tasks to the edge servers. An increased number of antennas leads to higher channel gain, resulting in a smaller violated penalty.

Fig. 7 demonstrates the impact of the energy coefficient on the min-max energy consumption. The figure considers three different tasks with computation-to-transmission data ratios of the encryption and decryption algorithm at values of 50, 75, and 100 (i.e., $c_k^{\text{en}}/b_k = c_k^{\text{de}}/b_k = \{50, 75, 100\}$). It is observed that the energy consumption is susceptible to both the computation-to-transmission data ratios and the energy coefficients.

Fig. 8 presents the minimum BEP experienced by eavesdroppers for the data transmitted by all users across different time slots. Users without PLE exhibit a minimum BEP at eavesdroppers that exceeds the defined threshold of 0.9. In contrast, users benefiting from PLE protection demonstrate a significantly lower minimum error probability at eavesdroppers, highly below the threshold. If PLE protection is not applied to these users, the wireless resources must ensure an increase in the minimum BEP at eavesdroppers to reach the threshold for all users. However, this task may become impossible if certain users' channel cannot be improved or if the channel between these users and eavesdroppers are inherently favorable to eavesdropping.

Fig. 9 illustrates the energy savings achieved with varying configurations of users and IRSs. It can be observed from the figure that the optimal number of users for maximizing energy savings depends on the system parameter (i.e., the number of IRS) as well as the evaluation criteria (i.e., mean, max). Without using IRS, the transmission relies solely on direct communication, which typically suffers from higher path loss and interference, thus the energy saving is also limited. When the number of users is high, the allocation of wireless and computing resources at the edge server for supporting and executing users' offloaded tasks becomes limited. Consequently,

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3607599

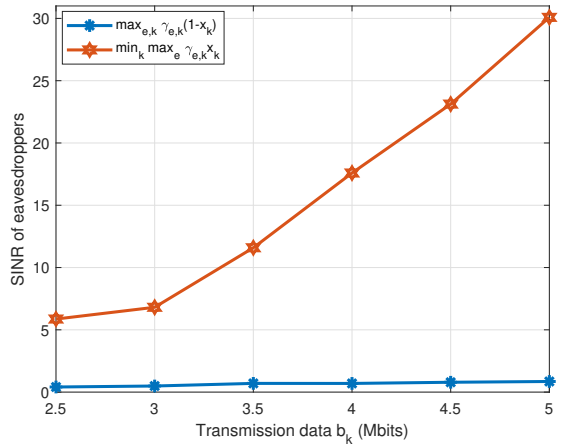

Fig. 10: The impact of PLE on the SINR of eavesdroppers.

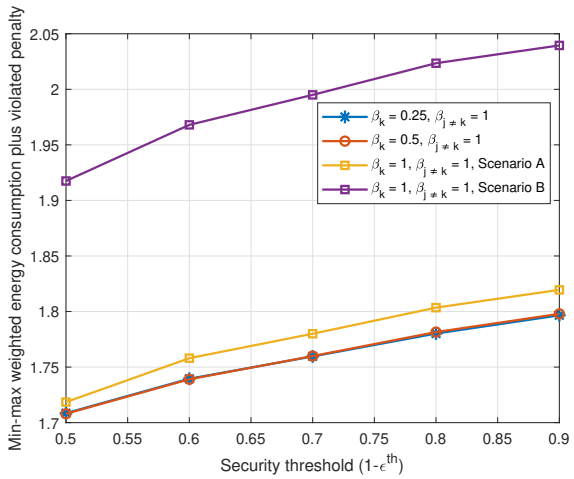

Fig. 11: The impact of the energy weights and security threshold.

the percentage of energy savings decreases correspondingly.

In Fig. 10, we compare the maximum SINR attained by eavesdroppers for signals from non-encrypted users with the minimum SINR achieved by eavesdroppers for signals from encrypted users. As the amount of transmission data increases, users protected by the PLS criterion continue to experience secure transmission thanks to the small SINR observed by eavesdroppers. Conversely, the PLE mechanism safeguards users whose signals, when intercepted by eavesdroppers, are of substantial strength. The escalation in SINR attained by eavesdroppers with the increase in required transmission data renders more users unprotected by the PLS criterion. Consequently, the encryption energy increases to ensure the protection of these users.

Fig. 11 illustrates the impact of energy weights and the security threshold (i.e., $1-\epsilon^{\text{th}}$) on the min-max weighted energy consumption, where the data size $b_k = 5$ Mbits. To examine the effect of energy weights, we assume that the weights for all users $j \neq k$ are set to 1 (i.e., $\beta_{j\neq k} = 1$), where k denotes the user with the highest energy consumption when executing tasks locally (i.e., the worst-case user). As shown in Fig. 11, when β_k is sufficiently small (e.g., $\beta_k = 0.25$ or $\beta_k = 0.5$) and $\beta_k \left(\xi_k^{\text{Lo}} + \xi_k^{\text{Tx}}\right) < \left(\xi_j^{\text{Lo}} + \xi_j^{\text{Tx}}\right)$, the min-max weighted energy consumption tends to prioritize minimizing the energy consumption of user j. Therefore, the energy consumption values for the curves labeled '$\beta_k = 0.25, \beta_{j\neq k} = 1$' and '$\beta_k = 0.5, \beta_{j\neq k} = 1$' are almost identical. However, when $\beta_k = 1$, the optimization problem seeks to minimize the energy consumption of user k. Since user k represents the worst-case scenario, this leads to an increase in the overall min-max weighted energy consumption. In Fig. 11, the curve labeled '$\beta_k = 1, \beta_{j\neq k} = 1$, Scenario A' shows the min-max weighted energy consumption when the control variables are optimized with $\beta_k = 1, \beta_{j\neq k} = 1$. In contrast, the curve labeled $\beta_k = 1, \beta_{j\neq k} = 1$, Scenario B' shows the result when the control variables are optimized using $\beta_k = 0.5, \beta_{j\neq k} = 1$, but evaluated under $\beta_k = 1, \beta_{j\neq k} = 1$. Overall, in Fig. 11, the weighted energy consumption approach helps reduce the energy usage of the worst-case user by approximately 10 %. Furthermore, as the security threshold increases, more users require PLE protection to meet the security requirements. Consequently, the overall energy consumption increases due to the additional computational overhead introduced by encryption algorithms.

VI. Conclusions

In this paper, we have proposed a novel criterion for the joint PLS and PLE based on the bit error probability attained by eavesdroppers. We have formulated and solved the joint offloading decision, encryption decision, IRS's coefficient, MIMO detection matrix, and computing and wireless resource allocation to minimize the maximum weighted energy consumption in a MEC system. Numerical results confirm that our proposed solution can meet the security requirement while minimizing the energy consumption from the users' perspective in a secure MEC network. Notably, the framework demonstrates high performance across scenarios, especially for data rates around 3.5 Mbits per second, typical of real-world applications like high-definition video streaming and augmented reality, and real-time sensor data processing in IoT systems, achieving up to 69% energy savings without compromising latency. In future work, we will investigate robust optimization techniques that explicitly account for CSI uncertainty to enhance the resilience of the proposed framework under more realistic conditions, particularly when perfect CSI is unavailable.

References

[1] T. T. Nguyen, L. B. Le, and Q. Le-Trung, "Computation offloading in MIMO based mobile edge computing systems under perfect and imperfect CSI estimation," *IEEE Trans. Services Computing*, vol. 14, no. 6, pp. 2011–2025, 2019.

[2] T. T. Nguyen, V. N. Ha, L. B. Le, and R. Schober, "Joint data compression and computation offloading in hierarchical fog-cloud systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 293–309, 2019.

[3] C. Pham, D. T. Nguyen, Y. Njah, N. H. Tran, K. K. Nguyen, and M. Cheriet, "Share-to-run IoT services in edge cloud computing," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 497–509, 2021.

[4] E. El Haber, M. Elhattab, C. Assi, S. Sharafeddine, and K. K. Nguyen, "Latency and reliability aware edge computation offloading via an intelligent reflecting surface," *IEEE Commun. Letters*, vol. 25, no. 12, pp. 3947–3951, 2021.

[5] T. M. Ho and K.-K. Nguyen, "Joint server selection, cooperative offloading and handover in multi-access edge computing wireless network: A

deep reinforcement learning approach," *IEEE Trans. Mobile Computing*, vol. 21, no. 7, pp. 2421–2435, 2020.

[6] M. Wang, S. Shi, D. Zhang, C. Wu, and Y. Wang, "Joint computation offloading and resource allocation for MIMO-NOMA assisted multi-user MEC systems," *IEEE Trans. Commun.*, 2023.

[7] P. Chen, B. Lyu, Y. Liu, H. Guo, and Z. Yang, "Multi-IRS assisted wireless-powered mobile edge computing for internet of things," *IEEE Trans. Green Commun. Network.*, vol. 7, no. 1, pp. 130–144, 2022.

[8] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and J. Wang, "Joint MU-MIMO precoding and resource allocation for mobile-edge computing," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1639–1654, 2020.

[9] X. Qi, M. Peng, and H. Zhang, "Joint mmWave beamforming and resource allocation in NOMA-MEC network for internet of things," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4969–4980, 2023.

[10] F. Naeem, M. Ali, G. Kaddoum, C. Huang, and C. Yuen, "Security and privacy for reconfigurable intelligent surface in 6G: A review of prospective applications and challenges," *IEEE Open J. Commun. Society*, vol. 4, pp. 1196–1217, 2023.

[11] R. Kaur, B. Bansal, S. Majhi, S. Jain, C. Huang, and C. Yuen, "A survey on reconfigurable intelligent surface for physical layer security of next-generation wireless communications," *IEEE open journal of vehicular technology*, vol. 5, pp. 172–199, 2024.

[12] A. Chorti, A. N. Barreto, S. Köpsell, M. Zoli, M. Chafii, P. Sehier, G. Fettweis, and H. V. Poor, "Context-aware security for 6G wireless: The role of physical layer security," *IEEE Commun. Standards Mag.*, vol. 6, no. 1, pp. 102–108, 2022.

[13] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-noise-aided secure MIMO wireless communications via intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7851–7866, 2020.

[14] X. He, R. Jin, and H. Dai, "Physical-layer assisted secure offloading in mobile-edge computing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4054–4066, 2020.

[15] Y. Liu, W. Wang, H.-H. Chen, F. Lyu, L. Wang, W. Meng, and X. Shen, "Physical layer security assisted computation offloading in intelligently connected vehicle networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3555–3570, 2021.

[16] Y. Chen, K. Zhao, H. Zhang, and S. He, "A comprehensive physical layer security mechanism for mobile edge computing," *IEEE Internet Things J.*, 2023.

[17] L. P. Qian, Y. Wu, N. Yu, D. Wang, F. Jiang, and W. Jia, "Energy-efficient multi-access mobile edge computing with secrecy provisioning," *IEEE Transactions on Mobile Computing*, vol. 22, no. 1, pp. 237–252, 2021.

[18] Z. Tang, T. Hou, Y. Liu, J. Zhang, and L. Hanzo, "Physical layer security of intelligent reflective surface aided NOMA networks," *IEEE Trans. Vehicular Technol.*, vol. 71, no. 7, pp. 7821–7834, 2022.

[19] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Letters*, vol. 8, no. 5, pp. 1410–1414, 2019.

[20] Y. Xiu, J. Zhao, C. Yuen, Z. Zhang, and G. Gui, "Secure beamforming for multiple intelligent reflecting surfaces aided mmwave systems," *IEEE Commun. Letters*, vol. 25, no. 2, pp. 417–421, 2020.

[21] Y. Gao, Z. Wang, Y. Zhang, W. Lu, J. Tang, N. Zhao, and F. Gao, "Multi-IRS-aided secure communication in UAV-MEC networks," *IEEE Trans. Veh. Technol.*, 2025.

[22] H. Zhang, Y. Huang, Z. Zhang, K. Guo, Z. Lin, and X. Lu, "RIS-assisted green and secure symbiotic UAV-MEC network," *IEEE Trans. Commun.*, 2025.

[23] W. Wang, W. Ni, H. Tian, and L. Song, "Intelligent omni-surface enhanced aerial secure offloading," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5007–5022, 2022.

[24] S. Mao, L. Liu, N. Zhang, M. Dong, J. Zhao, J. Wu, and V. C. Leung, "Reconfigurable intelligent surface-assisted secure mobile edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6647–6660, 2022.

[25] B. L. J. Liao, W. Wu, and Y. Li, "Intelligent reflecting surface assisted secure computation of wireless powered MEC system," *IEEE Trans. Mobile Computing*, 2023.

[26] B. Li, W. Wu, Y. Li, and W. Zhao, "Intelligent reflecting surface and artificial-noise-assisted secure transmission of MEC system," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 11 477–11 488, 2021.

[27] T. Sadig, M. Maleki, N. H. Tran, and H. R. Bahrami, "An encryption-aware PHY security framework for 4-node Gaussian wiretap channels with joint power constraint," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7837–7850, 2020.

[28] L. Xiao, S. Hong, S. Xu, H. Yang, and X. Ji, "IRS-aided energy-efficient secure WBAN transmission based on deep reinforcement learning," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4162–4174, 2022.

[29] D. Puthal, X. Wu, N. Surya, R. Ranjan, and J. Chen, "SEEN: A selective encryption method to ensure confidentiality for big sensing data streams," *IEEE Trans. Big Data*, vol. 5, no. 3, pp. 379–392, 2017.

[30] W. Wang, M. Hempel, D. Peng, H. Wang, H. Sharif, and H.-H. Chen, "On energy efficient encryption for video streaming in wireless sensor networks," *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 417–426, 2010.

[31] M. Mitev, A. Chorti, H. V. Poor, and G. Fettweis, "What physical layer security can do for 6G security," *IEEE Open J. Veh. Technol.*, 2023.

[32] D. Wang, P. Ren, J. Cheng, and Y. Wang, "Achieving full secrecy rate with energy-efficient transmission control," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5386–5400, 2017.

[33] Z. Li, M. Hua, Q. Wang, and Q. Song, "Weighted sum-rate maximization for multi-irs aided cooperative transmission," *IEEE Wireless Commun. Letters*, vol. 9, no. 10, pp. 1620–1624, 2020.

[34] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-irs-assisted full-duplex systems with hardware impairments," *IEEE Trans. Cognitive Commun. Network.*, vol. 7, no. 2, pp. 466–481, 2021.

[35] D. Gündüz, M. A. Wigger, T.-Y. Tung, P. Zhang, and Y. Xiao, "Joint source–channel coding: Fundamentals and recent progress in practical designs," *Proceedings of the IEEE*, 2024.

[36] C. E. Shannon, "A mathematical theory of communication," *The Bell System Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[37] Y. Liu, H.-H. Chen, and L. Wang, "Physical layer security for next generation wireless networks: Theories, technologies, and challenges," *IEEE Commun. Surveys & Tuts.*, vol. 19, no. 1, pp. 347–376, 2016.

[38] C.-H. Nguyen, Y. M. Saputra, D. T. Hoang, D. N. Nguyen, V.-D. Nguyen, Y. Xiao, and E. Dutkiewicz, "Encrypted data caching and learning framework for robust federated learning-based mobile edge computing," *IEEE/ACM Transactions on Networking*, vol. 32, no. 3, pp. 2705–2720, 2024.

[39] R. Venkataramanan, V. N. Swamy, and K. Ramchandran, "Low-complexity interactive algorithms for synchronization from deletions, insertions, and substitutions," *IEEE Trans. Informat. Theory*, vol. 61, no. 10, pp. 5670–5689, 2015.

[40] E. M. Abadi, C. Goken, C. Ozturk, and S. Gezici, "Optimal power allocation and optimal linear encoding for parameter estimation in the presence of a smart eavesdropper," *IEEE Trans. Signal Process.*, vol. 70, pp. 4093–4108, 2022.

[41] B. Fournier, V. V. T. Tong, and G. Guette, "Accurate measurement of the energy consumption of security functions." in *SECRYPT*, 2021, pp. 487–494.

[42] S. Posea, "Renewal periods for cryptographic keys," Ph.D. dissertation, Eindhoven University of Technology, 2012.

[43] L. Jin, X. Xu, S. Han, X. Chi, S. Lv, P. Zhang, and C. Liu, "Computation offloading outage probability analysis and min-max fairness optimization in RIS-assisted MEC system," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4615–4627, 2022.

[44] Z. Wan, W. Jiang, J. Nie, D. Niyato, C. Pan, and Z. Xiong, "Min-max fairness based joint optimal designfor IRS-assisted MEC systems," *IEEE Trans. Veh. Technol.*, 2024.

[45] L. You, J. Xiong, Y. Huang, D. W. K. Ng, C. Pan, W. Wang, and X. Gao, "Reconfigurable intelligent surfaces-assisted multiuser MIMO uplink transmission with partial CSI," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5613–5627, 2021.

[46] M. Sadek, A. Tarighat, and A. H. Sayed, "Active antenna selection in multiuser MIMO communications," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1498–1510, 2007.

[47] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, and M. Elkashlan, "Many-to-many matching with externalities for device-to-device communications," *IEEE Wireless Commun. letters*, vol. 6, no. 1, pp. 138–141, 2016.

[48] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel & Scientific Computations*, vol. 11, no. 4, pp. 351–368, 2003.

[49] 3GPP, "Edge computing," Jun. 2023, accessed: 2025-07-07. [Online]. Available: https://www.3gpp.org/technologies/edge-computing

[50] T. T. Nguyen and K.-K. Nguyen, "Channel estimation in IRS-assisted multi-cast multi-group communication systems," in *IEEE Int. Conf. Commun. (ICC)*. IEEE, 2023, pp. 1–6.

[51] M. Mitchell, *An introduction to genetic algorithms*. MIT press, 1998.

[52] Y. Rabinovich and A. Wigderson, "Techniques for bounding the convergence rate of genetic algorithms," *Random Structures & Algorithms*, vol. 14, no. 2, pp. 111–138, 1999.

[53] T. Bäck, J. N. Kok, and G. Rozenberg, *Handbook of natural computing*. Springer, Heidelberg, 2012.