A Cooperative Deception Strategy for Covert Communication in Presence of a Multi-Antenna Adversary

Jiangbo Si, Senior Member, IEEE, Zizhen Liu, Zan Li, Senior Member, IEEE, Hang Hu, Lei Guan, Chao Wang, Member, IEEE, and Naofal Al-Dhahir, Fellow, IEEE

Abstract—Covert transmission is investigated for a cooperative deception strategy, where a cooperative jammer (Jammer) tries to attract a multi-antenna adversary (Willie) and degrade the adversary’s reception ability for the signal from a transmitter (Alice). For this strategy, we formulate an optimization problem to maximize the covert rate when three different types of channel state information (CSI) are available. The total power is optimally allocated between Alice and Jammer subject to the Kullback-Leibler (KL) divergence constraint, which can be expressed analytically and be widely used as a covertness measurement. Different from the existing literature, in our proposed strategy, we also determine the optimal transmission power at the jammer when Alice is silent, while existing works always assume that the jammer’s power is fixed. Specifically, we apply the S-procedure to convert infinite constraints into linear-matrix-inequalities (LMI) constraints. When statistical CSI at Willie is available, we convert double integration to single integration using asymptotic approximation and substitution method. Finally, our simulation results show that for the proposed strategy, the covert rate is increased with the number of antennas at Willie. Moreover, compared to the benchmark, our proposed strategy is more robust in the presence of imperfect CSI.

Index Terms—Cooperative deception, covert transmission, CSI, multi-antenna adversary, power allocation.

I. INTRODUCTION

COVERT transmission is an emerging high-security communication technique to transmit confidential information without being discovered by an adversary, which is different from traditional communication security techniques and physical layer security (PLS) techniques [1], [2]. Traditional communication security relies on encryption systems. However, with the continuous advancement of technology, encrypted information may be deciphered by adversaries, and there is a risk of information leakage. Physical layer security relies on the characteristics of the wireless channel to realize secure transmission [3]. Recently, some new methods have been proposed for PLS. A learning-aided content-based image transmission scheme is proposed in [4], where fountain-based packet delivery and a deep neural network (DNN) are deployed to reinforce the system’s security. Furthermore, to mitigate hardware impairments and the malicious attack, a deep learning based method is proposed to guarantee the security of wireless secret key generation in [5] and [6]. However, in PLS, the communication link can be exposed to the eavesdroppers. Therefore, as users become more concerned about privacy, both industry and academia are paying more attention to covert communication [7].

In the previous research on covert communication, researchers studied the performance of covert communication for various channel models, and proved the existence of the Square Root Law (SRL), i.e., \[ \lim_{n \to \infty} \frac{\sqrt{n}}{n} = 0 \] [8], [9], [10], which was first discovered in the AWGN channel by Bash et al. [8]. To increase the covert transmission capacity, many researchers exploited the uncertainties of the system in order to reduce the correct detection probability of the adversary. These uncertainties include noise, channel, power, and transmission slot uncertainty [11], [12], [13], [14]. The authors proved that when noise uncertainty obeys a certain distribution, the covert communication rate can be improved [11]. The influence of channel uncertainty was studied in the Binary Symmetric Channel (BSC), and it also helped to increase the covert rate [12]. Additionally, a random transmit power was able to enhance covert communications performance [13]. Moreover, the authors showed that a positive covert rate was realized with the help of Alice’s transmission slot uncertainty [14]. Since the uncertainties of the channel itself were relatively limited and hard to control, researchers further used jamming or relaying nodes to transmit artificial noise (AN) for increasing the noise uncertainty and

Authorized licensed use limited to: University of Wollongong. Downloaded on July 04,2023 at 07:08:12 UTC from IEEE Xplore. Restrictions apply.
obtaining better transmission performance [15], [16], [17]. A friendly uniformed jammer generated AN to help the covert transmission between Alice and Bob in [15] and [16]. In addition, for the two-hop wireless relaying systems, a full-duplex relay transmitted the jamming signal towards the adversary [17].

Most existing works investigate covert transmission under the assumption that the channel state information (CSI) of all links is perfectly known at Alice [8], [11], [12], [13], [14], [15], [16], [17]. However, due to the channel estimation error and the hostility of an adversary, it is challenging for Alice to obtain all links’ instantaneous CSI. Thus, a few papers studied the covert transmission performance under imperfect CSI and statistical CSI. Several schemes were proposed to improve the covert transmission performance in [18], [19], [20] and [21]. With imperfect CSI and instantaneous CSI, optimal beamforming and power allocation schemes were investigated in [18], where a regular user was deployed to cover Alice’s covert transmission. For the imperfect CSI case, a robust beamforming scheme was proposed to maximize the covert rate over a multiple-input single-output (MISO) system [19]. With the statistical CSI at the adversary, the authors of [20] found that the performance of covert transmission improved significantly with the help of an Intelligent Reflecting Surface (IRS). Also, the performance of covert transmission assisted by IRS was studied under both imperfect CSI and statistical CSI in [21].

With the maturity of multi-antenna technology, several works have focused on the covert transmission performance for multi-antenna devices, such as [18], [22] and [23]. Deploying multiple antennas at Alice, an increase in the covert rate was achieved with increasing the number of antennas [18]. When a cooperative jammer was equipped with multiple antennas, beamforming was used in [22] to maximize the covert transmission performance. However, to the best of our knowledge, few works investigated the covert performance when multiple antennas were deployed at the adversary. It is shown in [23] that a slight increase in the number of the adversary’s antennas can significantly reduce the covert rate. However, this work did not consider the jammer’s effect on the covert performance and the cooperation between the jammer and Alice is neglected.

Motivated by this background, to improve the covert rate in presence of a multi-antenna adversary, we propose a novel cooperation deception strategy where Alice and the jammer cooperate in deceiving the multi-antenna adversary. Specifically, while Alice is silent, the jammer injects AN. When Alice transmits the message covertly, the jammer adjusts the transmission power to attract the adversary’s attention and cover Alice’s transmission. Our proposed deception strategy is motivated by one of thirty-six schemes in Sun Tzu’s Art of War: “make a feint to the east but attack in the west”. Thus, when Willie is attracted by the Jammer and neglects the signal from Alice, it naturally cannot estimate the CSI of the Alice-Wille link. In [24] and [25], a receiver (a jammer) is deceived to jam the victim frequency band and maintain the legitimate user communication at other frequency bands. In this paper, the receiver (Willie) is deceived to focus on the signal direction from the Jammer and the covert communication from different direction can be maintained. Namely, [24] and [25] deceived the receiver in the frequency domain, and our paper deceived the receiver in the space domain. With three different kinds of CSI, the transmission power at Alice and the jammer are optimally allocated to maximize the covert rate under the Kullback-Leibler (KL) divergence constraint which provides an upper bound on the total variation between the likelihood function under the null hypothesis and the likelihood function under the alternative hypothesis, and can be expressed analytically. To further explore the pros and cons of our proposed strategy, we also study the strategy where the jammer does not deceive the adversary as a benchmark. The main contributions of this paper are summarized as follows:

- We propose a cooperative deception strategy for covert transmission in presence of a multi-antenna adversary, where a jammer is used to attract the adversary’s attention in the space domain and cover the Alice’s transmission. As a result, the adversary adjusts it’s weighting vector towards the jammer’s signal direction, while it cannot focus its attention towards Alice’s signal direction. To the best of our knowledge, this is the first work that uses a deception strategy to effectively improve the covert rate, filling the gap in previous multi-antenna Willie research.
- For our proposed deception strategy, under the both hypotheses that Alice is silent and not, the transmission power at the jammer and Alice are jointly optimized to maximize the covert rate under three kinds of CSI, i.e., instantaneous CSI, imperfect CSI, and statistical CSI.
- Different methods are deployed to solve the optimization problems for covert rate maximization. For the instantaneous CSI case, we maximize the covert rate under both the covert equality and the total power constraints. For the imperfect CSI case, to solve the non-convex problem, the S-procedure is used to transform infinite constraints into LMI constraints. Furthermore, for the statistical CSI case, we derive the statistical constraint expression under the help of asymptotic approximation which ignores the noise power.
- To gain a deep insight into the performance of our proposed deception strategy, the covert rate without jammer deception, with jammer deception but fixed power, with jammer deception but uniform distributed transmit power are studied as a benchmark. Numerical results show that our proposed strategy can achieve a higher covert rate than the benchmarks, and for the deception strategy the number of antennas at adversary plays a positive effect on the covert rate. Moreover, the proposed deception strategy is more robust than the benchmarks under the imperfect CSI case.

The remainder of this paper is organized as follows. Section II describes the system model. The cooperative deception strategy and the covert constraint are illustrated in detail. In Section III, the covert rate with cooperative deception strategy is investigated under three different kinds of CSI. A special case is studied as a bench-
mark in Section IV. Numerical results are presented and discussed in Section V. Finally, we conclude this paper in Section VI.

Notations: $(\cdot)^H$ denotes the conjugate transpose; The operator $|\cdot|$ denotes the absolute value; $||\cdot||$ denotes the Frobenius norm; $\text{Re}(\cdot)$ denote the real part of a complex random variable (RV); $\exp(\sigma^2)$ denotes exponential distribution having mean $\sigma^2$; $\arg(\cdot)$ denotes the argument of a complex number; $\circ$ denotes the circularly symmetric Gaussian distribution with mean $\mu$ and variance $\sigma^2$.

## II. SYSTEM MODEL

As shown in Fig. 1, Alice tries to transmit messages to Bob without being detected by an adversary (Willie), who tries to detect whether there exists a transmission from Alice or not. Moreover, Jammer, Alice’s coordinator, tries to confuse Willie and helps to cover Alice’s transmission.

In this model, we consider a scenario where Willie uses $M > 1$ antennas for detection, whereas Alice, Jammer and Bob are equipped with a single antenna. This assumption is reasonable for military and national security scenario, where due to size and complexity constraints, Alice, Bob and Jammer are equipped with a single antenna, while Willie is responsible for the national security, and is equipped with multiple antennas. Each channel is assumed to follow the Rayleigh fading model. The instantaneous CSI of the Alice $\rightarrow$ Bob and Jammer $\rightarrow$ Bob links are, respectively, denoted by $h_{ab} \sim \mathcal{CN}(0, \sigma_{ab}^2)$ and $h_{jb} \sim \mathcal{CN}(0, \sigma_{jb}^2)$. In addition, $h_{aw} \in \mathbb{C}^{M \times 1} \sim \mathcal{CN}(0, \sigma_{aw}^2 I)$ and $h_{jw} \in \mathbb{C}^{M \times 1} \sim \mathcal{CN}(0, \sigma_{jw}^2 I)$ denote the instantaneous CSI of the Alice $\rightarrow$ Willie and Jammer $\rightarrow$ Willie links, respectively. The total maximum transmission power is $P_{\text{max}}$, and the power can be allocated to Alice and Jammer due to their coordinated transmission. In addition, the key variables are given in Table I.

### A. Cooperative Deception Strategy

We assume that Alice transmits $L$ symbols in a transmission time slot with the symbol index of $l$, and Jammer injects AN at each time slot. There exists a reliable link, such as wired link, between Alice and Jammer to exchange the information of when and how to inject AN. In addition, time synchronization between Alice and Jammer can be realized by using the GPS or BEIDOU navigation system [26]. Let $\mathcal{H}_0$ denotes the hypothesis that Alice does not transmit confidential information to Bob, while $\mathcal{H}_1$ denotes the hypothesis that Alice transmits the confidential messages to Bob [27]. From Willie’s perspective, the transmitted signals from Alice and Jammer are given by

$$x_a^l = \begin{cases} 0, & \mathcal{H}_0, \\ x_{b1}^l, & \mathcal{H}_1. \end{cases}$$

and $x_j^l = \begin{cases} x_{j,0}^l, & \mathcal{H}_0, \\ x_{j,1}^l, & \mathcal{H}_1, \end{cases}$ (1)

where $x_{j,0}^l$ and $x_{j,1}^l$ are transmitted information with unity power from Alice, and Jammer under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. As mentioned before, the transmit power satisfies $P_a + P_{j,1} \leq P_{\text{max}}$ under $\mathcal{H}_1$, and also $P_{j,0} \leq P_{\text{max}}$ under $\mathcal{H}_0$. Thus, the received signal at Bob and Willie are respectively given by

$$y_{b}^l = \begin{cases} \sqrt{P_{j,0}} h_{jb} x_{j,0}^l + n_{b1}^l, & \mathcal{H}_0, \\ \sqrt{P_{a} h_{ab} x_{b1}^l} + \sqrt{P_{j,1}} h_{jb} x_{j,1}^l + n_{b1}^l, & \mathcal{H}_1, \end{cases}$$

$$y_{w}^l = \begin{cases} \sqrt{P_{j,0}} h_{jw} x_{j,0}^l + n_{w1}^l, & \mathcal{H}_0, \\ \sqrt{P_{a} h_{aw} x_{b1}^l} + \sqrt{P_{j,1}} h_{jw} x_{j,1}^l + n_{w1}^l, & \mathcal{H}_1, \end{cases}$$

(2)

where $n_{b1}^l \sim \mathcal{CN}(0, \sigma_b^2)$ is the noise at Bob and $n_{w1}^l \sim \mathcal{CN}(0, \sigma_w^2 I)$ is the background noise at Willie. While obtaining the instantaneous channel state information (CSI) of the jamming link without knowledge of $x_{j,0}^l$ remains challenging, we assume that Willie has optimal detection performance for its concerned channel. Consequently, we assume that Willie knows the instantaneous CSI of both the Alice-Willie link and the Jammer-Willie link. Willie utilizes blind signal processing to estimate the channel state information [28] and applies maximum ratio combining (MRC) for diversity gain. Furthermore, we assume that Willie is equipped with multiple antennas and always focuses on the signal direction with

| Table I: Summary of Key Variables |
|-----------------------------------|
| Notation | Description |
| $h_{aw}$ | Instantaneous CSI between node $a$ and Willie, $a \in \{\text{Alice} (a) , \text{Jammer} (j)\}$ |
| $\Delta h_{aw}$ | Estimation error of $h_{aw}, a \in \{\text{Alice} (a) , \text{Jammer} (j)\}$ |
| $h_{jw}$ | Instantaneous CSI between node $j$ and Willie, $j \in \{\text{Alice} (a) , \text{Jammer} (j)\}$ |
| $\mu_w$ | MRC weight vector of Willie |
| $h_{wb}$ | Instantaneous CSI between node $w$ and Bob, $w \in \{\text{Alice} (a) , \text{Jammer} (j)\}$ |
| $\sigma^2_w$ | The noise variance at node $w, w \in \{\text{Alice} (a) , \text{Jammer} (j)\}$ |
| $P_{\text{max}}$ | The transmission power constraint at Alice and Jammer |
| $P_a$ | The transmission power of Alice |
| $P_{j,1}$ | The transmission power constraint of Jammer under $\mathcal{H}_1, j \in \{0, 1\}$ |
| $P_{j,0}$ | The transmission power of Jammer under $\mathcal{H}_0, j \in \{0, 1\}$ |
| $\mathcal{C}$ | A required threshold of detection error probability for covert transmission |
| $\mathcal{C}_2$ | The detection error probability at Willie |
| $\mathcal{C}_3$ | A required threshold of detection error probability for covert transmission |
| $\sigma_{aw}$ | The noise variance at node $a$ |
| $\sigma_{jw}$ | The noise variance at node $j$ |

Fig. 1. System model.
the maximum received power. In this manner, we can only conclude that Willie is effectively deceived into focusing on the signal from the Jammer and ignoring the signal from Alice, when the condition that the signal power from the Jammer exceeds that from Alice is met.

To explore the benefits of a cooperative deception strategy in this scenario, we assume that the Jammer initiates transmission of an interference signal before Alice transmits confidential messages. In burst communications, the transmission duration typically lasts for only a few microseconds. Under the assumption that the Jammer transmits first, it is challenging for warden to adjust its attention to Alice’s transmission in a short time. Consequently, Willie is more likely to use MRC for the Jammer than for Alice. Thus, the MRC weight vector is given by \( w_j \), and the actual received signal at Willie after using MRC is given by

\[
\hat{y}_w^l = u_w^l y_w^l = \begin{cases} \sqrt{P_{j,0}} |h_{jw}|^2 x_{j,0} + \frac{h_{jw}^H}{|h_{jw}|} n_w^l, & H_0 \\ \sqrt{P_{a}} |h_{aw}|^2 x_{j,0} + \sqrt{P_{j,1}} |h_{jw}|^2 x_{j,1} + \frac{h_{jw}^H}{|h_{jw}|} n_w^l, & H_1 \end{cases}
\]

When Willie uses MRC to receive the signal from Jammer, the signal from Alice is received by Willie with random combining [29]. As a result, the multi-antenna diversity gain is given by \( u_w^l = \frac{h_{jw}^H}{|h_{jw}|} \), and the actual received signal at Willie after using MRC is given by

\[
|\hat{y}_w|^2 = \begin{cases} P_{j,0} |h_{jw}|^2 + \sigma_w^2, & H_0 \\ P_a |h_{aw}|^2 + P_{j,1} |h_{jw}|^2 + \sigma_w^2, & H_1 \end{cases}
\]

where \( h_{aw} \) is the element of \( h_{aw} \).

**B. Covert Constraint**

In general, the prior probabilities of hypotheses \( H_0 \) and \( H_1 \) are assumed to be equal [15]. Hence, the detection error probability (DEP) \( \xi \) can be expressed as

\[
\xi = P_{MD} + P_{FA} = \Pr (D_0|H_1) + \Pr (D_1|H_0),
\]

where \( P_{MD} \) and \( P_{FA} \) denote, respectively, the miss detection probability and the false alarm probability. \( D_1 \) and \( D_0 \) indicate whether the transmission from Alice to Bob is present or not. When DEP is larger than a predetermined threshold \( 1 - \varepsilon \), the transmission can be treated as covert, e.g. the covert constraint is \( \xi \geq 1 - \varepsilon \). With some mathematical calculations and derivations, \( P_{MD} \) and \( P_{FA} \) are, respectively, given as [18]

\[
P_{MD} = \frac{\gamma \left( L, \frac{\sigma_w^2}{|h_{aw}|^2 P_a + |h_{jw}|^2 P_{j,1} + \sigma_w^2} \right) \theta^*}{\Gamma (L)},
\]

\[
P_{FA} = 1 - \frac{\gamma \left( L, \frac{\sigma_w^2}{|h_{aw}|^2 P_a + |h_{jw}|^2 P_{j,1} + \sigma_w^2} \right) \theta^*}{\Gamma (L)},
\]

where \( \gamma \) is the lower incomplete Gamma function, \( \Gamma (x) \) is the complete Gamma function, and \( \theta^* \) is the optimal threshold of Willie’s detector which is given by

\[
\theta^* = \frac{L \lambda_0 \lambda_1}{\lambda_1 - \lambda_0} \ln \frac{\lambda_1}{\lambda_0},
\]

where \( \lambda_0 = \frac{|h_{jw}|^2 P_{j,0} + \sigma_w^2}{|h_{jw}|^2 P_{j,1} + \sigma_w^2} \) and \( \lambda_1 = \frac{|h_{aw}|^2 P_a + |h_{jw}|^2 P_{j,1} + \sigma_w^2}{|h_{aw}|^2 P_a + |h_{jw}|^2 P_{j,1} + \sigma_w^2} \). From (5) and (6), we find that using \( P_{MD} + P_{FA} \geq 1 - \varepsilon \) as a constraint would make the problem difficult to solve due to the special functions involved. Thus, we use another expression of DEP.

For the \( l^{\text{th}} \) symbol, the probabilities of \( H_0 \) and \( H_1 \) are \( P_0 \) and \( P_1 \), respectively, and the corresponding probability density functions are \( p_0(x) \) and \( p_1(x) \). Willie detects symbols in a time slot, and each symbol is independent of the other symbols. Thus, for Willie, \( P \{H_0 \text{ is true} \} = (P_0)^L \), \( P \{H_1 \text{ is true} \} = (P_1)^L \). Both \( \hat{y}_w^l \) and \( y_w^l \) can be used for calculating \( p_0(x) \) and \( p_1(x) \) [17], and the former one is used in this paper, which is expressed as

\[
p_0 (\hat{y}_w^l) = \frac{1}{\pi \lambda_0} \exp \left( -\frac{|\hat{y}_w^l|^2}{\lambda_0} \right),
\]

\[
p_1 (\hat{y}_w^l) = \frac{1}{\pi \lambda_1} \exp \left( -\frac{|\hat{y}_w^l|^2}{\lambda_1} \right).
\]

Different from (5), let

\[
\xi = 1 - V_T (P_0^L, P_1^L),
\]

where \( V_T (x_1, x_2) \) is the total variation between \( x_1 \) and \( x_2 \). However, \( V_T (P_0^L, P_1^L) \) is mathematically intractable in some scenarios [8], such as delay constrained communication or short packet communication. Meanwhile, Kullback-Leibler (KL) divergence offers tight bounds on \( V_T (P_0^L, P_1^L) \), and can be expressed analytically [30]. Thus, the KL divergence is used in our paper. Then, using Pinsker’s inequality [31], we have

\[
V_T (P_0^L, P_1^L) \leq \frac{1}{2} D (P_0^L || P_1^L) = \sqrt{\frac{L}{2} D (P_0 || P_1)},
\]

\[
V_T (P_0^L, P_1^L) \leq \frac{1}{2} D (P_1^L || P_0^L) = \sqrt{\frac{L}{2} D (P_1 || P_0)},
\]

where \( D (P_0 || P_1) \) is the Kullback-Leibler (KL) divergence [13] from \( p_0(x) \) to \( p_1(x) \). Since the solution method under \( D (P_0 || P_1) \) and \( D (P_1 || P_0) \) are consistent, this paper takes \( D (P_0 || P_1) \) as an example which is given by

\[
D (P_0 || P_1) = \ln \frac{\lambda_1}{\lambda_0} + \frac{\lambda_0}{\lambda_1} - 1.
\]

Substituting (10) into (9), the covert constraint can be expressed as

\[
D (P_0^L || P_1^L) = LD (P_0 || P_1) \leq 2\varepsilon^2.
\]

We use (12) instead of (5) to express the covert constraint due to low complexity.
C. Different CSI Scenarios

We assume that due to the cooperative relationship between Alice, Jammer, and Bob, Alice and Jammer can estimate Bob’s CSI accurately. However, it is challenging for Alice to obtain the CSI of Willie’s links if Willie is silent and unfriendly to Alice. Thus, we consider the following three scenarios:

1) Scenario 1 (Instantaneous CSI): As assumed in many papers, Willie is a legitimate user to Alice and Jammer, while it is hostile to Bob. In this case, Alice and Jammer know instantaneous CSI of the Alice-to-Willy link and the Jammer-to-Willy link and use it to assist the covert transmission from Alice [18], [21]. Moreover, the covert rate with instantaneous CSI can be an upper bound over that with imperfect CSI and statistical CSI. Hence, it is necessary to investigate the covert rate under this scenario.

2) Scenario 2 (Imperfect CSI): Consider a more realistic scenario where Willie has only limited cooperation with Alice and Jammer. In this case, we assume that the CSI of Alice-to-Willy link and Jammer-to-Willy link are imperfect. More specifically, in the presence of channel estimation errors, the imperfect CSI elements of Willie’s links can be modeled as

\[ h_{aw} = \hat{h}_{aw} + \Delta h_{aw}, \quad \text{and} \quad h_{jw} = \hat{h}_{jw} + \Delta h_{jw}, \quad (13) \]

where \( \hat{h}_{aw} \) and \( \hat{h}_{jw} \) are the elements of the estimated CSI of Alice-to-Willy link and Jammer-to-Willy link, respectively. The corresponding CSI error coefficients are \( \Delta h_{aw} \) and \( \Delta h_{jw} \), which are characterized by the norm-bounded model, i.e.,

\[ |\Delta h_{aw}|^2 \leq v_{aw}^2, \quad \text{and} \quad |\Delta h_{jw}|^2 \leq u_{jw}^2, \quad (14) \]

where \( v_{aw}^2 > 0 \) and \( u_{jw}^2 > 0 \) denote the norm-bound of the channel estimation errors between Alice and Willie, and that between Jammer and Willie, respectively.

3) Scenario 3 (Statistical CSI): Further, assuming that Willie is an enemy adversary and deliberately hides himself, then Alice and Jammer only know the statistical CSI for Willie’s links, i.e., variances \( \sigma_{aw}^2 \) and \( \sigma_{jw}^2 \).

III. OPTIMAL POWER ALLOCATION FOR DIFFERENT CSI

In this section, we investigate the optimal power allocation at Alice and Jammer for different CSI scenarios where the cooperative deception strategy is applied, and Jammer transmits AN to attract Willie’s attention. As a result, Willie uses MRC to combine the signal from Jammer, and can only use random combining scheme to receive the signal from Alice. Thus, the advantage of multiple antennas cannot be used to detect Alice’s transmission, and Alice’s signal is received with a single antenna gain. In the following, we discuss the optimal power allocation for three different CSI.

A. Instantaneous CSI Scenario

In this section, we maximize the covert rate by optimizing the power allocation at Alice and Jammer under \( H_0 \) and \( H_1 \) when all links’ instantaneous CSI are available. Specifically, for the proposed strategy, the covert rate \( R_b \) is maximized subject to the perfect covert and total power constraints, and the optimization problem can be formulated as

\[
\begin{align*}
\max_{P_a, P_{j,1}, P_{j,0}} & \quad R_b, \\
s.t. & \quad D(\mathbb{P}_0|\mathbb{P}_1) = 0, \\
& \quad P_a |h_{aw}|^2 \leq P_{j,1} |h_{jw}|^2, \\
& \quad P_{j,1} + P_a \leq P_{\max}, \\
& \quad P_{j,0} \leq P_{\max}, \\
& \quad P_{j,1} \leq P_{j,0} \leq 2P_{j,1}.
\end{align*}
\]

(15a)

(15b)

(15c)

(15d)

(15e)

where \( R_b \) is obtained (\( 1 + \frac{P_{j,1}|h_{jw}|^2}{P_a|h_{aw}|^2 + \sigma_j^2} \)). It is possible to achieve perfect covert transmission with instantaneous CSI by adjusting the transmit power. Thus (15b) is an equality constraint, (15c) implies that the power received by Willie from Jammer under \( H_1 \) is greater than the power received from Alice, and Willie can be deceived to focus the signal from Jammer\(^1\). Otherwise, Willie will perceive that the power sent by Alice is much stronger and use MRC on Alice’s transmission, which will make the deception invalid.

Since the only one root of \( f(x) = \ln x + \frac{1}{x} - 1 \) is \( x = 1 \), (15b) holds only if \( \lambda_0 = \lambda_1 \), i.e. (15b) can be restated as

\[
P_a |h_{aw}|^2 = (P_{j,0} - P_{j,1}) |h_{jw}|^2 \quad (16)
\]

Substitute (16) into (15c), we get the following inequality:

\[
P_{j,1} \leq P_{j,0} \leq 2P_{j,1}.
\]

(17)

To simplify the calculation, like the assumption of multi-user channels in [34] and [35], we assume that the Jammer-to-Willy link follows the same Rayleigh distribution with mean 0 and variance \( \sigma_j^2 \) at each antenna. According to the Central Limit Theorem (CLT) [36], only when \( M \) is large, \( |h_{jw}|^2 \approx M |h_{jw}|^2 \) is satisfied, where \( h_{jw} \) is the element of \( h_{jw} \) which also follows the same channel distribution. In addition, the element of \( h_{jw} \) is considered as \( h_{jw,k} (1 \leq k \leq M) \). For simplicity, we omit the subscript \( k \) in \( h_{jw,k} \) and write the element of \( h_{jw} \) as \( h_{jw} \). Meanwhile, from the expression of \( R_b \) and (16), we find that maximizing \( R_b \) is equivalent to maximizing \( P_a \), which is also equivalent to maximizing \( P_{j,0} - P_{j,1} \). Thus (17) can be transformed into \( P_{j,0} = 2P_{j,1} \) and (16) can be transformed into

\[
P_a = P_{j,1}M |h_{jw}|^2 |h_{aw}|^2.
\]

(18)

Discussion: From (18), we discuss two cases according to the relationship between \( |h_{aw}|^2 \) and \( |h_{jw}|^2 \) as follows:

1) Case of \( |h_{aw}|^2 \geq |h_{jw}|^2 \): When \( M \leq \frac{|h_{aw}|^2}{|h_{jw}|^2} \), \( P_a \) increases with \( P_{j,1} \) and \( M \). Considering (15e), we can derive the optimal power at Alice and Jammer as

\[
P_a^* = M \max \frac{|h_{jw}|^2}{2|h_{aw}|^2}, \quad P_{j,1}^* = \frac{P_{\max}}{2} \quad \text{and} \quad P_{j,0}^* = P_{\max}.
\]

(19)

\(^1\)In this paper, we assume that Willie can use MRC for only one user, and cannot use MRC for both Jammer and Alice simultaneously [32], [33].
When $M > |h_{aw}|^2$, restricted by (19), $P_{2,1}$ cannot be increased any more. By combining $P_{2,1} + P_a = P_{\text{max}}$ and (18), we get
\[
P_a = \frac{P_{\text{max}} M |h_{jw}|^2}{|h_{aw}|^2 + M |h_{jw}|^2}, \quad P_{2,1} = 1 - P_a, \quad P_{2,0}^* = 2 (1 - P_a^*),
\]
(20)

To achieve perfect covert and the maximum covert rate, when $M \leq |h_{aw}|^2$, $P_{2,1}$ and $P_{2,0}$ are fixed, while $P_a$ increases linearly with $M$. However, when $M > |h_{aw}|^2$, $P_{2,0}$ and $P_{2,1}$ decrease as $P_a$ increases.

2) Case of $|h_{aw}|^2 < |h_{jw}|^2$: Since $M \in \mathbb{N}^+$, $M > |h_{aw}|^2$ is the only case to consider and the optimal power allocation is the same as (20) that $P_{2,0}$ and $P_{2,1}$ decrease as $P_a$ increases.

B. Imperfect CSI Scenario

The instantaneous CSI scenario in the previous section is the ideal case. In contrast, the imperfect CSI scenario considering the channel estimation error [37] is more general. Hence, in this section, we formulate the optimization problems under the imperfect CSI scenario. Due to the errors in channel estimation, perfect covert transmission is difficult to achieve [7]. We used imperfect covert constraints $D(\mathbb{P}_0||\mathbb{P}_1) \leq \frac{2\varepsilon^2}{M}$ instead of the perfect one $D(\mathbb{P}_0||\mathbb{P}_1) = 0$, and channel bounded error models (13) (14) are considered in the constraints. Mathematically, the optimization problem is formulated as
\[
\max_{P_a, P_{2,0}, P_{2,1}} P_a, \quad \text{s.t.} \quad D(\mathbb{P}_0||\mathbb{P}_1) \leq \frac{2\varepsilon^2}{M},
\]
(21a)

\[
P_a |h_{aw}|^2 \leq P_{2,1} |h_{jw}|^2, \quad h_{aw} = \tilde{h}_{aw} + \Delta h_{aw}, \quad |\Delta h_{aw}|^2 \leq \sigma_w^2, \quad h_{jw} = \tilde{h}_{jw} + \Delta h_{jw}, \quad |\Delta h_{jw}|^2 \leq \sigma_w^2,
\]
(21b)

\[
(\tilde{h}_{jw})^H \tilde{h}_{jw} - P_a \sigma_{jw}^2 \leq 0.
\]
(15d), (15e).

From (11), let $\Delta x_{10} = x, x > 0$, (21b) can be expressed as

\[
\ln x + 1 - 2 \frac{\varepsilon^2}{M} \leq 0, \quad x > 0.
\]

Taking the derivative with respect to $f(x) = \ln x + 1 - 2 \frac{\varepsilon^2}{M}$, we get $f'(x) = \frac{2\varepsilon^2}{x}$. Since when $0 < x < 1$, $f'(x) < 0$, $f(x)$ is monotonically decreasing. Similarly, $f(x)$ increases monotonically when $x > 1$. The minimum point of $f(x)$ is $x = 1$, and $f(1) < 0$. It is clear that $f(x)$ has two zeros, which are $0 < a < 1$ and $b > 1$, and when $a < x < b$, $f(x) < 0$.

(21b) can be equivalently written as
\[
\tilde{a} \leq \frac{P_a}{P_{2,0}} |h_{aw}|^2 + \frac{P_{2,1}}{P_{2,0}} |h_{jw}|^2 + \sigma_w^2 \leq \tilde{b},
\]
(22)

where $0 < \tilde{a} < 1$ and $\tilde{b} > 1$ is the root of $\ln x + 1 - 1 - 2 \frac{\varepsilon^2}{M} > 0$. Since Willie uses MRC to receive the Jammer’s signal, there is multi-antenna gain at Willie for the Jammer’s transmission. Therefore, with the help of Triangle Inequality, the estimation error model for $h_{jw}$ can be approximately transformed from (21e) to
\[
h_{jw} = \hat{h}_{jw} + \Delta h_{jw}, \quad |\Delta h_{jw}|^2 \leq Mu_{jw}^2,
\]
(23)

Notice that the error models in (21d) and (23) lead to an infinite number of constraints which makes the optimization problem intractable. Thus, the S-procedure is employed for the inequality constraints to convert non-Linear-matrix-inequalities (non-LMI) conditions into LMI [38]. More specifically, with the help of the S-procedure, the boundary constraints combined with (21c) and (22) are transformed into LMI. It is noted that the left half inequality of (22) is equivalently re-expressed as
\[
(aP_{2,0} - P_{2,1}) \left( \hat{h}_{jw} + \Delta h_{jw} \right)^H \left( \hat{h}_{jw} + \Delta h_{jw} \right) - P_a \left( \hat{h}_{aw} + \Delta h_{aw} \right)^H \left( \hat{h}_{aw} + \Delta h_{aw} \right) + (\tilde{a} - 1) \sigma_w^2 \leq 0,
\]
(24)

Noted that equations (24) satisfy the function
\[
f_m(x) = x^H A_m x + 2 \Re \{ b_m^H x \} + c_m,
\]
(25)

where $m \in \{1, 2, 3\}, x \in \mathbb{C}^{(M+2)\times1}, A_m \in \mathbb{C}^{M+2}, b_m \in \mathbb{C}^{(M+2)\times1}$ and $c_m \in \mathbb{R}^{1\times1}$. Let $e_M$ be a vector of all 1s consisting of $M$ is For (24), these parameters are given by $x = [\Delta h_{jw} \Delta h_{aw} 1]^H$, $A_1 = \text{diag} ((\tilde{a}P_{2,0} - P_{2,1}) \cdot e_M, -P_a, 0)$, $b_1 = ([\tilde{a}P_{2,0} - P_{2,1}] \hat{h}_{jw} - P_a \hat{h}_{aw}]^H$ and $c_1 = ([\tilde{a}P_{2,0} - P_{2,1}] \hat{h}_{jw}^2 - P_a \hat{h}_{aw}^2 + (\tilde{a} - 1) \sigma_w^2$. By using matrix formulation, (24) can be written as
\[
\begin{bmatrix}
(aP_{2,0} - P_{2,1}) I_M & 0 & (aP_{2,0} - P_{2,1}) \hat{h}_{jw} \\
0 & -P_a & -P_a \hat{h}_{aw} \\
(\tilde{a}P_{2,0} - P_{2,1}) \hat{h}_{jw}^H & -P_a \hat{h}_{aw} & c_1
\end{bmatrix} \leq 0.
\]
(26)

Meanwhile, (21c) and the right half inequality of (22) can be transformed into formulations similar to (26), and the error bound constraints can also be transformed as
\[
\text{diag} (e_M, 1, -M u_{jw}^2 - v_{aw}^2) \preceq 0.
\]
(27)

Then, according to the S-procedure [9], the implication (27) $\Rightarrow$ (26), (27) $\Rightarrow$ (21c) and (27) $\Rightarrow$ the right half inequality of (22) hold if and only if there exist variables $\eta_1 > 0, \eta_2 > 0$ and $\eta_3 > 0$, respectively, such that
\[
\begin{bmatrix}
(\eta_1 - \zeta) I_M & 0 & -\zeta \hat{h}_{jw} \\
0 & \eta_1 + P_a & P_a \hat{h}_{aw} \\
-\zeta \hat{h}_{jw}^H & P_a \hat{h}_{aw} & \eta_1 (-M u_{jw}^2 - v_{aw}^2) - c_1
\end{bmatrix} \succeq 0,
\]
(28)

where $\zeta = \tilde{a}P_{2,0} - P_{2,1}$,
\[
\begin{bmatrix}
(\eta_2 - \zeta) I_M & 0 & -\zeta \hat{h}_{jw} \\
0 & \eta_2 - P_a & -P_a \hat{h}_{aw} \\
-\zeta \hat{h}_{jw}^H & -P_a \hat{h}_{aw} & \eta_2 (-M u_{jw}^2 - v_{aw}^2) - c_2
\end{bmatrix} \succeq 0,
\]
(29)
where $\zeta = -\bar{b}P_{j,0} + P_{j,1}$ and $c_2 = (-\bar{b}P_{j,0} + P_{j,1}) \|\mathbf{h}_{jw}\|^2 + P_a \|\hat{h}_{aw}\|^2 - (\bar{b} - 1) \sigma_w^2$. 

\[
\begin{bmatrix}
(\eta_3 + P_{j,1}) \mathbf{I}_M & 0 \\
0 & P_{j,1} \mathbf{I}_M
\end{bmatrix} 
\begin{bmatrix}
\mathbf{h}_{jw} \\
\mathbf{h}_{jw}^H
\end{bmatrix} 
\begin{bmatrix}
\eta_3 - P_a & P_{j,1} \hat{h}_{jw} \\
-P_a \hat{h}_{aw} & -P_a \hat{h}_{aw}
\end{bmatrix} 
\mathbf{h}_{jw}^H 
\eta_3 (-M u_j^2 - \sigma_w^2 - c_3) \geq 0,
\]

(30)

where $c_3 = -P_{j,1} \|\hat{h}_{jw}\|^2 + P_a \|\hat{h}_{aw}\|^2$.

Therefore, we obtain the following optimization of (21) as

\[
\max_{P_a, P_{j,0}, P_{j,1}} P_a, \quad \text{s.t. (15d), (15e), (28), (29), (30)}.
\]

(31)

It is clear that (31) is a typical LMI optimal problem. Both the objective function and constraint functions are convex. Thus, the optimal solution for power allocation exists. Unlike Scenario 1, the analytical solution to the problem cannot be obtained. Using the optimization toolbox in MATLAB\textsuperscript{b}, we can compute the numerical solutions of $P_a$, $P_{j,0}$ and $P_{j,1}$, and the maximum covert rate $R_b$ can be obtained.

### C. Statistical CSI Scenario

In high-security scenarios, such as military communications, Willie may be an enemy device and has an adversarial relationship with Alice, Jammer, and Bob. In this case, it is impractical to assume that Alice knows Willie’s instantaneous CSI or even imperfect CSI. Therefore, we discuss the situation where only the statistical CSI, e.g., $\sigma_w^2$ and $\sigma_{jw}^2$, are known. Mathematically, according to (16), the optimization problem is formulated as

\[
\max_{P_a, P_{j,0}, P_{j,1}} P_a, \quad \text{s.t.} \quad \mathbb{E}\{D(\mathbb{P}_0||\mathbb{P}_1)\} \leq \frac{2e^2}{L},
\]

(32a)

\[
P_a \sigma_w^2 \leq P_{j,1} M \sigma_{jw}^2,
\]

(32b)

\[
(15d), (15e).
\]

(32c)

\[
\mathbb{E}\{D(\mathbb{P}_0||\mathbb{P}_1)\} = \ln \frac{\|\mathbf{h}_{aw}\|^2 P_a + \|\mathbf{h}_{jw}\|^2 P_{j,1} + \sigma_w^2}{\|\mathbf{h}_{jw}\|^2 P_{j,0} + \sigma_w^2} \right. 
\]

\[
+ \frac{\|\mathbf{h}_{jw}\|^2 P_{j,0} + \sigma_w^2}{\|\mathbf{h}_{aw}\|^2 P_a + \|\mathbf{h}_{jw}\|^2 P_{j,1} + \sigma_w^2} - 1.
\]

(33)

Since Alice and Jammer only know $\sigma_w^2$ and $\sigma_{jw}^2$, the statistical expectation of $D(\mathbb{P}_0||\mathbb{P}_1)$ is obtained by integrating over $\|\mathbf{h}_{aw}\|^2$ and $\|\mathbf{h}_{jw}\|^2$. Considering that each channel is assumed to follow the Rayleigh fading model, $\|\mathbf{h}_{aw}\|^2$ follows an exponential distribution with a parameter of $\frac{1}{\sigma_w^2}$. The probability density function (PDF) of $\|\mathbf{h}_{aw}\|^2$ is $f_{\|\mathbf{h}_{aw}\|^2}(x) = \frac{1}{\sigma_w^2} \exp\left(-\frac{x}{\sigma_w^2}\right)$. Similarly, since $\|\mathbf{h}_{jw}\|^2$ is the sum of $M$ independent exponential RV, it follows a Gamma distribution. The PDF of $\|\mathbf{h}_{jw}\|^2$ is $f_{\|\mathbf{h}_{jw}\|^2}(y) = (\sigma_{jw}^2)^{-M} \exp\left(-\frac{y}{\sigma_{jw}^2}\right)$. Thus, $\mathbb{E}\{D(\mathbb{P}_0||\mathbb{P}_1)\}$ is calculated as

\[
\mathbb{E}\{D(\mathbb{P}_0||\mathbb{P}_1)\} = \int_0^\infty \int_0^\infty D(\mathbb{P}_0||\mathbb{P}_1) f_{\|\mathbf{h}_{aw}\|^2}(x) f_{\|\mathbf{h}_{jw}\|^2}(y)\,dxdy
\]

\[
= \int_0^\infty \int_0^\infty \left[ \ln \frac{x P_a + y P_{j,1} + \sigma_w^2}{y P_{j,0} + \sigma_w^2} + \frac{y P_{j,0} + \sigma_w^2}{x P_a + y P_{j,1} + \sigma_w^2} - 1 \right] 
\]

\[
\left[ (\sigma_{jw}^2)^{-M} \exp\left(-\frac{y}{\sigma_{jw}^2}\right) \right] \,dxdy. \quad (34)
\]

Since there are logarithmic, exponential, and power functions in (34), this double integral is complicated and intractable. To solve this problem, we analyzed $D(\mathbb{P}_0||\mathbb{P}_1)$ asymptotically and converted the double integral into a single integral by substitution, which significantly reduces the computational complexity and makes this problem solvable.

Asymptotic: For $D(\mathbb{P}_0||\mathbb{P}_1)$, the denominator and numerator of both fractions in (33) contain the noise power $\sigma_w^2$, which can be ignored when the noise power is considered to be very small compared to the jamming power. Since $\|\mathbf{h}_{aw}\|^2 > 0$ and $\|\mathbf{h}_{jw}\|^2 > 0$, $D(\mathbb{P}_0||\mathbb{P}_1)$ can be asymptotically expressed as

\[
D(\mathbb{P}_0||\mathbb{P}_1) = \ln \frac{x P_a + y P_{j,1}}{y P_{j,0} + x P_a} + \frac{y P_{j,0}}{y P_{j,1} + x P_a} - 1
\]

\[
= \ln \frac{P_a + P_{j,1}}{P_{j,0}} + \frac{P_{j,0}}{P_{j,1}} - 1. \quad (35)
\]

Let $\frac{x}{y} = z$, and the above formula is simplified as

\[
D(\mathbb{P}_0||\mathbb{P}_1) = \ln \frac{P_a + P_{j,1}}{P_{j,0}} + \frac{P_{j,0}}{P_{j,1}} - 1
\]

Double Integral Conversion: In the following, we employ a property of the F-distribution to obtain the PDF of RV $z$, and convert (34) into a single integral operation.

Property 1 (F-Distribution): Let $X \sim N(\mu_x, \delta_x^2)$ and $Y \sim N(\mu_y, \delta_y^2)$ be independent, Sample $(X_1, X_2, \ldots, X_{n_1})$ and sample $(Y_1, Y_2, \ldots, Y_{n_2})$ from $X$ and $Y$, respectively, we have

\[
F = \sum_{i=1}^{n_1} (X_i - \mu_x)^2 \sum_{j=1}^{n_2} (Y_j - \mu_y)^2 \sim F(n_1, n_2).
\]

(36)

Since the exponential distribution and the Gamma distribution are both special chi-squared distributions, with the help of Property 1, we have $\frac{2M \sigma_w^2}{\sigma_{jw}^2} z \sim F(2, 2M)$, where the factors of 2 and $2M$ are the degrees of freedom of $\|\mathbf{h}_{aw}\|^2$ and $\|\mathbf{h}_{jw}\|^2$, respectively. Thus, the PDF of $z$ can be written as

\[
f_Z(z) = \frac{M \sigma_{jw}^2}{\sigma_w^2} \left( \frac{1}{B(1, M)} \right) \left( 1 + \frac{M \sigma_{jw}^2}{\sigma_w^2} z \right)^{-(M+1)}
\]

\[
= M \left( \frac{\sigma_w^2}{\sigma_{jw}^2} \right)^M \left( \frac{\sigma_{jw}^2}{\sigma_w^2} z + 1 \right)^{-(M+1)}. \quad (37)
\]

where $B(x, y)$ is the beta function (Euler’s integral of the first kind) and $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. Instead, with the help of (37), let $\alpha = \frac{P_a}{P_{\text{max}}}$, $\beta = \frac{P_{j,0}}{P_{\text{max}}}$ and $\gamma = \frac{P_{j,1}}{P_{\text{max}}}$. $\mathbb{E}\{D(\mathbb{P}_0||\mathbb{P}_1)\}$ can
be further expressed as
\[
E \{ D (P_0 || P_1) \} = \int_0^\infty D (P_0 || P_1) f_Z (z) \, dz
\]
\[
= M \left( \frac{\sigma_{aw}^2}{\sigma_{jw}^2} \right)^M \int_0^\infty \left[ \ln \frac{az + \gamma}{\beta} + \frac{\beta}{az + \gamma} - 1 \right] \times \left( \frac{\sigma_{aw}^2}{\sigma_{jw}^2} \right) \, dz. \tag{38}
\]

The details of deriving \( E \{ D (P_0 || P_1) \} \) are presented in Appendix A. With the help of Appendix A and \( b = \frac{\sigma_{aw}^2}{\sigma_{jw}^2} \), we finally get the expression of \( E \{ D (P_0 || P_1) \} \) as
\[
E \{ D (P_0 || P_1) \} = \frac{\ln \gamma + \frac{\alpha}{\gamma} M^{-1} F_1 \left( 1, 1; 1 + M; \frac{\gamma - \alpha b}{\gamma} \right) + \frac{\beta}{M} \frac{M + 1}{2} F_1 \left( 1, 1; 2 + M; \frac{\gamma - \alpha b}{\gamma} \right)}{M + 1} - 1. \tag{39}
\]

**Discussion:** To further simplify the problem, the relationship between the transmission powers, especially between \( P_{j,0} \) and \( P_{j,1} \), is explored. Performing the partial derivative of (39) with respect to \( \gamma \), the result is given in (40), as shown at the bottom of the next page. The detail of derivation is given in Appendix B. Setting \( \frac{\partial}{\partial \gamma} E \{ D (P_0 || P_1) \} = 0 \), we discover that when \( M > 1 \), the second, fourth and fifth terms of the equation all tend to 0. Thus, \( \gamma^* \approx \beta \) is the root of \( \frac{\partial}{\partial \gamma} E \{ D (P_0 || P_1) \} = 0 \), regardless of \( \alpha \). Moreover, when \( \gamma < \gamma^* \), \( \frac{\partial}{\partial \gamma} E \{ D (P_0 || P_1) \} < 0 \), when \( \gamma > \gamma^* \), \( \frac{\partial}{\partial \gamma} E \{ D (P_0 || P_1) \} > 0 \). Thus, \( \gamma^* \) is the minimum point of the function \( E \{ D (P_0 || P_1) \} \). Therefore, the optimization problem has an optimal solution at \( P_{j,0} = P_{j,1} \). Finally, considering \( P_a + P_{j,1} \leq P_{\text{max}} \), we get \( P_a = P_{\text{max}} - P_{j,1} = P_{\text{max}} - P_{j,0} \). After that, (39) is asymptotically simplified as
\[
E \{ D (P_0 || P_1) \} = \frac{\alpha}{\gamma} b M^{-1} F_1 \left( 1, 1; 1 + M; \frac{\gamma - \alpha b}{\gamma} \right) + M \frac{M + 1}{2} F_1 \left( 1, 1; 2 + M; \frac{\gamma - \alpha b}{\gamma} \right) - 1, \tag{41}
\]
where \( \alpha = 1 - \gamma \). Meanwhile, by taking the second partial derivative of \( \alpha \) in (41), it is proved that (41) is a convex function. Therefore, (32) is converted from a multivariable optimization problem to a univariate optimization problem which can be solved by CVX in Matlab®. Observing (41), it is clear that for the same statistical CSI, the power allocated to Alice is increased with \( M \).

**IV. Special Case**

As the performance benchmark of the proposed cooperative deception strategy, we study the scenario where Jammer does not participate in cooperative deception. In this case, Willie will use MRC on Alice under \( H_0 \). Since Jammer has no deceptive effect under \( H_0 \), and \( P_{j,1} > 0 \) will consume Alice’s transmit power and cause interference to Bob’s receptions. Therefore, we assume that in this case, Jammer only transmits AN under \( H_0 \), and Alice only transmits covert information at \( H_1 \). Under this assumption, the covert transmission performance under three different CSI scenarios is studied, and the covert rate is also maximized through power allocation.

**A. Instantaneous CSI Scenario**

In this scenario, similar to (15), the optimization problem is expressed as follows
\[
\max_{P_a, P_{j,0}} R_b^c, \tag{42a}
\]
\[\text{s.t.} \quad (15b), \quad P_a \leq P_{\text{max}}, \quad P_{j,0} \leq P_{\text{max}}. \tag{42b}\]

Since the transmission strategy has changed, the expression of \( D (P_0 || P_1) \) is also changed as \( D (P_0 || P_1) = \ln \frac{P_a || h_{aw} \|^2 + \sigma_{aw}^2}{P_{j,0} || h_{jw} \|^2 + \sigma_{jw}^2} + P_{j,0} || h_{aw} \|^2 + \sigma_{aw}^2 - 1 \). Thus, (15b) can be transformed into the form
\[
P_a || h_{aw} \|^2 = P_{j,0} || h_{jw} \|^2, \quad \text{i.e.,} \quad P_a || h_{aw} \|^2 = P_{j,0} || h_{jw} \|^2. \tag{43}\]

**Discussion:** It can be seen from (43) that the covert transmission performance in this case has nothing to do with the number of Willie’s antennas \( M \), but is only related to the transmission power as well as the channel gain, \( || h_{aw} \|^2 \) and \( || h_{jw} \|^2 \).

1) Case of \( || h_{aw} \|^2 \geq || h_{jw} \|^2 \): To maximize Alice’s transmit power, \( P_{j,0} \) should be \( P_{\text{max}} \), so that we get
\[
P_a^* = \frac{P_{\text{max}} || h_{jw} \|^2}{|| h_{aw} \|^2} \text{ and } P_{j,0}^* = P_{\text{max}}. \tag{44}\]

2) Case of \( || h_{aw} \|^2 < || h_{jw} \|^2 \): Since \( P_{j,0} \) participates in optimization, in this case, \( P_a \) can reach \( P_{\text{max}} \) by adjusting \( P_{j,0} \), which means that
\[
P_a^* = P_{\text{max}} \text{ and } P_{j,0}^* = \frac{P_{\text{max}} || h_{aw} \|^2}{|| h_{jw} \|^2}. \tag{45}\]

From (45), since \( P_{j,1} = 0 \) in the benchmark and \( P_a = P_{\text{max}} \), the covert rate can achieve its maximum value, \( R_{\text{max}} = \log (1 + \frac{P_{\text{max}} || h_{aw} \|^2}{\sigma_{aw}^2}) \).

**B. Imperfect CSI Scenario**

Under imperfect CSI, the optimization problem is formulated as
\[
\max_{P_a, P_{j,0}} P_a \tag{46}\]
\[\text{s.t.} \quad (21b), (21d), (21e), P_a \leq P_{\text{max}}, \quad P_{j,0} \leq P_{\text{max}}. \]

where (21b) can be equivalently written as
\[
\bar{a} P_{j,0} || h_{jw} \|^2 - P_a || h_{aw} \|^2 + (\bar{a} - 1) \sigma_{aw}^2 \leq 0, \tag{47a}\]
\[\text{and} \quad -\bar{b} P_{j,0} || h_{jw} \|^2 + P_a || h_{aw} \|^2 - (\bar{b} - 1) \sigma_{aw}^2 \leq 0. \tag{47b}\]

Similar to the approximation of cooperative deception, (21d) and (21e) can be transformed to
\[
h_{aw} = \hat{h}_{aw} + \Delta h_{aw}, \quad || \Delta h_{aw} \|^2 \leq M v_{aw}^2, \tag{48}\]
\[h_{jw} = \hat{h}_{jw} + \Delta h_{jw}, \quad || \Delta h_{jw} \|^2 \leq M v_{jw}^2, \tag{49}\]
which can be rewritten in matrix form as
\[
\begin{pmatrix}
\phi_1 + P_a \\
0 \\
-P_a \hat{h}_{aw}
\end{pmatrix}
\begin{pmatrix}
I_M \\
-a P_{j,0} H \\
-\varphi 
\end{pmatrix}
\geq 0.
\]

(49)

Meanwhile, (47a) and (47b) can be transformed in matrix form similar to (26).

Thus, use the deformation of the S-Procedure to deal with (49), (47a) and (47b). The implications (49)\rightarrow(47a) and (49)\rightarrow(47b) hold if and only if there exist variables \(\varphi > 0\) and \(\phi_2 > 0\) such that
\[
\begin{pmatrix}
\phi_2 - P_a \\
0 \\
P_a \hat{h}_{aw}
\end{pmatrix}
\begin{pmatrix}
I_M \\
-a P_{j,0} H \\
-\varphi 
\end{pmatrix}
\geq 0.
\]

(50)

and
\[
\begin{pmatrix}
\phi_2 - P_a \\
0 \\
P_a \hat{h}_{aw}
\end{pmatrix}
\begin{pmatrix}
I_M \\
-a P_{j,0} H \\
-\varphi 
\end{pmatrix}
\geq 0.
\]

(51)

where \(\varphi = -M v_{aw}^2 - M u_{jw}^2\), \(d_1 = -P_a h_{aw}^2\) + \(\tilde{a} P_{j,0} \hat{h}_{aw}^2\) and \(d_2 = P_a h_{aw}^2\) - \(\tilde{b} P_{j,0} \hat{h}_{aw}^2\). Finally, (46) can be written as
\[
\max_{P_a, P_{j,0}} P_a \\
s.t. (50), (51), P_a \leq P_{\text{max}}, P_{j,0} \leq P_{\text{max}}.
\]

(52)

Like (31), (52) also can be solved using the optimization toolbox in Matlab.

C. Statistical CSI Scenario

For this scenario, the optimization problem is expressed as
\[
\max_{P_a, P_{j,0}} P_a, \\
s.t. \mathbb{E}\{D(P_0\|P_1)\} \leq \frac{2 e^2}{L}, \\
P_{j,0}, P_a \leq P_{\text{max}}.
\]

(53)

Although under this scenario, \(P_{j,0} = 0\), due to the existence of \(\|h_{aw}\|^2\) and \(\|h_{jw}\|^2\), we also have the problem of double integration. Therefore, refer to the method in Section III-C and ignore \(\sigma_w^2\). Moreover, let \(\|h_{aw}\|^2 = x\), \(\|h_{jw}\|^2 = y\), \(\alpha = z\) and \(\frac{P_a}{P_{j,0}} = \alpha\), then the covert constraint \(D(P_0\|P_1)\) can be expressed as
\[
D(P_0\|P_1) = \ln \frac{x P_a}{y P_{j,0}} + \frac{y P_{j,0}}{x P_a} - 1 \\
= \ln (\alpha z') + (\alpha z')^{-1} - 1.
\]

(54)

Similar to (38), with the help of (37), we have \(\frac{\sigma_{jw}^2}{\sigma_{aw}^2} z' \sim F(2M, 2M)\), where \(2M\) is the degree of freedom of \(\|h_{aw}\|^2\) and \(\|h_{jw}\|^2\). The PDF of \(z'\) can be written as
\[
\begin{align*}
 f_{z'}(z') &= \frac{1}{B(M, M)} \times \left(\frac{\sigma_{aw}^2}{\sigma_{jw}^2}\right)^M (z')^{-M-1} \\
&\quad \times \left(\frac{\sigma_{aw}^2}{\sigma_{jw}^2} + z'\right)^{-2M}.
\end{align*}
\]

(55)

Let \(b' = \frac{\sigma_{aw}^2}{\sigma_{jw}^2}\), then \(\mathbb{E}\{D(P_0\|P_1)\}\) can be further expressed as
\[
\begin{align*}
\mathbb{E}\{D(P_0\|P_1)\} &= \int_0^\infty \! D(P_0\|P_1) f_{z'}(z') \, dz' \\
&= \frac{b'^M}{B(M, M)} (A' + B' + C').
\end{align*}
\]

(56)

With the help of the following equation [39]
\[
\int_0^\infty \! x^{\mu-1} \, dx = \frac{\mu}{\nu} B(\nu, \mu), \quad \text{where} \quad |\arg(\beta)| < \pi, \Re \nu > \Re \mu > 0
\]

(57)

we get
\[
\begin{align*}
A' &= \int_0^\infty \! \ln (\alpha z') (z')^{-M-1} (b' + z')^{-2M} \, dz' \\
B' &= \int_0^\infty \! \alpha^{-1} z'^{-2M} (b' + z')^{-2M} \, dz' \\
C' &= -\int_0^\infty \! z'^{-M-1} (b' + z')^{-2M} \, dz' = -\frac{B(M, M)}{b'^M}.
\end{align*}
\]

(58)

(59)

(60)

Although \(\mathbb{E}\{D(P_0\|P_1)\}\) is a multivariate optimization problem, the only constraint with variable coupling is only related to \(\alpha = \frac{P_a}{P_{j,0}}\). Thus, (53) can be transformed into a univariate \(\alpha\) optimization problem. The convexity of the constraints can be proved using the second-order derivative, and the optimal solution can be obtained using CVX.

V. NUMERICAL RESULT

In this section, we present and discuss the numerical performance results for our proposed transmission strategy. Monte-Carlo simulations are performed using 10,000 samples. To gain deep insight into our work, we consider the following three benchmarks:

1) Benchmark 1 (No Deception): In this case, Jammer is not participating in deception, e.g., \(P_{j,0} = 0\). We only do power allocation for \(P_a\) and \(P_{j,0}\). The relevant important derivations can be found in Section IV.
2) Benchmark 2 (Deception But Fixed Power): Considering a deception strategy that the power of Jammer is the same at $\mathcal{H}_0$ and $\mathcal{H}_1$ [40], [41], e.g., $P_{j,0} = P_{j,1}$. In this case, if $P_a > 0$, the received power at Willie must be different at $\mathcal{H}_0$ and $\mathcal{H}_1$. Thus, even if the instantaneous CSI is known, it is impossible to achieve the perfect covert transmission. Instead, we set $\varepsilon = 0.05$ under the instantaneous CSI.

3) Benchmark 3 (Deception But Uniform Distributed Transmit Power): Considering a deception strategy in which Jammer continuously transmits uniform distributed power, since Willie does not know each realization of $P_j$ although he knows the distribution, the covert constraint should be changed from $D(\bar{P}_0|\bar{P}_1) \leq \frac{\epsilon}{2}$ to $\xi^* \geq 1 - \varepsilon$ [13]. With the help of APPENDIX B in [13], $P_{FA}$ and $P_{MD}$ are calculated successfully. Due to the uncertainty of the transmit power of Jammer, we also set $\varepsilon = 0.05$ instead of $\varepsilon = 0$ under the instantaneous CSI.

By comparing with these benchmarks, the reliability of the proposed cooperative deception strategy will be further illustrated. In our simulations, the path-loss model is $PL = (\frac{P_L0}{d_0^\mu})$ dBm, where $P_L0 = -30$ dB, $d_0 = 1$m, and $\mu$ is the path-loss exponent. We set the number of symbols in a time slot $L = 50$, $P_{\text{max}} = 2$W, $\sigma_w^2 = \sigma_j^2 = -90$dBm [21], [42]. The distance from Alice to Bob is $d_{ab} = 50$m and the distance from Jammer to Bob is $d_{jb} = 90$m. The corresponding path-loss exponent is $\mu_{ab} = 2.5$ and $\mu_{jb} = 3$. The path-loss exponent of Alice-to-Wilkie link and Bob-to-Wilkie link are $\mu_{aw} = 3$ and $\mu_{wj} = 3.5$, respectively. In the following, we present some numerical examples of how different parameters, $d_{jw}$, $d_{aw}$, $\|\Delta h_{aw}\|^2$, and $M$, affect the covert transmission performance.

A. Discussion for Scenario 1

In this case, Alice knows the instantaneous CSI and Alice’s transmission can be completely covert. Fig. 2 illustrates the impact of $M$ on the transmission power at Alice and Jammer when $d_{aw} < d_{jw}$. It can be found when $M$ is small, $P_a$ increases linearly with $M$ while $P_{j,0}$ and $P_{j,1}$ remain unchanged. This is because Alice uses MRC for Jammer, and Willie’s detection of Jammer becomes stronger with the increase of $M$, which is more helpful for Alice’s transmission. However, when $M$ is large, $P_a$ increases slightly while $P_{j,0}$ and $P_{j,1}$ drop gradually. This behavior can be explained using (15d) and (16), where the received power for $\mathcal{H}_0$ is required to be equal to that for $\mathcal{H}_0$ due to the perfect covert constraint. Also, the simulation matches our analysis in the Discussion of Section III-A. In addition, for the same $M$, we find that the increase of $d_{aw}$ weakens Willie’s detection of Alice’s transmission, which is the same as traditional recognition. As a result, $P_a$ increases and the covert rate is improved.

Figure 3 shows the impact of $M$ on the transmission power at Alice and Jammer when $d_{aw} > d_{jw}$. When $M$ increases, Jammer’s transmit power decreases, and Alice’s transmission power increases gradually. In this case, Willie has a strong detection for Jammer while a weak detection for Alice. Comparing Fig. 2 with Fig. 3, we find that the transmission power at Alice in the case of $d_{aw} > d_{jw}$ is larger than that of $d_{aw} < d_{jw}$. This is due to the fact that in the case of $d_{aw} > d_{jw}$, less power is used to deceive Willie and more power can be allocated to Alice for covert transmission. With the help of Jammer’s deception, the covert transmission performance is enhanced. Moreover, when the value of $d_{aw}$ is large, under the covert constraint at Willie, more power can be allocated to Alice, and the covert rate is increased.

In Fig. 4, we investigate the covert rate versus $M$ when $d_{aw} < d_{jw}$ with different transmission strategies. We find that the covert performance of the proposed strategy is not necessarily better than that of all the benchmarks. When $M$ is small, the covert rate of the proposed strategy is less than that of Benchmark 1. Because more power is allocated to Jammer detection, the covert power allocation is enhanced. Moreover, when the value of $d_{aw}$ is large, under the covert constraint at Willie, more power can be allocated to Alice, and the covert rate is increased.
they both apply deception strategies and $\varepsilon = 0.05$. This is because our proposed strategy optimally allocates $P_a$, $P_{j,0}$ and $P_{j,1}$, while benchmarks 2 and 3 optimally allocate only $P_a$ and $P_{j,1}$. The proposed strategy allows Jammer to assist Alice’s transmission greatly, maximizing the advantage of the deception strategy.

Figure 5 illustrates the covert rate versus $M$ when $d_{aw} > d_{jw}$ with different transmission strategies under both approximation and exact values for $h_{jw}$. The exact values for $h_{jw}$ with different strategies are plotted with asterisk sign. It can be seen that the exact and approximate values are close, especially for larger $M$, which confirms the validity of our approximation $|h_{jw}|^2 \approx M|h_{jw}|^2$. The covert performance of the proposed strategy gradually approaches that of Benchmark 1 as $M$ increases. Willie’s detection of Alice under $H_1$ is weaker than that of Jammer under $H_0$. As a result, for Benchmark 1, even though Alice transmits with high power and MRC is used to receive Alice’s signal, Alice’s power received at Willie in $H_1$ is less than Jammer’s power received at Willie in $H_0$. Moreover, without the noise caused by Jammer, the covert rate of Benchmark 1 reaches its upper bound. Hence, when $d_{aw} > d_{jw}$, the deception strategy cannot work better than Benchmark 1 which matches our analysis in the Discussion of Section IV-A. Furthermore, taking into account Fig.4, we conclude that for the instantaneous CSI case, when Alice is close to Willie and $M$ is large, the proposed strategy outperforms Benchmark 1. Comparing with Benchmark 2 and Benchmark 3, whether it is $d_{aw} < d_{jw}$ or $d_{aw} > d_{jw}$, our proposed strategy is always better.

Figure 6 shows the relationship between covert rate and $d_{aw}$ under the same path-loss exponent. When Alice is close to Willie, Alice’s allowed transmit power is constrained significantly, and that constraint is more obvious in Benchmark 1 than in the proposed strategy. This is because in the proposed strategy, the jammer attracts multi-antenna Willie, reducing Willie’s attention on Alice’s transmission. Thus, the effect of $d_{aw}$ increasing on Alice’s transmission is reduced. In Fig. 6, it is clear that $M$ has no impact on the performance of Benchmark 1. In Benchmark 1, Willie uses MRC reception for Alice at $H_1$ and for Jammer at $H_0$, respectively. Thus, $M$ has the same effect on the detection of $H_0$ and $H_1$, which matches our analysis in the Discussion of Section IV-A. Moreover, for the proposed strategy, the covert rate increases as $M$ increases, which matches the behavior shown in Fig. 4 and Fig. 5.

B. Discussion for Scenario 2

In this scenario, Alice’s transmission can not be completely covert, instead, the constraint is $D(\mathbb{P}_0||\mathbb{P}_1) \leq \frac{2e}{L}$. In this subsection, we discuss how the channel estimation errors $|\Delta h_{aw}|^2$, $|\Delta h_{jw}|^2$ and $M$ affect the performance.

The impacts of $M$ on $R_0$ with imperfect CSI under both approximation and exact values of $|\Delta h_{jw}|^2$ are shown in Fig. 7. The covert rate with exact value of $|\Delta h_{jw}|^2$ are plotted with asterisk sign. When $M$ is large, the approximation result is close to the exact value, which confirms that the approximation in (23) is accurate. We find that when $M$ increases, the covert rate under imperfect CSI does not increase as fast as that under perfect CSI. When the number of antennas $M$ increases, the number of elements $\Delta h_{jw}$ increases. Then, due to the elements in $\Delta h_{jw}$ all satisfying $|\Delta h_{jw}|^2 \leq u_{jw}^2$, as $M$ increases, more estimation errors are introduced and the error on each channel increases. As a result,
the effects of channel estimation errors increase slightly with the increase of $M$. Additionally, for the same $\varepsilon$, comparing Benchmark 1 to the proposed strategy, we can arrive at the same conclusion as that under instantaneous CSI, that the proposed strategy can improve the covert rate when $d_{aw} < d_{jw}$ and $M$ is large. Besides, different from the case of perfect CSI, when $d_{aw} > d_{jw}$ and $M$ is large, the performance of the proposed strategy is also better than that of Benchmark 1, which indicates that the advantage of the proposed strategy is significantly under imperfect CSI compared to the case of perfect CSI.

Figure 8 investigates the covert rate against CSI estimation errors of both Alice-to-Willie and Jammer-to-Willie links. When $d_{aw} = d_{jw}$, we find that for the proposed strategy, the influence of $u_{aw}^2$ is greater than that of $v_{jw}^2$, which is explained by (23). However, for Benchmark 1, the change of $u_{aw}^2$ and $v_{jw}^2$ have the same effect on the performance when $d_{aw} = d_{jw}$. In this case, Willie uses MRC for Jammer under $\mathcal{H}_0$ and for Alice under $\mathcal{H}_1$, respectively. Thus the influence of imperfect CSI is increased to $M v_{aw}^2$ and $M v_{jw}^2$. In addition, the case of $d_{aw} > d_{jw}$ is also illustrated in Fig. 8. We can find that, for the proposed strategy, even if $d_{aw}$ increases, the influence of $v_{jw}^2$ is still less than that of $u_{aw}^2$ for $M \gg 1$. However, for Benchmark 1, without the help of Jammer’s deception, the influence of $v_{jw}^2$ increases since $d_{aw}$ increases in this case. Moreover, we conclude that for the case of $M > 1$, the covert rate decreases as the channel estimation errors increase and the proposed strategy can achieve a higher covert rate than Benchmark 1 when the channel estimation error is severe.

C. Discussion for Scenario 3

In Scenario 3, same as Scenario 2, Alice’s transmission can not be completely covert. We discuss how the detection threshold $\varepsilon$, the distance of Alice to Willie $d_{aw}$, and $M$ affect the covert communication performance under statistical CSI.

Figure 9 illustrates that when $P_{j,0}$ and $P_{j,1}$ are fixed, how Alice’s transmit power $P_a$ affects the statistical covert constraint $\mathbb{E}\{D(P_0|P_1)\}$ under different $M$. When $P_a$ increases, $\mathbb{E}\{D(P_0|P_1)\}$ first decreases and then increases which shows that it has a minimum value. This is because when Jammer’s transmit power is fixed, if Alice’s transmit power is too small, Willie’s received power at $\mathcal{H}_1$ is much less than that at $\mathcal{H}_0$. As Alice’s transmit power increases, this gap narrows and $\mathbb{E}\{D(P_0|P_1)\}$ decreases. When Alice’s transmit power continues to increase, Willie’s receive power at $\mathcal{H}_1$ is gradually greater than $\mathcal{H}_0$, and $\mathbb{E}\{D(P_0|P_1)\}$ increases. Moreover, when $M$ increases, the minimum point becomes larger. In this case, Alice can transmit the messages covertly with greater power. In addition, the asymptotic and exact data of $\mathbb{E}\{D(P_0|P_1)\}$ are shown in Fig. 9. It can be seen that when $\sigma_0^2$ is small, the asymptotic and exact data are highly consistent which confirms that the asymptotic result in Section III-C. is feasible.

In Fig. 10, the covert rate for the deception strategy and the benchmarks are plotted when the statistical CSI is available. We can find that when $M$ increases, the covert rate of different deception strategies increase while that of Benchmark
1 decreases. Furthermore, compared to the instantaneous and imperfect CSI cases, for the statistical CSI case, the deception strategies outperform Benchmark 1 with a smaller number of antennas. For Benchmark 1, the statistical CSI of $h_{jw}$ and $h_{wj}$ affect the covert constraint under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. When $M$ increases, the negative effect of only knowing statistical CSI also increases, which greatly affects the performance of Benchmark 1. However, for the deception strategies, the power detected by Willie under $\mathcal{H}_0$ and $\mathcal{H}_1$ are both mainly affected by the statistical CSI of $h_{jw}$, which greatly reduces the impact of only knowing statistical CSI on performance. Therefore, when $M$ increases, the covert rate increases slightly. This behavior verifies our analysis in the Discussion of Section III-C. When $M$ is quite small, same as the instantaneous and imperfect CSI case, due to the power constraints on $P_{j,1}$, the covert rate of Benchmark 1 is greater than that of the proposed strategy. However, As $M$ increases, the deception strategies increase in terms of covert rate while Benchmark 1 decreases.

**VI. CONCLUSION**

In this paper, we propose a cooperative deception strategy for covert transmission in presence of a multi-antenna adversary. Numerical results show that for the proposed deception strategy, the increase of the number of antennas has a positive effect on the covert rate under three different CSI scenarios.

**APPENDIX A**

**DERIVATION OF $\mathbb{E}\left\{ D(P_0||P_1) \right\}$**

Let $b = \frac{\sigma^2_i}{\sigma^2_{jw}}$, (38) is split into three parts, and $\mathbb{E}\left\{ D(P_0||P_1) \right\} = M b^M (A + B + C)$, where

$$A = \frac{1}{M} \int_0^\infty \ln \frac{\alpha z + \gamma}{\beta} (b + z)^{-M} d\zeta,$$

$$B = \frac{1}{M} \int_0^\infty \frac{\beta}{\alpha z + \gamma} (b + z)^{-M} d\zeta,$$

$$C = - \frac{1}{M} \int_0^\infty (b + z)^{-M} d\zeta.$$

Integrating (61a) by parts, we can get

$$A = \frac{1}{M} \int_0^\infty \ln \frac{\alpha z + \gamma}{\beta} (b + z)^{-M} d\zeta = \frac{1}{M} \int_0^\infty \frac{\alpha z + \gamma}{\beta} (b + z)^{-M} d\zeta = \frac{1}{M} \int_0^\infty \frac{1}{z + \frac{\alpha}{\beta}} (b + z)^{-M} d\zeta.$$

With the help of the following identity [39]

$$\int_0^\infty x^{\nu - 1} (\beta + x)^{-\mu} (\gamma + x)^{-\theta} dx = \beta^{-\gamma} \beta^\nu B(\nu, \mu + \nu + \theta) F_1 (\mu, \nu; \mu + \nu; 1 - \frac{\gamma}{\beta}),$$

$$[\arg\beta < \pi, |\arg\gamma| < \pi, \text{Re} \nu > 0, \text{Re} \mu > \text{Re} (\nu - \theta)]$$

After some algebraic operations, (62) can be expressed as

$$A = M^{-1} b^{-M} \ln \frac{\gamma}{\beta} + \frac{\alpha}{\gamma} M^{-2} b^{-1 - M} F_1 (1, 1; 1 + M; 1 - \frac{b \alpha}{\gamma}).$$

where $F_1 (\alpha, \beta; \gamma; z)$ is the Hypergeometric Function. Similarly, (61b) can be expressed as follow

$$B = \frac{\beta}{\alpha} \int_0^\infty \frac{1}{z + \frac{\alpha}{\beta}} (b + z)^{-M - 1} d\zeta = \frac{\beta}{\gamma} (M + 1)^{-1} b^{-M} F_1 (1, 1; 2 + M; 1 - \frac{b \alpha}{\gamma}).$$

In addition, after some operations, the final expression of (61c) can be obtained as

$$C = \frac{1}{M} (b + z)^{-M} \bigg|_0^\infty = -M^{-1} b^{-M}.$$
Thus, the second term of (67) is given by
\[
\frac{\partial}{\partial \gamma} \left\{ 2F_1 \left( 1, 1; 1 + M; \frac{\gamma - \alpha b}{\gamma} \right) \right\} \\
= \frac{\partial}{\partial \gamma} \left\{ \frac{\gamma}{\gamma} \int_0^t (1 - t)^{M-1} \left( 1 - \frac{\gamma - \alpha b}{\gamma} \right)^{-1} dt \right\} \\
= \frac{\alpha b}{\gamma^2} (M + 1)^{-1} 2F_1 \left( 2, 2; 2 + M; \frac{\gamma - \alpha b}{\gamma} \right).
\]
(69)

The derivation of the fourth term of (67) is similar to (69).

Finally, the expression of \( \frac{\partial}{\partial \gamma} \mathbb{E} \left[ D \left( P_0 | P_1 \right) \right] \) can be obtained.

REFERENCES

[1] Y. Zhang, Y. Shen, X. Jiang, and S. Kasahara, “Mode selection and spectrum partition for D2D inband communications: A physical layer security perspective,” IEEE Trans. Commun., vol. 67, no. 1, pp. 623–638, Jan. 2019.
[2] M. Letafati, A. Khezami, K. Wong, and Md. J. Piran, “A lightweight secure and resilient transmission scheme for the Internet of Things in the presence of a hostile jammer,” IEEE Internet Things J., vol. 8, no. 6, pp. 4373–4388, Mar. 2021.
[3] Y. Huang, J. Wang, C. Zhong, T. Q. Duong, and G. K. Karagiannidis, “Secure transmission in cooperative relaying networks with multiple antennas,” IEEE Trans. Wireless Commun., vol. 15, no. 10, pp. 6843–6856, Oct. 2016.
[4] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, “On learning-assisted content-based secure image transmission for delay-aware systems with randomly-distributed eavesdroppers,” IEEE Trans. Commun. (GLOBECOM), Dec. 2021, pp. 1–6.
[5] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, “Content-based medical image transmission against randomly-distributed passive eavesdroppers,” in Proc. IEEE Int. Conf. Commn. Workshops (ICC Workshops), Jun. 2021, pp. 1–7.
[6] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, “Deep learning for hardware-impaired wireless secret key generation with man-in-the-middle attacks,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2022, pp. 25–29.
[7] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, “Gaussian signalling for covert communications,” IEEE Trans. Wireless Commun., vol. 18, no. 7, pp. 3542–3553, Jul. 2019.
[8] B. A. Bash, D. Goeckel, and D. Towsley, “Limits of reliable communication with low probability of detection on AWGN channels,” IEEE J. Sel. Areas Commun., vol. 31, no. 9, pp. 1921–1930, Sep. 2013.
[9] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, “Achieving undetectable communication,” IEEE J. Sel. Topics Signal Proc., vol. 9, no. 7, pp. 1195–1205, Oct. 2015.
[10] L. Wang, G. W. Wornell, and L. Zheng, “Fundamental limits of communication with low probability of detection,” IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.
[11] B. He, S. Yan, X. Zhou, and V. K. N. Lau, “On covert communication with noise uncertainty,” IEEE Commun. Lett., vol. 21, no. 4, pp. 941–944, Apr. 2017.
[12] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, “Reliable deniable communication with channel uncertainty,” in Proc. IEEE Inf. Theory Workshop (ITW), Nov. 2014, pp. 30–34.
[13] S. Yan, B. He, X. Zhou, Y. Cong, and A. L. Swindlehurst, “Delay-intolerant covert communications with either fixed or random transmit power,” IEEE Trans. Inf. Forensics Security, vol. 14, no. 1, pp. 129–140, Jan. 2019.
[14] B. A. Bash, D. Goeckel, and D. Towsley, “Covert communication gains from adversary’s ignorance of transmission time,” IEEE Trans. Wireless Commun., vol. 15, no. 12, pp. 8394–8405, Dec. 2016.
[15] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, “Covert communication in the presence of an uninformed jammer,” IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 6193–6206, Sep. 2017.
[16] T. V. Sobers, B. A. Bash, D. Goeckel, S. Guha, and D. Towsley, “Covert communication with the help of an uninformed jammer achieves positive rate,” in Proc. 49th Asilomar Conf. Signals, Syst. Comput., Nov. 2015, pp. 625–629.
R. Zhang, X. Chen, M. Liu, N. Zhao, X. Wang, and A. Nallanathan, “UAV relay assisted cooperative jamming for covert communications over Rician fading,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7936–7941, Jul. 2022.

M. V. Jamali and H. Mahdavifar, “Covert millimeter-wave communication: Design strategies and performance analysis,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3691–3704, Jun. 2022.

I. S. Gradshteyn and I. M. Ryzhik, “8–9-special functions,” in *Table of Integrals, Series, and Products*, 7th ed., A. Jeffrey, D. Zwillinger, I. Gradshteyn, and I. Ryzhik, Eds. Boston, MA, USA: Academic, 2007, pp. 859–1048.

**Jiangbo Si** (Senior Member, IEEE) received the B.S. degree in communications engineering and the M.S. and Ph.D. degrees in communication and information systems from Xidian University, Xi’an, China, in 2003, 2006, and 2010, respectively. Since 2010, he has been a fixed Researcher with the State Key Laboratory of Integrated Services Networks, Xidian University, where he is currently a Professor. Up to now, he has been in charge of many projects, including those with the National Natural Science Foundation of China and the Natural Science Foundation of Shaanxi Province. In 2019, he was a Visiting Scholar with The University of Texas at Dallas, USA. As the main person, he has participated in Major National Science and Technology Projects and National 863 Research Projects. His research interests include physical layer security, cooperative spectrum sensing, and covert transmission.

**Lei Guan** received the B.S. degree in communications engineering and the Ph.D. degree in military communication from Xidian University, Xi’an, China, in 2009 and 2014, respectively. He is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks (ISN), Xidian University. His current research interests include frequency hopping systems, spectrum sensing, and covert communication.

**Chao Wang** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in information and communication engineering from Xi’an Jiaotong University, Xi’an, China, in 2008, 2013, and 2016, respectively. He is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University. His current research interests include 5G&6G wireless networks and key technologies, physical-layer security, covert communications, and deep learning and its application in the optimization of wireless communications. He was a recipient of the Excellent Doctoral Dissertation Award of Shaanxi Province in 2018 and the China Institute of Communications in 2016, and the Best Paper Awards at the IEEE ICC 2014 and IEEE SAGC 2020, respectively.

**Naofal Al-Dhahir** (Fellow, IEEE) received the Ph.D. degree from Stanford University. He is currently an Erik Jonsson Distinguished Professor and the ECE Associate Head of UT-Dallas. He was a Principal Member of Technical Staff with the GE Research Center and AT&T Shannon Laboratory from 1994 to 2003. He is a co-inventor of 43 issued patents, the coauthor of about 525 papers, and a co-recipient of five IEEE best paper awards. He is a fellow of AAIA. He received the 2019 IEEE COMSOC SPCC Technical Recognition Award, the 2021 Qualcomm Faculty Award, and the 2022 IEEE COMSOC RCC Technical Recognition Award. He served as the Editor-in-Chief for IEEE TRANSACTIONS ON COMMUNICATIONS from January 2016 to December 2019. He is a fellow of the U.S. National Academy of Inventors and a member of the European Academy of Sciences and Arts.

**Zan Li** (Senior Member, IEEE) received the B.S. degree in communications engineering and the M.S. and Ph.D. degrees in communication and information systems from Xidian University, Xi’an, China, in 1998, 2001, and 2004, respectively. She is currently a Professor with the State Key Laboratory of Integrated Services Networks (ISN), Xidian University. Her research interests include wireless communications and signal processing, including weak signal detection, spectrum sensing, and cooperative communication.

**Hang Hu** received the B.S. degree in telecommunications engineering from Xidian University, Xi’an, China, in 2010, and the M.S. and Ph.D. degrees in information and communications engineering from the College of Communications Engineering, PLA University of Science and Technology, Nanjing, China, in 2012 and 2016, respectively. He is currently an Associate Professor with the College of Information and Navigation, Air Force Engineering University, Xi’an. He received the National Post-Doctoral Program for Innovative Talents in 2017 and awarded Excellent Doctoral Dissertation of PLA. His current research interests include intelligent military communication technology, space-air communication networks, and green communications.