Low-Latency and Secure Computation Offloading Assisted by Hybrid Relay-Reflecting Intelligent Surface

Khac-Hoang Ngo*, Nhan Thanh Nguyen[†], Thinh Quang Dinh[‡], Trong-Minh Hoang[§], and Markku Juntti[†]

*Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden
[†]Centre for Wireless Communications, University of Oulu, P.O.Box 4500, FI-90014, Finland
[‡]Center of Excellence for Wearables Research and Development, Fossil Group, Inc., Vietnam
[§]Posts and Telecommunications Institute of Technology, Hanoi, Viet Nam
Emails: ngok@chalmers.se, {nhan.nguyen, markku.juntti}@oulu.fi, thinhdinh@ieee.org, hoangtrongminh@ptit.edu.vn

Abstract—Recently, the hybrid relay-reflecting intelligent surface (HRRIS) has been introduced as a spectral- and energy-efficient architecture to assist wireless communication systems. In the HRRIS, a single or few active relay elements are deployed along with a large number of passive reflecting elements, allowing it to not only reflect but also amplify the incident signals. In this work, we investigate the potential of the HRRIS in aiding the computation offloading in a single-user mobile edge computing system. The objective is to minimize the offloading latency while ensuring the secrecy of user data against a malicious eavesdropper. We develop efficient solutions to this latency minimization problem based on alternating optimization. Through numerical results, we show that the deployment of the HRRIS can result in a considerable reduction in latency. Furthermore, the latency reduction gain offered by the HRRIS is much more significant than that of the conventional reconfigurable intelligent surface (RIS).

I. Introduction

Edge computing is a novel paradigm to develop communication and computation infrastructures for the Internet of Things. It overcomes the high latency and low bandwidth drawbacks of centralized cloud computing by extending the cloud's capacities to near-user network edges [1]. In a mobile edge computing (MEC) system, the mobile users offload their entire or partial computation tasks to proximate MEC servers via wireless links. Thus, computation offloading suffers from fading and attenuation in the wireless medium [2]. This may entail a higher latency than the local execution, and hinder the advantage of MEC, especially for users located far from edge nodes [3]. Therefore, efficient methods to improve the user-server communications is needed.

A new technology called reconfigurable intelligent surface (RIS) has emerged as a promising solution to enhance the wireless communication capacity by reflecting radio waves in preferred directions [4]–[7]. RIS is comprised of a controller and a large number of passive reflecting elements which enable passive beamforming without requiring any radio frequency (RF) chains [6], [8]. However, the main limitation of RIS is that the passive reflection limits the beamforming gains. Recently, a novel concept of hybrid relay-reflecting intelligent surface (HRRIS) has been proposed to overcome this limitation [9]–[11]. In the HRRIS, a few elements are equipped with power amplifiers to serve as active relays. It was shown that HRRIS, even with a single active element, significantly

improves both the system spectral and energy efficiency (SE/EE) with respect to RIS [9]–[11]. Due to its potential, HRRIS can be deployed to assist the computation offloading in MEC. RIS-assisted MEC system has been investigated recently in, e.g., [3], [12]. However, to our best knowledge, the applications of HRRIS for computation offloading have not been considered in the literature.

Besides latency minimization, another challenge in MEC computation offloading is to guarantee the secrecy of user data under the presence of malicious eavesdroppers in the system. Various techniques have been proposed to maximize the secrecy communication rate such as jamming with artificial noise and beamforming schemes [13]–[15]. However, when the legitimate communication channel is weaker than the eavesdropping channel, the achievable secrecy rate is significantly limited, even with the aforementioned techniques [16]. RIS has been shown to be an efficient solution to tackle this challenge [16]–[20]. However, with the powerful reflecting/relaying capabilities, HRRIS is expected to provide significant improvement in the secrecy rate compared with the conventional passive RIS.

In this paper, we aim at minimizing the latency in computation offloading while guaranteeing secrecy in a single-user HRRIS-aided MEC system. We assume that the user has a local computing resource and offloads a fraction of the tasks to an edge node (EN). A multi-antenna eavesdropper receives the user's signal and attempts to decode the data. To prevent that, the EN informs the user to transmit at a secrecy rate lower than the maximal achievable rate. For the latency minimization, we propose a joint design of the receive combining vector, the HRRIS's reflecting/relaying coefficients (communication parameters), and the computation offloading volume (computation parameter). To this end, we develop an alternating optimization approach that efficiently solves the challenging latency minimization problem. In particular, we consider both the fixed HRRIS and the dynamic HRRIS schemes. In the former, the positions and number of the active elements are unchanged; in contrast, those in the latter can be dynamically changed and optimized according to the channel condition. We numerically evaluate the latency of these schemes and compare it with that of the conventional RIS. Our simulation results show that the HRRIS consistently outperforms the fixed HRRIS, and both provide significant latency reduction with respect to RIS.
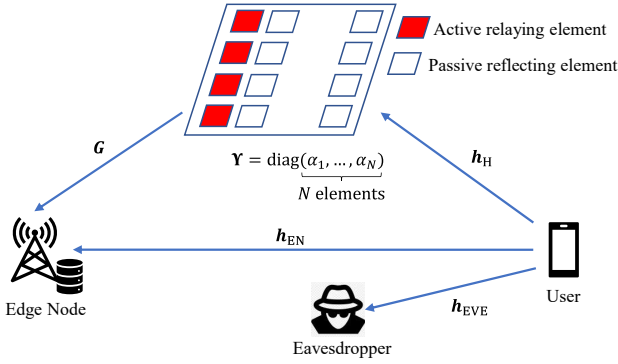
Fig. 1. The HRRIS-assisted communication and computation offloading system in the presence of an eavesdropper.

II. SYSTEM MODEL

We consider a single-user system in the presence of a malicious eavesdropper, as illustrated in Fig. 1. The user has computation tasks to partially offload to an EN. An HRRIS is deployed in the system to assist the offloading so that the computation result is returned to the user with low latency, while the user data is protected against the eavesdropper. The model is detailed in the following.

A. Communication Model

We assume that the user, the EN, and the eavesdropper are equipped with a single antenna, M antennas, and E antennas, respectively. The HRRIS has N elements, including A active relaying elements and $N - A$ passive reflecting elements. We note that a passive reflecting coefficient has the optimal amplitude of unity to maximize the received signal power [21]; therefore, a passive element only adjusts the incident signal's phase. In contrast, an active one can not only tune the phase but also amplify the signal power. We denote the index set of the positions of the A active elements by $\mathcal{A} \subset [N] \triangleq \{1, 2, \ldots, N\}$. Let $\alpha_n = |\alpha_n| e^{j\theta_n}$ denote the relay/reflection coefficient of the n-th element, where $|\alpha_n|$ and $\theta_n \in [\pi, 2\pi)$ represent the amplitude and phase shift, respectively. Note that $|\alpha_n| = 1, \forall n \notin \mathcal{A}$. Let $\boldsymbol{\alpha} \triangleq [\alpha_1 \ldots \alpha_N]^\mathsf{T}$. We denote $\boldsymbol{\Upsilon} \triangleq \operatorname{diag}(\alpha_1, \ldots, \alpha_N)$ and decompose $\boldsymbol{\Upsilon}$ as $\boldsymbol{\Upsilon} = \boldsymbol{\Phi} + \boldsymbol{\Psi}$ where $\boldsymbol{\Phi} = \operatorname{diag}(\phi_1, \ldots, \phi_N)$ and $\boldsymbol{\Psi} = \operatorname{diag}(\psi_1, \ldots, \psi_N)$ with $\phi_n = \alpha_n \mathbb{1}\{n \notin \mathcal{A}\}$ and $\psi_n = \alpha_n \mathbb{1}\{n \in \mathcal{A}\}$, where $\mathbb{1}\{\cdot\}$ is the indicator function. That is, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ contain the passive and active coefficients, respectively.

Let $\boldsymbol{h}_{\mathrm{EN}} \in \mathbb{C}^M$, $\boldsymbol{h}_{\mathrm{H}} \in \mathbb{C}^N$, and $\boldsymbol{h}_{\mathrm{EVE}} \in \mathbb{C}^E$ denote the channel vectors between the user and the EN, the HRRIS, and the eavesdropper, respectively. Let $\boldsymbol{G} \in \mathbb{C}^{M \times N}$ denote the channel matrix from the EN to the HRRIS. We assume that these channel coefficients are quasi-static, i.e., remain unchanged during the whole computation offloading circle. Furthermore, we assume that the eavesdropper knows $\boldsymbol{h}_{\mathrm{EVE}}$; the EN knows perfectly $\boldsymbol{h}_{\mathrm{EN}}$, $\boldsymbol{h}_{\mathrm{H}}$, and $\boldsymbol{G}$ but only knows an estimate $\hat{\boldsymbol{h}}_{\mathrm{EVE}}$ of $\boldsymbol{h}_{\mathrm{EVE}}$. Following the deterministic uncertainty model [22], [23], we assume that $\frac{\|\boldsymbol{h}_{\mathrm{EVE}} - \hat{\boldsymbol{h}}_{\mathrm{EVE}}\|}{\|\hat{\boldsymbol{h}}_{\mathrm{EVE}}\|} \leq \epsilon$, where ϵ is an upper bound of the relative estimation error.

Denote the transmit power and transmitted signal of the user by P and s, respectively. The incident signal at the HRRIS is

$\sqrt{P}\boldsymbol{h}_{\mathrm{H}}s + \boldsymbol{z}_{\mathrm{H}}$, where $\boldsymbol{z}_{\mathrm{H}} \sim \mathcal{CN}(0, \sigma_{\mathrm{H}}^2 \mathbf{I}_N)$ is the noise vector. The transmit power of the HRRIS's active elements is given as $P_{\mathrm{a}}(\boldsymbol{\alpha}) \triangleq \operatorname{tr}\left(\boldsymbol{\Psi}(P\boldsymbol{h}_{\mathrm{H}}\boldsymbol{h}_{\mathrm{H}}^\mathsf{H} + \sigma_{\mathrm{H}}^2 \mathbf{I}_N)\boldsymbol{\Psi}^\mathsf{H}\right)$. The received signal at the EN is given by

$$\boldsymbol{y} = \sqrt{P}(\boldsymbol{h}_{\mathrm{EN}} + \boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{h}_{\mathrm{H}})s_k + \boldsymbol{G}\boldsymbol{\Psi}\boldsymbol{z}_{\mathrm{H}} + \boldsymbol{z}_{\mathrm{EN}} \quad (1)$$
$$= \sqrt{P}\boldsymbol{h}s + \boldsymbol{z}, \quad (2)$$

where $\boldsymbol{h} \triangleq \boldsymbol{h}_{\mathrm{EN}} + \boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{h}_{\mathrm{H}}$ denotes the effective user-EN channel, $\boldsymbol{z}_{\mathrm{EN}} \sim \mathcal{CN}(0, \sigma_{\mathrm{EN}}^2 \mathbf{I}_M)$ is the additive noise at the EN, and $\boldsymbol{z} \triangleq \boldsymbol{G}\boldsymbol{\Psi}\boldsymbol{z}_{\mathrm{H}} + \boldsymbol{z}_{\mathrm{EN}}$ is the total effective noise. For notational simplicity, we assume that $\sigma_{\mathrm{H}}^2 = \sigma_{\mathrm{EN}}^2 = \sigma^2$, and thus $\boldsymbol{z} \sim \mathcal{CN}\left(0, \sigma^2 \boldsymbol{Q}\right)$ where $\boldsymbol{Q} \triangleq \mathbf{I}_M + \boldsymbol{G}\boldsymbol{\Psi}\boldsymbol{\Psi}^\mathsf{H}\boldsymbol{G}^\mathsf{H}$.

We assume that the EN employs a linear combining vector $\boldsymbol{w} \in \mathbb{C}^M$, and thus, the signal is recovered at the EN as $\hat{s} = \boldsymbol{w}^\mathsf{H}\boldsymbol{y}$. Then, the effective signal-to-interference-plus-noise ratio (SINR) is given by $\gamma(\boldsymbol{w}, \boldsymbol{\alpha}) = \frac{P|\boldsymbol{w}^\mathsf{H}\boldsymbol{h}|^2}{\sigma^2 \boldsymbol{w}^\mathsf{H}\boldsymbol{Q}\boldsymbol{w}}$. Therefore, the maximal achievable rate is given by

$$R(\boldsymbol{w}, \boldsymbol{\alpha}) = W \log_2[1 + \gamma(\boldsymbol{w}_k, \boldsymbol{\alpha})], \quad (3)$$

where W is the uplink bandwidth.

The eavesdropper also receives the user's signal and attempts to decode the data.[1] The rate that the eavesdropper can achieve from user k's signal, i.e., the leakage rate, is $R_{\mathrm{EVE}} = W \log_2\left(1 + \frac{P\|\boldsymbol{h}_{\mathrm{EVE}}\|^2}{\sigma_{\mathrm{EVE}}^2}\right)$, where σ_{EVE}^2 is the noise variance at the eavesdropper. The secrecy rate for user k is then given by $R(\boldsymbol{w}, \boldsymbol{\alpha}) - R_{\mathrm{EVE}}$. Since the EN only knows $\hat{\boldsymbol{h}}_{\mathrm{EVE}}$, it computes an upper bound $\bar{R}_{\mathrm{EVE}}$ on the leakage rate:

$$\bar{R}_{\mathrm{EVE}} \triangleq \log_2\left(1 + \frac{P(1+\epsilon)^2 \|\hat{\boldsymbol{h}}_{\mathrm{EVE}}\|^2}{\sigma_{\mathrm{EVE}}^2}\right). \quad (4)$$

To guarantee secrecy, the EN informs the user to transmit at a lower bound on the secrecy rate which is given by

$$R_{\mathrm{s}}(\boldsymbol{w}, \boldsymbol{\alpha}) = (R(\boldsymbol{w}, \boldsymbol{\alpha}) - \bar{R}_{\mathrm{EVE}})^+. \quad (5)$$

B. Computing Model

The user has some computational tasks and might offload a certain fraction or all of their tasks to the EN. Thus, s is the offloading signal. Let L, ℓ, and ν denote the total number of bits to be processed, the number of bits offloaded to the edge server, and the number of CPU cycles required to process one bit, respectively.

- *Local computing:* Denote the computational capability of the user by f^l CPU cycles/second. Then the time required for the local computation at user k is given by $D^l = \frac{(L-\ell)\nu}{f^l}$.
- *Edge computing:* We denote the computational capability of the MEC as f^e CPU cycles/second. Since the computation results are typically of small size, the delay due to feedback of these results to the user is assumed to be negligible. Therefore, the total latency imposed by the computation offloading and the edge computing is given by $D^e = \frac{\ell}{R_{\mathrm{s}}(\boldsymbol{w}, \boldsymbol{\alpha})} + \frac{\ell\nu}{f^e}$.

The overall latency is imposed by the maximum latency between the local and edge computing, that is,

$$D(\boldsymbol{w}, \boldsymbol{\alpha}, \ell) = \max\{D^l, D^e\}.$$

---

[1] We assume that the eavesdropper does not receive the signal from the HRRIS.

C. Problem Formulation

We aim at minimizing the offloading latency by jointly optimizing the computation offloading volume ℓ, the combining vector $\boldsymbol{w}$, and the HRRIS's coefficients $\boldsymbol{\alpha}$. In the fixed HRRIS architecture, the positions of the active elements, i.e., $\mathcal{A}$, are known. Thus, the latency minimization problem is given as

$$(\mathcal{P}_{\text{fixed}}) \quad \underset{\boldsymbol{w}, \boldsymbol{\alpha}, \ell}{\text{minimize}} \; D(\boldsymbol{w}, \boldsymbol{\alpha}, \ell) \tag{6a}$$

$$\text{subject to} \quad |\alpha_n| = 1, n \notin \mathcal{A}, \tag{6b}$$

$$P_{\text{a}}(\boldsymbol{\alpha}) \leq P_{\text{a}}^{\max}, \tag{6c}$$

$$\ell \in \{0, 1, \ldots, L\}, \tag{6d}$$

where $P_{\text{a}}^{\max}$ is the power budget of the HRRIS. Considering the dynamic HRRIS scheme, $\mathcal{A}$ is unknown and thus, it is cast as a design parameter in the latency minimization with the dynamic HRRIS, i.e.,

$$(\mathcal{P}_{\text{dynamic}}) \quad \underset{\boldsymbol{w}, \mathcal{A}, \boldsymbol{\alpha}, \ell}{\text{minimize}} \; D(\boldsymbol{w}, \boldsymbol{\alpha}, \ell) \tag{7a}$$

$$\text{subject to} \quad (6b), (6c), (6d),$$

$$|\mathcal{A}| \leq A, \mathcal{A} \subset [N]. \tag{7b}$$

In both problems $(\mathcal{P}_{\text{fixed}})$ and $(\mathcal{P}_{\text{dynamic}})$, $\boldsymbol{w}$ and $\boldsymbol{\alpha}$ are coupled. Furthermore, the objective function $D(\boldsymbol{w}, \boldsymbol{\alpha}, \ell)$ is segmented and nonconvex. Therefore, finding optimal solution is challenging. In the following, we develop efficient solutions to these optimization problems.

III. EFFICIENT ALTERNATING OPTIMIZATION SOLUTION

We employ the alternating optimization approach. Specifically, we will alternately solve for ℓ and $\{\boldsymbol{w}, \boldsymbol{\alpha}\}$ while fixing the other, as presented in the following subsections.

A. Optimization of ℓ

Given $\{\boldsymbol{w}, \boldsymbol{\alpha}\}$, ℓ is the solution to $\underset{l \in \{0, 1, \ldots, L\}}{\text{minimize}} \max(D^l, D^e)$. Notice that D^l is monotonically decreases while D^e monotonically increases with ℓ. Therefore, $\max(D^l, D^e)$ is minimized when $D^l = D^e$, which holds for

$$\ell = \hat{\ell}^{\star} = \frac{L \nu R_{\text{s}}(\boldsymbol{w}, \boldsymbol{\alpha}) f^e}{f^e f^l + \nu R_{\text{s}}(\boldsymbol{w}, \boldsymbol{\alpha})(f^e + f^l)}. \tag{8}$$

We integerize ℓ and obtain the optimal offloading volume

$$\ell^{\star} = \arg \min_{\hat{\ell} \in \{\lfloor \hat{\ell}^{\star} \rfloor, \lceil \hat{\ell}^{\star} \rceil\}} D(\boldsymbol{w}, \boldsymbol{\alpha}, \hat{\ell}). \tag{9}$$

From (8), we see that whenever the secrecy rate $R_{\text{s}}(\boldsymbol{w}, \boldsymbol{\alpha})$ is zero, we have $\ell^{\star} = 0$. That is to say, if secure communication cannot be guaranteed, the user should not offload and execute the computation tasks locally instead.

B. Joint Optimization of $\boldsymbol{w}$ and $\boldsymbol{\alpha}$

We next jointly optimize $\boldsymbol{w}$ and $\boldsymbol{\alpha}$ given that $\ell = \ell^{\star}$. We have seen that with $\ell = \ell^{\star}$, it holds that $D^l \approx D^e$. Therefore, we can replace the objective function $D(\boldsymbol{w}, \boldsymbol{\alpha}, l)$ by D^e. In doing so, the optimization of $\boldsymbol{w}$ and $\boldsymbol{\alpha}$ is equivalent to the maximization of the secrecy rate

$$(\mathcal{P}_{\text{secrecy}}) \quad \underset{\boldsymbol{w}, \boldsymbol{\alpha}}{\text{maximize}} \; R_{\text{s}}(\boldsymbol{w}_k, \boldsymbol{\alpha}), \tag{10}$$

where it is implicit in this subsection that the constraints are (6b), (6c) for fixed HRRIS and (6b), (6c), (7b) for dynamic HRRIS. Recall that $R_{\text{s}}(\boldsymbol{w}, \boldsymbol{\alpha}) = (R(\boldsymbol{w}, \boldsymbol{\alpha}) - \bar{R}_{\text{EVE}})^+$. The leakage rate $\bar{R}_{\text{EVE}}$ is independent of $(\boldsymbol{w}, \boldsymbol{\alpha})$, and it is expected that for the optimal solution, $R(\boldsymbol{w}, \boldsymbol{\alpha}) > \bar{R}_{\text{EVE}}$. Therefore, we replace $R_{\text{s}}(\boldsymbol{w}, \boldsymbol{\alpha})$ by $R(\boldsymbol{w}_k, \boldsymbol{\alpha})$ in $(\mathcal{P}_{\text{secrecy}})$, which leads to the maximization of the SINR, i.e.,

$$(\mathcal{P}_{\text{SINR}}) \; \underset{\boldsymbol{w}, \boldsymbol{\alpha}}{\text{maximize}} \left(\gamma(\boldsymbol{w}, \boldsymbol{\alpha}) = \frac{P|\boldsymbol{w}^{\mathsf{H}}(\boldsymbol{h}_{\text{EN}} + \boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{h}_{\text{H}})|^2}{\sigma^2 \boldsymbol{w}^{\mathsf{H}}(\mathbf{I}_M + \boldsymbol{G}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathsf{H}}\boldsymbol{G}^{\mathsf{H}})\boldsymbol{w}} \right).$$

1) Solution to $\boldsymbol{w}$: Given $\boldsymbol{\alpha}$, the optimal $\boldsymbol{w}$ maximizing the objective function in $(\mathcal{P}_{\text{SINR}})$ can be found after some simple manipulations as

$$\boldsymbol{w}^{\star} = \sqrt{\frac{P}{\sigma^2}} (\mathbf{I}_M + \boldsymbol{G}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathsf{H}}\boldsymbol{G}^{\mathsf{H}})^{-1} (\boldsymbol{h}_{\text{EN}} + \boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{h}_{\text{H}}). \tag{11}$$

2) Solution to $\boldsymbol{\alpha}$: In the case where there is no active element, i.e., $|\alpha_i| = 1, \forall i \in [N]$, the phases of $\boldsymbol{\alpha}$ can be optimized as in [3, Sec. IV]. Specifically, the SINR becomes

$$\frac{P|\boldsymbol{w}^{\mathsf{H}}(\boldsymbol{h}_{\text{EN}} + \boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{h}_{\text{H}})|^2}{\sigma^2 \|\boldsymbol{w}\|^2}$$

$$\leq \frac{P}{\sigma^2 \|\boldsymbol{w}\|^2} \left(|\boldsymbol{w}^{\mathsf{H}}\boldsymbol{h}_{\text{EN}}| + \sum_{n=1}^{N} |h_{\text{H},n}\boldsymbol{w}^{\mathsf{H}}\boldsymbol{g}_n| \right)^2, \tag{12}$$

where $h_{\text{H},n}$ is the n-th entry of $\boldsymbol{h}_{\text{H}}$ and $\boldsymbol{g}_n$ is the n-th column of $\boldsymbol{G}$. The equality in (12) occurs if $\arg\{\boldsymbol{w}^{\mathsf{H}}\boldsymbol{h}_{\text{EN}}\} = \arg\{\boldsymbol{w}^{\mathsf{H}}\boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{h}_{\text{H}}\}$. Accordingly, the phases of the HRRIS's coefficients can be obtained as

$$\boldsymbol{\theta}^{\star} = \arg\{\boldsymbol{w}^{\mathsf{H}}\boldsymbol{h}_{\text{EN}}\} - \arg\{\text{diag}\{\boldsymbol{w}^{\mathsf{H}}\boldsymbol{G}\}\boldsymbol{h}_{\text{H}}\}, \tag{13}$$

where $\arg\{\cdot\}$ returns the (element-wise) phases of a complex number or vector. When there is at least one active element, i.e., $A > 0$, both the phases and the amplitudes of $\{\alpha_i\}_{i \in \mathcal{A}}$ need to be optimized. We propose to first optimize the phases as in (13), then optimize the amplitudes as follows.

a) Fixed HRRIS: The phases obtained in (13) result in $|\boldsymbol{w}^{\mathsf{H}}(\boldsymbol{h}_{\text{EN}} + \boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{h}_{\text{H}})|^2 = \left(|\boldsymbol{w}^{\mathsf{H}}\boldsymbol{h}_{\text{EN}}| + \sum_{n=1}^{N} |\alpha_n h_{\text{H},n}\boldsymbol{w}^{\mathsf{H}}\boldsymbol{g}_n| \right)^2$. Thus, the SINR can be expanded as

$$\frac{\gamma(\boldsymbol{w}, \boldsymbol{\alpha})}{P/\sigma^2} = \frac{|\alpha_n|^2 a_n + |\alpha_n| b_n + c_n}{|\alpha_n|^2 u_n + v_n}, \tag{14}$$

for any $n \in \mathcal{A}$, where

$$a_n \triangleq |h_{\text{H},n}|^2 |\boldsymbol{w}^{\mathsf{H}}\boldsymbol{g}_n|^2, \tag{15}$$

$$c_n \triangleq \left(|\boldsymbol{w}^{\mathsf{H}}\boldsymbol{h}_{\text{EN}}| + \sum_{i=1, i \neq n}^{N} |\alpha_i h_{\text{H},i}\boldsymbol{w}^{\mathsf{H}}\boldsymbol{g}_i| \right)^2, \tag{16}$$

$$b_n \triangleq 2|h_{\text{H},n}\boldsymbol{w}^{\mathsf{H}}\boldsymbol{g}_n|\sqrt{c_n}, \tag{17}$$

$$u_n \triangleq |\boldsymbol{w}^{\mathsf{H}}\boldsymbol{g}_n|^2, \tag{18}$$

$$v_n \triangleq \|\boldsymbol{w}\|^2 + \sum_{i \in \mathcal{A}, i \neq n} |\alpha_i|^2 |\boldsymbol{w}^{\mathsf{H}}\boldsymbol{g}_i|^2. \tag{19}$$

Constraint (6c) on the amplitudes is equivalent to

$$P_{\text{a}}(\boldsymbol{\alpha}) = \sum_{n \in \mathcal{A}} |\alpha_n|^2 \xi_n \leq P_{\text{a}}^{\max} \tag{20}$$

where $\xi_n \triangleq \sigma^2 + P|h_{\text{H},n}|^2$. Let $\tilde{P}_{\text{a},n} \triangleq \sum_{i \in \mathcal{A}, i \neq n} |\alpha_i|^2 \xi_i$, $n \in \mathcal{A}$, which is a constant if $\{|\alpha_i|\}_{i \in \mathcal{A}, n \neq n}$ are fixed. The

3

constraint (20) can be written as $|\alpha_n| \leq \sqrt{\frac{P_{\text{a}}^{\max} - \tilde{P}_{\text{a},n}}{\xi_n}}$. We alternatively optimize the amplitude of each active element while keeping the others fixed by solving

$$\underset{|\alpha_n|}{\text{maximize}} \quad \frac{|\alpha_n|^2 a_n + |\alpha_n| b_n + c_n}{|\alpha_n|^2 u_n + v_n} \tag{21a}$$

$$\text{subject to} \quad |\alpha_n| \leq \sqrt{\frac{P_{\text{a}}^{\max} - \tilde{P}_{\text{a},n}}{\xi_n}}. \tag{21b}$$

The solution to (21) is given in closed form as

$$|\alpha_n|^\star = \min \left\{ \sqrt{\frac{P_{\text{a}}^{\max} - \tilde{P}_{\text{a},n}}{\xi_n}}, \frac{d_n}{b_n} + \sqrt{\frac{d_n^2}{b_n^2} + \frac{v_n}{u_n}} \right\}, \tag{22}$$

where $d_n \triangleq |h_{\text{H},n}|^2 v_n - c_n$. We note that $|\alpha_n|^\star$ is not necessarily the maximal amplitude constraint $\sqrt{\frac{P_{\text{a}}^{\max} - \tilde{P}_{\text{a},n}}{\xi_n}}$.

b) Dynamic HRRIS: For dynamic HRRIS, the set of active elements $\mathcal{A}$ subject to (7b) is also a design parameter. Determining both $\mathcal{A}$ and $\{|\alpha_i|\}_{i \in \mathcal{A}}$ is challenging since the active elements are involved in both the numerator and denominator of the SINR. Notice that

$$\gamma(\boldsymbol{w}, \boldsymbol{\alpha}) \leq \bar{\gamma}(\boldsymbol{w}, \boldsymbol{\alpha}) \triangleq \frac{P\left(|\boldsymbol{w}^{\text{H}}\boldsymbol{h}_{\text{EN}}| + \sum_{n=1}^{N} |\alpha_n h_{\text{H},n} \boldsymbol{w}^{\text{H}} \boldsymbol{g}_n|\right)^2}{\sigma^2 \|\boldsymbol{w}\|^2}.$$

$$= \frac{P(|\alpha_n|^2 a_n + |\alpha_n| b_n + c_n)}{\sigma^2 \|\boldsymbol{w}\|^2} \tag{23}$$

for any $n \in \mathcal{A}$. The equality occurs when $\mathcal{A} = \emptyset$. The smaller the term $\sigma^2 \boldsymbol{w}^{\text{H}} \boldsymbol{G} \boldsymbol{\Psi} \boldsymbol{\Psi}^{\text{H}} \boldsymbol{G}^{\text{H}} \boldsymbol{w}$ is, the tighter the bound is. Observe that $\sigma^2 \boldsymbol{w}^{\text{H}} \boldsymbol{G} \boldsymbol{\Psi} \boldsymbol{\Psi}^{\text{H}} \boldsymbol{G}^{\text{H}} \boldsymbol{w}$ is small when the noise variance is small, the path loss is large, and/or the number of active elements is small. This motivates us to simplify the optimization to maximizing $\bar{\gamma}(\boldsymbol{w}, \boldsymbol{\alpha})$.

Observe that an active element only improves the performance with respect to a passive one if $|\alpha_n| > 1$ since it causes performance loss due to signal attenuation otherwise. Thus, in the dynamic HRRIS, we have $|\alpha_n| > 1, \forall n \in \mathcal{A}$. Therefore, we ignore the $o(|\alpha_n|^2)$ terms, i.e., the terms involves $|\alpha_n|$ and the constant, in $\bar{\gamma}(\boldsymbol{w}, \boldsymbol{\alpha})$ and focus on the second-order term. In doing so, we arrive at maximizing the terms $Pa_n|\alpha_n|^2$. We consider the following problem

$$\underset{\mathcal{A}, \{|\alpha_n|\}_{n \in \mathcal{A}}}{\text{maximize}} \quad \sum_{n \in \mathcal{A}} \log(1 + Pa_n|\alpha_n|^2), \tag{24a}$$

$$\text{subject to} \quad \sum_{n \in \mathcal{A}} |\alpha_n|^2 \xi_n \leq P_{\text{a}}^{\max} \text{ and (7b).} \tag{24b}$$

Let $p_n \triangleq |\alpha_n|^2 \xi_n$ denote the power allocated to the n-th active element, $n \in \mathcal{A}$. The optimization (24) can be rewritten as

$$\underset{\mathcal{A}, \{p_n\}_{n \in \mathcal{A}}}{\text{maximize}} \quad \sum_{n \in \mathcal{A}} \log\left(1 + \frac{Pa_n}{\xi_n} p_n\right), \tag{25a}$$

$$\text{subject to} \quad \sum_{n \in \mathcal{A}} p_n \leq P_{\text{a}}^{\max} \text{ and (7b).} \tag{25b}$$

From (25), we determine the optimal set of active elements $\mathcal{A}^\star$ as the set of indices of A largest values in $\left\{\frac{Pa_1}{\xi_1}, \ldots, \frac{Pa_N}{\xi_N}\right\}$. Given $\mathcal{A}^\star$, the optimal solution for $\{p_n\}$ is the well-known water-filling solution $p_n^\star = \left(\frac{1}{\mu} - \frac{\xi_n}{Pa_n}\right)^+$ with μ satisfying

---

Algorithm 1 Latency minimization for HRRIS-aided secure MEC.

Require: $\boldsymbol{h}_{\text{EVE}}, \boldsymbol{h}_{\text{EN}}, \boldsymbol{h}_{\text{H}},$ and $\boldsymbol{G}$.
Ensure: $\{\ell^\star, \boldsymbol{w}^\star, \boldsymbol{\alpha}^\star\}$.
1: Randomly generate $\boldsymbol{\alpha}$ satisfying constraints (6b) and (6c).
2: **while** objective value has not converged **do**
3:     Compute $\boldsymbol{w}^\star$ based on (11).
4:     Compute $\boldsymbol{\theta}^\star = [\theta_1^\star \ldots \theta_N^\star]$ based on (13).
5:     **if** fixed HRRIS is deployed **then**
6:         Obtain $\{|\alpha_n^\star|\}_{n \in \mathcal{A}}$ based on (22). Set $|\alpha_n^\star| = 1, \forall n \in [N] \backslash \mathcal{A}$.
7:     **else if** dynamic HRRIS is deployed **then**
8:         Obtain $\{|\alpha_n^\star|\}_{n \in [N]}$ based on (26).
9:     **end if**
10:    Set $\alpha_n^\star = |\alpha_n^\star| e^{j\theta_n^\star}, n \in [N]$.
11: **end while**
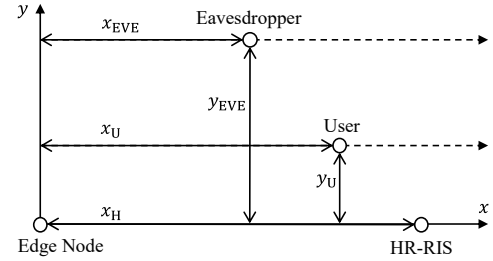12: Compute $\ell^\star$ based on (9).

Fig. 2. Locations of the EN, user, HRRIS, and eavesdropper.

$\sum_{n \in \mathcal{A}} p_n \leq P_{\text{a}}^{\max}$. It follows that the amplitudes of the HRRIS's elements are given by

$$|\alpha_n^\star| = \begin{cases} \max\left\{\sqrt{\frac{1}{\mu \xi_n} - \frac{1}{Pa_n}}, 1\right\}, & \text{for } n \in \mathcal{A}^\star, \\ 1, & \text{otherwise.} \end{cases} \tag{26}$$

C. Overall Joint Computing and Communication Design

The proposed joint computing and communication design scheme is summarized in Algorithm 1. We at first randomly generate the HRRIS coefficients satisfying constraints (6b) and (6c). Then, in steps 2–11, the solutions to combining vector $\boldsymbol{w}^\star$ and HRRIS coefficients $\boldsymbol{\alpha}^\star$ are alternatively updated. More specifically, $\boldsymbol{w}^\star$ is obtained based on (11), whereas, the HRRIS coefficients have the phases derived based on (13) and the amplitudes derived in (22) or (26) for the fixed or dynamic HRRIS schemes, respectively. Finally, when the update process terminated, i.e., when the convergence is reached, the optimal number of offloaded bits, i.e., $\ell^\star$, is obtained in step 12.

IV. NUMERICAL RESULTS

In this section, we numerically evaluate the performance of the proposed HRRIS-assisted computation offloading schemes. We assume that uniform linear arrays (ULAs) are deployed at the EN, user, and eavesdropper, while a uniform planar array (UPA) of N elements is deployed at the HRRIS. We assume half-wavelength distancing between array elements at all nodes. The locations of the nodes are illustrated in a two-dimensional coordinate in Fig. 2. The EN, HRRIS, user, and eavesdropper are located at $(0,0)$, $(x_{\text{H}}, 0)$, $(x_{\text{H}}, y_{\text{H}})$, and $(x_{\text{EVE}}, y_{\text{EVE}})$, respectively. From these coordinates, the distance between the nodes can be easily computed using the Pythagorean theorem.

4

For the large-scale fading, the path loss of a link distance d is given by $\beta(d) = \beta_0 \left(\frac{d}{1\text{m}}\right)^{-\eta}$, where β_0 is the path loss at the reference distance of 1 meter (m), and η is the path loss exponent. For small-scale fading, we assume the Rician fading channel model. Thus the small-scale fading channel between the HRRIS and the EN is modeled as $\tilde{\boldsymbol{G}} = \sqrt{\frac{\kappa}{1+\kappa}}\tilde{\boldsymbol{G}}^{\text{LoS}} + \sqrt{\frac{1}{1+\kappa}}\tilde{\boldsymbol{G}}^{\text{NLoS}}$, where $\tilde{\boldsymbol{G}}^{\text{LoS}}$ and $\tilde{\boldsymbol{G}}^{\text{NLoS}}$ represent the line-of-sight (LoS) and non-LoS (NLoS) components, respectively, and κ is the Rician factor. The NLoS component is modeled by the Rayleigh fading with independent entries following $\mathcal{CN}(0,1)$. The LoS component is given by the product of the array response vectors which are computed from the azimuth and elevated angle-of-departure (AoD) and angle-of-arrival (AoA), as detailed in [11, Sec. VI]. Finally, the channel between the HRRIS and the EN is obtained as $\boldsymbol{G} = \sqrt{\beta(x_{\text{H}})}\tilde{\boldsymbol{G}}$. The channels of other links are modeled similarly.

For comparison, we consider the following schemes: 1) conventional passive RIS with random phase, 2) conventional passive RIS with optimized phase according to (13), 3) optimized fixed HRRIS, and 4) optimized dynamic HRRIS. Note that the power consumption of the HRRIS schemes arises also from the active elements. For a fair comparison, we fix the same total power $P_{\text{tot}} = P + P_{\text{a}}^{\max}$ for all schemes. For reference, we also show the local computing latency $\frac{L\nu}{f^l}$.

In the following, unless otherwise stated, we set $M = 5$, $E = 1$, $N = 50$, $A = 1$, $P_{\text{tot}} = 30$ dBm, $P_{\text{a}}^{\max} = 0$ dBm, $\sigma^2 = \sigma_{\text{EVE}}^2 = -80$ dBm, $\{x_{\text{H}}, x_{\text{U}}, y_{\text{U}}, x_{\text{EVE}}, y_{\text{EVE}}\} = \{50, 45, 2, 30, 9\}$ m, $\beta_0 = -30$ dB, $W = 1$ MHz, and $\epsilon = 0.1$. The path loss exponents of the user-EN, user-HRRIS, user-eavesdropper, and HRRIS-EN links are given by 3.5, 2.2, 2.8, and 2.2, respectively. The Rician factors of these links are set to $0, 1, 0$, and 100, respectively. The computing parameters are given as $L = 300$ Kb, $\nu = 750$ CPU cycles/bit, $f^l = 5 \times 10^8$ cycles/s, and $f^e = 20 \times 10^9$ cycles/s.

First, in Fig. 3, we investigate the latency achieved with the aforementioned schemes for different locations of the user. We vary x_{U} in $[10, 100]$ m, i.e., let the user move along the x-axis in Fig. 2. It is observed that the latency of all schemes is close to that of local computing when the user is closest to the eavesdropper. This confirms that when secure communication is not guaranteed, the user should only execute the computation tasks with local computing resources rather than exposing the data while offloading. On the other hand, when the user comes close to the EN or to the HRRIS, the latency induced by RIS with optimized phase or by HRRIS is minimized. This is because the direct link user-EN or the user-HRRIS link is strong, allowing the users to transmit at a high secrecy rate and offload most/all of the tasks to the computationally powerful EN. RIS with random phase only slightly reduces the latency even when the user is close to the RIS. The dynamic HRRIS outperforms the fixed HRRIS, and both significantly reduce the latency with respect to RIS, especially when the user is far from the EN.

Next, in Fig. 4, we plot the latency as a function of N and A. In Fig. 4(a), we observe again that RIS with random phase, even with a large number of elements, only slightly improves upon local computing. The latency induced by RIS
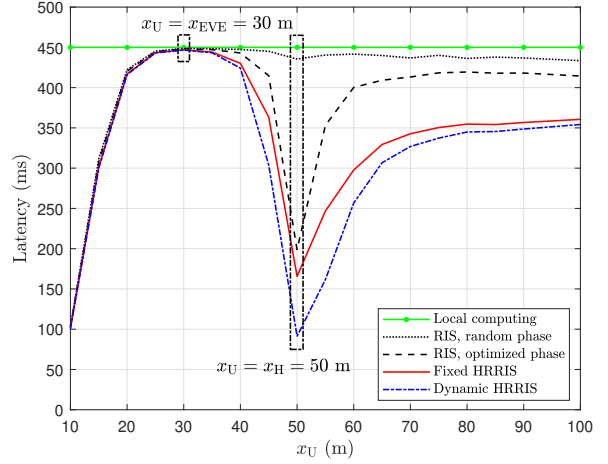

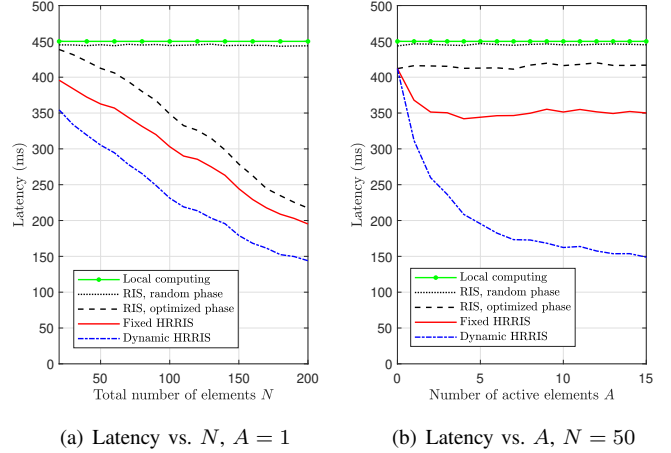Fig. 3. Latency vs. location of the user.

(a) Latency vs. N, $A = 1$ (b) Latency vs. A, $N = 50$

Fig. 4. Latency vs. total number of RIS/HRRIS elements N and the number of active elements A.

with optimized phase and by HRRIS decreases as N increases. Interestingly, while the gain of the fixed HRRIS with respect to RIS with optimized phase shrinks for large N, the dynamic HRRIS can maintain a high gain. Fig. 4(b) shows that the latency of both the fixed and dynamic HRRIS saturate as the number of active elements A increases, but the floor for the dynamic HRRIS is much lower than that for the fixed HRRIS.

Furthermore, we investigate the latency improvement of the HRRIS when the edge/local computing capability varies in Fig. 5. We still observe that the HRRIS achieves the lowest latency. In Fig. 5(a), the latency is reduced drastically when the edge computing capability f^e increases from a small value, whereas only a minor reduction in the latency can be seen when f^e is large. This has been observed for the RIS in [3]. The reason is that the latency imposed by edge computing dominates when f^e is small, whereas the latency imposed by communication dominates when f^e is large. In Fig. 5(b), we see that the latency of all schemes decreases proportional to $\frac{1}{f^l}$. The relative gain of HRRIS remains as f^l increases.

Finally, Fig. 6 depicts the latency as a function of the number of eavesdropper's antennas E for two values of $P_{\text{a}}^{\max}$, namely, 0 and 10 dBm. Naturally, the latency of all schemes increases as the eavesdropper becomes more capable. When $P_{\text{a}}^{\max} =$
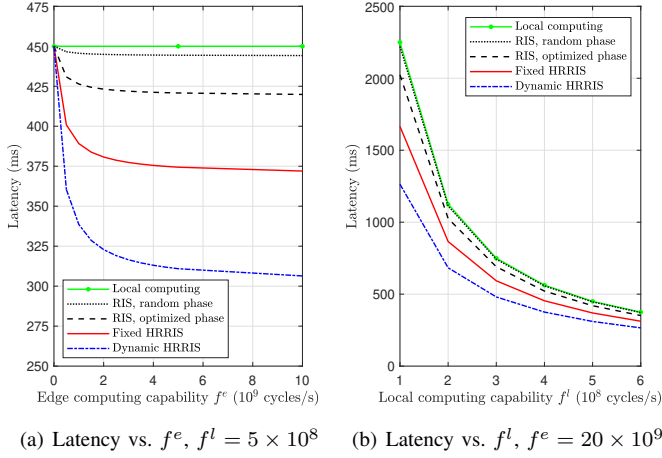
5

(a) Latency vs. f^e, $f^l = 5 \times 10^8$ (b) Latency vs. f^l, $f^e = 20 \times 10^9$

Fig. 5. Latency vs. edge computing capability f^e and local computing capability f^l. The EN has $M = 5$ antennas.

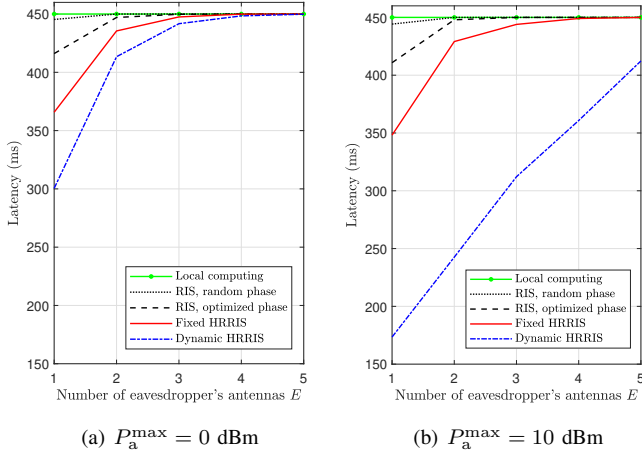(a) $P_{\mathrm{a}}^{\max} = 0$ dBm (b) $P_{\mathrm{a}}^{\max} = 10$ dBm

Fig. 6. Latency vs. number of eavesdropper's antennas E.

0 dBm, none of the schemes can improve the latency upon local computing if the eavesdropper has a comparable number of antennas as the EN. The situation remains similar for RIS and fixed HRRIS if $P_{\mathrm{a}}^{\max}$ goes up to 10 dBm. However, the dynamic HRRIS can effectively exploit the extra power to reduce significantly the latency, achieving approximately 9 % gain with respect to other schemes for $P_{\mathrm{a}}^{\max} = 10$ dBm when the eavesdropper has as many antennas as the EN.

V. CONCLUSION

In this paper, we investigate a single-user MEC system assisted by an HRRIS. We aim at minimizing the user's offloading latency while ensuring sucure transmission against a malicious eavesdropper. To this end, we develop a novel alternating optimization method by sequentially optimizing the surface's configuration, the receive combining vector, and the computation offloading volume. By exploring the structure of HRRIS, our numerical results show that introducing as few as one active relaying element in the surface can reduce significantly the latency in comparison with conventional RIS, especially if the position of the active element can be dynamically changed. Possible future extensions include investigating a multi-user system where resource allocation between the users also plays a key role.

REFERENCES

[1] Q. D. La, M. V. Ngo, T. Q. Dinh, T. Q. Quek, and H. Shin, "Enabling intelligence in fog computing to achieve energy and latency reduction," *Digital Communications and Networks*, vol. 5, no. 1, pp. 3–9, Sep. 2019, Artificial Intelligence for Future Wireless Communications and Networking.

[2] C. Jiang, X. Cheng, H. Gao, X. Zhou, and J. Wan, "Toward computation offloading in edge computing: A survey," *IEEE Access*, vol. 7, pp. 131 543–131 558, Aug. 2019.

[3] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Topics Signal Process.*, vol. 38, no. 11, pp. 2666–2682, Nov. 2020.

[4] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[5] J. He, N. T. Nguyen, R. Schroeder, V. Tapio, J. Kokkoniemi, and M. Juntti, "Channel estimation and hybrid architectures for RIS-assisted communications," *arXiv preprint arXiv:2104.07115*, 2021.

[6] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, Jan. 2021.

[7] N. T. Nguyen, L. V. Nguyen, T. Huynh-The, D. H. Nguyen, A. L. Swindlehurst, and M. Juntti, "Machine learning-based reconfigurable intelligent surface-aided MIMO systems," *arXiv preprint arXiv:2105.00347*, 2021.

[8] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Commun. Surveys Tuts.*, May 2021, early access.

[9] N. T. Nguyen *et al.*, "Hybrid relay-reflecting intelligent surface-aided wireless communications: Opportunities, challenges, and future perspectives," *arXiv preprint arXiv:2104.02039*, 2021.

[10] N. T. Nguyen, Q.-D. Vu, K. Lee, and M. Juntti, "Spectral efficiency optimization for hybrid relay-reflecting intelligent surface," in *Proc. IEEE Int. Conf. Commun. Workshop*, 2021.

[11] ——, "Hybrid Relay-Reflecting Intelligent Surface-Assisted Wireless Communication," *arXiv preprint arXiv:2103.03900*, 2021.

[12] X. Hu, C. Masouros, and K.-K. Wong, "Reconfigurable intelligent surface aided mobile edge computing: From optimization-based to location-only learning-based solutions," *IEEE Trans. Commun.*, Mar. 2021.

[13] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.

[14] L. Liu, R. Zhang, and K.-C. Chua, "Secrecy wireless information and power transfer with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1850–1863, Apr. 2014.

[15] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1550–1573, Feb. 2014.

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

[17] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82 599–82 612, Jun. 2019.

[18] 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, Dec. 2020.

[19] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[20] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, Sep. 2019.

[21] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Commun. Conf.*, Dec. 2018, pp. 1–6.

[22] M. R. A. Khandaker, C. Masouros, and K.-K. Wong, "Constructive interference based secure precoding: A new dimension in physical layer security," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2256–2268, Sep. 2018.

[23] 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, Jun. 2020.