Randomized RIS Signal Watermarking in FutureG Millimeter-Wave Wireless Communications

Farshad Soleiman, Edward Kwao, Xuan Chen, and Kai Zeng
Department of Electrical and Computer Engineering
Wireless Cyber Center
George Mason University, Fairfax, VA, U.S.A.
Emails: {fsoleima, ekwao, xchen38, kzeng2}@gmu.edu

***Abstract*—Reconfigurable Intelligent Surfaces (RISs) dynamically optimize the radio frequency (RF) environment for enhanced signal quality in FutureG millimeter-wave wireless communication. Despite their benefits, RISs can be exploited for malicious purposes, posing threats to the availability, integrity, and security of FutureG wireless communication systems. Conventional data encryption or authentication methods at data layer are not useful against malicious RIS attacks that are launched at physical (PHY) layer. To defend against malicious RIS attacks, we propose a novel RIS signal watermarking scheme that enables the authentication of a RIS deflected signal or the signal path that involves a legitimate RIS. The proposed scheme allows a legitimate RIS to embed a random bit into a deflected RF data symbol by conducting a differential binary phase shift keying (DBPSK) modulation in the middle of the RF symbol. In order to extract both the watermark and data with high reliability, we present a sequential decoding scheme that first decodes the watermark using non-coherent decoding and then data decoding from the recovered data symbol. We derive the bit error rates (BER) performance of both watermark and data decoding under a additive white Gaussian noise channel model. Analytical results show the efficiency of our proposed scheme.**

***Index Terms*—Watermark, millimeter wave communication, RIS, bit error rate**

I. Introduction

Reconfigurable Intelligent Surfaces (RISs) are designed to dynamically control and customize the radio frequency (RF) environment, which is considered one of the promising technologies to address the blockage and coverage problems in future generation (FutureG) millimeter-wave (mmWave) wireless communication systems. RISs are typically passive planar metasurfaces comprised of a large number of reflecting elements. By individually controlling the phase shift and reflecting coefficient of each element, RIS can steer or scatter the incident signal into one or multiple directions [1], [2].

On the one hand, the low cost and low power properties of RIS facilitate ease of deployment in scale. On the other hand, it is easy for the attacker to use this technology to program/control the RF environment in favor of malicious goals as well, e.g., gaining eavesdropping advantages, degrading wireless link quality, poisoning channel estimation, etc.

Malicious attackers can take advantage of this wireless channel programmability and increase the stealthiness of an adversarial act by skilfully deploying adversarial RISs in the environment. An adversarial actor with a malicious RIS can carry out over-the-air attacks that aim at the data link, link establishment or monitor the communication through itself or by redirecting the signal to an eavesdropping entity. As reported in [3], a malicious RIS is capable of altering the channel between transmitter and receiver during channel estimation and data transmission, resulting in high symbol error rate. The malicious RIS can manipulate the beam-searching process, leading to different constructive or destructive interference at receiver. A malicious RIS can also introduce destructive interference to the line-of-sight path by taking advantage of the transmitter sidelobe signal [4].

The fundamental reason for the success of the aforementioned malicious RIS-based attacks lies in the lack of ability to authenticate the RIS signal or differentiate legitimate RIS signals from malicious ones. Since the attacker is manipulating the RF environment/signal, conventional data encryption or authentication mechanisms at data layer are not useful to defend against such new attacks. To thwart the malicious RIS-based attacks, we propose a novel watermarking scheme that takes advantage of the reconfigurability of RIS to embed a watermark (i.e., a random bit stream) into the data signal when it travels through a legitimate RIS. This watermark can be generated through a shared secret between the legitimate sender, receiver, and RIS, thus is unpredictable by a malicious RIS. At the receiver side, the receiver can examine whether the expected watermark resides in the RF data symbols to differentiate a signal deflected from a legitimate RIS from a malicious one since a malicious RIS is unable to embed the expected watermark into the RF signal.

Since a RIS does not generate signals, it has to embed the watermark on-the-fly when the RF signal is deflecting from the RIS. Therefore, existing transmitter based RF signal watermarking schemes, such as constellation dithering [5]–[8], cannot be applied. Furthermore, the watermark embedding should not significantly affect the data decoding performance. To address these challenges, we propose an idea of embedding the watermark into each RF data symbol signal based on differential binary phase shift keying (DBPSK). The basic idea is to embed a bit '1' into the RF data symbol by shifting the phase of the second half of the data symbol by π and embed a bit '0' without shifting. Then at the receiver side, by examining whether there is a phase shift in the middle of the RF data symbol, the watermark can be exacted. To decode the data, we first recover the RF data symbol by reversing the effect of the watermark, then use a coherent detector. We call this method as a sequential decoding scheme.

The proposed RIS watermark embeding and decoding

scheme can be served as a building block to secure initial access, beam sweeping, and data communication processes for RIS-assisted FutureG mmWave wireless communications.

We summarize the the main contributions of this work as follows:

- We propose a RIS signal (or signal path) authentication technique that embeds DBPSK watermark in each RF data symbol. The watermark can be extracted using non-coherent detection without requiring the knowledge of the carrier frequency or channel state information.
- We propose and analyze the performance of a sequential decoding scheme. We derive the bit error rate (BER) performance for data and watermark decoding under two cases: DBPSK watermark embedded in BPSK data and QPSK data, respectively.
- Finally, we discuss future research topics and directions that can build upon this RIS signal watermark building block.

II. Related Work

There are existing physical layer authentication and watermarking mechanisms for active transmitter authentication. We point out the difference of our RIS watermarking schemes from the existing techniques as follows.

1) Difference and advantage over active transmitter RF fingerprinting: There are two major RF fingerprinting mechanisms: hardware-based and channel/location-based [9]. Hardware-based RF fingerprinting usually requires a high-end signal analyzer to extract subtle hardware-impairment-induced fingerprints and it is sensitive to environment noise and mobility. Channel or location-based RF fingerprinting usually requires knowledge of the channel state/statistics (which requires a training phase) and cannot work well in an unknown environment [10]. Our proposed RIS signal watermarking methods work in mobile and unknown environments and do not require a high-end signal analyzer. So they enjoy high applicability with low deployment overhead.

2) Difference from transmitter-based watermarking: Existing RF watermarking techniques focus on embedding a watermark at the transmitter side by applying the concept of constellation dithering [5]–[7]. However, dithering methods are constrained to low-order data modulation, e.g., QPSK, and their performance tends to degrade under higher-order data modulation [11]. Our watermarking schemes do not have such a constraint. The intuition behind it is that we use differential modulation and decoding within a data symbol, which allows us to introduce large perturbations, e.g., flip the phase of the signal, to achieve high watermarking decoding probability with almost negligible impact on data communication performance.

3) Difference from RIS-enhanced channel-based RF fingerprinting: There are few works [12], [13] on using RIS to enhance the channel diversity or controllability for transmitter authentication. However, the method proposed in [13] requires a pre-training step and can only be applied in static environments. It requires a channel estimation step at the beginning of a 4-step challenge-response protocol in [12]. It also requires high channel reciprocity and is unlikely to work well in dynamic environments. Furthermore, these works aim to authenticate the transmitter but not the RIS signal.

4) Difference from backscatter communication: Backscatter communication can code information into a known instance signal or ambient RF signal [14]. However, it usually requires the knowledge of the instance signal or a copy of the ambient RF signal to decode the information. Our watermarking scheme does not have such a requirement and it can decode both the data and watermark information.

5) Difference from existing RIS modulation schemes: There are a few recent works on RIS-based information transfer and symbiotic communication [15]–[17]. They either require CSI [16] for demodulation/decoding or incur high computational complexity (i.e., super exponential to the number of transmission antennas or RIS reflecting patterns). Although low-complexity non-coherent detection methods are proposed in [15], [17], they utilize spatial modulation that requires multiple antennas (i.e., RF chains) at the receiver and identify antenna index (spatial) information using energy detection which cannot achieve desirable performance under low SNR.

III. System Model and Problem Definition

We consider a mmWave network with one base station (BS), one legitimate RIS, one malicious RIS, and one user equipment (UE) as illustrated in Fig. 1. Assuming BS has N_b antennas and UE has a single antenna. RIS has M passive elements and its coefficient matrix is denoted as $\mathbf{\Phi} = diag(\gamma_1 e^{j\theta_1}, \gamma_2 e^{j\theta_2}, ..., \gamma_M e^{j\theta_M})$, where $\gamma_m \in [0, 1]$ and $\theta_m \in [0, 2\pi]$ $(1 \leq m \leq M)$ represent the amplitude reflection coefficient and phase shift associated with the m-th passive element of the RIS, respectively.

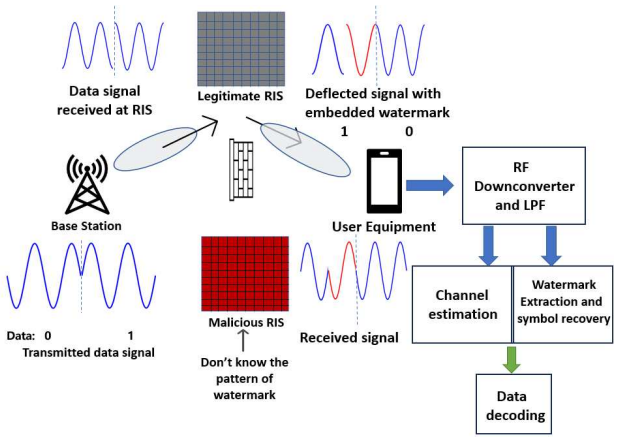

Fig. 1. Illustration of RIS Signal Watermark Embedding and Decoding

For simplicity and better illustration of our idea, we focus on the single RIS reflection path in this paper. This model can be extended to more complex models by considering other scattering paths or line-of-sight path as well as multiple antennas at the UE side.

Let $\boldsymbol{H}_{br} \in \mathbb{C}^{(M \times N_b)}$ and $\boldsymbol{h}_{ru} \in \mathbb{C}^{(M \times 1)}$ denote the BS-to-RIS and RIS-to-UE channel matrices, respectively. They are usually represented by geographic channel models [18]. Consider the donwlink communication from BS to UE through RIS

and denote the beamforming vector at the BS as $\boldsymbol{w} \in \mathbb{C}^{(N_b,1)}$, under perfect channel equalization at the UE side, the received baseband signal y given a transmitted baseband symbol x (e.g., x is either 1 or -1 for BPSK modulation) can be represented as

$$y(t) = \sqrt{\boldsymbol{2P}}\boldsymbol{h}_{ru}^T \boldsymbol{\Phi} \boldsymbol{H}_{br} \boldsymbol{w} x \cdot g(t) + n(t), \quad (1)$$

where $\boldsymbol{P}$ denotes the average transmission power per symbol, $n(t)$ is a white Gaussian noise process, and $g(t)$ is a gate function defined below:

$$g(t) = \begin{cases} 1 & \text{if } 0 \le t \le T_s \\ 0 & \text{otherwise} \end{cases},$$

where T_s is the symbol duration. We assume the legitimate BS, RIS, and UE share a secret which allows them to agree on a randomized watermark bit stream that is unpredictable by the malicious RIS. This secret could be generated using the preloaded shared secret between the BS and UE (e.g., in cellular networks, BS and UE share a master secret) and that between BS and RIS (assuming a legitimate RIS is either deployed by the network operator or operated by a trusted third party, so a shared secret can be deployed on BS and RIS). We also assume the BS and RIS are time synchronized through a control channel or a shared clock (e.g., GPS clock).

Attacking Model: We assume the RIS watermark scheme is public, allowing a malicious RIS to potentially make random guesses and embed their own watermark into its deflected signal. The malicious RIS could even be more powerful than the legitimate RIS and is capable of recording and replaying the signal. The major goal of the malicious RIS is to manipulate the RF environment in favor of its own purpose, such as gaining eavesdropping advantages by alluring the transmitter to point/beamform the signal to itself through offering a high-quality signal path during the beam sweeping/searching or channel estimation process. The denial-of-service (e.g., jamming attack) is out of the scope of this work.

In the subsequent section, we introduce our watermarking scheme, designed to enable the legitimate RIS to embed a randomized watermark into the deflected RF signal on-the-fly.

IV. Randomized RIS Signal Watermarking

The objective of RIS signal watermark embedding is to embed a randomized watermark into the deflected RF signal so that the receiver (UE) is able to extract both watermark and data with high reliability. Note that, the receiver does not know either data or watermark in prior, so it is non-trivial to embed a watermark without corrupting the data.

A. Challenges

Note that it is non-trivial to authenticate the RIS signal at the receiver/UE side even a secret is shared by BS, RIS, and UE. There are three challenges. First, we cannot simply apply the existing transmitter signal watermarking scheme to authenticate RIS signal, i.e., for the receiver to verify whether the signal does travel through a legitimate RIS but not a malicious one. Transmitter signal watermarking (e.g., constellation dithering) can only be used to authenticate the transmitter signal, but not the RIS signal since a malicious RIS can also deflect/redirect the transmitter signal that carries the legitimate watermark embedded by the legitimate transmitter. So when a receiver receives such a signal, it still cannot tell whether it is deflected by a legitimate RIS or malicious one. Second, data layer authentication mechanisms such as message authentication code cannot be applied either to solve the RIS signal authentication problem. The data layer authentication can only prove the data (i.e., decoded bits) *not the signal* is sent by the transmitter. Furthermore, RIS does not have a full RF chain so cannot decode the signal and append a message authentication code to the data. Third, the signal watermark embedded at RIS should have negligible effect on the data decoding performance at the receiver. That is, the receiver needs to extract the watermark and decode the data both at low BER. Our proposed randomized RIS signal watermarking scheme solves these challenges and is described below.

B. Randomized Watermark Embedding at RIS

Our idea is motivated by the observation that the modulated RF signal typically contains several tens or hundreds of carrier wave (i.e., sinusoidal wave) cycles within a single data symbol duration. For example, if the carrier frequency is 28GHz, for a BPSK signal at 1Gpbs, each symbol has 28 cycles. This redundancy offers an opportunity for us to embed a randomized watermark into the RF signal based on DBPSK when it travels through a legitimate RIS.

The idea is to keep the first half of the data symbol intact and change the phase of the second half of the data symbol signal based on DBPSK. For instance, if a watermark bit is '1', we flip the phase (or shift by π) of the second half of the data signal. Then the received watermarked signal at the receiver becomes:

$$y'(t) = \sqrt{\boldsymbol{2P}} h_{\text{ru}}^T \Phi H_{br} w x \cdot (g(2t) - g(2t - T_s)) + n(t) \quad (2)$$

On the other hand, if a watermark bit is '0', we have no phase shift on the second half of the signal (same as intact signal).

$$y'(t) = y(t) \quad (3)$$

An example is illustrated in Fig. 1, where the BS is transmitting a data bit stream "01". Assume that the BS applies BPSK modulation and each symbol/bit has a duration of two carrier cycles. When the RF signal arrives at the legitimate RIS, it inserts 1 bit watermark into each data symbol based on DBPSK. When the watermark is bit '1', RIS flips the phase of the second half of the RF data symbol; otherwise, keep the second half intact. Note that the first half of the RF data symbol is not be changed by watermark embedding no matter what bit information is embedded.

The watermark embedding should not affect the beamforming directions at the RIS. That is, if the RIS is deflecting the signal towards UE, embedding watermark should not change it. To achieve this, we can shift the phase of each passive element on RIS by π simultaneously. Since each symbol is very short, the channel can be considered static during one symbol duration. If the channel is not changed, according to Eq (1), given a RIS configuration (i.e., pointing the deflected signal to UE), shifting the phase of each RIS element by π simultaneously will shift the phase of the received signal by π without changing the beamforming direction of the RIS. It

should be noticed that shifting the phase of each RIS element by a common value will not change the beamforming direction or gain of the deflected signal.

The watermark can be embedded in any data symbol, but should not be embedded in preamble or pilot symbols that are used for channel estimation. As we will analyze in the following section, although the watermark decoding does not require channel state information, the data decoding still needs it.

C. Randomized Watermark Generation

The watermark bit string is generated randomly by a cryptographic pseudorandom number generator based on a fresh nonce and shared secret between BS, RIS, and UE for each use. A potential implementation is as follows. Before the channel estimation or beam searching process, the BS generates a fresh nonce and sends it to the legitimize RIS over a secured control channel. The randomized watermark is generated based on the nonce and the shared secret between the BS and RIS. If the BS and UE have already established a secure channel (e.g., the UE has already passed authentication and key agreement (AKA) process when it joined the network), the BS can send the expected watermark to UE securely. Then the UE can compare the decoded watermark and the expected watermark in the later RIS signal authentication stage. If it is in the initial access phase, a secure channel may not have been established between the BS and UE. Then the UE can extract the watermark and send it back with the beam searching results to the BS for a post checking at the BS.

D. Watermark and Data Decoding

At the receiver side, we extract the watermark and decode the data in a sequential way as illustrated in Fig. 1. We first examine each data symbol signal to see if there is a phase shift in the middle of the symbol using a differential/non-coherent detector. If there is, the watermark bit '1' is decoded, otherwise, bit '0' is decoded. For the data decoding, we can recover the second half of the data symbol based on the decoded watermark (i.e., flip the second half of the symbol by π if watermark bit '1' is decoded, otherwise keep the second half unchanged) and decode the data over the recovered symbol signal.

Note that for data decoding, we still need channel state information, which can be estimated using intact preamble or pilot symbols.

For sequential watermark extraction and data decoding, we can extract the watermark using efficient non-coherent detection. Due to the efficiency of non-coherent detection, only negligible delay would be introduced for the data decoding. The error rate at which the watermark is extracted will have an impact on data decoding performance. The effectiveness of this sequential decoding schemes is analyzed in the following section.

V. Performance and security Analysis

We analyze the watermark BER and data BER under two repreentative scenarios: DBPSK watermark embedded in BPSK data and QPSK data, respectively, under an Additive White Gaussian Noise (AWGN) model. To unify the analysis, we assume that the symbol duration (T_s) and the total energy per symbol (E_s) are identical for both QPSK and BPSK data modulation.

A. DBPSK Watermark Embedded in BPSK Data Symbol

Note that the sequential decoding extracts the watermark first and decode the data over the recovered data symbol signal depending on the watermark decoding result. Therefore, the data decoding performance is affected by the watermark decoding performance. The BER of data decoding can be expressed as:

$$P_r(\overline{D}) = P_r(\overline{D}|\overline{W}) \cdot P_r(\overline{W}) + P_r(\overline{D}|W) \cdot P_r(W), \quad (4)$$

where $P_r(\overline{D}|\overline{W})$ is the data BER given an unsuccessful watermark extraction, $P_r(\overline{W})$ is the probability of wrong watermark decoding/extraction, $P_r(\overline{D}|W)$ is the data BER given a successful watermark decoding, and $P_r(W)$ is the probability of correct watermark decoding. Note that $P_r(W) = 1 - P_r(\overline{W})$. To obtain $P_r(\overline{D})$, we derive $P_r(\overline{D}|\overline{W})$, $P_r(\overline{W})$, and $P_r(\overline{D}|W)$ as follows.

1) Derivation of $P_r(\overline{D}|\overline{W})$: If the watermark extraction goes wrong, the data symbol signal will not be correctly recovered. If the watermark bit is '0' (i.e., no phase shift in the middle of the deflected signal) but is detected as '1', the data symbol signal will be wrongly recovered by shifting the second half of the symbol. Similarly, if the watermark bit is '1' (i.e., there is indeed a phase shift of π in the middle of the deflected signal) but is detected as '0', the data symbol signal will be wrongly recovered without shifting the second half of the symbol back. Assume an optimal receiver is used for the BPSK data signal decoding based on a matched filter with an impulse response $h(t) = g(t)$ (i.e., identical to the gate function). [19]

To simplify analysis, for each data symbol signal, we ignore the fixed beamforming and channel gain ($h_{\text{ru}}^T \Phi H_{br} w$). Assuming an incorrect watermark decoding, after channel equalization, the received BPSK RF signal entering into the downconverter and matched filter can be represented as:

$$z(t) = \sqrt{\mathbf{2P}}x \cdot (g(2t) - g(2t - T_s)) \cdot cos(2\pi f_c t) + n'(t) \quad (5)$$

Therefore, the output after donwconverting $z(t)$ through a mixer with a local oscillator (LO) input of $\sqrt{\frac{2}{T_s}} cos(2\pi f_c t)$ and then through a matched filter, sampled at T_s is:

$$\begin{aligned} &X(T_s) \\ &= \int_{<T_s>} z(\tau) \cdot \sqrt{\frac{2}{T_s}} cos(2\pi f_c \tau) h(t-\tau) d\tau \\ &= \int_{<T_s/2>} \sqrt{\frac{2}{T_s}} cos(2\pi f_c \tau) \sqrt{2P} x \cdot g\big(2t\big) cos(2\pi f_c \tau) d\tau \\ &- \int_{<T_s/2>} \sqrt{\frac{2}{T_s}} cos(2\pi f_c \tau) \sqrt{2P} x \cdot g\big(2t - T_s\big) cos(2\pi f_c \tau) d\tau \\ &+ \int_{<T_s>} \sqrt{\frac{2}{T_s}} cos(2\pi f_c \tau) n'(\tau)\, d\tau = W \end{aligned} \quad (6)$$

where f_c is the carrier frequency and $n'(\tau)$ is a passband white Gaussian noise with zero mean and auto correlation function

of $\frac{N_0}{2} \cdot \delta(t-u)$. The initial two terms after the second equal sign in Eq. (6) cancel each other due to a phase difference of π between the first half and second half of the data symbol signal. The third term is equal to W, which is a Gaussian random variable following $\mathcal{N}(0, \frac{N_0}{2})$.

At the receiver side employing an optimum BPSK receiver, the sampled output of the matched filter is input into a comparator with a threshold of zero. When the sampled output signal is greater than zero, it is detected as bit '1'; otherwise, it is detected as bit '0'. Given that the sampled output of the matched filter follows $\mathcal{N}(0, \frac{N_0}{2})$ no mater what BPSK data symbol is transmitted, we can obtain that

$$P_r(\overline{D}|\overline{W}) = \int_0^\infty \frac{1}{\sqrt{2\pi \cdot \frac{N_0}{2}}} exp(-\frac{x^2}{2 \cdot \frac{N_0}{2}})dx = \frac{1}{2}. \quad (7)$$

2) Derivation of $P_r(\overline{W})$: Note that for a data symbol, no matter whether it represents bit '1' or '0', its RF signal is a constant sinusoidal waveform for the whole symbol duration. The watermark uses the first half of the data symbol RF signal as the reference signal and embed the information in the second half of the data symbol RF signal using DBPSK. Therefore, $P_r(\overline{W})$ is equal to the BER of DBPSK [20] with a symbol energy $\frac{E_s}{2}$ that is half of the data symbol energy.

$$P_r(\overline{W}) = \frac{1}{2}\exp\left(-\frac{E_s}{2N_0}\right) \quad (8)$$

3) Derivation of $P_r(\overline{D}|W)$: If the watermark is correctly decoded, the data symbol will be fully recovered. Therefore, $P_r(\overline{D}|W)$ is equal to the BER of conventional BPSK.

$$P_r(\overline{D}|W) = Q\left(\sqrt{\frac{2E_s}{N_0}}\right), \quad (9)$$

where $Q(\cdot)$ is the q-function: $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt$.

Finally, by substituting Eqs. (7), (8), and (9) into Eq. (4), we obtain the data BER under the scenario of DBPSK watermark embedded in BPSK data symbol as

$$P_r(\overline{D}) = \frac{1}{2}\times\frac{1}{2}\exp\left(-\frac{E_s}{2N_0}\right) + Q\left(\sqrt{\frac{2E_s}{N_0}}\right)\left(1-\frac{1}{2}\exp\left(-\frac{E_s}{2N_0}\right)\right) \quad (10)$$

B. DBPSK Watermark Embedded in QPSK Data Symbol

QPSK modulation can be considered as the combination of two parallel BPSK modulations with two carriers offset in $\frac{\pi}{2}$ [21]. When a DBPSK watermark is wrongly decoded, the first half and second half the data symbol RF signal will have a phase difference of π as discussed in the previous subsection. Therefore, the $P_r(\overline{D}|\overline{W})$ under this scenario is also $\frac{1}{2}$ as derived in Eq. (7).

Similarly, the BER of the DBPSK watermark embedded in the QPSK data symbol is the same as that (Eq. (8)) of DBPSK watermark embedding in BPSK data symbol. The reason behind that is no matter what modulation scheme the data transmission is using, the RF sinusoidal waveform for one data symbol duration is constant. That is, the amplitude or phase parameter of the data symbol does not affect the DBPSK watermark decoding since the it only depends on the relative phase change between the first half and and second half of the RF data symbol signal.

Similar to the derivation of Eq. (9), $P_r(\overline{D}|W)$ under this scenario is the BER of conventional QPSK:

$$P_r(\overline{D}|W) = Q\left(\sqrt{\frac{E_s}{N_0}}\right) \quad (11)$$

The data BER under the scenario of DBPSK watermark embedded in QPSK data symbol is

$$P_r(\overline{D}) = \frac{1}{2}\times\frac{1}{2}\exp\left(-\frac{E_s}{2N_0}\right) + Q\left(\sqrt{\frac{E_s}{N_0}}\right)\left(1-\frac{1}{2}\exp\left(-\frac{E_s}{2N_0}\right)\right) \quad (12)$$

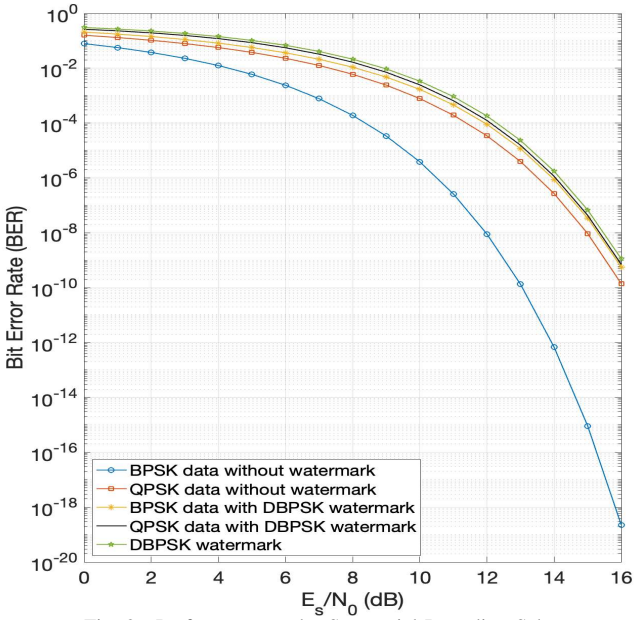

Fig. 2. Performance under Sequential Decoding Scheme

C. Analytical Results

Fig. 2 shows the analytical BER performance of watermarking and non-watermarking cases under different signal-to-noise ratios (SNRs). By comparing the data BER under watermark and non-watermark cases, we observe an interesting result: although embedding watermark does sacrifice the data decoding performance, the performance degradation is not significant and such degradation is smaller for QPSK data than BPSK data. It is about 3.5dB SNR performance loss for BPSK data and 1dB loss for QPSK data when embedding the DBPSK watermark. The performance degradation becomes smaller when SNR is increased. The conjecture is that the asymptotic BER performance of QPSK data decoding will be the same for watermarking and non-watermarking cases, while it will be 3dB performance difference for the BPSK data decoding performance with and without watermark. The

decoding performance of both BPSK and QPSK data with watermark is better than that of DBPSK watermark.

D. Security Analysis

Our randomized RIS watermarking scheme is secure against random guess and signal replaying attacks.

Since the watermark bit string is generated randomly by a pseudo random function based on a fresh nonce and shared secret between BS, RIS, and UE for each use, it is computationally infeasible for an attacker to guess out the randomized watermark. So if a malicious RIS generates its own watermark and embeds it into the RF signal, it will not pass the verification process at the receiver.

An advanced attacker could record the signal deflected by a legitimate RIS and replay signal in a later communication session. However, since the watermark is randomized for each use, this replay attack will fail too.

Note that, since the watermarking scheme is public, an eavesdropper will be able to extract the watermark as well. However, it does not compromise the security of the watermarking scheme since the watermark is randomized for each use.

VI. Conclusion and Future Work

In this paper, a watermark embedding scheme and sequential decoding method are proposed to authenticate a signal path through a legitimate RIS in the presence of malicious RIS. The differential modulation is applied to embed watermark. Non-coherent detection is adopted to decode watermark and coherent detection is used to recover the whole data. The watermark extraction and data decoding can be accomplished in both sequential way and parallel way. We consider the situation of BPSK and QPSK data and derive their probability of error.

We believe this work proposed a viable building block for securing RIS-assisted mmWave communication and provided a novel way to authenticate a programmable RF environment. We hope it will stimulate a lot of interesting follow-up research. In the future, we will investigate the performance of the watermarking scheme under different data and watermark modulation schemes, e.g., DQPSK watermark embedded in M-ary phase shift keying or QAM (quadrature amplitude modulation) data. For the OFDM system, we can embed the watermark into the time samples using a similar way. Building an analytical model for the watermark and data decoding performance for the OFDM system will be our future work. Practical impairments and constraints such as synchronization offsets between BS and RIS and the watermark embedding rate at the RIS need to be considered for real-world system implementation. We also plan to implement and test the watermarking scheme on a mmWave tested. The watermarking scheme can be integrated with beam management functions to protect the beam sweeping process and the scenario with multiple legitimate and malicious RISs deserves further investigation.

VII. Acknowledgement

This work is supported by National Science Foundation (NSF) through the Secure and Trustworthy Cyberspace (SaTC) Program under Grant No. 2318796.

References

[1] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Communications Surveys Tutorials*, vol. 23, no. 3, pp. 1546–1577, 2021.

[2] G. C. Trichopoulos, P. Theofanopoulos, B. Kashyap, A. Shekhawat, A. Modi, T. Osman, S. Kumar, A. Sengar, A. Chang, and A. Alkhateeb, "Design and Evaluation of Reconfigurable Intelligent Surfaces in Real-World Environment," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 462–474, 2022.

[3] Z. Shaikhanov, F. Hassan, H. Guerboukha, D. Mittleman, and E. Knightly, "Metasurface-in-the-middle attack: from theory to experiment," in *Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2022, pp. 257–267.

[4] B. Sadhu, Y. Tousi, J. Hallin, S. Sahl, S. Reynolds, Ö. Renström, K. Sjögren, O. Haapalahti, N. Mazor, B. Bokinge *et al.*, "7.2 a 28ghz 32-element phased-array transceiver ic with concurrent dual polarized beams and 1.4 degree beam-steering resolution for 5g communication," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, 2017, pp. 128–129.

[5] J. Ma, J. Chen, and G. Wu, "Robust watermarking via multidomain transform over wireless channel: Design and experimental validation," *IEEE Access*, vol. 10, pp. 92 284–92 293, 2022.

[6] H. Huang and L. Zhang, "Reliable and Secure Constellation Shifting Aided Differential Radio Frequency Watermark Design for NB-IoT Systems," *IEEE Communications Letters*, vol. 23, no. 12, pp. 2262–2265, 2019.

[7] Z. Xu and W. Yuan, "Watermark BER and Channel Capacity Analysis for QPSK-Based RF Watermarking by Constellation Dithering in AWGN Channel," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1068–1072, 2017.

[8] K. Grzesiak and Z. Piotrowski, "From constellation dithering to noma multiple access: Security in wireless systems," *Sensors*, vol. 21, no. 8, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/8/2752

[9] N. Wang, L. Jiao, P. Wang, W. Li, and K. Zeng, "Machine learning-based spoofing attack detection in mmWave 60GHz IEEE 802.11 ad networks," in *IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2020, pp. 2579–2588.

[10] R. Che and H. Chen, "Channel state information based indoor fingerprinting localization," *Sensors*, vol. 23, no. 13, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/13/5830

[11] K. Grzesiak and Z. Piotrowski, "From Constellation Dithering to NOMA Multiple Access: Security in Wireless Systems," *Sensors*, vol. 21, no. 8, p. 2752, 2021.

[12] S. Tomasin, H. Zhang, A. Chorti, and H. V. Poor, "Challenge-Response Physical Layer Authentication over Partially Controllable Channels," *IEEE Communications Magazine*, vol. 60, no. 12, pp. 138–144, 2022.

[13] S. Rajendran, Z. Sun, F. Lin, and K. Ren, "Injecting Reliable Radio Frequency Fingerprints Using Metasurface for the Internet of Things," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1896–1911, 2021.

[14] X. Liu, Z. Chi, W. Wang, Y. Yao, P. Hao, and T. Zhu, "Verification and Redesign of OFDM Backscatter," in *Proceedings of the 18th USENIX Symposium on Networked Systems Design and Implementation*, 2021, pp. 939–953.

[15] M. Wu, X. Lei, X. Zhou, Y. Xiao, X. Tang, and R. Q. Hu, "Reconfigurable intelligent surface assisted spatial modulation for symbiotic radio," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 12 918–12 931, 2021.

[16] Q. Li, M. Wen, L. Xu, and K. Li, "Reconfigurable intelligent surface-aided number modulation for symbiotic active/passive transmission," *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[17] X. Jin, X. Li, Z. Wu, and M. Wen, "Ris-aided joint transceiver space shift keying reflection modulation," *IEEE Communications Letters*, pp. 1–1, 2023.

[18] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for ris-aided mmwave mimo systems," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[19] J. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill, 2008. [Online]. Available: https://books.google.com/books?id=ABSmAQAACAAJ

[20] S. Gupta and G. Wassson, "Performance of ber for bpsk and dpsk (coherent and non-coherent) modulation in turbo-coded ofdm with channel equalization," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, no. 1, 2013.

[21] M. Viswanathan, "Simulation of digital communication systems using matlab," *Mathuranathan Viswanathan at Smashwords*, 2013.