SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

June 2024, Vol. 67, Iss. 6, 162304:1–162304:15
https://doi.org/10.1007/s11432-023-3971-7

Beamforming design for covert broadcast communication with hidden adversary

Weiyu CHEN[1], Haiyang DING[2], Shilian WANG[1*], Junshan LUO[1] & Fengkui GONG[3]

[1]*College of Electronic Science and Technology, National University of Defense Technology, Changsha* 410073*, China;*
[2]*College of Information and Communication, National University of Defense Technology, Wuhan* 430019*, China;*
[3]*State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an* 710071*, China*

Received 1 October 2023/Revised 27 December 2023/Accepted 5 March 2024/Published online 27 May 2024

Abstract This paper explores covert broadcast communication in a challenging situation in which the transmitter, Alice, faces uncertainty regarding the position of the passive adversary, Willie. In this situation, controlling the signal exposure becomes difficult. Specifically, this paper focuses on the cases where Alice can delimit the suspicious areas where Willie may be located. The suspicious areas can have arbitrary number and shapes. Under the Rician fading model, the analytical expression of Willie's detection performance is derived. Then, the transmit power, the beamforming direction, and the number of channel uses are jointly designed to minimize the optimal Willie's detection performance among all the possible positions of Willie, on the premise of satisfying receivers' quality-of-service requirements. To tackle the formulated knotty problem with an infinite number of irregular and discontinuous Willie's possible positions, the continuous nature of the antenna pattern is exploited to convert the original problem into the one with a finite number of possible positions, by sampling the suspicious areas. Then, different algorithms are developed to address the resultant non-convex optimization problem for single-antenna Willie and multi-antenna Willie, respectively. Numerical results demonstrate significant performance gains of the proposed scheme over the widely-adopted maximum ratio transmission scheme, which indicates that simply adjusting the main lobe towards the receiver is indeed far away from optimum and shows the importance of leveraging Alice's limited knowledge of Willie's position.

Keywords low-probability-of-detection communication, multiple antennas, covertness performance, uncertain position, parameters optimization

1 Introduction

Covert communication aims to prevent wireless communication signals from being discovered by adversaries, in the sense of high probability of error detection [1]. Based on covert communication techniques, the content, time, and place of the communication behavior are all concealed, which satisfies strong security requirements. In the classic model of covert communication, transmitter Alice seeks to send information to receiver Bob without being discovered by adversary Willie. The fundamental principle is to make Willie's observations when Alice is not transmitting similar to those when Alice is transmitting. One approach to achieve this goal is to increase and utilize Willie's uncertainty about his observations. The sources of uncertainty include noise variance [2], transmission time [3], and interference power [4].

In practice, Willie's uncertainty may be limited. Worse still, it is generally unknown for Alice, and thus can only be conservatively estimated. An alternative and effective approach to confuse Willie is to apply the beamforming technique to reduce signal leakage. In this field, Ref. [5] studied the received signal-to-noise ratio (SNR) maximization problems in multi-input single-output (MISO) systems under covertness constraint. It was revealed that when Alice perfectly knows Willie's channel state information (CSI), zero leakage can be achieved. When Willie's CSI is imprecisely known, the maximum transmit power for covertness was proved, whereas for Rayleigh fading channels with Willie's statistical CSI only, the maximum ratio transmission (MRT) was proved to be optimal. For multi-input multi-output (MIMO) covert transmission, Ref. [6] proposed adopting the full-duplex architecture at Bob to generate jamming

* Corresponding author (email: wangsl@nudt.edu.cn)

and confuse Willie. The beamforming and the jamming were jointly designed based on the statistical CSI of Willie's channel. Furthermore, Ref. [7] proposed using the reconfigurable intelligent surface to enhance covertness in MIMO systems, where global CSI is assumed to be available.

The above studies regarding the beamforming designs rely on the CSI of Willie's channel, which is, however, difficult to obtain in practice, because Willie only needs to observe and thus is often a passive node. Note that even the statistical CSI is still difficult to obtain, which indeed requires the knowledge of Willie's position. In this regard, Refs. [8, 9] proposed a novel metric, covert threat region, to evaluate the covertness for different Willie's positions. Specifically, for each given Willie's position, the optimal covert throughput was obtained by parameters optimization. Then, the covert threat region is defined as the set of Willie's positions that the optimal throughput is below a given quality-of-service (QoS) requirement. However, the results cannot guide the beamforming design for uncertain Willie's position, because the covert threat region is based on different optimal parameters for different given Willie's positions. Ref. [10] considered a rectangular region where Willie may be located and investigated robust parameters optimization under Rayleigh fading, wherein the MRT is adopted.

To sum up, regarding Alice's knowledge of Willie's channel, both perfect CSI model and estimated CSI model with bounded/statistical error do not match well with a passive Willie. Willie may even hide himself in order to hinder Alice from knowing statistical CSI. Fortunately, in practice, Alice may delimit the areas where Willie may be located (e.g., by reconnaissance in military applications). Fully exploiting this information may still significantly enhance covertness. For such, this work aims to design the beamforming when Alice does not know the exact position of passive Willie and can only delimit the suspicious areas, whose number and the shapes can be arbitrary. In such adverse but practical scenarios, the beamforming designs have not been investigated yet, to the best knowledge of the authors. To be more specific, this paper focuses on finite-block-length MISO broadcast communication in Rician fading channels, and the main contributions are summarized as follows.

(1) A strict lower bound of the covert probability, which is defined as the probability that Willie's minimal sum of false alarm and miss detection probabilities is larger than Alice's desired value, is derived for arbitrarily given Alice's beamforming vector, Willie's position, and number of Willie's antennas.

(2) The transmit power, the beamforming direction, and the number of channel uses are jointly optimized to maximize the minimal covert probability among all the possible positions of a single-antenna Willie, while satisfying Bobs' QoS requirements and the delay constraint. Specifically, the continuous nature of the antenna pattern is exploited to convert the original knotty problem with an infinite number of irregular and discontinuous Willie's possible positions into one with the finite number of possible positions. Then, the alternating optimization method, the semidefinite relaxation (SDR) method, and the minorization-maximization (MM) algorithm are invoked to solve the problem.

(3) For multi-antenna Willie, the successive convex approximation (SCA) algorithm and a backtracking method are invoked to tackle the complicated generalized Marcum Q function involved in the covert probability expression and to maximize the minimal covert probability among Willie's possible positions.

The rest of the paper is organized as follows. Section 2 presents the system model and analyzes Willie's detection performance as well as Bobs' decoding performance. In Section 3, the covert probability maximization algorithm for single-antenna Willie is developed, whereas multi-antenna Willie is considered in Section 4. Section 5 presents representative numerical results. Finally, Section 6 concludes the paper.

Notations. $\Pr\{A\}$ denotes the probability of event A. $\iota \triangleq \sqrt{-1}$ is the imaginary unit. $\otimes$ represents the Kronecker product operator. $\mathcal{CN}(\boldsymbol{u}, \boldsymbol{\Sigma})$ denotes complex Gaussian distribution with mean vector $\boldsymbol{u}$ and covariance matrix $\boldsymbol{\Sigma}$. $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. $\boldsymbol{X} \succcurlyeq 0$ represents that matrix $\boldsymbol{X}$ is positive semidefinite. $||\boldsymbol{x}||$ denotes the Euclidean norm of vector $\boldsymbol{x}$. $\mathrm{tr}(\cdot)$ and $|\cdot|$ are the trace and the modulus operators, respectively. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ represent transpose and Hermitian transpose, respectively.

2 System model and performance analyses

2.1 System model

As shown in Figure 1, in the considered system model, transmitter Alice expects to broadcast information to M receivers (Bob 1, Bob 2, ..., Bob M) without being discovered by a hidden adversary Willie. Alice is equipped with a uniform planar array (UPA) with N_t antennas to focus her transmit signals and suppress the signals leakage. Conservatively, we assume that each Bob is equipped with a single receive antenna,

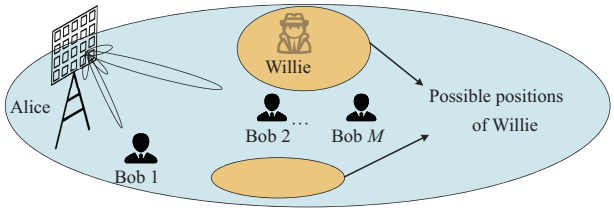

Figure 1 (Color online) Illustration of the considered covert broadcast communication systems.

whereas Willie is equipped with N_w antennas for detection. We focus on delay-intolerant transmission in quasi-static Rician fading channels. Denote n as the number of symbols in once transmission. In the i-th symbol period ($i = 1, 2, \ldots, n$), the received signals at Bob m ($m = 1, 2, \ldots, M$) can be written as

$$y_{b,m}(i) = \sqrt{d_{b,m}^{-\alpha}}\boldsymbol{h}_{b,m}^{\mathrm{T}}\boldsymbol{w}x(i) + n_{b,m}(i), \tag{1}$$

where $d_{b,m}$ denotes the distance from Alice to Bob m, α stands for the path loss exponent, $\boldsymbol{w} \in \mathbb{C}^{N_t \times 1}$ represents the beamforming vector, $n_{b,m}(i) \sim \mathcal{CN}\left(0, \sigma_b^2\right)$ denotes the noise, whereas $x(i) \sim \mathcal{CN}\left(0, 1\right)$ denotes the base-band symbol. We follow the common assumption that a secret codebook is shared by Alice and Bobs in advance [11][1)], whereas Willie only knows the distribution of $x(i)$. $\boldsymbol{h}_{b,m} \in \mathbb{C}^{N_t \times 1}$ denotes the normalized channel coefficients from Alice to Bob m, which is modeled as

$$\boldsymbol{h}_{b,m} = \sqrt{\frac{K_{b,m}}{K_{b,m}+1}}\bar{\boldsymbol{h}}_{b,m}\mathrm{e}^{\iota\theta_{b,m}} + \sqrt{\frac{1}{K_{b,m}+1}}\hat{\boldsymbol{h}}_{b,m}, \tag{2}$$

where $K_{b,m}$ denotes the Rician K-factor. The first term of the right-hand side (RHS) of (2) is the deterministic line-of-sight (LoS) component, where $\theta_{b,m}$ is the phase of the LoS component of the channel coefficient from Alice's first antenna to Bob m's antenna, whereas $\bar{\boldsymbol{h}}_{b,m}$ is modeled as $\bar{\boldsymbol{h}}_{b,m} = \boldsymbol{v}_x\left(\psi_{b,m}, \phi_{b,m}\right) \otimes \boldsymbol{v}_z\left(\psi_{b,m}\right)$, with $\psi_{b,m}$ and $\phi_{b,m}$ being the polar and the azimuthal angles of Bob m with respect to Alice, respectively [12]. Herein, $\boldsymbol{v}_x\left(\psi, \phi\right) \triangleq [1\ \mathrm{e}^{-\frac{\iota 2\pi d_x \sin(\psi)\cos(\phi)}{\lambda}}\ \cdots\ \mathrm{e}^{-\frac{\iota(N_{t,x}-1)2\pi d_x \sin(\psi)\cos(\phi)}{\lambda}}]^{\mathrm{T}}$ and $\boldsymbol{v}_z(\psi) \triangleq [1\ \mathrm{e}^{-\frac{\iota 2\pi d_z \cos(\psi)}{\lambda}}\ \cdots\ \mathrm{e}^{-\frac{\iota(N_{t,z}-1)2\pi d_z \cos(\psi)}{\lambda}}]^{\mathrm{T}}$ are the steering vectors, where λ is the wave length, $N_{t,x}$ and $N_{t,z}$ are the number of transmit antennas in the two directions of the UPA ($N_t = N_{t,x} \times N_{t,z}$), respectively, whereas d_x and d_z represent the corresponding antenna spacings. The second term of the RHS of (2) is the non-LoS component, where $\hat{\boldsymbol{h}}_{b,m}$ follows $\mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{I}_{N_t}\right)$.

Conservatively, we assume that Willie knows the time framework of the possible transmission and his equivalent channel coefficients[2)]. Then, Willie's detection can be modeled by a binary hypothesis test as

$$\begin{aligned} &H_0 : \boldsymbol{y}_w(i) = \boldsymbol{n}_w(i),\ i = 1, 2, \ldots, n, \\ &H_1 : \boldsymbol{y}_w(i) = \sqrt{d_w^{-\alpha}}\boldsymbol{H}_w\boldsymbol{w}x(i) + \boldsymbol{n}_w(i),\ i = 1, 2, \ldots, n, \end{aligned} \tag{3}$$

where null hypothesis H_0 denotes that Alice is not transmitting, alternate hypothesis H_1 denotes that Alice is transmitting, d_w denotes the distance from Alice to Willie, $\boldsymbol{n}_w(i) \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma_w^2\boldsymbol{I}_{N_w}\right)$ is the noise, and $\boldsymbol{H}_w \triangleq [\boldsymbol{h}_{w,1}\ \boldsymbol{h}_{w,2}\ \cdots\ \boldsymbol{h}_{w,N_w-1}\ \boldsymbol{h}_{w,N_w}]^{\mathrm{T}} \in \mathbb{C}^{N_w \times N_t}$ denotes the channel coefficients from Alice to Willie, with $\boldsymbol{h}_{w,k} \in \mathbb{C}^{N_t \times 1}$ ($k = 1, 2, \ldots, N_w$) being the normalized channel coefficients from Alice to Willie's k-th antenna. Similar to the channel coefficients from Alice to Bobs, herein, $\boldsymbol{h}_{w,k}$ is modeled as $\boldsymbol{h}_{w,k} = \sqrt{\frac{K_w}{K_w+1}}\bar{\boldsymbol{h}}_w\mathrm{e}^{\iota\theta_{w,k}} + \sqrt{\frac{1}{K_w+1}}\hat{\boldsymbol{h}}_{w,k}$, where K_w is the Rician K-factor, $\theta_{w,k}$ denotes the phase of the LoS component of the channel coefficient from Alice's first antenna to Willie's k-th antenna, $\bar{\boldsymbol{h}}_w$ is the same for different k due to the far field assumption[3)], and $\hat{\boldsymbol{h}}_{w,k}$ follows $\mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{I}_{N_t}\right)$ as before.

1) This assumption is widely adopted to overcome the difficulty of covert communication [11] due to the imbalanced targets of Bob and Willie. Specifically, Willie merely determines whether the communication behavior exists, whereas Bob needs to recover the information. Note that Alice and Bob only need to pre-share a short secret sequence, and then they can generate the codebook by using key stream generators.

2) Herein, Willie's equivalent channel coefficients refer to the product of $\sqrt{d_w^{-\alpha}}$, $\boldsymbol{H}_w$, and $\boldsymbol{w}$ in (3).

3) The far field assumption requires that Alice and Willie's antenna spacing is much smaller than the distance between Alice and Willie. As will be shown, Willie's array configuration is reflected in parameter $\theta_{w,k}$ but does not influence the following analytical results, due to the far field assumption. In addition, the assumption of the UPA at Alice is only for the ease of presenting the channel coefficients. Alice can also have arbitrary array configuration, in which case only the steering vector should be modified.

We focus on the scenario where Alice does not know Willie's exact position but knows all the possible positions where Willie may be located[4] as well as the Rician K-factor of Willie's channel[5]. Mathematically, Alice knows a set $\{\bar{\boldsymbol{h}}_w^{(j)}, d_w^{(j)}, K_w^{(j)} | j \in J\} \triangleq \mathbb{W}$, where J denotes Willie's possible positions. Note that $\bar{\boldsymbol{h}}_w^{(j)}$ is merely determined by the angle of departure from Alice to position j. As for the CSI of Bobs' channels, two possible cases need to be considered. In case 1, Alice knows $\boldsymbol{h}_{b,m}$ for each m, which happens when Bobs do not need to hide themselves and can periodically transmit pilot. In case 2, Alice only knows the statistical CSI $\left\{\bar{\boldsymbol{h}}_{b,m}, d_{b,m}, K_{b,m} \middle| m = 1, 2, \ldots, M\right\}$, which happens when Bobs are not allowed to transmit pilot. As case 2 is more complicated, this paper focuses on case 2, and the developed algorithms can be reduced to case 1 by modifying the QoS constraint in the formulated problems.

2.2 Analyses of the detection performance at Willie

Willie's detection performance is measured by the sum of false alarm probability and miss detection probability $\Pr(\mathcal{D}_1|H_0)+\Pr(\mathcal{D}_0|H_1) \triangleq \xi$ [6–10]. Following Lemmas 1 and 2 of [11], it can be demonstrated that Willie's optimal detection performance can be given by

$$\xi^* = 1 - \frac{\gamma\left(n, \frac{n\mathcal{T}^*}{\sigma_w^2}\right)}{\Gamma(n)} + \frac{\gamma\left(n, \frac{n\mathcal{T}^*}{d_w^{-\alpha}||\boldsymbol{H}_w\boldsymbol{w}||^2+\sigma_w^2}\right)}{\Gamma(n)}, \tag{4}$$

where $\mathcal{T}^* = \frac{(d_w^{-\alpha}||\boldsymbol{H}_w\boldsymbol{w}||^2+\sigma_w^2)\sigma_w^2}{d_w^{-\alpha}||\boldsymbol{H}_w\boldsymbol{w}||^2} \ln(\frac{d_w^{-\alpha}||\boldsymbol{H}_w\boldsymbol{w}||^2+\sigma_w^2}{\sigma_w^2})$. Herein, $\Gamma(z) \triangleq \int_0^\infty \mathrm{e}^{-t}t^{z-1}\mathrm{d}t$ is the Gamma function, whereas $\gamma(a,z) \triangleq \int_0^z t^{a-1}\mathrm{e}^{-t}\mathrm{d}t$ is the lower incomplete Gamma function.

Note that Alice cannot guarantee her desired value of ξ^* due to the random and unknown non-LoS components of Willie's channel. For such, the covert probability $\mathcal{P}_\xi \triangleq \Pr\{\xi^* \geqslant \xi_{\text{expect}}\}$ is adopted as the covertness metric for optimizing Alice's parameters, where ξ_{expect} $(0 < \xi_{\text{expect}} < 1)$ denotes the desired covertness level. Unfortunately, it is difficult to obtain the closed-form expression of $\mathcal{P}_\xi$ due to the incomplete Gamma functions in (4). To address this, a strict lower bound of the covert probability is derived so as to make a conservative design, whereas $\mathcal{P}_\xi$ will be used for numerical evaluation in Section 5. Specifically, denote $\mathbb{P}_0$ and $\mathbb{P}_1$ as the distribution of Willie's observations under H_0 and H_1, respectively. As per [13], equation $\xi^* = 1 - \mathcal{V}_T(\mathbb{P}_0, \mathbb{P}_1)$ holds, where $\mathcal{V}_T(\mathbb{P}_0, \mathbb{P}_1)$ denotes the total variation distance between $\mathbb{P}_0$ and $\mathbb{P}_1$. Furthermore, by invoking Pinsker's inequality $\mathcal{V}_T(\mathbb{P}_0, \mathbb{P}_1) \leqslant \sqrt{\frac{1}{2}D(\mathbb{P}_0||\mathbb{P}_1)}$ [14, Lemma 11.6.1], where $D(\mathbb{P}_0||\mathbb{P}_1)$ denotes the Kullback-Leibler divergence from $\mathbb{P}_0$ to $\mathbb{P}_1$, we have

$$\mathcal{P}_\xi = \Pr\{\mathcal{V}_T(\mathbb{P}_0, \mathbb{P}_1) \leqslant 1 - \xi_{\text{expect}}\} \geqslant \Pr\left\{D(\mathbb{P}_0||\mathbb{P}_1) \leqslant 2(1-\xi_{\text{expect}})^2\right\} \triangleq \mathcal{P}_{\text{KL}}. \tag{5}$$

For $\mathcal{P}_{\text{KL}}$, we present Proposition 1, the proof of which is presented in Appendix A.

Proposition 1. Denote the only solution of equation

$$n\left[\ln\left(\frac{x+\sigma_w^2}{\sigma_w^2}\right) - \frac{x}{x+\sigma_w^2}\right] = 2(1-\xi_{\text{expect}})^2 \tag{6}$$

with respect to x as ϵ, which can be efficiently determined by the bisection method because the left-hand side of (6) monotonically increases with x. Then, we have

$$\mathcal{P}_{\text{KL}} = 1 - Q_{N_w}\left(\sqrt{2K_wN_w\left|\bar{\boldsymbol{h}}_w^{\text{T}}\frac{\boldsymbol{w}}{||\boldsymbol{w}||}\right|^2}, \sqrt{\frac{2\epsilon(K_w+1)}{d_w^{-\alpha}||\boldsymbol{w}||^2}}\right), \tag{7}$$

where $Q_\nu(a,b) = \frac{1}{a^{\nu-1}}\int_b^\infty t^\nu \mathrm{e}^{-\frac{t^2+a^2}{2}} I_{\nu-1}(at)\,\mathrm{d}t$ is the generalized Marcum Q-function [15], with $I_v(z) = \left(\frac{1}{2}z\right)^v \sum_{k=0}^\infty \frac{\left(\frac{1}{4}z^2\right)^k}{k!\Gamma(v+k+1)}$ being the modified Bessel function of the first kind.

4) Conservatively, if a specific position is suspicious, in the sense that whether Willie may be there is unsure, the position can be included as a possible position.

5) Rician K-factor can be obtained by measuring the wireless propagation environment in advance. If the Rician K-factor is uncertain, we can expand the set $\mathbb{W}$ by considering multiple different Rician K-factors, so as to conservatively design the parameters.

Remark 1. As per [15], $Q_\nu(a,b)$ monotonically increases with a and ν for all $a \geqslant 0$ and $b, \nu > 0$. In addition, it monotonically decreases with b for all $a, b \geqslant 0$ and $\nu > 0$. Meanwhile, note that $\frac{\boldsymbol{w}}{||\boldsymbol{w}||}$ indeed is the direction of the beamforming vector, whereas $||\boldsymbol{w}||^2 \triangleq P_t$ is the transmit power. Also, note that ϵ monotonically decreases with n. As a result, it follows from (7) that $\mathcal{P}_{\mathrm{KL}}$ will increase if Willie has less antennas, or if Alice adjusts the null lobe towards Willie, decreases the transmit power, decreases the number of channel uses, or stays away from Willie, which accords with intuitions.

2.3 Analyses of the decoding performance at Bob

For conciseness, index m is omitted and we focus on the decoding performance at a specific receiver. For finite-block-length transmission, the decoding error probability cannot be ignored [16]. For such, the effective throughput is adopted as the metric for decoding performance, which can be written as [16]

$$C = nR(1-\delta). \tag{8}$$

Herein, R denotes the transmission rate in terms of nats per channel use. δ represents the decoding error probability, which can be well approximated by [16]

$$\delta \approx Q\left(\frac{\sqrt{n}\left(\ln\left(1+\gamma_b\right)-R\right)}{\sqrt{1-\left(\gamma_b+1\right)^{-2}}}\right). \tag{9}$$

Herein, $\gamma_b = d_b^{-\alpha}\left|\boldsymbol{h}_b^{\mathrm{T}}\boldsymbol{w}\right|^2/\sigma_b^2$ is the received SNR. As per (8) and (9), C increases with n. Meanwhile, as per remark 1, $\mathcal{P}_{\mathrm{KL}}$ decreases when n increases. This makes n an important parameter to be optimized. However, expression (9) is too complicated for further analysis. To address this, we define $\upsilon \triangleq \mathrm{e}^R - 1$ and $\kappa \triangleq \sqrt{\frac{n}{2\pi(\mathrm{e}^{2R}-1)}}$, and further adopt approximation [17, 18]

$$\delta \approx \begin{cases} 1, & \gamma_b < \upsilon - \dfrac{1}{2\kappa}, \\ \dfrac{1}{2} - \kappa\left(\gamma_b - \upsilon\right), & \upsilon - \dfrac{1}{2\kappa} \leqslant \gamma_b < \upsilon + \dfrac{1}{2\kappa}, \\ 0, & \upsilon + \dfrac{1}{2\kappa} \leqslant \gamma_b. \end{cases} \tag{10}$$

Furthermore, note that decoding error is disastrous in covert broadcast communication with multiple receivers, because both Bobs' feedback and Alice's re-transmission occupy additional resources and increase the risk of exposing the communication behavior. Therefore, according to (10), inequality $\gamma_b \geqslant \upsilon + \frac{1}{2\kappa}$ is adopted as the QoS requirement, in which case the received SNR is large enough and the decoding error probability is deemed small enough. Next, because the achievable rate in covert communication is normally small, we have $\upsilon \approx R$ and $\kappa \approx \frac{\sqrt{n}}{2\sqrt{\pi R}}$ [19]. As a result, the QoS requirement can be written as

$$\left|\boldsymbol{h}_b^{\mathrm{T}}\boldsymbol{w}\right|^2 \geqslant \frac{\sigma_b^2}{d_b^{-\alpha}}\left(R + \frac{\sqrt{\pi R}}{\sqrt{n}}\right). \tag{11}$$

Suppose that D nats covert information needed to be transmitted. We can further rewrite (11) as

$$\left|\boldsymbol{h}_b^{\mathrm{T}}\boldsymbol{w}\right|^2 \geqslant \frac{\sigma_b^2}{d_b^{-\alpha}}\left(\frac{D + \sqrt{\pi D}}{n}\right). \tag{12}$$

3 Covert probability maximization for single-antenna Willie

As Willie's actual position is unknown, we aim to maximize the minimal $\mathcal{P}_{\mathrm{KL}}$ among possible Willie's positions. Before considering multi-antenna Willie, this section focuses on single-antenna Willie, in which case a tight upper bound of the Marcum Q function can be invoked to develop simpler algorithms.

3.1 Problem formulation and conversion

Alice cannot guarantee the QoS requirement (12) due to the random and unknown non-LoS components of Bobs' channels. Alternatively, the probability of reliable reception is adopted as the QoS constraint, which can be written as

$$\Pr\left\{\left|\boldsymbol{h}_{b,m}^{\mathrm{T}}\boldsymbol{w}\right|^2 \geqslant \frac{\sigma_b^2}{d_{b,m}^{-\alpha}}\left(\frac{D+\sqrt{\pi D}}{n}\right)\right\} \geqslant \eta \ (m=1,2,\ldots,M),$$

where $\eta \in (0,1)$ denotes the desired probability. Next, following a similar derivation process as that in Appendix A, we have

$$\Pr\left\{\left|\boldsymbol{h}_{b,m}^{\mathrm{T}}\boldsymbol{w}\right|^2 \geqslant \frac{\sigma_b^2}{d_{b,m}^{-\alpha}}\left(\frac{D+\sqrt{\pi D}}{n}\right)\right\} = Q_1\left(\sqrt{2K_{b,m}\left|\bar{\boldsymbol{h}}_{b,m}^{\mathrm{T}}\frac{\boldsymbol{w}}{||\boldsymbol{w}||}\right|^2}, \sqrt{\frac{2\sigma_b^2\left(\frac{D+\sqrt{\pi D}}{n}\right)(K_{b,m}+1)}{d_{b,m}^{-\alpha}\,||\boldsymbol{w}||^2}}\right).$$

As a result, the optimization problem can be given by

$$\begin{aligned}
\mathrm{P1}: &\min_{\boldsymbol{w},n}\max_{j\in J}\left\{Q_1\left(\sqrt{2K_w^{(j)}\left|\left(\bar{\boldsymbol{h}}_w^{(j)}\right)^{\mathrm{T}}\frac{\boldsymbol{w}}{||\boldsymbol{w}||}\right|^2}, \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}\,||\boldsymbol{w}||^2}}\right)\right\}\\
\text{s.t. (a) } &Q_1\left(\sqrt{2K_{b,m}\left|\bar{\boldsymbol{h}}_{b,m}^{\mathrm{T}}\frac{\boldsymbol{w}}{||\boldsymbol{w}||}\right|^2}, \sqrt{\frac{2\sigma_b^2\left(\frac{D+\sqrt{\pi D}}{n}\right)(K_{b,m}+1)}{d_{b,m}^{-\alpha}\,||\boldsymbol{w}||^2}}\right) \geqslant \eta,\ \forall m=1,2,\ldots,M,\\
&\text{(b) } ||\boldsymbol{w}||^2 \leqslant P_t^{\max},\\
&\text{(c) } n \leqslant n_{\max},\ n\in\mathbb{N}^+,
\end{aligned} \tag{13}$$

where $P_t^{\max}$ denotes the transmit power limit, whereas $n_{\max}$ is the maximum allowable number of channel uses due to the delay constraint. Original problem P1 is difficult to solve for two reasons. Firstly, the areas where Willie may be located are generally irregular, and meanwhile Willie's possible positions are generally infinite, which makes it difficult to handle the max operator in the objective function. Secondly, the Marcum Q function involves complicated integral. To tackle the first difficulty, we note that with a given beamforming vector $\boldsymbol{w}$, the beamforming gain at a specific position and that at the adjacent positions do not differ too much. Therefore, we propose to uniformly sample the areas where Willie may be located, and maximize the minimal $\mathcal{P}_{\mathrm{KL}}$ among the sample positions[6]. To tackle the second difficulty, a tight upper bound $Q_1(a,b) \leqslant \exp\left(-\frac{(b-a)^2}{2}\right)$ [20] is invoked to simplify the Marcum Q function. As a result, the objective of the optimization problem is reformulated as

$$\max_{\boldsymbol{w},n}\min_{j\in J'}\left\{\sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}\,||\boldsymbol{w}||^2}} - \sqrt{2K_w^{(j)}\left|\left(\bar{\boldsymbol{h}}_w^{(j)}\right)^{\mathrm{T}}\frac{\boldsymbol{w}}{||\boldsymbol{w}||}\right|^2}\right\}, \tag{14}$$

where $J' \subseteq J$ denotes the sample positions. Indeed, inequality $Q_1(a,b) \leqslant \exp\left(-\frac{(b-a)^2}{2}\right)$ is guaranteed only when $b > a$ [20]. Nonetheless, on one hand, as per the monotonicity properties of the Marcum Q function mentioned in remark 1, the original problem P1 tends to decrease a and increase b. On the other hand, the reformulated objective also tends to decrease a and increase b. Therefore, the reformulated problem has the same optimization trend as the original problem, and meanwhile their solutions are expected to satisfy inequality $b > a$. Note that even if inequality $b > a$ is not satisfied, the reformulated problem tends to enhance covertness, since decreasing a and increasing b correspond to decreasing $||\boldsymbol{w}||^2$, n, and $|(\bar{\boldsymbol{h}}_w^{(j)})^{\mathrm{T}}\frac{\boldsymbol{w}}{||\boldsymbol{w}||}|^2$, which means decreasing the transmit power, decreasing the number of channel uses, and adjusting the null lobe towards Willie.

6) If the areas where Willie may be located are regular and meanwhile have a special relationship with respect to Alice's position, one may simplify the original objective function. Nonetheless, such a case rarely happens in practice.

Even with reformulated objective (14), P1 is still difficult to solve due to the complicated Marcum Q function in the QoS constraint. To address this, we note that variable $\boldsymbol{w}$ appears in the form of either $\frac{\boldsymbol{w}}{||\boldsymbol{w}||}$ or $||\boldsymbol{w}||^2$. Meanwhile, in the QoS constraint, $||\boldsymbol{w}||^2$ and variable n are in the same component of the Marcum Q function, whereas $\frac{\boldsymbol{w}}{||\boldsymbol{w}||}$ merely appears in the other component. This motivates us to decompose beamforming vector $\boldsymbol{w}$ into two optimization variables, the beamforming direction $\boldsymbol{w}_{\rm d} \triangleq \frac{\boldsymbol{w}}{||\boldsymbol{w}||}$ and the transmit power $P_t \triangleq ||\boldsymbol{w}||^2$. As a result, problem P1 is converted into

$$\begin{aligned}
\mathrm{P1(a)}: \max_{\boldsymbol{w}_{\rm d},P_t,n} \ & \min_{j\in J'} \left\{ \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}} - \sqrt{2K_w^{(j)}\left|\left(\bar{\boldsymbol{h}}_w^{(j)}\right)^{\rm T}\boldsymbol{w}_{\rm d}\right|^2} \right\} \\
\text{s.t. (a) } & Q_1\left(\sqrt{2K_{b,m}\left|\bar{\boldsymbol{h}}_{b,m}^{\rm T}\boldsymbol{w}_{\rm d}\right|^2}, \sqrt{\frac{2\sigma_b^2\left(\frac{D+\sqrt{\pi D}}{n}\right)(K_{b,m}+1)}{d_{b,m}^{-\alpha}P_t}}\right) \geqslant \eta,\ \forall m=1,2,\ldots,M, \\
& \text{(b) } 0<P_t\leqslant P_t^{\max}, \\
& \text{(c) } n\leqslant n_{\max},\ n\in\mathbb{N}^+, \\
& \text{(d) } ||\boldsymbol{w}_{\rm d}||^2=1.
\end{aligned} \tag{15}$$

Now, with the help of the monotonicity properties of the Marcum Q function, variables $\boldsymbol{w}_{\rm d}$ and $\{P_t, n\}$ can be optimized alternately, which is elaborated as follows.

3.2 Optimize P_t and n with given $\boldsymbol{w}_{\rm d}$

Because $Q_\nu(a,b)$ monotonically decreases with b for all $a,b\geqslant 0$ and $\nu>0$ [15], for given $\boldsymbol{w}_{\rm d}$, the QoS constraint can be rewritten as

$$\frac{2\sigma_b^2(\frac{D+\sqrt{\pi D}}{n})(K_{b,m}+1)}{d_{b,m}^{-\alpha}P_t} \leqslant \chi_m^*,\ \forall m=1,2,\ldots,M, \tag{16}$$

where χ_m^* is the only solution of equation

$$Q_1\left(\sqrt{2K_{b,m}\left|\bar{\boldsymbol{h}}_{b,m}^{\rm T}\boldsymbol{w}_{\rm d}\right|^2}, \sqrt{\chi}\right)=\eta \tag{17}$$

with respect to χ and can be efficiently obtained by the bisection method. As a result, the sub-problem of optimizing P_t and n can be written as

$$\begin{aligned}
\mathrm{P-Sub1}: \max_{P_t,n} \ & \min_{j\in J'} \left\{ \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}} - \sqrt{2K_w^{(j)}\left|\left(\bar{\boldsymbol{h}}_w^{(j)}\right)^{\rm T}\boldsymbol{w}_{\rm d}\right|^2} \right\} \\
\text{s.t. (a) } & 2\sigma_b^2\left(\frac{D+\sqrt{\pi D}}{n}\right)\max_{m=1,2,\ldots,M}\left\{\frac{K_{b,m}+1}{d_{b,m}^{-\alpha}\chi_m^*}\right\} \leqslant P_t \leqslant P_t^{\max}, \\
& \text{(b) } n\leqslant n_{\max}, n\in\mathbb{N}^+.
\end{aligned} \tag{18}$$

Next, by noting that the objective function of P-Sub1 monotonically decreases with P_t, we can conclude that for any given n, the optimal P_t equals

$$P_t^* = 2\sigma_b^2\left(\frac{D+\sqrt{\pi D}}{n}\right)\max_{m=1,2,\ldots,M}\left\{\frac{K_{b,m}+1}{d_{b,m}^{-\alpha}\chi_m^*}\right\}. \tag{19}$$

Inserting (19) into P-Sub1, we arrive at the following one-dimension integer optimization problem:

$$\mathrm{P-Sub1(a)}: \max_{n} \ \min_{j\in J'} \left\{ \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t^*}} - \sqrt{2K_w^{(j)}\left|\left(\bar{\boldsymbol{h}}_w^{(j)}\right)^{\rm T}\boldsymbol{w}_{\rm d}\right|^2} \right\}$$

$$\text{s.t. (a) } P_t^* \leqslant P_t^{\max}, \qquad \text{(b) } n \leqslant n_{\max}, n \in \mathbb{N}^+. \tag{20}$$

Recall that ϵ is a monotonically decreasing implicit function with respect to n, which does not have a closed-form expression and is determined by the bisection method. Meanwhile, it follows from (19) that P_t^* also monotonically decreases with n. Thus, the monotonicity properties of the objective function of P-Sub1(a) with respect to n is unknown. In addition, n is indeed an integer optimization variable. To address the above difficulties, one dimension search is conducted to find the optimal n. Then, the optimal transmit power can be directly determined by inserting the optimal n into (19).

3.3 Optimize $\boldsymbol{w}_{\mathrm{d}}$ with given $\boldsymbol{P_t}$ and $\boldsymbol{n}$

Because $Q_\nu(a,b)$ monotonically increases with a for all $a \geqslant 0$ and $b, \nu > 0$ [15], for given P_t and given n, the QoS constraint can be rewritten as

$$2K_{b,m}\left|\bar{\boldsymbol{h}}_{b,m}^{\mathrm{T}}\boldsymbol{w}_{\mathrm{d}}\right|^2 \geqslant \vartheta_m^*,\ \forall m = 1,2,\ldots,M, \tag{21}$$

where ϑ_m^* is the only solution of equation

$$Q_1\left(\sqrt{\vartheta}, \sqrt{\frac{2\sigma_b^2\left(\frac{D+\sqrt{\pi D}}{n}\right)(K_{b,m}+1)}{d_{b,m}^{-\alpha}P_t}}\right) = \eta \tag{22}$$

with respect to ϑ and can be obtained by the bisection method. Next, note that $\boldsymbol{w}_{\mathrm{d}}$ appears in the form of either $||\boldsymbol{w}_{\mathrm{d}}||^2$ or $\left|\boldsymbol{h}^{\mathrm{T}}\boldsymbol{w}_{\mathrm{d}}\right|^2$. This motivates us to invoke the SDR method [21]. Specifically, define $\bar{\boldsymbol{H}}_{b,m} \triangleq (\bar{\boldsymbol{h}}_{b,m}\bar{\boldsymbol{h}}_{b,m}^{\mathrm{H}})^{\mathrm{T}}$, $\bar{\boldsymbol{H}}_w^{(j)} \triangleq (\bar{\boldsymbol{h}}_w^{(j)}(\bar{\boldsymbol{h}}_w^{(j)})^{\mathrm{H}})^{\mathrm{T}}$, and $\boldsymbol{W} \triangleq \boldsymbol{w}_{\mathrm{d}}\boldsymbol{w}_{\mathrm{d}}^{\mathrm{H}}$. By using $|\boldsymbol{h}^{\mathrm{T}}\boldsymbol{w}_{\mathrm{d}}|^2 = \mathrm{tr}((\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}})^{\mathrm{T}}\boldsymbol{W})$ and then by relaxing the rank-one constraint, the sub-problem of optimizing $\boldsymbol{w}_{\mathrm{d}}$ is formulated as

$$\begin{aligned}
\mathrm{P-Sub2}: \max_{\boldsymbol{W}}\ & f(\boldsymbol{W}) \triangleq \min_{j\in J'}\left\{\sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}} - \sqrt{2K_w^{(j)}\mathrm{tr}(\bar{\boldsymbol{H}}_w^{(j)}\boldsymbol{W})}\right\} \\
\text{s.t. } & \text{(a) } 2K_{b,m}\mathrm{tr}\left(\bar{\boldsymbol{H}}_{b,m}\boldsymbol{W}\right) \geqslant \vartheta_m^*,\ \forall m = 1,2,\ldots,M, \\
& \text{(b) } \mathrm{tr}(\boldsymbol{W}) = 1, \\
& \text{(c) } \boldsymbol{W} \succcurlyeq 0.
\end{aligned} \tag{23}$$

The constraints of P-Sub2 are affine. However, the objective function is not concave. To address this, the MM algorithm is invoked. In each iteration of the MM algorithm, the objective function is replaced by a surrogate function. For a maximization problem, the surrogate function is a lower bound and first-order local approximation of the original objective function. Then, the original problem is solved by iteratively solving a series of surrogate problems [22]. In our case, the surrogate function is constructed by invoking $\sqrt{x} \leqslant \sqrt{x_0} + \frac{1}{2}x_0^{-\frac{1}{2}}(x - x_0)$. Then, in the t-th iteration, the optimization problem is formulated as

$$\begin{aligned}
\mathrm{P-Sub2(a)}: \max_{\boldsymbol{W}} \min_{j\in J'}\ & \left\{\sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}} - \left(\sqrt{2K_w^{(j)}c_{j,t}} + \frac{K_w^{(j)}\left(\mathrm{tr}(\bar{\boldsymbol{H}}_w^{(j)}\boldsymbol{W}) - c_{j,t}\right)}{\sqrt{2K_w^{(j)}c_{j,t}}}\right)\right\} \\
\text{s.t. } & \text{(a) } 2K_{b,m}\mathrm{tr}\left(\bar{\boldsymbol{H}}_{b,m}\boldsymbol{W}\right) \geqslant \vartheta_m^*,\ \forall m = 1,2,\ldots,M, \quad \text{(b) } \mathrm{tr}(\boldsymbol{W}) = 1, \quad \text{(c) } \boldsymbol{W} \succcurlyeq 0,
\end{aligned} \tag{24}$$

where we define $c_{j,t} \triangleq \mathrm{tr}(\bar{\boldsymbol{H}}_w^{(j)}\boldsymbol{W}_{t-1})$, with $\boldsymbol{W}_{t-1}$ being the solution of $\boldsymbol{W}$ in the $(t-1)$-th iteration of the MM algorithm. P-Sub2(a) is a convex optimization problem and can be solved by the CVX tool.

Before executing the MM algorithm, a feasible initial $\boldsymbol{W}_0$ is needed, which is obtained by heuristically modifying P-Sub2 as the following convex optimization problem and using the CVX.

$$\begin{aligned}
\mathrm{P-Sub2(b)}: \min_{\boldsymbol{W}} \max_{j\in J'}\ & \left\{2K_w^{(j)}\mathrm{tr}(\bar{\boldsymbol{H}}_w^{(j)}\boldsymbol{W})\right\} \\
\text{s.t. } & \text{(a) } 2K_{b,m}\mathrm{tr}\left(\bar{\boldsymbol{H}}_{b,m}\boldsymbol{W}\right) \geqslant \vartheta_m^*,\ \forall m = 1,2,\ldots,M, \quad \text{(b) } \mathrm{tr}(\boldsymbol{W}) = 1, \quad \text{(c) } \boldsymbol{W} \succcurlyeq 0.
\end{aligned} \tag{25}$$

Algorithm 1 Covert probability maximization algorithm for single-antenna Willie

1: Initialize transmit power $P_t^{(0)} = P_{\max}$, number of channel uses $n^{(0)} = n_{\max}$, iteration index of alternating optimization $t_a = 1$.
2: Generate $\{\boldsymbol{w}_{\rm d}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}\}$ by steps 5–13.
3: **repeat**
4: $t_a = t_a + 1$;
5: Compute ϑ_m^* by the bisection method as per (22) based on $\{P_t, n\} = \{P_t^{(t_a-1)}, n^{(t_a-1)}\}$;
6: Initialize iteration index of the MM algorithm $t_m = 0$; Generate $\boldsymbol{W}_0$ by solving P-Sub2(b);
7: **repeat**
8: $t_m = t_m + 1$; generate $\boldsymbol{W}_{t_m}$ by solving P-Sub2(a);
9: **until** $f(\boldsymbol{W}_{t_m}) - f(\boldsymbol{W}_{t_m-1}) < \rho\,|f(\boldsymbol{W}_{t_m-1})|$;
10: Conduct eigenvalue decomposition $\boldsymbol{W}_{t_m} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\rm H}$;
11: Generate N_G beamforming directions $\boldsymbol{w}_r = \frac{\boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{r}}{||\boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{r}||}$ and compute the corresponding χ_m^* by bisection method as per (17);
12: For each $\boldsymbol{w}_r$, conduct one-dimension search of n for solving P-Sub1(a) and use (19) to calculate the optimal P_t;
13: Select the feasible candidate solution that maximizes $\mathcal{P}_{\rm KL}^*$ as the solution of the t_a-th iteration $\{\boldsymbol{w}_{\rm d}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}\}$;
14: **until** $\mathcal{P}_{\rm KL}^*(\boldsymbol{w}_{\rm d}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}) - \mathcal{P}_{\rm KL}^*(\boldsymbol{w}_{\rm d}^{(t_a-1)}, P_t^{(t_a-1)}, n^{(t_a-1)}) < \rho\mathcal{P}_{\rm KL}^*(\boldsymbol{w}_{\rm d}^{(t_a-1)}, P_t^{(t_a-1)}, n^{(t_a-1)})$;
15: **return** $\{\boldsymbol{w}_{\rm d}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}\}$.

Note that simply dropping the objective function of P-Sub2 can also generate $\boldsymbol{W}_0$. However, by adopting the objective function of P-Sub2(b), it is expected to obtain a good initial solution, since the optimization process of P-Sub2(b) tends to adjust the null lobe towards Willie.

Generally, the solution $\boldsymbol{W}^*$ obtained by iteratively solving P-Sub2(a) is not rank-one. The Gaussian randomization method is used to recover the final rank-one solution [21]. Specifically, we firstly compute the eigenvalue decomposition of $\boldsymbol{W}^*$ as $\boldsymbol{W}^* = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\rm H}$, and then generate multiple candidate beamforming directions $\boldsymbol{w}_r = \boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{r}/||\boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{r}||$, where $\boldsymbol{r}$ follows $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_t})$.

Note that the candidate directions $\boldsymbol{w}_r$, combined with the given P_t and n, do not necessarily satisfy the QoS constraint. Such a problem is typically solved by discarding the infeasible candidate solutions, or by intuitively modifying the solutions into feasible ones [21]. Then, the feasible candidate solution that optimizes the objective function is selected as the solution. In our case, we note that: (a) it may happen that $\boldsymbol{w}_r$ merely slightly violates the QoS constraint but leads to good covertness performance; (b) it may happen that $\boldsymbol{w}_r$ leads to bad covertness performance but is still potential because Bobs' received power is much more than the QoS requirements. Therefore, to fully utilize all the obtained candidate directions, we incorporate the alternating optimization of P_t and n into the selection of the final rank-one solution, which is elaborated in the following part.

3.4 Overall algorithm and complexity analyses

The whole algorithm for solving P1 is summarized in Algorithm 1, where N_G denotes the Gaussian randomization times, ρ is the convergence tolerance, whereas $\mathcal{P}_{\rm KL}^*(\boldsymbol{w}_{\rm d}, P_t, n)$ denotes the minimal $\mathcal{P}_{\rm KL}$ among J' based on solution $\{\boldsymbol{w}_{\rm d}, P_t, n\}$. Step 12 incorporates the alternating optimization of P_t and n into the selection of the final rank-one solution, as mentioned above. Steps 1 and 2 are for generating an initial feasible solution before alternating optimization. Note that due to the complexity of P1, we cannot determine whether it is feasible in advance. Also note that because P1 is infeasible if and only if the QoS requirements and resources constraints cannot be satisfied simultaneously, which is irrelevant to Willie, Alice should keep silence when P1 is infeasible. Due to the same reason, in step 1, P_t and n are initialized by their maximum allowable values, based on which the MM algorithm is executed to find the initial feasible beamforming direction.

The main computational complexity of the proposed algorithm is due to steps 11–13, because the Gaussian randomization technique requires multiple realizations. Specifically, the computational complexity of step 11 is $\mathcal{O}(N_t^2 + M(\log((\chi_{\rm up} - \chi_{\rm low})/\epsilon_1) + N_t))$, where $[\chi_{\rm low}, \chi_{\rm up}]$ and ϵ_1 denote the range and the solution accuracy of the bisection method, respectively. The computational complexity of step 12 is mainly due to the calculation of the objective function of P-Sub1(a) and thus is $\mathcal{O}(|J'|(n_{\max} + N_t))$, where $|J'|$ is the cardinality of J'. Note that in P-Sub1(a), the values of ϵ corresponding to different n can be calculated in advance. Finally, step 13 involves at most $n_{\max}|J'|$ times of calculation of the Marcum Q function. As a result, denote the iteration times of the alternating optimization as T_a, the computational complexity of Algorithm 1 can be written as $\mathcal{O}(T_a N_G[N_t^2 + M(\log((\chi_{\rm up} - \chi_{\rm low})/\epsilon_1) + N_t) + |J'|(n_{\max} + N_t)])$. A larger N_G leads to stabler results and has a higher probability of obtaining better results. In practice, N_G can be selected by jointly considering the acceptable delay for computation and the required performance.

4 Covert probability maximization for multi-antenna Willie

4.1 Problem formulation and conversion

In this subsection, we turn to multi-antenna Willie. As before, we aim to maximize the minimal $\mathcal{P}_{\mathrm{KL}}$ among Willie's possible positions. To tackle the optimization problem, as in Subsection 3.1, we introduce transmit power $P_t \triangleq ||\boldsymbol{w}||^2$ and beamforming direction $\boldsymbol{w}_{\mathrm{d}} \triangleq \frac{\boldsymbol{w}}{||\boldsymbol{w}||}$ as optimization variables, and replace the irregular and infinite set J by sample positions J'. Then, it follows from Proposition 1 that the optimization problem can be written as

$$\begin{aligned}
\mathrm{P2}: \min_{\boldsymbol{w}_{\mathrm{d}},P_t,n} \max_{j\in J'} & \left\{ Q_{N_w}\left(\sqrt{2K_w^{(j)}N_w\left|\left(\bar{\boldsymbol{h}}_w^{(j)}\right)^{\mathrm{T}}\boldsymbol{w}_{\mathrm{d}}\right|^2}, \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}}\right)\right\} \\
\text{s.t. (a) } & Q_1\left(\sqrt{2K_{b,m}\left|\bar{\boldsymbol{h}}_{b,m}^{\mathrm{T}}\boldsymbol{w}_{\mathrm{d}}\right|^2}, \sqrt{\frac{2\sigma_b^2\left(\frac{D+\sqrt{\pi D}}{n}\right)(K_{b,m}+1)}{d_{b,m}^{-\alpha}P_t}}\right) \geqslant \eta,\ \forall m=1,2,\ldots,M, \\
& \text{(b) } 0\leqslant P_t\leqslant P_t^{\max}, \\
& \text{(c) } n\leqslant n_{\max},\ n\in\mathbb{N}^+, \\
& \text{(d) } ||\boldsymbol{w}_{\mathrm{d}}||^2=1.
\end{aligned} \tag{26}$$

As before, by using the monotonicity properties of the generalized Marcum Q function, variables $\boldsymbol{w}_{\mathrm{d}}$ and $\{P_t, n\}$ can be optimized alternately. Herein, the approach for optimizing $\{P_t, n\}$ is similar to that in Subsection 3.2. The only difference is that the objective function of P-Sub1(a) should be replaced by the objective function of P2. Thus, the details are omitted for brevity.

As for the sub-problem of optimizing $\boldsymbol{w}_{\mathrm{d}}$, to the best knowledge of the authors, for $N_w \neq 1$, the existing upper bounds of the generalized Marcum Q function [20,23] are too complicated to simplify the objective function, which hinders us from adopting the same approach as before. Alternatively, an algorithm based on SCA and a backtracking method is invoked, which is elaborated in what follows.

4.2 Optimize $\boldsymbol{w}_{\mathrm{d}}$ with given $\boldsymbol{P_t}$ and $\boldsymbol{n}$

We firstly apply (21) and the SDR technique to transform the sub-problem of optimizing $\boldsymbol{w}_{\mathrm{d}}$, which leads to the following problem:

$$\begin{aligned}
\mathrm{P-Sub3}: \min_{\boldsymbol{W}}\ & g(\boldsymbol{W}) \triangleq \max_{j\in J'}\left\{ Q_{N_w}\left(\sqrt{2K_w^{(j)}N_w\mathrm{tr}(\bar{\boldsymbol{H}}_w^{(j)}\boldsymbol{W})}, \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}}\right)\right\} \\
\text{s.t. } & \text{(a) } 2K_{b,m}\mathrm{tr}\left(\bar{\boldsymbol{H}}_{b,m}\boldsymbol{W}\right)\geqslant\vartheta_m^*,\ \forall m=1,2,\ldots,M, \\
& \text{(b) } \mathrm{tr}\left(\boldsymbol{W}\right)=1, \\
& \text{(c) } \boldsymbol{W}\succcurlyeq 0,
\end{aligned} \tag{27}$$

where we define $\bar{\boldsymbol{H}}_{b,m} \triangleq (\bar{\boldsymbol{h}}_{b,m}\bar{\boldsymbol{h}}_{b,m}^{\mathrm{H}})^{\mathrm{T}}$, $\bar{\boldsymbol{H}}_w^{(j)} \triangleq (\bar{\boldsymbol{h}}_w^{(j)}(\bar{\boldsymbol{h}}_w^{(j)})^{\mathrm{H}})^{\mathrm{T}}$, and $\boldsymbol{W} \triangleq \boldsymbol{w}_{\mathrm{d}}\boldsymbol{w}_{\mathrm{d}}^{\mathrm{H}}$ as before. Until now, all the constraints are affine, whereas the objective function is highly complicated. To address this, the SCA algorithm and a backtracking method are invoked. Specifically, in each iteration of the SCA algorithm, the objective function is replaced by a surrogate function, which is convex and is a first-order local approximation of the original objective function. To proceed, denote the solution of $\boldsymbol{W}$ in the $(t-1)$-th iteration of the SCA algorithm as $\boldsymbol{W}_{t-1}$ and define $a_{j,t} \triangleq 2K_w^{(j)}N_w\mathrm{tr}(\bar{\boldsymbol{H}}_w^{(j)}\boldsymbol{W}_{t-1})$. By invoking $\frac{d^n}{da^n}\left[Q_\nu\left(\sqrt{a},b\right)\right] = (-2)^{-n}\sum_{p=0}^{n}(-1)^p\binom{n}{p}Q_{\nu+p}\left(\sqrt{a},b\right)$ [24], the first-order Taylor expansion of $Q_{N_w}\left(\sqrt{a}, \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}}\right)$ with respect to $a=a_{j,t}$ can be given by

$$\widetilde{Q}_{j,t}\left(a\right) = Q_{N_w}\left(\sqrt{a_{j,t}}, \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}}\right) + \chi_{j,t}\left(a-a_{j,t}\right), \tag{28}$$

Algorithm 2 Covert probability maximization algorithm for multi-antenna Willie

Initialize transmit power $P_t^{(0)} = P_{\max}$, number of channel uses $n^{(0)} = n_{\max}$, iteration index of alternating optimization $t_a = 1$;
Generate $\{\boldsymbol{w}_{\mathrm{d}}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}\}$ by steps 5–13;
repeat
  $t_a = t_a + 1$;
  Compute ϑ_m^* by the bisection method as per (22) based on $\{P_t, n\} = \{P_t^{(t_a-1)}, n^{(t_a-1)}\}$;
  Initialize iteration index of the SCA algorithm $t_m = 0$; Generate $\boldsymbol{W}_0$ by solving P-Sub2(b);
  repeat
    $t_m = t_m + 1$; generate $\widetilde{\boldsymbol{W}}_{t_m}$ by solving P-Sub3(a); generate $\boldsymbol{W}_{t_m}$ by backtracking as per (30);
  until $g(\boldsymbol{W}_{t_m-1}) - g(\boldsymbol{W}_{t_m}) < \rho g(\boldsymbol{W}_{t_m-1})$;
  Conduct eigenvalue decomposition $\boldsymbol{W}_{t_m} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}$;
  Generate N_G beamforming directions $\boldsymbol{w}_r = \frac{\boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{r}}{||\boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{r}||}$ and compute the corresponding χ_m^* by bisection method as per (17);
  For each $\boldsymbol{w}_r$, conduct one-dimension search of n and use (19) to calculate the optimal P_t;
  Select the feasible candidate solution that maximizes $\mathcal{P}_{\mathrm{KL}}^*$ as the solution of the t_a-th iteration $\{\boldsymbol{w}_{\mathrm{d}}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}\}$;
until $\mathcal{P}_{\mathrm{KL}}^*(\boldsymbol{w}_{\mathrm{d}}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}) - \mathcal{P}_{\mathrm{KL}}^*(\boldsymbol{w}_{\mathrm{d}}^{(t_a-1)}, P_t^{(t_a-1)}, n^{(t_a-1)}) < \rho\mathcal{P}_{\mathrm{KL}}^*(\boldsymbol{w}_{\mathrm{d}}^{(t_a-1)}, P_t^{(t_a-1)}, n^{(t_a-1)})$;
return $\{\boldsymbol{w}_{\mathrm{d}}^{(t_a)}, P_t^{(t_a)}, n^{(t_a)}\}$.

where

$$\chi_{j,t} = -\frac{1}{2}\left[Q_{N_w}\left(\sqrt{a_{j,t}}, \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}}\right) - Q_{N_w+1}\left(\sqrt{a_{j,t}}, \sqrt{\frac{2\epsilon(K_w^{(j)}+1)}{(d_w^{(j)})^{-\alpha}P_t}}\right)\right].$$

Then, in the t-th iteration, the surrogate problem is formulated as

$$\begin{aligned}\mathrm{P-Sub3(a)}: &\min_{\boldsymbol{W}} \max_{j\in J'} \left\{\widetilde{Q}_{j,t}\left(2K_w^{(j)}N_w \mathrm{tr}(\bar{\boldsymbol{H}}_w^{(j)}\boldsymbol{W})\right)\right\} \\ &\text{s.t.} \quad \text{(a)}\ 2K_{b,m}\mathrm{tr}\left(\bar{\boldsymbol{H}}_{b,m}\boldsymbol{W}\right) \geqslant \vartheta_m^*,\ \forall m = 1,2,\ldots,M, \qquad \text{(b)}\ \mathrm{tr}(\boldsymbol{W}) = 1, \qquad \text{(c)}\ \boldsymbol{W} \succcurlyeq 0.\end{aligned} \tag{29}$$

As per the chain rule for derivatives, the surrogate objective function is a first-order approximation of the original objective function $g(\boldsymbol{W})$ at $\boldsymbol{W} = \boldsymbol{W}_{t-1}$. P-Sub3(a) is convex and can be solved by the CVX.

Note that, unlike the MM algorithm, the SCA algorithm itself does not guarantee convergence. This problem is solved by the backtracking method [25]. Specifically, denote $\widetilde{\boldsymbol{W}}_t$ as the solution of P-Sub3(a) in the t-th iteration, define $\beta \in (0,1)$ as the backtracking factor, and denote δ_t as the maximum value in set $\{\beta^q\}_{q=0,1,2,\ldots}$ that satisfies

$$g(\boldsymbol{W}_{t-1}) > g\left(\boldsymbol{W}_{t-1}(1-\delta_t) + \widetilde{\boldsymbol{W}}_t\delta_t\right). \tag{30}$$

Then, we adopt $\boldsymbol{W}_t = \boldsymbol{W}_{t-1}(1-\delta_t) + \widetilde{\boldsymbol{W}}_t\delta_t$ as the final solution of the t-th iteration of the SCA algorithm. Because the original objective function is lower bounded by zero and inequality (30) guarantees improvement in each iteration, the SCA algorithm now guarantees convergence thanks to the backtracking method. Note that as long as $\widetilde{\boldsymbol{W}}_t \neq \boldsymbol{W}_{t-1}$ (otherwise, the algorithm converges), δ_t definitely exists. This is explained as follows. On one hand, the gradient of the original objective function at $\boldsymbol{W} = \boldsymbol{W}_{t-1}$ is the same as that of the surrogate function. On the other hand, the surrogate function is convex. Thus, direction $\widetilde{\boldsymbol{W}}_t - \boldsymbol{W}_{t-1}$ is the direction of descent of both P-Sub3(a) and P-Sub3 at $\boldsymbol{W} = \boldsymbol{W}_{t-1}$.

As before, a feasible $\boldsymbol{W}_0$ and a final rank-one solution are required before and after executing the SCA algorithm, respectively, which are obtained in the same way as in Subsection 3.3. The whole algorithm for solving P2 is summarized in Algorithm 2. Note that the computational complexity of Algorithm 2 is the same with Algorithm 1, because the procedures for generating and selecting the beamforming directions, which account for the major computational complexity of the two algorithms, remain unchanged.

5 Numerical results and discussion

This section presents representative numerical results to validate the theoretical results and to demonstrate the performance of the proposed algorithms. Without loss of generality, three Bobs and two areas where Willie may be located are considered. Specifically, as shown in Figure 2, Alice's first antenna is located at $(0, 0, 20)$ in the meter for ease of presentation. Bobs' positions are set at $(100\sin 60°, 100\cos 60°, 0)$,

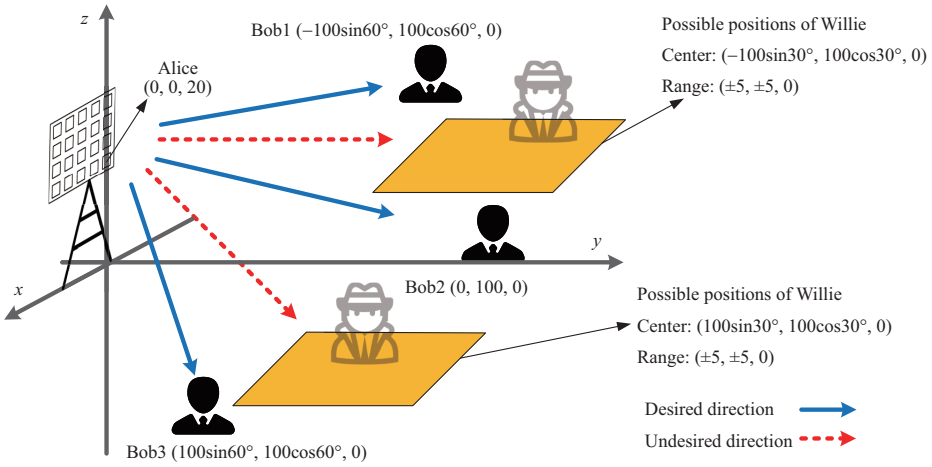

Figure 2 (Color online) Setup of the positions in the simulations.

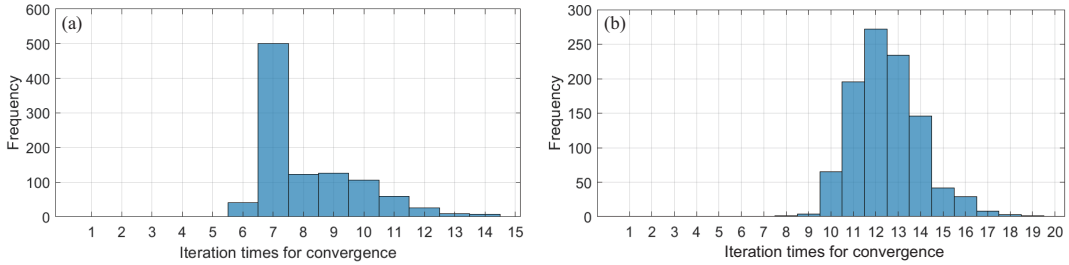

Figure 3 (Color online) Distribution of required iteration times. (a) $N_w = 1$; (b) $N_w = 2$.

$(0, 100, 0)$, and $(-100\sin 60°, 100\cos 60°, 0)$ in meter, respectively. Two square areas of side length 10 meters are set as the areas where Willie may be located, whose centers are $(100\sin 30°, 100\cos 30°, 0)$ and $(-100\sin 30°, 100\cos 30°, 0)$ in meter, respectively. The antenna spacings of the UPA at Alice are set as $d_x = d_z = 0.03$ m, which correspond to the half wave length of frequency 5 GHz. The number of the transmit antennas in the x direction is set as $N_{t,x} = 4$. Unless otherwise specified, the remaining parameters are set as $N_{t,z} = 8$, $n_{\max} = 500$, $D = 40$ nats, $P_t^{\max} = 1$ Watt, $\sigma_b^2 = \sigma_w^2 = -40$ dBm, $\alpha = 3$, $\xi_{\text{expect}} = 0.95$, $\eta = 0.95$, $\beta = 0.8$, $N_G = 1000$. The Rician K-factors for all the channels are set as 10[7)].

For ease of presentation, we denote S as the spacing between the uniformly sampled positions in the areas where Willie may be located. Unless otherwise specified, when executing the proposed algorithms, we set $S = 10$ meters, which means that the eight vertexes of the two square areas are selected as J'. The impacts of the selection of J' will be discussed later. On the other hand, it is expected that we can numerically simulate the minimal $\mathcal{P}_\xi$ among all the possible Willie's positions based on the optimization solutions, which directly reflects the effectiveness of the proposed schemes. This is done by setting $S = 1$ meter and using the densely sampled positions to represent all the possible positions, as it is impossible to determine the minimum among an infinite number of positions. To avoid ambiguities, in the rest of the paper, the spacing adopted during the execution of the algorithms is denoted by S_{Ana}, whereas the spacing adopted in numerical evaluations is denoted by S_{Sim}.

Figures 3(a) and (b) show the convergence behaviors of Algorithms 1 and 2, respectively, in terms of the distribution of the required iteration times for convergence. The results are based on 1000 times execution of the algorithms, which show that the required iteration times are stable and acceptable.

Figures 4(a) and (b) show the performance of the proposed algorithms. Herein, $\mathcal{P}_\xi^*$ denotes the minimal $\mathcal{P}_\xi$ among sampled positions, whereas $\mathcal{P}_{\text{KL}}^*$ denotes the minimal $\mathcal{P}_{\text{KL}}$ among sampled positions. 'Ana. $1 - \mathcal{P}_{\text{KL}}^*$' represents the analytical value of $1 - \mathcal{P}_{\text{KL}}^*$ among the sample positions that adopted in the execution of the algorithms, which is obtained by (7) based on the optimization solutions. 'Sim. $1 - \mathcal{P}_{\text{KL}}^*$' and 'Sim. $1 - \mathcal{P}_\xi^*$' denote the simulation values of $1 - \mathcal{P}_{\text{KL}}^*$ and $1 - \mathcal{P}_\xi^*$ among the sample positions that considered in numerical evaluations, respectively, which are obtained by averaging over the random

7) Due to the space limit, the performance under different Rician K-factors is not shown. Generally, larger Rician K-factors lead to better covertness performance, as Alice knows nothing about the non-LoS components of the channels.

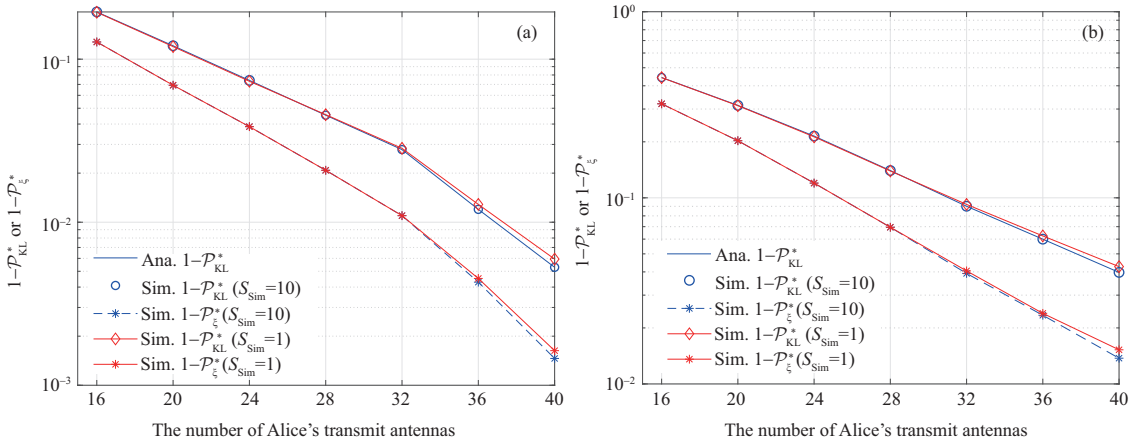

Figure 4 (Color online) Covertness performance. (a) $S_{\mathrm{Ana}} = 10$, $N_w = 1$; (b) $S_{\mathrm{Ana}} = 10$, $N_w = 2$.

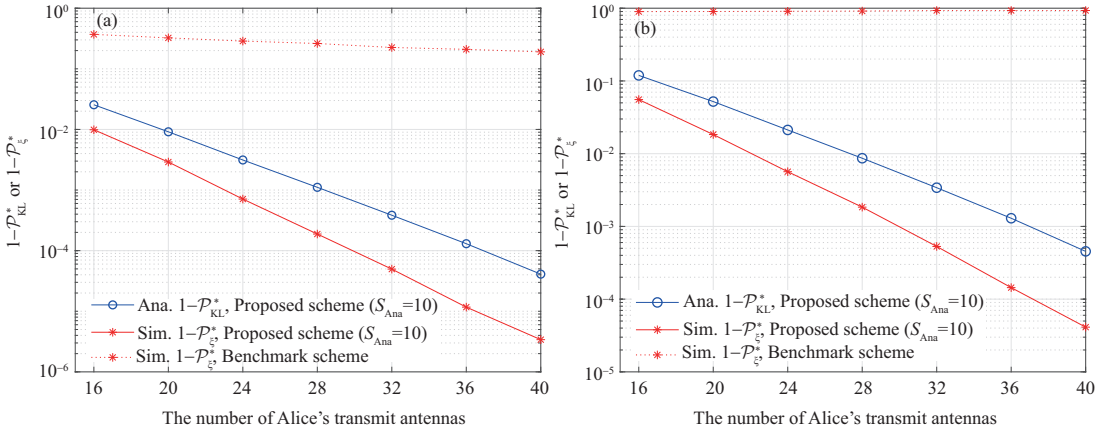

Figure 5 (Color online) Comparisons with MRT. (a) $S_{\mathrm{Sim}} = 1$, $N_w = 1$; (b) $S_{\mathrm{Sim}} = 1$, $N_w = 2$.

realizations of Willie's channel based on the optimization solutions. Several observations are drawn from the figures: (1) The covertness performance can be dramatically improved by increasing the number of Alice's transmit antennas, since using more transmit antennas can produce sharper and more flexible beams; (2) For $S_{\mathrm{Sim}} = S_{\mathrm{Ana}}$, 'Sim. $1 - \mathcal{P}^*_{\mathrm{KL}}$' matches well with 'Ana. $1 - \mathcal{P}^*_{\mathrm{KL}}$', which validates the analytical results; (3) As the number of Alice's transmit antennas increases, the minimal covert probability among the sample positions of $S = 1$ becomes smaller than that of $S = 10$. This is because, as the number of transmit antennas increases, the formed beams can be sharper, which leads to larger differences in the detection performance within a certain area. In this case, one should consider a denser J' when executing the algorithms, so as to effectively maximize the minimal covert probability among all Willie's possible positions.

Figures 5(a) and (b) compare the performance of the proposed scheme with the MRT benchmark. The later is widely adopted as the beamforming strategy in the existing works regarding multi-antenna covert communications. Herein, to make a fair comparison, only one receiver is considered, whose position is set at $(0, 100, 0)$ in meter. Specifically, the benchmark scheme adopts $\bar{\boldsymbol{h}}_b^* / \left|\left|\bar{\boldsymbol{h}}_b\right|\right|$ as the beamforming direction [9], whereas the transmit power and the number of channel uses are set as their optimal values in the same way as the proposed algorithms. As can be observed from the figures, the covert probability of the proposed scheme is significantly larger than that of the benchmark. This result indicates that simply adjusting the main lobe towards Bob is indeed far away from optimum, as the sidelobes' directions are not under control. This phenomenon indeed reveals the importance of making full use of Alice's limited knowledge of Willie's position, as done by the proposed scheme.

6 Concluding remarks

For the covert broadcast communication scenarios where Alice only knows the suspicious areas where Willie may be located, this paper jointly designed the transmit power, the beamforming direction, and the number of channel uses, so as to maximize the minimal possible covert probability. The results showed that the widely-adopted MRT strategy is indeed far away from optimum, and highlighted the importance of leveraging the available information about Willie's position. An important future direction is to consider different data streams and different beamforming vectors for different receivers.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62171445, 62371457, 62201590) and Youth Innovation Team Project of Shaanxi Higher Education Institutions (Grant No. LGNY-25).

References

1 Lu K, Liu H, Zeng L, et al. Applications and prospects of artificial intelligence in covert satellite communication: a review. Sci China Inf Sci, 2023, 66: 121301
2 He B, Yan S, Zhou X, et al. On covert communication with noise uncertainty. IEEE Commun Lett, 2017, 21: 941–944
3 Bash B A, Goeckel D, Towsley D. Covert communication gains from adversary's ignorance of transmission time. IEEE Trans Wireless Commun, 2016, 15: 8394–8405
4 He B, Yan S, Zhou X, et al. Covert wireless communication with a Poisson field of interferers. IEEE Trans Wireless Commun, 2018, 17: 6005–6017
5 Yu X C, Luo Y, Chen W. Covert communication with beamforming over MISO channels in the finite blocklength regime. Sci China Inf Sci, 2021, 64: 192303
6 Wang C, Li Z, Ng D W K. Covert rate optimization of millimeter wave full-duplex communications. IEEE Trans Wireless Commun, 2022, 21: 2844–2861
7 Chen X, Zheng T X, Dong L, et al. Enhancing MIMO covert communications via intelligent reflecting surface. IEEE Wireless Commun Lett, 2022, 11: 33–37
8 Lin Y D, Jin L, Huang K Z, et al. Threat region development of covert wireless communication based on 3D beamforming. Sci Sin Inform, 2021, 51: 1360–1374
9 Lin Y, Jin L, Huang K, et al. Covert threat region analysis of 3-D location-based beamforming in Rician channel. IEEE Wireless Commun Lett, 2022, 11: 1253–1257
10 Forouzesh M, Azmi P, Mokari N, et al. Covert communication using null space and 3D beamforming: uncertainty of Willie's location information. IEEE Trans Veh Technol, 2020, 69: 8568–8576
11 Shahzad K, Zhou X, Yan S. Covert wireless communication in presence of a multi-antenna adversary and delay constraints. IEEE Trans Veh Technol, 2019, 68: 12432–12436
12 Lu A A, Gao X, Meng X, et al. Omnidirectional precoding for 3D massive MIMO with uniform planar arrays. IEEE Trans Wireless Commun, 2020, 19: 2628–2642
13 Lehmann E, Romano J. Testing Statistical Hypotheses. 3rd ed. New York: Springer, 2005
14 Cover T M, Thomas J A. Elements of Information Theory, 2nd ed. Hoboken: John Wiley & Sons, 2006
15 Sun Y, Baricz Á, Zhou S. On the monotonicity, log-concavity, and tight bounds of the generalized Marcum and Nuttall Q-functions. IEEE Trans Inform Theor, 2010, 56: 1166–1186
16 Yan S, He B, Zhou X, et al. Delay-intolerant covert communications with either fixed or random transmit power. IEEE TransInformForensic Secur, 2019, 14: 129–140
17 Yu B, Cai Y, Wu D. Joint access control and resource allocation for short-packet-based mMTC in status update systems. IEEE J Sel Areas Commun, 2021, 39: 851–865
18 Wang C, Li Z, Zheng T X, et al. Intelligent reflecting surface-aided full-duplex covert communications: information freshness optimization. IEEE Trans Wireless Commun, 2023, 22: 3246–3263
19 Lu X, Yang W, Yan S, et al. Covertness and timeliness of data collection in UAV-aided wireless-powered IoT. IEEE Int Things J, 2022, 9: 12573–12587
20 Simon M K, Alouini M S. Exponential-type bounds on the generalized Marcum Q-function with application to error probability analysis over fading channels. IEEE Trans Commun, 2000, 48: 359–366
21 Luo Z, Ma W, So A, et al. Semidefinite relaxation of quadratic optimization problems. IEEE Signal Process Mag, 2010, 27: 20–34
22 Sun Y, Babu P, Palomar D P. Majorization-minimization algorithms in signal processing, communications, and machine learning. IEEE Trans Signal Process, 2017, 65: 794–816
23 Li R, Kam P Y, Fu H. New representations and bounds for the generalized Marcum Q-function via a geometric approach, and an application. IEEE Trans Commun, 2010, 58: 157–169
24 Brychkov Y A. On some properties of the Marcum Q function. Integral Transforms Spec Functions, 2012, 23: 177–182
25 Razaviyayn M. Successive convex approximation: analysis and applications. Dissertation for the Doctoral Degree. Minneapolis: University of Minnesota, 2014. 19–21

Appendix A Proof of Proposition 1

We first derive the expression of $D\left(\mathbb{P}_0||\mathbb{P}_1\right) = \int f\left(\boldsymbol{Y}|H_0\right)\ln\left(\frac{f(\boldsymbol{Y}|H_0)}{f(\boldsymbol{Y}|H_1)}\right)\mathrm{d}\boldsymbol{Y}$, where $f\left(\boldsymbol{Y}|H_q\right)$ $(q=1,2)$ denotes the probability density function of Willie's observations $\boldsymbol{Y} \triangleq [\boldsymbol{y}_w(1), \boldsymbol{y}_w(2), \ldots, \boldsymbol{y}_w(n)]$ under hypothesis H_q. Since $\boldsymbol{y}_w(i)$ is independent for different i, we have

$$\begin{aligned} f\left(\boldsymbol{Y}|H_q\right) &= \prod_{i=1}^{n}\frac{1}{\pi^{N_w}\left|\boldsymbol{\Sigma}_q\right|}\exp\left(-\boldsymbol{y}_w^{\mathrm{H}}(i)\boldsymbol{\Sigma}_q^{-1}\boldsymbol{y}_w(i)\right) = \pi^{-N_w n}\left|\boldsymbol{\Sigma}_q\right|^{-n}\exp\left(-\sum_{i=1}^{n}\mathrm{tr}\left(\boldsymbol{\Sigma}_q^{-1}\boldsymbol{y}_w(i)\boldsymbol{y}_w^{\mathrm{H}}(i)\right)\right) \\ &= \pi^{-N_w n}\left|\boldsymbol{\Sigma}_q\right|^{-n}\exp\left(-\mathrm{tr}\left(\boldsymbol{\Sigma}_q^{-1}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}\right)\right), \end{aligned} \tag{A1}$$

where $\boldsymbol{\Sigma}_0 = \sigma_w^2 \boldsymbol{I}_{N_w}$ and $\boldsymbol{\Sigma}_1 = d_w^{-\alpha} \boldsymbol{H}_w \boldsymbol{w} (\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}} + \sigma_w^2 \boldsymbol{I}_{N_w}$ are the covariance matrixes of $\boldsymbol{y}_w(i)$ under H_0 and H_1, respectively. Next, by noting that a rank-one matrix of order N_w has at least $N_w - 1$ zero eigenvalues, we can conclude that the eigenvalues of $d_w^{-\alpha} \boldsymbol{H}_w \boldsymbol{w} (\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}}$ are $\{d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2, \underbrace{0, \ldots, 0}_{N_w - 1}\}$. Furthermore, note that if λ is an eigenvalue of matrix $\boldsymbol{A}$, then $\lambda + \sigma^2$ is an eigenvalue of $\boldsymbol{A} + \sigma^2 \boldsymbol{I}$. As a result, we can conclude that the determinant of $\boldsymbol{\Sigma}_1$ is $|\boldsymbol{\Sigma}_1| = \left(d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2 + \sigma_w^2\right) \sigma_w^{2(N_w - 1)}$. On the other hand, by invoking the Woodbury Matrix identity for matrix inversion, one can conclude that $\boldsymbol{\Sigma}_1^{-1} = \frac{1}{\sigma_w^2} \boldsymbol{I}_{N_w} - \frac{\boldsymbol{H}_w \boldsymbol{w} (\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}}}{\sigma_w^2 (\frac{\sigma_w^2}{d_w^{-\alpha}} + ||\boldsymbol{H}_w \boldsymbol{w}||^2)}$. Based on the above results, we have

$$\begin{aligned} D(\mathbb{P}_0 || \mathbb{P}_1) &= \int f(\boldsymbol{Y}|H_0) \left(n \ln \left(\frac{d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2 + \sigma_w^2}{\sigma_w^2} \right) - \frac{1}{\sigma_w^2} \frac{||(\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}} \boldsymbol{Y}||^2}{\frac{\sigma_w^2}{d_w^{-\alpha}} + ||\boldsymbol{H}_w \boldsymbol{w}||^2} \right) \mathrm{d}\boldsymbol{Y} \\ &= n \ln \left(\frac{d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2 + \sigma_w^2}{\sigma_w^2} \right) - \frac{\frac{d_w^{-\alpha}}{\sigma_w^2}}{d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2 + \sigma_w^2} \int f(\boldsymbol{Y}|H_0) \left| \left| (\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}} \boldsymbol{Y} \right| \right|^2 \mathrm{d}\boldsymbol{Y}. \end{aligned} \tag{A2}$$

Next, note that $\int f(\boldsymbol{Y}|H_0) ||(\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}} \boldsymbol{Y}||^2 \mathrm{d}\boldsymbol{Y}$ is indeed the expectation of $||(\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}} \boldsymbol{Y}||^2$ under H_0. Meanwhile, note that under H_0, we have $||(\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}} \boldsymbol{Y}||^2 = \sum_{i=1}^n |(\boldsymbol{H}_w \boldsymbol{w})^{\mathrm{H}} \boldsymbol{n}_w(i)|^2 = \frac{||\boldsymbol{H}_w \boldsymbol{w}||^2 \sigma_w^2}{2} \chi_{2n}^2$, where χ_{2n}^2 represents a chi-squared random variable with $2n$ degrees of freedom. Therefore, it follows from (A2) that

$$D(\mathbb{P}_0 || \mathbb{P}_1) = n \left(\ln \left(\frac{d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2 + \sigma_w^2}{\sigma_w^2} \right) - \frac{d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2}{d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2 + \sigma_w^2} \right). \tag{A3}$$

Expression (A3) is highly non-linear for further analysis. To address this, we note that it is monotonically increasing with respect to $d_w^{-\alpha} ||\boldsymbol{H}_w \boldsymbol{w}||^2$. Therefore, we have

$$\mathcal{P}_{\mathrm{KL}} \triangleq \Pr \left\{ D(\mathbb{P}_0 || \mathbb{P}_1) \leqslant 2 (1 - \xi_{\mathrm{expect}})^2 \right\} = \Pr \left\{ ||\boldsymbol{H}_w \boldsymbol{w}||^2 \leqslant \frac{\epsilon}{d_w^{-\alpha}} \right\}, \tag{A4}$$

where ϵ denotes the only solution of (6) with respect to x and can be efficiently obtained by the bisection method. To proceed, we expand $||\boldsymbol{H}_w \boldsymbol{w}||^2$ as

$$||\boldsymbol{H}_w \boldsymbol{w}||^2 = \sum_{k=1}^{N_w} \left| \boldsymbol{h}_{w,k}^{\mathrm{T}} \boldsymbol{w} \right|^2 = \sum_{k=1}^{N_w} \left| \sqrt{\frac{K_w}{K_w + 1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w} + \sqrt{\frac{1}{K_w + 1}} \hat{\boldsymbol{h}}_{w,k}^{\mathrm{T}} \boldsymbol{w} \right|^2. \tag{A5}$$

Note that $\boldsymbol{h}_{w,k}^{\mathrm{T}} \boldsymbol{w}$ follows $\mathcal{CN}\left(\sqrt{\frac{K_w}{K_w+1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w}, \frac{1}{K_w+1} ||\boldsymbol{w}||^2\right)$, with its real part R_k and imaginary part I_k following real Gaussian distributions $\mathcal{N}\left(\mathrm{real}\left(\sqrt{\frac{K_w}{K_w+1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w}\right), \frac{1}{2} \frac{1}{K_w+1} ||\boldsymbol{w}||^2\right)$ and $\mathcal{N}\left(\mathrm{imag}\left(\sqrt{\frac{K_w}{K_w+1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w}\right), \frac{1}{2} \frac{1}{K_w+1} ||\boldsymbol{w}||^2\right)$, respectively. Note that R_k and I_k are independent from each other and independent for different k. Next, we rewrite (A5) as

$$||\boldsymbol{H}_w \boldsymbol{w}||^2 = \sum_{k=1}^{N_w} \left(R_k^2 + I_k^2 \right) \triangleq \left(\frac{1}{2} \frac{1}{K_w + 1} ||\boldsymbol{w}||^2 \right) \sum_{k=1}^{N_w} \left(R_k'^2 + I_k'^2 \right), \tag{A6}$$

where $R_k' \sim \mathcal{N}\left(\frac{\mathrm{real}\left(\sqrt{\frac{K_w}{K_w+1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w}\right)}{\sqrt{\frac{1}{2} \frac{1}{K_w+1} ||\boldsymbol{w}||^2}}, 1\right)$ and $I_k' \sim \mathcal{N}\left(\frac{\mathrm{imag}\left(\sqrt{\frac{K_w}{K_w+1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w}\right)}{\sqrt{\frac{1}{2} \frac{1}{K_w+1} ||\boldsymbol{w}||^2}}, 1\right)$ are independent from each other and independent for different k. As a result, it follows from (A6) that

$$||\boldsymbol{H}_w \boldsymbol{w}||^2 = \left(\frac{1}{2} \frac{1}{K_w + 1} ||\boldsymbol{w}||^2 \right) \Omega, \tag{A7}$$

where Ω follows noncentral Chi-Squared distribution [15] with $2N_w$ degrees of freedom and noncentrality parameter

$$\lambda = \sum_{k=1}^{N_w} \left[\left(\frac{\mathrm{real}\left(\sqrt{\frac{K_w}{K_w+1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w}\right)}{\sqrt{\frac{1}{2} \frac{1}{K_w+1} ||\boldsymbol{w}||^2}} \right)^2 + \left(\frac{\mathrm{imag}\left(\sqrt{\frac{K_w}{K_w+1}} \mathrm{e}^{\iota \theta_{w,k}} \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w}\right)}{\sqrt{\frac{1}{2} \frac{1}{K_w+1} ||\boldsymbol{w}||^2}} \right)^2 \right] = \sum_{k=1}^{N_w} \frac{\frac{K_w}{K_w+1} \left| \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w} \right|^2}{\frac{1}{2} \frac{1}{K_w+1} ||\boldsymbol{w}||^2} = 2 K_w N_w \frac{\left| \bar{\boldsymbol{h}}_w^{\mathrm{T}} \boldsymbol{w} \right|^2}{||\boldsymbol{w}||^2}. \tag{A8}$$

The cumulative distribution function of Ω is $F_\Omega(x) = 1 - Q_{N_w}\left(\sqrt{2 K_w N_w \left| \bar{\boldsymbol{h}}_w^{\mathrm{T}} \frac{\boldsymbol{w}}{||\boldsymbol{w}||} \right|^2}, \sqrt{x}\right)$. As a result, it follows from (A4) and (A7) that

$$\mathcal{P}_{\mathrm{KL}} = \Pr \left\{ \Omega \leqslant \frac{2\epsilon (K_w + 1)}{d_w^{-\alpha} ||\boldsymbol{w}||^2} \right\} = 1 - Q_{N_w} \left(\sqrt{2 K_w N_w \left| \bar{\boldsymbol{h}}_w^{\mathrm{T}} \frac{\boldsymbol{w}}{||\boldsymbol{w}||} \right|^2}, \sqrt{\frac{2\epsilon (K_w + 1)}{d_w^{-\alpha} ||\boldsymbol{w}||^2}} \right), \tag{A9}$$

which completes the proof.