STAR-RIS Enhanced Finite Blocklength Transmission for Uplink NOMA Networks

Suyu Lv, Xiaodong Xu, Senior Member, IEEE, Shujun Han, Member, IEEE, Yuanwei Liu, Senior Member, IEEE
Ping Zhang, Fellow, IEEE and Arumugam Nallanathan, Fellow, IEEE

Abstract—A simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) assisted uplink non-orthogonal multiple access (NOMA) framework for finite blocklength (FBL) transmission is proposed. Considering the different communication requirements of Internet of Things devices (IoTDs), a novel design to achieve high-rate and low-error is proposed. Two operating protocols for STAR-RIS are considered, namely energy splitting (ES) and mode switching (MS). 1) For STAR-RIS with ES, an alternating optimization (AO) algorithm is proposed to handle the highly-coupled integer programming problem. More particularly, a low-complexity received-signal-strength-based device pairing scheme is proposed. Based on the given device pair, the closed-form solutions for the power allocation problem are obtained. The transmitting and reflecting coefficient optimization problem is solved by exploiting the successive convex approximation and semidefinite relaxation methods. 2) For STAR-RIS with MS, a double-layer penalty-based (DLPB) algorithm is proposed to tackle the newly introduced binary amplitude constraints. Numerical results reveal that: i) the proposed AO and DLPB algorithms can converge within a few iteration times; ii) the FBL transmission performance can be improved by employing the proposed STAR-RIS framework compared with conventional transmitting/reflecting-only RISs; iii) NOMA is capable of enhancing FBL rate while guaranteeing the reliability constraints compared with orthogonal multiple access.

Index Terms—Finite blocklength (FBL) transmission, Internet of things (IoT), non-orthogonal multiple access (NOMA), simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS).

I. INTRODUCTION

The flourishing Internet of Things (IoT) technology will provide an enabling platform for the ubiquitous smart devices and sensors, making it an integral part for the realization of intelligent communication systems. It is predicted that by 2030, about 50 billion IoT devices (IoTDs) will be in service worldwide [1], which will form a tremendous network of interconnected devices from wearables to driverless vehicles. Profiting from this, our world is developing towards a more intelligent direction, covering various important fields such as smart home, smart healthcare, intelligent agriculture and intelligent manufacturing [2]–[4].

On the one hand, IoT technologies serving different industries also have different communication objectives and requirements, such as large bandwidth, high data rate, stringent reliability and end-to-end latency [5], [6]. On the other hand, the demand for massive IoT connections also makes the scarcity of spectrum resources increasingly prominent. Therefore, the novel design for satisfying various objectives and massive connections of IoT networks is becoming increasingly urgent. Furthermore, IoT technologies may be deployed in a variety of complex environments and the presence of high-rises, trees and other objects may cause the obstruction of line-of-sight (LoS) link, especially in urban or indoor environments. In order to overcome these challenges, reconfigurable intelligent surface (RIS) can be recognized as a potential green solution to improve communication quality by intelligently adjusting the wireless radio environment with extremely low energy consumption.

Traditionally, RISs consist of an array of low-cost passive reflecting elements, in which each element independently reflects the incident signal after adjusting the electromagnetic wave phase. However, the limitation of conventional RISs lies in its reflecting-only ability, which means that the transmitter and the receiver should be located on the same side of RIS, resulting in half-space coverage constraints. For solving this problem, a new paradigm named simultaneously transmitting and reflecting RIS (STAR-RIS) has been proposed, where the incident signals can be reflected and transmitted to both sides of the surface concurrently, thus achieving a full-space intelligent propagation environment [7]–[10]. Three practical protocols for operating STAR-RISs in wireless networks, namely energy splitting (ES), mode switching (MS), and time switching (TS), were proposed by the authors in [7].

A. Prior Works

Strict latency requirements and massive access requirements are key challenges for IoT. Aiming at realizing extremely low packet error probability and ultra-low latency for IoT networks, finite blocklength (FBL) transmission is viewed as a feasible solution, which has attracted considerable attention in recent years [11]–[15]. Moreover, non-orthogonal multiple access (NOMA) is considered a key remedy to provide massive connectivity by allowing multiple devices to share the same time-frequency resources simultaneously, and eliminating inter-device interferences at the receivers through
successive interference cancellation (SIC) technology [16],[17]. For improving user fairness and spectral efficiency of IoT networks, the authors in [18] introduced NOMA into FBL transmission, and investigated the tradeoff among the throughput, latency and reliability. Considering the joint decoding of short packets in NOMA-based networks, the authors in [19] studied the optimization of decoding error probability and power allocation factor. To address the problem that wireless signals are easily blocked by densely deployed equipment in factory automation scenarios, the authors in [20]–[22] proposed to utilize RIS to provide an alternative link, thus enhancing the FBL transmission rate and reliability. The authors in [23] combined RIS and NOMA into IoT systems to meet the needs of high energy efficiency, low power consumption and wide coverage. Aiming at maximizing the network sum rate or energy efficiency for RIS-assisted NOMA systems, the joint resource optimization problems were investigated in [24]–[26]. However, although the combination of RIS and NOMA has obvious advantages in enhancing capacity, it still has the limitations of half-space smart environment.

STAR-RISs have natural advantages in NOMA-based communication systems because of their distinctive ability to expand the channel differences by reconstructing the wireless propagation environment on both sides, which encourages the research on the integration of STAR-RIS and NOMA technologies [27]–[30]. The studies in [27] and [28] confirmed the superiority of STAR-RIS-NOMA in improving system sum rate compared with the conventional RIS and STAR-RIS-NOMA. A simultaneous signal-enhancement-and-cancellation-based design was proposed for STAR-RIS enhanced NOMA joint-transmission-coordinated-multipoint networks in [29], aiming at eliminating the inter-cell interference while boosting the desired signals simultaneously. Under FBL regime, the authors in [30] provided block error rate (BLER) analysis for STAR-RIS-aided downlink NOMA communication system. However, when applied to uplink communication systems, the incident signals come from both sides of the STAR-RISs, which will inevitably cause energy loss and information leakage. Focusing on uplink NOMA systems with the assistance of STAR-RISs, the authors in [31]–[33] investigated the hardware and signal model for the dual-sided STAR-RISs, the transmit power minimization problem and the secrecy transmission problem, respectively. In order to explore the potential of exploiting STAR-RIS to enhance throughput in uplink NOMA systems, the authors in [34] theoretically derived the ergodic sum rate under channel estimation errors and hardware impairments, while the authors in [35] investigated the average rate maximization problem with movable STAR-RIS. However, these researches are only aimed at one single performance metric, which cannot be applicable to IoT networks with diverse communication requirements.

B. Motivation and Contributions

On the one hand, massive IoT devices will generate a huge number of short data packets for uplink transmission, which requires high-efficiency multiple access to alleviate the pressure of scarce spectrum resources. On the other hand, IoT devices have various communication demands, which requires novel system design to satisfy specific needs in different application scenarios. Although the STAR-RIS-assisted NOMA system has received widespread attention, the research on dual-sided STAR-RIS for uplink transmission is still in its infancy, which motivates us to develop this paper. More specifically, there is no research considering that users on STAR-RIS’s two sides have different communication requirements, especially for the uplink-transmission-dominant networks with FBL constraints. Therefore, we focus on the FBL transmission performance enhancement in STAR-RIS-aided uplink NOMA IoT networks, aiming at maximizing the achievable sum rate of the IoTDSs on the reflection space, while ensuring the reliability of the IoTDSs on the transmission space. The main contributions of our works are listed as follows:

1) We propose a STAR-RIS-assisted uplink NOMA communication framework for IoT networks, where IoTDSs on reflection space (R-IoTDSs) are transmission performance-oriented, aiming to enhance achievable FBL rate, while the IoTDSs on transmission space (T-IoTDSs) aim at guaranteeing certain reliability under FBL constraints.

2) For STAR-RIS with ES protocol, we propose an alternating optimization (AO) algorithm to perform the resource allocation in an iterative manner, aiming to achieve high-rate and low-error (HRLE) for R-IoTDSs and T-IoTDSs respectively. Specifically, we first propose a received signal strength (RSS) based device pairing scheme, then obtain the closed-form expressions for the IoTDSs’ transmission power allocation problem. Finally, the STAR-RIS coefficient optimization problem is solved by the successive convex approximation (SCA) and semi-definite relaxation (SDR) methods.

3) For STAR-RIS with MS protocol, we propose a double-layer penalty-based (DLPB) algorithm to handle the newly introduced binary constraints. More particularly, the inner layer is for solving the resource allocation optimization problem with relaxed binary amplitude coefficient constraints, while the outer layer is for the binary amplitude constraint recovery.

4) Numerical results reveal that: i) the proposed AO and DLPB algorithms can achieve convergence within a few iterations; ii) the proposed STAR-RIS framework can obtain a higher sum FBL rate compared with the conventional transmitting/reflecting-only RISs; iii) NOMA scheme outperforms the OMA scheme in improving the achievable FBL rate while guaranteeing reliability constraints; iv) ES protocol of STAR-RIS has better FBL transmission performance than MS protocol.

The organization for the rest of this paper is as follows. In section II, we propose a STAR-RIS-assisted uplink NOMA framework for FBL transmission. In section III, we propose an AO algorithm for STAR-RIS with ES protocol to achieve HRLE design. Based on the AO algorithm, we propose a DLPB algorithm for STAR-RIS with MS protocol in section IV. The simulation results and conclusions are given in sections V and VI, respectively. The notations used in this paper are presented in Table I.
TABLE I

<table>
<thead>
<tr>
<th>Notations</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOTD m</td>
<td>m-th IOTD on R-Space</td>
</tr>
<tr>
<td>IOTD n</td>
<td>n-th IOTD on T-Space</td>
</tr>
<tr>
<td>$\mathcal{M}$</td>
<td>set of IOTDs on R/Space</td>
</tr>
<tr>
<td>$\mathcal{N}$</td>
<td>number of IOTDs on R/Space</td>
</tr>
<tr>
<td>K</td>
<td>number of STAR-RIS elements</td>
</tr>
<tr>
<td>$x_{m,n}$</td>
<td>indicator of device pairing</td>
</tr>
<tr>
<td>$h_0$</td>
<td>path loss at the reference distance $d = 1m$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>path loss exponent</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>path gain factor</td>
</tr>
<tr>
<td>$p_m/p_n$</td>
<td>transmission power of IOTD m/IOTD n</td>
</tr>
<tr>
<td>$L_m/L_n$</td>
<td>short-packet blocklength</td>
</tr>
<tr>
<td>$\varepsilon_{m/n}$</td>
<td>tolerable decoding error probability</td>
</tr>
<tr>
<td>$R_{m/n}$</td>
<td>target transmission rate of IOTD n</td>
</tr>
<tr>
<td>$\Theta/R$</td>
<td>diagonal matrix whose diagonal elements are $x_i$</td>
</tr>
<tr>
<td>$\text{Diag}(X)$</td>
<td>vector whose elements are the diagonal elements of X</td>
</tr>
<tr>
<td>$Q(\cdot)$</td>
<td>Gaussian Q-function</td>
</tr>
<tr>
<td>$[\cdot]$</td>
<td>max $[\cdot, 0]$</td>
</tr>
</tbody>
</table>

The STAR-RIS consists of a uniform planar array (UPA) with $K = K_x, K_y$ elements, where $K_x$ and $K_y$ are the number of elements along the x-axis and y-axis, respectively. The uplink transmission services are dominant in IoT networks, and signals are transmitted from IOTDs on both sides to D with the assistance of STAR-RIS. Denote the transmitted data symbols of IOTD m ($m \in \mathcal{M}$) and IOTD n ($n \in \mathcal{N}$) as $s_m$ and $s_n$, respectively. Disregarding the effect of path loss and fading, the signals of IOTD m and IOTD n radiating from the k-th element towards R-Space and towards T-Space, as shown in Fig. 2, can be given as follows

$$s_{k,R} = \sqrt{\beta_k^R} e^{j \theta_{k,R}} \sqrt{p_m s_m + \sqrt{\beta_k^R} e^{j \theta_{k,R}}} \sqrt{p_m s_n}, \quad (1a)$$

$$s_{k,T} = \sqrt{\beta_k^T} e^{j \theta_{k,T}} \sqrt{p_m s_m + \sqrt{\beta_k^T} e^{j \theta_{k,T}}} \sqrt{p_n s_n}, \quad (1b)$$

where $p_m$ and $p_n$ denote the transmission power of IOTD m and IOTD n respectively. $\beta_k^R$, $\beta_k^T$, $\Theta_k^R$ and $\Theta_k^T$ denote the amplitude coefficient of reflection for R-Space (transmission from R/Space to T/Space) of the dual-sided STAR-RIS. Specifically, $\beta_k^R, \beta_k^T, \Theta_k^R, \Theta_k^T \in [0, 1]$ for STAR-RIS operating with energy splitting (ES) protocol, and $\beta_k^R, \beta_k^T, \Theta_k^R, \Theta_k^T \in \{0, 1\}$ for STAR-RIS operating with mode switching (MS) protocol, satisfying the principle of energy conservation, i.e., $\beta_k^T + \beta_k^R = 1$ and $\beta_k^T + \beta_k^R = 0$ [1]. Similarly, $\theta_{k,R}, \theta_{k,T} \in [0, 2\pi]$ denote the phase shift coefficient of the dual-sided STAR-RIS.

Symmetrical dual-sided elements are considered [31]. Thus, we have $\beta_k^R, \beta_k^T, \Theta_k^R, \Theta_k^T \in [0, 2\pi]$ respectively hereinafter, denoting the amplitude coefficient of reflection/transmission and the phase shift coefficient of reflection/transmission. Accordingly, (1) can be reformulated as

$$s_{k,R} = \sqrt{\beta_k^R} e^{j \theta_{k,R}} \sqrt{p_m s_m + \sqrt{\beta_k^R} e^{j \theta_{k,R}}} \sqrt{p_n s_n}, \quad (2a)$$

$$s_{k,T} = \sqrt{\beta_k^T} e^{j \theta_{k,T}} \sqrt{p_m s_m + \sqrt{\beta_k^T} e^{j \theta_{k,T}}} \sqrt{p_n s_n}, \quad (2b)$$

A. STAR-RIS Model

STAR-RIS consists of a uniform planar array (UPA) with $K = K_x, K_y$ elements, where $K_x$ and $K_y$ are the number of elements along the x-axis and y-axis, respectively. Since uplink transmission services are dominant in IoT networks, we consider signals are transmitted from IOTDs on both sides to D with the assistance of STAR-RIS. Denote the transmitted data symbols of IOTD m ($m \in \mathcal{M}$) and IOTD n ($n \in \mathcal{N}$) as $s_m$ and $s_n$, respectively. Disregarding the effect of path loss and fading, the signals of IOTD m and IOTD n radiating from the k-th element towards R-Space and towards T-Space, as shown in Fig. 2, can be given as follows

$$s_{k,R} = \sqrt{\beta_k^R} e^{j \theta_{k,R}} \sqrt{p_m s_m + \sqrt{\beta_k^R} e^{j \theta_{k,R}}} \sqrt{p_n s_n}, \quad (1a)$$

$$s_{k,T} = \sqrt{\beta_k^T} e^{j \theta_{k,T}} \sqrt{p_m s_m + \sqrt{\beta_k^T} e^{j \theta_{k,T}}} \sqrt{p_n s_n}, \quad (1b)$$

where $p_m$ and $p_n$ denote the transmission power of IOTD m and IOTD n respectively. $\beta_k^R$, $\beta_k^T$, $\Theta_k^R$, and $\Theta_k^T$ denote the amplitude coefficient of reflection for R-Space (transmission from R/Space to T/Space) of the dual-sided STAR-RIS. Specifically, $\beta_k^R, \beta_k^T, \Theta_k^R, \Theta_k^T \in [0, 1]$ for STAR-RIS operating with energy splitting (ES) protocol, and $\beta_k^R, \beta_k^T, \Theta_k^R, \Theta_k^T \in \{0, 1\}$ for STAR-RIS operating with mode switching (MS) protocol, satisfying the principle of energy conservation, i.e., $\beta_k^R + \beta_k^T = 1$ and $\beta_k^R + \beta_k^T = 1$ [7]. Similarly, $\theta_{k,R}, \theta_{k,T} \in [0, 2\pi]$ denote the phase shift coefficient of the dual-sided STAR-RIS.

Symmetrical dual-sided elements are considered [31]. Thus, we have $\beta_k^R = \beta_k^T, \beta_k^R = \beta_k^T, \Theta_k^R = \Theta_k^T, \Theta_k^R = \Theta_k^T$, $\beta_k^R = \beta_k^T$, $\Theta_k^R = \Theta_k^T$, and $\beta_k^R = \beta_k^T$, which are abbreviated as $\beta_k$, $\beta_k$, $\Theta_k^R$, and $\Theta_k$ respectively hereinafter, denoting the amplitude coefficient of reflection/transmission and the phase shift coefficient of reflection/transmission. Accordingly, (1) can be reformulated as

$$s_{k,R} = \sqrt{\beta_k^R} e^{j \theta_{k,R}} \sqrt{p_m s_m + \sqrt{\beta_k^R} e^{j \theta_{k,R}}} \sqrt{p_n s_n}, \quad (2a)$$

$$s_{k,T} = \sqrt{\beta_k^T} e^{j \theta_{k,T}} \sqrt{p_m s_m + \sqrt{\beta_k^T} e^{j \theta_{k,T}}} \sqrt{p_n s_n}, \quad (2b)$$

Although the scenario considered in this article is specific, it provides the possibility for STAR-RIS to support users with different communication requirements. Based on the proposed framework, more STAR-RIS-aided wireless systems with various communication requirements can be designed.
where $\beta_k^+, \beta_k^- \in [0, 1]$ for ES, $\beta_k^+, \beta_k^- \in \{0, 1\}$ for MS, $\beta_k^+ + \beta_k^- = 1$, and $\theta_k^+, \theta_k^- \in [0, 2\pi)$. Note that $D$ is located at $R$-Space, thus the signal $s_{r,k}$ is the main consideration. Denote the reflection-coefficient matrix and the transmission-coefficient matrix of STAR-RIS as $\Theta_r \in \mathbb{C}^{K \times K}$ and $\Theta_t \in \mathbb{C}^{K \times K}$ respectively, given by

$$\Theta_r = \text{diag} \left( \sqrt{\beta_1^+} e^{j\theta_1^+}, \ldots, \sqrt{\beta_K^+} e^{j\theta_K^+}, \ldots, \sqrt{\beta_1^-} e^{j\theta_1^-}, \ldots, \sqrt{\beta_K^-} e^{j\theta_K^-} \right),$$

$$\Theta_t = \text{diag} \left( \sqrt{\beta_1^+} e^{j\theta_1^+}, \ldots, \sqrt{\beta_K^+} e^{j\theta_K^+}, \ldots, \sqrt{\beta_1^-} e^{j\theta_1^-}, \ldots, \sqrt{\beta_K^-} e^{j\theta_K^-} \right).$$

\(3a\) \hspace{1cm} \(3b\)

B. Channel Model

Uplink NOMA scheme is adopted, where one R-IoTD and one T-IoTD are considered to share a same resource block. Define a binary matching coefficient $x_{m,n}$, where $x_{m,n} = 1$ indicates that $m$ and $n$ are on a same NOMA cluster, sharing the same time-and-frequency resource block, $x_{m,n} = 0$, otherwise. Without loss of generality, we consider that IoTD $m$ and IoTD $n$ are on a same NOMA cluster. To investigate the pairing strategy of one device on each side of STAR-RIS, we consider the case that $M = N$ in this paper.

The direct channel from IoTD $m$ to $D$ is denoted as $h_{m,d}$, which is modeled as Rician fading channels,

$$h_{m,d} = h_0(d_{m,d})^{-\alpha} \sqrt{\frac{\kappa}{1 + \kappa}} h_{\text{LoS}, m,d} + \sqrt{\frac{1}{1 + \kappa}} h_{\text{NLoS}, m,d},$$

where $h_0$ denotes the path loss at the reference distance $d = 1m$, $d_{m,d}$ is the distance between IoTD $m$ and $D$, $\alpha \geq 2$ is the path loss exponent, and $\kappa$ is the Rician factor. $h_{\text{LoS}, m,d}$ is the deterministic LoS component, $h_{\text{NLoS}, m,d}$ is the non-LoS (NLoS) component which is modeled as Rayleigh fading. The channels from IoTD $m$ to STAR-RIS, from IoTD $n$ to STAR-RIS and from STAR-RIS to $D$ are denoted by $g_{m,s}$, $g_{n,s}$, and $g_{s,d}$ respectively, which can be generated by Rician fading mode similarly. Therefore, we can obtain the equivalent cascaded $m$-RIS-D channel and $n$-RIS-D channel as

$$c_m = h_{m,d} + g_{m,s} \Theta_r g_{s,d},$$

$$c_n = g_{n,s} \Theta_t g_{s,d}. \hspace{1cm} (5a) \hspace{1cm} (5b)$$

With given $(m, n)$, the received superimposed signals at $D$ can be given by

$$y = c_m \sqrt{p_{m}} s_m + c_n \sqrt{p_{n}} s_n + \omega, \hspace{1cm} (6)$$

where $\omega \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) at $D$. The signal-to-interference plus noise ratio (SINR) for decoding $s_z, z \in \{m, n\}$ at $D$ is

$$\gamma_z = \frac{\rho_z |c_z|^2}{\lambda_z p_z |c_z|^2 + \rho_z (1 - \lambda_z) p_z |c_z|^2 + \sigma^2}. \hspace{1cm} (7)$$

If $z = m$, then $\bar{z} = n$, else $\bar{z} = m$. $\lambda_z \in \{0, 1\}$ indicates the decoding order of $s_m$ and $s_n$ at $D$, $\lambda_m + \lambda_n = 1$. If $p_m |c_m|^2 \geq p_n |c_n|^2$, then $\lambda_m = 1$, $\lambda_n = 0$, else $\lambda_m = 0$, $\lambda_n = 1$. $\theta_z \in [0, 1]$ denotes the residual interference factor caused by imperfect SIC. Specifically, $\theta_z = 0$ means perfect interference cancellation, $\theta_z = 1$ means the failure of interference cancellation.\footnote{We can observe from expression (7) that the number of devices sharing a same resource block is restricted in the considered system, and the BS only performs interference cancellation once for each NOMA cluster, which will not cause excessive accumulated residual interference. Moreover, with the development of SIC technology and the improvement of hardware capabilities, the value of residual interference $\theta_{\text{R}}$ can be extremely small. Therefore, to evaluate upper bound of the FBL transmission performance, we consider perfect SIC in this paper, which means $\theta_{\text{R}} = 0$ in (7). We might extend our work to imperfect SIC in the future.}

C. Performance Metrics for FBL Transmission

FBL transmission is considered in the IoT scenario. The data generated by IoT devices usually only contains 20-250 bytes \cite{36}, so FBL will increase the proportion of intended data in the transmitted packet, thus improving the spectral efficiency. Furthermore, FBL transmission is an effective technology to reduce the transmission delay and processing delay, thus realizing the high timeliness of information for delay-sensitive IoT applications. However, FBL has a negative effect on the channel coding gain. In this regard, Shannon’s theory based on an infinite blocklength (IFBL) assumption will overestimate the achievable rate and underestimate the error probability of the systems operating under FBL constraints. Therefore, aiming at satisfying the communication requirements with FBL constraints, the following performance metrics are considered.

1) Achievable FBL rate: Considering the given packet length $L_m$ and tolerable decoding error probability $\epsilon_m$, the maximal rate for IoTD $m$ transmitting the message to $D$ reliably can be characterized by the achievable FBL rate, given by \cite{37}

$$R_m = \log_2 \left( 1 + \gamma_m \right) - \kappa_m \sqrt{1 - \frac{1}{(1 + \gamma_m)^2}}, \hspace{1cm} (8)$$

where $\kappa_m = Q^{-1}(\epsilon_m)$ is a constant value when the tolerable decoding error probability $\epsilon_m$ and blocklength $L_m$ are given, $Q^{-1}(\cdot)$ denotes the inverse of Gaussian Q-function $Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

2) Decoding error probability: For IoTD $n$ on the $T$-Space, with a given packet length $L_m$ and target rate $R_{n, \text{target}}$, the decoding error probability can be calculated as

$$\epsilon_n = Q \left( \frac{\log_2 \left( 1 + \gamma_n \right) - R_{n, \text{target}} \sqrt{L_m \ln 2}}{\sqrt{1 - \frac{1}{(1 + \gamma_n)^2}}} \right). \hspace{1cm} (9)$$

Note that for the IoTDs on the same subchannel, the equation $L_m = L_m$ should hold. In this paper, we aim to maximize the sum FBL rate of R-IoTDs while guaranteeing the decoding error constraints of T-IoTDs, namely, high rate and low error (HRLE).

III. HRLE Design with ES Protocol

In this section, we will propose an alternating optimization algorithm to realize high-rate for R-IoTDs and low-error for T-IoTDs for STAR-RIS with ES protocol. Specifically, HRLE
will be achieved through jointly optimizing device pairing, the transmission power of IoTDs, and the amplitude response and phase shift coefficient of STAR-RIS, with the mathematical form given by

\[ \text{(P1)}: \max_{\{x_p, p_r, \Theta_u, \Theta_n\}} \sum_{m \in \mathcal{M}, n \in \mathcal{N}} x_{m,n} R_m, \] (10a)\]

subject to
\[ x_{m,n} \in \{0, 1\}, \forall m \in \mathcal{M}, n \in \mathcal{N}, \] (10b)\]
\[ \sum_{m \in \mathcal{M}} x_{m,n} \leq 1, \sum_{n \in \mathcal{N}} x_{m,n} \leq 1, \] (10c)\]
\[ \lambda_m = 1, \lambda_n = 0, \text{if } p_m |c_m|^2 \geq p_n |c_n|^2, \]
\[ \lambda_m = 0, \lambda_n = 1, \text{if } p_m |c_m|^2 < p_n |c_n|^2, \] (10d)\]
\[ p_z \leq P_z, \forall z \in \{r, t\}, \] (10e)\]
\[ \beta_k^r, \beta_k^t \in [0,1], \beta_k^r + \beta_k^t = 1, \forall k \in \mathcal{K}, \] (10f)\]
\[ \theta_k^r, \theta_k^t \in [0, 2\pi], \forall k \in \mathcal{K}, \] (10g)\]
\[ \varepsilon_n \leq \varepsilon_{n,\max}, \forall n \in \mathcal{N}, \] (10h)\]

where (10b)-(10e) are the constraints for device pairing, (10f) is the decoding order constraint, (10e) is the maximum transmission rate of IoTDs, (10f) is the amplitude constraints for STAR-RIS with ES protocol, (10g) is the phase shift coefficient constraint of STAR-RIS, and (10h) is the maximum tolerable decoding error probability of IoTDs on the T-Space.

The proposed optimization problem is intractable due to the high coupling of optimization variables and the non-convexity of the objective function and the constraints. Furthermore, the device pairing index \( x_{m,n} \) can only be 0 or 1, which makes the proposed optimization problem a tough mixed integer programming (MIP) problem. In this section, we propose an alternating optimization based algorithm to solve (10) in an iterative manner.

### A. Received-Signal-Strength-Based Device Pairing Scheme

In this subsection, we propose an algorithm to solve the device pairing problem with the given transmission power of IoTDs, and amplitude response and phase shift coefficient of STAR-RIS. The optimality proof of the proposed scheme is given under the condition of high SINR. Under such circumstances, the product of transmission power and equivalent cascaded channel gain, \( p_z |c_z|^2, z \in \{r, t\} \), is determined values. Actually, \( p_z |c_z|^2 \) is the received signal strength of \( s_z \) at \( D \). The device pairing problem can be written as

\[ \text{(P1.1)}: \max_{\{x_p, p_r, \Theta_u, \Theta_n\}} \sum_{m \in \mathcal{M}, n \in \mathcal{N}} x_{m,n} R_m, \] (11a)\]

subject to
\[ (10b)-(10d). \] (11b)\]

### Remark 1
If \( D \) first decodes IoTD \( m \)'s signal and then \( D \) decodes IoTD \( n \)'s signal, the device pairing scheme will not have an effect on the achievable rate of IoTD \( m \) since \( s_n \) as interference has been eliminated according to the SIC principle.

### Theorem 1
Without loss of generality, we consider that
\[ p_1 |c_1|^2 \geq p_2 |c_2|^2 \geq \cdots \geq p_M |c_M|^2; \]
\[ p_1 |c_1|^2 \geq p_2 |c_2|^2 \geq \cdots \geq p_N |c_N|^2. \] Under the high SINR conditions, \( p_1 |c_1|^2 \geq p_2 |c_2|^2 \geq \cdots \geq p_M |c_M|^2 \), and \( p_1 |c_1|^2 \geq p_2 |c_2|^2 \geq \cdots \geq p_N |c_N|^2 \), the optimization goal is to maximize the sum transmission rate of IoTDs on the R-Space, and the decoding order will have a significant impact on the rate calculation. Therefore, according to the different decoding orders, we will discuss the power allocation scheme in the following two cases.

---

**Algorithm 1 RSS-Based Device Pairing Scheme**

**Input:** \( \mathcal{M}_u, \mathcal{N}_u. \)

**Output:** \( x^*. \)

1. repeat
   2. Find \( m = \max_{m \in \mathcal{M}_u} p_m |c_m|^2, n = \min_{n \in \mathcal{N}_u} p_n |c_n|^2; \)
   3. if \( p_m |c_m|^2 \geq p_n |c_n|^2 \) then
   4. Allocate IoTD \( m \) and IoTD \( n \) to the same pair;
   5. Let \( x_{m,n} = 1, \mathcal{M}_u = \mathcal{M}_u \setminus m, \mathcal{N}_u = \mathcal{N}_u \setminus n; \)
   6. \( \lambda_m = 1, \lambda_n = 0; \)
   7. else
   8. Find \( m = \min_{m \in \mathcal{M}_u} p_m |c_m|^2; \)
   9. Allocate IoTD \( m \) and IoTD \( n \) to the same pair;
   10. Let \( x_{m,n} = 1, \mathcal{M}_u = \mathcal{M}_u \setminus m, \mathcal{N}_u = \mathcal{N}_u \setminus n; \)
   11. \( \lambda_m = 0, \lambda_n = 1; \)
   12. end if
   13. until \( \mathcal{M}_u = \emptyset, \mathcal{N}_u = \emptyset. \)

If \( D \) first decodes IoTD \( m \)'s signal and then decodes IoTD \( n \)'s signal, the optimal device pairing scheme that maximizes the sum transmission rate of IoTDs on R-Space is that IoTD \( m \in \mathcal{M}_u \) with maximum \( p_m |c_m|^2 \) and IoTD \( n \in \mathcal{N}_u \) with minimum \( p_n |c_n|^2 \) are assigned to the same cluster. That is, the optimal device pairing scheme should be

\[ \Pi_{opt} = \{(1, r, N), (2, r, N - 1), \ldots, (M, 1)\}. \] (12)\]

**Proof.** See Appendix A.

---

**B. Power Allocation Algorithm**

In this subsection, we solve the uplink transmission power allocation problem under fixed device pairing, and STAR-RIS coefficient, which can be expressed as

\[ \text{(P1.2)}: \max_{\{p_r, p_t\}} \sum_{m \in \mathcal{M}, n \in \mathcal{N}} x_{m,n} R_m, \] (13a)\]

subject to
\[ (10d), (10e), (10h). \] (13b)\]

The optimization goal is to maximize the sum transmission rate of IoTDs on the R-Space, and the decoding order will have a significant impact on the rate calculation. Therefore, according to the different decoding orders, we will discuss the power allocation scheme in the following two cases.
Theorem 2. To maximize the sum transmission rate of IoTDs on the R-Space while satisfying the maximum decoding error constraints of IoTDs on the T-Space, the minimum SINR of IoTD $n$ should meet the following conditions,

$$
\gamma_n^* = \min \left\{ \frac{1}{\gamma_n} \left[ \log_2 \left( 1 + \gamma_n \right) - \frac{Q^{-1}(z_{n,max})}{\sqrt{c_m \ln 2}} \right] \right. 
- \frac{Q^{-1}(z_{n,max})}{\sqrt{c_m \ln 2}} \sqrt{1 - \frac{1}{(1+\gamma_n)^2}} - R_{n,target} = 0 \right\}.
$$

(14)

Case 1: When IoTD $m$ is the stronger user and IoTD $n$ is the weaker user, that is when the decoding order is $\lambda_m = 1$ and $\lambda_n = 0$, the closed-form expressions of optimal transmission power for IoTD $m$ and IoTD $n$ can be given by

$$
p_m^* = P_{m,max},
$$

(15a)

$$
p_n^* = \min \left\{ \frac{\gamma_n^* \sigma^2}{|c_n|^2}, P_{n,max} \right\}.
$$

(15b)

Proof. Case 1: If IoTD $m$ is the strong user, its signal is decoded by the BS first, where IoTD $n$’s signal is regarded as interferences. Therefore, in order to maximize the achievable rate of IoTD $m$, the transmission power of IoTD $m$ should be as large as possible, that is,

$$
p_m^* = P_{m,max}.
$$

(16a)

At the same time, the transmission power of IoTD $n$ should be the minimum value under the condition that the maximum decoding error probability constraint is met, and should not exceed its maximum power. Then we have

$$
p_n^* = \min \left\{ \frac{\gamma_n^* \sigma^2}{|c_n|^2}, P_{n,max} \right\}.
$$

(18)

Case 2: If IoTD $m$ is the weaker user, the IoTD $n$’s signal has been eliminated when the BS decodes $s_m$. Thus, IoTD $n$’s transmission power will not affect IoTD $m$’s achievable rate. However, to increase the achievable rate of IoTD $m$ by enhancing the transmission power while still meeting the error probability constraint of IoTD $n$, the transmission power of IoTD $n$ should be as large as possible, that is,

$$
p_n^* = P_{n,max}.
$$

(19)

Moreover, IoTD $m$’s power should be the maximum value that brings interference to IoTD $n$ but still meets IoTD $n$’s minimum SINR constraint, i.e.,

$$
\gamma_n^* = \frac{P_n^*|c_n|^2}{P_m|c_m|^2 + \sigma^2}.
$$

(20)

Though some algebraic transformation and substituting $p_n^* = P_{n,max}$ into (20), and taking the maximum power of IoTDs into consideration, we can obtain

$$
p_m^* = \min \left\{ \frac{P_{n,max}|c_n|^2}{\gamma_n^*|c_m|^2} - \frac{\sigma^2}{|c_m|^2}, P_{m,max} \right\}.
$$

(21)

Remark 2. In a single iteration, constraint (10h) may not be satisfied even if IoTD $n$ transmits signals with its maximum power, which can be caused by poor channel conditions or the unoptimized STAR-RIS coefficient. This problem can be solved by iterative optimization, and the optimization problem will eventually converge to the optimal solution meeting the constraints.

C. STAR-RIS Coefficient Optimization For ES

The STAR-RIS amplitude response and phase shift coefficient optimization problem will be solved in this subsection, with the given device pairing scheme and IoTDs’ transmission power. The mathematical form of STAR-RIS coefficient optimization can be written as

$$
(P1.3) : \max_{\{\Theta_r, \Theta_t\}} \sum_{m,n \in N} x_{m,n} R_m,
$$

s.t. (10d), (10f), (10g), (10h).

(22a)

Due to the non-convexity of the objective function and constraints, (P3) is a non-convex optimization problem. It can be easily proved that (8) is the difference between two monotonically increasing concave functions, which makes it difficult to judge its concavity or convexity. Utilizing first-Taylor Expansion, we rewrite the function $g(\gamma_m) = \sqrt{1 - \frac{1}{(1+\gamma_m)^2}}$ included in the second term of (8) into a linear function as follows,

$$
\tilde{g}(\gamma_m, \tilde{\gamma}_m) = g(\gamma_m) + g'(\gamma_m)(\gamma_m - \tilde{\gamma}_m),
$$

(23)

where $g'(\gamma_m) = \frac{1}{(1+\gamma_m)^3(1+\gamma_m)^{-1}}$ is the first derivative of $g(\gamma_m)$. $\tilde{\gamma}_m$ is the feasible point. Then the achievable transmission rate of IoTD $m$ can be given by

$$
\bar{R}_m = \log_2 (1 + \gamma_m) - \epsilon_m \left[ g(\gamma_m) + g'(\gamma_m)(\gamma_m - \tilde{\gamma}_m) \right],
$$

(24)

which is an increasing concave function with respect to $\gamma_m$. However, $\bar{R}_m$ is still non-concave to the optimization variables $\Theta_r$ and $\Theta_t$. Thus, we deal with this problem through the following methods.

Let

$$
\mathbf{u}_z = \text{Diag}(\Theta_z) = 
\begin{bmatrix}
\sqrt{E}\mathbf{e}^{j\Theta_{z1}}, & \cdots, & \sqrt{E}\mathbf{e}^{j\Theta_{zK}}
\end{bmatrix}^T,
$$

$z \in \{r, t\}$

denote the amplitude response and phase shift coefficient vector of the STAR-RIS. Introduce a variable

$$
\Phi_m = \text{diag}(g_{m,s_l}^H \mathbf{g}_{s,d}) \in \mathbb{C}^{K \times 1},
$$

then the cascaded channel from IoTD $m$ to the destination BS via STAR-RIS can be rewritten as $g_{m,s_l}^H \Theta_r \mathbf{g}_{s,d} = \mathbf{u}_r^H \Phi_m$. Thus, the equivalent combined channel gain can be expressed as

$$
|h_{m,d} + g_{m,s_l}^H \Theta_r \mathbf{g}_{s,d}|^2 = |h_{m,d} + \mathbf{u}_r^H \Phi_m|^2
$$

$$
= \mathbf{u}_r^H \Phi_m \Phi_m^H \mathbf{u}_r + \mathbf{u}_r^H \Phi_m h_{m,d} + h_{m,d} \mathbf{u}_r^H \Phi_m + |h_{m,d}|^2.
$$

(25)
Let $Q_m = \left[ \frac{\Phi_m \Phi_m^H}{h_{m,d} \Phi_m^H} \right] \in \mathbb{C}^{(K+1) \times (K+1)}$, $\mathcal{M}$, and $\tilde{u}_r = \left[ \begin{array}{c} u_r \\ 1 \end{array} \right] \in \mathbb{C}^{(K+1) \times 1}$, then we have

$$|h_{m,d} + g_m^H \Theta_g \mathbf{s}_d|^2 = \tilde{u}_r^H Q_m \tilde{u}_r + |h_{m,d}|^2.$$ Introduce another variable $U_r = \tilde{u}_r^H Q_m \tilde{u}_r \in \mathbb{C}^{(K+1) \times (K+1)}$, then we can get $\tilde{u}_r^H Q_m \tilde{u}_r = \text{Tr} \left( Q_m \tilde{u}_r \tilde{u}_r^H \right) = \text{Tr} \left( Q_m U_r \right)$, satisfying $U_r \succeq 0$ and rank $(U_r) = 1$, Thus

$$|c_m|^2 = |h_{m,d}|^2 + \text{Tr} \left( Q_m U_r \right). \quad (26)$$

Similarly,

$$|c_n|^2 = \text{Tr} \left( Q_n U_i \right), \quad (27)$$

where $Q_n = \Phi_n \Phi_n^H \in \mathbb{C}^{K \times K}$ with $\Phi_n = \text{diag} \left( \{ g_{n,s}^H \} \right), \mathbf{s}_d \in \mathbb{C}^{K \times 1}, U_i = \mathbf{u}_i \mathbf{u}_i^H \in \mathbb{C}^{K \times K}$ satisfying $U_i \succeq 0$ and rank $(U_i) = 1$. Furthermore, due to the coupling relationship between the amplitude coefficients of reflection and transmission, the following conditions should be met,

$$U_r(k, k) + U_i(k, k) = 1, \forall k \in K. \quad (28)$$

Accordingly, the received SINRs of IoT-D’s $m$’s and $n$’s signal at the BS can be respectively reshaped as

$$\gamma_m = \frac{p_m \text{Tr} \left( Q_m U_r \right)}{|h_{m,d}|^2 + \text{Tr} \left( Q_m U_r \right) + \sigma^2}, \quad (29a)$$

$$\gamma_n = \frac{p_n \text{Tr} \left( Q_n U_i \right)}{|h_{m,d}|^2 + \text{Tr} \left( Q_n U_i \right) + \sigma^2}. \quad (29b)$$

To handle the non-convexity of $\tilde{R}_m$, we introduce two slack variables $\alpha = [\alpha_1, \ldots, \alpha_m]^T$ and $\delta = [\delta_1, \ldots, \delta_n]^T$, denote $\tilde{\gamma} = [\tilde{\gamma}_1, \ldots, \tilde{\gamma}_m, \ldots, \tilde{\gamma}_n]^T$, and reformulate the objective function as the following form

$$f(\alpha, \delta, U_r, U_i, \tilde{\gamma}) = \sum_{m \in \mathcal{M}} \left\{ \log_2 \left( 1 + \kappa_m \right) - \kappa_m \left( g(\tilde{\gamma}_m) + g^* (\tilde{\gamma}_m) \right) \right\}, \quad (30)$$

where $\alpha$ and $\delta$ should satisfy the following constraints,

$$\alpha_m \leq \frac{p_m \text{Tr} \left( Q_m U_r \right)}{|h_{m,d}|^2 + \text{Tr} \left( Q_m U_r \right) + \sigma^2}, \quad (31a)$$

$$\delta_m \geq -\frac{p_n \text{Tr} \left( Q_n U_i \right) + \sigma^2}{\lambda_m \lambda_n \text{Tr} \left( Q_m U_r \right)}, \quad (31b)$$

**Lemma 1.** Given $\tilde{\gamma}$, $f(\alpha, \delta, U_r, U_i, \tilde{\gamma})$ is a tight lower bound of the original objective function $\sum_{m \in \mathcal{M}} R_m$.

**Proof.** Since $g(\gamma_m) = \sqrt{1 - \frac{1}{(1 + \gamma_m)^2}}$ is concave, based on the first-order condition for a concave function, we have $g(\gamma_m) + g^* (\gamma_m) \geq g(\gamma_m)$, which means $\log_2 (1 + \gamma_m) - \kappa_m (g(\gamma_m) + g^* (\gamma_m)) \geq \log_2 (1 + \gamma_m) - \kappa_m g(\gamma_m) = R_m$. It is easy to tell that the right side of inequality (31a) and (31b) is actually $\gamma_m$. Thus, when the constraints of (31a) and (31b) are satisfied, $f(\alpha_m, \delta_m, U_r, U_i, \tilde{\gamma}_m) \leq R_m$, i.e., $f(\alpha, \delta, U_r, U_i, \tilde{\gamma}) \leq \sum_{m \in \mathcal{M}} R_m$. Moreover, when $\gamma_m = \tilde{\gamma}_m$, $\alpha_m = \gamma_m$, and $\delta_m = \gamma_m$, we have $f(\alpha, \delta, U_r, U_i, \tilde{\gamma}) \geq \sum_{m \in \mathcal{M}} R_m$, which implies that $f(\alpha, \delta, U_r, U_i, \tilde{\gamma})$ is a tighter lower bound of $\sum_{m \in \mathcal{M}} R_m$.

Therefore, the original optimization problem (P3) can be rewritten as

$$\max_{\alpha, \delta, U_r, U_i} f(\alpha, \delta, U_r, U_i, \tilde{\gamma}), \quad (32a)$$

s.t. (10b), (28), (31),

$$\lambda_m = 1, \lambda_n = 0, \quad \text{if } p_m \text{Tr} \left( Q_n U_i \right) \leq p_m \left| h_{m,d} \right|^2 + \text{Tr} \left( Q_m U_r \right), \quad (32b)$$

$$\lambda_m = 0, \lambda_n = 1, \quad \text{if } p_n \text{Tr} \left( Q_m U_r \right) > p_m \left| h_{m,d} \right|^2 + \text{Tr} \left( Q_m U_r \right), \quad (32c)$$

$$U_z \succeq 0, \forall z \in \{ r, t \}, \quad (32d)$$

$$\text{rank} \left( U_z \right) = 1, z \in \{ r, t \}. \quad (32e)$$

The decoding error constraint (10b) is intractable due to the existence of Gaussian Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. Since the Gaussian Q-function is monotonically decreasing with $x, \varepsilon \leq Q(x)$ is equivalent to $x \geq Q^\top (\varepsilon)$. Thus, with given maximum tolerable decoding error probability $\varepsilon_{\text{max}}$ and targeted transmission rate $R_{\text{target}}$, we recast (10b) into the following form

$$\frac{\log_2 (1 + \gamma_m) - R_{\text{target}}}{\sqrt{1 - (1 + \gamma_m)^{-2}}} \geq Q^\top (\varepsilon_{\text{max}}) \triangleq \kappa_m, \quad (33)$$

which is equivalent to

$$\log_2 (1 + \gamma_m) - R_{\text{target}} \geq \kappa_m \sqrt{1 - \frac{1}{(1 + \gamma_m)^2}}. \quad (34)$$

To tackle the non-convexity of (34), we first construct approximate linear upper bounds of $\sqrt{1 - \frac{1}{(1 + \gamma_m)^2}}$ by utilizing first-order Taylor expansion

$$\mathcal{F}(\gamma_m, \bar{\gamma}_m) \triangleq \sqrt{1 - \frac{1}{(1 + \bar{\gamma}_m)^2}} = \frac{1}{(1 + \bar{\gamma}_m)^2} \left( 1 + \frac{1}{2} \left( \gamma_m - \bar{\gamma}_m \right) \right), \quad (35)$$

where $\bar{\gamma}_m$ is the optimal solution obtained in the previous iteration. Thus, we can rewrite (34) as

$$\log_2 (1 + \gamma_m) - R_{\text{target}} \geq \kappa_m \mathcal{F}(\gamma_m, \bar{\gamma}_m), \quad n \in \mathcal{N}. \quad (36)$$

Further, to handle the non-convexity of (36), we introduce the slack variables $\{ \epsilon_{n, u}, \epsilon_{n, d}, \theta_n \}$, which satisfy the following constraints

$$2^{n_{\text{max}}} \leq p_m \left| h_{m,d} \right|^2 + \text{Tr} \left( Q_m U_r \right) + \lambda_m p_n \text{Tr} \left( Q_n U_i \right) + \sigma^2, \quad (37a)$$

$$2^{n_{\text{max}}} \geq \lambda_m p_n \text{Tr} \left( Q_n U_i \right) + \sigma^2, \quad (37b)$$

$$\epsilon_{n, u} - \epsilon_{n, d} - R_{\text{target}} \geq \theta_n, \quad (37c)$$

$$\theta_n \geq \kappa_n \mathcal{F}(\gamma_m, \bar{\gamma}_m). \quad (37d)$$
Through some algebraic operations, we can rewrite (37d) into
\[ \gamma_n \leq \frac{\varrho_n}{\kappa_n} - (1 + \tilde{\gamma}_n) (\hat{\gamma}_n^2 + 2\tilde{\gamma}_n) + \tilde{\gamma}_n \leq \frac{\varrho_n}{\kappa_n} - \varpi, \]
which can be approximately transformed into the following linear form
\[ p_n \text{Tr} \left( Q_n U_R \right) \]
\[ \leq \frac{2}{\kappa_n} - \varpi \left[ \lambda_m p_m \left[ h_m d^2 + \text{Tr} \left( Q_m U_R \right) \right] + \sigma^2 \right] + \frac{1}{n_m} \left[ h_m d^2 + \text{Tr} \left( Q_m U_R \right) \right] + \sigma^2 \left( \varrho_n - \tilde{\gamma}_n \right) \]
\[ + \left( \frac{2}{\kappa_n} - \varpi \right) \lambda_m p_m Q^H \left( U_t - U_R \right). \] 
(39)
However, (37b) is still a non-convex constraint. Utilizing first-order Taylor expansion, we transform it as
\[ 2 \varpi + 2 \varpi \ln 2 \left( \varpi - \varpi \right) \geq \lambda_m p_m \text{Tr} \left( Q_n U_t \right) + \sigma^2. \] 
(40)
To this end, we have transformed the original non-convex constraint (10b) into several convex constraints, i.e., (37a), (37c), (39) and (40). It is noteworthy that to the existence of the unit rank constraints, rank \( \left( U_{z} \right) = 1, z \in \{ r, t \} \) in (P3.1), the optimization problem is still non-convex. Thus, by neglecting the rank one constraints, (P3.1) becomes a concave semi-definite programming (SDP) problem. In addition, by reshaping the constraints (31), we can obtain the reformulated optimization problem
\[ \max_{\left\{ \alpha, \delta, \left( U_{r} \right), \left( U_{t} \right), \left( U_{r,t} \right), \left( U_{r,t} \right) \right\}} f \left( \alpha, \delta, \left( U_{r} \right), \left( U_{t} \right), \left( \gamma \right) \right), \] 
(41a)
subject to
\[ \alpha_m \leq \lambda_m p_m \text{Tr} \left( Q_n U_t \right) + \sigma^2, \] 
(41b)
\[ p_m \left[ h_m d^2 + \text{Tr} \left( Q_m U_R \right) \right] \leq \delta_m \left( \lambda_m p_m Q^H \left( U_t - U_R \right) + \sigma^2 \right). \] 
(41c)
However, constraints (41b) and (41c) are still non-convex. Therefore, by utilizing the first-order Taylor expansion of multivariate functions, we can transform constraints (41b) and (41c) into the following linear formulations
\[ \hat{\alpha}_m \left( \lambda_m p_m \text{Tr} \left( Q_n U_t \right) + \sigma^2 \right) + \bar{\alpha}_m \lambda_m p_m Q^H \left( U_t - U_R \right) \]
\[ + \left( \lambda_m p_m \text{Tr} \left( Q_n U_t \right) + \sigma^2 \right) \left( \alpha_m - \hat{\alpha}_m \right) \leq p_m \left[ h_m d^2 + \text{Tr} \left( Q_m U_R \right) \right], \]
(42a)
\[ \hat{\delta}_m \left( \lambda_m p_m \text{Tr} \left( Q_n U_t \right) + \sigma^2 \right) + \bar{\delta}_m \lambda_m p_m Q^H \left( U_t - U_R \right) \]
\[ + \left( \lambda_m p_m \text{Tr} \left( Q_n U_t \right) + \sigma^2 \right) \left( \delta_m - \hat{\delta}_m \right). \] 
(42b)
Accordingly, (P3.2) is transformed into a standard concave problem, which can be efficiently solved via the well-known CVX tool. The rank-one constraint can be recovered by the Gaussian randomization method.

Algorithm 2 Proposed AO algorithm for STAR-RIS with ES Protocol

Input: \( x^*, \ p^*_r, \ p^*_t, \ \Theta^*_r, \ \Theta^*_t \), maximum iteration times \( J_{\text{max}} \), convergence threshold \( \xi \).
Output: \( x^*, \ p^*_r, \ p^*_t, \ \Theta^*_r, \ \Theta^*_t \).
1. Initialization: \( j = 1, x^{(1)}_r, \ p^{(1)}_r, \ p^{(1)}_t, \ \Theta^{(1)}_r, \ \Theta^{(1)}_t \).
2. Repeat
3. For given \( p^{(j)}_r, \ p^{(j)}_t, \ \Theta^{(j)}_r \), and \( \Theta^{(j)}_t \), update the device pairing scheme according to Algorithm 1, and obtain \( \Theta^{(j+1)} \).
4. For given \( x^{(j+1)} \), \( \Theta^{(j+1)}_r \) and \( \Theta^{(j+1)}_t \), calculate and update \( p^{(j+1)}_r, \ p^{(j+1)}_t \), according to (15) and (16);
5. For given \( x^{(j+1)} \), \( \Theta^{(j+1)}_r \), and \( \Theta^{(j+1)}_t \), solve the convex optimization problem (P3.2) in (41) to obtain \( U^{(j+1)}_r \) and \( U^{(j+1)}_t \);
6. Using the Gaussian randomization method to recover the optimal STAR-RIS coefficient \( \Theta^{(j+1)}_r \) and \( \Theta^{(j+1)}_t \) from \( U^{(j+1)}_r \) and \( U^{(j+1)}_t \);
7. Update \( \alpha^{(j)} = \alpha^{(j+1)} \), \( \delta^{(j)} = \delta^{(j+1)} \), \( \epsilon^{(j)}_d = \epsilon^{(j+1)}_d \), \( \varrho^{(j)} = \varrho^{(j+1)} \), \( \varpi^{(j)} = \varpi^{(j+1)} \), \( \gamma^{(j)} = \gamma^{(j+1)} \), \( \gamma^{(j)} = \gamma^{(j+1)} \), \( \gamma^{(j)} = \gamma^{(j+1)} \);
8. \( j \leftarrow j + 1 \);
9. Until \( j = J_{\text{max}} \) or \( \left| \sum_{m \in M} R^{(j)}_m - \sum_{m \in M} R^{(j-1)}_m \right| \leq \xi \).

Lemma 2. Denote \( \alpha^*, \delta^*, \ U_r^*, \ U_t^* \) and \( \xi \) are the optimal solutions to (P3.2), \( \sum_{m \in M} R^*_m \) are calculated with the obtained \( \gamma^* \), and \( \sum_{m \in M} R^*_m \) are calculated with \( \xi \). With given \( \xi \), we have \( \sum_{m \in M} R^*_m \geq \sum_{m \in M} R^*_m \).

Proof. Based on Lemma 1, we have \( \sum_{m \in M} R^*_m \geq f \left( \alpha^*, \delta^*, \ U_r^*, \ U_t^*, \ xi \right) \). When \( \alpha^* = \gamma \), and \( \delta^* = \xi \), we have \( f \left( \alpha^*, \delta^*, \ U_r^*, \ U_t^*, \ xi \right) = \sum_{m \in M} R^*_m \). Furthermore, since \( \alpha^*, \delta^*, \ U_r^* \) and \( U_t^* \) are the optimal solutions, we have \( f \left( \alpha^*, \delta^*, \ U_r^*, \ U_t^*, \ xi \right) \geq f \left( \alpha^*, \delta^*, \ U_r^*, \ U_t^*, \ xi \right) \). Therefore, \( \sum_{m \in M} R^*_m \geq \sum_{m \in M} R^*_m \). 

D. Overall Algorithm for ES

We propose an alternating optimization (AO) based algorithm to perform the joint optimization for maximizing the sum FBL rate of R-IoTDs, as summarized in Algorithm 2. Specifically, we first propose an RSS-based device pairing scheme, then obtain the closed-form solution of power allocation. The STAR-RIS reflecting and transmitting coefficient optimization problem is solved by transforming it into a convex problem. Finally, the solution to the initial joint optimization problem is obtained by an alternative and iterative method, where the initial point of each iteration is the solution in the previous iteration.

The device pairing scheme involves finding the maximum or minimum values of \( M \) and \( N \) (\( M = N \)), leading to the computational complexity of \( O(2M^2) \). The computational complexity of solving the power allocation problem is \( O(1) \) with the given device pair since the closed-form solution is obtained. Thus, the computational complexity of
power allocation for all device pairs can be computed as $O(2M)$. The computational complexity of solving the STAR-RIS coefficient optimization problem mainly comes from the SDP problem, which can be approximately calculated as $O\left(\log \left(\frac{1}{\epsilon} \right) \left(2K^{4.5} + 21MK^{3.5} \right) \right)$, where $\zeta > 0$ is the convergence accuracy of SDP [38, Theorem 3.12]. Therefore, the overall polynomial complexity of the proposed algorithm can be approximately calculated as $O\left(I_2 \left(2M^2 + \log \left(\frac{1}{\epsilon} \right) \left(2K^{4.5} + 21MK^{3.5} \right) \right) \right)$, with $I_2$ representing the number of iterations required for the convergence of Algorithm 2.

IV. HRLE DESIGN WITH MS PROTOCOL

In this section, we will solve the HRLE design problem for STAR-RIS with MS protocol. In contrast to ES scheme, the MS protocol only allows the amplitude response value of STAR-RIS to be 0 or 1 instead of any value from 0 to 1, which has the advantage of easy implementation. However, the binary constraint will also make the optimization problem intractable due to its non-convexity. The mathematical form of the HRLE problem with MS protocol can be expressed as

$$\max_{\{x, p, \Theta, \Theta_t, \Theta_i\}} \sum_{m \in M, n \in \mathcal{N}} x_{m,n} R_m,$$

subject to:

1. $10b - 10f, 10g, 10h$,
2. $\beta_{k}^{\ell}, \beta_{k}^{t} \in \{0, 1\}, \beta_{k}^{\ell} + \beta_{k}^{t} = 1, \forall k \in \mathcal{K},$

where (43c) is the amplitude constraint for STAR-RIS with MS protocol. By comparing (10) and (43), we can see that the difference only exists in the constraints on the STAR-RIS amplitude coefficient. Therefore, the proposed RSS-based device pairing scheme in Section III-A and the obtained closed-form solutions for power allocation in Section III-A are still applicable. Thus, we focus on addressing the STAR-RIS coefficient optimization problem in this section.

A. STAR-RIS Coefficient Optimization For MS

With given device pairs and transmission power, the transmitting and reflecting coefficient optimization problem for STAR-RIS with MS protocol can be given by

$$\max_{\{\Theta, \Theta_t, \Theta_i\}} \sum_{m \in M, n \in \mathcal{N}} x_{m,n} R_m,$$

subject to:

1. $10d, 10f, 10g, 10h$,
2. $\beta_{k}^{\ell}, \beta_{k}^{t} \in \{0, 1\}, \forall k \in \mathcal{K}.$

It is easy to tell that compared with the STAR-RIS coefficient optimization problem (22) with ES protocol, (44) involves the additional non-convex binary constraints (44c). Therefore, we only need to focus on dealing with this new difficulty since other non-convex terms can be handled by a similar method given in Section III-C. First, we transform the binary constraint (44c) into its equivalent form

$$\beta_{k}^{\ell} (\beta_{k}^{t} - 1) = 0, \forall z \in \{t, r\}, k \in \mathcal{K}.$$ (45)

By adopting the similar processing method of the objective function and other constraints in Section III-C, and introducing (45) as a penalty form into the objective function in (41), the reformulated optimization problem is

$$\max_{\{\alpha, \delta, U_r, U_t \}} f(\alpha, \delta, U_r, U_t) + \eta \sum_{z \in \{r, t\}} \sum_{k \in \mathcal{K}} \beta_{k}^{z} (\beta_{k}^{z} - 1),$$

subject to:

1. $28, 32c, 32d, 37a, 37c,$
2. $39, 40, 42a, 42b,$

where $\eta > 0$ is the penalty factor.

Lemma 3. The equality constraint (45) can be satisfied as $\eta \to \infty$.

Proof. Please refer to [39].

However, the resulting optimization problem (P3.1') is still non-convex. Thus, we construct the lower bound of the penalty form by utilizing the first-order Taylor expansion

$$\beta_{k}^{z} (\beta_{k}^{z} - 1) \geq \beta_{k}^{z} (\beta_{k}^{z} - 1)^2 + 2 \beta_{k}^{z} \left(\beta_{k}^{z} - \beta_{k}^{t} \right),$$

where $\beta_{k}^{z}$ is a feasible point in each iteration of the SCA method. By introducing (47) into the objective function of (46), we can obtain the reconstructed optimization problem

$$\max_{\{\alpha, \delta, U_r, U_t \}} f(\alpha, \delta, U_r, U_t) + \eta \sum_{z \in \{r, t\}} \sum_{k \in \mathcal{K}} \left(2 \beta_{k}^{z} - 1 \right) \beta_{k}^{z} (\beta_{k}^{z} - 1)^2,$$

subject to:

1. $28, 32c, 32d, 37a, 37c,$
2. $39, 40, 42a, 42b,$

which is convex and can be efficiently solved via standard convex problem solvers such as CVX.

B. Overall Algorithm for MS

To solve the sum FBL rate maximization problem for STAR-RIS with MS protocol, we propose a double-layer penalty-based (DLPB) algorithm as concluded in Algorithm 3. Specifically, the inner loop is AO with introduced penalty terms, where the device-pairing, power allocation and relaxed STAR-RIS coefficient are optimized in an alternative manner similar to Algorithm 2, while the outer loop is for the binary amplitude constraint recovery. Similar to the analysis given in Section III-D, the overall computational complexity can be calculated as $O\left(I_3 \left(I_2 \left(2M^2 + \log \left(\frac{1}{\epsilon} \right) \left(2K^{4.5} + 21MK^{3.5} \right) \right) \right) \right)$, where $I_3$ denotes the number of iterations required for the convergence of Algorithm 3.

C. Comparison of AO and DLPB

To illustrate the proposed algorithms clearly, we summarize them into schematic diagrams. The proposed AO algorithm for ES protocol is shown in Fig. 3 (a), where three sub-problems
Algorithm 3 Proposed DLPB Algorithm for STAR-RIS with MS Protocol

Input: $x, p_r^1, p_t^1, \Theta_r, \Theta_t, \text{maximum iteration times } J_{\text{max}}, L_{\text{max}}, \text{convergence threshold } \xi, \xi_2.

Output: $x^*, p_r^*, p_t^*, \Theta_r^*, \Theta_t^*$.

1. Initialization: $j = 1, l = 1, x^{(1)}, p_r^{(1)}, p_t^{(1)}, \Theta_r^{(1)}, \Theta_t^{(1)},\text{penalty factor } \eta, \text{discount factor } 0 < c < 1$.

2. repeat
3. Perform steps 2-9 in Algorithm 2. Specifically, step 5 becomes solving the problem (P3.2) in (48).
4. Update $x^{(j)}, p_r^{(j)}, p_t^{(j)}, \Theta_r^{(j)}, \Theta_t^{(j)}$ with the current solutions $x^{(j)}, p_r^{(j)}, p_t^{(j)}, \Theta_r^{(j)}, \Theta_t^{(j)}, j = 1$.
5. Update $\eta^{(l)} = \eta^{(l-1)}/c_l$.
6. $l \leftarrow l + 1$.
7. until $\max_{k,z} \beta_k^z(l) (\beta_z^z(l-1) - 1) - \max_{k,z} \beta_k^{z+1} (\beta_z^{z+1} - 1) \leq \xi_2$.

Fig. 3. Illustration for the proposed algorithms.

are solved alternatively until the convergence condition or maximum number of iterations is reached. The proposed DLPB algorithm for MS protocol is shown in Fig. 3 (b). Specifically, the inner layer of the DLPB algorithm is similar to the AO algorithm with introduced penalty terms, while the outer layer is for the binary constraint recovery by exploiting penalty-based optimization.

V. SIMULATION RESULTS

In this section, we evaluate the performance of our proposed AO algorithm and DLPB algorithm for the STAR-RIS-assisted uplink NOMA system via simulations. Consider a three-dimensional coordinate system, where the BS and the STAR-RIS are respectively located at (0, 0, 10) meter (m) and (40, 0, 5) m. The R-IoTDs are randomly located in a circle with a radius of 5 m centered on (35, 0, 0) m, and the T-IoTDs are randomly located in a circle with a radius of 3 m centered on (45, 0, 5) m. The simulation results are obtained through an average of more than 500 channel realizations. The parameters used in the simulations are given in Table II. A stringent tolerable decoding error probability, i.e., $10^{-3}$, is considered to meet the requirements of mission-critical IoT services [41]. Moreover, a relatively short blocklength, i.e., 200 channel uses, is considered to reduce the transmission latency while ensuring reliability to a certain degree. To evaluate the effectiveness of the proposed algorithms, we compare them with those schemes of information-directed branch-and-prune (IDBP) algorithm proposed in [42], conventional reflecting/transmitting-only RISs with $K/2$ elements (C-RIS), random amplitude and phase coefficient at STAR-RIS (Random-STARS), OMA scheme, fixed transmission power at IoTDs with their maximum power (Fixed-P).

Table II: Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>noise power $\sigma^2$ (dBm)</td>
<td>-10</td>
</tr>
<tr>
<td>path loss at the reference distance $d$ = 1 m</td>
<td>-30dB</td>
</tr>
<tr>
<td>path loss exponents $\alpha_{m,n}$, $\alpha_{o,n}$, $\alpha_{d,n}$</td>
<td>2.5</td>
</tr>
<tr>
<td>path loss exponents $\alpha_{m,n}$, $\alpha_{o,n}$, $\alpha_{d,n}$</td>
<td>4</td>
</tr>
<tr>
<td>Rician factor</td>
<td>3dB</td>
</tr>
<tr>
<td>number of STAR-RIS elements</td>
<td>20</td>
</tr>
<tr>
<td>number of IoTs on R/IO-Space</td>
<td>2</td>
</tr>
<tr>
<td>maximum transmission power of IoTs (dBm)</td>
<td>23</td>
</tr>
<tr>
<td>blocklength $L$ (channel uses)</td>
<td>200</td>
</tr>
<tr>
<td>tolerable decoding error probability $\epsilon$</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>convergence threshold $\xi$</td>
<td>$10^{-4}$</td>
</tr>
</tbody>
</table>

Fig. 4. Convergence performance of the AO algorithm for STAR-RIS with ES protocol.

3 In this paper, we focus on the beamforming design of small-scale STAR-RIS networks to reveal the potential of STAR-RIS-assisted uplink NOMA transmission. However, by partitioning the STAR-RIS into several tiles, a scalable algorithm can be utilized to handle the high computational complexity caused by massive elements of STAR-RIS [40], which can provide inspiration and guidance for our future works on large-scale STAR-RIS.
Fig. 5. Converged error probability of T-IoTDs obtained by the AO algorithm (rate $r = 0.1$).

Fig. 6. Convergence performance of the DLPB algorithm for STAR-RIS with MS protocol.

Fig. 7. The achievable sum rate of R-IoTDs versus number of STAR-RIS elements.

In Fig. 6, we evaluate the convergence performance of the proposed DLPB algorithm for STAR-RIS with MS protocol. Note that the depicted figure is the outer loop iteration, so the number of times required for convergence of the algorithm is not significantly more than that of ES protocol. Similarly, more iteration times are required for the algorithm convergence when $K = 30$ compared with $K = 20$, and the achievable rate of R-IoTDs will be improved by increasing the number of STAR-RIS elements or relaxing the target rate constraint of T-IoTDs. In addition, by comparing Fig. 6 with Fig. 4, we can tell that compared with ES protocol, there is a performance loss when the STAR-RIS adopts MS protocol, which is caused by the constrained amplitude coefficient.

The relationship between the achievable sum rate of R-IoTDs and the number of STAR-RIS elements is depicted in Fig. 7. It can be seen that the achievable sum rates obtained by the proposed AO algorithm and the IDBP algorithm are almost coincident. Since the effectiveness of IDBP algorithm has been verified by the simulation results in [42], where the performance obtained by IDBP is close to the exhaustive search with significantly reduced complexity, we can conclude that the proposed AO algorithm is also near-optimal. When the Fixed-P scheme is adopted, that is, when each IoTD sends signals at its maximum power, the sum rate almost does not change with the number of STAR-RIS elements. This shows the importance of the joint optimization and cooperation of various communication resources, otherwise, it is difficult to exploit the potential performance gains brought by more STAR-RIS elements. Except for the Fixed-P scheme, the sum rate of R-IoTDs obtained by all other schemes increases with the number of STAR-RIS elements due to the higher array gain. Compared with the conventional RISs, random STAR-RIS coefficient, OMA and Fixed-P schemes, the achievable sum rate of R-IoTDs obtained by the proposed AO algorithm for STAR-RIS with ES protocol increased by 11.75%, 23.65%, 85.96% and 226.56% respectively when $K = 24$, which verifies the effectiveness of the proposed scheme. To be specific, due to the fixed number of transmitting and reflecting elements, the conventional RISs cannot exploit the same degrees of freedom as STAR-RIS, leading to a worse ability to intelligently adjust the wireless radio environment.
Fig. 8. The achievable sum rate of R-IoTDs versus maximum power of IoTDs.

![Graph](image)

Fig. 9. The achievable sum rate of R-IoTDs versus blocklength.

![Graph](image)

Furthermore, Fig. 7 also shows that STAR-RIS with ES protocol can achieve a better transmission performance than that with MS protocol at the cost of implement complexity.

We compare the achievable sum rate of R-IoTDs versus the maximum power of IoTDs under different algorithms in Fig. 8. Regardless of the maximum transmission power of IoTDs, the proposed scheme almost coincides with IDBP algorithm, and is always superior to other comparison schemes. As expected, we can see that the achievable sum rate of R-IoTDs enhances with the increase of maximum transmission power of IoTDs, except for the Fixed-P scheme. In order to obtain better FBL transmission performance of R-IoTDs, the transmission power of T-IoTDs of other schemes will be selected as a minimum value that guarantees the decoding correctness constraints, no matter what the maximum power is. However, the Fixed-P scheme increases the transmission power of R-IoTDs and T-IoTDs at the same time, which results in a less obvious increase in the transmission rate of R-IoTDs due to higher interference. In addition, with the increase of transmission power, the performance gap between OMA scheme and the proposed scheme becomes more and more obvious, because NOMA can make better use of the difference in the received signal strength to obtain greater performance gain than OMA.

The impact of short-packet blocklength is depicted in Fig. 9. The proposed AO scheme for STAR-RIS with ES protocol always has the maximum sum rate, which confirms the superiority of the proposed scheme. Furthermore, the achievable sum rate of all schemes increases with the blocklength, and gradually approaches IFBL transmission. As shown in Fig. 9, the performance gap between FBL and the IFBL transmission is 4.91% when \( L = 100 \), and decreases to 0 when \( L = 10^5 \). Although the transmission scheme with IFBL can improve the achievable rate of the system, it will also cause greater end-to-end delay, which is unfriendly to latency-sensitive IoT services. It is not difficult to tell from Fig. 9 that the transmission rate reaches a relatively stable value when the blocklength reaches a certain level, and there is almost no use to increase the transmission rate by increasing the blocklength. The performance tradeoff between achievable transmission rate and latency encourages us to investigate the optimal blocklength, which may be our future research direction.

The FBL transmission performance of R-IoTDs under different T-IoTDs’ maximum decoding error constraints and required target rates is shown in Fig. 10. Except for OMA scheme, the achievable sum rates of R-IoTDs increase with T-IoTDs’ maximum decoding error constraint while decrease with the required target rate. More communication resources, including transmission power and STAR-RIS amplitude response, will be needed to satisfy the stricter T-IoTDs’ requirements (higher required target rate), leading to the decline of FBL transmission performance of R-IoTDs, while it is the opposite when the constraint is looser (higher tolerable decoding error probability). For OMA scheme, the change of the maximum decoding error constraint or the target rate of the T-IoTDs does not have a significant impact on the achievable sum rate of R-IoTDs, because the IoTDs on both sides of the STAR-RIS occupy different subchannels and will not have inter-device interferences. Furthermore, the proposed scheme always has the best transmission performance compared with all those schemes, confirming the effectiveness of adopting STAR-RIS and NOMA in FBL transmission systems.

VI. CONCLUSIONS

In this article, a STAR-RIS-assisted uplink NOMA transmission framework for FBL transmission was proposed, where the IoTDs on two sides of the STAR-RIS have different communication requirements. For improving the FBL trans-
mission performance of the R-IoTDs while satisfying the decoding reliability of the T-IoTDs, a sum rate maximization problem was formulated. For STAR-RIS with ES protocol, an AO algorithm was proposed to solve the highly-coupled non-convex optimization problem alternatively. For STAR-RIS with MS protocol, a DLBP algorithm was proposed to tackle the newly introduced binary amplitude constraints. The simulation results show that the proposed STAR-RIS framework outperforms the conventional transmitting/reflecting-only RISs. Furthermore, adopting NOMA can enhance the FBL transmission performance compared with the OMA schemes. In addition, the tolerable decoding error probabilities and the targeted transmission rates have a significant impact on the performance of FBL transmission, which inspires us to investigate the tradeoff between capacity, latency and reliability.

**APPENDIX A**

**Proof of Theorem 1**

According to Theorem 1, for any subsets \( \{w_1, w_2\} \) and \( \{w'_1, w'_2\} \) satisfying \( p_1|c_{1r}|^2 \geq p_2|c_{1r}|^2 \) and \( p_1|c_{1r}|^2 \geq p_2|c_{2r}|^2 \), the optimal device clustering scheme for maximizing sum rate of IoTDs on R-Space is the following

\[
\Pi_1 = \{(w_1, w'_1), (w_2, w'_2)\}. \quad (49)
\]

Accordingly, the sum rate of \( w_1 \) and \( w_2 \) can be given by

\[
R_{\text{sum}} = \log_2 \left(1 + \frac{p_1|c_{1r}|^2}{p_{2r}|c_{2r}|^2 + 1}\right) + \log_2 \left(1 + \frac{p_2|c_{1r}|^2}{p_1|c_{1r}|^2 + 1}\right) - \kappa_1 - \kappa_2.
\]

(50)

Assume the contrary, if \( \Pi_1 = \{(w_1, w'_2), (w_2, w'_1)\} \) is not the optimal set, then the device pairing scheme is given by

\[
\Pi_2 = \{(w_1, w'_1), (w_2, w'_2)\}. \quad (51)
\]

Accordingly, the sum transmission rate of \( w_1 \) and \( w_2 \) is

\[
R'_{\text{sum}} = \log_2 \left(1 + \frac{p_1|c_{1r}|^2}{p_{1r}|c_{1r}|^2 + 1}\right) + \log_2 \left(1 + \frac{p_2|c_{2r}|^2}{p_{2r}|c_{2r}|^2 + 1}\right) - \kappa_1 - \kappa_2.
\]

(52)

Under a high SINR regime, we consider that the short-packet loss terms can be regarded as constant values, i.e., \( \kappa_1 \) and \( \kappa_2 \). Then we have

\[
R_{\text{sum}} - R'_{\text{sum}} = \log_2 \left(\frac{p_1|c_{1r}|^2 + p_2|c_{2r}|^2 + 1}{p_{2r}|c_{2r}|^2 + 1}\right) + \log_2 \left(\frac{p_2|c_{2r}|^2 + p_{1r}|c_{1r}|^2 + 1}{p_{2r}|c_{2r}|^2 + 1}\right) - \log_2 \left(\frac{p_1|c_{1r}|^2 + p_{1r}|c_{1r}|^2 + 1}{p_{2r}|c_{2r}|^2 + 1}\right) - \log_2 \left(\frac{p_2|c_{2r}|^2 + p_{2r}|c_{2r}|^2 + 1}{p_{2r}|c_{2r}|^2 + 1}\right).
\]

(53)

To simplify writing, introduce variables 
\( a = p_1|c_{1r}|^2 + p_{2r}|c_{2r}|^2 + 1 \), 
\( b = p_2|c_{2r}|^2 + p_{1r}|c_{1r}|^2 + 1 \), and 
\( c = p_{1r}|c_{1r}|^2 + p_{2r}|c_{2r}|^2 + 1 \). Obviouly, we have \( a + b = c + d \). Furthermore,

\[
[\frac{a - b}{c - d}]
\]

(54)

where \( (1) \) holds due to \( p_1|c_{1r}|^2 - p_2|c_{2r}|^2 \geq 0 \) and \( (p_2|c_{2r}|^2 - p_{1r}|c_{1r}|^2 \geq 0 \), because if \( x \geq 0 \) and \( y \geq 0 \), then \( |x - y| \leq |x + y| \). Thus, we can get the following inequality

\[
(a + b)^2 - (a - b)^2 \geq (c + d)^2 - (c - d)^2 \quad \Rightarrow ab \geq cd.
\]

(55)

Therefore,

\[
R_{\text{sum}} - R'_{\text{sum}} = \log_2 \frac{ab}{cd} \geq 0,
\]

(56)

which means that the device pairing scheme

\[
\Pi_1 = \{(w_1, w'_1), (w_2, w'_2)\}
\]

is superior to

\[
\Pi_2 = \{(w_1, w'_1), (w_2, w'_2)\}.
\]

**REFERENCES**


