Covert Communication with Channel-State Information at the Transmitter

Si-Hyeon Lee, *Member, IEEE,* Ligong Wang, *Member, IEEE,* Ashish Khisti, *Member, IEEE,* and Gregory W. Wornell, *Fellow, IEEE,*

***Abstract*—We consider the problem of covert communication over a state-dependent channel, where the transmitter has causal or noncausal knowledge of the channel states. Here, "covert" means that a warden on the channel should observe similar statistics when the transmitter is sending a message and when it is not. When a sufficiently long secret key is shared between the transmitter and the receiver, we derive closed-form formulas for the maximum achievable covert communication rate ("covert capacity") for discrete memoryless channels and, when the transmitter's channel-state information (CSI) is noncausal, for additive white Gaussian noise (AWGN) channels. For certain channel models, including the AWGN channel, we show that the covert capacity is positive with CSI at the transmitter, but is zero without CSI. 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 [2]–[5] 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 [4]. For the AWGN channel, the latter situation applies [2], [4].

The role played by *channel uncertainties* in covert communications has been studied in some recent works. In particular, [6]–[8] consider the situation where the noise level (or crossover probability) of the channel is random, remains constant throughout the entire communication duration, and is unknown to the warden. In this case, it is difficult for the warden to tell whether what it observes is signal or noise. As a consequence, positive covert communication rates are achievable on certain channel models (binary symmetric channels are considered in [6] and AWGN channels in [7], [8]) which, without the unknown-noise-level assumption, only allow square-root scaling for covert communication.

The current work studies the benefit of channel uncertainties for covert communications in a different context. We consider channels with a random state that is independent and identically distributed (IID) across different channel uses. Clearly, if a channel state sequence is not known to any terminal, then it can be treated as part of the channel statistics, reducing the problem to the one studied in previous works. Hence, in general, an IID unknown parameter cannot help the communicating parties to communicate covertly. In the current work, we assume that the state sequence is known to the transmitter, either causally or noncausally, as channel-state information (CSI), but unknown to the receiver and the warden. As one motivating application consider a scenario where a noise source or jammer continuously emits IID random noise. If the jammer is friendly and reveals the predetermined noise symbols to the transmitter, then the transmitter can use this knowledge as CSI. Another scenario is when the path delay from the jammer to the receiver and the warden is larger than the total path delay from the jammer to the transmitter and from the transmitter to the receiver and the warden so that the transmitter knows the jammer's signal in advance and utilizes it as CSI. In the literature, such difference in path delays has motivated the study of lookahead relay channels [9], [10], [11, Chapter 16.9].

We study the maximum achievable rate for covert communication, which we call the "covert capacity," in this case. We derive closed-form formulas for the covert capacity, when the transmitter and the receiver share a sufficiently long secret key. We also derive upper bounds on the minimum length of the secret key needed to achieve the covert capacity. We do not have a good lower bound on this minimum secret-key length; we briefly comment on this in the concluding section. Our converse proofs are based on classical techniques, while covertness is accounted for with help of continuity properties of certain information quantities. Our achievability proof for the noncausal case is based on "likelihood encoding" employed in [12] rather than standard Gelfand-Pinsker coding

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2017R1D1A1B03034492), and in part by Korea Electric Power Corporation(Grant number:R18XA01).

S.-H. Lee is with the Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, South Korea 37673 (e-mail: sihyeon@postech.ac.kr). L. Wang is with ETIS–Université Paris Seine, Université de Cergy-Pontoise, ENSEA, CNRS, Cergy-Pontoise, France (e-mail: ligong.wang@ensea.fr). A. Khisti is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S, Canada (e-mail: akhisti@comm.utoronto.ca). G. W. Wornell is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA (e-mail: gww@mit.edu). The material in this paper was presented in part at IEEE ISIT 2017 [1].

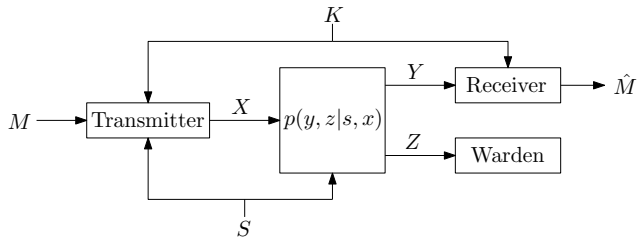

Figure 1. State-dependent discrete memoryless channel

[13], because the former admits easier covertness analysis. For the binary symmetric channel (BSC) and the AWGN channel, in certain parameter ranges, we show the covert capacity to be positive with CSI at the transmitter. (Recall that, without channel state, the covert capacity is zero for both channels in all parameter ranges.) The intuition behind achieving a positive rate for the example of AWGN channel (Section VI-B) where additive interference signal is known to the transmitter as state information is that we can reduce the interference power by utilizing the knowledge of the interference so that a message signal can be transmitted with a non-vanishing power corresponding to the reduced amount of interference power. We note that the present paper extends our conference paper [1] in the following aspects: (i) the case of causal CSI is studied in addition to the noncausal case, whereas only the latter case is studied in [1]; (ii) the proof of Lemma 5, which is omitted in [1] due to space limit, is provided; and (iii) achievability results for DMC are adapted for AWGN channels in a more rigorous way.

Our work is closely related to some works in steganography [14]–[16]. In steganography, the transmitter is given some data, called the "cover text," and attempts to embed its message by modifying the cover text. As pointed out in [16], the cover text can be seen as CSI that is noncausally known to the transmitter. The main difference between such problems and our setting is the following. In steganography it is normally assumed that no noise is imposed on the "stegotext"—the data after modification by the transmitter, hence, conditional on the states (i.e., the cover text), the channel is noiseless. In our setting, the channel has both states and noise.

The rest of this paper is arranged as follows: Section II formally defines the covert communication problem; Section III states the main results for DMCs; Sections IV and V prove the converse and achievability parts of the main results, respectively; Section VI applies the results to BSCs and AWGN channels; and Section VII concludes the paper with some remarks.

II. Problem Formulation

A state-dependent DMC in Fig. 1

$$(\mathcal{X}, \mathcal{S}, \mathcal{Y}, \mathcal{Z}, P_S, P_{Y,Z|S,X}) \tag{1}$$

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 $Q_0(\cdot) = \sum_{s\in\mathcal{S}} P_S(s) P_{Z|S,X}(\cdot|s, x_0)$ and let $Q_0^{\times n}(\cdot)$ denote the n-fold product of Q_0. 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 transmitter and the receiver are assumed to share a secret key K uniformly distributed over a set $\mathcal{K}$. The state sequence is assumed to be unknown to the receiver and the warden, but available to the transmitter. We consider two cases, where the state sequence is known to the transmitter causally and noncausally, respectively. For causal CSI, an $(|\mathcal{M}|, |\mathcal{K}|, n)$ code consists of an encoder at the transmitter that maps (M, K, S^i) to $X_i \in \mathcal{X}$ for $i \in [1 : n]$, and a decoder at the receiver that maps (Y^n, K) to $\hat{M} \in \mathcal{M}$. For noncausal CSI, 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 $\widehat{P}_{Z^n}$ denote the distribution observed by the warden when the transmitter is sending a message.[1] The warden's optimal hypothesis test satisfies $\alpha + \beta \geq 1 - \sqrt{D(\widehat{P}_{Z^n} \| Q_0^{\times n})}$ (see [17]). Hence, covertness is guaranteed if $D(\widehat{P}_{Z^n} \| Q_0^{\times n})$ is negligible. At this point, note that an input symbol x with $\text{supp}(P_Z(\cdot|x)) \notin \text{supp}(Q_0)$, where supp denotes the support set of a distribution, should not be transmitted with nonzero probability because otherwise $D(\widehat{P}_{Z^n} \| Q_0^{\times n})$ becomes infinity. Hence, by dropping such input symbols, we assume that

$$\text{supp}(Q_0) = \mathcal{Z}. \tag{2}$$

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* is defined as the supremum of all achievable covert rates and denoted by C_{c} and C_{nc} for the cases with causal CSI and with noncausal CSI, respectively.

III. Main Results for DMCs

In this section, we present upper and lower bounds on the covert capacity of DMCs with causal and with noncausal CSI at the transmitter. The proofs of the upper and lower bounds are provided in Sections IV and V, respectively.

[1]Note that $\widehat{P}_{Z^n}$ depends on the code used for the communication and is in general not IID.

A. Causal CSI at the Transmitter

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

$$C_{\mathrm{c}} \leq \max I(V;Y) \tag{3}$$

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

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

$$C_{\mathrm{c}} \geq \max I(V;Y) \tag{4}$$

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

$$I(V;Z) - I(V;Y) < R_K. \tag{5}$$

B. Noncausal CSI at the Transmitter

Theorem 3. *For $R_K \geq 0$ and $B \geq 0$, the covert capacity with noncausal CSI at the transmitter is upper-bounded as*

$$C_{\mathrm{nc}} \leq \max(I(U;Y) - I(U;S)) \tag{6}$$

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 $\mathrm{E}[b(X)] \leq B$.

Theorem 4. *For $R_K \geq 0$ and $B \geq 0$, the covert capacity with noncausal CSI at the transmitter is lower-bounded as*

$$C_{\mathrm{nc}} \geq \max(I(U;Y) - I(U;S)) \tag{7}$$

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

$$I(U;Z) - I(U;Y) < R_K. \tag{8}$$

Remark 1. *If we restrict U and S to be independent in Theorems 3 and 4, the bounds fall back to those in Theorems 1 and 2.*

Remark 2. *For the case with causal CSI (resp. noncausal CSI), if R_K is large enough so that* (5) *(resp.* (8)*) holds under the joint distribution that achieves the maximum on the right-hand side of* (3) *(resp.* (6)*), then Theorems 1 and 2 (resp. Theorems 3 and 4) establish the covert capacity as the right-hand side of* (3) *(resp.* (6)*). Furthermore, if under this joint distribution $I(V;Z) < I(V;Y)$ (resp. $I(U;Z) < I(U;Y)$), then no secret key is needed to achieve the covert capacity.*

Remark 3. *Let us consider the special case of $Y = Z$ as in [4]. Then, in the absence of CSI, the covert capacity can be positive if and only if x_0 is redundant [4], i.e., $P_{Z|X}(\cdot|x_0) \in \mathrm{conv}\{P_{Z|X}(\cdot|x') \colon x' \in \mathcal{X}, x' \neq x_0\}$ where conv denotes the convex hull. In the presence of CSI, the covert capacity can be positive even though x_0 is not redundant. Examples include channels with additive state where some fraction of state can be subtracted through appropriate precoding so that the transmitter can send message symbol (corresponding to V for the causal case and to U for the noncausal case) generated by taking into account the effect of subtracted state. In Section VI, we show such examples.*

IV. Proof of Upper Bounds

In this section, we prove the converse part of our main results, i.e., Theorems 1 and 3. Let us first define the following functions of nonnegative A and B:

$$C_{\mathrm{c}}(A,B) = \max_{\substack{P_V,\ x(v,s):\\ \mathrm{E}[b(X)]\leq B,\ D(P_Z\|Q_0)\leq A}} I(V;Y)$$

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

In the proof of Theorems 1 and 3, we use the following lemma, which is proven at the end of this section.

Lemma 5. *The functions $C_{\mathrm{c}}(A,B)$ and $C_{\mathrm{nc}}(A,B)$ are nondecreasing in each of A and B, and concave and continuous in (A,B).*

Now we are ready to prove Theorems 1 and 3.

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 causal CSI [11] without a covertness constraint:

$$nR \overset{(a)}{\leq} I(M;Y^n|K) + n\epsilon_n \tag{9}$$

$$= \sum_{i=1}^{n} I(M;Y_i|K,Y^{i-1}) + n\epsilon_n \tag{10}$$

$$\leq \sum_{i=1}^{n} I(M,K,Y^{i-1};Y_i) + n\epsilon_n \tag{11}$$

$$\leq \sum_{i=1}^{n} I(M,K,Y^{i-1},S^{i-1};Y_i) + n\epsilon_n \tag{12}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(M,K,Y^{i-1},S^{i-1},X^{i-1};Y_i) + n\epsilon_n \tag{13}$$

$$\overset{(c)}{=} \sum_{i=1}^{n} I(M,K,S^{i-1},X^{i-1};Y_i) + n\epsilon_n \tag{14}$$

$$\overset{(d)}{=} \sum_{i=1}^{n} I(V_i;Y_i) + n\epsilon_n \tag{15}$$

for $\epsilon_n \to 0$ and $V_i := (M,K,S^{i-1})$. Here, (a) follows by applying Fano's inequality from the reliability constraint; (b) and (d) because X^{i-1} is a function of (M,K,S^{i-1}); and (c) since $Y^{i-1} - (M,K,S^{i-1},X^{i-1}) - Y_i$ forms a Markov chain.

Now we utilize the definition and the property of $C_{\mathrm{c}}(A,B)$ to further bound the right-hand side of (15):

$$nR \leq \sum_{i=1}^{n} I(V_i;Y_i) + n\epsilon_n \tag{16}$$

$$\overset{(a)}{\leq} \sum_{i=1}^{n} C_{\mathrm{c}}(D(\widehat{P}_{Z_i}\|Q_0), \mathrm{E}[b(X_i)]) + n\epsilon_n \tag{17}$$

$$\overset{(b)}{\leq} nC_{\mathrm{c}}\left(\frac{1}{n}\sum_{i=1}^{n} D(\widehat{P}_{Z_i}\|Q_0), \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[b(X_i)]\right) + n\epsilon_n \tag{18}$$

where (a) is because X_i is a function of V_i and S_i and the Markov chain $V_i - (X_i,S_i) - (Y_i,Z_i)$ holds and (b) is

due to the concavity of $C_{\mathrm{c}}(A,B)$. Recall from Lemma 5 that $C_{\mathrm{c}}(A,B)$ is non-decreasing in each of A and B. According to the input cost constraint, there exists $\delta_n \to 0$ such that $\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[b(X_i)] \leq B+\delta_n$. On the other hand, from the covertness constraint, there exists $\delta'_n \to 0$ such that $D(\widehat{P}_{Z^n}\|Q_0^{\times n}) \leq \delta'_n$, while as in [4] we have

$$\begin{aligned}
&D(\widehat{P}_{Z^n}\|Q_0^{\times n})\\
&= -H(Z^n) + \mathrm{E}_{\widehat{P}_{Z^n}}\left[\log\frac{1}{Q_0^{\times n}(Z^n)}\right] && (19)\\
&= -\sum_{i=1}^{n} H(Z_i|Z^{i-1}) + \mathrm{E}_{\widehat{P}_{Z^n}}\left[\log\frac{1}{Q_0(Z_i)}\right] && (20)\\
&= -\sum_{i=1}^{n} H(Z_i|Z^{i-1}) + \mathrm{E}_{\widehat{P}_{Z_i}}\left[\log\frac{1}{Q_0(Z_i)}\right] && (21)\\
&\geq -\sum_{i=1}^{n} H(Z_i) + \mathrm{E}_{\widehat{P}_{Z_i}}\left[\log\frac{1}{Q_0(Z_i)}\right] && (22)\\
&= \sum_{i=1}^{n} D(\widehat{P}_{Z_i}\|Q_0). && (23)
\end{aligned}$$

Hence, (18) implies

$$R \leq C_{\mathrm{c}}\left(\frac{\delta'_n}{n}, B+\delta_n\right) + \epsilon_n. \tag{24}$$

Note that the right-hand side of (24) approaches $C_{\mathrm{c}}(0,B)$ as n tends to infinity due to the continuity of $C_{\mathrm{c}}(A,B)$, from which follows the condition $P_Z = Q_0$. Finally, the cardinality bound on $\mathcal{U}$ follows by applying the support lemma [11]. □

Proof of Theorem 3. 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 $\lim_{n\to\infty} D(\widehat{P}_{Z^n}\|Q_0^{\times n}) = 0$.

We start with the proof steps used for channels with noncausal CSI [13] without a covertness constraint:

$$\begin{aligned}
nR &\overset{(a)}{\leq} I(M;Y^n|K) + n\epsilon_n && (25)\\
&= \sum_{i=1}^{n} I(M;Y_i|K,Y^{i-1}) + n\epsilon_n && (26)\\
&\leq \sum_{i=1}^{n} I(M,K,Y^{i-1};Y_i) + n\epsilon_n && (27)\\
&= \sum_{i=1}^{n} I(M,K,Y^{i-1},S_{i+1}^n;Y_i)\\
&\quad - \sum_{i=1}^{n} I(Y_i;S_{i+1}^n|M,K,Y^{i-1}) + n\epsilon_n && (28)\\
&\overset{(b)}{=} \sum_{i=1}^{n} I(M,K,Y^{i-1},S_{i+1}^n;Y_i)\\
&\quad - \sum_{i=1}^{n} I(Y^{i-1};S_i|M,K,S_{i+1}^n) + n\epsilon_n && (29)\\
&\overset{(c)}{=} \sum_{i=1}^{n} I(M,K,Y^{i-1},S_{i+1}^n;Y_i)\\
&\quad - \sum_{i=1}^{n} I(M,K,Y^{i-1},S_{i+1}^n;S_i) + n\epsilon_n && (30)\\
&= \sum_{i=1}^{n} (I(U_i;Y_i) - I(U_i;S_i)) + n\epsilon_n && (31)
\end{aligned}$$

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 $C_{\mathrm{nc}}(A,B)$ to further bound the right-hand side of (31):

$$\begin{aligned}
nR &\leq \sum_{i=1}^{n} (I(U_i;Y_i) - I(U_i;S_i)) + n\epsilon_n && (32)\\
&\leq \sum_{i=1}^{n} C_{\mathrm{nc}}(D(\widehat{P}_{Z_i}\|Q_0), \mathrm{E}[b(X_i)]) + n\epsilon_n && (33)\\
&\overset{(a)}{\leq} nC_{\mathrm{nc}}\left(\frac{1}{n}\sum_{i=1}^{n} D(\widehat{P}_{Z_i}\|Q_0), \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[b(X_i)]\right) + n\epsilon_n && (34)
\end{aligned}$$

where (a) is due to the concavity of $C_{\mathrm{nc}}(A,B)$. Recall from Lemma 5 that $C_{\mathrm{nc}}(A,B)$ is non-decreasing in each of A and B. According to the input cost constraint, there exists $\delta_n \to 0$ such that $\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[b(X_i)] \leq B+\delta_n$. On the other hand, due to the chain of inequalities (19)-(23) from the covertness constraint, there exists $\delta'_n \to 0$ such that $\sum_{i=1}^{n} D(\widehat{P}_{Z_i}\|Q_0) \leq \delta'_n$. Hence, (34) implies

$$R \leq C_{\mathrm{nc}}\left(\frac{\delta'_n}{n}, B+\delta_n\right) + \epsilon_n. \tag{35}$$

Note that the right-hand side of (35) approaches $C_{\mathrm{nc}}(0,B)$ as n tends to infinity due to the continuity of $C_{\mathrm{nc}}(A,B)$, 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 [11]. □

Proof of Lemma 5. Let us show that $C_{\mathrm{nc}}(A,B)$ is non-decreasing in each of A and B, and concave and continuous in (A,B). It can be proved in a similar manner that the same statement holds for $C_{\mathrm{c}}(A,B)$.

First, $C_{\mathrm{nc}}(A,B)$ is non-decreasing in each of A and B since increasing A or B can only enlarge the set of feasible $P_{U|S}P_{X|U,S}$.

Second, to show the concavity, fix arbitrary (A_1,B_1) and (A_2,B_2) and let $P_{U_1|S}P_{X_1|U_1,S}$ and $P_{U_2|S}$ $P_{X_2|U_2,S}$ denote the corresponding conditional PMFs that achieve the maxima of $C_{\mathrm{nc}}(A_1,B_1)$ and $C_{\mathrm{nc}}(A_2,B_2)$, respectively. Let Y_1 and Z_1 (resp. Y_2 and Z_2) denote the channel outputs at the receiver and the warden, respectively, corresponding to $P_{U_1|S}P_{X_1|U_1,S}$ (resp. $P_{U_2|S}P_{X_2|U_2,S}$). Let Q denote a random variable independent of U_1, U_2, and S, which takes value 1 with probability λ and value 2 with probability $1-\lambda$.

Define $U' = (Q,U_Q)$. Let X', Y', and Z' denote the channel input at the transmitter and the channel outputs at the receiver and the warden, respectively, corresponding to $P_{X'|U',S} = P_{X_Q|U_Q,S}$. Note that $P_{X'} = \lambda P_{X_1} + (1-\lambda)P_{X_2}$

and $P_{Z'} = \lambda P_{Z_1} + (1-\lambda) P_{Z_2}$. Then,

$$\begin{aligned}\mathrm{E}[b(X')] &= \sum_{x\in\mathcal{X}} P_{X'}(x) b(x) \\ &= \lambda \sum_{x\in\mathcal{X}} P_{X_1}(x) b(x) + (1-\lambda) \sum_{x\in\mathcal{X}} P_{X_2}(x) b(x) \\ &\le \lambda B_1 + (1-\lambda) B_2 \end{aligned} \tag{36}$$

and

$$\begin{aligned} D(P_{Z'} \| Q_0) &\le \lambda D(P_{Z_1} \| Q_0) + (1-\lambda) D(P_{Z_2} \| Q_0) \\ &\le \lambda A_1 + (1-\lambda) A_2 \end{aligned} \tag{37}$$

because the relative entropy is convex in the first argument. Hence, it follows that

$$\begin{aligned} & C_{\mathrm{nc}}(\lambda(A_1, B_1) + (1-\lambda)(A_2, B_2)) \\ &\ge I(U'; Y') - I(U'; S) && (38) \\ &\ge I(U_Q, Q; Y') - I(U_Q, Q; S) && (39) \\ &\overset{(a)}{\ge} I(U_Q; Y'|Q) - I(U_Q; S|Q) && (40) \\ &= \lambda I(U_1; Y_1) + (1-\lambda) I(U_2; Y_2) \\ &\quad - \lambda I(U_1; S) - (1-\lambda) I(U_2; S) && (41) \\ &= \lambda C_{\mathrm{nc}}(A_1, B_1) + (1-\lambda) C_{\mathrm{nc}}(A_2, B_2), && (42) \end{aligned}$$

where (a) is because Q and S are independent. Hence, we conclude that $C_{\mathrm{nc}}(A, B)$ is concave.

Lastly, $C_{\mathrm{nc}}(A, B)$ is continuous in (A, B) since both the objective function $I(U;Y) - I(U;S)$ and the constraint functions $\mathrm{E}[b(X)]$ and $D(P_Z \| Q_0)$ are continuous in $P_{U|S} P_{X|U,S}$ as long as (2) is satisfied. □

V. Proof of Lower Bounds

In this section, we prove the achievability parts of our main results, i.e., Theorems 2 and 4. To prove the achievability part for the case with causal CSI at the transmitter, we employ the Shannon's strategy [18] and use the soft covering theorem [19, Theorem 4], [20, Corollary VII.4] for the covertness analysis. For the case with noncausal CSI at the transmitter, our scheme is based on multicoding [11], [13], but instead of performing the joint-typicality check to find a codeword that *seemingly* follows a joint distribution with the state sequence, we use likelihood encoding employed in [12] since it admits easier covertness analysis.

Proof of Theorem 2. Fix $\epsilon > 0$. Further fix P_V and $x(v, s)$ 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 a $v^n(k, m)$ according to $\prod_{i=1}^n P_V(v_i)$. These constitute the codebook $\mathcal{C}$.

2) Encoding at the transmitter: Given state sequence s^n, secret key k, and message m, the encoder transmits x^n where $x_i = x(v_i(k, m), s_i)$.

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

$$(v^n(k, \hat{m}), y^n) \in \mathcal{T}_\epsilon^{(n)}. \tag{43}$$

Otherwise it declares an error. Here $\mathcal{T}_\epsilon^{(n)}$ denotes the (strongly) typical set [21].

4) Covertness analysis: By the soft covering theorem [19, Theorem 4], [20, Lemma IV.1], we have $\mathrm{E}_{\mathcal{C}}[D(\widehat{P}_{Z^n} \| Q_0^{\times n})] \overset{n\to\infty}{\longrightarrow} 0$ if[2]

$$R + R_K > I(V; Z). \tag{44}$$

5) Reliability and input cost analysis: By the standard error analysis, it can be shown that the probability of error averaged over the random codebook $\mathcal{C}$ tends to zero as n tends to infinity if

$$R < I(V; Y). \tag{45}$$

Furthermore, by the typical average lemma [11],

$$\begin{aligned} & \mathrm{E}_{\mathcal{C},M,K,S^n}[b(X^n)] \\ &= P(X^n \notin T_\epsilon^{(n)}) \, \mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n) | X^n \notin T_\epsilon^{(n)} \right] \\ &\quad + P(X^n \in T_\epsilon^{(n)}) \, \mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n) | X^n \in T_\epsilon^{(n)} \right] && (46) \\ &\le P(X^n \notin T_\epsilon^{(n)}) B_{\max} + B, && (47) \end{aligned}$$

where $B_{\max} := \max_{x\in\mathcal{X}} b(x)$. Note that $P(X^n \notin T_\epsilon^{(n)}) \to 0$ as n tends to infinity. Hence, we have

$$\limsup_{n\to\infty} \mathrm{E}_{\mathcal{C},M,K,S^n}[b(X^n)] \le B. \tag{48}$$

In summary, if (44) and (45) are satisfied, then there must exist a sequence of codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$, $\lim_{n\to\infty} \mathrm{E}_{M,K,S^n}[b(X^n)] \le B$, and $\lim_{n\to\infty} D(P_{Z^n} \| Q_0^{\times n}) = 0$. By applying the Fourier-Mozkin elimination [11] to (44) and (45), we complete the proof. □

Proof of Theorem 4. 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

$$g(l | s^n, k, m) = \frac{P_{S|U}^{\times n}(s^n | u^n(k, m, l))}{\sum_{l' \in [1:2^{nR'}]} P_{S|U}^{\times n}(s^n | u^n(k, m, l'))}. \tag{49}$$

The encoder randomly generates l according to (49) 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

$$(u^n(k, \hat{m}, l), y^n) \in \mathcal{T}_\epsilon^{(n)} \tag{50}$$

for some $l \in [1 : 2^{nR'}]$; if no such unique $\hat{m}$ can be found, it declares an error.

4) Covertness analysis: For covertness analysis, we use the following lemma, which is proven at the end of this section.

[2]The soft covering theorem [19, Theorem 4], [20, Lemma IV.1] states that if we generate $2^{n\tilde{R}}$ codewords from $\mathcal{V}^n$ in an i.i.d. manner according to P_V, then the induced output distribution $\widehat{P}_{Z^n}$ by sending a uniformly chosen codeword through a memoryless channel $P_{Z|V}$ and the desired output distribution $Q_Z^{\times n} \sim \prod_{i=1}^n Q_Z$, where $Q_Z = \sum_V P_V P_{Z|V}$, satisfy $\mathrm{E} \| \widehat{P}_{Z^n} - Q_Z^{\times n} \| \overset{n\to\infty}{\longrightarrow} 0$ provided that $\tilde{R} > I(V; Z)$. Since the encoding procedure in Section V-2 induces a memoryless channel from V to Z, we can apply the soft covering theorem. Note that $\tilde{R}$ and Q_Z in the statement of soft covering theorem correspond to $R + R_K$ and Q_0 in our coding scheme, respectively, and $\mathrm{E} \| \widehat{P}_{Z^n} - Q_0^{\times n} \| \overset{n\to\infty}{\longrightarrow} 0$ implies $\mathrm{E}[D(\widehat{P}_{Z^n} \| Q_0^{\times n})] \overset{n\to\infty}{\longrightarrow} 0$.

Lemma 6. *For the codebook generation and encoding procedure described above, if $R' > I(U;S)$ and $R + R_K + R' > I(U;Z)$, then*

$$\mathrm{E}_{\mathcal{C}}\left[D(\widehat{P}_{Z^n}\|P_Z^{\times n})\right] \overset{n\to\infty}{\longrightarrow} 0. \tag{51}$$

Now, let

$$R' > I(U;S) \tag{52}$$

$$R + R_K + R' > I(U;Z). \tag{53}$$

Because $P_{U|S}$ and $x(u,s)$ are chosen to satisfy $P_Z = Q_0$, Lemma 6 implies that

$$\mathrm{E}_{\mathcal{C}}[D(\widehat{P}_{Z^n}\|Q_0^{\times n})] \overset{n\to\infty}{\longrightarrow} 0. \tag{54}$$

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 (49) at the encoder. The error event $\{\hat{M} \neq M\}$ occurs only if at least one of the following events occurs:

$$\mathcal{E}_1 := \{(U^n(K,M,L),S^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\} \tag{55}$$

$$\mathcal{E}_2 := \{(U^n(K,M,L),Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\} \tag{56}$$

$$\mathcal{E}_3 := \{(U^n(K,m,l),Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m \neq M \text{ and } l \in [1:2^{nR'}]\}. \tag{57}$$

Hence, the probability of error is bounded as

$$P(\hat{M} \neq M) \le P(\mathcal{E}_1) + P(\mathcal{E}_1^c \cap \mathcal{E}_2) + P(\mathcal{E}_3). \tag{58}$$

Now we bound each term on the right-hand side of (58). The first term $P(\mathcal{E}_1)$ tends to zero as n tends to infinity due to [12, Lemma 2], as long as (52) is satisfied. Next, note that

$$\mathcal{E}_1^c = \{(U^n(K,M,L),S^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}. \tag{59}$$

By the conditional typicality lemma [11], $P(\mathcal{E}_1^c \cap \mathcal{E}_2)$ 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 [11] provided

$$R + R' < I(U;Y). \tag{60}$$

In summary, the probability of error averaged over the random codebook $\mathcal{C}$ tends to zero as n tends to infinity if (52), (53), and (60) are satisfied.

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

$$P(\mathcal{E}_1) = P\{(U^n(K,M,L),S^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\} \tag{61}$$

$$= P((U^n(K,M,L),X^n,S^n) \notin \mathcal{T}_{\epsilon'}^{(n)}) \overset{n\to\infty}{\longrightarrow} 0. \tag{62}$$

Note that if $x^n \in \mathcal{T}_{\epsilon'}^{(n)}$, then $b(x^n) \le B$ by the typical average lemma [11]. Hence,

$$\begin{aligned}&\mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)\right]\\ &= P(\mathcal{E}_1)\,\mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)|\mathcal{E}_1\right]\\ &\quad + P(\mathcal{E}_1^c)\,\mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)|\mathcal{E}_1^c\right] \qquad (63)\\ &\le P(\mathcal{E}_1)B_{\max} + P(\mathcal{E}_1^c)B, \qquad (64)\end{aligned}$$

where $B_{\max} := \max_{x\in\mathcal{X}} b(x)$. By (62), the right-hand side of (64) approaches B as n tends to infinity. Hence, we have

$$\limsup_{n\to\infty} \mathrm{E}_{\mathcal{C},M,K,S^n}\left[b(X^n)\right] \le B. \tag{65}$$

In summary, if (52), (53), and (60) 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 [11] to (52), (53), and (60), we complete the proof. □

Remark 4. *The wiretap channels with noncausal CSI at the transmitter was first studied by Chen and Vinck [22]. For the secrecy metric of fractional equivocation, they proposed an achievability scheme that combines wiretap coding with Gelfand-Pinsker coding [13] that incorporates joint typicality encoding at the transmitter. For the covertness constraint considered in the current paper, likelihood encoding admits easier analysis as it chooses the codeword in a fully probabilistic manner so that the behavior of $\widehat{P}_{Z^n}$ can be analyzed more easily in the covertness analysis (Lemma 6).*

Remark 5. *Our scheme for the case with noncausal CSI at the transmitter is similar to that in [23] for wiretap channels with noncausal CSI at the transmitter under the semantic-security metric requiring negligible information leakage for all message distributions. The coding scheme in [23] incorporates superposition coding; the inner codebook is for a random index and the outer codebook is for the message and another random index. To compare that scheme with ours, let us consider the special case of the scheme in [23] where the rate of the inner codebook is set to zero and let R and R' denote the rates of the message and the random index in the outer codebook. For our scheme, let us consider the special case of $R_K = 0$ and sufficiently large B, i.e., no secret key and no input cost constraint. Then the codebook generation and the encoding procedure of the scheme [23] become the same as our scheme. As shown in the proof of Theorem 4, reliability at the receiver is satisfied if $R' > I(U;S)$ and $R + R' < I(U;Y)$, but covertness requires $R + R' > I(U;Z)$ and $P_Z = Q_0$ while semantic security requires $R' > I(U;Z)$.*

Proof of Lemma 6: The proof follows similar lines to [12, Section VII-A]. As in [12, Section VII-A], it can be checked that, to prove (51), it suffices to show that the total variation (TV) distance approaches zero:

$$\mathrm{E}_{\mathcal{C}}\,\|\widehat{P}_{Z^n} - P_Z^{\times n}\|_{\mathrm{TV}} \overset{n\to\infty}{\longrightarrow} 0. \tag{66}$$

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

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

Let $\Gamma^{(\mathcal{C})}_{S^n,Z^n}$ and $\Gamma^{(\mathcal{C})}_{Z^n}$ denote the marginals of $\Gamma^{(\mathcal{C})}_{K,M,L,U^n,S^n,Z^n}$ with respect to (S^n,Z^n) and Z^n, respectively. Using the triangle inequality for the TV distance, we upper-bound the left-hand side of (66) as

$$\begin{aligned}&\mathrm{E}_{\mathcal{C}}\,\|\widehat{P}_{Z^n} - P_Z^{\times n}\|_{\mathrm{TV}}\\ &\le \mathrm{E}_{\mathcal{C}}\,\|\widehat{P}_{Z^n} - \Gamma^{(\mathcal{C})}_{Z^n}\|_{\mathrm{TV}} + \mathrm{E}_{\mathcal{C}}\,\|\Gamma^{(\mathcal{C})}_{Z^n} - P_Z^{\times n}\|_{\mathrm{TV}}. \qquad (67)\end{aligned}$$

From the soft covering theorem [19, Theorem 4], [20, Corollary VII.4], the second term on the right-hand side of (67) 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 (67), note that

$$\mathrm{E}_{\mathcal{C}}\,\|\widehat{P}_{Z^n} - \Gamma^{(\mathcal{C})}_{Z^n}\|_{\mathrm{TV}} \le \mathrm{E}_{\mathcal{C}}\,\|\widehat{P}_{S^n,Z^n} - \Gamma^{(\mathcal{C})}_{S^n,Z^n}\|_{\mathrm{TV}}. \tag{68}$$

By applying the same analysis as in [12, Section VII-A], the right-hand side of (68) decays to zero as $n \to \infty$ if $R' > I(U;S)$. ■

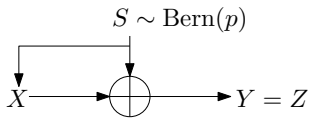

Figure 2. Binary symmetric channel with CSI at the transmitter

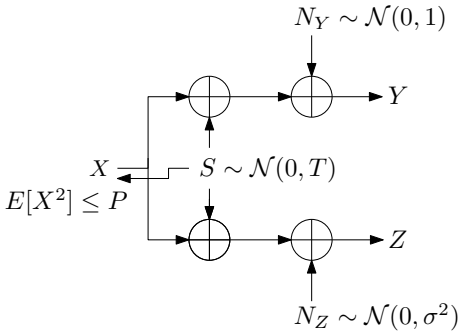

Figure 3. AWGN channel with CSI at the transmitter

VI. Examples

In this section, we show two examples where the covert capacity of a channel is zero in the absence of CSI at the transmitter, but is positive with CSI.

A. The Binary Symmetric Channel

Consider a channel in Fig. 2 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

$$Y = Z = X \oplus S. \tag{69}$$

Assume that $x_0 = 0$ and $R_K > 0$.

Using Theorems 1 and 2 one can check that, with causal CSI, the optimal choice is $V = Y = Z$ having the Bernoulli distribution of parameter p. This gives

$$C_{\text{c}} = H_{\text{b}}(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}. \tag{70}$$

Furthermore, it can be checked that $C_{\text{nc}} = C_{\text{c}}$. Note that, without CSI, covert communication cannot have a positive rate [3], [4] on this channel.

B. The AWGN Channel

Consider an AWGN channel in Fig. 3 where the channel outputs at the receiver and the warden are given as

$$Y = X + S + N_Y \tag{71}$$

$$Z = X + S + N_Z, \tag{72}$$

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 causally or 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 $\text{E}[X^2] \le P$. 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 and denoted by C_{c} and C_{nc} for causal and noncausal CSI cases, respectively.

The following theorems show that the covert capacity can be positive for the AWGN channel both with causal CSI and with noncausal CSI at the transmitter. In the following, we define

$$\gamma^* := \min\left\{1, \frac{P}{2T}\right\} \tag{73a}$$

$$T^* := (1-\gamma^*)^2 T \tag{73b}$$

$$P^* := T - T^*. \tag{73c}$$

Theorem 7. *If*

$$R_K > \frac{1}{2}\log\left(1 + \frac{P^*}{T^*+\sigma^2}\right) - \frac{1}{2}\log\left(1 + \frac{P^*}{T^*+1}\right), \tag{74}$$

the covert capacity with causal CSI at the transmitter is lower-bounded as

$$C_{\text{c}} \ge \frac{1}{2}\log\left(1 + \frac{P^*}{T^*+1}\right). \tag{75}$$

Theorem 8. *If the secret key rate satisfies* (76)*, the covert capacity is given by*

$$C_{\text{nc}} = \frac{1}{2}\log\left(1 + P^*\right). \tag{77}$$

Remark 6. *If the warden's channel is degraded, i.e., $\sigma^2 > 1$, a secret key is not needed to achieve the rates* (75) *and* (77) *for the cases with causal CSI and with noncausal CSI at the transmitter, respectively.*

Remark 7. *Let us assume that R_K is sufficiently large so that* (74) *and* (76) *are satisfied. If $T^* = 0$, i.e., $T \le \frac{P}{2}$, then $C_{\text{c}} = C_{\text{nc}}$. On the other hand, if $T \to \infty$, it follows $P^* \to P$. Then, C_{nc} approaches $\frac{1}{2}\log(1+P)$, which is the capacity of the channel* (71) *with noncausal CSI at the transmitter and without a covertness constraint.*

We prove Theorems 7 and 8 by adapting our DMC results in Theorems 2, 3, and 4. In the achievability proofs of Theorems 7 and 8, we reduce the interference power to make room for message transmission. We set the channel input to have the form of $X = X^* - \gamma^* S$ where X^* is independent of S, so that $\gamma^* S$ is subtracted from S when X is sent. Then, we regard X^* as the input for the channel with reduced interference power of $(1-\gamma^*)^2 T$, i.e., T^*. To satisfy the covertness constraint, X^* must have power $T - T^* = P^*$. Note that the choice of γ^* in (73a) ensures that the power constraint of X is satisfied, i.e.,

$$\text{E}[X^2] = \text{E}[X^{*2}] + \gamma^{*2} T \tag{78}$$

$$= T - (1-\gamma^*)^2 T + \gamma^{*2} T \tag{79}$$

$$= 2\gamma^* T \tag{80}$$

$$\le P. \tag{81}$$

For the case with causal CSI, the right-hand side of (75) is achieved by letting $V = X^*$ and treating interference as noise at the receiver. For the case with noncausal CSI, the right-hand side of (77) is achieved by choosing U as in "dirty paper coding" [24].

In the following we first prove Theorem 8.

Achievability proof of Theorem 8. We modify the proof of Theorem 4 so that it applies to the Gaussian case with a

$$R_K > \frac{1}{2}\log\left(1+\frac{(P^*+\frac{P^*}{P^*+1}T^*)^2}{(P^*+(\frac{P^*}{P^*+1})^2T^*)(P^*+T^*+\sigma^2)-(P^*+\frac{P^*}{P^*+1}T^*)^2}\right) - \frac{1}{2}\log\left(1+\frac{(P^*+\frac{P^*}{P^*+1}T^*)^2}{(P^*+(\frac{P^*}{P^*+1})^2T^*)(P^*+T^*+1)-(P^*+\frac{P^*}{P^*+1}T^*)^2}\right), \tag{76}$$

power constraint. Roughly speaking, our idea is to "quantize" at the decoder but not at the encoder. We choose a conditional probability density function (PDF) of U given S and a mapping from (U,S) to X via the following:

$$X^* \sim \mathcal{N}(0,P^*), \text{ independent of } S \tag{82}$$

$$U = X^* + \frac{P^*}{P^*+1}(1-\gamma^*)S \tag{83}$$

$$X = U - \frac{P^*+\gamma^*}{P^*+1}S = X^* - \gamma^* S. \tag{84}$$

We then employ the same encoding procedure as in Theorem 4, except that PMFs are now replaced by PDFs. Next, as an additional step for the encoder, we fix some small positive ϵ and check whether the resulting input sequence x^n satisfies the power constraint

$$\sum_{i=1}^{n} x_i^2 \le n(P+\epsilon) \tag{85}$$

or not. Denote by $\mathcal{E}_4$ the event that (85) is *not* satisfied. When $\mathcal{E}_4$ occurs, we replace x^n by the all-zero sequence. By similar analysis as in [12] one can show that the probability of $\mathcal{E}_4$ tends to zero as n tends to infinity for all positive ϵ. Clearly, in the limit where ϵ approaches zero, our encoding scheme above satisfies the given power constraint.

For covertness analysis, we adapt the proof for the DMC case as follows. Let $\widehat{P}_{Z^n}$ denote the distribution at the warden generated by the above coding scheme, and let $\bar{P}_{Z^n}$ denote the distribution generated by this scheme but *without the additional step* of replacing those codewords not satisfying (85) with the all-zero sequence. It is clear from the proof of Lemma 6 that it can be applied to PDFs without a maximum-cost constraint, so we can write, similarly to (66), that

$$\mathrm{E}_{\mathcal{C}} \|\bar{P}_{Z^n} - Q_0^{\times n}\|_{\mathrm{TV}} \overset{n\to\infty}{\longrightarrow} 0. \tag{86}$$

Next fix a codebook $\mathcal{C}$ and let P_1 denote the distribution resulting from conditioning the corresponding $\bar{P}_{Z^n}$ on the event $\bar{\mathcal{E}}_4$, and P_2 that on $\mathcal{E}_4$. Then $\bar{P}_{Z^n} = (1-P(\mathcal{E}_4|\mathcal{C}))P_1 + P(\mathcal{E}_4|\mathcal{C})P_2$ and $\widehat{P}_{Z^n} = (1-P(\mathcal{E}_4|\mathcal{C}))P_1 + P(\mathcal{E}_4|\mathcal{C})Q_0^{\times n}$. We have

$$\|\widehat{P}_{Z^n} - Q_0^{\times n}\|_{\mathrm{TV}} = (1-P(\mathcal{E}_4|\mathcal{C}))\|P_1 - Q_0^{\times n}\|_{\mathrm{TV}} \tag{87}$$

$$\le \|P_1 - Q_0^{\times n}\|_{\mathrm{TV}}. \tag{88}$$

On the other hand

$$\|\bar{P}_{Z^n} - Q_0^{\times n}\|_{\mathrm{TV}} \ge (1-P(\mathcal{E}_4|\mathcal{C}))\|P_1 - Q_0^{\times n}\|_{\mathrm{TV}} - P(\mathcal{E}_4|\mathcal{C})\|P_2 - Q_0^{\times n}\|_{\mathrm{TV}} \tag{89}$$

$$\ge \|P_1 - Q_0^{\times n}\|_{\mathrm{TV}} - 2P(\mathcal{E}_4|\mathcal{C}). \tag{90}$$

Combining (88) and (90) we obtain

$$\mathrm{E}_{\mathcal{C}} \|\widehat{P}_{Z^n} - Q_0^{\times n}\|_{\mathrm{TV}} \le \mathrm{E}_{\mathcal{C}} \|\bar{P}_{Z^n} - Q_0^{\times n}\|_{\mathrm{TV}} + 2P(\mathcal{E}_4). \tag{91}$$

By (86), (91), and the fact that $P(\mathcal{E}_4)$ tends to zero as $n \to \infty$, we know that the left-hand side of (91) tends to zero as $n \to \infty$. Because $\widehat{P}_{Z^n}$ is absolutely continuous with respect to $Q_0^{\times n}$, this further implies that (see [12, Section VII-A])

$$D\left(\widehat{P}_{Z^n}\middle\|Q_0^{\times n}\right) \overset{n\to\infty}{\longrightarrow} 0. \tag{92}$$

We next describe the decoder and analyze its probability of making an error. To this end, we first quantize the random variables S, U, and Y. A partition $\mathcal{P}$ 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

$$P\left([U]_{\mathcal{P}} = \begin{cases}\sup P_i & \text{if } \sup P_i < \infty\\ \inf P_i & \text{otherwise}\end{cases}\right) = P(U \in P_i). \tag{93}$$

Similarly, S (resp. Y) is quantized by partition $\tilde{\mathcal{P}}$ (resp. $\mathcal{P}'$) and its quantization is denoted by $[S]_{\tilde{\mathcal{P}}}$ (resp. $[Y]_{\mathcal{P}'}$). The decoder considers the above quantizations of the received sequence y^n and every u^n in the codebook, and performs typicality decoding as in the proof of Theorem 4. The event $\mathcal{E}_4$ defined above, which has vanishing probability as $n \to \infty$, can be taken into account as an additional error event. Then the conditions (52) and (60) become

$$R' > I([U]_{\mathcal{P}}; [S]_{\tilde{\mathcal{P}}}) \tag{94}$$

$$R + R' < I([U]_{\mathcal{P}}; [Y]_{\mathcal{P}'}). \tag{95}$$

As we refine the partitions $\mathcal{P}$, $\tilde{\mathcal{P}}$, and $\mathcal{P}'$, $I([U]_{\mathcal{P}}; [S]_{\tilde{\mathcal{P}}})$ approaches $I(U;S)$ and $I([U]_{\mathcal{P}}; [Y]_{\mathcal{P}'})$ approaches $I(U;Y)$ according to [25, Section 8.6].

We thus conclude that our coding scheme will succeed if (7) and (8) hold for the chosen PDFs. Computing these expressions explicitly completes the achievability proof of Theorem 8. □

Converse proof of Theorem 8. First, by examining the proof of Theorem 3, we see that it also applies to the Gaussian channel. Fix the conditional distribution $P_{U|S}$ and the mapping $x(u,s)$ that achieve the maximum in (6). Recall that they satisfy $P_Z = Q_0$ and $\mathrm{E}[X^2] \le P$. Let $\tilde{P} := \mathrm{Var}(X)$ and $\Lambda := \mathrm{E}[XS]$. It follows that

$$I(U;Y) - I(U;S) \le I(U;Y,S) - I(U;S) \tag{96}$$

$$= I(U;Y|S) \tag{97}$$

$$\le I(X,U;Y|S) \tag{98}$$

$$\overset{(a)}{=} I(X;Y|S) \tag{99}$$

$$= h(X+N_Y|S) - h(N_Y) \tag{100}$$

$$\overset{(b)}{\le} \frac{1}{2}\log\left(1+\tilde{P}-\frac{\Lambda^2}{T}\right), \tag{101}$$

where (a) is due to the Markov chain $U-(X,S)-Y$ and (b) is from [11, Problem 2.7]. Recall the condition $P_Z = Q_0$, which implies

$$T+\sigma^2 = \mathrm{Var}(X+S+N_Z) \tag{102}$$

$$= \tilde{P} + T + 2\Lambda + \sigma^2, \tag{103}$$

therefore we must have $\Lambda = -\frac{\tilde{P}}{2}$. Hence (101) implies

$$I(U;Y) - I(U;S) \le \frac{1}{2}\log\left(1+\tilde{P}-\frac{\tilde{P}^2}{4T}\right). \tag{104}$$

Note that $\tilde{P} \leq P$ and

$$\arg\max_{0\leq\tilde{P}\leq P}\left(\tilde{P}-\frac{\tilde{P}^2}{4T}\right)=\min\{P,2T\}. \quad (105)$$

Thus, we have

$$C \leq \frac{1}{2}\log\left(1+\min\{P,2T\}-\frac{(\min\{P,2T\})^2}{4T}\right), \quad (106)$$

which concludes the proof. □

Proof of Theorem 7. We can adapt Theorem 2 to the Gaussian case with a power constraint through a quantization argument that is similar to the one in the achievability proof of Theorem 8. By letting $V \sim \mathcal{N}(0, P^*)$ and $X = V - \gamma^* S$ in Theorem 2, Theorem 7 is proved. □

VII. Concluding Remarks

We have shown that causal and noncausal CSI at the transmitter can sometimes help it to communicate covertly at a postive rate over channels which, without CSI, obey the "square-root law" for covert communications. Computable single-letter formulas for the maximum achievable covert-communication rate (assuming that a sufficiently long key is available) have been derived. This work, from a different perspective to that of recent works [6]–[8], shows that channel statistics unknown to the warden can help the communicating parties to communicate covertly.

There are many channels over which, even with the help of CSI, covert communication cannot have a positive rate (Remark 3 contains simple examples). For some of these channels, CSI may help to improve the scaling constant of the maximum amount of information that can be covertly communicated with respect to the square root of the total number of channel uses. We have not investigated this possibility in the current paper.

So far, we have not been able to prove upper bounds on the minimum secret-key length required to achieve the covert capacity that match the lower bounds (5) and (8), except when the warden has a weaker channel than the intended receiver, in which case this length is zero. This key-length problem may be related to the secrecy capacity of the wiretap channel with causal or noncausal CSI at the transmitter [23], [26], which, to the best of our knowledge, is not yet completely solved.

References

[1] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with noncausal channel-state information at the transmitter," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2830–2834.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delays," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3413–3431, Oct 2007.

[10] H. Chang, S. Y. Chung, and S. Kim, "Interference channel with a causal relay under strong and very strong interference," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 859–865, Feb 2014.

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

[12] 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.

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

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

[15] 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.

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

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

[18] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, 1958.

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

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

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

[22] Y. Chen and A. J. H. Vinck, "Wiretap channel with side information," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 395–402, Jan 2008.

[23] 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.

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

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

[26] Y. K. Chia and A. E. Gamal, "Wiretap channel with causal state information," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2838–2849, May 2012.

Si-Hyeon Lee (S'08-M'13) is an assistant professor in the Department of Electrical Engineering at the Pohang University of Science and Technology (POSTECH), Pohang, South Korea. She received the B.S. (summa cum laude) and Ph.D. degrees in Electrical Engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2007 and 2013, respectively. From 2014 to 2016, she was a Postdoctoral Fellow in the Department of Electrical and Computer Engineering at the University of Toronto, Toronto, Canada. Her research interests include network information theory, physical layer security, and wireless communication systems.

Ligong Wang (S'08-M'12) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004, and the M.Sc. and Dr.Sc. degrees in electrical engineering from ETH Zurich, Switzerland, in 2006 and 2011, respectively. In the years 2011–2014 he was a Postdoctoral Associate at the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, MA, USA. He is now a researcher (chargé de recherche) with CNRS, France, and is affiliated with ETIS laboratory in Cergy-Pontoise. His research interests include classical and quantum information theory, physical-layer security, and digital, in particular optical communications.

Ashish Khisti (S'02-M'08) received his BASc Degree (2002) in Engineering Sciences (Electrical Option) from University of Toronto, and his S.M and Ph.D. Degrees in Electrical Engineering from the Massachusetts Institute of Technology. Between 2009-2015, he was an assistant professor in the Electrical and Computer Engineering department at the University of Toronto. He is presently an associate professor, and holds a Canada Research Chair in the same department. He is a recipient of an Ontario Early Researcher Award, the Hewlett-Packard Innovation Research Award and the Harold H. Hazen teaching assistant award from MIT. He presently serves as an associate editor for IEEE Transactions on Information Theory and is also a guest editor for the Proceedings of the IEEE (Special Issue on Secure Communications via Physical-Layer and Information-Theoretic Techniques).

Gregory W. Wornell (S'83-M'91-SM'00-F'04) received the B.A.Sc. degree from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering and computer science, in 1985, 1987 and 1991, respectively.

Since 1991 he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the Department of Electrical Engineering and Computer Science. At MIT he leads the Signals, Information, and Algorithms Laboratory within the Research Laboratory of Electronics. He is also chair of Graduate Area I (information and system science, electronic and photonic systems, physical science and nanotechnology, and bioelectrical science and engineering) within the EECS department's doctoral program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, the University of California, Berkeley, CA, and Hewlett-Packard Laboratories, Palo Alto, CA.

His research interests and publications span the areas of signal processing, information theory, digital communication, statistical inference, and information security, and include architectures for sensing, learning, computing, communication, and storage; systems for computational imaging, vision, and perception; aspects of computational biology and neuroscience; and the design of wireless networks. He has been involved in the Information Theory and Signal Processing societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching.