Controller Manipulation Attack on Reconfigurable Intelligent Surface Aided Wireless Communication

Siddharth Sankar Acharjee*  Arpan Chattopadhyay†

Abstract—In this paper, we introduce a new attack called controller manipulation attack (CMA) on a Reconfigurable Intelligent Surface (RIS) assisted communication system between a transmitter and a receiver. An attacker has the potential to manipulate the RIS controller and modify the phase shift induced by the RIS elements. The goal of the attacker is to minimize the data rate at the receiver, subject to a constraint on the attack detection probability at the receiver. We consider two different attack detection models: (i) composite hypothesis testing based attack detection in a given fading block for known channel gains, and (ii) SNR moment based detection over possibly multiple fading blocks. In the first case, a simple energy detector turns out to be uniformly most powerful (UMP) and the attack against this energy detector is designed via a novel optimization formulation and a semidefinite relaxation based solution. In the second case, we consider threshold detection using moments of SNR; various SNR moments under no attack are obtained analytically for large RIS and then used to formulate the attack design problem as a linear program. Finally, numerical results illustrate the performance and tradeoffs associated with the attack schemes, and also demonstrate their efficacy.

Index Terms—RIS, 6G, wireless communication, physical layer security, controller manipulation attack.

I. INTRODUCTION

The last two decades have seen tremendous activities towards developing techniques that exploit the inherent randomness of the wireless propagation medium. Such research efforts traditionally focused on optimizing the transmission and reception schemes. Reconfigurable metasurfaces have shown the potential to be useful as a tool for controlling the wireless propagation medium. Certain electromagnetic properties of these surfaces can be electronically controlled, which helps in their use in passive beamforming for wireless communication without using additional transmit antennas and RF chains. Some well-known implementations of RIS include reflect arrays and software defined metasurfaces [1]–[5], and this technology is envisioned to play a key role in 6G communications.

The RIS consists of multiple reflecting elements; and phase shift induced at each element can be controlled by an RIS micro-controller; see Figure 1. By simultaneously adjusting the phase shifts of all elements, the RIS can fully control the strength and direction of the reflected electromagnetic waves. Active transmit beamforming at the transmitter and passive reflect beamforming at the RIS can be used to optimise performance, such as transmit power reduction and energy efficiency maximisation [6], [7].

Meanwhile, physical layer security has been intensively explored as a counterpart to upper layer encryption approaches for safeguarding wireless security. A significant number of such works focus on enhancing the secrecy performance against a eavesdropper; e.g., [8] that combines transmit beamforming with artificial noise in an RIS based setup, [9] and [10] for using RIS to maximise the secrecy rate, and [11] for lowering the base station transmit power using RIS under a secrecy rate constraint. Wireless transmissions are also subject to jamming attacks; a traditional jammer uses its own energy source to broadcast powerful noise signals to the victim node. However, the authors of [12] suggested an RIS based jammer that makes use of the signal of victim system. Another attack is pilot contamination attack (PCA), where the pilot sequence is manipulated by a eavesdropper using RIS, leading to incorrect channel state information at the transmitter and consequently significant signal leakage to the eavesdropper during the succeeding data transmission phase; see [13]. On the other hand, false data injection (FDI) attack [14]–[16] is a potential threat to networked control systems, where the data flowing in the system is manipulated by an external stealthy attacker to disrupt the estimation and control operations.

In this paper, we propose a new physical layer attack called controller manipulation attack (CMA) on RIS-assisted communication, where an external attacker can manipulate the phase shift induced by the RIS during the data transmission phase. In spirit, CMA is a combination of PCA and FDI; unlike PCA, CMA works on the data and not on the pilot symbols, and unlike FDI, CMA operates at the physical layer. CMA can be carried by an attacker that can either hack the RIS microcontroller or infect it by a malware. As seen later in the paper, CMA results in significant throughput loss to the communication system.

The goal of this paper is to develop a mathematical model for design of stealth CMA. We summarise our major contributions as follows:

1) We propose the CMA attack for the first time, and provide its mathematical model.
2) For the case where the multiple received data symbols in a single fading block are used to detect CMA, we show that a simple energy detector is uniformly most powerful (UMP). In the process, we discover a certain interesting relation between the received signal variance, false alarm probability and attack detection probability.

3) The energy detector is leveraged to formulate an optimization problem for attack design to minimize the received signal to noise ratio (SNR) subject to detection constraint. This non-convex problem is converted to a rank-constrained optimization problem, which is sub-optimally solved via a combination of semidefinite relaxation and Gaussian randomization.

4) For detection over multiple fading blocks, we assume known channel only to the attacker; the detector at the receiver employs a threshold based detection rule on the difference between the empirical SNR moments and the SNR moments under no attack. The attacker’s problem is shown to be a linear program. In the process, we derive closed form expressions of the moment generating function and the first and second order moments of the received SNR under no attack.

II. SYSTEM MODEL

We consider a wireless communication system where a single antenna transmitter (Tx) communicates with a single antenna receiver (Rx) and is assisted by an RIS consisting of N elements. A flat fading model is assumed for all the channels involved. Furthermore, we assume that the line-of-sight (LOS) path is not present between the transmitter and receiver due to severe blockage, and the signals that are reflected from the RIS more than once have negligible strength [6]. We denote the channel between the transmitter and the kth RIS element by \( h_k \), whereas the channel between the kth RIS element and the receiver is denoted by \( g_k \).

We consider independent Rayleigh fading along the lines of [17], i.e., \( h_k = \sqrt{c_k}h_k \) and \( g_k = \sqrt{c_k}g_k \), where \( h_k \sim CN(0,1) \), \( g_k \sim CN(0,1) \), and \( c_k, c_g \) represent the corresponding path losses.

Let \( h_k = \alpha_k e^{i\phi_k} \) and \( g_k = \beta_k e^{i\psi_k} \) where \( \alpha_k = |h_k| \) and \( \beta_k = |g_k| \) represent the channel amplitudes following Rayleigh distribution, while \( \phi_k \) and \( \psi_k \) are the equivalent channel phases, uniformly distributed between \([0, 2\pi]\).

Let the phase shift induced by the kth RIS element be \( \phi_k \). The effect of the RIS can be captured by a matrix \( \Phi = \text{diag}\{e^{i\phi_1}, ..., e^{i\phi_N}\} \) and let \( \Omega = [\phi_1, \phi_2, ..., \phi_N]^T \) be a vector consisting of the corresponding RIS phase shifts.

The baseband signal at the receiver is given by [18]:

\[
y = \sqrt{P} \sum_{k=1}^{N} g_k e^{i\phi_k} h_k x + w
\]

Here \( x \sim CN(0,1) \) represents the transmitted signal drawn from a Gaussian codebook, \( P \) denotes the transmit power, and \( w \sim CN(0, \sigma_w^2) \) denotes the additive white Gaussian noise (AWGN) at the receiver. We can equivalently write \( y \) as [19]:

\[
y = \sqrt{P} g^H \Phi h x + w
\]

where \( h = [h_1, h_2, ..., h_N]^H \) and \( g = [g_1, g_2, ..., g_N]^H \). Let \( \Delta_k = \phi_k + \psi_k + \theta_k \).

The received SNR can be expressed as \( \Gamma = \frac{\kappa}{\sigma_w^2} \) where \( \kappa = P/\sigma_w^2 \), which can further be rewritten as:

\[
\Gamma = \kappa \sum_{k=1}^{N} \alpha_k \beta_k \exp\{j(\phi_k + \psi_k + \theta_k)\}^2
\]

We consider two phase shift models:

Continuous phase shift model: Here, the RIS elements are able to generate any arbitrary phase shift between \([0, 2\pi]\). If the phase shifts generated by the combined propagation channel through the RIS is accurately known, the reflektor phases of each element are adjusted by the RIS microcontroller to achieve zero phase errors. Obviously, when the RIS sets its phases as \( \phi_k^* = - (\theta_k + \psi_k) \) optimum SNR is achieved at the receiver. We denote \( \Phi_0 = \text{diag}\{e^{i\phi_1^*}, ..., e^{i\phi_N^*}\} \) as a matrix corresponding to the optimum phase shifts at the RIS.

Discrete phase shift model: Here the RIS elements are able to realise only a finite, discrete set of phase shifts. Typically, for each element, the collection of possible phase shifts is given by \( D = \{0, \Delta \phi, ..., \Delta \phi(M - 1)\} \), where \( \Delta \phi = 2\pi/M \) and \( M = 2^b \) for \( b \in \mathbb{Z}_+ \); see [7], [17]. Let \( \Phi_k \in D \) be the phase shift induced by the kth element of the RIS. We define the quantization error for the kth element as \( \delta_k = \phi_k - \phi_k^* \), where \( \phi_k^* \) is the optimum phase shift under the continuous phase shift model. Since the Rayleigh fading channel gains \( g_k \) and \( h_k \) are independent across \( k \), and the elements in \( D \) are spaced in regular intervals, \( \delta_k \) is independently and uniformly distributed in \((-\tau, \tau)\), where \( \tau = \pi/2^b \).

We consider two scenarios for the attacker who adversarially controls \( \Omega \) in order to hamper the communication. In the first scenario, attack detection is carried out by the receiver in each fading block at a much faster timescale, and, taking this into account, the attacker seeks to minimise the data rate; see Section III. Here the channel gains are assumed to be known to the attacker and the detector. In the second scenario, attack detection is carried out over multiple fading blocks or at a much slower timescale, and the attacker seeks to minimise the ergodic data rate at the receiver; see Section IV.

III. ATTACK FOR A GIVEN FADING BLOCK

In this section, we assume that the receiver and the attacker have perfect knowledge of the channel gains \( \{h_k, g_k\}_{1 \leq k \leq N} \) over a quasi-static flat fading block under consideration, and \( P = 1 \). The transmitted symbols are drawn from a Gaussian codebook, and hence \( x \sim CN(0,1) \), \( w \sim CN(0, \sigma_w^2) \). Obviously, \( y \sim CN(0, \sigma_w^2 + |g^H \Phi h|^2) \). We define \( \sigma^2 = \sigma_w^2 + |g^H \Phi h|^2 \) and \( \sigma_0^2 = \sigma_w^2 + |g^H \Phi_0 h|^2 \) as the received signal variances under attack and under no attack, respectively. We consider continuous phase shift model in this section.

Lemma 1. \( \sigma^2 \leq \sigma_0^2 \) if \( \Phi \neq \Phi_0 \). Equality holds if \( \Phi = \Phi_0 \).

Proof. See Appendix A

Theorem 1 provides an intuition that any non-optimal phase shift at the RIS results in a smaller signal variance than the optimal case; this observation crucially helps in designing the CMA detector. However, it has to be noted that \( \{\phi_k^* + c\}_{1 \leq k \leq N} \) for any constant \( c \) also maximizes the variance of
Lemma 3. \( s \) simplifies to an energy detector; when the received signal \( \sigma_{\text{opt}} \) from the perspective of an attacker, we seek to find an optimum \( \sigma^2 \) which is realizable via RIS phase shift, minimizes the data rate at the receiver, and achieves a low detection probability.

A. Attack detection via hypothesis testing

We formulate the detection problem as a binary hypothesis testing problem. Let us assume that \( K \) symbols are transmitted over a fading block, and the receiver detects a possible CMA at the end of the fading block. The two hypotheses are:

\[
H_0 : \Phi = \Phi_0 \\
H_1 : \Phi \neq \Phi_0
\]

Clearly, the null hypothesis means that there is no attack, and the alternate hypothesis implies an attack. Since the receiver receives \( K \) i.i.d observations \( \{y_i\}_{1 \leq i \leq K} \) in a fading block, where \( y_i \sim \mathcal{CN}(0, \sigma^2) \), the two hypotheses can be equivalently represented as:

\[
H_0 : \sigma = \sigma_0 \\
H_1 : \sigma < \sigma_0
\]

Lemma 2. The optimum likelihood ratio test reduces to

\[
W = \sum_{i=1}^{K} \frac{||y_i||^2}{\eta^2} \leq 2 \eta' \quad \forall \eta' > 0 \text{ is a threshold.}
\]

Proof. See Appendix [3]

In Lemma 2, we observe that the likelihood ratio test simplifies to an energy detector; when the received signal power (or SNR) is low, the detector declares an attack.

Lemma 3. If \( y_i \sim \mathcal{CN}(0, \sigma^2) \) for any \( \sigma > 0 \), then \( 2 \lambda W \) follows the Chi-squared distribution (\( \chi^2 \) distribution) with \( 2K \) degrees of freedom, where \( \lambda = \frac{1}{\sigma^2} \) and \( W = \sum_{i=1}^{K} ||y_i||^2 \).

Proof. See Appendix [4]

The probability of false alarm (PFA) is defined as, \( P_{FA} = \mathbb{P}(W \leq \eta' | H_0) \). Lemma 3 helps us in deriving \( P_{FA} \) of the detector and thus helps in attack design. Let \( \rho \) represent the significance level of a test and \( R_{2K, \rho} \) denote the inverse cumulative distribution function of the chi-squared probability density function (PDF) with \( 2K \) degrees of freedom, evaluated at a probability \( \rho \in [0, 1] \).

Theorem 1. The likelihood ratio test described in Lemma 2 is a Universally Most Powerful (UMP) test with a threshold \( \eta' = R_{2K, \rho} \sigma^2 \).

Proof. See Appendix [5]

B. Designing the attack against the UMP detector

Now we design an attack strategy against the UMP test, assuming that the attacker has knowledge of the threshold \( \eta' \). The probability of detection \( (P_D) \) can be written as follows:

\[
\mathbb{P}(2\lambda W \leq 2\lambda \eta' | H_1)
\]

From the perspective of an attacker, we seek to find an optimum \( \sigma^2 \) which is realizable via RIS phase shift, minimizes the data rate at the receiver, and achieves a low detection probability.

Theorem 2. The value of \( \sigma^2 \) for a particular probability of detection \( \xi \) and a known threshold determined by \( \rho \) as per Theorem 1 is given by \( \sigma^2 = \frac{R_{2K, \rho} \sigma^2}{R_{2K, \xi} \sigma_0^2} \).

Proof. See Appendix [6]
optimal solution to (9) as \( \bar{s} = Q \sqrt{\sum f} \), here \( f \) denotes a random vector which is independently generated from \( CN(0, I_{N+1}) \). Finally, a candidate solution to (7) is given by:

\[
\Omega = \arg \left( \frac{\bar{s}}{\bar{s}(N + 1)} \right)_{(1:N)}
\]

(11)

where \( \bar{s}(N + 1) \) denotes the \( N + 1 \)th element of \( \bar{s} \) and \( [\bar{s}]_{1:N} \) represents the first \( N \) elements of the vector \( \bar{s} \). Hence, by using (11), a candidate value of \( \Omega \) is obtained. However, if this candidate \( \Omega \) does not satisfy all constraints in (7), then this solution is discarded. By generating a large number of i.i.d. samples for \( f \) and evaluating only the feasible ones, we can determine the best \( s \) which satisfies the constraints in (8) with high probability; otherwise, this is a sub-optimal solution with high precision [22]. The entire mechanism is summarised in Algorithm 1.

Algorithm 1

Input: \( g, h \), the parameter \( \nu \), a large integer \( E \).

Output: \( \Omega \).

1. Solve (10) and derive the value of \( S \).
2. for \( e = 1, 2, 3, \ldots, E \) do
   3. Sample \( \bar{f}(e) \) from \( CN(0, I_{N+1}) \).
   4. Compute \( \Omega(e) \) using (11).
   5. If \( \Omega(e) \) is a feasible solution of (7), evaluate the objective function of (7) at \( \Omega(e) \).
3. end for
4. Find best feasible solution from \( \{ \Omega(e) \}_{1 \leq e \leq E} \) for (7).

IV. ATTACK OVER MULTIPLE FADING BLOCKS

In this section, we design the attack under the discrete phase shift model, though a similar treatment is possible for the continuous phase shift model as well. We assume that the fading in the wireless links is fast and i.i.d. Rayleigh across fading blocks. The detector at the receiver collects \( \{y_1, y_2, \ldots, y_T\} \) over \( T \) fading blocks, using which it has to decide whether an attack has occurred or not. If the detector knows the channel gains at each time, then it can simply employ an energy detector as in Section III and the theory for attack detection and design remains unchanged, subject to some scaling operations. However, in this section, we assume that the attacker knows \( \{k_h, k_g\}_{1 \leq k \leq N} \) in each fading block, but the receiver does not know it. Hence, the detector seeks to check whether the empirical distribution of SNR over \( T \) fading blocks matches with the SNR distribution under no attack. This can be done by performing a goodness-of-fit test, but we consider a simpler detector that only compares a few empirical moments of received SNR with their desired values; this helps in restricting the number of constraints in the optimization problem of the attacker in Section IV-B.

After receiving \( \{y_1, y_2, \ldots, y_T\} \), the detector computes the empirical moments of SNR, denoted by \( SNR_l \) for the \( l \)th moment, and declares that an attack has happened if and only if \( SNR_l - SNR_l^\prime \geq \zeta_l \), where \( \zeta_l \) represents a threshold and \( SNR_l \) denotes the \( l \)th moment of SNR under no attack. This condition can be checked for multiple values of \( l \). In this paper, we focus only on \( l = 1, 2 \), though our results can be extended for any \( l \). In Section IV-A, we first derive \( SNR_l \) in closed form for both continuous and discrete phase shift models, which is later used in Section IV-B for attack design.

A. Moments of SNR under no attack

Let us consider Rayleigh channels as discussed in Section II and define the moment generating function (M.G.F) of SNR \( \Gamma^* \) as \( M_{\Gamma^*}(t) = \mathbb{E}(e^{t \Gamma^*}) \). The \( l \)th moment of SNR is \( SNR_l = \mathbb{E}((\Gamma^*)^l) = \frac{d^l M_{\Gamma^*}(t)}{dt^l} \bigg|_{t=0} \).

1) Continuous phase shift model: We note from (3) that the maximum SNR is expressed as \( \Gamma^* = \tilde{r}_l \sum_{k=1}^{N} \alpha_k \beta_k |^2 = \tilde{r}_l |Z|^2 \), where \( Z = \sum_{k=1}^{N} \alpha_k \beta_k \).

Theorem 3. For large number of reflecting elements (\( N \rightarrow \infty \)), by central limit theorem the coefficient \( Z = \sum_{k=1}^{N} \alpha_k \beta_k \) can be approximated by a random variable with distribution \( N(\mu_Z, \sigma_Z^2) \), where \( \mu_Z = N \sqrt{\epsilon_k \epsilon_g \frac{\pi}{2}} \), \( \sigma_Z^2 = N \epsilon_k \epsilon_g (1 - \frac{\gamma}{16}) \). The SNR can be approximated by a non-central chi-squared distribution with one degree of freedom, also its PDF, M.G.F and moments are expressed as follows:

\[
f_{l^*}(\gamma) = \frac{1}{2\sqrt{2\pi \sigma^2}} e^{-\frac{(\gamma - \mu)^2}{2\sigma^2}}
\]

\[
M_{l^*}(t) = e^{\gamma t} (1 - 2\epsilon \sigma t)^{-1/2}
\]

\[
E(\Gamma^*) = N \kappa \epsilon \epsilon_g \left( 1 + \frac{\pi^2}{16}(N - 1) \right) + \frac{\pi^2}{8}(N - 1) + \frac{16 - \pi^2}{8} + \frac{\pi^2(2N - 1)(16 - \pi^2)}{128}
\]

Proof. See Appendix F.

2) Discrete phase shift model: By (3), the SNR received under discrete phase shift model is:

\[
\Gamma^* = \tilde{r}_l \sum_{k=1}^{N} \alpha_k \beta_k e^{j\delta_k} \]

(12)

where \( V \) is distributed by \( \Phi \), where \( \omega \) is distributed by \( \Phi \).

Theorem 4. For large number of reflecting elements (\( N \rightarrow \infty \)), by central limit theorem \( V = \nu + jV_{I} \) can be approximated to a complex Gaussian distribution. The real and imaginary parts \( V_R \) and \( V_I \) are independent with \( V_R \sim N(\mu, \sigma_R^2) \) and \( V_I \sim N(0, \sigma_I^2) \), where \( \mu = \frac{\pi}{4} \sqrt{\epsilon_k \epsilon_g \varphi \delta_k(1)} \), \( \sigma_R^2 = \frac{\varphi \delta_k(1)}{2N}(1 + \varphi \delta_k(2)) \) and \( \sigma_I^2 = \frac{\varphi \delta_k(1)}{4N}(1 - \varphi \delta_k(1)) \). Also, defining \( \theta |V|^2 = 4\sigma^2 V_R \) and \( k |V|^2 = \frac{\varphi \delta_k}{4\sigma V} \), the SNR can be approximated by a Gamma distribution with its PDF, M.G.F and moments expressed as follows:

\[
f_{l^*}(\gamma) = \frac{\gamma^{k_{V^2} - 1} e^{-\frac{k_{V^2} \gamma^2}{2}}}{(\gamma N^{2} \sigma_{V^2})^{k_{V^2}/2}}
\]

\[
M_{l^*}(t) = (1 - t \theta |V|^2)^{-1/2}
\]

\[
E(\Gamma^*) = \frac{N^2 \pi^2}{16} e^{|\epsilon_k | \varphi \delta_k(1)^2}
\]

\[
E(\Gamma^*) = e^{\epsilon_k^2} \left( \frac{N^3 \pi^2}{8} (1 + \varphi \delta_k(2)) + \frac{\pi^2 \varphi \delta_k(1)^2}{256} \right)
\]
Proof. See Appendix A.

Using Theorem 4 we can find out any moment of SNR under the discrete phase shift model.

B. Attack design: optimization problem formulation

Let $s$ be a typical state of the attacker, which consists of the instantaneous channel gains $\{h_k, g_k\}_{1 \leq k \leq N}$ available to the attacker. However, we discretize the state space to obtain a finite number of states. A generic action $a$ of the attacker is a vector of phase shifts $[\phi_1, \ldots, \phi_N] \in D^N$; the attacker takes an action in each fading block. The signal to noise ratio under a state-action pair is denoted by $SNR(s, a)$, and this can be computed by the attacker using $\hat{\phi}$. Moreover, $SNR_t$ can be calculated by an attacker using either Theorem 4 for large $N$ or by brute-force computation for small $N$. The attacker seeks to minimize the ergodic data rate at the receiver while roughly preserving the moments of SNR.

Let $\pi(s)$ denote the probability of occurrence of channel state $s$. We consider probabilistic action selection by the RIS at each time; let $p(a|s)$ denote the probability of choosing action $a$ under state $s$ on part of the attacker. The attacker seeks to minimize the ergodic data rate at the receiver, while preserving the moments of the received SNR. The attacker’s problem can be cast as the following optimization problem:

$$\min_{p(a|s)\forall a,s} \sum_s \pi(s) \sum_a p(a|s) \log(1 + SNR(s, a))$$

subject to

$$p(a|s) \geq 0 \quad \forall a, s; \quad \sum_a p(a|s) = 1 \quad \forall s$$

(13)

Let us consider $m$ actions and $n$ states. We denote $P \triangleq [p_1, \ldots, p_N]^T$, where $p_k \triangleq [\pi(s_k)1_m]^T$, and $1_m$ is an all-1 column vector of dimension $m \times 1$. Let $x \triangleq [x_1, \ldots, x_k, \ldots, x_N]^T$, where $x_k \triangleq [p(a_1|s_k), \ldots, p(a_m|s_k)]$. Also, $C \triangleq \text{diag}(t_1, \ldots, t_k, \ldots, t_N)$ where $t_k \triangleq [SNR(s_k, a_1), \ldots, SNR(s_k, a_m)]$, $K \triangleq \text{diag}(u_1, \ldots, u_k, \ldots, u_N)$ where $u_k \triangleq [\log(1 + SNR(s_k, a_1)), \ldots, \log(1 + SNR(s_k, a_m))]$, and $R \triangleq C^T$. The optimization problem (13) can be rewritten as a linear optimization problem as follows:

$$\min_{x} \quad p^T K x$$

subject to

$$x \geq 0, \quad Ax = 1_{n \times 1}, \quad p^T C x \leq SNR_1 + \zeta_1, \quad p^T C x \geq SNR_1 - \zeta_1$$

$$p^T R x \leq SNR_2 + \zeta_2, \quad p^T R x \geq SNR_2 - \zeta_2$$

(14)

The matrix $A$ is a logical matrix used to represent the constraint $\sum_a p(a|s) = 1 \forall s$, and its entries are chosen accordingly. This linear program can be solved by using traditional solvers, and its solution can be used by the attacker.

V. NUMERICAL RESULTS

We consider a Rayleigh fading model, i.e., $h_k, g_k \sim \mathcal{CN}(0, 1)$. We also consider $P = 30$ dBm, $\sigma_w^2 = -10$ dBm.

UMP test over a given fading block: We consider $N = 64$ and $K = 50$, and evaluate the performance of the proposed attack scheme over 10000 independently generated channel realizations. For various false alarm rates, we evaluate the threshold ($\eta'$) and use it to find $\sigma^2$ for a given probability of detection. Further, using Algorithm 1 we find out the optimum phase shift matrix ($\Omega$) and employ it to attack the generated instances. Fig 2 shows that, as the probability of detection is increased, the attacker can decrease the data rate by a considerably large value which is also intuitive. Additionally, we observe that as the probability of false alarm increases for a fixed detection probability, the percentage decrease in data rate also reduces. This is due to the fact that, as the false alarm increases, the detector becomes more cautious and this leaves limited opportunity for the attacker to reduce the data rate.

SNR based detector over multiple fading blocks: We consider an RIS with $N = 8$ and four possible phase shifts ($b = 2$). All results are averaged over 100 independent channel realizations. Table I shows that less stringent constraints result in more reduction in the ergodic data rate.

However, it is important to note that, the number of variables in (13) is $2^{bN}$. This necessitates development of low complexity algorithms for large $N$.

VI. CONCLUSION

In this paper, we have analytically designed controller manipulation attack against UMP energy detector over a single fading block and SNR based detector over multiple fading blocks. In the first case, a higher detection tolerance of the attacker enables it to cause more harm to the received data rate. Also, for the SNR based detection over multiple fading blocks, higher detection tolerance results in a much reduced ergodic data rate. In future, we will develop low complexity attack schemes for multiple fading blocks.
Appendix A
Proof of Lemma [1]
We know, \( \Lambda_k = \phi_k + \psi_k + \theta_k \). Now,
\[
\sigma^2 = \sigma_w^2 + |g^H \Phi h|^2 = \sigma_w^2 + \sum_{k=1}^{N} \alpha_k \beta_k e^{j \Lambda_k} |^2
\]
Obviously, \( \sigma^2 \) achieves the maximum variance \( \sigma_0^2 \) only if \( \Lambda_k = 0 \) for all \( 1 \leq k \leq N \), i.e., if \( \Phi = \Phi_0 \).

Appendix B
Proof of Lemma [2]
The likelihood ratio is:
\[
L_{\alpha}(y) = \frac{p_{\sigma}(y; H_1)}{p_{\sigma_0}(y; H_0)} = \frac{1}{\frac{1}{\pi \sigma_0^2 \epsilon_k}} e^{-\frac{\sum_{k=1}^{K} ||y_i||^2}{\sigma_0^2}}
\]
Taking logarithm, the likelihood ratio test becomes:
\[
\ln L_{\sigma}(y) = K \ln \frac{\sigma_0^2}{\sigma^2} + \sum_{i=1}^{K} \frac{||y_i||^2}{\sigma^2} - \frac{H_1}{\sigma^0} - \frac{H_0}{\sigma^0} \approx \eta
\]

Appendix C
Proof of Lemma [3]
Let \( y_i \sim \mathcal{CN}(0, \tilde{\sigma}^2) \) for any arbitrary \( \tilde{\sigma}^2 > 0 \). Now, \( ||y||^2 = Y_R^2 + Y_I^2 \), where \( Y_R \) and \( Y_I \) are real and imaginary parts of \( y_i \). Obviously, \( Y_R \sim \mathcal{N}(0, \tilde{\sigma}^2/2) \), \( Y_I \sim \mathcal{N}(0, \tilde{\sigma}^2/2) \). Since \( Y_R^2 \) and \( Y_I^2 \) are chi-squared distributed random variables with one degree of freedom, hence \( ||y||^2 \) is an exponentially distributed random variable with parameter \( \lambda = 1/\tilde{\sigma}^2 \).

Appendix D
Proof of Theorem [1]
From Lemma [2] we obtain:
\[
P_{FA} = \mathbb{P}(W \leq \eta' | H_0) = \mathbb{P}(2\lambda_0 W \leq \eta' | H_0) = \rho \quad (15)
\]
where \( \lambda_0 = 1/\sigma_0^2 \) and \( \eta' = 2\lambda_0 \eta' \). Using (15), the detector chooses the threshold \( \eta' \) based on a required probability of false alarm \( \rho \). We have \( \eta' = R_{2K, \rho} \) and hence the threshold \( \eta' = \eta''/2\lambda_0 \) is independent of the value of the parameter \( \sigma^2 \) (and \( \Phi \)) and the test is UMP.

Appendix E
Proof of Theorem [2]
We are given that \( \xi = \mathbb{P}(2\lambda W \leq 2\lambda \eta' | H_1) \). Now, we know that \( 2\lambda \eta' = R_{2K, \xi} \), where \( \lambda = 1/\sigma^2 \) and \( \eta' = R_{2K, \xi} / 2\lambda_0 \). This yields:
\[
\frac{R_{2K, \rho}}{2\lambda_0} = \frac{R_{2K, \xi}}{2\lambda} \quad \Rightarrow \quad \frac{R_{2K, \rho} \sigma_0^2}{R_{2K, \xi}} = \frac{\sigma_0^2}{\sigma^2}
\]

Appendix F
Proof of Theorem [3]
Note that, \( \alpha_k \) and \( \beta_k \) are independent Rayleigh distributed random variables and \( X_k = \alpha_k \beta_k \). Also, \( \mathbb{E}(X_k) = \mu_x = \sqrt{\epsilon_k \gamma \sigma_0^2} \) and \( \text{Var}(X_k) = \sigma_x^2 = \varepsilon_k \epsilon_\gamma ^2 (1 - \pi^2 / 16) \). By applying Central Limit Theorem (CLT) for large \( N \) we have:
\[
\Xi_N \sim \mathcal{Z} - N \mathcal{Z} \quad \text{d} \mathcal{N}(0, 1)
\]
Hence, for large \( N \), we can write the mean and variance of \( Z \) as \( \mu_z = N \sqrt{\epsilon_k \gamma \sigma_0^2} \) and \( \sigma_z^2 = N \varepsilon_k \epsilon_\gamma ^2 (1 - \pi^2 / 16) \). Also, since \( \text{SNR} (\Gamma^*) = \eta(\sum_{k=1}^{K} \alpha_k \beta_k) = \eta(\tilde{Z})^2 \), the SNR follows a non-central chi-squared distribution with one degree of freedom. Hence, we can write its P.D.F as:
\[
f_{\Gamma^*}(\gamma) = \frac{1}{2 \sqrt{2 \pi \kappa_\gamma^2 \gamma}} \left( e^{-(\sqrt{\gamma} - \mu_\gamma)^2 / \gamma} + e^{-(\sqrt{\gamma} + \mu_\gamma)^2 / \gamma} \right)
\]
and its M.G.F as:
\[
M_{\Gamma^*}(t) = e^{t \sqrt{\frac{\mu_\gamma}{\gamma}} (1 - 2 \kappa_t \mu_\gamma^2) / 2} (1 - 2 \tilde{\kappa} \sigma_\gamma^2)^{-t / 2}
\]
Hence, the first moment \( S \tilde{N} R_1 \) and second moment \( S \tilde{N} R_2 \) of SNR can be expressed as:
\[
\mathbb{E}(\Gamma^*) = \tilde{\kappa} \mathbb{E}(Z^2) = \tilde{\kappa} (\mu_z^2 + \sigma_z^2) = N \tilde{\kappa} \varepsilon_\gamma \sigma_0^2 (1 + \frac{\pi^2}{16} (N - 1))
\]
\[
\mathbb{E}(\Gamma^*^2) = N^2 \tilde{\kappa}^2 e_{\gamma^2}^2 \left( 1 + \frac{\pi^4}{256} (N - 1)^2 + \frac{\pi^2}{8} (N - 1) + \frac{16 - \pi^2}{8} \right) \left( \frac{\pi^2 (2N - 1) (16 - \pi^2)}{128} \right)
\]

Appendix G
Proof of Theorem [4]
Let us define \( V = \frac{1}{\lambda_0} \sum_{k=1}^{K} V_k \), where \( V_k = \alpha_k \beta_k e^{j \delta_k} \). We observe that as the distribution of \( \delta_k \) is symmetric around the origin, and hence its characteristic function \( \varphi_{\delta_k}(\omega) \) is always real. Hence, \( \mu = \mathbb{E}(V_k) = \mathbb{E}(\alpha_k) \mathbb{E}(\beta_k) \mathbb{E}(e^{j \delta_k}) = \frac{\pi}{4} \sqrt{\varepsilon_k \gamma} \varphi_{\delta_k}(1) \). Similarly, variance \( \nu = \mathbb{E}(V_k e^{j \delta_k}) (1 - \frac{\pi^2}{16} \varphi_{\delta_k}(1)^2) \) and pseudo-variance \( \bar{\rho} = \mathbb{E}(V_k (\varphi_{\delta_k}(2) - \frac{16}{\pi^2} \varphi_{\delta_k}(1)^2)^2) \). By applying
central limit theorem, for large $N$ we can approximate the distribution of $V$ as $\mathcal{CN}(\mu, \frac{\sigma}{\sqrt{N}}, \frac{\sigma}{\sqrt{N}})$. Let $V_R$ and $V_I$ denote the real and imaginary parts of $V$, respectively. We can write:

$$
\text{Cov}(V_R, V_I) = \frac{1}{2} \text{Im}(\frac{\nu}{N} + \frac{\rho}{N}) = 0.
$$

Since $V_R$ and $V_I$ are jointly Gaussian and uncorrelated, they are mutually independent. Also, $V_R \sim \mathcal{N}(\mu_{V_R}, \sigma_{V_R}^2)$ and $V_I \sim \mathcal{N}(0, \sigma_{V_I}^2)$, where $\mu_{V_R} = \mu$ and $\sigma_{V_R}^2 = \frac{1}{2} \text{Re}(\frac{\nu}{N} + \frac{\rho}{N}) = \frac{\epsilon_h \epsilon_g}{2N} (1 + \phi_\delta_k (2) - \frac{\pi^2}{8} \phi_\delta_k (1)^2)$ (18)

$$
\sigma_{V_I}^2 = \frac{1}{2} \text{Re}(\frac{\nu}{N} - \frac{\rho}{N}) = \frac{\epsilon_h \epsilon_g}{2N} (1 - \phi_\delta_k (2))
$$

By using the results derived in [25, Theorem 1, Appendix B] and [26] we can approximate the distribution of $|V|^2$ (for large $N$) as a gamma distribution with shape parameter $k_{V^2} = \frac{\mu_{V_R}^2}{\sigma_{V_R}^2}$ and scale parameter as $\theta_{V^2} = 4\sigma_{V_R}^2$. From (13) we write the SNR as $\Gamma^* = \bar{\kappa} N^2 |V|^2$. After rewriting, we obtain $\Gamma^* = \varrho |V|^2$, where $\varrho = \bar{\kappa} N^2$ and $\varrho > 0$. Hence, we can infer from above that as $|V|^2$ follows gamma distribution hence the SNR also follows a gamma distribution with shape parameter as $k_{\Gamma^*} = k_{V^2}$ and scale parameter as $\theta_{\Gamma^*} = \varrho \theta_{V^2}$. As a result, we can write the PDF of SNR as:

$$
\tilde{f}_{\Gamma^*}(\gamma) = \frac{\gamma^{k_{\Gamma^*}-1} e^{-\frac{\gamma}{\theta_{\Gamma^*}^2}}}{\theta_{\Gamma^*}^2 \Gamma(k_{\Gamma^*})} = \frac{\gamma^{k_{V^2}-1} e^{-\frac{\gamma}{\theta_{V^2}^2}}}{(\bar{\kappa} N^2 \varrho)^{k_{V^2}} \Gamma(k_{V^2})}
$$

Hence, the M.G.F of $|V|^2$ is $M_{|V|^2}(t) = (1 - t \theta_{|V|^2})^{-k_{V^2}}$, where $\theta_{|V|^2} = \frac{\epsilon_h \epsilon_g (1 + \phi_\delta_k (2) - \frac{\pi^2}{8} \phi_\delta_k (1)^2)}{N}$ is large for large $N$. Further, the M.G.F of SNR $\Gamma^*$ can be expressed as $M_{\Gamma^*}(t) = (1 - t \varrho \theta_{|V|^2})^{-k_{V^2}}$. [23, Chapter 5, Table 5-2].

The $i^{th}$ moment of SNR is given by:

$$
SNR_i = E \left( \Gamma^i \right) = M_{\Gamma^*}^{(i)}(0) = \frac{d^i M_{\Gamma^*}}{dt^i} \bigg|_{t=0}
$$

Hence, the first moment $SNR_1$ and second moment $SNR_2$ are:

$$
E(\Gamma^*) = \varrho k_{V^2} \theta_{V^2} = \bar{\kappa} N^2 \mu_{V_R}^2 = \frac{N^2 \bar{\kappa} R^2}{16} \frac{\epsilon_h \epsilon_g \phi_\delta_k (1)^2}{16}
$$

$$
E(\Gamma^{*2}) = \varrho^2 k_{V^2} \theta_{V^2} ^2 (1 + k_{V^2}) = \bar{\kappa} N^4 \mu_{V_R}^2 (4 \sigma_{V_R}^2 + \mu_{V_R}^2)
$$

$$
= \frac{\epsilon_h \epsilon_g}{4} \left( N^3 \pi^2 - \frac{\pi^2 \phi_\delta_k (1)^2}{8} \right) + \frac{N^4 \pi^4 \phi_\delta_k (1)^4}{256}
$$