An Extensible Covert Communication Scheme Over the AWGN Channel With Feedback

Hangmei Rao* and Ligong Wang†
yqr@std.uestc.edu.cn, ligong.wang@ensea.fr

*National Key Laboratory of Science and Technology on Communications,
University of Electronic Science and Technology of China, Chengdu 611731, China
†ETIS Laboratory—CY Cergy Paris University, ENSEA, CNRS, 95000 Cergy-Pontoise, France

***Abstract*—We consider a problem of communication over a physically-degraded additive white Gaussian noise wiretap channel. A covertness constraint is imposed, which says that the output at the wiretapper must statistically resemble pure noise. Previous works have shown that the total amount of information that can be transmitted over n uses of the channel is proportional to $\sqrt{n}$, but the communication schemes in those works are designed to transmit only one message of predetermined length. We propose a feedback-aided scheme that, at the cost of less-than-optimal message length, is easily extensible to transmitting multiple messages and an unbounded amount of information.**

I. Introduction

Recently, much progress has been made on the theory of covert communication; see [1]–[12] and references therein. The present paper concerns covert communication over the additive white Gaussian noise (AWGN) channel. Specifically, we consider a physically-degraded memoryless wiretap channel [13] with input-output relations

$$Y = X + V \tag{1a}$$

$$Z = X + V + W, \tag{1b}$$

where all random variables take values in $\mathbb{R}$. Here, X denotes the channel input, Y the output at the intended receiver, Z the output at the wiretapper (sometimes called the "warden" in the covert communication literature), V is the additive noise at the intended receiver, and W the additive noise from the intended receiver to the wiretapper. Both V and W are zero-mean Gaussian random variables: V has variance 1 and W has variance $\sigma^2 - 1$, where $\sigma^2 > 1$; V and W are independent.

To send a message to the intended receiver, the transmitter sends a sequence of symbols $x^n = (x_1, \ldots, x_n)^\intercal$ over n uses of the above channel. The covertness constraint says that the output at the warden must be statistically close to pure noise: for some given constant $\delta > 0$, the distribution of Z^n, denoted by $\mathcal{P}_{Z^n}$, must satisfy

$$D\Big(\mathcal{P}_{Z^n}\Big\|\mathcal{N}\big(\mathbf{0}^n, \sigma^2 \mathsf{I}^n\big)\Big) \le \delta, \tag{2}$$

where $D(\cdot\|\cdot)$ denotes the Kullback-Leibler (KL) divergence between two distributions, and $\mathcal{N}\big(\mathbf{0}^n, \sigma^2\mathsf{I}^n\big)$ denotes n independent and identical Gaussian distributions, each of mean 0 and variance σ^2.

It is known that the maximum amount of information that can be covertly communicated over the above AWGN channel scales like the square-root of the total number of channel uses [1], [4]. More precisely, to a first-order approximation, for large n, the number of information nats that can be communicated over n uses of the channel (1) subject to the constraint (2) is

$$\sqrt{n\delta} \cdot \sigma^2. \tag{3}$$

To attain such a scaling behavior, the input power per channel use is chosen as a function of n [4], [7]:

$$P \approx \sqrt{\frac{\delta}{n}} \cdot 2\sigma^2. \tag{4}$$

The scheme described above is "one shot," as it is designed to be used to send one message of predetermined length, and not to be used repeatedly. Indeed, if the transmitter sent k messages of length (3) using input power (4), then the total KL divergence over all kn channel uses would become approximately $k\delta$, hence the communication activity would be more likely to be detected by the warden.

In order to maintain the possibility of sending additional messages at a later time without violating (2), the transmitter needs to reserve some of its "covertness budget." For example, it can choose to send a message of length (3) over $2n$ instead of n channel uses, using an input power equaling half of (4), and resulting in KL divergence $\frac{\delta}{2}$; it can send the next message of the same length over $4n$ channel uses, using one quarter of the power (4), resulting in an additional KL divergence of $\frac{\delta}{4}$; and so on. In this approach, every message requires a different input power and, consequently, a different codebook.[1]

In this work, we propose a different approach, where the power of the ith input symbol, $i \in \mathbb{Z}^+$, is a decreasing function of i alone. A proper choice of this decreasing function will guarantee that, under independent Gaussian signaling (at the chosen powers), (2) will be satisfied even when $n \to \infty$,

This research was conducted while visiting the Signal and Information Processing Laboratory, ETH Zurich, Switzerland. The work of H. Rao was supported in part by the National Natural Science Foundation of China under Grant No. U19B2014 and No. 62001094, China Postdoctoral Science Foundation under Grant No. 2020M683290 and No. 2021T140095, and Fundamental Research Funds for the Central Universities under Grant No. ZYGX2020J030.l.

[1] Additionally, the input power depends on the lengths of the messages. This is true also for transmitting a single message.

hence the encoder no longer needs to adjust the input powers depending on the number of messages and their lengths.

We then consider coding strategies using the chosen input powers and with the help of feedback.[2] Specifically, we assume that there is noiseless feedback from the intended receiver to the transmitter, and that the warden cannot observe this feedback. The presence of feedback allows us to adapt the ideas of Schalkwijk and Kailath [15] to design a simple coding scheme. A short secret key shared between the transmitter and the receiver is sufficient to guarantee covertness of the scheme, where the length of the key does not depend on the length of the message. Moreover, the coding scheme is easily extensible, in the sense that, to account for additional messages to be communicated, or to adjust the lengths of the messages, the encoder and the decoder only need slight modification.

As we shall see, the number of nats that can be transmitted using our scheme grows with n (total number of channel uses) more slowly than (3), i.e., the scheme is suboptimal for sending a fixed number of nats. This compromise is unavoidable; we shall discuss this in more detail later.

Our scheme relies heavily on the assumptions that the channel is physically degraded, and that feedback is noiseless and not observable by the warden. This may limit the usefulness of the scheme in its current form. We leave it as a potential direction for future works to overcome this limitation.

The rest of this paper is organized as follows. Section II presents our choice of input powers; Section III presents and analyzes the coding scheme for a single message; and Section IV discusses how to transmit a second message.

II. Choice of Input Powers

We choose a sequence of powers P_i, $i \in \mathbb{Z}^+$, and assume (within this section) that the random variables $X_1, X_2, \ldots$ are mutually independent, with X_i having the Gaussian distribution $\mathcal{N}(0, P_i)$. Our choice is guided by two considerations: (2) should be satisfied for all $n \in \mathbb{Z}^+$, and $I(X^n; Y^n)$ should tend to infinity as $n \to \infty$. We shall choose

$$P_i = \alpha \cdot i^{-\beta}, \quad i \geq 2, \tag{5}$$

where β can take any value in $(0.5, 1]$ and α will be determined later on as a function of β and P_1. The choice of the powers is obviously not unique, and we do not claim (5) to be optimal in any sense. To verify (2) and to find an appropriate value for α, we use the following bound (note that Z_i has the distribution $\mathcal{N}\big(0, P_i + \sigma^2\big)$):

$$D\Big(\mathcal{P}_{Z^n} \Big\| \mathcal{N}\big(\mathbf{0}^n, \sigma^2 \mathsf{I}^n\big)\Big) \leq \sum_{i=1}^{\infty} D\big(\mathcal{N}\big(0, P_i + \sigma^2\big) \big\| \mathcal{N}\big(0, \sigma^2\big)\big) \tag{6}$$

$$= \sum_{i=1}^{\infty} \frac{1}{2}\left(\frac{P_i}{\sigma^2} - \log\left(1 + \frac{P_i}{\sigma^2}\right)\right) \tag{7}$$

$$\leq \frac{P_1^2}{4\sigma^4} + \frac{1}{4}\sum_{i=2}^{\infty}\Big(\frac{\alpha}{i^\beta \sigma^2}\Big)^2 \tag{8}$$

$$\leq \frac{P_1^2}{4\sigma^4} + \frac{1}{4}\int_1^{\infty}\left(\frac{\alpha}{\zeta^\beta \sigma^2}\right)^2 \mathrm{d}\zeta \tag{9}$$

$$= \frac{P_1^2}{4\sigma^4} + \frac{\alpha^2}{4\sigma^4(2\beta - 1)}. \tag{10}$$

Therefore, (2) is satisfied even as $n \to \infty$ if we choose $P_1 < 2\sigma^2\sqrt{\delta}$ and $\alpha = \sqrt{(2\beta - 1)(4\sigma^4\delta - P_1^2)}$.

One can calculate $I(X^n; Y^n)$ for the above choice of P_is: for large n,

$$I(X^n; Y^n) \approx \begin{cases} \dfrac{\alpha}{2(1-\beta)} \cdot n^{1-\beta}, & \beta < 1 \\ \dfrac{\alpha}{2}\log n, & \beta = 1. \end{cases} \tag{11}$$

This gives us a first idea how much information can be communicated over n channel uses, as we explain in more detail below.

Remark 1: If we fix the input powers as in (5) for a specific β, i.e., if we impose expected-power constraints $\mathsf{E}\big[X_i^2\big] \leq P_i$, $i \in \mathbb{Z}^+$ (we can now ignore the covertness constraint (2)), then the number of nats that can be reliably communicated over n channel uses cannot asymptotically exceed the right-hand side of (11). This is true both with and without feedback, as can be proven along the lines of the converse proofs in [4]. We expect that one can prove a direct coding theorem without feedback, which says that (11) is asymptotically achievable using input powers (5), again following the approach in [4]; the proof might however be cumbersome due to the nonconstant power. A direct result with feedback will be proven in the next section.

Remark 2: Recall that we must choose $\beta > 0.5$, which implies that the right-hand side of (11) grows more slowly than $\sqrt{n}$ for all permissible choices of β. Thus, using the input powers (5) cannot achieve the maximum possible message length given by (3). This suboptimality is unavoidable if one wishes to be able to send an unlimited number of messages. Indeed, at any time n, if the transmitter wishes to maintain the possibility of sending more messages in the future, then (2) must hold with strict inequality, resulting in less-than-maximum total information nats up to time n.

III. Communication Scheme With Feedback

We describe a covert communication scheme aided by noiseless feedback from the intended receiver to the transmitter as well as a short secret key, neither of which is available to the warden. The time-i input X_i is a function of the message M, all past outputs Y^{i-1}, and the key S, which is uniformly distributed over a set $\mathcal{S} = \{1, \ldots, |\mathcal{S}|\}$. The scheme is a variation of Schalkwijk and Kailath [15] (see also [16]). Let $P_2, P_3, \ldots$ be chosen as in (5). For brevity, we shall only consider $\beta \in (0.5, 1)$ (i.e., $\beta \neq 1$).

Encoding. To send a message $m \in \mathcal{M}$, the transmitter first sends

$$X_1 = \theta_m \triangleq \frac{2m - |\mathcal{M}| - 1}{2|\mathcal{M}|} \cdot \lambda, \tag{12}$$

[2] In a rather different setting, using feedback in covert secret key generation was studied in [14].

where λ is a constant that we shall specify later. After receiving the feedback Y_1, it recovers $V_1 = Y_1 - X_1$. The remaining transmissions will be devoted to conveying V_1 to the receiver with high accuracy. However, for covertness considerations, the transmitter first uses the secret key S to map V_1 to a new random variable:

$$U_2 = \Phi^{-1}\left(\Phi(V_1) + \frac{S}{|\mathcal{S}|} \bmod 1\right), \tag{13}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard Gaussian, and where the operation "a mod 1" removes the maximum integer from a (in the above case, it outputs a if $a < 1$, and outputs $a-1$ if $a \geq 1$). For the second channel use, the transmitter sends

$$X_2 = \sqrt{P_2}\, U_2. \tag{14}$$

At time $i \geq 3$, it sends

$$X_i = \sqrt{\frac{P_i}{\rho_i}}\, U_i, \tag{15}$$

where

$$\rho_i = \left(\prod_{j=2}^{i-1}(1+P_j)\right)^{-1} \tag{16}$$

$$\begin{aligned} U_i &= U_{i-1} - \frac{\sqrt{\rho_{i-1}P_{i-1}}\,Y_{i-1}}{1+P_{i-1}} && (17)\\ &= U_2 - \sum_{j=2}^{i-1}\frac{\sqrt{\rho_j P_j}\,Y_j}{1+P_j} && (18)\\ &= \frac{1}{\prod_{j=2}^{i-1}(1+P_j)}U_2 - \sum_{k=2}^{i-1}\frac{\sqrt{\rho_k P_k}}{\prod_{j=k}^{i-1}(1+P_j)}V_k. && (19)\end{aligned}$$

The purpose of the operation (13) is to use the secret key to "randomize" V_1 to obtain a new standard Gaussian random variable U_2. To see that U_2 is indeed standard Gaussian, note that $\Phi(V_1)$ is uniformly distributed on $(0,1)$, therefore so is $\left(\Phi(V_1) + \frac{S}{|\mathcal{S}|} \bmod 1\right)$. For the "randomize" part, we have the following lemma, which will be used in the covertness analysis.

Lemma 3: Let V_1 and U_2 be as above, and let N be independent of V_1 and have the distribution $\mathcal{N}\big(0,\tau^2\big)$ for some $\tau^2 > 0$. Then, for all $|\mathcal{S}| > 7$,

$$I(U_2; V_1 + N) \leq \frac{1}{|\mathcal{S}|}\left(\frac{4\log|\mathcal{S}| - 2\log(2\pi)}{\tau^2} + 4\right). \tag{20}$$

Proof: Omitted. ■

Decoding. After n channel uses, the decoder computes

$$\hat{U}_2 = \sum_{i=2}^{n}\frac{\sqrt{\rho_i P_i}\,Y_i}{1+P_i}. \tag{21}$$

Then, using the secret key S, it computes

$$\hat{V}_1 = \Phi^{-1}\left(\Phi(\hat{U}_2) - \frac{S}{|\mathcal{S}|} \bmod 1\right), \tag{22}$$

and

$$\hat{\Theta} = Y_1 - \hat{V}_1. \tag{23}$$

It then outputs the message $\hat{m}$ that minimizes $|\hat{\Theta} - \theta_{\hat{m}}|$.

Analysis of covertness. The warden outputs are given by

$$\begin{aligned} Z_1 &= \theta_M + V_1 + W_1 && (24)\\ Z_i &= \sqrt{\frac{P_i}{\rho_i}}\left(\frac{1}{\prod_{j=2}^{i-1}(1+P_j)}U_2 - \sum_{k=2}^{i-1}\frac{\sqrt{\rho_k P_k}}{\prod_{j=k}^{i-1}(1+P_j)}V_k\right)\\ &\quad + V_i + W_i, \qquad i \geq 2. && (25)\end{aligned}$$

Note that $U_2, V_2, V_3, \ldots$ are IID standard Gaussian random variables. It follows that $Z_2, Z_3, \ldots$ are jointly Gaussian. Calculation shows that Z_i, $i \geq 2$, has the distribution $\mathcal{N}\big(0, P_i + \sigma^2\big)$. Furthermore,

$$\mathsf{E}[Z_i Z_j] = 0, \quad i,j = 2,3,\ldots,\ i \neq j. \tag{26}$$

Since $Z_2, Z_3, \ldots$ are jointly Gaussian, (26) implies they are mutually independent. We then have

$$\begin{aligned} &D\Big(\mathcal{P}_{Z_2^n}\Big\|\mathcal{N}\big(\mathbf{0}^{n-1}, \sigma^2\mathsf{I}^{n-1}\big)\Big)\\ &\quad = \sum_{i=2}^{n} D\big(\mathcal{N}\big(0, P_i + \sigma^2\big)\big\|\mathcal{N}\big(0,\sigma^2\big)\big) && (27)\\ &\quad \leq \frac{\alpha^2}{4\sigma^4(2\beta - 1)}, && (28)\end{aligned}$$

where the last step follows from our calculations in (10). By the chain rule of the KL divergence,

$$\begin{aligned} D\Big(\mathcal{P}_{Z^n}\Big\|\mathcal{N}\big(\mathbf{0}^n, \sigma^2\mathsf{I}^n\big)\Big) &= D\Big(\mathcal{P}_{Z_2^n}\Big\|\mathcal{N}\big(\mathbf{0}^{n-1}, \sigma^2\mathsf{I}^{n-1}\big)\Big)\\ &\quad + D\Big(\mathcal{P}_{Z_1|Z_2^n}\Big\|\mathcal{N}\big(0,\sigma^2\big)\Big). && (29)\end{aligned}$$

Note that $Z_1 \;-\!\!\circ\!\!-\; U_2 \;-\!\!\circ\!\!-\; Z_2^n$ forms a Markov chain, therefore

$$\begin{aligned} D\Big(\mathcal{P}_{Z_1|Z_2^n}\Big\|\mathcal{N}\big(0,\sigma^2\big)\Big) &\leq D\Big(P_{Z_1|U_2}\Big\|\mathcal{N}\big(0,\sigma^2\big)\Big) && (30)\\ &= I(U_2;Z_1) + D\big(P_{Z_1}\big\|\mathcal{N}\big(0,\sigma^2\big)\big). && (31)\end{aligned}$$

We can now apply Lemma 3 to obtain, for $|\mathcal{S}| > 7$,

$$\begin{aligned} I(U_2;Z_1) &= I(U_2;\theta_M + V_1 + W_1) && (32)\\ &\leq I(U_2; V_1 + W_1) && (33)\\ &\leq \frac{1}{|\mathcal{S}|}\left(\frac{4\log|\mathcal{S}| - 2\log(2\pi)}{\sigma^2 - 1} + 4\right). && (34)\end{aligned}$$

For the second term on the right-hand side of (31) we have

$$\begin{aligned} D\big(P_{Z_1}\big\|\mathcal{N}\big(0,\sigma^2\big)\big) &\leq \mathsf{E}\big[D\big(\mathcal{N}\big(\theta_M,\sigma^2\big)\big\|\mathcal{N}\big(0,\sigma^2\big)\big)\big] && (35)\\ &\leq \frac{\lambda^2}{24\sigma^2}. && (36)\end{aligned}$$

Combining (28), (31), (34), and (36), we obtain

$$\begin{aligned} D\Big(\mathcal{P}_{Z_1^n}\Big\|\mathcal{N}\big(\mathbf{0}^n, \sigma^2\mathsf{I}^n\big)\Big) &\leq \frac{1}{|\mathcal{S}|}\left(\frac{4\log|\mathcal{S}| - 2\log(2\pi)}{\sigma^2 - 1} + 4\right)\\ &\quad + \frac{\lambda^2}{24\sigma^2} + \frac{\alpha^2}{4\sigma^4(2\beta-1)}. && (37)\end{aligned}$$

By choosing $|\mathcal{S}|$ to be large (the choice does not depend on $|\mathcal{M}|$), we can make the first term on the right-hand side of (37) close to zero. As we shall see later on, the value of λ does not affect the decoding error probability in the limit where $n \to \infty$, thus we can choose λ to be close to zero, thereby making the second term on the right-hand side of (37) also small. Therefore, there exist choices of $|\mathcal{S}|$ and λ to satisfy (2) as long as

$$\alpha < \sqrt{(2\beta - 1)\delta} \cdot 2\sigma^2. \tag{38}$$

Analysis of error probability. By (12), the values $\theta_1, \ldots, \theta_{|\mathcal{M}|}$ have a minimum distance of $\frac{\lambda}{|\mathcal{M}|}$, so the decoder makes an error only if $|\hat{\Theta} - \theta_m| \geq \frac{\lambda}{2|\mathcal{M}|}$, which is equivalent to

$$|V_1 - \hat{V}_1| \geq \frac{\lambda}{2|\mathcal{M}|}. \tag{39}$$

To bound the probability of (39), we consider three (not mutually exclusive) cases. The first case is where $|V_1|$ is large. Specifically, for some small positive ω, consider the event[3]

$$\mathcal{E}_1\colon\ \Phi(V_1) \in [0, \omega] \cup [1 - \omega, 1]. \tag{40}$$

Since $\Phi(V_1)$ is a uniform random variable over $[0, 1]$ and is independent of S, we have

$$\Pr(\mathcal{E}_1) = 2\omega. \tag{41}$$

The second case is

$$\mathcal{E}_2\colon\ |\Phi(U_2) - \Phi(\hat{U}_2)| \geq \frac{\omega}{2}. \tag{42}$$

We have the following bound:

$$\begin{align}
|\Phi(U_2) - \Phi(\hat{U}_2)| &\leq |U_2 - \hat{U}_2| \cdot \max_{u \in \mathbb{R}} \phi(u) \tag{43}\\
&= \frac{|U_2 - \hat{U}_2|}{\sqrt{2\pi}}, \tag{44}
\end{align}$$

where $\phi(\cdot)$ denotes the probability density function of the standard Gaussian distribution. It then follows that

$$\Pr(\mathcal{E}_2) \leq \Pr\left(|U_2 - \hat{U}_2| \geq \sqrt{\frac{\pi}{2}}\omega\right). \tag{45}$$

The third case $\mathcal{E}_3$ is where $\mathcal{E}_1$ and $\mathcal{E}_2$ are both false and (39) is true. When $\mathcal{E}_1$ and $\mathcal{E}_2$ are both false, we have $\Phi(V_1) - \Phi(\hat{V}_1) = \Phi(U_2) - \Phi(\hat{U}_2)$, so

$$\begin{align}
|V_1 - \hat{V}_1| &\leq \frac{|\Phi(V_1) - \Phi(\hat{V}_1)|}{\min_{u\colon \Phi(u) \in [\omega/2, 1-\omega/2]} \phi(u)} \tag{46}\\
&= |\Phi(V_1) - \Phi(\hat{V}_1)| \cdot \sqrt{2\pi} \cdot e^{\frac{a^2}{2}} \tag{47}\\
&= |\Phi(U_2) - \Phi(\hat{U}_2)| \cdot \sqrt{2\pi} \cdot e^{\frac{a^2}{2}} \tag{48}\\
&\leq |U_2 - \hat{U}_2| \cdot e^{\frac{a^2}{2}}, \tag{49}
\end{align}$$

where $a \triangleq Q^{-1}(\frac{\omega}{2})$, with $Q(\cdot)$ denoting the Q-function associated with the standard Gaussian distribution. We thus have

$$\Pr(\mathcal{E}_3) \leq \Pr\left(|U_2 - \hat{U}_2| \geq \frac{\lambda e^{-\frac{a^2}{2}}}{2|\mathcal{M}|}\right). \tag{50}$$

We can write $U_2 - \hat{U}_2$ as

$$U_2 - \hat{U}_2 = \frac{1}{\prod_{j=2}^n (1 + P_j)} U_2 - \sum_{i=2}^n \frac{\sqrt{\rho_i P_i}}{\prod_{j=i}^n (1 + P_j)} V_i. \tag{51}$$

It has a zero-mean Gaussian distribution, with

$$\operatorname{Var}\left(U_2 - \hat{U}_2\right) = \frac{1}{\prod_{i=2}^n (1 + P_i)}. \tag{52}$$

We thus have

$$\Pr(\mathcal{E}_3) \leq 2\, Q\left(\frac{\sqrt{\prod_{i=2}^n (1 + P_i)}}{2|\mathcal{M}|} \cdot \lambda e^{-\frac{a^2}{2}}\right). \tag{53}$$

Hence, for any $\omega > 0$ (recall that a is a function of ω), $\Pr(\mathcal{E}_3) \to 0$ as $n \to \infty$ provided we choose $|\mathcal{M}|$ to be any function of n that satisfies

$$\lim_{n \to \infty} \frac{\sqrt{\prod_{i=2}^n (1 + P_i)}}{|\mathcal{M}|} = \infty. \tag{54}$$

By such a choice, $\Pr(\mathcal{E}_2)$ will also approach zero as $n \to \infty$; recall (45). Since ω can be chosen to be arbitrarily close to zero, which in turn makes $\Pr(\mathcal{E}_1)$ close to zero, we conclude that the overall error probability can be made arbitrarily small provided that $|\mathcal{M}|$ satisfies (54). Consider the logarithm of the numerator in (54), with the choice (5) for some $\beta \in (0.5, 1)$:

$$\begin{align}
\log \sqrt{\prod_{i=2}^n (1 + P_i)} &= \frac{1}{2} \sum_{i=2}^n \log\left(1 + \frac{\alpha}{i^\beta}\right) \tag{55}\\
&\geq \frac{1}{2} \int_2^n \log\left(1 + \frac{\alpha}{\zeta^\beta}\right) d\zeta \tag{56}\\
&\geq \frac{1}{2} \int_2^n \left(\frac{\alpha}{\zeta^\beta} - \frac{\alpha^2}{2\zeta^{2\beta}}\right) d\zeta \tag{57}\\
&= \frac{1}{2}\left(\frac{\alpha}{1 - \beta}(n^{1-\beta} - 2^{1-\beta})\right.\nonumber\\
&\qquad \left. - \frac{\alpha^2}{2 - 4\beta}(n^{1-2\beta} - 2^{1-2\beta})\right). \tag{58}
\end{align}$$

If we choose, for any $\epsilon > 0$,

$$\log |\mathcal{M}| = (1 - \epsilon) \cdot \frac{\alpha}{2(1 - \beta)} n^{1-\beta}, \tag{59}$$

then (54) can be satisfied, and the error probability can indeed be made arbitrarily small.

Summarizing the above analyses and recalling our choice of α in (38), we have the following result.

Theorem 4: Fix $\beta \in (0.5, 1)$. For any $\epsilon > 0$, there exists a sequence of feedback- and secret-key-aided communication schemes, all of which satisfy the covertness condition (2), whose error probability approaches zero as $n \to \infty$, and which over n channel uses can transmit a message drawn from a set $\mathcal{M}(n)$ with

$$\log |\mathcal{M}(n)| \geq (1 - \epsilon)\sigma^2 \cdot \sqrt{\delta} \cdot \frac{\sqrt{2\beta - 1}}{1 - \beta} \cdot n^{1-\beta}. \tag{60}$$

The part in (60) that depends on β is $\frac{\sqrt{2\beta-1}}{1-\beta} \cdot n^{1-\beta}$. We plot it for different values of β in Fig. 1. As can be seen, choosing

[3] This event does not necessarily lead to a decoding error, but we shall treat it as an error event to obtain an upper bound on the error probability.

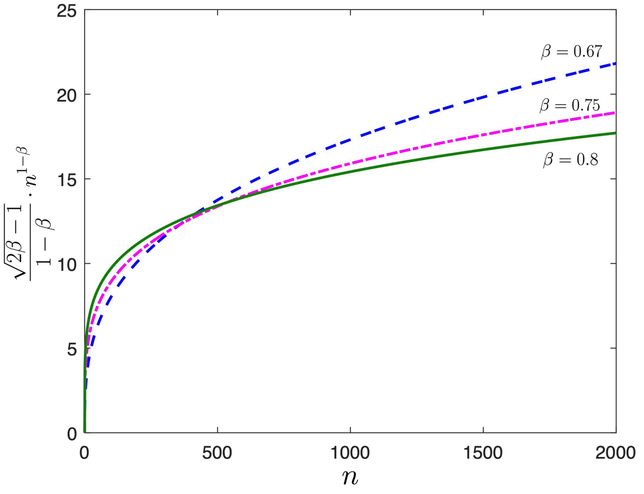

Fig. 1. Maximum possible throughput divided by $\sqrt{\delta}\cdot\sigma^2$ as a function of n, the number of channel uses, for different values of β.

β closer to 0.5 yields better asymptotic growth, but smaller throughput for small n due to the multiplicative factor.

IV. Adding a Second Message

Suppose now that the transmitter must send two messages to the intended receiver, but that it cannot combine the two messages into one long message. (The latter may be because that, at the time when transmission of the first message started, the transmitter was not aware of the content or even the existence of the second message.) We discuss two approaches, both of which can be extended to sending multiple (i.e., more than two) messages. For brevity, we shall skip some mathematical details to focus on presenting the main ideas.

Let the first message be M drawn from the set $\mathcal{M}$ as in the previous section, and let the second message be M' drawn from the set $\mathcal{M}'$.

A. Second message after completion of the first message

In this approach, the transmitter sends the second message using the same procedure as for sending the first message, except with different input powers. Suppose that transmission of the first message is completed at time ℓ. To send the second message, the transmitter first sends

$$X_{\ell+1} = \frac{2m' - |\mathcal{M}'| - 1}{2|\mathcal{M}'|}\cdot\lambda', \tag{61}$$

where λ' is an appropriately chosen function of ℓ. The encoder then computes $V_{\ell+1} = Y_{\ell+1} - X_{\ell+1}$ using the feedback, uses a secret key (which it shares with the receiver) to randomize $V_{\ell+1}$ in a similar way to (13) to obtain $U_{\ell+2}$, and sends $X_{\ell+2} = \sqrt{P_{\ell+2}}\,U_{\ell+2}$. The scheme proceeds in the same way as when sending the first message, using input powers $P_{\ell+2}, P_{\ell+3}, \ldots$. One can show that, with an appropriately chosen λ', the total KL divergence (over the transmission of both messages) is smaller than δ. Furthermore, the maximum possible value for $\log|\mathcal{M}| + \log|\mathcal{M}'|$ is again given by the right-hand side of (60), where n is now the total number of channel uses for both messages.

B. Superimposing the second message onto the first message

Suppose that, at time j during the transmission of the first message M (using the procedure described in Section III), the transmitter learns that it must send another message M' to the receiver. Instead of waiting for the transmission of M to complete, it can immediately superimpose M' onto M. To do so, it computes

$$\tilde{\theta}_{m,m'} = \theta_m + \frac{2m' - |\mathcal{M}'| - 1}{2|\mathcal{M}||\mathcal{M}'|}\cdot\lambda. \tag{62}$$

It then continues transmission as if it had been sending $\tilde{\theta}_{m,m'}$ instead of θ_m from the beginning. To this end, it computes

$$\tilde{V}_1 = Y_1 - \tilde{\theta}_{m,m'} \tag{63}$$

$$\tilde{U}_2 = \Phi^{-1}\left(\Phi(\tilde{V}_1) + \frac{S}{|\mathcal{S}|} \bmod 1\right), \tag{64}$$

where S is the same secret key as used in (13). To start superimposing m' onto m at time j, the transmitter does the following. Instead of computing (17) with $i = j$, it computes

$$\tilde{U}_j = (\tilde{U}_2 - U_2) + U_{j-1} - \frac{\sqrt{\rho_{j-1}P_{j-1}}\,Y_{j-1}}{1 + P_{j-1}}, \tag{65}$$

and sends $\tilde{X}_j = \sqrt{\frac{P_j}{\rho_j}}\,\tilde{U}_j$. For all $i \geq j+1$, it computes $\tilde{U}_i$ according to (17) with U_{i-1} replaced by $\tilde{U}_{i-1}$, and sends $\tilde{X}_i = \sqrt{\frac{P_i}{\rho_i}}\,\tilde{U}_i$. Note that, for all $i \geq j$, (18) now becomes

$$\tilde{U}_i = \tilde{U}_2 - \sum_{j=2}^{i-1} \frac{\sqrt{\rho_j P_j}\,Y_j}{1 + P_j}. \tag{66}$$

The decoder again computes $\hat{\Theta}$ using (21)–(23). It then outputs the pair $(\hat{m}, \hat{m}')$ that minimizes $|\hat{\Theta} - \tilde{\theta}_{\hat{m},\hat{m}'}|$.

The decoding error probability can be bounded in a similar way as in Section III, and is relatively straightforward. Rigorous analysis of covertness is more involved; the main difference from the one-message case comes from the jth channel use:

$$\tilde{U}_j = (\tilde{U}_2 - U_2) + U_j, \tag{67}$$

where the additional term $(\tilde{U}_2 - U_2)$ adds to the total KL divergence. The difference is significant if the decoder's estimation error of U_2 at time j—which in fact equals U_j—is small, i.e., if the decoder is "almost ready" to declare the outcome; it is negligible if U_j is large. Hence the superimposing approach is feasible if transmission of the first message is "far from finished." We shall elaborate on this fact in a full-length paper, which is under preparation.

References

[1] B. A. Bash, D. Goekel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 1921–1930, Sept. 2013.

[2] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. Inform. Theory*, (Istanbul, Turkey), July 10–15 2013.

[3] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nature Communications*, vol. 6, no. 8620, 2015.

[4] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inform. Theory*, vol. 62, pp. 3493–3503, June 2016.

[5] M. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inform. Theory*, vol. 62, pp. 2334–2354, May 2016.

[6] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE Trans. Inform. Forensics and Security*, vol. 13, no. 9, pp. 2310–2319, 2018.

[7] L. Wang, "On Gaussian covert communication in continuous time," *EURASIP J. App. Sig. Proc.*, Dec. 2019.

[8] K. Arumugam and M. Bloch, "Covert communication over a k-user multiple access channel," *IEEE Trans. Inform. Theory*, vol. 66, pp. 7020–7044, Nov. 2019.

[9] Q. Zhang, M. Bakshi, and S. Jaggi, "Covert communication with polynomial computational complexity," *IEEE Trans. Inform. Theory*, vol. 66, pp. 1354–1384, Mar. 2019.

[10] M. Tahmasbi and M. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Trans. Inform. Theory*, vol. 65, pp. 2190–2212, Apr. 2019.

[11] L. Wang, "Covert communication over the Poisson channel," *IEEE J. Select. Areas Inform. Theory*, vol. 2, pp. 23–31, Mar. 2021.

[12] H. Zivari-Fard, M. Bloch, and A. Nosratinia, "Keyless covert communication via channel state information." To app. in *IEEE Trans. Inform. Theory*, 2021.

[13] A. D. Wyner, "The wiretap channel," *Bell System Techn. J.*, vol. 54, pp. 1355–1387, 1975.

[14] M. Tahmasbi and M. Bloch, "Covert secret key generation with an active warden," *IEEE Trans. Inform. Forensics and Security*, vol. 15, pp. 1026–1039, 2020.

[15] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback I: No bandwidth constraint," *IEEE Trans. Inform. Theory*, vol. 12, pp. 172–182, Apr. 1966.

[16] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inform. Theory*, vol. 56, pp. 6–17, Jan. 2010.