Keyless Covert Communication in the Presence of Non-causal Channel State Information

Hassan ZivariFard[†], Matthieu Bloch[††] and Aria Nosratinia[†]
[†] The University of Texas at Dallas, Richardson, TX, USA, Email: {hassan, aria}@utdallas.edu
[††] Georgia Institute of Technology, Atlanta, GA, USA, Email: {matthieu.bloch}@ece.gatech.edu

***Abstract*—We consider the problem of covert communication over a state-dependent channel, for which the transmitter and the legitimate receiver have non-causal access to the channel state information. Covert communication with respect to an adversary, referred to as the "warden," is one in which the distribution induced during communication at the channel output observed by the warden is identical to the output distribution conditioned on an inactive channel-input symbol. Covert communication involves fooling an adversary in part by a proliferation of codebooks; for reliable decoding at the legitimate receiver the codebook uncertainty is removed via a shared secret key that is unavailable to the warden. Unlike earlier work in state-dependent covert communication, we do not assume the availability of a shared key at the transmitter and legitimate receiver. Rather, a shared randomness is extracted at the transmitter and the receiver from the channel state, in a manner that keeps the shared randomness secret from the warden despite the influence of the channel state on the warden's output. An inner bound on the covert capacity, in the absence of an externally provided secret key, is derived.**

I. Introduction

Covert communication refers to scenarios in which reliable communication over a channel must occur while simultaneously ensuring that a separate channel output at a node called the warden has a distribution identical to that induced by an inactive channel symbol [1]–[4]. For discrete memoryless channels (DMC), the inactive input symbol is denoted x_0. It is known that in a point-to-point DMC without state, if the output distribution (at the warden) induced by x_0 is a convex combination of the output distributions generated by the other input symbols, then it is possible to achieve a positive rate; otherwise the number of possible bits that can be reliably and covertly communicated over n channel transmissions scales at most as $O(\sqrt{n})$. This result has motivated the study of other models in which positive rates are achievable.

Of particular relevance to the present work, Lee et al. [5] have considered the problem of covert communication over a state dependent channel in which the channel state is known to the transmitter but unknown to the receiver and the warden. The authors derived the covert capacity when the transmitter and the receiver share a sufficiently long secret key and also derived a lower bound on the minimum length of the secret key needed to achieve the covert capacity. Given the presence of a channel state, one can wonder if covert communication with positive rate is still possible *without* requiring an external secret key. In particular, several works demonstrate the benefits of exploiting common randomness and channel states to generate secret keys. For instance, the problem of stealth secret key generation from correlated sources has been studied in [6], [7] and covert secret key generation has been studied in [8], [9]. Most importantly, the usefulness of exploiting states for secrecy has been extensively investigated. The discrete memoryless wiretap channel with random states was first studied by Chen and Han Vinck [10], who studied a scenario in which the CSI is available only at the encoder. They established a lower bound on the secrecy-capacity based on a combination of wiretap coding with Gelfand-Pinsker coding. Generally speaking, the coding scheme with Channel State Information (CSI) outperforms the one without CSI, because perfect knowledge of the CSI not only enables the transmitter to beamform its signal toward the legitimate receiver but also provides a source of common randomness from which to generate a common secret key and enhance the secrecy capacity. Khisti et al. [11] studied the problem of secret key generation from the channel state and established the secret key capacity. Chia and El Gamal studied the state-dependent wiretap channel with weak secrecy, proposing a scheme in which the transmitter and the receiver extract a key from the state, and protect the confidential message via a one-time-pad driven with the extracted key [12] (see also [13] and [14]). Han and Sasaki [15] subsequently extended the coding scheme to achieve strong secrecy. Goldfeld et al. [16] proposed a superposition coding scheme for the problem of transmitting a semantically secure message over a state dependent channel while the channel state is available non-causally at the transmitter.

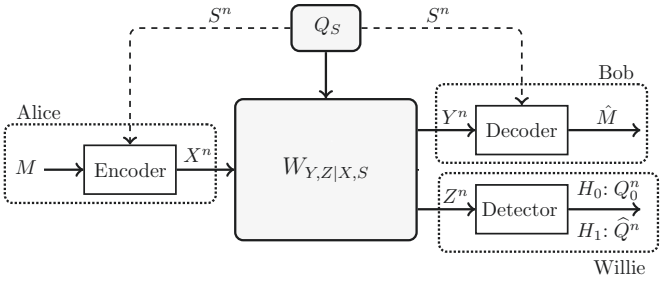

Fig. 1. Model of Covert Communication Over State Dependent DMC

We study here the problem of covert communications over

The work of H. ZivariFard and A. Nosratinia is supported by National Science Foundation (NSF) grant 1514050. The work of M. R. Bloch is supported by National Science Foundation (NSF) grant 1527387.

a state dependent discrete memoryless channel with channel state available non-causally at both the transmitter and the receiver (see Fig. 1). The main innovation of this work is to show that the channel state can be used to simultaneously and efficiently accomplish two necessary tasks: using the channel state for a Gelfand-Pinsker style encoding to assist covert communication while also extracting shared randomness at the two legitimate terminals that is kept secret from the warden, to resolve the multiple codebooks that are necessary for covert communication. The shared randomness extracted from the state effectively takes the place of the external secret key in other models, thus generalizes and expands the applicability of covert communication. We derive an inner bound on the covert capacity of this problem, as well as a condition that shows when this scheme results in a positive rate.

II. Preliminaries

Throughout this paper, random variables are denoted by capital letters and their realizations by lower case letters. The set of $\epsilon-$strongly jointly typical sequences of length n, according to $P_{X,Y}$, is denoted by $\mathcal{T}_\epsilon^{(n)}(P_{X,Y})$. For convenience, whenever there is no danger of confusion, typicality will reference the random variables rather than the distribution, e.g., we write $\mathcal{T}_\epsilon^{(n)}(X,Y)$ or $\mathcal{T}_\epsilon^{(n)}(X|Y)$. Superscripts denote the dimension of a vector, e.g., X^n. The integer set $\{1,\dots,M\}$ is denoted by $[\![1,M]\!]$, and $X_{[i:j]}$ indicates the set $\{X_i, X_{i+1},\dots,X_j\}$. The cardinality of a set is denoted by $|\cdot|$. The total variation between probability mass functions (pmfs) is defined as $||q-p||_1 = \frac{1}{2}\sum_x |p-q|$ and the Kullback-Leibler (KL) divergence between pmfs is defined as $\mathbb{D}(P||Q) = \sum_x p(x)\log\frac{p(x)}{q(x)}$. The support of a probability distribution P is denoted by $\text{supp}(P)$. The n-fold product distribution constructed from the same distribution P is denoted P^n. $P_X \approx Q_X$ indicates $\mathbb{D}(P_X||Q_X) < \epsilon$ (or $\|P_X - Q_X\|_1 < \epsilon$).

Consider a discrete memoryless state-dependent channel as shown in Fig. 1. The finite sets $(\mathcal{X},\mathcal{S},\mathcal{Y},\mathcal{Z})$ and the transition probability distribution $W_{Y,Z|X,S}$ are the constitutive components of this channel. Here, $\mathcal{X}$ is the channel input, $\mathcal{Y}$ and $\mathcal{Z}$ are the channel outputs at the legitimate receiver and the warden, respectively. All of these alphabets are finite. Let $x_0 \in \mathcal{X}$ be a symbol corresponding to the case in which the transmitter is not communicating with the receiver. We assume that the channel state is independent and identically distributed (i.i.d.) and drawn according to P_S. Define $Q_0(\cdot) = \sum_{s\in\mathcal{S}} Q_S(s)W_{Z|X,S}(\cdot|x_0,s)$ and let $Q_0^n = \prod_{i=1}^n Q_0$. We define a non-negative cost $g(x)$ for every channel input $x \in \mathcal{X}$. The average cost of an input sequence $x^n \in \mathcal{X}^n$ is $g(x^n) = \frac{1}{n}\sum_{i=1}^n g(x_i)$. The channel state is available non-causally at both the transmitter and the receiver (see Fig. 1) while the warden does not have access to it. An $(|\mathcal{M}|,n)$ code consists of an encoder that maps (M,S^n) to $X^n \in \mathcal{X}^n$, and a decoder at the receiver that maps Y^n to $\hat{M} \in \mathcal{M}$. The transmitter and the receiver want to design a code that is both reliable and covert. The code is defined to be reliable if the probability of error $P_e^{(n)} = P(\hat{M} \neq M)$ goes to zero when $n \to \infty$. The code is covert if the warden cannot determine whether communication is happening (hypothesis H_1) or not (hypothesis H_0). The probabilities of false alarm (accepting H_1 when the transmitter is not sending a meaningful information) and missed detection (accepting H_0 when the transmitter is sending a meaningful information), are denoted by α and β, respectively. We know that a blind test without looking at the channel output satisfies $\alpha+\beta=1$. If P_{Z^n} denotes the distribution induced at the warden's output when the transmitter sends a message, the optimal hypothesis test by the warden satisfies $\alpha+\beta \geq 1-\sqrt{\mathbb{D}(P_{Z^n}||Q_0^n)}$ [17]. Therefore, to show that the communication is covert it is sufficient to show that $\mathbb{D}(P_{Z^n}||Q_0^n) \to 0$. Note that $\text{supp}(Q_0) = \mathcal{Z}$ otherwise $\mathbb{D}(P_{Z^n}||Q_0^n) \to \infty$. Consequently, our goal is to design a sequence of $(2^{nR},n)$ codes such that

$$\lim_{n\to\infty} P_e^{(n)} \to 0, \tag{1}$$

$$\lim_{n\to\infty} \mathbb{D}(P_{Z^n}||Q_0^n) \to 0, \tag{2}$$

$$\lim_{n\to\infty} \mathbb{E}[g(X^n)] \leq G. \tag{3}$$

We define the covert capacity as the supremum of all achievable covert rates and denote it by C_{NC}.

III. Main Result

Our main result is a lower bound for the covert capacity, obtained by designing a coding scheme in which the transmitter not only generates a key from S but also selects its codeword according to S by using a likelihood encoder [18].

Theorem 1. *The covert capacity of DMC $W_{Y,Z|S,X}$ with noncausal CSI at the transmitter and the receiver is lower-bounded as*

$$C_{\text{NC}} \geq \max_{p(u),\, x(u,s)} \mathbb{I}(U;S,Y) - \mathbb{I}(U;S), \tag{4}$$

where the maximum is distributions of the form

$$Q_S P_{U|S} P_{X|U,S} W_{YZ|X,S}, \tag{5}$$

such that

$$\mathbb{H}(S|Z) \geq \mathbb{I}(U;S,Z) - \mathbb{I}(U;S,Y), \tag{6}$$

$\mathbb{E}[g(X)] \leq G$, *and* $Q_Z = Q_0$.

Remark 1. *Theorem 1 suggests that the key rate that we extract from channel state (i.e. $H(S|Z)$) should be at least as large as the Right Hand Side (RHS) of the* (6)*.*

Proof. We adapt a block-Markov encoding scheme over $B > 0$ consecutive and dependent blocks. This scheme involves the transmission of $B-1$ independent messages over B channel block transmissions each of length r, indexed by $j = 1,2,...,B$, such that $n = rB$. Here, we assume B and $B-1$ are fixed and sufficiently large positive integers. Therefore, the warden's channel output observation Z^n can be described as $Z^n = (Z_1^r,\dots,Z_B^r)$ in which each Z_j^r for $j \in [\![1,B]\!]$ indicates the channel output in block j. Hence, the induced distribution by the coding scheme at output of the

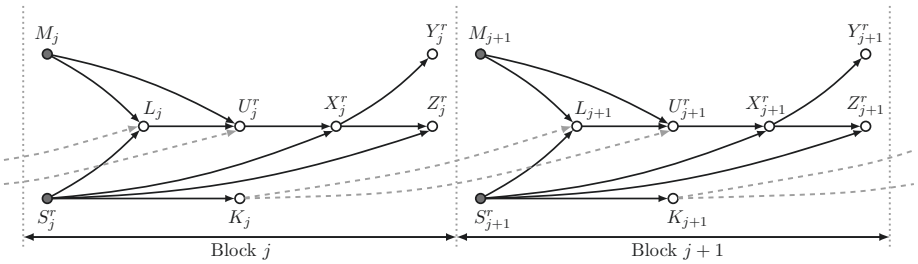

Fig. 2. Functional dependence graph for the block-Markov encoding scheme

warden is $P_{Z^n} \triangleq P_{Z_1^r,\dots,Z_B^r}$ and the target output distribution is $Q_Z^n = \prod_{j=1}^B Q_Z^r$. Note that

$$\begin{aligned}\mathbb{D}(P_{Z^n}||Q_Z^n) &= \mathbb{D}(P_{Z_1^r\dots Z_B^r}||Q_Z^{rB}) \\ &= \sum_{j=1}^{B} \mathbb{D}(P_{Z_j^r|Z_{j+1}^{B,r}}||Q_Z^r|P_{Z_{j+1}^{B,r}}) \\ &= \sum_{j=1}^{B} [\mathbb{D}(P_{Z_j^r}||Q_Z^r) + \mathbb{D}(P_{Z_j^r|Z_{j+1}^{B,r}}||P_{Z_j^r}|P_{Z_{j+1}^{B,r}})] \\ &= \sum_{j=1}^{B} [\mathbb{D}(P_{Z_j^r}||Q_Z^r) + \mathbb{I}(Z_j^r; Z_{j+1}^{B,r})], \end{aligned} \tag{7}$$

where $Z_{j+1}^{B,r} = \{Z_{j+1}^r, \dots Z_B^r\}$. Therefore, to show $\mathbb{D}(P_{Z^n}||Q_Z^n) \xrightarrow[n\to\infty]{} 0$ we need to show that $\forall j \in \{1,\dots B\}$ both $\mathbb{D}(P_{Z_j^r}||Q_Z^r) \xrightarrow[r\to\infty]{} 0$ and $\mathbb{I}(Z_j^r; Z_{j+1}^{B,r}) \xrightarrow[r\to\infty]{} 0$. As shown in the following, we achieve this by constructing a code that approximates Q_Z^r in each block and show that the dependencies across blocks, created by block-Markov coding, can be eliminated. The random code generation is as follows:

Fix $P_U(u)$, $P_{U|S}(u|s)$, $x(u,s)$, $P_{S,U,X,Y,Z}(s,u,x,y,z) = Q_S(s)P_{U|S}(u|s)\mathbb{1}_{\{x(u,s)=x\}}W_{Y,Z|X,S}$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $Q_Z = Q_0$ and $\mathbb{E}[g(X)] \le \frac{G}{1+\epsilon_2}$.

Codebook for Key Generation: For each block $j \in [\![1, B]\!]$, to generate an efficient key K_j of rate R_K by using the channel state S_j^r, we create a function $\Phi(S_j^r) \in [\![1, 2^{rR_K}]\!]$ through random binning. The key $K_j = \Phi(S_j^r)$ obtained in block $j \in [\![1, B]\!]$ is used to assist the encoder in the next block.

Codebook for Message Transmission: For each block $j \in [\![1, B]\!]$ and for each $k_{j-1} \in [\![1, 2^{rR_k}]\!]$, $m_j \in [\![1, 2^{rR}]\!]$, and $\ell_j \in [\![1, 2^{rR'}]\!]$, randomly and independently generate $2^{rR'}$ codewords $u^r(k_{j-1}, m_j, \ell_j)$ according to $\prod_{i=1}^r p(u_i)$. These constitute the codebook $\mathcal{C}_n$. The indices (k_{j-1}, m_j, ℓ_j) can be seen as a two layer binning. We define an ideal Probability Mass Function (PMF) for codebook $\mathcal{C}_n$ as

$$\begin{aligned}&\Gamma^{(\mathcal{C}_n)}_{K_{j-1},M_j,L_j,U^r,S_j^r,Z_j^r,K_j}(k_{j-1}, m_j, \ell_j, \tilde{u}^r, s_j^r, z_j^r, k_j) = \\ &2^{-r(R_k+R+R')} \times \mathbb{1}_{\{u^r(k_{j-1},m_j,\ell_j)=\tilde{u}^r\}} \\ &\times P_{S|U}^r(s_j^r|\tilde{u}^r) \times W_{Z|U,S}^r(s_j^r, z_j^r|\tilde{u}^r) \times \mathbb{1}_{\{\Phi(s_j^r)=k_j\}}. \end{aligned} \tag{8}$$

where $W_{Z|U,S}$ is the marginal distribution of $W_{Y,Z|U,S}$ in (5).

Encoding: In the first block, we choose a codeword randomly and transmit it over the channel and use the CSI to generate a key k_1 for the next block. In block $j \in [\![2, B]\!]$, the codeword U^r will be chosen based on message, key and the current CSI; simultaneously we generate a key from CSI for the next block.

For the first block, the transmitter generates a key from CSI and the encoder chooses (k_0, m_1, ℓ_1) uniformly at random and finds codeword $u^r(k_0, m_1, \ell_1)$ according to these indices. It then transmits a codeword x^r where $x_i = x(u_i(k_0, m_1, \ell_1), s_{1,i})$.

For block $j \in [\![2, B]\!]$, to send message m_j according to the generated key k_{j-1} from the previous block and the CSI of the current block, the encoder selects index ℓ_j from bin (k_{j-1}, m_j) according to the distribution

$$f(\ell_j|s_j^r, k_{j-1}, m_j) = \frac{P_{S|U}^r\left(s_j^r|u^r(k_{j-1}, m_j, \ell_j)\right)}{\sum\limits_{\ell' \in [\![1,2^{rR'}]\!]} P_{S|U}^r\left(s_j^r|u^r(k_{j-1}, m_j, \ell')\right)}, \tag{9}$$

where $p_{S|U}^r$ is defined from the ideal PMF in (8). Each coordinate of the transmitted signal is a function of the state, as well as the corresponding sample of the transmitter codeword u_i, i.e., $x_i = x(u_i(k_{j-1}, m_j, \ell_j), s_i)$.

For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$\begin{aligned}&P^{(\mathcal{C}_n)}_{K_{j-1},M_j,L_j,U^r,S_j^r,Z_j^r,K_j}(k_{j-1}, m_j, \ell_j, \tilde{u}^r, s_j^r, z_j^r, k_j) = \\ &Q_S^r(s_j^r) \times 2^{-r(R_k+R)} \times f(\ell_j|s_j^r, k_{j-1}, m_j) \\ &\times \mathbb{1}_{\{u^r(k_{j-1},m_j,\ell_j)=\tilde{u}^r\}} \times W_{Z|S,U}^r(z_j^r|s_j^r, \tilde{u}^r) \times \mathbb{1}_{\{\Phi(s_j^r)=k_j\}}. \end{aligned} \tag{10}$$

Covertness Analysis: Now, we show that this coding scheme guarantees that $\mathbb{E}_{\mathcal{C}_n}||P_{Z^n}^{(\mathcal{C}_n)} - Q_Z^n||_1 \xrightarrow[n\to\infty]{} 0$ and therefore $\mathbb{E}_{\mathcal{C}_n}[\mathbb{D}(P_{Z^n}^{(\mathcal{C}_n)}||Q_Z^n)] \xrightarrow[n\to\infty]{} 0$ [19, eq. (323)]. From the expansion in (7), note that for every $j \in [\![2, B]\!]$,

$$\begin{aligned}\mathbb{I}(Z_j^r; Z_{j+1}^{B,r}) &\le \mathbb{I}(Z_j^r; K_j, Z_{j+1}^{B,r}) \\ &\overset{(a)}{=} \mathbb{I}(Z_j^r; K_j), \end{aligned} \tag{11}$$

where (a) is because $Z_j^r - K_j - Z_{j+1}^{B,r}$ forms a Markov chain as seen in the functional dependence graph depicted in Fig. 2. Next, note that

$$\begin{aligned}\mathbb{I}(Z_j^r; K_j) &= \mathbb{D}(P^{(\mathcal{C}_n)}_{Z_j^r,K_j}||P^{(\mathcal{C}_n)}_{Z_j^r}P^{(\mathcal{C}_n)}_{K_j}) \\ &\overset{(b)}{\le} \mathbb{D}(P^{(\mathcal{C}_n)}_{Z_j^r,K_j}||Q_Z^r Q_{K_j}), \end{aligned} \tag{12}$$

where Q_{K_j} is the uniform distribution on $[\![1, 2^{rR_K}]\!]$ and (b) follows from

$$\mathbb{D}(P_{Z_j^r,K_j}||P_{Z_j^r}P_{K_j}) = \mathbb{D}(P_{Z_j^r,K_j}||Q_Z^r Q_{K_j})$$

$$- \mathbb{D}(P_{K_j}||Q_{K_j}) - \mathbb{D}(P_{Z_j^r}||Q_Z^r). \quad (13)$$

Therefore by combining (12) and (7), we have

$$\mathbb{D}(P_{Z^n}^{(\mathcal{C}_n)}||Q_Z^n) \leq 2\sum_{j=1}^{B} \mathbb{D}(P_{Z_j^r,K_j}^{(\mathcal{C}_n)}||Q_Z^r Q_{K_j}). \quad (14)$$

By using the triangle inequality we have:

$$\mathbb{E}_{\mathcal{C}_n}||P_{Z_j^r,K_j}^{(\mathcal{C}_n)} - Q_Z^r Q_{K_j}||_1 \leq \mathbb{E}_{\mathcal{C}_n}||P_{Z_j^r,K_j}^{(\mathcal{C}_n)} - \Gamma_{Z_j^r,K_j}^{(\mathcal{C}_n)}||_1 + \mathbb{E}_{\mathcal{C}_n}||\Gamma_{Z_j^r,K_j}^{(\mathcal{C}_n)} - Q_Z^r Q_{K_j}||_1. \quad (15)$$

To bound the first term on the RHS of (15) note that

$$\Gamma_{M_j K_{j-1}}^{(\mathcal{C}_n)} = 2^{-r(R+R_k)} = P_{M_j K_{j-1}}^{(\mathcal{C}_n)}, \quad (16)$$

$$\Gamma_{K_j|S_j^r}^{(\mathcal{C}_n)} = P_{K_j|S_j^r}^{(\mathcal{C}_n)} \quad (17)$$

$$\Gamma_{L_j|M_j,K_{j-1},S_j^r}^{(\mathcal{C}_n)} = f(\ell_j|s_j^r, k_{j-1}, m_j) = P_{L_j|M_j,K_{j-1},S_j^r}^{(\mathcal{C}_n)}, \quad (18)$$

$$\Gamma_{U^r|M_j,K_{j-1},L_j,S_j^r}^{(\mathcal{C}_n)} = \mathbb{1}_{\{u^r(k_{j-1},m_j,\ell_j)=\tilde{u}^r\}} = P_{U^r|M_j,K_{j-1},L_j,S_j^r}^{(\mathcal{C}_n)}, \quad (19)$$

$$\Gamma_{Z_j^r|M_j,K_{j-1},L_j,S_j^r,U^r}^{(\mathcal{C}_n)} = W_{Z|U,S}^r = P_{Z_j^r|M_j,K_{j-1},L_j,S_j^r,U^r}^{(\mathcal{C}_n)}, \quad (20)$$

where (18) is due to (9). Hence,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}||P_{Z_j^r,K_j}^{(\mathcal{C}_n)} - \Gamma_{Z_j^r,K_j}^{(\mathcal{C}_n)}||_1 \\
&\leq \mathbb{E}_{\mathcal{C}_n}||P_{S_j^r,M_j,K_{j-1},L_j,U^r,Z_j^r,K_j}^{(\mathcal{C}_n)} - \Gamma_{S_j^r,M_j,K_{j-1},L_j,U^r,Z_j^r,K_j}^{(\mathcal{C}_n)}||_1 \\
&= \mathbb{E}_{\mathcal{C}_n}||P_{S_j^r,K_{j-1},M_j}^{(\mathcal{C}_n)} - \Gamma_{S_j^r,K_{j-1},M_j}^{(\mathcal{C}_n)}||_1 \\
&\overset{(c)}{=} \mathbb{E}_{\mathcal{C}_n}||Q_S^r - \Gamma_{S_j^r|K_{j-1}=1,M_j=1}^{(\mathcal{C}_n)}||_1, \quad (21)
\end{aligned}$$

where (c) follows from the symmetry of the codebook construction with respect to M and K_{j-1}. Based on the soft covering lemma [20, Corollary VII.5] the RHS of (21) vanishes if

$$R' > \mathbb{I}(U;S). \quad (22)$$

To bound the second term on the RHS of (15) note that

$$\Gamma_{Z_j^r,K_j}^{(\mathcal{C}_n)}(z_j^r, k_j) = \sum_{s_j^r}\sum_{k_{j-1}}\sum_{m_j}\sum_{\ell} \frac{1}{2^{r(R_k+R+R')}} W_{S,Z|U}^r(s_j^r, z_j^r|u^r(k_{j-1}, m_j, \ell_j)) \times \mathbb{1}_{\{\Phi(s_j^r)=k_j\}}. \quad (23)$$

where $W_{S,Z|U}^r = P_{S|U}^r W_{Z|U,S}^r$.

Using Pinsker's inequality, it is sufficient to bound $\mathbb{E}_{\mathcal{C}_n}[\mathbb{D}(\Gamma_{Z_j^r,K_j}^{(\mathcal{C}_n)}||Q_Z^r Q_{K_j})]$ as in (24) shown at the top of next page. In (26) $\mu_{U,S,Z} = \min_{(u,s,z)\in(\mathcal{U},\mathcal{S},\mathcal{Z})} P_{U,S,Z}(u,s,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $r \to \infty$ then $\Psi_2 \to 0$ and if we choose $R_K = \mathbb{H}(S|Z) - \epsilon$, Ψ_1 goes to zero when r grows if

$$R_k + R + R' > \mathbb{I}(U;S,Z), \quad (27)$$

$$R_k + R + R' > \mathbb{I}(U;Z) \quad (28)$$

where (28) is redundant because of (27).

Decoding and Error Probability Analysis: By following the same steps as in [5], the probability of error vanishes when n grows if

$$R + R' \leq \mathbb{I}(U;S,Y). \quad (29)$$

Input Cost Analysis: The proof follows similar lines to [5, Sec V].

The region in Theorem 1 is derived by remarking that the scheme requires $R_K \geq R_k$ and applying Fourier-Motzkin to (22), (27), and (29). □

REFERENCES

[1] A. B. Bash, D. Goeckel, A. Khisti, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[2] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2945–2949.

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

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

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

[6] P.-H. Lin, C. R. Janda, and E. A. Jorswieck, "Stealthy secret key generation," in *Proc. IEEE Global Conf. on Signal and Info. Processing (GlobalSIP)*, Montreal, QC, Canada, Mar. 2017, pp. 492–496.

[7] P.-H. Lin, C. R. Janda, E. A. Jorswieck, and R. F. Schaefer, "Stealthy keyless secret key generation from degraded sources," in *51st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Apr. 2018, pp. 14–18.

[8] M. Tahmasbi and M. R. Bloch, "Covert secret key generation," in *Proc. IEEE Conf. on Commun. and Network Security (CNS)*, Las Vegas, NV, USA, Dec. 2017, pp. 540–544.

[9] ——, "Framework for covert and secret key expansion over classical-quantum channels," *Phys. Rev. A*, vol. 99, p. 052329, May 2019.

[10] Y. Chen and A. J. Han Vinck, "Wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 395–402, Jan. 2008.

[11] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret key agreement using asymmetry in channel state knowledge," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Seoul, South Korea, Jul. 2009, pp. 2286–2290.

[12] Y.-K. Chia and A. El Gamal, "Wiretap channel with causal state information," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2838–2849, May 2012.

[13] H. Fujita, "On the secrecy capacity of wiretap channels with side information at the transmitter," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 11, pp. 2441–2452, Nov. 2016.

[14] A. Sonee and G. A. Hodtani, "Wiretap channel with strictly causal side information at encoder," in *Proc. Iran Workshop on commun. and Info. Theory (IWCIT)*, Tehran, Iran, May 2014, pp. 1–6.

[15] T. S. Han and M. Sasaki, "Wiretap channels with causal state information: Strong secrecy," *available at https://arxiv.org/abs/1708.00422*, Aug. 2017.

[16] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Wiretap channels with random states non-causally available at the encoder," *available at https://arxiv.org/abs/1608.00743*, Aug. 2016.

[17] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. New York, NY, USA: Springer-Verlag, 2005.

[18] C. S. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1836–1849, Apr. 2016.

[19] I. Sason and S. Verdú, "f-divergence inequalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

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

$$\mathbb{E}_{\mathcal{C}_n}[\mathbb{D}(\Gamma^{(\mathcal{C}_n)}_{Z_j^r,K_j}||Q_Z^r Q_{K_j})] = \mathbb{E}_{\mathcal{C}_n}\Big[\sum_{z_j^r,k_j}\Gamma^{(\mathcal{C}_n)}(z_j^r,k_j)\log(\frac{\Gamma^{(\mathcal{C}_n)}(z_j^r,k_j)}{Q_Z^r(z_j^r)Q_{K_j}(k_j)})\Big]$$

$$= \mathbb{E}_{\mathcal{C}_n}\Bigg[\sum_{k_j,z_j^r}\sum_{k_{j-1}}\sum_{m_j}\sum_{\ell_j}\frac{1}{2^{r(R_k+R+R')}}\sum_{s_j^r}W^r_{S,Z|U}(s_j^r,z_j^r|u^r(k_{j-1},m_j,\ell_j))\mathbb{1}_{\{\Phi(s_j^r)=k_j\}}$$
$$\times\log\Bigg(\frac{\sum_{\tilde{s}_j^r}\sum_{k'_{j-1}}\sum_{m'_j}\sum_{\ell'_j}W^r_{S,Z|U}(\tilde{s}_j^r,z_j^r|u^r(k'_{j-1},m'_j,\ell'_j))\mathbb{1}_{\{\Phi(\tilde{s}_j^r)=k_j\}}}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}\Bigg)\Bigg]$$

$$\overset{(g)}{\leq}\sum_{k_j,z_j^r}\sum_{k_{j-1}}\sum_{m_j}\sum_{\ell_j}\frac{1}{2^{r(R_k+R+R')}}\sum_{s_j^r}\sum_{u^r(k_{j-1},m_j,\ell_j)}\Gamma^{(\mathcal{C}_n)}_{S^r,Z^r,U^r}(s_j^r,z_j^r,u^r(k_{j-1},m_j,\ell_j))\sum_{\Phi(s_j^r)}\frac{1}{2^{rR_K}}\mathbb{1}_{\{\Phi(s_j^r)=k_j\}}$$
$$\times\log\mathbb{E}_{\setminus((k_{j-1},m_j,\ell_j),\Phi(s_j^r))}\Bigg[\frac{\sum_{\tilde{s}_j^r}\sum_{k'_{j-1}}\sum_{m'_j}\sum_{\ell'_j}W^r_{S,Z|U}(\tilde{s}_j^r,z_j^r|u^r(k'_{j-1},m'_j,\ell'_j))\mathbb{1}_{\{\Phi(\tilde{s}_j^r)=k_j\}}}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}\Bigg]$$

$$\leq\sum_{k_j,z_j^r}\sum_{k_{j-1}}\sum_{m_j}\sum_{\ell_j}\frac{1}{2^{r(R_k+R+R')}}\sum_{s_j^r}\sum_{u^r(k_{j-1},m_j,\ell_j)}\Gamma^{(\mathcal{C}_n)}_{S^r,Z^r,U^r}(s_j^r,z_j^r,u^r(k_{j-1},m_j,\ell_j))\times\frac{1}{2^{rR_K}}$$
$$\times\log\Bigg(\frac{W^r_{S,Z|U}(s_j^r,z_j^r|u^r(k_{j-1},m_j,\ell_j))}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}+\sum_{(k'_{j-1},m'_j,\ell'_j)\neq(k_{j-1},m_j,\ell_j)}\frac{Q^r_{S,Z}(s_j^r,z_j^r)}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}$$
$$+\sum_{\tilde{s}_j^r\neq s_j^r}\frac{W^r_{S,Z|U}(\tilde{s}_j^r,z_j^r|u^r(k_{j-1},m_j,\ell_j))}{2^{r(R_k+R+R')}Q_Z^r(z_j^r)}+1\Bigg)$$

$$=\Psi_1+\Psi_2, \tag{24}$$

$$\Psi_1=\sum_{k_j}\sum_{k_{j-1}}\sum_{m_j}\sum_{\ell_j}\frac{1}{2^{r(R_k+R+R'+R_K)}}\sum_{(s_j^r,z_j^r,u^r(k_{j-1},m_j,\ell_j))\in T_\epsilon^{(n)}}\Gamma^{(\mathcal{C}_n)}_{S^r,Z^r,U^r}(s_j^r,z_j^r,u^r(k_{j-1},m_j,\ell_j))$$
$$\times\log\Bigg(\frac{W^r_{S,Z|U}(s_j^r,z_j^r|u^r(k'_{j-1},m'_j,\ell'_j))}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}+\sum_{(k'_{j-1},m'_j,\ell'_j)\neq(k_{j-1},m_j,\ell_j)}\frac{Q^r_{S,Z}(s_j^r,z_j^r)}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}$$
$$+\sum_{\tilde{s}_j^r\neq s_j^r}\frac{W^r_{S,Z|U}(\tilde{s}_j^r,z_j^r|u^r(k_{j-1},m_j,\ell_j))}{2^{r(R_k+R+R')}Q_Z^r(z_j^r)}+1\Bigg)$$

$$\leq\log\big(\frac{2^{rR_K}2^{-r(1-\epsilon)\mathbb{H}(S,Z|U)}}{2^{r(R_k+R+R')}2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{rR_K}2^{-r(1-\epsilon)\mathbb{H}(S,Z)}}{2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U)}}{2^{r(R_k+R+R')}2^{-r(1+\epsilon)\mathbb{H}(Z)}}+1\big) \tag{25}$$

$$\Psi_2=\sum_{k_j}\sum_{k_{j-1}}\sum_{m_j}\sum_{\ell_j}\frac{1}{2^{r(R_k+R+R'+R_K)}}\sum_{(s_j^r,z_j^r,u^r(k_{j-1},m_j,\ell_j))\notin T_\epsilon^{(n)}}\Gamma^{(\mathcal{C}_n)}_{S^r,Z^r,U^r}(s_j^r,z_j^r,u^r(k_{j-1},m_j,\ell_j))$$
$$\times\log\Bigg(\frac{W^r_{S,Z|U}(s_j^r,z_j^r|u^r(k'_{j-1},m'_j,\ell'_j))}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}+\sum_{(k'_{j-1},m'_j,\ell'_j)\neq(k_{j-1},m_j,\ell_j)}\frac{Q^r_{S,Z}(s_j^r,z_j^r)}{2^{r(R_k+R+R'-R_K)}Q_Z^r(z_j^r)}$$
$$+\sum_{\tilde{s}_j^r\neq s_j^r}\frac{W^r_{S,Z|U}(\tilde{s}_j^r,z_j^r|u^r(k_{j-1},m_j,\ell_j))}{2^{r(R_k+R+R')}Q_Z^r(z_j^r)}+1\Bigg)$$

$$\leq 2|S||Z||U|e^{-r\varepsilon\mu_{S,Z,U}}r\log(\frac{3}{\mu_Z}+1) \tag{26}$$