Covert Communication with Noncausal Channel-State Information at the Transmitter

Si-Hyeon Lee, Ligong Wang, Ashish Khisti, Gregory W Wornell

Covert Communication with Noncausal Channel-State Information at the Transmitter

Si-Hyeon Lee*, Ligong Wang†, Ashish Khisti‡, and Gregory W. Wornell§
*POSTECH, Pohang, South Korea (e-mail:sihyeon@postech.ac.kr)
†ETIS–Université Paris Seine, Université de Cergy-Pontoise, ENSEA, CNRS, France (e-mail: ligong.wang@ensea.fr)
‡University of Toronto, Toronto, Canada (e-mail:akhisti@comm.utoronto.ca)
§Massachusetts Institute of Technology, Cambridge, USA (e-mail: gww@mit.edu)

***Abstract*—We consider the problem of covert communication over a state-dependent channel, where the transmitter has noncausal knowledge of the channel states. Here, "covert" means that the probability that a warden on the channel can detect the communication must be small. In contrast with traditional models without noncausal channel-state information at the transmitter, we show that covert communication can be possible with positive rate. We derive closed-form formulas for the maximum achievable covert communication rate ("covert capacity") in this setting for discrete memoryless channels as well as additive white Gaussian noise channels. We also derive lower bounds on the rate of the secret key that is needed for the transmitter and the receiver to achieve the covert capacity.**

I. Introduction

Covert communication [1]–[4] refers to scenarios where the transmitter and the receiver must keep the warden (eavesdropper) from discovering the fact that they are using the channel to communicate. Specifically, the signals observed by the warden must be statistically close to the signals when the transmitter is switched off. For additive white Gaussian noise (AWGN) channels, the transmitter being switched off is usually modeled by it always sending zero; for discrete memoryless channels (DMCs), this is modeled by it sending a specially designated "no input" symbol x_0. For a DMC, if the output distribution at the warden generated by x_0 is a convex combination of the output distributions generated by the other input symbols, then a positive covert communication rate is achievable; otherwise the maximum amount of information that can be covertly communicated scales like the square root of the total number of channel uses [3]. For the AWGN channel, the latter situation applies [1], [3].

The role played by *channel uncertainties* in covert communications has been studied in some recent works. In particular, [5]–[7] consider the situation where the noise level in the channel is unknown to the warden, and show that, in this case, positive covert communication rates are achievable on certain channel models (binary channels are considered in [5] and AWGN channels in [6], [7]) which otherwise only allow square-root scaling for covert communication. An important assumption common to [5]–[7] is that the unknown parameter, e.g., noise power, remains the same throughout the entire communication duration. This makes it difficult for the warden to tell whether what it observes is signal power or noise power.

The current work studies the benefit of channel uncertainties for covert communications in a different context. We consider channels with an unknown parameter that is independent and identically distributed (IID) across different channel uses, and that is known to the transmitter noncausally as channel-state information (CSI).[1] We study the maximum achievable rate for covert communication, which we call the "covert capacity" of such channels. For both DMCs and AWGN channels, we derive closed-form formulas for the covert capacity, as well as lower bounds for the length of the secret key shared between the transmitter and the receiver that may be needed to achieve the covert capacity. Our achievability proofs are based on "likelihood encoding" employed in [8]–[10] rather than standard Gelfand-Pinsker coding [11], because the former admits easier covertness analysis. Our converse proof is based on that of [11], taking into account covertness requirements.

For some DMCs (Example 1) and for the AWGN channel (Theorem 5), our results show that the covert capacity is zero without CSI but positive with noncausal CSI at the transmitter. This, to some extent, confirms again that channel statistics unknown to the warden can help the transmitter to communicate covertly.

Our work is closely related to some works in steganography [12]–[14]. As noted in [14], the covertext in steganography can be seen as CSI that is noncausally known to the transmitter. However, in steganography it is normally assumed that no noise is imposed on the stegotext, hence, conditional on the states (i.e., the covertext), the channel is noiseless. In our setting, the channel has both states and noise.

II. Problem Formulation

A state-dependent DMC $(\mathcal{X}, \mathcal{S}, \mathcal{Y}, \mathcal{Z}, P_S, P_{Y,Z|S,X})$ consists of channel input alphabet $\mathcal{X}$, state alphabet $\mathcal{S}$, channel output alphabets $\mathcal{Y}$ and $\mathcal{Z}$ at the receiver and the warden, respectively, state probability mass function (PMF) P_S, and channel law $P_{Y,Z|S,X}$. All alphabets are finite. Let $x_0 \in \mathcal{X}$ be a "no input" symbol that is sent when no communication takes place. Define $Q_0(\cdot) = \sum_{s\in\mathcal{S}} P_S(s)P_{Z|S,X}(\cdot|s,$

[1]Note that, if an IID channel parameter is not known to any terminal, then it can be treated as part of the channel statistics, and generally cannot help the communicating parties to communicate covertly. One can see that the same is true for an ergodic random parameter having a coherence time that is much shorter than the communication duration. Hence the assumption of the transmitter having CSI is crucial to our model.

x_0) and let $Q_0^{\times n}(\cdot)$ denote the n-fold product of Q_0. We assume $\text{supp}(Q_0) = \mathcal{Z}$ where $\text{supp}(\cdot)$ denotes the support set of a distribution. The state sequence S^n is assumed to be IID, hence the warden observes Z^n distributed according to $Q_0^{\times n}(\cdot)$ if no communication takes place over n channel uses. We define a nonnegative cost $b(x)$ for each input symbol $x \in \mathcal{X}$. The average input cost of $x^n \in \mathcal{X}^n$ is defined as $b(x^n) = \frac{1}{n}\sum_{i=1}^n b(x_i)$.

The state sequence is assumed to be noncausally known to the transmitter, but unknown to the receiver and the warden.[2] Furthermore, the transmitter and the receiver are assumed to share a secret key K uniformly distributed over a set $\mathcal{K}$. An $(|\mathcal{M}|, |\mathcal{K}|, n)$ code consists of an encoder at the transmitter that maps (M, K, S^n) to $X^n \in \mathcal{X}^n$, and a decoder at the receiver that maps (Y^n, K) to $\hat{M} \in \mathcal{M}$.

The transmitter and the receiver aim at constructing a code that is both reliable and covert. As usual, their code is reliable if the probability of error $P_e^{(n)} = P(\hat{M} \neq M)$ is negligible. Their code is covert if it is hard for the warden to determine whether the transmitter is sending a message (hypothesis H_1) or not (hypothesis H_0). Let α and β denote the probabilities of false alarm (accepting H_1 when the transmitter is not sending a message) and missed detection (accepting H_0 when the transmitter is sending a message), respectively. Note that a blind test satisfies $\alpha + \beta = 1$. Let $\widehat{P}$ denote the distribution when the transmitter is sending a message.[3] The warden's optimal hypothesis test satisfies $\alpha + \beta \geq 1 - \sqrt{D(\widehat{P}_{Z^n} \| Q_0^{\times n})}$ (see [16]). Hence, covertness is guaranteed if $D(\widehat{P}_{Z^n} \| Q_0^{\times n})$ is negligible.

Let $\mathcal{K} = [1 : 2^{nR_K}]$ and $\mathcal{M} = [1 : 2^{nR}]$ for $R_K \geq 0$ and $R \geq 0$. For given $R_K \geq 0$ and $B \geq 0$, a covert rate of R is said to be achievable if there exists a sequence of $(2^{nR}, 2^{nR_K}, n)$ codes that simultaneously satisfies the input cost constraint $\limsup_{n\to\infty} \mathrm{E}_{M,K,S^n}[b(X^n)] \leq B$, reliability constraint $\lim_{n\to\infty} P_e^{(n)} = 0$, and covertness constraint $\lim_{n\to\infty} D(\widehat{P}_{Z^n} \| Q_0^{\times n}) = 0$. The *covert capacity* C is defined as the supremum of all achievable covert rates.

III. Main Results for DMCs

The following two theorems present an upper and a lower bound on the covert capacity, respectively. The proofs of these theorems are provided in Sections IV and V, respectively.

Theorem 1. *For $R_K \geq 0$ and $B \geq 0$, the covert capacity is upper-bounded as*

$$C \leq \max(I(U;Y) - I(U;S)) \tag{1}$$

where the maximum is over conditional PMF $P_{U|S}$ and function $x(u,s)$ such that $|\mathcal{U}| \leq \min(|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| + |\mathcal{S}| - 3, |\mathcal{X}| \cdot |\mathcal{S}|)$, $P_Z = Q_0$ and $\mathrm{E}[b(X)] \leq B$.

[2]In the full version [15] of this paper, the case where the state sequence is causally known to the transmitter is also considered.

[3]Note that $\widehat{P}_{Z^n}$ depends on the code used for the communication and is in general not IID. On the other hand, $\widehat{P}_{S^n} = P_S^{\times n}$ since the state sequence is generated independently of whether communication is taking place or not.

Theorem 2. *For $R_K \geq 0$ and $B \geq 0$, the covert capacity is lower-bounded as*

$$C \geq \max(I(U;Y) - I(U;S)) \tag{2}$$

where the maximum is over conditional PMF $P_{U|S}$ and function $x(u,s)$ such that $|\mathcal{U}| \leq \min(|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| + |\mathcal{S}| - 2, |\mathcal{X}| \cdot |\mathcal{S}| + 1)$, $P_Z = Q_0$, $\mathrm{E}[b(X)] \leq B$, and

$$I(U;Z) - I(U;Y) < R_K. \tag{3}$$

Remark 1. *If R_K is large enough so that* (3) *holds under the joint distribution that achieves the maximum on the right-hand side of* (1)*, then Theorems 1 and 2 establish the covert capacity as the right-hand side of* (1)*. Furthermore, if under this joint distribution $I(U;Z) < I(U;Y)$, then no secret key is needed to achieve the covert capacity.*

Example 1. *Consider a channel where $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{S}$ are all binary, and where P_S is the Bernoulli distribution of parameter $p \in (0, 0.5)$. The channel law is $Y = Z = X \oplus S$. Assume that $R_K > 0$.*

Using Theorems 1 and 2 one can check that the optimal choice is $U = Y = Z$ having the Bernoulli distribution of parameter p, independently of S. This gives $C = H_\mathrm{b}(p) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$.

Note that, without CSI, the channel in Example 1 is a binary symmetric channel, over which covert communication cannot have a positive rate [2], [3].

IV. Proof of Upper Bound (Theorem 1)

Let us first define the following function:

$$C(A,B) = \max_{\substack{P_{U|S}, P_{X|U,S}: \\ \mathrm{E}[b(X)] \leq B,\ D(P_Z \| Q_0) \leq A}} (I(U;Y) - I(U;S)).$$

In the proof of Theorem 1, we use the following property of $C(A,B)$, which is proven in the full version [15] of this paper.

Lemma 3. *The function $C(A,B)$ is non-decreasing in each of A and B, and concave and continuous in (A,B).*

Proof of Theorem 1. For $R_K \geq 0$ and $B \geq 0$, consider any sequence of $(2^{nR}, 2^{nR_K}, n)$ codes that simultaneously satisfies the input cost constraint $\limsup_{n\to\infty} \mathrm{E}_{M,K,S^n}[b(X^n)] \leq B$, reliability constraint $\lim_{n\to\infty} P_e^{(n)} = 0$, and covertness constraint $\lim_{n\to\infty} D(\widehat{P}_{Z^n} \| Q_0^{\times n}) = 0$.

Let us start with the proof steps used for channels with noncausal CSI [11] without a covertness constraint:

$$\begin{aligned}
nR &\overset{(a)}{\leq} I(M; Y^n | K) + n\epsilon_n && (4)\\
&= \sum_{i=1}^n I(M; Y_i | K, Y^{i-1}) + n\epsilon_n && (5)\\
&\leq \sum_{i=1}^n I(M, K, Y^{i-1}; Y_i) + n\epsilon_n && (6)\\
&= \sum_{i=1}^n I(M, K, Y^{i-1}, S_{i+1}^n; Y_i) \\
&\qquad - \sum_{i=1}^n I(Y_i; S_{i+1}^n | M, K, Y^{i-1}) + n\epsilon_n && (7)
\end{aligned}$$

$$\stackrel{(b)}{=} \sum_{i=1}^{n} I(M,K,Y^{i-1},S_{i+1}^{n};Y_i) - \sum_{i=1}^{n} I(Y^{i-1};S_i|M,K,S_{i+1}^{n}) + n\epsilon_n \tag{8}$$

$$\stackrel{(c)}{=} \sum_{i=1}^{n} I(M,K,Y^{i-1},S_{i+1}^{n};Y_i) - \sum_{i=1}^{n} I(M,K,Y^{i-1},S_{i+1}^{n};S_i) + n\epsilon_n \tag{9}$$

$$= \sum_{i=1}^{n} (I(U_i;Y_i) - I(U_i;S_i)) + n\epsilon_n \tag{10}$$

for $\epsilon_n \to 0$ and $U_i := (M,K,Y^{i-1},S_{i+1}^{n})$. Here, (a) follows by applying Fano's inequality from the reliability constraint; (b) by Csiszár's sum identity; and (c) because S_i and (M,K,S_{i+1}^{n}) are independent.

Now we utilize the definition and the property of $C(A,B)$ to further bound the right-hand side of (10):

$$nR \le \sum_{i=1}^{n} (I(U_i;Y_i) - I(U_i;S_i)) + n\epsilon_n \tag{11}$$

$$\le \sum_{i=1}^{n} C(D(\widehat{P}_{Z_i}\|Q_0), \mathrm{E}[b(X_i)]) + n\epsilon_n \tag{12}$$

$$\stackrel{(a)}{\le} nC\left(\frac{1}{n}\sum_{i=1}^{n} D(\widehat{P}_{Z_i}\|Q_0), \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[b(X_i)]\right) + n\epsilon_n \tag{13}$$

where (a) is due to the concavity of $C(A,B)$. Recall from Lemma 3 that $C(A,B)$ is non-decreasing in each of A and B. According to the input cost constraint, there exists $\delta_n \to 0$ such that $\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[b(X_i)] \le B + \delta_n$. On the other hand, from the covertness constraint, there exists $\delta_n' \to 0$ such that $D(\widehat{P}_{Z^n}\|Q_0^{\times n}) \le \delta_n'$, while we have

$$D(\widehat{P}_{Z^n}\|Q_0^{\times n}) = -H(Z^n) + \mathrm{E}_{\widehat{P}_{Z^n}}\left[\log\frac{1}{Q_0^{\times n}(Z^n)}\right] \tag{14}$$

$$= -\sum_{i=1}^{n} H(Z_i|Z^{i-1}) + \mathrm{E}_{\widehat{P}_{Z^n}}\left[\log\frac{1}{Q_0(Z_i)}\right] \tag{15}$$

$$= -\sum_{i=1}^{n} H(Z_i|Z^{i-1}) + \mathrm{E}_{\widehat{P}_{Z_i}}\left[\log\frac{1}{Q_0(Z_i)}\right] \tag{16}$$

$$\ge -\sum_{i=1}^{n} H(Z_i) + \mathrm{E}_{\widehat{P}_{Z_i}}\left[\log\frac{1}{Q_0(Z_i)}\right] \tag{17}$$

$$= \sum_{i=1}^{n} D(\widehat{P}_{Z_i}\|Q_0). \tag{18}$$

Hence, (13) implies

$$R \le C\left(\frac{\delta_n'}{n}, B+\delta_n\right) + \epsilon_n. \tag{19}$$

Note that the right-hand side of (19) approaches $C(0,B)$ as n tends to infinity due to the continuity of $C(A,B)$, from which follows the condition $P_Z = Q_0$. Further, because $I(U;Y) - I(U;S)$ is convex in the conditional distribution $P_{X|U,S}$, it suffices to maximize it over functions $x(u,s)$ instead of $P_{X|S,U}$. Finally, the cardinality bound on $\mathcal{U}$ follows by applying the support lemma [17]. □

V. Proof of Lower Bound (Theorem 2)

Fix $\epsilon > \epsilon' > 0$. Further fix $P_{U|S}$ and $x(u,s)$ such that $P_Z = Q_0$ and $\mathrm{E}[b(X)] \le \frac{B}{1+\epsilon'}$.

1) Codebook generation: For each $k \in [1 : 2^{nR_K}]$ and $m \in [1 : 2^{nR}]$, randomly and independently generate $2^{nR'}$ codewords $u^n(k,m,l)$, $l \in [1 : 2^{nR'}]$ according to $\prod_{i=1}^{n} P_U(u_i)$. These constitute the codebook $\mathcal{C}$.

2) Encoding at the transmitter: Given state sequence s^n, secret key k, and message m, evaluate the likelihood

$$g(l|s^n,k,m) = \frac{P_{S|U}^{\times n}(s^n|u^n(k,m,l))}{\sum_{l'\in[1:2^{nR'}]} P_{S|U}^{\times n}(s^n|u^n(k,m,l'))}. \tag{20}$$

The encoder randomly generates l according to (20) and transmits x^n where $x_i = x(u_i(k,m,l),s_i)$.

3) Decoding at the receiver: Upon receiving y^n, with access to the secret key k, the decoder declares that $\hat{m}$ is sent if it is the unique message such that

$$(u^n(k,\hat{m},l), y^n) \in \mathcal{T}_\epsilon^{(n)} \tag{21}$$

for some $l \in [1 : 2^{nR'}]$; otherwise it declares an error. Here $\mathcal{T}_\epsilon^{(n)}$ denotes the (strongly) typical set [18].

4) Covertness analysis: For covertness analysis, we use the following lemma, which is proven in Appendix A.

Lemma 4. *For the codebook generation and encoding procedure described in Sections V-1 and V-2, respectively, if $R' > I(U;S)$ and $R + R_K + R' > I(U;Z)$, then*

$$\mathrm{E}_{\mathcal{C}}\left[D(\widehat{P}_{Z^n}\|P_Z^{\times n})\right] \stackrel{n\to\infty}{\longrightarrow} 0. \tag{22}$$

Now, let

$$R' > I(U;S) \tag{23}$$

$$R + R_K + R' > I(U;Z). \tag{24}$$

Because $P_{U|S}$ and $x(u,s)$ are chosen to satisfy $P_Z = Q_0$, Lemma 4 implies that

$$\mathrm{E}_{\mathcal{C}}[D(\widehat{P}_{Z^n}\|Q_0^{\times n})] \stackrel{n\to\infty}{\longrightarrow} 0. \tag{25}$$

5) Reliability analysis: Consider the probability of error averaged over the randomly generated codebook $\mathcal{C}$. Let M and $\hat{M}$ denote the transmitted and decoded messages, respectively, and let L denote the index generated according to (20) at the encoder. The error event $\{\hat{M} \ne M\}$ occurs only if at least one of the following events occurs:

$$\mathcal{E}_1 := \{(U^n(K,M,L), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\} \tag{26}$$

$$\mathcal{E}_2 := \{(U^n(K,M,L), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\} \tag{27}$$

$$\mathcal{E}_3 := \{(U^n(K,m,l), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m \ne M \text{ and } l \in [1 : 2^{nR'}]\}. \tag{28}$$

Hence, the probability of error is bounded as

$$P(\hat{M} \ne M) \le P(\mathcal{E}_1) + P(\mathcal{E}_1^c \cap \mathcal{E}_2) + P(\mathcal{E}_3). \tag{29}$$

Now we bound each term on the right-hand side of (29). The first term $P(\mathcal{E}_1)$ tends to zero as n tends to infinity due to [9, Lemma 2], as long as (23) is satisfied. Next, note that

$$\mathcal{E}_1^c = \{(U^n(K,M,L), S^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}. \tag{30}$$

By the conditional typicality lemma [17], $P(\mathcal{E}_1^c \cap \mathcal{E}_2)$ tends to zero as n tends to infinity.

Lastly, $P(\mathcal{E}_3)$ tends to zero as n tends to infinity by the packing lemma [17] provided

$$R + R' < I(U;Y). \tag{31}$$

In summary, the probability of error averaged over the random codebook $\mathcal{C}$ tends to zero as n tends to infinity if (23), (24), and (31) are satisfied.

6) Input cost analysis: In the reliability analysis, it is shown that

$$P(\mathcal{E}_1) = P\{(U^n(K,M,L), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\} \tag{32}$$

$$= P((U^n(K,M,L), X^n, S^n) \notin \mathcal{T}_{\epsilon'}^{(n)}) \overset{n\to\infty}{\longrightarrow} 0. \tag{33}$$

Note that if $x^n \in \mathcal{T}_{\epsilon'}^{(n)}$, then $b(x^n) \le B$ by the typical average lemma [17]. Hence,

$$\begin{aligned}
&\mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)\right] \\
&= P(\mathcal{E}_1)\,\mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)|\mathcal{E}_1\right] \\
&\quad + P(\mathcal{E}_1^c)\,\mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)|\mathcal{E}_1^c\right] &(34)\\
&\le P(\mathcal{E}_1)B_{\max} + P(\mathcal{E}_1^c)B, &(35)
\end{aligned}$$

where $B_{\max} := \max_{x\in\mathcal{X}} b(x)$. By (33), the right-hand side of (35) approaches B as n tends to infinity. Hence, we have

$$\limsup_{n\to\infty} \mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)\right] \le B. \tag{36}$$

In summary, if (23), (24), and (31) are satisfied, then there must exist a sequence of codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$, $\lim_{n\to\infty} \mathrm{E}_{M,K,S^n}\left[b(X^n)\right] \le B$, and $\lim_{n\to\infty} D(P_{Z^n}\|Q_0^{\times n}) = 0$. By applying the Fourier-Mozkin elimination [17] to (23), (24), and (31), we complete the proof.

VI. The AWGN Channel

Consider an AWGN channel where the channel outputs at the receiver and the warden are given as

$$Y = X + S + N_Y \tag{37}$$

$$Z = X + S + N_Z, \tag{38}$$

respectively, where X is the channel input from the transmitter, $S \sim \mathcal{N}(0,T)$ is the external interference that is known to the transmitter noncausally but unknown to the receiver and the warden, and $N_Y \sim \mathcal{N}(0,1)$ and $N_Z \sim \mathcal{N}(0,\sigma^2)$, $\sigma^2 > 0$, are additive Gaussian noises. Let P denote the input power constraint at the transmitter, so the input must satisfy $\mathrm{E}[X^2] \le P$. The "no input" symbol is 0, hence the warden observes Z^n distributed according to $Q_0^{\times n}$, where $Q_0 = \mathcal{N}(0, T+\sigma^2)$, when no communication takes place over n channel uses. The transmitter and the receiver are assumed to share a secret key of rate R_K. The covertness constraint is again given by $\lim_{n\to\infty} D(\widehat{P}_{Z^n}\|Q_0^{\times n}) = 0$. The covert capacity of this channel is defined in the same way as in Section II.

The following theorem establishes the covert capacity when the secret key rate is sufficiently large.

Theorem 5. *Let* $P^* := \min\{P, 2T\} - \frac{(\min\{P,2T\})^2}{4T}$. *If*

$$R_K > \log \frac{P^* + (1-\gamma)^2 T + \sigma^2}{P^*\sigma^2 + \frac{1+P^*\sigma^2}{1+P^*}(1-\gamma)^2 T + \sigma^2}, \tag{39}$$

where $\gamma = \min\{1, \frac{P}{2T}\}$, *then the covert capacity is given by* $C = \frac{1}{2}\log(1+P^*)$.

Remark 2. *If the warden's channel is degraded, i.e., if* $\sigma^2 \ge 1$, *a secret key is not needed.*

Remark 3. *Assume that the secret key rate is sufficiently large. As* $T \to \infty$, *the covert capacity approaches* $\frac{1}{2}\log(1+P)$, *which is the capacity of the channel* (37) *with noncausal CSI and without covertness constraint. On the other hand, when* $T = 0$, *a positive covert-communication rate is not achievable, which is consistent with [1], [3].*

Converse proof of Theorem 5. It can be checked that Theorem 1 applies to the AWGN channel with $b(x) = x^2$. Fix $P_{U|S}$ and $x(u,s)$ that achieve the maximum in (1). Note that $P_Z = Q_0$ and $\mathrm{E}[X^2] \le P$. Let $\tilde{P} := \mathrm{Var}(X)$ and $\Lambda := \mathrm{E}[XS]$. It follows that

$$\begin{aligned}
I(U;Y) - I(U;S) &\le I(U;Y,S) - I(U;S) &(40)\\
&= I(U;Y|S) &(41)\\
&\le I(X,U;Y|S) &(42)\\
&\overset{(a)}{=} I(X;Y|S) &(43)\\
&= h(X+N_Y|S) - h(N_Y) &(44)\\
&\overset{(b)}{\le} \frac{1}{2}\log\left(1 + \tilde{P} - \frac{\Lambda^2}{T}\right), &(45)
\end{aligned}$$

where (a) is due to the Markov chain $U - (X,S) - Y$ and (b) is from [17, Problem 2.7]. Recall the condition $P_Z = Q_0$, which implies

$$\begin{aligned}
T + \sigma^2 &= \mathrm{Var}(X + S + N_Z) &(46)\\
&= \tilde{P} + T + 2\Lambda + \sigma^2, &(47)
\end{aligned}$$

therefore we must have $\Lambda = -\frac{\tilde{P}}{2}$. Hence (45) implies

$$I(U;Y) - I(U;S) \le \frac{1}{2}\log\left(1 + \tilde{P} - \frac{\tilde{P}^2}{4T}\right). \tag{48}$$

Note that $\tilde{P} \le P$ and $\arg\max_{0\le\tilde{P}\le P}\left(\tilde{P} - \frac{\tilde{P}^2}{4T}\right) = \min\{P, 2T\}$. Thus, we have

$$C \le \frac{1}{2}\log\left(1 + \min\{P,2T\} - \frac{(\min\{P,2T\})^2}{4T}\right), \tag{49}$$

which concludes the proof. □

Achievability proof of Theorem 5. We can adapt Theorem 2 for the Gaussian case with a power constraint, as explained in Remark 4. We basically perform dirty paper coding (DPC) [19], after reducing the interference power to make room for message transmission. Hence, we let $X = X' - \gamma S$ where γS is subtracted from X to reduce the interference power and treat $X' \sim \mathcal{N}(0,P')$ as the channel input for performing DPC over the equivalent channel $Y = X' + (1-\gamma)S + N_Y$.

Formally, this corresponds to the following choice of $P_{U|S}$ and $x(u,s)$ in Theorem 2:

$$X' \sim \mathcal{N}(0,P'), \text{ independent of } S \tag{50}$$

$$U = X' + \frac{P'}{P'+1}(1-\gamma)S \tag{51}$$

$$X = X' - \gamma S. \tag{52}$$

The parameters γ and P' are determined from the following:

$$P' + \gamma^2 T \le P \tag{53}$$

$$P' + (1-\gamma)^2 T = T, \tag{54}$$

where the inequality comes from the power constraint $\mathrm{E}[X^2] \leq P$ and the equality is due to the covertness constraint $P_Z = Q_0$. From the above, we choose $\gamma = \min\{1, \frac{P}{2T}\}$ and $P' = P^*$.

By applying the aforementioned choice of $P_{U|S}$ and $x(u,s)$ in Theorem 2, we obtain $C \geq \frac{1}{2}\log(1+P^*)$ when the secret key rate satisfies

$$R_K > I(U;Z) - I(U;Y) \tag{55}$$

$$= \log \frac{P^* + (1-\gamma)^2 T + \sigma^2}{P^*\sigma^2 + \frac{1+P^*\sigma^2}{1+P^*}(1-\gamma)^2 T + \sigma^2}, \tag{56}$$

which completes the proof. □

Remark 4. *We adapt Theorem 2 for the Gaussian case with a power constraint as follows. We employ the same encoding procedure while PMFs are replaced with probability density functions induced by the channel* (37)-(38) *and the choice* (50)-(52)*. Then, the covertness analysis proceeds in the same way as in the discrete case and hence the covertness condition would be satisfied if* (23) *and* (24) *are satisfied.*

For the decoding procedure, we cannot use the notion of strong typicality as in (21) *for continuous alphabet. Instead, the decoder considers quantized versions of U and Y. A partition of $\mathcal{P}$ of $\mathcal{U}$ is a finite collection of disjoint sets P_i such that $\cup_i P_i = \mathcal{U}$. The quantization of U by $\mathcal{P}$ is denoted as $[U]_{\mathcal{P}}$ and defined by*

$$P\left([U]_{\mathcal{P}} = \begin{cases} \sup P_i \text{ if } \sup P_i < \infty \\ \inf P_i \text{ otherwise} \end{cases}\right) = P(U \in P_i).$$

For partitions $\mathcal{P}$ and $\mathcal{P}'$ of $\mathcal{U}$ and $\mathcal{Y}$, respectively, the decoder performs joint typicality check for $[U]^n_{\mathcal{P}}$ and $[Y]^n_{\mathcal{P}'}$ with respect to the joint distribution $p([U]_{\mathcal{P}}, [Y]_{\mathcal{P}'})$. Similarly, let us consider quantized version $[S]_{\tilde{\mathcal{P}}}$ of S by partition $\tilde{\mathcal{P}}$ for the joint typicality in (26) *and consider $[X]_{\mathcal{P},\tilde{\mathcal{P}}} := x([U]_{\mathcal{P}}, [S]_{\tilde{\mathcal{P}}})$ for the joint typicality in* (33)*. Then, the condition* (31) *becomes $R + R' < I([U]_{\mathcal{P}}; [Y]_{\mathcal{P}'})$. As we refine the partitions $\mathcal{P}$ and $\mathcal{P}'$, $I([U]_{\mathcal{P}}; [Y]_{\mathcal{P}'})$ approaches to $I(U;Y)$ according to [20, Section 8.6]. Furthermore, the input cost constraint is asymptotically satisfied as the partitions are refined given that the function $x(u,s)$ is continuous in each of u and s.*

Appendix A
Proof of Lemma 4

The proof follows similar lines to [9, Section VII-A]. As in [9, Section VII-A], it can be checked that, to prove (22), it suffices to show that the total variation (TV) distance approaches zero:

$$\mathrm{E}_{\mathcal{C}} \left\| \widehat{P}_{Z^n} - P_Z^{\times n} \right\|_{\mathrm{TV}} \overset{n\to\infty}{\longrightarrow} 0. \tag{57}$$

To evaluate the TV distance, define the ideal PMF for codebook $\mathcal{C}$ as follows:

$$\Gamma^{(\mathcal{C})}(k,m,l,u^n,s^n,z^n) =$$

$$2^{-n(R_K+R+R')} \mathbb{1}_{u^n(k,m,l)=u^n} P_{S|U}^{\times n}(s^n|u^n) P_{Z|U,S}^{\times n}(z^n|u^n,s^n).$$

Using the triangle inequality for the TV distance, we upper-bound the left-hand side of (57) as

$$\mathrm{E}_{\mathcal{C}} \left\| \widehat{P}_{Z^n} - P_Z^{\times n} \right\|_{\mathrm{TV}}$$

$$\leq \mathrm{E}_{\mathcal{C}} \left\| \widehat{P}_{Z^n} - \Gamma^{(\mathcal{C})}_{Z^n} \right\|_{\mathrm{TV}} + \mathrm{E}_{\mathcal{C}} \left\| \Gamma^{(\mathcal{C})}_{Z^n} - P_Z^{\times n} \right\|_{\mathrm{TV}}. \tag{58}$$

From the soft covering theorem [21, Theorem 4], [22, Corollary VII.4], the second term on the right-hand side of (58) decays to zero as $n \to \infty$ if $R_K + R + R' > I(U;Z)$. For the first term on the right-hand side of (58), note that

$$\mathrm{E}_{\mathcal{C}} \left\| \widehat{P}_{Z^n} - \Gamma^{(\mathcal{C})}_{Z^n} \right\|_{\mathrm{TV}} \leq \mathrm{E}_{\mathcal{C}} \left\| \widehat{P}_{S^n,Z^n} - \Gamma^{(\mathcal{C})}_{S^n,Z^n} \right\|_{\mathrm{TV}}. \tag{59}$$

By applying the same analysis as in [9, Section VII-A], the right-hand side of (59) decays to zero as $n \to \infty$ if $R' > I(U;S)$.

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, no. 9, 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] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 3493–3503, June 2016.

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

[5] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. Inform. Theory Workshop (ITW)*, Hobart, Australia, Nov. 2–5, 2014.

[6] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1195–1205, Oct 2015.

[7] T. V. Sobers, B. A. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert communication in the presence of an uninformed jammer," [Online]. Available: http://arxiv.org/abs/1608.00698.

[8] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1836–1849, April 2016.

[9] Z. Goldfeld, G. Kramer, H. H. Permuter, and P. Cuff, "Strong secrecy for cooperative broadcast channels," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 469–495, Jan 2017.

[10] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Wiretap channels with random states non-causally available at the encoder," [Online]. Available: http://arxiv.org/abs/1608.00743.

[11] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, pp. 19–31, 1980.

[12] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 2009.

[13] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions,," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2706–2722, June 2008.

[14] I. Ezzeddine and P. Moulin, "Achievable rates for queue-based timing stegocodes," in *Proc. Inform. Theory Workshop (ITW)*, 2009.

[15] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," in preparation.

[16] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*. New York: Springer Verlag., 2005.

[17] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[18] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[19] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[20] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991.

[21] T. S. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[22] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, Nov 2013.