Defensive-RIS against Eavesdropping and False Data Injection relying on Non-Reciprocal Links

Chen Hu, Kun; Popovski, Petar

Published in:
2024 IEEE International Mediterranean Conference on Communications and Networking, MeditCom 2024

DOI (link to publication from Publisher):
10.1109/MeditCom61057.2024.10621367

Publication date:
2024

Document Version
Accepted author manuscript, peer reviewed version

Defensive-RIS against Eavesdropping and False Data Injection relying on Non-Reciprocal Links

Kun Chen-Hu and Petar Popovski
Department of Electronic Systems, Aalborg University, Denmark
E-mails: {kchenhu, petarp}@es.aau.es

***Abstract*—Reconfigurable Intelligent Surfaces (RISs) have excellent beam steering capabilities, which is applied to enhance wireless communication systems as a customizable signal reflector. Recently, these surfaces are used to disrupt the existing communication systems by introducing new types of vulnerability to the physical layer, commonly known as *RIS-In-The-Middle (RITM) attack*. An adversary uses RIS to replace the direct channel between two transceivers, and hence it can eavesdrop on all exchanged data by the legitimate users, but also perform a false data injection to the receiver. This work proposes anti-attack techniques based on a non-reciprocal channel produced by a defensive-RIS (D-RIS). The proposed precoding and combining methods and the channel estimation procedure for a non-reciprocal link are robust against potential adversaries while keeping the existing advantages of the RIS. We analyse the robustness of the system against attacks in terms of achievable secrecy rate and probability of detecting fake data. We believe that this defensive role of RIS can be a basis for new protocols and algorithms in the area.**

***Index Terms*—defensive, meta-surface, non-reciprocity, OFDM, RIS.**

I. Introduction

Reconfigurable intelligent surface (RIS) [1]–[4] is an artificial panel made of low-cost passive meta-materials that exhibit configurable electromagnetic properties. RIS has been extensively applied to enhance wireless communication systems, as it can transform traditional wireless networks into smart radio environments by manipulating the propagation environment. It can enlarge the coverage distance of the signal by providing an alternative propagation path with a better channel condition and/or enhance the data rate of the existing links.

On the flip side, these panels can be easily camouflaged in the environment, such that the RIS technology can be also adversely employed to disrupt the existing communication systems by introducing new types of vulnerability to the physical layer. *RIS-In-The-Middle (RITM) attack* [5]–[11] consists of using a RIS to divert and usurp an existing established communication link between two legitimate transceivers thanks to its advanced control and manipulation of electromagnetic waves at the physical layer. Note that, several works in the literature have assumed that the employed cryptographic system is typically not robust enough against the adversary. It has been shown experimentally [5], [6] that an adversarial RIS panel is not only able to eavesdrop on all the information exchanged by these legitimate transceivers, but it is also performing a false data injection, by reflecting either a corrupted or faked version of the received information. Moreover, no traces are left by the adversarial panel, making it impossible to be detected by the traditional existing procedures in the physical layer.

As a solution to partially alleviate these security issues, the deployment of defensive-RIS (D-RIS) is used to produce an alternative high-gain path. If the size of the D-RIS is significantly larger than the adversarial RIS, the signal manipulated by the adversary is eclipsed by the legitimate one. The theoretical analysis performance of this solution [7], [8] is given in terms of achievable secrecy rate (ASR). However, they cannot guarantee that the size of the legitimate RIS is going to be always larger and prevent data eavesdropping. Other works targeted to prevent the eavesdroppers by using narrow directive beams in the transmitter and D-RIS [9]–[11], and hence spatially filter the adversary. The joint optimization problem to compute these precoders is typically assuming the availability of the instantaneous [9] or statistical [10] channel state information (CSI) of the adversary, or the geographical localization of the adversary [11]. These may be hard to justify as adversaries can be easily camouflaged in the environment. Later, experiments showed that a mmWave system can be easily disrupted by the RITM attack [6], even using narrow beam widths. Moreover, in the hypothetical case that eavesdropping was solved, the adversary can still perform a false data injection to trick the receiver, and this issue, to the best of our knowledge, has not been addressed yet. Consequently, these new potential threats produced by the environmental manipulation at the RIS need to be carefully explored.

To the best of our knowledge, these security and privacy issues related to eavesdropping and false data injection by an unknown adversary have not been thoroughly addressed in previous research works. Therefore, the main objective of this paper consists of developing novel defensive techniques in the physical layer relying on the non-reciprocal channels produced by a D-RIS in the absence of any information related to the adversary. Since the direct channel between the legitimate devices is reciprocal, the adversary will replicate an alternative reciprocal one with better channel propagation conditions. Our key idea is to induce a non-reciprocal channel between the legitimate transceivers. This non-reciprocal channel response may be obtained in real-time by properly configuring the phase configurations, and the CSI of the DL and UL can be seen as a pair of keys for each D-RIS link. To exploit the non-reciprocity, both legitimate entities must properly precode the data stream to be transmitted and combine the received symbols using their unique combination of the CSI

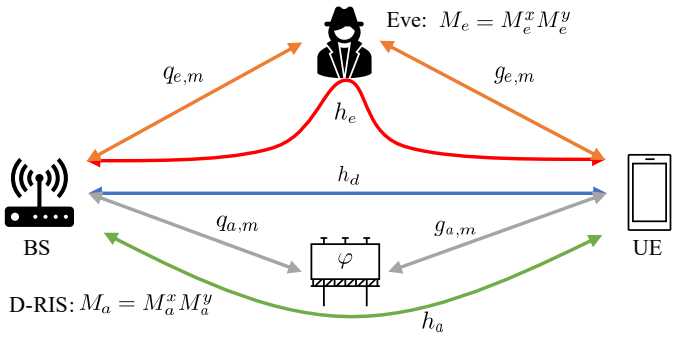

Fig. 1. Illustration of the considered system model. There is a legitimate link between BS and UE assisted by a D-RIS. Then, Eve is a third illegitimate entity, that is interested in jeopardizing the legitimate link.

of both links. Hence, any illegitimate intruder enhanced by an adversarial RIS cannot replicate this non-reciprocal channel since it does not have the secret CSI. The receiver can trust that the received information is sent by a legitimate entity, as it can filter out illegitimate transmissions and detect potential adversaries by using the CSI of the non-reciprocal channel.

The remainder of the paper is organized as follows. Section II introduces the system model of the considered D-RIS in a RITM attack scenario. Sections III provides the novel defensive technique based on a non-reciprocal link via D-RIS. Section IV analyses the system performance in terms of ASR and the probability of detecting fake symbol. Section V presents several numerical results for the proposed architecture, providing an assessment of the achieved performance. Finally, in Section VI, the conclusions are disclosed.

II. System Model of the Proposed Defensive Scenario under RITM Attack

A BS is serving a particular fixed UE of interest, both are equipped with a single omnidirectional antenna. This pattern corresponds to a pessimistic case as compared to [5]–[11] since any adversary can attack the legitimate communication link from any geographical position. A D-RIS is deployed to not only improve the quality of the channel between BS and UE by providing a new alternative one, but it also enhances the security by neutralizing potential adversaries, see Fig. 1. The D-RIS is equipped by a uniform planar array (UPA), whose number of elements is given by $M_a = M_a^x M_a^y$, where M_a^x and M_a^y are the number of elements in the x and y-axis, respectively. In addition to the legitimate entities (BS, UE and D-RIS), there is an adversary Eve, who is interested in jeopardizing the legitimate link between the BS and UE. According to [5], Eve is also equipped with another adversarial RIS, which is capable of manipulating the impinging signal and reflecting an alternative fake one. For simplicity, it is assumed that the hardware specifications related to the RIS of Eve are similar to the D-RIS, where the number of elements of the UPA is denoted by M_e. In this work, any information related to Eve is considered to be unknown by legitimate entities, such as geographical positions, RIS size, CSI, etc.

The chosen waveform corresponds to the well-known orthogonal frequency division multiplexing (OFDM), which is deployed in the current 5G [12]. It is considered that a slot is the minimum allocable resource in the system, which is built by K subcarriers and N consecutive OFDM symbols. A time-division duplexing (TDD) scheme is adopted [13], where the DL and UL transmissions are allocated in different OFDM symbols within one slot. Since all involved entities (BS, UE, D-RIS, and Eve) are fixed in a specific position within a delimited area, it is assumed that the coherence time of all reciprocal channels is long enough to cope with, at least, one slot. Moreover, it is considered that the cyclic prefix (CP) is long enough to absorb the multi-path effects of all channels. Consequently, without loss of generality and for the sake of the space, a particular subcarrier of the OFDM symbol, out of K, is taken into consideration.

At the receiver, after discarding the CP and performing the discrete Fourier transform (DFT), the received signal of the UE and BS at the n-th OFDM symbol and the subcarrier of interest can be modelled as

$$y_{u,n} = h_{r,n} x_{b,n} + w_{u,n}, \quad n \in \mathcal{N}^{\text{DL}}, \tag{1}$$

$$y_{b,n} = h_{r,n} x_{u,n} + w_{b,n}, \quad n \in \mathcal{N}^{\text{UL}}, \tag{2}$$

respectively, where $x_{b,n}$ and $x_{u,n}$ are the transmitted data symbols from the BS and UE at n-th OFDM symbol and the subcarrier of interest, respectively. $w_{u,n}$ and $w_{b,n}$ account for the additive white Gaussian noise (AWGN) of the UE and BS at n-th OFDM symbol and the subcarrier of interest, respectively, whose distribution follows $\mathcal{CN}\left(0, \sigma_w^2\right)$. $\mathcal{N}^{\text{DL}}$ and $\mathcal{N}^{\text{UL}}$ are the two subsets that contain the OFDM symbol indexes for the DL and UL transmissions within one slot, respectively. $h_{r,n}$ is the resulting channel frequency response between BS$\leftrightarrow$UE at n-th OFDM symbol and the subcarrier of interest, which can be expressed as

$$h_{r,n} = h_d + h_{a,n} + h_{e,n}, \quad n \in \mathcal{N}, \tag{3}$$

$$\mathcal{N} = \{0, 1, \cdots, N\} = \mathcal{N}^{\text{DL}} \cup \mathcal{N}^{\text{UL}}, \quad \mathcal{N}^{\text{DL}} \cap \mathcal{N}^{\text{UL}} = \emptyset, \tag{4}$$

where $\mathcal{N}$ is the full set that contains all the N OFDM symbol indexes. h_d corresponds to the frequency channel response of the direct link between the BS$\leftrightarrow$UE at the subcarrier of interest, whose distribution follows $\mathcal{CN}\left(0, \sigma_d^2\right)$. $h_{a,n}$ is the alternative cascaded channel frequency response [14] between the legitimate entities via D-RIS (BS$\leftrightarrow$D-RIS$\leftrightarrow$UE) at the n-th OFDM symbol and the subcarrier of interest, which can be modelled as

$$h_{a,n} = \sum_{m=1}^{M_a} \exp\left(j\varphi_{a,m,n}\right) q_{a,m} g_{a,m}, \quad n \in \mathcal{N}, \tag{5}$$

$$q_{a,m} \sim \mathcal{CN}\left(0, \sigma_{qa}^2\right), \quad g_{a,m} \sim \mathcal{CN}\left(0, \sigma_{ga}^2\right), \tag{6}$$

where $\varphi_{a,m,n}$ is the tunable phase response of the m-th element of the D-RIS at the n-th OFDM symbol, $q_{a,m}$ and $g_{a,m}$ are the reciprocal channel frequency response of BS$\leftrightarrow$D-RIS and D-RIS$\leftrightarrow$UE at n-th OFDM symbol and the interested subcarrier, respectively. Additionally, the channel coefficients among different passive elements of the RIS are considered

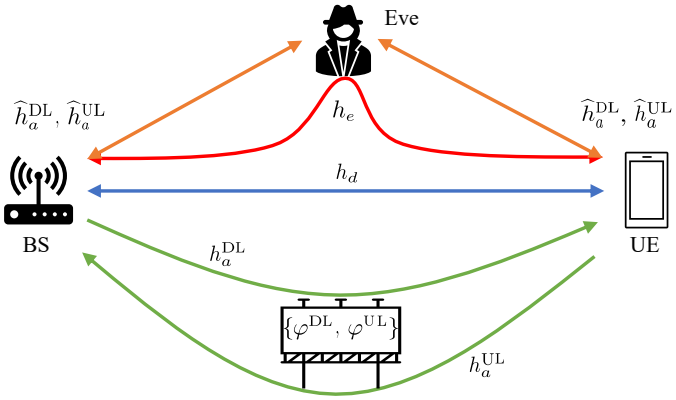

Fig. 2. Proposed non-reciprocal channel via D-RIS.

to be independently and identically distributed to simplify the notation. Analogously to D-RIS, h_e represents the cascaded frequency channel response between the legitimate entities via Eve (BS↔Eve↔UE), which can be also modelled by (5), replacing the sub-index a by e. Both RISs require a training period [1], [2] to find out their respective best phase configurations ($\varphi_{a,m,n}$ and $\varphi_{e,m,n}$), and ensure that their resulting channel gains are much higher than the direct one ($|h_{a,n}|, |h_{e,n}| >> |h_d|$).

The channel between the legitimate entities via D-RIS ($h_{a,n}$) can be turned into a non-reciprocal one (see Fig. 2) thanks to the proper configuration of the tunable phase configurations given in (5), which can be decomposed as

$$\varphi_{a,m,n} = \varphi_{a,m} + \varphi_{a,n}, \quad 1 \leq m \leq M_a, \quad 1 \leq n \leq N, \quad (7)$$

where $\varphi_{a,m}$ is the static phase at the m-th passive element, which is typically designed to create an alternative propagation path between the BS and UE [1]–[4], meanwhile $\varphi_{a,n}$ corresponds to the common dynamic phase which is the same for all passive elements of the D-RIS. This time-varying phase at the D-RIS can be tuned by the scheduler at the BS, where its value is changed according to the DL and UL OFDM symbol periods as

$$\varphi_{a,n} = \begin{cases} \varphi_a^{\mathrm{DL}} & n \in \mathcal{N}^{\mathrm{DL}} \\ \varphi_a^{\mathrm{UL}} & n \in \mathcal{N}^{\mathrm{UL}} \end{cases} \rightarrow h_{a,n} = \begin{cases} h_a^{\mathrm{DL}} & n \in \mathcal{N}^{\mathrm{DL}} \\ h_a^{\mathrm{UL}} & n \in \mathcal{N}^{\mathrm{UL}} \end{cases}, \quad (8)$$

where $\varphi_a^{\mathrm{DL}}, \varphi_a^{\mathrm{UL}}$ are the common time-varying phase configurations for the DL and UL, respectively, and $h_a^{\mathrm{DL}}, h_a^{\mathrm{UL}}$ account for the cascaded channel frequency response between the legitimate entities via D-RIS at the subcarrier of interested for the DL and UL, respectively. This non-reciprocal channel created by the D-RIS will be exploited to provide a reliable network between legitimate entities since it allows the detection of potential adversaries.

III. Proposed Defensive System against RITM Attack based on Non-Reciprocal Channel via RIS

A novel defensive system against RITM attack relying on a non-reciprocal channel via RIS is proposed, which is capable of solving all the relevant security and privacy issues, such as eavesdropping and false data injection. This no-reciprocity property makes the D-RIS path completely different from the existing reciprocal ones (direct and Eve's path), and hence this difference will ensure that legitimate entities will exchange their information through the desired D-RIS link, providing a reliable physical link to the higher layers.

The proposed defensive system assumes that both BS and UE have CSI of both DL and UL channels via D-RIS. At the transmitter, the legitimate entities are performing a precoding by combining the CSI of these two links. At the receiver, the effect of the precoder is partially mitigated by the channel itself, and the other part left should be compensated by the combiner at the receiver. The advantage of this procedure is twofold, on one hand the the precoder is more robust than since Eve requires even more time to find out the CSI of both links, making the eavesdropping process significantly harder. On the other hand, the combining checks the integrity of the received symbols and verifies that they are coming from the expected entity.

By assuming that it is assumed that the CSI is perfectly obtained to ease the notation. The precoders at the BS ($v_{b,n}$) and UE ($v_{u,n}$) can be computed as

$$v_{b,n} = v_{b,n}^{\mathrm{DL}} v_{b,n}^{\mathrm{UL}} = \left(h_a^{\mathrm{DL}}\right)^* \exp\left(j\theta_a^{\mathrm{UL}}\right), \quad n \in \mathcal{N}^{\mathrm{DL}}, \quad (9)$$

$$v_{u,n} = v_{u,n}^{\mathrm{UL}} v_{u,n}^{\mathrm{DL}} = \left(h_a^{\mathrm{UL}}\right)^* \exp\left(j\theta_a^{\mathrm{DL}}\right), \quad n \in \mathcal{N}^{\mathrm{UL}}, \quad (10)$$

$$\theta_a^{\mathrm{UL}} = \angle\left(h_a^{\mathrm{UL}}\right), \quad \theta_a^{\mathrm{DL}} = \angle\left(h_a^{\mathrm{DL}}\right), \quad (11)$$

respectively, where $v_{b,n}^{\mathrm{DL}}$ and $v_{b,n}^{\mathrm{UL}}$ are the precoders at the BS obtained from the CSI of DL and UL, respectively. $v_{u,n}^{\mathrm{UL}}$ and $v_{u,n}^{\mathrm{DL}}$ correspond to the precoders at the UE obtained from the CSI of UL and DL, respectively. θ_a^{DL} and θ_a^{UL} account for the phase components of the D-RIS channel in the DL and UL, respectively. Note that (9)-(11) correspond to a modified version of the well-known MRT precoding [15], where an additional phase ($\theta_a^{\mathrm{UL}}, \theta_a^{\mathrm{DL}}$) is embedded in the precoders of the BS and UE ($v_{b,n}$ and $v_{u,n}$, respectively).

Taking into consideration (9)-(11), the received symbols at the UE and BS, given in (1) and (2) respectively, can be rewritten as

$$y_{u,n} = \left|h_a^{\mathrm{DL}}\right|^2 v_{b,n}^{\mathrm{UL}} x_{b,n} + \left|h_{e,n}\right|^2 x_{e,b,n} + w_u, \quad n \in \mathcal{N}^{\mathrm{DL}}, \quad (12)$$

$$y_{b,n} = \left|h_a^{\mathrm{UL}}\right|^2 v_{u,n}^{\mathrm{DL}} x_{u,n} + \left|h_{e,n}\right|^2 x_{e,u,n} + w_b, \quad n \in \mathcal{N}^{\mathrm{UL}}, \quad (13)$$

respectively. Before performing the symbol decision, a combining based on phase rotation is performed to the received signals at the UE and BS as

$$z_{u,n} = \exp\left(-j\theta_a^{\mathrm{UL}}\right) y_{u,n} = \left|h_a^{\mathrm{DL}}\right|^2 x_{b,n} + \exp\left(-j\theta_a^{\mathrm{UL}}\right)\left(\left|h_{e,n}\right|^2 x_{e,b,n} + w_{u,n}\right), \quad n \in \mathcal{N}^{\mathrm{DL}}, \quad (14)$$

$$z_{b,n} = \exp\left(-j\theta_a^{\mathrm{DL}}\right) y_{b,n} = \left|h_a^{\mathrm{UL}}\right|^2 x_{u,n} + \exp\left(-j\theta_a^{\mathrm{DL}}\right)\left(\left|h_{e,n}\right|^2 x_{e,u,n} + w_{b,n}\right), \quad n \in \mathcal{N}^{\mathrm{UL}}, \quad (15)$$

respectively. Note that the combining corresponds to a phase rotation which does not enhance the noise and interference.

The proposed precoders, given in (9)-(11), make the eavesdropping process more difficult since Eve must search two different precoders, enlarging the execution time of the optimization algorithm. Moreover, the use of additional terms $v_{b,n}^{\mathrm{UL}}$ and $v_{u,n}^{\mathrm{DL}}$ in the computation of precoding matrices will force the receiver to perform an additional combining, given in (14)-(15). This will allow the receiver not only to check the identity of the transmitter but also will avoid the demodulation of the injected false data stream reflected by Eve ($x_{e,b,n}$ and $x_{e,u,n}$) since the combiner is rotating these fake symbols, regardless the amount of signal strength of both original and fake data streams. Consequently, the false data injection by Eve is even more difficult than the eavesdropping process, since Eve not only has to discover the two precoders ($v_{b,n}$ and $v_{u,n}$), but also must find out the terms $v_{b,n}^{\mathrm{UL}}$ and $v_{u,n}^{\mathrm{DL}}$, which is additionally enhancing the difficulty of the optimization problem.

IV. Analysis of the Achievable Secrecy Rate (ASR) and Probability of Fake Symbol Detection

This section provides the analysis of the performance the proposed defensive system. On one hand, the ASR is analysed for the eavesdropping stage, it accounts for the amount of information eavesdropped by Eve as compared to the information received by the legitimate receiver. On the other hand, the probability of fake data detection is obtained for the false data injection stage. Additionally, its performance is also compared to the hypothetical case based on the traditional reciprocal channel via RIS, showing that our proposed system is significantly better.

A. Eavesdropping: ASR

In this case, Eve is only listening to the data information transmitted by both legitimate entities of the link (BS and UE). The achievable rate of any communication of interest can be obtained as

$$C_i = \log_2\left(1+\rho_i\right), \quad i \in \{d, e, a\}, \tag{16}$$

where N_p is the number of OFDM symbols devoted to the transmission of reference signals, η_i and ρ_i correspond to the efficiency and SNR of the link of interest, respectively, and i is a token.

Then, the ASR [16] between the direct and Eve links measures the amount of information eavesdropped by Eve as compared to the legitimate link, which can defined as

$$E_d = \begin{cases} C_d - C_e & \rho_d > \rho_e \\ 0 & \rho_d \leq \rho_e \end{cases}, \quad \rho_e \in \{\rho_{eq}, \rho_{eg}\}, \tag{17}$$

$$\rho_d = \frac{\mathbb{E}\left\{|h_d|^2\right\}}{\sigma_w^2} = \frac{\sigma_d^2}{\sigma_w^2}, \quad \mathbb{E}\left\{|x_{b,n}|^2\right\} = \mathbb{E}\left\{|x_{u,n}|^2\right\} = 1, \tag{18}$$

$$\rho_{eq} = \frac{\mathbb{E}\left\{|q_e|^2\right\}}{\sigma_w^2} = M_e\frac{\sigma_{q_e}^2}{\sigma_w^2}, \quad \rho_{eg} = \frac{\mathbb{E}\left\{|g_e|^2\right\}}{\sigma_w^2} = M_e\frac{\sigma_{g_e}^2}{\sigma_w^2}, \tag{19}$$

where ρ_d is the SNR of the directly link between BS$\leftrightarrow$UE, while ρ_{eq} and ρ_{eg} correspond to the SNRs of the links between BS$\leftrightarrow$Eve and UE$\leftrightarrow$Eve, respectively. Regarding the efficiency of the system, $N_p = 1$ since the CSI can be obtained in the UL and reused in the DL due to the channel reciprocity property.

Typically, $E_d = 0$, because the SNR of Eve is always better than the direct link since the channel propagation of the former is better than the latter. Moreover, the high number of passive elements of the RIS, equipped by Eve, will enhance the received signal of the link. To circumvent this issue, the D-RIS is introduced to provide also an alternative link between the legitimate entities with a higher SNR, and hence, the ASR and its corresponding SNR can be obtained as

$$E_a = \begin{cases} C_a - C_e & \rho_a > \rho_e \\ 0 & \rho_a \leq \rho_e \end{cases}, \quad \rho_e \in \{\rho_{eq}, \rho_{eg}\}, \tag{20}$$

$$\rho_a = \frac{|h_a|^2}{\sigma_w^2} = M_a\frac{\sigma_{q_a}^2\sigma_{g_a}^2}{\sigma_w^2}, \tag{21}$$

respectively. Note that average gain of the cascaded channel of D-RIS is obtained by approximating its probability density function as a normal distribution, making use the Central Limit Theorem (CLT) [17]. This approximation can be also employed for the adversarial RIS equipped by Eve. By comparing (17)-(19) with (20)-(21), to improve the ASR not only increasing the number of passive elements of D-RIS is relevant to enhance the C_a, but it also requires to mitigate the information eavesdropped by Eve, which consists on minimizing C_e.

To avoid Eve accessing the information transmitted by the legitimate links, precoding is required to mask the transmitted information. In the case of using a reciprocal channel provided D-RIS, $N_p = 4$ since both BS and UE must transmit reference signals to allow the estimation of the CSI without the need of feeding back it. Consequently, (20) can be rewritten as

$$E_{ar} = \frac{N_r}{N}C_a + \left(1-\frac{N_r}{N}\right)\left(C_a - C_e\right) = C_a - \left(1-\frac{N_r}{N}\right)C_e, \tag{22}$$

where E_{ar} is the ASR of a reciprocal channel, N_r is the amount of time measured in OFDM symbols that Eve requires to find out the used precoder or the CSI of the reciprocal channel by the legitimate links. On one hand, the first term of (22) accounts for the ASR of the link when Eve has not discovered the used precoders yet, and hence, she is not able to obtain any information. On the other hand, the second term of (22) points out the ASR once Eve has found out the precoders, and therefore, the term C_e is penalizing the expression.

According to our proposed non-reciprocal channel ($N_p = 5$), (22) can be further improved by

$$E_{an} = C_a - \left(1-\frac{N_n}{N}\right)C_e, \tag{23}$$

where E_{an} is the ASR of a non-reciprocal channel, N_n is the amount of time measured in OFDM symbols that Eve requires to find out the used precoders by the legitimate links in a non-reciprocal channel. According to (9)-(11), Eve must search two

different precoders to allow her to eavesdrop on the transmitted information by both BS and UE. Hence, the proposed system based on a non-reciprocal channel, given in (23), doubles the required searching time of Eve to the reciprocal case given in (22) ($N_n = 2N_r$).

The final expression for the ASR for the non-reciprocal channel for the D-RIS can be expressed as

$$E_{an} = \eta_a \left(\log_2 (1 + \rho_a) - \left(1 - \frac{N_n}{N} \right) \log_2 (1 + \rho_e) \right), \tag{24}$$

where $\rho_e \in \{\rho_{eq}, \rho_{eg}\}$.

B. Data Manipulation: Probability of Fake Symbol Detection

Later, Eve can also inject false data to all legitimate receivers. Hence, this receiver is receiving both the original and false symbols sent by the legitimate transmitter and Eve, respectively. Depending on the signal strength of both signals and the noise, the receiver may wrongly decode the fake signal injected by Eve instead of the original one.

The probability of fake symbol detection of the proposed system can be defined as

$$P_r = \frac{N'_n}{N} P_1 + \left(1 - \frac{N'_n}{N} \right) P_2, \tag{25}$$

where P_1 and P_2 are the probabilities of decoding the fake message sent by Eve before and after finding out the employed combiners, respectively. N'_n accounts for the time required by Eve not only to obtain the used precoders by BS ($v_{b,n}$) and UE ($v_{u,n}$), but she must also compute the two terms of each precoder to find out the combiners ($v^{\text{UL}}_{b,n}$ and $v^{\text{DL}}_{u,n}$). Therefore, the proposed defensive system is even more robust against the false data injection rather than eavesdropping since it satisfies that $N'_n >> N_n$.

It is clear that before finding out the combiner, the legitimate receiver will not be able to decode the false message ($P_1 = 0$). Hence, (25) can be simplified as

$$P_r = \left(1 - \frac{N'_n}{N} \right) P_2. \tag{26}$$

In order to demodulate the injected signal by Eve at any legitimate receiver, its power should be higher than power threshold as, hence the resilience probability can be defined as

$$P_2 = Pr \left(|h_{e,n}|^2 > \beta \right), \tag{27}$$

where β is the power threshold, and given the scenario depicted in Fig. 1, it can be calculated as

$$\beta = M_a \sigma^2_{q_a} \sigma^2_{g_a} + \sigma^2_d + \sigma^2_w, \tag{28}$$

where the first term accounts for the average gain of the cascaded channel of D-RIS, obtained in (21). Therefore, substituting (28) in (25), the resilience probability can be computed as

$$\begin{aligned} P_2 &= Pr \left(|h_{e,n}|^2 > M_a \sigma^2_{q_a} \sigma^2_{g_a} + \sigma^2_d + \sigma^2_w \right) \\ &= \exp \left(- \frac{M_a \sigma^2_{q_a} \sigma^2_{g_a} + \sigma^2_d + \sigma^2_w}{M_e \sigma^2_{q_e} \sigma^2_{g_e}} \right). \end{aligned} \tag{29}$$

TABLE I
SIMULATION PARAMETERS

BS loc.	$(0, 0, 3)$	$\mathbf{P_{max}}$	-30 dBm	$\mathbf{M_e}$	2000
UE loc.	$(20, 0, 1)$	$\eta_\mathbf{a}$	$0.82, 0.86$	$\eta_\mathbf{d}$	0.91
D-RIS loc.	$(10, 5, 3)$	$\eta_\mathbf{s}$	$0.05 - 0.6$	$\mathbf{K}$	600
Eve loc.	$(10, -5, 3)$	$\mathbf{M_a}$	$[1 - 8] \times 10^3$	$\mathbf{N}$	22

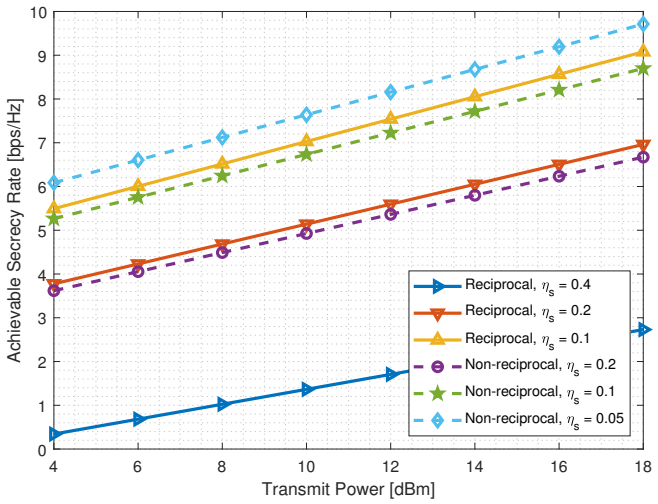

Fig. 3. Comparison of the ASR for eavesdropping between reciprocal and non-reciprocal cases for $M_a = 2000$.

V. Performance Evaluation

In this section, several numerical results are provided to show the performance of the proposed defensive system as compared to the existing reciprocal one. A summary of the simulation parameters is given in Table I. The channel propagation model adopted corresponds to the factory scenario given in [18]. Specifically, the direct channel between BS↔UE is assumed to be non-line-of-sigh (NLOS) since some obstacles may obstruct the direct ray. Other channels via D-RIS and Eve are considered to be line-of-sigh (LOS) since they provide a significantly better alternative path to the direct one. The geographical location of all entities is specified in Cartesian coordinates (x, y, z) m. Additionally, let us define η_s which is the percentage of the slot that Eve is jeopardizing the legitimate link since she has already found out the precoders/combiners, and it can be defined as

$$\eta_s = \begin{cases} 1 - N_r/N & \text{Reciprocal chan.} \\ 1 - N_n/N & \text{Non-recipr. chan. (eavesdropping)} \\ 1 - N'_n/N & \text{Non-recipr. chan. (data manipulation)} \end{cases}. \tag{30}$$

A. Evaluation of Eavesdropping

In Fig. 3, the comparison of the ASR between the non-reciprocal and reciprocal channels is given by considering different values of η_s. It is assumed that the time required by Eve to find out the precoding matrices for the non-reciprocal case is half of the reciprocal one ($N_{sr} = 2N_{sn}$). The proposed non-reciprocal case (dotted lines) has a better performance than the reciprocal one (solid lines) when the Eve is disrupting the link. Despite the proposed non-reciprocal

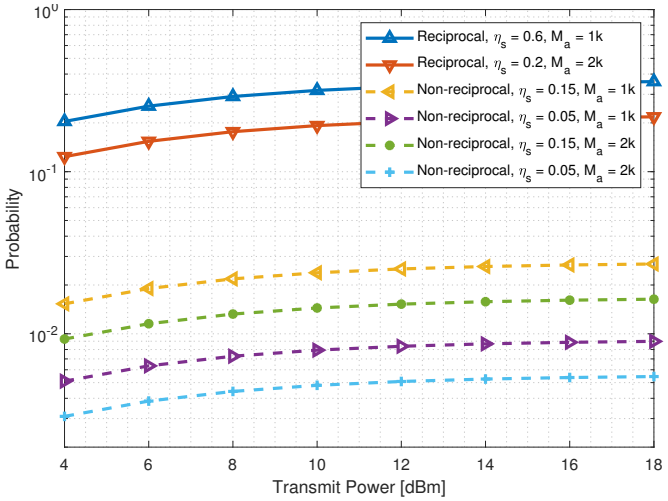

Fig. 4. Comparison of the probability of false data detection for different values of η_s and M_a.

defensive system has a lower performance than the reciprocal one from the theoretical perspective in Fig. **??** due to the CEP (η_a), the former is more robust against attacks since the system has a better protection as a consequence of using different precoders and combiners, and therefore, Eve requires more time find out the employed precoders and combiners to break the communication link.

B. Evaluation of False Data Injection

In Figs. 4 a comparison in terms of probability of fake symbol detection between the reciprocal and non-reciprocal cases for the data manipulation scenario is given. In Fig. 4, the proposed defensive system based on a non-reciprocal channel is again always better than the reciprocal one via RIS even though it requires only one more OFDM symbol to transmit the reference signals. The reason behind these results is due to the robust design of the precoders and combiners, making the intrusion more difficult for the adversarial RIS. Note that either increasing the number of passive elements at D-RIS (M_a) or enlarging the time required to find out the precoders/combiners will lower the probability of detecting the fake symbols reflected by Eve, however the latter is more effective than the former since it is able to reduce the probability by one order of magnitude.

VI. Conclusions

A novel defensive system against RITM attacks relying on a non-reciprocal channel via D-RIS is detailed in this work to provide a more reliable and secure communication system. The proposed technique is not only able to prevent eavesdropping, but it also is capable of checking the integrity of injected symbols by a potential adversary without assuming any knowledge of them, unlike the existing works. The employed precoding and combining techniques based on the combination of the CSI of a non-reciprocal channel are robust against eavesdropping and false data injection since its disruption is significantly harder than existing techniques. Consequently, the integration of this defensive system is a key element in the evolution towards 6G since it protects the existing communication system, safeguarding the privacy.

Acknowledgments

The work of K. Chen-Hu and P. Popovski have been funded by the Villum Investigator Grant "WATER" from the Velux Foundation, Denmark.

References

[1] J. Wang, W. Tang, S. Jin, C.-K. Wen, X. Li, and X. Hou, "Hierarchical codebook-based beam training for RIS-assisted mmwave communication systems," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3650–3662, Jun. 2023.

[2] K. Chen-Hu, G. C. Alexandropoulos, and A. g. Armada, "Differential data-aided beam training for RIS-empowered multi-antenna communications," *IEEE Access*, vol. 10, pp. 113 200–113 213, Oct. 2022.

[3] V. Croisfelt, F. Saggese, I. Leyva-Mayorga, R. Kotaba, G. Gradoni, and P. Popovski, "Random access protocol with channel oracle enabled by a reconfigurable intelligent surface," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 12, pp. 9157–9171, 2023.

[4] V. Shahiri, H. Behroozi, A. Kuhestani, and K.-K. Wong, "Reconfigurable intelligent surface-assisted secret key generation under spatially correlated channels in quasi-static environments," *IEEE Internet Things J.*, pp. 1–1, 2024.

[5] M. Wei, H. Zhao, V. Galdi, L. Li, and T. J. Cui, "Metasurface-enabled smart wireless attacks at the physical layer," *Nature Electronics*, vol. 6, pp. 610–618, Aug. 2023.

[6] Z. Shaikhanov, F. Hassan, H. Guerboukha, D. M. Mittleman, and E. W. Knightly, "Metasurface-in-the-middle attack: From theory to experiment," *Proc. 15th ACM WiSec*, pp. 257–267, 2022.

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

[8] W. Shi, J. Xu, W. Xu, C. Yuen, A. L. Swindlehurst, and C. Zhao, "On secrecy performance of RIS-assisted MISO systems over Rician channels with spatially random eavesdroppers," *IEEE Trans. Wirel. Commun.*, pp. 1–1, 2024.

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

[10] G. C. Alexandropoulos, K. D. Katsanos, M. Wen, and D. B. Da Costa, "Counteracting eavesdropper attacks through reconfigurable intelligent surfaces: A new threat model and secrecy rate optimization," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1285–1302, 2023.

[11] J. Luo, F. Wang, S. Wang, H. Wang, and D. Wang, "Reconfigurable intelligent surface: Reflection design against passive eavesdropping," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 5, pp. 3350–3364, May 2021.

[12] *5G-NR; Physical channels and modulation (Release 17)*, 3GPP Std. 38.211, 2022.

[13] W. Kim, H. Ji, and B. Shim, "Channel aware sparse transmission for ultra low-latency communications in TDD systems," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1175–1186, Feb. 2020.

[14] H. Chung and S. Kim, "Location-aware beam training and multi-dimensional ANM-based channel estimation for RIS-aided mmwave systems," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 1, pp. 652–666, Jan. 2024.

[15] S. Jin, X. Liang, K.-K. Wong, X. Gao, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 3, pp. 1480–1491, 2015.

[16] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 356–360.

[17] J. A. Rice, *Mathematical Statistics and Data Analysis.*, 3rd ed. Belmont, CA: Duxbury Press., 2006.

[18] *Study on channel model for frequencies from 0.5 to 100 GHz (Release 17)*, 3GPP Std. 38.901, 2022.