Rate-Splitting Multiple Access for Intelligent Reflecting Surface-Aided Secure Transmission

Ying Gao, Qingqing Wu, Wen Chen, *Senior Member, IEEE*, and Derrick Wing Kwan Ng, *Fellow, IEEE*

***Abstract*— In this letter, we study a rate-splitting multiple access (RSMA)-based intelligent reflecting surface (IRS)-aided multi-user multiple-input single-output (MISO) secure communication system with a potential eavesdropper (Eve). Aiming to maximize the minimum secrecy rate (SR) among all the legitimate users (LUs), a design problem for jointly optimizing the transmit beamforming with artificial noise (AN), the IRS beamforming, and the secrecy common rate allocation is formulated. Since the design problem is highly non-convex with coupled optimization variables, we develop a computationally efficient algorithm based on the alternating optimization (AO) technique to solve it suboptimally. Numerical results demonstrate that the proposed design can significantly improve the max-min SR over the benchmark schemes without IRS or adopting other multiple access techniques. In particular, employing the RSMA strategy can substantially reduce the required numbers of IRS elements for achieving a target level of secrecy performance compared with the benchmark schemes.**

***Index Terms*— Rate-splitting multiple access, intelligent reflecting surface, physical layer security, max-min fairness.**

I. Introduction

DUE to the capability of reconfiguring the wireless propagation environment, intelligent reflect surface (IRS) has recently drawn considerable research attention [1], [2], [3]. In particular, IRS exhibits great potential in enhancing the physical layer security (PLS) for wireless networks. By intelligently adapting the IRS elements' phase shifts, the reflected and direct signals can be superimposed constructively at legitimate users (LUs) while destructively at potential eavesdroppers (Eves) [4]. Initial studies on IRS-aided secure wireless communications, e.g., [5], [6], considered a simple setting with only one LU. Notably, the simulation results in [6] confirmed the benefit of exploiting artificial noise (AN). For a more general case of multi-antenna wireless networks with multiple LUs, two major transmission schemes, i.e., multi-user linear precoding (MU-LP) and power-domain non-orthogonal multiple access (NOMA), have been thoroughly investigated in previous works (see e.g., [7], [8], [9], [10]). On the other hand, rate-splitting multiple access (RSMA), bridging conventional MU-LP and NOMA, has been recently advocated as a promising transmission strategy to suppress multi-user interference [11]. Rate-splitting (RS) relies on the split of user messages into common and private parts, superimposed transmission from the transmitter, and successive interference cancellation (SIC) at the receivers. Particularly, the common message can not only provide extra degrees of freedom but also act as AN to enhance the quality of security provisioning [12]. Existing research, e.g., [11], [12], [13] and references therein, has demonstrated that RSMA outperforms MU-LP and NOMA in several aspects such as energy efficiency (EE), robustness, user fairness, PLS, etc.

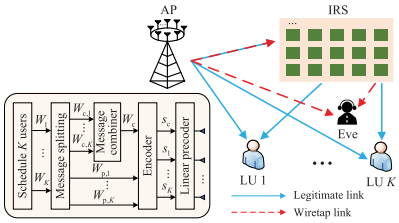

Fig. 1. Illustration of IRS-aided secure transmission with 1-layer RS.

To reap the advantages of both IRS and RSMA, some recent works have studied the amalgamation of them under different criteria, e.g., EE maximization [14], max-min fairness [15], and outage performance [16]. However, to the best of the authors' knowledge, there is no study on integrating RSMA with IRS for security performance enhancement. With the consideration of PLS, the corresponding problem formulation and resource allocation design for IRS-aided RSMA are quite different from those in e.g., [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. This is not only because RSMA allows to better manage the inter-user interference by applying the principle of RS, but also because the common message can serve as AN to reduce the potential information leakage to Eves. These factors complicate the interference management in the design of IRS's reflection, too. Therefore, the resource allocation designs proposed in e.g., [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] cannot be directly applied to IRS-aided RSMA. Besides, from the perspective of RSMA, it is still unknown whether the IRS can further enhance the PLS for RSMA-based systems. From the perspective of IRS, it remains unclear whether RSMA can evidently reduce the required surface size of IRS when a certain secrecy performance level is desired, which is of great practical interests for the implementation of IRSs.

Motivated by the above discussions, this letter investigates a RSMA-based secure downlink communication system consisting of an IRS, a multiple-antenna access point (AP), multiple single-antenna LUs, and a single-antenna Eve, as depicted in Fig. 1. As an initial study, we consider the low-complexity 1-layer RS strategy, where a single common stream is shared by all the LUs and a single layer of

Manuscript received 6 October 2022; revised 28 October 2022; accepted 21 November 2022. Date of publication 24 November 2022; date of current version 13 February 2023. The work was supported by SRG2020-00024-IOTSC, National key project 2020YFB1807700 and 2018YFB1801102, Shanghai Kewei 20JC1416502 and 22JC1404000, Pudong PKX2021-D02, NSFC 62071296, and the Australian Research Council's Discovery Project (DP210102169). The associate editor coordinating the review of this letter and approving it for publication was L. Bariah. *(Corresponding author: Qingqing Wu.)*

Ying Gao is with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau, China (e-mail: yinggao@um.edu.mo).

Qingqing Wu and Wen Chen are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 201210, China (e-mail: wu.qq1010@gmail.com; wenchen@sjtu.edu.cn).

Derrick Wing Kwan Ng is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Kensington, NSW 2052, Australia (e-mail: w.k.ng@unsw.edu.au).

Digital Object Identifier 10.1109/LCOMM.2022.3224499

SIC is required at each LU [11]. A minimum secrecy rate (SR) maximization problem is formulated, where the transmit precoder and the AN covariance matrix at the AP, the IRS phase shifts, and the secrecy common rate allocation are jointly optimized. Due to the non-smoothness and non-convexity of the formulated problem, we first propose an equivalent transformation to the problem at hand to arrive a tractable reformulation. Then, we propose a computationally efficient alternating optimization (AO) algorithm to obtain a suboptimal solution by decomposing the resulting problem into two subproblems that are alternatingly solved until convergence is achieved. Simulation results show that remarkable security performance enhancement can be achieved by our proposed IRS-aided RSMA transmission strategy as compared to the cases without IRS or adopting MU-LP and NOMA schemes. Moreover, RSMA is more attractive to space-constrained environments as it needs fewer IRS elements than MU-LP and NOMA for achieving a given performance level.

Notations: $\|\cdot\|_*$ and $\|\cdot\|_2$ represent the sum of the singular values and the maximum singular value of a matrix, respectively. $\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{y})$ denotes the partial gradient of function $f(\mathbf{x},\mathbf{y})$ with respect to (w.r.t.) vector $\mathbf{x}$.

II. System Model and Problem Formulation

As illustrated in Fig. 1, an IRS equipped with N reflecting elements is deployed in a secure communication system, where an M-antenna AP serves K single-antenna LUs in the presence of a single-antenna Eve. Let $\mathcal{N} \triangleq \{1,\cdots,N\}$ and $\mathcal{K} \triangleq \{1,\cdots,K\}$ be the sets of reflecting elements and LUs, respectively. Denoted by $\mathbf{\Theta} = \text{diag}\left(e^{\jmath\theta_1},\cdots,e^{\jmath\theta_N}\right)$ the reflection-coefficient matrix at the IRS, with $\theta_n \in [0,2\pi)$ representing the phase shift introduced by the n-th IRS element. Let $\mathbf{G} \in \mathbb{C}^{N\times M}$, $\mathbf{h}_{d,k}^H \in \mathbb{C}^{1\times M}$, $\mathbf{h}_{d,\text{e}}^H \in \mathbb{C}^{1\times M}$, $\mathbf{h}_{r,k}^H \in \mathbb{C}^{1\times N}$, and $\mathbf{h}_{r,\text{e}}^H \in \mathbb{C}^{1\times N}$ denote the channel coefficients from the AP to the IRS, from the AP to LU k, from the AP to Eve, from the IRS to LU k, and from the IRS to Eve, respectively. Then, we denote the cascaded AP-IRS-LU k and AP-IRS-Eve channels as $\mathbf{Q}_k = \text{diag}\big(\mathbf{h}_{r,k}^H\big)\mathbf{G}$ and $\mathbf{Q}_\text{e} = \text{diag}\big(\mathbf{h}_{r,\text{e}}^H\big)\mathbf{G}$, respectively.

The 1-layer RS strategy [11] mentioned in Section I is implemented at the AP. In addition, AN is transmitted simultaneously with the data streams to combat the potential eavesdropping by Eve. The transmitted signal is given by $\boldsymbol{x} = \mathbf{w}_\text{c} s_\text{c} + \sum_{k\in\mathcal{K}} \mathbf{w}_k s_k + \mathbf{z}$, where $s_\text{c} \sim \mathcal{CN}(0,1)$ and $s_k \sim \mathcal{CN}(0,1)$ denote the common stream for all the LUs and the private stream for LU k only, respectively, which are statistically uncorrelated and precoded by the precoding vectors $\mathbf{w}_\text{c} \in \mathbb{C}^{M\times 1}$ and $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$. Besides, $\mathbf{z} \sim \mathcal{CN}(\mathbf{0},\mathbf{Z})$ is the AN vector, with $\mathbf{Z} \in \mathbb{H}^M, \mathbf{Z} \succeq \mathbf{0}$ being its covariance matrix. We assume that the AN is unknown to both the LUs and Eve. Then, the received signal at LU k or Eve can be written as $y_j = \big(\mathbf{h}_{d,j}^H + \mathbf{h}_{r,j}^H \mathbf{\Theta}\mathbf{G}\big)\boldsymbol{x} + n_j = \mathbf{v}^H \mathbf{H}_j \boldsymbol{x} + n_j$, $j \in \mathcal{K}\cup\{\text{e}\}$, where $\mathbf{v} = [\mathbf{u};1]$ with $\mathbf{u} = \left[e^{\jmath\theta_1},\cdots,e^{\jmath\theta_N}\right]^H$, $\mathbf{H}_j = \left[\mathbf{Q}_j;\mathbf{h}_{d,j}^H\right]$, and n_j is the zero-mean additive white Gaussian noise (AWGN) with variance σ_j^2, respectively.

Following the 1-layer RS decoding order [11], the achievable rate of decoding the common stream s_c at LU k or Eve in bits/second/Hertz (bps/Hz) can be expressed as $R_{\text{c},j} = \log_2(1+\gamma_{\text{c},j})$, where $\gamma_{\text{c},j} = \left|\mathbf{v}^H\mathbf{H}_j\mathbf{w}_\text{c}\right|^2 \Big/ \Big(\sum_{i\in\mathcal{K}}\left|\mathbf{v}^H\mathbf{H}_j\mathbf{w}_i\right|^2 + \text{tr}\left(\mathbf{H}_j^H\mathbf{v}\mathbf{v}^H\mathbf{H}_j\mathbf{Z}\right) + \sigma_j^2\Big)$, $j \in \mathcal{K}\cup\{\text{e}\}$. Particularly, the transmission rate of s_c shall not exceed $R_\text{c} = \min_{k\in\mathcal{K}}\{R_{\text{c},k}\}$ so that all the LUs can successfully decode s_c. After decoding s_c, LU k removes it from the received signal via SIC and then decodes the private stream s_k. Hence, the achievable rate of decoding s_k at LU k in bps/Hz is given by $R_{\text{p},k} = \log_2(1+\gamma_{\text{p},k})$, with $\gamma_{\text{p},k} = \left|\mathbf{v}^H\mathbf{H}_k\mathbf{w}_k\right|^2 \Big/ \Big(\sum_{i\in\mathcal{K}\setminus\{k\}}\left|\mathbf{v}^H\mathbf{H}_k\mathbf{w}_i\right|^2 + \text{tr}\left(\mathbf{H}_k^H\mathbf{v}\mathbf{v}^H\mathbf{H}_k\mathbf{Z}\right) + \sigma_k^2\Big)$. On the other hand, to enable the common message to act as AN for degrading the signal-to-interference-plus-noise ratio (SINR) of each private message at Eve, we need to prevent s_c from being decoded by Eve. For this purpose, the condition $R_{\text{c,e}} < R_\text{c}$ should be satisfied if $\mathbf{w}_\text{c} \neq \mathbf{0}$. As such, the achievable rate of decoding s_k at Eve in bps/Hz can be expressed as $R_{\text{p,e}\to k} = \log_2(1+\gamma_{\text{p,e}\to k})$, where $\gamma_{\text{p,e}\to k} = \left|\mathbf{v}^H\mathbf{H}_\text{e}\mathbf{w}_k\right|^2 \Big/ \Big(\left|\mathbf{v}^H\mathbf{H}_\text{e}\mathbf{w}_\text{c}\right|^2 + \sum_{i\in\mathcal{K}\setminus\{k\}}\left|\mathbf{v}^H\mathbf{H}_\text{e}\mathbf{w}_i\right|^2 + \text{tr}\left(\mathbf{H}_\text{e}^H\mathbf{v}\mathbf{v}^H\mathbf{H}_\text{e}\mathbf{Z}\right) + \sigma_\text{e}^2\Big)$. Consequently, the achievable SR of LU k in bps/Hz is given by $R_k^\text{sec} = r_{\text{c},k}^\text{sec} + [R_{\text{p},k} - R_{\text{p,e}\to k}]^+$, where $[x]^+ = \max(x,0)$ and $r_{\text{c},k}^\text{sec}$ denotes the non-negative secrecy common rate allocated to LU k. In addition, the non-negative optimization variables $\{r_{\text{c},k}^\text{sec}\}$ need to satisfy the condition $\sum_{k\in\mathcal{K}} r_{\text{c},k}^\text{sec} \le R_\text{c} - R_{\text{c,e}}$, which also implies that $R_{\text{c,e}} \le R_\text{c}$.

Targeting at maximizing the minimum SR among all the LUs, we formulate the joint design of the transmit precoder and the AN covariance matrix at the AP, the phase shifts at the IRS, and the secrecy common rate allocated to each LU as follows:

$$
\begin{aligned}
(\text{P1}):\ \max_{\mathbf{w},\mathbf{Z}\in\mathbb{H}^M,\mathbf{v},\mathbf{r}_\text{c}^\text{sec}}\ & \min_{k\in\mathcal{K}}\ \left\{r_{\text{c},k}^\text{sec} + [R_{\text{p},k} - R_{\text{p,e}\to k}]^+\right\} && \text{(1a)}\\
\text{s.t.}\ & \sum\nolimits_{k\in\mathcal{K}} r_{\text{c},k}^\text{sec} \le R_\text{c} - R_{\text{c,e}}, && \text{(1b)}\\
& \|\mathbf{w}_\text{c}\|^2 + \sum\nolimits_{k\in\mathcal{K}}\|\mathbf{w}_k\|^2 + \text{tr}\left(\mathbf{Z}\right) \le P_\text{max}, && \text{(1c)}\\
& |[\mathbf{v}]_n| = 1,\quad \forall n\in\mathcal{N},\ [\mathbf{v}]_{N+1} = 1, && \text{(1d)}\\
& \mathbf{Z}\succeq\mathbf{0},\ r_{\text{c},k}^\text{sec} \ge 0,\quad \forall k\in\mathcal{K}, && \text{(1e)}
\end{aligned}
$$

where $\mathbf{w} \triangleq \{\mathbf{w}_\text{c},\mathbf{w}_1,\cdots,\mathbf{w}_K\}$ and $\mathbf{r}_\text{c}^\text{sec} \triangleq \{r_{\text{c},1}^\text{sec},\cdots,r_{\text{c},K}^\text{sec}\}$. Furthermore, P_max in (1c) represents the maximum transmit power at the AP, (1d) ensures the unit-modulus constraints on the phase shifts, and (1e) imposes the semidefinite and non-negativity constraints on $\mathbf{Z}$ and $\{r_{\text{c},k}^\text{sec}\}$, respectively. Note that (P1) is an intractable non-smooth and non-convex problem because of the non-smoothness introduced by the operator $[\cdot]^+$, the coupled optimization variables in (1a) and (1b), and the non-convex unit-modulus constraints in (1d). Hence, it is challenging, if not impossible, to solve (P1) optimally.

III. Solution to Problem (P1)

To facilitate the solution design, we define $\mathbf{W}_l = \mathbf{w}_l\mathbf{w}_l^H$, $\forall l \in \mathcal{L} \triangleq \{\text{c}\}\cup\mathcal{K}$ and $\mathbf{V} = \mathbf{v}\mathbf{v}^H$, where $\mathbf{W}_l \succeq \mathbf{0}$, $\text{rank}\left(\mathbf{W}_l\right) \le 1$, $\forall l\in\mathcal{L}$, $\mathbf{V}\succeq\mathbf{0}$, and $\text{rank}\left(\mathbf{V}\right)\le 1$. Besides, we introduce auxiliary variables t and $\mathbf{r}_\text{p}^\text{sec} \triangleq \{r_{\text{p},1}^\text{sec},\cdots,r_{\text{p},K}^\text{sec}\}$ and then equivalently convert (P1) into the following form:

$$
(\text{P2}):\ \max_{\mathcal{A}}\ t \qquad \text{(2a)}
$$

$$\text{s.t. } r_{\mathrm{c},k}^{\mathrm{sec}}+r_{\mathrm{p},k}^{\mathrm{sec}}\geq t,\quad \forall k\in\mathcal{K}, \tag{2b}$$

$$(f_{\mathrm{p},k}-g_{\mathrm{p},k})-(f_{\mathrm{e}}-g_{\mathrm{p,e}\to k})\geq r_{\mathrm{p},k}^{\mathrm{sec}},\quad \forall k\in\mathcal{K}, \tag{2c}$$

$$\sum\nolimits_{k\in\mathcal{K}} r_{\mathrm{c},k}^{\mathrm{sec}}\leq (f_{\mathrm{c},k}-g_{\mathrm{c},k})-(f_{\mathrm{e}}-g_{\mathrm{c,e}}),\quad \forall k\in\mathcal{K}, \tag{2d}$$

$$\sum\nolimits_{l\in\mathcal{L}} \mathrm{tr}\left(\mathbf{W}_l\right)+\mathrm{tr}\left(\mathbf{Z}\right)\leq P_{\max}, \tag{2e}$$

$$[\mathbf{V}]_{n,n}=1,\quad \forall n\in\mathcal{N}\cup\{N+1\}, \tag{2f}$$

$$\mathbf{Z}\succeq\mathbf{0},\mathbf{W}_l\succeq\mathbf{0},\mathrm{rank}\left(\mathbf{W}_l\right)\leq 1,\quad \forall l\in\mathcal{L}, \tag{2g}$$

$$\mathbf{V}\succeq\mathbf{0},\mathrm{rank}\left(\mathbf{V}\right)\leq 1, \tag{2h}$$

$$r_{\mathrm{c},k}^{\mathrm{sec}}\geq 0, r_{\mathrm{p},k}^{\mathrm{sec}}\geq 0,\quad \forall k\in\mathcal{K}, \tag{2i}$$

where $\mathcal{A}\triangleq\{\{\mathbf{W}_l\in\mathbb{H}^M\},\mathbf{Z}\in\mathbb{H}^M,\mathbf{V}\in\mathbb{H}^{N+1},\mathbf{r}_{\mathrm{c}}^{\mathrm{sec}},\mathbf{r}_{\mathrm{p}}^{\mathrm{sec}},t\}$, $f_{\mathrm{p},k}\triangleq\log_2\left(\sum_{i\in\mathcal{K}}\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{W}_i\right)+\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{Z}\right)+\sigma_k^2\right)$, $g_{\mathrm{p},k}\triangleq\log_2\left(\sum_{i\in\mathcal{K}\setminus\{k\}}\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{W}_i\right)+\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{Z}\right)+\sigma_k^2\right)$, $f_{\mathrm{e}}\triangleq\log_2\left(\sum_{i\in\mathcal{L}}\mathrm{tr}\left(\mathbf{H}_{\mathrm{e}}^H\mathbf{V}\mathbf{H}_{\mathrm{e}}\mathbf{W}_i\right)+\mathrm{tr}\left(\mathbf{H}_{\mathrm{e}}^H\mathbf{V}\mathbf{H}_{\mathrm{e}}\mathbf{Z}\right)+\sigma_{\mathrm{e}}^2\right)$, $g_{\mathrm{p,e}\to k}\triangleq\log_2\left(\sum_{i\in\mathcal{L}\setminus\{k\}}\mathrm{tr}\left(\mathbf{H}_{\mathrm{e}}^H\mathbf{V}\mathbf{H}_{\mathrm{e}}\mathbf{W}_i\right)+\mathrm{tr}\left(\mathbf{H}_{\mathrm{e}}^H\mathbf{V}\mathbf{H}_{\mathrm{e}}\mathbf{Z}\right)+\sigma_{\mathrm{e}}^2\right)$, $f_{\mathrm{c},k}\triangleq\log_2\left(\sum_{i\in\mathcal{L}}\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{W}_i\right)+\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{Z}\right)+\sigma_k^2\right)$, $g_{\mathrm{c},k}\triangleq\log_2\left(\sum_{i\in\mathcal{K}}\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{W}_i\right)+\mathrm{tr}\left(\mathbf{H}_k^H\mathbf{V}\mathbf{H}_k\mathbf{Z}\right)+\sigma_k^2\right)$, and $g_{\mathrm{c,e}}\triangleq\log_2\left(\sum_{i\in\mathcal{K}}\mathrm{tr}\left(\mathbf{H}_{\mathrm{e}}^H\mathbf{V}\mathbf{H}_{\mathrm{e}}\mathbf{W}_i\right)+\mathrm{tr}\left(\mathbf{H}_{\mathrm{e}}^H\mathbf{V}\mathbf{H}_{\mathrm{e}}\mathbf{Z}\right)+\sigma_{\mathrm{e}}^2\right)$. Although (P2) is smooth, it is still non-convex w.r.t. $\mathcal{A}$. To tackle (P2), we resort to the widely used AO method [5], [6], [7], [8], [9], [10]. Specifically, we alternatingly optimize $\mathcal{A}\backslash\{\mathbf{V}\}$ and $\mathbf{V}$ until convergence is achieved, with details given below.

A. Optimizing $\mathcal{A}\backslash\{\mathbf{V}\}$ for Given $\mathbf{V}$

For any given $\mathbf{V}$, the subproblem w.r.t. $\mathcal{A}\backslash\{\mathbf{V}\}$ is given by

$$\max_{\mathcal{A}\backslash\{\mathbf{V}\}}\ t\quad \text{s.t. (2b)}-\text{(2e)},\text{(2g)},\text{(2i)}. \tag{3}$$

Note that the functions $f_{\mathrm{p},k}$, $g_{\mathrm{p},k}$, f_{e}, $g_{\mathrm{p,e}\to k}$, $f_{\mathrm{c},k}$, $g_{\mathrm{c},k}$, and $g_{\mathrm{c,e}}$ are all jointly concave w.r.t. their corresponding variables. Although the concavity of $g_{\mathrm{p},k}$, f_{e}, and $g_{\mathrm{c},k}$ makes the constraints in (2c) and (2d) non-convex, it facilitates the application of the iterative successive convex approximation (SCA) technique [15]. For ease of notation, let $\mathbf{W}$, $\hat{\mathbf{W}}$, and $\tilde{\mathbf{W}}$ denote the collections of the variables $\{\mathbf{W}_i\}_{\forall i\in\mathcal{L}}$, $\{\mathbf{W}_i\}_{\forall i\in\mathcal{K}}$, and $\{\mathbf{W}_i\}_{\forall i\in\mathcal{K}\setminus\{k\}}$, respectively. Then, given the local feasible points $\tilde{\mathbf{W}}^r\triangleq\{\mathbf{W}_i^r\}_{\forall i\in\mathcal{K}\setminus\{k\}}$ and $\mathbf{Z}^r$ in the r-th iteration, the differentiable concave function $g_{\mathrm{p},k}$ is globally upper-bounded by its first-order Taylor expansion, i.e.,

$$\begin{aligned} g_{\mathrm{p},k}(\tilde{\mathbf{W}},\mathbf{Z})&\leq g_{\mathrm{p},k}(\tilde{\mathbf{W}}^r,\mathbf{Z}^r)\\ &+\sum\nolimits_{i\in\mathcal{K}\setminus\{k\}}\mathrm{tr}\left(\left(\nabla_{\mathbf{W}_i}g_{\mathrm{p},k}\left(\tilde{\mathbf{W}}^r,\mathbf{Z}^r\right)\right)^H\left(\mathbf{W}_i-\mathbf{W}_i^r\right)\right)\\ &+\mathrm{tr}\left(\left(\nabla_{\mathbf{Z}}g_{\mathrm{p},k}\left(\tilde{\mathbf{W}}^r,\mathbf{Z}^r\right)\right)^H\left(\mathbf{Z}-\mathbf{Z}^r\right)\right)\triangleq g_{\mathrm{p},k}^r(\tilde{\mathbf{W}},\mathbf{Z}). \end{aligned} \tag{4}$$

Similarly, we can obtain the global upper bounds, denoted by $f_{\mathrm{e}}^r(\mathbf{W},\mathbf{Z})$ and $g_{\mathrm{c},k}^r(\hat{\mathbf{W}},\mathbf{Z})$, of f_{e} and $g_{\mathrm{c},k}$, respectively, whose expressions are similar to that given in (4) and are therefore omitted for brevity. Then, by replacing $g_{\mathrm{p},k}$, f_{e}, and $g_{\mathrm{c},k}$ with their respective global upper bounds, we can replace (2c) and (2d) with the following convex constraints:

$$(f_{\mathrm{p},k}-g_{\mathrm{p},k}^r(\tilde{\mathbf{W}},\mathbf{Z}))-(f_{\mathrm{e}}^r(\mathbf{W},\mathbf{Z})-g_{\mathrm{p,e}\to k})\geq r_{\mathrm{p},k}^{\mathrm{sec}},\quad \forall k\in\mathcal{K}, \tag{5}$$

$$\sum\nolimits_{k\in\mathcal{K}} r_{\mathrm{c},k}^{\mathrm{sec}}\leq (f_{\mathrm{c},k}-g_{\mathrm{c},k}^r(\hat{\mathbf{W}},\mathbf{Z}))-(f_{\mathrm{e}}^r(\mathbf{W},\mathbf{Z})-g_{\mathrm{c,e}}),\quad \forall k\in\mathcal{K}, \tag{6}$$

respectively. However, the rank constraints in (2g) make the problem at hand still non-convex. Thus, we drop the rank constraints by applying semidefinite relaxation (SDR) [5], [6], [7] that yields

$$\max_{\mathcal{A}\backslash\{\mathbf{V}\}}\ t \tag{7a}$$

$$\text{s.t. (2b)},(5),(6),\text{(2e)},\text{(2i)},\mathbf{Z}\succeq\mathbf{0},\mathbf{W}_l\succeq\mathbf{0},\quad \forall l\in\mathcal{L}. \tag{7b}$$

By direct inspection, problem (7) is a convex semidefinite program (SDP) that can be efficiently solved by off-the-shelf solvers such as CVX. Particularly, the adopted SDR is tight for problem (7). The detailed proof is similar to that of [7, Theorem 1] and we omit it due to the space limitation.

B. Optimizing $\mathbf{V}$ for Given $\mathcal{A}\backslash\{\mathbf{V}\}$

For any given $\mathcal{A}\backslash\{\mathbf{V}\}$, (P2) is reduced to a feasibility-check problem. Inspired by [1], we introduce "residual" variables Δt, $\Delta\mathbf{r}_{\mathrm{c}}^{\mathrm{sec}}\triangleq\{\Delta r_{\mathrm{c},1}^{\mathrm{sec}},\cdots,\Delta r_{\mathrm{c},K}^{\mathrm{sec}}\}$, and $\Delta\mathbf{r}_{\mathrm{p}}^{\mathrm{sec}}\triangleq\{\Delta r_{\mathrm{p},1}^{\mathrm{sec}},\cdots,\Delta r_{\mathrm{p},K}^{\mathrm{sec}}\}$, and then transform the subproblem into the following problem

$$\max_{\mathbf{V}\in\mathbb{H}^{N+1},\Delta t,\Delta\boldsymbol{r}_{\mathrm{c}}^{\mathrm{sec}},\Delta\boldsymbol{r}_{\mathrm{p}}^{\mathrm{sec}}}\ t+\Delta t \tag{8a}$$

$$\text{s.t. } r_{\mathrm{c},k}^{\mathrm{sec}}+\Delta r_{\mathrm{c},k}^{\mathrm{sec}}+r_{\mathrm{p},k}^{\mathrm{sec}}+\Delta r_{\mathrm{p},k}^{\mathrm{sec}}\geq t+\Delta t,\quad \forall k\in\mathcal{K}, \tag{8b}$$

$$(f_{\mathrm{p},k}-g_{\mathrm{p},k})-(f_{\mathrm{e}}-g_{\mathrm{p,e}\to k})\geq r_{\mathrm{p},k}^{\mathrm{sec}}+\Delta r_{\mathrm{p},k}^{\mathrm{sec}},\quad \forall k\in\mathcal{K}, \tag{8c}$$

$$\sum\nolimits_{k\in\mathcal{K}} r_{\mathrm{c},k}^{\mathrm{sec}}+\Delta r_{\mathrm{c},k}^{\mathrm{sec}}\leq (f_{\mathrm{c},k}-g_{\mathrm{c},k})-(f_{\mathrm{e}}-g_{\mathrm{c,e}}),\quad \forall k\in\mathcal{K}, \tag{8d}$$

$$\Delta t\geq 0,\Delta r_{\mathrm{c},k}^{\mathrm{sec}}\geq 0,\Delta r_{\mathrm{p},k}^{\mathrm{sec}}\geq 0,\quad \forall k\in\mathcal{K}, \tag{8e}$$

$$\text{(2f)},\text{(2h)}, \tag{8f}$$

which is more efficient than the original subproblem when it comes to the converged solution (please refer to [1] for a detailed explanation). Similar to problem (3), the non-convexity of problem (8) stems from the concavity of $g_{\mathrm{p},k}$, f_{e}, and $g_{\mathrm{c},k}$ as well as the non-convex rank constraint in (2h). As in the previous subsection, we employ the SCA method to tackle this problem. To be specific, by applying the first-order Taylor expansion at the given local feasible point $\mathbf{V}^r$ in the r-th iteration to $g_{\mathrm{p},k}$, we obtain $g_{\mathrm{p},k}(\mathbf{V})\leq g_{\mathrm{p},k}(\mathbf{V}^r)+\mathrm{tr}\left(\left(\nabla_{\mathbf{V}}g_{\mathrm{p},k}(\mathbf{V}^r)\right)^H\left(\mathbf{V}-\mathbf{V}^r\right)\right)\triangleq g_{\mathrm{p},k}^r(\mathbf{V})$. Similarly, f_{e} and $g_{\mathrm{c},k}$ are globally upper-bounded by their respective first-order Taylor expansions at $\mathbf{V}^r$, denoted by $f_{\mathrm{e}}^r(\mathbf{V})$ and $g_{\mathrm{c},k}^r(\mathbf{V})$, respectively. Accordingly, the non-convex constraints

(8c) and (8d) can be approximated as

$$(f_{\mathrm{p},k} - g^r_{\mathrm{p},k}(\mathbf{V})) - (f^r_{\mathrm{e}}(\mathbf{V}) - g_{\mathrm{p,e}\to k}) \geq r^{\mathrm{sec}}_{\mathrm{p},k} + \Delta r^{\mathrm{sec}}_{\mathrm{p},k}, \quad \forall k \in \mathcal{K}, \tag{9}$$

$$\sum\nolimits_{k\in\mathcal{K}} r^{\mathrm{sec}}_{\mathrm{c},k} + \Delta r^{\mathrm{sec}}_{\mathrm{c},k} \leq (f_{\mathrm{c},k} - g^r_{\mathrm{c},k}(\mathbf{V})) - (f^r_{\mathrm{e}}(\mathbf{V}) - g_{\mathrm{c,e}}), \quad \forall k \in \mathcal{K}. \tag{10}$$

The only remaining obstacle to solving problem (8) is the non-convex rank constraint in (2h). Recall that when optimizing $\{\mathbf{W}_l\}$ in the previous subsection, we dropped the rank constraints and showed the tightness of SDR. However, dropping the rank constraint w.r.t. $\mathbf{V}$ in problem (8) cannot guarantee a rank-one optimal solution. Thus, instead of applying SDR, we exploit the penalty-based method [7] to handle the rank constraint. Specifically, $\mathrm{rank}\,(\mathbf{V}) \leq 1$ is equivalent to $\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \leq 0$. Then, we incorporate the constraint $\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \leq 0$ into the objective function (8a) by introducing a positive penalty parameter ρ and we replace the non-convex constraints (8c) and (8d) with their convex subsets (9) and (10), which yields the following problem

$$\min_{\mathbf{V}\in\mathbb{H}^{N+1},\Delta t,\Delta \boldsymbol{r}^{\mathrm{sec}}_{\mathrm{c}},\Delta \boldsymbol{r}^{\mathrm{sec}}_{\mathrm{p}}} \; -t - \Delta t + \frac{1}{2\rho}\left(\|\mathbf{V}\|_* - \|\mathbf{V}\|_2\right) \tag{11a}$$

$$\text{s.t. (8b), (9), (10), (8e), (2f)}, \mathbf{V} \succeq \mathbf{0}. \tag{11b}$$

According to [7, Proposition 2], problem (11) admits a rank-one solution when ρ is sufficiently small. Note that the convexity of $\|\mathbf{V}\|_2$ makes problem (11) still non-convex, which motivates us to replace $\|\mathbf{V}\|_2$ by its first-order Taylor expansion-based lower bound. By doing so, we can approximate problem (11) as

$$\min_{\mathbf{V}\in\mathbb{H}^{N+1},\Delta t,\Delta \boldsymbol{r}^{\mathrm{sec}}_{\mathrm{c}},\Delta \boldsymbol{r}^{\mathrm{sec}}_{\mathrm{p}}} \; -t - \Delta t + \frac{1}{2\rho}\Big(\|\mathbf{V}\|_* - \|\mathbf{V}^r\|_2 - \mathrm{tr}\left(\boldsymbol{\lambda}^r_{\max}\left(\boldsymbol{\lambda}^r_{\max}\right)^H(\mathbf{V}-\mathbf{V}^r)\right)\Big) \tag{12a}$$

$$\text{s.t. (8b), (9), (10), (8e), (2f)}, \mathbf{V} \succeq \mathbf{0}, \tag{12b}$$

where $\boldsymbol{\lambda}^r_{\max}$ is the eigenvector that corresponds to the largest eigenvalue of $\mathbf{V}^r$. Problem (12) is a convex SDP and existing solvers (e.g., CVX) can be utilized to solve it optimally.

C. Computational Complexity Analysis

Regarding complexity, obviously, it is dominated by solving SDPs (7) and (12) in each iteration. The computational complexity of solving (7) is $\mathcal{O}\left(\sqrt{M}\log\frac{1}{\varepsilon}\left(KM^3+K^2M^2+K^3\right)\right)$ and that of solving (12) is $\mathcal{O}\left(\sqrt{N}\log\frac{1}{\varepsilon}\left(N^4+KN^3+K^2N^2+K^3\right)\right)$, where $\varepsilon > 0$ represents the required solution accuracy in each iteration [7]. Hence, the computational complexity of each iteration of the proposed algorithm is about $\mathcal{O}\Big(\log\frac{1}{\varepsilon}\big(\sqrt{M}(KM^3+K^2M^2+K^3)+\sqrt{N}(N^4+KN^3+K^2N^2+K^3)\big)\Big)$.

IV. Simulation Results

This section evaluates the effectiveness of the proposed algorithm via simulations. We consider a two-dimensional (2D) coordinate setup and the locations of the AP, the IRS, and Eve are set as $(0,0)$, $(50,0)$, and $(45,0)$ in meters (m), respectively. The large-scale path loss model in [1] is adopted with the path loss at the reference distance of 1 m being -30 dB and the path loss exponents being 2.2 for the cascaded links while 3.5 for the direct links [9]. Furthermore, we adopt the Rician fading model for the cascaded links with a Rician factor of 3 dB while the Rayleigh fading model for the direct links [1]. Unless further specified, other parameters are set as $\sigma_j^2 = -80$ dBm [1], $\forall j \in \mathcal{K}\cup\{\mathrm{e}\}$, $\rho = 5\times10^{-4}$ [7], and $\varepsilon = 10^{-4}$ [1]. For comparison, we adopt two other transmission schemes, namely, MU-LP [7], [8], [9] and NOMA [10].

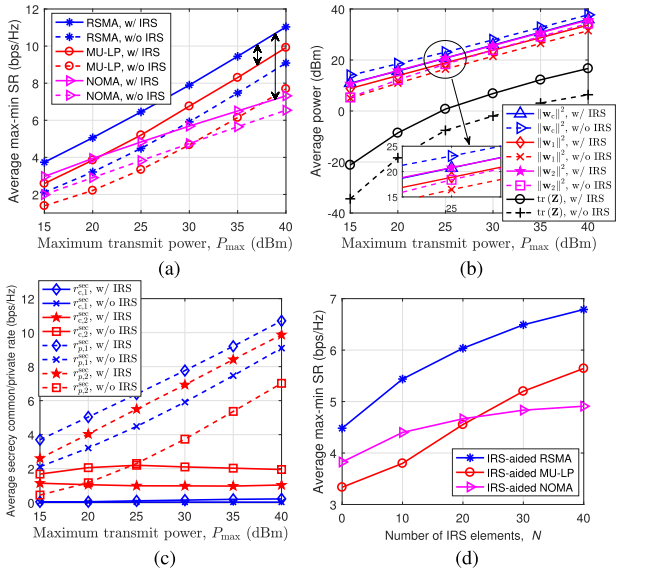

Fig. 2. (a) Average max-min SR versus $P_{\max}$; (b) Average transmit power allocation of the RSMA scheme versus $P_{\max}$; (c) Average secrecy common/private rate allocation of the RSMA scheme versus $P_{\max}$; (d) Average max-min SR versus N.

Fig. 2(a) depicts the average max-min SR of different strategies versus $P_{\max}$. Here, we set $M = 2$, $N = 30$, and $K = 2$ with LUs 1 and 2 being located at $(0,20)$ and $(50,5)$ in m, respectively. Two cases with and without the IRS are considered. As can be seen, the IRS-aided designs are obviously superior to their counterparts without the IRS. This is expected since the IRS phase shifts can be properly adjusted not only to enhance the signal strength at the LUs while degrading that at Eve, but also to decrease the interference power at the LUs while increasing that at Eve. It is also observed that our proposed IRS-aided RSMA scheme significantly outperforms the other two strategies. To acquire more insights, we plot the transmit power allocation and the secrecy common/private rate allocation obtained by the RSMA scheme in Figs. 2(b) and 2(c), respectively. Fig. 2(b) shows that the transmit power allocation changes greatly after the introduction of the IRS, which corroborates that the IRS is capable of reconfiguring the wireless propagation environment. From Fig. 2(c), it can be seen that since the obtained secrecy private rate of LU 2 is much smaller than that of LU 1, almost all of the achievable secrecy common rate is allocated to LU 2 to improve its total SR, thereby ensuring the secrecy fairness between LUs 1 and 2. Moreover, we note that with increasing $P_{\max}$, the secrecy private rates increase rapidly, while the secrecy common rates change slightly. A possible explanation is that as Eve's received power from $\mathbf{w}_{\mathrm{c}}$ becomes significant

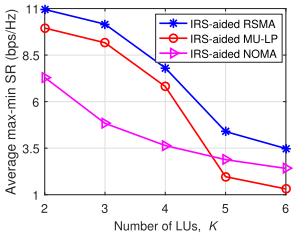

Fig. 3. Average max-min SR versus K.

in the high $P_{\max}$ regime, the common message contributes little to the total achievable secrecy common rate. Yet, it can serve as AN and contribute much to the individual secrecy private rates. Besides, we observe from Fig. 2(a) that in the case without IRS, the max-min SR of the MU-LP scheme is noticeably higher than that of its NOMA counterpart when $P_{\max}$ is larger than about 33 dBm. This is consistent with the result in [17], which states that for two-user MISO broadcast channels, MU-LP strictly outperforms NOMA at high signal-to-noise ratio (SNR) as the max-min fair (MMF) multiplexing gain of MU-LP is twice that of NOMA. When deploying an IRS, however, the MU-LP scheme performs better than the NOMA scheme when $P_{\max}$ is larger than about 23 dBm. The reason is that the IRS can lead to a high SNR even when $P_{\max}$ is not that large.

Following the same setup in Fig. 2(a), we investigate in Fig. 2(d) the impact of the number of IRS elements on the system SR performance with $P_{\max} = 25$ dBm. It is seen that the proposed IRS-aided RSMA scheme always obtains the highest max-min SR. This is understandable since 1-layer RS is always a super-scheme of MU-LP regardless of the value of K while NOMA is a sub-scheme of 1-layer RS in a two-user case (although it is not the case when $K > 2$) [17]. It is also notable that the performance of the IRS-aided MU-LP scheme exceeds that of its NOMA counterpart when N becomes large. This is because increasing N with a fixed $P_{\max}$ can be viewed as increasing $P_{\max}$ with a fixed N that results in an increase in SNR. Finally, we observe that when a certain level of performance is desired, the surface size of the IRS required by the RSMA scheme is much smaller than that required by the other two schemes. This suggests that the IRS-aided RSMA scheme is more appealing to space-limited scenarios.

Fig. 3 plots the average max-min SR of different schemes versus K with $M = 4$, $N = 20$, and $P_{\max} = 35$ dBm. The locations of LUs 1 and 2 are identical to those set in Fig. 2(a) and LUs $3-6$ are located at $(0,-20)$, $(50,-5)$, $(55,0)$, and $(-20,0)$ in m, respectively. It is observed that the proposed IRS-aided RSMA scheme still performs the best in both considered underloaded and overloaded regimes. This observation is in line with the theoretical and numerical results in [17] that 1-layer RS always achieves the same or higher MMF multiplexing gains and rates than MU-LP and NOMA. Another observation is that when K increases from 4 to 5, the max-min SR achieved by the IRS-aided MU-LP scheme drops sharply and is no longer higher than that achieved by the IRS-aided NOMA scheme. This is because the MMF multiplexing gain of MU-LP is 1 if $K \leq M$ and 0 otherwise, while that of NOMA is $1/K$ whenever $K \geq 2$ [17].

V. Conclusion

This letter investigated the potential performance gain of integrating RSMA with IRS in a secure communication system. A computationally efficient iterative algorithm was developed to guarantee the secrecy fairness among the LUs, where the transmit/reflect beamforming with AN and the secrecy common rate allocation were jointly optimized. Simulation results showed that our proposed IRS-aided RSMA strategy can achieve a higher secrecy performance enhancement than the existing IRS-aided MU-LP and NOMA schemes. Possible extensions include the case of imperfect CSI [9], adopting the generalized RS strategy [11], and investigating whether AN is needed in a secure RSMA-based system (especially in the case of multiple Eves), which are left for future work.

References

[1] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
[2] D. Li, "Ergodic capacity of intelligent reflecting surface-assisted communication systems with phase errors," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1646–1650, Aug. 2020.
[3] Z. Lin et al., "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 3717–3724, Aug. 2022.
[4] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.
[5] M. Cui et al., "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.
[6] X. Guan et al., "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 778–782, Jun. 2020.
[7] X. Yu et al., "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.
[8] H. Niu et al., "Weighted sum secrecy rate maximization using intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6170–6184, Sep. 2021.
[9] H. Niu et al., "Robust design for intelligent reflecting surface assisted secrecy SWIPT network," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4133–4149, Jun. 2022.
[10] Z. Zhang et al., "Robust and secure communications in intelligent reflecting surface assisted NOMA networks," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 739–743, Mar. 2021.
[11] Y. Mao et al., "Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–54, May 2018.
[12] H. Fu et al., "Robust secure beamforming design for two-user downlink MISO rate-splitting systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8351–8365, Dec. 2020.
[13] Y. Mao et al., "Max-min fairness of K-user cooperative rate-splitting in MISO broadcast channel with user relaying," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6362–6376, Oct. 2020.
[14] Z. Yang et al., "Energy efficient rate splitting multiple access (RSMA) with reconfigurable intelligent surface," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jul. 2020, pp. 1–6.
[15] H. Fu, S. Feng, and D. W. Kwan Ng, "Resource allocation design for IRS-aided downlink MU-MISO RSMA systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6.
[16] A. Bansal, K. Singh, B. Clerckx, C.-P. Li, and M.-S. Alouini, "Rate-splitting multiple access for intelligent reflecting surface aided multi-user communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9217–9229, Sep. 2021.
[17] B. Clerckx et al., "Is NOMA efficient in multi-antenna networks? A critical look at next generation multiple access techniques," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1310–1343, 2021.