2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) | 979-8-3503-9318-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/SPAWC60668.2024.10694006

Beyond-Diagonal RIS Attacks on Physical Layer Key Generation

Haoyu Wang*, Josef Nossek†, A. Lee Swindlehurst*
*Center for Pervasive Communications and Computing, University of California Irvine, CA, USA.
† Department of Electrical and Computer Engineering, Technical University of Munich, Germany.
Email: haoyuw30@uci.edu, josef.a.nossek@tum.de, swindle@uci.edu.

***Abstract*—While reconfigurable intelligent surface (RIS) technology shows great promise for wireless communication, an adversary using such technology can threaten wireless performance. This paper explores an RIS-based attack on time-division duplex (TDD) based wireless systems that use channel reciprocity for physical layer key generation (PLKG). We demonstrate that deploying a non-reciprocal RIS with a non-symmetric "beyond diagonal" (BD) phase shift matrix can compromise channel reciprocity and thus break key consistency. The attack can be achieved without transmission of signal energy, channel state information (CSI), and synchronization with the legitimate system, and thus it is difficult to detect and counteract. We propose a physically consistent BD-RIS model and verify the impact of its attack on the secret key rate (SKR) of the legitimate system via simulations. Moreover, we provide a heuristic approach for optimizing the BD-RIS configuration to realize a more severe attack in cases where some partial knowledge of the channel state information is available. Our results demonstrate that such channel reciprocity attacks can significantly decrease the SKR of the legitimate system.**

***Index Terms*—channel reciprocity, passive jamming, reconfigurable intelligent surfaces, physical layer key generation.**

I. INTRODUCTION

Ensuring secure data transmission in wireless systems is vital to protect sensitive information from unauthorized access and eavesdropping. Traditional cryptographic techniques rely on exchanging keys through dedicated secure channels or employing computationally intensive algorithms, which may not always be practical or efficient, especially in dynamic or resource-constrained environments. Recently, physical layer security has emerged as a promising approach to address these challenges by leveraging the unique characteristics of the wireless channel for secure key generation [1].

Existing physical layer key generation (PLKG) schemes commonly rely on the assumption of channel reciprocity (CR) in time-division duplex (TDD) wireless systems, where the same frequency band is shared for both uplink (UL) and downlink (DL) transmissions, using dynamic allocation of the UL and DL time slots. CR-based PLKG capitalizes on the reciprocity of the wireless radio-frequency (RF) channel inherent in TDD systems to collaboratively generate shared secret keys directly from their channel responses without complex cryptographic protocols [2], [3]. This process involves estimating the UL and DL channel matrices, correlating them to identify common features, and extracting secure key material from the correlated channel coefficients. The resulting shared keys can be used to establish secure communication channels, authenticate devices, and encrypt data transmissions, providing a robust defense against various security threats, including eavesdropping and unauthorized access.

The advent of Reconfigurable Intelligent Surfaces (RIS) presents a novel avenue to further augment the performance of PLKG [4]. An RIS consists of a large number of elements with tunable electromagnetic responses [5]. When an RIS is present, the effective channel comprises both the direct link and the link induced by the RIS [4]. Thus, the efficacy of PLKG in this scenario is influenced by the reciprocity of the RIS-induced link, necessitating a symmetrical structure and circuitry for the RIS, with reflection properties that do not change with the angles of incidence and departure [6], [7]. The reciprocity of the overall channel could be destroyed by deploying a malicious RIS that does not satisfy these conditions [8]–[11].

The idea of using an RIS to compromise the reciprocity of wireless channels has been the study of recent research. In [8], [9], the RIS introduces pilot contamination by randomly changing the RIS phase shifts between the uplink pilots and the downlink data transmission. Since the RIS thus breaks the channel reciprocity, the precoder designed based on the uplink pilots is no longer optimal and the downlink performance is degraded. Similarly, [10], [11] discuss how to implement attacks on PLKG in TDD-based wireless system by rapidly changing the RIS behavior to break the channel reciprocity.

While the methods in [8]–[11] do not require CSI, the attacking RIS has to be synchronized with the legitimate system so that it knows when the uplink and downlink transmissions occur. To overcome this drawback, we introduce an alternative approach based on the use of an RIS that is physically constructed to be non-reciprocal without dynamically changing its response. This architecture is based on the so-called "beyond-diagonal" RIS (BD-RIS) concept discussed in [12], [13]. The elements of a standard BD-RIS are interconnected and allow for a flow of energy between the array elements, resulting in a non-diagonal reflection matrix – hence the name "BD." While not diagonal, the reflection matrix of a BD-RIS is still symmetric, and hence reciprocal. However, if non-reciprocal devices are added to the BD-RIS interconnections, a non-

This work was supported by the U.S. National Science Foundation (NSF) under grants ECCS-2030029 and CNS-2107182.

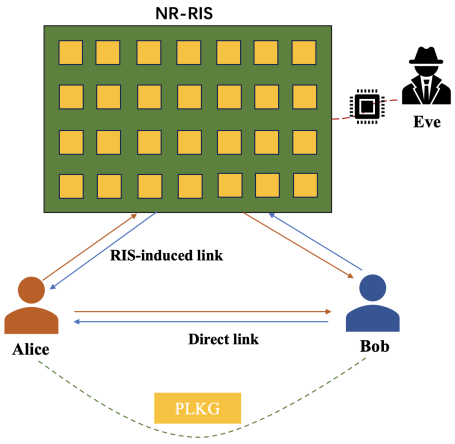

Fig. 1. Illustration of PLKG in a TDD communication system under a BD-RIS channel reciprocity attack during channel probing.

symmetric reflection matrix can be implemented, which leads to non-reciprocal (NR) RIS reflections. This novel approach is referred to as a Channel Reciprocity AttaCK (CRACK), and was proposed in [14]. CRACK based on an NR BD-RIS will inherently produce non-reciprocal uplink and downlink channels without the need for any synchronization nor CSI from the legitimate system, and thus is difficult to detect and defend against. In this paper, we investigate the impact of CRACK on the secret key rate of CR-based PLKG.

The outline of the paper is as follows. In Section II, we derive a mathematical model to show how a BD-RIS can be used to implement attacks on CR-based PLKG that sabotage the channel reciprocity and thus decrease the secret key rate. In particular, we propose a physically consistent BD-RIS architecture using RF circulators that produces a non-symmetric reflection matrix. In Section III, we derive the secret key rate of the legitimate system under CRACK and design heuristic approaches that determine the BD-RIS reflection matrix in cases where either no channel state information (CSI) is available, or when only knowledge of the line-of-sight (LoS) components of the RIS-cascaded channels is available. Section IV presents the results of several simulation studies that demonstrate the degradation that an NR BD-RIS can provide without CSI, and also how information about only the LoS components and statistical CSI can significantly increase the performance loss. Finally, Section V concludes the paper.

II. System Model

We consider PLKG for the communication system illustrated in Fig. 1, where Alice and Bob are two legitimate users deploying a standard TDD protocol and seeking to establish a shared cryptographic key from their channel observations. However, a malicious NR BD-RIS composed of N elements is present that aims to stealthily degrade the PLGK process. We use $h_{ab}, \mathbf{h}_{ar} \in \mathbb{C}^{N\times 1}$ and $\mathbf{h}_{rb} \in \mathbb{C}^{N\times 1}$, respectively, to denote the flat fading channels between Alice and Bob, between Alice and the BD-RIS, and between the BD-RIS and Bob. According to channel reciprocity, the corresponding

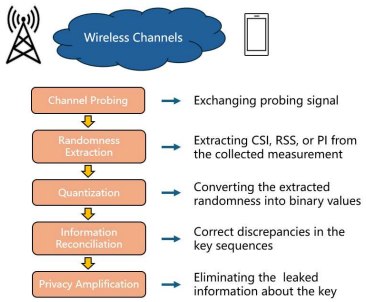

Fig. 2. Outline of the physical layer key generation process.

opposite-direction channels would be their transposes. Next, we will briefly introduce the process of CR-based PLKG, and then explain how a NR BD-RIS poses a severe threat to PLKG.

A. CR-based PLKG

CR-PLKG exploits the inherent randomness and reciprocity of the wireless channel between two communicating parties to generate a shared secret key. First, Alice and Bob exchange known probe signals, and then each party uses the received signal and the probe signal for CSI estimation. Assume that Alice and Bob transmit pilot signals at t_1 and t_2, respectively, so that the signals received at Bob and Alice can be written as

$$y_{ab}(t_1) = h_{ab}(t_1)s(t_1) + n_b(t_1) \tag{1}$$

$$y_{ba}(t_2) = h_{ba}(t_2)s(t_2) + n_a(t_2), \tag{2}$$

where $s(t_1)$ and $s(t_2)$ are the known probe signals and $n_b(t_1)$ and $n_a(t_2)$ respectively denote independent additive white Gaussian noise at Bob and Alice. Bob and Alice then implement channel estimation based on the received signals. In each channel probing round, the time difference $t_2 - t_1$ is smaller than the channel coherence time, guaranteeing a high correlation between the channel observations.

After channel probing, the estimated channels are converted into bit-strings that are appropriate for cryptographic key generation. The remaining operations are summarized in Fig. 2. This procedure aligns with established methods for PLKG and are therefore not specifically discussed in our work.

B. PLKG under BD-RIS CRACK

We assume a malicious attacker attempts to degrade CR-PLKG for the legitimate system by deploying an NR BD-RIS that destroys the reciprocity between the uplink and downlink channels. Due to the existence of the NR BD-RIS, the *actual* channels between Alice and Bob will be

$$h^*_{ab}(t_1) = \mathbf{h}^T_{ar}(t_1)(\mathbf{\Phi} - \mathbf{I})\mathbf{h}_{rb}(t_1) + h_{ab}(t_1), \tag{3}$$

$$h^*_{ba}(t_2) = \mathbf{h}^T_{rb}(t_2)(\mathbf{\Phi} - \mathbf{I})\mathbf{h}_{ar}(t_2) + h_{ba}(t_2), \tag{4}$$

where the identity matrix $\mathbf{I}$ is included in agreement with the physically consistent multiport network model for the RIS [15], [16]. Since the phase shift matrix $\mathbf{\Phi}$ for an NR BD-RIS is neither diagonal nor symmetric, we have $\mathbf{\Phi} \neq \mathbf{\Phi}^T$ [14], and thus the uplink and downlink channels will

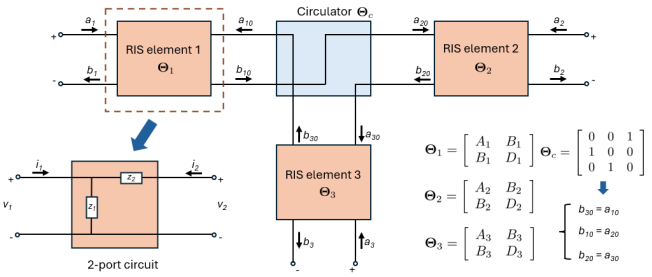

Fig. 3. Illustration of 3-element group-connected NR BD-RIS.

not be reciprocal since in general $\mathbf{h}_{ar}^T(t_1)(\mathbf{\Phi}-\mathbf{I})\mathbf{h}_{rb}(t_1) \neq \mathbf{h}_{rb}^T(t_2)(\mathbf{\Phi}-\mathbf{I})\mathbf{h}_{ar}(t_2)$. The reduced correlation between the channel observations will impact the randomness extraction needed for the subsequent key generation process.

For this paper, we will consider a particular NR BD-RIS implementation that employs a circulator together with a group-connected BD-RIS architecture with three elements per group, as depicted in Fig. 3. We assume a two-port model for each RIS element, and that each element i is composed of two tunable impedances Z_{1i} and Z_{2i}. For a reference impedance R, the scattering matrix for each element is given by

$$\mathbf{\Theta} = \frac{1}{\Delta}\begin{bmatrix} Z_1Z_2 - RZ_2 - R^2 & 2Z_1R \\ 2Z_1R & Z_1Z_2 + RZ_2 - R^2 \end{bmatrix}, \quad (5)$$

where $\Delta = Z_1Z_2 + 2Z_1R + Z_2R + R^2$ and for simplicity we have dropped the subscript on $\mathbf{\Theta}, Z_1$, and Z_2. A 3-port circulator connects the three elements as shown in the figure with a scattering matrix denoted by $\mathbf{\Theta}_c$. After some straightforward calculation, the 3×3 scattering (reflection) matrix for this 3-element group can be shown to satisfy Eq. (6). The full NR BD-RIS is assumed to be composed of G such groups, defined by $\mathbf{\Phi} = \text{diag}\{\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_g, \cdots, \mathbf{\Phi}_G\}$ and thus has a total of $N = 3G$ elements.

III. Secret Key Rate under BD-RIS CRACK

To evaluate the effect of the NR BD-RIS on PLKG, we analyze the secret key rate of the legitimate system introduced in Section II, for a fixed choice of $\mathbf{\Phi}$. The theoretical results reveal the impact of various system variables, including the number of reflecting elements at the ND-RIS, the RIS induced-LoS channels, and the path loss exponent. In the following section, we will explore the impact of these on the attack performance.

A. Assumed Channel Models

In the following theoretical analysis and the simulations in the next section, we will adopt the following common channel models: Rayleigh fading for the direct channel, and Rician fading for all RIS-related channels. Mathematically, the channel between Alice and Bob is given by

$$h_{ab} = \sqrt{\alpha_{ab}}\widetilde{h}_{ab}, \quad (7)$$

where $\alpha_{ab} = \rho d_{ab}^{-\iota_{ab}}$ is the distance-dependent large-scale path-loss, ρ is the path loss at the reference distance $d_0 = 1$m, d_{ab} is the distance between Alice and Bob, and ι_{ab} is the path loss exponent of the Alice-to-Bob link. The channel coefficient $\widetilde{h}_{ab} \in \mathbb{C}^{1\times 1}$ is drawn from a complex Gaussian distribution with zero mean and unit variance. The Rician channel $\mathbf{h}_{ar}$ between Alice and the BD-RIS is expressed as

$$\mathbf{h}_{ar} = \sqrt{\alpha_{ar}}\left(\sqrt{\frac{\kappa_{ar}}{1+\kappa_{ar}}}\overline{\mathbf{h}}_{ar} + \sqrt{\frac{1}{1+\kappa_{ar}}}\widetilde{\mathbf{h}}_{ar}\right), \quad (8)$$

where κ_{ar} is the Rician factor, $\overline{\mathbf{h}}_{ar}$ is the line-of-sight (LoS) channel component, and the elements of the non-LoS (NLoS) term $\widetilde{\mathbf{h}}_{ar}$ follow an i.i.d. complex Gaussian distribution with zero mean and unit variance. Similarly, the Rician Bob-RIS link is described as

$$\mathbf{h}_{rb} = \sqrt{\alpha_{rb}}\left(\sqrt{\frac{\kappa_{rb}}{1+\kappa_{rb}}}\overline{\mathbf{h}}_{rb} + \sqrt{\frac{1}{1+\kappa_{rb}}}\widetilde{\mathbf{h}}_{rb}\right). \quad (9)$$

To simplify the channel description, we will define $k_1 = \frac{k_{ar}}{1+k_{ar}}$, $k_2 = \frac{1}{1+k_{ar}}$, $k_3 = \frac{k_{rb}}{1+k_{rb}}$, and $k_4 = \frac{1}{1+k_{rb}}$.

B. Secret Key Rate Analysis

In general, the number of secure bits yielded by PLKG is given by the mutual information between the channels h_{ab}^* and h_{ba}^* conditioned on the eavesdropping channel h_e. This can be expressed as $R = I\left(h_{ab}^*; h_{ba}^* | h_{ae}; h_{be}\right)$, where h_{ae} and h_{be} are the channels observed by the eavesdropper. In practice, the distance from Eve to the legitimate user is likely much greater than the carrier wavelength, so we assume h_{ae} and h_{be} to be independent of the legitimate channels. Then the secret key rate (SKR) can be represented as [11]

$$\begin{aligned} R &= I\left(h_{ab}^*; h_{ba}^* | h_{ae}; h_{be}\right) \\ &= \log\left(\frac{K_{ab}K_{ba}}{\det\left(\mathbf{K}_{ab}\right)}\right), \end{aligned} \quad (10)$$

where K_{ab} and K_{ba} are respectively the covariances of h_{ab}^* and h_{ba}^*, and $\mathbf{K}_{ab}$ is the total correlation matrix given by

$$\mathbf{K}_{ab} = \mathbb{E}\begin{bmatrix} h_{ab}^* h_{ab}^{*H} & h_{ab}^* h_{ba}^{*H} \\ h_{ba}^* h_{ab}^{*H} & h_{ba}^* h_{ba}^{*H} \end{bmatrix}. \quad (11)$$

Subsequently, K_{ab} and K_{ba} are calculated as

$$\begin{aligned} K_{ab} &= \mathbb{E}\left\{h_{ab}^* h_{ab}^{*H}\right\} = \mathbb{E}\left\{\left|\mathbf{h}_{ar}^T(\mathbf{\Phi}-\mathbf{I})\mathbf{h}_{rb}\right|^2 + h_{ab}h_{ab}^H + \sigma^2\right\} \\ &= \alpha_{ar}\alpha_{rb}\left\{k_1k_3\left|\overline{\mathbf{h}}_{ar}^T(\mathbf{\Phi}-\mathbf{I})\overline{\mathbf{h}}_{rb}\right|^2 + k_1k_4\|\overline{\mathbf{h}}_{ar}^T(\mathbf{\Phi}-\mathbf{I})\|^2\right. \\ &\quad \left. + k_2k_3\|(\mathbf{\Phi}-\mathbf{I})\overline{\mathbf{h}}_{rb}\|^2 + k_2k_4\text{Tr}\left\{\mathbf{\Phi}^*\right\}\right\} + \alpha_{ab} + \sigma^2, \end{aligned} \quad (12)$$

and

$$\begin{aligned} K_{ba} &= \mathbb{E}\left\{h_{ba}^* h_{ba}^{*H}\right\} = \mathbb{E}\left\{\left|\mathbf{h}_{rb}^T(\mathbf{\Phi}-\mathbf{I})\mathbf{h}_{ar}\right|^2 + h_{ab}h_{ab}^H + \sigma^2\right\} \\ &= \alpha_{ar}\alpha_{rb}\left\{k_1k_3\left|\overline{\mathbf{h}}_{rb}^T(\mathbf{\Phi}-\mathbf{I})\overline{\mathbf{h}}_{ar}\right|^2 + k_1k_4\|(\mathbf{\Phi}-\mathbf{I})\overline{\mathbf{h}}_{ar}\|^2\right. \\ &\quad \left. + k_2k_3\|\overline{\mathbf{h}}_{rb}^T(\mathbf{\Phi}-\mathbf{I})\|^2 + k_2k_4\text{Tr}\left\{\mathbf{\Phi}^*\right\}\right\} + \alpha_{ab} + \sigma^2, \end{aligned} \quad (13)$$

where $\mathbf{\Phi}^* = (\mathbf{\Phi}-\mathbf{I})(\mathbf{\Phi}-\mathbf{I})^H$ and σ^2 is due to the channel estimation error [11].

$$\mathbf{\Phi}_g = \begin{bmatrix} A_1 + \frac{B_1^2 D_2 D_3}{1-D_1 D_2 D_3} & B_1 B_2 D_3 + \frac{B_1 B_2 D_1 D_2 D_3^2}{1-D_1 D_2 D_3} & B_1 B_3 + \frac{B_1 B_3 D_1 D_2 D_3}{1-D_1 D_2 D_3} \\ B_1 B_2 + \frac{B_1 B_2 D_1 D_2 D_3}{1-D_1 D_2 D_3} & A_2 + \frac{B_2^2 D_1 D_3}{1-D_1 D_2 D_3} & B_2 B_3 D_1 + \frac{B_2 B_3 D_1^2 D_2 D_3}{1-D_1 D_2 D_3} \\ B_1 B_3 D_2 + \frac{B_1 B_3 D_1 D_2^2 D_3}{1-D_1 D_2 D_3} & B_2 B_3 + \frac{B_2 B_3 D_1 D_2 D_3}{1-D_1 D_2 D_3} & A_3 + \frac{B_3^2 D_1 D_2}{1-D_1 D_2 D_3} \end{bmatrix}. \tag{6}$$

$$\begin{aligned} K_{C1} &= \mathbb{E}\left\{h_{ab}^* h_{ba}^{*H}\right\} = \mathbb{E}\left\{\mathbf{h}_{ar}^T(\mathbf{\Phi}-\mathbf{I})\mathbf{h}_{rb}\left(\mathbf{h}_{rb}^T(\mathbf{\Phi}-\mathbf{I})\mathbf{h}_{ar}\right)^H + h_{ab}h_{ab}^H\right\} \\ &= \mathbb{E}\left\{\mathbf{h}_{ar}^T(\mathbf{\Phi}-\mathbf{I})\mathbf{h}_{rb}\mathbf{h}_{rb}^H(\widehat{\mathbf{\Phi}}-\mathbf{I})\widehat{\mathbf{h}}_{ar}\right\} + \alpha_{ab} = \mathbb{E}\left\{\mathrm{Tr}\left\{(\mathbf{\Phi}-\mathbf{I})\mathbb{E}\left\{\mathbf{h}_{rb}\mathbf{h}_{rb}^H\right\}(\widehat{\mathbf{\Phi}}-\mathbf{I})\widehat{\mathbf{h}}_{ar}\mathbf{h}_{ar}^T\right\}\right\} + \alpha_{ab} \\ &= \alpha_{ar}\alpha_{rb}\mathrm{Tr}\left\{(\mathbf{\Phi}-\mathbf{I})\left(k_4\mathbf{I}+k_3\overline{\mathbf{h}}_{rb}\overline{\mathbf{h}}_{rb}^H\right)(\widehat{\mathbf{\Phi}}-\mathbf{I})\left(k_2\mathbf{I}+k_1\widehat{\overline{\mathbf{h}}}_{ar}\overline{\mathbf{h}}_{ar}^T\right)\right\} + \alpha_{ab}, \end{aligned} \tag{14}$$

$$K_{C2} = \alpha_{ar}\alpha_{rb}\mathrm{Tr}\left\{(\mathbf{\Phi}-\mathbf{I})\left(k_2\mathbf{I}+k_1\overline{\mathbf{h}}_{ar}\overline{\mathbf{h}}_{ar}^H\right)(\widehat{\mathbf{\Phi}}-\mathbf{I})\left(k_4\mathbf{I}+k_3\widehat{\overline{\mathbf{h}}}_{rb}\overline{\mathbf{h}}_{rb}^T\right)\right\} + \alpha_{ab}, \tag{15}$$

When calculating $\mathbf{K}_{ab}$, we denote the cross-terms as $K_{C1} = \mathbb{E}\left\{h_{ab}^* h_{ba}^{*H}\right\}$ and $K_{C2} = \mathbb{E}\left\{h_{ba}^* h_{ab}^{*H}\right\}$. The term K_{C1} can be calculated as in (14), where $\widehat{\mathbf{h}}$ is the conjugate of $\mathbf{h}$. Similarly, K_{C2} can be calculated as in (15). Substituting (12)-(15) into (10), we can obtain the exact expression for the secret key rate.

The expression derived above provides the secret key rate assuming a fixed NR BD-RIS configuration $\mathbf{\Phi}$. Later we will see in the simulation section that, even without any instantaneous CSI to design $\mathbf{\Phi}$, CRACK can significantly degrade the secret key rate of the legitimate system. On the other hand, if the NR BD-RIS possesses statistical CSI and information about the LoS components of the RIS-induced links, the choice of $\mathbf{\Phi}$ can be optimized in order to inflict greater performance loss.

IV. Numerical Results

This section presents numerical results to assess the performance of the proposed BD-RIS CRACK approach introduced in Section II. We examine the secret key rate of the TDD-based system detailed in Section III. We assume the BD-RIS is comprised of a linear array with elements spaced one-half wavelength apart. Alice is located at the 3D coordinates (15m, 0m, 1.8m), Bob is located at (15m, 30m, 1.8m), and the NR BD-RIS is deployed at the location (0m, 25m, 5m), closer to Bob. The path-loss exponents of the Alice-RIS and Bob-RIS links are set as $\iota_{a,r} = 2.5$ and $\iota_{b,r} = 2$, respectively. Unless otherwise specified, the path-loss exponent of the Alice-Bob link is $\iota_{r,b} = 3$ and the transmit power of the probing signal is $P_k = 30$ dBm. Several different cases for the number of BD-RIS elements are investigated, as listed in Table I together with other simulation parameters. In particular, note that we assume that $R = 1$, $Z_{2i} = 0$, and that Z_{1i} is purely imaginary.

A. Benchmark Schemes

The proposed BD-RIS CRACK implementation will be compared with the following benchmark approaches:

* **Without BD-RIS**: No RIS is present.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Number of RIS elements N	[60,120,180,240,300,360]
Impedance Z_{1i}	[-5,5]
Path loss ρ	-20 dB
Noise power σ^2	3.98×10^{-10} W
Rician factor $\kappa_{a,r}$	3
Rician factor $\kappa_{b,r}$	3

* **BD-RIS CRACK**: The values of Z_{1i} are chosen randomly.
* **Heuristic Algorithm (HA)**: Assumes knowledge of the LoS component of the RIS channels, randomly generates 300 impedance values and chooses the values such that the resulting $\mathbf{\Phi}$ maximizes the difference between the uplink and downlink LoS BS-RIS-User channels $\beta|\overline{\mathbf{h}}_{a,r}^T(\mathbf{\Phi}-\mathbf{I})\overline{\mathbf{h}}_{r,b} - \overline{\mathbf{h}}_{r,b}^T(\mathbf{\Phi}-\mathbf{I})\overline{\mathbf{h}}_{a,r}|^2$, where $\beta = \frac{\alpha_{a,r}\alpha_{r,b}\kappa_{a,r}\kappa_{r,b}}{(1+\kappa_{a,r})(1+\kappa_{r,b})}$.
* **HA1**: Assumes no CSI, and chooses the impedances from a randomly generated set such that $\mathbf{\Phi}$ maximizes the difference between $\mathbf{\Phi}$ and $\mathbf{\Phi}^T$.
* **Non-Diagonal (ND)-RIS**: Assumes the idealized non-reciprocal model of [13], [14] where all energy entering one RIS element is channeled to another arbitrary element. This configuration is not possible using the physically motivated model of Fig. 3.

B. PLKG on CRACK

Here we study the average secret key rate of BD-RIS CRACK averaged over 3000 random BD-RIS channels. Fig. 4 shows the average SKR as a function of the number of RIS elements and emphasizes the impact of the RIS-User channel path loss on CRACK performance. As N increases, CRACK can dramatically reduce the SKR. A higher path loss leads to a reduced cascaded RIS channel gain, resulting in a weaker CRACK impact. With knowledge of the statistical CSI and LoS channels, the NR BD-RIS can be designed to realize a more severe attack than the random scheme and achieves

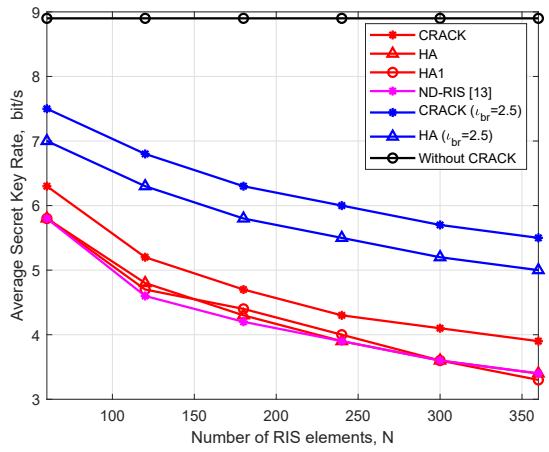

Fig. 4. Average secret key rate versus different number of RIS elements.

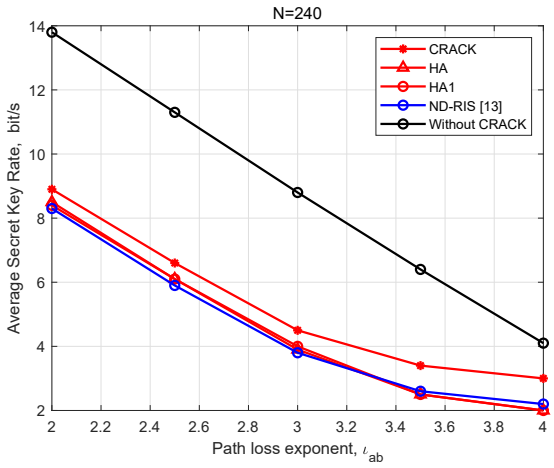

Fig. 5. Average secret key rate versus ι_{ab}.

a result close to that of the idealized ND-RIS. In addition, HA1 shows that the impact of the attack can also be enhanced by simply increasing the asymmetry of the reflecting matrix, without any CSI.

In Fig. 5, we show the average secret key rate versus the path loss of the direct channel h_{ab} for $N = 240$. As ι_{ab} increases, the SKR experiences a sharp decrease for all schemes that were tested. As expected, both HA and HA1 achieve a greater performance degradation than the Random approach and their performance is near to that of the idealized ND-RIS model. When the quality of h_{ab} is poor ($\iota_{ab} = 4$), BD-RIS CRACK can reduce the SKR to a very low level, posing a significant threat to the legitimate system.

V. Conclusions

In this study, we introduced a novel approach referred to as BD-RIS CRACK, to impair the physical layer key generation performance in TDD communication systems. CRACK operates using a non-reciprocal BD-RIS architecture that disrupts the presumed reciprocity of the legitimate channel. Such attacks can be executed without the need for CSI of the RIS-related links, and the attacks are effective without knowledge of the synchronization between the uplink and downlink of the legitimate system. We also proposed a physically consistent BD-RIS model involving a group connected BD-RIS architecture comprising three RIS elements connected through a circulator. Simulations demonstrate that this simple non-reciprocal BD-RIS architecture can achieve a significant reduction in the secret key rate of the legitimate system using simple heuristic methods with or without CSI.

References

[1] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, "Key generation from wireless channels: A review," *IEEE Access*, vol. 4, pp. 614–626, 2016.

[2] N. Patwari, J. Croft, S. Jana, and S. K. Kasera, "High-rate uncorrelated bit extraction for shared secret key generation from channel measurements," *IEEE Transactions on Mobile Computing*, vol. 9, no. 1, pp. 17–30, 2010.

[3] Y. Peng, P. Wang, W. Xiang, and Y. Li, "Secret key generation based on estimated channel state information for TDD-OFDM systems over fading channels," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5176–5186, 2017.

[4] G. Li, C. Sun, W. Xu, M. D. Renzo, and A. Hu, "On maximizing the sum secret key rate for reconfigurable intelligent surface-assisted multiuser systems," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 211–225, 2022.

[5] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, June 2021.

[6] W. Chen, L. Bai, W. Tang, S. Jin, W. X. Jiang, and T. J. Cui, "Angle-dependent phase shifter model for reconfigurable intelligent surfaces: Does the angle-reciprocity hold?" *IEEE Communications Letters*, vol. 24, no. 9, pp. 2060–2064, 2020.

[7] Z. Wei, B. Li, and W. Guo, "Adversarial reconfigurable intelligent surface against physical layer key generation," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2368–2381, 2023.

[8] X. Luo and H. Zhu, "IRS-based TDD reciprocity breaking for pilot decontamination in massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 102–106, Jan. 2021.

[9] H. Huang, Y. Zhang, H. Zhang, Y. Cai, A. L. Swindlehurst, and Z. Han, "Disco intelligent reflecting surfaces: Active channel aging for fully-passive jamming attack," *IEEE Transactions on Wireless Communications*, vol. 23, no. 1, pp. 806–819, 2024.

[10] L. Hu, G. Li, H. Luo, and A. Hu, "On the RIS manipulating attack and its countermeasures in physical-layer key generation," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–5.

[11] G. Li, P. Staat, H. Li, M. Heinrichs, C. Zenger, R. Kronberger, H. Elders-Boll, C. Paar, and A. Hu, "RIS-jamming: Breaking key consistency in channel reciprocity-based key generation," *arXiv preprint arXiv:2303.07015*, 2023.

[12] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2311–2324, 2023.

[13] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. M. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *IEEE Trans. Vehic. Tech.*, vol. 71, no. 6, pp. 6367–6383, June 2022.

[14] H. Wang, Z. Han, and A. L. Swindlehurst, "Channel reciprocity attacks using intelligent surfaces with Non-Diagonal phase shifts," *IEEE Open Journal of the Communications Society, Early Access*, pp. 1–1, 2024.

[15] J. A. Nossek, D. Semmler, M. Joham, and W. Utschick, "Modelling of wireless links with reconfigurable intelligent surfaces using multiport network analysis," in *Proc. 27th Int'l Workshop on Smart Antennas (WSA)*, 2024.

[16] ——, "Physically consistent modelling of wireless links with reconfigurable intelligent surfaces using multiport network analysis," *IEEE Wireless Communications Letters (to appear)*, 2024.