Is Gaussian Signalling Optimal for Covert Communications?

Shihao Yan[†], Yirui Cong[‡], Stephen Hanly[†], and Xiangyun Zhou[*]
[†]School of Engineering, Macquarie University, Sydney, NSW, Australia
[‡]College of Intelligence Science and Technology, National University of Defense Technology, Changsha, China
[*]Research School of Engineering, Australian National University, Canberra, ACT, Australia
Email: shihao.yan@mq.edu.au, congyirui11@nudt.edu.cn, stephen.hanly@mq.edu.au, xiangyun.zhou@anu.edu.au

***Abstract*—While Gaussian signalling is assumed in many studies on covert communications, its optimality has not been carefully investigated. In this work, we examine this optimality by considering the approach of upper bounding $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y}))$ as the covert communication constraint, where $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y}))$ is the Kullback-Leibler divergence from $p_0(\mathbf{y})$ to $p_1(\mathbf{y})$, $p_0(\mathbf{y})$ and $p_1(\mathbf{y})$ are the likelihood functions of the observation $\mathbf{y}$ at the warden under the null hypothesis (no covert transmission) and alternative hypothesis (a covert transmission occurs), respectively. Considering additive white Gaussian noise at both the receiver and the warden, we prove that Gaussian signalling is not optimal in terms of maximizing the mutual information of transmitted and received signals for covert communications with $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) \leq 2\epsilon^2$ as the constraint. We also explicitly show a skew-normal signalling can outperform Gaussian signalling in terms of achieving higher mutual information subject to the same covertness constraint $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) \leq 2\epsilon^2$.**

I. Introduction

People and organizations are becoming more dependent on cyber-physical systems to share private information (e.g., location, physiological information for e-health). Against this background, concerns about the security and privacy of wireless communications in many cyber-physical systems are believed to be the biggest barrier to the widespread adoption of these systems. To guarantee strong security and/or privacy (e.g., to avoid exposing location information), it is often not sufficient to only protect the content of wireless communications, but it is also required to hide the very existence of wireless transmissions. As the last line of defence in wireless communication security, hiding wireless transmissions is also explicitly desired by government and military bodies, e.g., it is desirable for a stealth fighter or an unmanned aerial vehicle to be able to hide itself from enemies while communicating with its military bases. However, hiding wireless transmissions cannot be achieved by existing cryptography or information-theoretic security technologies [1–3], since these technologies only protect the contents (i.e., what are transmitted) of wireless communications. Against this background, covert communication is emerging as a new technique to achieve a strong security and privacy in wireless communications (i.e., hiding wireless transmissions) [4–6].

Spread spectrum was invented a century ago with the original purpose of hiding military wireless transmissions (by spreading transmit power into noise) [7]. However, the covertness achieved has never been established theoretically, because there was no fundamental understanding on when or how often spread spectrum fails to hide wireless transmissions. As such, hiding wireless transmissions using spread spectrum cannot be guaranteed. Due to the unproven and unguaranteed performance, the main use of spread spectrum deviated from achieving covertness to obtaining high reliability and high data rate in the last two decades. Motivated by the ever-increasing desire of strong security and/or privacy, cutting-edge research on wireless communication security has called for a rethinking and generalization of spread spectrum (the only existing solution to hiding wireless transmissions) at a more fundamental level, which inspires the emergence of a new security paradigm termed covert communications. Research on covert communications focuses on the fundamental limit of hiding wireless transmissions, in terms of the amount of information that can be conveyed covertly wireless networks (e.g., [4, 8]). This enables covert communication techniques to achieve the proven and guaranteed security and/or privacy.

In covert communications, a transmitter (Alice) desires to transmit information to a legitimate receiver (Bob) without being detected by a warden (Willie), who is collecting observations to detect this transmission. Besides the fundamental limits of covert communications [4], some works focused on the design and performance analysis of covert communications in practical application scenarios, for example, by considering unknown background noise power [9], ignorance of transmission time [10], noise uncertainty [11], delay constraints [12], random interferers [13], uninformed jamming [14], relay networks [15, 16], and full-duplex technology [17, 18].

In covert communications, for an optimal detector at Willie, we have $\xi^* = 1 - \mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y}))$, where ξ^* is the minimum detection error probability and $\mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y}))$ is the total variation between the likelihood function $p_0(\mathbf{y})$ of the observation $\mathbf{y}$ under the null hypothesis (when Alice does not transmit to Bob) and the likelihood function $p_0(\mathbf{y})$ under the alternative hypothesis (when Alice transmits to Bob). Due to the mathematically intractable expressions for $\mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y}))$, Kullback-Leibler (KL) divergence (i.e., relative entropy) has been widely adopted to limit the detection performance at Willie in the literature of covert communications. Specifically, as per the Pinsker's inequality we have $\mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y})) \leq \sqrt{\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y}))/2}$, where $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y}))$ is the KL divergence from $p_0(\mathbf{y})$ to $p_1(\mathbf{y})$.

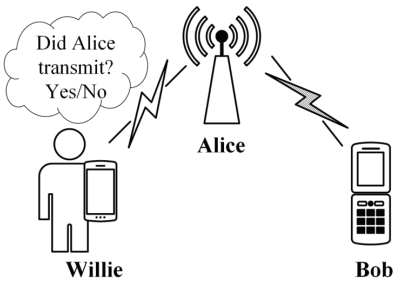

Fig. 1. Illustration of the system model for covert communications.

Then, the covert communication constraint $\xi^* \geq 1-\epsilon$ can be guaranteed by $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) \leq 2\epsilon^2$, where ϵ is a small value determining the required covertness.

In the literature, Gaussian signalling was widely adopted in covert communications, but its optimality has never been clarified in the context of covert communications, which mainly motivates this work. In this work, we tackle the optimality of Gaussian signalling with $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) \leq 2\epsilon^2$ as the covertness constraint. Specifically, we prove that Gaussian signalling is **not** optimal in terms of maximizing the mutual information between the transmitted signal vector $\mathbf{x}$ and the received signal vector $\mathbf{z}$ at the legitimate receiver, i.e., $I(\mathbf{x};\mathbf{z})$, subject to $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) \leq 2\epsilon^2$ for covert communications. We explicitly show that a skew-normal signalling strategy can achieve a higher $I(\mathbf{x};\mathbf{z})$ subject to $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) \leq 2\epsilon^2$ than Gaussian signalling.

II. System Model

A. Channel Model

The system model for covert communications is illustrated in Fig. 1, where each of Alice, Bob, and Willie is equipped with a single antenna. We assume the channel from Alice to Bob and the channel from Alice to Willie are only subject to AWGN. In this work, we assume that Alice transmits one real-valued symbol $\mathbf{x}[i]$ to Bob in the i-th channel use, while Willie is passively collecting one observation on Alice's transmission to detect whether or not Alice has transmitted the signal to Bob. We denote the AWGN at Bob and Willie in the i-th channel use as $\mathbf{n}_b[i]$ and $\mathbf{n}_w[i]$, respectively, where the elements of $\mathbf{n}_b$ or $\mathbf{n}_w$ are identically and independently distributed (i.i.d.). Thus, we have $\mathbf{n}_b[i] \sim \mathcal{N}(0,\sigma_b^2)$, $\mathbf{n}_w[i] \sim \mathcal{N}(0,\sigma_w^2)$, where σ_b^2 and σ_w^2 are the noise variances at Bob and Willie, respectively. In addition, we assume that $\mathbf{x}$, $\mathbf{n}_b$, and $\mathbf{n}_w$ are mutually independent and the number of channel uses (denoted by N) is sufficient large such that the elements of $\mathbf{x}$ are i.i.d.. We further assume that Alice's transmit power of $\mathbf{x}[i]$ is fixed and denoted as P_x, i.e., we have $\mathbb{E}\{|\mathbf{x}[i]|^2\} = P_x$.

B. Binary Hypothesis Testing at Willie

In order to detect the presence of covert communications, Willie must distinguish between the following two hypotheses:

$$\begin{cases} \mathcal{H}_0: \ \mathbf{y}[i] = \mathbf{n}_w[i], \\ \mathcal{H}_1: \ \mathbf{y}[i] = \mathbf{x}[i] + \mathbf{n}_w[i], \end{cases} \tag{1}$$

where $\mathcal{H}_0$ denotes the null hypothesis where Alice has not transmitted signals, $\mathcal{H}_1$ denotes the alternative hypothesis where Alice has transmitted, and $\mathbf{y}[i]$ is the received signal at Willie in the i-th channel use.

In general, the detection error probability is adopted to measure Willie's detection performance, which is defined as

$$\xi = \alpha + \beta, \tag{2}$$

where $\alpha \triangleq \Pr(\mathcal{D}_1|\mathcal{H}_0)$ is the false positive rate, $\beta \triangleq \Pr(\mathcal{D}_0|\mathcal{H}_1)$ is the miss detection rate, and $\mathcal{D}_1$ and $\mathcal{D}_0$ are the binary decisions that infer whether Alice's transmission is present or not, respectively. In covert communications, Willie's ultimate goal is to detect the presence of Alice's transmission with the minimum detection error probability ξ^*, which is achieved by using an optimal detector. Then, the actual covert communication constraint can be written as $\xi^* \geq 1-\epsilon$ for a given ϵ, where the value of ϵ is predetermined and is normally small in order to guarantee sufficient covertness.

For an optimal detector at Willie, we have [4, 19]

$$\xi^* = 1 - \mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y})) = 1 - \frac{1}{2}\|p_0(\mathbf{y}) - p_1(\mathbf{y})\|_1, \tag{3}$$

where $\mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y}))$ is the total variation between $p_0(\mathbf{y})$ and $p_0(\mathbf{y})$, $\|a-b\|_1$ is the $\mathcal{L}_1$ norm, and $p_0(\mathbf{y}) = f(\mathbf{y}|\mathcal{H}_0)$ and $p_1(\mathbf{y}) = f(\mathbf{y}|\mathcal{H}_1)$ are the likelihood functions of $\mathbf{y}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. In general, computing $\mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y}))$ analytically is intractable and thus Pinsker's inequality is normally adopted to upper bound it. Based on Pinsker's inequality, we have

$$\mathcal{V}_T(p_0(\mathbf{y}), p_1(\mathbf{y})) \leq \sqrt{\frac{1}{2}\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y}))}, \tag{4}$$

where $\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y}))$ is the KL divergence from $p_0(\mathbf{y})$ to $p_1(\mathbf{y})$, which is given by

$$\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) = \int_{\mathcal{Y}} p_0(\mathbf{y}) \log \frac{p_0(\mathbf{y})}{p_1(\mathbf{y})} d\mathbf{y}. \tag{5}$$

Following (3) and (4), it is also sufficient to guarantee

$$\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) \leq 2\epsilon^2, \tag{6}$$

in order to guarantee $\xi^* \geq 1-\epsilon$. Noting that the elements of $\mathbf{y}$ are i.i.d., we have

$$\mathcal{D}(p_0(\mathbf{y})||p_1(\mathbf{y})) = N \times \mathcal{D}(p_0(\mathbf{y}[i])||p_1(\mathbf{y}[i])), \tag{7}$$

where we recall that N is the total number of channel uses, which is sufficient large as assumed in this work.

C. Mutual Information

When Alice transmits $\mathbf{x}[i]$, the received signal at Bob in the i-th channel use is given by

$$\mathbf{z}[i] = \mathbf{x}[i] + \mathbf{n}_b[i]. \tag{8}$$

Then, the mutual information of $\mathbf{x}$ and $\mathbf{z}$ is given by

$$I(\mathbf{x};\mathbf{z}) = \int_{\mathcal{Z}}\int_{\mathcal{X}} p(\mathbf{x},\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z})}{p(\mathbf{x})p(\mathbf{z})} d\mathbf{x}d\mathbf{z}, \tag{9}$$

where $p(\mathbf{x},\mathbf{z})$ is the joint probability function of $\mathbf{x}$ and $\mathbf{z}$ and $p(\mathbf{z})$ is the marginal probability distribution of $\mathbf{z}$. For $\mathbf{n}_b[i] \sim \mathcal{N}(0,\sigma_b^2)$, $p(\mathbf{x}[i]) = \mathcal{N}(0,P)$ can maximize $I(\mathbf{x};\mathbf{z})$ subject to $\mathbb{E}[|\mathbf{x}[i]|^2] = P$ as per [19, Theorem 8.6.5]. This is the main reason why Gaussian signaling is widely adopted in the literature of covert communications (e.g., [4, 8]). Noting that the elements of $\mathbf{x}$ are i.i.d. and the elements of $\mathbf{z}$ are i.i.d., we have

$$I(\mathbf{x};\mathbf{z}) = N \times I(\mathbf{x}[i];\mathbf{z}[i]). \tag{10}$$

Considering (7) and (10), without loss of generality in this work we focus on one particular channel use, where we denote the particular element of $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{n}_b$, and $\mathbf{n}_w$ as x, y, z, n_b, and n_w, respectively, in order to seek the simplicity of presentations. As such, in the remainder of this work, we tackle whether the Gaussian signalling is optimal in terms of maximizing $I(x;z)$ subject to the covert communication constraint $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$.

III. Optimality of Gaussian Signalling

In this section, we analytically prove that Gaussian signaling is **not** optimal for the covert communication with $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ as the constraint. We present a skew-normal signalling strategy that can potentially outperform the Gaussian signalling in terms of achieving a higher $I(x;z)$ subject to $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$.

A. Gaussian Signalling is Not Optimal

In this subsection, we prove that Gaussian signaling is **not** optimal for covert communication with $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ as the constraint in the following theorem.

Theorem 1: Gaussian signaling, i.e., $p(x) = \mathcal{N}(m_x,\sigma_x^2)$, is not the solution to the following optimization problem

$$\begin{aligned}
\underset{p(x),P_x}{\operatorname{argmax}} \quad & I(x;z), && \text{(11a)}\\
\text{s.t.} \quad & \mathbb{E}[|x|^2] = P_x, && \text{(11b)}\\
& \int_{-\infty}^{\infty} p(x)dx = 1, && \text{(11c)}\\
& \mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2, && \text{(11d)}\\
& p(x) \geq 0, && \text{(11e)}
\end{aligned}$$

where m_x and σ_x^2 can take arbitrary values.

Proof: In order to prove Theorem 1, we next prove that Gaussian signalling is not in general the solution to the optimization problem given (11) in a special case, where Bob and Willie both experience the same level of AWGN. In this special case, we have n_w in (1) and n_b in (8) are i.i.d. and thus the pdf of z and the pdf of y under $\mathcal{H}_1$ are the same, i.e., we have $p(z) = p_1(y)$. As such, in the rest of this proof we use $p_1(y)$ to represent $p(z)$. Following (8) and noting that x is independent of n_b, we have

$$I(x;z) = h(z) - h(n_b), \tag{12}$$

where

$$h(z) = \int_{-\infty}^{\infty} p(z)\log\frac{1}{p(z)}dz = \int_{-\infty}^{\infty} p_1(y)\log\frac{1}{p_1(y)}dy \tag{13}$$

is the differential entropy of z and $h(n_b)$ is the differential entropy of n_b, which is not a function of $p(z)$ or $p_1(y)$. As such, in this special case to prove Theorem 1 we are going to prove that $p_1(y) = \mathcal{N}(0,P_y)$ is not the solution to the following optimization problem:

$$\begin{aligned}
\underset{p_1(y)}{\operatorname{argmax}} \quad & \int_{-\infty}^{\infty} p_1(y)\log\frac{1}{p_1(y)}dy, && \text{(14a)}\\
\text{s.t.} \quad & \int_{-\infty}^{\infty} p_1(y)dy = 1, && \text{(14b)}\\
& \int_{-\infty}^{\infty} y^2 p_1(y)dy = P_y, && \text{(14c)}\\
& \mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2, && \text{(14d)}\\
& p_1(y) \geq 0. && \text{(14e)}
\end{aligned}$$

In order to apply calculus of variations, following (14) we can write the functional as

$$\begin{aligned}
&\int_{-\infty}^{\infty} p_1(y)\log\frac{1}{p_1(y)}dy + \eta_0\left[\mathcal{D}(p_0(y)||p_1(y)) - 2\epsilon^2\right]\\
&+ \eta_1\left[\int_{-\infty}^{\infty} p_1(y)dy - 1\right] + \eta_2\left[\int_{-\infty}^{\infty} y^2 p_1(y)dy - P_y\right]\\
&= \int_{-\infty}^{\infty} \overline{\mathcal{L}}(y,p_1(y))dy - c,
\end{aligned} \tag{15}$$

where η_0, η_1, η_2, and η_3 are the Lagrange multipliers. Then, $\overline{\mathcal{L}}(y,p_1(y))$ in (15) is given by

$$\begin{aligned}
\overline{\mathcal{L}}(y,p_1(y)) = p_1(y)\log\frac{1}{p_1(y)} + \eta_0 p_0(y)\log\frac{p_0(y)}{p_1(y)}\\
+ \eta_1 p_1(y) + \eta_2 y^2 p_1(y),
\end{aligned} \tag{16}$$

and c is a constant determined by the Lagrange multipliers, $h(n_b)$, ϵ^2, and P_y, which is given by

$$c = h(n_b) + \eta_0 2\epsilon^2 + \eta_1 + \eta_2 P_y. \tag{17}$$

Following (16), the functional derivative (i.e., the first derivative of $\overline{\mathcal{L}}(y,p_1(y))$ with respect to $p_1(y)$) is given by

$$\frac{\partial\overline{\mathcal{L}}(y,p_1(y))}{\partial p_1(y)} = -\log p_1(y) - 1 - \eta_0\frac{p_0(y)}{p_1(y)} + \eta_1 + \eta_2 y^2. \tag{18}$$

Using the calculus of variations, a necessary condition for the optimal $p_1(y)$ in (14) is the existence of Lagrange multipliers such that the functional derivative given in (18) is zero [20]. As per [19, Theorem 8.6.5], $p_1(y) = \mathcal{N}(0,P_y)$ maximizes the mutual information between x and z subject to the constraints given in (14b), (14c), and (14e). As such, $p_1(y) = (N)(0,P_y)$ must satisfy

$$-\log p_1(y) - 1 + \eta_1^a + \eta_2^a y^2 = 0, \tag{19}$$

for two Lagrange multipliers η_1^a and η_2^a. If $p_1(y) = (N)(0, P_y)$ is the solution to the optimization problem given in (14), following (19) it must satisfy

$$-\eta_0 \frac{p_0(y)}{p_1(y)} + \eta_1^b + \eta_2^b y^2 = 0, \tag{20}$$

with $\eta_1^b = \eta_1 - \eta_1^a$ and $\eta_2^b = \eta_2 - \eta_2^a$. If (20) is satisfied, then $p_1(y)$ is given by

$$p_1(y) = \frac{\eta_0 p_0(y)}{\eta_1^b + \eta_2^b y^2}. \tag{21}$$

We note that in (21) the value of η_2^b cannot be zero. Otherwise, we will have $p_1(y) = \eta_0 p_0(y)/\eta_1^b$. In order to guarantee the pdf constraint (11c) with $p_1(y) = \eta_0 p_0(y)/\eta_1^b$, we would have $\eta_0/\eta_1^b = 1$, which cannot guarantee the power constraint (11b) simultaneously, since $P_y = P_x + \sigma_w^2 > \sigma_w^2$. As such, (21) with $\eta_2^b \neq 0$ indicates that the optimal signalling (if it exists) is not Gaussian, which completes the proof of Theorem 1. ■

B. Improving on Gaussian Signalling

In this subsection, we present the skew-normal distribution as a sometimes better $p(x)$ relative to the normal distribution (corresponding to the Gaussian signalling), where we consider the case in which the AWGN at Bob and Willie is i.i.d (i.e., r_w and n_b are i.i.d) such that the received signal at Willie y and the received signal at Bob z follow the same distribution.

If x follows a skew-normal distribution, the corresponding expression of $p(x)$ is given by [21]

$$p(x) = \frac{1}{\omega\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\omega^2}} \left[1 + \text{erf}\left(\frac{\theta(x-\mu)}{\omega\sqrt{2}}\right)\right], \tag{22}$$

where μ is the location parameter, ω is the scale parameter, θ is the skew parameter, and $\text{erf}(x)$ is the error function given by $\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$. We note that the normal distribution is recovered from (22) when $\theta = 0$ and the skewness increases as $|\theta|$ increases. In addition, the skew-normal distribution is right skewed relative to the normal distribution if $\theta > 0$ and is left skewed if $\theta < 0$. For the distribution given in (22), the mean and variance of x are, respectively, given by

$$\mathbb{E}\{x\} = \mu + \omega\delta\sqrt{\frac{2}{\pi}}, \tag{23}$$

$$\mathbb{E}\{|x - \mathbb{E}\{x\}|^2\} = \omega^2\left(1 - \frac{2\delta^2}{\pi}\right), \tag{24}$$

where $\delta = \theta/\sqrt{1+\theta^2}$. In this work, we focus on the skew-normal distribution with zero and P_x as the mean and variance, respectively. To this end, as per (23) and (24), for a given θ we have

$$\omega = \pm\sqrt{\frac{P_x}{1 - \frac{2\theta^2}{\pi(1+\theta^2)}}}, \tag{25}$$

$$\mu = -\omega\sqrt{\frac{2\theta^2}{\pi(1+\theta^2)}}. \tag{26}$$

We can vary the values of θ to obtain different skew-normal distributions with zero and P_x as the mean and variance, respectively, where the values of ω and μ are updated as per θ according to (25) and (26), respectively. This allows us to find a potential better $p(x)$ than the normal distribution in terms of achieving a higher $I(x;z)$ subject to the constraints given in (11b), (11c), (11d), and (11e), which will be confirmed in the numerical section.

In order to facilitate the calculation of the KL divergence from $p_0(y)$ to $p_1(y)$ and the mutual information between x and z, we derive the expression of $p_1(y)$ for the skew-normal $p(x)$ in the following proposition, which is also the expression of $p(z)$ for i.i.d. r_w and n_b.

***Proposition** 1:* For a skew-normal $p(x)$ with zero mean, variance P_x, and a non-zero skew parameter θ, following (1) the expression of $p_1(y)$ is derived as

$$\begin{aligned}
p_1(y) &= \frac{|\omega|}{\omega\sqrt{2\pi(\sigma_w^2+\omega^2)}} e^{-\frac{(y-\mu)^2}{2(\sigma_w^2+\omega^2)}} \\
&+ \frac{1}{\pi\sqrt{2\pi}\sigma_w^3\theta^2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1} e^{-\frac{(y-\mu)^2}{2\sigma_w^2}}}{(2k-1)(k-1)!} \left(\frac{(\sigma_w^2+\omega)^2}{\sigma_w^2\theta^2}\right)^{-\frac{1}{2}-k} \\
&\times \left[-\frac{\sigma_w\theta}{|\theta|}(\sigma_w^2+\omega^2)^{k+1}\Gamma(k)\,{}_1F_1\left(k, \frac{1}{2}, \frac{\omega^2(y-\mu)^2}{2\sigma_w^2(\sigma_w^2+\omega^2)}\right)\right. \\
&+ \frac{\sigma_w^{2k+3}\theta^3}{|\theta|^{-2k+1}} \left(\frac{\sigma_w^2+\omega^2}{\sigma^2\theta^2}\right)^{k+1} \Gamma(k)\,{}_1F_1\left(k, \frac{1}{2}, \frac{\omega^2(y-\mu)^2}{2\sigma_w^2(\sigma_w^2+\omega^2)}\right) \\
&+ 2\sqrt{2}\omega(\sigma_w^2+\omega^2)^{k+\frac{1}{2}}(y-\mu)\Gamma\left(k+\frac{1}{2}\right) \\
&\left.\times\,{}_1F_1\left(k+\frac{1}{2}, \frac{3}{2}, \frac{\omega^2(y-\mu)^2}{2\sigma_w^2(\sigma_w^2+\omega^2)}\right)\right],
\end{aligned} \tag{27}$$

where ${}_1F_1(a,b,z)$ is the Kummer confluent hypergeometric function.

Proof: Following (1), we have $y = x + n_w$ under $\mathcal{H}_1$ and noting $n_w \sim \mathcal{N}(0, \sigma_w^2)$ we have

$$p_1(y) = \frac{1}{\sqrt{2\pi}\sigma_w} \int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{2\sigma_w^2}} p(x) dx, \tag{28}$$

since x and n_w are independent. Then, substituting (22) into (28) we have

$$\begin{aligned}
p_1(y) &= \frac{1}{2\pi\sigma_w\omega} \int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{2\sigma_w^2} - \frac{(x-\mu)^2}{2\omega^2}} dx \\
&\quad + \frac{1}{2\pi\sigma_w\omega} \int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{2\sigma_w^2}} \text{erf}\left(\frac{\theta(x-\mu)}{\omega\sqrt{2}}\right) dx \\
&\overset{a}{=} \frac{|\omega|}{\omega\sqrt{2\pi(\sigma_w^2+\omega^2)}} e^{-\frac{(y-\mu)^2}{2(\sigma_w^2+\omega^2)}} \\
&+ \frac{\sqrt{2}}{\pi\sqrt{\pi}\sigma_w\theta} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{(2k-1)(k-1)!} \int_{-\infty}^{\infty} \chi^{2k-1} e^{-\frac{(y-\frac{\omega\sqrt{2}\chi}{\theta}-\mu)^2}{2\sigma_w^2}} d\chi,
\end{aligned} \tag{29}$$

where $\overset{a}{=}$ is achieved by setting $\chi = \frac{\theta(x-\mu)}{\omega\sqrt{2}}$ and with the aid of the following identity [22, Eq. (8.253.1)]

$$\text{erf}(\chi) = \frac{2}{\sqrt{\pi}} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\chi^{2k-1}}{(2k-1)(k-1)!}. \tag{30}$$

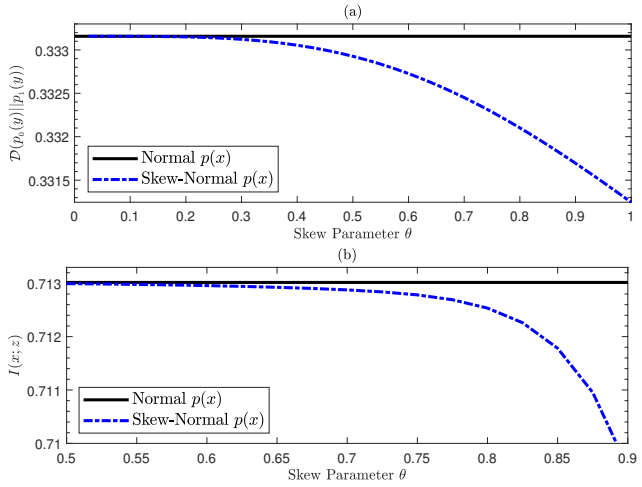

Fig. 2. The KL divergence $\mathcal{D}(p_0(y)||p_1(y))$ and mutual information $I(x;z)$ for skew-normal $p_1(y)$ with different values of the skew parameter θ, where $\sigma_b^2 = \sigma_w^2 = 0$dB and $P_x = 0$dB.

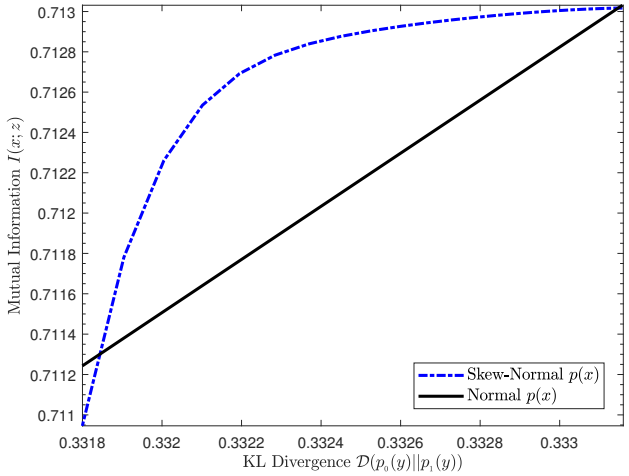

Fig. 3. The achieved mutual information $I(x;z)$ versus the associated KL divergence $\mathcal{D}(p_0(y)||p_1(y))$ for the skew-normal and normal $p(x)$.

Then, solving the resultant integrals in (29) leads to the desired result in (27), which completes the proof of Proposition 1. ■

Following Proposition 1, the KL divergence from $p_0(y)$ to $p_1(y)$ can be obtained by substituting (27) into (5). Since x and n_b are i.i.d, the mutual information between x and z can be written as

$$\begin{aligned} I(x;z) &= h(z) - h(n_b) \\ &= -\int_{-\infty}^{\infty} p(z)\log p(z)dz - \frac{1}{2}\log(2\pi e\sigma_b^2), \end{aligned} \tag{31}$$

where the expression for $p(z)$ is the same as that for $p_1(y)$ given in (27).

IV. Numerical Results

In this section, we first present the KL divergence $\mathcal{D}(p_0(y)||p_1(y))$ and mutual information $I(x;z)$ for a skew-normal signalling, which as shown can achieve a higher $I(x;z)$ subject to $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ than Gaussian signalling. This confirms that Gaussian signalling is not optimal for covert communications with $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ as the constraint. We then use $\xi^* \geq 1-\epsilon$ (i.e., $\mathcal{V}_T(p_0(y), p_1(y)) \leq \epsilon$) as the actual covert communication constraint and numerically show that a skew-normal $p(x)$ can achieve a higher mutual information $I(x;z)$ than the normal $p(x)$, which draws a more general conclusion that Gaussian signalling is not optimal for covert communications with $\xi^* \geq 1-\epsilon$ as the actual constraint.

In Fig. 2, we plot the KL divergence $\mathcal{D}(p_0(y)||p_1(y))$ and mutual information $I(x;z)$ for a skew-normal $p(x)$ with different skew parameters, where the mean and variance of x are fixed as 0 and P_x, respectively. From this figure, we observe that the skew-normal $p(x)$ can achieve a lower KL divergence $\mathcal{D}(p_0(y)||p_1(y))$ with some specific values of the skew parameter θ than the corresponding normal $p(x)$, although the former always achieves a lower mutual information $I(x;z)$ than the later. This provides the possibility that the skew-normal $p(x)$ achieves a higher $I(x;z)$ subject to $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ than the normal $p(x)$. To confirm this, we plot the achieved mutual information $I(x;z)$ versus the associated KL divergence $\mathcal{D}(p_0(y)||p_1(y))$ for skew-normal and normal $p(x)$ in Fig. 3. In order to plot Fig. 3, we fix $P_x = 0$dB for the skew-normal $p(x)$ and vary θ to generate different values of $I(x;z)$ and $\mathcal{D}(p_0(y)||p_1(y))$, while for the normal $p(x)$ we slightly vary P_x to obtain similar values of $I(x;z)$ and $\mathcal{D}(p_0(y)||p_1(y))$, since for the normal $p(x)$ there is a unique $I(x;z)$ and a unique $\mathcal{D}(p_0(y)||p_1(y))$ for each P_x. Noting that the equality in the constraint $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ for the normal $p(x)$ should be guaranteed, Fig. 3 confirms that the skew-normal $p(x)$ can achieve a higher $I(x;z)$ than the normal $p(x)$ subject to $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$. We note that in Fig. 3 the skew parameter θ is not optimized in terms of maximizing $I(x;z)$ subject to $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$. With optimized θ, the skew-normal $p(x)$ can achieve a higher $I(x;z)$ for a given $\mathcal{D}(p_0(y)||p_1(y))$.

Following a similar procedure of obtaining Fig. 3 but replacing the KL divergence $\mathcal{D}(p_0(y)||p_1(y))$ with the total variation $\mathcal{V}_T(p_0(y), p_1(y))$, we plot the achieved mutual information $I(x;z)$ versus $\mathcal{V}_T(p_0(y), p_1(y))$ in Fig. 4. From Fig. 4, we observe that the skew-normal $p_1(y)$ can achieve a higher $I(x;z)$ for some specific values of $\mathcal{V}_T(p_0(y), p_1(y))$ than the normal $p_1(y)$. Noting $\xi^* = 1 - \mathcal{V}_T(p_0(y), p_1(y))$, this observation indicates that Gaussian signalling is not optimal for covert communications with the actual constraint $\xi^* \geq 1-\epsilon$. As discussed in the Introduction, we note that the bounds determined by the KL divergences are still useful, since this total variation $\mathcal{V}_T(p_0(y), p_1(y))$ can only be numerically determined, while these bounds can provide mathematically tractable expressions on the performance of covert communications.

Considering Gaussian signalling, in Fig. 5 we plot the minimum detection error probability ξ^* and its lower bound determined by the KL divergence $\mathcal{D}(p_0(y)||p_1(y))$, versus the transmit power P_x for different AWGN power at Willie (i.e.,

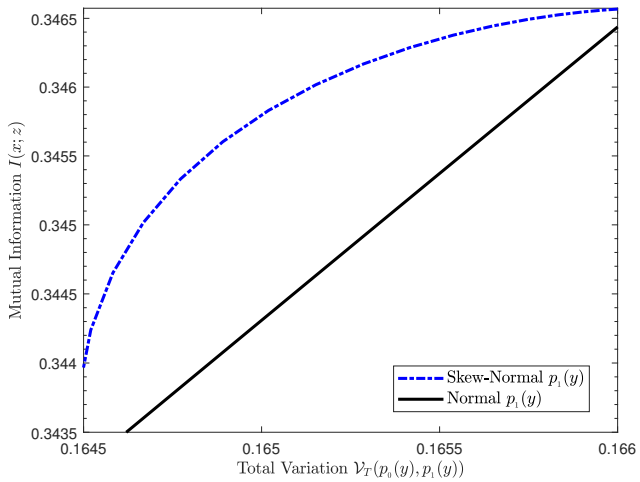

Fig. 4. The mutual information $I(x;z)$ versus the total variation $\mathcal{V}_T(p_0(y),p_1(y))$ for skew-normal and normal $p_1(y)$.

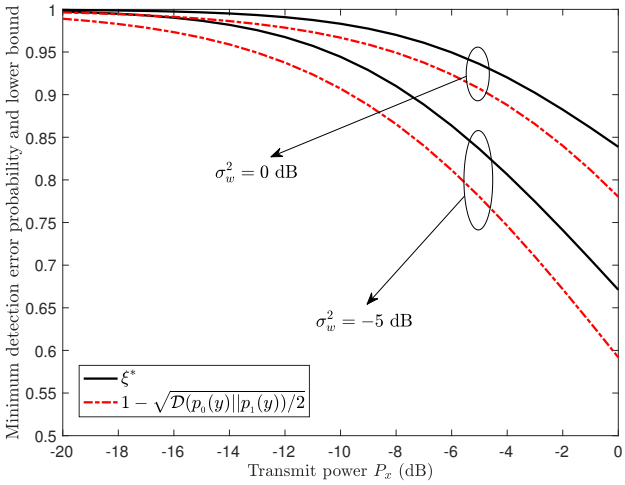

Fig. 5. The minimum detection error probability ξ^* and its two lower bounds versus the transmit power P_x for different values of σ_w^2.

σ_w^2). In this figure, we first observe that the lower bound is close to ξ^* when ξ^* is close to 1 for Gaussian signalling. We note that in the covert communication constraint the value of ϵ is usually very small, which enforces ξ^* being close to 1.

V. Conclusion

In this work, we proved that Gaussian signalling is not optimal for covert communications with $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ as the constraint, for which the optimal signalling will be tackled in our near future works. As we showed, a skew-normal $p(x)$ can achieve a higher $I(x;z)$ subject to $\mathcal{D}(p_0(y)||p_1(y)) \leq 2\epsilon^2$ than the normal $p(x)$. With $\xi^* \geq 1-\epsilon$ as the actual covert communication constraint, we numerically showed that Gaussian signalling is not optimal, since as we numerically showed a skew-normal signalling can achieve a better performance. It is noted that there are different approaches for defining the covert constraint. For example, one can use $\mathcal{D}(p_1(y)||p_0(y))$ instead of $\mathcal{D}(p_0(y)||p_1(y))$ in the covertness constraint, which may lead to different results, as discussed in [23]. Another important problem is the design and optimization of practical signalling and modulation for covert communications.

References

[1] J. Hu, Y. Cai, N. Yang, X. Zhou, and W. Yang, "Artificial-noise-aided secure transmission scheme with limited training and feedback overhead," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 193–205, Jan. 2017.

[2] F. Shu, X. Wu, J. Hu, J. Li, R. Chen, and J. Wang, "Secure and precise wireless transmission for random-subcarrier-selection-based directional modulation transmit antenna array," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 890–904, Apr. 2018.

[3] S. Yan, N. Yang, Ingmar Land, R. Malaney, and J. Yuan, "Three artificial-noise-aided secure transmission schemes in wiretap channels," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3669?3673, Apr. 2018.

[4] B. 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.

[5] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[6] S. Yan, B. He, Y. Cong, and X. Zhou, "Covert communication with finite blocklength in AWGN channels," in *Proc. IEEE ICC*, May 2017, pp. 1–6.

[7] M. K. Simon, Jim K. Omura, Robert A. Scholtz, and Barry K. Levitt, *Spread spectrum communications handbook*, McGraw-Hill, 1994.

[8] H. Wu, X. Liao, Y. Dang, Y. Shen, and X. Jiang, "Limits of covert communication on two-hop AWGN channels," in *Proc. International Conference on Networking and Network Applications*, Oct. 2017, pp. 42–47.

[9] D. Goeckel, B. Bash, S. Guha, and D. Towsley, "Covert communications when the warden does not know the background noise power," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 236–239, Feb. 2016.

[10] B. Bash, D. Goeckel, and D. Towsley, "Covert communication gains from adversarys ignorance of transmission time," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8394–8405, Dec. 2016.

[11] B. He, S. Yan, X. Zhou, and V. Lau, "On covert communication with noise uncertainty," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 941–944, Apr. 2017.

[12] 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.

[13] B. He, S. Yan, X. Zhou, and H. Jafarkhani, "Covert wireless communication with a poisson field of interferers," *IEEE Transactions on Wireless Commun.*, vol. 17, no. 9, pp. 6005–6017, Sep. 2018.

[14] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6193–6206, Sep. 2017.

[15] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J.Wang, "Covert communications in wireless relay networks," in *Proc. IEEE GlobeCOM*, Dec. 2017, pp. 1–6.

[16] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, "Covert communication achieved by a greedy relay in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4766–4779, Jul. 2018.

[17] J. Hu, K. Shahzad, S. Yan, X. Zhou, F. Shu, and J. Li, "Covert communications with a full-duplex receiver over wireless fading channels," in *Proc. IEEE ICC*, May 2018, pp. 1–6.

[18] K. Shahzad, X. Zhou, S. Yan, J. Hu, F. Shu, and J. Li, "Achieving covert wireless communications using a full-duplex receiver," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8517–8530, Dec. 2018.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Hoboken, NJ, 2002.

[20] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. New York: Dover, 2000.

[21] A. Azzalini, The Skew-Normal and Related Families. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[22] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 7th ed., Academic, San Diego, CA, 2007.

[23] S. Yan, Y. Cong, S. Hanly, and X. Zhou, "Gaussian signalling for covert communications," *arXiv:1807.00719*, pp. 1–11, Jul. 2018.